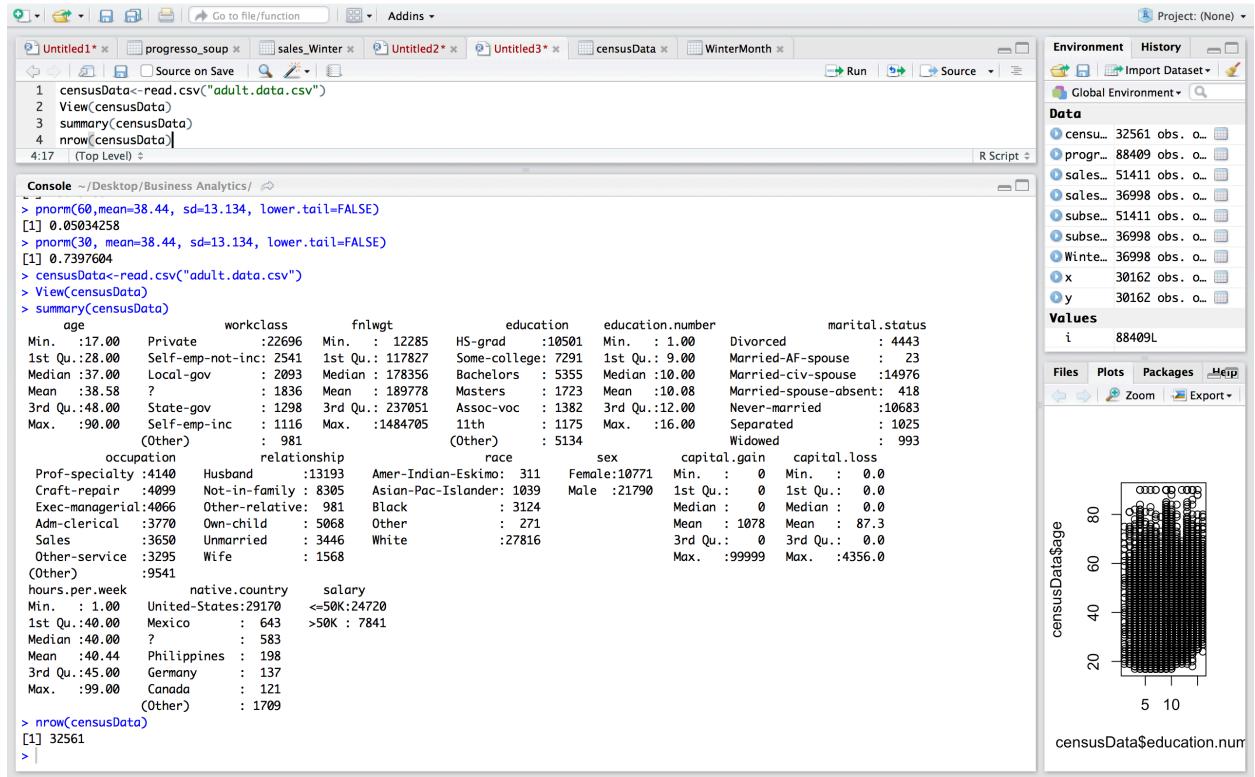


Business Analytics Assignment 2

Homework Part 1

1) Census Income Data Set in the UCI Machine Learning Repository

<https://archive.ics.uci.edu/ml/datasets/Census+Income>



2) Data exploration

- Missing and/or invalid values??

Checked for NA records

```

1 censusData<-read.csv("adult.data.csv")
2 View(censusData)
3 is.na(censusData)
4 sum(is.na(censusData))
5 which (is.na(censusData))
6
[reached getOption("max.print") -- omitted 31895 rows]
> sum(is.na(censusData))
[1] 0
> which (is.na(censusData))
integer(0)
>

```

The screenshot shows the RStudio interface with the code above in the script pane. The console pane displays the output of the code, showing that there are no missing values (sum(is.na(censusData)) returns 0) and no rows with missing values (which (is.na(censusData)) returns integer(0)). The environment pane on the right shows various objects defined in the session.

I could see missing data “?” in the summary for censusData.
Therefore, removed the records with missing value.

```

1 censusData<-read.csv("adult.data.csv")
2 View(censusData)
3 summary(censusData)
4 nrow(censusData)
5 is.na(censusData) = censusData=="?"
6 is.na(censusData) = censusData=="?"
7 censusData = na.omit(censusData)
8 nrow(censusData)
9 summary(censusData)
9:20 (Top Level) <

```

	age	workclass	fnlwgt	education	education.number
Min.	:17.00	Private	:2286	Min. : 13769	HS-grad : 9840
1st Qu.	:28.00	Self-emp-not-inc	:2499	1st Qu.: 117627	Some-college: 6678
Median	:37.00	Local-gov	:2067	Median : 178425	Bachelors : 5044
Mean	:38.44	State-gov	:1279	Mean : 189794	Masters : 1627
3rd Qu.	:47.00	Self-emp-inc	:1074	3rd Qu.: 237628	Assoc-voc : 1307
Max.	:90.00	Federal-gov	:943	Max. :1484705	11th : 1048
	(Other)		: 14	(Other) : 4618	Max. : 16.00

	marital.status	occupation	relationship	race	sex
Divorced	: 4214	Prof-specialty	:4038	Husband : 12463	Amer-Indian-Eskimo: 286
Married-AF-spouse	: 21	Craft-repair	:4030	Not-in-family : 7726	Asian-Pac-Islander: 895
Married-civ-spouse	:14065	Exec-managerial	:3992	Other-relative: 889	Black : 2817
Married-spouse-absent	: 370	Adm-clerical	:3721	Own-child : 4466	Other : 231
Never-married	: 9726	Sales	:3584	Unmarried : 3212	White : 25933
Separated	: 939	Other-service	:3212	Wife : 1406	
Widowed	: 827	(Other)	:7585		

	capital.gain	capital.loss	hours.per.week	native.country	salary
Min.	: 0	Min. : 0.00	Min. : 1.00	United-States:7504	<=50K:22654
1st Qu.	: 0	1st Qu.: 0.00	1st Qu.:40.00	Mexico : 610	>50K : 7508
Median	: 0	Median : 0.00	Median :40.00	Philippines : 188	
Mean	: 1092	Mean : 88.37	Mean :40.93	Germany : 128	
3rd Qu.	: 0	3rd Qu.: 0.00	3rd Qu.:45.00	Puerto-Rico : 109	

The screenshot shows the RStudio interface with the code above. The console pane shows the removal of records with missing values using na.omit() and the resulting summary statistics for the cleaned dataset. The environment pane on the right shows the updated state of the session.

- Summary statistics

```

1 censusData<-read.csv("adult.data.csv")
2 View(censusData)
3 summary(censusData)
4

```

```

> summary(censusData)
   age          workclass      fnlwgt      education      education.number
Min. :17.00    Private       :22696   Min. : 12285   HS-grad        :10501   Min. : 1.00
1st Qu.:28.00   Self-emp-not-inc: 2541   1st Qu.:117827  Some-college: 7291   1st Qu.: 9.00
Median :37.00   Local-gov     : 2093   Median :178356  Bachelors     : 5355   Median :10.00
Mean   :38.58   ?             : 1836   Mean   :189778  Masters       : 1723   Mean   :10.08
3rd Qu.:48.00   State-gov     :1298   3rd Qu.:237051  Assoc-voc     : 1382   3rd Qu.:12.00
Max.  :90.00   Self-emp-inc  :1116   Max.  :1484705  (Other)       : 5134   Max.  :16.00
   marital.status   occupation      relationship      race          sex
Divorced      : 4443  Prof-specialty :14140  Husband       :13193  Amer-Indian-Eskimo: 311   Female:10771
Married-AF-spouse : 23   Craft-repair   :4099  Not-in-family : 8305  Asian-Pac-Islander: 1039  Male :21790
Married-civ-spouse :14976 Exec-managerial:4066  Other-relative: 981   Black          : 3124
Married-spouse-absent: 418  Adm-clerical  :3770  Own-child     : 5068  Other          : 271
Never-married   :10683  Sales          :3650  Unmarried     : 3446  White          :27816
Separated       : 1025  Other-service  :3295  Wife          : 1568
Widowed         : 993   (Other)       : 9541
   capital.gain   capital.loss   hours.per.week native.country      salary
Min.   : 0.00000  Min.   :0.000000  Min.   : 0.00  United-States:29170  <=50K:24720
1st Qu.: 0.00000  1st Qu.:0.000000  1st Qu.: 0.00  Mexico       : 643  >50K : 7841
Median : 0.00000  Median :0.000000  Median : 0.00  Philippines  : 198
Mean   : 1078.00000  Mean   :40.440000  Mean   : 40.44  Germany     : 137
3rd Qu.: 0.00000  3rd Qu.:45.000000  3rd Qu.:45.00  Canada      : 121
Max.  :99999.00000  Max.  :4356.000000  Max.  :99.00  (Other)     :1709

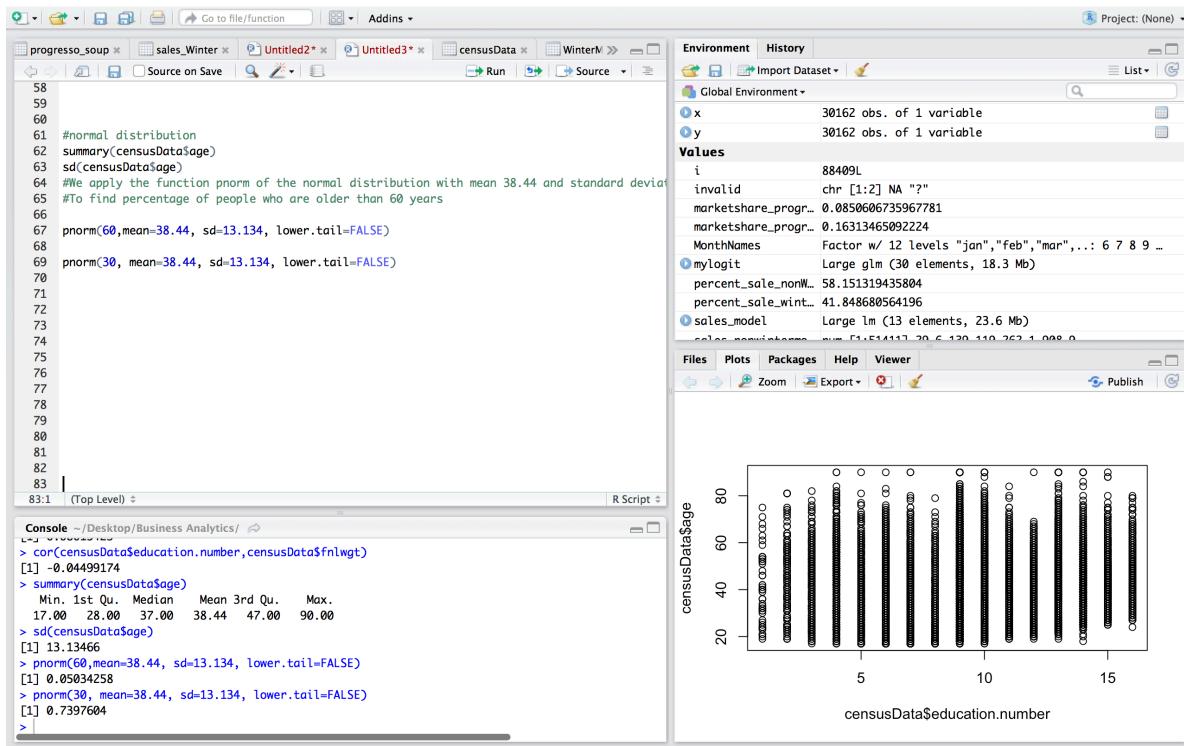
```

- Distributions

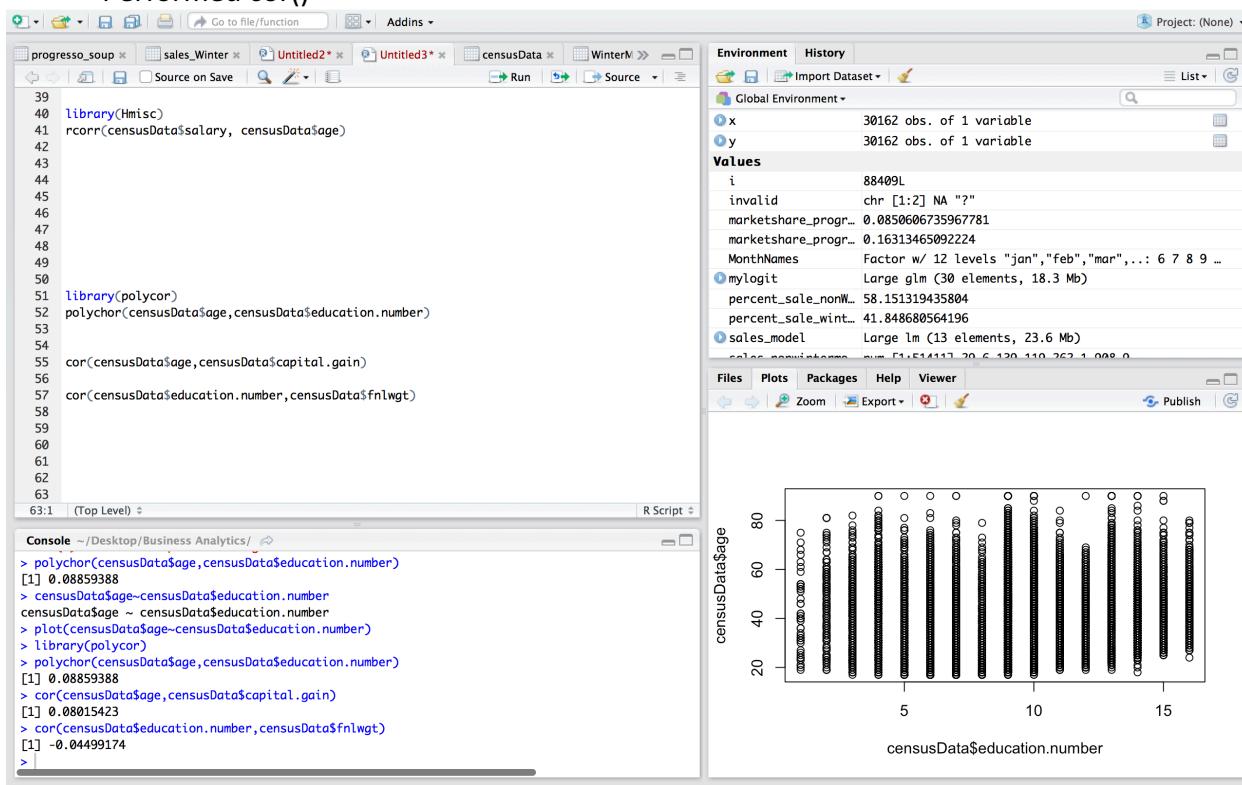
Normal distribution using 'pnorm'

To calculate percentage of people who are older than 60 years : 5.03%

To calculate percentage of people who are older than 30 years : 73.97%



- Correlations
- Polychoric Correlation
- Performed cor()



- Used Library Hmisc to implement rcorr
(To find correlation between variables that are not numeric)

The screenshot shows the RStudio interface. On the left, the R Script pane displays R code for calculating correlations and running a logistic regression model. The Console pane shows the output of these commands. On the right, the Environment pane lists global variables and their values. The Packages pane shows the installed R packages and their versions.

```

39 library(Hmisc)
40 rcorr(censusData$age, censusData$salary)
41
42
43
44
45
46
47
48
49
50
51
51:1 (Top Level) + R Script

```

```

> library(Hmisc)
> rcorr(censusData$age, censusData$salary)
      x   y
x 1.00 0.24
y 0.24 1.00

n= 30162

P
      x   y
x  0
y  0
> rcorr(censusData$salary, censusData$age)
      x   y
x 1.00 0.24
y 0.24 1.00

n= 30162

P
      x   y
x  0
y  0

```

Environment pane (Global Environment):

- x 30162 obs. of 1 variable
- y 30162 obs. of 1 variable
- Values
- i 88409L
- invalid chr [1:2] NA "?"
- marketshare_progr... 0.0850606735967781
- marketshare_progr... 0.16313465092224
- MonthNames Factor w/ 12 levels "jan","feb","mar",..: 6 7 8 9 ...
- mylogit Large glm (30 elements, 18.3 Mb)
- percent_sale_nonW... 58.151319435804
- percent_sale_wint... 41.848680564196
- sales_model Large lm (13 elements, 23.6 Mb)

Packages pane:

Name	Description	Version
<input type="checkbox"/> acepack	ace() and avas() for selecting regression transformations	1.3-3.3
<input type="checkbox"/> boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-18
<input type="checkbox"/> chron	Chronological Objects which can Handle Dates and Times	2.3-47
<input type="checkbox"/> class	Functions for Classification	7.3-14
<input type="checkbox"/> cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.4
<input type="checkbox"/> codetools	Code Analysis Tools for R	0.2-14
<input type="checkbox"/> colorspace	Color Space Manipulation	1.2-6
<input type="checkbox"/> compiler	The R Compiler Package	3.3.1
<input type="checkbox"/> data.table	Extension of Data.frame	1.9.6
<input checked="" type="checkbox"/> datasets	The R Datasets Package	3.3.1
<input type="checkbox"/> dichromat	Color Schemes for Dichromats	2.0-0
<input type="checkbox"/> digest	Create Compact Hash Digests of R Objects	0.6.10
<input type="checkbox"/> foreign	Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...	0.8-66

3) Prediction

- Run a logistic regression

Since there are two possible outcomes i.e. salary >50k or salary<=50k, I made use of Logistic Regression

Age, sex and education Number contributes a lot to determine the salary of a person.

Implication:

Salary increases by 0.045 if the age increases by one year

```

9 summary(censusData)
10 dim(censusData)
11 sapply(censusData, sd)
12 mylogit<-glm(salary~age + education.number + workclass + sex, data=censusData, family="binomial")
13 mylogit
14 summary(mylogit)
15

```

Call:
`glm(formula = salary ~ age + education.number + workclass + sex,
 family = "binomial", data = censusData)`

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4830	-0.6729	-0.4471	-0.0813	3.2559

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.486249	0.130891	-57.195	< 2e-16 ***
age	0.045452	0.001250	36.351	< 2e-16 ***
education.number	0.370784	0.0006934	53.475	< 2e-16 ***
workclass Local-gov	-0.425317	0.092833	-4.582	4.62e-06 ***
workclass Private	-0.375969	0.077786	-4.833	1.34e-06 ***
workclass Self-emp-inc	0.374233	0.103062	3.631	0.000282 ***
workclass Self-emp-not-inc	-0.658459	0.090865	-7.247	4.27e-13 ***
workclass State-gov	-0.685588	0.105194	-6.517	7.16e-11 ***
workclass Without-pay	-11.905249	77.128923	-0.154	0.877330
sex Male	1.329284	0.038475	34.549	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 33851 on 30161 degrees of freedom
Residual deviance: 26790 on 30152 degrees of freedom
AIC: 26810

Number of Fisher Scoring iterations: 11

Homework Part 2

- Create a dummy variable for “Winter” months defined as Oct, Nov, Dec, Jan & Feb. Use the “Month” variable to create this.

Created a Dummy Variable ‘winter_month’ representing months ‘oct, nov, dec, jan, feb’

The screenshot shows the RStudio interface with the following details:

- Project:** (None)
- Environment:** Shows variables like pr_88..., sa_51..., sa_36..., Wi_36... with their respective values.
- Data:** Shows the structure of the dataset, including variables like IRI_KEY, Month, Region, Low_Income, High_Income, Price.Campbell, Price.PL, Price.Progresso, Sales.Progresso, Category_Sales, and winter_month.
- Values:** Shows the first few rows of the dataset.
- Files:** Shows the R script file.
- Plots:** No plots are present.
- Console:** Displays the R code and its output. The output shows the dataset structure and the first few rows of the data frame.

```

1 url<- "https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/Week-03/Progresso_Soup_Hwk.csv"
2 progresso_soup<-read.csv(url)
3 str(progresso_soup)
4 progresso_soup$Month
5 #change the data type for Month to a factor variable
6 progresso_soup$Month<-factor(progresso_soup$Month)
7 str(progresso_soup)
8 MonthNames<-factor(progresso_soup$Month, levels=1:12, labels=c("jan","feb","mar","apr","may","jun","jul","aug","sep","oct","nov","dec"))
9 #creating a dummy variables for "Winter" months defined as Oct, Nov, Dec, Jan & Feb.
10 progresso_soup$winter_month<-as.logical(0)
11 str(progresso_soup)
12
13:1 (Top Level) 

```

The screenshot shows the RStudio interface with the following details:

- Project:** (None)
- Environment:** Shows variables like pr_88..., sa_51..., sa_36..., Wi_36... with their respective values.
- Data:** Shows the structure of the dataset, including variables like IRI_KEY, Month, Region, Low_Income, High_Income, Price.Campbell, Price.PL, Price.Progresso, Sales.Progresso, Category_Sales, and winter_month.
- Values:** Shows the first few rows of the dataset.
- Files:** Shows the R script file.
- Plots:** No plots are present.
- Console:** Displays the R code and its output. The output shows the dataset structure and the first few rows of the data frame. The 'winter_month' variable is now correctly assigned using a for loop.

```

1 url<- "https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/Week-03/Progresso_Soup_Hwk.csv"
2 progresso_soup<-read.csv(url)
3 str(progresso_soup)
4 progresso_soup$Month
5 #change the data type for Month to a factor variable
6 progresso_soup$Month<-factor(progresso_soup$Month)
7 str(progresso_soup)
8 MonthNames<-factor(progresso_soup$Month, levels=1:12, labels=c("jan","feb","mar","apr","may","jun","jul","aug","sep","oct","nov","dec"))
9 #creating a dummy variables for "Winter" months defined as Oct, Nov, Dec, Jan & Feb.
10 progresso_soup$winter_month<-as.logical(0)
11 str(progresso_soup)
12
13 for(i in 1:nrow(progresso_soup)){
14   if(progresso_soup$Month[i]==1 | progresso_soup$Month[i]==2 | progresso_soup$Month[i]==10 | progresso_soup$Month[i]==11 | progresso_soup$Month[i]==12)
15     progresso_soup$winter_month[i]<-as.logical(1)
16 }
17 View(progresso_soup)
18
12:1 (Top Level) 

```

The screenshot shows the RStudio interface. On the left is a data grid titled "progresso_soup" with columns: IRI_KEY, Month, Region, Low_Income, High_Income, Price.Campbell, Price.PL, Price.Progresso, Sales.Progresso, Category_Sales, and winter_month. The data consists of 26 rows of sales data for different months and regions. On the right is the "Environment" tab of the workspace, showing various global variables like pr, sa, i, and url. Below the workspace is the "System Library" pane.

Data Grid:

IRI_KEY	Month	Region	Low_Income	High_Income	Price.Campbell	Price.PL	Price.Progresso	Sales.Progresso	Category_Sales	winter_month	
1	200039	6	East	0	0	1.429	1.295	1.798	29.563	346.0352	FALSE
2	200039	7	East	0	0	1.220	1.140	1.842	139.033	2703.2897	FALSE
3	200039	8	East	0	0	1.485	1.162	1.868	118.970	1558.8480	FALSE
4	200039	9	East	0	0	1.016	1.104	1.535	262.096	5018.1365	FALSE
5	200039	10	East	0	0	1.259	1.073	1.876	137.846	3442.1859	TRUE
6	200039	11	East	0	0	1.398	1.161	0.925	1685.616	4476.7378	TRUE
7	200039	12	East	0	0	1.315	1.177	0.981	1382.987	5237.5357	TRUE
8	200039	1	East	0	0	1.365	1.182	0.918	1911.459	4965.7170	TRUE
9	200039	2	East	0	0	1.292	1.065	0.936	1462.487	4948.9620	TRUE
10	200039	3	East	0	0	1.365	0.990	1.133	908.917	4191.7657	FALSE
11	200039	4	East	0	0	1.400	0.969	1.643	221.784	1770.7023	FALSE
12	200039	5	East	0	0	1.275	1.047	1.633	316.536	2428.0682	FALSE
13	200039	6	East	0	0	1.290	1.108	1.645	212.377	2144.1987	FALSE
14	200039	7	East	0	0	1.454	1.058	1.980	76.626	1229.5095	FALSE
15	200039	8	East	0	0	1.420	1.192	1.912	60.188	1103.2815	FALSE
16	200039	9	East	0	0	1.414	1.084	1.176	945.323	3404.4959	FALSE
17	200039	10	East	0	0	1.274	1.110	0.915	1523.205	4571.3007	TRUE
18	200039	11	East	0	0	1.181	1.039	0.907	1718.239	5331.2294	TRUE
19	200039	12	East	0	0	1.241	1.090	1.361	632.196	4318.8184	TRUE
20	200039	1	East	0	0	0.964	1.126	0.878	1979.215	4352.6445	TRUE
21	200411	6	East	0	0	1.610	1.389	1.930	183.501	562.7724	FALSE
22	200411	7	East	0	0	1.647	1.335	1.903	854.008	2724.6475	FALSE
23	200411	8	East	0	0	1.618	1.366	1.888	725.352	2367.1353	FALSE
24	200411	9	East	0	0	1.211	1.339	1.640	1965.761	7041.4142	FALSE
25	200411	10	East	0	0	1.248	1.288	1.544	2461.457	7486.1501	TRUE

Showing 1 to 26 of 88,409 entries

Console:

```
> View(progresso_soup)
> |
```

- Compute the “Market Share” for Progresso (as percentage of total sales) in the Winter vs. non-Winter months using the variable created in (1).

The screenshot shows the RStudio interface. The top part displays an R script with code to subset the data into winter and non-winter months and calculate market share for Progresso. The bottom part shows the execution of this code in the console, resulting in market share values for winter and non-winter months.

R Script:

```
14 if(progresso_soup$Month[i]==1 | progresso_soup$Month[i]==2 | progresso_soup$Month[i]==10 | progresso_soup$Month[i]==11 | progresso_soup$Month[i]==12)
15 progresso_soup$winter_month[i]<-as.logical(1)
16 }
17 View(progresso_soup)
18
19 subset_wintermonths<-subset(progresso_soup, winter_month==TRUE)
20 subset_nonwintermonths<-subset(progresso_soup, winter_month==FALSE)
21 sales_wintermonths<-subset_wintermonths$Sales.Progresso
22 sales_nonwintermonths<-subset_nonwintermonths$Sales.Progresso
23 sum(sales_wintermonths)
24 marketshare_progresso_winter<-sum(sales_wintermonths)/sum(progresso_soup$Category_Sales)
25 marketshare_progresso_nonwinter<-sum(sales_nonwintermonths)/sum(progresso_soup$Category_Sales)
26 marketshare_progresso_winter
27 marketshare_progresso_nonwinter
28
29
30
31 |
```

Console:

```
> View(progresso_soup)
> subset_wintermonths<-subset(progresso_soup, winter_month==TRUE)
> subset_nonwintermonths<-subset(progresso_soup, winter_month==FALSE)
> subset_wintermonths<-subset(progresso_soup, winter_month==TRUE)
> subset_nonwintermonths<-subset(progresso_soup, winter_month==FALSE)
> sales_wintermonths<-subset_wintermonths$Sales.Progresso
> sales_nonwintermonths<-subset_nonwintermonths$Sales.Progresso
> sum(sales_wintermonths)
[1] 81833499
> subset_wintermonths<-subset(progresso_soup, winter_month==TRUE)
> subset_nonwintermonths<-subset(progresso_soup, winter_month==FALSE)
> sales_wintermonths<-subset_wintermonths$Sales.Progresso
> sales_nonwintermonths<-subset_nonwintermonths$Sales.Progresso
> marketshare_progresso_winter<-sum(sales_wintermonths)/sum(progresso_soup$Category_Sales)
> marketshare_progresso_nonwinter<-sum(sales_nonwintermonths)/sum(progresso_soup$Category_Sales)
> marketshare_progresso_winter
[1] 0.1631347
> marketshare_progresso_nonwinter
[1] 0.08506067
> |
```

Market Share for progresso is 16.31% in Winter months and 8.5% in non Winter Months.

- Develop a linear regression model to predict Progresso sales. Explain the results of the regression model (model strength, variable importance, relationship between the predictor and dependent variables). Use 1st tab in file.

Initial analysis to see how price structuring and price of competition impacts Progresso sales.

The screenshot shows the RStudio interface with the following details:

- Code Editor:** Contains R code for calculating market share and fitting a linear regression model.
- Console:** Displays the output of the R code, including the summary of the linear regression model.
- Environment:** Shows the current environment variables and their values.
- Files:** Shows the project files: `Untitled1*`, `progresso_soup`, `sales_Winter`, and `Untitled2*`.
- Plots:** No plots are visible.
- System Library:** Shows the installed packages: `t Boo`, `Fun`, `Fin`, and `Gro`.

```

25 marketshare_progresso_nonwinter<-sum(sales_nonwintermonths)/sum(progresso_soup$Category_Sales)
26 marketshare_progresso_winter
27 marketshare_progresso_nonwinter
28
29
30
31 #Linear regression model for progresso sales with respect to the price of competitors
32 Sales_Price_model <- lm(Sales.Progresso ~ Price.Progresso + Price.Campbell + Price.PL, data=progresso_soup)
33 summary(Sales_Price_model)
34
35
36
37
39:1 (Top Level) <

```

Console Output:

```

> Sales_Price_model <- lm(Sales.Progresso ~ Price.Progresso + Price.Campbell + Price.PL, data=progresso_soup)
> summary(Sales_Price_model)

Call:
lm(formula = Sales.Progresso ~ Price.Progresso + Price.Campbell +
    Price.PL, data = progresso_soup)

Residuals:
    Min      1Q  Median      3Q     Max 
-4348   -961   -325    455   50232 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3673.86    45.68 80.43 <2e-16 ***
Price.Progresso -3340.65   19.29 -173.20 <2e-16 ***
Price.Campbell  1204.43   32.54  37.02 <2e-16 ***
Price.PL       1356.51   34.60  39.21 <2e-16 ***  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1849 on 88405 degrees of freedom
Multiple R-squared:  0.2615, Adjusted R-squared:  0.2614 
F-statistic: 1.043e+04 on 3 and 88405 DF,  p-value: < 2.2e-16

```

predictor variables: Price.Progresso, Price.Combell, Price.PL (Competitor Prices)

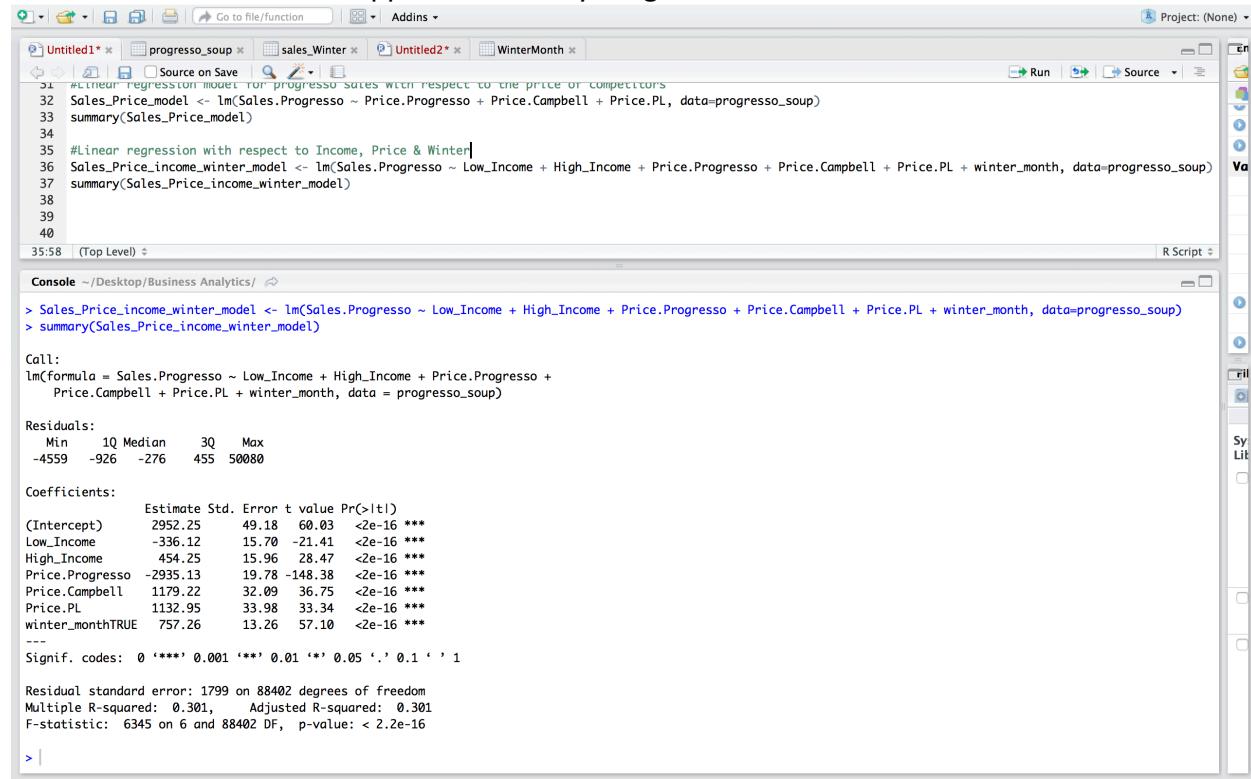
dependent variable: Sales.Progresso

Model Strength: R-squared: 0.26 indicates moderate strength of the model

Results represent progresso price has the strongest influence on sales (coefficient -3340 and t test = -173 > 1.96 for two tailed).

Price of competition also impacts Progresso Sales (increased price of competition = increase in Progresso sales)

The second Linear regression includes income, Price winter months
 It indicates Low income shoppers do not buy Progresso



```

32 #Linear regression model for progresso sales with respect to the price of competitors
33 Sales_Price_model <- lm(Sales.Progresso ~ Price.Progresso + Price.Campbell + Price.PL, data=progresso_soup)
34 summary(Sales_Price_model)
35
36 #Linear regression with respect to Income, Price & Winter
37 Sales_Price_income_winter_model <- lm(Sales.Progresso ~ Low_Income + High_Income + Price.Progresso + Price.Campbell + Price.PL + winter_month, data=progresso_soup)
38 summary(Sales_Price_income_winter_model)
39
40
41
42
43
44
45
46
47
48 | (Top Level) R Script

```

Console ~/Desktop/Business Analytics/

```

> Sales_Price_income_winter_model <- lm(Sales.Progresso ~ Low_Income + High_Income + Price.Progresso + Price.Campbell + Price.PL + winter_month, data=progresso_soup)
> summary(Sales_Price_income_winter_model)

Call:
lm(formula = Sales.Progresso ~ Low_Income + High_Income + Price.Progresso +
Price.Campbell + Price.PL + winter_month, data = progresso_soup)

Residuals:
    Min      1Q Median      3Q     Max 
-4559   -926   -276    455  50080 

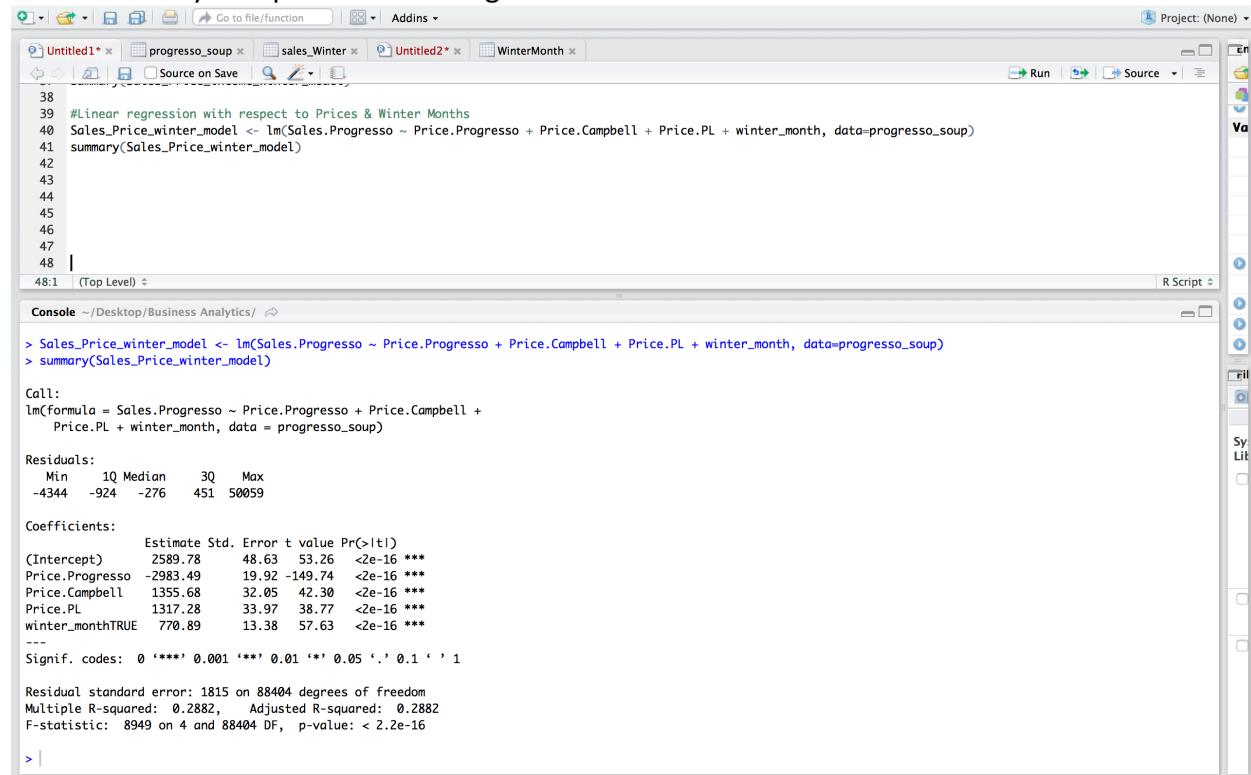
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2952.25    49.18   60.03 <2e-16 ***
Low_Income   -336.12   15.70  -21.41 <2e-16 ***
High_Income    454.25   15.96   28.47 <2e-16 ***
Price.Progresso -2935.13   19.78  -148.38 <2e-16 ***
Price.Campbell  1179.22   32.09   36.75 <2e-16 ***
Price.PL       1132.95   33.98   33.34 <2e-16 ***
winter_monthTRUE  757.26   13.26   57.10 <2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1799 on 88402 degrees of freedom
Multiple R-squared:  0.301, Adjusted R-squared:  0.301 
F-statistic:  6345 on 6 and 88402 DF,  p-value: < 2.2e-16

```

The third analysis represent that Progresso Sales increases in winter



```

38
39 #Linear regression with respect to Prices & Winter Months
40 Sales_Price_winter_model <- lm(Sales.Progresso ~ Price.Progresso + Price.Campbell + Price.PL + winter_month, data=progresso_soup)
41 summary(Sales_Price_winter_model)
42
43
44
45
46
47
48 | (Top Level) R Script

```

Console ~/Desktop/Business Analytics/

```

> Sales_Price_winter_model <- lm(Sales.Progresso ~ Price.Progresso + Price.Campbell + Price.PL + winter_month, data=progresso_soup)
> summary(Sales_Price_winter_model)

Call:
lm(formula = Sales.Progresso ~ Price.Progresso + Price.Campbell +
Price.PL + winter_month, data = progresso_soup)

Residuals:
    Min      1Q Median      3Q     Max 
-4344   -924   -276    451  50059 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2589.78    48.63   53.26 <2e-16 ***
Price.Progresso -2983.49   19.92  -149.74 <2e-16 ***
Price.Campbell  1355.68   32.05   42.30 <2e-16 ***
Price.PL       1317.28   33.97   38.77 <2e-16 ***
winter_monthTRUE  770.89   13.38   57.63 <2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1815 on 88404 degrees of freedom
Multiple R-squared:  0.2882, Adjusted R-squared:  0.2882 
F-statistic:  8949 on 4 and 88404 DF,  p-value: < 2.2e-16

```