Group 8 Project ITEC-620

Loc Le

2023-11-30

Packages

library(tree)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 0.3.4   
## ✔ tibble 3.1.6 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.1 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(dplyr)  
library(class)  
library(fastDummies)

## Thank you for using fastDummies!  
## To acknowledge our work, please cite the package:  
## Kaplan, J. & Schlegel, B. (2023). fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from Categorical Variables. Version 1.7.1. URL: https://github.com/jacobkap/fastDummies, https://jacobkap.github.io/fastDummies/.

library(FNN)

##   
## Attaching package: 'FNN'  
##   
## The following objects are masked from 'package:class':  
##   
## knn, knn.cv

**IMPORTANT**

* Use specific Amazon.csv file uploaded in OneDrive folder
* Run each code chunk one by one (do not run all)

Loading Kaggle Dataset (Saved as Amazon.csv)

Amazon <- read.csv("Amazon.csv")

Formatting Dataset

# ONLY RUN THIS CODE CHUNK ONCE  
  
Amazon <- Amazon[,-1] # Remove Timestamp Column

# Creates a table of each variable to check for weird values such as N/As, blanks, etc. and infrequently occuring values that can be lumped into an "Other" category per the professor's suggestion  
  
variable\_names <- names(Amazon)  
  
frequency\_tables <- list()  
  
for (variable in variable\_names) {  
 frequency\_tables[[variable]] <- table(Amazon[[variable]])  
}  
  
for (i in seq\_along(frequency\_tables)) {  
 cat("Variable:", names(frequency\_tables)[i], "\n")  
 print(frequency\_tables[[i]])  
 cat("\n")  
}

## Variable: age   
##   
## 3 12 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32   
## 1 1 2 5 4 7 4 5 30 17 123 40 36 27 17 9 9 8 9 19   
## 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 52 53   
## 7 48 15 16 14 4 4 12 4 4 6 5 34 12 6 3 1 5 1 2   
## 54 55 56 57 58 60 62 63 64 67   
## 6 1 8 2 1 3 1 1 1 2   
##   
## Variable: Gender   
##   
## Female Male Others Prefer not to say   
## 352 142 19 89   
##   
## Variable: Purchase\_Frequency   
##   
## Few times a month Less than once a month Multiple times a week   
## 203 124 56   
## Once a month Once a week   
## 107 112   
##   
## Variable: Purchase\_Categories   
##   
## Beauty and Personal Care   
## 106   
## Beauty and Personal Care;Clothing and Fashion   
## 46   
## Beauty and Personal Care;Clothing and Fashion;Home and Kitchen   
## 42   
## Beauty and Personal Care;Clothing and Fashion;Home and Kitchen;others   
## 8   
## Beauty and Personal Care;Clothing and Fashion;others   
## 12   
## Beauty and Personal Care;Home and Kitchen   
## 21   
## Beauty and Personal Care;Home and Kitchen;others   
## 5   
## Beauty and Personal Care;others   
## 7   
## Clothing and Fashion   
## 106   
## Clothing and Fashion;Home and Kitchen   
## 27   
## Clothing and Fashion;Home and Kitchen;others   
## 16   
## Clothing and Fashion;others   
## 14   
## Groceries and Gourmet Food   
## 14   
## Groceries and Gourmet Food;Beauty and Personal Care   
## 7   
## Groceries and Gourmet Food;Beauty and Personal Care;Clothing and Fashion   
## 10   
## Groceries and Gourmet Food;Beauty and Personal Care;Clothing and Fashion;Home and Kitchen   
## 14   
## Groceries and Gourmet Food;Beauty and Personal Care;Clothing and Fashion;Home and Kitchen;others   
## 32   
## Groceries and Gourmet Food;Beauty and Personal Care;Clothing and Fashion;others   
## 1   
## Groceries and Gourmet Food;Beauty and Personal Care;Home and Kitchen   
## 4   
## Groceries and Gourmet Food;Beauty and Personal Care;others   
## 3   
## Groceries and Gourmet Food;Clothing and Fashion   
## 6   
## Groceries and Gourmet Food;Clothing and Fashion;Home and Kitchen   
## 4   
## Groceries and Gourmet Food;Clothing and Fashion;Home and Kitchen;others   
## 3   
## Groceries and Gourmet Food;Clothing and Fashion;others   
## 2   
## Groceries and Gourmet Food;Home and Kitchen   
## 5   
## Groceries and Gourmet Food;Home and Kitchen;others   
## 6   
## Home and Kitchen   
## 24   
## Home and Kitchen;others   
## 9   
## others   
## 48   
##   
## Variable: Personalized\_Recommendation\_Frequency   
##   
## No Sometimes Yes   
## 251 229 122   
##   
## Variable: Browsing\_Frequency   
##   
## Few times a month Few times a week Multiple times a day   
## 199 249 77   
## Rarely   
## 77   
##   
## Variable: Product\_Search\_Method   
##   
## categories Filter Keyword others   
## 2 223 127 214 36   
##   
## Variable: Search\_Result\_Exploration   
##   
## First page Multiple pages   
## 160 442   
##   
## Variable: Customer\_Reviews\_Importance   
##   
## 1 2 3 4 5   
## 169 115 216 64 38   
##   
## Variable: Add\_to\_Cart\_Browsing   
##   
## Maybe No Yes   
## 248 138 216   
##   
## Variable: Cart\_Completion\_Frequency   
##   
## Always Never Often Rarely Sometimes   
## 47 21 158 72 304   
##   
## Variable: Cart\_Abandonment\_Factors   
##   
## Changed my mind or no longer need the item   
## 241   
## Found a better price elsewhere   
## 255   
## High shipping costs   
## 70   
## others   
## 36   
##   
## Variable: Saveforlater\_Frequency   
##   
## Always Never Often Rarely Sometimes   
## 54 59 156 82 251   
##   
## Variable: Review\_Left   
##   
## No Yes   
## 292 310   
##   
## Variable: Review\_Reliability   
##   
## Heavily Moderately Never Occasionally Rarely   
## 149 199 23 190 41   
##   
## Variable: Review\_Helpfulness   
##   
## No Sometimes Yes   
## 138 227 237   
##   
## Variable: Personalized\_Recommendation\_Frequency.1   
##   
## 1 2 3 4 5   
## 80 169 240 78 35   
##   
## Variable: Recommendation\_Helpfulness   
##   
## No Sometimes Yes   
## 172 273 157   
##   
## Variable: Rating\_Accuracy   
##   
## 1 2 3 4 5   
## 58 179 288 56 21   
##   
## Variable: Shopping\_Satisfaction   
##   
## 1 2 3 4 5   
## 115 194 209 67 17   
##   
## Variable: Service\_Appreciation   
##   
## . All the above   
## 1 1   
## Competitive prices Customer service   
## 182 2   
## Product recommendations Quick delivery   
## 185 1   
## User-friendly website/app interface Wide product selection   
## 80 150   
##   
## Variable: Improvement\_Areas   
##   
## .   
## 1   
## Add more familiar brands to the list   
## 1   
## App UI   
## 1   
## better app interface and lower shipping charges   
## 1   
## Customer service responsiveness   
## 217   
## I don't have any problem with Amazon   
## 1   
## I have no problem with Amazon yet but others tell me about the refund issues   
## 1   
## Irrelevant product suggestions   
## 1   
## Nil   
## 1   
## No problems with Amazon   
## 1   
## Nothing   
## 1   
## Product quality and accuracy   
## 159   
## Quality of product is very poor according to the big offers   
## 1   
## Reducing packaging waste   
## 133   
## Scrolling option would be much better than going to next page   
## 1   
## Shipping speed and reliability   
## 79   
## User interface   
## 2

# Removing rows with weird values  
  
Amazon <- Amazon[-c(23,120,383),]

# Gender  
  
Amazon <- Amazon %>%   
 mutate(Gender = str\_replace(Gender,   
 "Prefer not to say", "Others"))

# Lumping purchase subcategories into their parent categories to reduce total number of categories  
  
# Beauty and Personal Care  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Beauty and Personal Care;Clothing and Fashion", "Beauty and Personal Care"))  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Beauty and Personal Care;Home and Kitchen", "Beauty and Personal Care"))  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Beauty and Personal Care;others", "Beauty and Personal Care"))  
  
# Clothing and Fashion  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Clothing and Fashion;Home and Kitchen", "Clothing and Fashion"))  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Clothing and Fashion;others", "Clothing and Fashion"))  
  
# Groceries and Gourmet Food  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Groceries and Gourmet Food;Beauty and Personal Care", "Groceries and Gourmet Food"))  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Groceries and Gourmet Food;Clothing and Fashion", "Groceries and Gourmet Food"))  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Groceries and Gourmet Food;Home and Kitchen", "Groceries and Gourmet Food"))  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Groceries and Gourmet Food;Home and Kitchen;others", "Groceries and Gourmet Food"))  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Groceries and Gourmet Food;others", "Groceries and Gourmet Food"))  
  
# Home and Kitchen  
  
Amazon <- Amazon %>%   
 mutate(Purchase\_Categories = str\_replace(Purchase\_Categories,   
 "Home and Kitchen;others", "Home and Kitchen"))

# Lumping infrequent values in Service\_Appreciation to an "Other" category  
  
# All the above  
  
Amazon <- Amazon %>%   
 mutate(Service\_Appreciation = str\_replace(Service\_Appreciation,   
 "All the above", "Other"))  
  
# Customer service  
  
Amazon <- Amazon %>%   
 mutate(Service\_Appreciation = str\_replace(Service\_Appreciation,   
 "Customer service", "Other"))  
  
# Quick delivery  
  
Amazon <- Amazon %>%   
 mutate(Service\_Appreciation = str\_replace(Service\_Appreciation,   
 "Quick delivery", "Other"))

# Lumping similar values together as well as infrequent values in Improvement\_Areas to an "Other" category  
  
# Other  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "Add more familiar brands to the list", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "better app interface and lower shipping charges", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "Irrelevant product suggestions", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "I don't have any problem with Amazon", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "I have no problem with Amazon yet but others tell me about the refund issues", "Other"))  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "Nil", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "No problems with Amazon", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "Nothing", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "Scrolling option would be much better than going to next page", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "User interface", "Other"))  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "App UI", "Other"))  
  
# Product quality and accuracy  
  
Amazon <- Amazon %>%   
 mutate(Improvement\_Areas = str\_replace(Improvement\_Areas,   
 "Quality of product is very poor according to the big offers", "Product quality and accuracy"))

# Creates a new column called Purchase\_Frequency\_Num for transforming Purchase\_Frequency values into numeric values per the professor's suggestion  
  
Amazon$Purchase\_Frequency\_Num <- 0.5 # Less than once a month  
  
Amazon$Purchase\_Frequency\_Num[Amazon$Purchase\_Frequency == "Once a month"] <- 1  
  
Amazon$Purchase\_Frequency\_Num[Amazon$Purchase\_Frequency == "Few times a month"] <- 3  
  
Amazon$Purchase\_Frequency\_Num[Amazon$Purchase\_Frequency == "Once a week"] <- 5  
  
Amazon$Purchase\_Frequency\_Num[Amazon$Purchase\_Frequency == "Multiple times a week"] <- 7

Less than once a month = 0.5

Once a month = 1

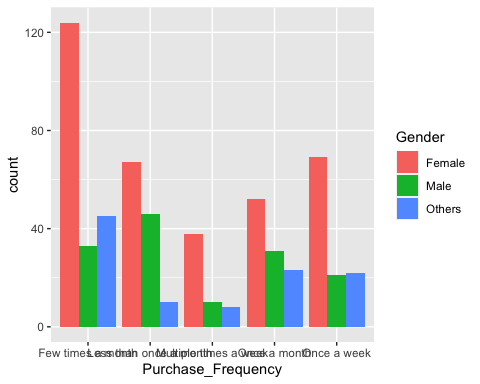
Few times a month = 3

Once a week = 5

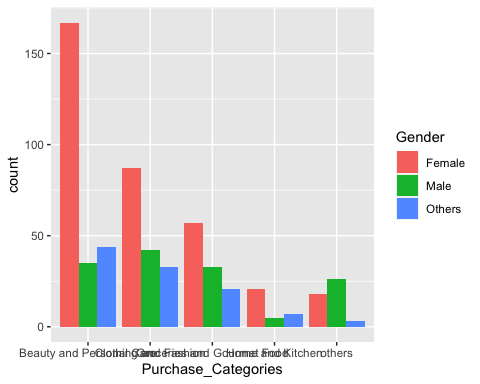
Multiple times a week = 7

Exploratory Data Analysis:

ggplot(data = Amazon, mapping = aes(x = Purchase\_Frequency, fill = Gender)) +  
 geom\_bar(position = "dodge")

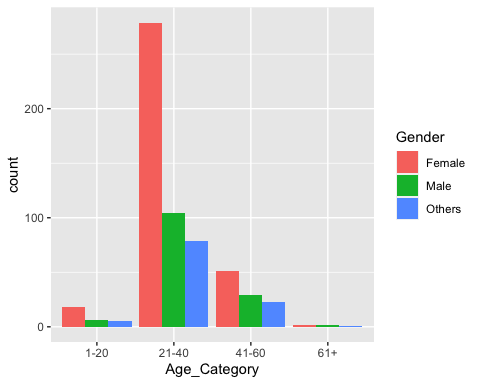
 Majority of Amazon customers shop there a few times a month and females are the most prominent shoppers.

ggplot(data = Amazon, mapping = aes(x = Purchase\_Categories, fill = Gender)) +  
 geom\_bar(position = "dodge")

 The most popular category at Amazon is Beauty and Personal Care and it is where the majority of females are shopping.

Amazon$Age\_Category <- ifelse(Amazon$age <= 20, "1-20",  
 ifelse(Amazon$age <= 40, "21-40",  
 ifelse(Amazon$age <= 60, "41-60", "61+")))

ggplot(data = Amazon, mapping = aes(x = Age\_Category, fill = Gender)) +  
 geom\_bar(position = "dodge")

 The majority of female shoppers are between the ages of 21-40

# ONLY RUN THIS CODE CHUNK ONCE  
  
Amazon <- Amazon[,-3] # Remove old Purchase\_Frequency column

# ONLY RUN THIS CODE CHUNK ONCE  
  
Amazon <- Amazon[,-23] # Remove Age\_Category column

# Creating dummy variables from categorical variables  
  
Amazon <- dummy\_cols(Amazon, select\_columns=c("Gender", "Purchase\_Categories", "Personalized\_Recommendation\_Frequency", "Browsing\_Frequency", "Product\_Search\_Method", "Search\_Result\_Exploration", "Add\_to\_Cart\_Browsing", "Cart\_Completion\_Frequency", "Cart\_Abandonment\_Factors", "Saveforlater\_Frequency", "Review\_Left", "Review\_Reliability", "Review\_Helpfulness", "Recommendation\_Helpfulness", "Service\_Appreciation", "Improvement\_Areas"))

# Remove old categorical variables  
Amazon <- Amazon[,-c(2:7,9:15,17,20,21)]

glimpse(Amazon)

## Rows: 599  
## Columns: 67  
## $ age <int> …  
## $ Customer\_Reviews\_Importance <int> …  
## $ Personalized\_Recommendation\_Frequency.1 <int> …  
## $ Rating\_Accuracy <int> …  
## $ Shopping\_Satisfaction <int> …  
## $ Purchase\_Frequency\_Num <dbl> …  
## $ Gender\_Female <int> …  
## $ Gender\_Male <int> …  
## $ Gender\_Others <int> …  
## $ `Purchase\_Categories\_Beauty and Personal Care` <int> …  
## $ `Purchase\_Categories\_Clothing and Fashion` <int> …  
## $ `Purchase\_Categories\_Groceries and Gourmet Food` <int> …  
## $ `Purchase\_Categories\_Home and Kitchen` <int> …  
## $ Purchase\_Categories\_others <int> …  
## $ Personalized\_Recommendation\_Frequency\_No <int> …  
## $ Personalized\_Recommendation\_Frequency\_Sometimes <int> …  
## $ Personalized\_Recommendation\_Frequency\_Yes <int> …  
## $ `Browsing\_Frequency\_Few times a month` <int> …  
## $ `Browsing\_Frequency\_Few times a week` <int> …  
## $ `Browsing\_Frequency\_Multiple times a day` <int> …  
## $ Browsing\_Frequency\_Rarely <int> …  
## $ Product\_Search\_Method\_categories <int> …  
## $ Product\_Search\_Method\_Filter <int> …  
## $ Product\_Search\_Method\_Keyword <int> …  
## $ Product\_Search\_Method\_others <int> …  
## $ `Search\_Result\_Exploration\_First page` <int> …  
## $ `Search\_Result\_Exploration\_Multiple pages` <int> …  
## $ Add\_to\_Cart\_Browsing\_Maybe <int> …  
## $ Add\_to\_Cart\_Browsing\_No <int> …  
## $ Add\_to\_Cart\_Browsing\_Yes <int> …  
## $ Cart\_Completion\_Frequency\_Always <int> …  
## $ Cart\_Completion\_Frequency\_Never <int> …  
## $ Cart\_Completion\_Frequency\_Often <int> …  
## $ Cart\_Completion\_Frequency\_Rarely <int> …  
## $ Cart\_Completion\_Frequency\_Sometimes <int> …  
## $ `Cart\_Abandonment\_Factors\_Changed my mind or no longer need the item` <int> …  
## $ `Cart\_Abandonment\_Factors\_Found a better price elsewhere` <int> …  
## $ `Cart\_Abandonment\_Factors\_High shipping costs` <int> …  
## $ Cart\_Abandonment\_Factors\_others <int> …  
## $ Saveforlater\_Frequency\_Always <int> …  
## $ Saveforlater\_Frequency\_Never <int> …  
## $ Saveforlater\_Frequency\_Often <int> …  
## $ Saveforlater\_Frequency\_Rarely <int> …  
## $ Saveforlater\_Frequency\_Sometimes <int> …  
## $ Review\_Left\_No <int> …  
## $ Review\_Left\_Yes <int> …  
## $ Review\_Reliability\_Heavily <int> …  
## $ Review\_Reliability\_Moderately <int> …  
## $ Review\_Reliability\_Never <int> …  
## $ Review\_Reliability\_Occasionally <int> …  
## $ Review\_Reliability\_Rarely <int> …  
## $ Review\_Helpfulness\_No <int> …  
## $ Review\_Helpfulness\_Sometimes <int> …  
## $ Review\_Helpfulness\_Yes <int> …  
## $ Recommendation\_Helpfulness\_No <int> …  
## $ Recommendation\_Helpfulness\_Sometimes <int> …  
## $ Recommendation\_Helpfulness\_Yes <int> …  
## $ `Service\_Appreciation\_Competitive prices` <int> …  
## $ Service\_Appreciation\_Other <int> …  
## $ `Service\_Appreciation\_Product recommendations` <int> …  
## $ `Service\_Appreciation\_User-friendly website/app interface` <int> …  
## $ `Service\_Appreciation\_Wide product selection` <int> …  
## $ `Improvement\_Areas\_Customer service responsiveness` <int> …  
## $ Improvement\_Areas\_Other <int> …  
## $ `Improvement\_Areas\_Product quality and accuracy` <int> …  
## $ `Improvement\_Areas\_Reducing packaging waste` <int> …  
## $ `Improvement\_Areas\_Shipping speed and reliability` <int> …

# Replace spaces in variables with underscores  
colnames(Amazon) <- gsub(" ", "\_", colnames(Amazon))

# Manual fix for this variable  
names(Amazon)[names(Amazon) == "Service\_Appreciation\_User-friendly\_website/app\_interface"] <- "Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface"

# Regression Analysis

set.seed(12345)  
training <- sample(1:nrow(Amazon), 0.6\*nrow(Amazon))  
ycol <- match("Purchase\_Frequency\_Num",colnames(Amazon))  
Amazon.training <- Amazon[training,-ycol]  
Amazon.training.results <- Amazon[training,ycol]  
Amazon.test <- Amazon[-training,-ycol]  
Amazon.test.results <- Amazon[-training,ycol]

# Linear Regression with all variables (removed variables are baselines)  
Amazon.reg.full <- lm(Purchase\_Frequency\_Num ~ .   
 -Gender\_Male   
 -Purchase\_Categories\_Beauty\_and\_Personal\_Care   
 - Personalized\_Recommendation\_Frequency\_No  
 - Browsing\_Frequency\_Multiple\_times\_a\_day  
 - Product\_Search\_Method\_categories  
 - Search\_Result\_Exploration\_Multiple\_pages  
 - Add\_to\_Cart\_Browsing\_Yes  
 - Cart\_Completion\_Frequency\_Never  
 - Cart\_Abandonment\_Factors\_Found\_a\_better\_price\_elsewhere  
 - Saveforlater\_Frequency\_Never  
 - Review\_Left\_Yes  
 - Review\_Reliability\_Never  
 - Review\_Helpfulness\_Yes  
 - Recommendation\_Helpfulness\_No  
 - Service\_Appreciation\_Wide\_product\_selection  
 - Improvement\_Areas\_Reducing\_packaging\_waste  
 ,data = Amazon[training,])  
  
Amazon.reg.full.predictions <- predict(Amazon.reg.full,Amazon)[-training]  
(mean((Amazon.test.results-Amazon.reg.full.predictions)^2))^0.5

## [1] 1.990002

RMSE of linear regression model with all variables = 1.990002

summary(Amazon.reg.full)

##   
## Call:  
## lm(formula = Purchase\_Frequency\_Num ~ . - Gender\_Male - Purchase\_Categories\_Beauty\_and\_Personal\_Care -   
## Personalized\_Recommendation\_Frequency\_No - Browsing\_Frequency\_Multiple\_times\_a\_day -   
## Product\_Search\_Method\_categories - Search\_Result\_Exploration\_Multiple\_pages -   
## Add\_to\_Cart\_Browsing\_Yes - Cart\_Completion\_Frequency\_Never -   
## Cart\_Abandonment\_Factors\_Found\_a\_better\_price\_elsewhere -   
## Saveforlater\_Frequency\_Never - Review\_Left\_Yes - Review\_Reliability\_Never -   
## Review\_Helpfulness\_Yes - Recommendation\_Helpfulness\_No -   
## Service\_Appreciation\_Wide\_product\_selection - Improvement\_Areas\_Reducing\_packaging\_waste,   
## data = Amazon[training, ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.0005 -1.1034 -0.1433 0.9682 5.3029   
##   
## Coefficients:  
## Estimate  
## (Intercept) 0.8502448  
## age 0.0062191  
## Customer\_Reviews\_Importance 0.1581668  
## Personalized\_Recommendation\_Frequency.1 0.1747385  
## Rating\_Accuracy -0.2443125  
## Shopping\_Satisfaction 0.0118960  
## Gender\_Female 0.5749352  
## Gender\_Others 0.5634569  
## Purchase\_Categories\_Clothing\_and\_Fashion -0.6053618  
## Purchase\_Categories\_Groceries\_and\_Gourmet\_Food -0.6125794  
## Purchase\_Categories\_Home\_and\_Kitchen -1.1055028  
## Purchase\_Categories\_others -0.9137327  
## Personalized\_Recommendation\_Frequency\_Sometimes 0.2061765  
## Personalized\_Recommendation\_Frequency\_Yes 0.7006645  
## Browsing\_Frequency\_Few\_times\_a\_month -0.0126202  
## Browsing\_Frequency\_Few\_times\_a\_week 0.0558355  
## Browsing\_Frequency\_Rarely -0.4396606  
## Product\_Search\_Method\_Filter 0.2054439  
## Product\_Search\_Method\_Keyword -0.1707587  
## Product\_Search\_Method\_others -0.1993489  
## Search\_Result\_Exploration\_First\_page -0.0992158  
## Add\_to\_Cart\_Browsing\_Maybe -0.0856382  
## Add\_to\_Cart\_Browsing\_No 0.6127614  
## Cart\_Completion\_Frequency\_Always 2.1016356  
## Cart\_Completion\_Frequency\_Often 1.0576198  
## Cart\_Completion\_Frequency\_Rarely 0.4427738  
## Cart\_Completion\_Frequency\_Sometimes 0.6907956  
## Cart\_Abandonment\_Factors\_Changed\_my\_mind\_or\_no\_longer\_need\_the\_item 0.1088111  
## Cart\_Abandonment\_Factors\_High\_shipping\_costs -0.0869809  
## Cart\_Abandonment\_Factors\_others 0.3070986  
## Saveforlater\_Frequency\_Always 0.2013356  
## Saveforlater\_Frequency\_Often -0.1589321  
## Saveforlater\_Frequency\_Rarely -0.5504996  
## Saveforlater\_Frequency\_Sometimes -0.3273684  
## Review\_Left\_No -0.5017806  
## Review\_Reliability\_Heavily 0.6740719  
## Review\_Reliability\_Moderately 0.6938192  
## Review\_Reliability\_Occasionally 0.6646928  
## Review\_Reliability\_Rarely -0.5142558  
## Review\_Helpfulness\_No 0.2658284  
## Review\_Helpfulness\_Sometimes 0.3133095  
## Recommendation\_Helpfulness\_Sometimes 0.3125752  
## Recommendation\_Helpfulness\_Yes 0.0005032  
## Service\_Appreciation\_Competitive\_prices 0.2460635  
## Service\_Appreciation\_Other -0.5523913  
## Service\_Appreciation\_Product\_recommendations 0.5681681  
## Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface -0.3062684  
## Improvement\_Areas\_Customer\_service\_responsiveness -0.0379210  
## Improvement\_Areas\_Other -0.1204651  
## Improvement\_Areas\_Product\_quality\_and\_accuracy -0.4421329  
## Improvement\_Areas\_Shipping\_speed\_and\_reliability -0.9433615  
## Std. Error  
## (Intercept) 1.2477364  
## age 0.0099780  
## Customer\_Reviews\_Importance 0.1001245  
## Personalized\_Recommendation\_Frequency.1 0.1073424  
## Rating\_Accuracy 0.1361945  
## Shopping\_Satisfaction 0.1268514  
## Gender\_Female 0.2511238  
## Gender\_Others 0.3240242  
## Purchase\_Categories\_Clothing\_and\_Fashion 0.2423512  
## Purchase\_Categories\_Groceries\_and\_Gourmet\_Food 0.2825529  
## Purchase\_Categories\_Home\_and\_Kitchen 0.4615775  
## Purchase\_Categories\_others 0.4043411  
## Personalized\_Recommendation\_Frequency\_Sometimes 0.2496874  
## Personalized\_Recommendation\_Frequency\_Yes 0.2782718  
## Browsing\_Frequency\_Few\_times\_a\_month 0.3504977  
## Browsing\_Frequency\_Few\_times\_a\_week 0.3281752  
## Browsing\_Frequency\_Rarely 0.4327315  
## Product\_Search\_Method\_Filter 0.2651386  
## Product\_Search\_Method\_Keyword 0.2682532  
## Product\_Search\_Method\_others 0.4681528  
## Search\_Result\_Exploration\_First\_page 0.2374526  
## Add\_to\_Cart\_Browsing\_Maybe 0.2733930  
## Add\_to\_Cart\_Browsing\_No 0.3100864  
## Cart\_Completion\_Frequency\_Always 0.7895752  
## Cart\_Completion\_Frequency\_Often 0.7221506  
## Cart\_Completion\_Frequency\_Rarely 0.7171716  
## Cart\_Completion\_Frequency\_Sometimes 0.7008467  
## Cart\_Abandonment\_Factors\_Changed\_my\_mind\_or\_no\_longer\_need\_the\_item 0.2104587  
## Cart\_Abandonment\_Factors\_High\_shipping\_costs 0.3302184  
## Cart\_Abandonment\_Factors\_others 0.4729057  
## Saveforlater\_Frequency\_Always 0.4661498  
## Saveforlater\_Frequency\_Often 0.4060210  
## Saveforlater\_Frequency\_Rarely 0.4338445  
## Saveforlater\_Frequency\_Sometimes 0.4057047  
## Review\_Left\_No 0.2130475  
## Review\_Reliability\_Heavily 0.6485121  
## Review\_Reliability\_Moderately 0.6285155  
## Review\_Reliability\_Occasionally 0.6274229  
## Review\_Reliability\_Rarely 0.6981347  
## Review\_Helpfulness\_No 0.3311045  
## Review\_Helpfulness\_Sometimes 0.2790641  
## Recommendation\_Helpfulness\_Sometimes 0.2734017  
## Recommendation\_Helpfulness\_Yes 0.3156521  
## Service\_Appreciation\_Competitive\_prices 0.2846260  
## Service\_Appreciation\_Other 1.3719085  
## Service\_Appreciation\_Product\_recommendations 0.3029427  
## Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface 0.3358443  
## Improvement\_Areas\_Customer\_service\_responsiveness 0.2660253  
## Improvement\_Areas\_Other 0.7624790  
## Improvement\_Areas\_Product\_quality\_and\_accuracy 0.3033437  
## Improvement\_Areas\_Shipping\_speed\_and\_reliability 0.3506599  
## t value  
## (Intercept) 0.681  
## age 0.623  
## Customer\_Reviews\_Importance 1.580  
## Personalized\_Recommendation\_Frequency.1 1.628  
## Rating\_Accuracy -1.794  
## Shopping\_Satisfaction 0.094  
## Gender\_Female 2.289  
## Gender\_Others 1.739  
## Purchase\_Categories\_Clothing\_and\_Fashion -2.498  
## Purchase\_Categories\_Groceries\_and\_Gourmet\_Food -2.168  
## Purchase\_Categories\_Home\_and\_Kitchen -2.395  
## Purchase\_Categories\_others -2.260  
## Personalized\_Recommendation\_Frequency\_Sometimes 0.826  
## Personalized\_Recommendation\_Frequency\_Yes 2.518  
## Browsing\_Frequency\_Few\_times\_a\_month -0.036  
## Browsing\_Frequency\_Few\_times\_a\_week 0.170  
## Browsing\_Frequency\_Rarely -1.016  
## Product\_Search\_Method\_Filter 0.775  
## Product\_Search\_Method\_Keyword -0.637  
## Product\_Search\_Method\_others -0.426  
## Search\_Result\_Exploration\_First\_page -0.418  
## Add\_to\_Cart\_Browsing\_Maybe -0.313  
## Add\_to\_Cart\_Browsing\_No 1.976  
## Cart\_Completion\_Frequency\_Always 2.662  
## Cart\_Completion\_Frequency\_Often 1.465  
## Cart\_Completion\_Frequency\_Rarely 0.617  
## Cart\_Completion\_Frequency\_Sometimes 0.986  
## Cart\_Abandonment\_Factors\_Changed\_my\_mind\_or\_no\_longer\_need\_the\_item 0.517  
## Cart\_Abandonment\_Factors\_High\_shipping\_costs -0.263  
## Cart\_Abandonment\_Factors\_others 0.649  
## Saveforlater\_Frequency\_Always 0.432  
## Saveforlater\_Frequency\_Often -0.391  
## Saveforlater\_Frequency\_Rarely -1.269  
## Saveforlater\_Frequency\_Sometimes -0.807  
## Review\_Left\_No -2.355  
## Review\_Reliability\_Heavily 1.039  
## Review\_Reliability\_Moderately 1.104  
## Review\_Reliability\_Occasionally 1.059  
## Review\_Reliability\_Rarely -0.737  
## Review\_Helpfulness\_No 0.803  
## Review\_Helpfulness\_Sometimes 1.123  
## Recommendation\_Helpfulness\_Sometimes 1.143  
## Recommendation\_Helpfulness\_Yes 0.002  
## Service\_Appreciation\_Competitive\_prices 0.865  
## Service\_Appreciation\_Other -0.403  
## Service\_Appreciation\_Product\_recommendations 1.875  
## Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface -0.912  
## Improvement\_Areas\_Customer\_service\_responsiveness -0.143  
## Improvement\_Areas\_Other -0.158  
## Improvement\_Areas\_Product\_quality\_and\_accuracy -1.458  
## Improvement\_Areas\_Shipping\_speed\_and\_reliability -2.690  
## Pr(>|t|)   
## (Intercept) 0.49611   
## age 0.53356   
## Customer\_Reviews\_Importance 0.11520   
## Personalized\_Recommendation\_Frequency.1 0.10458   
## Rating\_Accuracy 0.07382 .   
## Shopping\_Satisfaction 0.92535   
## Gender\_Female 0.02273 \*   
## Gender\_Others 0.08305 .   
## Purchase\_Categories\_Clothing\_and\_Fashion 0.01302 \*   
## Purchase\_Categories\_Groceries\_and\_Gourmet\_Food 0.03092 \*   
## Purchase\_Categories\_Home\_and\_Kitchen 0.01722 \*   
## Purchase\_Categories\_others 0.02453 \*   
## Personalized\_Recommendation\_Frequency\_Sometimes 0.40959   
## Personalized\_Recommendation\_Frequency\_Yes 0.01231 \*   
## Browsing\_Frequency\_Few\_times\_a\_month 0.97130   
## Browsing\_Frequency\_Few\_times\_a\_week 0.86501   
## Browsing\_Frequency\_Rarely 0.31042   
## Product\_Search\_Method\_Filter 0.43902   
## Product\_Search\_Method\_Keyword 0.52489   
## Product\_Search\_Method\_others 0.67054   
## Search\_Result\_Exploration\_First\_page 0.67636   
## Add\_to\_Cart\_Browsing\_Maybe 0.75431   
## Add\_to\_Cart\_Browsing\_No 0.04904 \*   
## Cart\_Completion\_Frequency\_Always 0.00818 \*\*  
## Cart\_Completion\_Frequency\_Often 0.14407   
## Cart\_Completion\_Frequency\_Rarely 0.53743   
## Cart\_Completion\_Frequency\_Sometimes 0.32507   
## Cart\_Abandonment\_Factors\_Changed\_my\_mind\_or\_no\_longer\_need\_the\_item 0.60551   
## Cart\_Abandonment\_Factors\_High\_shipping\_costs 0.79242   
## Cart\_Abandonment\_Factors\_others 0.51657   
## Saveforlater\_Frequency\_Always 0.66611   
## Saveforlater\_Frequency\_Often 0.69574   
## Saveforlater\_Frequency\_Rarely 0.20544   
## Saveforlater\_Frequency\_Sometimes 0.42034   
## Review\_Left\_No 0.01914 \*   
## Review\_Reliability\_Heavily 0.29943   
## Review\_Reliability\_Moderately 0.27050   
## Review\_Reliability\_Occasionally 0.29025   
## Review\_Reliability\_Rarely 0.46192   
## Review\_Helpfulness\_No 0.42268   
## Review\_Helpfulness\_Sometimes 0.26243   
## Recommendation\_Helpfulness\_Sometimes 0.25381   
## Recommendation\_Helpfulness\_Yes 0.99873   
## Service\_Appreciation\_Competitive\_prices 0.38798   
## Service\_Appreciation\_Other 0.68749   
## Service\_Appreciation\_Product\_recommendations 0.06167 .   
## Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface 0.36252   
## Improvement\_Areas\_Customer\_service\_responsiveness 0.88674   
## Improvement\_Areas\_Other 0.87457   
## Improvement\_Areas\_Product\_quality\_and\_accuracy 0.14599   
## Improvement\_Areas\_Shipping\_speed\_and\_reliability 0.00753 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.707 on 308 degrees of freedom  
## Multiple R-squared: 0.3914, Adjusted R-squared: 0.2926   
## F-statistic: 3.961 on 50 and 308 DF, p-value: 4.396e-14

# New model with insignificant variables removed  
  
Amazon.reg.reduced <- lm(Purchase\_Frequency\_Num ~ .   
 -Gender\_Male   
 -Purchase\_Categories\_Beauty\_and\_Personal\_Care   
 -Personalized\_Recommendation\_Frequency\_No  
 -Browsing\_Frequency\_Multiple\_times\_a\_day  
 -Product\_Search\_Method\_categories  
 -Search\_Result\_Exploration\_Multiple\_pages  
 -Add\_to\_Cart\_Browsing\_Yes  
 -Cart\_Completion\_Frequency\_Never  
 -Cart\_Abandonment\_Factors\_Found\_a\_better\_price\_elsewhere  
 -Saveforlater\_Frequency\_Never  
 -Review\_Left\_Yes  
 -Review\_Reliability\_Never  
 -Review\_Helpfulness\_Yes  
 -Recommendation\_Helpfulness\_No  
 -Service\_Appreciation\_Wide\_product\_selection  
 -Improvement\_Areas\_Reducing\_packaging\_waste  
 -age  
 -Customer\_Reviews\_Importance  
 -Personalized\_Recommendation\_Frequency.1  
 -Rating\_Accuracy  
 -Shopping\_Satisfaction  
 -Gender\_Others  
 -Personalized\_Recommendation\_Frequency\_Sometimes  
 -Browsing\_Frequency\_Few\_times\_a\_month  
 -Browsing\_Frequency\_Few\_times\_a\_week  
 -Browsing\_Frequency\_Rarely  
 -Product\_Search\_Method\_Filter  
 -Product\_Search\_Method\_Keyword  
 -Product\_Search\_Method\_others  
 -Search\_Result\_Exploration\_First\_page  
 -Add\_to\_Cart\_Browsing\_Maybe  
 -Cart\_Completion\_Frequency\_Often  
 -Cart\_Completion\_Frequency\_Rarely  
 -Cart\_Completion\_Frequency\_Sometimes  
 -Cart\_Abandonment\_Factors\_Changed\_my\_mind\_or\_no\_longer\_need\_the\_item  
 -Cart\_Abandonment\_Factors\_High\_shipping\_costs  
 -Cart\_Abandonment\_Factors\_others  
 -Saveforlater\_Frequency\_Always  
 -Saveforlater\_Frequency\_Often  
 -Saveforlater\_Frequency\_Rarely  
 -Saveforlater\_Frequency\_Sometimes  
 -Review\_Reliability\_Heavily  
 -Review\_Reliability\_Moderately  
 -Review\_Reliability\_Occasionally  
 -Review\_Reliability\_Rarely  
 -Review\_Helpfulness\_No  
 -Review\_Helpfulness\_Sometimes  
 -Recommendation\_Helpfulness\_Sometimes  
 -Recommendation\_Helpfulness\_Yes  
 -Service\_Appreciation\_Competitive\_prices  
 -Service\_Appreciation\_Other  
 -Service\_Appreciation\_Product\_recommendations  
 -Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface  
 -Improvement\_Areas\_Customer\_service\_responsiveness  
 -Improvement\_Areas\_Other  
 -Improvement\_Areas\_Product\_quality\_and\_accuracy  
 , data = Amazon[training, ])  
  
  
Amazon.reg.reduced.predictions <- predict(Amazon.reg.reduced,Amazon)[-training]  
(mean((Amazon.test.results-Amazon.reg.reduced.predictions)^2))^0.5

## [1] 1.979471

RMSE of linear regression model with removed variables = 1.979471

summary(Amazon.reg.reduced)

##   
## Call:  
## lm(formula = Purchase\_Frequency\_Num ~ . - Gender\_Male - Purchase\_Categories\_Beauty\_and\_Personal\_Care -   
## Personalized\_Recommendation\_Frequency\_No - Browsing\_Frequency\_Multiple\_times\_a\_day -   
## Product\_Search\_Method\_categories - Search\_Result\_Exploration\_Multiple\_pages -   
## Add\_to\_Cart\_Browsing\_Yes - Cart\_Completion\_Frequency\_Never -   
## Cart\_Abandonment\_Factors\_Found\_a\_better\_price\_elsewhere -   
## Saveforlater\_Frequency\_Never - Review\_Left\_Yes - Review\_Reliability\_Never -   
## Review\_Helpfulness\_Yes - Recommendation\_Helpfulness\_No -   
## Service\_Appreciation\_Wide\_product\_selection - Improvement\_Areas\_Reducing\_packaging\_waste -   
## age - Customer\_Reviews\_Importance - Personalized\_Recommendation\_Frequency.1 -   
## Rating\_Accuracy - Shopping\_Satisfaction - Gender\_Others -   
## Personalized\_Recommendation\_Frequency\_Sometimes - Browsing\_Frequency\_Few\_times\_a\_month -   
## Browsing\_Frequency\_Few\_times\_a\_week - Browsing\_Frequency\_Rarely -   
## Product\_Search\_Method\_Filter - Product\_Search\_Method\_Keyword -   
## Product\_Search\_Method\_others - Search\_Result\_Exploration\_First\_page -   
## Add\_to\_Cart\_Browsing\_Maybe - Cart\_Completion\_Frequency\_Often -   
## Cart\_Completion\_Frequency\_Rarely - Cart\_Completion\_Frequency\_Sometimes -   
## Cart\_Abandonment\_Factors\_Changed\_my\_mind\_or\_no\_longer\_need\_the\_item -   
## Cart\_Abandonment\_Factors\_High\_shipping\_costs - Cart\_Abandonment\_Factors\_others -   
## Saveforlater\_Frequency\_Always - Saveforlater\_Frequency\_Often -   
## Saveforlater\_Frequency\_Rarely - Saveforlater\_Frequency\_Sometimes -   
## Review\_Reliability\_Heavily - Review\_Reliability\_Moderately -   
## Review\_Reliability\_Occasionally - Review\_Reliability\_Rarely -   
## Review\_Helpfulness\_No - Review\_Helpfulness\_Sometimes - Recommendation\_Helpfulness\_Sometimes -   
## Recommendation\_Helpfulness\_Yes - Service\_Appreciation\_Competitive\_prices -   
## Service\_Appreciation\_Other - Service\_Appreciation\_Product\_recommendations -   
## Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface -   
## Improvement\_Areas\_Customer\_service\_responsiveness - Improvement\_Areas\_Other -   
## Improvement\_Areas\_Product\_quality\_and\_accuracy, data = Amazon[training,   
## ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6896 -1.3067 -0.2335 1.2064 4.8186   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 3.2150 0.2480 12.964  
## Gender\_Female 0.3823 0.2005 1.906  
## Purchase\_Categories\_Clothing\_and\_Fashion -0.6110 0.2407 -2.539  
## Purchase\_Categories\_Groceries\_and\_Gourmet\_Food -0.3585 0.2713 -1.321  
## Purchase\_Categories\_Home\_and\_Kitchen -1.8355 0.4313 -4.256  
## Purchase\_Categories\_others -1.7667 0.3733 -4.732  
## Personalized\_Recommendation\_Frequency\_Yes 0.5924 0.2377 2.492  
## Add\_to\_Cart\_Browsing\_No 0.8233 0.2387 3.449  
## Cart\_Completion\_Frequency\_Always 0.9689 0.3726 2.601  
## Review\_Left\_No -0.8048 0.1967 -4.091  
## Improvement\_Areas\_Shipping\_speed\_and\_reliability -0.9858 0.2753 -3.580  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## Gender\_Female 0.057426 .   
## Purchase\_Categories\_Clothing\_and\_Fashion 0.011555 \*   
## Purchase\_Categories\_Groceries\_and\_Gourmet\_Food 0.187339   
## Purchase\_Categories\_Home\_and\_Kitchen 2.68e-05 \*\*\*  
## Purchase\_Categories\_others 3.24e-06 \*\*\*  
## Personalized\_Recommendation\_Frequency\_Yes 0.013169 \*   
## Add\_to\_Cart\_Browsing\_No 0.000631 \*\*\*  
## Cart\_Completion\_Frequency\_Always 0.009703 \*\*   
## Review\_Left\_No 5.34e-05 \*\*\*  
## Improvement\_Areas\_Shipping\_speed\_and\_reliability 0.000392 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.791 on 348 degrees of freedom  
## Multiple R-squared: 0.2434, Adjusted R-squared: 0.2217   
## F-statistic: 11.2 on 10 and 348 DF, p-value: < 2.2e-16

# New model with insignificant variables removed (Iteration 2)  
  
Amazon.reg.reduced2 <- lm(Purchase\_Frequency\_Num ~ .   
 -Gender\_Male   
 -Purchase\_Categories\_Beauty\_and\_Personal\_Care   
 -Personalized\_Recommendation\_Frequency\_No  
 -Browsing\_Frequency\_Multiple\_times\_a\_day  
 -Product\_Search\_Method\_categories  
 -Search\_Result\_Exploration\_Multiple\_pages  
 -Add\_to\_Cart\_Browsing\_Yes  
 -Cart\_Completion\_Frequency\_Never  
 -Cart\_Abandonment\_Factors\_Found\_a\_better\_price\_elsewhere  
 -Saveforlater\_Frequency\_Never  
 -Review\_Left\_Yes  
 -Review\_Reliability\_Never  
 -Review\_Helpfulness\_Yes  
 -Recommendation\_Helpfulness\_No  
 -Service\_Appreciation\_Wide\_product\_selection  
 -Improvement\_Areas\_Reducing\_packaging\_waste  
 -age  
 -Customer\_Reviews\_Importance  
 -Personalized\_Recommendation\_Frequency.1  
 -Rating\_Accuracy  
 -Shopping\_Satisfaction  
 -Gender\_Others  
 -Personalized\_Recommendation\_Frequency\_Sometimes  
 -Browsing\_Frequency\_Few\_times\_a\_month  
 -Browsing\_Frequency\_Few\_times\_a\_week  
 -Browsing\_Frequency\_Rarely  
 -Product\_Search\_Method\_Filter  
 -Product\_Search\_Method\_Keyword  
 -Product\_Search\_Method\_others  
 -Search\_Result\_Exploration\_First\_page  
 -Add\_to\_Cart\_Browsing\_Maybe  
 -Cart\_Completion\_Frequency\_Often  
 -Cart\_Completion\_Frequency\_Rarely  
 -Cart\_Completion\_Frequency\_Sometimes  
 -Cart\_Abandonment\_Factors\_Changed\_my\_mind\_or\_no\_longer\_need\_the\_item  
 -Cart\_Abandonment\_Factors\_High\_shipping\_costs  
 -Cart\_Abandonment\_Factors\_others  
 -Saveforlater\_Frequency\_Always  
 -Saveforlater\_Frequency\_Often  
 -Saveforlater\_Frequency\_Rarely  
 -Saveforlater\_Frequency\_Sometimes  
 -Review\_Reliability\_Heavily  
 -Review\_Reliability\_Moderately  
 -Review\_Reliability\_Occasionally  
 -Review\_Reliability\_Rarely  
 -Review\_Helpfulness\_No  
 -Review\_Helpfulness\_Sometimes  
 -Recommendation\_Helpfulness\_Sometimes  
 -Recommendation\_Helpfulness\_Yes  
 -Service\_Appreciation\_Competitive\_prices  
 -Service\_Appreciation\_Other  
 -Service\_Appreciation\_Product\_recommendations  
 -Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface  
 -Improvement\_Areas\_Customer\_service\_responsiveness  
 -Improvement\_Areas\_Other  
 -Improvement\_Areas\_Product\_quality\_and\_accuracy  
 -Gender\_Female  
 -Purchase\_Categories\_Groceries\_and\_Gourmet\_Food  
 , data = Amazon[training, ])  
  
  
Amazon.reg.reduced2.predictions <- predict(Amazon.reg.reduced2,Amazon)[-training]  
(mean((Amazon.test.results-Amazon.reg.reduced2.predictions)^2))^0.5

## [1] 1.980209

RMSE of linear regression model with removed variables (second iteration) = 1.980209

This is higher than previously so it will not be used.

Therefore, Linear Regression Model 2 is the best

summary(Amazon.reg.reduced2)

##   
## Call:  
## lm(formula = Purchase\_Frequency\_Num ~ . - Gender\_Male - Purchase\_Categories\_Beauty\_and\_Personal\_Care -   
## Personalized\_Recommendation\_Frequency\_No - Browsing\_Frequency\_Multiple\_times\_a\_day -   
## Product\_Search\_Method\_categories - Search\_Result\_Exploration\_Multiple\_pages -   
## Add\_to\_Cart\_Browsing\_Yes - Cart\_Completion\_Frequency\_Never -   
## Cart\_Abandonment\_Factors\_Found\_a\_better\_price\_elsewhere -   
## Saveforlater\_Frequency\_Never - Review\_Left\_Yes - Review\_Reliability\_Never -   
## Review\_Helpfulness\_Yes - Recommendation\_Helpfulness\_No -   
## Service\_Appreciation\_Wide\_product\_selection - Improvement\_Areas\_Reducing\_packaging\_waste -   
## age - Customer\_Reviews\_Importance - Personalized\_Recommendation\_Frequency.1 -   
## Rating\_Accuracy - Shopping\_Satisfaction - Gender\_Others -   
## Personalized\_Recommendation\_Frequency\_Sometimes - Browsing\_Frequency\_Few\_times\_a\_month -   
## Browsing\_Frequency\_Few\_times\_a\_week - Browsing\_Frequency\_Rarely -   
## Product\_Search\_Method\_Filter - Product\_Search\_Method\_Keyword -   
## Product\_Search\_Method\_others - Search\_Result\_Exploration\_First\_page -   
## Add\_to\_Cart\_Browsing\_Maybe - Cart\_Completion\_Frequency\_Often -   
## Cart\_Completion\_Frequency\_Rarely - Cart\_Completion\_Frequency\_Sometimes -   
## Cart\_Abandonment\_Factors\_Changed\_my\_mind\_or\_no\_longer\_need\_the\_item -   
## Cart\_Abandonment\_Factors\_High\_shipping\_costs - Cart\_Abandonment\_Factors\_others -   
## Saveforlater\_Frequency\_Always - Saveforlater\_Frequency\_Often -   
## Saveforlater\_Frequency\_Rarely - Saveforlater\_Frequency\_Sometimes -   
## Review\_Reliability\_Heavily - Review\_Reliability\_Moderately -   
## Review\_Reliability\_Occasionally - Review\_Reliability\_Rarely -   
## Review\_Helpfulness\_No - Review\_Helpfulness\_Sometimes - Recommendation\_Helpfulness\_Sometimes -   
## Recommendation\_Helpfulness\_Yes - Service\_Appreciation\_Competitive\_prices -   
## Service\_Appreciation\_Other - Service\_Appreciation\_Product\_recommendations -   
## Service\_Appreciation\_User\_friendly\_website\_or\_app\_interface -   
## Improvement\_Areas\_Customer\_service\_responsiveness - Improvement\_Areas\_Other -   
## Improvement\_Areas\_Product\_quality\_and\_accuracy - Gender\_Female -   
## Purchase\_Categories\_Groceries\_and\_Gourmet\_Food, data = Amazon[training,   
## ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.4257 -1.3259 -0.3247 1.2876 4.9841   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 3.3642 0.1782 18.878  
## Purchase\_Categories\_Clothing\_and\_Fashion -0.5325 0.2248 -2.369  
## Purchase\_Categories\_Home\_and\_Kitchen -1.7294 0.4265 -4.055  
## Purchase\_Categories\_others -1.7640 0.3577 -4.932  
## Personalized\_Recommendation\_Frequency\_Yes 0.5615 0.2387 2.352  
## Add\_to\_Cart\_Browsing\_No 0.7763 0.2384 3.256  
## Cart\_Completion\_Frequency\_Always 0.9741 0.3750 2.598  
## Review\_Left\_No -0.8159 0.1975 -4.131  
## Improvement\_Areas\_Shipping\_speed\_and\_reliability -1.0023 0.2760 -3.632  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## Purchase\_Categories\_Clothing\_and\_Fashion 0.018402 \*   
## Purchase\_Categories\_Home\_and\_Kitchen 6.19e-05 \*\*\*  
## Purchase\_Categories\_others 1.26e-06 \*\*\*  
## Personalized\_Recommendation\_Frequency\_Yes 0.019234 \*   
## Add\_to\_Cart\_Browsing\_No 0.001240 \*\*   
## Cart\_Completion\_Frequency\_Always 0.009781 \*\*   
## Review\_Left\_No 4.52e-05 \*\*\*  
## Improvement\_Areas\_Shipping\_speed\_and\_reliability 0.000323 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.803 on 350 degrees of freedom  
## Multiple R-squared: 0.2288, Adjusted R-squared: 0.2111   
## F-statistic: 12.98 on 8 and 350 DF, p-value: < 2.2e-16

KNN

# Find optimal value of k  
best.k <- -1  
RMSE <- -1  
best.RMSE <- 99999999  
for (i in seq(from=1, to=100, by=2)) {  
 set.seed(12345)  
 Amazon.knn <- knn.reg(Amazon.training, Amazon.test, Amazon.training.results, k=i)  
 RMSE <- (mean((Amazon.knn$pred - Amazon.test.results)^2))^0.5  
 if (RMSE < best.RMSE) {  
 best.k <- i  
 best.RMSE <- RMSE  
 }  
}  
print(paste("The optimal value of k is",best.k,"with a RMSE of",best.RMSE))

## [1] "The optimal value of k is 25 with a RMSE of 1.95250309432106"

Amazon.knn <- knn.reg(Amazon.training, Amazon.test, Amazon.training.results, k=25)  
(mean((Amazon.knn$pred - Amazon.test.results)^2))^0.5

## [1] 1.952503

Regression Tree:

best.mindev <- -1  
RMSE <- -1  
best.RMSE <- 99999999  
for (i in seq(from=0.0005, to=0.05, by=0.0005)) {  
 Amazon.tree <- tree(Purchase\_Frequency\_Num ~ ., data=Amazon[training,], mindev=i)  
 Amazon.tree.predictions <- predict(Amazon.tree,Amazon)[-training]  
 RMSE <- (mean((Amazon.test.results-Amazon.tree.predictions)^2))^0.5  
 if (RMSE < best.RMSE) {  
 best.mindev <- i  
 best.RMSE <- RMSE  
 }  
}  
print(paste("The optimal value of mindev is",best.mindev,"with a RMSE of",best.RMSE))

## [1] "The optimal value of mindev is 0.045 with a RMSE of 2.03884342739826"

Amazon.tree <- tree(Purchase\_Frequency\_Num ~., data=Amazon[training,],mindev=0.045)  
Amazon.tree.predictions <- predict(Amazon.tree,Amazon)[-training]  
(mean((Amazon.test.results-Amazon.tree.predictions)^2))^0.5

## [1] 2.038843

plot(Amazon.tree)  
text(Amazon.tree, cex=0.6)

