SYRACUSE UNIVERSITY
SCHOOL OF INFORMATION STUDIES

IST 687 – Applied Data Science

Data Analysis of Hyatt Group of Hotels

GROUP 3
Aiyappa Uthaiah
Amandeep Sachdeva
Chandni Maishery
Dheeraj Pawar
Sriram M Manohar
Yomeena Khan

Introduction

Our team of data analysts worked with the Hyatt group of hotels' data to devise major business questions, identify problems and provide solutions to maximise their revenue generation. The vision of this project was to analyse the bulk of data and work towards identifying key factors for increasing the Net Promoter Score (NPS) for the group of hotels. The Net Promoter Score (NPS) is a means to depict how satisfied a customer is with the facilities provided at the hotel. Analyzing the NPS would provide insights into how likely the customer is to visit again or recommend it to others. Our team had to work with a large amount of structured data that recorded both customer and hotel information for a span of a year. The critical steps that our team employed were:

- **Identifying key factors that influence the NPS**
- **Devising critical business questions**
- **Cleaning the data**
- **Calculating the NPS**
- **Visualizing data graphically**
- **Developing models for predictive analysis**
- **Providing calculated recommendations to improve overall quality of services being provided**

Data Cleaning

The data set comprised of 15,711,552 observations and 236 attributes spanning from February 2014 to January 2015. So our biggest challenge was to optimize the data and clean the dataset for quality analysis. We decided to munge the data and focus our analysis on the first five months data i.e from Feb'14 to Jun'14. We created a data frame for each month with only 59 attributes consisting of all important factors. Individual dataset contained approximately 1 million observations. While exploring the data, we found that the dataset contained too many NA & blank values. As we were focusing our analysis on the **NPS_Type** & **Likelihood_to_recommend** attributes, we decided to eliminate all the rows that included Null values or NA's in these columns. This significantly reduced the data size. Imputing mean/ median in place of NAs and blank values would lead to skewed results hence our decision to eliminate the NA's. Additionally, we decided to focus our analysis on hotels situated in United States and make recommendations specific to the hotels in the US.

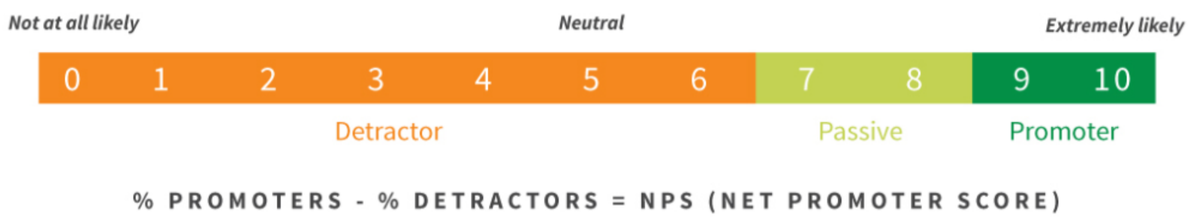
What Is Net Promoter Score?

NPS measures customer experience and predicts business growth. It provides core measurement for customer experience for various businesses around the world.

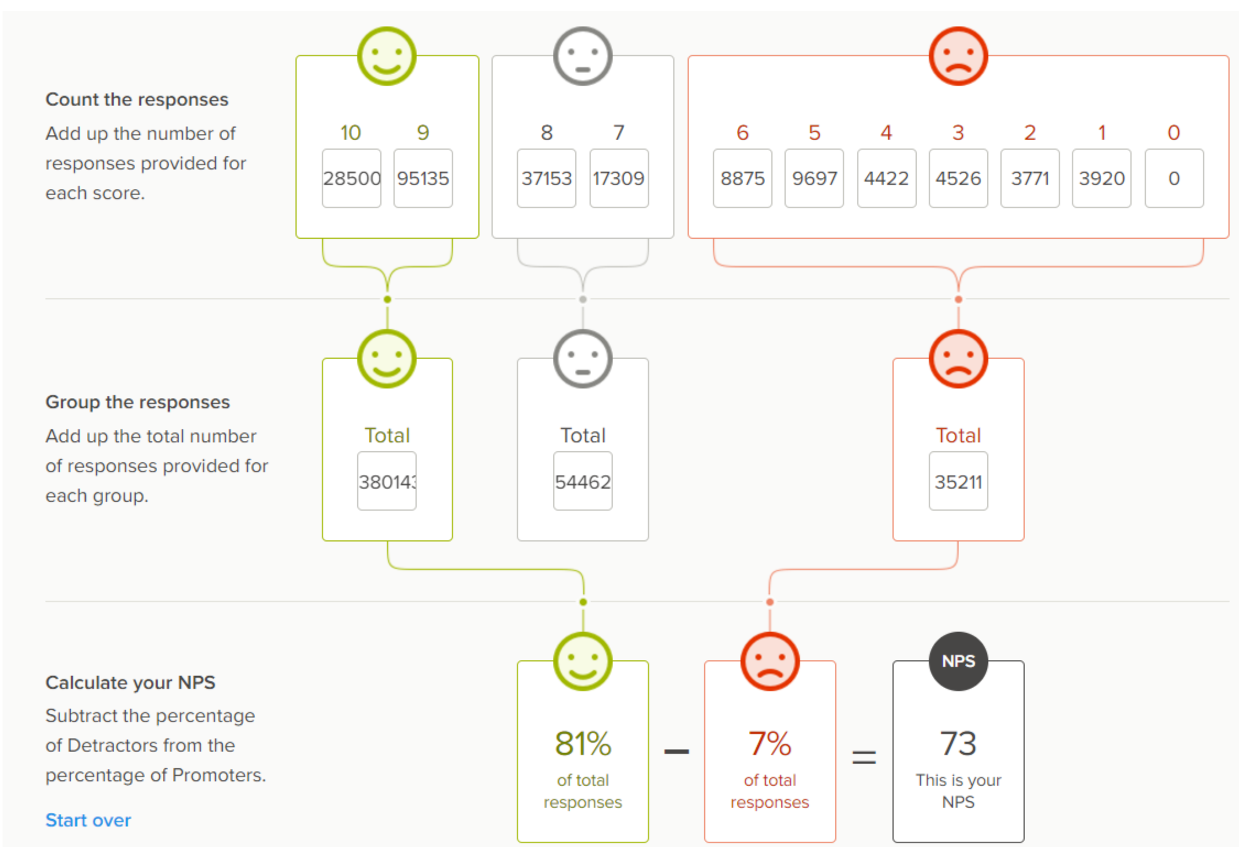
There are three types of responses:

- **Detractors**
- **Passive**
- **Promoters**

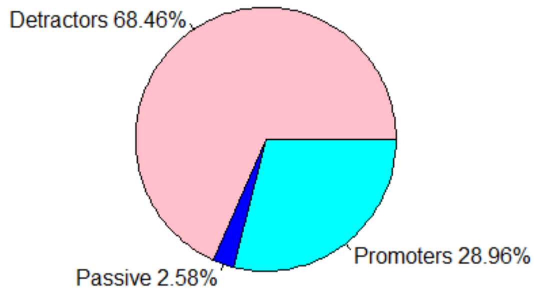
While Promoters have a response of 9 or greater, Passive have a response of 7 or 8 and Detractors have a a response of 6 and lesser. This is pictorially represented below:



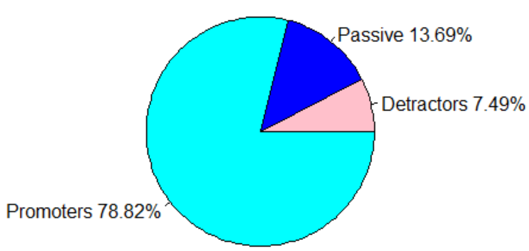
The below figure represents NPS of the dataset we used for our analysis:



For the data given for Feb to June, the overall percentage of detractors exceeds that of promoters. The pie chart shows that around 69% customers were not likely to recommend the Hotels and around 29% were satisfied with the services and would recommend the Hotels.



We reduced the data by eliminating certain columns and then created a pie chart to see the percentage of promoters and detractors. As seen in the pie chart, about 8% of the customers would not recommend Hyatt Group of Hotels which means they were not satisfied with the services provided. Majority of the customers, about 79% were satisfied with the services and were likely to recommend the Hotels.



Business Questions

Identifying the business questions was instrumental in evaluating what customers liked in a hotel, the services customers expected and were provided, services customers expected and were not available and areas of improvement among others. Framing the right business questions provides insights into what hotels can do to improve the NPS. The most critical business questions that our team identified were:

What was the distribution of detractors in each of the states in the United states?

Is there a correlation between the value of NPS and the state in which the hotel is located?

What are the services provided by a hotel that influence the NPS?

What are some interesting facts about the various group of hotels under Hyatt?

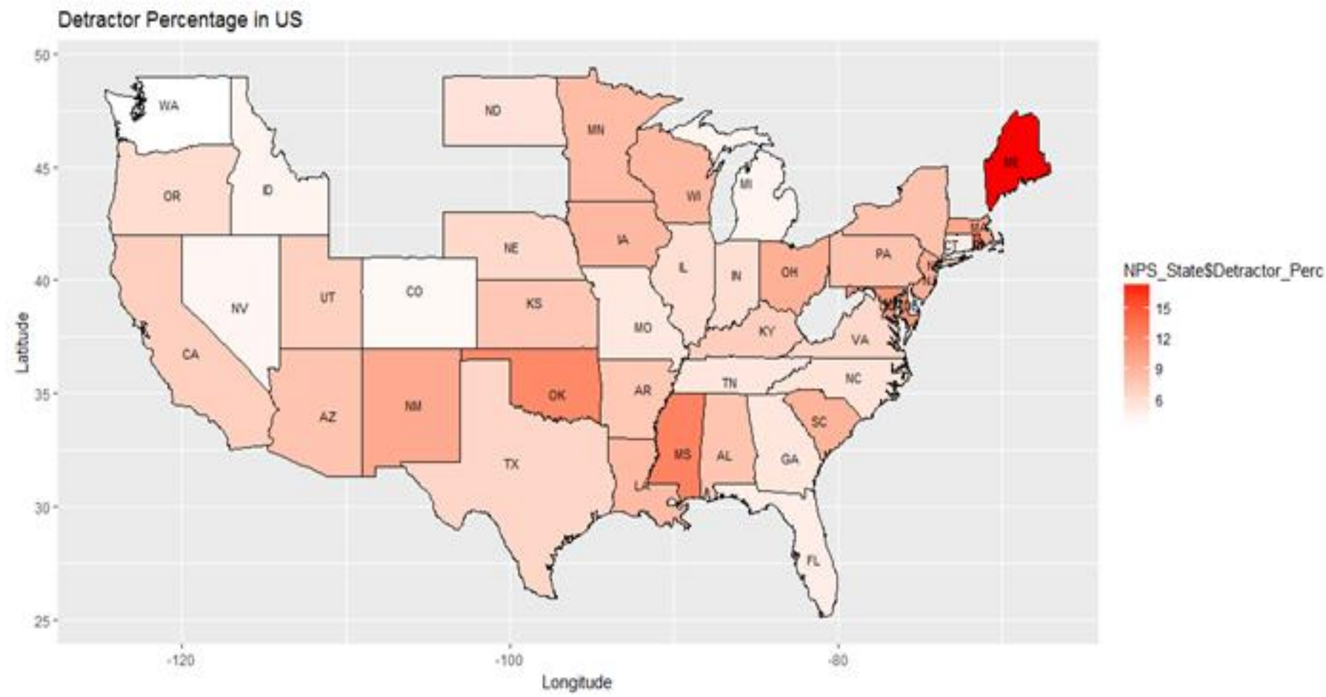
Visualization

To address the business questions, we started our visualizing data by calculating NPS for each state in United States. We plotted it on the US map to understand the stretch of Detractors and Promoters.

Detractor Percentage

Code:

```
us <- map_data("state")
mp_det <- ggplot(NPS_State, aes(map_id =  NPS_State$State_PL))
mp_det <- mp_det + geom_map(map = us, aes(fill = NPS_State$Detractor_Perc), color =
"black", na.rm = TRUE)
mp_det <- mp_det + expand_limits(x = us$long, y = us$lat)
mp_det <- mp_det + geom_text(aes(label=NPS_State$stateabb, x = NPS_State$long, y =
NPS_State$lat), size = 3, color = "black")
mp_det <- mp_det + scale_fill_gradient(low = "white", high = "red", guide = "colorbar")
mp_det <- mp_det + ggtitle("Detractor Percentage in US") +
  labs(x = "Longitude", y = "Latitude")
mp_det
```

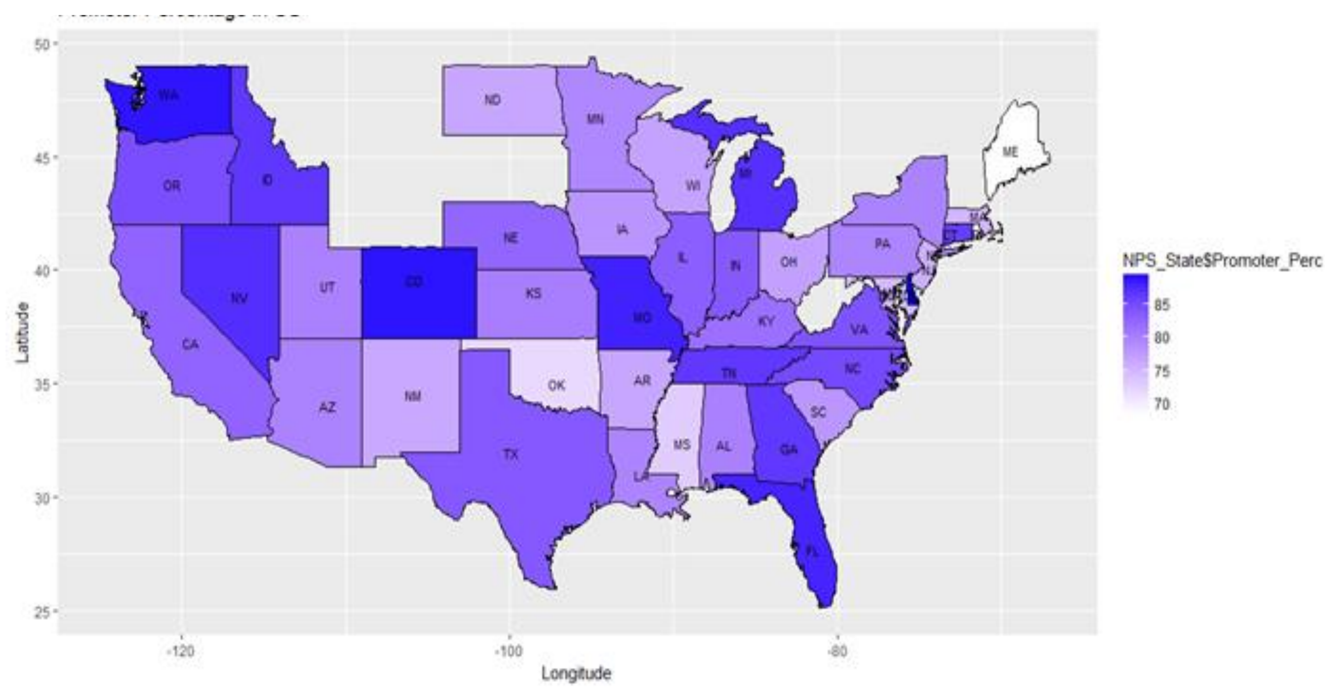


We plotted the percentage of detractors and promoters in US, state-wise to determine which state had to improve their services better to improve their NPS score.
We see that Oklahoma, Mississippi, New Mexico and Maine are the few states with a relatively high detractor percentage.

Promoter Percentage

Code:

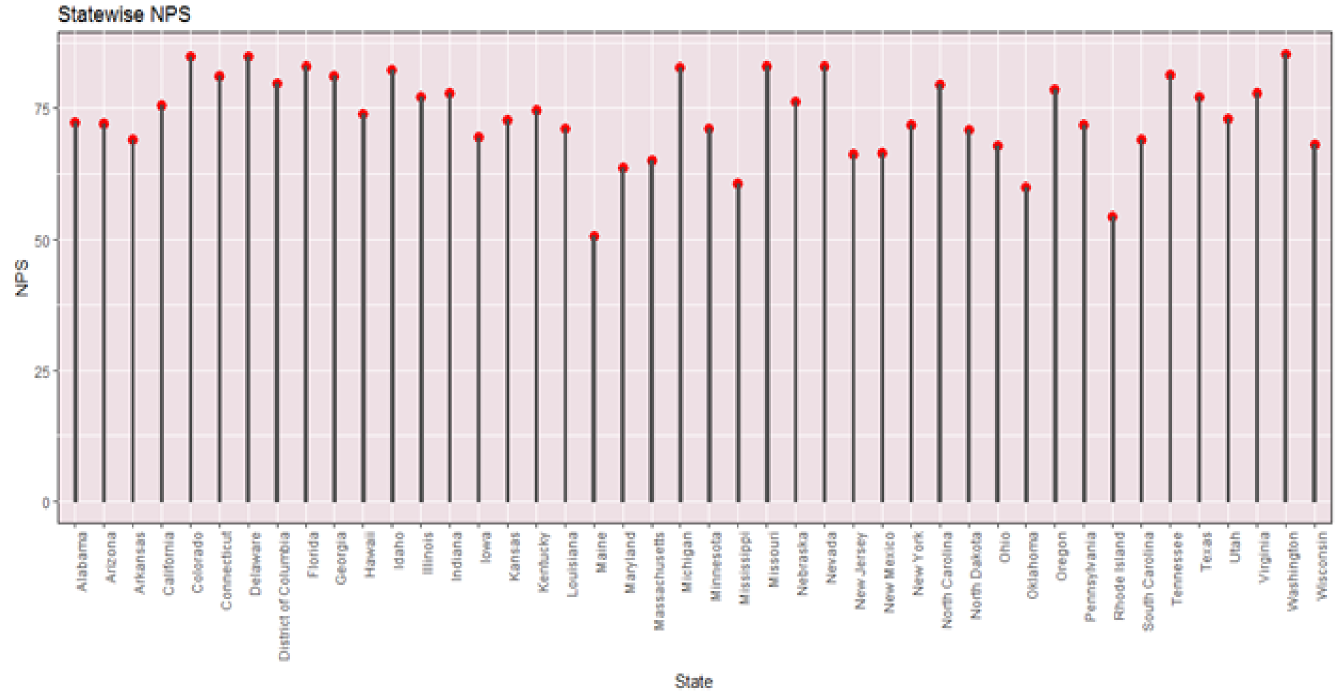
```
mp_prom <- ggplot(NPS_State, aes(map_id =  NPS_State$State_PL))
mp_prom <- mp_prom + geom_map(map = us, aes(fill = NPS_State$Promoter_Perc), color =
"black", na.rm = TRUE)
mp_prom <- mp_prom + expand_limits(x = us$long, y = us$lat)
mp_prom <- mp_prom + geom_text(aes(label=NPS_State$stateabb, x = NPS_State$long, y =
NPS_State$lat), size = 3, color = "black")
mp_prom <- mp_prom + scale_fill_gradient(low = "white", high = "blue", guide = "colorbar")
mp_prom <- mp_prom + ggtitle("Promoter Percentage in US") +
  labs(x = "Longitude", y = "Latitude")
mp_prom
```



From the map, we can see that Washington, Missouri, Colorado, Michigan, Florida are the ones which have high promoter percentage compared to the other states.

We plotted a graph to show individual NPS for the Hyatt group of hotels in each state to precisely understand how each state fared. It is apparent that:

- **The hotels in the state of Washington have the maximum value of NPS. This is an indication that hotels in the state of Washington cater to the needs of the customers**
- **The best followed by Colorado and Delaware**
- **Maine has lowest NPS of 50.23**
- **The average NPS of United States is 73.69** (which can be seen figure 2)

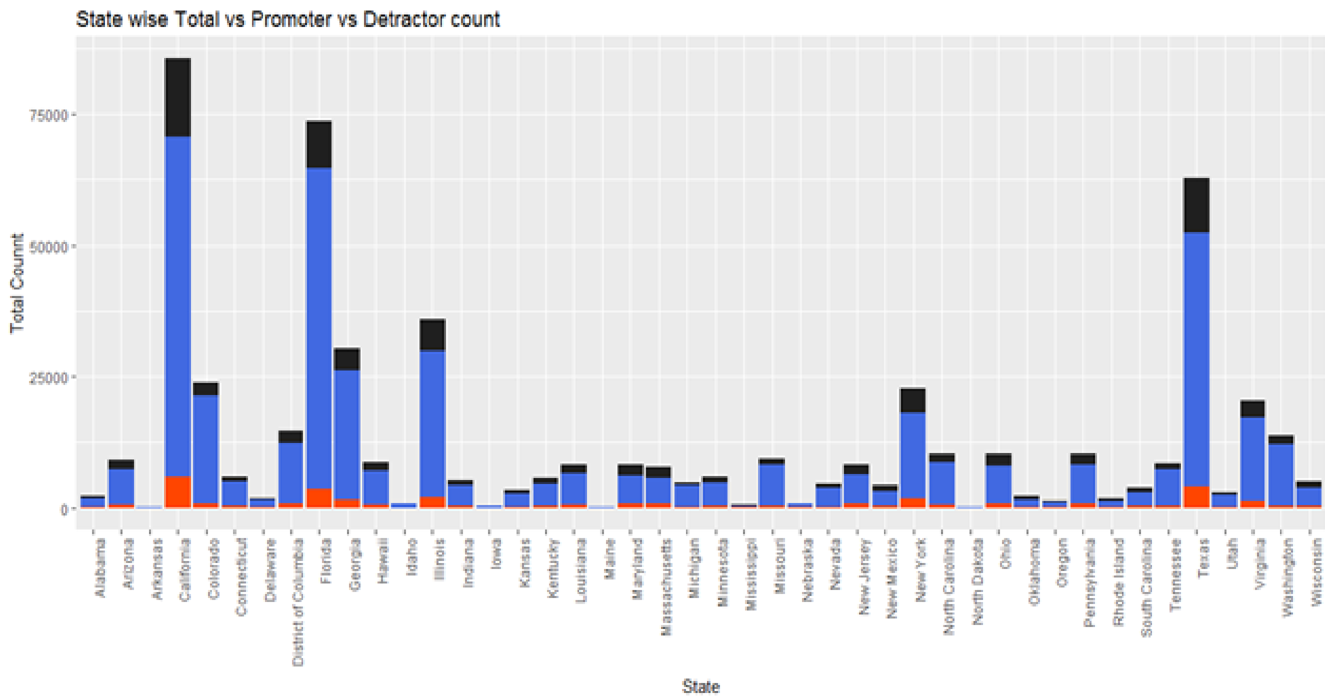


Code:

```
plot_nps1 <- ggplot(NPS_5, aes(x= State_PL))
plot_nps1 <- plot_nps1 + geom_point(aes(y = NPS_5$NPS), size = 3, color = "red1")
plot_nps1 <- plot_nps1 + geom_col(aes(y = NPS_5$NPS), width = .1, color = "gray26")
plot_nps1 <- plot_nps1 + theme(axis.text.x = element_text(angle = 90,hjust = 1))
plot_nps1 <- plot_nps1 + ggtitle("Statewise NPS") +
  labs(x = "State", y = "NPS")
plot_nps1 <- plot_nps1 + theme(panel.background = element_rect(fill = 'lavenderblush2', colour
= 'black'))
plot_nps1
```

The plot is a combination of the total number of responses (black + blue + orange), the total number of promoters (blue + orange) and the total number of detractors (orange). The graph gives us a visual representation of the proportion of promoters and detractors in comparison to the total reviews.

California, **Florida** and **Texas** have the most number of responses.
Arkansas, **Maine** and **North Dakota** have the least number of responses.



Code:

```
plot1 <- ggplot(NPS_Dummy5, aes(x= State_PL))
plot1 <- plot1 + geom_col(aes(y = NPS_Dummy5$Total_Voters), fill = "gray12")
plot1 <- plot1 + geom_col(aes(y = NPS_Dummy5$No_of_promoters), fill = "royalblue")
plot1 <- plot1 + geom_col(aes(y = NPS_Dummy5$No_of_detractors), fill = "orangered")
plot1 <- plot1 + ggtitle("State wise Total vs Promoter vs Detractor count") +
          labs(x = "State", y = "Total Count")
plot1 <- plot1 + theme(axis.text.x = element_text(angle = 90, hjust = 1))
Plot1
```

Favorable Brand

When we analysed the dataset, we found that the Hyatt hotels had 8 brands.

- Andaz
- Grand Hyatt
- Hyatt
- Hyatt House
- Hyatt Place
- Hyatt Regency
- Hyatt Ziva
- Park Hyatt

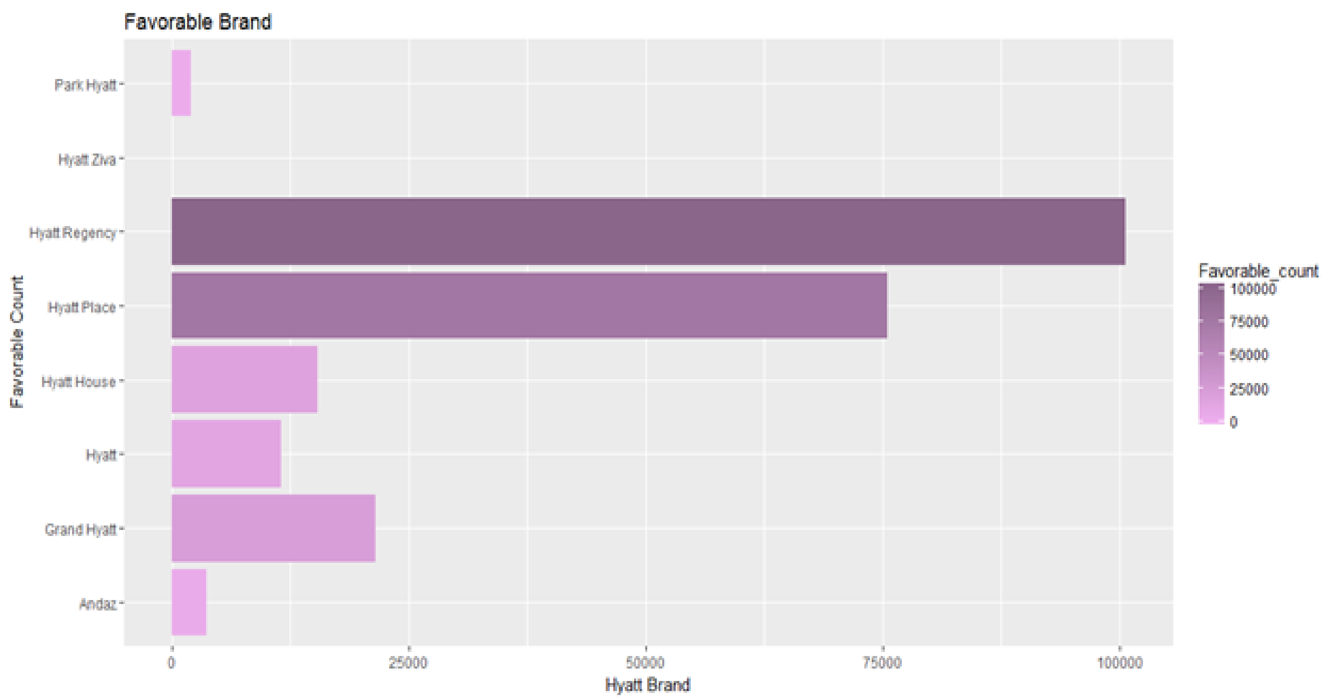
We wanted to analyse this aspect a bit more in terms of how much did these brands contributed to the overall count. We used these brands to understand what percentage of promoters, detractors or passive customers are inclined towards which kind of brand. This analysis narrowed down to regions so as to keep the classification less complicated to regional level. Here we have used the column BRAND_PL so as to understand its relationship with the column REGION_PL.

The bar graph shows the count of people who voted for each brand of Hyatt hotel in US. It can be seen that there is vast difference between individual brands.

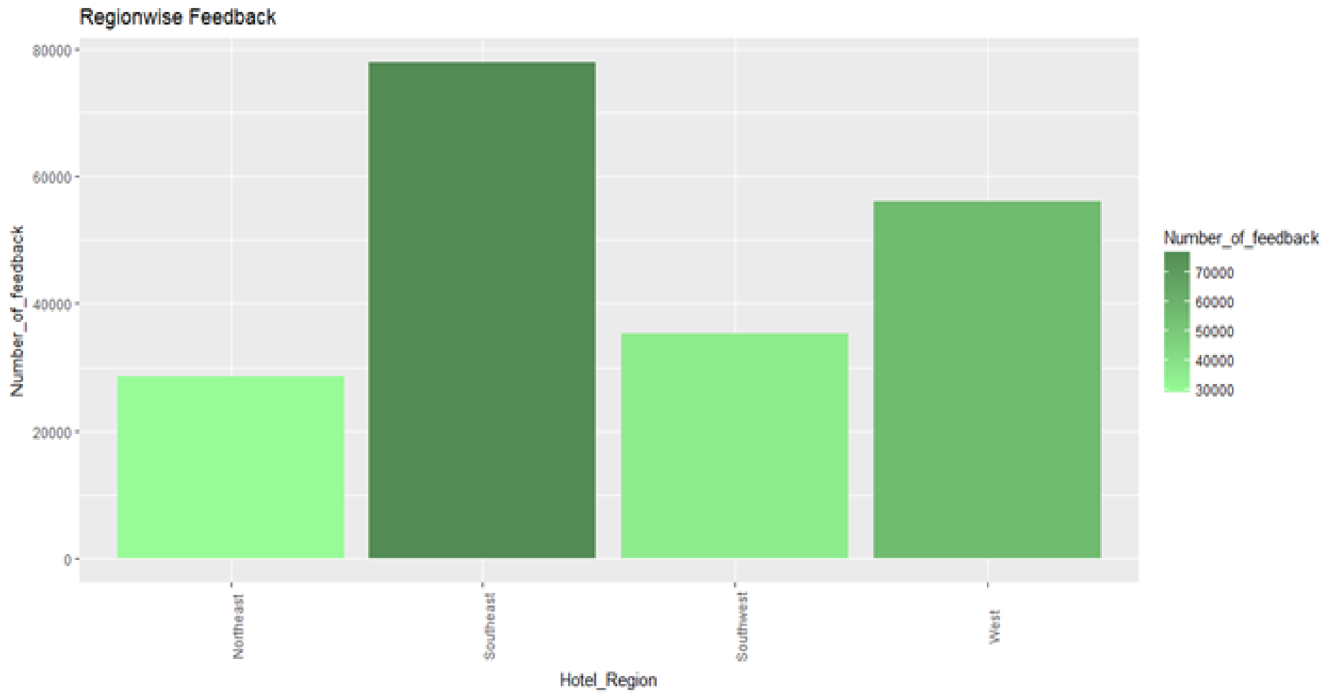
- **Hyatt Regency has maximum favorable count of 100000**
- Followed by Hyatt Place (almost one-fourth less of Hyatt Regency).
- Hyatt Ziva has no favorable count (more analysis is required to understand impact on this brand)

Code:

```
gnew <- ggplot(data=favbrand, aes(x=Hotel_Brand, y=Favorable_count, fill=Favorable_count))
gnew <- gnew + geom_bar(stat="identity")
gnew <- gnew + scale_fill_gradient(low = "plum2", high = "plum4", guide = "colorbar") +
coord_flip()
gnew <- gnew + ggtitle("Favorable Brand") +
  labs(x = "Favorable Count", y = "Hyatt Brand")
gnew
```



The above analysis shows the Hyatt Regency has most number of favorable count and Park Hyatt has the least. When we narrowed the analysis to different regions in US, we found that the feedback for Hyatt Regency was maximum in Southeast region.

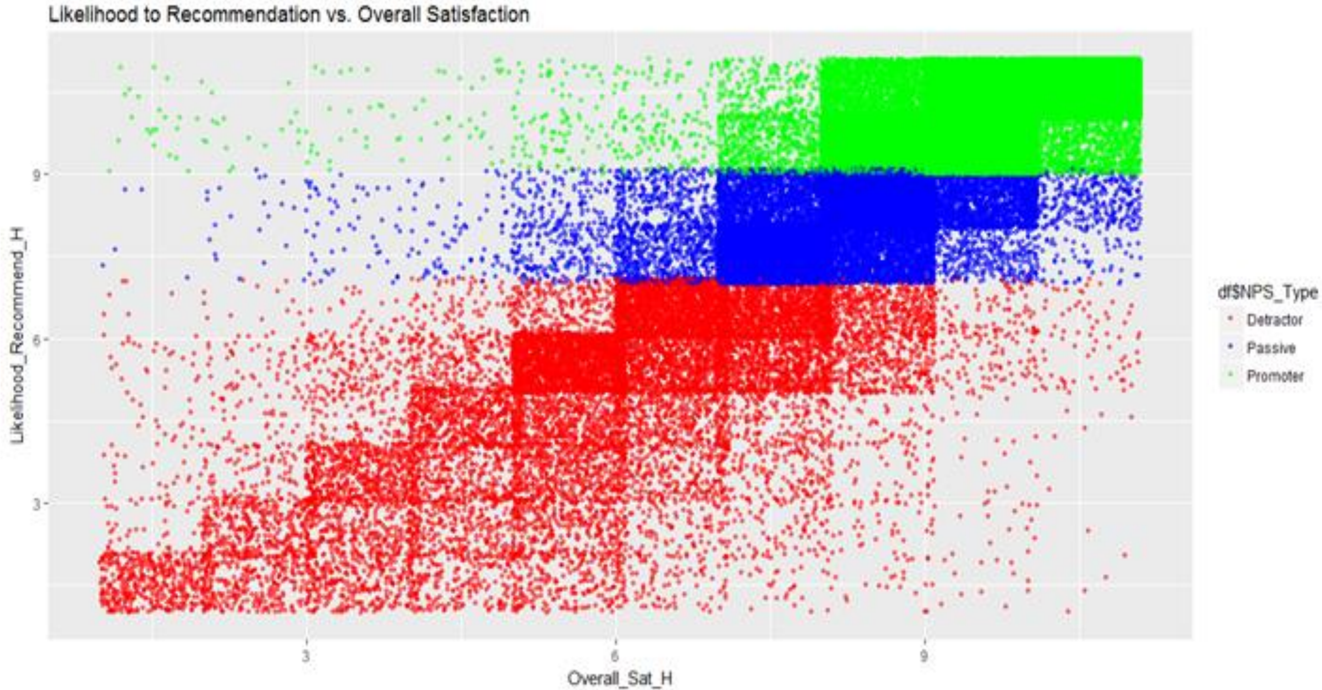


After selecting the brand that we need to focus on, we compared that with number of factors to find obvious, but verified findings.

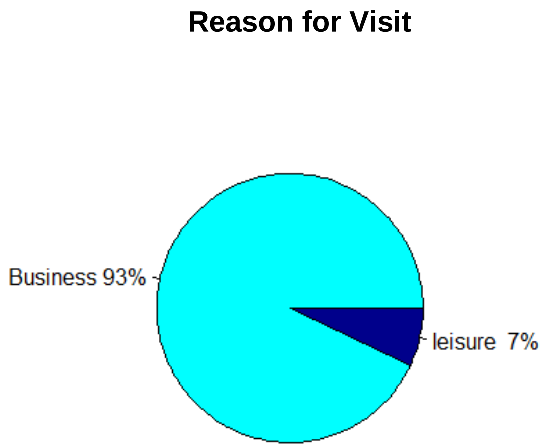
Likelihood to Recommend Vs Overall satisfaction

```
g <- ggplot(data=df, aes(x=Overall_Sat_H))
g <- g + geom_point(aes(y=Likelihood_Recommend_H, color= df$NPS_Type), size = 1, alpha =
0.5)
g <- g + scale_color_manual(breaks = c("Detractor", "Passive", "Promoter"),
                    values=c("red", "blue", "green"))
g <- g + ggtitle("Likelihood to Recommendation vs. Overall Satisfaction")
g
```

Based on the given data, for region wise analysis we figured, the main factor to calculate NPS is Likelihood to Recommendation. From the graph we can see that it is high then people are likely to recommend the hotel. Here overall satisfaction is an independent variable.



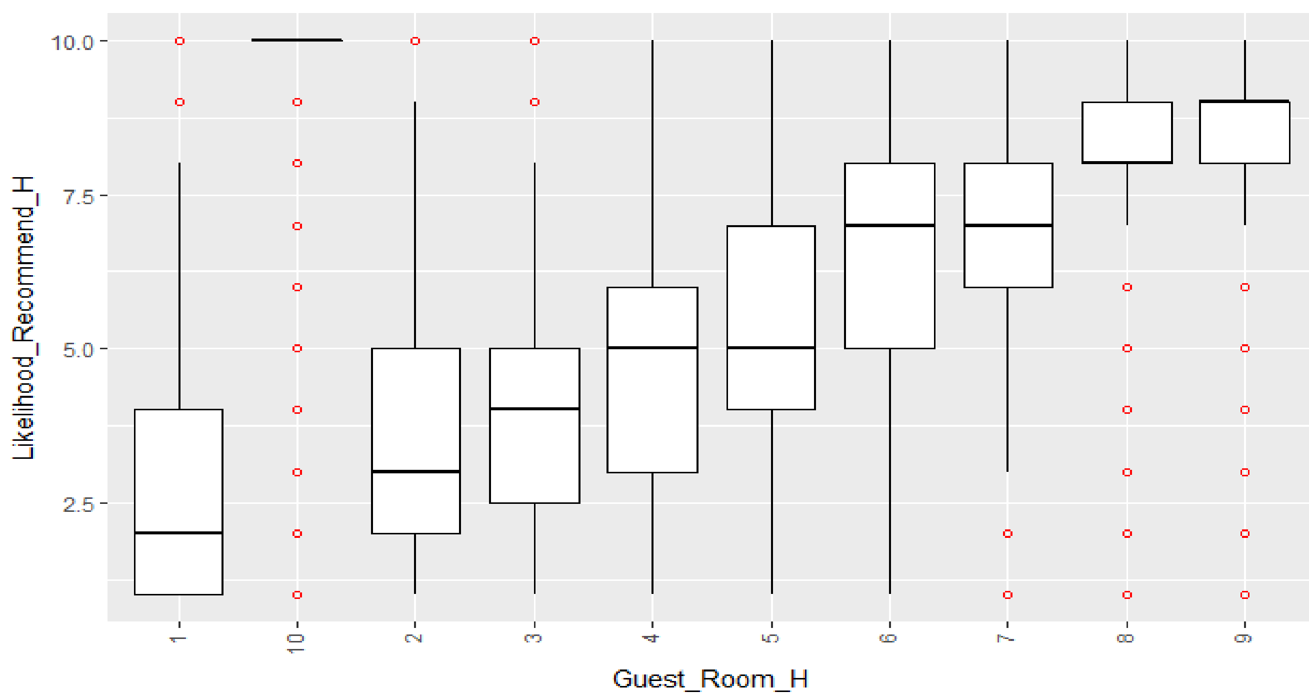
We used a pie chart to divide the reason of visit of customers: Business and Leisure and find out which one has a higher percentage. The percentage of people visiting for business is 93% and for leisure is 7%. We have further segregated the reasons of visit and visualized them using bar graphs to show the customers whose reason for visit is a combination of business and leisure and customers who did not wish to answer.




```
slices <- c(93,7)
lbls<- c( "Business", "leisure ")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
```

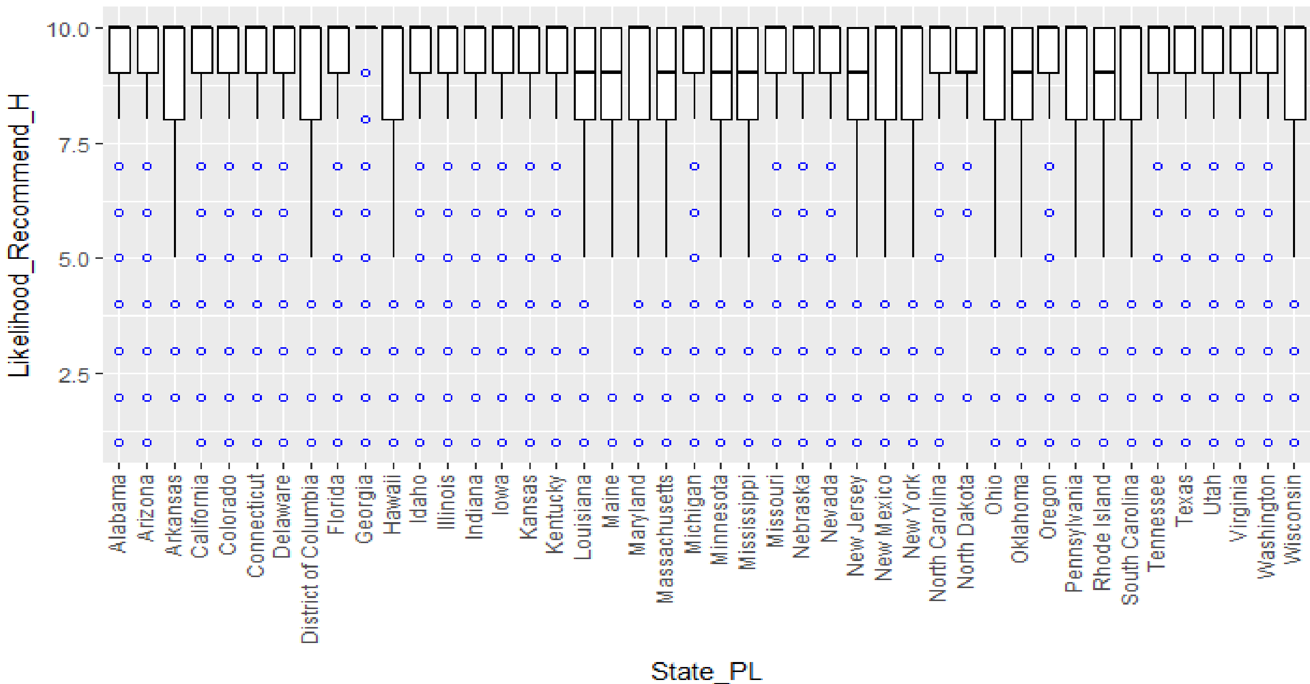
colors1<- c("cyan", "darkblue")
pie(slices,labels = lbls, col=colors1, main=" reason for visit ")

We used box plots to show the relationship between type of guest room and likelihood of recommendation. Guests who rated the room between 1-7 were less likely to promote the Hotels, while guests who gave a rating of 8 or above to the rooms, were featured to be promoters of the Hotels.



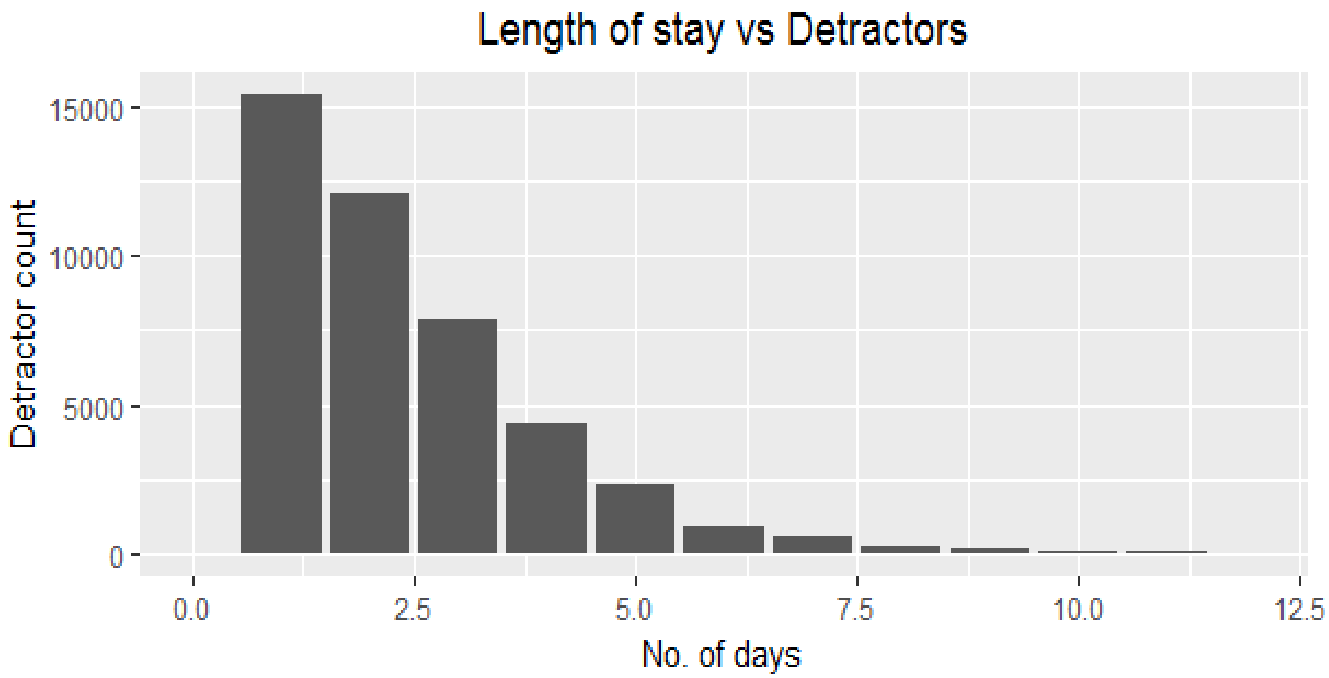
datanew <- WorkingDS2
datanew$Likelihood_Recommend_H <-as.numeric(datanew$Likelihood_Recommend_H)
p <- ggplot(datanew, aes(Guest_Room_H, Likelihood_Recommend_H))
p + geom_boxplot(outlier.colour = "red", outlier.shape = 1,fill = "white", colour = "black")
+theme(axis.text.x = element_text(angle=90, hjust=1,vjust=0.3))

The box plot for the relationship between the State in which the Hotel is located and Likelihood of Recommendation shows that most of the states have a mean of 10 for likelihood of recommendation. Georgia has the maximum number of outliers. States like Louisiana, Maine, Mississippi, Rhode Island, Minnesota can work on their services to increase the likelihood of recommendation.



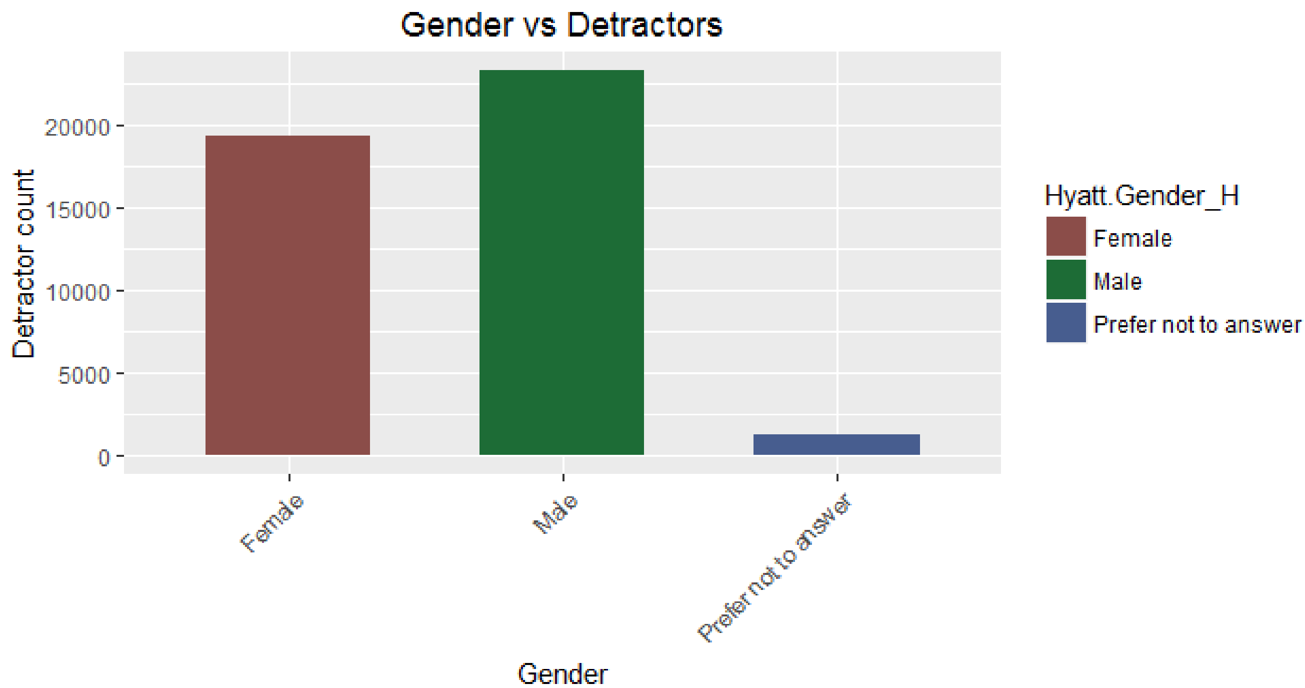

```
datanew 1<- WorkingDS2
datanew1$Likelihood_Recommend_H <-as.numeric(datanew1$Likelihood_Recommend_H)
p <- ggplot(datanew1, aes(State_PL, Likelihood_Recommend_H))
p + geom_boxplot(outlier.colour = "blue", outlier.shape = 1,fill = "white", colour = "black")
+theme(axis.text.x = element_text(angle=90, hjust=1,vjust=0.3))
```

The plot between the number of days a customer stays at a hotel versus the total count of detractors gives us an indication that the detractor responses are comparatively higher when people stay for a shorter amount of duration. In order to reduce the detractor count, the hotels must probably focus on the providing the best possible services to the guests making stay for less than 3 days. Also, hotels must provide attractive offers so that customers reside for a longer time and they can be more satisfied by utilizing the available services to the fullest extent.



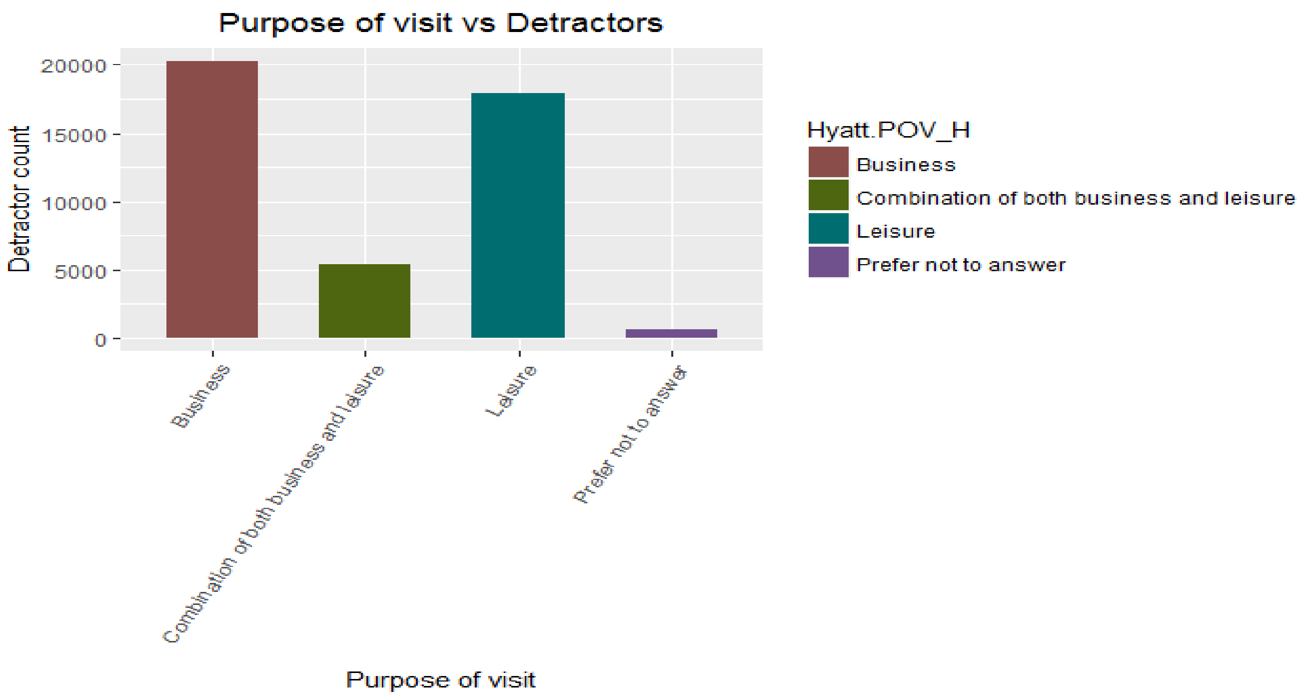

```
NPS_Length_of_stay_Plot <-
ggplot(subset(df.cat[df.cat$Hyatt.LENGTH_OF_STAY_C!='',],Hyatt.NPS_Type=="Detractor"),
aes(x= Hyatt.LENGTH_OF_STAY_C)) + geom_bar() + ggtitle("Length of stay vs Detractors") +
labs(x="No. of days", y="Detractor count") + theme(plot.title = element_text(hjust = 0.5)) +
xlim(0, 12)
NPS_Length_of_stay_Plot
```

The bar graph of Gender versus Detractor count shows that ‘Males’ are likely to give poor rating compared to ‘Females’. In order to change this trend, hotel management can focus more on the services that are more closely related to men.



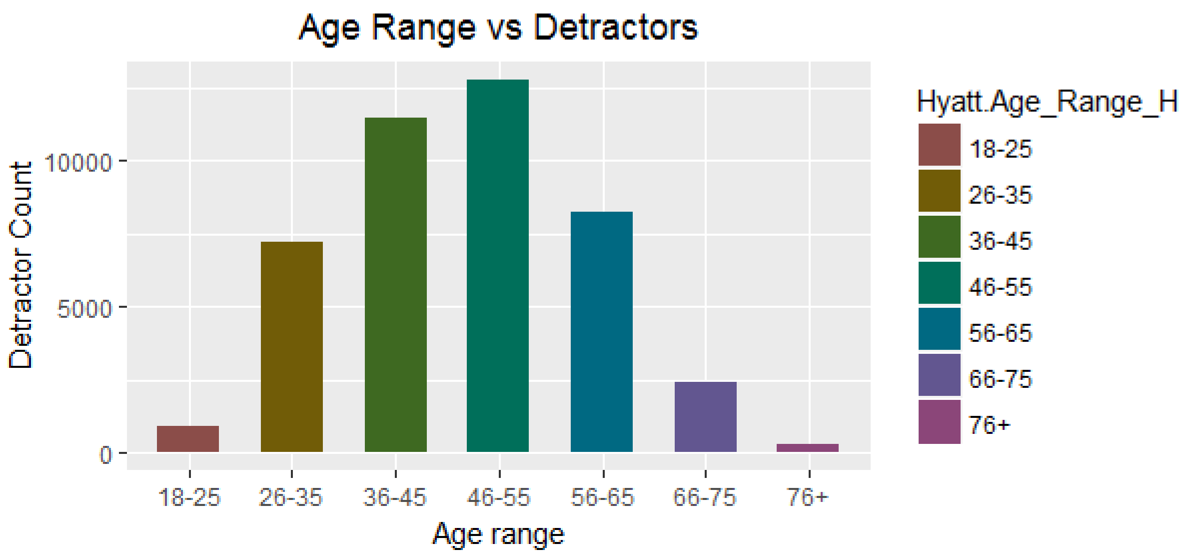

```
NPS_Gender_Plot <-
ggplot(subset(df.cat[df.cat$Hyatt.Gender_H!='',],Hyatt.NPS_Type=="Detractor"), aes(x=
Hyatt.Gender_H, fill= Hyatt.Gender_H))
NPS_Gender_Plot <- NPS_Gender_Plot + geom_bar(position = "dodge", width = 0.6) +
ggtitle("Gender vs Detractors") + labs(x="Gender", y="Detractor count") +
scale_fill_hue(c=45,l=40)+ theme(plot.title = element_text(hjust = 0.5), axis.text.x =
element_text(angle = 45, hjust = 1))
NPS_Gender_Plot
```

When we plotted a bar plot for the purpose of visit against the detractor count, it was observed that the people visiting hotels with ‘Business’ as their primary purpose were mostly responsible for detractor ratings followed by purpose as ‘Leisure’ and then combination of both.’ The probable suggestion for this problem would be to inculcate facilities such as meetings, conference room and other similar options which are primarily used by corporate customers. This way, the visiting guest ratings might be in more favor as promoters.




```
NPS_POV_H_Plot <-
ggplot(subset(df.cat[df.cat$Hyatt.POV_H!='',],Hyatt.NPS_Type=="Detractor"), aes(x=
Hyatt.POV_H,fill=Hyatt.POV_H)) + geom_bar(position="dodge", width = 0.6) + ggtitle("Purpose
of visit vs Detractors") + labs(x="Purpose of visit", y="Detractor count")
+scale_fill_hue(c=45,l=40)+ theme(plot.title = element_text(hjust = 0.5), axis.text.x =
element_text(angle = 60, hjust = 1))
NPS_POV_H_Plot
```

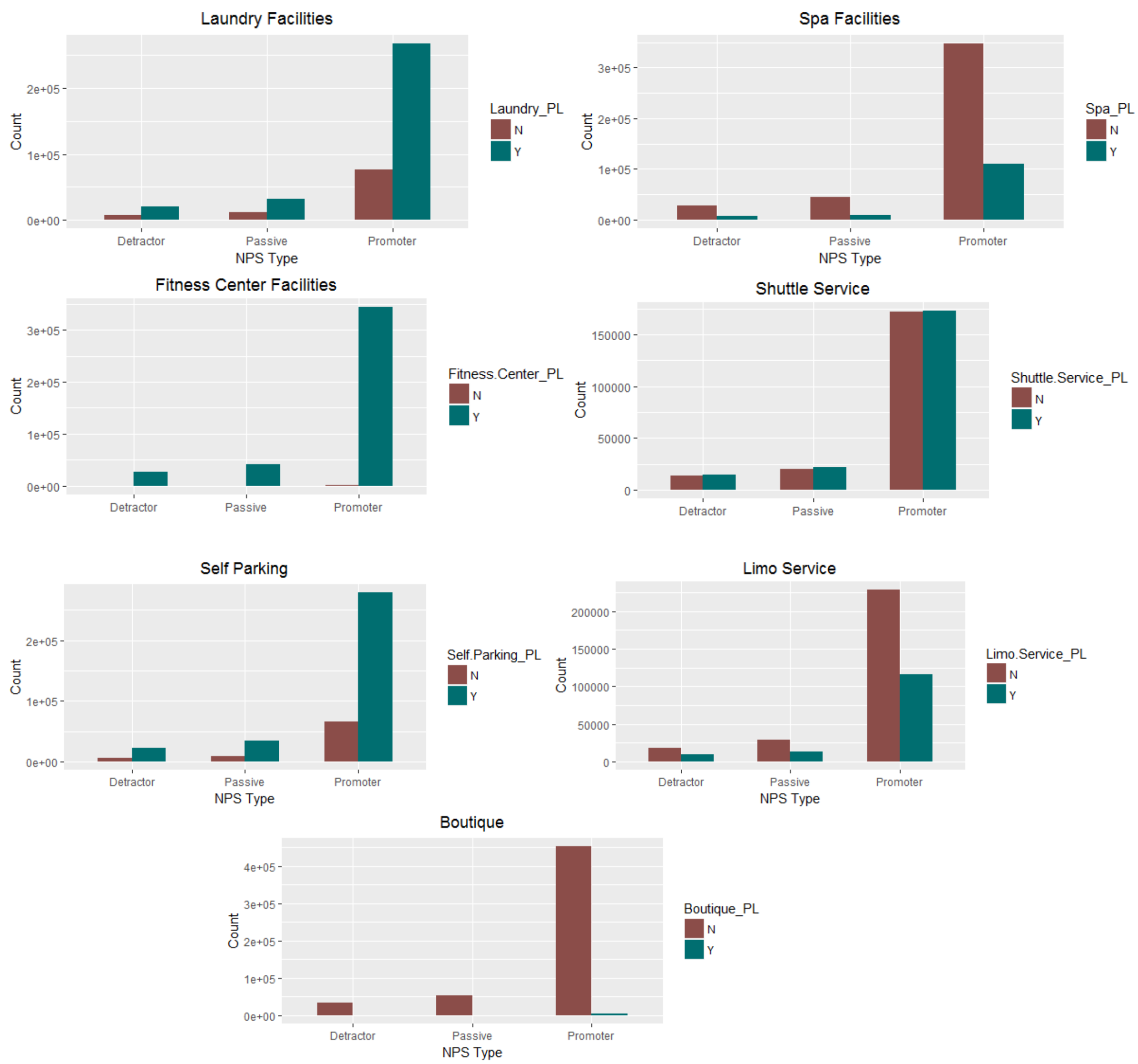
The plot which shows the split up of number of detractors according to the age of customers, gives us an indication that guests between the ages of 36 and 55 are mostly dissatisfied and more likely to be under the category of detractor ratings. The hotel facilities which are more commonly used by people belonging to these age groups can be stressed upon so that the overall ratings can be improved and detractors proportion can be reduced.




```
NPS_Age_Range_H_Plot <-
ggplot(subset(df.cat[df.cat$Hyatt.Age_Range_H!='',],Hyatt.NPS_Type=="Detractor"), aes(x=
Hyatt.Age_Range_H,fill=Hyatt.Age_Range_H)) + geom_bar(position="dodge", width = 0.6) +
ggtitle("Age Range vs Detractors") + labs(x="Age range", y="Detractor Count")
+scale_fill_hue(c=45,l=40)+ theme(plot.title = element_text(hjust = 0.5))
NPS_Age_Range_H_Plot
```

Hotel Facilities:

To know the extent to which a particular hotel facility can contribute to the promoter or detractor NPS type, we plotted bar charts for the most recorded responses facilities. Based on the observations, we can see that the location where laundry facilities are present affects the overall detractor score indirectly, since the promoter count is more where the hotel provides laundry amenities. However, when we view the graph in case of ‘Spa service’, the promoter count is high in spite of the service not being present at the hotels. Thus, we can make a statement that spa facility does not comparatively affect the overall detractor score and hotels need not focus more on it. Similar conclusion can be observed in the case of ‘Fitness center’, ‘Shuttle service’, ‘Boutique’ and ‘Limo service’ facilities.




```
#Laundry_PL
NPS_Laundry_Plot <-ggplot(Hyatt_Feb2Jun[Hyatt_Feb2Jun$Laundry_PL!='',],
aes(NPS_Type,fill=Laundry_PL)) + geom_bar(position="dodge", width = 0.6) + ggtitle("Laundry
Facilities") + scale_fill_hue(c=45,l=40) + labs(x="NPS Type", y="Count") + theme(plot.title =
element_text(hjust = 0.5))
NPS_Laundry_Plot

#Spa_Plot
NPS_Spa_Plot <-ggplot(Hyatt_Feb2Jun[Hyatt_Feb2Jun$Spa_PL!='',],
aes(NPS_Type,fill=Spa_PL )) + geom_bar(position="dodge", width = 0.6) + ggtitle("Spa
Facilities")+ scale_fill_hue(c=45,l=40) + labs(x="NPS Type", y="Count") + theme(plot.title =
element_text(hjust = 0.5))
NPS_Spa_Plot
```

```
#Fitness.Center_PL
NPS_Fitness.Center_PL_Plot <-ggplot(Hyatt_Feb2Jun[Hyatt_Feb2Jun$Fitness.Center_PL!='',],
aes(NPS_Type,fill=Fitness.Center_PL)) + geom_bar(position="dodge", width = 0.6) +
ggtitle("Fitness Center Facilities")+ scale_fill_hue(c=45,l=40) + labs(x="NPS Type", y="Count")
+ theme(plot.title = element_text(hjust = 0.5))
NPS_Fitness.Center_PL_Plot

#Shuttle Services
NPS_Shuttle.Service_PL_Plot <-
ggplot(Hyatt_Feb2Jun[Hyatt_Feb2Jun$Shuttle.Service_PL!='',],
aes(NPS_Type,fill=Shuttle.Service_PL )) + geom_bar(position="dodge", width = 0.6) +
ggtitle("Shuttle Service")+ scale_fill_hue(c=45,l=40) + labs(x="NPS Type", y="Count") +
theme(plot.title = element_text(hjust = 0.5))
NPS_Shuttle.Service_PL_Plot

#Self Parking
NPS_Self.Parking_PL_Plot <-ggplot(Hyatt_Feb2Jun[Hyatt_Feb2Jun$Self.Parking_PL!='',],
aes(NPS_Type,fill=Self.Parking_PL )) + geom_bar(position="dodge", width = 0.6) + ggtitle("Self
Parking")+ scale_fill_hue(c=45,l=40) + labs(x="NPS Type", y="Count") + theme(plot.title =
element_text(hjust = 0.5))
NPS_Self.Parking_PL_Plot

#Limo Service
NPS_Limo.Service_PL_Plot <-ggplot(Hyatt_Feb2Jun[Hyatt_Feb2Jun$Limo.Service_PL!='',],
aes(NPS_Type,fill=Limo.Service_PL )) + geom_bar(position="dodge", width = 0.6) +
ggtitle("Limo Service")+ scale_fill_hue(c=45,l=40) + labs(x="NPS Type", y="Count") +
theme(plot.title = element_text(hjust = 0.5))
NPS_Limo.Service_PL_Plot


#Boutique
NPS_Boutique_PL_Plot <-ggplot(Hyatt_Feb2Jun[Hyatt_Feb2Jun$Boutique_PL!='',],
aes(NPS_Type,fill=Boutique_PL )) + geom_bar(position="dodge", width = 0.6) +
ggtitle("Boutique")+ scale_fill_hue(c=45,l=40) + labs(x="NPS Type", y="Count") +
theme(plot.title = element_text(hjust = 0.5))
NPS_Boutique_PL_Plot
```

Linear Modelling

With our data of 5 months for the region of United States, we conducted Linear Modelling on hotels where the customers have rated Hyatt hotels on seven factors:

- Guest room satisfaction metric
- Tranquility metric
- Condition of hotel metric
- Quality of customer service metric
- Staff cared metric
- Internet satisfaction metric
- Quality of check-in process

Likelihood to recommend is the dependent variable and the above factors are independent variables. We ran Linear modelling individually for each factor and then for all the factors together. The results are below:

Independent Variable	Dependent variable	R Squared Value
Likelihood to Recommend	Guest room satisfaction metric	0.6064
Likelihood to Recommend	Tranquility metric	0.4219
Likelihood to Recommend	Condition of hotel metric	0.5775
Likelihood to Recommend	Quality of customer service metric	0.567

Likelihood to Recommend	Staff cared metric	0.4646
Likelihood to Recommend	Internet satisfaction metric	0.158
Likelihood to Recommend	Quality of check-in process	0.2776
Likelihood to Recommend	· Guest room satisfaction metric · Tranquility metric · Condition of hotel metric · Quality of customer service metric · Staff cared metric · Internet satisfaction metric · Quality of check-in process	0.6982

We see that R squared is highest when we consider all the factors compared to when we consider each factor individually.

```
> lmall <- lm(formula = Likelihood_Recommend_H ~ Guest_Room_H + Tranquility_H + Condition_Hotel_H
+             + Customer_SVC_H + Staff_Cared_H + Internet_Sat_H + Check_In_H, data = Hyatt_Feb2Jun)
> summary(lmall)

Call:
lm(formula = Likelihood_Recommend_H ~ Guest_Room_H + Tranquility_H +
    Condition_Hotel_H + Customer_SVC_H + Staff_Cared_H + Internet_Sat_H +
    Check_In_H, data = Hyatt_Feb2Jun)

Residuals:
    Min      1Q  Median      3Q     Max
-8.9787  0.0213  0.0213  0.1858  6.4976

Coefficients:
                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)       -2.024351   0.019682 -102.854   <2e-16 ***
Guest_Room_H       0.320398   0.002681  119.524   <2e-16 ***
Tranquility_H      0.126699   0.001858   68.191   <2e-16 ***
Condition_Hotel_H  0.201981   0.003012   67.050   <2e-16 ***
Customer_SVC_H     0.369465   0.003616  102.164   <2e-16 ***
Staff_Cared_H      0.136593   0.003018   45.266   <2e-16 ***
Internet_Sat_H     0.041119   0.001205   34.128   <2e-16 ***
Check_In_H         0.004050   0.002335    1.734   0.0829 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8413 on 180940 degrees of freedom
  (366218 observations deleted due to missingness)
Multiple R-squared:  0.6982,    Adjusted R-squared:  0.6982
F-statistic: 5.979e+04 on 7 and 180940 DF,  p-value: < 2.2e-16
```

The ‘significance codes’ associated to ‘Quality of Check-in process’ was pretty low so we ran the model again without ‘Quality of Check-in process’. The results are as below:

```
Call:
lm(formula = Likelihood_Recommend_H ~ Guest_Room_H + Tranquility_H +
    Condition_Hotel_H + Customer_SVC_H + Staff_Cared_H + Internet_Sat_H,
    data = Hyatt_Feb2Jun)

Residuals:
    Min      1Q  Median      3Q     Max
-8.9776  0.0224  0.0224  0.1858  7.4902

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       -2.033465   0.018474 -110.07   <2e-16 ***
Guest_Room_H       0.319925   0.002682  119.27   <2e-16 ***
Tranquility_H      0.126294   0.001861   67.88   <2e-16 ***
Condition_Hotel_H  0.200213   0.002985   67.08   <2e-16 ***
Customer_SVC_H     0.374476   0.003501  106.95   <2e-16 ***
Staff_Cared_H      0.139358   0.002983   46.72   <2e-16 ***
Internet_Sat_H     0.040836   0.001206   33.85   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.848 on 181827 degrees of freedom
  (365332 observations deleted due to missingness)
Multiple R-squared:    0.7, Adjusted R-squared:    0.7
F-statistic: 7.071e+04 on 6 and 181827 DF,  p-value: < 2.2e-16
```

After excluding the 'Quality of check-in process' we see that the R squared value increased to 0.7.

The R squared value for linear modelling is 0.7 which means that the combination of all the six factors account for 70% of the rating for 'Likelihood to Recommend'. The 'significance codes' associated to each estimate is as rated at three stars which means that the p-value is highly significant.

Association Rules

We used association rules mining for predicting the factors or the amenities or both which could have the maximum impact on the NPS score of Hyatt Hotels. We set the RHS in the model as NPS type detractors. This enabled us to analyze what were the factors which were most likely associated with the detractors. Based on the model we generated some rules and picked the initial 40 with the highest value of lift.

Code:

load("C:/Users/sachd/Downloads/FinalDataset.RData")

```
Hyatt_Feb2Jun<- Hyatt_Feb2Jun[-nrow(Hyatt_Feb2Jun),]
Hyatt_Feb2Jun<- Hyatt_Feb2Jun[-nrow(Hyatt_Feb2Jun),]

install.packages("arules")
library(arules)

install.packages("arulesViz")
library(arulesViz)

install.packages("scales")
library(scales)

Hyatt_Feb2Jun <- Hyatt_Feb2Jun[,-1]
Hyatt_Feb2Jun <- Hyatt_Feb2Jun[,-ncol(Hyatt_Feb2Jun)]
Hyatt_Feb2Jun <- Hyatt_Feb2Jun[,-ncol(Hyatt_Feb2Jun)]
Hyatt_Feb2Jun <- Hyatt_Feb2Jun[,-ncol(Hyatt_Feb2Jun)]

Hyatt_Feb2Junaa<- Hyatt_Feb2Jun[,-3]
Hyatt_Feb2Junaa<- Hyatt_Feb2Junaa[,-3]
Hyatt_Feb2Junaa<- Hyatt_Feb2Junaa[,-12:-18]
Hyatt_Feb2Junaa<- Hyatt_Feb2Junaa[,-13]


rule1 <- apriori(Hyatt_Feb2Junaa, parameter=list(support=0.005, confidence=0.05), appearance
= list(rhs=c("Hyatt.NPS_Type=Detractor"), default="lhs"))
rule1
goodrules <- rule1[quality(rule1)$lift>1]
goodrules
inspect(goodrules)
rules.sorted<- sort(goodrules, by="lift")
rules.sorted
inspect(rules.sorted[1:40])
plot(rule1)
```

As per the model and our support and confidence values, it generated 2540 rules. We then filtered only those rules which had a confidence of more than 1. It generated 1505 rules. Sorting was then done on the basis of decreasing order of lift and the first 40 rules were taken.
The rules generated are as follows:

	lhs	rhs	support	confidence	lift
[1]	{Hyatt.Guest_Room_H=3, Hyatt.Resort_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005326259	0.8598251	12.89152
[2]	{Hyatt.Guest_Room_H=3, Hyatt.Golf_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005397956	0.8583734	12.86976
[3]	{Hyatt.Guest_Room_H=3, Hyatt.Casino_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005509484	0.8582775	12.86832
[4]	{Hyatt.POV_CODE_C=BUSINESS, Hyatt.Guest_Room_H=3}	{Hyatt.NPS_Type=Detractor}	0.005034693	0.8582292	12.86760
[5]	{Hyatt.Guest_Room_H=3, Hyatt.Boutique_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005515857	0.8577800	12.86086
[6]	{Hyatt.Guest_Room_H=3}	{Hyatt.NPS_Type=Detractor}	0.005581180	0.8571079	12.85078
[7]	{Hyatt.Guest_Room_H=3, Hyatt.Ski_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005552502	0.8571077	12.85078
[8]	{Hyatt.Status_H=COMPLETED,	{Hyatt.NPS_Type=Detractor}	0.005579587	0.8570729	12.85026

	Hyatt.Guest_Room_H=3}				
[9]	{Hyatt.Guest_Room_H=3, Hyatt.Conference_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005557282	0.8570025	12.84920
[10]	{Hyatt.Guest_Room_H=3, Hyatt.Indoor.Corridors_PL=Y}	{Hyatt.NPS_Type=Detractor}	0.005421855	0.8563161	12.83891
[11]	{Hyatt.Guest_Room_H=4, Hyatt.Country_PL=United States}	{Hyatt.NPS_Type=Detractor}	0.005230664	0.7924210	11.88092
[12]	{Hyatt.Guest_Room_H=4, Hyatt.Resort_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005845661	0.7895416	11.83775
[13]	{Hyatt.POV_CODE_C=BUSINESS, Hyatt.Guest_Room_H=4}	{Hyatt.NPS_Type=Detractor}	0.005587553	0.7859704	11.78421
[14]	{Hyatt.Guest_Room_H=4, Hyatt.Casino_PL=N}	{Hyatt.NPS_Type=Detractor}	0.006057564	0.7856995	11.78014
[15]	{Hyatt.Guest_Room_H=4, Hyatt.Ski_PL=N}	{Hyatt.NPS_Type=Detractor}	0.006103769	0.7856850	11.77993
[16]	{Hyatt.Guest_Room_H=4, Hyatt.Golf_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005914171	0.7854422	11.77629
[17]	{Hyatt.Guest_Room_H=4, Hyatt.Indoor.Corridors_PL=Y}	{Hyatt.NPS_Type=Detractor}	0.005990648	0.7852966	11.77410
[18]	{Hyatt.Guest_Room_H=4}	{Hyatt.NPS_Type=Detractor}	0.006134041	0.7844336	11.76116
[19]	{Hyatt.Guest_Room_H=4, Hyatt.Conference_PL=N}	{Hyatt.NPS_Type=Detractor}	0.006116515	0.7844299	11.76111
[20]	{Hyatt.Guest_Room_H=4, Hyatt.Boutique_PL=N}	{Hyatt.NPS_Type=Detractor}	0.006068717	0.7843904	11.76052
[21]	{Hyatt.Status_H=COMPLETED, Hyatt.Guest_Room_H=4}	{Hyatt.NPS_Type=Detractor}	0.006132447	0.7843896	11.76050
[22]	{Hyatt.Condition_Hotel_H=5, Hyatt.Spa.F.B.offering_PL=}	{Hyatt.NPS_Type=Detractor}	0.006286993	0.7770776	11.65087
[23]	{Hyatt.Condition_Ho	{Hyatt.NPS_Type=Detrac	0.006286993	0.7770776	11.65087

	tel_H=5, Hyatt.Spa.online.booking_PL=}	tor}			
[24]	{Hyatt.Condition_Hotel_H=5, Hyatt.Spa_PL=N}	{Hyatt.NPS_Type=Detractor}	0.006006580	0.7741273	11.60664
[25]	{Hyatt.Condition_Hotel_H=5, Hyatt.Country_PL=United States}	{Hyatt.NPS_Type=Detractor}	0.006921110	0.7735043	11.59730
[26]	{Hyatt.Condition_Hotel_H=5, Hyatt.Mini.Bar_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005076118	0.7725509	11.58300
[27]	{Hyatt.Condition_Hotel_H=5, Hyatt.Fitness.Trainer_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005308733	0.7677419	11.51090
[28]	{Hyatt.Condition_Hotel_H=5, Hyatt.All.Suites_PL=N}	{Hyatt.NPS_Type=Detractor}	0.005998614	0.7672712	11.50385
[29]	{Hyatt.Condition_Hotel_H=5, Hyatt.Resort_PL=N}	{Hyatt.NPS_Type=Detractor}	0.007910523	0.7671508	11.50204
[30]	{Hyatt.Condition_Hotel_H=5, Hyatt.Restaurant_PL=Y}	{Hyatt.NPS_Type=Detractor}	0.006172279	0.7668250	11.49715
[31]	{Hyatt.POV_CODE_C=BUSINESS, Hyatt.Condition_Hotel_H=5}	{Hyatt.NPS_Type=Detractor}	0.007556820	0.7659884	11.48461
[32]	{Hyatt.Condition_Hotel_H=5, Hyatt.Bell.Staff_PL=Y}	{Hyatt.NPS_Type=Detractor}	0.005135068	0.7650131	11.46999
[33]	{Hyatt.Status_H=COMPLETED, Hyatt.Condition_Hotel_H=5}	{Hyatt.NPS_Type=Detractor}	0.008315210	0.7648007	11.46680
[34]	{Hyatt.Condition_Hotel_H=5, Hyatt.Casino_PL=N}	{Hyatt.NPS_Type=Detractor}	0.008182970	0.7647409	11.46591
[35]	{Hyatt.Condition_Hotel_H=5}	{Hyatt.NPS_Type=Detractor}	0.008315210	0.7646886	11.46512
[36]	{Hyatt.Condition_Hotel_H=5, Hyatt.Indoor.Corridors_PL=Y}	{Hyatt.NPS_Type=Detractor}	0.008069848	0.7644129	11.46099
[37]	{Hyatt.Condition_Hotel_H=5, Hyatt.Golf_PL=N}	{Hyatt.NPS_Type=Detractor}	0.007994965	0.7642400	11.45840
[38]	{Hyatt.Condition_Hotel_H=5, Hyatt.Ski_PL=N}	{Hyatt.NPS_Type=Detractor}	0.008251480	0.7640897	11.45614

[39]	{Hyatt.Condition_Hotel_H=5, Hyatt.Elevators_PL=Y}	{Hyatt.NPS_Type=Detractor}	0.006502083	0.7640891	11.45613
[40]	{Hyatt.Condition_Hotel_H=5, Hyatt.Boutique_PL=N}	{Hyatt.NPS_Type=Detractor}	0.008202089	0.7637982	11.45177

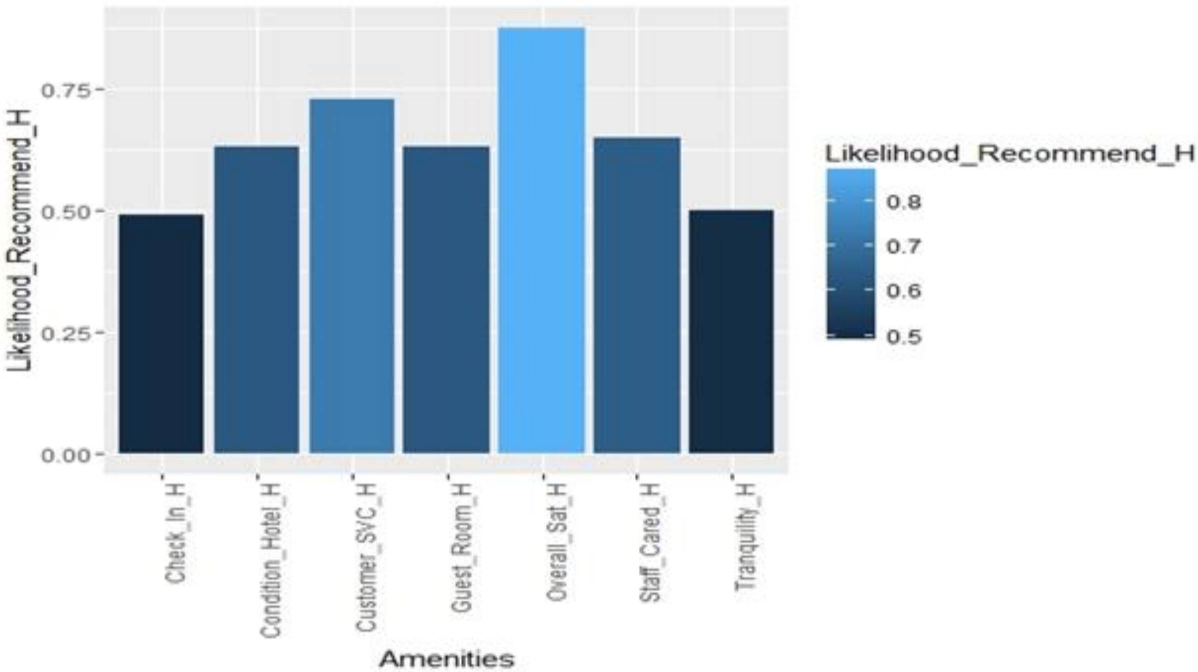
We see that the following factors most affect the detractors:

- Guest Room
- Resort
- Golf
- Casino
- Conference
- Mini Bar
- Fitness trainer
- Indoor Corridors
- Ski

Based on the combination of these factors on the basis of the model we can make suggestions to improve the NPS.

SVM

We decided to analyze how the Likelihood to recommendation factor is dependent on other factors. It is crucial to understand the relationship between the most important parameter i.e likelihood to recommend and several other rated parameters for e.g. overall satisfaction, tranquility, condition of the hotel, check in process, customer service and the staff cared. The below graph visually helps us to understand the same:



So, here we can see from analysing Model 1 for which all we have considered for analysis was the above amenities (except overall satisfaction) and we were able to draw a conclusion based on our findings that if a customer just utilizes these amenities the Hotel is able to categorize him either as a promoter or a detractor correctly 68.4% of the times.

Code:

```
svmModel1 <- ksvm(Likelihood_Recommend_H ~ Guest_Room_H + Tranquility_H +
Condition_Hotel_H + Customer_SVC_H + Staff_Cared_H, data=trainData, kernel="rbfdot",
kpar="automatic", C=5,cross=3, prob.model=TRUE)
```

```
Support Vector Machine object of class "ksvm"

SV type: eps-svr  (regression)
 parameter : epsilon = 0.1  cost C = 5

Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  0.250860078974567

Number of Support Vectors : 5473

Objective Function Value : -17093.43
Training error : 0.464973
Cross validation error : 5.735783
Laplace distr. width : 6.666672
```

Second model uses Likelihood to recommend and Overall Satisfaction. This model gives an accuracy of 84%

svmModel2 <- ksvm(Likelihood_Recommend_H ~ Overall_Sat_H, data=trainData, kernel="rbfdot", kpar="automatic", C=5, cross=3, prob.model=TRUE)

```
Support Vector Machine object of class "ksvm'

SV type: eps-svr  (regression)
 parameter : epsilon = 0.1  cost C = 5

Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  6.34515070287732

Number of Support Vectors : 3303

Objective Function Value : -10580.25
Training error : 0.325135
Cross validation error : 3.788106
Laplace distr. width : 7.039444
```

Hence, we can summarize that Likelihood to recommend is dependent on Overall Satisfaction.

Recommendations to the Hotel:

From the above analysis, following are the recommendations suggested:

1. The state of Washington has least percent of detractors and provides high quality of amenities to their customers. So, Maine, which has maximum percent of detractors, can provide similar services to its customers to improve their NPS.
2. Amenities such as Shuttle Service, Limo Service, Spa Services, and Boutique are unavailable in brands like Andaz and Park Hyatt. Implementing these services may improve overall NPS score.
3. Arkansas, Maine, and North Dakota can work on their surveying strategies to get more feedback on their hotels. Trigger more specific customer surveys at targeted customer touch-points and after key customer events, such as point of sale or just after the customer has engaged with the website or visited the store.
4. Andaz brand needs to improve on low Likelihood to Recommend, overall Satisfaction, and other amenities.
5. Male customers traveling for business between the age of 46-55 years should be focused-on by providing perks and allowances in order to gain customer loyalty and improve NPS of the hotel overall.
6. Provide a high quality, consistent service that leads to good customer experience and increases the likelihood to recommend to others.