HYATT HOTEL CHAIN

NET PROMOTER SCORE IMPROVEMENT & RECOMMENDATIONS

[IST 687 GROUP PROJECT]

GROUP MEMBERS:

JINGJING ZHANG

SHIVANSHI BAJPAI

MANDEEP RANA

ANUJ JAIN

MANISH DALAL

Table of Contents

1. INTRODUCTION

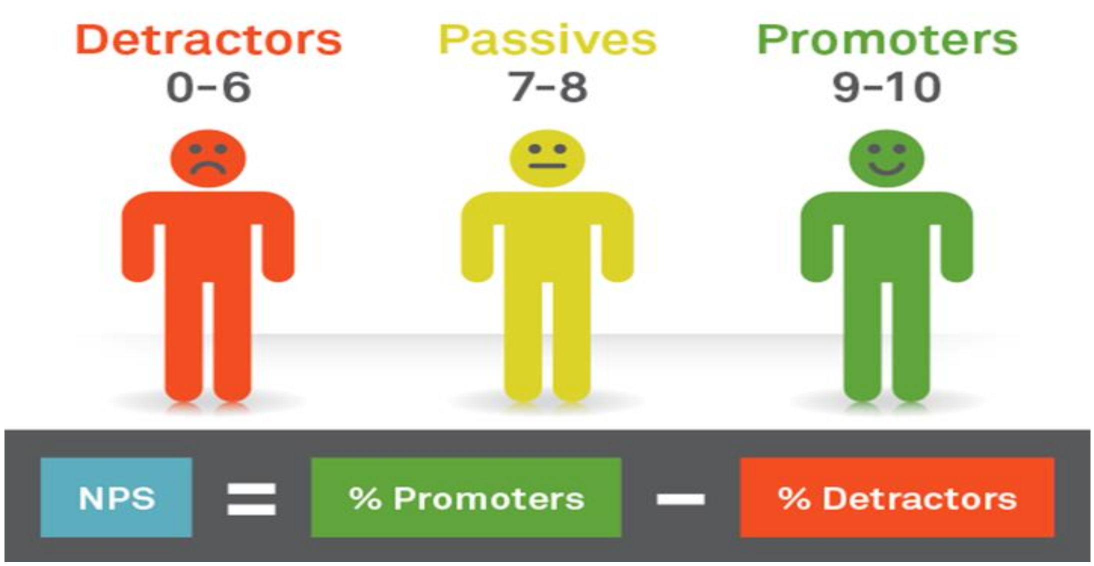
This document reports the analysis that we carried on the customer data of Hyatt Chain of Hotels. Customer satisfaction plays the key role in the success of any business. Hyatt Corporation conducts in-depth surveys with former customers collecting information on individual hotel experiences, geographic location data and their likelihood to recommend their hotel. These surveys help the business to determine which aspects of their hotels are satisfactory or which need to improve. The survey also helps the hotel to determine whether a customer is likely to be a "promoter" or "detractor" for Hyatt, with complacent people being labeled as "passive". In this document, we will describe our understanding, business questions about the dataset, our solutions to those questions, analysis and implementation. We would conclude by determining the factors that would help the Hyatt Brand of Hotels to improve their Net Promoter Score (NPS).

Net Promotor Score is a customer loyalty metric that is corelated with revenue growth. It indicates how likely the customer would recommend the hotel brand to a friend/relative/colleague. Respondents are grouped as follows:

Promoters (score 9-10) are loyal enthusiasts who will keep buying and refer others, fueling growth.

Detractors (score 0-6) are unhappy customers who can damage your brand and impede growth through negative word-of-mouth.

Passives (score 7-8) are satisfied but unenthusiastic customers who are vulnerable to competitive offerings.



2. BUSINESS QUESTIONS

We are aiming to answer following business questions based on the analysis of the dataset.

1. How are the Hyatt hotels in the USA performing overall?
2. What is the Net Promoter Score (NPS) distribution of Hyatt hotels in the USA?
3. How are the Hyatt hotel brands performing in the USA and determining which is the worst performing brand?
4. How is the performance of the Hyatt Hotels in the worst performing brand in the USA overall?
5. What are the customer opinions factors that are important for customer’s likelihood to recommend the worst performing Hyatt hotel brand?
6. Are the factors we found out in question number 5, really impacting the customer’s likelihood to recommend the worst performing brand?
7. Which hotel facilities are required to enhance the NPS value of the worst performing Hyatt hotel brand?

3. DATA CLEANSE/MUNGE/PREPARATION

The dataset was divided into 13 csv files based on the month wise from January 2014 to January 2015. In total, the information was well over 10GB in size, and much of it was filled with empty fields or unfinished responses. Each spreadsheet contained 237 variables, many of which were placeholders for non-existent questions or improperly named. To make the data ready for the analysis purpose, we performed the basic ETL process along with the cleansing of the dataset.

out-201402.csv	20.09.2016 14:17	1.21 GB
out-201403.csv	21.09.2016 06:11	1.4 GB
out-201404.csv	21.09.2016 06:20	1.36 GB
out-201405.csv	21.09.2016 07:40	1.43 GB
out-201406.csv	21.09.2016 09:10	1.4 GB
out-201407.csv	21.09.2016 09:52	1.39 GB
out-201408.csv	21.09.2016 17:03	1.45 GB
out-201409.csv	21.09.2016 17:04	1.4 GB
out-201410.csv	21.09.2016 18:28	1.5 GB
out-201411.csv	20.09.2016 14:21	1.36 GB
out-201412-short.csv	21.09.2016 18:23	8 KB
out-201412.csv	20.09.2016 14:18	1.26 GB
out-201501.csv	20.09.2016 14:13	1.13 GB

To efficiently operate with the dataset, we selected last four months of the dataset, i.e. Oct’2014, Nov’2014, Dec’2014 and Jan’2015 creating a 2 GB collection of entries that we could use for our computation and analysis. We chose this period of the dataset for the following reasons.

These four months basically denote holiday season with the customers traveling around the globe during the vacation. This period also includes many business travels as the new contracts for the next year is signed around this time. As the major portion of hotels business is coming from these four months as it is the holiday time and includes many business travels we would be analyzing these four months to suggest what hotels should do to improve their NPS. Our team decided to focus on the performance of US hotels. Thus, we only selected customer responses from the US hotels and deleted all the rows that came from the surveys in the non-US hotels. The dataset contained a lot of redundant values for e.g. NA’s and blank values. Additionally, many of the attributes were not in the desired format and would need to be converted. This would need to be treated before any adequate analysis could be carried out. Since our focus for the analysis is NPS which is determined by Likelihood_Recommend_H, it would be unnecessary to include the rows that did not have NPS_Type and Likelihood_Recommend_H. Thus, we deleted all the rows that did not have any values for these variables. Analyzing the guest specific data would help us understand what guests think about Hyatt Hotel Chain. Thus, our target analysis was focused on the customer feedback captured for the Hyatt Chain of hotels.

To understand the customer opinions about the hotel facilities, we considered factors like condition of hotels, customer service, guest room satisfaction, etc. For customer opinion columns like Check_In_H, Internet_Sat_H, Staff_Cared_H, Tranquility_H, Customer_SVC_H, Condition_Hotel_H, Guest_Room_H we replaced the NA values with the mean value of corresponding column group by its corresponding hotel. For example, if the customer's internet satisfaction data for HR La Jolla Hotel is 1, 2, NA,4,5, we substitute this NA with 3 (the mean of 1,2,4,5). Finally, we substituted the hotels without the Guest.NPS.Goal_PL values with the mean Guest.NPS.Goal_PL values of all the hotels in the nation, which is approximately 60.

As the team started to work with the deluge of data, it soon became apparent that we would need to prioritize attributes to work with. The dataset contains around 237 variables varying from demographic data of the Customers, Hotel location and amenities data, check-in information of the customers, revenue data and so much more. We needed to prioritize the data which we consider important to our analysis and filter out the rest of the data. So, after brainstorming within the team, we decided to perform our analysis based on the following columns:

Column Name	Definition
GP_Tier_H	GP tier of the guest
Guest NPS Goal_PL	Hotel's NPS goals
City_PL	City in which the hotel is located
Country_PL	Country in which the hotel is located
NPS_Type	Indicates if the guest's HySat responses mark them as a promoter, a passive, or a detractor
Likelihood_Recommend_H	Likelihood to recommend metric; value on a 1 to 10 scale
Guest_Room_H	Guest room satisfaction metric; value on a 1 to 10 scale
Tranquility_H	Tranquility metric; value on a 1 to 10 scale
Condition_Hotel_H	Condition of hotel metric; value on a 1 to 10 scale
Customer_SVC_H	Quality of customer service metric; value on a 1 to 10 scale
Staff_Cared_H	Staff cared metric; value on a 1 to 10 scale
Internet_Sat_H	Internet satisfaction metric; value on a 1 to 10 scale
Check_In_H	Quality of the check in process metric; value on a 1 to 10 scale
Bell Staff_PL	Flag indicating if the hotel has bell staff
Convention_PL	Flag indicating if the hotel has convention space
Dry-Cleaning_PL	Flag indicating if the hotel has dry-cleaning
Fitness Center_PL	Flag indicating if the hotel has a fitness center
Business Center_PL	Flag indicating if the hotel has a business center
Golf_PL	Flag indicating if the hotel is near a golf space
Laundry_PL	Flag indicating if the hotel has laundry space
Limo Service_PL	Flag indicating if the hotel has limo service
Restaurant_PL	Flag indicating if the hotel has onsite restaurants
Shuttle Service_PL	Flag indicating if the hotel has shuttle service
Spa_PL	Flag indicating if the hotel has a spa
Valet Parking_PL	Flag indicating if the hotel has valet parking
Hotel Name-Long_PL	Full hotel name
Property Latitude_PL	Latitude of the hotel's location
Property Longitude_PL	Longitude of the hotel's location
Class_PL	Class
Relationship_PL	Relationship of the hotel with Hyatt Corporation
Booking_Channel	Defined booking channel as per the NPS analysis
Internet_Dissat_Lobby_H	Detailed questions about the available internet
Internet_Dissat_Slow_H	Detailed questions about the available internet
Internet_Dissat_Expensive_H	Detailed questions about the available internet
Internet_Dissat_Connectivity_H	Detailed questions about the available internet
Internet_Dissat_Billing_H	Detailed questions about the available internet
Internet_Dissat_Wired_H	Detailed questions about the available internet
Room_Dissat_Internet_H	Detailed questions about the available internet
Status_H	Status of the survey
Room_Type_H	Guest's room type code
Language_H	The language that the survey was taken in
Age_Range_H	Guest's age range
Gender_H	Guest's gender
45 Guest_Country_H	Guest's country
46 Brand_PL	Hotel's brand
47 POV_CODE_C	Purpose of visit
48 State_PL	State in which the hotel is located

Below is the data cleansing R code taking into consideration data for for only Jan’2015 month.

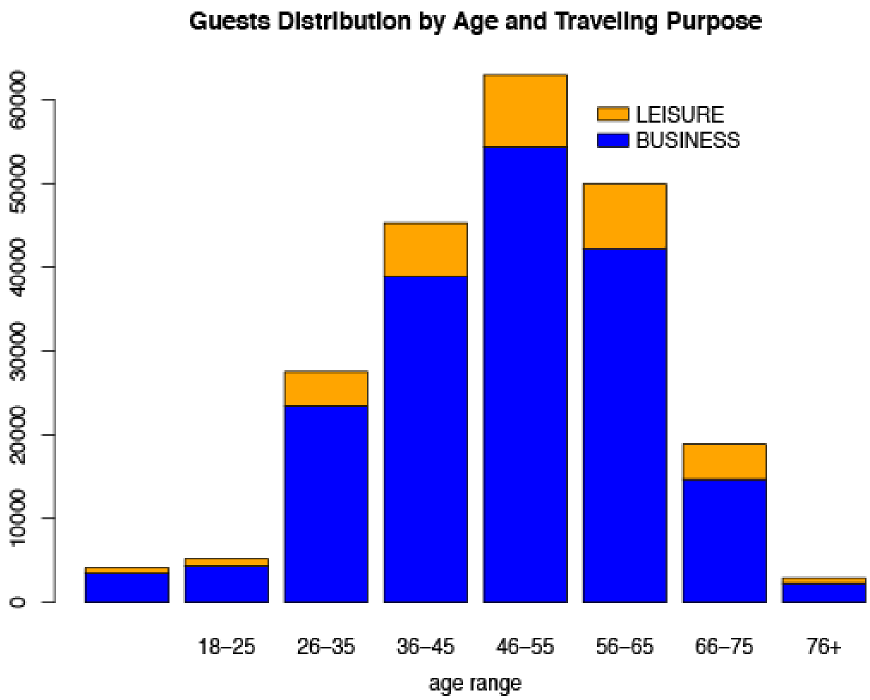
DATA CLEANSING R CODE

```
# clean and merge datasets

data.fname<-file.choose()

JanData <- readData[readData$Country_PL == "United States" ,]

str(JanData)

# delete the rows that did not have NPS_Type and LTR

unique(JanData$NPS_Type)

JanData <- JanData[-which(JanData$NPS_Type == ""), ]

str(JanData)

sum(is.na(JanData$Likelihood_Recommend_H))

str(JanData) # 23672 rows

# replace NA with mean value of the same hotel

JanData$Guest_Room_H <- with(JanData, ave(Guest_Room_H, Hotel.Name.Long_PL, FUN = function(x)
replace(x, is.na(x), mean(x, na.rm = TRUE))))

sum(is.na(JanData$Guest_Room_H))

# tranquility

JanData$Tranquility_H <- with(JanData, ave(Tranquility_H, Hotel.Name.Long_PL,

                              FUN = function(x) replace(x, is.na(x), mean(x, na.rm = TRUE))))

sum(is.na(JanData$Tranquility_H))

# hotel condition

JanData$Condition_Hotel_H <- with(JanData, ave(Condition_Hotel_H, Hotel.Name.Long_PL,

                              FUN = function(x) replace(x, is.na(x), mean(x, na.rm = TRUE))))

sum(is.na(JanData$Condition_Hotel_H))

# Quality of customer service metric; value on a 1 to 10 scale

JanData$Customer_SVC_H <- with(JanData, ave(Customer_SVC_H, Hotel.Name.Long_PL,

                              FUN = function(x) replace(x, is.na(x), mean(x, na.rm = TRUE))))

sum(is.na(JanData$Customer_SVC_H))


# Staff cared metric; value on a 1 to 10 scale

JanData$Staff_Cared_H <- with(JanData, ave(Staff_Cared_H, Hotel.Name.Long_PL,

                              FUN = function(x) replace(x, is.na(x), mean(x, na.rm = TRUE))))

sum(is.na(JanData$Staff_Cared_H))
```

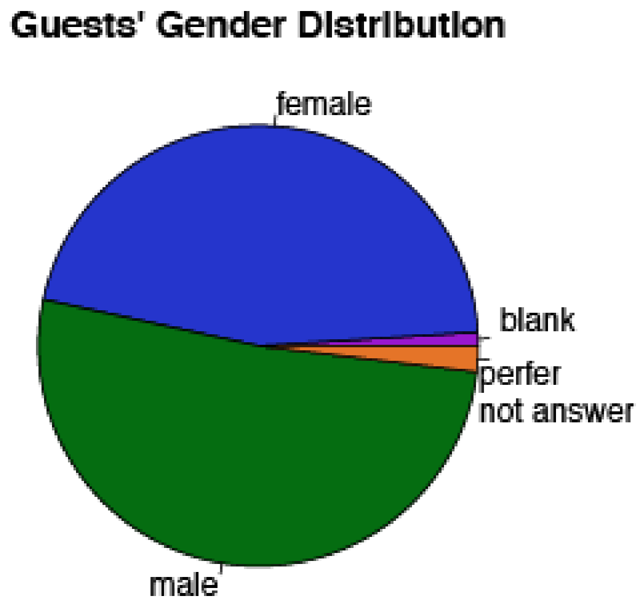
```
#NPS goals

JanData$Check_In_H <- with(JanData, ave(Guest.NPS.Goal_PL, Hotel.Name.Long_PL,

                         FUN = function(x) replace(x, is.na(x), mean(x, na.rm = TRUE))))

sum(is.na(JanData$Guest.NPS.Goal_PL))

JanData$Guest.NPS.Goal_PL

# check the new dataset

View(JanData)
```

4. DESCRIPTIVE STATISTICAL ANALYSIS

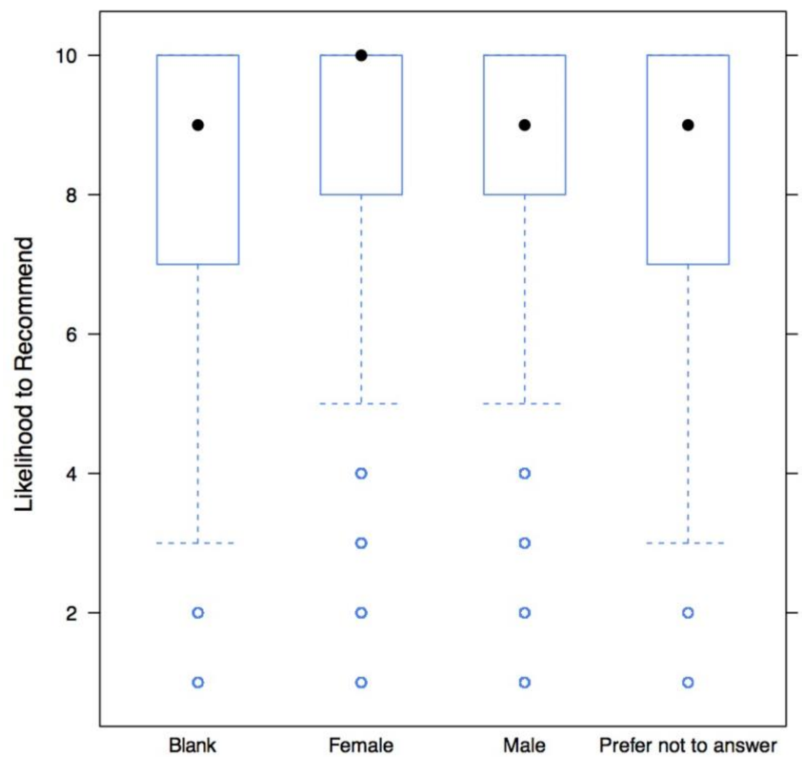
The final dataset we work on ranges from October 2014 to January 2015. We're interested in this part of dataset mainly because we want to focus on the Hyatt corporation's performance during the holiday season. By applying a statistical analysis of the four months, we found 84.7% of all the guests were traveling for business, illustrated as the figure below. Most (73%) of all the guests' age range stays in 36-65, with 29% of all the guests are 46-55.



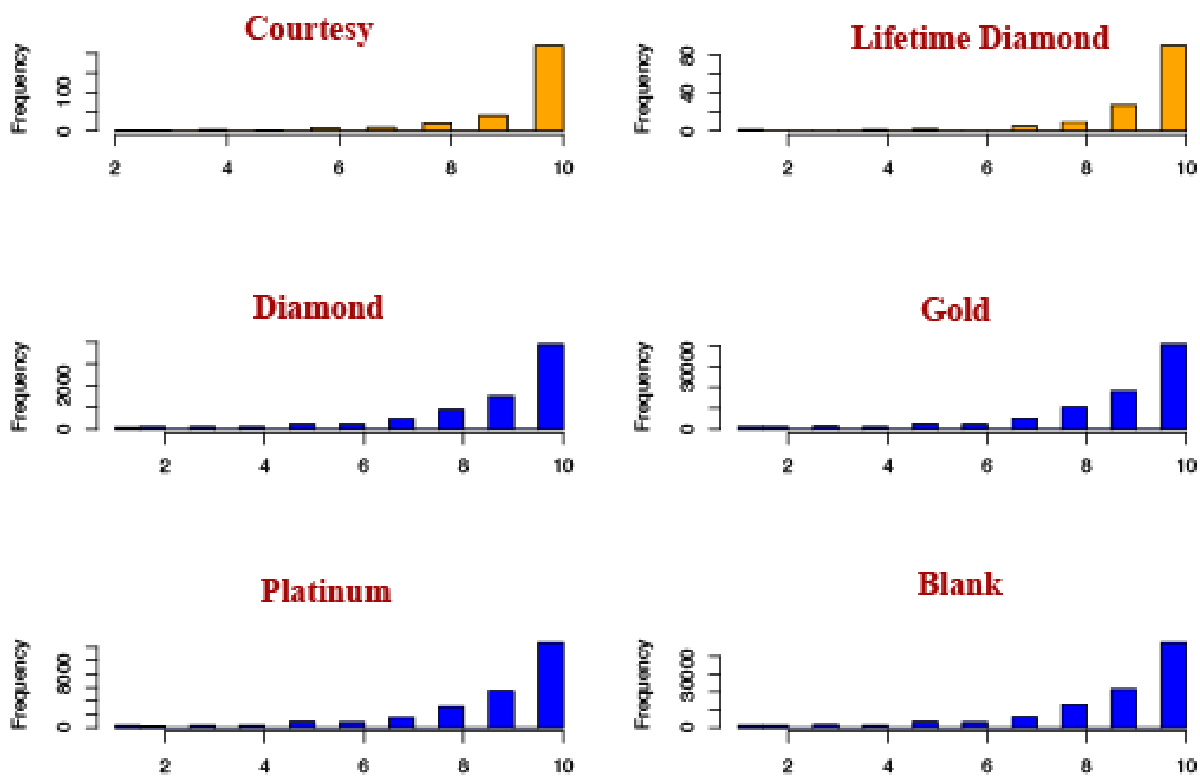
For these four months, 86% of all the hotel guests come from the USA out of 154 countries, with 45% female and 51% male guests.



To learn whether gender has an influence on the LTR, we plot the LTR distribution among different gender groups, as shown below. The black dot represents the median LTR and the blue box means the range of LTR. The plot shows that the female guests are more likely to recommend with median LTR of 10, while male guests and the guests who prefer not to declare their gender are less likely to recommend. We validate the above analysis by computing the percentage of promoters for gender groups: there is no significant difference between female and male guests on LTR. It is notable that the guests who prefer not to declare their gender have a much lower likelihood to promote and a higher likelihood to detract.

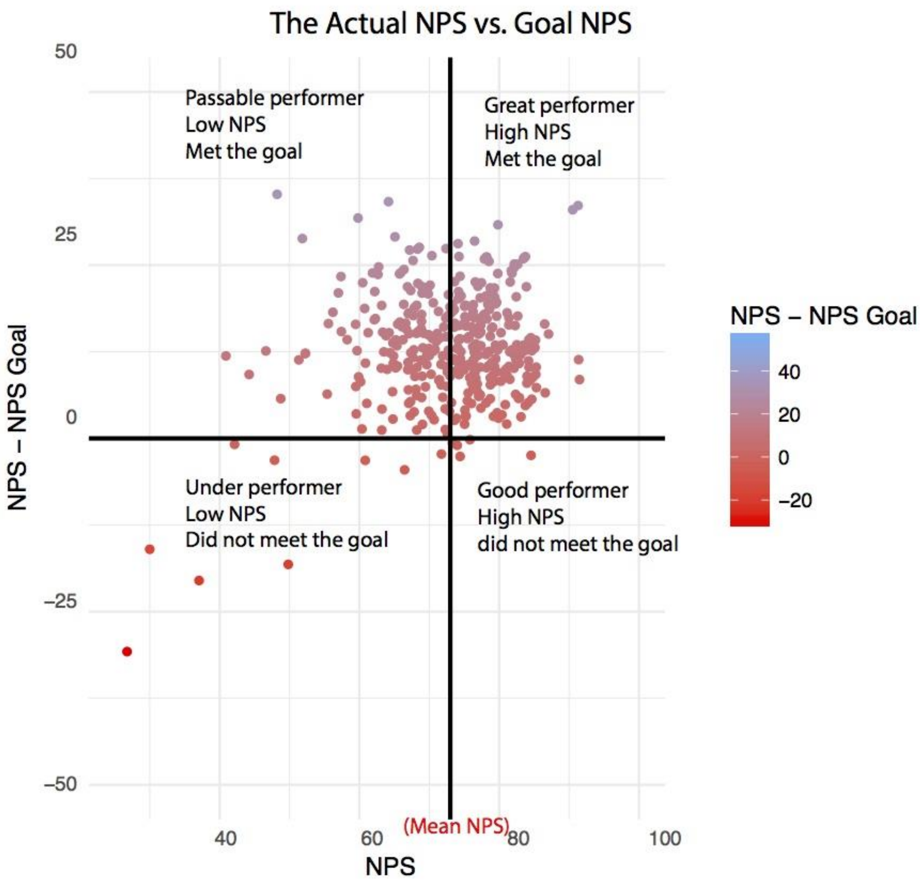


The Hyatt membership has five categories: platinum, gold, diamond, lifetime diamond, and courtesy. The LTR distribution among the guests who belong to different groups of membership categories listed as below. The guests who have a courtesy membership or a lifetime diamond membership have a much higher likelihood to recommend. These two groups of guests are frequent ones and have a high loyalty. Thus we recommend Hyatt corporation should invest in member loyalty program to hold the promoters, which would further improve the corporation's reputation, improve NPS score, and increase the corporate revenues ultimately.



5. BUSINESS QUESTIONS AND RECOMMENDATIONS

5.1 QUESTION 1: How are the Hyatt hotels in USA performing overall?



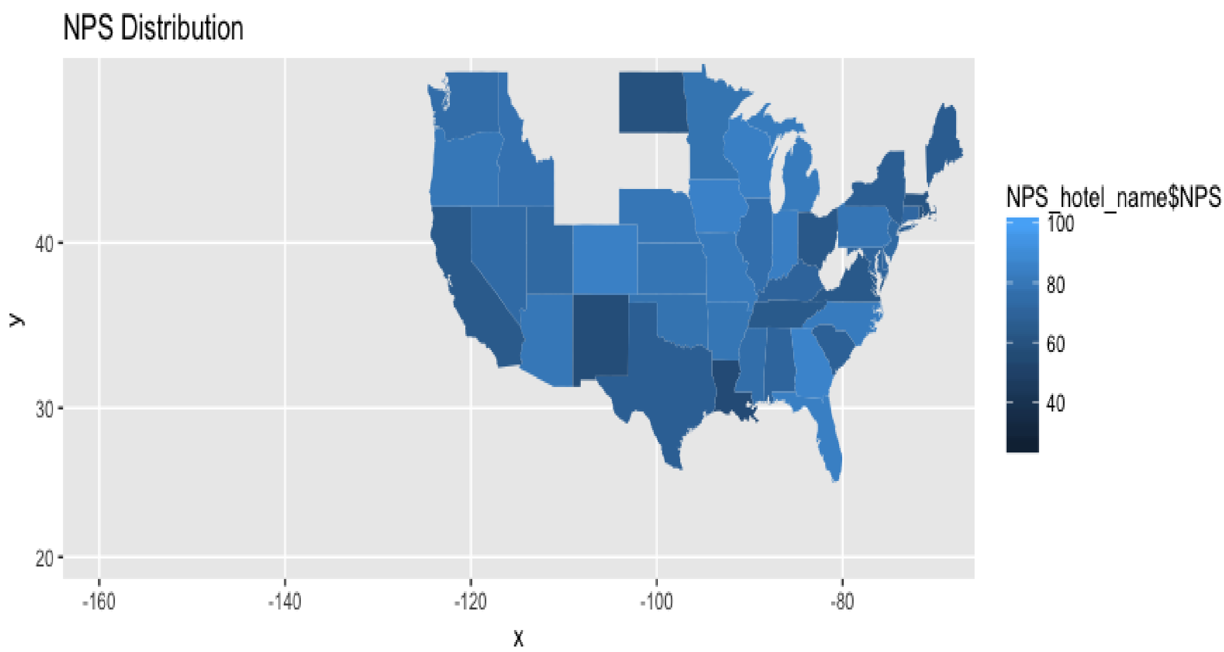
The 396 hotels within the United States that belong to 7 brands are in 44 states, with the hotels in California, Florida and Texas accommodating the largest number of guests. All the hotels in the United States during the four months performed significantly well, with mean NPS of 73. Specifically,

Priority efforts must be laid on the Under Performing Hotels as they are really dragging the overall NPS of the Hyatt brand hotels in the US down, and their low goal and low performances could be hurting the Hyatt brand. Business models of the Great Performing Hotels should be taken as an example and we would suggest further analysis on why they are doing so great.

R CODE:

```
#------------ US Hotels performance -----------------#

NPS_us <- aggregate(f2$f1.Likelihood_Recommend_H,list(f2$f1.Hotel.Name.Long_PL),
                    FUN=function(x){
                      y <- (sum(x[x>=9])-sum(x[x<=6]))*100/sum(x);
                      return(y)})

View(NPS_us)


NPSgoal_us <- aggregate(f2$f1.Guest.NPS.Goal_PL,list(f2$f1.Hotel.Name.Long_PL),mean)

NPSdiff_us <- NPS_us$x-NPSgoal_us$x

NPSdf_us <- data.frame(NPS_us$Group.1,NPS_us$x,NPSgoal_us$x,NPSdiff_us)

colnames(NPSdf_us) <- c("hotel_name","NPS","NPS_goal","NPS-NPS_goal")

View(NPSdf_us)


library(ggplot2)

us_performance <- ggplot(NPSdf_us, aes(x=NPS, y=NPS-NPS_goal)) +

  geom_point(aes(color=NPS-NPS_goal)) +

  scale_colour_gradient(low = "red")+

  lims(x=c(25,100),y=c(-50,50)) +

  theme_minimal() +

  coord_fixed() +

  geom_vline(xintercept = 68) + geom_hline(yintercept = 0)

us_performance
```

5.2 QUESTION 2: What is the Net Promoter Score (NPS) distribution of Hyatt hotels in USA?



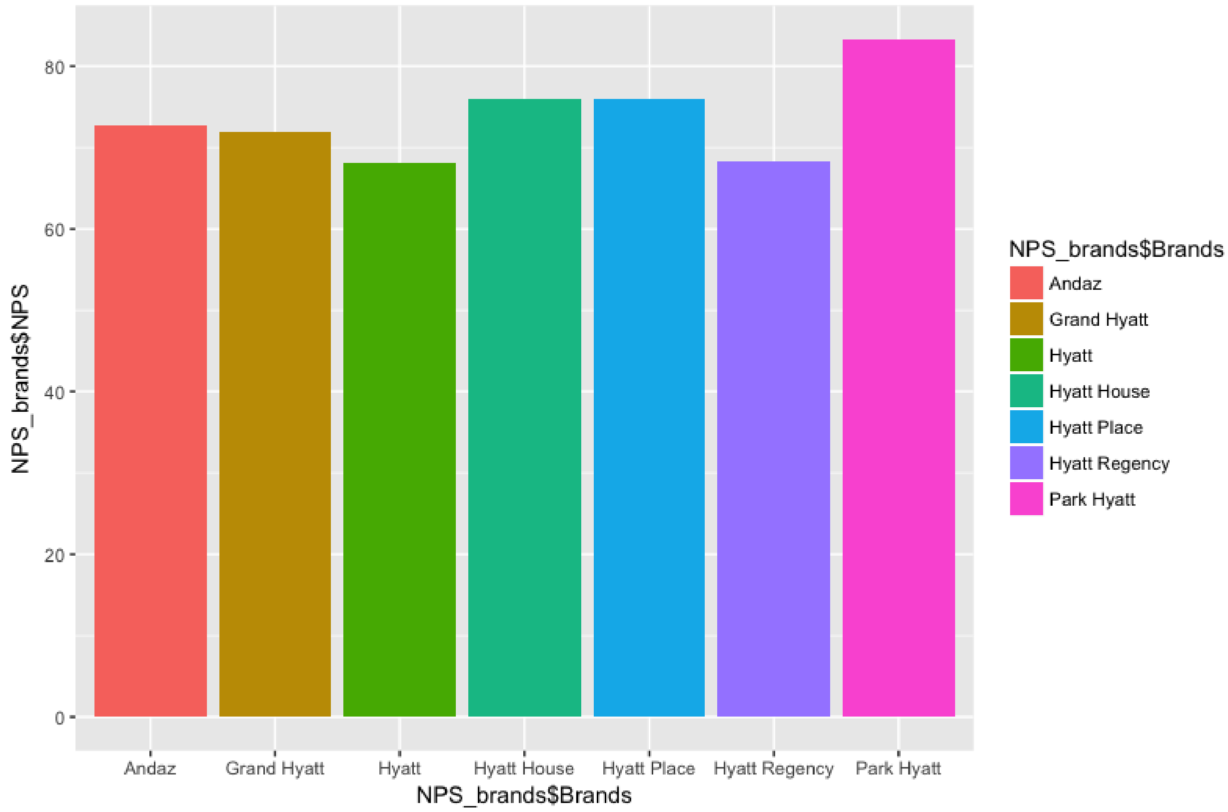
INSIGHTS

The above map showcases the NPS distribution of Hyatt hotels states wise over the USA. From the above map, we can perceive that states like Florida, California and Texas perform well and accommodated the largest number of guests. All the hotels in the USA performed well in four months (Oct'14, Nov'14, Dec'14 and Jan'15) with mean NPS of 73. Also, we observe that states like North Dakota, New Mexico, Louisiana are under performers acting as drainers for the Hyatt Corporation business.

R CODE:

```
library(ggplot2)

LTRhotel <- aggregate(f2$f1.Likelihood_Recommend_H,list(f2$f1.Hotel.Name.Long_PL),mean)

NPS_hotel_name <-
aggregate(f2$f1.Likelihood_Recommend_H,list(f2$f1.State_PL,f2$f1.Property.Latitude_PL,f2$f1.Property.Long
itude_PL),

              FUN=function(x){

               y <- (sum(x[x>=9])-sum(x[x<=6]))*100/sum(x);

               return(y)})

View(NPS_hotel_name)

NPS_hotel_name$Group.1<-tolower(NPS_hotel_name$Group.1)

names(NPS_hotel_name)<-c('State_PL','Lat_PL','Long_PL','NPS')

library(openintro)

us=map_data("state")

us

map.HotColor<- ggplot(NPS_hotel_name,aes(map_id=NPS_hotel_name$State_PL))

map.HotColor<-map.HotColor+geom_map(map=us,aes(fill=NPS_hotel_name$NPS))

map.HotColor<-map.HotColor+expand_limits(x=NPS_hotel_name$Long_PL,y=NPS_hotel_name$Lat_PL)

map.HotColor<-map.HotColor+coord_map()+ggtitle("NPS Distribution")

map.HotColor
```

5.3 QUESTION 3: How are the Hyatt hotel brands performing in USA and determining which is the worst performing brand?



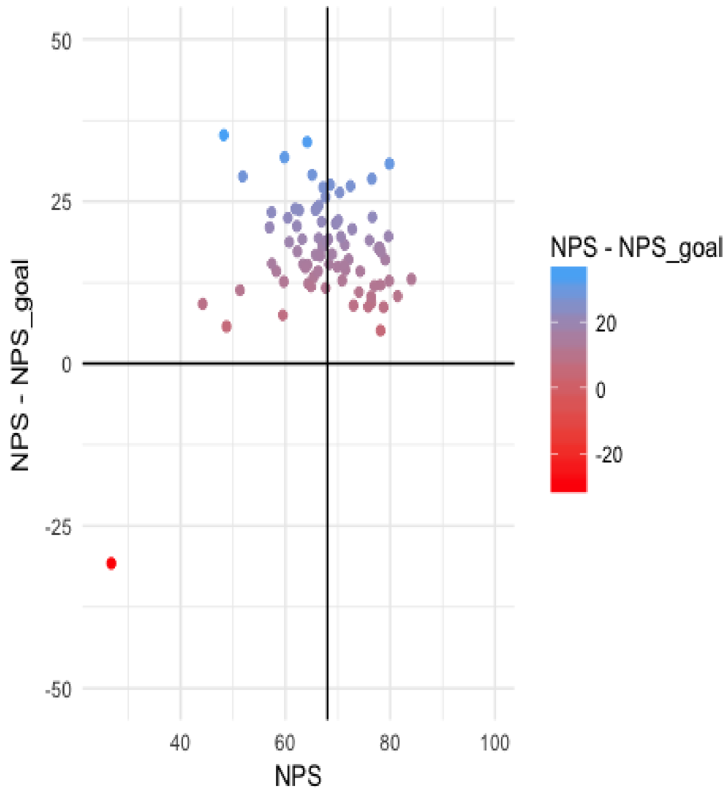
INSIGHTS

Park Hyatt is a great performing brand with high NPS value. Performance of other brands like Hyatt House and Hyatt Place is quite good. Brands like Andaz and Grand Hyatt are performing satisfactorily as compared to worst performers like Hyatt and Hyatt Regency. As Hyatt Regency brand is not performing well, we select this brand for our further analysis. It has enough data for analysis that is a total of 82,887 records.

R CODE:

```
NPS_brands <-aggregate(f2$f1.Likelihood_Recommend_H,list(f2$f1.Brand_PL),
                   FUN=function(x){
                     y <- (sum(x[x>=9])-sum(x[x<=6]))*100/sum(x);
                     return(y)})
names(NPS_brands) <- c("Brands","NPS")
#View(NPS_brands)
library(ggplot2)
g <- ggplot(NPS_brands, aes(x= NPS_brands$Brands, y = NPS_brands$NPS, fill = NPS_brands$Brands))+
geom_bar(stat="identity")
g
# Best perfomring is Park Hyatt and least performing is Hyatt Regency(this brand because has more data)
nrow(f2[f2$f1.Brand_PL=="Hyatt Regency", ])
nrow(f2[f2$f1.Brand_PL=="Hyatt", ])
nrow(f2[f2$f1.Brand_PL=="Park Hyatt", ])


# we selected Hyatt Regency because it is least perfomring and has more data
# performance of Hyatt Regency in USA overall
```

5.4 QUESTION 4: How is the performance of the Hyatt Hotels in the worst performing brand in the USA?



INSIGHTS:

From the previous question, we got an insight that Hyatt Regency is the worst performing brand. With the further predictive analysis on the performance of all the hotels of Hyatt Regency brand, we can observe that all the hotels are 'good' and 'passive' performers except for 'The Concourse Hotel at LA International Airport' (red point on the plot above). From the above plot, we can conclude that this hotel is the worst performer of all the Hyatt Regency brand hotels. The NPS of Concourse Hotel is 26.8%. Thus the Concourse hotel is now undergoing renovation of about $65 million.

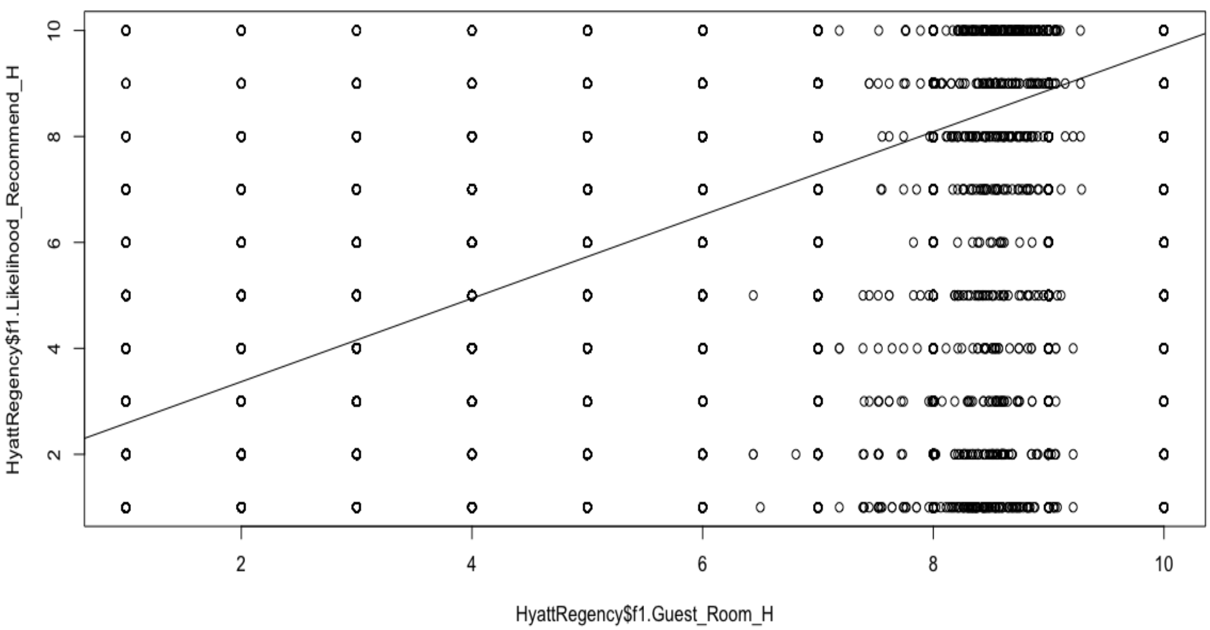
R CODE:

```
# Hyatt Regency performance

#average of each hotels of regency brand

HyattRegency <-f2[f2$f1.Brand_PL=="Hyatt Regency", ]

LTR_regency <-
aggregate(HyattRegency$f1.Likelihood_Recommend_H,list(HyattRegency$f1.Hotel.Name.Long_PL),mean)

NPS_regency <-
aggregate(HyattRegency$f1.Likelihood_Recommend_H,list(HyattRegency$f1.Hotel.Name.Long_PL),

             FUN=function(x){

              y <- (sum(x[x>=9])-sum(x[x<=6]))*100/sum(x);

              return(y)})

View(NPS_regency)

NPSgoal_regency <-
aggregate(HyattRegency$f1.Guest.NPS.Goal_PL,list(HyattRegency$f1.Hotel.Name.Long_PL),mean)

NPSdiff_regency <- NPS_regency$x-NPSgoal_regency$x

NPSdf_regency <- data.frame(NPS_regency$Group.1,NPS_regency$x,NPSgoal_regency$x,NPSdiff_regency)

colnames(NPSdf_regency) <- c("hotel_name","NPS","NPS_goal","NPS-NPS_goal")

View(NPSdf_regency)

# plotting all the regency hotels in USA

# underperformer: The Concourse Hotel at Los Angeles Intl Airport

#NPSdf_regency[index,]

library(ggplot2)

regency_performance <- ggplot(NPSdf_regency, aes(x=NPS, y=NPS-NPS_goal)) +

  geom_point(aes(color=NPS-NPS_goal)) +

  scale_colour_gradient(low = "red")+

  lims(x=c(25,100),y=c(-50,50)) +

  theme_minimal() +

  coord_fixed() +

  geom_vline(xintercept = 68) + geom_hline(yintercept = 0)

regency_performance
```

```
# mean of NPS of entire regency hotels in USA

mean(NPSdf_regency$NPS)

#number fo regency hotel in USA that is not performing well

sum(NPSdf_regency$`NPS-NPS_goal`<0)

# index of least performing hotel that is concourse hotel LA

index <-which(NPSdf_regency$`NPS-NPS_goal`<0)

index
```

5.5 QUESTION 5: What are the customer opinions factors that are important for customer's likelihood to recommend the worst performing Hyatt hotel brand?

To determine the customer opinion factors that are important for customers likelihood to recommend we performed linear modeling of the guest opinions on Guest Room, Hotel Condition, Customer Service, Staff Cared on LTR for the Hyatt Regency hotel. The modelling result is statistically significant since the p-value is less than 2.2e-16 and multiple R-squared value is 0.6729. Linear modeling of the guest opinions on Guest Room, Hotel Condition, Customer Service, Staff Care on LTR for the Concourse Hotel was also perform. The modeling result is statistically significant since the p-value is less than 2.2e-16 and multiple R-squared value is 0.7228.



INSIGHTS:

The statistical analysis indicates that the quality of check in process is not a significant customer opinion factor to determine the NPS. Hotels should focus on Customer Service, Hotel Condition and Guest Room to improve their NPS. Guest room rating is the main reason accounting for the negative NPS.

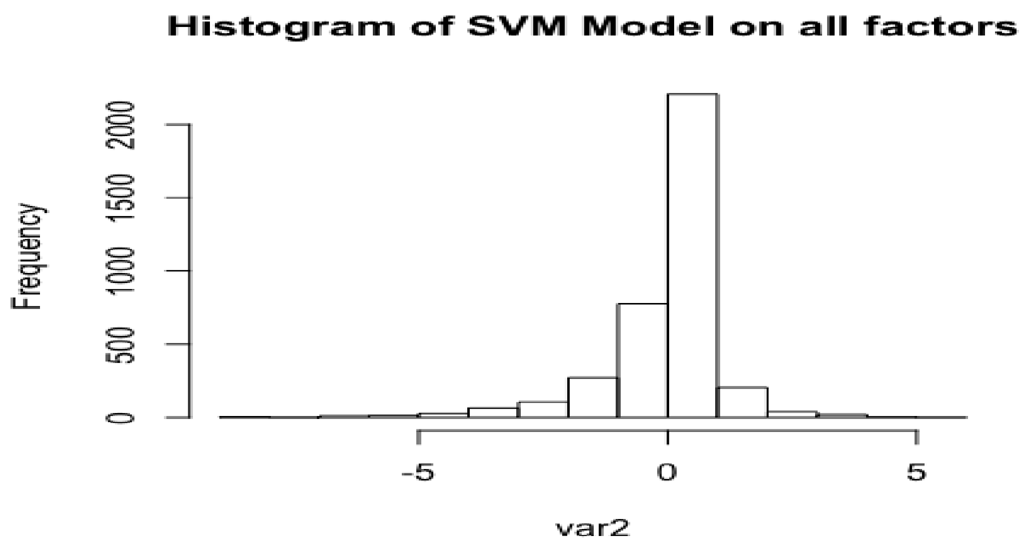
R CODE:

```
m1_hyattReg <- lm(formula = f1.Likelihood_Recommend_H ~ f1.Guest_Room_H+ f1.Tranquility_H+
f1.Condition_Hotel_H+
                f1.Customer_SVC_H+ f1.Staff_Cared_H+f1.Internet_Sat_H+f1.Check_In_H , data = HyattRegency)
summary(m1_hyattReg)

#Check_In is not significant
m2_hyattReg <- lm(formula = f1.Likelihood_Recommend_H ~ f1.Guest_Room_H+  f1.Condition_Hotel_H+
                f1.Customer_SVC_H+ f1.Staff_Cared_H, data = HyattRegency)
summary(m2_hyattReg)
#Internet is not significant
# m3_hyattReg <- lm(formula = f1.Likelihood_Recommend_H ~ f1.Guest_Room_H+ f1.Tranquility_H+
f1.Condition_Hotel_H+
#               f1.Customer_SVC_H+ f1.Staff_Cared_H , data = HyattRegency)
# summary(m3_hyattReg)

# m4_hyattReg <- lm(formula = f1.Likelihood_Recommend_H ~ f1.Guest_Room_H+ f1.Tranquility_H+
f1.Condition_Hotel_H+
#               f1.Customer_SVC_H +f1.Internet_Sat_H+f1.Check_In_H , data = HyattRegency)
# summary(m4_hyattReg)
#
# m5_hyattReg <- lm(formula = f1.Likelihood_Recommend_H ~ f1.Guest_Room_H+ f1.Condition_Hotel_H+
f1.Staff_Cared_H +f1.Customer_SVC_H  , data = HyattRegency)
# summary(m5_hyattReg)

# Guest_Room is important because it is giving 0.5292 which is the highest value compared to the other 2
m3_hyattReg <- lm(formula = f1.Likelihood_Recommend_H ~ f1.Guest_Room_H, data = HyattRegency)
summary(m3_hyattReg)

plot( HyattRegency$f1.Guest_Room_H, HyattRegency$f1.Likelihood_Recommend_H) +
abline(lm(f1.Likelihood_Recommend_H ~ f1.Guest_Room_H, data= HyattRegency))

m4_hyattReg <- lm(formula = f1.Likelihood_Recommend_H ~  f1.Condition_Hotel_H,data = HyattRegency)
summary(m4_hyattReg)

m5_hyattReg <- lm(formula = f1.Likelihood_Recommend_H ~  f1.Customer_SVC_H, data = HyattRegency)
summary(m5_hyattReg)
```

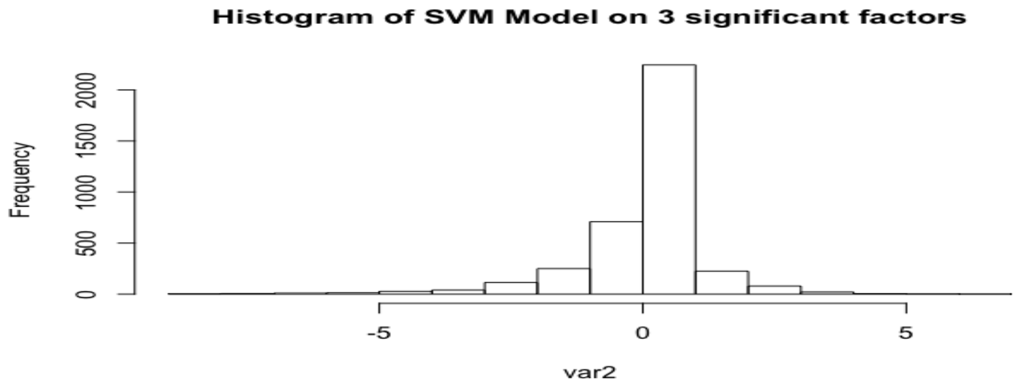
5.6 Question 6: SVM validation

In order to prove the results that we obtained in the Business question 5, we performed SVM and KSVM validation by training the model and then predicting the results by giving test data as an input to the trained model. The results were validated by plotting histogram and determining the error that is the difference between the actual and predicted values against the majority of the data points. Therefore, from the below 3 histograms, we can observe that major data points lie in the area where there is minor difference (between 0 and 1) between the actual and predicted values, which means that the model is accurate. We performed analysis on 3 cases:

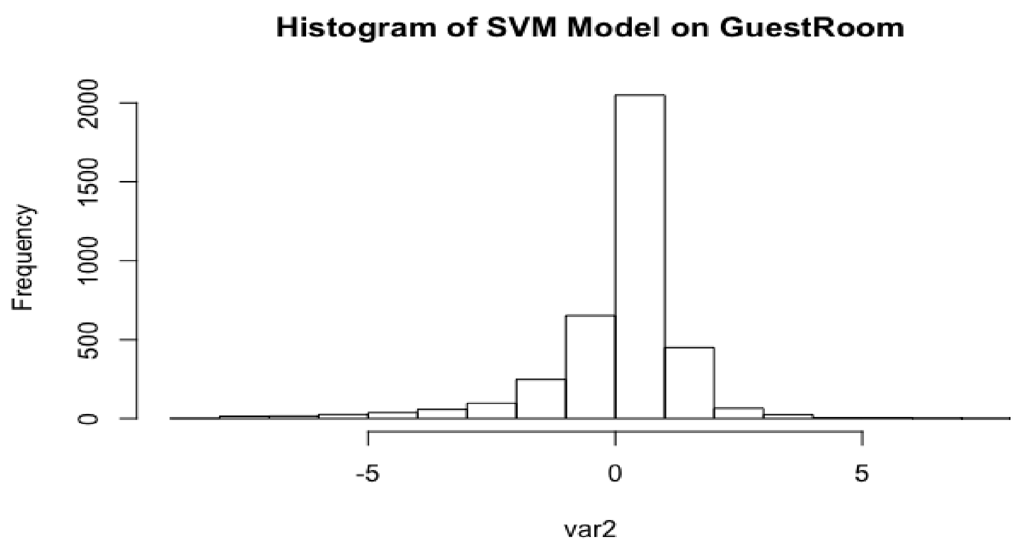
1. Validating whether customer opinions like Guest Room, Hotel Condition, Customer SVC, Tranquility, Internet and Staff Cared are really influencing the likelihood to recommend the Hyatt Regency Hotels



2. Validating whether the 3 customer opinions like Customer SVC, Guest Room and Hotel Condition are really the most significant parameters that the Hyatt Regency Hotels should focus on in order to enhance their NPS score



3. Validating whether Guest Room customer opinion is the most important opinions out of the other 7 opinions that the Hyatt Regency Hotels should focus on to improve the NPS making customers more loyal to their hotels



R CODE:

```
HyattRegency_validation <- HyattRegency[,c("f1.Likelihood_Recommend_H",
                              "f1.Guest_Room_H","f1.Condition_Hotel_H", "f1.Customer_SVC_H")]

HyattRegency_validation <- HyattRegency[,c("f1.Likelihood_Recommend_H",
                              "f1.Guest_Room_H","f1.Condition_Hotel_H",
"f1.Customer_SVC_H","f1.Staff_Cared_H","f1.Tranquility_H","f1.Internet_Sat_H")]

HyattRegency_validation <- HyattRegency[,c("f1.Likelihood_Recommend_H",
                              "f1.Guest_Room_H")]

nrow(HyattRegency)

nrows <- nrow(HyattRegency_validation)

Na <- sum(is.na(HyattRegency_validation))

Na

jl <- sum(is.na(HyattRegency_validation$f1.Guest_Room_H))

jl

HyattRegency_validation <- na.omit(HyattRegency_validation)

random.indexes <- sample(1:nrows,replace=FALSE)

#View(random.indexes)

# create a 2/3 cutpoint and round the number

cutpoint<-floor(nrows/4*3)

# check the 2/3 cutpoint

cutpoint

# create train data set, which contains the first 2/3 of overall data

mydata.train <-HyattRegency_validation[random.indexes[1:cutpoint],]

mydata.train

nrow(mydata.train)

# check the train dataset

str(mydata.train)

# create test data, which contains the left 1/3 of the overall data

mydata.test <-HyattRegency_validation[random.indexes[(cutpoint+1):nrows],]

mydata.test <- na.omit(mydata.test)

mydata.train <- na.omit(mydata.train)
```

```
X <- sum(is.na(mydata.train))

X

Y <- sum(is.na(mydata.test))

Y

nrow(mydata.test)

nrow(mydata.train)

# check the test dataset

str(mydata.test)

library(kernlab)

#install.packages("kernlab")

#install.packages("e1071")

#install.packages("ggplot2")

#install.packages("gridExtra")

#install.packages("caret")

#install.packages("Metrics")

library("Metrics")

library("kernlab")

library("e1071")

library("ggplot2")

library("gridExtra")

library("caret")

svmOutput <- svm(f1.Likelihood_Recommend_H ~ f1.Guest_Room_H + f1.Condition_Hotel_H +
f1.Customer_SVC_H,

          data=mydata.train)

svmOutput

svmpred <- predict(svmOutput,mydata.test)

svmpred

table(svmpred)

actual <- mydata.test$f1.Likelihood_Recommend_H

actual
```

```
length(actual)

length(svmpred)

var2 <- as.numeric(actual) - as.numeric(svmpred)

View(var2)

var2 <- na.omit(var2)

hist(var2, main = "Histogram of SVM Model on 3 significant factors")

# on all the factors

svmOutputall <- svm(f1.Likelihood_Recommend_H ~ f1.Guest_Room_H + f1.Condition_Hotel_H +
f1.Customer_SVC_H+f1.Staff_Cared_H+f1.Tranquility_H+f1.Internet_Sat_H,

            data=mydata.train)

svmOutputall

svmpred <- predict(svmOutputall,mydata.test)

svmpred

table(svmpred)

actual <- mydata.test$f1.Likelihood_Recommend_H

actual

length(actual)

length(svmpred)

var2 <- as.numeric(actual) - as.numeric(svmpred)

View(var2)

var2 <- na.omit(var2)

hist(var2, main = "Histogram of SVM Model on all factors")

# on guest room

svmOutputguest <- svm(f1.Likelihood_Recommend_H ~ f1.Guest_Room_H,

              data=mydata.train)

svmOutputguest

svmpred <- predict(svmOutputguest,mydata.test)

svmpred
```

```
table(svmpred)

actual <- mydata.test$f1.Likelihood_Recommend_H

actual

length(actual)

length(svmpred)

var2 <- as.numeric(actual) - as.numeric(svmpred)

View(var2)

var2 <- na.omit(var2)

hist(var2, main = "Histogram of SVM Model on GuestRoom")
```

KSVM Validation:

We performed Kernel SVM to analyze whether the 3 customer opinions that is Hotel Condition, Customer SVC and Guest Room are really significant and important for the Hyatt Regency Hotels. Our model proved that these 3 parameters are really important for NPS score with training error of 0.29, giving us a good accuracy.

R CODE:

```
#KSVM
modelksvm <-ksvm(f1.Likelihood_Recommend_H ~
                 f1.Condition_Hotel_H+
                 f1.Guest_Room_H+
                 f1.Customer_SVC_H,
               data=mydata.train,
               kernel="rbfdot",
               kpar="automatic",
               C=30,
               cross=50)

modelksvm
unique(f2$f1.Likelihood_Recommend_H)

predictedksvm <-predict(modelksvm,mydata.test)
predictedksvm
mydata.test
nrow(mydata.test)

mydata.test
nrow(mydata.test)

E <- sum(is.na(predictedksvm))
E

table(predictedksvm)
results <- table(predictedksvm, mydata.test$f1.Likelihood_Recommend_H)

nrow(predictedksvm)
totalcorrect = results[1,1]+results[2,2]
totalcorrect = (results[1,1]+results[2,2])/(results[1,1]+results[1,2]+results[2,1]+results[2,2])*100
totalintest = nrow(mydata.test)
totalcorrect/totalintest

table(predictedksvm, mydata.test$f1.Condition_Hotel_H)
table(predictedksvm, mydata.test$f1.Guest_Room_H)
table(predictedksvm, mydata.test$f1.Staff_Cared_H)
table(predictedksvm, mydata.test$f1.Customer_SVC_H)


# check the predicted value
predictedksvm
scatterchartgraphksvm = ggplot(data = mydata.test,aes(x=mydata.test$f1.Guest_Room_H,
```

y=mydata.test$f1.Condition_Hotel_H)) + geom_point(aes(size= predictedksvm,   color= mydata.test$f1.Likelihood_Recommend_H))
#
scatterchartgraphksvm

OUTPUT:

Support Vector Machine object of class "ksvm"

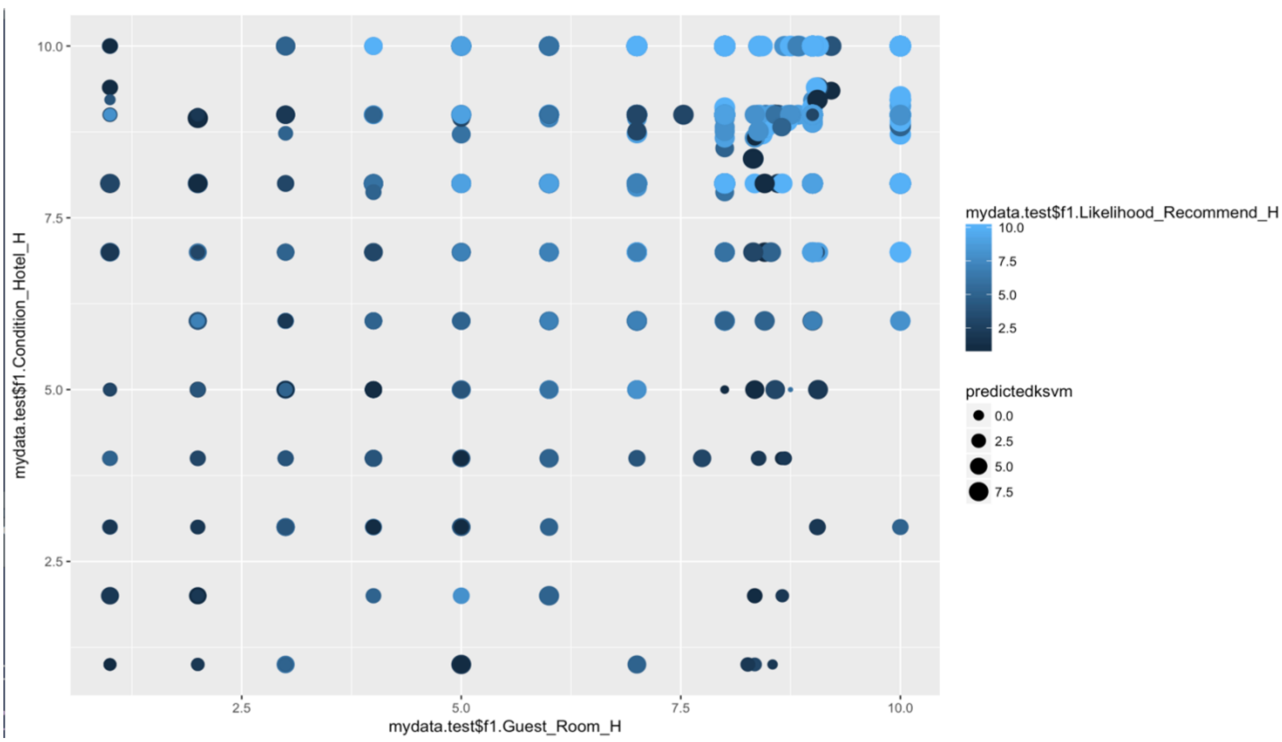
SV type: eps-svr (regression)
parameter : epsilon = 0.1 cost C = 30

Gaussian Radial Basis kernel function.
Hyperparameter : sigma = 1.03934647072841

Number of Support Vectors : 5578

Objective Function Value : -66959.66
Training error : 0.292785
Cross validation error : 1.568224

PLOT:



5.7 Question 7: Which hotel facilities are required to enhance the NPS value of the worst performing Hyatt hotel brand (Hyatt Regency)?

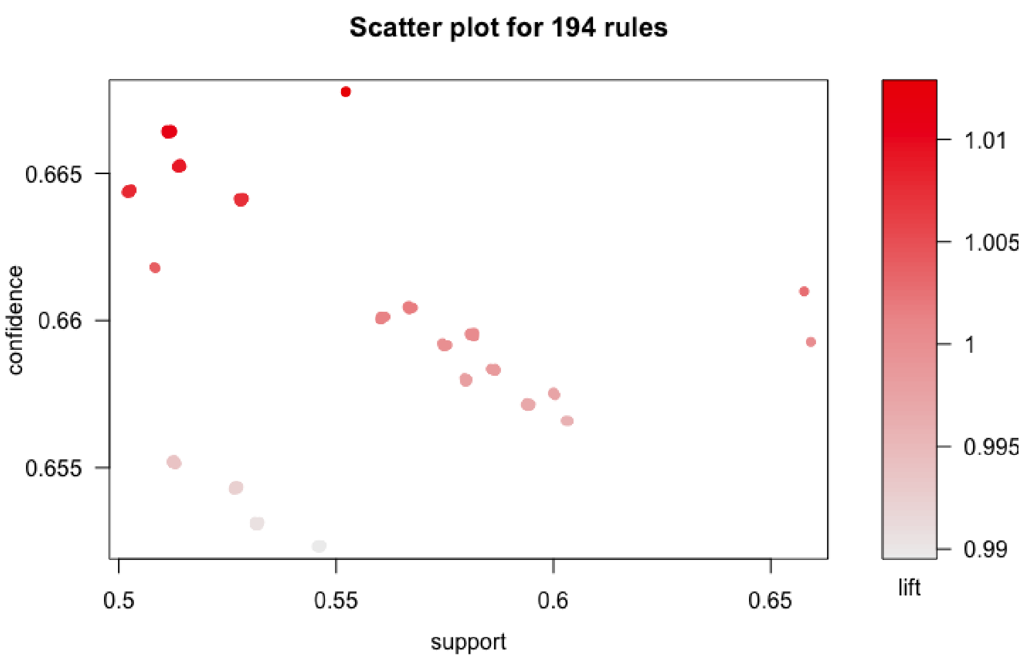
Association rule mining, is a technique for efficiently identifying association rules with good predictive power from a dataset. In theory, any combination of items could form an association rule. We employed several statistical measures to differentiate rules which are likely to be predictive from those which are not. These measures are:

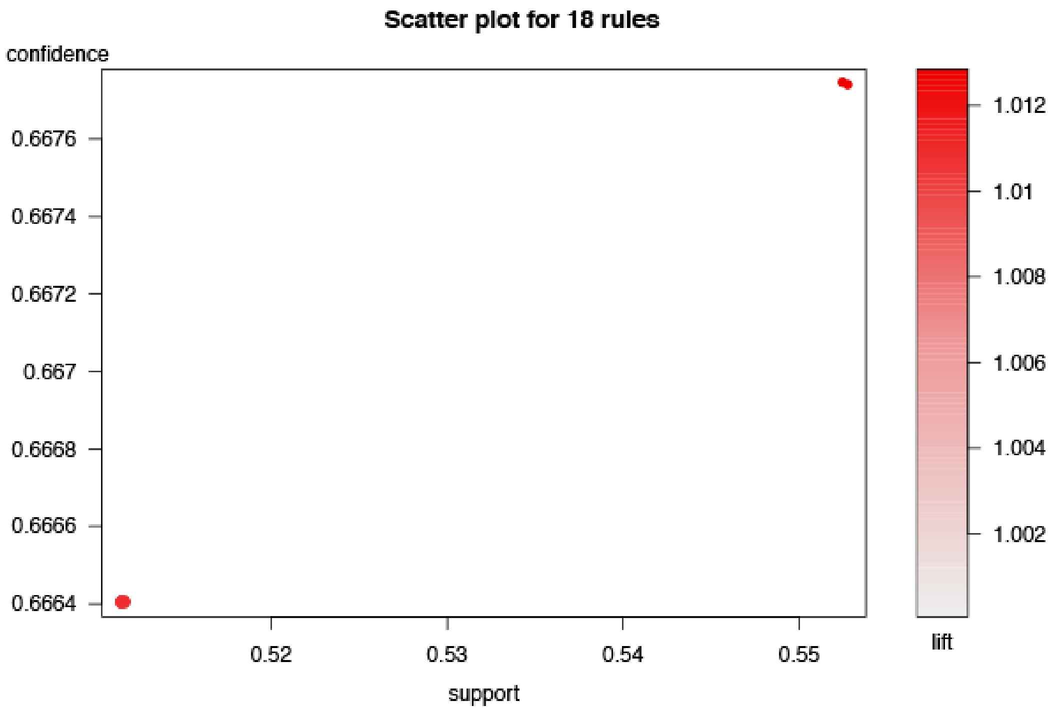
Support: the proportion of times that a pairing occurs across all observations.

Confidence: how frequently a pair occurs among all the times when the first item is present.

Lift: novel pairs.

With "support" and "confidence" set as 0.5 and 0.65 respectively, the Arules modeling returns 194 rules. The number is too large to examine manually. So we used the plot() command of the arulesViz package to generate better insights. We focus on a subset of rules that have lift higher than 1.01.





From the rules which lift value is greater than 1.01, a total of 18 pairs were identified that exceeded our confidence and support threshold. The top 5 rules based on the confidence and support value as listed below.

```
> inspect(sort(goodrules_promoters, by = c('lift','confidence')))
     lhs                         rhs                   support confidence     lift count
[1]  {Convention_PL=Y}        => {NPS_Type=Promoter} 0.5526078  0.6677455 1.012780 45804
[2]  {Convention_PL=Y,
      Restaurant_PL=Y}        => {NPS_Type=Promoter} 0.5526078  0.6677455 1.012780 45804
[3]  {Laundry_PL=Y,
      Valet.Parking_PL=Y}     => {NPS_Type=Promoter} 0.5115881  0.6664047 1.010747 42404
[4]  {Dry.Cleaning_PL=Y,
      Laundry_PL=Y,
      Valet.Parking_PL=Y}     => {NPS_Type=Promoter} 0.5115881  0.6664047 1.010747 42404
[5]  {Fitness.Center_PL=Y,
      Laundry_PL=Y,
      Valet.Parking_PL=Y}     => {NPS_Type=Promoter} 0.5115881  0.6664047 1.010747 42404
[6]  {Bell.Staff_PL=Y,
      Laundry_PL=Y,
      Valet.Parking_PL=Y}     => {NPS_Type=Promoter} 0.5115881  0.6664047 1.010747 42404
[7]  {Laundry_PL=Y,
      Restaurant_PL=Y,
      Valet.Parking_PL=Y}     => {NPS_Type=Promoter} 0.5115881  0.6664047 1.010747 42404
[8]  {Dry.Cleaning_PL=Y,
      Fitness.Center_PL=Y,
      Laundry_PL=Y,
      Valet.Parking_PL=Y}     => {NPS_Type=Promoter} 0.5115881  0.6664047 1.010747 42404
[9]  {Bell.Staff_PL=Y,
      Dry.Cleaning_PL=Y,
      Laundry_PL=Y,
      Valet.Parking_PL=Y}     => {NPS_Type=Promoter} 0.5115881  0.6664047 1.010747 42404
[10] {Dry.Cleaning_PL=Y,
      Laundry_PL=Y,
      Restaurant_PL=Y,
      Valet.Parking_PL=Y}     => {NPS_Type=Promoter} 0.5115881  0.6664047 1.010747 42404
```

Thus, we recommend Hyatt Regency should add or make improvements in amenities like Convention Center, Business Center, Golf, Restaurant, Valet Parking, and Fitness Center.