HYATT – TO GO OR NOT TO GO?

IST687 – APPLIED DATA SCIENCE

ABSTRACT

In this document, we describe our analysis of the customer feedback data of the Hyatt group of hotels. We describe the techniques used, implementation methodologies and finally, predicted the factors that contribute towards the net promoter score.

TEAM MEMBERS

ABHIMANYU TOMAR

AKANSHA SINGH

ARPITHA SRIDHARA

KRITHIKA KEDALASARA SOPPINA

SHIVALI NAIK

WAMSI VISWANATH

Table of Contents

Introduction

Customer feedback is a critical indicator for the popularity of hotels, eventually influencing the revenue of hotels via word-of-mouth. **Customer feedback is so important** because it provides marketers and business owners with insight that they can use to improve their business, products and/or overall customer experience.

What we did

- We started by creating a dataset - to perform data analysis
- Random sampling – We combined a sample of 50,000 rows from each month (total – 12 months) and used the resulting dataset for analysis
  - Why? A random sample that spans across the whole year might give us a more accurate value of NPS
- Also, 3 month period (April, May, June)
  - Large sample of consistent data
- Number of Detractors, Promoters and Passives in our dataset
  - We have 11.66% detractors – Why? Let's see.

Business Questions

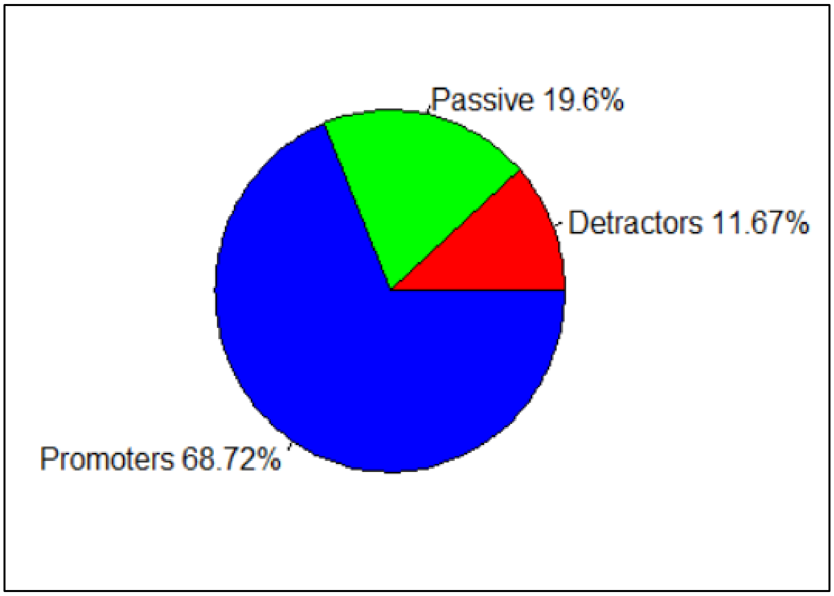
We answered following business questions on the analysis of the dataset.

- Which attributes influence NPS type the most?
- Which lead to the biggest positive/negative impact on NPS?
- Which country to focus on that will affect the NPS score?
- Which country/state should the client consider on improving the NPS score?
- What purpose of visit for the customer has a great impact on the NPS score?

Calculating number of promoters, detractors and passives in the dataset

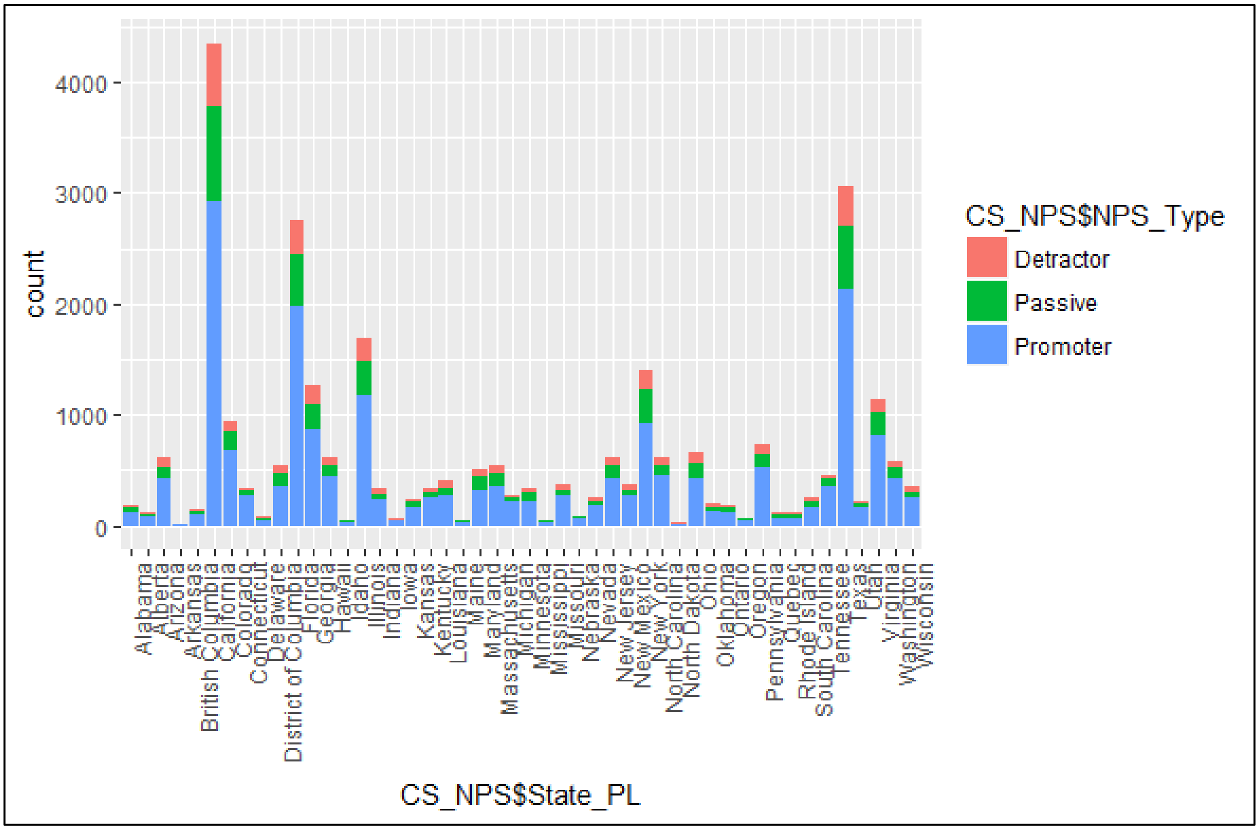
```
#Calculating number of detractors, promoters, passives


dataset = read.csv("C://Users//ARPAM//Desktop//R//Project//RData.csv",header = TRUE,  sep=",",na.strings = c("", "NA"))

#dataset$Likelihood_Recommend_H <-as.numeric(surveydatanew6$Likelihood_Recommend_H)
surveydatanewD<- subset(dataset,Likelihood_Recommend_H < 6 | Likelihood_Recommend_H == 6 )
surveydatanewPass <- subset(dataset, Likelihood_Recommend_H == 7 | Likelihood_Recommend_H == 8)
surveydatanewP<- subset(dataset,Likelihood_Recommend_H > 8)
sumD<- (sum(table(surveydatanewD$Likelihood_Recommend_H)))
sumPass <- (sum(table(surveydatanewPass$Likelihood_Recommend_H)))
sumP <- (sum(table(surveydatanewP$Likelihood_Recommend_H)))

#Pie Chart
lbls1 = c("Detractors", "Passive", "Promoters")
slices1 <- c(sumD, sumPass,sumP)
pct1 <- round(slices1/sum(slices1)*100,digits=2)
lbls1 <- paste(lbls1, pct1) # add percents to labels
lbls1 <- paste(lbls1,"%",sep="") # ad % to labels
pie(slices1,labels=lbls1,col= rainbow(length(lbls1)))
```

Number of Detractors, Promoters and Passives in our dataset



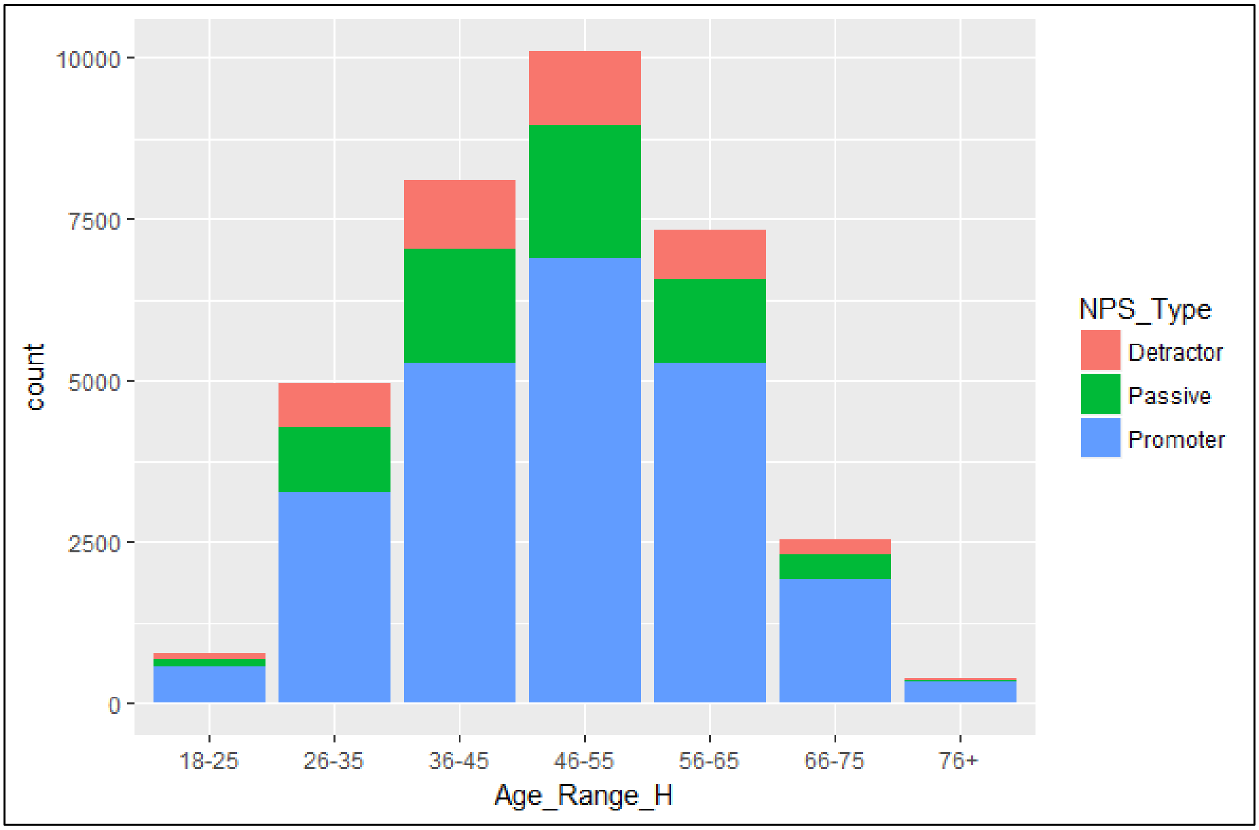
Promoter and Detractor State Wise

```
#Promoter, Detractor State Wise
CleanSubset <- dataset[-which(is.na(dataset$State_PL)), ]
CS_NPS <- CleanSubset[-which(is.na(CleanSubset$NPS_Type)),]
#Summary of the data statewise:
g2 <- ggplot(data=CS_NPS, aes(fill= CS_NPS$NPS_Type, x=CS_NPS$State_PL)) + geom_bar()
g2 <- g2 + theme(axis.text.x =element_text(angle = 90, hjust = 1))
g2
```



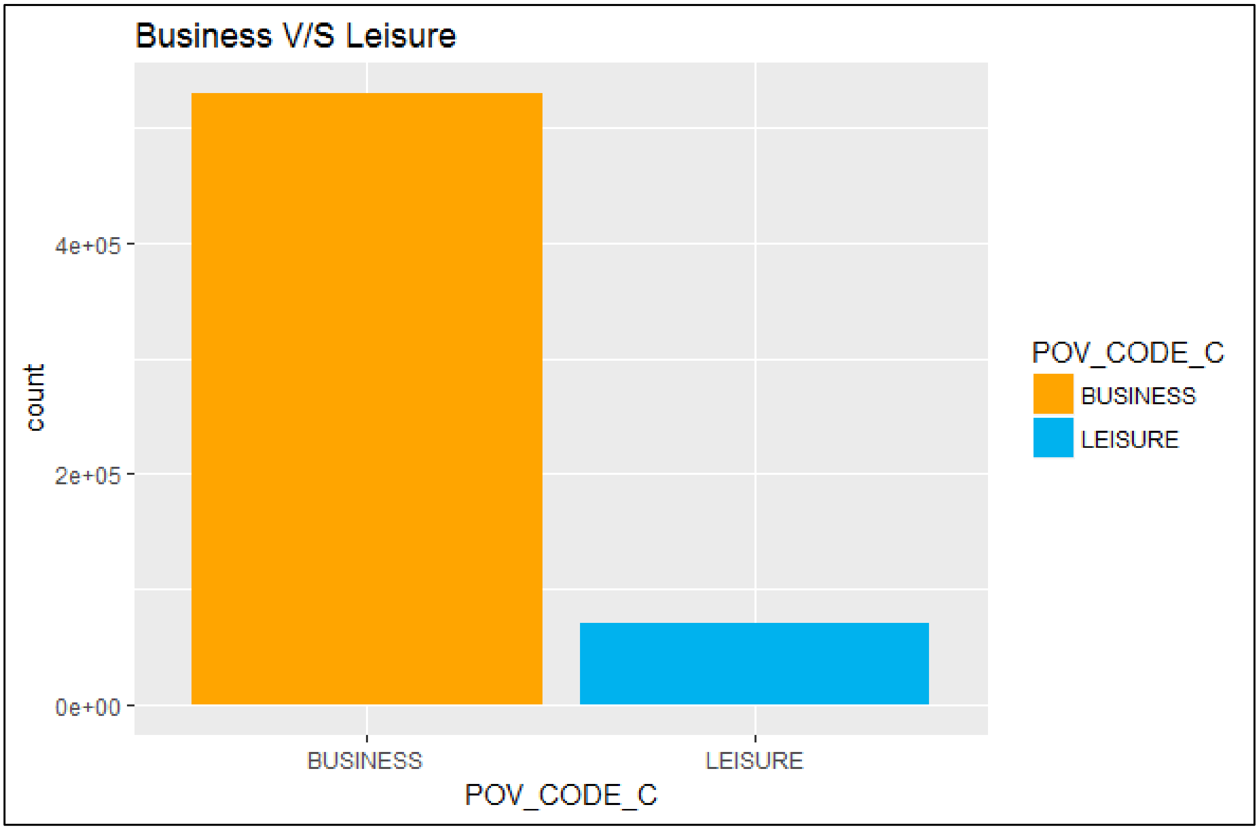
Summary of Data Age Wise

```
#Summary of data age wise
CleanSubset1 <- dataset[-which(is.na(dataset$Age_Range_H)), ]
CS2 <- CleanSubset1[-which(is.na(CleanSubset1$NPS_Type)), ]
Plot1 <- ggplot(data=CS2,aes(x=Age_Range_H,fill=NPS_Type))+geom_bar()
Plot1
```



Plotting the Data for Business or Leisure

```
#Plotting the data for Business or Leisure
g <- ggplot(data= dataset, aes(x= POV_CODE_C,fill=POV_CODE_C))
g <- g + geom_bar() + labs(title = "Business V/S Leisure")
g<- g + scale_fill_manual(values=c("orange","deepskyblue2","firebrick"))
g
```



Association Rules Mining

Code snippet

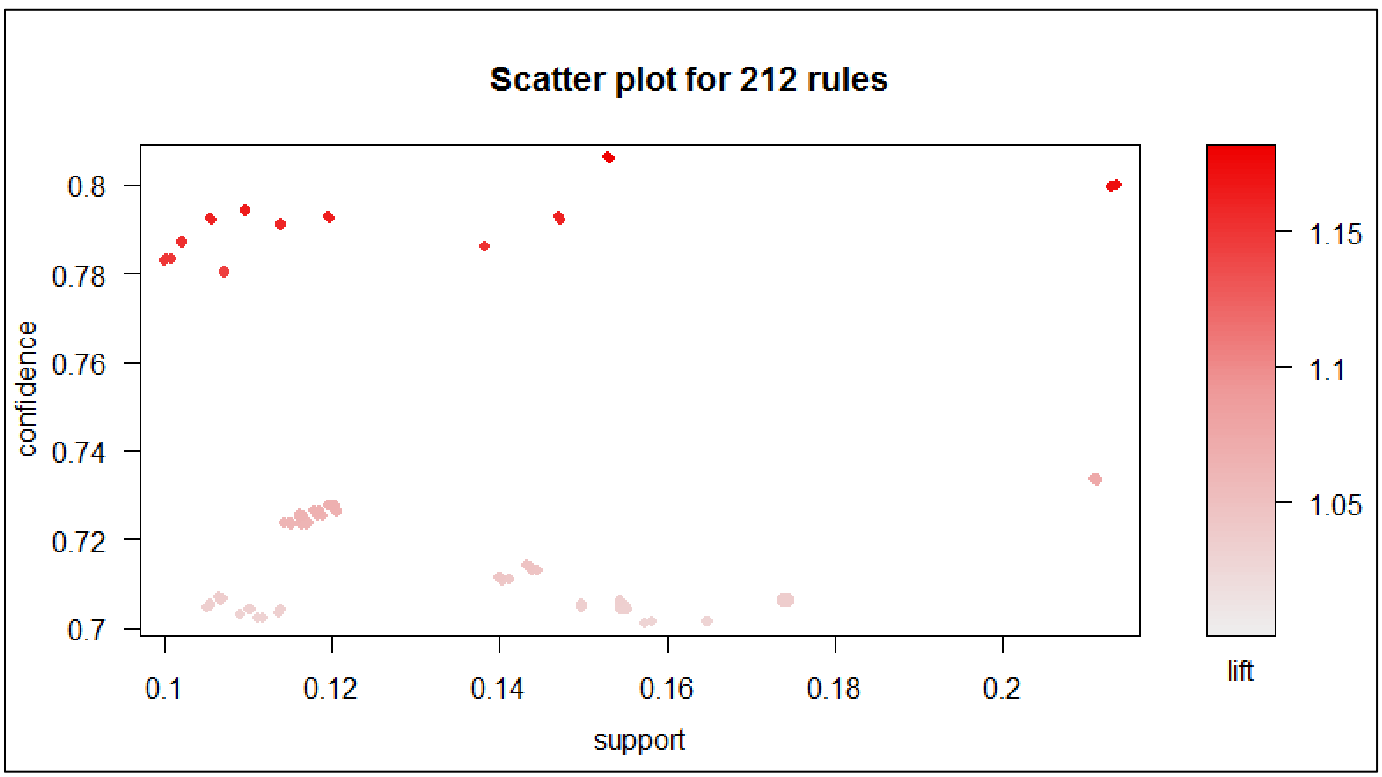
```
#Reading Dataset
dataset = read.csv("C://Users//ARPAM//Desktop//R//Project//RData1.csv",header = TRUE)

#Creating a subset of required columns
df <- dataset[,c("POV_CODE_C","ROOM_TYPE_CODE_R","NPS_Type","Spa_Used_H","Business.Center_PL","Mini.Bar_PL","Restaurant_PL","Laundry_PL","Internet_Sat_H","Shuttle.Service_PL",
                 "Valet.Parking_PL")]
dfBusiness <- subset(df,df$POV_CODE_C=="BUSINESS")
table(dfBusiness$POV_CODE_C)

#Cleaning SubSet
#Assigning NA to blank rows in NPS_Type so that it can be eliminated
dfBusiness$NPS_Type[dfBusiness$NPS_Type==""] <- NA
#CleanDF<- na.omit(dfBusiness)
CleanDF <- dfBusiness[-which(is.na(dfBusiness$NPS_Type)), ]

#Internet Satisfaction Cleaning
CleanDF$Internet_Sat_H[CleanDF$Internet_Sat_H==1 | CleanDF$Internet_Sat_H==2 | CleanDF$Internet_Sat_H==3 | CleanDF$Internet_Sat_H==4 | CleanDF$Internet_Sat_H==5 |
                         CleanDF$Internet_Sat_H==6] <- "Poor"
CleanDF$Internet_Sat_H[CleanDF$Internet_Sat_H==8 | CleanDF$Internet_Sat_H==7] <- "Good"
CleanDF$Internet_Sat_H[CleanDF$Internet_Sat_H==10 | CleanDF$Internet_Sat_H==9] <- "Excellent"
CleanDF$Internet_Sat_H <- as.factor(CleanDF$Internet_Sat_H)

# Association Rules Mining
library(arulesViz)
#Analyzing for NPS_Type=Promoter
rules.positive<-apriori(CleanDF, parameter= list(support=0.1, confidence=0.7),appearance=list(rhs='NPS_Type=Promoter',default='lhs'))
inspect(rules.positive)
plot(rules.positive)
sorted_rules_positive <- sort(rules.positive,decreasing=TRUE, by="lift")
inspect(sorted_rules_positive)

#Analyzing for NPS_Type=Detractor
rules.negative<-apriori(CleanDF, parameter= list(support=0.04, confidence=0.1),appearance=list(rhs='NPS_Type=Detractor',default='lhs'))
inspect(rules.negative)
summary(rules.negative)
sorted_rules_negative <- sort(rules.negative,decreasing=FALSE, by="lift")
inspect(sorted_rules_negative)
plot(sorted_rules_negative)
```

We performed association rules mining in order to observe the trends that contributed to detractors when the purpose of visit was for “Business”. Using this analysis, we wanted to know what Business men/women wanted most out of Hyatt. The parameters considered for this purpose were:

1. POV_CODE_C – “Business”, “Leisure”, “Combination of both”, “I don’t know”
2. NPS_Type – “Promoter”, “Detractor”,” Neutral”
3. Spa_Used_H – “No”, “Yes”
4. Business.Center_PL – “N”, “Y”
5. Mini.Bar_PL – “N”, “Y”
6. Restaurant_PL
7. Laundry_PL - “N”, “Y”
8. Internet_Sat_H – Scores out of 10; It was converted to Poor, Good, Excellent (**Discretization of continuous values**) based on the score given
9. Shuttle.Service_PL
10. Valet.Parking_PL

The rules were sorted in decreasing order of their lift values, so that the top rules could be evaluated. The support and confidence parameters were altered so that we got enough rules to inspect (approx. 150 to 250) but not too many, that would clutter the analysis.

The factors that contributed towards positive NPS score were:



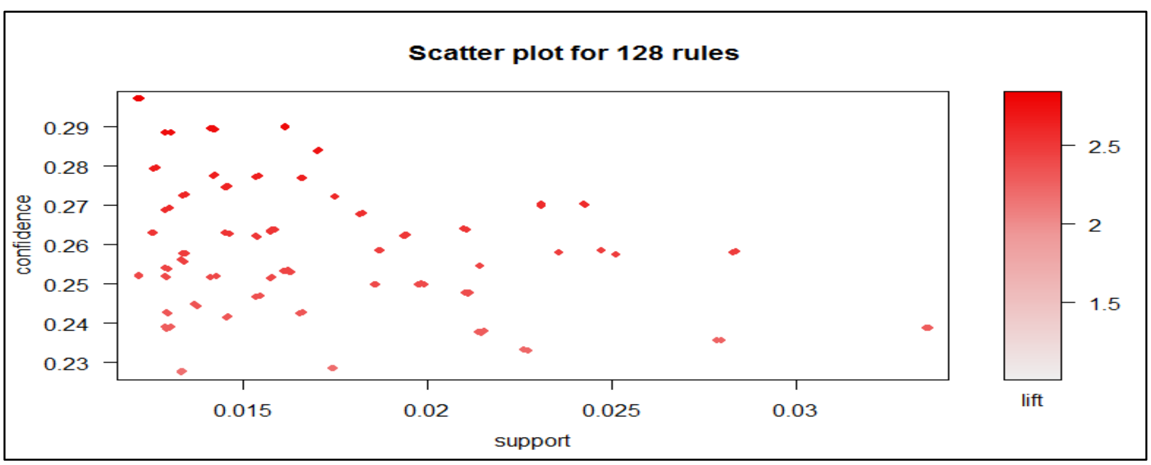
The top ten rules were as follows:

```
      lhs                                rhs                    support   confidence lift
[1]   {Spa_Used_H=,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1530290  0.8061863 1.180865
[2]   {POV_CODE_C=BUSINESS,
       Spa_Used_H=,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1530290  0.8061863 1.180865
[3]   {Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.2131003  0.7995670 1.171169
[4]   {POV_CODE_C=BUSINESS,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.2131003  0.7995670 1.171169
[5]   {Mini.Bar_PL=N,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1098252  0.7943054 1.163462
[6]   {POV_CODE_C=BUSINESS,
       Mini.Bar_PL=N,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1098252  0.7943054 1.163462
[7]   {Laundry_PL=Y,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1198032  0.7927240 1.161146
[8]   {POV_CODE_C=BUSINESS,
       Laundry_PL=Y,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1198032  0.7927240 1.161146
[9]   {Business.Center_PL=Y,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1473613  0.7923358 1.160577
[10]  {POV_CODE_C=BUSINESS,
       Business.Center_PL=Y,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1473613  0.7923358 1.160577
[11]  {Business.Center_PL=Y,
       Mini.Bar_PL=N,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1057865  0.7921220 1.160264
[12]  {POV_CODE_C=BUSINESS,
       Business.Center_PL=Y,
       Mini.Bar_PL=N,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1057865  0.7921220 1.160264
[13]  {Business.Center_PL=Y,
       Laundry_PL=Y,
       Internet_Sat_H=Excellent} => {NPS_Type=Promoter} 0.1135585  0.7908296 1.158371
```

We observed from the above rules that the availability of a Business Center, Laundry and Excellent High Speed Wi-Fi contributed towards a high NPS score from the Business people.

We wanted to look at the factors that gave rise to detractors, as we could provide recommendations to improve those.

Following were our observations:



The top rules were:

```
     lhs                          rhs                        support confidence     lift
[1]  {Spa_Used_H=,
      Business.Center_PL=Y,
      Mini.Bar_PL=N,
      Restaurant_PL=Y,
      Laundry_PL=Y,
      Internet_Sat_H=Poor}   => {NPS_Type=Detractor} 0.01215067  0.2970297 2.831530
[2]  {POV_CODE_C=BUSINESS,
      Spa_Used_H=,
      Business.Center_PL=Y,
      Mini.Bar_PL=N,
      Restaurant_PL=Y,
      Laundry_PL=Y,
      Internet_Sat_H=Poor}   => {NPS_Type=Detractor} 0.01215067  0.2970297 2.831530
[3]  {Business.Center_PL=Y,
      Mini.Bar_PL=N,
      Restaurant_PL=Y,
      Laundry_PL=Y,
      Internet_Sat_H=Poor}   => {NPS_Type=Detractor} 0.01620089  0.2898551 2.763136
[4]  {POV_CODE_C=BUSINESS,
      Business.Center_PL=Y,
      Mini.Bar_PL=N,
      Restaurant_PL=Y,
      Laundry_PL=Y,
      Internet_Sat_H=Poor}   => {NPS_Type=Detractor} 0.01620089  0.2898551 2.763136
[5]  {Business.Center_PL=Y,
      Mini.Bar_PL=N,
      Laundry_PL=Y,
      Internet_Sat_H=Poor,
      Valet.Parking_PL=Y}    => {NPS_Type=Detractor} 0.01417578  0.2892562 2.757427
[6]  {Business.Center_PL=Y,
      Mini.Bar_PL=N,
      Restaurant_PL=Y,
      Laundry_PL=Y,
      Internet_Sat_H=Poor,
      Valet.Parking_PL=Y}    => {NPS_Type=Detractor} 0.01417578  0.2892562 2.757427
```

```
[7]     {POV_CODE_C=BUSINESS,
         Business.Center_PL=Y,
         Mini.Bar_PL=N,
         Laundry_PL=Y,
         Internet_Sat_H=Poor,
         Valet.Parking_PL=Y}    => {NPS_Type=Detractor} 0.01417578  0.2892562 2.757427
[8]     {POV_CODE_C=BUSINESS,
         Business.Center_PL=Y,
         Mini.Bar_PL=N,
         Restaurant_PL=Y,
         Laundry_PL=Y,
         Internet_Sat_H=Poor,
         Valet.Parking_PL=Y}    => {NPS_Type=Detractor} 0.01417578  0.2892562 2.757427
[9]     {Spa_Used_H=,
         Business.Center_PL=Y,
         Mini.Bar_PL=N,
         Laundry_PL=Y,
         Internet_Sat_H=Poor}   => {NPS_Type=Detractor} 0.01296071  0.2882883 2.748200
[10]    {POV_CODE_C=BUSINESS,
         Spa_Used_H=,
         Business.Center_PL=Y,
         Mini.Bar_PL=N,
         Laundry_PL=Y,
         Internet_Sat_H=Poor}   => {NPS_Type=Detractor} 0.01296071  0.2882883 2.748200
```

An interesting rule pattern was also found for the shuttle service:

```
[33]  {Business.Center_PL=Y,
       Laundry_PL=Y,
       Internet_Sat_H=Poor,
       Shuttle.Service_PL=N} => {NPS_Type=Detractor} 0.01296071  0.2689076 2.563447
[34]  {POV_CODE_C=BUSINESS,
       Business.Center_PL=Y,
       Laundry_PL=Y,
       Internet_Sat_H=Poor,
       Shuttle.Service_PL=N} => {NPS_Type=Detractor} 0.01296071  0.2689076 2.563447
```

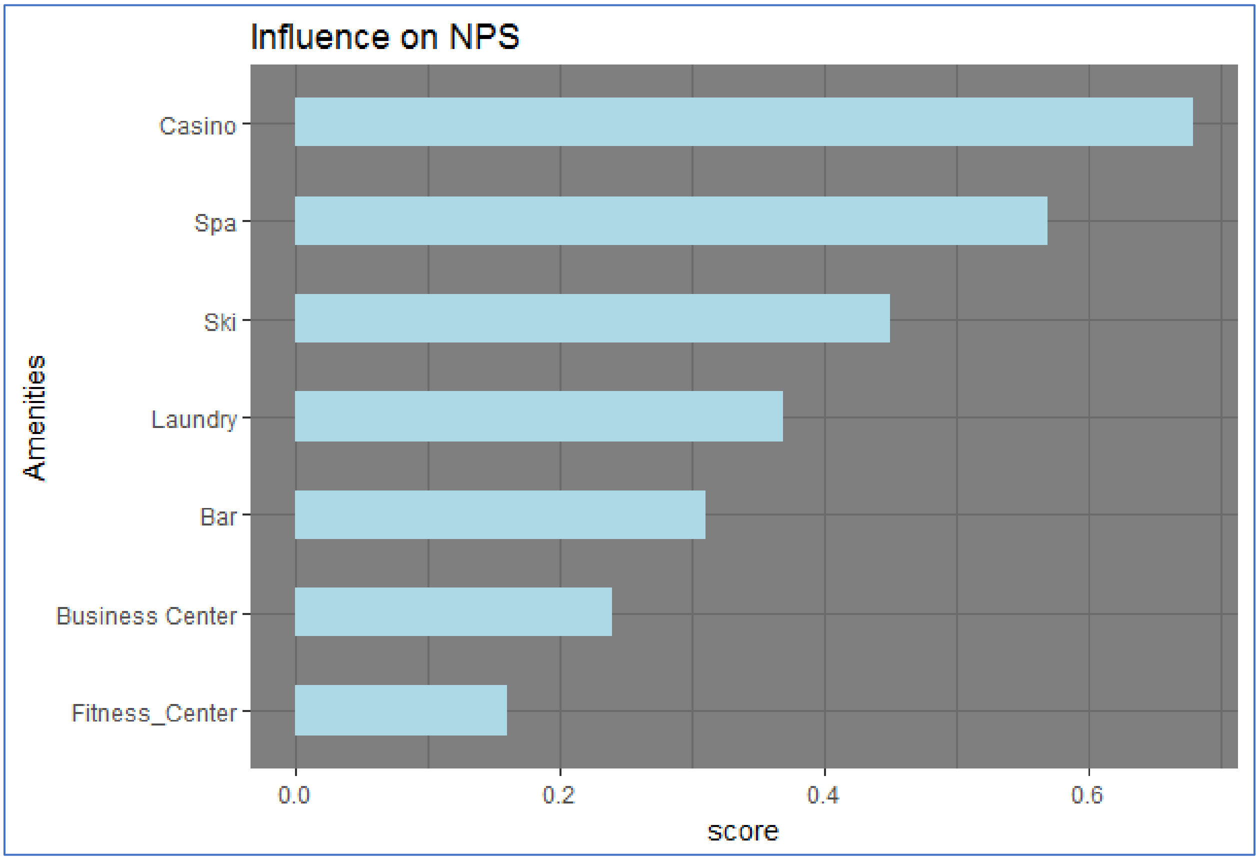
Based on these rules we can see three areas that have caused discomfort to the customers:

- Lack of mini-bar
- Lack of high speed Wi-Fi
- Lack of Shuttle Service

Our Recommendations

- Partner with nearby airport/ train stations to provide shuttle service for people who come to the hotel for a business visit
- Enable HIGH speed Wi-Fi in the hotel – Especially at the lobby and the business center
- A mini-bar can be made available for the guests either in their room or at the restaurant

NPS - Amenities Influencers



Most Influential factors:

- Casino
- Spa

Least Influential factor:

- Fitness Center
- Business Center

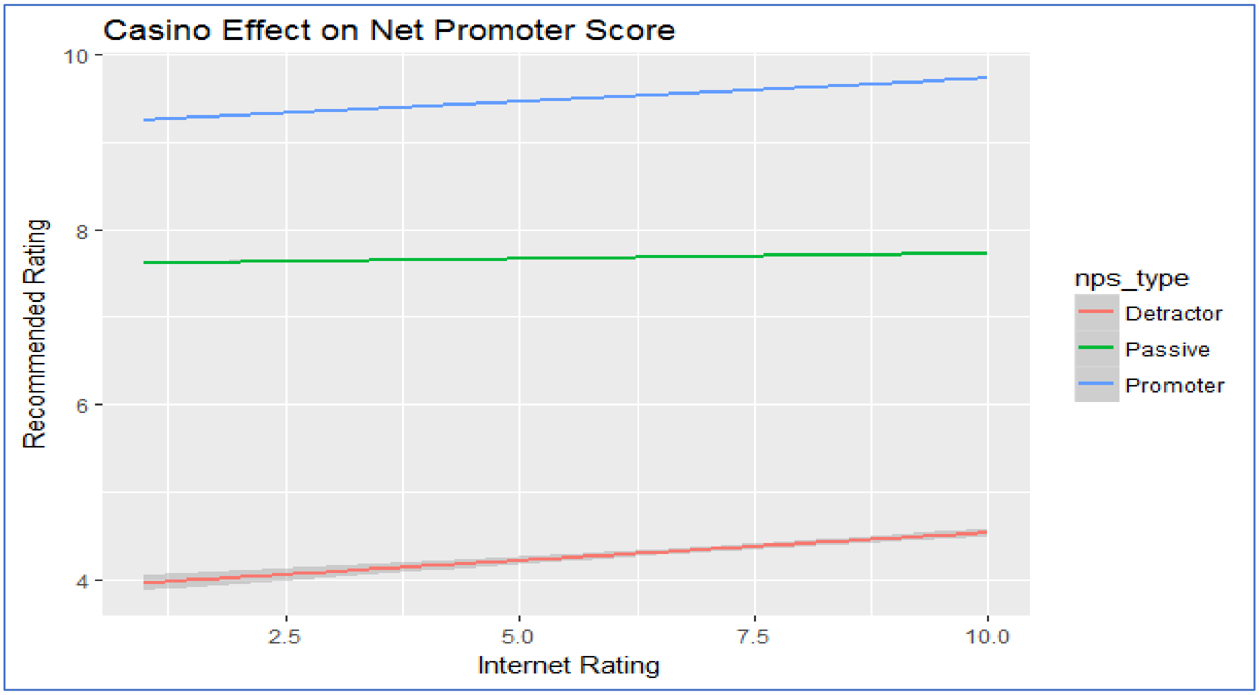
NPS - Services Influencers

Code snippet

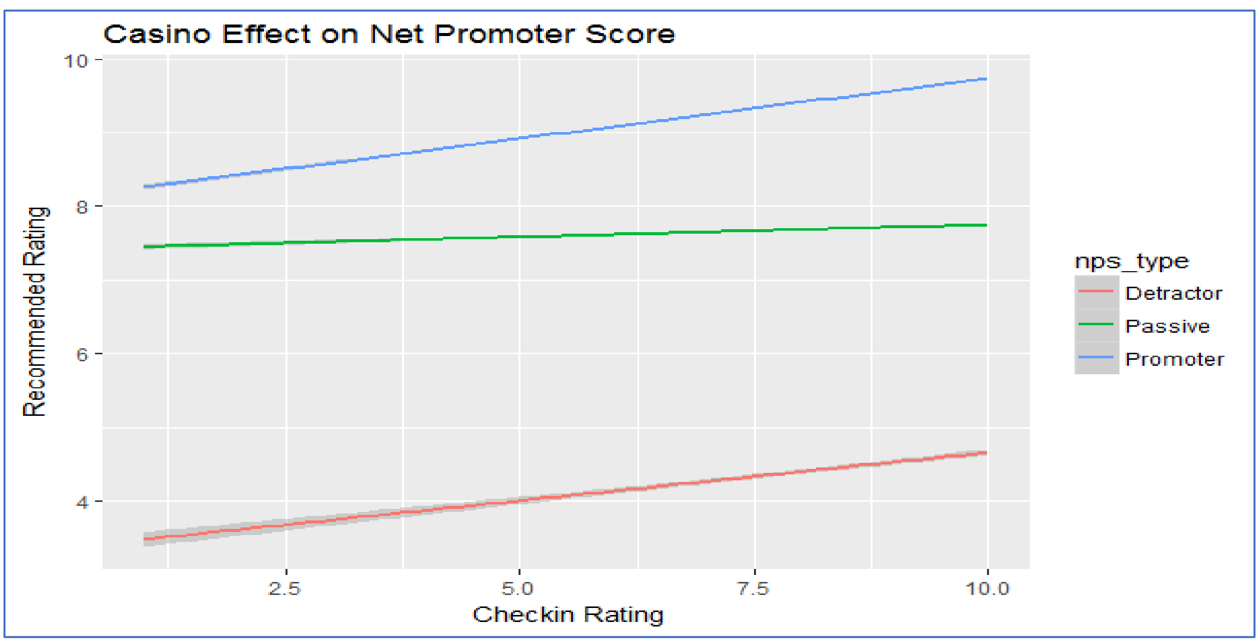
```
#plotting the linear regression
ggplot(df_quarter2, aes(x=internet_rating, y=recommend_rating, color=nps_type))  +
geom_smooth(method = "lm") + ylab("Recommended Rating") +   xlab(" Internet Rating") + ggtitle("Casino Effect on Net Promoter Score")

ggplot(df_quarter2, aes(x=checkin_rating, y=recommend_rating, color=nps_type))  +
geom_smooth(method = "lm") + ylab("Recommended Rating") +   xlab(" Checkin Rating") + ggtitle("Casino Effect on Net Promoter Score")
```

Most Influential factors:

- Internet Service



- Check-in rating



Our Recommendation:

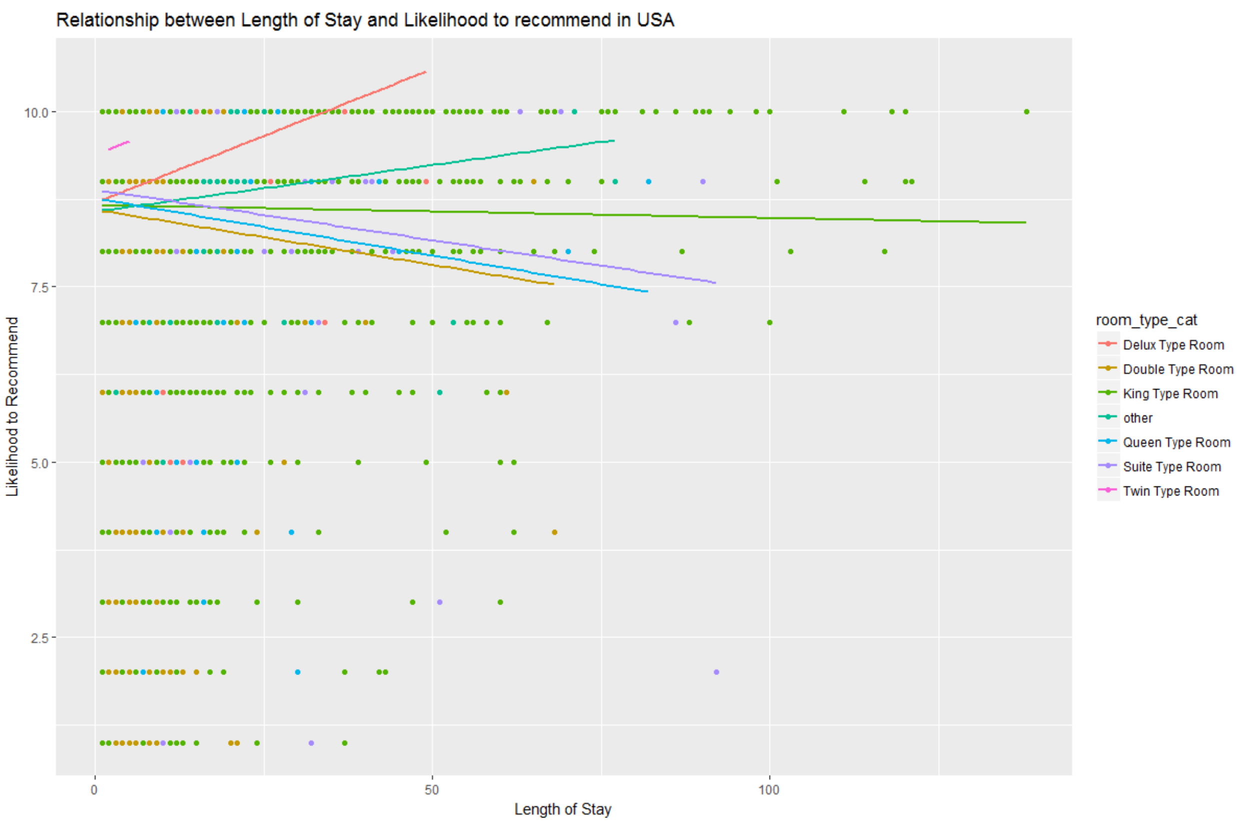
Better internet service and ease of check-in has positive impact on NPS score

Linear Modeling

Does the length of stay affect the likelihood to recommend?

We used the Linear Modelling, to analyze the relationship between length of stay and likelihood to recommend based on room type. At first, we tried to find out the relationship from all countries but we could not get the satisfying result. Then we narrow down our dataset to country US and 3 months (quarter 2) data from the whole dataset – April, May and June. We have concentrated on the surveys by the hotel such as Food and Beverages experience, room satisfaction and customer service satisfaction to provide support our analysis.

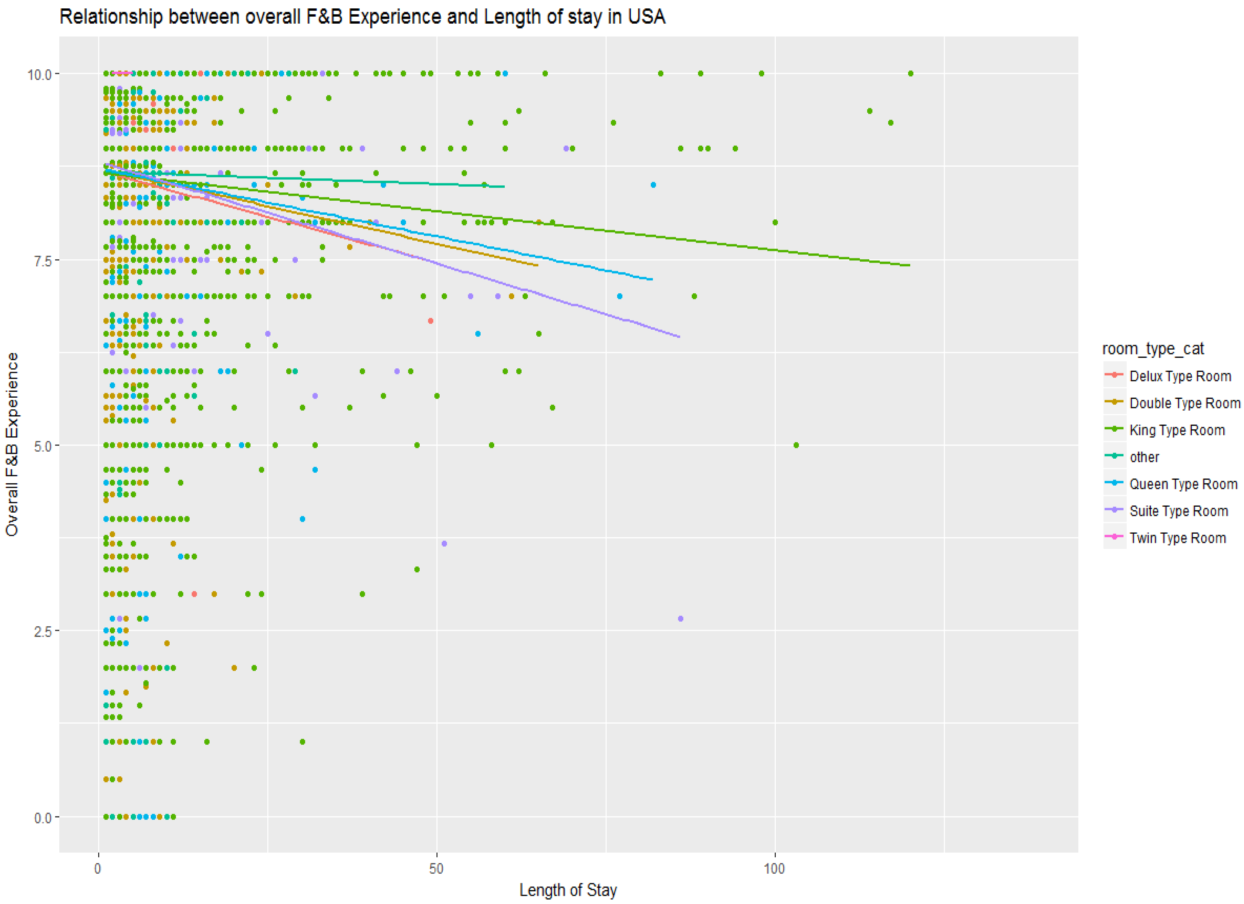
Relationship between Length of Stay & Likelihood to recommend



- Therefore, we try to see the relationship between Length of stay and rating in the F&B experience, categorized by the type of room
- On plotting the relation, we observe that for room types King, Double, Queen, Suite and Twin, the likelihood decreases as the length of stay increases

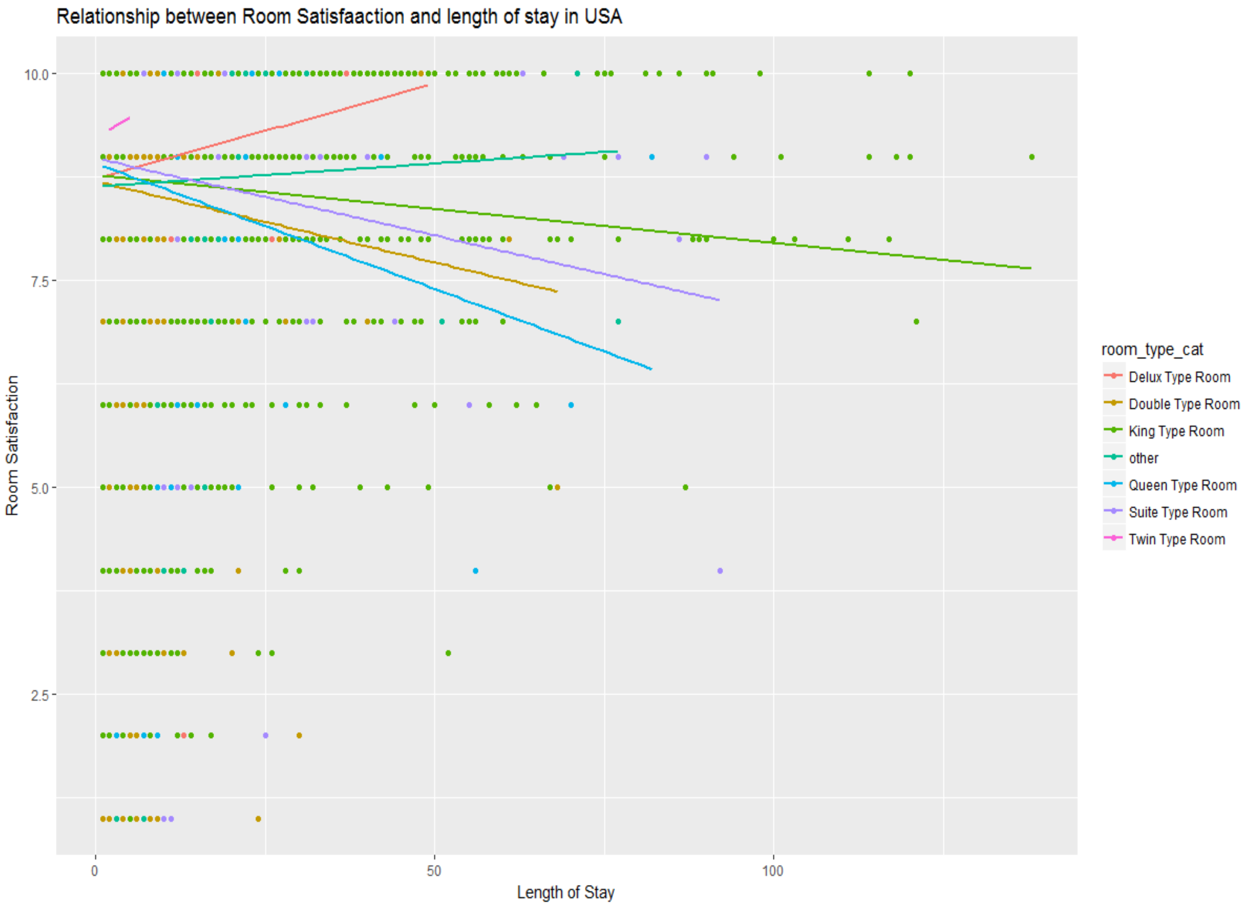
To support above analysis, we see what factors cause an unhappy guest when they stay for a longer duration

Relationship between Length of Stay & rating on Food & Beverages Experience



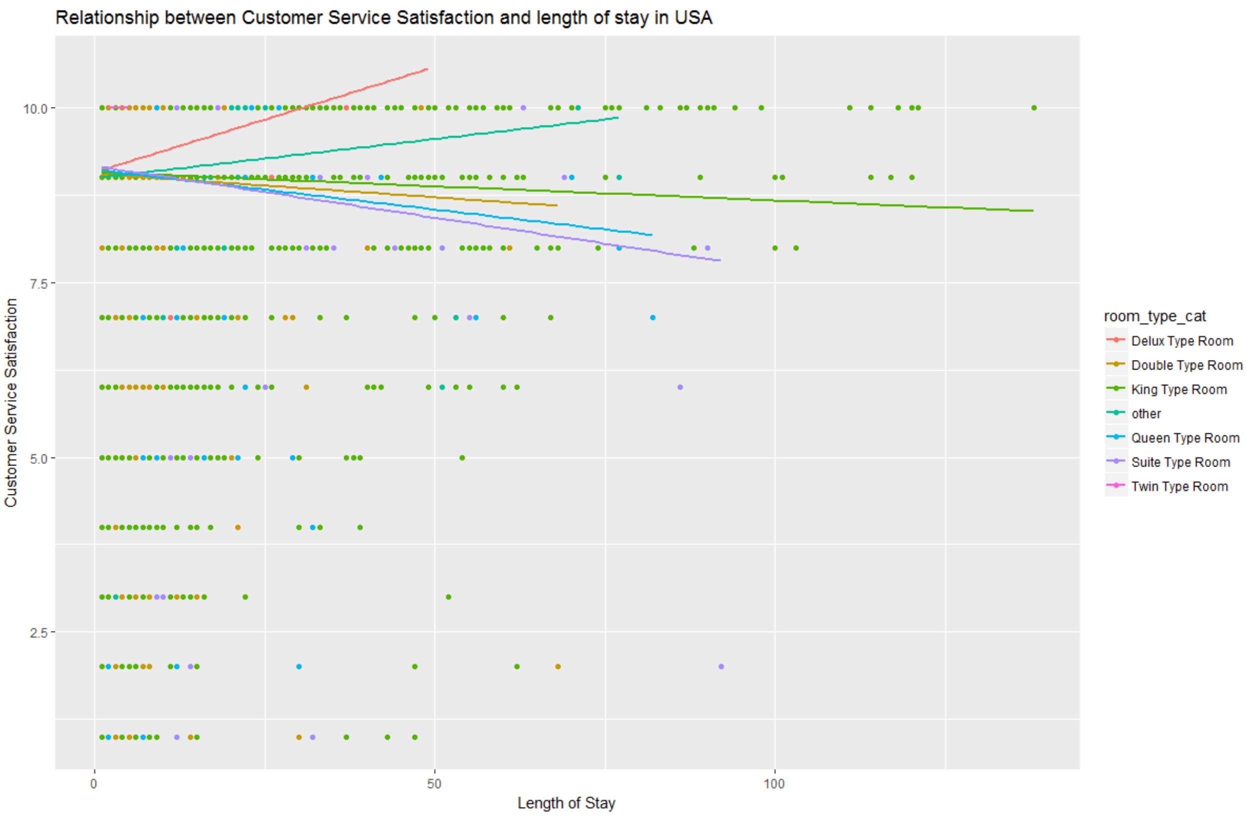
- Therefore, we try to see the relationship between Length of stay and rating in the F&B experience, categorized by the type of room
- On plotting the relation, we observe that for room types King, Double, Queen, Suite and Twin, the likelihood decreases as the length of stay increases

Relationship between Lengths of Stay & Room Satisfaction



- Next, we study the relationship between Length of stay and room satisfaction, categorized by the type of room
- On plotting the relation, we observe that for room types Double, Queen, Suite and King, the likelihood decreases as the length of stay increases
- In fact, Queen Type room has a steep decrease in room satisfaction with a long stay

Relationship between Lengths of Stay & Customer Service Satisfaction



- Next, we study the relationship between Length of stay and Customer Service satisfaction, categorized by the type of room
- On plotting the relation, we observe that for room types Double, Queen and Suite the likelihood decreases as the length of stay increases

Our Recommendations

- When guests stay for a loner duration, the food and beverages options must be increased and well spread across a wide range of days in order to keep the guests from getting bored of the menu
- Room maintenance must be monitored and intermittent feedback must be taken from the guests to keep track of their satisfaction

Code snippet

```
num_row <- nrow(hyatt_usa)
for (i in 1:num_row)
{

  if (grepl("king", hyatt_usa$ROOM_TYPE_DESCRIPTION_C[i], ignore.case = TRUE, perl = FALSE, fixed = FALSE, useBytes = FALSE) == TRUE)
  {
    hyatt_usa$room_type_cat[i] <- "King Type Room"
  }
  else if (grepl("Queen", hyatt_usa$ROOM_TYPE_DESCRIPTION_C[i], ignore.case = TRUE, perl = FALSE, fixed = FALSE, useBytes = FALSE) == TRUE)
  {
    hyatt_usa$room_type_cat[i] <- "Queen Type Room"
  }
  else if (grepl("Twin", hyatt_usa$ROOM_TYPE_DESCRIPTION_C[i], ignore.case = TRUE, perl = FALSE, fixed = FALSE, useBytes = FALSE) == TRUE)
  {
    hyatt_usa$room_type_cat[i] <- "Twin Type Room"
  }
  else if (grepl("Double", hyatt_usa$ROOM_TYPE_DESCRIPTION_C[i], ignore.case = TRUE, perl = FALSE, fixed = FALSE, useBytes = FALSE) == TRUE)
  {
    hyatt_usa$room_type_cat[i] <- "Double Type Room"
  }
  else if (grepl("Suite", hyatt_usa$ROOM_TYPE_DESCRIPTION_C[i], ignore.case = TRUE, perl = FALSE, fixed = FALSE, useBytes = FALSE) == TRUE)
  {
    hyatt_usa$room_type_cat[i] <- "Suite Type Room"
  }
  else if (grepl("Delux", hyatt_usa$ROOM_TYPE_DESCRIPTION_C[i], ignore.case = TRUE, perl = FALSE, fixed = FALSE, useBytes = FALSE) == TRUE)
  {
    hyatt_usa$room_type_cat[i] <- "Delux Type Room"
  }
  else
  {
    hyatt_usa$room_type_cat[i] <- "other"
  }
}
```

In the original dataset, there are lot room type for King, Queen etc. We segregate all the King type rooms to 'King Type Room' to clearly visualize the result and trend of the hotel room type.

Random sampling

Random sampling is a part of the sampling technique in which each sample has an equal probability of being chosen. A sample chosen randomly is meant to be an unbiased representation of the total population. If for some reasons, the sample does not represent the population, the variation is called a sampling error.

Why did we chose Random sampling?

The full dataset has data for 12 months. The full data is too huge to run in R studio on personal computers. The best way to analyze the data is to pick sample from each month and merge them and created a data frame. We have selected 50000 sample from each month and merged them to get a sample size of 600000 observations.

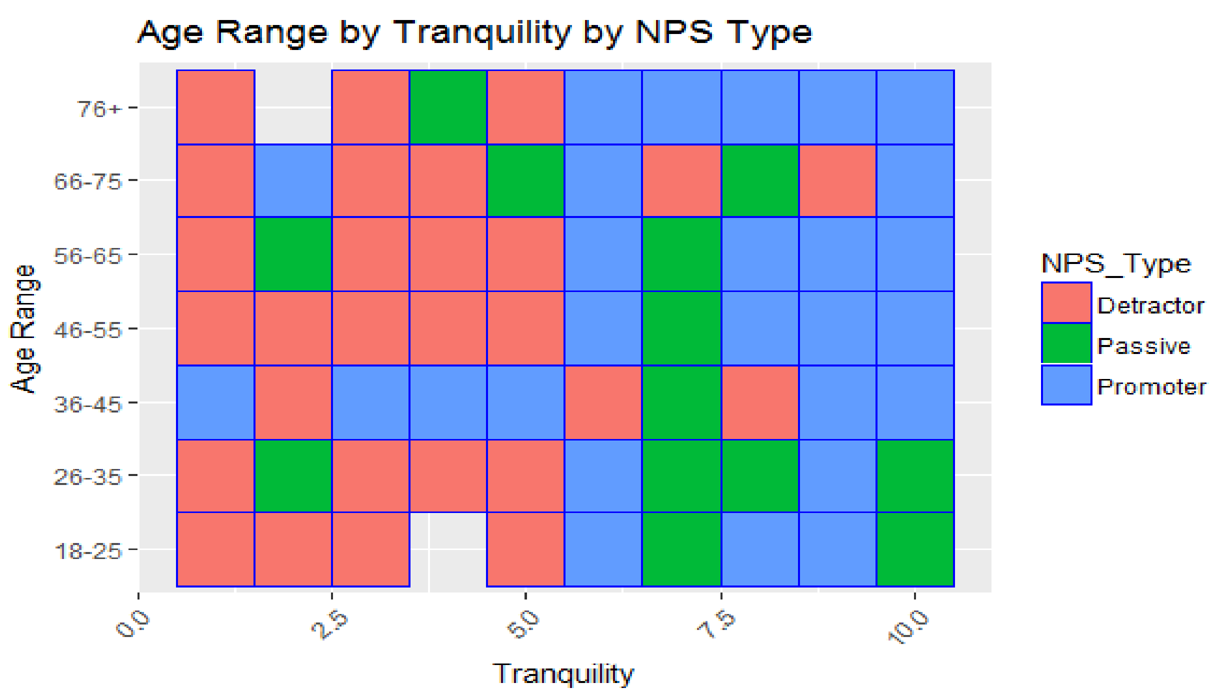
Random sampling source code

```r
df <- read.csv('out-201402.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201402-s.csv', x=df)
gc()
df <- read.csv('out-201403.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201403-s.csv', x=df)
gc()
df <- read.csv('out-201404.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201404-s.csv', x=df)
gc()
df <- read.csv('out-201405.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201405-s.csv', x=df)
gc()
df <- read.csv('out-201406.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201406-s.csv', x=df)
gc()
df <- read.csv('out-201407.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201407-s.csv', x=df)
gc()
df <- read.csv('out-201408.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201408-s.csv', x=df)
gc()
df <- read.csv('out-201409.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201409-s.csv', x=df)
gc()
df <- read.csv('out-201410.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201410-s.csv', x=df)
gc()
df <- read.csv('out-201411.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201411-s.csv', x=df)
gc()
df <- read.csv('out-201412.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201412-s.csv', x=df)
gc()
```

```
df <- read.csv('out-201501.csv')
df <- df[sample(nrow(df), 50000), ]
write.csv(file='out-201501-s.csv', x=df)
gc()

df1 <- read.csv('out-201402-s.csv')
df2 <- read.csv('out-201403-s.csv')
df3 <- read.csv('out-201404-s.csv')
df4 <- read.csv('out-201405-s.csv')
df5 <- read.csv('out-201406-s.csv')
df6 <- read.csv('out-201407-s.csv')
df7 <- read.csv('out-201408-s.csv')
df8 <- read.csv('out-201409-s.csv')
df9 <- read.csv('out-201410-s.csv')
df10 <- read.csv('out-201411-s.csv')
df11 <- read.csv('out-201412-s.csv')
df12 <- read.csv('out-201501-s.csv')
df <- rbind(df1, df2, df3, df4, df5, df6, df7, df8, df9, df10, df11, df12)

# delete any extra columns that got generated in the beginning during the process
df <- df[, -c(0:1)]
# write full year sample data
write.csv(file='full_year_sample.csv', x=df, row.names = FALSE)
```

Tranquility by Age Range

Below you will find our analysis using parameters such as tranquility, and age range data contained in the survey and NPS Type. Analysis explains how tranquility affects to the overall NPS Type for different age range Customers. We have used random sampling to conduct this analysis.



For our analysis, we have used ggplot and plotted the heat map to show the results.

Business Question - **can we improve NPS by accommodating guests of certain age range with quieter rooms?**

Our Observation

- Older age guests (age 75 or above) have liked quieter hotels/rooms.
- Guests in age range 45 to 75 also have tendency of liking quieter hotels/rooms

Our Recommendations

Allocating quieter rooms to older guests will improve Net Promoter Score.

- The recommendation can be implemented by providing guests an option in their online booking to hotels
- Accommodate older guests in quieter rooms when guests show up for check-ins in their respective hotels

Codes for Tranquility by Age Range analysis

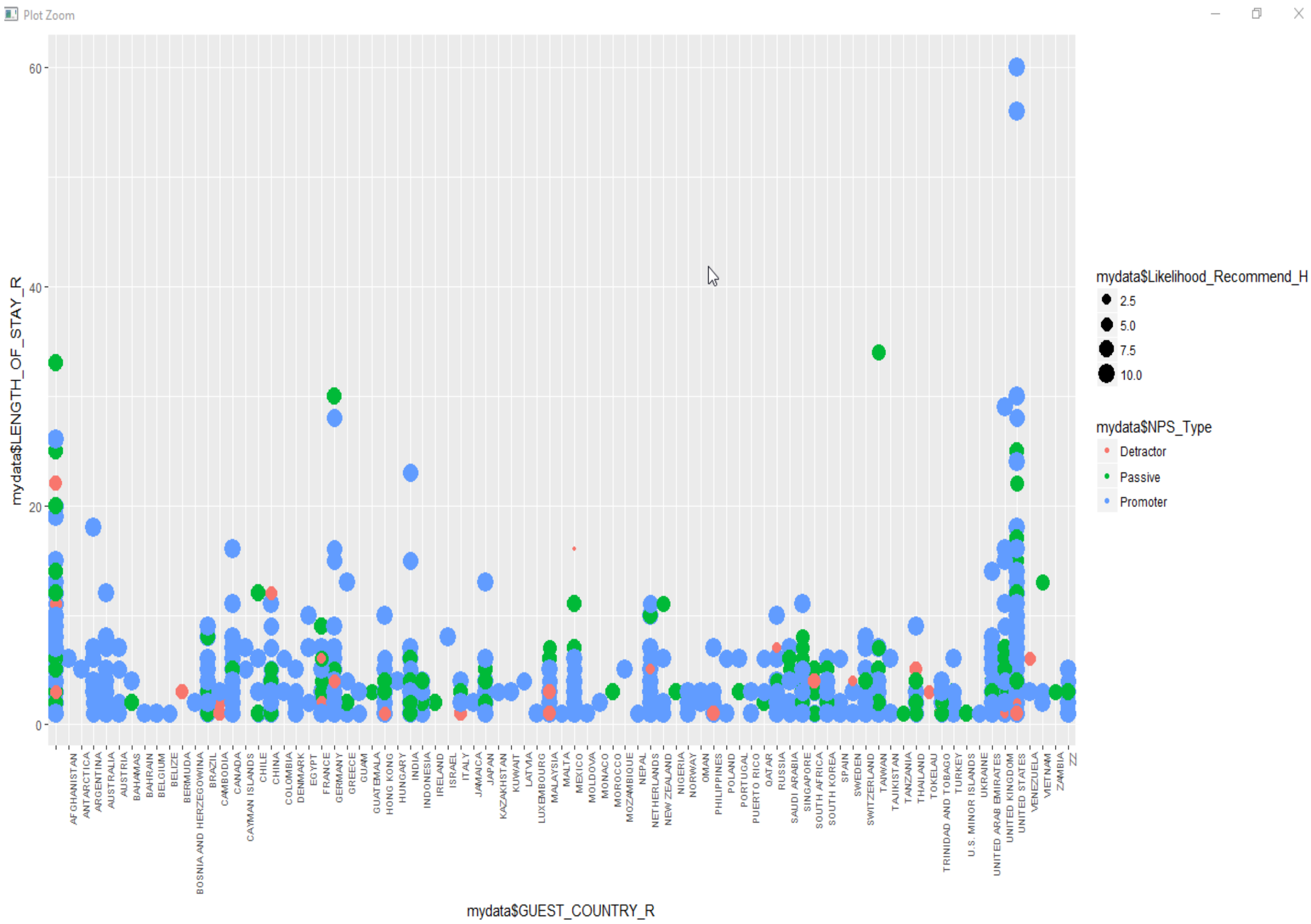
Reading the sample data csv file and copy the columns we want to analyze into vectors

```
df <- read.csv('full_year_sample.csv')
NPS_Type <- df$NPS_Type
Tranquility <- df$Tranquility_H
Age_Range <- df$Age_Range_H
# create a small dataframe from the above vectors
dfs <- data.frame(Tranquility, Age_Range, NPS_Type)
# remove NA and blank rows
dfs[dfs==""] <- NA
dfs <- na.omit(dfs)

# age tranquilty heat map
ggplot(dfs, aes(x = Tranquility, y = Age_Range)) + geom_tile(aes(fill = NPS_Type), colour = "blue") +
ggtitle("Age Range by Tranquility by NPS Type") + xlab("Tranquility") + ylab("Age Range") + theme(axis.text.x = element_text(angle=45, hjust=1))
```

Analysis

- Based on the sampled data set choose the target columns which were:
- Guest_Country
- NPS_Type
- Likelihood_to_recommend
- Length_of_stay
- Cleaning the NA from Likelihood_to_recommend
- Based value of rmse we predicted Likelihood_to_recommend on Length of stay
- The red dots can be seen only at the lower range where people stayed below 5 days
- The higher number of days people stayed the more the the Likelihood to recommend score

```r
mydata<-read.csv("C:/Users/shnaik/Copysample.csv")
colnames(mydata)
mydata<-mydata[,-233:-237]
mydata<-mydata[,-138:-231]
mydata<-mydata[,-67:-136]
mydata<-mydata[,-2:-58]
mydata<-mydata[,-3:-7]
mydata<-mydata[,-4]
head(mydata,10)

#Removing rows with Likelihood to recommend is Na
mydata<-mydata[!is.na(mydata$Likelihood_Recommend_H),]

install.packages("ggplot2")
library("ggplot2")
g<-ggplot(mydata,aes(x=mydata$GUEST_COUNTRY_R,y=mydata$LENGTH_OF_STAY_R))
g<-g+geom_point(aes(color=mydata$NPS_Type))
g <- g + theme(axis.text.x =element_text(angle = 90, hjust = 1,size = 7))
g
str(mydata)
lm<-(mydata$Likelihood_Recommend_H~mydata$LENGTH_OF_STAY_R)
Summary(lm)
```



Suggestion to improve NPS score

- Make the visitor stay more number of days as this will enable them to take more advantage of all the facilities in the hotels and more likely are they to give a better score.
- To make the customers stay more numbers of days there can be certain discount offers made available for a stay period of six and above this can be also combined with various lunch, dinner and spa expenses being covered
- Giving some anniversary or birthday discounts so that people plan a longer vacation

Conclusion

By conducting the analysis on a span of 12 Months data achieved by random sampling and also, Quarter-2 data consisting of April, May, and June, we can infer that the below factors are essential for improvement of the Hotel Brand value:

• Casino and Spa

• Internet and Check-in Service

• Length of stay and Room Location

• Tranquility

Lessons Learnt

- Scrutinizing large amount of data
- Working as a team of diverse group of people
- Understanding from team member’s view point
- Learning throughout the class and implementing the lesson learnt on data sets given in the start
- We were guided by the talented Professors and TA throughout the process that motivated us to give our best in the final project
- We learned the value of team work and utilized it to our full potential
- We learned the basic concepts of R programming, Data Analysis and the Kanban workflow process

References

Byers, T. (2015) Ggplot 2.0.0. Available at: https://blog.rstudio.org/2015/12/21/ggplot2-2-0-0/ (Accessed: April 27th, 2017)

CheckMarket (2011) Net promoter score (NPS) - use, application and pitfalls. Available at: https://www.checkmarket.com/blog/net-promoter-score/ (Accessed: April 27th, 2017)

Geom_boxplot. Ggplot2 2.1.0 (no date) Available at: http://docs.ggplot2.org/current/geom_boxplot.html (Accessed: 16 December 2016)

Geom_boxplot. Ggplot2 2.1.0 (no date) Available at: http://docs.ggplot2.org/current/geom_boxplot.html (Accessed: April 27th, 2017)

Godwin, H. (2011) Merge all files in a directory using R into a single dataframe. Available at: https://www.r-bloggers.com/merge-all-files-in-a-directory-using-r-into-a-single-dataframe/ (Accessed: April 27th, 2017)

Legends (ggplot2) (no date) Available at: http://www.cookbook-r.com/Graphs/Legends_(ggplot2)/ (Accessed: April 27th, 2017)

Robk, R.K. - (2014) Quick-r: Pie charts. Available at: http://www.statmethods.net/graphs/pie.html (Accessed: April 27th, 2017)

Systems, S. (2016) What is net promoter? Available at: https://www.netpromoter.com/know/ (Accessed: April 27th, 2017)