IST 687
HYATT GROUP DATA ANALYSIS

-Nidhi Rakheja
-Pranay Lulla
-Vivek Veeraraghavan

TABLE OF CONTENTS

Introduction:

Hyatt group has a dataset of size 15 GB and approximately 3 million records for the year 2014. Few columns give information about the guests such as guest title and guest preferred language. There are some columns about the hotel like location, hotel name, spa and casino. The dataset also has a column which identifies the guest as a promoter, detractor or passive. This information is useful to the management as it helps them understand the customer's experience at the hotel.

The NPS is a value which helps the management decide whether the customer was satisfied with the hotel and its facilities or not. In our analysis, we have selected columns like Tranquility, Hotel Condition, Bell Staff, Business Center and others that may have an impact on someone staying at the hotel to recommend their experience to others or give bad reviews which will lead to people not booking a room in the hotel.

Types of customers:

Those who respond with a score of 9 to 10 are called Promoters, and are considered likely to exhibit value-creating behaviors, such as buying more, remaining customers for longer, and making more positive referrals to other potential customers.

Those who respond with a score of 0 to 6 are labeled Detractors, and are believed to be less likely to exhibit the value-creating behaviors.

Responses of 7 and 8 are labeled Passives, and their behavior falls in the middle of Promoters and Detractors and are usually of neutral opinion and wouldn't mind exploring their options.

Impact on the business:

Along with the technical analysis, it is important to understand the business aspect as well. Hyatt group is an international brand and its main aim is to increase its revenue. The main way to increase revenue for a hotel is to increase its customer base. The hotel takes feedback from its customers about the facilities and services provided by them. From the feedback provided, they can decide on what to improve. If they receive good feedback then can be assured that the customers will recommend their friends, relatives and colleagues to book rooms in the hotel.

Business Questions:

1. What are the facilities that affect the likelihood to recommend?
2. What is the importance of amenities for people travelling for business purpose?
3. What facilities are available in the Hotels which have maximum Promoters in comparison to the Hyatt Hotels with Detractors
4. Importance of Overall F&B, Customer care, Guest satisfaction and Hotel condition on NPS_type?
5. State and Country that has maximum Promoters, Detractors and Passives

Steps followed in analyzing the data:

- Exploring the data
- Cleaning the data
- Calculating the NPS
- Visualize the data based on columns that impact the recommendation
- Use modelling techniques
- Give useful insights based on the analysis

.

Exploring and cleaning the dataset:

The complete dataset has data for each of the twelve months in a separate CSV file of about 1GB in size. We selected the months of February, April, August and December to get a holistic view and cover all the seasons to perform our analysis. We read the dataset into R Studio using the import function in it. There was a lot of missing data that either needed to be replaced or removed. The data was divided in two parts mainly for United States and the rest of the world. In the data for United States we chose the columns that we thought were relevant and would affect the ratings for a hotel. We decided to only use columns that had at least 80% of the data present to avoid manipulating the data too much. Then we filtered the data for all the values which existed for NPS_Type to make sure we do not miss out any of the obvious data.Three subsets were created each for Promoter, Detractor and Passive. In each subset the matrix value columns were replaced by the mean of only respective NPS type. This step was performed for all the three types of customers. To not increase manipulation, all NAs in flag type columns were omitted.

To be able to use flag type and matrix in both numerical and categorical form; Matrix values were divided in three categories namely HIGH (8-10), MEDIUM(5-7) and LOW(0-4) and Flags were changed with Y being 1 and N being 0.

Further subsets were created for separate modeling and visualizations depending on the needs.

Calculating the NPS:

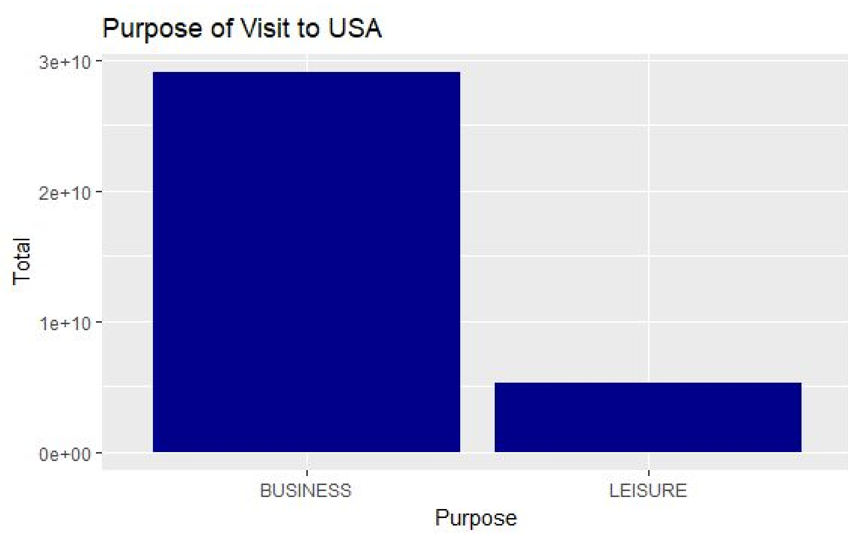
Net Promoter or **Net Promoter Score** (**NPS**) is a management tool that can be used to gauge the loyalty of a firm's customer relationships. It serves as an alternative to traditional customer satisfaction research and claims to be correlated with revenue growth. NPS can be as low as −100 (everybody is a detractor) or as high as +100 (everybody is a promoter). An NPS that is positive is felt to be good, and an NPS of +50 is excellent. In our analysis we used the NPS type column to calculate the NPS value for promoters, detractors and passive respectively.

We have used the following formula to calculate the NPS:
((Total Promoters - Total Detractors)/(Total Respondents)) * 100

We have plotted the map of USA based on the NPS goal and also on the NPS that we have calculated. We found that the difference between the goal and the actual NPS value is high i .e the actual NPS is very less compared to the goal value.

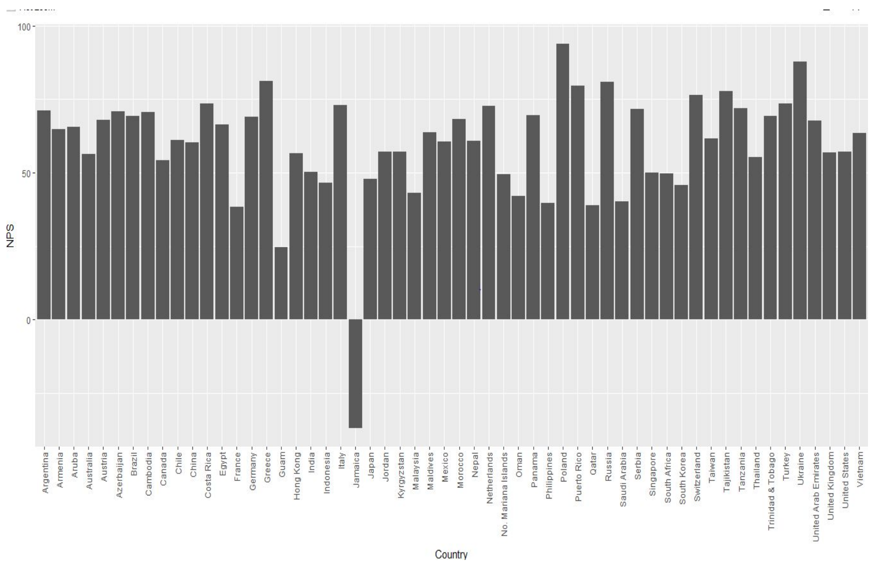
- **<u>Visualizing the data for descriptive statistics:</u>**

We have made a bar plot to visualize people's purpose of visit to the United States.



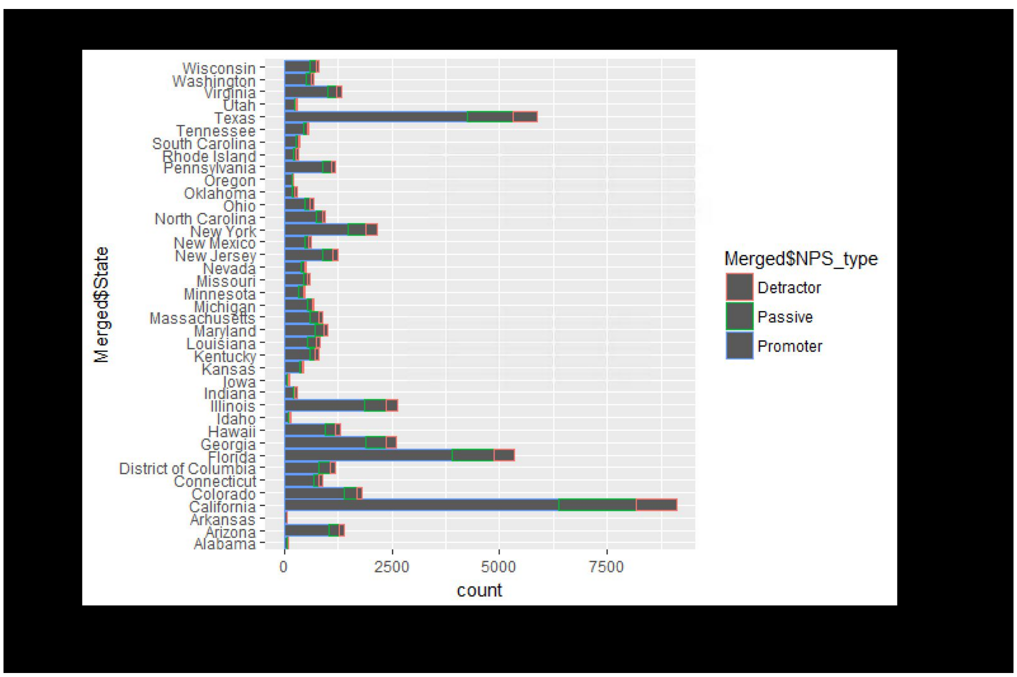
From this we can see that people visit the United States for business purposes a lot more than they do for leisure purposes.

A bar plot to visualize the calculated NPS for each country.



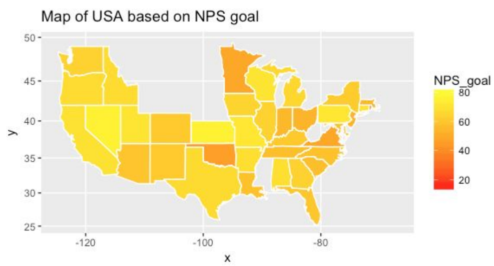
From this map, we can infer that Poland is the country with the highest NPS and Guam is the one with the lowest NPS. Jamaica had a negative NPS.

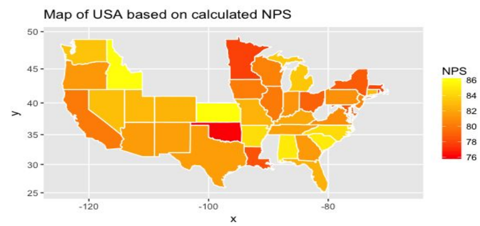
A bar plot to see the number of promoters and detractors for every state in the United States.



From this plot, California and Texas have the most number of promoters as well as detractors.

A state map of United States with NPS goal and calculated NPS to compare the improvement required by each state.





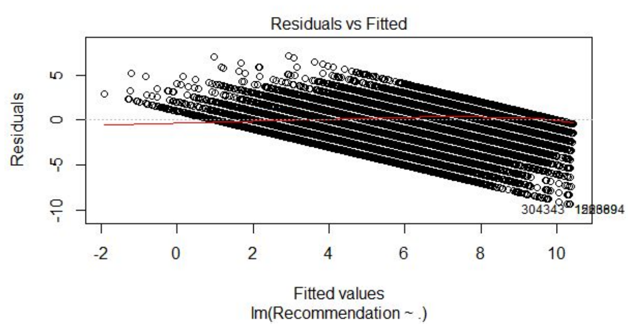
- **<u>Modeling techniques:</u>**

Linear Modeling:
To answer:
What are the facilities that affect the likelihood to recommend?

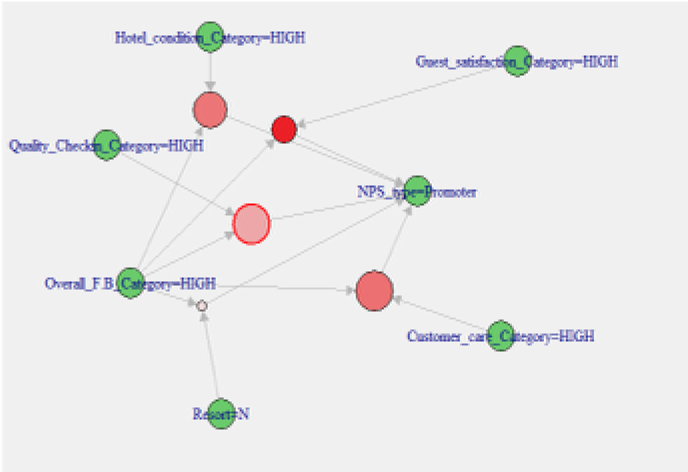
Linear modeling to find out which facilities have an impact on the likelihood to recommend by a customer.

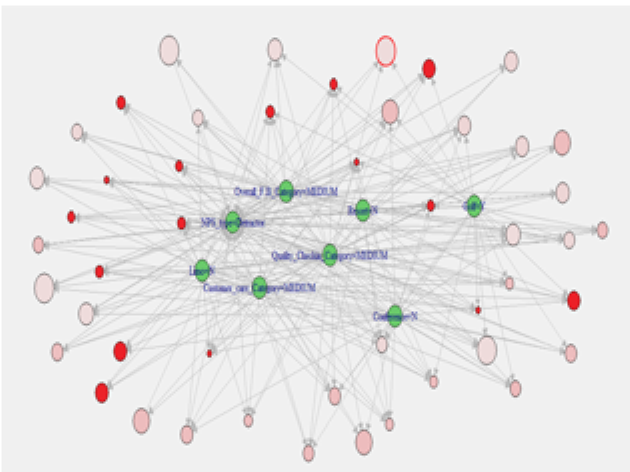
```
Call:
lm(formula = Recommendation ~ ., data = linear_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.3488 -0.3280  0.1320  0.4664  7.0646 

Coefficients:
                    Estimate Std. Error  t value Pr(>|t|)    
(Intercept)        -2.914981   0.019127 -152.401  < 2e-16 ***
Guest_satisfaction  0.310772   0.002117  146.821  < 2e-16 ***
Hotel_condition     0.198338   0.002501   79.296  < 2e-16 ***
Customer_service    0.257126   0.002377  108.170  < 2e-16 ***
Customer_care       0.260104   0.002952   88.110  < 2e-16 ***
Quality_Checkin     0.016802   0.002636    6.374 1.84e-10 ***
Overall_F.B         0.281152   0.002251  124.920  < 2e-16 ***
ConventionY        -0.055074   0.004955  -11.114  < 2e-16 ***
FitnessY            0.020846   0.005345    3.900 9.61e-05 ***
ResortY             0.092410   0.009784    9.445  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.032 on 185141 degrees of freedom
Multiple R-squared:  0.739,	Adjusted R-squared:  0.739 
F-statistic: 5.825e+04 on 9 and 185141 DF,  p-value: < 2.2e-16
```

From the above model, the guest satisfaction column is the most important for likelihood to recommend. Customer service, customer care and overall food and beverages column also impact the likelihood to recommend to a certain extent.
The value of R squared is 0.739.
Plotted the residual vs fitted graph to visualize the result of the linear model.



<u>Associative Rules:</u>
To Answer:
What is the importance of amenities for people travelling for business purpose?
Effect of amenities and facilities on likelihood of a client being a promoter or detractor?





SVM:

To answer:

Importance of Overall F&B, Customer care, Guest satisfaction and Hotel condition on NPS_type?

SVM model was used to predict the number of promoter, detractors and passives on the test data based on the trained data. Training data was 2/3rd of the dataset (subset which had relevant columns about hotels in the United States) ; the remaining 1/3rd was the test data where to validate the model.

```
Detractor   Passive  Promoter
     7443     10183     44091
> result_svm <- table(predicted_SVM, SVM_data_subset.test$NPS_type)
> print(result_svm)

predicted_SVM Detractor Passive Promoter
    Detractor      6101    1114      228
    Passive        1016    7687     1480
    Promoter        793    2932    40366
```

```
> Correct_svm <- (result_svm[1,1]+result_svm[2,2]+result_svm[3,3])/(result_svm[1,1]+result_svm[1,2]+result_svm[1,3]+result_svm[2,1]+
result_svm[2,2]+result_svm[2,3]+result_svm[3,1]+result_svm[3,2]+result_svm[3,3])*100
> Correct_svm
[1] 87.74568
>
```

NPS_type was taken as the dependent variable and the model obtained 87.7% accuracy

KSVM:

KSVM model was used to predict the number of promoter, detractors and passives on the test data based on the trained data. Training data was 2/3rd of the dataset (subset which had relevant columns about hotels in the United States) ; the remaining 1/3rd was the test data where to validate the model.

```
> ksvm(NPS_type ~ .,data=train_data, cost=0.1, scale=FALSE)
Support Vector Machine object of class "ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 1

Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  0.583333333333333

Number of Support Vectors : 43009

Objective Function Value : -15927.39 -8953.494 -28654.79
Training error : 0.119006

> Accuracy
[1] 88.20626
```

	pred	Var2	Freq
1	Detractor	Detractor	5930
2	Passive	Detractor	1120
3	Promoter	Detractor	759
4	Detractor	Passive	744
5	Passive	Passive	8011
6	Promoter	Passive	2888
7	Detractor	Promoter	153
8	Passive	Promoter	1729
9	Promoter	Promoter	40383

NPS_type was taken as the dependant variable and the model obtained 88.2% accuracy.

Recommendations:

- Customer service, Guest_room_Satisfaction and Overall_F&B are important factors of satisfaction.
- Business centers and Convention centers must be added or improved for people travelling for business.
- People coming for business purpose are more than that of leisure purpose, so focus should be more on improving the amenities which are required by people coming for

business purpose.

- Internet connection and conference are not important in determining the NPS, since the customers can work on a basic internet connection.
- Fitness centers must be well maintained for leisure purpose
- Golf course must be in a good condition.

Lessons Learnt:

- Understood the various modeling and visualization techniques
- Impact of NPS on the business of the Hyatt Group.
- Strengths and weaknesses of all the team members.

Challenges Faced:

- To decide the columns which would have an impact on the likelihood to recommend.
- Cleaning the dataset i.e. to deal with the NAs and blank columns.
- Decide on a particular time when all the team members could meet.

R CODE

DATA CLEANING

```
Feb_data <-
out_201402[,c(19,23,24,137:145,147,163,167,168,169,171,175,176,179,187,191,201:205,208,2
10,213,218,232)]

#April data
April_data <-
out_201404[,c(19,23,24,137:145,147,163,167,168,169,171,175,176,179,187,191,201:205,208,2
10,213,218,232)]

#August data
August_data <-
out_201408[,c(19,23,24,137:145,147,163,167,168,169,171,175,176,179,187,191,201:205,208,2
10,213,218,232)]

#December data
December_data <-
out_201412[,c(19,23,24,137:145,147,163,167,168,169,171,175,176,179,187,191,201:205,208,2
10,213,218,232)]

#Merge data
Merged <- rbind(Feb_data, April_data,August_data,December_data)

colnames(Merged) <-
c("Duration","Purpose","Rate","Recommendation","Satisfaction","Guest_satisfaction","Tranquilit
y","Hotel_condition","Customer_service","Customer_care","Internet","Quality_Checkin","Overall
_FandB","Hotel_Name","City","State","Region","Country","Latitude","Longitude","NPS_goal","Me
eting_space","Region","Bell_staff","Business_center","Casino","Conference","Convention","Golf"
,"Limo","Fitness","Resort","NPS_type")


#Promoters in USA
dataUSA_subset_promoters <-data.frame(subset(data_subset_usa_NPS_Type
,data_subset_usa_NPS_Type$NPS_type=="Promoter"))

for(i in 4:13){
  dataUSA_subset_promoters[is.na(dataUSA_subset_promoters[,i]), i] <-
floor(mean(dataUSA_subset_promoters[,i], na.rm = TRUE))
}

dataUSA_subset_promoters <- data.frame(na.omit(dataUSA_subset_promoters))
#Promoters in USA(126,351)

#Detractors in USA (30,508)
```

```
dataUSA_subset_detractor <-data.frame(subset(data_subset_usa_NPS_Type
,data_subset_usa_NPS_Type$NPS_type=="Detractor"))

for(i in 4:13){
  dataUSA_subset_detractor[is.na(dataUSA_subset_detractor[,i]), i] <-
floor(mean(dataUSA_subset_detractor[,i], na.rm = TRUE))
}

dataUSA_subset_detractor <- data.frame(na.omit(dataUSA_subset_detractor))
#Detractors in USA

#Passives in USA
dataUSA_subset_passive <-data.frame(subset(data_subset_usa_NPS_Type
,data_subset_usa_NPS_Type$NPS_type=="Passive"))

for(i in 4:13){
  dataUSA_subset_passive[is.na(dataUSA_subset_passive[,i]), i] <-
floor(mean(dataUSA_subset_passive[,i], na.rm = TRUE))
}

dataUSA_subset_passive <- data.frame(na.omit(dataUSA_subset_passive))

#Passives in USA


USA_data<-rbind(dataUSA_subset_promoters,dataUSA_subset_detractor,dataUSA_subset_pa
ssive)

##################################################################################


data_subset_ROW <- data.frame(subset (Merged, Merged$Country != "United States"))

Merged1 <- data.frame(subset(data_subset_ROW, data_subset_ROW$NPS_type!="NA"))
Merged1 <-data.frame(Merged1[c(-16,-17)])
#Number of PromotersROW(41,269)
dataROW_NPSType_subset_promoters <-data.frame(subset(Merged1
,Merged1$NPS_type=="Promoter"))

for(i in 4:13){
  dataROW_NPSType_subset_promoters[is.na(dataROW_NPSType_subset_promoters[,i]), i] <-
floor(mean(dataROW_NPSType_subset_promoters[,i], na.rm = TRUE))
}
```

```
dataROW_NPSType_subset_promoters <-
data.frame(na.omit(dataROW_NPSType_subset_promoters))
#Number of PromotersROW(31,224)

#Detractors in ROW (6,656)
dataROW_NPSType_subset_detractor <-data.frame(subset(Merged1
,Merged1$NPS_type=="Detractor"))

for(i in 4:13){
  dataROW_NPSType_subset_detractor[is.na(dataROW_NPSType_subset_detractor[,i]), i] <-
floor(mean(dataROW_NPSType_subset_detractor[,i], na.rm = TRUE))
}

dataROW_NPSType_subset_detractor <-
data.frame(na.omit(dataROW_NPSType_subset_detractor))
#Detractors in USA(4,617)

#Passives in ROW(13,891)
dataROW_NPSType_subset_passive <-data.frame(subset(Merged1
,Merged1$NPS_type=="Passive"))

for(i in 4:13){
  dataROW_NPSType_subset_passive[is.na(dataROW_NPSType_subset_passive[,i]), i] <-
floor(mean(dataROW_NPSType_subset_passive[,i], na.rm = TRUE))
}

dataROW_NPSType_subset_passive <-
data.frame(na.omit(dataROW_NPSType_subset_passive))

#Passives in ROW(10,293)

Merged1<- rbind
(dataROW_NPSType_subset_promoters,dataROW_NPSType_subset_detractor,dataROW_NP
SType_subset_passive)
View(Merged1)


##############################################################################

USA_data_value_category <-USA_data[c(6:13)]
```

```
USA_data_value_category$Overall_F.B <-as.numeric(USA_data_value_category$Overall_F.B)



USA_data_value_category$Guest_satisfaction_category
<-(ifelse(USA_data_value_category$Guest_satisfaction>=8,"HIGH",
ifelse(USA_data_value_category$Guest_satisfaction>=5 &
USA_data_value_category$Guest_satisfaction<8,"MEDIUM",
 ifelse(USA_data_value_category$Guest_satisfaction<5,"LOW",NA))))



USA_data_value_category$Tranquility_category
<-(ifelse(USA_data_value_category$Tranquility>=8,"HIGH",
                                        ifelse(USA_data_value_category$Tranquility>=5 &
USA_data_value_category$Tranquility<8,"MEDIUM",
                                              ifelse(USA_data_value_category$Tranquility<5,"LOW",NA))))

  USA_data_value_category$Hotel_condition_category
<-(ifelse(USA_data_value_category$Hotel_condition>=8,"HIGH",
                                             ifelse(USA_data_value_category$Hotel_condition>=5 &
USA_data_value_category$Hotel_condition<8,"MEDIUM",

ifelse(USA_data_value_category$Hotel_condition<5,"LOW",NA))))

   USA_data_value_category$Customer_service_category
<-(ifelse(USA_data_value_category$Customer_service>=8,"HIGH",
                                              ifelse(USA_data_value_category$Customer_service>=5 &
USA_data_value_category$Customer_service<8,"MEDIUM",

ifelse(USA_data_value_category$Customer_service<5,"LOW",NA))))

  USA_data_value_category$Customer_care_category
<-(ifelse(USA_data_value_category$Customer_care>=8,"HIGH",
                                           ifelse(USA_data_value_category$Customer_care>=5 &
USA_data_value_category$Customer_care<8,"MEDIUM",

ifelse(USA_data_value_category$Customer_care<5,"LOW",NA))))

   USA_data_value_category
$Interne_categoryt<-(ifelse(USA_data_value_category$Internet>=8,"HIGH",
                                       ifelse(USA_data_value_category$Internet>=5 &
USA_data_value_category$Internet<8,"MEDIUM",
```

```
                                             ifelse(USA_data_value_category$Internet<5,"LOW",NA))))

  USA_data_value_category$Quality_Checkin_category
<-(ifelse(USA_data_value_category$Quality_Checkin>=8,"HIGH",
                                             ifelse(USA_data_value_category$Quality_Checkin>=5 &
USA_data_value_category$Quality_Checkin<8,"MEDIUM",

ifelse(USA_data_value_category$Quality_Checkin<5,"LOW",NA))))

   USA_data_value_category$Overall_F.B_category
<-(ifelse(USA_data_value_category$Overall_F.B>=8,"HIGH",
                                          ifelse(USA_data_value_category$Overall_F.B>=5 &
USA_data_value_category$Overall_F.B<8,"MEDIUM",

ifelse(USA_data_value_category$Overall_F.B<5,"LOW",NA))))




###############################Separation of NPS-type USA_data and statewise NPS
calculation  ###################
install.packages("sqldf")
library(sqldf)
Temp_NPS_P <- sqldf('select count(*) as Total_Promoters ,State
                from USA_data where NPS_type = "Promoter" Group By
                State')
Temp_NPS_P
Temp_NPS_D <- sqldf('select count(*) as Total_Detractors ,State
                from USA_data where NPS_type = "Detractor" Group By
                State')
Temp_NPS_D

Texas <- which(USA_data$State=="Texas")
Type <- subset(USA_data$NPS_type, USA_data$State=="Texas")
tapply(Texas, Type, length)


Temp_NPS_P_D <- sqldf('select a.Total_Promoters, b.Total_Detractors, a.State from
                  Temp_NPS_P a LEFT JOIN Temp_NPS_D b ON a.State = b.State')
```

```
Temp_NPS_Pa <- sqldf('select count(*) as Total_Passives ,State from
                USA_data where NPS_Type = "Passive" Group By
                State')
Temp_NPS_Pa

Temp_NPS_P_D_Pa <- sqldf('select a.Total_Promoters, a.Total_Detractors, b.Total_Passives,
                   a.State from Temp_NPS_P_D a LEFT JOIN Temp_NPS_Pa b ON
a.State=b.State')
Total_Respondents <-
Temp_NPS_P_D_Pa$Total_Promoters+Temp_NPS_P_D_Pa$Total_Detractors+Temp_NPS_P
_D_Pa$Total_Passives
Percentage_Promoter <- Temp_NPS_P_D_Pa$Total_Promoters *100/Total_Respondents
Percentage_Detractor <- Temp_NPS_P_D_Pa$Total_Detractors *100/Total_Respondents
Percentage_Passive <- Temp_NPS_P_D_Pa$Total_Passives *100/ Total_Respondents

##Promoters###
ggplot(subset(USA_data,NPS_type=="Promoter"),aes(x=State)) +
  geom_bar() + ggtitle("Promoter Count") +
  labs(x = "State", y = "Promoter Count") + coord_flip()

###Detractors####
ggplot(subset(USA_data, NPS_type=="Detractor"),aes(x=State)) +
  geom_bar() + ggtitle("Detractor Count") +
  labs(x = "State", y = "Detractor Count") + coord_flip()


NPS_State <-
((Temp_NPS_P_D_Pa$Total_Promoters+Temp_NPS_P_D_Pa$Total_Detractors)/
Total_Respondents)*100
Temp_NPS_P_D_Pa$Percentage_Promoters <- Percentage_Promoter
Temp_NPS_P_D_Pa$Percentage_Detractors <- Percentage_Detractor
Temp_NPS_P_D_Pa$Percentage_Passives <- Percentage_Passive
Temp_NPS_P_D_Pa$Total <- Total_Respondents
Temp_NPS_P_D_Pa$NPS <- NPS_State
View(Temp_NPS_P_D_Pa)
```

```
usMapData <- map_data("state")
###Map of USA based on NPS goal###
mapNPS <- ggplot(USA_data, aes(map_id = tolower(State), fill=NPS_goal))
mapNPS <- mapNPS + geom_map(map = usMapData, color="white")
#Forming states using lat and long based on NPS goal
mapNPS <- mapNPS+ expand_limits(x = usMapData$long, y = usMapData$lat)
mapNPS <- mapNPS +coord_map() + ggtitle("Map of USA based on NPS goal")
mapNPS <- mapNPS +scale_fill_continuous(low ="red", high="yellow")
mapNPS
```

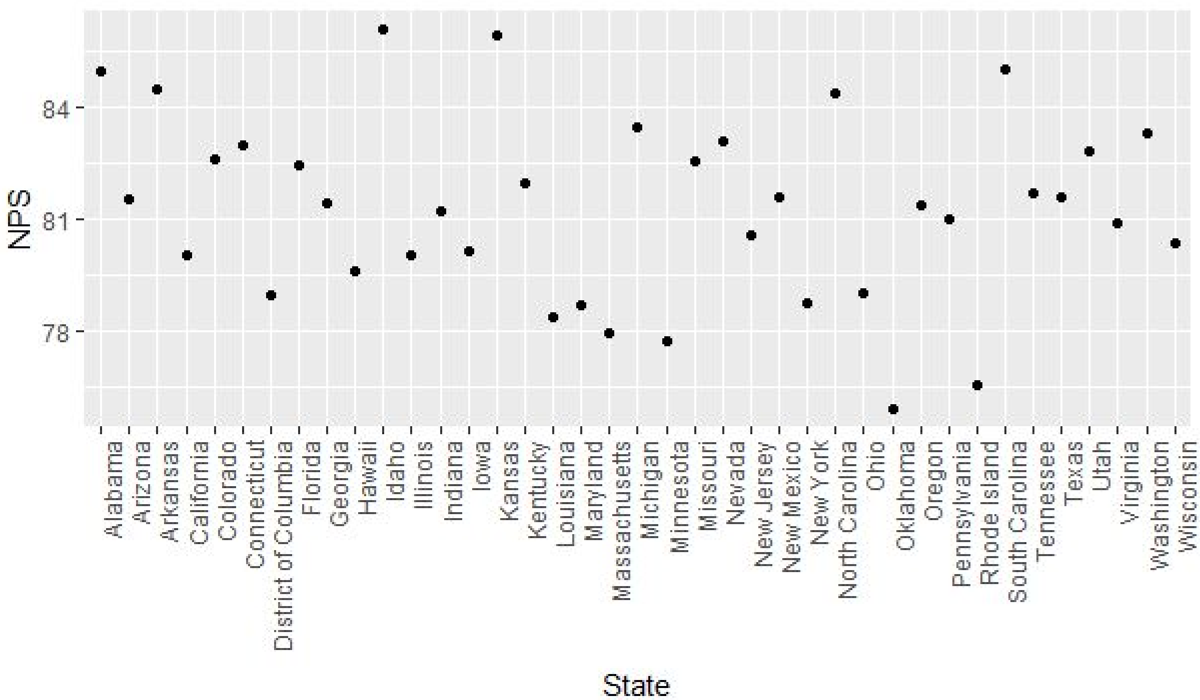
```
usMapData <- map_data("state")
###Map of USA based on  current NPS
###
mapNPScalculated <- ggplot(Temp_NPS_P_D_Pa, aes(map_id = tolower(State), fill=NPS))
mapNPScalculated <- mapNPScalculated + geom_map(map = usMapData, color="white")
#Forming states using lat and long based on current NPS
mapNPScalculated <- mapNPScalculated+ expand_limits(x = usMapData$long, y =
usMapData$lat)
mapNPScalculated <- mapNPScalculated +coord_map() + ggtitle("Map of USA based on
calculated NPS ")
mapNPScalculated <- mapNPScalculated +scale_fill_continuous(low ="red", high="yellow")
mapNPScalculated
```

#########KSVM############
ksvm_data <- subset(USA_data, select=c(2,23:37))
random.indexes <- sample(1:nrow(ksvm_data))
cutPoint2_3 <- floor(nrow(ksvm_data)/3*2)

#Creating test and train datasets for future computation
SVM_data.train <- ksvm_data[random.indexes[1:cutPoint2_3],]
SVM_data.test <- ksvm_data[random.indexes[(cutPoint2_3+1):nrow(ksvm_data)],]
ksvm_model <- ksvm(NPS_type ~ .,data=SVM_data.train, kernel = "rbfdot",kpar="automatic", C=5, cross=10,prob.model=T)
pred <- predict(ksvm_model, SVM_data.test)

pred_NPS_type <- data.frame(pred, SVM_data.test$NPS_type)
result <- table(pred,SVM_data.test$NPS_type)
View(result)
Accuracy <- (result[1,1]+result[2,2])/(result[1,1]+result[1,2]+result[2,1]+result[2,2])*100
Accuracy

```
> ksvm(NPS_type ~ .,data=train_data, cost=0.1, scale=FALSE)
Support Vector Machine object of class "ksvm"

SV type: C-svc  (classification)
 parameter : cost C = 1

Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  0.583333333333333

Number of Support Vectors : 43009

Objective Function Value : -15927.39 -8953.494 -28654.79
Training error : 0.119006
```




```
#########SVM##################
SVM_data <- subset(USA_data, select= c(23:38))


SVM_data$NPS_type_category <-(ifelse(SVM_data$NPS_type=="Promoter",1,
                        ifelse(SVM_data$NPS_type=="Passive",2,
                            ifelse(SVM_data$NPS_type=="Detractor",3,NA))))

SVM_data_subset <-subset(SVM_data,select=c(1,2,3,5,8,9,10,11,12,14,16))

random.indexes <- sample(1:nrow(SVM_data_subset))
cutPoint2_3 <- floor(nrow(SVM_data_subset)/3*2)

#Creating test and train datasets for future computation
SVM_data_subset.train <- SVM_data_subset[random.indexes[1:cutPoint2_3],]
SVM_data_subset.test <-
SVM_data_subset[random.indexes[(cutPoint2_3+1):nrow(SVM_data_subset)],]
```

```
SVM_model <-svm(as.factor(NPS_type) ~., data=SVM_data_subset.train)
predicted_SVM <- predict(SVM_model,SVM_data_subset.test)
summary(predicted_SVM)

result_svm <- table(predicted_SVM, SVM_data_subset.test$NPS_type)
print(result_svm)
Correct_svm <-
(result_svm[1,1]+result_svm[2,2]+result_svm[3,3])/(result_svm[1,1]+result_svm[1,2]+result_svm[1,3]+result_svm[2,1]+result_svm[2,2]+result_svm[2,3]+result_svm[3,1]+result_svm[3,2]+result_svm[3,3])*100
Correct_svm
```

```
Detractor    Passive   Promoter
     7443      10183      44091
> result_svm <- table(predicted_SVM, SVM_data_subset.test$NPS_type)
> print(result_svm)

predicted_SVM Detractor Passive Promoter
    Detractor      6101    1114      228
    Passive        1016    7687     1480
    Promoter        793    2932    40366
```

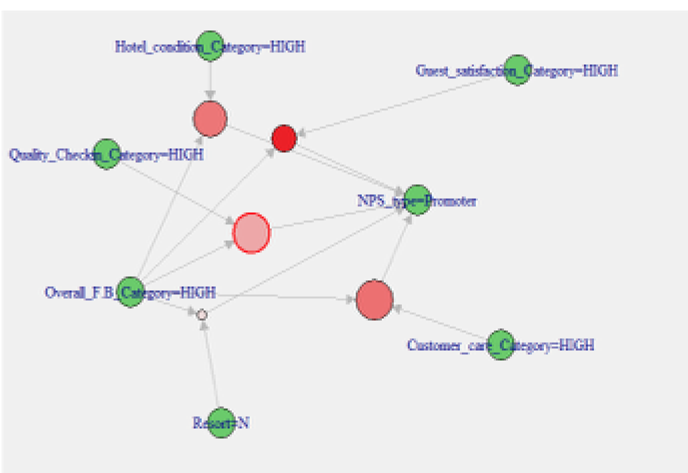
```
> Correct_svm
[1] 87.74568
```

###################Associative Rules##############################

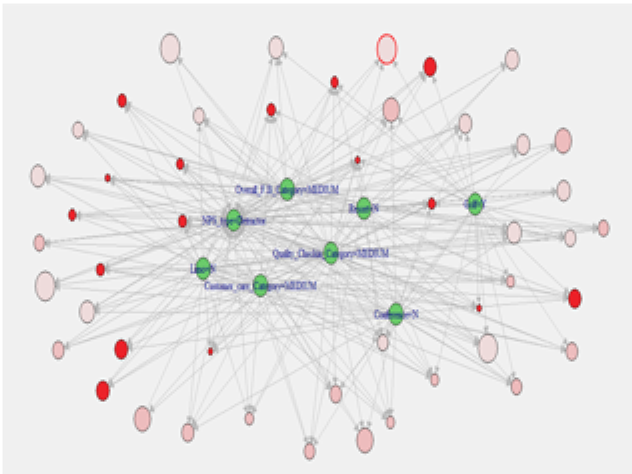
```
install.packages("arules")
install.packages("arulesViz")
library(arulesViz)
library(arules)

USA_data$Guest_satisfaction_Category <-
USA_data_value_category$Guest_satisfaction_category
USA_data$Hotel_condition_Category <- USA_data_value_category$Hotel_condition_category
USA_data$Overall_F.B_Category <- USA_data_value_category$Overall_F.B_category
USA_data$Customer_care_Category <- USA_data_value_category$Customer_care_category
```

```
USA_data$Quality_Checkin_Category <-
USA_data_value_category$Quality_Checkin_category
View(USA_data)
Rules_data <- subset(USA_data, select= c(2,23:37))
Rules_data <- replace(Rules_data, TRUE, lapply(Rules_data, factor))

RuleSet <- apriori(Rules_data,parameter =list(support=0.5,confidence=0.9), appearance =
list(rhs= c("NPS_type=Promoter", "NPS_type=Detractor")))
inspect(RuleSet)
RuleSet_df <- data.frame(LHS=labels(lhs(RuleSet)),RHS=labels(rhs(RuleSet)),
quality(RuleSet))
View(RuleSet_df)
GoodRules <- RuleSet[quality(RuleSet)$lift >1.2]
GoodRulesGraph<-
plot(GoodRules,method="graph",measure="support",shading="lift",interactive=TRUE)
inspect(GoodRules)

Rules_df <- data.frame(LHS=labels(lhs(GoodRules)),RHS=labels(rhs(GoodRules)),
quality(GoodRules))
View(Rules_df)
Rules_P <- Rules_df[Rules_df$RHS=='{NPS_Type=Promoter}']
View(Rules_P)
```

PROMOTERS



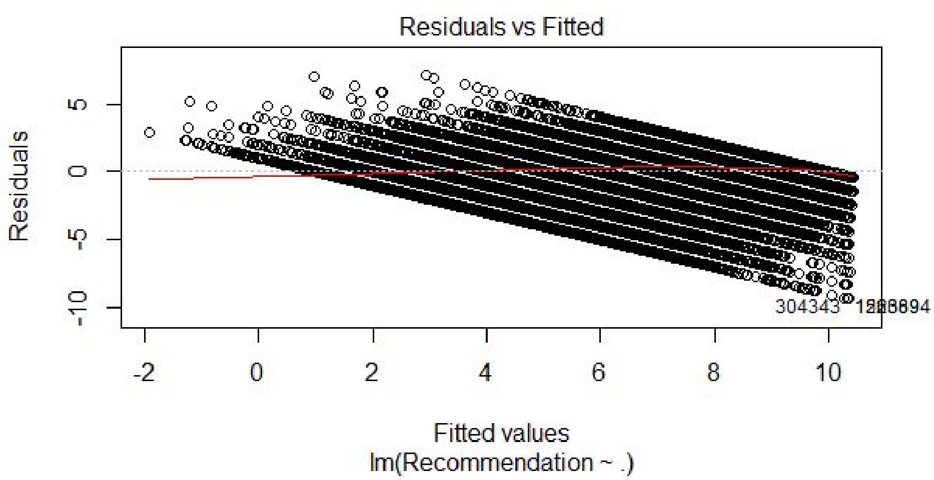
DETRACTORS

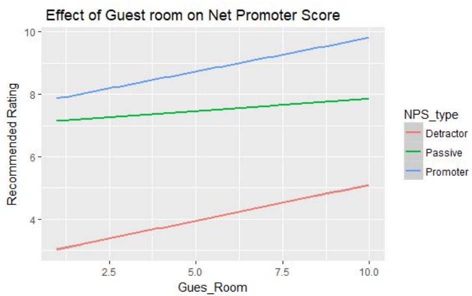


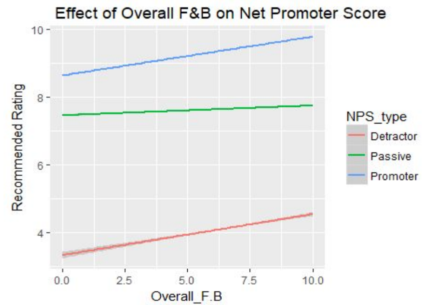
```
######################Linear Modelling###########################
linear_data <- subset(USA_data, select= c(4,6,8,9,10,12,13,27,30,31))
linear_final <- lm(Recommendation ~., data=linear_data)
#plot(linear_final)
summary(linear_final)
ggplot(USA_data, aes(x=Guest_satisfaction, y=Recommendation, color=NPS_type)) +
  geom_smooth(method = "lm") + ylab("Recommended Rating") +  xlab(" Gues_Room ") +
ggtitle(" Effect of Guest room on Net Promoter Score")
ggplot(USA_data, aes(x=Overall_F.B, y=Recommendation, color=NPS_type)) +
  geom_smooth(method = "lm") + ylab("Recommended Rating") +  xlab(" Overall_F.B ") +
ggtitle(" Effect of Overall F&B on Net Promoter Score")


ggplot(linear_final, aes(Recommendation, Overall_F.B)) + geom_point()


library(ggplot2)
GG <-ggplot(data=USA_data,aes(x=Guest_satisfaction,y=Recommendation))
ScatterPLot<- GG +geom_point(aes(colour=NPS_type),shape=19,alpha=0.5,position =
position_jitter(w=0.5,h=0.5))
ScatterPLot
GG <-ggplot(data=USA_data,aes(x=Customer_service,y=Recommendation))
ScatterPLot<- GG +geom_point(aes(colour=NPS_type),shape=19,alpha=0.5,position =
position_jitter(w=0.5,h=0.5))
ScatterPLot
GG <-ggplot(data=USA_data,aes(x=Overall_F.B,y=Recommendation))
```

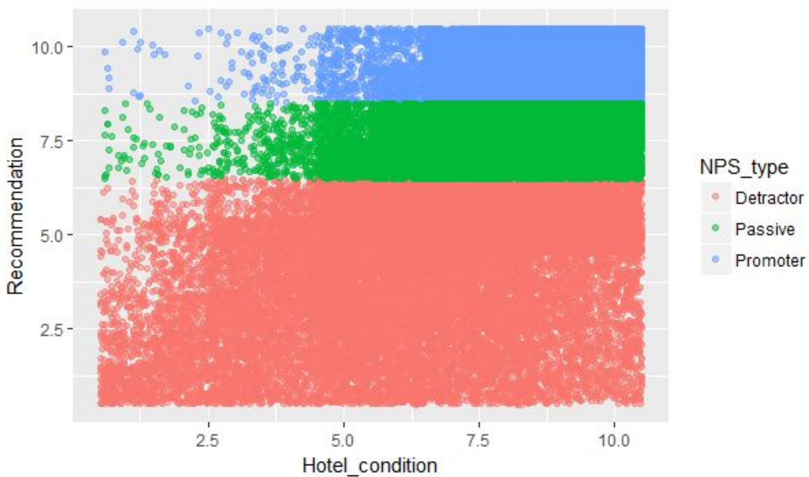
ScatterPLot<- GG +geom_point(aes(colour=NPS_type),shape=19,alpha=0.5,position = position_jitter(w=0.5,h=0.5))
ScatterPLot

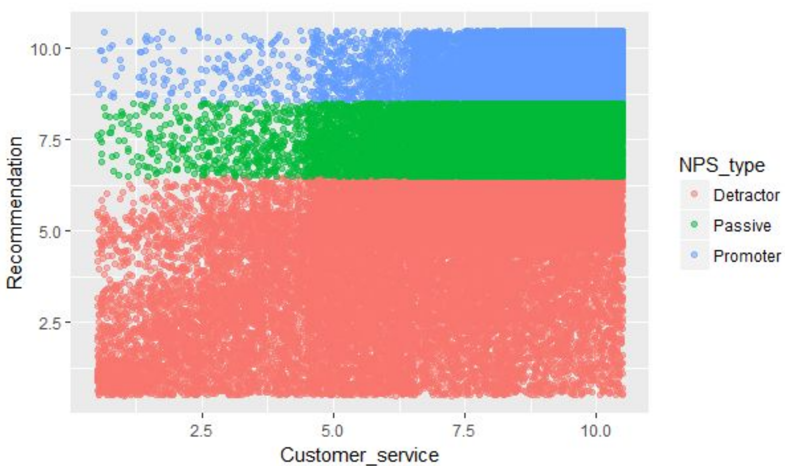
```
Call:
lm(formula = Recommendation ~ ., data = linear_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.3488 -0.3280  0.1320  0.4664  7.0646 

Coefficients:
                     Estimate Std. Error  t value Pr(>|t|)    
(Intercept)         -2.914981   0.019127 -152.401  < 2e-16 ***
Guest_satisfaction   0.310772   0.002117  146.821  < 2e-16 ***
Hotel_condition      0.198338   0.002501   79.296  < 2e-16 ***
Customer_service     0.257126   0.002377  108.170  < 2e-16 ***
Customer_care        0.260104   0.002952   88.110  < 2e-16 ***
Quality_Checkin      0.016802   0.002636    6.374 1.84e-10 ***
Overall_F.B          0.281152   0.002251  124.920  < 2e-16 ***
ConventionY         -0.055074   0.004955  -11.114  < 2e-16 ***
FitnessY             0.020846   0.005345    3.900 9.61e-05 ***
ResortY              0.092410   0.009784    9.445  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.032 on 185141 degrees of freedom
Multiple R-squared:  0.739,     Adjusted R-squared:  0.739 
F-statistic: 5.825e+04 on 9 and 185141 DF,  p-value: < 2.2e-16
```

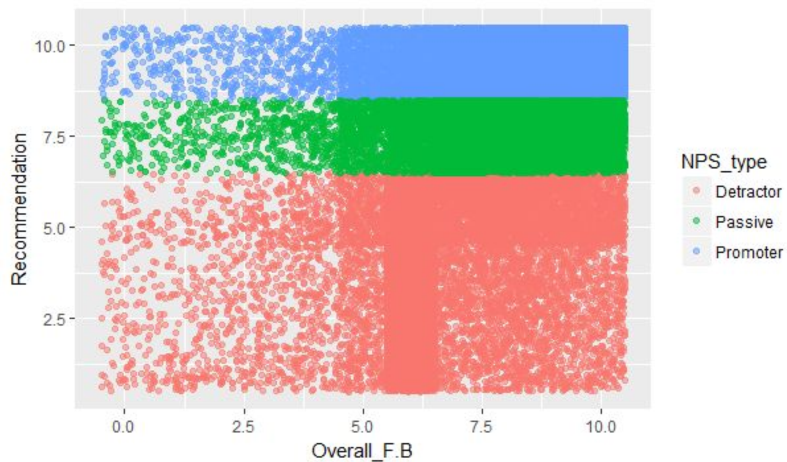



Effect of Guest room on Net Promoter Score
Recommended Rating
Gues_Room
NPS_type
Detractor
Passive
Promoter


Effect of Overall F&B on Net Promoter Score
Recommended Rating
Overall_F.B
NPS_type
Detractor
Passive
Promoter


Recommendation
Hotel_condition
NPS_type
Detractor
Passive
Promoter


Recommendation
Customer_service
NPS_type
Detractor
Passive
Promoter


NPS_type
Detractor
Passive
Promoter
Recommendation
10.0
7.5
5.0
2.5
0.0
2.5
5.0
7.5
10.0
Overall_F.B