Analysis of Sales Report of a Clothes Manufacturing Outlet

- Using R ; <**Domain:** Retail>

Harish Ganesan

Jan 10, 2021

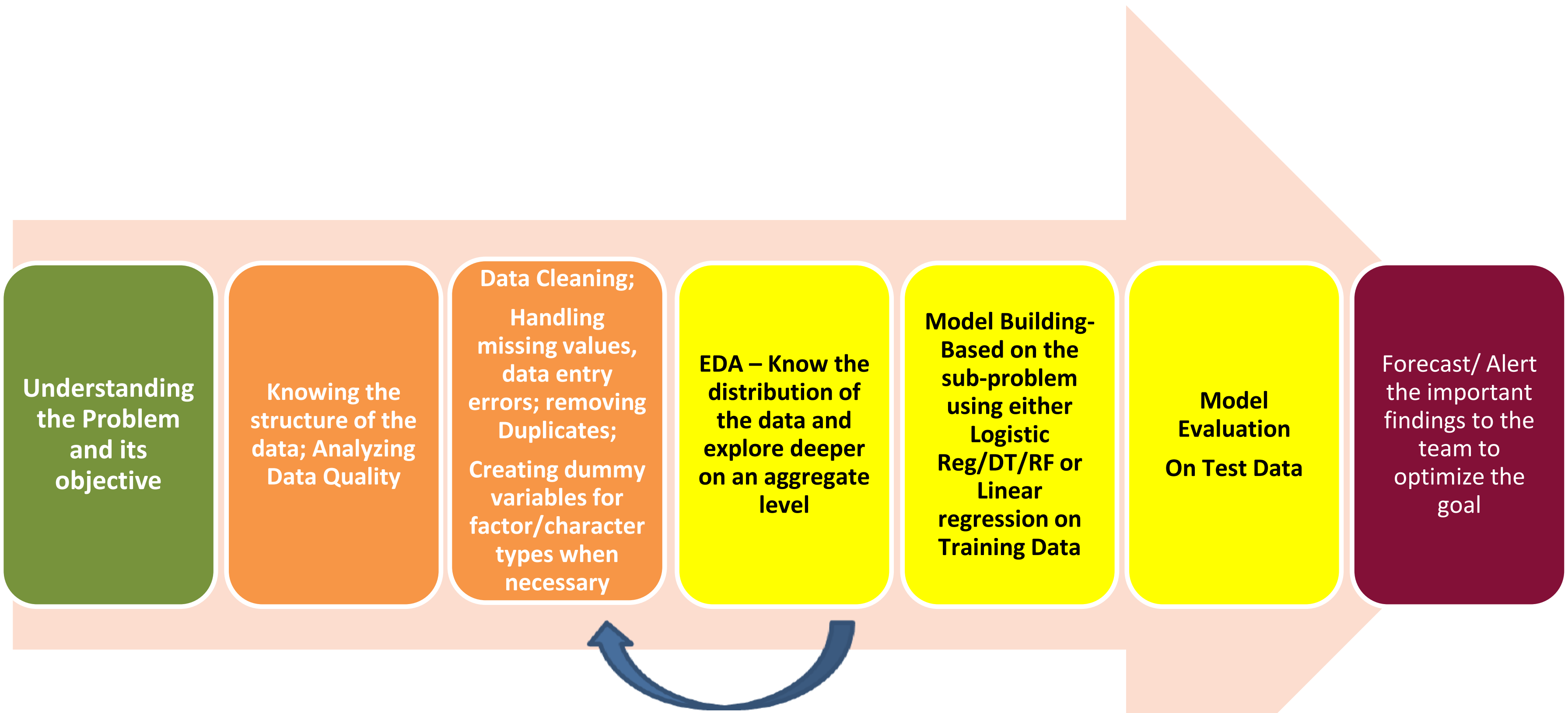
Preface

- The first few pages will only contain the results pertaining to respective sub-problems of the project
- Appendix will contain the full R code that was used to observe these findings along with comments/logs wherever necessary.

Background and Objective

- A high-end fashion retail store is looking to expand its products. It wants to understand the market and find the current trends in the industry. It has a database of all products with attributes, such as style, material, season, and the sales of the products over a period of two months.

Project Workflow



Iteratively going back to Data cleaning step because there are a lot of columns and its best to handle the big dataset as multiple sub-problems using divide and conquer approach str(ATT_DS) // Attribute DataSet ; nrow(ATT_DS) [1] 500 ; ncol(ATT_DS) [1] 14

```
tibble [500 x 14] (S3: tbl_df/tbl/data.frame)
 $ Dress_ID      : num [1:500] 1.01e+09 1.21e+09 1.19e+09 9.66e+08 8.76e+08 ...
 $ Style         : chr [1:500] "Sexy" "Casual" "vintage" "Brief" ...
 $ Price         : chr [1:500] "Low" "Low" "High" "Average" ...
 $ Rating        : num [1:500] 4.6 0 0 4.6 4.5 0 0 0 0 0 ...
 $ Size          : chr [1:500] "M" "L" "L" "L" ...
 $ Season        : chr [1:500] "Summer" "Summer" "Automn" "Spring" ...
 $ NeckLine      : chr [1:500] "o-neck" "o-neck" "o-neck" "o-neck" ...
 $ SleeveLength  : chr [1:500] "sleevless" "Petal" "full" "full" ...
 $ waiseline     : chr [1:500] "empire" "natural" "natural" "natural" ...
 $ Material      : chr [1:500] "null" "microfiber" "polyster" "silk" ...
 $ FabricType    : chr [1:500] "chiffon" "null" "null" "chiffon" ...
 $ Decoration    : chr [1:500] "ruffles" "ruffles" "null" "embroidary" ...
 $ Pattern Type  : chr [1:500] "animal" "animal" "print" "print" ...
 $ Recommendation: num [1:500] 1 0 0 1 0 0 0 0 1 1 ...
```

- **Data entry errors** – need to be rectified e.g. wollen vs woolen, sattin vs satin, summer vs Summer etc. need to be converted as one level. *<highlighted issues in red>*
- **Pool data** in levels that have **insufficient information**. Lets assume a threshold of 10. If any group has less than 10 rows then we will group all that as Other group
- nrow(distinct(ATT_DS)) - #[1] 499 - we can see that there is a duplicate, that has to be removed from the dataset

FabricType <Different Levels>; sort(table(ATT_DS$FabricType), decreasing = T)

null	chiffon	broadcloth	worsted	jersey	shiffon	sattin	batik	Corduroy
265	135	31	19	12	9	6	2	2
dobby	poplin	tulle	wollen	flannael	flannel	knitted	knitting	lace
2	2	2	2	1	1	1	1	1
organza	other	satin	terry	woolen	sum(table(ATT_DS$FabricType)) # 499 # % of Null/none in the total column FabricType > 100*(265/499) # [1] 53.10621			
1	1	1	1	1				

Decoration <Different Levels>; sort(table(ATT_DS$Decoration), decreasing = T)

null	lace	sashes	beading	applique	hollowout	ruffles	bow	sequined
235	70	42	22	21	21	17	15	14
button	embroidary	Pockets	flowers	crystal	rivet	ruched	draped	feathers
6	5	5	4	3	3	3	2	2
none	plain	Cascading	pearls	pleat	Tassel	Tiered	sum(table(ATT_DS$Decoration)) # 499; # % of Null/none in the total column Decoration 100*237/499 # [1] 47.49499	
2	2	1	1	1	1	1		

Pattern Type <Different Levels>; sort(table(ATT_DS$`Pattern Type`), decreasing = T)

solid	null	print	Patchwork	animal	striped	dot	geometric
203	108	71	48	21	17	14	5
leopard	plaid	floral	Character	leapord	none	splice	
3	3	2	1	1	1	1	
sum(table(ATT_DS$`Pattern Type`)) # 499 # % of Null/none in the total column `Pattern Type` > 100*109/499 # [1] 21.84369							

Material <Different Levels>; sort(table(ATT_DS$Material), decreasing = T)

cotton	Null	polyster	silk	chiffonfabric	mix	nylon	rayon
152	127	99	26	25	12	10	10
milksilk	Spandex	cashmere	acrylic	linen	lycra	microfiber	other
5	5	4	3	3	3	3	2
Shiffon	Viscos	knitting	lace	modal	model	sill	wool
2	2	1	1	1	1	1	1
sum(table(ATT_DS$Material)) # 499 # % of Null/none in the total column Material 100*127/499 #[1] 25.45							

waiseline <Different Levels>; sort(table(ATT_DS$waiseline), decreasing = T)

natural	empire	null	dropped	princess
304	104	86	4	1
sum(table(ATT_DS$waiseline)) # 499 # % of Null/none in the total column waiseline 100*86/499 # [1] 17.23				

SleeveLength <Different Levels>; sort(table(ATT_DS$SleeveLength), decreasing = T)								
sleevless	full	short	halfsleeve	threequarter	thressqatar	sleeveless	capsleeves	sleeevless
223	97	96	35	17	10	5	3	3
cap-sleeves	NULL	butterfly	half	Petal	sleveless	threequater	turndowncollor	urndowncollor
2	2	1	1	1	1	1	1	1
# sum(table(ATT_DS$SleeveLength)) # [1] 500 ; # % of Null/none in the total column SleeveLength 100*2/500 # [1] 0.4								

NeckLine <Different Levels>; sort(table(ATT_DS$NeckLine), decreasing = T)								
o-neck	v-neck	slash-neck	boat-neck	Sweetheart	turndowncollor	bowneck	peterpan-collor	sqare-collor
271	124	25	19	14	13	10	6	5
open	NULL	Scoop	backless	halter	mandarin-collor	ruffled	sweetheart	
3	2	2	1	1	1	1	1	
# sum(table(ATT_DS$NeckLine)) # [1] 499 ; # % of Null/none in the total column NeckLine 100*2/499 # [1] 0.4								

Season <Different Levels>; sort(table(ATT_DS$Season), decreasing = T)							
Summer	Spring	Winter	Automn	winter	Autumn	spring	summer
159	122	99	61	46	8	2	1
sum(table(ATT_DS$Season)) # 498 # % of Null/none in the total column Season 100*0/498 #[1] 0; 2 Nas/ Missing values							

Size <Different Levels>; sort(table(ATT_DS$Size), decreasing = T)						
free	L	M	s	S	small	XL
173	96	177	1	37	1	15
sum(table(ATT_DS$Size)) # 500 # % of Null/none in the total column Size 100*0/500 #[1] 0;						

Price <Different Levels>; sort(table(ATT_DS$Price), decreasing = T)						
Average	Low	low	Medium	very-high	high	High
252	129	45	30	21	15	6
sum(table(ATT_DS$Price)) # 498 # % of Null/none in the total column Price 100*0/498 #[1] 0; 2 Nas/ Missing values						

Style <Different Levels>; sort(table(ATT_DS$Style), decreasing = T)

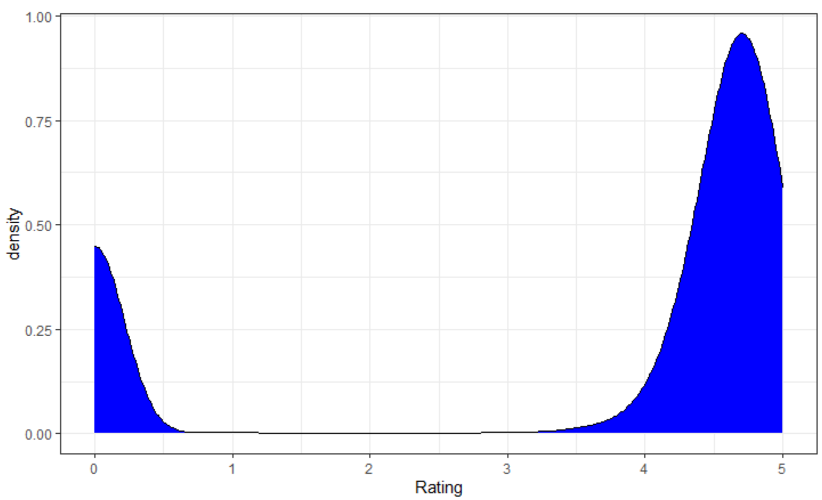
Casual	Sexy	party	cute	vintage	bohemian	Brief
232	69	51	45	25	24	18
work	Novelty	sexy	Flare	fashion	OL	
17	8	7	2	1	1	

sum(table(ATT_DS$Style)) # 500 # % of Null/none in the total column Style 100*0/500 #[1] 0;

Rating <Different Levels>; table(ATT_DS$Rating)

0	1	3	3.5	3.6	3.7	4	4.1	4.2
120	1	1	1	1	2	7	5	6
4.3	4.4	4.5	4.6	4.7	4.8	4.9	5	
20	27	34	54	84	57	25	55	

sum(table(ATT_DS$Rating)) # [1] 500 ; # % of Null/none in the total column Rating 100*0/500 # [1] 0



Rating Class <Different Levels>;

<=4	>4
133	367

```
ggplot(ATT_DS, aes(Rating)) + geom_density(fill="blue")+theme_bw()
# most of the data is 0 -1 or >4, hence best to create RatingClass
RatingClass = cut(Rating,c(-0.1,4,max(Rating)), labels = c("<=4",">4"))
```

More Data Cleaning

- nrow(distinct(DS_DS)) # 500 - no duplicates as a combination of all columns but
- length(levels(DS_DS$Dress_ID)) # Dress_id is only 475, meaning 25 records have more than one same dress_id - #[1] 475
- # an example of one such instance is Dress_id 560474456
- # We can handle this either of two ways - 1) Taking the average of the duplicate rows or
- # 2) taking the maximum value of the rows by Dress_id , assuming there was a glitch while entering and
- # max value was supposed to be the final value but instead of making more assumptions,
- # lets take the average which is a decent starting point for such dress_ids
- If there were any Nas first they were changed to 0 and then average was taken

Goal1: To automate the process of recommendations, the store needs to analyze the given attributes of the product, like the style, season, etc., and come up with a model to predict the recommendation of products (in binary output – 0 or 1) accordingly

- For this goal, we can use **Logistic Regression/ Decision Tree/ Random Forest** etc. and **compare** the **metrics** of these algorithms and use the one that has **best Balanced Accuracy.** The metrics of DT/RF are only mentioned here and full details are in appendix in R code with their results as comments. For the solution of this goal 1, here, we are going to explain Logistic alone in full as that was having decent balanced accuracy.
- Logistic regression, also called a logit model, is used to model dichotomous/binary outcome variables. Since we have to automate recommendation which is either 0 or 1 any of these 3 algorithms can be used. Lets start with logistic first.
- #But before doing logit model, lets see vif (**to check for multicollinearity**) for which we will use alias(lm(Recommendation~.))
- #''alias'' refers to the variables that are linearly dependent on others (i.e. cause perfect multicollinearity).
- # The **autocorrelated variables are** these mentioned below (found by using alias(mylm))
- # **SeasonAutumn** = 1 - Seasonspring - Seasonsummer – Seasonwinter
- # **Pricevery_high** = 1 - Priceaverage - Pricehigh - Pricelow – Pricemedium

- # **Style_work_dmy** = 1 -Style_bohemian_dmy -Style_brief_dmy -Style_casual_dmy -Style_cute_dmy -Style_fashion_dmy -Style_novelty_dmy -Style_OL_dmy -Style_party_dmy -Style_sexy_dmy -Style_vintage_dmy
- # NeckLine_mandarincollor_dmy = Style_OL_dmy

- # **NeckLine_open_dmy** = 1 + Style_fashion_dmy -Material_acrylic_dmy -Material_cashmere_dmy -Material_chiffon_dmy -Material_cotton_dmy -Material_knitting_dmy -Material_linen_dmy -Material_lycra_dmy -Material_microfiber_dmy
- # -Material_milksilk_dmy -Material_mix_dmy -Material_modal_dmy -Material_nylon_dmy -Material_other_dmy -Material_polyster_dmy - Material_rayon_dmy - Material_silk_dmy - Material_spandex_dmy
- # - Material_viscos_dmy

- # **NeckLine_vneck_dmy** = - Style_fashion_dmy -Style_OL_dmy + Material_acrylic_dmy + Material_cashmere_dmy + Material_chiffon_dmy + Material_cotton_dmy + Material_knitting_dmy + Material_linen_dmy + Material_lycra_dmy
- # + Material_microfiber_dmy + Material_milksilk_dmy + Material_mix_dmy + Material_modal_dmy + Material_nylon_dmy + Material_other_dmy + Material_polyster_dmy + Material_rayon_dmy
- # + Material_silk_dmy + Material_spandex_dmy + Material_viscos_dmy - NeckLine_boatneck_dmy - NeckLine_bowneck_dmy - NeckLine_halter_dmy - NeckLine_oneck_dmy - NeckLine_peterpancollor_dmy
- # - NeckLine_ruffled_dmy - NeckLine_Scoop_dmy - NeckLine_slashneck_dmy -NeckLine_squarecollor_dmy - NeckLine_sweetheart_dmy - NeckLine_turndowncollor_dmy

Answer Goal1: Results of Logit:

Positive influence on Recommendation – Spring season, nylon, cotton and other material.

Negative effect on the response variable – Short sleeve length, Prices (low, high, average) and bow-neck, o-neck neckline. So the positive factors tend to increase the recommendation and the negative factors have a depreciating influence on recommending a dress.

```
summary(mylogit)
Call:
glm(formula = Recommendation ~ Rating + SleeveLengthshort + Seasonspring +  Priceaverage + Pricehigh + Pricelow +
    Style_cute_dmy + Style_OL_dmy + Material_cashmere_dmy + Material_cotton_dmy + Material_nylon_dmy +
    Material_other_dmy + NeckLine_bowneck_dmy + NeckLine_oneck_dmy + NeckLine_peterpancollor_dmy +
    NeckLine_squarecollor_dmy + NeckLine_turndowncollor_dmy + Material_acrylic_dmy,
    family = binomial(link = "logit"), data = Training1)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9113  -0.9863  -0.4891   1.0733   2.3082
Coefficients:                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.4921     0.4945   0.995 0.319611
Rating                        0.1057     0.0625   1.692 0.090679 .
SleeveLengthshort            -0.9575     0.3514  -2.725 0.006432 ** // significant
Seasonspring                  0.9848     0.2812   3.502 0.000461 *** very significant
Priceaverage                 -1.6305     0.4479  -3.641 0.000272 *** very significant
Pricehigh                    -2.0190     0.6764  -2.985 0.002836 ** // significant
Pricelow                     -1.3934     0.4615  -3.019 0.002533 ** // significant
Style_cute_dmy                0.7472     0.4118   1.814 0.069629 . //marginally significant
Style_OL_dmy                -17.6394  2399.5448  -0.007 0.994135
Material_cashmere_dmy        18.1105  1677.7192   0.011 0.991387
Material_cotton_dmy           0.5812     0.3055   1.902 0.057112 . //marginally significant
Material_nylon_dmy            1.7934     0.9386   1.911 0.056030 . //marginally significant
Material_other_dmy            0.6699     0.3043   2.202 0.027677 *  significant
NeckLine_bowneck_dmy         -2.3582     1.2365  -1.907 0.056492 . //marginally significant
NeckLine_oneck_dmy           -0.4957     0.2571  -1.928 0.053840 . //marginally significant
NeckLine_peterpancollor_dmy -16.3605  1063.6762  -0.015 0.987728
NeckLine_squarecollor_dmy   -16.4364  1320.2093  -0.012 0.990067
NeckLine_turndowncollor_dmy  -1.4237     0.8558  -1.664 0.096197 .
Material_acrylic_dmy        -16.2223  1272.7943  -0.013 0.989831
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 476.2  on 349  degrees of freedom
Residual deviance: 398.0  on 331  degrees of freedom
AIC: 436

Number of Fisher Scoring iterations: 15
```

There are insignificant terms too in this model, meaning, they could be influential if interactions maybe considered. Tried pooling them to the error degrees of freedom but AIC values started Increasing. So these should be left in the model. For full code, Please refer to the appendix.

Metrics

LOGIT RESULTS – Training <70 percent of Total data>

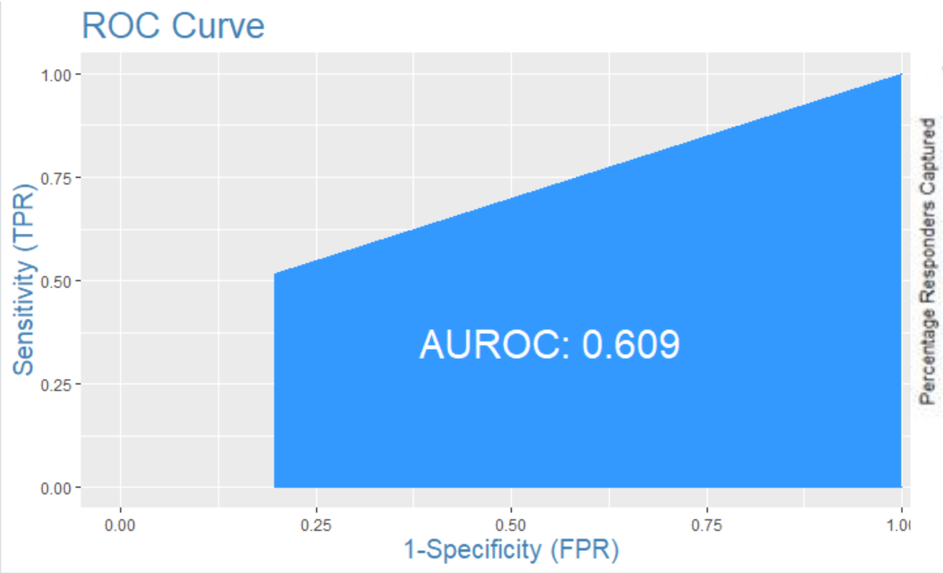
```
caret::confusionMatrix(a, positive = '1')
Confusion Matrix and Statistics

      predicted
actual   0   1
     0 163  40
     1  71  76

               Accuracy : 0.6829
                 95% CI : (0.6313, 0.7313)
    No Information Rate : 0.6686
    P-Value [Acc > NIR] : 0.306371

                  Kappa : 0.3295

 Mcnemar's Test P-Value : 0.004407

            Sensitivity : 0.6552
            Specificity : 0.6966
         Pos Pred Value : 0.5170
         Neg Pred Value : 0.8030
             Prevalence : 0.3314
         Detection Rate : 0.2171
   Detection Prevalence : 0.4200
      Balanced Accuracy : 0.6759

       'Positive' Class : 1
```

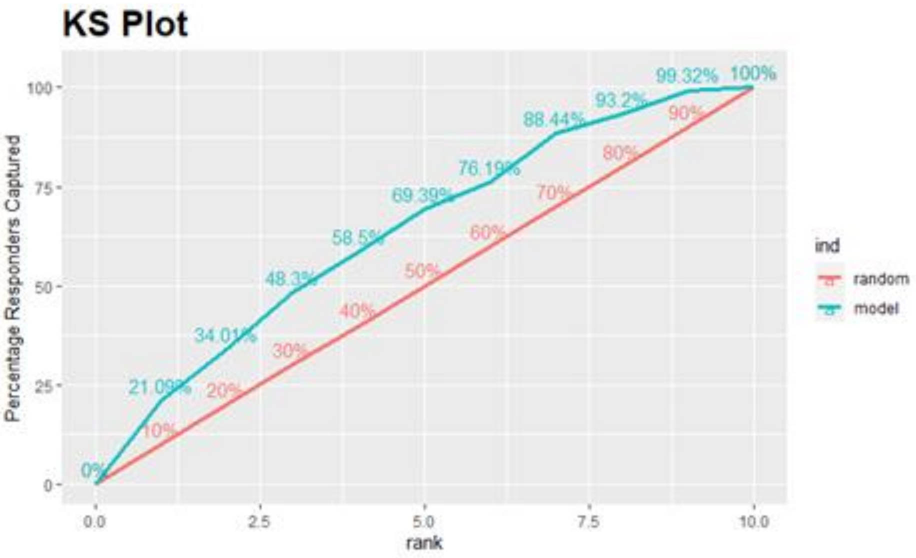
LOGIT RESULTS – Test Data <30 percent of Total data>

```
caret::confusionMatrix(atest, positive = '1')
Confusion Matrix and Statistics

      predicted
actual  0  1
     0 69 17
     1 40 23

               Accuracy : 0.6174
                 95% CI : (0.5343, 0.6958)
    No Information Rate : 0.7315
    P-Value [Acc > NIR] : 0.999119

                  Kappa : 0.176

 Mcnemar's Test P-Value : 0.003569

            Sensitivity : 0.5750
            Specificity : 0.6330
         Pos Pred Value : 0.3651
         Neg Pred Value : 0.8023
             Prevalence : 0.2685
         Detection Rate : 0.1544
   Detection Prevalence : 0.4228
      Balanced Accuracy : 0.6040

       'Positive' Class : 1
```

LOGISTIC Reg. Model - performance measurement visually

Area Under the ROC curve (AUC) is an aggregated metric that evaluates how well a logistic regression model classifies positive and negative outcomes at all possible cutoffs. AUC 0.61 means there is 61% chance that model will be able to distinguish between positive class and negative class correctly.



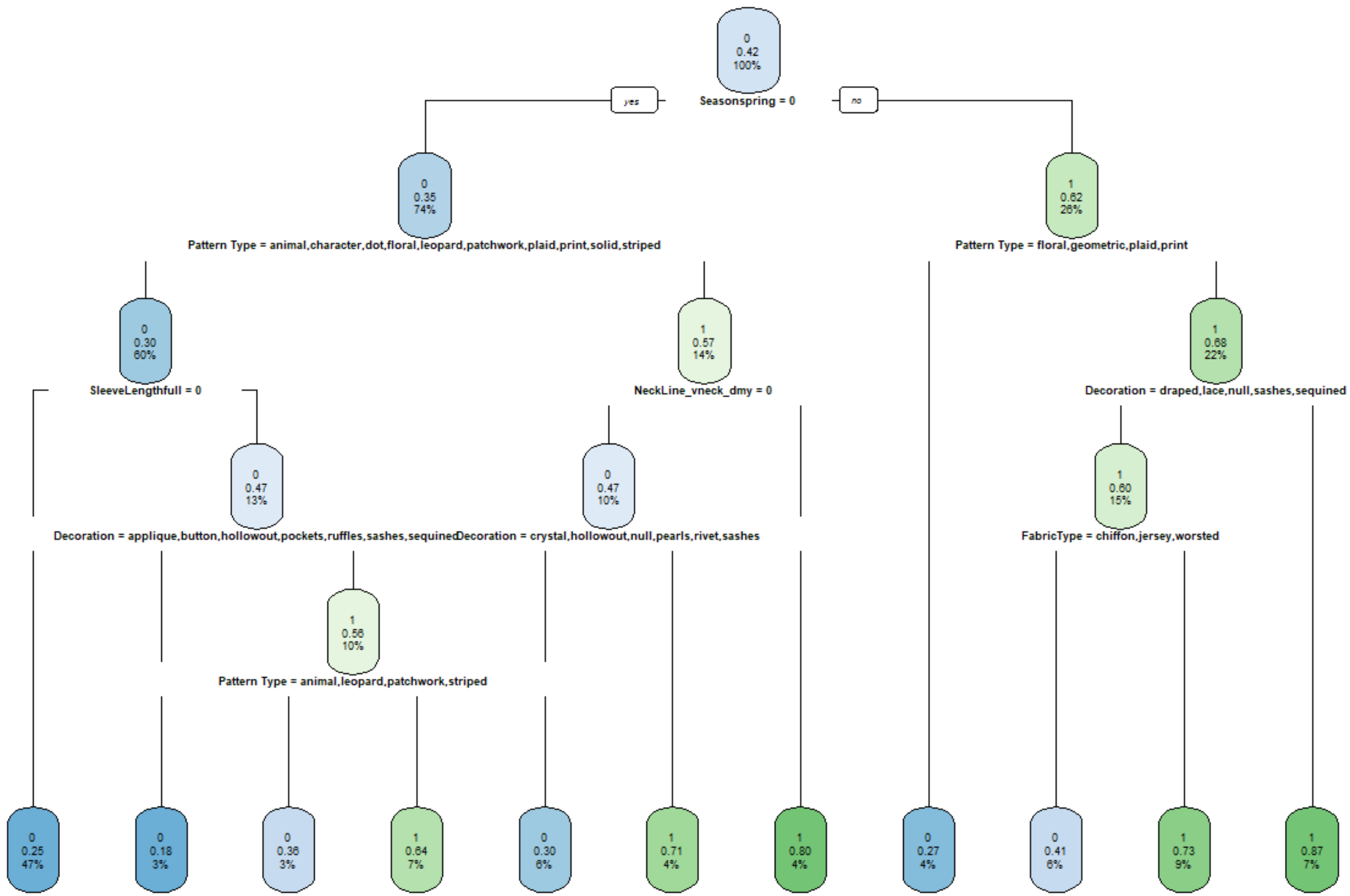


KS Plot plots the lift is capturing the responders (Ones) against the random case where we don't use the model.

The **more curvier** (higher) the model curve, the **better is the model**.

Decision Tree

fit = rpart(Recommendation ~ ., method ="class", data = Training1,
control = rpart.control(minsplit = 30))
rpart.plot(fit, cex=0.55)



Additional ways of solving Goal 1

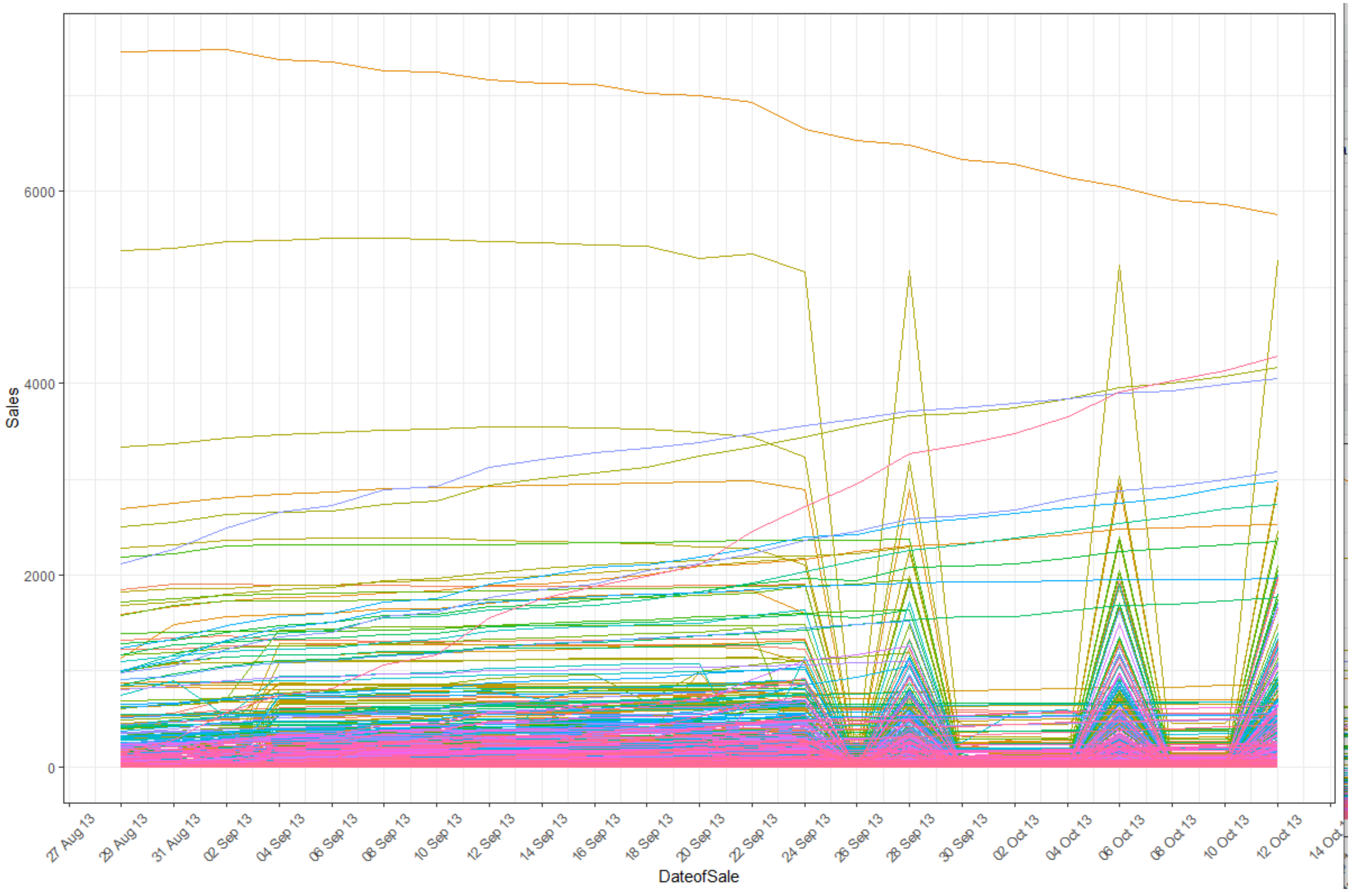
Decision Tree Metric on Test Data

- caret::confusionMatrix(a, positive ="1")
- # Confusion Matrix and Statistics
- #
- # predicted
- # actual 0 1
- # 0 70 16
- # 1 39 24
- #
- # Accuracy : 0.6309
- # 95% CI : (0.548, 0.7084)
- # No Information Rate : 0.7315
- # P-Value [Acc > NIR] : 0.997297
- #
- # Kappa : 0.2049
- #
- # Mcnemar's Test P-Value : 0.003012
- #
- # Sensitivity : 0.6000
- # Specificity : 0.6422
- # Pos Pred Value : 0.3810
- # Neg Pred Value : 0.8140
- # Prevalence : 0.2685
- # Detection Rate : 0.1611
- # Detection Prevalence : 0.4228
- # Balanced Accuracy : **0.6211**
- #
- # 'Positive' Class : 1

Random Forest Metric on Test Data

- caret::confusionMatrix(b, positive ="1")
- # Confusion Matrix and Statistics
- #
- # predicted
- # actual 0 1
- # 0 64 22
- # 1 31 32
- #
- # Accuracy : 0.6443
- # 95% CI : (0.5618, 0.7209)
- # No Information Rate : 0.6376
- # P-Value [Acc > NIR] : 0.4692
- #
- # Kappa : 0.257
- #
- # Mcnemar's Test P-Value : 0.2718
- #
- # Sensitivity : 0.5962
- # Specificity : 0.6737
- # Pos Pred Value : 0.5079
- # Neg Pred Value : 0.7442
- # Prevalence : 0.3624
- # Detection Rate : 0.2148
- # Detection Prevalence : 0.4228
- # Balanced Accuracy : **0.6331** **<best balanced accuracy>**
- #
- # 'Positive' Class : 1

Goal 2: In order to stock the inventory, the store wants to analyze the sales data and predict the trend of total sales for each dress for an extended period of three more alternative days.



Observations – given data EDA

- Spaghetti plot of observed Sales data – We see some Sale spike on 28Sep/6Oct/12Oct (maybe promotion days?)
- Also we observe one dress sales is consistently decreasing which has the most sales among all dresses.

Forecasting Algorithms explored

ETS Exponential Smoothing Algorithm

- (used as a benchmark algorithm to forecast here)
- fit_ets <- ets(Y)
- checkresiduals(fit_ets) # Residual SD :74776.39 which is high

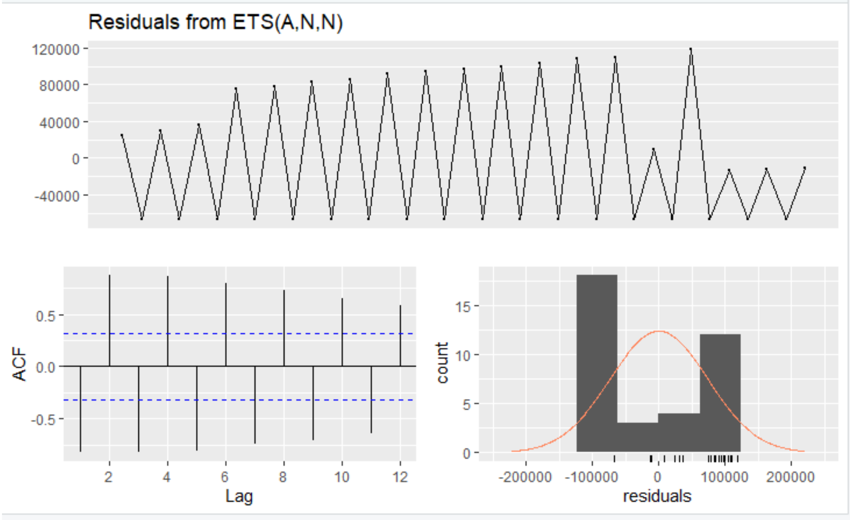
ARIMA

- fit_arima <- summary(auto.arima(Y, d=1,stepwise = F,approximation = F,trace = T)) #residual sqrt(788991890) ~ 28089
- ARIMA decreased the residuals drastically

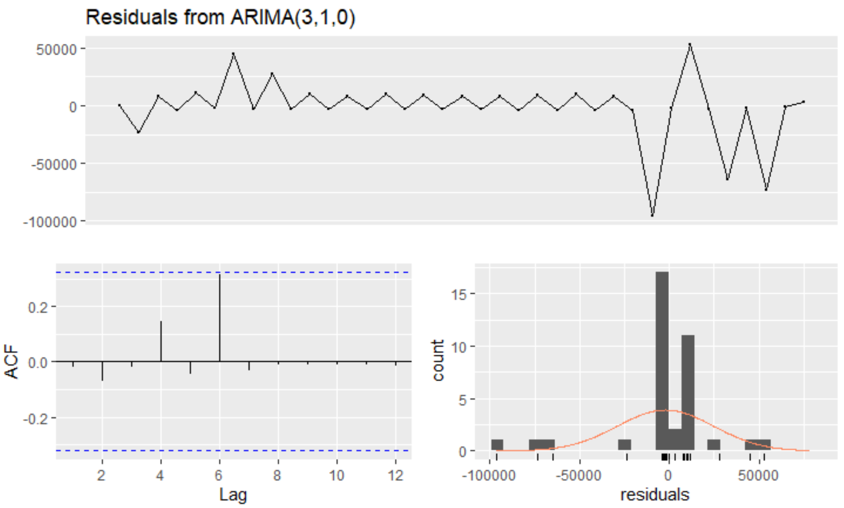
Lets check how ACF or autocorrelation function performed for these in the next page.
Autocorrelation measures the linear relation between lagged values of a time series.
In the ACF figure (correlogram), we would want the spikes within the blue dashed line,
which is the case for Arima, which essentially is good and better than ETS.

Residual Plots - Forecasting

ETS



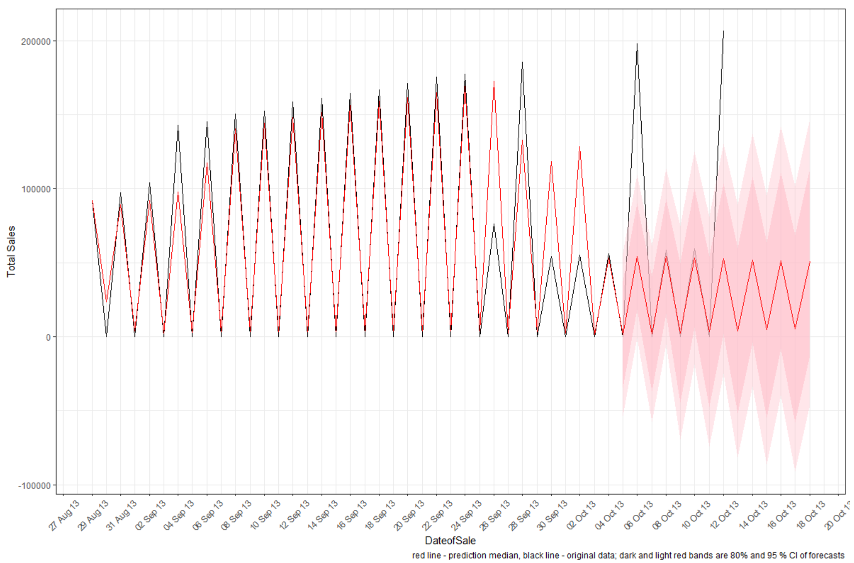
ARIMA

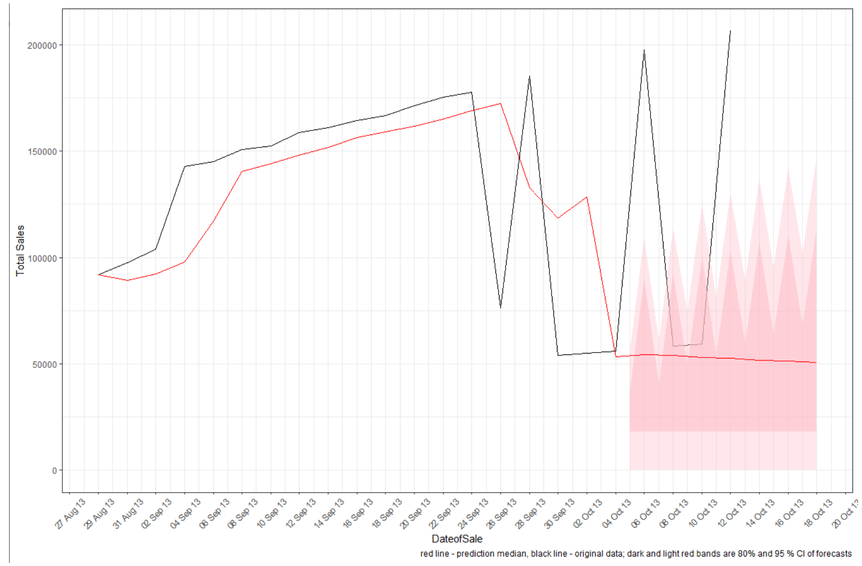


Total Sale of all Dresses forecasted for 3 more alternate days using 85:15 train-test split

The spikes on alternate days drop to zero because we don’t have information on those days

By plotting just alternate days where the information was present, we get

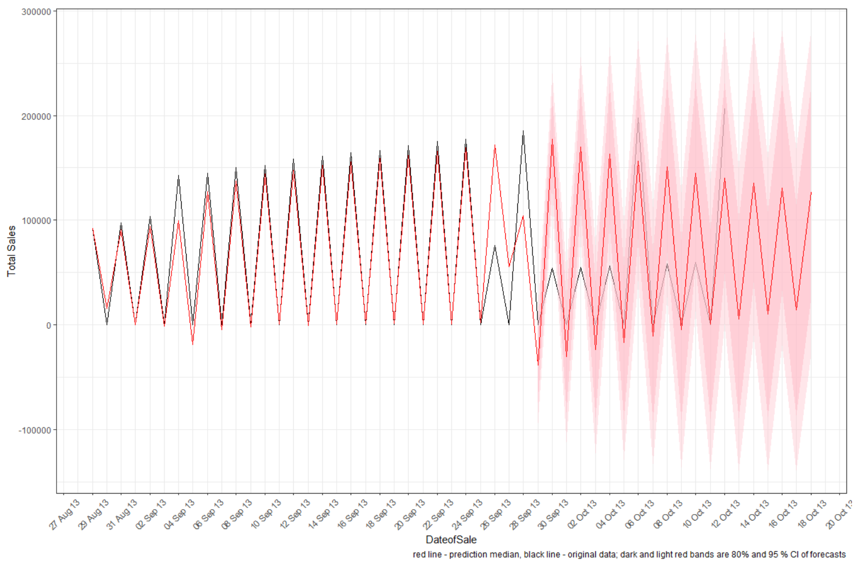




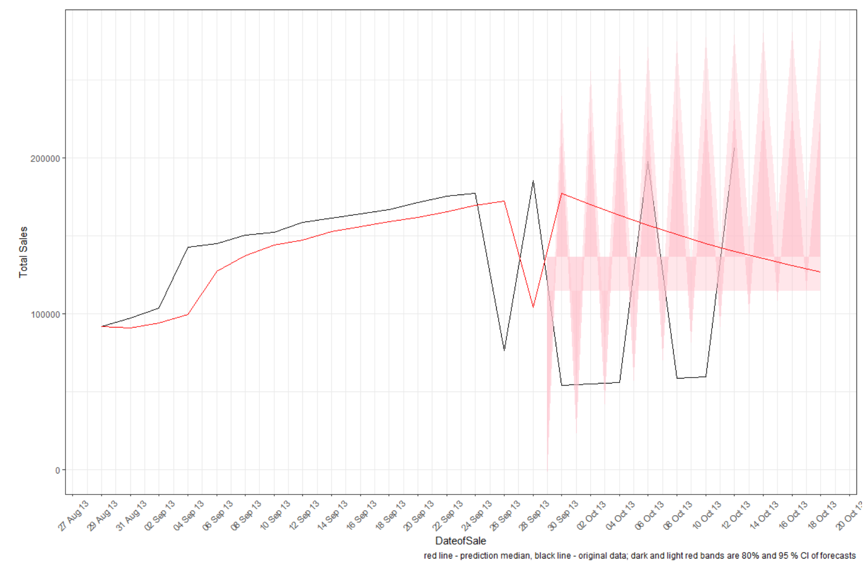
Overall, if we see the predictions are okay but not capturing well on 3 big spikes . Those days could possibly be promotion days where some discount was offered or some national holiday etc. Two possible ways we could improve this prediction i.e. by using another Train:Test split or using another algorithm called Prophet (facebook). Still we haven’t solved the goal 2 which was trends asked on an Individual level of every Dress. So for that lets try Prophet assuming that was the best algorithm among the 3. It will capture the promotion/discount days very well.

Total Sale of all Dresses forecasted for 3 more alternate days using 70:30 train-test split

The spikes on alternate days drop to zero because we don't have information on those days



By plotting just alternate days where the information was present, we get



This looks much better than previous forecast with 85:15 train test split. Still we see that had we known a prior event like 27 Sep date (some public holiday etc.) we could have predicted the drop accurately for 26th Sep instead of 28th Sep . In this split we are over-predicting 30 Sep,2Oct and 4th Oct. whereas in the previous split we were more underpredicting the important sales day . So based on the biz decision if under-predicting or over-predicting is more important to the company, we could decide which split to use. In the next page lets see Prophet usage where we will solve the goal 2 (task 2 i.e. on an individual level prediction for 3 more alternate days)

Prophet R execution

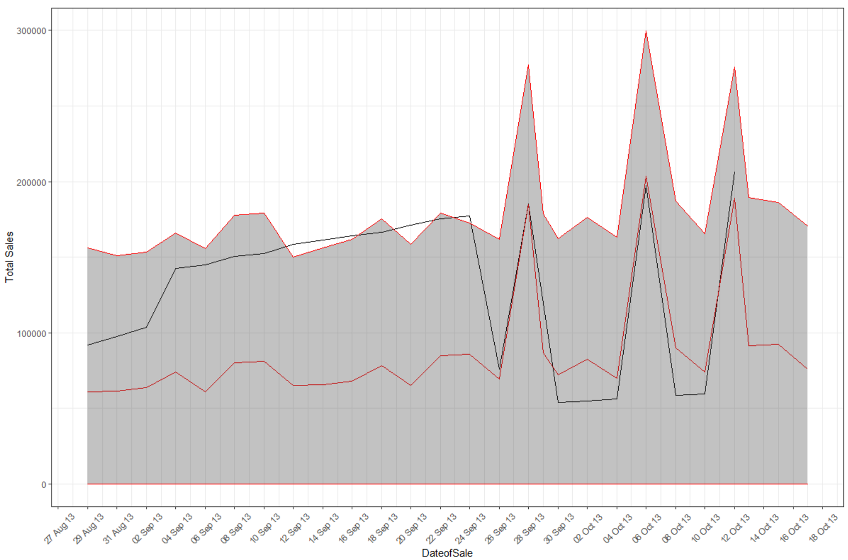
Without Promotion days coded

```
library(prophet)        library(purrr)
#out of the 23 date records for every dress id, lets consider 70 percent as training and rest as testing records
train_index <-  1: floor(0.7 * 23 ) # nrow(input) is changed to n as we have already calc for each SKU
test_index <- setdiff(1:23, train_index) # nrow(input) is changed to n as we have already calc for each SKU
DS_DS_T  <- DS_DS_T %>% group_by(Dress_ID) %>% mutate(SNO = 1:n(), TRAINTESTFIL = ifelse(SNO <= train_index,"Train","Test")) %>% ungroup()
train <- dplyr::filter(DS_DS_T,TRAINTESTFIL =="Train")
test <- dplyr::filter(DS_DS_T,!TRAINTESTFIL =="Train")
train_all <- train %>%
      group_by(DateofSale,SNO,TRAINTESTFIL) %>%
      summarise(Sales = sum(Sales,na.rm = T))  %>%
      ungroup() %>% dplyr::select(1,2,4,3) %>%
      padr::pad(interval = "day") %>%
      tidyr::fill(SNO,TRAINTESTFIL) %>%
      mutate(Sales = if_else(is.na(Sales), 0, Sales))
test_all <- test %>%
      group_by(DateofSale,SNO,TRAINTESTFIL) %>%
      summarise(Sales = sum(Sales,na.rm = T))  %>%
      ungroup() %>% dplyr::select(1,2,4,3) %>%
      padr::pad(interval = "day") %>%
      tidyr::fill(SNO,TRAINTESTFIL) %>% mutate(Sales = if_else(is.na(Sales), 0, Sales))
d1_all <- train_all %>% rename(ds = DateofSale, y = Sales) %>% # for prophet to work Must have columns ds (date type) and y, the time series.
      nest() %>%  mutate(m = map(data, prophet)) %>%
      mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
      mutate(forecast = map2(m, future, predict))
 d_all <- d1_all %>% unnest(forecast) %>%
      dplyr::select(ds, yhat, yhat_lower,yhat_upper)
 indi_ds_all <- bind_rows(train_all,test_all) %>% mutate(ds = DateofSale)
indi_ds2_all <- left_join(d_all,indi_ds_all) %>% mutate(SNO = 1:n(),
           DateofSale = as.Date(as.character(ds)))
 ggplot(data = indi_ds2_all %>% filter(Sales >0) , aes(x = DateofSale)) +
      geom_line(aes(x = DateofSale, y = Sales)) +
      scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
      theme_bw() + theme(legend.title = element_blank(),
                  axis.text.x  = element_text(angle=45, vjust=0.5)) +
      theme(legend.position = "none") +  ylab("Total Sales")+ geom_line(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |  (indi_ds2_all$SNO %% 2 ==0 & is.na(indi_ds2_all$Sales)),] , aes(y = yhat), color = "red") + geom_ribbon(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) | (indi_ds2_all$SNO %% 2 ==0 & is.na(indi_ds2_all$Sales)) ,], aes(ymin=yhat_lower, ymax=yhat_upper),
            color = "red", alpha =0.3)
```

With Promotion days coded

```
#now using promotion days to factor in some days to improve the prediction
promotion_days <- data.table::data.table(
      holiday = 'sale discount days',
      ds=as.Date(c('2013-09-28',
               '2013-10-06', '2013-10-12'
      )),
      lower_window = 0, # looks back at 0 days before the last day to shop and get it before sale
      upper_window = 0
)
d1_all <- train_all %>% rename(ds = DateofSale, y = Sales) %>% # for prophet to work Must have columns ds (date type) and y, the time series.
      nest() %>%
      mutate(m = map(data, ~prophet(.,holidays = promotion_days))) %>%
      mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
      mutate(forecast = map2(m, future, predict))
 d_all <- d1_all %>%
      unnest(forecast) %>%
      dplyr::select(ds, yhat, yhat_lower,yhat_upper) # in case if we wish to see all variables we can comment this

indi_ds_all <- bind_rows(train_all,test_all) %>% mutate(ds = DateofSale)
indi_ds2_all <- left_join(d_all,indi_ds_all) %>%
      mutate(SNO = 1:n(),DateofSale = as.Date(as.character(ds)))

 ggplot(data = indi_ds2_all %>%
           filter(Sales >0) ,
      aes(x = DateofSale)) +geom_line(aes(x = DateofSale, y = Sales)) +
      scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
      theme_bw() + theme(legend.title = element_blank(),
                  axis.text.x  = element_text(angle=45, vjust=0.5)) +
      theme(legend.position = "none") +
      geom_line(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                          (indi_ds2_all$SNO %% 2 ==0 & is.na(indi_ds2_all$Sales)),],
           aes(y = yhat), color = "red") +
      ylab("Total Sales")+geom_ribbon(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                          (indi_ds2_all$SNO %% 2 ==0 & is.na(indi_ds2_all$Sales)) ,],
            aes(ymin=yhat_lower, ymax=yhat_upper),
            color = "red", alpha =0.3)

###now also using day seasonality in addition to promotional campaign days
d1_all <- train_all %>% rename(ds = DateofSale, y = Sales) %>% # for prophet to work Must have columns ds (date type) and y, the time series.
      nest() %>%
      mutate(m = map(data, ~prophet(.,holidays = promotion_days,daily.seasonality=TRUE))) %>%
      mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
      mutate(forecast = map2(m, future, predict))

d_all <- d1_all %>%
      unnest(forecast) %>%
      dplyr::select(ds, yhat, yhat_lower,yhat_upper) # in case if we wish to see all variables we can comment this


indi_ds_all <- bind_rows(train_all,test_all) %>% mutate(ds = DateofSale)
indi_ds2_all <- left_join(d_all,indi_ds_all) %>% mutate(SNO = 1:n(), DateofSale = as.Date(as.character(ds)))


ggplot(data = indi_ds2_all %>%  filter(Sales >0) ,
      aes(x = DateofSale)) +   geom_line(aes(x = DateofSale, y = Sales)) +
      scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
      theme_bw() + theme(legend.title = element_blank(), axis.text.x  = element_text(angle=45, vjust=0.5)) +
theme(legend.position = "none") +
      geom_line(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                          (indi_ds2_all$SNO %% 2 ==0 & is.na(indi_ds2_all$Sales)),]
           ,
           aes(y = yhat), color = "red") +   ylab("Total Sales")+  geom_ribbon(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                          (indi_ds2_all$SNO %% 2 ==0 & is.na(indi_ds2_all$Sales)) ,],
            aes(ymin=yhat_lower, ymax=yhat_upper),  color = "red", alpha =0.3)
```

Prophet – Trend Total Sales



- We have captured the event (promotion) days very well using this algorithm.
- The cons are that it is not predicting that well before these days. Arima was better on days leading to event days. This model is kind of over-fitting.
- So depending on the business needs, if we were to understand what is more important i.e. overfitting or under-fitting around these event days, we could choose Arima or this accordingly. Our goal 2 of predicting on an individual level for 3 more additional alternate days is completed (please check the appendix for full code related to that .)
- Previous page included this algorithm on all data (all dresses)
- The code in this page only lists the important piece pertaining to individual level prediction. i.e. nesting by Dress_Id

Answer Goal 2: on individual level

```
d1 <- train %>% rename(ds = DateofSale, y = Sales) %>%
# for prophet to work Must have columns ds (date type) and y, the time series.
    nest(-Dress_ID) %>%
    mutate(m = map(data, prophet)) %>%
    mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
    mutate(forecast = map2(m, future, predict))
d <- d1 %>%
    unnest(forecast) %>%
    dplyr::select(ds, Dress_ID, yhat, yhat_lower,yhat_upper)
```

Goal/Task 3

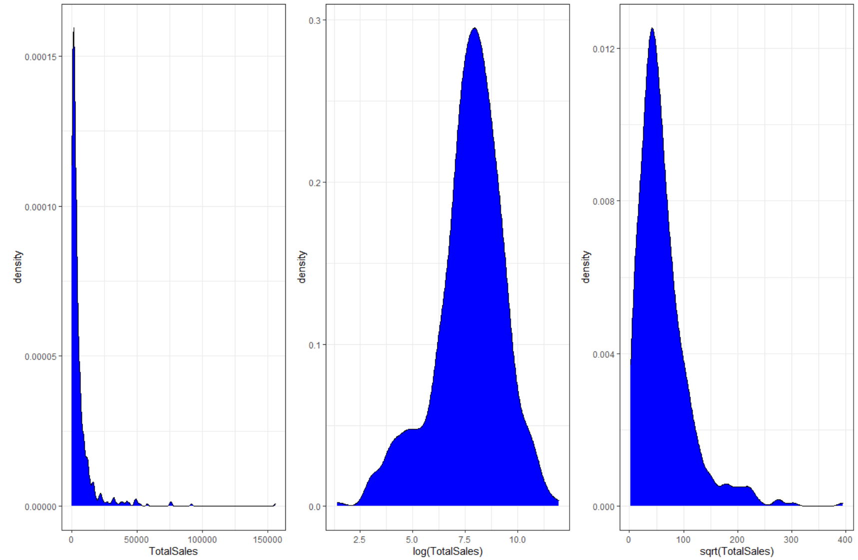
To decide the pricing for various upcoming clothes, they wish to find how the style, season, and material affect the sales of a dress and if the style of the dress is more influential than its price.

- For this task, lets consider the total sales per dress id and do an Linear Regression/ANOVA
- # ANOVA is the statistical model that you use to predict a continuous outcome on the basis of one or more categorical predictor variables
- # we can use lm itself because we have already created dummy variables for categorical variables
- lets explore the response variable TotalSales distribution and see if it is normal; If not use the transformed variable whichever is
- Log transformed looks best. Hence lets model that
- The **autocorrelated variables that will be removed are** these mentioned below (found by using alias(mylm))
- ***SeasonAutumn*** = 1 - Seasonspring - Seasonsummer – Seasonwinter
- ***Pricevery_high*** = 1 - Priceaverage - Pricehigh - Pricelow - Pricemedium
- ***Style_work_dmy*** = 1 -Style_bohemian_dmy -Style_brief_dmy -Style_casual_dmy -Style_cute_dmy -Style_fashion_dmy -Style_novelty_dmy -Style_OL_dmy -Style_party_dmy -Style_sexy_dmy -Style_vintage_dmy
- Even after removing these there were some factors that had high VIF >10. These were removed in a step-wise fashion until we had VIF values <5. A high VIF indicates multicollinearity still exists.
- Lets consider alpha value by default 0.05 for all regression problems in this project.

__Answer Goal3__:

- Based on the final model with 2-way interactions and just considering these 4 factors asked in the goal3, we see that **season doesn't affect as it has been pooled to error degrees of freedom and low, medium price, cashmere, chiffon, lycra , rayon materials as individual effects are significant. Also high-priced nylon, low-priced cute-style, medium-priced silk material, low-priced rayon material and high-priced cute style affects the total sales in a significant way. Vintage style and low-priced nylon are marginally significant.**
- The ones in green above are influencing in a positive way increasing the sales and the ones in red are decreasing the sales based on the coefficients of the regression.
- **Style of the dress is NOT more influential than its price.** (please see next page linear reg. results as some levels of price have more significance than all style levels, based on p-value).

Distribution of Response variable



Please note that Adj Rsq here is only 17% and p-val <0.05. So could be biased estimates here. We see that clearly later in the solution of Goal4/5 where rating and other dress attributes are also factored in the model and Adj Rsq. actually improves. Since we didn't consider all factors affecting the model, this one is a biased model. But in the context of comparing style vs price, price is more significant based on p-value.

Goal 3: Regression results

First step of Lin Reg., after removing VIF >10 factors; Modeling log(TotalSales) ~

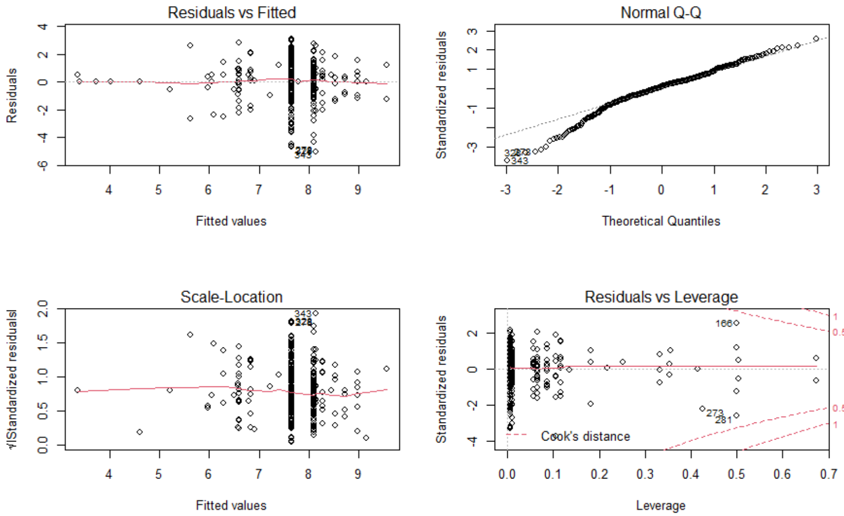
```
Residuals:
    Min      1Q  Median      3Q     Max 
-4.6612 -0.7140  0.1533  0.8534  3.6154 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)              7.76788    0.29476  26.354   <2e-16 ***
Seasonspring            -0.09386    0.30462  -0.308   0.7582    
Seasonsummer            -0.06357    0.28142  -0.226   0.8214    
Seasonwinter            -0.07403    0.28616  -0.259   0.7960    
Pricehigh               -0.44247    0.42253  -1.047   0.2958    
Pricelow                 0.50079    0.19605   2.554   0.0111 *  
Pricemedium             -0.95238    0.38777  -2.456   0.0146 *  
Style_bohemian_dmy      -0.41881    0.45753  -0.915   0.3607    
Style_brief_dmy          0.26802    0.45611   0.588   0.5572    
Style_cute_dmy          -0.51047    0.33960  -1.503   0.1338    
Style_fashion_dmy       -1.64986    1.53763  -1.073   0.2841    
Style_flare_dmy          1.68939    1.69459   0.997   0.3196    
Style_novelty_dmy        0.19035    0.59774   0.318   0.7504    
Style_OL_dmy            -0.68001    1.57857  -0.431   0.6669    
Style_party_dmy         -0.05456    0.30561  -0.179   0.8584    
Style_sexy_dmy          -0.03518    0.23846  -0.148   0.8828    
Style_vintage_dmy        0.53679    0.38586   1.391   0.1652    
Material_acrylic_dmy    -0.25884    1.10752  -0.234   0.8154    
Material_cashmere_dmy   -2.06425    0.93461  -2.209   0.0279 *  
Material_chiffon_dmy     0.57357    0.42747   1.342   0.1806    
Material_knitting_dmy   -0.04894    1.54582  -0.032   0.9748    
Material_linen_dmy       0.22779    0.91018   0.250   0.8025    
Material_lycra_dmy      -1.98517    1.10440  -1.798   0.0732 .  
Material_microfiber_dmy  0.21157    0.89776   0.236   0.8138    
Material_milksilk_dmy    0.09030    0.78106   0.116   0.9080    
Material_mix_dmy         0.18214    0.71806   0.254   0.7999    
Material_modal_dmy      -1.68259    1.09342  -1.539   0.1248    
Material_nylon_dmy      -1.41114    0.72656  -1.942   0.0530 .  
Material_other_dmy      -0.06351    0.22760  -0.279   0.7804    
Material_polyster_dmy    0.01582    0.24481   0.065   0.9485    
Material_rayon_dmy      -0.64360    0.57862  -1.112   0.2668    
Material_silk_dmy       -0.62744    0.40819  -1.537   0.1253    
Material_spandex_dmy    -0.70575    0.79951  -0.883   0.3781    
Material_viscos_dmy     -0.14776    1.10108  -0.134   0.8933    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.523 on 318 degrees of freedom
Multiple R-squared:  0.1506,	Adjusted R-squared:  0.06242 
F-statistic: 1.708 on 33 and 318 DF,  p-value: 0.011
```

Last step of Reg., that also contains 2-way interaction.

```
Residuals:
    Min      1Q  Median      3Q     Max 
-5.0113 -0.6128  0.1188  0.8238  3.1419 

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)                            7.6462     0.1079  70.870  < 2e-16 ***
Pricelow                               0.4485     0.1736   2.583  0.01022 *  significant
Pricemedium                           -1.0526     0.3809  -2.764  0.00604 ** very sig.
Style_cute_dmy                         0.5006     0.4774   1.049  0.29508    
Style_fashion_dmy                     -1.5917     1.4372  -1.108  0.26887    
Style_vintage_dmy                      0.6219     0.3498   1.778  0.07635 .  Marginally sig.
Material_cashmere_dmy                 -3.9085     1.4372  -2.720  0.00688 ** very sig.
Material_chiffon_dmy                   0.8835     0.4233   2.087  0.03763 *  significant
Material_lycra_dmy                    -3.6208     1.4372  -2.519  0.01223 *  significant
Material_modal_dmy                    -1.5983     1.0174  -1.571  0.11714    
Material_rayon_dmy                    -1.3560     0.6141  -2.208  0.02793 *  significant
Material_nylon_dmy:Pricehigh          -4.2450     1.4372  -2.954  0.00336 ** very sig.
Style_cute_dmy:Pricehigh              -2.1728     0.9581  -2.268  0.02399 *  significant
Pricelow:Style_cute_dmy               -1.7716     0.6717  -2.637  0.00875 ** very sig,
Pricelow:Material_lycra_dmy            3.3093     2.0342   1.627  0.10471    
Pricelow:Material_nylon_dmy           -1.5121     0.8393  -1.802  0.07251 .  Marginally sig.
Pricelow:Material_rayon_dmy            2.8191     1.1935   2.362  0.01876 *  significant
Pricemedium:Style_cute_dmy             1.3510     1.2869   1.050  0.29454    
Pricemedium:Material_cashmere_dmy      2.9381     1.7961   1.636  0.10282    
Pricemedium:Material_silk_dmy         -3.2381     1.1954  -2.709  0.00710 ** very sig.
Style_cute_dmy:Material_chiffon_dmy   -1.6226     0.9969  -1.628  0.10454    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.433 on 331 degrees of freedom
Multiple R-squared:  0.2173,	Adjusted R-squared:   0.17 
F-statistic: 4.595 on 20 and 331 DF,  p-value: 0.000000001098
```

The adj Rsq 0.17 tells the predictive ability of this model and its not great. Just means we haven't considered all the factors to predict this. The p-val. associated with the model is significant <<<<0.05 (1.1^10-9) indicating we have a significant model nevertheless. It could very well be an under-specified biased model considering only these factors asked in this task but to compare which factors among these impact more , we can consider the p-value of these and numerically sort them in orders of significance with the least values being most significant. The p-values > 0.05 are the ones which are insignificant. Some are slightly >0.05 but not by a lot, so we consider them as marginally significant.

Goal 3: Results (continued)



- The residual plot indicates there are some influential outliers (Cook's distance) and the q-q plot indicates normal plot overall with some heavy tail where the outliers are perhaps dragging and make it bad.
- But overall the residual plots look good as we don't find any patterns and the residual is homoscedastic. So for this task , we are ok with the regression results for the time being. We will run another run after all the tasks are done to see if we can improve this i.e. removing the heavy tail

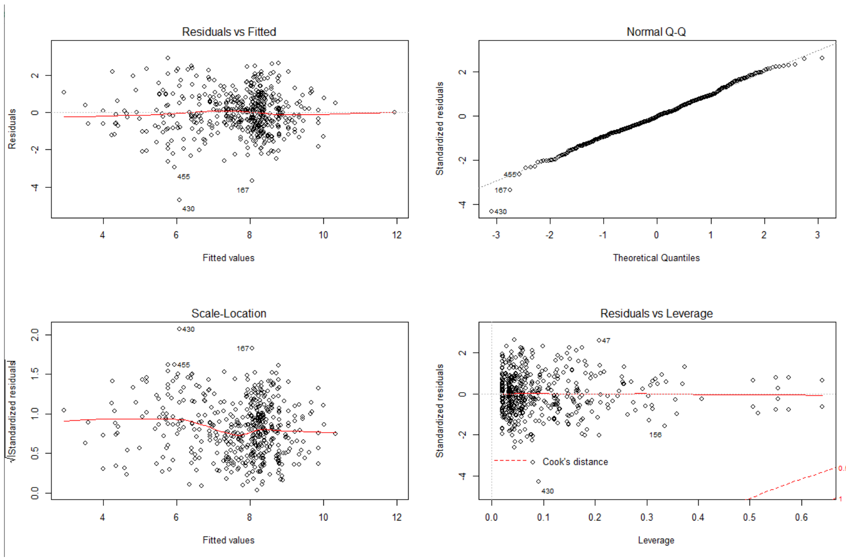
```
pred1 <- predict(totsale_lm.19_c.15, newdata = Testing1)
rmse <- sqrt(sum((exp(pred1) - Testing1$TotalSales)^2)/length(Testing1$TotalSales))
c(RMSE = rmse, R2=summary(totsale_lm.19_c.15)$r.squared)
```

- As the last step, we will predict the 'test' observation and will see the comparison between predicted response and actual response value.
- RMSE explains on an average how much of the predicted value will be from the actual value.
- Based on RMSE = 19749.31, we can conclude that on an average predicted value will be off by 19749.31 from the actual value.
- It is not a great model because we only considered the few factors asked in Task 3 for the modeling of Total Sales

Goal4: Also, to increase sales, the management wants to analyze the attributes of dresses and find which are the leading factors affecting the sale of a dress.
Goal5: To regularize the rating procedure and find its efficiency, the store wants to find if the rating of the dress affects the total sales

- Lets consider both the goal/tasks together and evaluate one single regression. Also instead of just considering these factors alone for regression and to remove bias, lets consider even price , styles as other factors affecting sales as that was significant earlier.
- By using step()/stepAIC(), we found the final model to be used and we have analyzed with 2-way interactions and without interactions as well.
- And the model with 2-way interaction is better and considered as final for now based on the Adj Rsq.

Residual Plots of Overall Total Sales Model with 2 factor interactions- **Looks good**



- Q-Q plot indicates the data is normal with the exception of outliers indicated by cook's D
- The residuals vs fitted points and other plots indicate homoscedastic (equal) variance as there is no pattern and it is evenly distributed
- There are few outliers indicated by Cooks distance but overall the model satisfies the linear regression assumptions and is good to use albeit the Adj. Rsq is just over 50% indicating further factors need to be explored or maybe 3 way interaction level need to be thought of as well.
- But this model of predicting sales has improved considerably after exploring one very significant factor (Rating) which was not evaluated earlier.

- Results of regression in next page

Coefficients: Main Effects

NOTE: Coefficients of the interaction terms are in the next page

	Estimate	Std. Error	t value	Pr(>|t|)			
(Intercept)	8.713	0.5205	16.74	< 2e-16	***		***INFERENCE***
SleeveLengththreequarter	-0.5254	0.5333	-0.985	0.325058		*insignificant*	
SleeveLengthsleeveless	-1.1716	0.4923	-2.38	0.017735	*	***significant***	***Sleeveless decreases sales***
SleeveLengthshort	-0.8836	0.5013	-1.763	0.078654	.	*marginally significant*	
SleeveLengthhalfsleeve	-1.2042	0.5307	-2.269	0.023746	*	***significant***	***Half-sleeve decreases sales***
SleeveLengthfull	-2.397	0.6759	-3.546	0.000432	***	***very significant***	***Full-sleeve decreases sales***
Priceaverage	0.5847	0.1928	3.033	0.002562	**	***very significant***	***Price low, average increases sales***
Pricelow	0.7513	0.2193	3.426	0.000669	***	***very significant***	
Style_sexy	-0.1063	0.2289	-0.465	0.64251		*insignificant*	
Material_cashmere	0.9235	0.949	0.973	0.33104		*insignificant*	
Material_linen	1.2258	0.8503	1.442	0.150101		*insignificant*	
Material_mix	0.6347	0.6103	1.04	0.29891		*insignificant*	
Material_modal	-1.8147	0.8221	-2.207	0.027799	*	***significant***	***Modal material decreases sales***
NeckLine_bowneck	-1.9092	0.5653	-3.377	0.000796	***	***very significant***	***Bow-neck decreases sales***
NeckLine_ruffled	3.1834	1.1657	2.731	0.006567	**	***very significant***	***Ruffled neckline increases sales***
Size_L	1.7671	0.5204	3.396	0.000746	***	***very significant***	
Size_M	0.1931	0.1426	1.354	0.176538		*insignificant*	***Large and small size increase sales***
Size_S	1.2033	0.3962	3.037	0.002529	**	***very significant***	
FabricType_null	-0.2452	0.1222	-2.006	0.045492	*	***significant***	***Fabric type if null decreases sales***
FabricType_worsted	0.2006	0.3493	0.574	0.56614		*insignificant*	
Pattern_Type_null	-1.7803	0.5011	-3.553	0.000422	***	***very significant***	***Pattern Type if null decreases sales***
RATINGLE4	-1.7407	0.1973	-8.822	< 2e-16	***	***very significant***	***Rating below 4 decreases sales***

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.147 on 447 degrees of freedom
Multiple R-squared:  0.5575,	Adjusted R-squared:  0.507
F-statistic: 11.04 on 51 and 447 DF,  p-value: < 2.2e-16
```

Coefficients: 2-way interaction

//the positive significant coefficients (estimate) improve sales and the negative significant coefficients decrease sales

	Estimate	Std. Error	t value	Pr(>|t|)		
SleeveLengththreequarter:Size_L	-2.5566	0.7306	-3.5	0.000513	***	***very significant***
SleeveLengththreequarter:RATINGLE4	-1.1852	0.5407	-2.192	0.028897	*	*significant*
SleeveLengthsleeveless:NeckLine_bowneck	1.778	0.7834	2.269	0.023718	*	*significant*
SleeveLengthsleeveless:Size_L	-1.5114	0.5538	-2.729	0.006596	**	***very significant***
SleeveLengthsleeveless:Size_S	-0.8113	0.5091	-1.593	0.11176		*insignificant*
SleeveLengthsleeveless:Pattern_Type_null	1.3994	0.5198	2.692	0.007364	**	***very significant***
SleeveLengthsleeveless:RATINGLE4	-0.6317	0.2571	-2.458	0.014368	*	***significant***
SleeveLengthshort:Size_L	-1.13	0.5891	-1.918	0.055729	.	*marginally significant*
SleeveLengthshort:Size_S	-1.1741	0.6218	-1.888	0.059641	.	*marginally significant*
SleeveLengthshort:Pattern_Type_null	1.9552	0.6129	3.19	0.001523	**	***very significant***
SleeveLengthhalfsleeve:FabricType_worsted	1.9841	1.3743	1.444	0.149507		*insignificant*
SleeveLengthhalfsleeve:Pattern_Type_null	1.8378	0.6336	2.901	0.003909	**	***very significant***
SleeveLengthfull:Priceaverage	1.0262	0.4533	2.264	0.024057	*	***significant***
SleeveLengthfull:Pricelow	0.938	0.4607	2.036	0.042335	*	***significant***
SleeveLengthfull:Size_L	-0.9438	0.6315	-1.495	0.135738		*insignificant*
SleeveLengthfull:FabricType_null	0.4693	0.297	1.58	0.11471		*insignificant*
SleeveLengthfull:Pattern_Type_null	1.9306	0.6355	3.038	0.002523	**	***very significant***
Priceaverage:Material_mix	1.0899	0.7394	1.474	0.1412		*insignificant*
Priceaverage:Size_S	-0.717	0.4405	-1.628	0.104307		*insignificant*
Priceaverage:Material_cashmere	-3.0039	1.2782	-2.35	0.019203	*	***significant***
Pricelow:RATINGLE4	1.094	0.2674	4.091	0.000051	***	***very significant***
Style_sexy:Size_L	1.1392	0.4464	2.552	0.011047	*	***significant***
Style_sexy:Size_M	0.6697	0.3466	1.932	0.053946	.	*marginally significant*
Style_sexy:Size_S	1.1038	0.6073	1.818	0.069782	.	*marginally significant*
Material_linen:Size_M	2.105	1.4613	1.441	0.150422		*insignificant*
Material_mix:Size_L	-1.4219	0.8185	-1.737	0.083026	.	*marginally significant*
Material_mix:Size_M	-2.0617	0.9718	-2.121	0.034432	*	***significant***
Size_S:FabricType_worsted	-3.7085	1.2779	-2.902	0.003892	**	***very significant***
FabricType_worsted:RATINGLE4	-1.4966	0.7627	-1.962	0.050367	.	*marginally significant*
Pattern_Type_null:RATINGLE4	-0.7449	0.2999	-2.484	0.013362	*	***significant***

Appendix

- R Code that generated these results is mentioned in this section.
- The results of R code are added as R comments using # whenever possible (i.e. for anything except plots/simulations).

```
"-------------------------------------------------------------------------
# Purpose          : Project 1 (Analysis of Sales Report of a Clothes Manufacturing
Outlet)
# Date             : 12/14/2020 (creation); last edited date: 01/10/2021
# Author           : Harish Ganesan
# Input file(s)    : C:/Users/hganesan/Documents/training/DS_R_lms_proj/Dress
Sales.xlsx
#                    C:/Users/hganesan/Documents/training/DS_R_lms_proj/Attribute
DataSet.xlsx
# Output file(s)   :
-----------------------------------------------------------------------------"
library(tidyverse)
library(lubridate)
library(scales)
#library(qdap)
library(tm)
library(usdm)

any_column_NA <- function(x){
        any(is.na(x))
}
replace_NA_0 <- function(x){
        if_else(is.na(x),0,x)
}

# Dress Sales DS_DS
path_src1 <- "C:/Users/hganesan/Documents/training/DS_R_lms_proj/Dress Sales.xlsx"
#path_src1 <- "C:/Users/hganesan/Documents/DS_R_lms_proj/Dress Sales.xlsx"

DS_DSh <- readxl::read_excel(path = path_src1 , col_types ="guess", guess_max = 5000,
skip = 0, col_names = F, n_max = 1)
DS_DS <- readxl::read_excel(path = path_src1 , col_types ="guess", guess_max = 5000,
skip = 0, col_names = T)
DS_DSh1 <- DS_DSh %>%
           mutate_at(vars(2,3,10:18), as_date, format = '%d/%m/%Y')
DS_DSh1 <- DS_DSh1 %>%
           mutate_if(is.POSIXct ,as.character) %>%
           mutate_if(is.Date,as.character)
# while checking the values of DS_DSh1, spotted a potential data entry error - while
all columns are talking about 2013 year,
# one alone stands out
# hence correcting that below

DS_DSh1[,22] <- gsub("2010", "2013", DS_DSh1[,22])
str(DS_DSh1)
colnames(DS_DS) <- DS_DSh1[1,]

rm(DS_DSh,DS_DSh1)
str(DS_DS) # some are character types even though they are numeric
DS_DS <- DS_DS%>%
         mutate_if(is.character,as.numeric) %>% # converting some character columns as
numeric
         mutate(Dress_ID = factor(Dress_ID))


# Attribute DataSet ATT_DS
path_src2 <- "C:/Users/hganesan/Documents/training/DS_R_lms_proj/Attribute
DataSet.xlsx"
#path_src2 <- "C:/Users/hganesan/Documents/DS_R_lms_proj/Attribute DataSet.xlsx"
ATT_DS <- readxl::read_excel(path = path_src2 , col_types ="guess", guess_max = 5000,
skip = 0, col_names = T)

#************************ ATT_DS: Factor 1 below*************-
#************************ ----------------------*************-
table(ATT_DS$FabricType)
#satin, sattin can be clubbed; similarly flannael flannel can be grouped, woolen and
wollen can be grouped, knitted and knitting can be grouped

# batik broadcloth    chiffon   Corduroy      dobby   flannael    flannel     jersey
knitted   knitting       lace       null
# 2         31        135          2          2          1          1         12
1          1             1        265
```

```
# organza      other     poplin      satin     sattin    shiffon      terry      tulle
wollen     woolen    worsted
# 1         1         2          1          6          9          1          2
2          1         19

ATT_DS <- ATT_DS %>% mutate( FabricType = ifelse( tolower(FabricType) %in%
c("flannael"),"flannel",
                                                  ifelse( tolower(FabricType) %in%
c("knitting"),"knitted",
                                                          ifelse( tolower(FabricType)
%in% c("sattin"),"satin",
                                                                  ifelse(
tolower(FabricType) %in% c("wollen"),"woolen",
                                                                          ifelse(
tolower(FabricType) %in% c("shiffon"),"chiffon",

FabricType))))))

table(ATT_DS$FabricType)

# batik broadcloth    chiffon   Corduroy      dobby    flannel     jersey    knitted
lace       null    organza      other
# 2          31       144          2          2          2         12          2
1        265          1          1
# poplin      satin        terry      tulle     woolen    worsted
# 2          7            1          2          3         19
sum(table(ATT_DS$FabricType)) #  499

# % of Null/NA/none in the total column FabricType > 100*(265/499)
# [1]  53.10621

#************************ ATT_DS: Factor 2 below*************-
#************************ ----------------------*************-
table(ATT_DS$Decoration)

# applique    beading        bow     button  cascading    crystal     draped
embroidary   feathers    flowers  hollowout       lace
# 21         22         15          6          1          3          2          5
2          4         21          70
# none       null     pearls      plain      pleat    pockets      rivet     ruched
ruffles     sashes   sequined     tassel
# 2         235          1          2          1          5          3          3
17         42         14          1
# Tiered
# 1
sum(table(ATT_DS$Decoration)) #  499
# % of Null/NA/none in the total column Decoration > 100*237/499
# [1] 47.49499

#************************ ATT_DS: Factor 3 below*************-
#************************ ----------------------*************-
table(ATT_DS$`Pattern Type`)

# animal character       dot    floral geometric   leapord   leopard      none
null patchwork     plaid     print     solid
# 21         1         14         2         5         1         3          1
108        48         3        71       203
# splice   striped
# 1        17
ATT_DS <- ATT_DS %>% mutate( `Pattern Type` = ifelse( tolower(`Pattern Type`) %in%
c("leapord"),"leopard",`Pattern Type`))

table(ATT_DS$`Pattern Type`)
# animal character       dot    floral geometric   leopard      none      null
patchwork     plaid     print     solid    splice
# 21       1         14         2         5         4         1       108        48
3        71       203         1
# striped
# 17

sum(table(ATT_DS$`Pattern Type`)) #  499
```

```
# % of Null/NA/none in the total column `Pattern Type` > 100*109/499
# [1] 21.84369

#************************ ATT_DS: Factor 4 below*************-
#************************ ----------------------*************-

table(ATT_DS$Material)
#modal is the fabric not model. similarly shiffon shd be chiffon

# acrylic      cashmere chiffonfabric        cotton      knitting          lace
linen         lycra    microfiber
# 3             4            25           152              1             1
3             3            3
# milksilk           mix          modal          model          null          nylon
other      polyster          rayon
# 5                 12             1             1          127            10
2            99            10
# shiffon          silk           sill       spandex        viscos          wool
# 2            26             1            5             2             1

ATT_DS <- ATT_DS %>% mutate( Material = ifelse( tolower(Material) %in%
c("shiffon"),"chiffonfabric",
                                                ifelse( tolower(Material) %in%
c("model"),"modal",
                                                        ifelse( tolower(Material) %in%
c("sill"),"silk",
                                                        Material))))

table(ATT_DS$Material)

# acrylic      cashmere chiffonfabric        cotton      knitting          lace
linen         lycra    microfiber
# 3             4            27           152              1             1
3             3            3
# milksilk     mix          modal          null          nylon          other
polyster         rayon          silk
# 5            12             2           127            10             2
99           10            27
# spandex        viscos          wool
# 5             2             1


# % of Null/none in the total column Material  100*127/499 #[1] 25.45


#************************ ATT_DS: Factor 5 below*************-
#************************ ----------------------*************-

table(ATT_DS$waiseline)

# dropped   empire  natural     null princess
#       4      104      304       86        1

# % of Null/NA/none in the total column waiseline 100*86/499
# [1] 17.23

#************************ ATT_DS: Factor 6 below*************-
#************************ ----------------------*************-

table(ATT_DS$SleeveLength)

# butterfly    cap-sleeves     capsleeves           full           half     halfsleeve
NULL          Petal            short
#       1              2              3            97              1             35
2             1            96
# sleeevless     sleeveless      sleevless      sleveless   threequarter
threequater    thressqatar turndowncollor  urndowncollor
#      3              5            223              1             17             1
10              1             1

ATT_DS <- ATT_DS %>% mutate( SleeveLength = ifelse( tolower(SleeveLength) %in%
c("capsleeves"),"cap-sleeves",
```

```
                                                 ifelse( tolower(SleeveLength) %in%
c("half"),"halfsleeve",
                                                         ifelse( tolower(SleeveLength)
%in% c("sleeevless","sleevless","sleveless"),"sleeveless",
                                                                 ifelse(
tolower(SleeveLength) %in% c("threequater","thressqatar"),"threequarter",
                                                                         ifelse(
tolower(SleeveLength) %in% c("urndowncollor"),"turndowncollor",

SleeveLength))))))
table(ATT_DS$SleeveLength)
# butterfly    cap-sleeves     full     halfsleeve          NULL          Petal
short     sleeveless   threequarter
# 1            5             97            36              2              1
96           232            28
# turndowncollor
# 2

# sum(table(ATT_DS$SleeveLength)) # [1] 500
# % of Null/none in the total column SleeveLength 100*2/500 # [1] 0.4


#************************ ATT_DS: Factor 7 below*************-
#************************ ---------------------*************-

table(ATT_DS$NeckLine)

# backless       boat-neck         bowneck          halter mandarin-collor
NULL          o-neck            open
#      1              19             10               1               1
2             271            3
# peterpan-collor         ruffled           Scoop      slash-neck    square-collor
sweetheart      Sweetheart  turndowncollor
#          6               1               2              25               5
1             14              13
# v-neck
# 124
ATT_DS <- ATT_DS %>% mutate( NeckLine = ifelse( tolower(NeckLine) %in% c("sqare-
collor"),"square-collor",
                                                     ifelse(NeckLine %in%
c("Sweetheart"),"sweetheart",
                                                            NeckLine)))

table(ATT_DS$NeckLine)

# backless   boat-neck         bowneck          halter mandarin-collor           NULL
o-neck           open
# 1             19             10               1               1               2
271            3
# peterpan-collor       ruffled  Scoop      slash-neck   square-collor      sweetheart
turndowncollor   v-neck
# 6                    1          2              25               5             15
13              124

# % of Null/none in the total column NeckLine > 100*2/499
# [1] 0.4

#************************ ATT_DS: Factor 8 below*************-
#************************ ---------------------*************-

table(ATT_DS$Season)
#
# Automn Autumn spring Spring summer Summer winter Winter
# 61      8      2    122      1    159     46     99

ATT_DS <- ATT_DS %>% mutate( Season = ifelse( tolower(Season) %in%
c("automn"),"Autumn",
                                              ifelse( tolower(Season) %in%
c("spring"),"spring",
                                                      ifelse( tolower(Season) %in%
c("summer"),"summer",
```

```
                                                                ifelse( tolower(Season)
%in% c("winter"),"winter",
                                                         Season)))))


table(ATT_DS$Season)
# Autumn spring summer winter
#   69    124    160    145

#************************ ATT_DS: Factor 9 below*************-
#************************ ----------------------*************-

table(ATT_DS$Size)
# In the following s S and small all indicate the same, we need to format the data
with values (small,s) and
# make it one standard value S

# free     L     M     s     S small    XL
#  173    96   177     1    37     1    15

ATT_DS <- ATT_DS %>% mutate( Size = ifelse( tolower(Size) %in%
c("s","small"),"S",Size))

table(ATT_DS$Size)
# Now the following is alright
# free    L    M    S   XL
#   173   96  177   39   15

#************************ ATT_DS: Factor 10 below*************-
#************************ ----------------------*************-

table(ATT_DS$Price)

# Average      high      High       low       Low    Medium very-high
#    252        15         6        45       129        30        21

ATT_DS <- ATT_DS %>% mutate( Price = ifelse( tolower(Price) %in% c("high"),"high",
                                              ifelse( tolower(Price) %in%
c("low"),"low",
                                                      ifelse( tolower(Price) %in%
c("medium"),"medium",
                                                              ifelse( tolower(Price)
%in% c("average"),"average",
                                                                      Price)))))
table(ATT_DS$Price)
# average  high       low    medium very-high
# 252        21       174        30        21


#************************ ATT_DS: Factor 11 below*************-
#************************ ----------------------*************-

table(ATT_DS$Style)

# bohemian    Brief   Casual     cute   fashion    Flare  Novelty       OL    party
sexy     Sexy  vintage     work
#     24       18      232       45         1        2        8        1       51
7       69       25       17

ATT_DS <- ATT_DS %>% mutate( Style = ifelse( tolower(Style) %in% c("sexy"),"sexy",
                                              ifelse( tolower(Style) %in%
c("novelty"),"novelty",
                                                      ifelse( tolower(Style) %in%
c("brief"),"brief",
                                                              ifelse( tolower(Style)
%in% c("casual"),"casual",
                                                                              ifelse(
tolower(Style) %in% c("flare"),"flare",
                                                                      Style))))))
table(ATT_DS$Style)
```

```
# bohemian    brief   casual     cute  fashion    flare  novelty       OL    party
sexy  vintage     work
#     24       18      232       45        1        2        8        1       51
76       25       17


#*********************** ATT_DS: Factor 12 below*************-
#*********************** --------------------*************-
table(ATT_DS$Rating)
# 0     1   3 3.5 3.6 3.7   4 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9   5
# 120   1   1   1   1   2   7   5   6  20  27  34  54  84  57  25  55

## -(a)
# Checking for NAs and replacing (if possible) with measures of central tendency
median/mean (continuous) and  mode (categorical/discrete)
# (we can replace continuous missing values by median that is more resistant to
outliers and
# for categorical missing like this below, we can use mode to replace the missing
value)

lapply(ATT_DS,function(x) { length(which(is.na(x)))})
# $Dress_ID # [1] 0 #
# $Style # [1] 0 #
# $Price # [1] 2 # categorical column that can be replaced by Mode
# $Rating # [1] 0 #
# $Size # [1] 0 #
# $Season # [1] 2 # categorical column that can be replaced by Mode
# $NeckLine # [1] 1 #
# $SleeveLength # [1] 0 #
# $waiseline # [1] 1 #
# $Material # [1] 1 #
# $FabricType # [1] 1 #
# $Decoration # [1] 1 #
# $`Pattern Type` # [1] 1 #
# $Recommendation # [1] 0

# Price,Season, NeckLine, waiseline, Material, FabricType, Decoration, Pattern Type
all have NAs

## -(b)
# Checking for null or none values which is similar to NA, just that they are NA's
that are predefined as null or none

lapply(ATT_DS,function(x) { length(which(x %in% c("NULL","null","none")))})

# $Dress_ID # [1] 0 #
# $Style # [1] 0 #
# $Price # [1] 0 #
# $Rating # [1] 0 #
# $Size # [1] 0 #
# $Season # [1] 0 #
# $NeckLine # [1] 2 #
# $SleeveLength # [1] 2 #
# $waiseline # [1] 86 #
# $Material # [1] 127 #
# $FabricType # [1] 265 #
# $Decoration # [1] 237 #
# $`Pattern Type` # [1] 109 #
# $Recommendation # [1] 0

# Based on list -(a) above, we are going to replace the NAs in some columns that can
be replaced which are less in number

filter(ATT_DS,is.na(ATT_DS$Price ))
# A tibble: 2 x 14
# Dress_ID Style Price Rating Size  Season NeckLine SleeveLength waiseline Material
FabricType Decoration `Pattern Type` Recommendation
# <dbl>    <chr> <chr>  <dbl> <chr> <chr>  <chr>    <chr>        <chr>     <chr>
<chr>      <chr>      <chr>             <dbl>
#1 6.63e8  party NA       4.8 free  winter o-neck   sleeveless   empire    null
null       embroidary null              1
```

```
#2 1.09e9  party NA       4.5 L     summer NA       full         NA        NA       NA
NA         NA                  1

table(filter(ATT_DS,ATT_DS$Style =="party" )$Style)

# party
# 51

table((filter(ATT_DS,ATT_DS$Style =="party" ) )$Price)

# average      high    medium very-high
#     18         9         5        17
sum(table((filter(ATT_DS,ATT_DS$Style =="party" ) )$Price))
# [1] 49

# So of the 51 party Style records, 2 NA values can be replaced by the mode of the
Price column (average ~ 18 occurring most)
# since it is a categorical missing value and the number of records to replace is very
less.
# If this case was similar and like Fabric type i.e. having > 50% of missing values,
# we would not be that comfortable assigning mode because that is a lot of assumption
replacing the values



filter(ATT_DS,is.na(ATT_DS$Season))
# A tibble: 2 x 14
#    Dress_ID Style Price Rating Size  Season NeckLine SleeveLength waiseline Material
FabricType Decoration `Pattern Type` Recommendation
#    <dbl>    <chr> <chr>  <dbl> <chr> <chr>  <chr>    <chr>        <chr>     <chr>
<chr>      <chr>      <chr>                   <dbl>
# 1 929797706 casu~  low      0   free  NA     o-neck   full          natural   cotton
null       null       patchwork          0
# 2 751364623 party aver~    4.8 L     NA     sweethe~ sleeveless   empire    null
null       pleat      null                    1

table(ATT_DS$Season)

# Autumn spring summer winter
# 69    124    160    145

sum(table(ATT_DS$Season)) #[1] 498

# So of the 500 Season records, 2 NA values can be replaced by the mode of the Season
column (summer ~ 160 occurring most)
# since it is a categorical missing value and the number of records to replace is very
less.


filter(ATT_DS,is.na(ATT_DS$NeckLine))
# A tibble: 1 x 14
#       Dress_ID Style Price Rating Size  Season NeckLine SleeveLength waiseline
Material FabricType Decoration `Pattern Type` Recommendation
#       <dbl>    <chr> <chr>  <dbl> <chr> <chr>  <chr>    <chr>        <chr>     <chr>
<chr>      <chr>      <chr>                   <dbl>
#1    1.09e9     party NA       4.5 L     summer NA       full         NA        NA
NA         NA         NA           1

table(ATT_DS$NeckLine)

# backless       boat-neck      bowneck          halter mandarin-collor         NULL
o-neck           open
# 1              19               10               1               1               2
271              3
# peterpan-collor       ruffled           Scoop      slash-neck   square-collor
sweetheart  turndowncollor       v-neck
#        6                  1               2              25               5
15              13             124

sum(table(ATT_DS$NeckLine)) #[1] 499
```

```
ATT_DS %>% group_by(NeckLine) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n) %>%
top_n(3)
# `summarise()` ungrouping output (override with `.groups` argument)
# Selecting by n
# # A tibble: 3 x 2
# NeckLine      n
# <chr>      <int>
# 1 o-neck      271
# 2 v-neck      124
# 3 slash-neck   25

# So of the 500 NeckLine records, 1 NA value and 2 NULL can be replaced by the mode of
the NeckLine column (o-neck ~ 271 occurring most)
# since it is a categorical missing value and the number of records to replace is very
less.


ATT_DS %>% group_by(waiseline) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n)
# `summarise()` ungrouping output (override with `.groups` argument)
# Selecting by n
# A tibble: 3 x 2
#  waiseline     n
#  <chr>     <int>
# 1 natural     304
# 2 empire      104
# 3 null         86
# 4 dropped       4
# 5 princess      1
# 6 NA            1
# https://discuss.analyticsvidhya.com/t/what-should-be-the-allowed-percentage-of-
missing-values/2456
# Theoretically, 25 to 30% is the maximum missing values are allowed, beyond which we
might want to drop the variable from analysis if that is not that important of a
variable.
# Based on my experience i have usually used central tendencies if the amount of
assumption was less than 10-15 percent and the above link
# also backs my experience and intuition with mine being a less bias prone and more
strict assumption.
# I am not ok converting the null waiseline as  natural value as 17.2 % (86/500) of it
is missing.
# For this project i set the threshold as 15%. Anything less than 15% of missing data,
i am ok with using central tendency

# So of the 500 waiseline records, 1 NA value can be replaced by the mode of the
waiseline column (natural ~ 304 occurring most)
# since it is a categorical missing value and the number of records to replace is very
less.
# Also i would test the model with these 86 null as natural values later as a
sensitivity analysis and
# see for biases or tradeoffs we may have to take based on the selection/assumption of
data

ATT_DS %>% group_by(Material) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n) %>%
top_n(3)  # 25.4 % of null
# A tibble: 3 x 2
# Material     n
# <chr>    <int>
# 1 cotton     152
# 2 null       127
# 3 polyster    99
ATT_DS %>% group_by(Material) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n) %>%
slice_tail()
# `summarise()` ungrouping output (override with `.groups` argument)
# # A tibble: 1 x 2
# Material     n
# <chr>    <int>
# 1 NA           1

ATT_DS %>% group_by(FabricType) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n) %>%
top_n(3)
# FabricType     n
# <chr>      <int>
```

```
# 1 null         265
# 2 chiffon      144
# 3 broadcloth    31

ATT_DS %>% group_by(FabricType) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n) %>%
slice_tail()
#`summarise()` ungrouping output (override with `.groups` argument)
# A tibble: 1 x 2
# FabricType     n
# <chr>      <int>
# 1 NA             1

ATT_DS %>% group_by(Decoration) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n) %>%
top_n(3)
# A tibble: 3 x 2
# Decoration     n
# <chr>      <int>
# 1 null         235
# 2 lace          70
# 3 sashes        42

ATT_DS %>% group_by(Decoration) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n) %>%
slice_tail()
#`summarise()` ungrouping output (override with `.groups` argument)
# A tibble: 1 x 2
# Decoration     n
# <chr>      <int>
# 1 NA             1

ATT_DS %>% group_by(`Pattern Type`) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n)
%>% top_n(3) # 21.6 % 108/500 is null
# A tibble: 3 x 2
# `Pattern Type`     n
# <chr>          <int>
# 1 solid            203
# 2 null             108
# 3 print             71

ATT_DS %>% group_by(`Pattern Type`) %>% summarise(n=n()) %>% ungroup() %>% arrange(-n)
%>% slice_tail()
# `Pattern Type`     n
# <chr>          <int>
# 1 NA                 1


ATT_DS %>% group_by(SleeveLength) %>% summarise(n= n()) %>% ungroup() %>% arrange(-n)
%>% top_n(3)

ATT_DS %>% group_by(SleeveLength) %>% summarise(n= n()) %>% ungroup() %>% arrange(-n)
# 2 NULLS can be replaced by the mode sleeveless
# A tibble: 10 x 2
# SleeveLength       n
# <chr>          <int>
# 1 sleeveless       232
# 2 full              97
# 3 short             96
# 4 halfsleeve        36
# 5 threequarter      28
# 6 cap-sleeves        5
# 7 NULL               2
# 8 turndowncollor     2
# 9 butterfly          1
# 10 Petal              1


ATT_DS <- ATT_DS %>% mutate(Price = ifelse(is.na(Price),"average",Price),
                            Season = ifelse(is.na(Season),"summer",Season),
                            NeckLine = ifelse(is.na(NeckLine) | toupper(NeckLine)
=="NULL","o-neck",NeckLine),
                            SleeveLength = ifelse(is.na(SleeveLength) |
toupper(SleeveLength) =="NULL","sleeveless",
                                                  SleeveLength),
```

```
                            waiseline = ifelse(is.na(waiseline),"natural",waiseline),
                            Material = ifelse(is.na(Material),"cotton",Material),
                            FabricType = ifelse(is.na(FabricType),"null",FabricType),
# even though NA as is or NA changed to null is not that useful, we are trying to
group these as same values
                            Decoration = ifelse(is.na(Decoration),"null",Decoration),
# even though NA as is or NA changed to null is not that useful, we are trying to
group these as same values
                            `Pattern Type` = ifelse(is.na(`Pattern
Type`),"solid",`Pattern Type`))



table(ATT_DS$Recommendation)

# 0   1
# 290 210


str(ATT_DS)
table(ATT_DS$Recommendation) # 210 have a recommendation as 1; 290 have a
recommendation as 0
# 0   1
#290 210

table(is.na(ATT_DS$Dress_ID))
# FALSE
# 500
table(is.na(DS_DS$Dress_ID))
# FALSE
# 500

nrow(distinct(ATT_DS))
#[1] 499 - we can see that there is a duplicate, that has to be removed

ATT_DS <- distinct(ATT_DS,.keep_all = T)

nrow(distinct(DS_DS)) # no duplicates as a combination of all columns but

length(levels(DS_DS$Dress_ID)) # Dress_id is only 475, meaning 25 records have more
than one same dress_id
#[1] 475

# an example of one such instance is Dress_id 560474456

# We can handle this either of two ways - 1) Taking the average of the duplicate rows
or
# 2) taking the maximum value of the rows by Dress_id , assuming there was a glitch
while entering and
# max value was supposed to be the final value but instead of making more assumptions,
# lets take the average which is a decent starting point for such dress_ids


DS_DS <- DS_DS %>% mutate_if(any_column_NA,replace_NA_0) # changing NAs as 0

DS_DS <- aggregate(. ~Dress_ID, data=DS_DS, mean, na.rm=T) # only 475 unique Dress ID
remain in this dataset

# we can use the Dress_ID as primary key
ATT_DS$Dress_ID <- sapply(ATT_DS$Dress_ID,factor)


ATT_DS <- ATT_DS %>% left_join(DS_DS)

ATT_DS <- ATT_DS %>% mutate(Recommendation = as.factor(Recommendation))

str(ATT_DS)

# now even though there are 499 unique combinations only 475 dress ids are there.
meaning again 24 have multiple records where one or more columns vary
# 560474456 again can be used as one such example where it was purchased in diff
season and the material used for the dress was different
```

```
# Logistic regression, also called a logit model, is used to model dichotomous/binary
outcome variables
library(aod)
library(car)
table(ATT_DS$Recommendation)
# 1:210 0:289
summary(ATT_DS)

#xtabs(~ Recommendation + Style + Price  , data = ATT_DS)
# + Size + Rating + Season + NeckLine + SleeveLength + waiseline + Material +
FabricType + Decoration + `Pattern Type`


#estimate a logistic regression model using the glm (generalized linear model)
function
# we are going to assume alpha = 0.05

#ATT_DS$Rating <- factor(ATT_DS$Rating)
# commenting this - we convert Rating to a factor to indicate that Rating should be
treated as a categorical variable just to see how this plays out

#But before doing logit model, lets see vif (to check for multicollinearity) for which
we will use lm()

# ATT_DS2 <- ATT_DS %>%
#             mutate(Style_bohemian_dmy = ifelse(Style =="bohemian", 1, 0),
#                    Style_brief_dmy = ifelse(Style =="brief", 1, 0),
#                    Style_casual_dmy = ifelse(Style =="casual", 1, 0),
#                    Style_cute_dmy = ifelse(Style =="cute", 1, 0),
#                    Style_fashion_dmy = ifelse(Style =="fashion", 1, 0),
#                    Style_flare_dmy = ifelse(Style =="flare", 1, 0),
#                    Style_novelty_dmy = ifelse(Style =="novelty", 1, 0),
#                    Style_OL_dmy = ifelse(Style =="OL", 1, 0),
#                    Style_party_dmy = ifelse(Style =="party", 1, 0),
#                    Style_sexy_dmy = ifelse(Style =="sexy", 1, 0),
#                    Style_vintage_dmy = ifelse(Style =="vintage", 1, 0),
#                    Style_work_dmy = ifelse(Style =="work", 1, 0),
#                    NeckLine_backless_dmy = ifelse(NeckLine =="backless", 1, 0),
#                    NeckLine_boatneck_dmy = ifelse(NeckLine =="boat-neck", 1, 0),
#                    NeckLine_bowneck_dmy = ifelse(NeckLine =="bowneck", 1, 0),
#                    NeckLine_halter_dmy = ifelse(NeckLine =="halter", 1, 0),
#                    NeckLine_mandarincollor_dmy = ifelse(NeckLine =="mandarin-collor",
1, 0),
#                    NeckLine_oneck_dmy = ifelse(NeckLine =="o-neck", 1, 0),
#                    NeckLine_open_dmy = ifelse(NeckLine =="open", 1, 0),
#                    NeckLine_peterpancollor_dmy = ifelse(NeckLine =="peterpan-collor",
1, 0),
#                    NeckLine_ruffled_dmy = ifelse(NeckLine =="ruffled", 1, 0),
#                    NeckLine_Scoop_dmy = ifelse(NeckLine =="Scoop", 1, 0),
#                    NeckLine_slashneck_dmy = ifelse(NeckLine =="slash-neck", 1, 0),
#                    NeckLine_squarecollor_dmy = ifelse(NeckLine =="square-collor", 1,
0),
#                    NeckLine_sweetheart_dmy = ifelse(NeckLine =="sweetheart", 1, 0),
#                    NeckLine_turndowncollor_dmy = ifelse(NeckLine =="turndowncollor",
1, 0),
#                    NeckLine_vneck_dmy = ifelse(NeckLine =="v-neck", 1, 0)
#                           ) %>% dplyr::select(-c("NeckLine","Style"))


# usdm::vif((ATT_DS2 %>% dplyr::select(c(36:62))))

mylm <- lm(Recommendation ~
                   Style + Price + Size + Rating + Season + NeckLine + SleeveLength +
                   waiseline + Material + FabricType + Decoration + `Pattern Type`,
           data = ATT_DS)

alias( mylm )
#''alias'' refers to the variables that are linearly dependent on others (i.e. cause
perfect multicollinearity).
```

```
# Since NeckLinemandarin-collor is linearly dependent on StyleOL we are going to use
stepAIC to remove unimportant factors and reduce the number of factors using MASS
package stepAIC()

# Complete :
#                        (Intercept) Stylebrief Stylecasual Stylecute Stylefashion
Styleflare Stylenovelty StyleOL Styleparty Stylesexy
# NeckLinemandarin-collor 0           0          0           0         0            0
0            1       0        0
#                        Stylevintage Stylework Pricehigh Pricelow Pricemedium
Pricevery-high SizeL SizeM SizeS SizeXL Rating
# NeckLinemandarin-collor 0            0         0         0        0           0
0     0     0     0      0
#                        Seasonspring Seasonsummer Seasonwinter NeckLineboat-neck
NeckLinebowneck NeckLinehalter NeckLineo-neck
# NeckLinemandarin-collor 0            0            0            0                 0
0              0
#                         NeckLineopen NeckLinepeterpan-collor NeckLineruffled
NeckLineScoop NeckLineslash-neck NeckLinesquare-collor
# NeckLinemandarin-collor 0            0                       0               0
0                  0
#                        NeckLinesweetheart NeckLineturndowncollor NeckLinev-neck
SleeveLengthcap-sleeves SleeveLengthfull
# NeckLinemandarin-collor 0                  0                      0              0
0
#                        SleeveLengthhalfsleeve SleeveLengthNULL SleeveLengthPetal
SleeveLengthshort SleeveLengthsleeveless
# NeckLinemandarin-collor 0                      0                0                 0
0
#                        SleeveLengththreequarter SleeveLengthturndowncollor
waiselineempire waiselinenatural waiselinenull
# NeckLinemandarin-collor 0                        0                          0
0                0
#                        waiseline princess Materialcashmere Materialchiffonfabric
Materialcotton Materialknitting Materiallace
# NeckLinemandarin-collor 0                0                0                     0
0                0
#                        Materiallinen Materiallycra Materialmicrofiber
Materialmilksilk Materialmix Materialmodal Materialnull
# NeckLinemandarin-collor 0             0             0                  0
0           0             0
#                        Materialnylon Materialother Materialpolyster Materialrayon
Materialsilk Materialspandex Materialviscos
# NeckLinemandarin-collor 0             0             0                0             0
0               0
#                        Materialwool FabricTypebroadcloth FabricTypechiffon
FabricTypeCorduroy FabricTypedobby FabricTypeflannel
# NeckLinemandarin-collor 0            0                    0                 0
0               0
#                        FabricTypejersey FabricTypeknitted FabricTypelace
FabricTypenull FabricTypeorganza FabricTypeother
# NeckLinemandarin-collor 0                0                 0              0
0                 0
#                        FabricTypepoplin FabricTypesatin  FabricTypeterry
FabricTypetulle FabricTypewoolen
# NeckLinemandarin-collor 0                0                          0             0
0
#                        FabricTypeworsted Decorationbeading Decorationbow
Decorationbutton Decorationcascading Decorationcrystal
# NeckLinemandarin-collor 0                 0                 0             0
0                   0
#                        Decorationdraped Decorationembroidary Decorationfeathers
Decorationflowers Decorationhollowout Decorationlace
# NeckLinemandarin-collor 0                0                    0                  0
0                   0
#                        Decorationnone Decorationnull Decorationpearls Decorationplain
Decorationpleat Decorationpockets
# NeckLinemandarin-collor 0              0              0                0
0               0
#                        Decorationrivet Decorationruched Decorationruffles
Decorationsashes Decorationsequined Decorationtassel
```

```
# NeckLinemandarin-collor 0               0                 0                  0
0                  0
#                       DecorationTiered `Pattern Type`character `Pattern Type`dot
`Pattern Type`floral `Pattern Type`geometric
# NeckLinemandarin-collor 0                0                       0                 0
0
#                       `Pattern Type`leopard `Pattern Type`none `Pattern Type`null
`Pattern Type`patchwork `Pattern Type`plaid
# NeckLinemandarin-collor 0                     0                  0
0                       0
#                       `Pattern Type`print `Pattern Type`solid `Pattern Type`splice
`Pattern Type`striped
# NeckLinemandarin-collor 0                   0                   0
0


# since we have many categorical values, we are changing them to dichotomous dummy
variables

ATT_DS2 <- ATT_DS %>% mutate( SleeveLengththreequarter =
ifelse(SleeveLength=="threequarter",1,0),
                              SleeveLengthsleeveless =
ifelse(SleeveLength=="sleeveless",1,0),
                              SleeveLengthshort= ifelse(SleeveLength=="short",1,0),
                              SleeveLengthhalfsleeve =
ifelse(SleeveLength=="halfsleeve",1,0),
                              SleeveLengthfull = ifelse(SleeveLength=="full",1,0),
                              SleeveLengthcap_sleeves = ifelse(SleeveLength=="cap-
sleeves",1,0),
                              Seasonspring= ifelse(Season =="spring",1,0),
                              Seasonsummer = ifelse(Season =="summer",1,0),
                              Seasonwinter = ifelse(Season =="winter",1,0),
                              SeasonAutumn = ifelse(Season =="Autumn",1,0),
                              Priceaverage= ifelse(Price =="average",1,0),
                              Pricehigh= ifelse(Price =="high",1,0),
                              Pricelow= ifelse(Price =="low",1,0),
                              Pricemedium= ifelse(Price =="medium",1,0),
                              Pricevery_high= ifelse(Price =="very-high",1,0),
                              Style_bohemian_dmy = ifelse(Style =="bohemian", 1, 0),
                              Style_brief_dmy = ifelse(Style =="brief", 1, 0),
                              Style_casual_dmy = ifelse(Style =="casual", 1, 0),
                              Style_cute_dmy = ifelse(Style =="cute", 1, 0),
                              Style_fashion_dmy = ifelse(Style =="fashion", 1, 0),
                              Style_flare_dmy = ifelse(Style =="flare", 1, 0),
                              Style_novelty_dmy = ifelse(Style =="novelty", 1, 0),
                              Style_OL_dmy = ifelse(Style =="OL", 1, 0),
                              Style_party_dmy = ifelse(Style =="party", 1, 0),
                              Style_sexy_dmy = ifelse(Style =="sexy", 1, 0),
                              Style_vintage_dmy = ifelse(Style =="vintage", 1, 0),
                              Style_work_dmy = ifelse(Style =="work", 1, 0),

                              Material_acrylic_dmy = ifelse(Material =="acrylic", 1,
0),
                              Material_cashmere_dmy = ifelse(Material =="cashmere", 1,
0),
                              Material_chiffon_dmy = ifelse(Material
=="chiffonfabric", 1, 0),
                              Material_cotton_dmy = ifelse(Material =="cotton", 1, 0),
                              Material_knitting_dmy = ifelse(Material =="knitting", 1,
0),
                              Material_linen_dmy = ifelse(Material =="linen", 1, 0),
                              Material_lycra_dmy = ifelse(Material =="lycra", 1, 0),
                              Material_microfiber_dmy = ifelse(Material
=="microfiber", 1, 0),
                              Material_milksilk_dmy = ifelse(Material =="milksilk", 1,
0),
                              Material_mix_dmy = ifelse(Material =="mix", 1, 0),
                              Material_modal_dmy = ifelse(Material =="modal", 1, 0),
                              Material_nylon_dmy = ifelse(Material =="nylon", 1, 0),
                              Material_other_dmy = ifelse(Material %in%
c("other","null"), 1, 0),
```

```
                              Material_polyster_dmy = ifelse(Material =="polyster", 1,
0),
                              Material_rayon_dmy = ifelse(Material =="rayon", 1, 0),
                              Material_silk_dmy = ifelse(Material =="silk", 1, 0),
                              Material_spandex_dmy = ifelse(Material =="spandex", 1,
0),
                              Material_viscos_dmy = ifelse(Material =="viscos", 1, 0),
                              NeckLine_backless_dmy = ifelse(NeckLine =="backless", 1,
0),
                              NeckLine_boatneck_dmy = ifelse(NeckLine =="boat-neck",
1, 0),
                              NeckLine_bowneck_dmy = ifelse(NeckLine =="bowneck", 1,
0),
                              NeckLine_halter_dmy = ifelse(NeckLine =="halter", 1, 0),
                              NeckLine_mandarincollor_dmy = ifelse(NeckLine
=="mandarin-collor", 1, 0),
                              NeckLine_oneck_dmy = ifelse(NeckLine =="o-neck", 1, 0),
                              NeckLine_open_dmy = ifelse(NeckLine =="open", 1, 0),
                              NeckLine_peterpancollor_dmy = ifelse(NeckLine
=="peterpan-collor", 1, 0),
                              NeckLine_ruffled_dmy = ifelse(NeckLine =="ruffled", 1,
0),
                              NeckLine_Scoop_dmy = ifelse(NeckLine =="Scoop", 1, 0),
                              NeckLine_slashneck_dmy = ifelse(NeckLine =="slash-neck",
1, 0),
                              NeckLine_squarecollor_dmy = ifelse(NeckLine =="square-
collor", 1, 0),
                              NeckLine_sweetheart_dmy = ifelse(NeckLine
=="sweetheart", 1, 0),
                              NeckLine_turndowncollor_dmy = ifelse(NeckLine
=="turndowncollor", 1, 0),
                              NeckLine_vneck_dmy = ifelse(NeckLine =="v-neck", 1, 0) )
%>%
        dplyr::select(-
c("Price","Season","SleeveLength","NeckLine","Style","Material"))


library(MASS)
library(caret)

# first creating training and testing of data so that we can validate model later
set.seed(123)
inTrain <- createDataPartition(ATT_DS$Recommendation,p = 0.7, list = FALSE)


Training = ATT_DS2[inTrain,]
Testing = ATT_DS2[-inTrain,]

sum(as.numeric(as.character(ATT_DS$Recommendation)))
#[1] 210

sum(as.numeric(as.character(Training$Recommendation)))
#[1] 147

Training1 <- Training %>% dplyr::select(c(2:8,32:91))
Testing1 <- Testing %>% dplyr::select(c(2:8,32:91))

mylm <- lm(as.numeric(Recommendation) ~ .
                ,
           data = Training1)

alias( mylm )

vif(mylm)
#Error in vif.default(mylm) : there are aliased coefficients in the model

# The autocorrelated variables are these mentioned below (found by using alias(mylm))

# SeasonAutumn  = 1 - Seasonspring  - Seasonsummer - Seasonwinter

# Pricevery_high  =  1 - Priceaverage - Pricehigh - Pricelow - Pricemedium
```

```
# Style_work_dmy  = 1 -Style_bohemian_dmy  -Style_brief_dmy  -Style_casual_dmy  -
Style_cute_dmy  -Style_fashion_dmy  -Style_novelty_dmy  -Style_OL_dmy  -
Style_party_dmy  -Style_sexy_dmy  -Style_vintage_dmy

# NeckLine_mandarincollor_dmy  =  Style_OL_dmy

# NeckLine_open_dmy  = 1 + Style_fashion_dmy  -Material_acrylic_dmy -
Material_cashmere_dmy -Material_chiffon_dmy -Material_cotton_dmy -
Material_knitting_dmy -Material_linen_dmy -Material_lycra_dmy -Material_microfiber_dmy
# -Material_milksilk_dmy -Material_mix_dmy  -Material_modal_dmy -Material_nylon_dmy  -
Material_other_dmy - Material_polyster_dmy - Material_rayon_dmy - Material_silk_dmy -
Material_spandex_dmy
# - Material_viscos_dmy

# NeckLine_vneck_dmy  =  - Style_fashion_dmy -Style_OL_dmy  + Material_acrylic_dmy +
Material_cashmere_dmy + Material_chiffon_dmy + Material_cotton_dmy +
Material_knitting_dmy + Material_linen_dmy + Material_lycra_dmy
# + Material_microfiber_dmy + Material_milksilk_dmy + Material_mix_dmy  +
Material_modal_dmy + Material_nylon_dmy  + Material_other_dmy +  Material_polyster_dmy
+ Material_rayon_dmy
# + Material_silk_dmy  + Material_spandex_dmy +  Material_viscos_dmy  -
NeckLine_boatneck_dmy - NeckLine_bowneck_dmy - NeckLine_halter_dmy -
NeckLine_oneck_dmy - NeckLine_peterpancollor_dmy
# - NeckLine_ruffled_dmy - NeckLine_Scoop_dmy - NeckLine_slashneck_dmy  -
NeckLine_squarecollor_dmy - NeckLine_sweetheart_dmy - NeckLine_turndowncollor_dmy


# In the below glm we have to remove the autocorrelated variables, which we will do
after the following
fit0 = glm(Recommendation ~ ., data= Training1, family = binomial(link = "logit"))
# Got a warning message when i ran the above - it indicates there are too many
variables in the model
# So lets compare correlation between the factors to remove the highly correlated
variables and reduce
# Warning message:
#         glm.fit: fitted probabilities numerically 0 or 1 occurred


fit0 <- update(fit0, .~. -SeasonAutumn -Pricevery_high - Style_work_dmy
 - NeckLine_mandarincollor_dmy - NeckLine_open_dmy - NeckLine_vneck_dmy)

#we still got the following warning for the above update
# Warning message:
#         glm.fit: fitted probabilities numerically 0 or 1 occurred
# since we explored and saw two NAs for the following factors,
# we are going to exclude that as well from the model as these are not significant
factors
# as all of the 350 records are 0
# > table(Training1$NeckLine_backless_dmy)
#
# 0
# 350
# > table(Training1$Style_flare_dmy)
#
# 0
# 350

fit0 <- update(fit0, .~. -NeckLine_backless_dmy - Style_flare_dmy)
# Warning message:
#         glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(fit0)
# since there are too many factors lets first use stepAIC to reduce this to only
significant factors


step1 = stepAIC(fit0, direction = "both")

#Last iteration result #Step:  AIC=436
#
# Step:  AIC=436
# Recommendation ~ Rating + SleeveLengthshort + Seasonspring +
```

```
#         Priceaverage + Pricehigh + Pricelow + Style_cute_dmy + Style_OL_dmy +
#         Material_cashmere_dmy + Material_cotton_dmy + Material_nylon_dmy +
#         Material_other_dmy + NeckLine_bowneck_dmy + NeckLine_oneck_dmy +
#         NeckLine_peterpancollor_dmy + NeckLine_squarecollor_dmy +
#         NeckLine_turndowncollor_dmy + Material_acrylic_dmy
#
# Df Deviance    AIC
# <none>                            398.00 436.00
# - Style_OL_dmy                1   400.60 436.60
# + Style_fashion_dmy           1   396.72 436.72
# + Style_brief_dmy             1   396.81 436.81
# + NeckLine_halter_dmy         1   396.81 436.81
# + NeckLine_ruffled_dmy        1   396.89 436.89
# - Rating                      1   400.93 436.93
# + Material_milksilk_dmy       1   396.96 436.96
# + Style_bohemian_dmy          1   397.01 437.01
# + Material_linen_dmy          1   397.09 437.09
# + SleeveLengthfull            1   397.12 437.12
# + Material_modal_dmy          1   397.12 437.12
# - Material_acrylic_dmy        1   401.19 437.19
# + Seasonwinter                1   397.23 437.23
# + NeckLine_boatneck_dmy       1   397.29 437.29
# - NeckLine_turndowncollor_dmy 1   401.30 437.30
# + Material_knitting_dmy       1   397.31 437.31
# + NeckLine_sweetheart_dmy     1   397.32 437.32
# + NeckLine_slashneck_dmy      1   397.33 437.33
# - Style_cute_dmy              1   401.33 437.33
# + Style_party_dmy             1   397.44 437.44
# + SleeveLengththreequarter    1   397.45 437.45
# + Style_sexy_dmy              1   397.52 437.52
# + Material_rayon_dmy          1   397.54 437.54
# + SleeveLengthhalfsleeve      1   397.55 437.55
# - NeckLine_squarecollor_dmy   1   401.64 437.64
# + SleeveLengthcap_sleeves     1   397.65 437.65
# - Material_cotton_dmy         1   401.67 437.67
# + Style_casual_dmy            1   397.73 437.73
# - NeckLine_oneck_dmy          1   401.74 437.74
# + NeckLine_Scoop_dmy          1   397.78 437.78
# + Material_viscos_dmy         1   397.80 437.80
# + Seasonsummer                1   397.84 437.84
# + Material_mix_dmy            1   397.84 437.84
# - Material_nylon_dmy          1   401.89 437.89
# + Material_spandex_dmy        1   397.89 437.89
# + Material_polyster_dmy       1   397.91 437.91
# + Style_novelty_dmy           1   397.92 437.92
# + Style_vintage_dmy           1   397.93 437.93
# + Material_chiffon_dmy        1   397.95 437.95
# + Material_silk_dmy           1   397.96 437.96
# + SleeveLengthsleeveless      1   397.98 437.98
# + Material_lycra_dmy          1   397.98 437.98
# + Material_microfiber_dmy     1   397.98 437.98
# + Pricemedium                 1   397.99 437.99
# - Material_other_dmy          1   402.91 438.91
# - NeckLine_bowneck_dmy        1   403.21 439.21
# - NeckLine_peterpancollor_dmy 1   403.72 439.72
# - Material_cashmere_dmy       1   403.79 439.79
# + waiseline                   4   394.36 440.36
# + Size                        4   395.68 441.68
# - SleeveLengthshort           1   406.05 442.05
# - Pricehigh                   1   407.51 443.51
# - Pricelow                    1   407.93 443.93
# + `Pattern Type`             12   383.91 445.91
# - Seasonspring                1   410.62 446.62
# - Priceaverage                1   412.68 448.68
# + FabricType                 14   387.24 453.24
# + Decoration                 21   383.82 463.82

#Just checking VIF again as a check
mylm <- lm(as.numeric(Recommendation) ~ Rating + SleeveLengthshort + Seasonspring +
                              Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Style_OL_dmy +
```

```
                              Material_cashmere_dmy + Material_cotton_dmy +
Material_nylon_dmy +
                              Material_other_dmy + NeckLine_bowneck_dmy +
NeckLine_oneck_dmy +
                              NeckLine_peterpancollor_dmy + NeckLine_squarecollor_dmy
+
                              NeckLine_turndowncollor_dmy + Material_acrylic_dmy
                       ,
           data = Training1)
vif(mylm)
# Rating           SleeveLengthshort                Seasonspring
# 1.065692                    1.061912                    1.074698
# Priceaverage                   Pricehigh                    Pricelow
# 2.873957                    1.463455                    2.866266
# Style_cute_dmy                Style_OL_dmy       Material_cashmere_dmy
# 1.059598                    1.045989                    1.107724
# Material_cotton_dmy          Material_nylon_dmy          Material_other_dmy
# 1.321626                    1.078019                    1.314816
# NeckLine_bowneck_dmy          NeckLine_oneck_dmy NeckLine_peterpancollor_dmy
# 1.105460                    1.183940                    1.043706
# NeckLine_squarecollor_dmy NeckLine_turndowncollor_dmy        Material_acrylic_dmy
# 1.037134                    1.068606                    1.023906


# now since we don't have VIF >10 , we can assume these as predictors for glm

mylogit <- glm(Recommendation ~
                       Rating + SleeveLengthshort + Seasonspring +
                       Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Style_OL_dmy +
                       Material_cashmere_dmy + Material_cotton_dmy +
Material_nylon_dmy +
                       Material_other_dmy + NeckLine_bowneck_dmy + NeckLine_oneck_dmy
+
                       NeckLine_peterpancollor_dmy + NeckLine_squarecollor_dmy +
                       NeckLine_turndowncollor_dmy + Material_acrylic_dmy

               # Was trying the following in the past - not needed now as the above is
deemed final significant factors to start with
                       # Priceaverage +
                       # Pricehigh +
                       # Pricelow +
                       # Pricemedium +
                       # Pricevery_high +
                       # Seasonspring +
                       # Seasonsummer +
                       # Seasonwinter +
                       # SeasonAutumn +
                       # SleeveLengththreequarter +
                       # SleeveLengthsleeveless +
                       # SleeveLengthshort +
                       # SleeveLengthhalfsleeve +
                       # SleeveLengthfull +
                       # SleeveLengthcap_sleeves
               ,
               data = Training1, family = binomial(link = "logit"))


summary(mylogit)
#
# Call:
#         glm(formula = Recommendation ~ Rating + SleeveLengthshort + Seasonspring +
#                       Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Style_OL_dmy +
#                       Material_cashmere_dmy + Material_cotton_dmy + Material_nylon_dmy
+
#                       Material_other_dmy + NeckLine_bowneck_dmy + NeckLine_oneck_dmy +
#                       NeckLine_peterpancollor_dmy + NeckLine_squarecollor_dmy +
#                       NeckLine_turndowncollor_dmy + Material_acrylic_dmy, family =
binomial(link = "logit"),
#             data = Training1)
#
```

```
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -1.9113  -0.9863  -0.4891   1.0733   2.3082
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)                  0.4921     0.4945   0.995 0.319611
# Rating                       0.1057     0.0625   1.692 0.090679 .
# SleeveLengthshort           -0.9575     0.3514  -2.725 0.006432 **
#         Seasonspring                  0.9848     0.2812   3.502 0.000461 ***
#         Priceaverage                 -1.6305     0.4479  -3.641 0.000272 ***
#         Pricehigh                    -2.0190     0.6764  -2.985 0.002836 **
#         Pricelow                     -1.3934     0.4615  -3.019 0.002533 **
#         Style_cute_dmy                0.7472     0.4118   1.814 0.069629 .
# Style_OL_dmy               -17.6394  2399.5448  -0.007 0.994135
# Material_cashmere_dmy       18.1105  1677.7192   0.011 0.991387
# Material_cotton_dmy          0.5812     0.3055   1.902 0.057112 .
# Material_nylon_dmy           1.7934     0.9386   1.911 0.056030 .
# Material_other_dmy           0.6699     0.3043   2.202 0.027677 *
#         NeckLine_bowneck_dmy         -2.3582     1.2365  -1.907 0.056492 .
# NeckLine_oneck_dmy          -0.4957     0.2571  -1.928 0.053840 .
# NeckLine_peterpancollor_dmy  -16.3605  1063.6762  -0.015 0.987728
# NeckLine_squarecollor_dmy    -16.4364  1320.2093  -0.012 0.990067
# NeckLine_turndowncollor_dmy   -1.4237     0.8558  -1.664 0.096197 .
# Material_acrylic_dmy        -16.2223  1272.7943  -0.013 0.989831
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.2  on 349  degrees of freedom
# Residual deviance: 398.0  on 331  degrees of freedom
# AIC: 436
#
# Number of Fisher Scoring iterations: 15
vif(mylogit)

# Rating           SleeveLengthshort                Seasonspring
# 1.054771                   1.057073                     1.041390
# Priceaverage                  Pricehigh                     Pricelow
# 3.423224                   1.615719                     3.436006
# Style_cute_dmy                Style_OL_dmy       Material_cashmere_dmy
# 1.061445                   1.000000                     1.000000
# Material_cotton_dmy          Material_nylon_dmy          Material_other_dmy
# 1.373702                   1.082158                     1.340865
# NeckLine_bowneck_dmy          NeckLine_oneck_dmy NeckLine_peterpancollor_dmy
# 1.039433                   1.120444                     1.000000
# NeckLine_squarecollor_dmy NeckLine_turndowncollor_dmy        Material_acrylic_dmy
# 1.000000                   1.055415                     1.000000

# no multicollinear factors (as vif <10) - which is good
#now we are going to pool the most insignificant term into the error degrees of
freedom

fitA = update(mylogit, .~. -Style_OL_dmy)
summary(fitA)
# glm(formula = Recommendation ~ Rating + SleeveLengthshort + Seasonspring +
#             Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cashmere_dmy +
#             Material_cotton_dmy + Material_nylon_dmy + Material_other_dmy +
#             NeckLine_bowneck_dmy + NeckLine_oneck_dmy + NeckLine_peterpancollor_dmy
+
#             NeckLine_squarecollor_dmy + NeckLine_turndowncollor_dmy +
#             Material_acrylic_dmy, family = binomial(link = "logit"),
#     data = Training1)
#
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -1.8592  -0.9845  -0.5002   1.0707   2.3027
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
```

```
# (Intercept)                    0.34667    0.47978   0.723 0.469952
# Rating                         0.11254    0.06234   1.805 0.071023 .
# SleeveLengthshort             -0.93085    0.34939  -2.664 0.007717 **
#        Seasonspring                   0.99478    0.28005   3.552 0.000382 ***
#        Priceaverage                  -1.51381    0.43483  -3.481 0.000499 ***
#        Pricehigh                     -1.89949    0.66761  -2.845 0.004438 **
#        Pricelow                      -1.27682    0.44875  -2.845 0.004437 **
#        Style_cute_dmy                 0.75404    0.41060   1.836 0.066291 .
# Material_cashmere_dmy         18.18322 1673.79498   0.011 0.991332
# Material_cotton_dmy            0.54066    0.30371   1.780 0.075045 .
# Material_nylon_dmy             1.79219    0.93712   1.912 0.055820 .
# Material_other_dmy             0.67130    0.30329   2.213 0.026868 *
#        NeckLine_bowneck_dmy          -2.31276    1.22527  -1.888 0.059086 .
# NeckLine_oneck_dmy            -0.48007    0.25648  -1.872 0.061240 .
# NeckLine_peterpancollor_dmy  -16.31541 1062.88934  -0.015 0.987753
# NeckLine_squarecollor_dmy    -16.37607 1331.12285  -0.012 0.990184
# NeckLine_turndowncollor_dmy   -1.37713    0.85106  -1.618 0.105633
# Material_acrylic_dmy         -16.23737 1267.73560  -0.013 0.989781
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.2  on 349  degrees of freedom
# Residual deviance: 400.6  on 332  degrees of freedom
# AIC: 436.6
#
# Number of Fisher Scoring iterations: 15

vif(fitA)
# Rating           SleeveLengthshort                Seasonspring
# 1.061180                   1.053520                    1.041168
# Priceaverage                  Pricehigh                    Pricelow
# 3.247463                   1.572441                    3.264356
# Style_cute_dmy       Material_cashmere_dmy         Material_cotton_dmy
# 1.061364                   1.000000                    1.367792
# Material_nylon_dmy          Material_other_dmy        NeckLine_bowneck_dmy
# 1.082365                   1.339567                    1.037920
# NeckLine_oneck_dmy NeckLine_peterpancollor_dmy   NeckLine_squarecollor_dmy
# 1.123985                   1.000000                    1.000000
# NeckLine_turndowncollor_dmy        Material_acrylic_dmy
# 1.053661                   1.000000

fitB = update(fitA, .~. -Material_cashmere_dmy)
summary(fitB)
# Call:
#        glm(formula = Recommendation ~ Rating + SleeveLengthshort + Seasonspring +
#                   Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cotton_dmy +
#                   Material_nylon_dmy + Material_other_dmy + NeckLine_bowneck_dmy +
#                   NeckLine_oneck_dmy + NeckLine_peterpancollor_dmy +
NeckLine_squarecollor_dmy +
#                   NeckLine_turndowncollor_dmy + Material_acrylic_dmy, family =
binomial(link = "logit"),
#            data = Training1)
#
# Deviance Residuals:
#        Min       1Q   Median       3Q      Max
# -1.8934  -0.9870  -0.5844   1.0852   2.2515
#
# Coefficients:
#        Estimate Std. Error z value Pr(>|z|)
# (Intercept)                    0.54754    0.47564   1.151 0.249660
# Rating                         0.09176    0.06111   1.502 0.133216
# SleeveLengthshort             -0.95207    0.34801  -2.736 0.006223 **
#        Seasonspring                   0.95360    0.27778   3.433 0.000597 ***
#        Priceaverage                  -1.55369    0.43373  -3.582 0.000341 ***
#        Pricehigh                     -1.96566    0.66635  -2.950 0.003179 **
#        Pricelow                      -1.33749    0.44831  -2.983 0.002851 **
#        Style_cute_dmy                 0.69720    0.40843   1.707 0.087819 .
# Material_cotton_dmy            0.47177    0.29970   1.574 0.115456
# Material_nylon_dmy             1.71483    0.93623   1.832 0.067005 .
```

```
# Material_other_dmy             0.60293    0.30043   2.007 0.044759 *
#         NeckLine_bowneck_dmy          -1.48781    0.92506  -1.608 0.107762
# NeckLine_oneck_dmy            -0.49381    0.25503  -1.936 0.052835 .
# NeckLine_peterpancollor_dmy  -16.34696 1063.39050  -0.015 0.987735
# NeckLine_squarecollor_dmy    -16.41475 1334.86486  -0.012 0.990189
# NeckLine_turndowncollor_dmy   -1.39579    0.85092  -1.640 0.100937
# Material_acrylic_dmy         -16.25407 1280.40697  -0.013 0.989872
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 406.58  on 333  degrees of freedom
# AIC: 440.58
#
# Number of Fisher Scoring iterations: 15

vif(fitB)
# Rating           SleeveLengthshort                Seasonspring
# 1.054818                   1.054156                     1.039644
# Priceaverage                  Pricehigh                    Pricelow
# 3.287177                   1.580767                     3.305692
# Style_cute_dmy         Material_cotton_dmy          Material_nylon_dmy
# 1.051321                   1.349244                     1.078984
# Material_other_dmy        NeckLine_bowneck_dmy          NeckLine_oneck_dmy
# 1.324937                   1.044119                     1.131673
# NeckLine_peterpancollor_dmy   NeckLine_squarecollor_dmy NeckLine_turndowncollor_dmy
# 1.000000                   1.000000                     1.052936
# Material_acrylic_dmy
# 1.000000


fitC = update(fitB, .~. -NeckLine_squarecollor_dmy)
summary(fitC)
# Call:
#         glm(formula = Recommendation ~ Rating + SleeveLengthshort + Seasonspring +
#                      Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cotton_dmy +
#                      Material_nylon_dmy + Material_other_dmy + NeckLine_bowneck_dmy +
#                      NeckLine_oneck_dmy + NeckLine_peterpancollor_dmy +
NeckLine_turndowncollor_dmy +
#                      Material_acrylic_dmy, family = binomial(link = "logit"),
#             data = Training1)
#
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -1.8942  -0.9898  -0.5997   1.0740   2.2593
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)                    0.45653    0.46807   0.975 0.329389
# Rating                         0.08383    0.06080   1.379 0.167927
# SleeveLengthshort             -0.98408    0.34822  -2.826 0.004712 **
#         Seasonspring                  0.97022    0.27692   3.504 0.000459 ***
#         Priceaverage                 -1.49451    0.42317  -3.532 0.000413 ***
#         Pricehigh                    -1.88071    0.65997  -2.850 0.004376 **
#         Pricelow                     -1.25546    0.43834  -2.864 0.004182 **
#         Style_cute_dmy                0.71755    0.40791   1.759 0.078559 .
# Material_cotton_dmy            0.46934    0.29748   1.578 0.114626
# Material_nylon_dmy             1.73300    0.93587   1.852 0.064063 .
# Material_other_dmy             0.63440    0.29877   2.123 0.033724 *
#         NeckLine_bowneck_dmy         -1.42779    0.92270  -1.547 0.121767
# NeckLine_oneck_dmy            -0.44909    0.25342  -1.772 0.076375 .
# NeckLine_peterpancollor_dmy  -16.27854 1065.20921  -0.015 0.987807
# NeckLine_turndowncollor_dmy   -1.33417    0.84929  -1.571 0.116201
# Material_acrylic_dmy         -16.25203 1276.21595  -0.013 0.989840
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
```

```
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 410.18  on 334  degrees of freedom
# AIC: 442.18
#
# Number of Fisher Scoring iterations: 15

vif(fitC)
# Rating           SleeveLengthshort                Seasonspring
# 1.050076                    1.061667                    1.039557
# Priceaverage                   Pricehigh                    Pricelow
# 3.155762                    1.550552                    3.174590
# Style_cute_dmy         Material_cotton_dmy         Material_nylon_dmy
# 1.052199                    1.340407                    1.078509
# Material_other_dmy        NeckLine_bowneck_dmy           NeckLine_oneck_dmy
# 1.316823                    1.041896                    1.129126
# NeckLine_peterpancollor_dmy NeckLine_turndowncollor_dmy        Material_acrylic_dmy
# 1.000000                    1.050858                    1.000000

fitD = update(fitC, .~. -Material_acrylic_dmy)
summary(fitD)
# Call:
#         glm(formula = Recommendation ~ Rating + SleeveLengthshort + Seasonspring +
#                     Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cotton_dmy +
#                     Material_nylon_dmy + Material_other_dmy + NeckLine_bowneck_dmy +
#                     NeckLine_oneck_dmy + NeckLine_peterpancollor_dmy +
NeckLine_turndowncollor_dmy,
#             family = binomial(link = "logit"), data = Training1)
#
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -1.8960  -0.9796  -0.6046   1.0872   2.2685
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)                   0.43909    0.46550   0.943 0.345542
# Rating                        0.08414    0.06046   1.392 0.164046
# SleeveLengthshort            -0.95823    0.34757  -2.757 0.005835 **
#         Seasonspring                  0.95726    0.27556   3.474 0.000513 ***
#         Priceaverage                 -1.52055    0.42232  -3.600 0.000318 ***
#         Pricehigh                    -1.87801    0.65932  -2.848 0.004394 **
#         Pricelow                     -1.26409    0.43649  -2.896 0.003779 **
#         Style_cute_dmy                0.64548    0.40143   1.608 0.107844
# Material_cotton_dmy           0.50826    0.29635   1.715 0.086333 .
# Material_nylon_dmy            1.75680    0.93507   1.879 0.060273 .
# Material_other_dmy            0.67397    0.29773   2.264 0.023595 *
#         NeckLine_bowneck_dmy         -1.40320    0.91997  -1.525 0.127194
# NeckLine_oneck_dmy           -0.45402    0.25253  -1.798 0.072198 .
# NeckLine_peterpancollor_dmy -15.27298  645.73236  -0.024 0.981130
# NeckLine_turndowncollor_dmy  -1.33303    0.84996  -1.568 0.116801
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 413.45  on 335  degrees of freedom
# AIC: 443.45
#
# Number of Fisher Scoring iterations: 14

vif(fitD)
# Rating           SleeveLengthshort                Seasonspring
# 1.048515                    1.059441                    1.043693
# Priceaverage                   Pricehigh                    Pricelow
# 3.167561                    1.549416                    3.173094
# Style_cute_dmy         Material_cotton_dmy         Material_nylon_dmy
# 1.055153                    1.336955                    1.077464
# Material_other_dmy        NeckLine_bowneck_dmy           NeckLine_oneck_dmy
# 1.313407                    1.041336                    1.129971
# NeckLine_peterpancollor_dmy NeckLine_turndowncollor_dmy
# 1.000000                    1.051007
```

```
fitE = update(fitD, .~. -NeckLine_peterpancollor_dmy)
summary(fitE)
# Call:
#        glm(formula = Recommendation ~ Rating + SleeveLengthshort + Seasonspring +
#                    Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cotton_dmy +
#                    Material_nylon_dmy + Material_other_dmy + NeckLine_bowneck_dmy +
#                    NeckLine_oneck_dmy + NeckLine_turndowncollor_dmy, family =
binomial(link = "logit"),
#            data = Training1)
#
# Deviance Residuals:
#        Min       1Q   Median       3Q      Max
# -1.9019  -0.9833  -0.6134   1.1110   2.2867
#
# Coefficients:
#        Estimate Std. Error z value Pr(>|z|)
# (Intercept)                  0.35941    0.45860   0.784 0.433211
# Rating                       0.08481    0.05996   1.414 0.157239
# SleeveLengthshort           -1.00047    0.34676  -2.885 0.003911 **
#        Seasonspring                 0.98719    0.27471   3.594 0.000326 ***
#        Priceaverage                -1.50933    0.41590  -3.629 0.000284 ***
#        Pricehigh                   -1.81508    0.65550  -2.769 0.005623 **
#        Pricelow                    -1.23214    0.42978  -2.867 0.004145 **
#        Style_cute_dmy               0.62214    0.39614   1.570 0.116305
# Material_cotton_dmy          0.48039    0.29488   1.629 0.103294
# Material_nylon_dmy           1.77570    0.93457   1.900 0.057430 .
# Material_other_dmy           0.67125    0.29680   2.262 0.023719 *
#        NeckLine_bowneck_dmy        -1.33344    0.91776  -1.453 0.146245
# NeckLine_oneck_dmy          -0.38812    0.25072  -1.548 0.121619
# NeckLine_turndowncollor_dmy -1.25268    0.84858  -1.476 0.139889
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 418.83  on 336  degrees of freedom
# AIC: 446.83
#
# Number of Fisher Scoring iterations: 4

fitF = update(fitE, .~. -Rating)
summary(fitF)
# Call:
#        glm(formula = Recommendation ~ SleeveLengthshort + Seasonspring +
#                    Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cotton_dmy +
#                    Material_nylon_dmy + Material_other_dmy + NeckLine_bowneck_dmy +
#                    NeckLine_oneck_dmy + NeckLine_turndowncollor_dmy, family =
binomial(link = "logit"),
#            data = Training1)
#
# Deviance Residuals:
#        Min       1Q   Median       3Q      Max
# -2.0211  -0.9723  -0.6231   1.1324   2.1409
#
# Coefficients:
#        Estimate Std. Error z value Pr(>|z|)
# (Intercept)                   0.6061     0.4245   1.428 0.153322
# SleeveLengthshort            -0.9745     0.3441  -2.832 0.004629 **
#        Seasonspring                  0.9756     0.2731   3.573 0.000353 ***
#        Priceaverage                 -1.4678     0.4128  -3.556 0.000377 ***
#        Pricehigh                    -1.7805     0.6499  -2.740 0.006153 **
#        Pricelow                     -1.1723     0.4258  -2.753 0.005899 **
#        Style_cute_dmy                0.5822     0.3936   1.479 0.139120
# Material_cotton_dmy           0.4480     0.2930   1.529 0.126245
# Material_nylon_dmy            1.6741     0.9351   1.790 0.073396 .
# Material_other_dmy            0.6707     0.2960   2.266 0.023435 *
#        NeckLine_bowneck_dmy         -1.3135     0.9092  -1.445 0.148527
```

```
# NeckLine_oneck_dmy           -0.3488     0.2482  -1.406 0.159826
# NeckLine_turndowncollor_dmy  -1.2184     0.8508  -1.432 0.152095
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 420.87  on 337  degrees of freedom
# AIC: 446.87
#
# Number of Fisher Scoring iterations: 4

fitG = update(fitF, .~. -NeckLine_oneck_dmy)
summary(fitG)
# Call:
#         glm(formula = Recommendation ~ SleeveLengthshort + Seasonspring +
#                       Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cotton_dmy +
#                       Material_nylon_dmy + Material_other_dmy + NeckLine_bowneck_dmy +
#                       NeckLine_turndowncollor_dmy, family = binomial(link = "logit"),
#             data = Training1)
#
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -2.1183  -0.9698  -0.6140   1.1091   2.0829
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)                   0.4605     0.4096   1.124 0.260904
# SleeveLengthshort            -0.9707     0.3441  -2.821 0.004789 **
#         Seasonspring                 0.9641     0.2716   3.549 0.000386 ***
#         Priceaverage                -1.5378     0.4091  -3.759 0.000171 ***
#         Pricehigh                   -1.7284     0.6467  -2.672 0.007529 **
#         Pricelow                    -1.2619     0.4207  -3.000 0.002703 **
#         Style_cute_dmy               0.5657     0.3904   1.449 0.147308
# Material_cotton_dmy           0.4752     0.2917   1.629 0.103333
# Material_nylon_dmy            1.7296     0.9241   1.872 0.061264 .
# Material_other_dmy            0.7068     0.2946   2.399 0.016447 *
#         NeckLine_bowneck_dmy        -1.1242     0.9038  -1.244 0.213554
# NeckLine_turndowncollor_dmy  -1.0299     0.8431  -1.222 0.221874
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 422.84  on 338  degrees of freedom
# AIC: 446.84
#
# Number of Fisher Scoring iterations: 4

fitH = update(fitG, .~. -NeckLine_turndowncollor_dmy)
summary(fitH)
# Call:
#         glm(formula = Recommendation ~ SleeveLengthshort + Seasonspring +
#                       Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cotton_dmy +
#                       Material_nylon_dmy + Material_other_dmy + NeckLine_bowneck_dmy,
#             family = binomial(link = "logit"), data = Training1)
#
# Deviance Residuals:
#         Min      1Q  Median      3Q     Max
# -2.108  -1.018  -0.678   1.115   2.092
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)            0.4276     0.4080   1.048 0.294618
# SleeveLengthshort     -0.9665     0.3429  -2.819 0.004816 **
#         Seasonspring          0.9632     0.2703   3.563 0.000367 ***
#         Priceaverage         -1.5311     0.4082  -3.751 0.000176 ***
#         Pricehigh            -1.7018     0.6463  -2.633 0.008457 **
```

```
#         Pricelow              -1.2419     0.4194  -2.961 0.003065 **
#         Style_cute_dmy         0.5945     0.3897   1.525 0.127151
# Material_cotton_dmy    0.4382     0.2896   1.513 0.130163
# Material_nylon_dmy     1.7421     0.9239   1.886 0.059347 .
# Material_other_dmy     0.7166     0.2938   2.439 0.014731 *
#         NeckLine_bowneck_dmy  -1.0970     0.9063  -1.210 0.226099
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 424.58  on 339  degrees of freedom
# AIC: 446.58
#
# Number of Fisher Scoring iterations: 4

fitI = update(fitH, .~. -NeckLine_bowneck_dmy)
summary(fitI)

# Call:
#         glm(formula = Recommendation ~ SleeveLengthshort + Seasonspring +
#                     Priceaverage + Pricehigh + Pricelow + Style_cute_dmy +
Material_cotton_dmy +
#                     Material_nylon_dmy + Material_other_dmy, family = binomial(link
= "logit"),
#             data = Training1)
#
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -2.0859  -1.0132  -0.6814   1.1252   2.0960
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)           0.3544     0.4012   0.883 0.377060
# SleeveLengthshort    -0.9453     0.3412  -2.771 0.005592 **
#         Seasonspring          0.9611     0.2693   3.569 0.000359 ***
#         Priceaverage         -1.4878     0.4047  -3.676 0.000237 ***
#         Pricehigh            -1.6424     0.6435  -2.552 0.010708 *
#         Pricelow             -1.1995     0.4155  -2.887 0.003892 **
#         Style_cute_dmy        0.5915     0.3918   1.510 0.131081
# Material_cotton_dmy   0.4457     0.2887   1.544 0.122652
# Material_nylon_dmy    1.7636     0.9234   1.910 0.056131 .
# Material_other_dmy    0.7394     0.2939   2.516 0.011870 *
#         ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 426.24  on 340  degrees of freedom
# AIC: 446.24
#
# Number of Fisher Scoring iterations: 4

fitJ = update(fitI, .~. -Style_cute_dmy)
summary(fitJ)

# Call:
#         glm(formula = Recommendation ~ SleeveLengthshort + Seasonspring +
#                     Priceaverage + Pricehigh + Pricelow + Material_cotton_dmy +
#                     Material_nylon_dmy + Material_other_dmy, family = binomial(link
= "logit"),
#             data = Training1)
#
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -2.1006  -1.0098  -0.7043   1.0918   2.0414
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)           0.3996     0.3979   1.004 0.315148
```

```
# SleeveLengthshort    -0.8856     0.3358  -2.637 0.008366 **
#         Seasonspring          1.0070     0.2676   3.763 0.000168 ***
#         Priceaverage         -1.4648     0.4017  -3.647 0.000266 ***
#         Pricehigh            -1.5912     0.6397  -2.488 0.012863 *
#         Pricelow             -1.1791     0.4133  -2.853 0.004329 **
#         Material_cotton_dmy   0.4189     0.2870   1.459 0.144500
# Material_nylon_dmy    1.6709     0.9201   1.816 0.069370 .
# Material_other_dmy    0.6831     0.2901   2.355 0.018539 *
#         ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 428.52  on 341  degrees of freedom
# AIC: 446.52
#
# Number of Fisher Scoring iterations: 4

fitK = update(fitJ, .~. -Material_cotton_dmy)
summary(fitK)
#
# Call:
#         glm(formula = Recommendation ~ SleeveLengthshort + Seasonspring +
#                     Priceaverage + Pricehigh + Pricelow + Material_nylon_dmy +
#                     Material_other_dmy, family = binomial(link = "logit"), data =
Training1)
#
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -2.0590  -0.9471  -0.6520   1.1392   1.9469
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)          0.5379     0.3860   1.393 0.163506
# SleeveLengthshort   -0.8711     0.3348  -2.602 0.009264 **
#         Seasonspring         0.9517     0.2631   3.617 0.000298 ***
#         Priceaverage        -1.3991     0.3970  -3.524 0.000425 ***
#         Pricehigh           -1.5736     0.6379  -2.467 0.013626 *
#         Pricelow            -1.1071     0.4085  -2.710 0.006733 **
#         Material_nylon_dmy   1.4823     0.9089   1.631 0.102913
# Material_other_dmy   0.5022     0.2596   1.934 0.053067 .
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 430.66  on 342  degrees of freedom
# AIC: 446.66
#
# Number of Fisher Scoring iterations: 4


fitL = update(fitK, .~. -Material_nylon_dmy)
summary(fitL)
# Call:
#         glm(formula = Recommendation ~ SleeveLengthshort + Seasonspring +
#                     Priceaverage + Pricehigh + Pricelow + Material_other_dmy,
#             family = binomial(link = "logit"), data = Training1)
#
# Deviance Residuals:
#         Min       1Q   Median       3Q      Max
# -2.0408  -0.9661  -0.6689   1.1339   1.9321
#
# Coefficients:
#         Estimate Std. Error z value Pr(>|z|)
# (Intercept)          0.5515     0.3853   1.431 0.152338
# SleeveLengthshort   -0.8637     0.3333  -2.591 0.009568 **
#         Seasonspring         0.9381     0.2620   3.581 0.000342 ***
#         Priceaverage        -1.3864     0.3960  -3.501 0.000463 ***
#         Pricehigh           -1.3952     0.6180  -2.258 0.023969 *
```

```
#         Pricelow             -1.0712     0.4072  -2.631 0.008512 **
#         Material_other_dmy   0.4595     0.2576   1.784 0.074494 .
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 476.20  on 349  degrees of freedom
# Residual deviance: 433.52  on 343  degrees of freedom
# AIC: 447.52
#
# Number of Fisher Scoring iterations: 4

#Looks like we can't pool any more insignificant terms.
# Material_other_dmy   is marginally significant as it has p-val 0.07 which is not
that far from 0.05 threshold alpha value which we assumed above.
# all of the above independent terms are significant

# But fitL has AIC 447.52 when the last of stepaic was only 436. Hence we need to go
with that itself
# This was just done to check if we could do something to improvise.
# The only possible explanation could be we have just seen main effects
# Probably interaction is needed and some of the insignificant terms could be useful
to make something else significant.
# anyway for now lets consider this mylogit as the final logit run

mylogit$coefficients
# (Intercept)                      Rating           SleeveLengthshort
# 0.4921253                 0.1057431                  -0.9575346
# Seasonspring              Priceaverage                  Pricehigh
# 0.9847531                -1.6305465                  -2.0189868
# Pricelow              Style_cute_dmy               Style_OL_dmy
# -1.3933465                 0.7471881                  -17.6393730
# Material_cashmere_dmy         Material_cotton_dmy          Material_nylon_dmy
# 18.1105515                 0.5811792                   1.7934062
# Material_other_dmy        NeckLine_bowneck_dmy          NeckLine_oneck_dmy
# 0.6699423                -2.3582023                  -0.4957287
# NeckLine_peterpancollor_dmy   NeckLine_squarecollor_dmy NeckLine_turndowncollor_dmy
# -16.3604505               -16.4363868                  -1.4237090
# Material_acrylic_dmy
# -16.2223475


mylogit$fitted.values
Pred = predict(mylogit, newdata = Training1[,-7], type= "response") # checking this
with above

sum(mylogit$fitted.values - Pred) # just to show both are equal- diff is 0

#Now to predict the test data
Pred_Test = predict(mylogit, newdata = Testing1[,-7], type= "response")

#This is training datas below
Pred1 = ifelse(Pred <0.5,0,1) # making it binary - converting probability values as 0
and 1

Pred_Test1 = ifelse(Pred_Test <0.5,0,1) # making it binary - converting probability
values as 0 and 1

library(e1071)

a = table(Training1$Recommendation, Pred1, dnn = list('actual','predicted'))

#               predicted
# actual        0   1
# 0             163  40
# 1             71  76


caret::confusionMatrix(a, positive = '1')

predicted
```

```
actual   0   1
0 163  40
1  71  76
> caret::confusionMatrix(a, positive = '1')
Confusion Matrix and Statistics

# predicted
# actual   0   1
# 0 163  40
# 1  71  76
#
# Accuracy : 0.6829
# 95% CI : (0.6313, 0.7313)
# No Information Rate : 0.6686
# P-Value [Acc > NIR] : 0.306371
#
# Kappa : 0.3295
#
# Mcnemar's Test P-Value : 0.004407
#
#             Sensitivity : 0.6552
#             Specificity : 0.6966
#          Pos Pred Value : 0.5170
#          Neg Pred Value : 0.8030
#              Prevalence : 0.3314
#          Detection Rate : 0.2171
#    Detection Prevalence : 0.4200
#       Balanced Accuracy : 0.6759
#
#        'Positive' Class : 1


#Checking the testing dataset
atest = table(Testing1$Recommendation, Pred_Test1, dnn = list('actual','predicted'))
atest
caret::confusionMatrix(atest, positive = '1')

#install.packages("InformationValue")
library(InformationValue)

plotROC(actuals=Training1$Recommendation, predictedScores=Pred1)

#visually inspecting sensitivity plot
sensMat <- plotROC(actuals=Training1$Recommendation,  predictedScores=Pred1,
returnSensitivityMat = TRUE)
sensMat

#Sensitivity, also considered as the 'True Positive Rate' or 'recall' is the
# proportion of 'Events' (or 'Ones') correctly predicted by the model,
# for a given prediction probability cutoff score


# For a given probability score cutoff (threshold), precision or 'positive predictive
value'
# computes the proportion of the total events (ones) out of the total that were
predicted
# to be events (ones).

precision(actuals=Training1$Recommendation,
          predictedScores=Pred1)
# 0.6551724

# somersD computes how many more concordant than discordant pairs exist divided by the
# total number of pairs. Larger the Somers D value, better model's predictive ability
# Concordance is the percentage of predicted probability scores where the scores of
actual
# positive's are greater than the scores of actual negative's.
# It is calculated by taking into account the scores of all possible pairs of Ones
# and Zeros. If the concordance of a model is 100%, it means that,
# by tweaking the prediction probability cutoff, we could accurately predict
# all of the events and non-events
```

```
somersD(actuals=Training1$Recommendation,
        predictedScores=Pred) # we are using Pred instead of Pred1 (binary values)
because somersD uses actual prediction values
#0.4887571

ks_stat(actuals=Training1$Recommendation,
        predictedScores=Pred)
# [1] 0.3343

ks_plot(actuals=Training1$Recommendation,
        predictedScores=Pred)

# ks_plot plots the lift is capturing the responders (Ones) against the the random
case
# where we don't use the model. The more curvier (higher) the model curve,
# the better is the model.

#For example, from the above chart for instance, by targeting first 40% of the Dress,
# the model will be able to capture 58.5% of total responders(Ones),
# while without the model, you can expect to capture only 40% of responders by random
targeting.

# Based on the ks_plot and ROC curve, we have a decent model . Still we are going to
explore Decision Tree/RF
####Start DT

# first creating training and testing of data so that we can validate model later
set.seed(123)
inTrain <- createDataPartition(ATT_DS$Recommendation,p = 0.7, list = FALSE)

ATT_DS3 <- ATT_DS2 %>%
        mutate(RatingClass = cut(Rating,c(-0.1,4,max(Rating)),
                                 labels = c("<=4",">4")))%>%
        mutate_if(is.numeric,as.factor) %>%
        mutate_if(is.character,as.factor) # for DT lets use this dataset

Training = ATT_DS3[inTrain,]
Testing = ATT_DS3[-inTrain,]

# since rating is numeric and has 16 values, lets use cut and make it into few groups
to make it easier to view
# Training %>%
#         dplyr::select(c(2:8,32:91)) %>%
#              mutate(RatingClass = cut(Rating,c(-0.1,1,2,3,4,max(Rating)),
#                                       labels = c("0-1","1-2","2-3","3-4","4-5")))
%>%
#         dplyr::select(RatingClass) %>% table()

# 0-1 1-2 2-3 3-4 4-5
# 89   0   1   6 254
# Testing %>% dplyr::select(c(2:8,32:91)) %>%
#              mutate(RatingClass = cut(Rating,c(-0.1,1,2,3,4,max(Rating)),
#              labels = c("0-1","1-2","2-3","3-4","4-5"))) %>%
#         dplyr::select(RatingClass) %>% table()
#
# 0-1 1-2 2-3 3-4 4-5
# 32   0   0   5 112

#so max rating is either 4-5 or 0-1 with very few in 3-4; so lets cut it into two >=4
or less than 4
Training1 <- Training %>% dplyr::select(c(2:8,32:92)) %>%
                     dplyr::select(-c(Rating))

Testing1 <- Testing %>% dplyr::select(c(2:8,32:92)) %>%
        dplyr::select(-c(Rating))


# Now for DT, we need rpart and rpart.plot; Referred based on Dec 18 class of Kapil
library(rpart)
library(rpart.plot)
```

```
fit = rpart(Recommendation ~ ., method ="class", data = Training1) # method = class
for classification tree;
rpart.plot(fit, cex=0.5)

#suppose we wish to prune the tree we could use the option rpart.control below. but
testing the different values, the minsplit of 30 which was the one chosen above
without using the option by default seems to be the best model in this
fit = rpart(Recommendation ~ ., method ="class", data = Training1,
            control = rpart.control(minsplit = 40))#// minimum 30 data records need to
be there to split further

rpart.plot(fit, cex=0.5, extra = 1) # using absolute values
rpart.plot(fit, cex=0.7)


#now lets validate the model using Test data

Pred = predict(fit, newdata = Testing1[,-6], type="class") # without the dependent
column as best practice; Sourcing my model fit to it

table(Pred)
# Pred
# 0   1
# 109  40

a= table(Testing1$Recommendation, Pred, dnn = list("actual","predicted"))

caret::confusionMatrix(a, positive ="1")
# Confusion Matrix and Statistics
#
# predicted
# actual  0  1
# 0 70 16
# 1 39 24
#
# Accuracy : 0.6309
# 95% CI : (0.548, 0.7084)
# No Information Rate : 0.7315
# P-Value [Acc > NIR] : 0.997297
#
# Kappa : 0.2049
#
# Mcnemar's Test P-Value : 0.003012
#
#              Sensitivity : 0.6000
#              Specificity : 0.6422
#           Pos Pred Value : 0.3810
#           Neg Pred Value : 0.8140
#               Prevalence : 0.2685
#           Detection Rate : 0.1611
#     Detection Prevalence : 0.4228
#        Balanced Accuracy : 0.6211
#
#         'Positive' Class : 1

# Looking at the metrics (62.11% for balanced accuracy) of DT, it looks slightly
better than logistic regression

####end DT

#### Random Forest- lets try another algo as well
library(randomForest)

F1 = randomForest(Recommendation ~ ., data = Training1, ntree = 1000)
#Error in eval(predvars, data, env) : object 'Pattern Type' not found
#Guessing its not accepting spaces in the variable names- hence lets rename that in
both training and test

Training1 <- Training1 %>% rename(Pattern_Type=`Pattern Type`)
Testing1 <- Testing1 %>% rename(Pattern_Type=`Pattern Type`)
set.seed(2345)
F1 = randomForest(Recommendation ~ ., data = Training1, ntree = 3000) #3000 trees used
```

```
aa= predict(F1,Testing1,type ="vote")
aa1 = ifelse(aa[,2]>0.5,"Good","Bad")
table(Testing1$Recommendation,aa1)

aa = predict(F1, newdata = Testing1[,-6], type="class") # without the dependent column
as best practice; Sourcing my model fit to it

table(aa)
# Pred
# 0   1
# 95  54

b= table(Testing1$Recommendation, aa, dnn = list("actual","predicted"))

caret::confusionMatrix(b, positive ="1")
#
# Confusion Matrix and Statistics
#
# predicted
# actual  0  1
# 0 64 22
# 1 31 32
#
# Accuracy : 0.6443
# 95% CI : (0.5618, 0.7209)
# No Information Rate : 0.6376
# P-Value [Acc > NIR] : 0.4692
#
# Kappa : 0.257
#
# Mcnemar's Test P-Value : 0.2718
#
#            Sensitivity : 0.5926
#            Specificity : 0.6737
#         Pos Pred Value : 0.5079
#         Neg Pred Value : 0.7442
#             Prevalence : 0.3624
#         Detection Rate : 0.2148
#   Detection Prevalence : 0.4228
#      Balanced Accuracy : 0.6331
#
#       'Positive' Class : 1


####End Random Forest <best Balanced Accuracy/Sensitivity: 0.6331>;
#There are other metrics too (sensitivity/specificity).
#So depending on what is more important to us for the biz decision, we can choose that
as our model - Lets choose Log Reg as our model
#But Random forest overall performs best among these 3

#### Task 2: In order to stock the inventory, the store wants to analyze the sales
data and predict the trend of total sales for each dress for an extended period of
three more alternative days.


DS_DS_T <- DS_DS %>% group_by(Dress_ID) %>% gather(key = "DateofSale", value =
"Sales",

names(DS_DS)[2:length(names(DS_DS))] ) %>%
                ungroup() %>%
                arrange(Dress_ID)
DS_DS_T <- DS_DS_T %>% mutate(max3rdpart = str_split(DS_DS_T$DateofSale,"-")) %>%
        separate(max3rdpart, c("A", "B","C","D")) %>% dplyr::select(-c(A))
DS_DS_T <- DS_DS_T %>% group_by(Dress_ID) %>%
        mutate(lagD = lag(D)) %>%
        ungroup()
DS_DS_T <- DS_DS_T %>%
        mutate( DateofSale = ifelse((as.numeric(D) - as.numeric(lagD) <=0) &
!is.na(lagD),paste(B,D,C,sep='-'),DateofSale))
DS_DS_T <- DS_DS_T %>% dplyr::select(1:3)  %>% mutate(DateofSale =
as.Date(parse_date_time(DateofSale,orders =c("y-m-d"))))
```

```
ggplot(data = DS_DS_T , aes(x = DateofSale, y = Sales)) +
        geom_line(aes(x = DateofSale, y = Sales,color = factor(Dress_ID))) +
        #geom_line(col = "hotpink") +
        #ylim(0, 8000) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none")

library(forecast)
# DS_DS_T2 <- DS_DS_T %>%
#         group_by(Dress_ID)  %>% mutate(arima1 = nest(auto.arima(as.ts(Sales)))) %>%
#         ungroup()
# arima1


#out of the 23 date records for every dress id, lets consider 85 percent as training
and rest as testing records
#inTrain2 <- createDataPartition(DS_DS_T$Sales,p = 0.7, list = FALSE)
train_index <-  1: floor(0.85 * 23 ) # nrow(input) is changed to n as we have already
calc for each SKU
test_index <- setdiff(1:23, train_index) # nrow(input) is changed to n as we have
already calc for each SKU

DS_DS_T  <- DS_DS_T %>% group_by(Dress_ID) %>% mutate(SNO = 1:n(), TRAINTESTFIL =
ifelse(SNO <= train_index,"Train","Test")) %>% ungroup()


train <- dplyr::filter(DS_DS_T,TRAINTESTFIL =="Train")
test <- dplyr::filter(DS_DS_T,!TRAINTESTFIL =="Train")

summary(train$Sales)
# Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
# 0.0     5.0    73.0   275.3   275.0  7479.0

library(fpp2)

train_all <- train %>%
        group_by(DateofSale,SNO,TRAINTESTFIL) %>%
        summarise(Sales = sum(Sales,na.rm = T))  %>%
        ungroup() %>% dplyr::select(1,2,4,3) %>%
         padr::pad(interval = "day") %>%
        tidyr::fill(SNO,TRAINTESTFIL) %>%
        mutate(Sales = if_else(is.na(Sales), 0, Sales))

test_all <- test %>%
        group_by(DateofSale,SNO,TRAINTESTFIL) %>%
        summarise(Sales = sum(Sales,na.rm = T))  %>%
        ungroup() %>% dplyr::select(1,2,4,3) %>%
        padr::pad(interval = "day") %>%
        tidyr::fill(SNO,TRAINTESTFIL) %>%
        mutate(Sales = if_else(is.na(Sales), 0, Sales))

#Declaring train as a time-series data



Y <- ts(train_all[,3],start = c(2013, as.numeric(format(train$DateofSale[1], "%j"))),
#starting date
                end = c(2013,as.numeric(format(train$DateofSale[9025], "%j"))), # end
date from last available dt
                 frequency = 365)

#########preliminary analysis#####

#Time Plot
autoplot(Y) +
        ggtitle("Time Plot of Sales") +
        ylab("Total Sales")+ theme_bw()

ggplot(data = train_all , aes(x = DateofSale, y = Sales)) +
```

```
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none")

#There is a slight trend. Investigate transformation
# Take the first difference to remove the trend

DY <- diff(Y)

options(scipen=7) #Do not want scientific notation on plot axis
#Time Plot of the differenced data
autoplot(DY) +
        ggtitle("Time Plot of Change in Sales") +
        ylab("Total Sales") + theme_bw()

#Series appears trend-stationary.

# Next Need to investigate seasonality but since we only have just two months we can
jump to algos, commenting the following hence
# ggseasonplot(DY) +
#         ggtitle("Seasonal Plot: Change in Daily Total Sales") +
#         ylab("Total Sales")+
#         theme_bw()

#ggsubseriesplot(DY)
# Error in ggsubseriesplot(DY) :
#         Each season requires at least 2 observations. Your series length may be too
short for this graphic.

#######
#Use a benchmark method to forecast
# Lets use ETS method as our bench mark
#######

fit_ets <- ets(Y)
print(summary(fit_ets))
# ETS(A,N,N)
#
# Call:
#         ets(y = Y)
#
# Smoothing parameters:
#         alpha = 0.0001
#
# Initial states:
#         l = 67152.9864
#
# sigma:  74776.39
#
# AIC     AICc      BIC
# 967.9949 968.7222 972.8277
#
# Training set error measures:
#         ME     RMSE      MAE  MPE MAPE MASE       ACF1
# Training set -7.099539 72727.33 67310.57 -Inf  Inf  NaN -0.8248643
# ME     RMSE      MAE  MPE MAPE MASE       ACF1
# Training set -7.099539 72727.33 67310.57 -Inf  Inf  NaN -0.8248643

checkresiduals(fit_ets) # Residual SD :74776.39 which is high

# Ljung-Box test
#
# data:  Residuals from ETS(A,N,N)
# Q* = 205.63, df = 5, p-value < 2.2e-16
#
# Model df: 2.   Total lags used: 7


#######
#Fit arima model - which needs to have stationary data
```

```
#######
fit_arima <- auto.arima(Y, d=1,stepwise = F,approximation = F,trace = T) #residual
sqrt(788991890) ~ 28089
# ARIMA(0,1,0)                                    : 957.8362
# ARIMA(0,1,0)           with drift              : 960.0804
# ARIMA(0,1,1)                                    : Inf
# ARIMA(0,1,1)           with drift              : Inf
# ARIMA(0,1,2)                                    : Inf
# ARIMA(0,1,2)           with drift              : Inf
# ARIMA(0,1,3)                                    : Inf
# ARIMA(0,1,3)           with drift              : Inf
# ARIMA(0,1,4)                                    : Inf
# ARIMA(0,1,4)           with drift              : Inf
# ARIMA(0,1,5)                                    : Inf
# ARIMA(0,1,5)           with drift              : Inf
# ARIMA(1,1,0)                                    : 859.8935
# ARIMA(1,1,0)           with drift              : 862.2476
# ARIMA(1,1,1)                                    : 861.1228
# ARIMA(1,1,1)           with drift              : 863.5355
# ARIMA(1,1,2)                                    : 854.3237
# ARIMA(1,1,2)           with drift              : 856.8826
# ARIMA(1,1,3)                                    : Inf
# ARIMA(1,1,3)           with drift              : Inf
# ARIMA(1,1,4)                                    : Inf
# ARIMA(1,1,4)           with drift              : Inf
# ARIMA(2,1,0)                                    : 862.274
# ARIMA(2,1,0)           with drift              : 864.7811
# ARIMA(2,1,1)                                    : 861.284
# ARIMA(2,1,1)           with drift              : Inf
# ARIMA(2,1,2)                                    : 857.0318
# ARIMA(2,1,2)           with drift              : 859.778
# ARIMA(2,1,3)                                    : Inf
# ARIMA(2,1,3)           with drift              : Inf
# ARIMA(3,1,0)                                    : 849.3568
# ARIMA(3,1,0)           with drift              : 851.9142
# ARIMA(3,1,1)                                    : 852.0663
# ARIMA(3,1,1)           with drift              : 854.8084
# ARIMA(3,1,2)                                    : Inf
# ARIMA(3,1,2)           with drift              : Inf
# ARIMA(4,1,0)                                    : 852.0664
# ARIMA(4,1,0)           with drift              : 854.809
# ARIMA(4,1,1)                                    : 854.7237
# ARIMA(4,1,1)           with drift              : 857.7211
# ARIMA(5,1,0)                                    : 854.8037
# ARIMA(5,1,0)           with drift              : 857.7234
#
#
#
# Best model: ARIMA(3,1,0)

print(summary(fit_arima))
# Series: Y
# ARIMA(3,1,0)
#
# Coefficients:
#          ar1      ar2      ar3
# -0.9865  -0.5833  -0.5749
# s.e.   0.1321   0.1826   0.1267
#
# sigma^2 estimated as 788991890:  log likelihood=-420.03
# AIC=848.07   AICc=849.36   BIC=854.4
#
# Training set error measures:
#          ME     RMSE     MAE  MPE MAPE MASE        ACF1
# Training set -2162.108 26527.26 14634.1 -Inf  Inf  NaN -0.02084238
# ME     RMSE     MAE  MPE MAPE MASE        ACF1
# Training set -2162.108 26527.26 14634.1 -Inf  Inf  NaN -0.02084238

checkresiduals(fit_arima)
```

```
# data:  Residuals from ARIMA(3,1,0)
# Q* = 5.8982, df = 4, p-value = 0.2069
#
# Model df: 3.   Total lags used: 7

# Residual dropped from ~74K to ~28K which was good and the ACF plot falls between
# the Confidence interval which is also good and the distribution of the residuals
# is also normal around 0 which is good. So lets use arima as our model to forecast

#########Forecast with ARIMA model####
fcst <- forecast(fit_arima, h =14) # 14 more days to forecast from the last of
training date to get 3 more additional alternate days

autoplot(fcst)+
        theme_bw()
#Looks realistic as the previous few days sales was low, its forecasting low sales
like those first few days of oct
autoplot(fcst, include = 21) +
        theme_bw()  # last 21 days or 3 weeks before the prediction period (just
zooming the forecast plot)

check_pred <- bind_rows(train_all,test_all) %>%
                        padr::pad(interval = "day", end_val = as.Date("2013-10-18"))
check_pred <- check_pred %>%
        dplyr::mutate(Sales = if_else(is.na(Sales) & check_pred$DateofSale <= "2013-
10-12",
                                      0, Sales))


check_pred2 <- data.frame(DateofSale = check_pred$DateofSale,
                          Sales = c(fcst$fitted[1:37],fcst$mean[1:14]))
bandds <- data.frame(DateofSale = filter(check_pred, DateofSale >="2013-10-
05")$DateofSale, LOWER95 = as.numeric(fcst$lower[,2]),
           UPPER95 = as.numeric(fcst$upper[,2]))
bandds2 <- data.frame(DateofSale = filter(check_pred, DateofSale >="2013-10-
05")$DateofSale,
                     LOWER80 = as.numeric(fcst$lower[,1]),
                     UPPER80 = as.numeric(fcst$upper[,1]))

ggplot(data = check_pred , aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        geom_ribbon(data = bandds,aes(ymin = LOWER95, ymax = UPPER95), fill = "pink",
alpha = 0.4)+
        geom_ribbon(data = bandds2,aes(ymin = LOWER80, ymax = UPPER80), fill = "pink",
alpha = 0.6)+
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() +
        ylab("Total Sales")+
        theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
geom_line(data = check_pred2, aes(y = Sales),color = "red")+
        labs(caption = "red line - prediction median, black line - original data; dark
and light red bands are 80% and 95 % CI of forecasts")

ggplot(data = check_pred %>% filter(Sales >0) , aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        geom_ribbon(data = bandds,aes(ymin = max(LOWER95,0), ymax = UPPER95), fill =
"pink", alpha = 0.4)+
        geom_ribbon(data = bandds2,aes(ymin = max(LOWER80,0), ymax = UPPER80), fill =
"pink", alpha = 0.6)+
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        ylab("Total Sales")+
        geom_line(data = check_pred2[!as.numeric(row.names(check_pred2)) %% 2 ==0,],
                  aes( y = Sales),color = "red")+
        labs(caption = "red line - prediction median, black line - original data; dark
and light red bands are 80% and 95 % CI of forecasts")
```

```
# Based on the plot just looking at the alternate days (data provided to us),
# looks like the model considering the Train:Test split as 85% doesn't do well as per
the
# Test data as there is a huge fluctuation after the split i.e. after
# max(train$DateofSale) ~ "2013-10-04", there seems to be some promotion or some sale
# or discount or could be some event (public holiday etc.) depending on what location
this
# shop is in. So instead of considering train:test split as 85%, had we considered it
# as 90% or 70%, we might have been able to capture one or both of the two peaks of
the test
# i.e on the 6th oct, lets check once more with 70% train test split
#Repeating the same steps as above

#out of the 23 date records for every dress id, lets consider 70 percent as training
and rest as testing records
train_index <-  1: floor(0.7 * 23 ) # nrow(input) is changed to n as we have already
calc for each SKU
test_index <- setdiff(1:23, train_index) # nrow(input) is changed to n as we have
already calc for each SKU

DS_DS_T  <- DS_DS_T %>% group_by(Dress_ID) %>% mutate(SNO = 1:n(), TRAINTESTFIL =
ifelse(SNO <= train_index,"Train","Test")) %>% ungroup()


train <- dplyr::filter(DS_DS_T,TRAINTESTFIL =="Train")
test <- dplyr::filter(DS_DS_T,!TRAINTESTFIL =="Train")

summary(train$Sales)
# Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
# 0.0    16.0   101.0   305.2   318.0  7479.0


train_all <- train %>%
        group_by(DateofSale,SNO,TRAINTESTFIL) %>%
        summarise(Sales = sum(Sales,na.rm = T))  %>%
        ungroup() %>% dplyr::select(1,2,4,3) %>%
        padr::pad(interval = "day") %>%
        tidyr::fill(SNO,TRAINTESTFIL) %>%
        mutate(Sales = if_else(is.na(Sales), 0, Sales))

test_all <- test %>%
        group_by(DateofSale,SNO,TRAINTESTFIL) %>%
        summarise(Sales = sum(Sales,na.rm = T))  %>%
        ungroup() %>% dplyr::select(1,2,4,3) %>%
        padr::pad(interval = "day") %>%
        tidyr::fill(SNO,TRAINTESTFIL) %>%
        mutate(Sales = if_else(is.na(Sales), 0, Sales))

#Declaring train as a time-series data

Y <- ts(train_all[,3],start = c(2013, as.numeric(format(train$DateofSale[1], "%j"))),
#starting date
        end = c(2013,as.numeric(format(train$DateofSale[length(train$DateofSale)],
"%j"))), # end date from last available dt
        frequency = 365)

##########preliminary analysis#####

#Time Plot
autoplot(Y) +
        ggtitle("Time Plot of Sales") +
        ylab("Total Sales")+ theme_bw()

ggplot(data = train_all , aes(x = DateofSale, y = Sales)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none")

#There is a slight trend. Investigate transformation
# Take the first difference to remove the trend
```

```
DY <- diff(Y)

options(scipen=7) #Do not want scientific notation on plot axis
#Time Plot of the differenced data
autoplot(DY) +
        ggtitle("Time Plot of Change in Sales") +
        ylab("Total Sales") + theme_bw()

#Series appears trend-stationary.

# Next Need to investigate seasonality but since we only have just two months we can
jump to algos, commenting the following hence
# ggseasonplot(DY) +
#         ggtitle("Seasonal Plot: Change in Daily Total Sales") +
#         ylab("Total Sales")+
#         theme_bw()

#ggsubseriesplot(DY)
# Error in ggsubseriesplot(DY) :
#         Each season requires at least 2 observations. Your series length may be too
short for this graphic.

#######
#Use a benchmark method to forecast
# Lets use ETS method as our bench mark
#######

fit_ets <- ets(Y)
print(summary(fit_ets))
# ETS(A,N,N)
#
# Call:
#         ets(y = Y)
#
# Smoothing parameters:
#         alpha = 0.0001
#
# Initial states:
#         l = 74836.1411
#
# sigma:  78762.66
#
# AIC     AICc      BIC
# 809.3862 810.2751 813.6882
#
# Training set error measures:
#         ME     RMSE      MAE  MPE MAPE MASE       ACF1
# Training set -3.338684 76179.57 72414.18 -Inf  Inf  NaN -0.8806181
# ME     RMSE      MAE  MPE MAPE MASE       ACF1
# Training set -3.338684 76179.57 72414.18 -Inf  Inf  NaN -0.8806181

checkresiduals(fit_ets) # Residual SD :78762.66 which is high

# Ljung-Box test
#
# data:  Residuals from ETS(A,N,N)
# Q* = 160.67, df = 4, p-value < 2.2e-16
#
# Model df: 2.   Total lags used: 6


#######
#Fit arima model - which needs to have stationary data

#######
fit_arima <- auto.arima(Y, d=1,stepwise = F,approximation = F,trace = T) #residual
sqrt(810727537) ~ 28473.28
# ARIMA(0,1,0)                               : 801.8936
# ARIMA(0,1,0)             with drift        : 804.1821
# ARIMA(0,1,1)                               : Inf
# ARIMA(0,1,1)             with drift        : Inf
```

```
# ARIMA(0,1,2)                                : Inf
# ARIMA(0,1,2)             with drift         : Inf
# ARIMA(0,1,3)                                : Inf
# ARIMA(0,1,3)             with drift         : Inf
# ARIMA(0,1,4)                                : Inf
# ARIMA(0,1,4)             with drift         : Inf
# ARIMA(0,1,5)                                : Inf
# ARIMA(0,1,5)             with drift         : Inf
# ARIMA(1,1,0)                    : Inf
# ARIMA(1,1,0) with drift         : Inf
# ARIMA(1,1,1)                                : 709.2702
# ARIMA(1,1,1)             with drift         : 711.451
# ARIMA(1,1,2)                    : Inf
# ARIMA(1,1,2) with drift         : Inf
# ARIMA(1,1,3)                                : Inf
# ARIMA(1,1,3)             with drift         : Inf
# ARIMA(1,1,4)                    : Inf
# ARIMA(1,1,4)             with drift         : Inf
# ARIMA(2,1,0)                    : Inf
# ARIMA(2,1,0) with drift         : Inf
# ARIMA(2,1,1)                                : 709.8638
# ARIMA(2,1,1)             with drift         : Inf
# ARIMA(2,1,2)                                : Inf
# ARIMA(2,1,2)             with drift         : Inf
# ARIMA(2,1,3)                                : Inf
# ARIMA(2,1,3)             with drift         : Inf
# ARIMA(3,1,0)                    : Inf
# ARIMA(3,1,0) with drift         : Inf
# ARIMA(3,1,1)                    : Inf
# ARIMA(3,1,1) with drift         : Inf
# ARIMA(3,1,2)                    : Inf
# ARIMA(3,1,2) with drift         : Inf
# ARIMA(4,1,0)                    : Inf
# ARIMA(4,1,0) with drift         : Inf
# ARIMA(4,1,1)                                : Inf
# ARIMA(4,1,1)             with drift         : Inf
# ARIMA(5,1,0)                    : Inf
# ARIMA(5,1,0) with drift         : Inf
#
#
#
# Best model: ARIMA(1,1,1)

print(summary(fit_arima))
# Series: Y
# ARIMA(1,1,1)
#
# Coefficients:
#         ar1      ma1
# -0.9644  -0.5525
# s.e.   0.0381   0.1924
#
# sigma^2 estimated as 810727537:  log likelihood=-351.17
# AIC=708.35   AICc=709.27   BIC=712.55
#
# Training set error measures:
#         ME     RMSE      MAE MPE MAPE MASE       ACF1
# Training set 3013.953 27060.49 14414.72 NaN  Inf  NaN 0.09657674
# ME     RMSE      MAE MPE MAPE MASE       ACF1
# Training set 3013.953 27060.49 14414.72 NaN  Inf  NaN 0.09657674

checkresiduals(fit_arima)

# Ljung-Box test
#
# data:  Residuals from ARIMA(1,1,1)
# Q* = 3.471, df = 4, p-value = 0.4823
#
# Model df: 2.   Total lags used: 6

# Residual dropped from ~78K to ~28K which was good and the ACF plot falls between
# the Confidence interval which is also good and the distribution of the residuals
```

```
# is also normal around 0 which is good. So lets use arima as our model to forecast

########Forecast with ARIMA model####
fcst <- forecast(fit_arima, h =20) # 20 more days to forecast from the last of
training date to get 3 more additional alternate days

autoplot(fcst)+
        theme_bw()
#Looks realistic as the previous few days sales was low, its forecasting low sales
like those first few days of oct
autoplot(fcst, include = 21) +
        theme_bw()  # last 21 days or 3 weeks before the prediction period (just
zooming the forecast plot)

check_pred <- bind_rows(train_all,test_all) %>%
        padr::pad(interval = "day", end_val = as.Date("2013-10-18"))
check_pred <- check_pred %>%
        mutate(Sales = if_else(is.na(Sales) & check_pred$DateofSale <= "2013-10-12",
0, Sales))


check_pred2 <- data.frame(DateofSale = check_pred$DateofSale,
                          Sales = c(fcst$fitted[1:31],fcst$mean[1:20]))

bandds <- data.frame(DateofSale = filter(check_pred, DateofSale >="2013-09-
29")$DateofSale, LOWER95 = as.numeric(fcst$lower[,2]),
                     UPPER95 = as.numeric(fcst$upper[,2]))
bandds2 <- data.frame(DateofSale = filter(check_pred, DateofSale >="2013-09-
29")$DateofSale,
                      LOWER80 = as.numeric(fcst$lower[,1]),
                      UPPER80 = as.numeric(fcst$upper[,1]))

ggplot(data = check_pred , aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        geom_ribbon(data = bandds,aes(ymin = LOWER95, ymax = UPPER95), fill = "pink",
alpha = 0.4)+
        geom_ribbon(data = bandds2,aes(ymin = LOWER80, ymax = UPPER80), fill = "pink",
alpha = 0.6)+
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() +
        ylab("Total Sales")+
        theme(legend.title = element_blank(),
              axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        geom_line(data = check_pred2, aes(y = Sales),color = "red")+
        labs(caption = "red line - prediction median, black line - original data; dark
and light red bands are 80% and 95 % CI of forecasts")

ggplot(data = check_pred %>% filter(Sales >0) , aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        geom_ribbon(data = bandds,aes(ymin = max(LOWER95,0), ymax = UPPER95), fill =
"pink", alpha = 0.4)+
        geom_ribbon(data = bandds2,aes(ymin = max(LOWER80,0), ymax = UPPER80), fill =
"pink", alpha = 0.6)+
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        ylab("Total Sales")+
        geom_line(data = check_pred2[!as.numeric(row.names(check_pred2)) %% 2 ==0,],
                  aes( y = Sales),color = "red")+
        labs(caption = "red line - prediction median, black line - original data; dark
and light red bands are 80% and 95 % CI of forecasts")



#This looks much better than previous forecast with 85:15 train test split.
# Still we see that had we known a prior event like 27 Sep date (some public holiday
etc)
# we could have predicted the drop accurately for 26th Sep instead of 28th Sep
# In this split we are over-predicting 30 Sep,2Oct and 4th Oct. whereas in the
# previous split we were more underpredicting the important sales day
```

```
# So based on the biz decision if under-predicting or over-predicting is more
# important to the company, we could decide which split to use

# P.S for this Forecasting exercise sought help based on a Youtube video by Adam Check
# His explanation was fantastic for forecasting in R
# https://www.youtube.com/watch?v=dBNy_A6Zpcc

library(prophet)
library(purrr)

####Prophet Check -start
#out of the 23 date records for every dress id, lets consider 70 percent as training
and rest as testing records
train_index <-  1: floor(0.7 * 23 ) # nrow(input) is changed to n as we have already
calc for each SKU
test_index <- setdiff(1:23, train_index) # nrow(input) is changed to n as we have
already calc for each SKU

DS_DS_T  <- DS_DS_T %>% group_by(Dress_ID) %>% mutate(SNO = 1:n(), TRAINTESTFIL =
ifelse(SNO <= train_index,"Train","Test")) %>% ungroup()


train <- dplyr::filter(DS_DS_T,TRAINTESTFIL =="Train")
test <- dplyr::filter(DS_DS_T,!TRAINTESTFIL =="Train")

summary(train$Sales)
# Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
# 0.0    16.0   101.0   305.2   318.0  7479.0


train_all <- train %>%
        group_by(DateofSale,SNO,TRAINTESTFIL) %>%
        summarise(Sales = sum(Sales,na.rm = T))  %>%
        ungroup() %>% dplyr::select(1,2,4,3) %>%
        padr::pad(interval = "day") %>%
        tidyr::fill(SNO,TRAINTESTFIL) %>%
        mutate(Sales = if_else(is.na(Sales), 0, Sales))

test_all <- test %>%
        group_by(DateofSale,SNO,TRAINTESTFIL) %>%
        summarise(Sales = sum(Sales,na.rm = T))  %>%
        ungroup() %>% dplyr::select(1,2,4,3) %>%
        padr::pad(interval = "day") %>%
        tidyr::fill(SNO,TRAINTESTFIL) %>%
        mutate(Sales = if_else(is.na(Sales), 0, Sales))

d1_all <- train_all %>% rename(ds = DateofSale, y = Sales) %>% # for prophet to work
Must have columns ds (date type) and y, the time series.
        nest() %>%
        mutate(m = map(data, prophet)) %>%
        mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
        mutate(forecast = map2(m, future, predict))

d_all <- d1_all %>%
        unnest(forecast) %>%
        dplyr::select(ds, yhat, yhat_lower,yhat_upper) # in case if we wish to see all
variables we can comment this


indi_ds_all <- bind_rows(train_all,test_all) %>% mutate(ds = DateofSale)
indi_ds2_all <- left_join(d_all,indi_ds_all) %>%
        mutate(SNO = 1:n(),
               DateofSale = as.Date(as.character(ds)))


ggplot(data = indi_ds2_all %>%
               filter(Sales >0) ,
       aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
```

```
        theme(legend.position = "none") +
        ylab("Total Sales")+
        geom_line(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                                              (indi_ds2_all$SNO %% 2 ==0 &
is.na(indi_ds2_all$Sales)),]
                  ,
                  aes(y = yhat), color = "red") +
        geom_ribbon(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                                                (indi_ds2_all$SNO %% 2 ==0 &
is.na(indi_ds2_all$Sales)) ,],
                    aes(ymin=yhat_lower, ymax=yhat_upper),
                    color = "red", alpha =0.3)
#now using promotion days to factor in some days to improve the prediction
promotion_days <- data.table::data.table(
        holiday = 'sale discount days',
        ds=as.Date(c('2013-09-28',
                     '2013-10-06',
                     '2013-10-12'
        )),
        lower_window = 0, # looks back at 0 days before the last day to shop and get
it before sale
        upper_window = 0
)
d1_all <- train_all %>% rename(ds = DateofSale, y = Sales) %>% # for prophet to work
Must have columns ds (date type) and y, the time series.
        nest() %>%
        mutate(m = map(data, ~prophet(.,holidays = promotion_days))) %>%
        mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
        mutate(forecast = map2(m, future, predict))

d_all <- d1_all %>%
        unnest(forecast) %>%
        dplyr::select(ds, yhat, yhat_lower,yhat_upper) # in case if we wish to see all
variables we can comment this


indi_ds_all <- bind_rows(train_all,test_all) %>% mutate(ds = DateofSale)
indi_ds2_all <- left_join(d_all,indi_ds_all) %>%
        mutate(SNO = 1:n(),
               DateofSale = as.Date(as.character(ds)))


ggplot(data = indi_ds2_all %>%
               filter(Sales >0) ,
       aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        geom_line(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                                              (indi_ds2_all$SNO %% 2 ==0 &
is.na(indi_ds2_all$Sales)),]
                  ,
                  aes(y = yhat), color = "red") +
        ylab("Total Sales")+
        geom_ribbon(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                                                (indi_ds2_all$SNO %% 2 ==0 &
is.na(indi_ds2_all$Sales)) ,],
                    aes(ymin=yhat_lower, ymax=yhat_upper),
                    color = "red", alpha =0.3)

###now also using day seasonality in addition to promotional campaign days
d1_all <- train_all %>% rename(ds = DateofSale, y = Sales) %>% # for prophet to work
Must have columns ds (date type) and y, the time series.
        nest() %>%
        mutate(m = map(data, ~prophet(.,holidays =
promotion_days,daily.seasonality=TRUE))) %>%
        mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
        mutate(forecast = map2(m, future, predict))

d_all <- d1_all %>%
```

```
        unnest(forecast) %>%
        dplyr::select(ds, yhat, yhat_lower,yhat_upper) # in case if we wish to see all
variables we can comment this


indi_ds_all <- bind_rows(train_all,test_all) %>% mutate(ds = DateofSale)
indi_ds2_all <- left_join(d_all,indi_ds_all) %>%
        mutate(SNO = 1:n(),
               DateofSale = as.Date(as.character(ds)))


ggplot(data = indi_ds2_all %>%
               filter(Sales >0) ,
       aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        geom_line(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                                              (indi_ds2_all$SNO %% 2 ==0 &
is.na(indi_ds2_all$Sales)),]
                  ,
                  aes(y = yhat), color = "red") +
        ylab("Total Sales")+
        geom_ribbon(data = indi_ds2_all[!(indi_ds2_all$Sales ==0) |
                                                (indi_ds2_all$SNO %% 2 ==0 &
is.na(indi_ds2_all$Sales)) ,],
                    aes(ymin=0, ymax=yhat_upper),# yhat_lower has been changed to 0
because we are going with the assumption there are no returns allowed
                    color = "red", alpha =0.3)

gc() # to clear the memory dump

#####Check -end
###
### We still havent solved the objective which was based on individual Dress ID
##############################################################################

# For that lets explore Facebook's Prophet package to predict the Sales
# (Technically we need to use the algo which gives the least residuals or based on Biz
decision
# for different purposes different algos; Like if say the company wanted to predict
well the
# holiday/sale days well even accepting over-fitting but not under-fitting on those
days
# then our model choice would have been prophet here as that only captures the
promotion
# days well and also captures other days but the range of CI is wide. We would have
liked that
# the model even predict the non-sale days better, hence if the biz wanted to predict
only
# the non-sale days a lot better then probably we would have gone with Arima model
above
# but for didactic purposes lets assume prophet was the best algo better than arima
above overall)


#out of the 23 date records for every dress id, lets consider 70 percent as training
and rest as testing records
train_index <-  1: floor(0.7 * 23 ) # nrow(input) is changed to n as we have already
calc for each SKU
test_index <- setdiff(1:23, train_index) # nrow(input) is changed to n as we have
already calc for each SKU

DS_DS_T  <- DS_DS_T %>% group_by(Dress_ID) %>% mutate(SNO = 1:n(), TRAINTESTFIL =
ifelse(SNO <= train_index,"Train","Test")) %>% ungroup()


train <- dplyr::filter(DS_DS_T,TRAINTESTFIL =="Train")
test <- dplyr::filter(DS_DS_T,!TRAINTESTFIL =="Train")
```

```
d1 <- train %>% rename(ds = DateofSale, y = Sales) %>% # for prophet to work Must have
columns ds (date type) and y, the time series.
        nest(-Dress_ID) %>%
        mutate(m = map(data, prophet)) %>%
        mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
        mutate(forecast = map2(m, future, predict))

d <- d1 %>%
        unnest(forecast) %>%
        dplyr::select(ds, Dress_ID, yhat, yhat_lower,yhat_upper) # in case if we wish
to see all variables we can comment this
# d1 %>%
#       +     unnest(forecast) %>% names()
# [1] "Dress_ID"                   "data"                        "m"
# [4] "future"                     "ds"                          "trend"
# [7] "additive_terms"             "additive_terms_lower"        "additive_terms_upper"
# [10] "weekly"                     "weekly_lower"                "weekly_upper"
# [13] "multiplicative_terms"       "multiplicative_terms_lower"
"multiplicative_terms_upper"
# [16] "yhat_lower"                 "yhat_upper"                  "trend_lower"
# [19] "trend_upper"                "yhat"


indi_ds <- bind_rows(train,test) %>% mutate(ds = DateofSale)
indi_ds2 <- left_join(d,indi_ds) %>%
        group_by(Dress_ID) %>%
        mutate(SNO = 1:n(),
               DateofSale = as.Date(as.character(ds)))
gc()
#A call of gc causes a garbage collection to take place

#Checking randomly for two ids via plot
ggplot(data = indi_ds2 %>% filter(Sales >0, Dress_ID =="444282011") ,
       aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        geom_line(data = indi_ds2[(indi_ds2$SNO <=16 |
                                           (as.numeric(format(indi_ds2$DateofSale,
format = "%d")) %% 2 ==0 &
                                                    indi_ds2$SNO >=17)) &
                                          indi_ds2$Dress_ID =="444282011",]
                  ,
                  aes(y = yhat), color = "red") +
        geom_ribbon(data = indi_ds2[(indi_ds2$SNO <=16 |
                                             (as.numeric(format(indi_ds2$DateofSale,
format = "%d")) %% 2 ==0 &
                                                      indi_ds2$SNO >=17)) &
                                            indi_ds2$Dress_ID =="444282011",],
                    aes(ymin=yhat_lower, ymax=yhat_upper),
                    color = "red", alpha =0.3)


#Checking randomly for two ids via plot
ggplot(data = indi_ds2 %>% filter(Sales >0, Dress_ID =="549401113") ,
       aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        geom_line(data = indi_ds2[(indi_ds2$SNO <=16 |
                                           (as.numeric(format(indi_ds2$DateofSale,
format = "%d")) %% 2 ==0 &
                                                    indi_ds2$SNO >=17)) &
                                          indi_ds2$Dress_ID =="549401113",]
                  ,
                  aes(y = yhat), color = "red") +
        geom_ribbon(data = indi_ds2[(indi_ds2$SNO <=16 |
```

```
                                             (as.numeric(format(indi_ds2$DateofSale,
format = "%d")) %% 2 ==0 &
                                                      indi_ds2$SNO >=17)) &
                                            indi_ds2$Dress_ID =="549401113",],
                    aes(ymin=yhat_lower, ymax=yhat_upper),
                    color = "red", alpha =0.3)

# now lets explore as though the days where total sales was high as sale
days/promotion days
# even b4 forecasting we will feed these dates into the prophet algo
# this will help better estimate the peak threshold days

d2 <- train %>%
        rename(ds = DateofSale, y = Sales) %>% # for prophet to work Must have columns
ds (date type) and y, the time series.
        dplyr::select(-c(SNO,TRAINTESTFIL)) %>%
        nest(-Dress_ID) %>%
        mutate(m = map(data, ~prophet(.,holidays = promotion_days))) %>%
        mutate(future = map(m, make_future_dataframe, period = 20, freq = 'day')) %>%
        mutate(forecast = map2(m, future, predict))

d_pro <- d2 %>%
        unnest(forecast) %>%
        dplyr::select(ds, Dress_ID, yhat, yhat_lower,yhat_upper) # in case if we wish
to see all variables we can comment this

indi_ds_pro <- bind_rows(train,test) %>% mutate(ds = DateofSale)
indi_ds2_pro <- left_join(d_pro,indi_ds_pro) %>%
        group_by(Dress_ID) %>%
        mutate(SNO = 1:n(),
               DateofSale = as.Date(as.character(ds)))

#Checking randomly for two ids via plot
ggplot(data = indi_ds2_pro %>% filter(Sales >0, Dress_ID =="444282011") ,
       aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        geom_line(data = indi_ds2_pro[(indi_ds2_pro$SNO <=16 |
                                           (as.numeric(format(indi_ds2_pro$DateofSale,
                                                              format = "%d")) %% 2 ==0
&
                                                     indi_ds2_pro$SNO >=17)) &
                                          indi_ds2_pro$Dress_ID =="444282011",]
                  ,
                  aes(y = yhat), color = "red") +
        geom_ribbon(data = indi_ds2_pro[(indi_ds2_pro$SNO <=16 |

(as.numeric(format(indi_ds2_pro$DateofSale,
                                                                format = "%d")) %% 2
==0 &
                                                      indi_ds2_pro$SNO >=17)) &
                                            indi_ds2_pro$Dress_ID =="444282011",],
                    aes(ymin=yhat_lower, ymax=yhat_upper),
                    color = "red", alpha =0.3)


#Checking randomly for two ids via plot
ggplot(data = indi_ds2_pro %>% filter(Sales >0, Dress_ID =="549401113") ,
       aes(x = DateofSale)) +
        geom_line(aes(x = DateofSale, y = Sales)) +
        scale_x_date(date_labels = "%d %b %y", date_breaks = "2 days") +
        theme_bw() + theme(legend.title = element_blank(),
                           axis.text.x  = element_text(angle=45, vjust=0.5)) +
        theme(legend.position = "none") +
        geom_line(data = indi_ds2_pro[(indi_ds2_pro$SNO <=16 |
                                           (as.numeric(format(indi_ds2_pro$DateofSale,
                                                              format = "%d")) %% 2 ==0
&
                                                     indi_ds2_pro$SNO >=17)) &
```

```
                                          indi_ds2_pro$Dress_ID =="549401113",]
                  ,
                  aes(y = yhat), color = "red") +
        geom_ribbon(data = indi_ds2_pro[(indi_ds2_pro$SNO <=16 |

(as.numeric(format(indi_ds2_pro$DateofSale,
                                                          format = "%d")) %% 2
==0 &
                                                     indi_ds2_pro$SNO >=17)) &
                                           indi_ds2_pro$Dress_ID =="549401113",],
                    aes(ymin=yhat_lower, ymax=yhat_upper),
                    color = "red", alpha =0.3)

#Task 3:
# To decide the pricing for various upcoming clothes, they wish to find how the
# style, season, and material affect the sales of a dress and if the style of the
dress
# is more influential than its price.

# For this task, lets consider the total sales per dress id and do an Linear
Regression/ANOVA
# ANOVA is the statistical model that you use to predict a continuous outcome on the
basis
# of one or more categorical predictor variables
# we can use lm itself because we have already created dummy variables for categorical
variables


# lets assume alpha value as 0.05 and any p-value < alpha will be
# significant factots contributing to the sales

summ_totsales_ds <- DS_DS_T %>%
        group_by(Dress_ID) %>%
        summarise(TotalSales = sum(Sales,na.rm = T)) %>%
        ungroup()
ATT_DS3 <- left_join(ATT_DS2,summ_totsales_ds)

# lets explore the response variable TotalSales distribution and see if it is normal;
# if not use the transformed variable whichever is
g1 <- ggplot(ATT_DS3, aes(TotalSales)) + geom_density(fill="blue")+theme_bw() # it is
normal because we have more than 20 samples/records but just not like our usual bell
curve
g2<- ggplot(ATT_DS3, aes(log(TotalSales))) + geom_density(fill="blue")+theme_bw() #
more closer to normal
g3 <- ggplot(ATT_DS3, aes(sqrt(TotalSales))) + geom_density(fill="blue")+theme_bw()

gridExtra::grid.arrange(g1,g2,g3, ncol = 3)
# The coefficient for linear regression is calculated based on the sample data.
# The basic assumption here is that the sample is not biased and residuals are normal.
# relation between dependent and independent are linear
# residual variance is homoscedastic equal

# we could have technically used any of the above response variable but i prefer to
use log transformed as it is close to usual normal bell curve
# first creating training and testing of data so that we can validate model later
set.seed(123)
inTrain <- createDataPartition(ATT_DS3$TotalSales,
                               p = 0.7,
                               list = FALSE)


Training = ATT_DS3[inTrain,]
Testing = ATT_DS3[-inTrain,]

Training1 <- Training %>% dplyr::select(c(2:7,32:92))
Testing1 <- Testing %>% dplyr::select(c(2:7,32:92))
#just for removing multicollinearity lets use lm
totsale_lm <- lm(data = Training1,
                 log(TotalSales) ~ Seasonspring + Seasonsummer + Seasonwinter +
                         SeasonAutumn +
                         Priceaverage + Pricehigh + Pricelow + Pricemedium +
```

```
                         Pricevery_high + Style_bohemian_dmy + Style_brief_dmy +
Style_casual_dmy +
                         Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_novelty_dmy +
                         Style_OL_dmy + Style_party_dmy + Style_sexy_dmy +
Style_vintage_dmy +
                         Style_work_dmy + Material_acrylic_dmy + Material_cashmere_dmy
+
                         Material_chiffon_dmy + Material_cotton_dmy +
Material_knitting_dmy +
                         Material_linen_dmy + Material_lycra_dmy +
Material_microfiber_dmy +
                         Material_milksilk_dmy + Material_mix_dmy + Material_modal_dmy
+
                         Material_nylon_dmy + Material_other_dmy +
Material_polyster_dmy +
                         Material_rayon_dmy + Material_silk_dmy + Material_spandex_dmy
+
                         Material_viscos_dmy)

alias(totsale_lm) # indicates SeasonAutumn,Pricevery_high, Style_work_dmy need to be
removed due to autocorrelation
# The autocorrelated variables are these mentioned below (found by using alias(mylm))

# SeasonAutumn  = 1 - Seasonspring  - Seasonsummer - Seasonwinter

# Pricevery_high  =  1 - Priceaverage - Pricehigh - Pricelow - Pricemedium

# Style_work_dmy  = 1 -Style_bohemian_dmy  -Style_brief_dmy  -Style_casual_dmy  -
Style_cute_dmy  -Style_fashion_dmy  -Style_novelty_dmy  -Style_OL_dmy  -
Style_party_dmy  -Style_sexy_dmy  -Style_vintage_dmy

vif(totsale_lm)

#Error in vif.default(mylm) : there are aliased coefficients in the model



totsale_lm <- lm(data = Training1,
                 log(TotalSales) ~ Seasonspring + Seasonsummer + Seasonwinter +
                         Priceaverage + Pricehigh + Pricelow + Pricemedium +
                         Style_bohemian_dmy + Style_brief_dmy + Style_casual_dmy +
                         Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_novelty_dmy +
                         Style_OL_dmy + Style_party_dmy + Style_sexy_dmy +
Style_vintage_dmy +
                         Material_acrylic_dmy + Material_cashmere_dmy +
                         Material_chiffon_dmy + Material_cotton_dmy +
Material_knitting_dmy +
                         Material_linen_dmy + Material_lycra_dmy +
Material_microfiber_dmy +
                         Material_milksilk_dmy + Material_mix_dmy + Material_modal_dmy
+
                         Material_nylon_dmy + Material_other_dmy +
Material_polyster_dmy +
                         Material_rayon_dmy + Material_silk_dmy + Material_spandex_dmy
+
                         Material_viscos_dmy)
alias(totsale_lm) # no warning/err

vif(totsale_lm)
# Seasonspring           Seasonsummer           Seasonwinter           Priceaverage
# 2.542901               2.700566               2.744959               9.468094
# Pricehigh               Pricelow             Pricemedium      Style_bohemian_dmy
# 2.304913               9.418700               2.952570               2.231499
# Style_brief_dmy       Style_casual_dmy         Style_cute_dmy
Style_fashion_dmy
# 2.156559               8.607728               3.181639               1.100120
# Style_flare_dmy       Style_novelty_dmy           Style_OL_dmy
Style_party_dmy
# 1.314469               1.632623               1.165004               4.428099
```

```
# Style_sexy_dmy       Style_vintage_dmy    Material_acrylic_dmy
Material_cashmere_dmy
# 5.494190             2.502165             3.121383             4.220774
# Material_chiffon_dmy     Material_cotton_dmy   Material_knitting_dmy
Material_linen_dmy
# 15.938040            79.295952             2.052178             4.099462
# Material_lycra_dmy Material_microfiber_dmy   Material_milksilk_dmy
Material_mix_dmy
# 3.080968             4.146351             5.070427             6.244475
# Material_modal_dmy      Material_nylon_dmy      Material_other_dmy
Material_polyster_dmy
# 3.101582             7.396789            68.515930            62.652227
# Material_rayon_dmy       Material_silk_dmy    Material_spandex_dmy
Material_viscos_dmy
# 9.134801            18.669394             5.190389             3.082433

#A variance inflation factor(VIF) detects multicollinearity in regression analysis.
# Multicollinearity is when there's correlation between predictors
# (i.e. independent variables) in a model;
# >10 vif values associated factors need to be removed one at a time from the model
# to remove this multicollinearity

#removing Material_cotton_dmy first (~79.29 vif)


totsale_lm <- lm(data = Training1,
                 log(TotalSales) ~ Seasonspring + Seasonsummer + Seasonwinter +
                         Priceaverage + Pricehigh + Pricelow + Pricemedium +
                         Style_bohemian_dmy + Style_brief_dmy + Style_casual_dmy +
                         Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_novelty_dmy +
                         Style_OL_dmy + Style_party_dmy + Style_sexy_dmy +
Style_vintage_dmy +
                         Material_acrylic_dmy + Material_cashmere_dmy +
                         Material_chiffon_dmy + Material_knitting_dmy +
                         Material_linen_dmy + Material_lycra_dmy +
Material_microfiber_dmy +
                         Material_milksilk_dmy + Material_mix_dmy + Material_modal_dmy
+
                         Material_nylon_dmy + Material_other_dmy +
Material_polyster_dmy +
                         Material_rayon_dmy + Material_silk_dmy + Material_spandex_dmy
+
                         Material_viscos_dmy)
alias(totsale_lm) # no warning/err

vif(totsale_lm)
# Seasonspring           Seasonsummer           Seasonwinter           Priceaverage
# 2.510237             2.650694             2.709898             9.467003
# Pricehigh              Pricelow            Pricemedium      Style_bohemian_dmy
# 2.304793             9.417106             2.952483             2.231119
# Style_brief_dmy        Style_casual_dmy          Style_cute_dmy
Style_fashion_dmy
# 2.155900             8.604117             3.181527             1.100099
# Style_flare_dmy       Style_novelty_dmy           Style_OL_dmy
Style_party_dmy
# 1.314454             1.632414             1.164906             4.428093
# Style_sexy_dmy       Style_vintage_dmy    Material_acrylic_dmy
Material_cashmere_dmy
# 5.489841             2.502033             1.051401             1.120336
# Material_chiffon_dmy   Material_knitting_dmy      Material_linen_dmy
Material_lycra_dmy
# 1.135140             1.027128             1.084950             1.070396
# Material_microfiber_dmy   Material_milksilk_dmy        Material_mix_dmy
Material_modal_dmy
# 1.079631             1.040156             1.107579             1.025369
# Material_nylon_dmy      Material_other_dmy   Material_polyster_dmy
Material_rayon_dmy
# 1.344171             1.474553             1.543209             1.132127
# Material_silk_dmy    Material_spandex_dmy     Material_viscos_dmy
# 1.226671             1.090327             1.066874
```

```
# we can see that Priceaverage and Pricelow are highly correlated
# the more VIF increases, the less reliable regression results are going to be.

#We removed the priceaverage from the first model due to high vif (even though it was
less than 10 it was still >>5 highly correlated)
# we would have got

totsale_lm <- lm(data = Training1,
                 log(TotalSales) ~ Seasonspring + Seasonsummer + Seasonwinter +
                          Pricehigh + Pricelow + Pricemedium +
                         Style_bohemian_dmy + Style_brief_dmy + Style_casual_dmy +
                         Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_novelty_dmy +
                         Style_OL_dmy + Style_party_dmy + Style_sexy_dmy +
Style_vintage_dmy +
                         Material_acrylic_dmy + Material_cashmere_dmy +
                         Material_chiffon_dmy + Material_knitting_dmy +
                         Material_linen_dmy + Material_lycra_dmy +
Material_microfiber_dmy +
                         Material_milksilk_dmy + Material_mix_dmy + Material_modal_dmy
+
                         Material_nylon_dmy + Material_other_dmy +
Material_polyster_dmy +
                         Material_rayon_dmy + Material_silk_dmy + Material_spandex_dmy
+
                         Material_viscos_dmy)
alias(totsale_lm) # no warning/err

vif(totsale_lm)
# Seasonspring            Seasonsummer            Seasonwinter              Pricehigh
# 2.502160                2.650669                2.688577                1.178129
# Pricelow             Pricemedium      Style_bohemian_dmy         Style_brief_dmy
# 1.343921                1.227101                2.207769                2.155373
# Style_casual_dmy          Style_cute_dmy       Style_fashion_dmy
Style_flare_dmy
# 8.584700                3.176479                1.099847                1.313947
# Style_novelty_dmy            Style_OL_dmy        Style_party_dmy
Style_sexy_dmy
# 1.631870                1.164323                4.072482                5.478410
# Style_vintage_dmy    Material_acrylic_dmy   Material_cashmere_dmy
Material_chiffon_dmy
# 2.500271                1.051395                1.120334                1.132027
# Material_knitting_dmy      Material_linen_dmy      Material_lycra_dmy
Material_microfiber_dmy
# 1.027093                1.084856                1.045606                1.060663
# Material_milksilk_dmy        Material_mix_dmy      Material_modal_dmy
Material_nylon_dmy
# 1.040127                1.107303                1.025292                1.342131
# Material_other_dmy   Material_polyster_dmy      Material_rayon_dmy
Material_silk_dmy
# 1.474540                1.541561                1.128407                1.226608
# Material_spandex_dmy     Material_viscos_dmy
# 1.090199                1.039359
#Style casual still has high vif. Even though its not >10 but it is >>5, we try
removing that as well

totsale_lm <- lm(data = Training1,
                 log(TotalSales) ~ Seasonspring + Seasonsummer + Seasonwinter +
                         Pricehigh + Pricelow + Pricemedium +
                         Style_bohemian_dmy + Style_brief_dmy  +
                         Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_novelty_dmy +
                         Style_OL_dmy + Style_party_dmy + Style_sexy_dmy +
Style_vintage_dmy +
                         Material_acrylic_dmy + Material_cashmere_dmy +
                         Material_chiffon_dmy + Material_knitting_dmy +
                         Material_linen_dmy + Material_lycra_dmy +
Material_microfiber_dmy +
                         Material_milksilk_dmy + Material_mix_dmy + Material_modal_dmy
+
                         Material_nylon_dmy + Material_other_dmy +
Material_polyster_dmy +
```

```
                         Material_rayon_dmy + Material_silk_dmy + Material_spandex_dmy
+
                         Material_viscos_dmy)
alias(totsale_lm) # no warning/err

vif(totsale_lm)
# Seasonspring              Seasonsummer              Seasonwinter               Pricehigh
# 2.494244                  2.643206                  2.682305                  1.175266
# Pricelow               Pricemedium      Style_bohemian_dmy          Style_brief_dmy
# 1.325716                  1.222572                  1.129676                  1.122656
# Style_cute_dmy         Style_fashion_dmy          Style_flare_dmy
Style_novelty_dmy
# 1.154457                  1.016183                  1.234244                  1.056590
# Style_OL_dmy           Style_party_dmy           Style_sexy_dmy       Style_vintage_dmy
# 1.071022                  1.364656                  1.203716                  1.096055
# Material_acrylic_dmy   Material_cashmere_dmy    Material_chiffon_dmy
Material_knitting_dmy
# 1.051382                  1.119881                  1.131105                  1.027035
# Material_linen_dmy      Material_lycra_dmy Material_microfiber_dmy
Material_milksilk_dmy
# 1.062086                  1.045479                  1.033318                  1.039839
# Material_mix_dmy      Material_modal_dmy      Material_nylon_dmy
Material_other_dmy
# 1.095435                  1.024792                  1.341953                  1.462404
# Material_polyster_dmy      Material_rayon_dmy       Material_silk_dmy
Material_spandex_dmy
# 1.539354                  1.128220                  1.226605                  1.089550
# Material_viscos_dmy
# 1.039207
summary(totsale_lm)

# Call:
#         lm(formula = log(TotalSales) ~ Seasonspring + Seasonsummer +
#                    Seasonwinter + Pricehigh + Pricelow + Pricemedium +
Style_bohemian_dmy +
#                    Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                    Style_novelty_dmy + Style_OL_dmy + Style_party_dmy +
Style_sexy_dmy +
#                    Style_vintage_dmy + Material_acrylic_dmy + Material_cashmere_dmy
+
#                    Material_chiffon_dmy + Material_knitting_dmy + Material_linen_dmy
+
#                    Material_lycra_dmy + Material_microfiber_dmy +
Material_milksilk_dmy +
#                    Material_mix_dmy + Material_modal_dmy + Material_nylon_dmy +
#                    Material_other_dmy + Material_polyster_dmy + Material_rayon_dmy +
#                    Material_silk_dmy + Material_spandex_dmy + Material_viscos_dmy,
#            data = Training1)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6612 -0.7140  0.1533  0.8534  3.6154
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)              7.76788    0.29476  26.354   <2e-16 ***
#         Seasonspring            -0.09386    0.30462  -0.308   0.7582
# Seasonsummer            -0.06357    0.28142  -0.226   0.8214
# Seasonwinter            -0.07403    0.28616  -0.259   0.7960
# Pricehigh               -0.44247    0.42253  -1.047   0.2958
# Pricelow                 0.50079    0.19605   2.554   0.0111 *
#         Pricemedium             -0.95238    0.38777  -2.456   0.0146 *
#         Style_bohemian_dmy      -0.41881    0.45753  -0.915   0.3607
# Style_brief_dmy          0.26802    0.45611   0.588   0.5572
# Style_cute_dmy          -0.51047    0.33960  -1.503   0.1338
# Style_fashion_dmy       -1.64986    1.53763  -1.073   0.2841
# Style_flare_dmy          1.68939    1.69459   0.997   0.3196
# Style_novelty_dmy        0.19035    0.59774   0.318   0.7504
# Style_OL_dmy            -0.68001    1.57857  -0.431   0.6669
# Style_party_dmy         -0.05456    0.30561  -0.179   0.8584
# Style_sexy_dmy          -0.03518    0.23846  -0.148   0.8828
```

```
# Style_vintage_dmy        0.53679    0.38586   1.391   0.1652
# Material_acrylic_dmy    -0.25884    1.10752  -0.234   0.8154
# Material_cashmere_dmy   -2.06425    0.93461  -2.209   0.0279 *
#         Material_chiffon_dmy     0.57357    0.42747   1.342   0.1806
# Material_knitting_dmy   -0.04894    1.54582  -0.032   0.9748
# Material_linen_dmy       0.22779    0.91018   0.250   0.8025
# Material_lycra_dmy      -1.98517    1.10440  -1.798   0.0732 .
# Material_microfiber_dmy  0.21157    0.89776   0.236   0.8138
# Material_milksilk_dmy    0.09030    0.78106   0.116   0.9080
# Material_mix_dmy         0.18214    0.71806   0.254   0.7999
# Material_modal_dmy      -1.68259    1.09342  -1.539   0.1248
# Material_nylon_dmy      -1.41114    0.72656  -1.942   0.0530 .
# Material_other_dmy      -0.06351    0.22760  -0.279   0.7804
# Material_polyster_dmy    0.01582    0.24481   0.065   0.9485
# Material_rayon_dmy      -0.64360    0.57862  -1.112   0.2668
# Material_silk_dmy       -0.62744    0.40819  -1.537   0.1253
# Material_spandex_dmy    -0.70575    0.79951  -0.883   0.3781
# Material_viscos_dmy     -0.14776    1.10108  -0.134   0.8933
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.523 on 318 degrees of freedom
# Multiple R-squared:  0.1506,   Adjusted R-squared:  0.06242
# F-statistic: 1.708 on 33 and 318 DF,  p-value: 0.011

Training1s <- Training1 %>% dplyr::select(TotalSales,Seasonspring , Seasonsummer ,
Seasonwinter ,
                                           Pricehigh , Pricelow , Pricemedium ,
                                           Style_bohemian_dmy , Style_brief_dmy  ,
                                           Style_cute_dmy , Style_fashion_dmy ,
Style_flare_dmy , Style_novelty_dmy ,
                                           Style_OL_dmy , Style_party_dmy ,
Style_sexy_dmy , Style_vintage_dmy ,
                                           Material_acrylic_dmy ,
Material_cashmere_dmy ,
                                           Material_chiffon_dmy ,
Material_knitting_dmy ,
                                           Material_linen_dmy , Material_lycra_dmy ,
Material_microfiber_dmy ,
                                           Material_milksilk_dmy , Material_mix_dmy ,
Material_modal_dmy ,
                                           Material_nylon_dmy , Material_other_dmy ,
Material_polyster_dmy ,
                                           Material_rayon_dmy , Material_silk_dmy ,
Material_spandex_dmy ,
                                           Material_viscos_dmy)
totsale_lm.1 <- lm(data = Training1s,
                       log(TotalSales) ~.-Material_knitting_dmy) # has the highest p-
value- pool to the error deg of freedom as its insignificant
summary(totsale_lm.1)

# Call:
#         lm(formula = log(TotalSales) ~ . - Material_knitting_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6618 -0.7135  0.1532  0.8539  3.6155
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)              7.76757    0.29414  26.408   <2e-16 ***
# Seasonspring            -0.09387    0.30414  -0.309   0.7578
# Seasonsummer            -0.06340    0.28093  -0.226   0.8216
# Seasonwinter            -0.07463    0.28508  -0.262   0.7937
# Pricehigh               -0.44247    0.42187  -1.049   0.2951
# Pricelow                 0.50027    0.19505   2.565   0.0108 *
# Pricemedium             -0.95225    0.38714  -2.460   0.0144 *
# Style_bohemian_dmy      -0.41860    0.45677  -0.916   0.3601
# Style_brief_dmy          0.26847    0.45518   0.590   0.5557
# Style_cute_dmy          -0.51026    0.33900  -1.505   0.1333
# Style_fashion_dmy       -1.64973    1.53521  -1.075   0.2834
# Style_flare_dmy          1.68920    1.69193   0.998   0.3188
```

```
# Style_novelty_dmy        0.19045    0.59680   0.319   0.7498
# Style_OL_dmy            -0.67923    1.57591  -0.431   0.6668
# Style_party_dmy         -0.05448    0.30512  -0.179   0.8584
# Style_sexy_dmy          -0.03483    0.23784  -0.146   0.8837
# Style_vintage_dmy        0.53704    0.38517   1.394   0.1642
# Material_acrylic_dmy    -0.25829    1.10564  -0.234   0.8154
# Material_cashmere_dmy   -2.06343    0.93279  -2.212   0.0277 *
# Material_chiffon_dmy     0.57412    0.42645   1.346   0.1792
# Material_linen_dmy       0.22804    0.90872   0.251   0.8020
# Material_lycra_dmy      -1.98451    1.10248  -1.800   0.0728 .
# Material_microfiber_dmy  0.21216    0.89616   0.237   0.8130
# Material_milksilk_dmy    0.09103    0.77949   0.117   0.9071
# Material_mix_dmy         0.18255    0.71682   0.255   0.7991
# Material_modal_dmy      -1.68220    1.09164  -1.541   0.1243
# Material_nylon_dmy      -1.41064    0.72525  -1.945   0.0526 .
# Material_other_dmy      -0.06308    0.22683  -0.278   0.7811
# Material_polyster_dmy    0.01640    0.24373   0.067   0.9464
# Material_rayon_dmy      -0.64339    0.57767  -1.114   0.2662
# Material_silk_dmy       -0.62676    0.40698  -1.540   0.1245
# Material_spandex_dmy    -0.70522    0.79809  -0.884   0.3776
# Material_viscos_dmy     -0.14719    1.09922  -0.134   0.8936
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.521 on 319 degrees of freedom
# Multiple R-squared:  0.1506,   Adjusted R-squared:  0.06536
# F-statistic: 1.767 on 32 and 319 DF,  p-value: 0.007933

totsale_lm.2 <- update(totsale_lm.1, ~.-Material_polyster_dmy) # next highest p-value
removed

summary(totsale_lm.2)

# Call:
#        lm(formula = log(TotalSales) ~ Seasonspring + Seasonsummer +
#                     Seasonwinter + Pricehigh + Pricelow + Pricemedium +
Style_bohemian_dmy +
#                     Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                     Style_novelty_dmy + Style_OL_dmy + Style_party_dmy +
Style_sexy_dmy +
#                     Style_vintage_dmy + Material_acrylic_dmy + Material_cashmere_dmy
+
#                     Material_chiffon_dmy + Material_linen_dmy + Material_lycra_dmy +
#                     Material_microfiber_dmy + Material_milksilk_dmy +
Material_mix_dmy +
#                     Material_modal_dmy + Material_nylon_dmy + Material_other_dmy +
#                     Material_rayon_dmy + Material_silk_dmy + Material_spandex_dmy +
#                     Material_viscos_dmy, data = Training1s)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.6611 -0.7139  0.1525  0.8491  3.6090
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)              7.77491    0.27274  28.507   <2e-16 ***
# Seasonspring            -0.09406    0.30366  -0.310   0.7570
# Seasonsummer            -0.06601    0.27781  -0.238   0.8123
# Seasonwinter            -0.07287    0.28344  -0.257   0.7973
# Pricehigh               -0.44194    0.42114  -1.049   0.2948
# Pricelow                 0.49877    0.19349   2.578   0.0104 *
# Pricemedium             -0.95175    0.38647  -2.463   0.0143 *
# Style_bohemian_dmy      -0.41741    0.45572  -0.916   0.3604
# Style_brief_dmy          0.26584    0.45280   0.587   0.5575
# Style_cute_dmy          -0.51098    0.33831  -1.510   0.1319
# Style_fashion_dmy       -1.65447    1.53121  -1.080   0.2807
# Style_flare_dmy          1.68849    1.68926   1.000   0.3183
# Style_novelty_dmy        0.19333    0.59434   0.325   0.7452
# Style_OL_dmy            -0.68884    1.56698  -0.440   0.6605
# Style_party_dmy         -0.05459    0.30464  -0.179   0.8579
# Style_sexy_dmy          -0.03381    0.23698  -0.143   0.8866
```

```
# Style_vintage_dmy        0.53576    0.38410   1.395   0.1640
# Material_acrylic_dmy    -0.26314    1.10156  -0.239   0.8113
# Material_cashmere_dmy   -2.07287    0.92075  -2.251   0.0250 *
# Material_chiffon_dmy     0.56829    0.41691   1.363   0.1738
# Material_linen_dmy       0.22157    0.90221   0.246   0.8062
# Material_lycra_dmy      -1.99235    1.09461  -1.820   0.0697 .
# Material_microfiber_dmy  0.20570    0.88962   0.231   0.8173
# Material_milksilk_dmy    0.08395    0.77115   0.109   0.9134
# Material_mix_dmy         0.17540    0.70780   0.248   0.8044
# Material_modal_dmy      -1.68618    1.08833  -1.549   0.1223
# Material_nylon_dmy      -1.41708    0.71779  -1.974   0.0492 *
# Material_other_dmy      -0.06957    0.20500  -0.339   0.7346
# Material_rayon_dmy      -0.64913    0.57045  -1.138   0.2560
# Material_silk_dmy       -0.63316    0.39510  -1.603   0.1100
# Material_spandex_dmy    -0.71145    0.79147  -0.899   0.3694
# Material_viscos_dmy     -0.15382    1.09309  -0.141   0.8882
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.518 on 320 degrees of freedom
# Multiple R-squared:  0.1506,   Adjusted R-squared:  0.06827
# F-statistic:  1.83 on 31 and 320 DF,  p-value: 0.005631


totsale_lm.3 <- update(totsale_lm.2, ~. - Material_milksilk_dmy)

summary(totsale_lm.3)
# Call:
#        lm(formula = log(TotalSales) ~ Seasonspring + Seasonsummer +
#                     Seasonwinter + Pricehigh + Pricelow + Pricemedium +
Style_bohemian_dmy +
#                     Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                     Style_novelty_dmy + Style_OL_dmy + Style_party_dmy +
Style_sexy_dmy +
#                     Style_vintage_dmy + Material_acrylic_dmy + Material_cashmere_dmy
+
#                     Material_chiffon_dmy + Material_linen_dmy + Material_lycra_dmy +
#                     Material_microfiber_dmy + Material_mix_dmy + Material_modal_dmy +
#                     Material_nylon_dmy + Material_other_dmy + Material_rayon_dmy +
#                     Material_silk_dmy + Material_spandex_dmy + Material_viscos_dmy,
#           data = Training1s)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.6607 -0.7061  0.1520  0.8476  3.6087
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)              7.77768    0.27113  28.686   <2e-16 ***
# Seasonspring            -0.09565    0.30284  -0.316   0.7523
# Seasonsummer            -0.06729    0.27714  -0.243   0.8083
# Seasonwinter            -0.07310    0.28300  -0.258   0.7963
# Pricehigh               -0.44216    0.42049  -1.052   0.2938
# Pricelow                 0.49945    0.19309   2.587   0.0101 *
# Pricemedium             -0.95281    0.38575  -2.470   0.0140 *
# Style_bohemian_dmy      -0.41883    0.45483  -0.921   0.3578
# Style_brief_dmy          0.26430    0.45188   0.585   0.5590
# Style_cute_dmy          -0.51178    0.33771  -1.515   0.1306
# Style_fashion_dmy       -1.65595    1.52879  -1.083   0.2795
# Style_flare_dmy          1.68848    1.68666   1.001   0.3175
# Style_novelty_dmy        0.19140    0.59316   0.323   0.7471
# Style_OL_dmy            -0.69031    1.56451  -0.441   0.6593
# Style_party_dmy         -0.05526    0.30411  -0.182   0.8559
# Style_sexy_dmy          -0.03415    0.23659  -0.144   0.8853
# Style_vintage_dmy        0.53454    0.38335   1.394   0.1642
# Material_acrylic_dmy    -0.26472    1.09977  -0.241   0.8099
# Material_cashmere_dmy   -2.07470    0.91918  -2.257   0.0247 *
# Material_chiffon_dmy     0.56650    0.41595   1.362   0.1742
# Material_linen_dmy       0.21923    0.90056   0.243   0.8078
# Material_lycra_dmy      -1.99404    1.09281  -1.825   0.0690 .
# Material_microfiber_dmy  0.20397    0.88810   0.230   0.8185
```

```
# Material_mix_dmy         0.17423    0.70663   0.247   0.8054
# Material_modal_dmy      -1.68801    1.08653  -1.554   0.1213
# Material_nylon_dmy      -1.41824    0.71661  -1.979   0.0487 *
# Material_other_dmy      -0.07113    0.20418  -0.348   0.7278
# Material_rayon_dmy      -0.65004    0.56951  -1.141   0.2546
# Material_silk_dmy       -0.63491    0.39416  -1.611   0.1082
# Material_spandex_dmy    -0.71276    0.79016  -0.902   0.3677
# Material_viscos_dmy     -0.15566    1.09128  -0.143   0.8867
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.516 on 321 degrees of freedom
# Multiple R-squared:  0.1505,   Adjusted R-squared:  0.07114
# F-statistic: 1.896 on 30 and 321 DF,  p-value: 0.003937

totsale_lm.4 <- update(totsale_lm.3, ~.-Material_viscos_dmy)
summary(totsale_lm.4)

#
# Call:
#         lm(formula = log(TotalSales) ~ Seasonspring + Seasonsummer +
#                      Seasonwinter + Pricehigh + Pricelow + Pricemedium +
Style_bohemian_dmy +
#                      Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                      Style_novelty_dmy + Style_OL_dmy + Style_party_dmy +
Style_sexy_dmy +
#                      Style_vintage_dmy + Material_acrylic_dmy + Material_cashmere_dmy
+
#                      Material_chiffon_dmy + Material_linen_dmy + Material_lycra_dmy +
#                      Material_microfiber_dmy + Material_mix_dmy + Material_modal_dmy +
#                      Material_nylon_dmy + Material_other_dmy + Material_rayon_dmy +
#                      Material_silk_dmy + Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6595 -0.7026  0.1522  0.8479  3.6086
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)              7.77823    0.27069  28.735   <2e-16 ***
# Seasonspring            -0.09547    0.30237  -0.316   0.7524
# Seasonsummer            -0.06854    0.27657  -0.248   0.8044
# Seasonwinter            -0.07436    0.28243  -0.263   0.7925
# Pricehigh               -0.44047    0.41968  -1.050   0.2947
# Pricelow                 0.49821    0.19260   2.587   0.0101 *
# Pricemedium             -0.95230    0.38514  -2.473   0.0139 *
# Style_bohemian_dmy      -0.41892    0.45414  -0.922   0.3570
# Style_brief_dmy          0.26403    0.45119   0.585   0.5588
# Style_cute_dmy          -0.51241    0.33717  -1.520   0.1296
# Style_fashion_dmy       -1.65524    1.52646  -1.084   0.2790
# Style_flare_dmy          1.68553    1.68396   1.001   0.3176
# Style_novelty_dmy        0.19139    0.59226   0.323   0.7468
# Style_OL_dmy            -0.69011    1.56213  -0.442   0.6589
# Style_party_dmy         -0.06095    0.30102  -0.202   0.8397
# Style_sexy_dmy          -0.03669    0.23556  -0.156   0.8763
# Style_vintage_dmy        0.53445    0.38276   1.396   0.1636
# Material_acrylic_dmy    -0.26326    1.09805  -0.240   0.8107
# Material_cashmere_dmy   -2.07433    0.91777  -2.260   0.0245 *
# Material_chiffon_dmy     0.56765    0.41523   1.367   0.1726
# Material_linen_dmy       0.21910    0.89919   0.244   0.8076
# Material_lycra_dmy      -1.98931    1.09064  -1.824   0.0691 .
# Material_microfiber_dmy  0.20651    0.88657   0.233   0.8160
# Material_mix_dmy         0.17416    0.70555   0.247   0.8052
# Material_modal_dmy      -1.68668    1.08483  -1.555   0.1210
# Material_nylon_dmy      -1.41602    0.71535  -1.979   0.0486 *
# Material_other_dmy      -0.06910    0.20337  -0.340   0.7343
# Material_rayon_dmy      -0.64862    0.56856  -1.141   0.2548
# Material_silk_dmy       -0.63331    0.39340  -1.610   0.1084
# Material_spandex_dmy    -0.71271    0.78895  -0.903   0.3670
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#
# Residual standard error: 1.514 on 322 degrees of freedom
# Multiple R-squared:  0.1505,   Adjusted R-squared:  0.07396
# F-statistic: 1.967 on 29 and 322 DF,  p-value: 0.002711

totsale_lm.5 <- update(totsale_lm.4, ~.- Style_sexy_dmy )
summary(totsale_lm.5)

# Call:
#         lm(formula = log(TotalSales) ~ Seasonspring + Seasonsummer +
#                    Seasonwinter + Pricehigh + Pricelow + Pricemedium +
Style_bohemian_dmy +
#                    Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                    Style_novelty_dmy + Style_OL_dmy + Style_party_dmy +
Style_vintage_dmy +
#                    Material_acrylic_dmy + Material_cashmere_dmy +
Material_chiffon_dmy +
#                    Material_linen_dmy + Material_lycra_dmy + Material_microfiber_dmy
+
#                    Material_mix_dmy + Material_modal_dmy + Material_nylon_dmy +
#                    Material_other_dmy + Material_rayon_dmy + Material_silk_dmy +
#                    Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6879 -0.7089  0.1526  0.8560  3.6088
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.76641    0.25945  29.934   <2e-16 ***
# Seasonspring           -0.09124    0.30069  -0.303   0.7618
# Seasonsummer           -0.06440    0.27488  -0.234   0.8149
# Seasonwinter           -0.07143    0.28137  -0.254   0.7998
# Pricehigh              -0.43708    0.41848  -1.044   0.2971
# Pricelow                0.49511    0.19129   2.588   0.0101 *
# Pricemedium            -0.95108    0.38448  -2.474   0.0139 *
# Style_bohemian_dmy     -0.40851    0.44851  -0.911   0.3631
# Style_brief_dmy         0.27309    0.44674   0.611   0.5414
# Style_cute_dmy         -0.50504    0.33332  -1.515   0.1307
# Style_fashion_dmy      -1.64757    1.52336  -1.082   0.2803
# Style_flare_dmy         1.69709    1.67979   1.010   0.3131
# Style_novelty_dmy       0.20066    0.58836   0.341   0.7333
# Style_OL_dmy           -0.68245    1.55899  -0.438   0.6619
# Style_party_dmy        -0.05327    0.29651  -0.180   0.8575
# Style_vintage_dmy       0.54278    0.37844   1.434   0.1525
# Material_acrylic_dmy   -0.25796    1.09586  -0.235   0.8141
# Material_cashmere_dmy  -2.06626    0.91493  -2.258   0.0246 *
# Material_chiffon_dmy    0.57499    0.41193   1.396   0.1637
# Material_linen_dmy      0.22953    0.89534   0.256   0.7978
# Material_lycra_dmy     -2.00172    1.08609  -1.843   0.0662 .
# Material_microfiber_dmy  0.21303    0.88425   0.241   0.8098
# Material_mix_dmy        0.17878    0.70386   0.254   0.7997
# Material_modal_dmy     -1.67746    1.08158  -1.551   0.1219
# Material_nylon_dmy     -1.42000    0.71381  -1.989   0.0475 *
# Material_other_dmy     -0.06963    0.20303  -0.343   0.7319
# Material_rayon_dmy     -0.64685    0.56759  -1.140   0.2553
# Material_silk_dmy      -0.63362    0.39281  -1.613   0.1077
# Material_spandex_dmy   -0.70543    0.78638  -0.897   0.3704
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.511 on 323 degrees of freedom
# Multiple R-squared:  0.1504,   Adjusted R-squared:  0.07676
# F-statistic: 2.042 on 28 and 323 DF,  p-value: 0.001835

totsale_lm.6 <- update(totsale_lm.5, ~.-Style_party_dmy)

summary(totsale_lm.6)


# Call:
```

```
#         lm(formula = log(TotalSales) ~ Seasonspring + Seasonsummer +
#                    Seasonwinter + Pricehigh + Pricelow + Pricemedium +
Style_bohemian_dmy +
#                    Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                    Style_novelty_dmy + Style_OL_dmy + Style_vintage_dmy +
Material_acrylic_dmy +
#                    Material_cashmere_dmy + Material_chiffon_dmy + Material_linen_dmy
+
#                    Material_lycra_dmy + Material_microfiber_dmy + Material_mix_dmy +
#                    Material_modal_dmy + Material_nylon_dmy + Material_other_dmy +
#                    Material_rayon_dmy + Material_silk_dmy + Material_spandex_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
#         -4.6744 -0.6990  0.1573  0.8633  3.6098
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.76313    0.25842  30.040  < 2e-16 ***
# Seasonspring           -0.09919    0.29697  -0.334  0.73860
# Seasonsummer           -0.06836    0.27359  -0.250  0.80286
# Seasonwinter           -0.07674    0.27940  -0.275  0.78376
# Pricehigh              -0.45202    0.40952  -1.104  0.27051
# Pricelow                0.50358    0.18511   2.720  0.00687 **
# Pricemedium            -0.94964    0.38382  -2.474  0.01387 *
# Style_bohemian_dmy     -0.40182    0.44630  -0.900  0.36860
# Style_brief_dmy         0.28091    0.44395   0.633  0.52734
# Style_cute_dmy         -0.49486    0.32798  -1.509  0.13232
# Style_fashion_dmy      -1.64033    1.52055  -1.079  0.28149
# Style_flare_dmy         1.71464    1.67444   1.024  0.30660
# Style_novelty_dmy       0.20805    0.58605   0.355  0.72282
# Style_OL_dmy           -0.67529    1.55616  -0.434  0.66461
# Style_vintage_dmy       0.55114    0.37500   1.470  0.14261
# Material_acrylic_dmy   -0.25819    1.09423  -0.236  0.81362
# Material_cashmere_dmy  -2.05862    0.91257  -2.256  0.02475 *
# Material_chiffon_dmy    0.57321    0.41119   1.394  0.16427
# Material_linen_dmy      0.23413    0.89363   0.262  0.79349
# Material_lycra_dmy     -2.02267    1.07819  -1.876  0.06156 .
# Material_microfiber_dmy  0.20147    0.88059   0.229  0.81918
# Material_mix_dmy        0.18572    0.70175   0.265  0.79145
# Material_modal_dmy     -1.67445    1.07983  -1.551  0.12196
# Material_nylon_dmy     -1.42631    0.71188  -2.004  0.04595 *
# Material_other_dmy     -0.07593    0.19967  -0.380  0.70398
# Material_rayon_dmy     -0.65757    0.56360  -1.167  0.24417
# Material_silk_dmy      -0.63406    0.39221  -1.617  0.10693
# Material_spandex_dmy   -0.70097    0.78481  -0.893  0.37243
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.509 on 324 degrees of freedom
# Multiple R-squared:  0.1503,   Adjusted R-squared:  0.07952
# F-statistic: 2.123 on 27 and 324 DF,  p-value: 0.001222

totsale_lm.7 <- update(totsale_lm.6, ~.-Material_microfiber_dmy)

summary(totsale_lm.7)
# Call:
#         lm(formula = log(TotalSales) ~ Seasonspring + Seasonsummer +
#                    Seasonwinter + Pricehigh + Pricelow + Pricemedium +
Style_bohemian_dmy +
#                    Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                    Style_novelty_dmy + Style_OL_dmy + Style_vintage_dmy +
Material_acrylic_dmy +
#                    Material_cashmere_dmy + Material_chiffon_dmy + Material_linen_dmy
+
#                    Material_lycra_dmy + Material_mix_dmy + Material_modal_dmy +
#                    Material_nylon_dmy + Material_other_dmy + Material_rayon_dmy +
#                    Material_silk_dmy + Material_spandex_dmy, data = Training1s)
#
```

```
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6750 -0.6913  0.1562  0.8595  3.6064
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)           7.76514    0.25790  30.109  < 2e-16 ***
# Seasonspring         -0.09569    0.29614  -0.323  0.74680
# Seasonsummer         -0.06666    0.27309  -0.244  0.80731
# Seasonwinter         -0.07491    0.27888  -0.269  0.78840
# Pricehigh            -0.45377    0.40886  -1.110  0.26788
# Pricelow              0.50316    0.18483   2.722  0.00683 **
# Pricemedium          -0.95192    0.38313  -2.485  0.01347 *
# Style_bohemian_dmy   -0.40382    0.44556  -0.906  0.36544
# Style_brief_dmy       0.27876    0.44320   0.629  0.52981
# Style_cute_dmy       -0.49690    0.32738  -1.518  0.13004
# Style_fashion_dmy    -1.64404    1.51824  -1.083  0.27968
# Style_flare_dmy       1.71219    1.67196   1.024  0.30657
# Style_novelty_dmy     0.20554    0.58509   0.351  0.72560
# Style_OL_dmy         -0.67685    1.55387  -0.436  0.66342
# Style_vintage_dmy     0.54908    0.37435   1.467  0.14340
# Material_acrylic_dmy  -0.26067    1.09258  -0.239  0.81158
# Material_cashmere_dmy -2.06094    0.91119  -2.262  0.02437 *
# Material_chiffon_dmy   0.56985    0.41033   1.389  0.16586
# Material_linen_dmy     0.23156    0.89226   0.260  0.79539
# Material_lycra_dmy    -2.02713    1.07644  -1.883  0.06057 .
# Material_mix_dmy       0.18173    0.70051   0.259  0.79547
# Material_modal_dmy    -1.67795    1.07815  -1.556  0.12060
# Material_nylon_dmy    -1.42936    0.71071  -2.011  0.04513 *
# Material_other_dmy    -0.07900    0.19893  -0.397  0.69156
# Material_rayon_dmy    -0.66014    0.56266  -1.173  0.24156
# Material_silk_dmy     -0.63740    0.39137  -1.629  0.10436
# Material_spandex_dmy  -0.70575    0.78339  -0.901  0.36831
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.507 on 325 degrees of freedom
# Multiple R-squared:  0.1502,   Adjusted R-squared:  0.0822
# F-statistic: 2.209 on 26 and 325 DF,  p-value: 0.0008035

totsale_lm.8 <- update(totsale_lm.7, ~. -Material_acrylic_dmy)
summary(totsale_lm.8)

# Call:
#         lm(formula = log(TotalSales) ~ Seasonspring + Seasonsummer +
#                   Seasonwinter + Pricehigh + Pricelow + Pricemedium +
Style_bohemian_dmy +
#                   Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                   Style_novelty_dmy + Style_OL_dmy + Style_vintage_dmy +
Material_cashmere_dmy +
#                   Material_chiffon_dmy + Material_linen_dmy + Material_lycra_dmy +
#                   Material_mix_dmy + Material_modal_dmy + Material_nylon_dmy +
#                   Material_other_dmy + Material_rayon_dmy + Material_silk_dmy +
#                   Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6751 -0.6878  0.1572  0.8621  3.6062
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)           7.76519    0.25752  30.153  < 2e-16 ***
# Seasonspring         -0.09617    0.29571  -0.325  0.74522
# Seasonsummer         -0.06926    0.27247  -0.254  0.79950
# Seasonwinter         -0.07700    0.27834  -0.277  0.78222
# Pricehigh            -0.45422    0.40826  -1.113  0.26671
# Pricelow              0.50206    0.18450   2.721  0.00686 **
# Pricemedium          -0.95170    0.38258  -2.488  0.01336 *
# Style_bohemian_dmy   -0.40273    0.44489  -0.905  0.36601
# Style_brief_dmy       0.25905    0.43480   0.596  0.55174
# Style_cute_dmy       -0.49630    0.32690  -1.518  0.12993
```

```
# Style_fashion_dmy     -1.64149    1.51601  -1.083  0.27971
# Style_flare_dmy        1.71069    1.66953   1.025  0.30628
# Style_novelty_dmy      0.20672    0.58422   0.354  0.72369
# Style_OL_dmy          -0.67503    1.55160  -0.435  0.66381
# Style_vintage_dmy      0.54990    0.37379   1.471  0.14222
# Material_cashmere_dmy -2.05905    0.90983  -2.263  0.02429 *
# Material_chiffon_dmy   0.57158    0.40967   1.395  0.16390
# Material_linen_dmy     0.23238    0.89096   0.261  0.79440
# Material_lycra_dmy    -2.02534    1.07486  -1.884  0.06042 .
# Material_mix_dmy       0.18658    0.69920   0.267  0.78976
# Material_modal_dmy    -1.67485    1.07651  -1.556  0.12072
# Material_nylon_dmy    -1.42745    0.70964  -2.012  0.04509 *
# Material_other_dmy    -0.07639    0.19835  -0.385  0.70040
# Material_rayon_dmy    -0.65690    0.56168  -1.170  0.24305
# Material_silk_dmy     -0.63551    0.39072  -1.627  0.10481
# Material_spandex_dmy  -0.70492    0.78225  -0.901  0.36817
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.505 on 326 degrees of freedom
# Multiple R-squared:   0.15,    Adjusted R-squared:  0.08486
# F-statistic: 2.302 on 25 and 326 DF,  p-value: 0.0005185

totsale_lm.9 <- update(totsale_lm.8, ~. -Seasonsummer)
summary(totsale_lm.9)

# Call:
#         lm(formula = log(TotalSales) ~ Seasonspring + Seasonwinter +
#                    Pricehigh + Pricelow + Pricemedium + Style_bohemian_dmy +
#                    Style_brief_dmy + Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy +
#                    Style_novelty_dmy + Style_OL_dmy + Style_vintage_dmy +
Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_linen_dmy + Material_lycra_dmy +
#                    Material_mix_dmy + Material_modal_dmy + Material_nylon_dmy +
#                    Material_other_dmy + Material_rayon_dmy + Material_silk_dmy +
#                    Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6980 -0.6907  0.1494  0.8592  3.6145
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.71537    0.16681  46.253  < 2e-16 ***
# Seasonspring          -0.04669    0.22229  -0.210  0.83376
# Seasonwinter          -0.02749    0.19854  -0.138  0.88996
# Pricehigh             -0.45855    0.40732  -1.126  0.26108
# Pricelow               0.50079    0.18417   2.719  0.00689 **
# Pricemedium           -0.94563    0.38129  -2.480  0.01364 *
# Style_bohemian_dmy    -0.40632    0.44403  -0.915  0.36083
# Style_brief_dmy        0.26226    0.43400   0.604  0.54607
# Style_cute_dmy        -0.50429    0.32492  -1.552  0.12162
# Style_fashion_dmy     -1.66093    1.51191  -1.099  0.27277
# Style_flare_dmy        1.71555    1.66703   1.029  0.30419
# Style_novelty_dmy      0.21417    0.58265   0.368  0.71342
# Style_OL_dmy          -0.68079    1.54922  -0.439  0.66063
# Style_vintage_dmy      0.54098    0.37161   1.456  0.14641
# Material_cashmere_dmy -2.06278    0.90841  -2.271  0.02381 *
# Material_chiffon_dmy   0.57470    0.40890   1.405  0.16083
# Material_linen_dmy     0.25911    0.88346   0.293  0.76949
# Material_lycra_dmy    -2.02438    1.07331  -1.886  0.06017 .
# Material_mix_dmy       0.18805    0.69818   0.269  0.78784
# Material_modal_dmy    -1.69366    1.07243  -1.579  0.11524
# Material_nylon_dmy    -1.43197    0.70840  -2.021  0.04405 *
# Material_other_dmy    -0.07297    0.19761  -0.369  0.71218
# Material_rayon_dmy    -0.64511    0.55897  -1.154  0.24929
# Material_silk_dmy     -0.63323    0.39006  -1.623  0.10546
# Material_spandex_dmy  -0.70195    0.78104  -0.899  0.36945
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
```

```
# Residual standard error: 1.503 on 327 degrees of freedom
# Multiple R-squared:  0.1499,   Adjusted R-squared:  0.08747
# F-statistic: 2.402 on 24 and 327 DF,  p-value: 0.0003286

totsale_lm.10 <- update(totsale_lm.9, ~. -Seasonwinter)
summary(totsale_lm.10)

# Call:
#         lm(formula = log(TotalSales) ~ Seasonspring + Pricehigh + Pricelow +
#                    Pricemedium + Style_bohemian_dmy + Style_brief_dmy +
Style_cute_dmy +
#                    Style_fashion_dmy + Style_flare_dmy + Style_novelty_dmy +
#                    Style_OL_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_linen_dmy + Material_lycra_dmy +
#                    Material_mix_dmy + Material_modal_dmy + Material_nylon_dmy +
#                    Material_other_dmy + Material_rayon_dmy + Material_silk_dmy +
#                    Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6871 -0.6985  0.1484  0.8555  3.6157
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.70254    0.13851  55.612  < 2e-16 ***
# Seasonspring          -0.03508    0.20554  -0.171  0.86461
# Pricehigh             -0.46318    0.40534  -1.143  0.25400
# Pricelow               0.50497    0.18141   2.784  0.00569 **
# Pricemedium           -0.95275    0.37724  -2.526  0.01202 *
# Style_bohemian_dmy    -0.40173    0.44213  -0.909  0.36422
# Style_brief_dmy        0.25767    0.43208   0.596  0.55136
# Style_cute_dmy        -0.50422    0.32443  -1.554  0.12111
# Style_fashion_dmy     -1.64810    1.50681  -1.094  0.27486
# Style_flare_dmy        1.71750    1.66448   1.032  0.30290
# Style_novelty_dmy      0.21595    0.58164   0.371  0.71066
# Style_OL_dmy          -0.68834    1.54594  -0.445  0.65643
# Style_vintage_dmy      0.54197    0.37098   1.461  0.14500
# Material_cashmere_dmy -2.07270    0.90423  -2.292  0.02252 *
# Material_chiffon_dmy   0.57613    0.40816   1.412  0.15904
# Material_linen_dmy     0.27194    0.87728   0.310  0.75677
# Material_lycra_dmy    -2.03320    1.06982  -1.901  0.05824 .
# Material_mix_dmy       0.18363    0.69640   0.264  0.79219
# Material_modal_dmy    -1.68292    1.06802  -1.576  0.11605
# Material_nylon_dmy    -1.43270    0.70732  -2.026  0.04362 *
# Material_other_dmy    -0.07098    0.19679  -0.361  0.71857
# Material_rayon_dmy    -0.63870    0.55621  -1.148  0.25168
# Material_silk_dmy     -0.63558    0.38911  -1.633  0.10334
# Material_spandex_dmy  -0.70285    0.77985  -0.901  0.36811
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.5 on 328 degrees of freedom
# Multiple R-squared:  0.1498,   Adjusted R-squared:  0.0902
# F-statistic: 2.513 on 23 and 328 DF,  p-value: 0.0002008

totsale_lm.11 <- update(totsale_lm.10, ~. -Seasonspring)
summary(totsale_lm.11)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                    Style_bohemian_dmy + Style_brief_dmy + Style_cute_dmy +
Style_fashion_dmy +
#                    Style_flare_dmy + Style_novelty_dmy + Style_OL_dmy +
Style_vintage_dmy +
#                    Material_cashmere_dmy + Material_chiffon_dmy + Material_linen_dmy
+
#                    Material_lycra_dmy + Material_mix_dmy + Material_modal_dmy +
#                    Material_nylon_dmy + Material_other_dmy + Material_rayon_dmy +
#                    Material_silk_dmy + Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
```

```
# -4.6787 -0.7040  0.1500  0.8608  3.5900
#
# Coefficients:
#          Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.6975     0.1351  56.979  < 2e-16 ***
# Pricehigh             -0.4662     0.4043  -1.153  0.24973
# Pricelow               0.5040     0.1810   2.784  0.00569 **
# Pricemedium           -0.9562     0.3761  -2.542  0.01147 *
# Style_bohemian_dmy    -0.3982     0.4410  -0.903  0.36717
# Style_brief_dmy        0.2582     0.4314   0.598  0.54996
# Style_cute_dmy        -0.5086     0.3230  -1.575  0.11629
# Style_fashion_dmy     -1.6430     1.5043  -1.092  0.27553
# Style_flare_dmy        1.6966     1.6575   1.024  0.30678
# Style_novelty_dmy      0.2220     0.5797   0.383  0.70202
# Style_OL_dmy          -0.6798     1.5429  -0.441  0.65978
# Style_vintage_dmy      0.5466     0.3695   1.479  0.13999
# Material_cashmere_dmy  -2.0653     0.9019  -2.290  0.02265 *
# Material_chiffon_dmy    0.5705     0.4062   1.404  0.16114
# Material_linen_dmy     0.2770     0.8755   0.316  0.75190
# Material_lycra_dmy    -2.0452     1.0659  -1.919  0.05589 .
# Material_mix_dmy       0.1666     0.6882   0.242  0.80885
# Material_modal_dmy    -1.6773     1.0659  -1.574  0.11655
# Material_nylon_dmy    -1.4419     0.7043  -2.047  0.04142 *
# Material_other_dmy    -0.0743     0.1955  -0.380  0.70420
# Material_rayon_dmy    -0.6505     0.5511  -1.181  0.23865
# Material_silk_dmy     -0.6456     0.3841  -1.681  0.09369 .
# Material_spandex_dmy  -0.7313     0.7607  -0.961  0.33712
# ---
#          Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.498 on 329 degrees of freedom
# Multiple R-squared:  0.1497,    Adjusted R-squared:  0.09289
# F-statistic: 2.634 on 22 and 329 DF,  p-value: 0.0001203

totsale_lm.12 <- update(totsale_lm.11, ~. -Material_mix_dmy)
summary(totsale_lm.12)

# Call:
#          lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                     Style_bohemian_dmy + Style_brief_dmy + Style_cute_dmy +
Style_fashion_dmy +
#                     Style_flare_dmy + Style_novelty_dmy + Style_OL_dmy +
Style_vintage_dmy +
#                     Material_cashmere_dmy + Material_chiffon_dmy + Material_linen_dmy
+
#                     Material_lycra_dmy + Material_modal_dmy + Material_nylon_dmy +
#                     Material_other_dmy + Material_rayon_dmy + Material_silk_dmy +
#                     Material_spandex_dmy, data = Training1s)
#
# Residuals:
#          Min      1Q  Median      3Q     Max
# -4.6796 -0.7072  0.1470  0.8579  3.5859
#
# Coefficients:
#          Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.70260    0.13324  57.811   <2e-16 ***
# Pricehigh             -0.46837    0.40367  -1.160   0.2468
# Pricelow               0.49994    0.18003   2.777   0.0058 **
# Pricemedium           -0.96001    0.37527  -2.558   0.0110 *
# Style_bohemian_dmy    -0.39952    0.44035  -0.907   0.3649
# Style_brief_dmy        0.26887    0.42855   0.627   0.5308
# Style_cute_dmy        -0.50961    0.32247  -1.580   0.1150
# Style_fashion_dmy     -1.64816    1.50200  -1.097   0.2733
# Style_flare_dmy        1.69328    1.65510   1.023   0.3070
# Style_novelty_dmy      0.21802    0.57865   0.377   0.7066
# Style_OL_dmy          -0.68113    1.54064  -0.442   0.6587
# Style_vintage_dmy      0.55375    0.36773   1.506   0.1331
# Material_cashmere_dmy -2.06791    0.90051  -2.296   0.0223 *
# Material_chiffon_dmy   0.56702    0.40540   1.399   0.1629
# Material_linen_dmy     0.27188    0.87398   0.311   0.7559
# Material_lycra_dmy    -2.04828    1.06434  -1.924   0.0552 .
# Material_modal_dmy    -1.68046    1.06434  -1.579   0.1153
```

```
# Material_nylon_dmy    -1.44362    0.70321  -2.053   0.0409 *
# Material_other_dmy    -0.07859    0.19445  -0.404   0.6863
# Material_rayon_dmy    -0.65619    0.54977  -1.194   0.2335
# Material_silk_dmy     -0.64696    0.38347  -1.687   0.0925 .
# Material_spandex_dmy  -0.73412    0.75955  -0.967   0.3345
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.496 on 330 degrees of freedom
# Multiple R-squared:  0.1496,   Adjusted R-squared:  0.09547
# F-statistic: 2.764 on 21 and 330 DF,  p-value: 0.00007111

totsale_lm.13 <- update(totsale_lm.12, ~. -Material_linen_dmy)
summary(totsale_lm.13)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                    Style_bohemian_dmy + Style_brief_dmy + Style_cute_dmy +
Style_fashion_dmy +
#                    Style_flare_dmy + Style_novelty_dmy + Style_OL_dmy +
Style_vintage_dmy +
#                    Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                    Material_modal_dmy + Material_nylon_dmy + Material_other_dmy +
#                    Material_rayon_dmy + Material_silk_dmy + Material_spandex_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6822 -0.7114  0.1501  0.8525  3.5817
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.70892    0.13150  58.623  < 2e-16 ***
# Pricehigh             -0.47187    0.40297  -1.171  0.24244
# Pricelow               0.49562    0.17924   2.765  0.00601 **
# Pricemedium           -0.96416    0.37453  -2.574  0.01048 *
# Style_bohemian_dmy    -0.40212    0.43966  -0.915  0.36106
# Style_brief_dmy        0.26527    0.42781   0.620  0.53564
# Style_cute_dmy        -0.51175    0.32196  -1.590  0.11290
# Style_fashion_dmy     -1.65448    1.49981  -1.103  0.27077
# Style_flare_dmy        1.68895    1.65278   1.022  0.30758
# Style_novelty_dmy      0.21283    0.57762   0.368  0.71277
# Style_OL_dmy          -0.68330    1.53852  -0.444  0.65724
# Style_vintage_dmy      0.55024    0.36706   1.499  0.13482
# Material_cashmere_dmy -2.07147    0.89921  -2.304  0.02186 *
# Material_chiffon_dmy   0.56338    0.40468   1.392  0.16481
# Material_lycra_dmy    -2.05244    1.06281  -1.931  0.05432 .
# Material_modal_dmy    -1.68462    1.06281  -1.585  0.11391
# Material_nylon_dmy    -1.44560    0.70222  -2.059  0.04031 *
# Material_other_dmy    -0.08233    0.19381  -0.425  0.67128
# Material_rayon_dmy    -0.65924    0.54894  -1.201  0.23063
# Material_silk_dmy     -0.64883    0.38289  -1.695  0.09110 .
# Material_spandex_dmy  -0.73774    0.75842  -0.973  0.33140
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.494 on 331 degrees of freedom
# Multiple R-squared:  0.1493,   Adjusted R-squared:  0.09794
# F-statistic: 2.906 on 20 and 331 DF,  p-value: 0.00004155

totsale_lm.14 <- update(totsale_lm.13, ~. -Style_novelty_dmy )
summary(totsale_lm.14)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                    Style_bohemian_dmy + Style_brief_dmy + Style_cute_dmy +
Style_fashion_dmy +
#                    Style_flare_dmy + Style_OL_dmy + Style_vintage_dmy +
Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                    Material_nylon_dmy + Material_other_dmy + Material_rayon_dmy +
```

```
#                     Material_silk_dmy + Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6881 -0.7161  0.1503  0.8547  3.5769
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.71798    0.12901  59.825  < 2e-16 ***
# Pricehigh             -0.47784    0.40212  -1.188  0.23556
# Pricelow               0.48913    0.17814   2.746  0.00637 **
# Pricemedium           -0.95861    0.37374  -2.565  0.01076 *
# Style_bohemian_dmy    -0.40735    0.43886  -0.928  0.35398
# Style_brief_dmy        0.25942    0.42696   0.608  0.54387
# Style_cute_dmy        -0.51601    0.32133  -1.606  0.10926
# Style_fashion_dmy     -1.66354    1.49766  -1.111  0.26747
# Style_flare_dmy        1.68177    1.65052   1.019  0.30898
# Style_OL_dmy          -0.69792    1.53600  -0.454  0.64986
# Style_vintage_dmy      0.54379    0.36617   1.485  0.13846
# Material_cashmere_dmy -2.08424    0.89737  -2.323  0.02080 *
# Material_chiffon_dmy   0.55862    0.40395   1.383  0.16762
# Material_lycra_dmy    -2.05826    1.06131  -1.939  0.05330 .
# Material_modal_dmy    -1.69044    1.06131  -1.593  0.11216
# Material_nylon_dmy    -1.44749    0.70128  -2.064  0.03979 *
# Material_other_dmy    -0.08543    0.19338  -0.442  0.65895
# Material_rayon_dmy    -0.66269    0.54814  -1.209  0.22754
# Material_silk_dmy     -0.65221    0.38229  -1.706  0.08893 .
# Material_spandex_dmy  -0.74250    0.75733  -0.980  0.32759
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.492 on 332 degrees of freedom
# Multiple R-squared:  0.149,    Adjusted R-squared:  0.1003
# F-statistic: 3.059 on 19 and 332 DF,  p-value: 0.000024

totsale_lm.15 <- update(totsale_lm.14, ~. -Material_other_dmy)
summary(totsale_lm.15)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                     Style_bohemian_dmy + Style_brief_dmy + Style_cute_dmy +
Style_fashion_dmy +
#                     Style_flare_dmy + Style_OL_dmy + Style_vintage_dmy +
Material_cashmere_dmy +
#                     Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                     Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy +
#                     Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7507 -0.7035  0.1650  0.8629  3.6006
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.6952     0.1181  65.170  < 2e-16 ***
# Pricehigh              -0.5048     0.3970  -1.271  0.20445
# Pricelow                0.4868     0.1779   2.737  0.00653 **
# Pricemedium            -0.9584     0.3733  -2.568  0.01068 *
# Style_bohemian_dmy     -0.4101     0.4383  -0.936  0.35007
# Style_brief_dmy         0.2544     0.4263   0.597  0.55101
# Style_cute_dmy         -0.5169     0.3209  -1.611  0.10822
# Style_fashion_dmy      -1.6407     1.4950  -1.098  0.27321
# Style_flare_dmy         1.6739     1.6484   1.015  0.31064
# Style_OL_dmy           -0.6753     1.5333  -0.440  0.65991
# Style_vintage_dmy       0.5357     0.3653   1.467  0.14345
# Material_cashmere_dmy  -2.0616     0.8948  -2.304  0.02184 *
# Material_chiffon_dmy    0.5832     0.3996   1.460  0.14535
# Material_lycra_dmy     -2.0343     1.0586  -1.922  0.05551 .
# Material_modal_dmy     -1.6665     1.0586  -1.574  0.11640
# Material_nylon_dmy     -1.4168     0.6970  -2.033  0.04287 *
# Material_rayon_dmy     -0.6337     0.5435  -1.166  0.24449
# Material_silk_dmy      -0.6245     0.3767  -1.658  0.09824 .
```

```
# Material_spandex_dmy   -0.7183     0.7544  -0.952  0.34173
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.49 on 333 degrees of freedom
# Multiple R-squared:  0.1485,   Adjusted R-squared:  0.1025
# F-statistic: 3.226 on 18 and 333 DF,  p-value: 0.00001381

totsale_lm.16 <- update(totsale_lm.15, ~. -Style_OL_dmy)
summary(totsale_lm.16)

# Call:
#        lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                   Style_bohemian_dmy + Style_brief_dmy + Style_cute_dmy +
Style_fashion_dmy +
#                   Style_flare_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                   Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                   Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy +
#                   Material_spandex_dmy, data = Training1s)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.7501 -0.7027  0.1657  0.8620  3.5982
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.6945     0.1179  65.248  < 2e-16 ***
# Pricehigh             -0.5051     0.3965  -1.274  0.20358
# Pricelow               0.4865     0.1776   2.739  0.00650 **
# Pricemedium           -0.9951     0.3634  -2.738  0.00651 **
# Style_bohemian_dmy    -0.4063     0.4377  -0.928  0.35386
#              Style_brief_dmy         0.2553     0.4258   0.600  0.54922
# Style_cute_dmy        -0.5138     0.3205  -1.603  0.10985
# Style_fashion_dmy     -1.6401     1.4932  -1.098  0.27282
# Style_flare_dmy        1.6751     1.6464   1.017  0.30969
# Style_vintage_dmy      0.5385     0.3648   1.476  0.14081
# Material_cashmere_dmy  -2.0364     0.8919  -2.283  0.02305 *
# Material_chiffon_dmy    0.5832     0.3991   1.461  0.14487
# Material_lycra_dmy    -2.0335     1.0574  -1.923  0.05531 .
# Material_modal_dmy    -1.6656     1.0574  -1.575  0.11614
# Material_nylon_dmy    -1.4174     0.6961  -2.036  0.04254 *
# Material_rayon_dmy    -0.6346     0.5429  -1.169  0.24323
# Material_silk_dmy     -0.6202     0.3761  -1.649  0.10004
# Material_spandex_dmy   -0.7183     0.7535  -0.953  0.34117
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.488 on 334 degrees of freedom
# Multiple R-squared:  0.148,    Adjusted R-squared:  0.1046
# F-statistic: 3.413 on 17 and 334 DF,  p-value: 0.000007727

totsale_lm.17 <- update(totsale_lm.16, ~. -Style_brief_dmy)
summary(totsale_lm.17)
# Call:
#        lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                   Style_bohemian_dmy + Style_cute_dmy + Style_fashion_dmy +
#                   Style_flare_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                   Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                   Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy +
#                   Material_spandex_dmy, data = Training1s)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.7641 -0.7162  0.1636  0.8660  3.5942
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.7086     0.1155  66.762  < 2e-16 ***
# Pricehigh             -0.5177     0.3956  -1.309  0.19155
# Pricelow               0.4840     0.1774   2.728  0.00671 **
# Pricemedium           -1.0053     0.3627  -2.772  0.00589 **
# Style_bohemian_dmy    -0.4197     0.4367  -0.961  0.33714
```

```
# Style_cute_dmy         -0.5239     0.3197  -1.639  0.10225
# Style_fashion_dmy      -1.6541     1.4915  -1.109  0.26823
# Style_flare_dmy         1.6658     1.6448   1.013  0.31191
# Style_vintage_dmy       0.5247     0.3637   1.443  0.15006
# Material_cashmere_dmy  -2.0437     0.8910  -2.294  0.02242 *
# Material_chiffon_dmy    0.5733     0.3984   1.439  0.15108
# Material_lycra_dmy     -2.0463     1.0561  -1.938  0.05352 .
# Material_modal_dmy     -1.6784     1.0561  -1.589  0.11295
# Material_nylon_dmy     -1.4221     0.6954  -2.045  0.04165 *
# Material_rayon_dmy     -0.6087     0.5406  -1.126  0.26104
# Material_silk_dmy      -0.6277     0.3755  -1.672  0.09553 .
# Material_spandex_dmy   -0.7285     0.7526  -0.968  0.33373
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.487 on 335 degrees of freedom
# Multiple R-squared:  0.1471,   Adjusted R-squared:  0.1063
# F-statistic: 3.611 on 16 and 335 DF,  p-value: 0.000004479

totsale_lm.18 <- update(totsale_lm.17, ~. -Style_bohemian_dmy)
summary(totsale_lm.18)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                     Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_vintage_dmy +
#                     Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                     Material_modal_dmy + Material_nylon_dmy + Material_rayon_dmy +
#                     Material_silk_dmy + Material_spandex_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7498 -0.7364  0.1655  0.8745  3.5953
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.6943     0.1145  67.206  < 2e-16 ***
# Pricehigh              -0.4953     0.3949  -1.254  0.21054
# Pricelow                0.4779     0.1773   2.696  0.00737 **
# Pricemedium            -1.0180     0.3624  -2.809  0.00526 **
# Style_cute_dmy         -0.5099     0.3193  -1.597  0.11130
# Style_fashion_dmy      -1.6398     1.4913  -1.100  0.27230
# Style_flare_dmy         1.8345     1.6352   1.122  0.26272
# Style_vintage_dmy       0.5428     0.3632   1.495  0.13593
# Material_cashmere_dmy  -2.0209     0.8906  -2.269  0.02389 *
# Material_chiffon_dmy    0.5864     0.3981   1.473  0.14169
# Material_lycra_dmy     -2.0289     1.0559  -1.922  0.05550 .
# Material_modal_dmy     -1.6611     1.0559  -1.573  0.11661
# Material_nylon_dmy     -1.5765     0.6765  -2.330  0.02039 *
# Material_rayon_dmy     -0.6539     0.5385  -1.214  0.22553
# Material_silk_dmy      -0.6132     0.3752  -1.634  0.10309
# Material_spandex_dmy   -0.7147     0.7524  -0.950  0.34284
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.487 on 336 degrees of freedom
# Multiple R-squared:  0.1447,   Adjusted R-squared:  0.1065
# F-statistic:  3.79 on 15 and 336 DF,  p-value: 0.00000315

totsale_lm.19 <- update(totsale_lm.18, ~. -Material_spandex_dmy )
summary(totsale_lm.19)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                     Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_vintage_dmy +
#                     Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                     Material_modal_dmy + Material_nylon_dmy + Material_rayon_dmy +
#                     Material_silk_dmy, data = Training1s)
#
```

```
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7421 -0.7303  0.1687  0.8714  3.6283
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.6866     0.1142  67.317  < 2e-16 ***
# Pricehigh              -0.4861     0.3947  -1.232  0.21894
# Pricelow                0.4726     0.1772   2.667  0.00801 **
# Pricemedium            -1.0098     0.3622  -2.788  0.00561 **
# Style_cute_dmy         -0.5360     0.3181  -1.685  0.09296 .
# Style_fashion_dmy      -1.6321     1.4911  -1.095  0.27447
# Style_flare_dmy         1.8331     1.6350   1.121  0.26300
# Style_vintage_dmy       0.5496     0.3630   1.514  0.13096
# Material_cashmere_dmy  -2.0187     0.8904  -2.267  0.02402 *
# Material_chiffon_dmy    0.6013     0.3977   1.512  0.13150
# Material_lycra_dmy     -2.0186     1.0556  -1.912  0.05670 .
# Material_modal_dmy     -1.6508     1.0556  -1.564  0.11882
# Material_nylon_dmy     -1.5674     0.6764  -2.317  0.02108 *
# Material_rayon_dmy     -0.6403     0.5383  -1.190  0.23502
# Material_silk_dmy      -0.5978     0.3748  -1.595  0.11161
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.487 on 337 degrees of freedom
# Multiple R-squared:  0.1424,   Adjusted R-squared:  0.1068
# F-statistic: 3.998 on 14 and 337 DF,  p-value: 0.00000215

totsale_lm.20 <- update(totsale_lm.19, ~. -Style_fashion_dmy)
summary(totsale_lm.20)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                    Style_cute_dmy + Style_flare_dmy + Style_vintage_dmy +
Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                    Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7326 -0.7398  0.1749  0.8810  3.6340
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.6770     0.1139  67.411  < 2e-16 ***
# Pricehigh              -0.4778     0.3947  -1.210  0.22694
# Pricelow                0.4811     0.1770   2.717  0.00692 **
# Pricemedium            -1.0015     0.3623  -2.764  0.00602 **
# Style_cute_dmy         -0.5321     0.3182  -1.672  0.09538 .
# Style_flare_dmy         1.8399     1.6354   1.125  0.26138
# Style_vintage_dmy       0.5563     0.3631   1.532  0.12639
# Material_cashmere_dmy  -2.0147     0.8907  -2.262  0.02433 *
# Material_chiffon_dmy    0.6057     0.3978   1.523  0.12881
# Material_lycra_dmy     -2.0133     1.0559  -1.907  0.05742 .
# Material_modal_dmy     -1.6454     1.0559  -1.558  0.12010
# Material_nylon_dmy     -1.5646     0.6766  -2.313  0.02134 *
# Material_rayon_dmy     -0.6357     0.5384  -1.181  0.23852
# Material_silk_dmy      -0.5971     0.3749  -1.593  0.11214
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.487 on 338 degrees of freedom
# Multiple R-squared:  0.1394,   Adjusted R-squared:  0.1063
# F-statistic: 4.211 on 13 and 338 DF,  p-value: 0.000001614

totsale_lm.21 <- update(totsale_lm.20, ~. -Style_flare_dmy)
summary(totsale_lm.21)
# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                    Style_cute_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
```

```
#                     Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy,
#             data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7389 -0.7425  0.1764  0.8747  3.6263
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.6833     0.1138  67.522  < 2e-16 ***
# Pricehigh              -0.5049     0.3941  -1.281  0.20103
# Pricelow                0.4663     0.1766   2.640  0.00868 **
# Pricemedium            -1.0084     0.3624  -2.783  0.00569 **
# Style_cute_dmy         -0.5307     0.3183  -1.667  0.09642 .
# Style_vintage_dmy       0.5536     0.3632   1.524  0.12837
# Material_cashmere_dmy  -2.0164     0.8910  -2.263  0.02427 *
# Material_chiffon_dmy    0.6062     0.3980   1.523  0.12864
# Material_lycra_dmy     -2.0122     1.0564  -1.905  0.05765 .
# Material_modal_dmy     -1.6443     1.0564  -1.557  0.12050
# Material_nylon_dmy     -1.2524     0.6173  -2.029  0.04325 *
# Material_rayon_dmy     -0.6350     0.5386  -1.179  0.23926
# Material_silk_dmy      -0.5892     0.3750  -1.571  0.11700
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.488 on 339 degrees of freedom
# Multiple R-squared:  0.1362,   Adjusted R-squared:  0.1056
# F-statistic: 4.453 on 12 and 339 DF,  p-value: 0.000001218


totsale_lm.22 <- update(totsale_lm.21, ~. -Material_rayon_dmy )
summary(totsale_lm.22)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                     Style_cute_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                     Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                     Material_nylon_dmy + Material_silk_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7252 -0.7511  0.1814  0.8884  3.6831
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.6696     0.1133  67.717  < 2e-16 ***
# Pricehigh              -0.5265     0.3939  -1.337  0.18226
# Pricelow                0.4695     0.1767   2.657  0.00826 **
# Pricemedium            -0.9916     0.3623  -2.737  0.00653 **
# Style_cute_dmy         -0.5738     0.3164  -1.814  0.07061 .
# Style_vintage_dmy       0.5292     0.3628   1.458  0.14565
# Material_cashmere_dmy  -2.0139     0.8915  -2.259  0.02452 *
# Material_chiffon_dmy    0.6287     0.3978   1.581  0.11491
# Material_lycra_dmy     -2.0001     1.0569  -1.892  0.05929 .
# Material_modal_dmy     -1.6322     1.0569  -1.544  0.12343
# Material_nylon_dmy     -1.2367     0.6175  -2.003  0.04599 *
# Material_silk_dmy      -0.5677     0.3747  -1.515  0.13069
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.489 on 340 degrees of freedom
# Multiple R-squared:  0.1326,   Adjusted R-squared:  0.1046
# F-statistic: 4.726 on 11 and 340 DF,  p-value: 0.0000009445

totsale_lm.23 <- update(totsale_lm.22, ~. -Pricehigh )
summary(totsale_lm.23)
# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                     Style_vintage_dmy + Material_cashmere_dmy + Material_chiffon_dmy
+
#                     Material_lycra_dmy + Material_modal_dmy + Material_nylon_dmy +
#                     Material_silk_dmy, data = Training1s)
```

```
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6867 -0.7465  0.1697  0.9067  3.7614
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.6311     0.1097  69.587  < 2e-16 ***
# Pricelow                0.5155     0.1735   2.971  0.00318 **
# Pricemedium            -0.9462     0.3611  -2.620  0.00918 **
# Style_cute_dmy         -0.6136     0.3153  -1.946  0.05250 .
# Style_vintage_dmy       0.5536     0.3628   1.526  0.12793
# Material_cashmere_dmy  -2.0057     0.8925  -2.247  0.02527 *
# Material_chiffon_dmy    0.6520     0.3978   1.639  0.10215
# Material_lycra_dmy     -1.9846     1.0581  -1.876  0.06156 .
# Material_modal_dmy     -1.6168     1.0581  -1.528  0.12743
# Material_nylon_dmy     -1.3089     0.6158  -2.126  0.03426 *
# Material_silk_dmy      -0.6172     0.3733  -1.653  0.09921 .
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.49 on 341 degrees of freedom
# Multiple R-squared:  0.1281,   Adjusted R-squared:  0.1025
# F-statistic: 5.008 on 10 and 341 DF,  p-value: 0.0000008435

totsale_lm.24 <- update(totsale_lm.23, ~. -Style_vintage_dmy)
summary(totsale_lm.24)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                    Material_modal_dmy + Material_nylon_dmy + Material_silk_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7239 -0.7327  0.1483  0.9165  3.7544
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)             7.6683     0.1071  71.578  < 2e-16 ***
# Pricelow                0.5011     0.1736   2.887  0.00414 **
# Pricemedium            -0.9470     0.3618  -2.617  0.00925 **
# Style_cute_dmy         -0.6438     0.3153  -2.042  0.04194 *
# Material_cashmere_dmy  -2.0423     0.8940  -2.285  0.02295 *
# Material_chiffon_dmy    0.6645     0.3985   1.667  0.09635 .
# Material_lycra_dmy     -2.0146     1.0599  -1.901  0.05819 .
# Material_modal_dmy     -1.6468     1.0599  -1.554  0.12119
# Material_nylon_dmy     -1.3389     0.6167  -2.171  0.03060 *
# Material_silk_dmy      -0.6372     0.3738  -1.705  0.08919 .
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.493 on 342 degrees of freedom
# Multiple R-squared:  0.1221,   Adjusted R-squared:  0.099
# F-statistic: 5.285 on 9 and 342 DF,  p-value: 0.0000009274

totsale_lm.25 <- update(totsale_lm.24, ~. -Material_modal_dmy)
summary(totsale_lm.25)
# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                    Material_nylon_dmy + Material_silk_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7146 -0.7659  0.1560  0.9231  3.7557
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
```

```
# (Intercept)            7.6590     0.1072  71.456  < 2e-16 ***
# Pricelow               0.4942     0.1739   2.842  0.00476 **
# Pricemedium           -0.9402     0.3625  -2.593  0.00991 **
# Style_cute_dmy        -0.6358     0.3159  -2.012  0.04496 *
# Material_cashmere_dmy  -2.0376    0.8958  -2.275  0.02355 *
# Material_chiffon_dmy    0.6754    0.3993   1.692  0.09163 .
# Material_lycra_dmy     -2.0018    1.0621  -1.885  0.06030 .
# Material_nylon_dmy     -1.3262    0.6179  -2.146  0.03255 *
# Material_silk_dmy      -0.6254    0.3745  -1.670  0.09584 .
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.496 on 343 degrees of freedom
# Multiple R-squared:  0.1159,   Adjusted R-squared:  0.09529
# F-statistic: 5.621 on 8 and 343 DF,  p-value: 0.000001035

totsale_lm.26 <- update(totsale_lm.25, ~. -Material_silk_dmy)
summary(totsale_lm.26)
# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                    Material_nylon_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7078 -0.7424  0.1140  0.9058  3.8365
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.6523     0.1074  71.258  < 2e-16 ***
# Pricelow               0.4363     0.1709   2.554  0.01109 *
# Pricemedium           -0.9927     0.3621  -2.742  0.00643 **
# Style_cute_dmy        -0.7099     0.3136  -2.264  0.02422 *
# Material_cashmere_dmy  -1.9958    0.8978  -2.223  0.02686 *
# Material_chiffon_dmy    0.7240    0.3993   1.813  0.07064 .
# Material_lycra_dmy     -1.9661    1.0646  -1.847  0.06564 .
# Material_nylon_dmy     -1.2905    0.6191  -2.084  0.03786 *
#         ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.5 on 344 degrees of freedom
# Multiple R-squared:  0.1087,   Adjusted R-squared:  0.09059
# F-statistic: 5.995 on 7 and 344 DF,  p-value: 0.000001328

totsale_lm.27 <- update(totsale_lm.26, ~. -Material_chiffon_dmy)
summary(totsale_lm.27)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Material_cashmere_dmy + Material_lycra_dmy + Material_nylon_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7321 -0.7137  0.0878  0.9473  3.7521
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.6766     0.1069  71.810  < 2e-16 ***
#         Pricelow               0.4498     0.1713   2.627  0.00901 **
#         Pricemedium           -1.0226     0.3629  -2.818  0.00512 **
#         Style_cute_dmy        -0.6498     0.3129  -2.077  0.03857 *
#         Material_cashmere_dmy  -2.0002    0.9007  -2.221  0.02703 *
#         Material_lycra_dmy     -1.9972    1.0680  -1.870  0.06233 .
#         Material_nylon_dmy     -1.3216    0.6209  -2.128  0.03402 *
#         ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.505 on 345 degrees of freedom
# Multiple R-squared:  0.1002,   Adjusted R-squared:  0.08455
# F-statistic: 6.403 on 6 and 345 DF,  p-value: 0.000002074
```

```
totsale_lm.28 <- update(totsale_lm.27, ~. -Material_lycra_dmy)
summary(totsale_lm.28)
# Call:
#        lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                   Material_cashmere_dmy + Material_nylon_dmy, data = Training1s)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.7215 -0.7073  0.0895  0.9601  3.7504
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)            7.6660     0.1071  71.554  < 2e-16 ***
#        Pricelow              0.4430     0.1718   2.578  0.01036 *
#        Pricemedium          -1.0137     0.3642  -2.783  0.00567 **
#        Style_cute_dmy       -0.6374     0.3140  -2.030  0.04309 *
#        Material_cashmere_dmy  -1.9955     0.9040  -2.207  0.02794 *
#        Material_nylon_dmy     -1.3075     0.6231  -2.098  0.03660 *
#        ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.511 on 346 degrees of freedom
# Multiple R-squared:  0.09108,  Adjusted R-squared:  0.07795
# F-statistic: 6.934 on 5 and 346 DF,  p-value: 0.000003458


#Even though we did 28 steps of linear regression, based on Adjusted R-squared,
# ~ 0.1068 for totsale_lm.19, is the max value. hence that needs to be used as our
model in terms of a better prediction
# even despite some insignificant factors in the model

# we can try one more additional run where we consider 2-level interactions of the run
19 main effects to see if we can better the model further
totsale_lm.19_b <- lm(formula = log(TotalSales) ~ (Pricehigh + Pricelow + Pricemedium
+
                              Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_vintage_dmy +
                              Material_cashmere_dmy + Material_chiffon_dmy +
Material_lycra_dmy +
                              Material_modal_dmy + Material_nylon_dmy +
Material_rayon_dmy +
                              Material_silk_dmy)^2, data = Training1s)

summary(totsale_lm.19_b) # looking at the results we are going to remove NA relations
# Call:
#        lm(formula = log(TotalSales) ~ (Pricehigh + Pricelow + Pricemedium +
#                                              Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy + Style_vintage_dmy +
#                                              Material_cashmere_dmy +
Material_chiffon_dmy + Material_lycra_dmy +
#                                              Material_modal_dmy +
Material_nylon_dmy + Material_rayon_dmy +
#                                              Material_silk_dmy)^2, data =
Training1s)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.9230 -0.5997  0.0340  0.8567  3.1675
#
# Coefficients: (70 not defined because of singularities)
# Estimate Std. Error t value Pr(>|t|)
# (Intercept)                           7.62058    0.11728  64.978  < 2e-16 ***
#        Pricehigh                             -0.08152    0.44956  -0.181
0.85622
# Pricelow                              0.54017    0.19515   2.768  0.00597 **
#        Pricemedium                           -0.97552    0.40315  -2.420
0.01609 *
#        Style_cute_dmy                         0.43795    0.51223   0.855
0.39321
# Style_fashion_dmy                    -1.56614    1.46230  -1.071  0.28498
# Style_flare_dmy                      -0.42613    2.06134  -0.207  0.83636
```

```
# Style_vintage_dmy                          0.82813    0.45486   1.821  0.06961 .
# Material_cashmere_dmy                     -3.88291    1.46230  -2.655  0.00832 **
#        Material_chiffon_dmy                         1.22057    0.62994   1.938
0.05356 .
# Material_lycra_dmy                        -3.59523    1.46230  -2.459  0.01448 *
#        Material_modal_dmy                          -1.03994    1.46230  -0.711
0.47750
# Material_nylon_dmy                         0.75781    1.46230   0.518  0.60466
# Material_rayon_dmy                        -1.53258    0.84967  -1.804  0.07223 .
# Material_silk_dmy                          0.92946    1.46230   0.636  0.52549
# Pricehigh:Pricelow                              NA         NA      NA       NA
# Pricehigh:Pricemedium                           NA         NA      NA       NA
# Pricehigh:Style_cute_dmy                  -1.55162    1.37305  -1.130  0.25931
# Pricehigh:Style_fashion_dmy                     NA         NA      NA       NA
# Pricehigh:Style_flare_dmy                       NA         NA      NA       NA
# Pricehigh:Style_vintage_dmy                     NA         NA      NA       NA
# Pricehigh:Material_cashmere_dmy                 NA         NA      NA       NA
# Pricehigh:Material_chiffon_dmy                  NA         NA      NA       NA
# Pricehigh:Material_lycra_dmy                    NA         NA      NA       NA
# Pricehigh:Material_modal_dmy                    NA         NA      NA       NA
# Pricehigh:Material_nylon_dmy              -4.89567    2.10979  -2.320  0.02095 *
#        Pricehigh:Material_rayon_dmy                -1.01067    2.45759  -0.411
0.68117
# Pricehigh:Material_silk_dmy               -1.16902    1.94204  -0.602  0.54763
# Pricelow:Pricemedium                            NA         NA      NA       NA
# Pricelow:Style_cute_dmy                   -1.50111    0.73678  -2.037  0.04244 *
#        Pricelow:Style_fashion_dmy                     NA         NA      NA
NA
# Pricelow:Style_flare_dmy                        NA         NA      NA       NA
# Pricelow:Style_vintage_dmy                -0.53901    0.87314  -0.617  0.53746
# Pricelow:Material_cashmere_dmy                  NA         NA      NA       NA
# Pricelow:Material_chiffon_dmy             -0.66052    0.82013  -0.805  0.42120
# Pricelow:Material_lycra_dmy                3.21770    2.07056   1.554  0.12118
# Pricelow:Material_modal_dmy               -1.15723    2.07056  -0.559  0.57663
# Pricelow:Material_nylon_dmy               -2.33598    1.69435  -1.379  0.16897
# Pricelow:Material_rayon_dmy                2.92956    1.34482   2.178  0.03012 *
#        Pricelow:Material_silk_dmy                  -1.22624    1.53181  -0.801
0.42401
# Pricemedium:Style_cute_dmy                 1.61246    1.36560   1.181  0.23858
# Pricemedium:Style_fashion_dmy                   NA         NA      NA       NA
# Pricemedium:Style_flare_dmy                     NA         NA      NA       NA
# Pricemedium:Style_vintage_dmy             -1.22903    1.57487  -0.780  0.43574
# Pricemedium:Material_cashmere_dmy          2.86102    1.83013   1.563  0.11898
# Pricemedium:Material_chiffon_dmy                NA         NA      NA       NA
# Pricemedium:Material_lycra_dmy                  NA         NA      NA       NA
# Pricemedium:Material_modal_dmy                  NA         NA      NA       NA
# Pricemedium:Material_nylon_dmy                  NA         NA      NA       NA
# Pricemedium:Material_rayon_dmy                  NA         NA      NA       NA
# Pricemedium:Material_silk_dmy             -3.96883    1.93675  -2.049  0.04127 *
#        Style_cute_dmy:Style_fashion_dmy               NA         NA      NA
NA
# Style_cute_dmy:Style_flare_dmy                  NA         NA      NA       NA
# Style_cute_dmy:Style_vintage_dmy                NA         NA      NA       NA
# Style_cute_dmy:Material_cashmere_dmy            NA         NA      NA       NA
# Style_cute_dmy:Material_chiffon_dmy       -1.77200    1.04277  -1.699  0.09024 .
# Style_cute_dmy:Material_lycra_dmy               NA         NA      NA       NA
# Style_cute_dmy:Material_modal_dmy               NA         NA      NA       NA
# Style_cute_dmy:Material_nylon_dmy               NA         NA      NA       NA
# Style_cute_dmy:Material_rayon_dmy          0.77181    1.75930   0.439  0.66117
# Style_cute_dmy:Material_silk_dmy          -0.69925    1.03735  -0.674  0.50076
# Style_fashion_dmy:Style_flare_dmy               NA         NA      NA       NA
# Style_fashion_dmy:Style_vintage_dmy             NA         NA      NA       NA
# Style_fashion_dmy:Material_cashmere_dmy         NA         NA      NA       NA
# Style_fashion_dmy:Material_chiffon_dmy          NA         NA      NA       NA
# Style_fashion_dmy:Material_lycra_dmy            NA         NA      NA       NA
# Style_fashion_dmy:Material_modal_dmy            NA         NA      NA       NA
# Style_fashion_dmy:Material_nylon_dmy            NA         NA      NA       NA
# Style_fashion_dmy:Material_rayon_dmy            NA         NA      NA       NA
# Style_fashion_dmy:Material_silk_dmy             NA         NA      NA       NA
# Style_flare_dmy:Style_vintage_dmy               NA         NA      NA       NA
# Style_flare_dmy:Material_cashmere_dmy           NA         NA      NA       NA
# Style_flare_dmy:Material_chiffon_dmy            NA         NA      NA       NA
```

```
# Style_flare_dmy:Material_lycra_dmy              NA         NA      NA       NA
# Style_flare_dmy:Material_modal_dmy              NA         NA      NA       NA
# Style_flare_dmy:Material_nylon_dmy              NA         NA      NA       NA
# Style_flare_dmy:Material_rayon_dmy              NA         NA      NA       NA
# Style_flare_dmy:Material_silk_dmy               NA         NA      NA       NA
# Style_vintage_dmy:Material_cashmere_dmy         NA         NA      NA       NA
# Style_vintage_dmy:Material_chiffon_dmy    -0.53098    1.64758  -0.322  0.74745
# Style_vintage_dmy:Material_lycra_dmy            NA         NA      NA       NA
# Style_vintage_dmy:Material_modal_dmy            NA         NA      NA       NA
# Style_vintage_dmy:Material_nylon_dmy            NA         NA      NA       NA
# Style_vintage_dmy:Material_rayon_dmy       0.05927    1.74346   0.034  0.97290
# Style_vintage_dmy:Material_silk_dmy             NA         NA      NA       NA
# Material_cashmere_dmy:Material_chiffon_dmy      NA         NA      NA       NA
# Material_cashmere_dmy:Material_lycra_dmy        NA         NA      NA       NA
# Material_cashmere_dmy:Material_modal_dmy        NA         NA      NA       NA
# Material_cashmere_dmy:Material_nylon_dmy        NA         NA      NA       NA
# Material_cashmere_dmy:Material_rayon_dmy        NA         NA      NA       NA
# Material_cashmere_dmy:Material_silk_dmy         NA         NA      NA       NA
# Material_chiffon_dmy:Material_lycra_dmy         NA         NA      NA       NA
# Material_chiffon_dmy:Material_modal_dmy         NA         NA      NA       NA
# Material_chiffon_dmy:Material_nylon_dmy         NA         NA      NA       NA
# Material_chiffon_dmy:Material_rayon_dmy         NA         NA      NA       NA
# Material_chiffon_dmy:Material_silk_dmy          NA         NA      NA       NA
# Material_lycra_dmy:Material_modal_dmy           NA         NA      NA       NA
# Material_lycra_dmy:Material_nylon_dmy           NA         NA      NA       NA
# Material_lycra_dmy:Material_rayon_dmy           NA         NA      NA       NA
# Material_lycra_dmy:Material_silk_dmy            NA         NA      NA       NA
# Material_modal_dmy:Material_nylon_dmy           NA         NA      NA       NA
# Material_modal_dmy:Material_rayon_dmy           NA         NA      NA       NA
# Material_modal_dmy:Material_silk_dmy            NA         NA      NA       NA
# Material_nylon_dmy:Material_rayon_dmy           NA         NA      NA       NA
# Material_nylon_dmy:Material_silk_dmy            NA         NA      NA       NA
# Material_rayon_dmy:Material_silk_dmy            NA         NA      NA       NA
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.458 on 316 degrees of freedom
# Multiple R-squared:  0.227,    Adjusted R-squared:  0.1414
# F-statistic: 2.652 on 35 and 316 DF,  p-value: 0.000004122


totsale_lm.19_c <- lm(formula = log(TotalSales) ~
                              (Pricehigh + Pricelow + Pricemedium +
                              Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_vintage_dmy +
                              Material_cashmere_dmy + Material_chiffon_dmy +
Material_lycra_dmy +
                              Material_modal_dmy + Material_nylon_dmy +
Material_rayon_dmy +
                              Material_silk_dmy) +
                              Pricehigh:Material_nylon_dmy + Pricehigh:Style_cute_dmy
+
                              Pricehigh:Material_rayon_dmy +
Pricehigh:Material_silk_dmy +
                              Pricelow:Style_cute_dmy + Pricelow:Style_vintage_dmy +
                              Pricelow:Material_chiffon_dmy +
Pricelow:Material_lycra_dmy +
                              Pricelow:Material_modal_dmy +
Pricelow:Material_nylon_dmy +
                      Pricelow:Material_rayon_dmy + Pricelow:Material_silk_dmy +
                      Pricemedium:Style_cute_dmy + Pricemedium:Style_vintage_dmy +
                      Pricemedium:Material_cashmere_dmy +
Pricemedium:Material_silk_dmy +
                      Style_cute_dmy:Material_chiffon_dmy +
Style_cute_dmy:Material_rayon_dmy +
                      Style_cute_dmy:Material_silk_dmy +
Style_vintage_dmy:Material_chiffon_dmy +
                      Style_vintage_dmy:Material_rayon_dmy
                      , data = Training1s)

# Call:
#         lm(formula = log(TotalSales) ~ (Pricehigh + Pricelow + Pricemedium +
```

```
#                                                   Style_cute_dmy + Style_fashion_dmy +
Style_flare_dmy + Style_vintage_dmy +
#                                                   Material_cashmere_dmy +
Material_chiffon_dmy + Material_lycra_dmy +
#                                                   Material_modal_dmy +
Material_nylon_dmy + Material_rayon_dmy +
#                                                   Material_silk_dmy) +
Pricehigh:Material_nylon_dmy + Pricehigh:Style_cute_dmy +
#                     Pricehigh:Material_rayon_dmy + Pricehigh:Material_silk_dmy +
#                     Pricelow:Style_cute_dmy + Pricelow:Style_vintage_dmy +
Pricelow:Material_chiffon_dmy +
#                     Pricelow:Material_lycra_dmy + Pricelow:Material_modal_dmy +
#                     Pricelow:Material_nylon_dmy + Pricelow:Material_rayon_dmy +
#                     Pricelow:Material_silk_dmy + Pricemedium:Style_cute_dmy +
#                     Pricemedium:Style_vintage_dmy + Pricemedium:Material_cashmere_dmy
+
#                     Pricemedium:Material_silk_dmy +
Style_cute_dmy:Material_chiffon_dmy +
#                     Style_cute_dmy:Material_rayon_dmy +
Style_cute_dmy:Material_silk_dmy +
#                     Style_vintage_dmy:Material_chiffon_dmy +
Style_vintage_dmy:Material_rayon_dmy,
#             data = Training1s)
#
# Residuals:
#          Min      1Q  Median      3Q     Max
# -4.9230 -0.5997  0.0340  0.8567  3.1675
#
# Coefficients:
#          Estimate Std. Error t value Pr(>|t|)
# (Intercept)                            7.62058    0.11728  64.978  < 2e-16 ***
# Pricehigh                             -0.08152    0.44956  -0.181  0.85622
# Pricelow                               0.54017    0.19515   2.768  0.00597 **
# Pricemedium                           -0.97552    0.40315  -2.420  0.01609 *
# Style_cute_dmy                         0.43795    0.51223   0.855  0.39321
# Style_fashion_dmy                     -1.56614    1.46230  -1.071  0.28498
# Style_flare_dmy                       -0.42613    2.06134  -0.207  0.83636
# Style_vintage_dmy                      0.82813    0.45486   1.821  0.06961 .
# Material_cashmere_dmy                 -3.88291    1.46230  -2.655  0.00832 **
# Material_chiffon_dmy                   1.22057    0.62994   1.938  0.05356 .
# Material_lycra_dmy                    -3.59523    1.46230  -2.459  0.01448 *
# Material_modal_dmy                    -1.03994    1.46230  -0.711  0.47750
# Material_nylon_dmy                     0.75781    1.46230   0.518  0.60466
# Material_rayon_dmy                    -1.53258    0.84967  -1.804  0.07223 .
# Material_silk_dmy                      0.92946    1.46230   0.636  0.52549
# Pricehigh:Material_nylon_dmy          -4.89567    2.10979  -2.320  0.02095 *
# Pricehigh:Style_cute_dmy              -1.55162    1.37305  -1.130  0.25931
# Pricehigh:Material_rayon_dmy          -1.01067    2.45759  -0.411  0.68117
# Pricehigh:Material_silk_dmy           -1.16902    1.94204  -0.602  0.54763
# Pricelow:Style_cute_dmy               -1.50111    0.73678  -2.037  0.04244 *
# Pricelow:Style_vintage_dmy            -0.53901    0.87314  -0.617  0.53746
# Pricelow:Material_chiffon_dmy         -0.66052    0.82013  -0.805  0.42120
# Pricelow:Material_lycra_dmy            3.21770    2.07056   1.554  0.12118
# Pricelow:Material_modal_dmy           -1.15723    2.07056  -0.559  0.57663
# Pricelow:Material_nylon_dmy           -2.33598    1.69435  -1.379  0.16897
# Pricelow:Material_rayon_dmy            2.92956    1.34482   2.178  0.03012 *
# Pricelow:Material_silk_dmy            -1.22624    1.53181  -0.801  0.42401
# Pricemedium:Style_cute_dmy             1.61246    1.36560   1.181  0.23858
# Pricemedium:Style_vintage_dmy         -1.22903    1.57487  -0.780  0.43574
# Pricemedium:Material_cashmere_dmy      2.86102    1.83013   1.563  0.11898
# Pricemedium:Material_silk_dmy         -3.96883    1.93675  -2.049  0.04127 *
# Style_cute_dmy:Material_chiffon_dmy   -1.77200    1.04277  -1.699  0.09024 .
# Style_cute_dmy:Material_rayon_dmy      0.77181    1.75930   0.439  0.66117
# Style_cute_dmy:Material_silk_dmy      -0.69925    1.03735  -0.674  0.50076
# Style_vintage_dmy:Material_chiffon_dmy -0.53098    1.64758  -0.322  0.74745
# Style_vintage_dmy:Material_rayon_dmy   0.05927    1.74346   0.034  0.97290
# ---
#          Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.458 on 316 degrees of freedom
# Multiple R-squared:  0.227,    Adjusted R-squared:  0.1414
# F-statistic: 2.652 on 35 and 316 DF,  p-value: 0.000004122
```

```
totsale_lm.19_c.1 <- update(totsale_lm.19_c, ~. -Style_vintage_dmy:Material_rayon_dmy)

# Call:
#         lm(formula = log(TotalSales) ~ Pricehigh + Pricelow + Pricemedium +
#                     Style_cute_dmy + Style_fashion_dmy + Style_flare_dmy +
Style_vintage_dmy +
#                     Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                     Material_modal_dmy + Material_nylon_dmy + Material_rayon_dmy +
#                     Material_silk_dmy + Pricehigh:Material_nylon_dmy +
Pricehigh:Style_cute_dmy +
#                     Pricehigh:Material_rayon_dmy + Pricehigh:Material_silk_dmy +
#                     Pricelow:Style_cute_dmy + Pricelow:Style_vintage_dmy +
Pricelow:Material_chiffon_dmy +
#                     Pricelow:Material_lycra_dmy + Pricelow:Material_modal_dmy +
#                     Pricelow:Material_nylon_dmy + Pricelow:Material_rayon_dmy +
#                     Pricelow:Material_silk_dmy + Pricemedium:Style_cute_dmy +
#                     Pricemedium:Style_vintage_dmy + Pricemedium:Material_cashmere_dmy
+
#                     Pricemedium:Material_silk_dmy +
Style_cute_dmy:Material_chiffon_dmy +
#                     Style_cute_dmy:Material_rayon_dmy +
Style_cute_dmy:Material_silk_dmy +
#                     Style_vintage_dmy:Material_chiffon_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9230 -0.5996  0.0412  0.8568  3.1678
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                             7.62032    0.11683  65.226  < 2e-16 ***
# Pricehigh                              -0.08125    0.44878  -0.181  0.85644
# Pricelow                                0.54044    0.19469   2.776  0.00583 **
# Pricemedium                            -0.97525    0.40243  -2.423  0.01594 *
# Style_cute_dmy                          0.43820    0.51137   0.857  0.39214
# Style_fashion_dmy                      -1.56588    1.45997  -1.073  0.28429
# Style_flare_dmy                        -0.42613    2.05809  -0.207  0.83610
# Style_vintage_dmy                       0.83217    0.43841   1.898  0.05859 .
# Material_cashmere_dmy                  -3.88265    1.45997  -2.659  0.00823 **
# Material_chiffon_dmy                    1.22081    0.62890   1.941  0.05312 .
# Material_lycra_dmy                     -3.59496    1.45997  -2.462  0.01433 *
# Material_modal_dmy                     -1.03968    1.45997  -0.712  0.47691
# Material_nylon_dmy                      0.75807    1.45997   0.519  0.60396
# Material_rayon_dmy                     -1.51850    0.74077  -2.050  0.04120 *
# Material_silk_dmy                       0.92973    1.45997   0.637  0.52470
# Pricehigh:Material_nylon_dmy           -4.89594    2.10645  -2.324  0.02074 *
# Pricehigh:Style_cute_dmy               -1.55186    1.37087  -1.132  0.25848
# Pricehigh:Material_rayon_dmy           -1.01069    2.45371  -0.412  0.68069
# Pricehigh:Material_silk_dmy            -1.16929    1.93896  -0.603  0.54691
# Pricelow:Style_cute_dmy                -1.50134    0.73558  -2.041  0.04208 *
# Pricelow:Style_vintage_dmy             -0.54304    0.86368  -0.629  0.52996
# Pricelow:Material_chiffon_dmy          -0.66074    0.81881  -0.807  0.42030
# Pricelow:Material_lycra_dmy             3.21743    2.06728   1.556  0.12062
# Pricelow:Material_modal_dmy            -1.15750    2.06728  -0.560  0.57593
# Pricelow:Material_nylon_dmy            -2.33625    1.69166  -1.381  0.16824
# Pricelow:Material_rayon_dmy             2.91548    1.27748   2.282  0.02314 *
# Pricelow:Material_silk_dmy             -1.22651    1.52937  -0.802  0.42317
# Pricemedium:Style_cute_dmy              1.61222    1.36343   1.182  0.23790
# Pricemedium:Style_vintage_dmy          -1.23307    1.56792  -0.786  0.43220
# Pricemedium:Material_cashmere_dmy       2.86076    1.82723   1.566  0.11843
# Pricemedium:Material_silk_dmy          -3.96910    1.93368  -2.053  0.04093 *
# Style_cute_dmy:Material_chiffon_dmy    -1.77217    1.04111  -1.702  0.08970 .
# Style_cute_dmy:Material_rayon_dmy       0.75775    1.70729   0.444  0.65747
# Style_cute_dmy:Material_silk_dmy       -0.69927    1.03572  -0.675  0.50007
# Style_vintage_dmy:Material_chiffon_dmy -0.53499    1.64076  -0.326  0.74459
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.455 on 317 degrees of freedom
# Multiple R-squared:  0.227,    Adjusted R-squared:  0.1441
```

```
# F-statistic: 2.738 on 34 and 317 DF,  p-value: 0.00000243

totsale_lm.19_c.2 <- update(totsale_lm.19_c.1, ~. - Pricehigh)
summary(totsale_lm.19_c.2)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Style_fashion_dmy + Style_flare_dmy + Style_vintage_dmy +
#                    Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                    Material_modal_dmy + Material_nylon_dmy + Material_rayon_dmy +
#                    Material_silk_dmy + Pricehigh:Material_nylon_dmy +
Pricehigh:Style_cute_dmy +
#                    Pricehigh:Material_rayon_dmy + Pricehigh:Material_silk_dmy +
#                    Pricelow:Style_cute_dmy + Pricelow:Style_vintage_dmy +
Pricelow:Material_chiffon_dmy +
#                    Pricelow:Material_lycra_dmy + Pricelow:Material_modal_dmy +
#                    Pricelow:Material_nylon_dmy + Pricelow:Material_rayon_dmy +
#                    Pricelow:Material_silk_dmy + Pricemedium:Style_cute_dmy +
#                    Pricemedium:Style_vintage_dmy + Pricemedium:Material_cashmere_dmy
+
#                    Pricemedium:Material_silk_dmy +
Style_cute_dmy:Material_chiffon_dmy +
#                    Style_cute_dmy:Material_rayon_dmy +
Style_cute_dmy:Material_silk_dmy +
#                    Style_vintage_dmy:Material_chiffon_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9225 -0.5984  0.0334  0.8580  3.1733
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                             7.6148     0.1126  67.602  < 2e-16 ***
# Pricelow                                0.5461     0.1919   2.846  0.00471 **
# Pricemedium                            -0.9695     0.4006  -2.420  0.01607 *
# Style_cute_dmy                          0.4432     0.5098   0.869  0.38534
# Style_fashion_dmy                      -1.5604     1.4574  -1.071  0.28515
# Style_flare_dmy                        -0.4261     2.0550  -0.207  0.83586
# Style_vintage_dmy                       0.8373     0.4368   1.917  0.05616 .
# Material_cashmere_dmy                  -3.8771     1.4574  -2.660  0.00820 **
# Material_chiffon_dmy                    1.2255     0.6274   1.953  0.05166 .
# Material_lycra_dmy                     -3.5895     1.4574  -2.463  0.01431 *
# Material_modal_dmy                     -1.0342     1.4574  -0.710  0.47848
# Material_nylon_dmy                      0.7636     1.4574   0.524  0.60070
# Material_rayon_dmy                     -1.5143     0.7393  -2.048  0.04135 *
# Material_silk_dmy                       0.9352     1.4574   0.642  0.52153
# Material_nylon_dmy:Pricehigh           -4.9772     2.0550  -2.422  0.01599 *
# Style_cute_dmy:Pricehigh               -1.6115     1.3287  -1.213  0.22607
# Material_rayon_dmy:Pricehigh           -1.0323     2.4471  -0.422  0.67342
# Material_silk_dmy:Pricehigh            -1.2294     1.9074  -0.645  0.51968
# Pricelow:Style_cute_dmy                -1.5087     0.7333  -2.057  0.04048 *
# Pricelow:Style_vintage_dmy             -0.5483     0.8619  -0.636  0.52508
# Pricelow:Material_chiffon_dmy          -0.6650     0.8172  -0.814  0.41645
# Pricelow:Material_lycra_dmy             3.2118     2.0639   1.556  0.12066
# Pricelow:Material_modal_dmy            -1.1631     2.0639  -0.564  0.57344
# Pricelow:Material_nylon_dmy            -2.3419     1.6888  -1.387  0.16650
# Pricelow:Material_rayon_dmy             2.9111     1.2753   2.283  0.02311 *
# Pricelow:Material_silk_dmy             -1.2337     1.5265  -0.808  0.41959
# Pricemedium:Style_cute_dmy              1.6029     1.3604   1.178  0.23957
# Pricemedium:Style_vintage_dmy          -1.2385     1.5652  -0.791  0.42939
# Pricemedium:Material_cashmere_dmy       2.8550     1.8242   1.565  0.11856
# Pricemedium:Material_silk_dmy          -3.9789     1.9300  -2.062  0.04005 *
# Style_cute_dmy:Material_chiffon_dmy    -1.7744     1.0395  -1.707  0.08879 .
# Style_cute_dmy:Material_rayon_dmy       0.7540     1.7046   0.442  0.65853
# Style_cute_dmy:Material_silk_dmy       -0.6868     1.0319  -0.666  0.50613
# Style_vintage_dmy:Material_chiffon_dmy -0.5394     1.6381  -0.329  0.74218
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.453 on 318 degrees of freedom
# Multiple R-squared:  0.2269,   Adjusted R-squared:  0.1467
```

```
# F-statistic: 2.829 on 33 and 318 DF,  p-value: 0.000001425

totsale_lm.19_c.3 <- update(totsale_lm.19_c.2, ~. -Style_flare_dmy)
summary(totsale_lm.19_c.3)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                    Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy +
#                    Material_nylon_dmy:Pricehigh + Style_cute_dmy:Pricehigh +
#                    Material_rayon_dmy:Pricehigh + Material_silk_dmy:Pricehigh +
#                    Pricelow:Style_cute_dmy + Pricelow:Style_vintage_dmy +
Pricelow:Material_chiffon_dmy +
#                    Pricelow:Material_lycra_dmy + Pricelow:Material_modal_dmy +
#                    Pricelow:Material_nylon_dmy + Pricelow:Material_rayon_dmy +
#                    Pricelow:Material_silk_dmy + Pricemedium:Style_cute_dmy +
#                    Pricemedium:Style_vintage_dmy + Pricemedium:Material_cashmere_dmy
+
#                    Pricemedium:Material_silk_dmy +
Style_cute_dmy:Material_chiffon_dmy +
#                    Style_cute_dmy:Material_rayon_dmy +
Style_cute_dmy:Material_silk_dmy +
#                    Style_vintage_dmy:Material_chiffon_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9225 -0.5984  0.0416  0.8580  3.1733
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                            7.6148     0.1125  67.703  < 2e-16 ***
# Pricelow                               0.5461     0.1916   2.850  0.00465 **
# Pricemedium                           -0.9695     0.4000  -2.424  0.01591 *
# Style_cute_dmy                         0.4432     0.5091   0.871  0.38463
# Style_fashion_dmy                     -1.5604     1.4552  -1.072  0.28442
# Style_vintage_dmy                      0.8373     0.4362   1.920  0.05579 .
# Material_cashmere_dmy                 -3.8771     1.4552  -2.664  0.00811 **
# Material_chiffon_dmy                   1.2255     0.6265   1.956  0.05131 .
# Material_lycra_dmy                    -3.5895     1.4552  -2.467  0.01417 *
# Material_modal_dmy                    -1.0342     1.4552  -0.711  0.47782
# Material_nylon_dmy                     0.5505     1.0321   0.533  0.59413
# Material_rayon_dmy                    -1.5143     0.7382  -2.051  0.04104 *
# Material_silk_dmy                      0.9352     1.4552   0.643  0.52091
# Material_nylon_dmy:Pricehigh          -4.7641     1.7770  -2.681  0.00772 **
# Style_cute_dmy:Pricehigh              -1.6115     1.3267  -1.215  0.22538
# Material_rayon_dmy:Pricehigh          -1.0323     2.4434  -0.422  0.67296
# Material_silk_dmy:Pricehigh           -1.2294     1.9045  -0.646  0.51905
# Pricelow:Style_cute_dmy               -1.5087     0.7322  -2.060  0.04017 *
# Pricelow:Style_vintage_dmy            -0.5483     0.8606  -0.637  0.52446
# Pricelow:Material_chiffon_dmy         -0.6650     0.8160  -0.815  0.41575
# Pricelow:Material_lycra_dmy            3.2118     2.0608   1.559  0.12010
# Pricelow:Material_modal_dmy           -1.1631     2.0608  -0.564  0.57287
# Pricelow:Material_nylon_dmy           -2.1288     1.3383  -1.591  0.11266
# Pricelow:Material_rayon_dmy            2.9111     1.2734   2.286  0.02290 *
# Pricelow:Material_silk_dmy            -1.2337     1.5242  -0.809  0.41889
# Pricemedium:Style_cute_dmy             1.6029     1.3583   1.180  0.23887
# Pricemedium:Style_vintage_dmy         -1.2385     1.5629  -0.792  0.42869
# Pricemedium:Material_cashmere_dmy      2.8550     1.8214   1.567  0.11800
# Pricemedium:Material_silk_dmy         -3.9789     1.9271  -2.065  0.03975 *
# Style_cute_dmy:Material_chiffon_dmy   -1.7744     1.0379  -1.710  0.08831 .
# Style_cute_dmy:Material_rayon_dmy      0.7540     1.7020   0.443  0.65805
# Style_cute_dmy:Material_silk_dmy      -0.6868     1.0303  -0.667  0.50549
# Style_vintage_dmy:Material_chiffon_dmy  -0.5394     1.6356  -0.330  0.74180
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.451 on 319 degrees of freedom
# Multiple R-squared:  0.2268,   Adjusted R-squared:  0.1493
# F-statistic: 2.925 on 32 and 319 DF,  p-value: 0.0000008248
```

```
totsale_lm.19_c.4 <- update(totsale_lm.19_c.3, ~. -
Style_vintage_dmy:Material_chiffon_dmy)
summary(totsale_lm.19_c.4)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                    Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy +
#                    Material_nylon_dmy:Pricehigh + Style_cute_dmy:Pricehigh +
#                    Material_rayon_dmy:Pricehigh + Material_silk_dmy:Pricehigh +
#                    Pricelow:Style_cute_dmy + Pricelow:Style_vintage_dmy +
Pricelow:Material_chiffon_dmy +
#                    Pricelow:Material_lycra_dmy + Pricelow:Material_modal_dmy +
#                    Pricelow:Material_nylon_dmy + Pricelow:Material_rayon_dmy +
#                    Pricelow:Material_silk_dmy + Pricemedium:Style_cute_dmy +
#                    Pricemedium:Style_vintage_dmy + Pricemedium:Material_cashmere_dmy
+
#                    Pricemedium:Material_silk_dmy +
Style_cute_dmy:Material_chiffon_dmy +
#                    Style_cute_dmy:Material_rayon_dmy +
Style_cute_dmy:Material_silk_dmy,
#             data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9276 -0.5993  0.0445  0.8572  3.1711
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                         7.6170     0.1121  67.932  < 2e-16 ***
# Pricelow                            0.5444     0.1913   2.846  0.00471 **
# Pricemedium                        -0.9715     0.3993  -2.433  0.01553 *
# Style_cute_dmy                      0.4461     0.5083   0.878  0.38081
# Style_fashion_dmy                  -1.5625     1.4532  -1.075  0.28308
# Style_vintage_dmy                   0.7992     0.4200   1.903  0.05795 .
# Material_cashmere_dmy              -3.8793     1.4532  -2.669  0.00798 **
# Material_chiffon_dmy                1.1466     0.5781   1.983  0.04818 *
# Material_lycra_dmy                 -3.5916     1.4532  -2.472  0.01397 *
# Material_modal_dmy                 -1.0363     1.4532  -0.713  0.47628
# Material_nylon_dmy                  0.5483     1.0306   0.532  0.59506
# Material_rayon_dmy                 -1.5069     0.7368  -2.045  0.04165 *
# Material_silk_dmy                   0.9331     1.4532   0.642  0.52128
# Material_nylon_dmy:Pricehigh       -4.7641     1.7745  -2.685  0.00764 **
# Style_cute_dmy:Pricehigh           -1.6176     1.3247  -1.221  0.22296
# Material_rayon_dmy:Pricehigh       -1.0263     2.4399  -0.421  0.67432
# Material_silk_dmy:Pricehigh        -1.2304     1.9019  -0.647  0.51812
# Pricelow:Style_cute_dmy            -1.5183     0.7306  -2.078  0.03851 *
# Pricelow:Style_vintage_dmy         -0.5106     0.8518  -0.600  0.54925
# Pricelow:Material_chiffon_dmy      -0.5931     0.7853  -0.755  0.45063
# Pricelow:Material_lycra_dmy         3.2135     2.0579   1.562  0.11939
# Pricelow:Material_modal_dmy        -1.1614     2.0579  -0.564  0.57290
# Pricelow:Material_nylon_dmy        -2.1271     1.3364  -1.592  0.11244
# Pricelow:Material_rayon_dmy         2.9033     1.2714   2.284  0.02305 *
# Pricelow:Material_silk_dmy         -1.2317     1.5221  -0.809  0.41898
# Pricemedium:Style_cute_dmy          1.5982     1.3564   1.178  0.23955
# Pricemedium:Style_vintage_dmy      -1.2005     1.5565  -0.771  0.44111
# Pricemedium:Material_cashmere_dmy   2.8570     1.8189   1.571  0.11723
# Pricemedium:Material_silk_dmy      -3.9786     1.9244  -2.067  0.03950 *
# Style_cute_dmy:Material_chiffon_dmy -1.7207    1.0236  -1.681  0.09374 .
# Style_cute_dmy:Material_rayon_dmy   0.7416     1.6992   0.436  0.66281
# Style_cute_dmy:Material_silk_dmy   -0.6817     1.0288  -0.663  0.50804
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.449 on 320 degrees of freedom
# Multiple R-squared:  0.2266,   Adjusted R-squared:  0.1517
# F-statistic: 3.024 on 31 and 320 DF,  p-value: 0.0000004805

totsale_lm.19_c.5 <- update(totsale_lm.19_c.4, ~. -Material_rayon_dmy:Pricehigh)
summary(totsale_lm.19_c.5)
```

```
# Call:
#        lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                   Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                   Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                   Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy +
#                   Material_nylon_dmy:Pricehigh + Style_cute_dmy:Pricehigh +
#                   Material_silk_dmy:Pricehigh + Pricelow:Style_cute_dmy +
Pricelow:Style_vintage_dmy +
#                   Pricelow:Material_chiffon_dmy + Pricelow:Material_lycra_dmy +
#                   Pricelow:Material_modal_dmy + Pricelow:Material_nylon_dmy +
#                   Pricelow:Material_rayon_dmy + Pricelow:Material_silk_dmy +
#                   Pricemedium:Style_cute_dmy + Pricemedium:Style_vintage_dmy +
#                   Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                   Style_cute_dmy:Material_chiffon_dmy +
Style_cute_dmy:Material_rayon_dmy +
#                   Style_cute_dmy:Material_silk_dmy, data = Training1s)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.9693 -0.5996  0.0484  0.8568  3.1716
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)                          7.6165     0.1120  68.019  < 2e-16 ***
# Pricelow                             0.5455     0.1910   2.856  0.00457 **
# Pricemedium                         -0.9697     0.3988  -2.431  0.01558 *
# Style_cute_dmy                       0.4883     0.4977   0.981  0.32728
# Style_fashion_dmy                   -1.5621     1.4513  -1.076  0.28260
# Style_vintage_dmy                    0.8000     0.4194   1.907  0.05738 .
# Material_cashmere_dmy               -3.8788     1.4513  -2.673  0.00791 **
# Material_chiffon_dmy                 1.1422     0.5772   1.979  0.04870 *
# Material_lycra_dmy                  -3.5912     1.4513  -2.474  0.01386 *
# Material_modal_dmy                  -1.0359     1.4513  -0.714  0.47591
# Material_nylon_dmy                   0.5488     1.0293   0.533  0.59427
# Material_rayon_dmy                  -1.5066     0.7359  -2.047  0.04143 *
# Material_silk_dmy                    0.9335     1.4513   0.643  0.52054
# Material_nylon_dmy:Pricehigh        -4.7641     1.7722  -2.688  0.00756 **
# Style_cute_dmy:Pricehigh            -1.9201     1.1110  -1.728  0.08492 .
# Material_silk_dmy:Pricehigh         -1.1293     1.8842  -0.599  0.54935
# Pricelow:Style_cute_dmy             -1.5701     0.7193  -2.183  0.02977 *
# Pricelow:Style_vintage_dmy          -0.5120     0.8507  -0.602  0.54764
# Pricelow:Material_chiffon_dmy       -0.5847     0.7840  -0.746  0.45636
# Pricelow:Material_lycra_dmy          3.2124     2.0553   1.563  0.11904
# Pricelow:Material_modal_dmy         -1.1625     2.0553  -0.566  0.57204
# Pricelow:Material_nylon_dmy         -2.1282     1.3347  -1.595  0.11180
# Pricelow:Material_rayon_dmy          2.9024     1.2698   2.286  0.02292 *
# Pricelow:Material_silk_dmy          -1.2403     1.5200  -0.816  0.41512
# Pricemedium:Style_cute_dmy           1.5357     1.3465   1.141  0.25491
# Pricemedium:Style_vintage_dmy       -1.2026     1.5545  -0.774  0.43971
# Pricemedium:Material_cashmere_dmy    2.8552     1.8165   1.572  0.11699
# Pricemedium:Material_silk_dmy       -3.9993     1.9213  -2.082  0.03817 *
# Style_cute_dmy:Material_chiffon_dmy -1.7439     1.0208  -1.708  0.08853 .
# Style_cute_dmy:Material_rayon_dmy    0.3378     1.4001   0.241  0.80953
# Style_cute_dmy:Material_silk_dmy    -0.6235     1.0181  -0.612  0.54068
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.447 on 321 degrees of freedom
# Multiple R-squared:  0.2262,   Adjusted R-squared:  0.1538
# F-statistic: 3.127 on 30 and 321 DF,  p-value: 0.000000282

totsale_lm.19_c.6 <- update(totsale_lm.19_c.5, ~. -Style_cute_dmy:Material_rayon_dmy)
summary(totsale_lm.19_c.6)

# Call:
#        lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                   Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                   Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                   Material_nylon_dmy + Material_rayon_dmy + Material_silk_dmy +
#                   Material_nylon_dmy:Pricehigh + Style_cute_dmy:Pricehigh +
```

```
#                  Material_silk_dmy:Pricehigh + Pricelow:Style_cute_dmy +
Pricelow:Style_vintage_dmy +
#                  Pricelow:Material_chiffon_dmy + Pricelow:Material_lycra_dmy +
#                  Pricelow:Material_modal_dmy + Pricelow:Material_nylon_dmy +
#                  Pricelow:Material_rayon_dmy + Pricelow:Material_silk_dmy +
#                  Pricemedium:Style_cute_dmy + Pricemedium:Style_vintage_dmy +
#                  Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                  Style_cute_dmy:Material_chiffon_dmy +
Style_cute_dmy:Material_silk_dmy,
#          data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9952 -0.5989  0.0464  0.8575  3.1730
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                         7.6151     0.1117  68.198  < 2e-16 ***
# Pricelow                            0.5464     0.1907   2.865  0.00444 **
# Pricemedium                        -0.9689     0.3982  -2.433  0.01551 *
# Style_cute_dmy                      0.5156     0.4839   1.066  0.28744
# Style_fashion_dmy                  -1.5607     1.4492  -1.077  0.28232
# Style_vintage_dmy                   0.7942     0.4181   1.899  0.05841 .
# Material_cashmere_dmy              -3.8774     1.4492  -2.676  0.00784 **
# Material_chiffon_dmy                1.1427     0.5764   1.982  0.04828 *
# Material_lycra_dmy                 -3.5898     1.4492  -2.477  0.01376 *
# Material_modal_dmy                 -1.0345     1.4492  -0.714  0.47585
# Material_nylon_dmy                  0.5502     1.0278   0.535  0.59278
# Material_rayon_dmy                 -1.4136     0.6258  -2.259  0.02455 *
# Material_silk_dmy                   0.9349     1.4492   0.645  0.51930
# Material_nylon_dmy:Pricehigh       -4.7641     1.7696  -2.692  0.00747 **
# Style_cute_dmy:Pricehigh           -1.8432     1.0629  -1.734  0.08383 .
# Material_silk_dmy:Pricehigh        -1.1686     1.8744  -0.623  0.53345
# Pricelow:Style_cute_dmy            -1.5899     0.7135  -2.228  0.02655 *
# Pricelow:Style_vintage_dmy         -0.5058     0.8490  -0.596  0.55179
# Pricelow:Material_chiffon_dmy      -0.5828     0.7829  -0.744  0.45715
# Pricelow:Material_lycra_dmy         3.2115     2.0523   1.565  0.11860
# Pricelow:Material_modal_dmy        -1.1634     2.0523  -0.567  0.57118
# Pricelow:Material_nylon_dmy        -2.1291     1.3327  -1.598  0.11112
# Pricelow:Material_rayon_dmy         2.8099     1.2087   2.325  0.02070 *
# Pricelow:Material_silk_dmy         -1.2384     1.5178  -0.816  0.41514
# Pricemedium:Style_cute_dmy          1.5174     1.3424   1.130  0.25916
# Pricemedium:Style_vintage_dmy      -1.1962     1.5520  -0.771  0.44140
# Pricemedium:Material_cashmere_dmy   2.8544     1.8139   1.574  0.11655
# Pricemedium:Material_silk_dmy      -3.9917     1.9182  -2.081  0.03823 *
# Style_cute_dmy:Material_chiffon_dmy -1.7647    1.0157  -1.737  0.08326 .
# Style_cute_dmy:Material_silk_dmy   -0.6493     1.0110  -0.642  0.52121
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.445 on 322 degrees of freedom
# Multiple R-squared:  0.226,    Adjusted R-squared:  0.1563
# F-statistic: 3.242 on 29 and 322 DF,  p-value: 0.000000156

totsale_lm.19_c.7 <- update(totsale_lm.19_c.6, ~. -Material_nylon_dmy)
summary(totsale_lm.19_c.7)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                  Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                  Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                  Material_rayon_dmy + Material_silk_dmy +
Material_nylon_dmy:Pricehigh +
#                  Style_cute_dmy:Pricehigh + Material_silk_dmy:Pricehigh +
#                  Pricelow:Style_cute_dmy + Pricelow:Style_vintage_dmy +
Pricelow:Material_chiffon_dmy +
#                  Pricelow:Material_lycra_dmy + Pricelow:Material_modal_dmy +
#                  Pricelow:Material_nylon_dmy + Pricelow:Material_rayon_dmy +
#                  Pricelow:Material_silk_dmy + Pricemedium:Style_cute_dmy +
#                  Pricemedium:Style_vintage_dmy + Pricemedium:Material_cashmere_dmy
+
```

```
#                    Pricemedium:Material_silk_dmy +
Style_cute_dmy:Material_chiffon_dmy +
#                    Style_cute_dmy:Material_silk_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9959 -0.6005  0.0444  0.8559  3.1665
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                          7.6216     0.1109  68.738  < 2e-16 ***
# Pricelow                             0.5399     0.1901   2.840  0.00479 **
# Pricemedium                         -0.9754     0.3976  -2.453  0.01469 *
# Style_cute_dmy                       0.5098     0.4832   1.055  0.29223
# Style_fashion_dmy                   -1.5672     1.4476  -1.083  0.27978
# Style_vintage_dmy                    0.7884     0.4175   1.888  0.05990 .
# Material_cashmere_dmy               -3.8839     1.4476  -2.683  0.00767 **
# Material_chiffon_dmy                 1.1375     0.5757   1.976  0.04901 *
# Material_lycra_dmy                  -3.5963     1.4476  -2.484  0.01348 *
# Material_modal_dmy                  -1.0410     1.4476  -0.719  0.47258
# Material_rayon_dmy                  -1.4175     0.6250  -2.268  0.02400 *
# Material_silk_dmy                    0.9284     1.4476   0.641  0.52173
# Material_nylon_dmy:Pricehigh        -4.2204     1.4476  -2.916  0.00380 **
# Style_cute_dmy:Pricehigh            -1.8413     1.0617  -1.734  0.08381 .
# Material_silk_dmy:Pricehigh         -1.1672     1.8724  -0.623  0.53347
# Pricelow:Style_cute_dmy             -1.5847     0.7127  -2.224  0.02687 *
# Pricelow:Style_vintage_dmy          -0.5000     0.8480  -0.590  0.55586
# Pricelow:Material_chiffon_dmy       -0.5781     0.7819  -0.739  0.46027
# Pricelow:Material_lycra_dmy          3.2180     2.0500   1.570  0.11745
# Pricelow:Material_modal_dmy         -1.1570     2.0500  -0.564  0.57288
# Pricelow:Material_nylon_dmy         -1.5789     0.8476  -1.863  0.06338 .
# Pricelow:Material_rayon_dmy          2.8137     1.2073   2.331  0.02039 *
# Pricelow:Material_silk_dmy          -1.2320     1.5160  -0.813  0.41701
# Pricemedium:Style_cute_dmy           1.5228     1.3408   1.136  0.25693
# Pricemedium:Style_vintage_dmy       -1.1904     1.5502  -0.768  0.44309
# Pricemedium:Material_cashmere_dmy    2.8609     1.8118   1.579  0.11532
# Pricemedium:Material_silk_dmy       -3.9856     1.9161  -2.080  0.03830 *
# Style_cute_dmy:Material_chiffon_dmy  -1.7614     1.0145  -1.736  0.08349 .
# Style_cute_dmy:Material_silk_dmy     -0.6481     1.0099  -0.642  0.52149
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.443 on 323 degrees of freedom
# Multiple R-squared:  0.2253,   Adjusted R-squared:  0.1582
# F-statistic: 3.355 on 28 and 323 DF,  p-value: 0.00000009214

totsale_lm.19_c.8 <- update(totsale_lm.19_c.7, ~.- Pricelow:Material_modal_dmy)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                    Material_rayon_dmy + Material_silk_dmy +
Material_nylon_dmy:Pricehigh +
#                    Style_cute_dmy:Pricehigh + Material_silk_dmy:Pricehigh +
#                    Pricelow:Style_cute_dmy + Pricelow:Style_vintage_dmy +
Pricelow:Material_chiffon_dmy +
#                    Pricelow:Material_lycra_dmy + Pricelow:Material_nylon_dmy +
#                    Pricelow:Material_rayon_dmy + Pricelow:Material_silk_dmy +
#                    Pricemedium:Style_cute_dmy + Pricemedium:Style_vintage_dmy +
#                    Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                    Style_cute_dmy:Material_chiffon_dmy +
Style_cute_dmy:Material_silk_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9966 -0.5961  0.0484  0.8586  3.1631
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
```

```
# (Intercept)                          7.6250     0.1106  68.941  < 2e-16 ***
# Pricelow                             0.5300     0.1891   2.803  0.00537 **
# Pricemedium                         -0.9788     0.3971  -2.465  0.01423 *
# Style_cute_dmy                       0.5071     0.4827   1.051  0.29422
# Style_fashion_dmy                   -1.5705     1.4460  -1.086  0.27824
# Style_vintage_dmy                    0.7853     0.4170   1.883  0.06058 .
# Material_cashmere_dmy               -3.8873     1.4460  -2.688  0.00755 **
# Material_chiffon_dmy                 1.1353     0.5751   1.974  0.04919 *
# Material_lycra_dmy                  -3.5996     1.4460  -2.489  0.01330 *
# Material_modal_dmy                  -1.6178     1.0239  -1.580  0.11508
# Material_rayon_dmy                  -1.4198     0.6244  -2.274  0.02363 *
# Material_silk_dmy                    0.9251     1.4460   0.640  0.52279
# Material_nylon_dmy:Pricehigh        -4.2238     1.4460  -2.921  0.00373 **
# Style_cute_dmy:Pricehigh            -1.8398     1.0605  -1.735  0.08374 .
# Material_silk_dmy:Pricehigh         -1.1650     1.8704  -0.623  0.53383
# Pricelow:Style_cute_dmy             -1.5767     0.7118  -2.215  0.02745 *
# Pricelow:Style_vintage_dmy          -0.4904     0.8469  -0.579  0.56298
# Pricelow:Material_chiffon_dmy       -0.5700     0.7810  -0.730  0.46601
# Pricelow:Material_lycra_dmy          3.2279     2.0477   1.576  0.11592
# Pricelow:Material_nylon_dmy         -1.5724     0.8466  -1.857  0.06418 .
# Pricelow:Material_rayon_dmy          2.8226     1.2059   2.341  0.01986 *
# Pricelow:Material_silk_dmy          -1.2224     1.5144  -0.807  0.42014
# Pricemedium:Style_cute_dmy           1.5266     1.3394   1.140  0.25523
# Pricemedium:Style_vintage_dmy       -1.1873     1.5486  -0.767  0.44380
# Pricemedium:Material_cashmere_dmy    2.8643     1.8099   1.583  0.11450
# Pricemedium:Material_silk_dmy       -3.9811     1.9140  -2.080  0.03832 *
# Style_cute_dmy:Material_chiffon_dmy  -1.7619     1.0134  -1.739  0.08306 .
# Style_cute_dmy:Material_silk_dmy     -0.6515     1.0088  -0.646  0.51888
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.442 on 324 degrees of freedom
# Multiple R-squared:  0.2246,   Adjusted R-squared:  0.1599
# F-statistic: 3.475 on 27 and 324 DF,  p-value: 0.00000005417

totsale_lm.19_c.9 <- update(totsale_lm.19_c.8, ~. -Pricelow:Style_vintage_dmy)
summary(totsale_lm.19_c.9)


# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                     Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                     Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                     Material_rayon_dmy + Material_silk_dmy +
Material_nylon_dmy:Pricehigh +
#                     Style_cute_dmy:Pricehigh + Material_silk_dmy:Pricehigh +
#                     Pricelow:Style_cute_dmy + Pricelow:Material_chiffon_dmy +
#                     Pricelow:Material_lycra_dmy + Pricelow:Material_nylon_dmy +
#                     Pricelow:Material_rayon_dmy + Pricelow:Material_silk_dmy +
#                     Pricemedium:Style_cute_dmy + Pricemedium:Style_vintage_dmy +
#                     Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                     Style_cute_dmy:Material_chiffon_dmy +
Style_cute_dmy:Material_silk_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9951 -0.6177  0.0743  0.8512  3.1557
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                          7.6323     0.1098  69.540  < 2e-16 ***
# Pricelow                             0.5064     0.1844   2.746  0.00638 **
# Pricemedium                         -0.9863     0.3965  -2.488  0.01336 *
# Style_cute_dmy                       0.4982     0.4820   1.034  0.30204
# Style_fashion_dmy                   -1.5779     1.4445  -1.092  0.27548
# Style_vintage_dmy                    0.6665     0.3626   1.838  0.06701 .
# Material_cashmere_dmy               -3.8947     1.4445  -2.696  0.00738 **
# Material_chiffon_dmy                 1.1481     0.5740   2.000  0.04634 *
# Material_lycra_dmy                  -3.6070     1.4445  -2.497  0.01302 *
# Material_modal_dmy                  -1.6134     1.0229  -1.577  0.11569
```

```
# Material_rayon_dmy                     -1.4040     0.6231  -2.253  0.02491 *
# Material_silk_dmy                       0.9177     1.4445   0.635  0.52567
# Material_nylon_dmy:Pricehigh           -4.2311     1.4445  -2.929  0.00364 **
# Style_cute_dmy:Pricehigh               -1.8416     1.0595  -1.738  0.08312 .
# Material_silk_dmy:Pricehigh            -1.1559     1.8684  -0.619  0.53656
# Pricelow:Style_cute_dmy                -1.5537     0.7099  -2.189  0.02934 *
# Pricelow:Material_chiffon_dmy          -0.5668     0.7802  -0.726  0.46806
# Pricelow:Material_lycra_dmy             3.2515     2.0452   1.590  0.11285
# Pricelow:Material_nylon_dmy            -1.5561     0.8453  -1.841  0.06653 .
# Pricelow:Material_rayon_dmy             2.8231     1.2047   2.343  0.01971 *
# Pricelow:Material_silk_dmy             -1.1998     1.5123  -0.793  0.42813
# Pricemedium:Style_cute_dmy              1.5381     1.3379   1.150  0.25114
# Pricemedium:Style_vintage_dmy          -1.0683     1.5333  -0.697  0.48647
# Pricemedium:Material_cashmere_dmy       2.8718     1.8080   1.588  0.11317
# Pricemedium:Material_silk_dmy          -3.9712     1.9120  -2.077  0.03859 *
# Style_cute_dmy:Material_chiffon_dmy    -1.7740     1.0122  -1.753  0.08062 .
# Style_cute_dmy:Material_silk_dmy       -0.6588     1.0077  -0.654  0.51374
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.44 on 325 degrees of freedom
# Multiple R-squared:  0.2238,   Adjusted R-squared:  0.1617
# F-statistic: 3.603 on 26 and 325 DF,  p-value: 0.00000003152

totsale_lm.19_c.10 <- update(totsale_lm.19_c.9, ~. -Material_silk_dmy:Pricehigh)
summary(totsale_lm.19_c.10)
# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                    Material_rayon_dmy + Material_silk_dmy +
Material_nylon_dmy:Pricehigh +
#                    Style_cute_dmy:Pricehigh + Pricelow:Style_cute_dmy +
Pricelow:Material_chiffon_dmy +
#                    Pricelow:Material_lycra_dmy + Pricelow:Material_nylon_dmy +
#                    Pricelow:Material_rayon_dmy + Pricelow:Material_silk_dmy +
#                    Pricemedium:Style_cute_dmy + Pricemedium:Style_vintage_dmy +
#                    Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                    Style_cute_dmy:Material_chiffon_dmy +
Style_cute_dmy:Material_silk_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9946 -0.6403  0.0862  0.8505  3.1550
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                             7.6331     0.1096  69.617  < 2e-16 ***
# Pricelow                                0.5039     0.1842   2.736  0.00657 **
# Pricemedium                            -0.9902     0.3961  -2.500  0.01291 *
# Style_cute_dmy                          0.4970     0.4815   1.032  0.30272
# Style_fashion_dmy                      -1.5787     1.4431  -1.094  0.27479
# Style_vintage_dmy                       0.6650     0.3623   1.836  0.06734 .
# Material_cashmere_dmy                  -3.8954     1.4431  -2.699  0.00731 **
# Material_chiffon_dmy                    1.1502     0.5735   2.006  0.04572 *
# Material_lycra_dmy                     -3.6078     1.4431  -2.500  0.01291 *
# Material_modal_dmy                     -1.6130     1.0219  -1.578  0.11544
# Material_rayon_dmy                     -1.3844     0.6217  -2.227  0.02665 *
# Material_silk_dmy                       0.2300     0.9216   0.250  0.80307
# Material_nylon_dmy:Pricehigh           -4.2319     1.4431  -2.932  0.00360 **
# Style_cute_dmy:Pricehigh               -1.9599     1.0411  -1.883  0.06065 .
# Pricelow:Style_cute_dmy                -1.5243     0.7077  -2.154  0.03198 *
# Pricelow:Material_chiffon_dmy          -0.5699     0.7794  -0.731  0.46521
# Pricelow:Material_lycra_dmy             3.2540     2.0433   1.593  0.11224
# Pricelow:Material_nylon_dmy            -1.5544     0.8444  -1.841  0.06656 .
# Pricelow:Material_rayon_dmy             2.8052     1.2032   2.331  0.02034 *
# Pricelow:Material_silk_dmy             -0.4946     0.9928  -0.498  0.61872
# Pricemedium:Style_cute_dmy              1.5852     1.3345   1.188  0.23574
# Pricemedium:Style_vintage_dmy          -1.0638     1.5318  -0.694  0.48790
# Pricemedium:Material_cashmere_dmy       2.8757     1.8063   1.592  0.11235
```

```
# Pricemedium:Material_silk_dmy        -3.2376     1.4985  -2.161  0.03146 *
# Style_cute_dmy:Material_chiffon_dmy  -1.7836     1.0111  -1.764  0.07867 .
# Style_cute_dmy:Material_silk_dmy     -0.7904     0.9841  -0.803  0.42243
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.439 on 326 degrees of freedom
# Multiple R-squared:  0.2228,   Adjusted R-squared:  0.1632
# F-statistic: 3.739 on 25 and 326 DF,  p-value: 0.00000001835

totsale_lm.19_c.11 <- update(totsale_lm.19_c.10,~. -Material_silk_dmy)
summary(totsale_lm.19_c.11)
# Call:
#        lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                   Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                   Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                   Material_rayon_dmy + Material_nylon_dmy:Pricehigh +
Style_cute_dmy:Pricehigh +
#                   Pricelow:Style_cute_dmy + Pricelow:Material_chiffon_dmy +
#                   Pricelow:Material_lycra_dmy + Pricelow:Material_nylon_dmy +
#                   Pricelow:Material_rayon_dmy + Pricelow:Material_silk_dmy +
#                   Pricemedium:Style_cute_dmy + Pricemedium:Style_vintage_dmy +
#                   Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                   Style_cute_dmy:Material_chiffon_dmy +
Style_cute_dmy:Material_silk_dmy,
#           data = Training1s)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.9952 -0.6181  0.0896  0.8547  3.1521
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)                           7.6360     0.1089  70.134  < 2e-16 ***
# Pricelow                              0.5020     0.1838   2.731  0.00665 **
# Pricemedium                          -0.9915     0.3955  -2.507  0.01265 *
# Style_cute_dmy                        0.4947     0.4807   1.029  0.30420
# Style_fashion_dmy                    -1.5816     1.4410  -1.098  0.27321
# Style_vintage_dmy                     0.6635     0.3617   1.834  0.06753 .
# Material_cashmere_dmy                -3.8983     1.4410  -2.705  0.00718 **
# Material_chiffon_dmy                  1.1465     0.5725   2.003  0.04604 *
# Material_lycra_dmy                   -3.6106     1.4410  -2.506  0.01271 *
# Material_modal_dmy                   -1.6149     1.0204  -1.583  0.11448
# Material_rayon_dmy                   -1.3960     0.6191  -2.255  0.02480 *
# Material_nylon_dmy:Pricehigh         -4.2348     1.4410  -2.939  0.00353 **
# Style_cute_dmy:Pricehigh             -1.9014     1.0129  -1.877  0.06139 .
# Pricelow:Style_cute_dmy              -1.5361     0.7051  -2.179  0.03008 *
# Pricelow:Material_chiffon_dmy        -0.5660     0.7781  -0.727  0.46751
# Pricelow:Material_lycra_dmy           3.2559     2.0403   1.596  0.11151
# Pricelow:Material_nylon_dmy          -1.5554     0.8432  -1.845  0.06601 .
# Pricelow:Material_rayon_dmy           2.8158     1.2007   2.345  0.01962 *
# Pricelow:Material_silk_dmy           -0.2732     0.4459  -0.613  0.54043
# Pricemedium:Style_cute_dmy            1.5650     1.3301   1.177  0.24021
# Pricemedium:Style_vintage_dmy        -1.0637     1.5296  -0.695  0.48728
# Pricemedium:Material_cashmere_dmy     2.8771     1.8037   1.595  0.11166
# Pricemedium:Material_silk_dmy        -3.0301     1.2449  -2.434  0.01547 *
# Style_cute_dmy:Material_chiffon_dmy  -1.7772     1.0093  -1.761  0.07922 .
# Style_cute_dmy:Material_silk_dmy     -0.7259     0.9481  -0.766  0.44447
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.437 on 327 degrees of freedom
# Multiple R-squared:  0.2227,   Adjusted R-squared:  0.1656
# F-statistic: 3.904 on 24 and 327 DF,  p-value: 0.000000009273

totsale_lm.19_c.12 <- update(totsale_lm.19_c.11, ~. -Pricelow:Material_silk_dmy)
summary(totsale_lm.19_c.12)

# Call:
#        lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                   Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
```

```
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                    Material_rayon_dmy + Material_nylon_dmy:Pricehigh +
Style_cute_dmy:Pricehigh +
#                    Pricelow:Style_cute_dmy + Pricelow:Material_chiffon_dmy +
#                    Pricelow:Material_lycra_dmy + Pricelow:Material_nylon_dmy +
#                    Pricelow:Material_rayon_dmy + Pricemedium:Style_cute_dmy +
#                    Pricemedium:Style_vintage_dmy + Pricemedium:Material_cashmere_dmy
+
#                    Pricemedium:Material_silk_dmy +
Style_cute_dmy:Material_chiffon_dmy +
#                    Style_cute_dmy:Material_silk_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.9966 -0.6256  0.0868  0.8480  3.1525
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                         7.6356     0.1088  70.198  < 2e-16 ***
#         Pricelow                            0.4724     0.1772   2.666  0.00805 **
#         Pricemedium                        -0.9950     0.3951  -2.519  0.01225 *
#         Style_cute_dmy                      0.4965     0.4803   1.034  0.30202
# Style_fashion_dmy                  -1.5812     1.4396  -1.098  0.27286
# Style_vintage_dmy                   0.6715     0.3611   1.859  0.06387 .
# Material_cashmere_dmy              -3.8980     1.4396  -2.708  0.00713 **
#         Material_chiffon_dmy                1.1457     0.5719   2.003  0.04597 *
#         Material_lycra_dmy                 -3.6103     1.4396  -2.508  0.01263 *
#         Material_modal_dmy                 -1.5997     1.0191  -1.570  0.11745
# Material_rayon_dmy                 -1.4072     0.6183  -2.276  0.02349 *
#         Material_nylon_dmy:Pricehigh       -4.2344     1.4396  -2.941  0.00350 **
#         Style_cute_dmy:Pricehigh           -1.8438     1.0076  -1.830  0.06817 .
# Pricelow:Style_cute_dmy            -1.5350     0.7044  -2.179  0.03004 *
#         Pricelow:Material_chiffon_dmy      -0.5356     0.7758  -0.690  0.49047
# Pricelow:Material_lycra_dmy         3.2854     2.0378   1.612  0.10787
# Pricelow:Material_nylon_dmy        -1.5255     0.8410  -1.814  0.07061 .
# Pricelow:Material_rayon_dmy         2.8569     1.1977   2.385  0.01763 *
#         Pricemedium:Style_cute_dmy          1.6211     1.3257   1.223  0.22227
# Pricemedium:Style_vintage_dmy      -1.0679     1.5282  -0.699  0.48516
# Pricemedium:Material_cashmere_dmy   2.8805     1.8020   1.599  0.11088
# Pricemedium:Material_silk_dmy      -2.9722     1.2401  -2.397  0.01710 *
#         Style_cute_dmy:Material_chiffon_dmy  -1.7785     1.0084  -1.764  0.07871 .
# Style_cute_dmy:Material_silk_dmy   -0.8917     0.9078  -0.982  0.32670
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.436 on 328 degrees of freedom
# Multiple R-squared:  0.2218,   Adjusted R-squared:  0.1672
# F-statistic: 4.065 on 23 and 328 DF,  p-value: 0.000000005182


totsale_lm.19_c.13 <- update(totsale_lm.19_c.12, ~. -Pricelow:Material_chiffon_dmy)
summary(totsale_lm.19_c.13)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                    Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                    Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                    Material_rayon_dmy + Material_nylon_dmy:Pricehigh +
Style_cute_dmy:Pricehigh +
#                    Pricelow:Style_cute_dmy + Pricelow:Material_lycra_dmy +
Pricelow:Material_nylon_dmy +
#                    Pricelow:Material_rayon_dmy + Pricemedium:Style_cute_dmy +
#                    Pricemedium:Style_vintage_dmy + Pricemedium:Material_cashmere_dmy
+
#                    Pricemedium:Material_silk_dmy +
Style_cute_dmy:Material_chiffon_dmy +
#                    Style_cute_dmy:Material_silk_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -5.0317 -0.6105  0.0954  0.8262  3.1449
#
```

```
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                        7.6432     0.1081  70.687  < 2e-16 ***
#         Pricelow                           0.4492     0.1738   2.584  0.01018 *
#         Pricemedium                       -1.0020     0.3946  -2.539  0.01157 *
#         Style_cute_dmy                     0.5240     0.4782   1.096  0.27403
# Style_fashion_dmy                 -1.5888     1.4384  -1.105  0.27017
# Style_vintage_dmy                  0.6857     0.3603   1.903  0.05789 .
# Material_cashmere_dmy             -3.9056     1.4384  -2.715  0.00697 **
# Material_chiffon_dmy               0.8808     0.4237   2.079  0.03840 *
# Material_lycra_dmy                -3.6179     1.4384  -2.515  0.01237 *
# Material_modal_dmy                -1.5957     1.0183  -1.567  0.11806
# Material_rayon_dmy                -1.4196     0.6175  -2.299  0.02214 *
# Material_nylon_dmy:Pricehigh      -4.2420     1.4384  -2.949  0.00342 **
# Style_cute_dmy:Pricehigh          -1.8844     1.0051  -1.875  0.06169 .
# Pricelow:Style_cute_dmy           -1.5920     0.6990  -2.278  0.02339 *
# Pricelow:Material_lycra_dmy        3.3086     2.0359   1.625  0.10509
# Pricelow:Material_nylon_dmy       -1.5099     0.8400  -1.797  0.07319 .
# Pricelow:Material_rayon_dmy        2.8848     1.1961   2.412  0.01642 *
# Pricemedium:Style_cute_dmy         1.5835     1.3235   1.196  0.23238
# Pricemedium:Style_vintage_dmy     -1.0828     1.5268  -0.709  0.47872
# Pricemedium:Material_cashmere_dmy  2.8875     1.8005   1.604  0.10974
# Pricemedium:Material_silk_dmy     -2.9823     1.2391  -2.407  0.01664 *
# Style_cute_dmy:Material_chiffon_dmy -1.7004   1.0012  -1.698  0.09039 .
# Style_cute_dmy:Material_silk_dmy  -0.8629     0.9061  -0.952  0.34165
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.434 on 329 degrees of freedom
# Multiple R-squared:  0.2207,   Adjusted R-squared:  0.1686
# F-statistic: 4.234 on 22 and 329 DF,  p-value: 0.000000002956

totsale_lm.19_c.14 <- update(totsale_lm.19_c.13, ~. -Pricemedium:Style_vintage_dmy)
summary(totsale_lm.19_c.14)


# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                     Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                     Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                     Material_rayon_dmy + Material_nylon_dmy:Pricehigh +
Style_cute_dmy:Pricehigh +
#                     Pricelow:Style_cute_dmy + Pricelow:Material_lycra_dmy +
Pricelow:Material_nylon_dmy +
#                     Pricelow:Material_rayon_dmy + Pricemedium:Style_cute_dmy +
#                     Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                     Style_cute_dmy:Material_chiffon_dmy +
Style_cute_dmy:Material_silk_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -5.0310 -0.6127  0.1134  0.8236  3.1410
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                        7.6471     0.1079  70.868  < 2e-16 ***
# Pricelow                           0.4475     0.1737   2.577  0.01041 *
# Pricemedium                       -1.0727     0.3815  -2.812  0.00522 **
# Style_cute_dmy                     0.5194     0.4778   1.087  0.27781
# Style_fashion_dmy                 -1.5927     1.4373  -1.108  0.26864
# Style_vintage_dmy                  0.6254     0.3498   1.788  0.07475 .
# Material_cashmere_dmy             -3.9095     1.4373  -2.720  0.00688 **
# Material_chiffon_dmy               0.8828     0.4233   2.085  0.03782 *
# Material_lycra_dmy                -3.6218     1.4373  -2.520  0.01221 *
# Material_modal_dmy                -1.5988     1.0175  -1.571  0.11709
# Material_rayon_dmy                -1.4120     0.6170  -2.289  0.02273 *
# Material_nylon_dmy:Pricehigh      -4.2459     1.4373  -2.954  0.00336 **
# Style_cute_dmy:Pricehigh          -1.8837     1.0043  -1.876  0.06160 .
# Pricelow:Style_cute_dmy           -1.5879     0.6984  -2.273  0.02364 *
# Pricelow:Material_lycra_dmy        3.3104     2.0344   1.627  0.10465
```

```
# Pricelow:Material_nylon_dmy          -1.5120     0.8394  -1.801  0.07257 .
# Pricelow:Material_rayon_dmy           2.8752     1.1951   2.406  0.01669 *
# Pricemedium:Style_cute_dmy            1.6352     1.3205   1.238  0.21647
# Pricemedium:Material_cashmere_dmy     2.9582     1.7964   1.647  0.10056
# Pricemedium:Material_silk_dmy        -2.9351     1.2364  -2.374  0.01817 *
# Style_cute_dmy:Material_chiffon_dmy  -1.7026     1.0005  -1.702  0.08974 .
# Style_cute_dmy:Material_silk_dmy     -0.8706     0.9054  -0.962  0.33694
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.433 on 330 degrees of freedom
# Multiple R-squared:  0.2195,   Adjusted R-squared:  0.1698
# F-statistic: 4.419 on 21 and 330 DF,  p-value: 0.000000001671

totsale_lm.19_c.15 <- update(totsale_lm.19_c.14, ~. -Style_cute_dmy:Material_silk_dmy)
summary(totsale_lm.19_c.15)

# Call:
#        lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_cute_dmy +
#                     Style_fashion_dmy + Style_vintage_dmy + Material_cashmere_dmy +
#                     Material_chiffon_dmy + Material_lycra_dmy + Material_modal_dmy +
#                     Material_rayon_dmy + Material_nylon_dmy:Pricehigh +
Style_cute_dmy:Pricehigh +
#                     Pricelow:Style_cute_dmy + Pricelow:Material_lycra_dmy +
Pricelow:Material_nylon_dmy +
#                     Pricelow:Material_rayon_dmy + Pricemedium:Style_cute_dmy +
#                     Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                     Style_cute_dmy:Material_chiffon_dmy, data = Training1s)
#
#       Residuals:
#         Min      1Q  Median      3Q     Max
#       -5.0113 -0.6128  0.1188  0.8238  3.1419
#
# Coefficients:
#                                      Estimate Std. Error t value Pr(>|t|)
# (Intercept)                           7.6462     0.1079  70.870  < 2e-16 ***
# Pricelow                              0.4485     0.1736   2.583  0.01022 *
# Pricemedium                          -1.0526     0.3809  -2.764  0.00604 **
# Style_cute_dmy                        0.5006     0.4774   1.049  0.29508
# Style_fashion_dmy                    -1.5917     1.4372  -1.108  0.26887
# Style_vintage_dmy                     0.6219     0.3498   1.778  0.07635 .
# Material_cashmere_dmy                -3.9085     1.4372  -2.720  0.00688 **
# Material_chiffon_dmy                  0.8835     0.4233   2.087  0.03763 *
# Material_lycra_dmy                   -3.6208     1.4372  -2.519  0.01223 *
# Material_modal_dmy                   -1.5983     1.0174  -1.571  0.11714
# Material_rayon_dmy                   -1.3560     0.6141  -2.208  0.02793 *
# Material_nylon_dmy:Pricehigh         -4.2450     1.4372  -2.954  0.00336 **
# Style_cute_dmy:Pricehigh             -2.1728     0.9581  -2.268  0.02399 *
# Pricelow:Style_cute_dmy              -1.7716     0.6717  -2.637  0.00875 **
# Pricelow:Material_lycra_dmy           3.3093     2.0342   1.627  0.10471
# Pricelow:Material_nylon_dmy          -1.5121     0.8393  -1.802  0.07251 .
# Pricelow:Material_rayon_dmy           2.8191     1.1935   2.362  0.01876 *
# Pricemedium:Style_cute_dmy            1.3510     1.2869   1.050  0.29454
# Pricemedium:Material_cashmere_dmy     2.9381     1.7961   1.636  0.10282
# Pricemedium:Material_silk_dmy        -3.2381     1.1954  -2.709  0.00710 **
# Style_cute_dmy:Material_chiffon_dmy  -1.6226     0.9969  -1.628  0.10454
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.433 on 331 degrees of freedom
# Multiple R-squared:  0.2173,   Adjusted R-squared:   0.17
# F-statistic: 4.595 on 20 and 331 DF,  p-value: 0.000000001098


totsale_lm.19_c.16 <- update(totsale_lm.19_c.15, ~. -Style_cute_dmy)

summary(totsale_lm.19_c.16)

# Call:
#        lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_fashion_dmy +
```

```
#                     Style_vintage_dmy + Material_cashmere_dmy + Material_chiffon_dmy
+
#                     Material_lycra_dmy + Material_modal_dmy + Material_rayon_dmy +
#                     Material_nylon_dmy:Pricehigh + Style_cute_dmy:Pricehigh +
#                     Pricelow:Style_cute_dmy + Pricelow:Material_lycra_dmy +
Pricelow:Material_nylon_dmy +
#                     Pricelow:Material_rayon_dmy + Pricemedium:Style_cute_dmy +
#                     Pricemedium:Material_cashmere_dmy + Pricemedium:Material_silk_dmy
+
#                     Style_cute_dmy:Material_chiffon_dmy, data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7257 -0.6143  0.1155  0.8254  3.1179
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                          7.6701     0.1055  72.734  < 2e-16 ***
# Pricelow                             0.4261     0.1723   2.472  0.01393 *
# Pricemedium                         -1.0752     0.3803  -2.827  0.00498 **
# Style_fashion_dmy                   -1.6157     1.4372  -1.124  0.26174
# Style_vintage_dmy                    0.6008     0.3493   1.720  0.08633 .
# Material_cashmere_dmy               -3.9325     1.4372  -2.736  0.00655 **
# Material_chiffon_dmy                 0.8725     0.4232   2.062  0.04003 *
# Material_lycra_dmy                  -3.6448     1.4372  -2.536  0.01167 *
# Material_modal_dmy                  -1.6111     1.0175  -1.583  0.11429
# Material_rayon_dmy                  -1.2851     0.6105  -2.105  0.03604 *
# Material_nylon_dmy:Pricehigh        -4.2690     1.4372  -2.970  0.00319 **
# Pricehigh:Style_cute_dmy            -1.7198     0.8554  -2.011  0.04518 *
# Pricelow:Style_cute_dmy             -1.3091     0.5068  -2.583  0.01022 *
# Pricelow:Material_lycra_dmy          3.3318     2.0344   1.638  0.10242
# Pricelow:Material_nylon_dmy         -1.5136     0.8394  -1.803  0.07227 .
# Pricelow:Material_rayon_dmy          2.7467     1.1917   2.305  0.02179 *
# Pricemedium:Style_cute_dmy           1.8507     1.1956   1.548  0.12258
# Pricemedium:Material_cashmere_dmy    2.9607     1.7962   1.648  0.10023
# Pricemedium:Material_silk_dmy       -3.2390     1.1956  -2.709  0.00710 **
# Material_chiffon_dmy:Style_cute_dmy  -1.2817     0.9425  -1.360  0.17480
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.433 on 332 degrees of freedom
# Multiple R-squared:  0.2147,    Adjusted R-squared:  0.1698
# F-statistic: 4.777 on 19 and 332 DF,  p-value: 0.000000000762

totsale_lm.19_c.17 <- update(totsale_lm.19_c.16, ~. -Style_fashion_dmy)
summary(totsale_lm.19_c.17)

# Call:
#         lm(formula = log(TotalSales) ~ Pricelow + Pricemedium + Style_vintage_dmy +
#                     Material_cashmere_dmy + Material_chiffon_dmy + Material_lycra_dmy
+
#                     Material_modal_dmy + Material_rayon_dmy +
Material_nylon_dmy:Pricehigh +
#                     Pricehigh:Style_cute_dmy + Pricelow:Style_cute_dmy +
Pricelow:Material_lycra_dmy +
#                     Pricelow:Material_nylon_dmy + Pricelow:Material_rayon_dmy +
#                     Pricemedium:Style_cute_dmy + Pricemedium:Material_cashmere_dmy +
#                     Pricemedium:Material_silk_dmy +
Material_chiffon_dmy:Style_cute_dmy,
#            data = Training1s)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.7170 -0.6385  0.1161  0.8312  3.1266
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                          7.6614     0.1052  72.819  < 2e-16 ***
# Pricelow                             0.4343     0.1723   2.521  0.01217 *
# Pricemedium                         -1.0669     0.3804  -2.805  0.00533 **
# Style_vintage_dmy                    0.6066     0.3494   1.736  0.08343 .
# Material_cashmere_dmy               -3.9238     1.4378  -2.729  0.00669 **
```

```
# Material_chiffon_dmy                0.8766     0.4234   2.071  0.03917 *
# Material_lycra_dmy                 -3.6361     1.4378  -2.529  0.01190 *
# Material_modal_dmy                 -1.6065     1.0179  -1.578  0.11546
# Material_rayon_dmy                 -1.2785     0.6107  -2.093  0.03706 *
# Material_nylon_dmy:Pricehigh       -4.2603     1.4378  -2.963  0.00326 **
# Pricehigh:Style_cute_dmy           -1.7133     0.8557  -2.002  0.04607 *
# Pricelow:Style_cute_dmy            -1.3093     0.5070  -2.582  0.01024 *
# Pricelow:Material_lycra_dmy         3.3236     2.0352   1.633  0.10339
# Pricelow:Material_nylon_dmy        -1.5131     0.8398  -1.802  0.07247 .
# Pricelow:Material_rayon_dmy         2.7405     1.1922   2.299  0.02214 *
# Pricemedium:Style_cute_dmy          1.8510     1.1961   1.548  0.12267
# Pricemedium:Material_cashmere_dmy   2.9524     1.7969   1.643  0.10132
# Pricemedium:Material_silk_dmy      -3.2388     1.1961  -2.708  0.00712 **
# Material_chiffon_dmy:Style_cute_dmy  -1.2798     0.9429  -1.357  0.17561
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.434 on 333 degrees of freedom
# Multiple R-squared:  0.2117,   Adjusted R-squared:  0.1691
# F-statistic: 4.968 on 18 and 333 DF,  p-value: 0.0000000005564


#Adj Rsq is starting to decrease so lets stop and consider totsale_lm.19_c.15 as best
model
# even though it has some insignificant terms


# Coefficients:
#                                    Estimate Std. Error t value Pr(>|t|)
# (Intercept)                         7.6462     0.1079  70.870  < 2e-16 ***
# Pricelow                            0.4485     0.1736   2.583  0.01022 *
# Pricemedium                        -1.0526     0.3809  -2.764  0.00604 **
# Style_cute_dmy                      0.5006     0.4774   1.049  0.29508
# Style_fashion_dmy                  -1.5917     1.4372  -1.108  0.26887
# Style_vintage_dmy                   0.6219     0.3498   1.778  0.07635 .
# Material_cashmere_dmy              -3.9085     1.4372  -2.720  0.00688 **
# Material_chiffon_dmy                0.8835     0.4233   2.087  0.03763 *
# Material_lycra_dmy                 -3.6208     1.4372  -2.519  0.01223 *
# Material_modal_dmy                 -1.5983     1.0174  -1.571  0.11714
# Material_rayon_dmy                 -1.3560     0.6141  -2.208  0.02793 *
# Material_nylon_dmy:Pricehigh       -4.2450     1.4372  -2.954  0.00336 **
# Style_cute_dmy:Pricehigh           -2.1728     0.9581  -2.268  0.02399 *
# Pricelow:Style_cute_dmy            -1.7716     0.6717  -2.637  0.00875 **
# Pricelow:Material_lycra_dmy         3.3093     2.0342   1.627  0.10471
# Pricelow:Material_nylon_dmy        -1.5121     0.8393  -1.802  0.07251 .
# Pricelow:Material_rayon_dmy         2.8191     1.1935   2.362  0.01876 *
# Pricemedium:Style_cute_dmy          1.3510     1.2869   1.050  0.29454
# Pricemedium:Material_cashmere_dmy   2.9381     1.7961   1.636  0.10282
# Pricemedium:Material_silk_dmy      -3.2381     1.1954  -2.709  0.00710 **
# Style_cute_dmy:Material_chiffon_dmy  -1.6226     0.9969  -1.628  0.10454
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.433 on 331 degrees of freedom
# Multiple R-squared:  0.2173,   Adjusted R-squared:   0.17
# F-statistic: 4.595 on 20 and 331 DF,  p-value: 0.000000001098

par(mfrow=c(2,2))
plot(totsale_lm.19_c.15)

# based on the residual plots and cooks distance,
#looks like there are couple of outliers hence heavy tail in Q-Q plot at lower
quantiles.
# For most part it is normal. If we remove the high leverage points and outliers,
# the model could get better ; but for now this is fine and residuals are
homoscedastic (equal variance)
# Adj Rsq also improved from the run totsale_lm.19 where we only had Main effects
after the addition of interaction variables
# is this the best model - no and we need to refine further adding more factors into
the model based on experimental design

# To decide the pricing for various upcoming clothes, they wish to find how the
```

```
# style, season, and material affect the sales of a dress and if the style of the
dress
# is more influential than its price -
# So the ones with p-value < alpha =0.05 are significant factors in our model,
# and affect the sales and the weightage how they affect is described by the
coefficient ;
# To answer if style is more influential than price, based on our model,
# the answer is no because two of the price levels ARE significant and
# only Style_vintage is marginally significant.

#validate the regression model using test data
# Till now we were checking training-error but the real goal of the model is to reduce
the
# testing error. As we already split the sample dataset into training and testing
dataset,
# we will use test dataset to evaluate the model that we have arrived upon.
# We will make a prediction based on 'totsale_lm.19_c.15' and will evaluate the model.
# As the last step, we will predict the 'test' observation and will see the comparison
# between predicted response and actual response value.
# RMSE explains on an average how much of the predicted value will be from the actual
value.
# Based on RMSE = 19749.31, we can conclude that on an average predicted value will be
off by 19749.31 from the actual value.
# It is not a great model because we only considered the few factors asked in Task 3
for the modeling of Total Sales. had we considered every possible reasonable factor
then we could have minimized RMSE
# But we have answered the objective of Task 3 based on what we have created so far
pred1 <- predict(totsale_lm.19_c.15, newdata = Testing1)
rmse <- sqrt(sum((exp(pred1) - Testing1$TotalSales)^2)/length(Testing1$TotalSales))
c(RMSE = rmse, R2=summary(totsale_lm.19_c.15)$r.squared)
# RMSE           R2
# 19749.3077309     0.2172935

par(mfrow=c(1,1))
plot(Testing1$TotalSales, exp(pred1)) # not that great

#now onto tasks 4 and 5
#Task4:
#Also, to increase sales, the management wants to analyze the attributes of dresses
and find which are the leading factors affecting the sale of a dress.

#we are yet to explore Size,Neckline,Sleeve length,Waistline,Fabric
type,Decoration,Pattern Type of the dress
# we have already expored the material style and season, so we can add the imp factors
from those later after completing the final task and then refine the model once at
that time


task4lm <- (lm(log(TotalSales) ~ factor(Size) + factor(waiseline) +
                              factor(FabricType) + factor(Decoration) +
                              factor(`Pattern Type`)+
                              SleeveLengththreequarter + SleeveLengthsleeveless +
                              SleeveLengthshort + SleeveLengthhalfsleeve +
SleeveLengthfull+
                              SleeveLengthcap_sleeves +
                              NeckLine_backless_dmy + NeckLine_boatneck_dmy +
                      NeckLine_bowneck_dmy + NeckLine_halter_dmy +
NeckLine_mandarincollor_dmy +
                      NeckLine_oneck_dmy + NeckLine_open_dmy +
NeckLine_peterpancollor_dmy +
                      NeckLine_ruffled_dmy + NeckLine_Scoop_dmy +
NeckLine_slashneck_dmy +
                      NeckLine_squarecollor_dmy + NeckLine_sweetheart_dmy +
                      NeckLine_turndowncollor_dmy + NeckLine_vneck_dmy
                              , data = ATT_DS3))

vif(task4lm)
# Error in vif.default(task4lm) :
#         there are aliased coefficients in the model
 alias(task4lm)
# Model :
#         log(TotalSales) ~ factor(Size) + factor(waiseline) + factor(FabricType) +
```

```
#          factor(Decoration) + factor(`Pattern Type`) + SleeveLengththreequarter +
#          SleeveLengthsleeveless + SleeveLengthshort + SleeveLengthhalfsleeve +
#          SleeveLengthfull + SleeveLengthcap_sleeves + NeckLine_backless_dmy +
#          NeckLine_boatneck_dmy + NeckLine_bowneck_dmy + NeckLine_halter_dmy +
#          NeckLine_mandarincollor_dmy + NeckLine_oneck_dmy + NeckLine_open_dmy +
#          NeckLine_peterpancollor_dmy + NeckLine_ruffled_dmy + NeckLine_Scoop_dmy +
#          NeckLine_slashneck_dmy + NeckLine_squarecollor_dmy + NeckLine_sweetheart_dmy
+
#          NeckLine_turndowncollor_dmy + NeckLine_vneck_dmy
#
# Complete :
#          (Intercept) factor(Size)L factor(Size)M factor(Size)S factor(Size)XL
# NeckLine_vneck_dmy  1           0             0             0             0
# factor(waiseline)empire factor(waiseline)natural factor(waiseline)null
# NeckLine_vneck_dmy  0                       0                        0
# factor(waiseline)princess factor(FabricType)broadcloth
# NeckLine_vneck_dmy  0                         0
# factor(FabricType)chiffon factor(FabricType)Corduroy factor(FabricType)dobby
# NeckLine_vneck_dmy  0                         0                         0
# factor(FabricType)flannel factor(FabricType)jersey factor(FabricType)knitted
# NeckLine_vneck_dmy  0                         0                       0
# factor(FabricType)lace factor(FabricType)null factor(FabricType)organza
# NeckLine_vneck_dmy  0                      0                      0
# factor(FabricType)other factor(FabricType)poplin factor(FabricType)satin
# NeckLine_vneck_dmy  0                       0                        0
# factor(FabricType)terry factor(FabricType)tulle factor(FabricType)woolen
# NeckLine_vneck_dmy  0                       0                       0
# factor(FabricType)worsted factor(Decoration)beading factor(Decoration)bow
# NeckLine_vneck_dmy  0                         0                         0
# factor(Decoration)button factor(Decoration)cascading factor(Decoration)crystal
# NeckLine_vneck_dmy  0                        0                           0
# factor(Decoration)draped factor(Decoration)embroidary
# NeckLine_vneck_dmy  0                        0
# factor(Decoration)feathers factor(Decoration)flowers
# NeckLine_vneck_dmy  0                          0
# factor(Decoration)hollowout factor(Decoration)lace factor(Decoration)none
# NeckLine_vneck_dmy  0                           0                      0
# factor(Decoration)null factor(Decoration)pearls factor(Decoration)plain
# NeckLine_vneck_dmy  0                      0                        0
# factor(Decoration)pleat factor(Decoration)pockets factor(Decoration)rivet
# NeckLine_vneck_dmy  0                       0                         0
# factor(Decoration)ruched factor(Decoration)ruffles factor(Decoration)sashes
# NeckLine_vneck_dmy  0                        0                         0
# factor(Decoration)sequined factor(Decoration)tassel factor(Decoration)Tiered
# NeckLine_vneck_dmy  0                          0                        0
# factor(`Pattern Type`)character factor(`Pattern Type`)dot
# NeckLine_vneck_dmy  0                               0
# factor(`Pattern Type`)floral factor(`Pattern Type`)geometric
# NeckLine_vneck_dmy  0                            0
# factor(`Pattern Type`)leopard factor(`Pattern Type`)none
# NeckLine_vneck_dmy  0                             0
# factor(`Pattern Type`)null factor(`Pattern Type`)patchwork
# NeckLine_vneck_dmy  0                          0
# factor(`Pattern Type`)plaid factor(`Pattern Type`)print
# NeckLine_vneck_dmy  0                           0
# factor(`Pattern Type`)solid factor(`Pattern Type`)splice
# NeckLine_vneck_dmy  0                           0
# factor(`Pattern Type`)striped SleeveLengththreequarter SleeveLengthsleeveless
# NeckLine_vneck_dmy  0                             0                        0
# SleeveLengthshort SleeveLengthhalfsleeve SleeveLengthfull
# NeckLine_vneck_dmy  0                 0                      0
# SleeveLengthcap_sleeves NeckLine_backless_dmy NeckLine_boatneck_dmy
# NeckLine_vneck_dmy  0                      -1                    -1
# NeckLine_bowneck_dmy NeckLine_halter_dmy NeckLine_mandarincollor_dmy
# NeckLine_vneck_dmy -1                   -1                  -1
# NeckLine_oneck_dmy NeckLine_open_dmy NeckLine_peterpancollor_dmy
# NeckLine_vneck_dmy -1                 -1                -1
# NeckLine_ruffled_dmy NeckLine_Scoop_dmy NeckLine_slashneck_dmy
# NeckLine_vneck_dmy -1                   -1                 -1
# NeckLine_squarecollor_dmy NeckLine_sweetheart_dmy NeckLine_turndowncollor_dmy
# NeckLine_vneck_dmy -1                        -1                      -1
#
```

```
task4lm <- (lm(log(TotalSales) ~ factor(Size) + factor(waiseline) +
                       factor(FabricType) + factor(Decoration) +
                       factor(`Pattern Type`)+
                       SleeveLengththreequarter + SleeveLengthsleeveless +
                       SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull+
                       SleeveLengthcap_sleeves +
                       NeckLine_backless_dmy + NeckLine_boatneck_dmy +
                       NeckLine_bowneck_dmy + NeckLine_halter_dmy +
NeckLine_mandarincollor_dmy +
                       NeckLine_oneck_dmy + NeckLine_open_dmy +
NeckLine_peterpancollor_dmy +
                       NeckLine_ruffled_dmy + NeckLine_Scoop_dmy +
NeckLine_slashneck_dmy +
                       NeckLine_squarecollor_dmy + NeckLine_sweetheart_dmy +
                       NeckLine_turndowncollor_dmy
               , data = ATT_DS3))

vif(task4lm)
# GVIF Df GVIF^(1/(2*Df))
# factor(Size)                  2.340937  4        1.112177
# factor(waiseline)             2.528051  4        1.122918
# factor(FabricType)           35.640902 17        1.110825
# factor(Decoration)          129.747578 24        1.106682
# factor(`Pattern Type`)        6.997702 13        1.077701
# SleeveLengththreequarter      8.204722  1        2.864389
# SleeveLengthsleeveless       33.948462  1        5.826531
# SleeveLengthshort            21.422838  1        4.628481
# SleeveLengthhalfsleeve        9.945179  1        3.153598
# SleeveLengthfull             22.041905  1        4.694881
# SleeveLengthcap_sleeves       2.355190  1        1.534663
# NeckLine_backless_dmy         2.040432  1        1.428437
# NeckLine_boatneck_dmy         1.423237  1        1.192995
# NeckLine_bowneck_dmy          1.257232  1        1.121264
# NeckLine_halter_dmy           2.033808  1        1.426117
# NeckLine_mandarincollor_dmy   1.078339  1        1.038431
# NeckLine_oneck_dmy            1.750278  1        1.322981
# NeckLine_open_dmy             1.297945  1        1.139274
# NeckLine_peterpancollor_dmy   1.127475  1        1.061826
# NeckLine_ruffled_dmy          1.151200  1        1.072940
# NeckLine_Scoop_dmy            4.169585  1        2.041956
# NeckLine_slashneck_dmy        1.310494  1        1.144768
# NeckLine_squarecollor_dmy     1.490239  1        1.220753
# NeckLine_sweetheart_dmy       1.941704  1        1.393450
# NeckLine_turndowncollor_dmy   1.217727  1        1.103507

task4lm <- update(task4lm, ~. - SleeveLengthsleeveless) # removing the vif >5
vif(task4lm)
# GVIF Df GVIF^(1/(2*Df))
# factor(Size)                  2.335705  4        1.111865
# factor(waiseline)             2.524140  4        1.122701
# factor(FabricType)           35.455822 17        1.110655
# factor(Decoration)          126.877531 24        1.106167
# factor(`Pattern Type`)        6.786349 13        1.076431
# SleeveLengththreequarter      1.357291  1        1.165028
# SleeveLengthshort             1.511971  1        1.229622
# SleeveLengthhalfsleeve        1.285793  1        1.133928
# SleeveLengthfull              1.449922  1        1.204127
# SleeveLengthcap_sleeves       1.063690  1        1.031353
# NeckLine_backless_dmy         2.040417  1        1.428432
# NeckLine_boatneck_dmy         1.422589  1        1.192724
# NeckLine_bowneck_dmy          1.255550  1        1.120513
# NeckLine_halter_dmy           2.033289  1        1.425934
# NeckLine_mandarincollor_dmy   1.078334  1        1.038429
# NeckLine_oneck_dmy            1.750260  1        1.322974
# NeckLine_open_dmy             1.297614  1        1.139129
# NeckLine_peterpancollor_dmy   1.127458  1        1.061818
# NeckLine_ruffled_dmy          1.148372  1        1.071621
# NeckLine_Scoop_dmy            4.168900  1        2.041788
# NeckLine_slashneck_dmy        1.308595  1        1.143938
# NeckLine_squarecollor_dmy     1.489482  1        1.220443
# NeckLine_sweetheart_dmy       1.938776  1        1.392399
```

```
 # NeckLine_turndowncollor_dmy   1.217720  1       1.103504
 # Now we are good have vif <5

 # Now we can continue the factor way of handling the categorical variables for lm
 # but it is more easier to remove the
 # insignificant variables if they were dummy variables -
 # hence lets convert these factors that are going to modeled as dummy
 # Before that let us also group (pool) the levels of the various factor/character
levels that
 # have insufficient
 # samples into other group as we can't do much with very few samples in the group.
 # This should have been done for the other groups which were already changed to dummy
values but
 # lets not change those for now as we have been using them so far as is. Just for
these
 # dress levels which are still character/factor we are going to convert now.
 # As a threshold, lets set 10 as the minimum count per group if they were to be
considered
 # as a group of their own else we will pool them to a value called other

 # table(ATT_DS3$Size) # not changing Size as all the levels have atleast 10 in each

 # free    L    M    S   XL
 # 173   96  177   38   15

 # table(ATT_DS3$waiseline) # lets change dropped and princess into a level called
other
 # dropped   empire  natural     null princess
 # 4      104      304       86        1

 # table(ATT_DS3$FabricType)
 # lets change batik , Corduroy, dobby, flannel, knitted, lace , organza, poplin,
 # satin, terry , tulle, woolen into a level called other
 #
 # batik broadcloth    chiffon   Corduroy      dobby    flannel     jersey    knitted
lace
 # 2         31        144          2           2          2         12          2
1
 # null    organza      other     poplin      satin      terry      tulle     woolen
worsted
 # 265          1          1          2           7          1          2          3
19


 # table(ATT_DS3$Decoration)
 # lets change button , cascading ,   crystal ,    draped ,embroidary ,  feathers,
flowers, none,
 # pearls, plain, pleat, pockets,rivet ,ruched, tassel , Tiered to a level called
other
 # applique    beading   bow     button  cascading    crystal     draped embroidary
feathers
 # 21        22         15          6          1          3          2          5
2
 # flowers  hollowout  lace       none       null     pearls      plain      pleat
pockets
 # 4        21         70          2        235          1          2          1
5
 # rivet     ruched    ruffles     sashes   sequined     tassel     Tiered
 # 3          3         17         42         14          1          1


 # table(ATT_DS3$`Pattern Type`)
 # lets change character, floral, geometric, leopard, none, plaid, splice into other
level
 # animal character    dot    floral geometric   leopard      none      null patchwork
 # 21        1        14        2         5         4         1       108        47
 # plaid     print     solid    splice   striped
 # 3        71       204         1        17


 summary(task4lm)
 # Call:
```

```
 #         lm(formula = log(TotalSales) ~ factor(Size) + factor(waiseline) +
 #                    factor(FabricType) + factor(Decoration) + factor(`Pattern Type`)
+
 #                    SleeveLengththreequarter + SleeveLengthshort +
SleeveLengthhalfsleeve +
 #                    SleeveLengthfull + SleeveLengthcap_sleeves +
NeckLine_backless_dmy +
 #                    NeckLine_boatneck_dmy + NeckLine_bowneck_dmy +
NeckLine_halter_dmy +
 #                    NeckLine_mandarincollor_dmy + NeckLine_oneck_dmy +
NeckLine_open_dmy +
 #                    NeckLine_peterpancollor_dmy + NeckLine_ruffled_dmy +
NeckLine_Scoop_dmy +
 #                    NeckLine_slashneck_dmy + NeckLine_squarecollor_dmy +
NeckLine_sweetheart_dmy +
 #                    NeckLine_turndowncollor_dmy, data = ATT_DS3)
 #


 # we are going to create SNO to merge the dataset again after stripping Dress ID and
Recommendation before undergoing the dummy variable creation
 ATT_DS3 <- ATT_DS3 %>% mutate(SNOMERGE = 1:n())
 TASK4_DS <- fastDummies::dummy_cols(dplyr::select(ATT_DS3, -
c(Dress_ID,Recommendation)) %>%
                rename(Pattern_Type=`Pattern Type`) %>%
                mutate(
                waiseline = ifelse(waiseline %in%
c("dropped","princess"),"other",waiseline),
                FabricType = ifelse(FabricType %in% c("batik","Corduroy",
                                                     "dobby", "flannel", "knitted",
                                                    "lace" , "organza", "poplin",
                                                    "satin", "terry" ,
                                                    "tulle", "woolen"),"other",
                                                    FabricType),
                Decoration = ifelse( Decoration %in% c("button" , "cascading" ,
                                                      "crystal" ,
                                                    "draped" ,"embroidary" ,
                                                    "feathers", "flowers","none",
                                                   "pearls", "plain", "pleat",
                                                   "pockets","rivet" ,
                                                    "ruched", "tassel" , "Tiered"),
                                                     "other",Decoration),
                Pattern_Type = ifelse(Pattern_Type %in%
                                             c("character", "floral", "geometric",
                                                "leopard", "none", "plaid",
"splice"),
                                                "other",Pattern_Type)

                                            ), remove_first_dummy = TRUE)
 TASK4_DS <- left_join(TASK4_DS
,dplyr::select(ATT_DS3,SNOMERGE,Dress_ID,Recommendation))

 #By default, dummy_cols() will make dummy variables from factor or character columns
only.
 #This is because in most cases those are the only types of data you want dummy
variables from.
 #If those are the only columns you want, then the function takes your data set as the
first
 #parameter and returns a data.frame with the newly created variables appended to the
end of the original data.
 #option for dummy_cols() is remove_first_dummy which by default is FALSE.
 #If TRUE, it removes the first dummy variable created from each column.
 #This is done to avoid multicollinearity in a multiple regression model caused by
included all dummy variables.

 task4lm <- lm(formula = log(TotalSales) ~ Size_L + Size_M +
    Size_S + Size_XL  +  waiseline_natural + waiseline_null +
            waiseline_other  + FabricType_chiffon + FabricType_jersey +
    FabricType_null + FabricType_other + FabricType_worsted +
            Decoration_beading +
    Decoration_bow  + Decoration_hollowout +
    Decoration_lace  + Decoration_null +
```

```
    Decoration_other + Decoration_ruffles + Decoration_sashes + Decoration_sequined +
    Pattern_Type_dot + Pattern_Type_other + Pattern_Type_null +
    Pattern_Type_patchwork + Pattern_Type_print +
    Pattern_Type_solid + Pattern_Type_striped +
                                SleeveLengththreequarter + SleeveLengthshort +
SleeveLengthhalfsleeve +
                                SleeveLengthfull + SleeveLengthcap_sleeves +
NeckLine_backless_dmy +
                                NeckLine_boatneck_dmy + NeckLine_bowneck_dmy +
NeckLine_halter_dmy +
                                NeckLine_mandarincollor_dmy + NeckLine_oneck_dmy +
NeckLine_open_dmy +
                                NeckLine_peterpancollor_dmy + NeckLine_ruffled_dmy +
NeckLine_Scoop_dmy +
                                NeckLine_slashneck_dmy + NeckLine_squarecollor_dmy +
NeckLine_sweetheart_dmy +
                                NeckLine_turndowncollor_dmy, data = TASK4_DS)

 vif(task4lm)
 # Size_L                      Size_M                       Size_S
 # 1.432387                    1.536966                     1.287723
 # Size_XL            waiseline_natural             waiseline_null
 # 1.184655                    1.798530                     1.881452
 # waiseline_other          FabricType_chiffon           FabricType_jersey
 # 1.133072                    4.741197                     1.476295
 # FabricType_null            FabricType_other          FabricType_worsted
 # 5.378450                    2.053439                     1.748252
 # Decoration_beading              Decoration_bow        Decoration_hollowout
 # 2.328712                    1.971131                     2.165474
 # Decoration_lace             Decoration_null            Decoration_other
 # 4.109546                    7.407472                     3.164365
 # Decoration_ruffles           Decoration_sashes         Decoration_sequined
 # 1.980412                    3.191992                     1.769016
 # Pattern_Type_dot           Pattern_Type_other          Pattern_Type_null
 # 1.929328                    1.900682                     5.407070
 # Pattern_Type_patchwork          Pattern_Type_print           Pattern_Type_solid
 # 3.295449                    4.150701                     7.196751
 # Pattern_Type_striped    SleeveLengththreequarter           SleeveLengthshort
 # 1.932287                    1.234200                     1.416055
 # SleeveLengthhalfsleeve            SleeveLengthfull     SleeveLengthcap_sleeves
 # 1.244727                    1.341314                     1.059962
 # NeckLine_backless_dmy       NeckLine_boatneck_dmy         NeckLine_bowneck_dmy
 # 1.062225                    1.232815                     1.192466
 # NeckLine_halter_dmy NeckLine_mandarincollor_dmy           NeckLine_oneck_dmy
 # 1.056616                    1.075163                     1.678604
 # NeckLine_open_dmy NeckLine_peterpancollor_dmy         NeckLine_ruffled_dmy
 # 1.100550                    1.121503                     1.147062
 # NeckLine_Scoop_dmy      NeckLine_slashneck_dmy   NeckLine_squarecollor_dmy
 # 1.123392                    1.218818                     1.154260
 # NeckLine_sweetheart_dmy NeckLine_turndowncollor_dmy
 # 1.408604                    1.172605

 #Removing high vif >5 variable out

 task4lm <- update(task4lm, ~. - Decoration_null)
 vif(task4lm)
 # Size_L                      Size_M                       Size_S
 # 1.432153                    1.536943                     1.285580
 # Size_XL            waiseline_natural             waiseline_null
 # 1.184613                    1.792448                     1.862631
 # waiseline_other          FabricType_chiffon           FabricType_jersey
 # 1.131783                    4.665047                     1.472194
 # FabricType_null            FabricType_other          FabricType_worsted
 # 5.256554                    2.046386                     1.746501
 # Decoration_beading              Decoration_bow        Decoration_hollowout
 # 1.333053                    1.305314                     1.157021
 # Decoration_lace            Decoration_other          Decoration_ruffles
 # 1.241436                    1.288986                     1.180160
 # Decoration_sashes          Decoration_sequined            Pattern_Type_dot
 # 1.238657                    1.133247                     1.908603
 # Pattern_Type_other            Pattern_Type_null      Pattern_Type_patchwork
 # 1.900682                    5.401288                     3.294295
```

```
# Pattern_Type_print          Pattern_Type_solid        Pattern_Type_striped
# 4.138395                    7.181514                    1.932034
# SleeveLengththreequarter          SleeveLengthshort      SleeveLengthhalfsleeve
# 1.230709                    1.414643                    1.244706
# SleeveLengthfull     SleeveLengthcap_sleeves       NeckLine_backless_dmy
# 1.331672                    1.059877                    1.062210
# NeckLine_boatneck_dmy        NeckLine_bowneck_dmy         NeckLine_halter_dmy
# 1.216127                    1.188428                    1.056391
# NeckLine_mandarincollor_dmy         NeckLine_oneck_dmy           NeckLine_open_dmy
# 1.075102                    1.677717                    1.082504
# NeckLine_peterpancollor_dmy       NeckLine_ruffled_dmy          NeckLine_Scoop_dmy
# 1.118218                    1.146450                    1.097606
# NeckLine_slashneck_dmy   NeckLine_squarecollor_dmy     NeckLine_sweetheart_dmy
# 1.218781                    1.152754                    1.408486
# NeckLine_turndowncollor_dmy
# 1.170170

task4lm <- update(task4lm, ~. -Pattern_Type_solid)
vif(task4lm)
# Size_L                      Size_M                      Size_S
# 1.431129                    1.534607                    1.285561
# Size_XL           waiseline_natural              waiseline_null
# 1.183945                    1.772984                    1.857362
# waiseline_other          FabricType_chiffon          FabricType_jersey
# 1.130897                    4.655324                    1.468109
# FabricType_null             FabricType_other          FabricType_worsted
# 5.250691                    2.037080                    1.746492
# Decoration_beading              Decoration_bow        Decoration_hollowout
# 1.329165                    1.303053                    1.154039
# Decoration_lace             Decoration_other          Decoration_ruffles
# 1.240753                    1.276107                    1.158506
# Decoration_sashes          Decoration_sequined            Pattern_Type_dot
# 1.221615                    1.128492                    1.343024
# Pattern_Type_other            Pattern_Type_null      Pattern_Type_patchwork
# 1.121031                    1.586395                    1.238263
# Pattern_Type_print         Pattern_Type_striped    SleeveLengththreequarter
# 1.273995                    1.134430                    1.230626
# SleeveLengthshort      SleeveLengthhalfsleeve            SleeveLengthfull
# 1.411245                    1.240625                    1.325156
# SleeveLengthcap_sleeves       NeckLine_backless_dmy       NeckLine_boatneck_dmy
# 1.059211                    1.061926                    1.214714
# NeckLine_bowneck_dmy          NeckLine_halter_dmy NeckLine_mandarincollor_dmy
# 1.188038                    1.055334                    1.074958
# NeckLine_oneck_dmy            NeckLine_open_dmy NeckLine_peterpancollor_dmy
# 1.666802                    1.081935                    1.117761
# NeckLine_ruffled_dmy          NeckLine_Scoop_dmy      NeckLine_slashneck_dmy
# 1.146380                    1.097594                    1.217812
# NeckLine_squarecollor_dmy     NeckLine_sweetheart_dmy NeckLine_turndowncollor_dmy
# 1.152707                    1.407439                    1.169171

task4lm <- update(task4lm, ~. - FabricType_null)

vif(task4lm)
# Size_L                      Size_M                      Size_S
# 1.423209                    1.520799                    1.279703
# Size_XL           waiseline_natural              waiseline_null
# 1.182514                    1.756499                    1.851313
# waiseline_other          FabricType_chiffon          FabricType_jersey
# 1.130793                    1.244453                    1.086278
# FabricType_other          FabricType_worsted          Decoration_beading
# 1.184775                    1.147812                    1.298203
# Decoration_bow        Decoration_hollowout             Decoration_lace
# 1.302618                    1.152715                    1.235841
# Decoration_other          Decoration_ruffles           Decoration_sashes
# 1.274963                    1.157802                    1.220103
# Decoration_sequined            Pattern_Type_dot          Pattern_Type_other
# 1.113802                    1.343023                    1.120951
# Pattern_Type_null      Pattern_Type_patchwork          Pattern_Type_print
# 1.585991                    1.238135                    1.263320
# Pattern_Type_striped    SleeveLengththreequarter           SleeveLengthshort
# 1.132847                    1.229244                    1.404493
# SleeveLengthhalfsleeve            SleeveLengthfull     SleeveLengthcap_sleeves
```

```
 # 1.201575                    1.315696                    1.058704
 # NeckLine_backless_dmy       NeckLine_boatneck_dmy        NeckLine_bowneck_dmy
 # 1.061902                    1.214600                    1.186221
 # NeckLine_halter_dmy NeckLine_mandarincollor_dmy          NeckLine_oneck_dmy
 # 1.053751                    1.072103                    1.651747
 # NeckLine_open_dmy NeckLine_peterpancollor_dmy        NeckLine_ruffled_dmy
 # 1.078326                    1.115863                    1.146354
 # NeckLine_Scoop_dmy      NeckLine_slashneck_dmy   NeckLine_squarecollor_dmy
 # 1.097109                    1.215168                    1.150057
 # NeckLine_sweetheart_dmy NeckLine_turndowncollor_dmy
 # 1.400279                    1.168974
 # now since we have vif <5, we are free from multicollinearity and we can proceed in
the modeling

 summary(task4lm)

 #
 # Call:
 #         lm(formula = log(TotalSales) ~ Size_L + Size_M + Size_S + Size_XL +
 #                    waiseline_natural + waiseline_null + waiseline_other +
FabricType_chiffon +
 #                    FabricType_jersey + FabricType_other + FabricType_worsted +
 #                    Decoration_beading + Decoration_bow + Decoration_hollowout +
 #                    Decoration_lace + Decoration_other + Decoration_ruffles +
 #                    Decoration_sashes + Decoration_sequined + Pattern_Type_dot +
 #                    Pattern_Type_other + Pattern_Type_null + Pattern_Type_patchwork
+
 #                    Pattern_Type_print + Pattern_Type_striped +
SleeveLengththreequarter +
 #                    SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
 #                    SleeveLengthcap_sleeves + NeckLine_backless_dmy +
NeckLine_boatneck_dmy +
 #                    NeckLine_bowneck_dmy + NeckLine_halter_dmy +
NeckLine_mandarincollor_dmy +
 #                    NeckLine_oneck_dmy + NeckLine_open_dmy +
NeckLine_peterpancollor_dmy +
 #                    NeckLine_ruffled_dmy + NeckLine_Scoop_dmy +
NeckLine_slashneck_dmy +
 #                    NeckLine_squarecollor_dmy + NeckLine_sweetheart_dmy +
NeckLine_turndowncollor_dmy,
 #            data = TASK4_DS)
 #
 # Residuals:
 #         Min      1Q  Median      3Q     Max
 # -5.6894 -0.8307  0.1907  1.0211  3.6866
 #
 # Coefficients:
 #         Estimate Std. Error t value Pr(>|t|)
 # (Intercept)                 7.2175387  0.2727557  26.462  < 2e-16 ***
 #         Size_L                      0.7042388  0.2140516   3.290 0.001080 **
 #         Size_M                      0.3814969  0.1823029   2.093 0.036935 *
 #         Size_S                      0.2502171  0.3016370   0.830 0.407240
 # Size_XL                     0.2617063  0.4504089   0.581 0.561501
 # waiseline_natural           0.1430612  0.1921065   0.745 0.456840
 # waiseline_null              0.1855573  0.2547932   0.728 0.466826
 # waiseline_other             0.4043898  0.7551188   0.536 0.592545
 # FabricType_chiffon          0.1795170  0.1741270   1.031 0.303111
 # FabricType_jersey          -0.1253488  0.4811578  -0.261 0.794585
 # FabricType_other           -0.2694214  0.3345035  -0.805 0.420989
 # FabricType_worsted         -0.0079227  0.3959226  -0.020 0.984044
 # Decoration_beading         -0.0704926  0.3925302  -0.180 0.857558
 # Decoration_bow              0.3094092  0.4727293   0.655 0.513110
 # Decoration_hollowout       -0.2944195  0.3781896  -0.778 0.436682
 # Decoration_lace            -0.0414212  0.2263997  -0.183 0.854914
 # Decoration_other           -0.1740817  0.2876327  -0.605 0.545334
 # Decoration_ruffles         -0.2449926  0.4195092  -0.584 0.559512
 # Decoration_sashes           0.2140791  0.2813764   0.761 0.447155
 # Decoration_sequined        -0.2607449  0.4520030  -0.577 0.564316
 # Pattern_Type_dot            0.0448991  0.4963397   0.090 0.927961
 # Pattern_Type_other          0.2106234  0.4127790   0.510 0.610119
 # Pattern_Type_null          -0.7904660  0.2162830  -3.655 0.000287 ***
 #         Pattern_Type_patchwork      -0.4031408  0.2694251  -1.496 0.135270
```

```
# Pattern_Type_print            0.0673335  0.2275506   0.296 0.767437
# Pattern_Type_striped         -0.0415812  0.4149635  -0.100 0.920226
# SleeveLengththreequarter     -0.3528284  0.3407233  -1.036 0.300974
# SleeveLengthshort             0.2415244  0.2126395   1.136 0.256623
# SleeveLengthhalfsleeve        0.1942378  0.2996442   0.648 0.517165
# SleeveLengthfull              0.0259397  0.2058078   0.126 0.899757
# SleeveLengthcap_sleeves      -0.2899727  0.7306529  -0.397 0.691651
# NeckLine_backless_dmy         1.5358116  1.6296701   0.942 0.346486
# NeckLine_boatneck_dmy        -0.3194972  0.4072784  -0.784 0.433174
# NeckLine_bowneck_dmy         -0.9631927  0.5496681  -1.752 0.080394 .
# NeckLine_halter_dmy           1.3682439  1.6234035   0.843 0.399771
# NeckLine_mandarincollor_dmy -2.0889568  1.6374790  -1.276 0.202709
# NeckLine_oneck_dmy            0.3986382  0.1826032   2.183 0.029541 *
#         NeckLine_open_dmy           -0.4976577  0.9500486  -0.524 0.600657
# NeckLine_peterpancollor_dmy -0.0756365  0.6854543  -0.110 0.912184
# NeckLine_ruffled_dmy          3.5827580  1.6932338   2.116 0.034895 *
#         NeckLine_Scoop_dmy           0.6773712  1.1724757   0.578 0.563735
# NeckLine_slashneck_dmy        0.0456119  0.3573807   0.128 0.898500
# NeckLine_squarecollor_dmy     0.2929464  0.7615235   0.385 0.700651
# NeckLine_sweetheart_dmy      -0.0001764  0.4901300   0.000 0.999713
# NeckLine_turndowncollor_dmy  0.1285134  0.4800481   0.268 0.789045
# ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.58 on 454 degrees of freedom
# Multiple R-squared:  0.1467,  Adjusted R-squared:  0.06398
# F-statistic: 1.774 on 44 and 454 DF,  p-value: 0.002304

task4lm.1 <- update(task4lm, ~. -NeckLine_sweetheart_dmy)
summary(task4lm.1)

#
# Call:
#         lm(formula = log(TotalSales) ~ Size_L + Size_M + Size_S + Size_XL +
#                    waiseline_natural + waiseline_null + waiseline_other +
FabricType_chiffon +
#                    FabricType_jersey + FabricType_other + FabricType_worsted +
#                    Decoration_beading + Decoration_bow + Decoration_hollowout +
#                    Decoration_lace + Decoration_other + Decoration_ruffles +
#                    Decoration_sashes + Decoration_sequined + Pattern_Type_dot +
#                    Pattern_Type_other + Pattern_Type_null + Pattern_Type_patchwork
+
#                    Pattern_Type_print + Pattern_Type_striped +
SleeveLengththreequarter +
#                    SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
#                    SleeveLengthcap_sleeves + NeckLine_backless_dmy +
NeckLine_boatneck_dmy +
#                    NeckLine_bowneck_dmy + NeckLine_halter_dmy +
NeckLine_mandarincollor_dmy +
#                    NeckLine_oneck_dmy + NeckLine_open_dmy +
NeckLine_peterpancollor_dmy +
#                    NeckLine_ruffled_dmy + NeckLine_Scoop_dmy +
NeckLine_slashneck_dmy +
#                    NeckLine_squarecollor_dmy + NeckLine_turndowncollor_dmy,
#            data = TASK4_DS)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -5.6894 -0.8307  0.1907  1.0211  3.6866
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                 7.217528   0.270763  26.656  < 2e-16 ***
#         Size_L                      0.704238   0.213792   3.294 0.001065 **
#         Size_M                      0.381492   0.181513   2.102 0.036126 *
#         Size_S                      0.250221   0.301118   0.831 0.406424
# Size_XL                     0.261707   0.449912   0.582 0.561068
# waiseline_natural           0.143063   0.191835   0.746 0.456197
# waiseline_null              0.185565   0.253580   0.732 0.464678
# waiseline_other             0.404371   0.752416   0.537 0.591234
# FabricType_chiffon          0.179514   0.173765   1.033 0.302112
# FabricType_jersey          -0.125348   0.480627  -0.261 0.794363
```

```
# FabricType_other           -0.269428   0.333661  -0.807 0.419807
# FabricType_worsted         -0.007927   0.395338  -0.020 0.984012
# Decoration_beading         -0.070534   0.374820  -0.188 0.850818
# Decoration_bow              0.309410   0.472208   0.655 0.512645
# Decoration_hollowout       -0.294422   0.377715  -0.779 0.436101
# Decoration_lace            -0.041418   0.226018  -0.183 0.854681
# Decoration_other           -0.174097   0.284330  -0.612 0.540642
# Decoration_ruffles         -0.244990   0.418968  -0.585 0.559009
# Decoration_sashes           0.214077   0.281005   0.762 0.446558
# Decoration_sequined        -0.260744   0.451499  -0.578 0.563882
# Pattern_Type_dot            0.044893   0.495479   0.091 0.927847
# Pattern_Type_other          0.210618   0.412038   0.511 0.609486
# Pattern_Type_null          -0.790482   0.211497  -3.738 0.000209 ***
#        Pattern_Type_patchwork      -0.403143   0.269083  -1.498 0.134772
# Pattern_Type_print          0.067334   0.227299   0.296 0.767186
# Pattern_Type_striped       -0.041575   0.414148  -0.100 0.920081
# SleeveLengththreequarter   -0.352820   0.339619  -1.039 0.299417
# SleeveLengthshort           0.241533   0.210947   1.145 0.252813
# SleeveLengthhalfsleeve      0.194246   0.298455   0.651 0.515479
# SleeveLengthfull            0.025947   0.204706   0.127 0.899194
# SleeveLengthcap_sleeves    -0.289961   0.729088  -0.398 0.691035
# NeckLine_backless_dmy       1.535856   1.623191   0.946 0.344551
# NeckLine_boatneck_dmy      -0.319489   0.406147  -0.787 0.431906
# NeckLine_bowneck_dmy       -0.963183   0.548385  -1.756 0.079693 .
# NeckLine_halter_dmy         1.368284   1.617820   0.846 0.398132
# NeckLine_mandarincollor_dmy -2.088948   1.635511  -1.277 0.202167
# NeckLine_oneck_dmy          0.398650   0.179374   2.222 0.026743 *
#        NeckLine_open_dmy           -0.497656   0.948990  -0.524 0.600252
# NeckLine_peterpancollor_dmy -0.075626   0.684127  -0.111 0.912026
# NeckLine_ruffled_dmy        3.582774   1.690787   2.119 0.034632 *
#        NeckLine_Scoop_dmy           0.677400   1.168538   0.580 0.562405
# NeckLine_slashneck_dmy      0.045632   0.352797   0.129 0.897144
# NeckLine_squarecollor_dmy   0.292963   0.759204   0.386 0.699764
# NeckLine_turndowncollor_dmy  0.128521   0.479107   0.268 0.788628
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.578 on 455 degrees of freedom
# Multiple R-squared:  0.1467,  Adjusted R-squared:  0.06604
# F-statistic: 1.819 on 43 and 455 DF,  p-value: 0.001655


task4lm.2 <- update(task4lm.1, ~. - FabricType_worsted)
summary(task4lm.2)

# Call:
#        lm(formula = log(TotalSales) ~ Size_L + Size_M + Size_S + Size_XL +
#                   waiseline_natural + waiseline_null + waiseline_other +
FabricType_chiffon +
#                   FabricType_jersey + FabricType_other + Decoration_beading +
#                   Decoration_bow + Decoration_hollowout + Decoration_lace +
#                   Decoration_other + Decoration_ruffles + Decoration_sashes +
#                   Decoration_sequined + Pattern_Type_dot + Pattern_Type_other +
#                   Pattern_Type_null + Pattern_Type_patchwork + Pattern_Type_print
+
#                   Pattern_Type_striped + SleeveLengththreequarter +
SleeveLengthshort +
#                   SleeveLengthhalfsleeve + SleeveLengthfull +
SleeveLengthcap_sleeves +
#                   NeckLine_backless_dmy + NeckLine_boatneck_dmy +
NeckLine_bowneck_dmy +
#                   NeckLine_halter_dmy + NeckLine_mandarincollor_dmy +
NeckLine_oneck_dmy +
#                   NeckLine_open_dmy + NeckLine_peterpancollor_dmy +
NeckLine_ruffled_dmy +
#                   NeckLine_Scoop_dmy + NeckLine_slashneck_dmy +
NeckLine_squarecollor_dmy +
#                   NeckLine_turndowncollor_dmy, data = TASK4_DS)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -5.6884 -0.8306  0.1911  1.0216  3.6868
```

```
 #
 # Coefficients:
 #         Estimate Std. Error t value Pr(>|t|)
 # (Intercept)                 7.21703    0.26931  26.798  < 2e-16 ***
 #         Size_L                      0.70437    0.21345   3.300 0.001043 **
 #         Size_M                      0.38183    0.18051   2.115 0.034950 *
 #         Size_S                      0.25068    0.29991   0.836 0.403682
 # Size_XL                     0.26140    0.44916   0.582 0.560870
 # waiseline_natural           0.14293    0.19152   0.746 0.455855
 # waiseline_null              0.18568    0.25323   0.733 0.463778
 # waiseline_other             0.40474    0.75136   0.539 0.590371
 # FabricType_chiffon          0.17985    0.17276   1.041 0.298406
 # FabricType_jersey          -0.12493    0.47965  -0.260 0.794625
 # FabricType_other           -0.26896    0.33250  -0.809 0.418979
 # Decoration_beading         -0.07049    0.37440  -0.188 0.850739
 # Decoration_bow              0.30978    0.47133   0.657 0.511352
 # Decoration_hollowout       -0.29415    0.37706  -0.780 0.435728
 # Decoration_lace            -0.04207    0.22346  -0.188 0.850768
 # Decoration_other           -0.17393    0.28389  -0.613 0.540413
 # Decoration_ruffles         -0.24486    0.41846  -0.585 0.558739
 # Decoration_sashes           0.21428    0.28052   0.764 0.445347
 # Decoration_sequined        -0.26084    0.45098  -0.578 0.563282
 # Pattern_Type_dot            0.04405    0.49316   0.089 0.928861
 # Pattern_Type_other          0.21112    0.41084   0.514 0.607590
 # Pattern_Type_null          -0.79030    0.21108  -3.744 0.000204 ***
 #         Pattern_Type_patchwork      -0.40370    0.26736  -1.510 0.131756
 # Pattern_Type_print          0.06739    0.22703   0.297 0.766743
 # Pattern_Type_striped       -0.04269    0.40996  -0.104 0.917112
 # SleeveLengththreequarter   -0.35251    0.33889  -1.040 0.298806
 # SleeveLengthshort           0.24180    0.21029   1.150 0.250805
 # SleeveLengthhalfsleeve      0.19424    0.29813   0.652 0.515032
 # SleeveLengthfull            0.02574    0.20423   0.126 0.899754
 # SleeveLengthcap_sleeves    -0.28950    0.72793  -0.398 0.691033
 # NeckLine_backless_dmy       1.53567    1.62138   0.947 0.344071
 # NeckLine_boatneck_dmy      -0.31937    0.40566  -0.787 0.431522
 # NeckLine_bowneck_dmy       -0.96275    0.54737  -1.759 0.079268 .
 # NeckLine_halter_dmy         1.36856    1.61599   0.847 0.397500
 # NeckLine_mandarincollor_dmy -2.08885    1.63371  -1.279 0.201690
 # NeckLine_oneck_dmy          0.39861    0.17917   2.225 0.026583 *
 #         NeckLine_open_dmy           -0.49730    0.94778  -0.525 0.600048
 # NeckLine_peterpancollor_dmy -0.07543    0.68331  -0.110 0.912151
 # NeckLine_ruffled_dmy        3.58280    1.68893   2.121 0.034432 *
 #         NeckLine_Scoop_dmy           0.67680    1.16687   0.580 0.562194
 # NeckLine_slashneck_dmy      0.04584    0.35226   0.130 0.896517
 # NeckLine_squarecollor_dmy   0.29307    0.75835   0.386 0.699336
 # NeckLine_turndowncollor_dmy  0.12825    0.47839   0.268 0.788757
 # ---
 #         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 #
 # Residual standard error: 1.576 on 456 degrees of freedom
 # Multiple R-squared:  0.1467,  Adjusted R-squared:  0.06808
 # F-statistic: 1.866 on 42 and 456 DF,  p-value: 0.001174

 task4lm.3 <- update(task4lm.2, ~. - Pattern_Type_dot)
 summary(task4lm.3)

 #
 # Call:
 #         lm(formula = log(TotalSales) ~ Size_L + Size_M + Size_S + Size_XL +
 #                    waiseline_natural + waiseline_null + waiseline_other +
FabricType_chiffon +
 #                    FabricType_jersey + FabricType_other + Decoration_beading +
 #                    Decoration_bow + Decoration_hollowout + Decoration_lace +
 #                    Decoration_other + Decoration_ruffles + Decoration_sashes +
 #                    Decoration_sequined + Pattern_Type_other + Pattern_Type_null +
 #                    Pattern_Type_patchwork + Pattern_Type_print +
Pattern_Type_striped +
 #                    SleeveLengththreequarter + SleeveLengthshort +
SleeveLengthhalfsleeve +
 #                    SleeveLengthfull + SleeveLengthcap_sleeves +
NeckLine_backless_dmy +
```

```
#                    NeckLine_boatneck_dmy + NeckLine_bowneck_dmy +
NeckLine_halter_dmy +
#                    NeckLine_mandarincollor_dmy + NeckLine_oneck_dmy +
NeckLine_open_dmy +
#                    NeckLine_peterpancollor_dmy + NeckLine_ruffled_dmy +
NeckLine_Scoop_dmy +
#                    NeckLine_slashneck_dmy + NeckLine_squarecollor_dmy +
NeckLine_turndowncollor_dmy,
#            data = TASK4_DS)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -5.6872 -0.8279  0.1911  1.0196  3.6881
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)                 7.21723    0.26901  26.829  < 2e-16 ***
#        Size_L                      0.70451    0.21322   3.304 0.001027 **
#        Size_M                      0.38285    0.17996   2.127 0.033922 *
#        Size_S                      0.25297    0.29849   0.847 0.397165
# Size_XL                     0.26266    0.44845   0.586 0.558356
# waiseline_natural           0.14354    0.19119   0.751 0.453189
# waiseline_null              0.18709    0.25247   0.741 0.459054
# waiseline_other             0.40584    0.75045   0.541 0.588909
# FabricType_chiffon          0.18106    0.17204   1.052 0.293152
# FabricType_jersey          -0.12533    0.47910  -0.262 0.793745
# FabricType_other           -0.26957    0.33207  -0.812 0.417333
# Decoration_beading         -0.07144    0.37385  -0.191 0.848541
# Decoration_bow              0.31977    0.45738   0.699 0.484829
# Decoration_hollowout       -0.29608    0.37604  -0.787 0.431482
# Decoration_lace            -0.04343    0.22269  -0.195 0.845447
# Decoration_other           -0.17633    0.28230  -0.625 0.532525
# Decoration_ruffles         -0.24266    0.41728  -0.582 0.561170
# Decoration_sashes           0.21177    0.27881   0.760 0.447913
# Decoration_sequined        -0.26187    0.45034  -0.582 0.561187
# Pattern_Type_other          0.20913    0.40979   0.510 0.610055
# Pattern_Type_null          -0.79245    0.20948  -3.783 0.000176 ***
#        Pattern_Type_patchwork      -0.40706    0.26441  -1.539 0.124375
# Pattern_Type_print          0.06433    0.22419   0.287 0.774294
# Pattern_Type_striped       -0.04420    0.40916  -0.108 0.914014
# SleeveLengththreequarter   -0.35082    0.33799  -1.038 0.299847
# SleeveLengthshort           0.24489    0.20721   1.182 0.237876
# SleeveLengthhalfsleeve      0.19449    0.29779   0.653 0.514023
# SleeveLengthfull            0.02568    0.20400   0.126 0.899879
# SleeveLengthcap_sleeves    -0.28926    0.72713  -0.398 0.690952
# NeckLine_backless_dmy       1.53881    1.61924   0.950 0.342448
# NeckLine_boatneck_dmy      -0.31693    0.40430  -0.784 0.433503
# NeckLine_bowneck_dmy       -0.96215    0.54673  -1.760 0.079106 .
# NeckLine_halter_dmy         1.37231    1.61369   0.850 0.395536
# NeckLine_mandarincollor_dmy -2.08842    1.63193  -1.280 0.201292
# NeckLine_oneck_dmy          0.39925    0.17883   2.233 0.026057 *
#        NeckLine_open_dmy          -0.49877    0.94661  -0.527 0.598516
# NeckLine_peterpancollor_dmy -0.07648    0.68246  -0.112 0.910824
# NeckLine_ruffled_dmy        3.61135    1.65661   2.180 0.029770 *
#        NeckLine_Scoop_dmy          0.67521    1.16547   0.579 0.562642
# NeckLine_slashneck_dmy      0.04508    0.35177   0.128 0.898096
# NeckLine_squarecollor_dmy   0.30301    0.74934   0.404 0.686135
# NeckLine_turndowncollor_dmy  0.12698    0.47766   0.266 0.790480
#
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.575 on 457 degrees of freedom
# Multiple R-squared:  0.1467,  Adjusted R-squared:  0.07011
# F-statistic: 1.916 on 41 and 457 DF,  p-value: 0.0008245

task4lm.4 <- update(task4lm.3, ~. -Pattern_Type_striped)
summary(task4lm.4)


# Call:
#        lm(formula = log(TotalSales) ~ Size_L + Size_M + Size_S + Size_XL +
```

```
 #                    waiseline_natural + waiseline_null + waiseline_other +
FabricType_chiffon +
 #                    FabricType_jersey + FabricType_other + Decoration_beading +
 #                    Decoration_bow + Decoration_hollowout + Decoration_lace +
 #                    Decoration_other + Decoration_ruffles + Decoration_sashes +
 #                    Decoration_sequined + Pattern_Type_other + Pattern_Type_null +
 #                    Pattern_Type_patchwork + Pattern_Type_print +
SleeveLengththreequarter +
 #                    SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
 #                    SleeveLengthcap_sleeves + NeckLine_backless_dmy +
NeckLine_boatneck_dmy +
 #                    NeckLine_bowneck_dmy + NeckLine_halter_dmy +
NeckLine_mandarincollor_dmy +
 #                    NeckLine_oneck_dmy + NeckLine_open_dmy +
NeckLine_peterpancollor_dmy +
 #                    NeckLine_ruffled_dmy + NeckLine_Scoop_dmy +
NeckLine_slashneck_dmy +
 #                    NeckLine_squarecollor_dmy + NeckLine_turndowncollor_dmy,
 #            data = TASK4_DS)
 #
 # Residuals:
 #         Min      1Q  Median      3Q     Max
 # -5.6843 -0.8266  0.1864  1.0147  3.6853
 #
 # Coefficients:
 #         Estimate Std. Error t value Pr(>|t|)
 # (Intercept)                 7.21148    0.26341  27.377  < 2e-16 ***
 #         Size_L                      0.70611    0.21247   3.323 0.000961 ***
 #         Size_M                      0.38339    0.17970   2.134 0.033409 *
 #         Size_S                      0.25503    0.29756   0.857 0.391846
 # Size_XL                     0.26400    0.44779   0.590 0.555779
 # waiseline_natural           0.14262    0.19080   0.748 0.455137
 # waiseline_null              0.18806    0.25204   0.746 0.455945
 # waiseline_other             0.40594    0.74964   0.542 0.588414
 # FabricType_chiffon          0.18017    0.17166   1.050 0.294456
 # FabricType_jersey          -0.12350    0.47828  -0.258 0.796362
 # FabricType_other           -0.27002    0.33168  -0.814 0.416007
 # Decoration_beading         -0.06829    0.37231  -0.183 0.854544
 # Decoration_bow              0.32243    0.45623   0.707 0.480096
 # Decoration_hollowout       -0.29239    0.37408  -0.782 0.434845
 # Decoration_lace            -0.04108    0.22138  -0.186 0.852882
 # Decoration_other           -0.17591    0.28197  -0.624 0.533032
 # Decoration_ruffles         -0.24153    0.41670  -0.580 0.562454
 # Decoration_sashes           0.21265    0.27839   0.764 0.445359
 # Decoration_sequined        -0.25787    0.44833  -0.575 0.565450
 # Pattern_Type_other          0.21171    0.40865   0.518 0.604661
 # Pattern_Type_null          -0.78900    0.20680  -3.815 0.000155 ***
 #         Pattern_Type_patchwork      -0.40466    0.26319  -1.538 0.124859
 # Pattern_Type_print          0.06795    0.22143   0.307 0.759062
 # SleeveLengththreequarter   -0.34887    0.33715  -1.035 0.301324
 # SleeveLengthshort           0.24730    0.20577   1.202 0.230057
 # SleeveLengthhalfsleeve      0.19544    0.29734   0.657 0.511328
 # SleeveLengthfull            0.02759    0.20302   0.136 0.891973
 # SleeveLengthcap_sleeves    -0.28812    0.72627  -0.397 0.691767
 # NeckLine_backless_dmy       1.54158    1.61729   0.953 0.341000
 # NeckLine_boatneck_dmy      -0.31585    0.40374  -0.782 0.434441
 # NeckLine_bowneck_dmy       -0.95978    0.54570  -1.759 0.079278 .
 # NeckLine_halter_dmy         1.37510    1.61174   0.853 0.394007
 # NeckLine_mandarincollor_dmy -2.08412    1.62969  -1.279 0.201597
 # NeckLine_oneck_dmy          0.39991    0.17853   2.240 0.025567 *
 # NeckLine_open_dmy          -0.49888    0.94559  -0.528 0.598040
 # NeckLine_peterpancollor_dmy -0.07569    0.68169  -0.111 0.911637
 # NeckLine_ruffled_dmy        3.61237    1.65479   2.183 0.029545 *
 # NeckLine_Scoop_dmy          0.67426    1.16418   0.579 0.562758
 # NeckLine_slashneck_dmy      0.04593    0.35130   0.131 0.896034
 # NeckLine_squarecollor_dmy   0.30440    0.74842   0.407 0.684399
 # NeckLine_turndowncollor_dmy  0.12976    0.47645   0.272 0.785471
 # ---
 #         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 #
 # Residual standard error: 1.573 on 458 degrees of freedom
 # Multiple R-squared:  0.1466,  Adjusted R-squared:  0.07211
```

```
# F-statistic: 1.968 on 40 and 458 DF,  p-value: 0.000572

task4lm.5 <- update(task4lm.4, ~. -NeckLine_peterpancollor_dmy)
summary(task4lm.5)

# Call:
#         lm(formula = log(TotalSales) ~ Size_L + Size_M + Size_S + Size_XL +
#                     waiseline_natural + waiseline_null + waiseline_other +
FabricType_chiffon +
#                     FabricType_jersey + FabricType_other + Decoration_beading +
#                     Decoration_bow + Decoration_hollowout + Decoration_lace +
#                     Decoration_other + Decoration_ruffles + Decoration_sashes +
#                     Decoration_sequined + Pattern_Type_other + Pattern_Type_null +
#                     Pattern_Type_patchwork + Pattern_Type_print +
SleeveLengththreequarter +
#                     SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
#                     SleeveLengthcap_sleeves + NeckLine_backless_dmy +
NeckLine_boatneck_dmy +
#                     NeckLine_bowneck_dmy + NeckLine_halter_dmy +
NeckLine_mandarincollor_dmy +
#                     NeckLine_oneck_dmy + NeckLine_open_dmy + NeckLine_ruffled_dmy +
#                     NeckLine_Scoop_dmy + NeckLine_slashneck_dmy +
NeckLine_squarecollor_dmy +
#                     NeckLine_turndowncollor_dmy, data = TASK4_DS)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -5.6840 -0.8261  0.1862  1.0166  3.6939
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                7.20766    0.26087  27.629  < 2e-16 ***
#         Size_L                     0.70743    0.21191   3.338 0.000911 ***
#         Size_M                     0.38338    0.17950   2.136 0.033225 *
#         Size_S                     0.25685    0.29679   0.865 0.387260
# Size_XL                    0.26625    0.44685   0.596 0.551573
# waiseline_natural          0.14364    0.19037   0.755 0.450920
# waiseline_null             0.18857    0.25172   0.749 0.454159
# waiseline_other            0.40916    0.74827   0.547 0.584780
# FabricType_chiffon         0.18045    0.17145   1.052 0.293140
# FabricType_jersey         -0.12322    0.47776  -0.258 0.796590
# FabricType_other          -0.26834    0.33098  -0.811 0.417929
# Decoration_beading        -0.06637    0.37151  -0.179 0.858292
# Decoration_bow             0.32341    0.45565   0.710 0.478205
# Decoration_hollowout      -0.29116    0.37352  -0.780 0.436078
# Decoration_lace           -0.04028    0.22103  -0.182 0.855472
# Decoration_other          -0.17471    0.28146  -0.621 0.535089
# Decoration_ruffles        -0.23939    0.41580  -0.576 0.565087
# Decoration_sashes          0.21335    0.27802   0.767 0.443241
# Decoration_sequined       -0.25641    0.44765  -0.573 0.567069
# Pattern_Type_other         0.21252    0.40815   0.521 0.602834
# Pattern_Type_null         -0.78936    0.20656  -3.822 0.000151 ***
#         Pattern_Type_patchwork      -0.40723    0.26188  -1.555 0.120630
# Pattern_Type_print         0.06819    0.22118   0.308 0.758003
# SleeveLengththreequarter  -0.35648    0.32975  -1.081 0.280232
# SleeveLengthshort          0.24540    0.20484   1.198 0.231530
# SleeveLengthhalfsleeve     0.19517    0.29701   0.657 0.511440
# SleeveLengthfull           0.02757    0.20280   0.136 0.891937
# SleeveLengthcap_sleeves   -0.28711    0.72543  -0.396 0.692453
# NeckLine_backless_dmy      1.54428    1.61537   0.956 0.339579
# NeckLine_boatneck_dmy     -0.31287    0.40241  -0.777 0.437277
# NeckLine_bowneck_dmy      -0.95749    0.54472  -1.758 0.079455 .
# NeckLine_halter_dmy        1.37706    1.60991   0.855 0.392793
# NeckLine_mandarincollor_dmy -2.08175    1.62779  -1.279 0.201586
# NeckLine_oneck_dmy         0.40290    0.17630   2.285 0.022753 *
#         NeckLine_open_dmy           -0.49646    0.94432  -0.526 0.599325
# NeckLine_ruffled_dmy       3.61684    1.65252   2.189 0.029123 *
#         NeckLine_Scoop_dmy           0.67749    1.16256   0.583 0.560341
# NeckLine_slashneck_dmy     0.04806    0.35040   0.137 0.890961
# NeckLine_squarecollor_dmy  0.30711    0.74722   0.411 0.681263
# NeckLine_turndowncollor_dmy  0.13247    0.47531   0.279 0.780596
# ---
```

```
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.571 on 459 degrees of freedom
# Multiple R-squared:  0.1466,  Adjusted R-squared:  0.07411
# F-statistic: 2.022 on 39 and 459 DF,  p-value: 0.0003917

task4lm.6 <- update(task4lm.5, ~. - SleeveLengthfull)
summary(task4lm.6)

# Call:
#         lm(formula = log(TotalSales) ~ Size_L + Size_M + Size_S + Size_XL +
#                    waiseline_natural + waiseline_null + waiseline_other +
FabricType_chiffon +
#                    FabricType_jersey + FabricType_other + Decoration_beading +
#                    Decoration_bow + Decoration_hollowout + Decoration_lace +
#                    Decoration_other + Decoration_ruffles + Decoration_sashes +
#                    Decoration_sequined + Pattern_Type_other + Pattern_Type_null +
#                    Pattern_Type_patchwork + Pattern_Type_print +
SleeveLengththreequarter +
#                    SleeveLengthshort + SleeveLengthhalfsleeve +
SleeveLengthcap_sleeves +
#                    NeckLine_backless_dmy + NeckLine_boatneck_dmy +
NeckLine_bowneck_dmy +
#                    NeckLine_halter_dmy + NeckLine_mandarincollor_dmy +
NeckLine_oneck_dmy +
#                    NeckLine_open_dmy + NeckLine_ruffled_dmy + NeckLine_Scoop_dmy +
#                    NeckLine_slashneck_dmy + NeckLine_squarecollor_dmy +
NeckLine_turndowncollor_dmy,
#            data = TASK4_DS)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -5.6699 -0.8233  0.1784  1.0110  3.6968
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                 7.21290    0.25774  27.985  < 2e-16 ***
#         Size_L                      0.70634    0.21153   3.339 0.000909 ***
#         Size_M                      0.38362    0.17930   2.140 0.032921 *
#         Size_S                      0.25894    0.29607   0.875 0.382257
# Size_XL                     0.27372    0.44299   0.618 0.536944
# waiseline_natural           0.14642    0.18907   0.774 0.439056
# waiseline_null              0.19324    0.24910   0.776 0.438287
# waiseline_other             0.41107    0.74734   0.550 0.582555
# FabricType_chiffon          0.17885    0.17087   1.047 0.295777
# FabricType_jersey          -0.12322    0.47725  -0.258 0.796384
# FabricType_other           -0.26432    0.32930  -0.803 0.422584
# Decoration_beading         -0.06532    0.37103  -0.176 0.860340
# Decoration_bow              0.32046    0.45465   0.705 0.481261
# Decoration_hollowout       -0.29142    0.37311  -0.781 0.435176
# Decoration_lace            -0.03885    0.22054  -0.176 0.860260
# Decoration_other           -0.17450    0.28115  -0.621 0.535132
# Decoration_ruffles         -0.24001    0.41533  -0.578 0.563639
# Decoration_sashes           0.21128    0.27731   0.762 0.446505
# Decoration_sequined        -0.25863    0.44687  -0.579 0.563037
# Pattern_Type_other          0.21265    0.40771   0.522 0.602222
# Pattern_Type_null          -0.79345    0.20414  -3.887 0.000117 ***
#         Pattern_Type_patchwork      -0.40491    0.26104  -1.551 0.121563
# Pattern_Type_print          0.06677    0.22070   0.303 0.762361
# SleeveLengththreequarter   -0.36514    0.32320  -1.130 0.259163
# SleeveLengthshort           0.23707    0.19525   1.214 0.225297
# SleeveLengthhalfsleeve      0.18754    0.29135   0.644 0.520093
# SleeveLengthcap_sleeves    -0.29400    0.72289  -0.407 0.684417
# NeckLine_backless_dmy       1.54452    1.61364   0.957 0.338988
# NeckLine_boatneck_dmy      -0.30866    0.40079  -0.770 0.441622
# NeckLine_bowneck_dmy       -0.95570    0.54398  -1.757 0.079606 .
# NeckLine_halter_dmy         1.37292    1.60790   0.854 0.393627
# NeckLine_mandarincollor_dmy -2.08031    1.62602  -1.279 0.201405
# NeckLine_oneck_dmy          0.40463    0.17565   2.304 0.021693 *
#         NeckLine_open_dmy           -0.49069    0.94236  -0.521 0.602822
# NeckLine_ruffled_dmy        3.62425    1.64986   2.197 0.028540 *
#         NeckLine_Scoop_dmy           0.68000    1.16117   0.586 0.558424
```

```
 # NeckLine_slashneck_dmy       0.04505    0.34932   0.129 0.897443
 # NeckLine_squarecollor_dmy    0.30846    0.74635   0.413 0.679584
 # NeckLine_turndowncollor_dmy  0.13726    0.47350   0.290 0.772035
 # ---
 #         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 #
 # Residual standard error: 1.57 on 460 degrees of freedom
 # Multiple R-squared:  0.1466,  Adjusted R-squared:  0.07608
 # F-statistic: 2.079 on 38 and 460 DF,  p-value: 0.000265

 # since there are a lot of params lets make use of stepAIC or step function to choose
a model for us in an automated way based on aic
 step(task4lm.6, direction = "both", test = "F")

 #just pasting last step from this
 # Step:  AIC=442.53
 # log(TotalSales) ~ Size_L + Size_M + FabricType_chiffon + Pattern_Type_null +
 #         Pattern_Type_patchwork + SleeveLengthshort + NeckLine_bowneck_dmy +
 #         NeckLine_oneck_dmy + NeckLine_ruffled_dmy
 #
 # Df Sum of Sq    RSS    AIC F value     Pr(>F)
 # <none>                            1163.7 442.53
 # - SleeveLengthshort            1     4.943 1168.7 442.65  2.0772   0.150152
 # - FabricType_chiffon           1     4.946 1168.7 442.65  2.0785   0.150026
 # - NeckLine_bowneck_dmy         1     5.076 1168.8 442.71  2.1330   0.144803
 # - Pattern_Type_patchwork       1     5.955 1169.7 443.08  2.5023   0.114326
 # + SleeveLengththreequarter     1     2.750 1161.0 443.35  1.1559   0.282840
 # + FabricType_other             1     2.640 1161.1 443.40  1.1094   0.292733
 # + NeckLine_mandarincollor_dmy  1     2.422 1161.3 443.49  1.0176   0.313589
 # + Size_S                       1     2.212 1161.5 443.58  0.9294   0.335504
 # + Decoration_sashes            1     2.205 1161.5 443.59  0.9264   0.336276
 # + Decoration_hollowout         1     2.016 1161.7 443.67  0.8468   0.357900
 # + Decoration_bow               1     1.521 1162.2 443.88  0.6388   0.424535
 # + NeckLine_halter_dmy          1     1.452 1162.3 443.91  0.6096   0.435334
 # + NeckLine_backless_dmy        1     1.413 1162.3 443.93  0.5933   0.441514
 # - Size_M                       1     8.113 1171.8 444.00  3.4093   0.065435 .
 # + SleeveLengthhalfsleeve       1     1.130 1162.6 444.05  0.4743   0.491330
 # + Decoration_ruffles           1     1.085 1162.6 444.07  0.4556   0.500002
 # + NeckLine_turndowncollor_dmy  1     0.971 1162.7 444.12  0.4076   0.523508
 # + Decoration_sequined          1     0.929 1162.8 444.14  0.3900   0.532607
 # + NeckLine_open_dmy            1     0.901 1162.8 444.15  0.3780   0.538971
 # + NeckLine_boatneck_dmy        1     0.811 1162.9 444.19  0.3402   0.559975
 # + Pattern_Type_other           1     0.683 1163.0 444.24  0.2865   0.592711
 # + NeckLine_squarecollor_dmy    1     0.672 1163.0 444.25  0.2819   0.595673
 # + Size_XL                      1     0.627 1163.1 444.27  0.2629   0.608342
 # + NeckLine_Scoop_dmy           1     0.613 1163.1 444.27  0.2572   0.612302
 # + SleeveLengthcap_sleeves      1     0.468 1163.2 444.33  0.1965   0.657724
 # + Decoration_lace              1     0.304 1163.4 444.40  0.1276   0.721133
 # + Decoration_other             1     0.272 1163.4 444.42  0.1139   0.735891
 # + Pattern_Type_print           1     0.187 1163.5 444.45  0.0784   0.779577
 # + Decoration_beading           1     0.156 1163.5 444.47  0.0656   0.797926
 # + waiseline_other              1     0.138 1163.6 444.47  0.0577   0.810269
 # + waiseline_natural            1     0.123 1163.6 444.48  0.0517   0.820282
 # + waiseline_null               1     0.104 1163.6 444.49  0.0435   0.834916
 # + NeckLine_slashneck_dmy       1     0.088 1163.6 444.50  0.0370   0.847461
 # + FabricType_jersey            1     0.079 1163.6 444.50  0.0333   0.855333
 # - NeckLine_ruffled_dmy         1    14.627 1178.3 446.77  6.1465   0.013503 *
 #         - Size_L                       1    22.661 1186.4 450.16  9.5222   0.002145
**
 #         - NeckLine_oneck_dmy           1    25.749 1189.5 451.45 10.8201   0.001076
**
 #         - Pattern_Type_null            1    43.731 1207.4 458.94 18.3760 0.00002185
***
 #         ---
 #         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 #
 # Call:
 #         lm(formula = log(TotalSales) ~ Size_L + Size_M + FabricType_chiffon +
 #                    Pattern_Type_null + Pattern_Type_patchwork + SleeveLengthshort +
 #                    NeckLine_bowneck_dmy + NeckLine_oneck_dmy +
NeckLine_ruffled_dmy,
 #            data = TASK4_DS)
```

```
 #
 # Coefficients:
 #         (Intercept)                  Size_L                  Size_M
FabricType_chiffon
 # 7.3162                0.5874                0.2921                0.2240
 # Pattern_Type_null  Pattern_Type_patchwork       SleeveLengthshort
NeckLine_bowneck_dmy
 # -0.7485               -0.3815                0.2568                -0.7339
 # NeckLine_oneck_dmy    NeckLine_ruffled_dmy
 # 0.4729                3.8682

 finaltask4 <- lm( log(TotalSales) ~ Size_L + Size_M + FabricType_chiffon +
Pattern_Type_null +
          Pattern_Type_patchwork + SleeveLengthshort + NeckLine_bowneck_dmy +
          NeckLine_oneck_dmy + NeckLine_ruffled_dmy, data = TASK4_DS)

 # Call:
 #         lm(formula = log(TotalSales) ~ Size_L + Size_M + FabricType_chiffon +
 #                    Pattern_Type_null + Pattern_Type_patchwork + SleeveLengthshort +
 #                    NeckLine_bowneck_dmy + NeckLine_oneck_dmy +
NeckLine_ruffled_dmy,
 #            data = TASK4_DS)
 #
 # Residuals:
 #         Min      1Q  Median      3Q     Max
 # -5.9299 -0.7943  0.1948  1.0321  3.5212
 #
 # Coefficients:
 #         Estimate Std. Error t value  Pr(>|t|)
 # (Intercept)              7.3162     0.1449  50.497   < 2e-16 ***
 # Size_L                   0.5874     0.1903   3.086   0.00215 ** <very significant>
 # Size_M                   0.2921     0.1582   1.846   0.06543 .  <marginally
significant>
 # FabricType_chiffon       0.2240     0.1554   1.442   0.15003    <insignificant by
itself>
 # Pattern_Type_null       -0.7485     0.1746  -4.287 0.0000219 *** <very significant>
 # Pattern_Type_patchwork  -0.3815     0.2412  -1.582   0.11433    <insignificant by
itself>
 # SleeveLengthshort        0.2568     0.1782   1.441   0.15015    <insignificant by
itself>
 # NeckLine_bowneck_dmy    -0.7339     0.5025  -1.460   0.14480    <insignificant by
itself>
 # NeckLine_oneck_dmy       0.4729     0.1438   3.289   0.00108 ** <very significant>
 # NeckLine_ruffled_dmy     3.8682     1.5602   2.479   0.01350 *  <significant>
 #         ---
 #         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 #
 # Residual standard error: 1.543 on 489 degrees of freedom
 # Multiple R-squared:  0.1237,  Adjusted R-squared:  0.1076
 # F-statistic: 7.669 on 9 and 489 DF,  p-value: 0.0000000001428

 # from the model looks like these are significant if we just consider the main effect
 # Pattern_Type_null,NeckLine_oneck_dmy,Size_L,NeckLine_ruffled_dmy,Size_M(marginally)
 # So the company can focus on these for campaigns etc. # Still Adj Rsq is just 10
percent
 # We need to use the some factors from Task3, this and next Task if significant and
have a cumulative model to predict the sales very well
 # we needed to do this on Training and validate on test data but since we know the
concept
 # and just to answer task4, we had considered the whole dataset itself for this task


 # we can see why the above was given as final model from step function even though
insig factors
 # were still present
 # by removing the most insig factor based on p-value we see that adj. rsq starts to
decrease at
 # this stage. Hence we are going to term the model as final at the above step
summary(update(finaltask4, ~. -SleeveLengthshort)) # just to show adj rsq is starting
to decrease if we remove more

# Call:
```

```
#         lm(formula = log(TotalSales) ~ Size_L + Size_M + FabricType_chiffon +
#                    Pattern_Type_null + Pattern_Type_patchwork + NeckLine_bowneck_dmy
+
#                    NeckLine_oneck_dmy + NeckLine_ruffled_dmy, data = TASK4_DS)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -5.9740 -0.8008  0.1870  1.0482  3.4481
#
# Coefficients:
#         Estimate Std. Error t value  Pr(>|t|)
# (Intercept)              7.3603     0.1418  51.919   < 2e-16 ***
#         Size_L                   0.6163     0.1895   3.253   0.00122 **
#         Size_M                   0.3055     0.1581   1.933   0.05384 .
# FabricType_chiffon       0.2242     0.1555   1.441   0.15012
# Pattern_Type_null       -0.7737     0.1739  -4.448 0.0000107 ***
#         Pattern_Type_patchwork  -0.3774     0.2414  -1.563   0.11858
# NeckLine_bowneck_dmy    -0.7578     0.5028  -1.507   0.13238
# NeckLine_oneck_dmy       0.4729     0.1439   3.286   0.00109 **
#         NeckLine_ruffled_dmy     4.0672     1.5558   2.614   0.00922 **
#         ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.544 on 490 degrees of freedom
# Multiple R-squared:   0.12,    Adjusted R-squared:  0.1056
# F-statistic:  8.35 on 8 and 490 DF,  p-value: 0.0000000001198


#Task5:
#To regularize the rating procedure and find its efficiency, the store wants to find
if the rating of the dress affects the total sales

#Task 5 solution:
# Lets consider ATT_DS4 dataset with RatingClass column since Rating as a factor
(categorical variable is more relevant here as there is a clear distinction about the
distribution of Rating Data)

ATT_DS4 <- TASK4_DS %>%
        mutate(RatingClass = cut(Rating,c(-0.1,4,max(Rating)),
                                 labels = c("<=4",">4")))%>%
        mutate_if(is.character,as.factor) %>%
        dplyr::select(-c(Recommendation,SNOMERGE,
                   Size, waiseline, FabricType, Decoration,
                  Pattern_Type, `2013-08-29`,`2013-08-31`,`2013-02-09`,
                  `2013-04-09`,`2013-06-09`,`2013-08-09`,
                  `2013-10-09`, `2013-12-09`, `2013-09-14`, `2013-09-16`,
                  `2013-09-18`,`2013-09-20`,`2013-09-22`,`2013-09-24`,
                  `2013-09-26`,`2013-09-28`,`2013-09-30`, `2013-02-10`,
                  `2013-04-10`, `2013-06-10`, `2013-08-10`, `2013-10-10`,
                  `2013-12-10` ))

ggplot(ATT_DS4, aes(Rating)) + geom_density(fill="blue")+theme_bw() # indicates more
of 0 -1 or >4, hence best to use RatingClass

table(ATT_DS4$RatingClass)
# <=4  >4
# 133 366

ATT_DS4 <- ATT_DS4 %>% mutate(RATINGLE4 = ifelse(RatingClass=="<=4",1,0),
                             RATINGGT4 = ifelse(RatingClass==">4",1,0) # these two
are not needed, just one of them is sufficient, but for consistency creating both of
these
)

# we again need to do Training and test but since we know the concept and
# just want to answer task5, lets consider the whole dataset itself for now
task5lm <- lm(log(TotalSales) ~ RATINGLE4 + RATINGGT4 ,data = ATT_DS4)

summary(task5lm)
# Call:
#         lm(formula = log(TotalSales) ~ RATINGLE4 + RATINGGT4, data = ATT_DS4)
#
```

```
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.8642 -0.9255 -0.0501  0.9041  4.6135
#
# Coefficients: (1 not defined because of singularities)
# Estimate Std. Error t value Pr(>|t|)
# (Intercept)  8.22763    0.07214  114.05   <2e-16 ***
#        RATINGLE4   -1.97718    0.13973  -14.15   <2e-16 ***
#        RATINGGT4         NA         NA      NA       NA
# ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.38 on 497 degrees of freedom
# Multiple R-squared:  0.2872,   Adjusted R-squared:  0.2857
# F-statistic: 200.2 on 1 and 497 DF,  p-value: < 2.2e-16

task5lm.1 <- update(task5lm, ~. -RATINGGT4)
summary(task5lm.1)

# Call:
#        lm(formula = log(TotalSales) ~ RATINGLE4, data = ATT_DS4)
#
# Residuals:
#        Min      1Q  Median      3Q     Max
# -4.8642 -0.9255 -0.0501  0.9041  4.6135
#
# Coefficients:
#        Estimate Std. Error t value Pr(>|t|)
# (Intercept)  8.22763    0.07214  114.05   <2e-16 ***
#        RATINGLE4   -1.97718    0.13973  -14.15   <2e-16 ***
#        ---
#        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.38 on 497 degrees of freedom
# Multiple R-squared:  0.2872,   Adjusted R-squared:  0.2857
# F-statistic: 200.2 on 1 and 497 DF,  p-value: < 2.2e-16
# based on the p-value of the model, rating of the dress affects the total sales as
its is very significant <0.05
# Adj Rsq of 28.57%; adding this factor alone to previous best model from task3 and
# significant factors from task 4 would have improved the model so much

### Now just to complete the model as a one whole one for the retail store, lets
# combine the significant factors from task3,4,5 and create a single model and see how
the
# adj rs is. we expect it to be better than every task step individually but lets
validate

#Additional Step done by me to predict sales overall
#also in one of the earlier steps we considered interactions as well to improve the
model
# but just lets take the main effects from these steps and create interactions all
over again
# and use step/stepAIC to do one single regression model for TotalSales

ADDIT_lm <- lm(data=ATT_DS4,log(TotalSales) ~. - Dress_ID - Rating -RATINGGT4 -
RatingClass)

step(ADDIT_lm, direction = "both", test = "F")

# Call:
#        lm(formula = log(TotalSales) ~ SleeveLengththreequarter +
SleeveLengthsleeveless +
#                   SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
#                   SleeveLengthcap_sleeves + Priceaverage + Pricelow +
Style_sexy_dmy +
#                   Material_linen_dmy + Material_mix_dmy + Material_modal_dmy +
#                   NeckLine_bowneck_dmy + NeckLine_ruffled_dmy + Size_L + Size_M +
#                   Size_S + FabricType_null + FabricType_worsted + Pattern_Type_null
+
#                   RATINGLE4 + Material_cashmere_dmy, data = ATT_DS4)
#
# Coefficients:
```

```
#         (Intercept)  SleeveLengththreequarter    SleeveLengthsleeveless
# 8.8275                   -1.6199                   -1.5815
# SleeveLengthshort    SleeveLengthhalfsleeve         SleeveLengthfull
# -1.0749                   -1.0387                   -1.3517
# SleeveLengthcap_sleeves              Priceaverage                  Pricelow
# -1.5542                    0.6858                    1.0835
# Style_sexy_dmy        Material_linen_dmy         Material_mix_dmy
# 0.3279                    1.2108                    0.7005
# Material_modal_dmy      NeckLine_bowneck_dmy      NeckLine_ruffled_dmy
# -1.7594                   -0.9290                    2.8326
# Size_L                    Size_M                    Size_S
# 0.6019                    0.2885                    0.3885
# FabricType_null        FabricType_worsted         Pattern_Type_null
# -0.1982                   -0.5554                   -0.5331
# RATINGLE4     Material_cashmere_dmy
# -1.9268                   -1.0137


TOTALSALES_FINAL <- lm(formula = log(TotalSales) ~ SleeveLengththreequarter +
SleeveLengthsleeveless +
                    SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
                    SleeveLengthcap_sleeves + Priceaverage + Pricelow + Style_sexy_dmy
+
                    Material_linen_dmy + Material_mix_dmy + Material_modal_dmy +
                    NeckLine_bowneck_dmy + NeckLine_ruffled_dmy + Size_L + Size_M +
                    Size_S + FabricType_null + FabricType_worsted + Pattern_Type_null
+
                    RATINGLE4 + Material_cashmere_dmy, data = ATT_DS4)
summary(TOTALSALES_FINAL)

# Call:
#         lm(formula = log(TotalSales) ~ SleeveLengththreequarter +
SleeveLengthsleeveless +
#                    SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
#                    SleeveLengthcap_sleeves + Priceaverage + Pricelow +
Style_sexy_dmy +
#                    Material_linen_dmy + Material_mix_dmy + Material_modal_dmy +
#                    NeckLine_bowneck_dmy + NeckLine_ruffled_dmy + Size_L + Size_M +
#                    Size_S + FabricType_null + FabricType_worsted + Pattern_Type_null
+
#                    RATINGLE4 + Material_cashmere_dmy, data = ATT_DS4)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -5.1765 -0.8716  0.0396  0.8234  3.3445
#
# Coefficients:
#         Estimate Std. Error t value     Pr(>|t|)
# (Intercept)               8.8275     0.6741  13.095      < 2e-16 ***
# SleeveLengththreequarter  -1.6199     0.6679  -2.425     0.015671 *
# SleeveLengthsleeveless    -1.5815     0.6309  -2.507     0.012521 *
# SleeveLengthshort         -1.0749     0.6385  -1.684     0.092907 .
# SleeveLengthhalfsleeve    -1.0387     0.6586  -1.577     0.115459
# SleeveLengthfull          -1.3517     0.6396  -2.113     0.035093 *
# SleeveLengthcap_sleeves   -1.5542     0.8360  -1.859     0.063621 .
# Priceaverage               0.6858     0.1741   3.939 0.0000940896 ***
# Pricelow                   1.0835     0.1928   5.620 0.0000000326 ***
# Style_sexy_dmy             0.3279     0.1617   2.028     0.043128 *
# Material_linen_dmy         1.2108     0.7281   1.663     0.097002 .
# Material_mix_dmy           0.7005     0.3732   1.877     0.061132 .
# Material_modal_dmy        -1.7594     0.8842  -1.990     0.047190 *
# NeckLine_bowneck_dmy      -0.9290     0.4036  -2.301     0.021797 *
# NeckLine_ruffled_dmy       2.8326     1.2516   2.263     0.024079 *
# Size_L                     0.6019     0.1625   3.703     0.000238 ***
# Size_M                     0.2885     0.1397   2.065     0.039495 *
# Size_S                     0.3885     0.2263   1.717     0.086608 .
# FabricType_null           -0.1982     0.1167  -1.699     0.090066 .
# FabricType_worsted        -0.5554     0.3075  -1.806     0.071556 .
# Pattern_Type_null         -0.5331     0.1460  -3.652     0.000289 ***
# RATINGLE4                 -1.9268     0.1289 -14.952      < 2e-16 ***
# Material_cashmere_dmy     -1.0137     0.6395  -1.585     0.113580
# ---
```

```
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.237 on 476 degrees of freedom
# Multiple R-squared:  0.4512,   Adjusted R-squared:  0.4258
# F-statistic: 17.79 on 22 and 476 DF,  p-value: < 2.2e-16

#with 2-level interactions # to improve the model further

TOTALSALES_FINAL_W_2INT <- lm(formula = log(TotalSales) ~

                               (SleeveLengththreequarter + SleeveLengthsleeveless +
                               SleeveLengthshort + SleeveLengthhalfsleeve +
SleeveLengthfull +
                               SleeveLengthcap_sleeves + Priceaverage + Pricelow +
Style_sexy_dmy +
                               Material_linen_dmy + Material_mix_dmy +
Material_modal_dmy +
                               NeckLine_bowneck_dmy + NeckLine_ruffled_dmy + Size_L +
Size_M +
                               Size_S + FabricType_null + FabricType_worsted +
Pattern_Type_null +
                               RATINGLE4 + Material_cashmere_dmy)^2, data = ATT_DS4)


step(TOTALSALES_FINAL_W_2INT, direction = "both", test = "F")
#Pasting last call from step func.
# Call:
#         lm(formula = log(TotalSales) ~ SleeveLengththreequarter +
SleeveLengthsleeveless +
#                    SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
#                    Priceaverage + Pricelow + Style_sexy_dmy + Material_linen_dmy +
#                    Material_mix_dmy + Material_modal_dmy + NeckLine_bowneck_dmy +
#                    NeckLine_ruffled_dmy + Size_L + Size_M + Size_S + FabricType_null
+
#                    FabricType_worsted + Pattern_Type_null + RATINGLE4 +
Material_cashmere_dmy +
#                    SleeveLengththreequarter:Size_L +
SleeveLengththreequarter:RATINGLE4 +
#                    SleeveLengthsleeveless:NeckLine_bowneck_dmy +
SleeveLengthsleeveless:Size_L +
#                    SleeveLengthsleeveless:Size_S +
SleeveLengthsleeveless:Pattern_Type_null +
#                    SleeveLengthsleeveless:RATINGLE4 + SleeveLengthshort:Size_L +
#                    SleeveLengthshort:Size_S + SleeveLengthshort:Pattern_Type_null +
#                    SleeveLengthhalfsleeve:FabricType_worsted +
SleeveLengthhalfsleeve:Pattern_Type_null +
#                    SleeveLengthfull:Priceaverage + SleeveLengthfull:Pricelow +
#                    SleeveLengthfull:Size_L + SleeveLengthfull:FabricType_null +
#                    SleeveLengthfull:Pattern_Type_null +
Priceaverage:Material_mix_dmy +
#                    Priceaverage:Size_S + Priceaverage:Material_cashmere_dmy +
#                    Pricelow:RATINGLE4 + Style_sexy_dmy:Size_L +
Style_sexy_dmy:Size_M +
#                    Style_sexy_dmy:Size_S + Material_linen_dmy:Size_M +
Material_mix_dmy:Size_L +
#                    Material_mix_dmy:Size_M + Size_S:FabricType_worsted +
FabricType_worsted:RATINGLE4 +
#                    Pattern_Type_null:RATINGLE4, data = ATT_DS4)
#
# Coefficients:
#         (Intercept)                        SleeveLengththreequarter
# 8.7130                                         -0.5254
# SleeveLengthsleeveless                               SleeveLengthshort
# -1.1716                                         -0.8836
# SleeveLengthhalfsleeve                                SleeveLengthfull
# -1.2042                                         -2.3970
# Priceaverage                                        Pricelow
# 0.5847                                          0.7513
# Style_sexy_dmy                             Material_linen_dmy
# -0.1063                                          1.2258
# Material_mix_dmy                             Material_modal_dmy
# 0.6347                                         -1.8147
```

```
# NeckLine_bowneck_dmy                          NeckLine_ruffled_dmy
# -1.9092                                        3.1834
# Size_L                                        Size_M
# 1.7671                                        0.1931
# Size_S                               FabricType_null
# 1.2033                                       -0.2452
# FabricType_worsted                             Pattern_Type_null
# 0.2006                                       -1.7803
# RATINGLE4                         Material_cashmere_dmy
# -1.7407                                        0.9235
# SleeveLengththreequarter:Size_L          SleeveLengththreequarter:RATINGLE4
# -2.5566                                       -1.1852
# SleeveLengthsleeveless:NeckLine_bowneck_dmy
SleeveLengthsleeveless:Size_L
# 1.7780                                       -1.5114
# SleeveLengthsleeveless:Size_S     SleeveLengthsleeveless:Pattern_Type_null
# -0.8113                                        1.3994
# SleeveLengthsleeveless:RATINGLE4                      SleeveLengthshort:Size_L
# -0.6317                                       -1.1300
# SleeveLengthshort:Size_S          SleeveLengthshort:Pattern_Type_null
# -1.1741                                        1.9552
# SleeveLengthhalfsleeve:FabricType_worsted
SleeveLengthhalfsleeve:Pattern_Type_null
# 1.9841                                        1.8378
# SleeveLengthfull:Priceaverage                     SleeveLengthfull:Pricelow
# 1.0262                                        0.9380
# SleeveLengthfull:Size_L            SleeveLengthfull:FabricType_null
# -0.9438                                        0.4693
# SleeveLengthfull:Pattern_Type_null                 Priceaverage:Material_mix_dmy
# 1.9306                                        1.0899
# Priceaverage:Size_S           Priceaverage:Material_cashmere_dmy
# -0.7170                                       -3.0039
# Pricelow:RATINGLE4                         Style_sexy_dmy:Size_L
# 1.0940                                        1.1392
# Style_sexy_dmy:Size_M                         Style_sexy_dmy:Size_S
# 0.6697                                        1.1038
# Material_linen_dmy:Size_M                       Material_mix_dmy:Size_L
# 2.1050                                       -1.4219
# Material_mix_dmy:Size_M                      Size_S:FabricType_worsted
# -2.0617                                       -3.7085
# FabricType_worsted:RATINGLE4                   Pattern_Type_null:RATINGLE4
# -1.4966                                       -0.7449

TOTALSALES_FINAL_W_2INT <- lm(formula = log(TotalSales) ~ SleeveLengththreequarter +
SleeveLengthsleeveless +
                    SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
                    Priceaverage + Pricelow + Style_sexy_dmy + Material_linen_dmy +
                    Material_mix_dmy + Material_modal_dmy + NeckLine_bowneck_dmy +
                    NeckLine_ruffled_dmy + Size_L + Size_M + Size_S + FabricType_null
+
                    FabricType_worsted + Pattern_Type_null + RATINGLE4 +
Material_cashmere_dmy +
                    SleeveLengththreequarter:Size_L +
SleeveLengththreequarter:RATINGLE4 +
                    SleeveLengthsleeveless:NeckLine_bowneck_dmy +
SleeveLengthsleeveless:Size_L +
                    SleeveLengthsleeveless:Size_S +
SleeveLengthsleeveless:Pattern_Type_null +
                    SleeveLengthsleeveless:RATINGLE4 + SleeveLengthshort:Size_L +
                    SleeveLengthshort:Size_S + SleeveLengthshort:Pattern_Type_null +
                    SleeveLengthhalfsleeve:FabricType_worsted +
SleeveLengthhalfsleeve:Pattern_Type_null +
                    SleeveLengthfull:Priceaverage + SleeveLengthfull:Pricelow +
                    SleeveLengthfull:Size_L + SleeveLengthfull:FabricType_null +
                    SleeveLengthfull:Pattern_Type_null + Priceaverage:Material_mix_dmy
+
                    Priceaverage:Size_S + Priceaverage:Material_cashmere_dmy +
                    Pricelow:RATINGLE4 + Style_sexy_dmy:Size_L + Style_sexy_dmy:Size_M
+
                    Style_sexy_dmy:Size_S + Material_linen_dmy:Size_M +
Material_mix_dmy:Size_L +
```

```
                    Material_mix_dmy:Size_M + Size_S:FabricType_worsted +
FabricType_worsted:RATINGLE4 +
                    Pattern_Type_null:RATINGLE4, data = ATT_DS4)

summary(TOTALSALES_FINAL_W_2INT)

# Call:
#         lm(formula = log(TotalSales) ~ SleeveLengththreequarter +
SleeveLengthsleeveless +
#                    SleeveLengthshort + SleeveLengthhalfsleeve + SleeveLengthfull +
#                    Priceaverage + Pricelow + Style_sexy_dmy + Material_linen_dmy +
#                    Material_mix_dmy + Material_modal_dmy + NeckLine_bowneck_dmy +
#                    NeckLine_ruffled_dmy + Size_L + Size_M + Size_S + FabricType_null
+
#                    FabricType_worsted + Pattern_Type_null + RATINGLE4 +
Material_cashmere_dmy +
#                    SleeveLengththreequarter:Size_L +
SleeveLengththreequarter:RATINGLE4 +
#                    SleeveLengthsleeveless:NeckLine_bowneck_dmy +
SleeveLengthsleeveless:Size_L +
#                    SleeveLengthsleeveless:Size_S +
SleeveLengthsleeveless:Pattern_Type_null +
#                    SleeveLengthsleeveless:RATINGLE4 + SleeveLengthshort:Size_L +
#                    SleeveLengthshort:Size_S + SleeveLengthshort:Pattern_Type_null +
#                    SleeveLengthhalfsleeve:FabricType_worsted +
SleeveLengthhalfsleeve:Pattern_Type_null +
#                    SleeveLengthfull:Priceaverage + SleeveLengthfull:Pricelow +
#                    SleeveLengthfull:Size_L + SleeveLengthfull:FabricType_null +
#                    SleeveLengthfull:Pattern_Type_null +
Priceaverage:Material_mix_dmy +
#                    Priceaverage:Size_S + Priceaverage:Material_cashmere_dmy +
#                    Pricelow:RATINGLE4 + Style_sexy_dmy:Size_L +
Style_sexy_dmy:Size_M +
#                    Style_sexy_dmy:Size_S + Material_linen_dmy:Size_M +
Material_mix_dmy:Size_L +
#                    Material_mix_dmy:Size_M + Size_S:FabricType_worsted +
FabricType_worsted:RATINGLE4 +
#                    Pattern_Type_null:RATINGLE4, data = ATT_DS4)
#
# Residuals:
#         Min      1Q  Median      3Q     Max
# -4.6935 -0.6899  0.0000  0.7239  2.9308
#
# Coefficients:
#         Estimate Std. Error t value Pr(>|t|)
# (Intercept)                                  8.7130     0.5205  16.740  < 2e-16 ***
# SleeveLengththreequarter                    -0.5254     0.5333  -0.985 0.325058
# SleeveLengthsleeveless                      -1.1716     0.4923  -2.380 0.017735 *
# SleeveLengthshort                           -0.8836     0.5013  -1.763 0.078654 .
# SleeveLengthhalfsleeve                      -1.2042     0.5307  -2.269 0.023746 *
# SleeveLengthfull                            -2.3970     0.6759  -3.546 0.000432 ***
# Priceaverage                                 0.5847     0.1928   3.033 0.002562 **
# Pricelow                                     0.7513     0.2193   3.426 0.000669 ***
# Style_sexy_dmy                              -0.1063     0.2289  -0.465 0.642510
# Material_linen_dmy                           1.2258     0.8503   1.442 0.150101
# Material_mix_dmy                             0.6347     0.6103   1.040 0.298910
# Material_modal_dmy                          -1.8147     0.8221  -2.207 0.027799 *
# NeckLine_bowneck_dmy                        -1.9092     0.5653  -3.377 0.000796 ***
# NeckLine_ruffled_dmy                         3.1834     1.1657   2.731 0.006567 **
# Size_L                                       1.7671     0.5204   3.396 0.000746 ***
# Size_M                                       0.1931     0.1426   1.354 0.176538
# Size_S                                       1.2033     0.3962   3.037 0.002529 **
# FabricType_null                             -0.2452     0.1222  -2.006 0.045492 *
# FabricType_worsted                           0.2006     0.3493   0.574 0.566140
# Pattern_Type_null                           -1.7803     0.5011  -3.553 0.000422 ***
# RATINGLE4                                   -1.7407     0.1973  -8.822  < 2e-16 ***
# Material_cashmere_dmy                        0.9235     0.9490   0.973 0.331040
# SleeveLengththreequarter:Size_L             -2.5566     0.7306  -3.500 0.000513 ***
# SleeveLengththreequarter:RATINGLE4          -1.1852     0.5407  -2.192 0.028897 *
# SleeveLengthsleeveless:NeckLine_bowneck_dmy  1.7780     0.7834   2.269 0.023718 *
# SleeveLengthsleeveless:Size_L               -1.5114     0.5538  -2.729 0.006596 **
# SleeveLengthsleeveless:Size_S               -0.8113     0.5091  -1.593 0.111760
```

```
# SleeveLengthsleeveless:Pattern_Type_null      1.3994     0.5198   2.692 0.007364 **
# SleeveLengthsleeveless:RATINGLE4             -0.6317     0.2571  -2.458 0.014368 *
# SleeveLengthshort:Size_L                     -1.1300     0.5891  -1.918 0.055729 .
# SleeveLengthshort:Size_S                     -1.1741     0.6218  -1.888 0.059641 .
# SleeveLengthshort:Pattern_Type_null           1.9552     0.6129   3.190 0.001523 **
# SleeveLengthhalfsleeve:FabricType_worsted     1.9841     1.3743   1.444 0.149507
# SleeveLengthhalfsleeve:Pattern_Type_null      1.8378     0.6336   2.901 0.003909 **
# SleeveLengthfull:Priceaverage                 1.0262     0.4533   2.264 0.024057 *
# SleeveLengthfull:Pricelow                     0.9380     0.4607   2.036 0.042335 *
# SleeveLengthfull:Size_L                      -0.9438     0.6315  -1.495 0.135738
# SleeveLengthfull:FabricType_null              0.4693     0.2970   1.580 0.114710
# SleeveLengthfull:Pattern_Type_null            1.9306     0.6355   3.038 0.002523 **
# Priceaverage:Material_mix_dmy                 1.0899     0.7394   1.474 0.141200
# Priceaverage:Size_S                          -0.7170     0.4405  -1.628 0.104307
# Priceaverage:Material_cashmere_dmy           -3.0039     1.2782  -2.350 0.019203 *
# Pricelow:RATINGLE4                            1.0940     0.2674   4.091 0.000051 ***
# Style_sexy_dmy:Size_L                         1.1392     0.4464   2.552 0.011047 *
# Style_sexy_dmy:Size_M                         0.6697     0.3466   1.932 0.053946 .
# Style_sexy_dmy:Size_S                         1.1038     0.6073   1.818 0.069782 .
# Material_linen_dmy:Size_M                     2.1050     1.4613   1.441 0.150422
# Material_mix_dmy:Size_L                      -1.4219     0.8185  -1.737 0.083026 .
# Material_mix_dmy:Size_M                      -2.0617     0.9718  -2.121 0.034432 *
# Size_S:FabricType_worsted                    -3.7085     1.2779  -2.902 0.003892 **
# FabricType_worsted:RATINGLE4                 -1.4966     0.7627  -1.962 0.050367 .
# Pattern_Type_null:RATINGLE4                  -0.7449     0.2999  -2.484 0.013362 *
#         ---
#         Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 1.147 on 447 degrees of freedom
# Multiple R-squared:  0.5575,   Adjusted R-squared:  0.507
# F-statistic: 11.04 on 51 and 447 DF,  p-value: < 2.2e-16

par(mfrow=c(2,2))
plot(TOTALSALES_FINAL_W_2INT)
```