Exploratory2

Harshita Cherukuri

November 24, 2018

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

train <- read\_csv("train.csv",col\_types=list(  
 Store = col\_integer(),  
 DayOfWeek= col\_integer(),  
 Date = col\_date(),  
 Customers = col\_integer(),  
 Open = col\_integer(),  
 Promo = col\_integer(),  
 StateHoliday = col\_character(),  
 SchoolHoliday = col\_integer()))  
  
test <- read\_csv("test.csv",col\_types=list(  
 Store = col\_integer(),  
 DayOfWeek= col\_integer(),  
 Date = col\_date(),  
 Sales = col\_integer(),  
 Customers = col\_integer(),  
 Open = col\_integer(),  
 Promo = col\_integer(),  
 StateHoliday = col\_character(),  
 SchoolHoliday = col\_integer()))

## Warning: The following named parsers don't match the column names: Sales,  
## Customers

store <- read\_csv("store.csv",col\_types=list(  
 Store = col\_integer(),  
 StoreType= col\_character(),  
 Assortment = col\_character(),  
 CompetitionDistance = col\_integer(),  
 CompetitionOpenSinceMonth = col\_integer(),  
 CompetitionOpenSinceYear = col\_integer(),  
 Promo2 = col\_integer(),  
 Promo2SinceYear = col\_character(),  
 PromoInterval = col\_character()))  
  
colnames(train)

## [1] "Store" "DayOfWeek" "Date" "Sales"   
## [5] "Customers" "Open" "Promo" "StateHoliday"   
## [9] "SchoolHoliday"

colnames(test)

## [1] "Id" "Store" "DayOfWeek" "Date"   
## [5] "Open" "Promo" "StateHoliday" "SchoolHoliday"

colnames(store)

## [1] "Store" "StoreType"   
## [3] "Assortment" "CompetitionDistance"   
## [5] "CompetitionOpenSinceMonth" "CompetitionOpenSinceYear"   
## [7] "Promo2" "Promo2SinceWeek"   
## [9] "Promo2SinceYear" "PromoInterval"

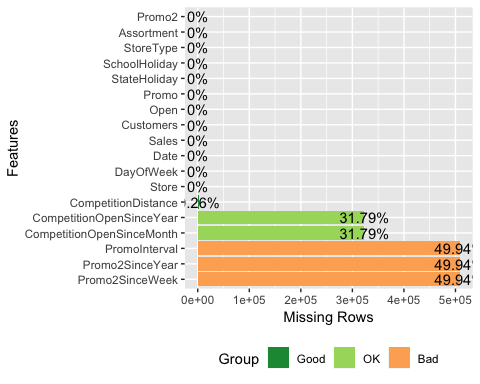
mergedTrain <- merge(train,store)  
mergedTest <- merge(test,store)

## Including Plots

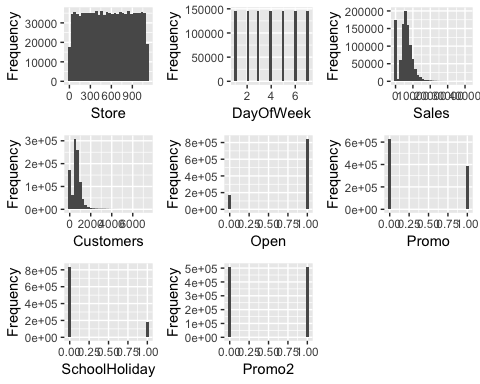
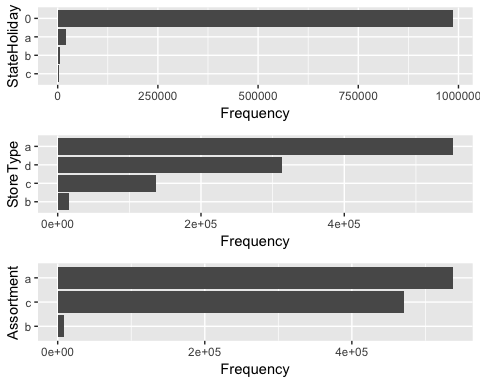
You can also embed plots, for example:



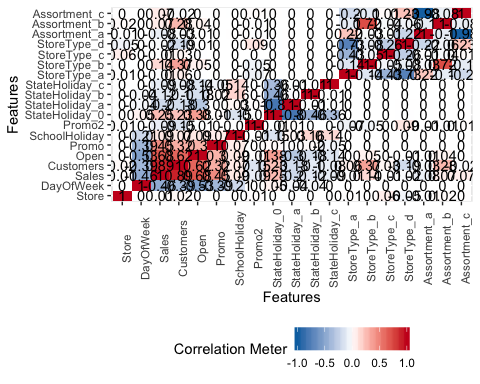
## Store DayOfWeek   
## 0 0   
## Date Sales   
## 0 0   
## Customers Open   
## 0 0   
## Promo StateHoliday   
## 0 0   
## SchoolHoliday StoreType   
## 0 0   
## Assortment CompetitionDistance   
## 0 2642   
## CompetitionOpenSinceMonth CompetitionOpenSinceYear   
## 323348 323348   
## Promo2 Promo2SinceWeek   
## 0 508031   
## Promo2SinceYear PromoInterval   
## 508031 508031



## 1 columns ignored with more than 50 categories.  
## Date: 942 categories



## 1 features with more than 20 categories ignored!  
## Date: 942 categories



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.