School of Information Studies
SYRACUSE UNIVERSITY

IST 687 – Applied Data Science

Lab Section M004 | Group 2

Data Analysis for Hyatt Hotels Corporation in the United States

Recommendations to Increase Revenue and Improve Customer Satisfaction

HYATT®

Submitted by:

Sumegha Arora | Akash Pillai | Austin Lewis | Abhishek Vernekar | Manan Shah

Table of Contents

1. Description

Hyatt Hotels Corporation is an American multinational owner, operator, and franchiser of hotels, resorts, and vacation properties. As of September 30, 2017, Hyatt has 739 properties in 57 countries. In 2017, Fortune magazine listed Hyatt as the 32nd-best U.S. Company to work for.

The project revolves around analyzing the data collected from customer evaluations and feedback, booking information, and overall customer satisfaction, in order to provide recommendations to regional hotel(s) in the United States. The recommendations simply answer business questions to increase revenue and provide suggestions for improvement at the hotel for enhanced customer satisfaction.

2. Project Scope and Objective

Scope of the project is based on the extensiveness of the data set and is two-fold. Firstly, location of hotels, which in this case is limited to all properties in the United States. Secondly, span of time for which data has been collected, which in this case is February 2014 through January 2015. The data set that has been specifically selected for analysis has been obtained after pre-processing.

Objective of the project is to deliver suggestions for best practices at Hyatt hotel locations with the main purpose of increasing revenue and customer satisfaction. The project will consist of using various data analysis models to generate correlational trends between various parameters of the data set and providing accurate results for solution of business questions that will be formulated subsequently.

3. Deliverables

Starting from the very first phase, the following are deliverables of the project:

- Initial processing of data for selection of most suitable subset of data that has been used for processing. NPS_Type was calculated for all major states and was the basis for selection of appropriate data.

- A clean data set consisting of no discrepancies, inaccurate data or any kind of inconsistencies. All NA values in the data set have been either replaced with mean values for all values of that column or have been omitted.
- Identification of correlational trends and association rules using Apriori algorithm for identification of parameters that induce increase in revenue. One action leading to another, performed multiple times, is an important factor.
- Using various algorithms such as KSVM (Structured Vector Machines) and Linear Regression for formulating actionable insights that can be implemented for enhancing business practices.
- Recommendations of improvements to facilities and services provided by the hotel, obtained from insights formulated in the above deliverable, to improve its performance and revenue.

4. Data Requisition

Data was made available to us by the course instructors. Initially, we downloaded data for the time period of October 2014. Before any kind of data munging, this data set consisted of approximately 1.4 million rows and 237 variables. After studying this data, the team decided to increase the limits of data to further months, that is, data for November 2014 through January 2015.

This data was extensively studied to determine usable variables. After initial analysis, the data set was forwarded to the preprocessing phase of the project for data munging.

Following is the code for data requisition:

```
Oct <- read.csv(file="out-201410.csv", header=TRUE, sep=",")
Nov <- read.csv(file="out-201411.csv", header=TRUE, sep=",")
Dec <- read.csv(file="out-201412.csv", header=TRUE, sep=",")
Jan <- read.csv(file="out-201501.csv", header=TRUE, sep=",")
```

Output obtained for the above code consists of data sets for the months from October 2014 through January 2015.

5. Data Preprocessing

Before preprocessing, the data set for October 2014 had 1433747 rows and 237 variables, November 2014 had 1303803 rows and 237 variables, December 2014 had 1216187 rows and 237 variables, and January 2015 had 1151316 rows and 237 variables. All these data sets contained information regarding various attributes of Hyatt hotels. The data sets contained numerous NA values, which were best dealt with by omitting them, as working on substantial values makes more sense than assuming some calculated values such as mean. Only the columns that are relevant to the analysis of data have been kept, while the remaining columns were eliminated.

In summary, the data preprocessing phase provides a subset of useable data for the project, consisting of only the columns that need to be worked with and eliminating Null/NA values. The code for data preprocessing is given below.

Following is the code for data preprocessing for October 2014:

```
Octdata <- Oct[-
c(1:18,20:55,57:65,67:82,84:106,108:136,148:167,172:181,183:195,197:198,199,201,203,204,209,210,211,213,214,
216,217,218,220,222,223,224,225,226,227,228:231,233:237)]
ncol(Octdata)
nrow(Octdata)
View(Octdata)
OctD <-Octdata[Octdata$Country_PL == "United States",]
OctD <-OctD[OctD$US.Region_PL == "Northeast",]
nrow(OctD)
#****************************** Null to NA ************************
nullToNA <- function(OctD){
  # split df into numeric & non-numeric functions
  a<-OctD[,sapply(df, is.numeric), drop = FALSE]
  b<-OctD[,sapply(df, Negate(is.numeric)), drop = FALSE]
  # Change empty strings to NA
  b<-b[lapply(b,function(x) levels(x) <- c(levels(x), NA) ),] # add NA level
  b<-b[lapply(b,function(x) x[x==""",]<- NA),]                # change Null to NA
  # Put the columns back together
  d<-cbind(a,b)
  d[, names(OctD)]
}
is.null(OctD)
View(OctD)
nrow(OctD)
#***************************************************************
OctCL<-OctD[complete.cases(OctD$Likelihood_Recommend_H),]
View(OctCL$Likelihood_Recommend_H)
nrow(OctCL)
is.na(OctCL$Likelihood_Recommend_H)
install.packages("NPS")
library("NPS")
OctNPS <- nps(OctCL$Likelihood_Recommend_H, breaks = list(0:6, 7:8, 9:10))
```

Following is the code for data preprocessing for November 2014:

```
ncol(Nov)
Novdata <- Nov[-
c(1:18,20:55,57:65,67:82,84:106,108:136,148:167,172:181,183:195,197:198,199,201,203,204,209,210,211,213,214,
216,217,218,220,222,223,224,225,226,227,228:231,233:237)]
ncol(Novdata)
nrow(Novdata)
View(Novdata)
NovD <-Novdata[Novdata$Country_PL == "United States",]
NovD <-NovD[NovD$US.Region_PL == "Northeast",]
View(NovD[,169])
nrow(NovD)
#***************************** Null to NA ************************
nullToNA <- function(NovD){
  # split df into numeric & non-numeric functions
  a<-NovD[,sapply(df, is.numeric), drop = FALSE]
  b<-NovD[,sapply(df, Negate(is.numeric)), drop = FALSE]
  # Change empty strings to NA
  b<-b[lapply(b,function(x) levels(x) <- c(levels(x), NA) ),] # add NA level
  b<-b[lapply(b,function(x) x[x==""",]<- NA),]                # change Null to NA
  # Put the columns back together
  d<-cbind(a,b)
  d[, names(NovD)]
}
is.null(NovD)
View(NovD)
nrow(NovD)
#**************************************************************
NovCL<-NovD[complete.cases(NovD$Likelihood_Recommend_H),]
View(NovCL$Likelihood_Recommend_H)
nrow(NovCL)
is.na(NovCL$Likelihood_Recommend_H)
install.packages("NPS")
library("NPS")
NovNPS <- nps(NovCL$Likelihood_Recommend_H, breaks = list(0:6, 7:8, 9:10))
NovNPS
```

Following is the code for data preprocessing for December 2014:

```
Decdata <- Dec[-
c(1:18,20:55,57:65,67:82,84:106,108:136,148:167,172:181,183:195,197:198,199,201,203,204,209,210,211,213,214,
216,217,218,220,222,223,224,225,226,227,228:231,233:237)]
ncol(Decdata)
nrow(Decdata)
View(Decdata)
DecD <-Decdata[Decdata$Country_PL == "United States",]
DecD <-DecD[DecD$US.Region_PL == "Northeast",]
nrow(DecD)
#***************************** Null to NA ************************
nullToNA <- function(DecD){
  # split df into numeric & non-numeric functions
  a<-DecD[,sapply(df, is.numeric), drop = FALSE]
  b<-DecD[,sapply(df, Negate(is.numeric)), drop = FALSE]
  # Change empty strings to NA
  b<-b[lapply(b,function(x) levels(x) <- c(levels(x), NA) ),] # add NA level
  b<-b[lapply(b,function(x) x[x==""",]<- NA),]                # change Null to NA
  # Put the columns back together
  d<-cbind(a,b)
```

```
  d[, names(DecD)]
}
is.null(DecD)
View(DecD)
nrow(DecD)
#***********************************************************
DecCL<-DecD[complete.cases(DecD$Likelihood_Recommend_H),]
View(DecCL$Likelihood_Recommend_H)
nrow(DecCL)
is.na(DecCL$Likelihood_Recommend_H)
install.packages("NPS")
library("NPS")
DecNPS <- nps(DecCL$Likelihood_Recommend_H, breaks = list(0:6, 7:8, 9:10))
DecNPS
```

Following is the code for data preprocessing for January 2015:

```
Jandata <- Jan[-
c(1:18,20:55,57:65,67:82,84:106,108:136,148:167,172:181,183:195,197:198,199,201,203,204,209,210,211,213,214,
216,217,218,220,222,223,224,225,226,227,228:231,233:237)]
ncol(Jandata)
nrow(Jandata)
View(Jandata)
JanD <-Jandata[Jandata$Country_PL == "United States",]
JanD <-JanD[JanD$US.Region_PL == "Northeast",]
nrow(JanD)
#**************************** Null to NA ************************
nullToNA <- function(JanD){
  # split df into numeric & non-numeric functions
  a<-JanD[,sapply(df, is.numeric), drop = FALSE]
  b<-JanD[,sapply(df, Negate(is.numeric)), drop = FALSE]
  # Change empty strings to NA
  b<-b[lapply(b,function(x) levels(x) <- c(levels(x), NA) ),] # add NA level
  b<-b[lapply(b,function(x) x[x=="",]<- NA),]                # change Null to NA
  # Put the columns back together
  d<-cbind(a,b)
  d[, names(JanD)]
}
is.null(JanD)
View(JanD)
nrow(JanD)
#***********************************************************
JanCL<-JanD[complete.cases(JanD$Likelihood_Recommend_H),]
View(JanCL$Likelihood_Recommend_H)
nrow(JanCL)
is.na(JanCL$Likelihood_Recommend_H)
install.packages("NPS")
library("NPS")
JanNPS <- nps(JanCL$Likelihood_Recommend_H, breaks = list(0:6, 7:8, 9:10))
JanNPS
```

Output of the above four code snippets consists of four clean data sets for the months from October 2014 through January 2015. These data sets consists of only the required variables and rows, and tackle the issue of NA values.

Data munging was also performed for transforming and mapping raw data into usable format for the project, enabling easy data analytics on it. After preprocessing for cleaning the data and data munging, the data set for October 2014 had 12074 rows and 33 variables, November 2014 had 10040 rows and 33 variables, December 2014 had 9567 rows and 33 variables, and January 2015 had 3684 rows and 33 variables. While the variables for all data sets remain constant, the amount of information varies with the number of rows available. Some examples of variables used in the data sets include Tranquility_H, Overall_Sat_H, Condition_Hotel_H, Customer_SVC_H, Staff_Cared_H, Guest_Room_H, Check_In_H, and F.B_Overall_Experience_H.

6. Initial Phase

Initial phase of the project started with selection of a single prime attribute for which the Net Promoter Score (NPS) can be calculated. Net Promoter Score is an index that measures the willingness of customers to recommend Hyatt hotels. It is a measure of the customers' overall satisfaction with the services provided to them and their loyalty towards the brand.

The attribute 'Likelihood_Recommend_H' which signifies the likelihood of a customer to recommend the hotel on a scale of 1-10, had been selected to calculate the NPS score. The following metrics were used:

Rating of 9-10: Promoters (fully satisfied customers)

Rating of 6-8: Passive (somewhat satisfied customers)

Rating of 1-5: Detractors (unsatisfied customers)

NPS Score quantifies the overall customer satisfaction that can be transcended into incorporation of a successful business model. Based on the NPS Score, regions within the United States were compared. On comparison of the North-East, North-West, South-East, and South-West regions, it was found that the North-East region had the lowest NPS Score, and hence this region was selected for further analysis.

7. Modeling

Various models have been used for accurate results of the information obtained from the data set. These models give an understandable representation of real-life information from the data sets. The following models have been implemented:

7.1 Association Rules

The Apriori Algorithm is an influential algorithm for mining frequent item sets for Boolean association rules. Apriori is designed to operate on data sets containing transactions.

The code for Apriori algorithm implemented in our project is as follows:

```
#************************************************** ARules*****************************************************
CombD_AP1 <- CombD12
View(CombD_AP1)
typeof(CombD_AP1$CombD1.Tranquility_H)
for(i in 1:ncol(CombD_AP1)){
  CombD_AP1[is.na(CombD_AP1[,i]), i] <- mean(CombD_AP1[,i], na.rm = TRUE)
}
View(CombD_AP1)
CombD_AP1$Tranquility_H <- ifelse(CombD_AP1$Tranquility_H < 6 , "low", ifelse(CombD_AP1$Tranquility_H >= 8,"high","medium") )
CombD_AP1$Guest_Room_H <- ifelse(CombD_AP1$Guest_Room_H < 6 , "low",ifelse(CombD_AP1$Guest_Room_H>=8,"high","medium") )
CombD_AP1$Condition_Hotel_H <- ifelse(CombD_AP1$Condition_Hotel_H < 6 , "low",ifelse(CombD_AP1$Condition_Hotel_H>=8,"high","medium") )
CombD_AP1$Customer_SVC_H <- ifelse(CombD_AP1$Customer_SVC_H < 6 , "low",ifelse(CombD_AP1$Customer_SVC_H>=8,"high","medium") )
CombD_AP1$Staff_Cared_H <- ifelse(CombD_AP1$Staff_Cared_H < 6 , "low",ifelse(CombD_AP1$Staff_Cared_H>=8,"high","medium") )
CombD_AP1$Check_In_H <- ifelse(CombD_AP1$Check_In_H < 6 , "low",ifelse(CombD_AP1$Check_In_H>=8,"high","medium") )
CombD_AP1$F.B_FREQ_H <- ifelse(CombD_AP1$F.B_FREQ_H < 2 , "low",ifelse(CombD_AP1$F.B_FREQ_H>=4,"high","medium") )
CombD_AP1$F.B_Overall_Experience_H <- ifelse(CombD_AP1$F.B_Overall_Experience_H < 6 , "low",ifelse(CombD_AP1$F.B_Overall_Experience_H>=8,"high","medium") )
CombD_AP1$Internet_Sat_H <- ifelse(CombD_AP1$Internet_Sat_H < 6 , "low",ifelse(CombD_AP1$Internet_Sat_H>=8,"high","medium") )
for(i in 1:ncol(CombD_AP1))
{
  CombD_AP1[,i] <- as.factor(CombD_AP1[,i])
}
View(CombD_AP1)
Sub_CombD_AP1 <- data.frame(CombD_AP1$NPS_Type, CombD_AP1$Guest_Room_H, CombD_AP1$Tranquility_H, CombD_AP1$Condition_Hotel_H, CombD_AP1$Customer_SVC_H, CombD_AP1$Staff_Cared_H, CombD_AP1$F.B_Overall_Experience_H, CombD_AP1$Internet_Sat_H, CombD_AP1$Business.Center_PL, CombD_AP1$Fitness.Center_PL, CombD_AP1$Convention_PL, CombD_AP1$Restaurant_PL)
View(Sub_CombD_AP1)
Sub_CombD_AP1[Sub_CombD_AP1==""] <- NA
Sub_CombD_AP1_N <- na.omit(Sub_CombD_AP1)
View(Sub_CombD_AP1_N)
CombD_AP1_rule_P <- apriori(Sub_CombD_AP1_N ,parameter = list(support=0.26,confidence=0.75),appearance = list(rhs="CombD_AP1.NPS_Type=Promoter",default="lhs"))
```

```
CombD_AP1_rule_P
CombD_AP1_rule_N <- apriori(Sub_CombD_AP1_N ,parameter = list(support=0.3,confidence=0.65),appearance =
list(rhs="CombD_AP1.NPS_Type=Detractor",default="lhs"))
CombD_AP1_rule_N
CombD_AP1Transaction <- as(CombD_AP1,'transactions')
itemFrequencyPlot(CombD_AP1Transaction,support=0.3,col='red')
inspect(CombD_AP1_rule) plot(CombD_AP1_rule_P)
goodRules <- CombD_AP1_rule_P[quality(CombD_AP1_rule_P)$lift>1.329]
 goodRules
  plot(goodRules)
#***************************************************************************
```

The output for Apriori algorithm implemented in our project is as follows:

```
> goodRules
set of 19 rules
> inspect(goodRules)
    lhs                                  rhs                           support confidence     lift count
[1]  {CombD_AP1.Guest_Room_H=high,
      CombD_AP1.Tranquility_H=high,
      CombD_AP1.Condition_Hotel_H=high,
      CombD_AP1.Customer_SVC_H=high,
      CombD_AP1.F.B_Overall_Experience_H=high,
      CombD_AP1.Convention_PL=N}                       => {CombD_AP1.NPS_Type=Promoter} 0.4054482
0.8789512 1.331065 10761
[2]  {CombD_AP1.Guest_Room_H=high,
      CombD_AP1.Tranquility_H=high,
      CombD_AP1.Condition_Hotel_H=high,
      CombD_AP1.Customer_SVC_H=high,
      CombD_AP1.Staff_Cared_H=high,
      CombD_AP1.F.B_Overall_Experience_H=high} => {CombD_AP1.NPS_Type=Promoter} 0.5624505  0.8800849
1.332782 14928
[3]  {CombD_AP1.Guest_Room_H=high,
      CombD_AP1.Tranquility_H=high,
      CombD_AP1.Condition_Hotel_H=high,
      CombD_AP1.Customer_SVC_H=high,
      CombD_AP1.Staff_Cared_H=high,
      CombD_AP1.F.B_Overall_Experience_H=high,
CombD_AP1.Convention_PL=N}              => {CombD_AP1.NPS_Type=Promoter} 0.4015674  0.8841878 1.338995
10658
[4]  {CombD_AP1.Guest_Room_H=high,
      CombD_AP1.Tranquility_H=high,
      CombD_AP1.Condition_Hotel_H=high,
      CombD_AP1.Customer_SVC_H=high,
      CombD_AP1.F.B_Overall_Experience_H=high,
      CombD_AP1.Fitness.Center_PL=Y,
      CombD_AP1.Convention_PL=N}             => {CombD_AP1.NPS_Type=Promoter} 0.4054482  0.8789512
1.331065 10761
[5]  {CombD_AP1.Guest_Room_H=high,
      CombD_AP1.Tranquility_H=high,
      CombD_AP1.Condition_Hotel_H=high,
      CombD_AP1.Customer_SVC_H=high,
      CombD_AP1.F.B_Overall_Experience_H=high,
      CombD_AP1.Internet_Sat_H=high,
      CombD_AP1.Convention_PL=N}             => {CombD_AP1.NPS_Type=Promoter} 0.4054482  0.8789512
1.331065 10761
[6]  {CombD_AP1.Guest_Room_H=high,
      CombD_AP1.Tranquility_H=high,
      CombD_AP1.Condition_Hotel_H=high,
      CombD_AP1.Customer_SVC_H=high,
      CombD_AP1.Staff_Cared_H=high,
```

```
     CombD_AP1.F.B_Overall_Experience_H=high,
     CombD_AP1.Business.Center_PL=Y}         => {CombD_AP1.NPS_Type=Promoter} 0.5317433  0.8789313
1.331035 14113
[7]  {CombD_AP1.Guest_Room_H=high,
     CombD_AP1.Tranquility_H=high,
     CombD_AP1.Condition_Hotel_H=high,
     CombD_AP1.Customer_SVC_H=high,
     CombD_AP1.Staff_Cared_H=high,
     CombD_AP1.F.B_Overall_Experience_H=high,
     CombD_AP1.Fitness.Center_PL=Y}          => {CombD_AP1.NPS_Type=Promoter} 0.5624505  0.8800849
1.332782 14928
[8]  {CombD_AP1.Guest_Room_H=high,
     CombD_AP1.Tranquility_H=high,
     CombD_AP1.Condition_Hotel_H=high,
     CombD_AP1.Customer_SVC_H=high,
     CombD_AP1.Staff_Cared_H=high,
     CombD_AP1.F.B_Overall_Experience_H=high,
     CombD_AP1.Internet_Sat_H=high}          => {CombD_AP1.NPS_Type=Promoter} 0.5624505  0.8800849
1.332782 14928
[9]  {CombD_AP1.Guest_Room_H=high,
     CombD_AP1.Tranquility_H=high,
     CombD_AP1.Condition_Hotel_H=high,
     CombD_AP1.Customer_SVC_H=high,
     CombD_AP1.Staff_Cared_H=high,
     CombD_AP1.F.B_Overall_Experience_H=high,
     CombD_AP1.Business.Center_PL=Y,
     CombD_AP1.Convention_PL=N}              => {CombD_AP1.NPS_Type=Promoter} 0.3708602  0.8828594
1.336983  9843
[10] {CombD_AP1.Guest_Room_H=high,
CombD_AP1.Tranquility_H=high,
CombD_AP1.Condition_Hotel_H=high,
CombD_AP1.Customer_SVC_H=high,
CombD_AP1.Staff_Cared_H=high,
CombD_AP1.F.B_Overall_Experience_H=high,
CombD_AP1.Fitness.Center_PL=Y,
CombD_AP1.Convention_PL=N}              => {CombD_AP1.NPS_Type=Promoter} 0.4015674  0.8841878 1.338995
10658

[11] {CombD_AP1.Guest_Room_H=high,
     CombD_AP1.Tranquility_H=high,
     CombD_AP1.Condition_Hotel_H=high,
     CombD_AP1.Customer_SVC_H=high,
     CombD_AP1.Staff_Cared_H=high,
     CombD_AP1.F.B_Overall_Experience_H=high,
     CombD_AP1.Internet_Sat_H=high,
     CombD_AP1.Convention_PL=N}              => {CombD_AP1.NPS_Type=Promoter} 0.4015674  0.8841878
1.338995 10658
[12] {CombD_AP1.Guest_Room_H=high,
     CombD_AP1.Tranquility_H=high,
     CombD_AP1.Condition_Hotel_H=high,
     CombD_AP1.Customer_SVC_H=high,
     CombD_AP1.F.B_Overall_Experience_H=high,
     CombD_AP1.Internet_Sat_H=high,
     CombD_AP1.Fitness.Center_PL=Y,
     CombD_AP1.Convention_PL=N}              => {CombD_AP1.NPS_Type=Promoter} 0.4054482  0.8789512
1.331065 10761
[13] {CombD_AP1.Guest_Room_H=high,
     CombD_AP1.Tranquility_H=high,
     CombD_AP1.Condition_Hotel_H=high,
     CombD_AP1.Customer_SVC_H=high,
     CombD_AP1.Staff_Cared_H=high,
     CombD_AP1.F.B_Overall_Experience_H=high,
```

```
    CombD_AP1.Business.Center_PL=Y,
    CombD_AP1.Fitness.Center_PL=Y}          => {CombD_AP1.NPS_Type=Promoter} 0.5317433  0.8789313
1.331035 14113
[14] {CombD_AP1.Guest_Room_H=high,
    CombD_AP1.Tranquility_H=high,
    CombD_AP1.Condition_Hotel_H=high,
    CombD_AP1.Customer_SVC_H=high,
    CombD_AP1.Staff_Cared_H=high,
    CombD_AP1.F.B_Overall_Experience_H=high,
    CombD_AP1.Internet_Sat_H=high,
    CombD_AP1.Business.Center_PL=Y}         => {CombD_AP1.NPS_Type=Promoter} 0.5317433  0.8789313
1.331035 14113
[15] {CombD_AP1.Guest_Room_H=high,
    CombD_AP1.Tranquility_H=high,
    CombD_AP1.Condition_Hotel_H=high,
    CombD_AP1.Customer_SVC_H=high,
    CombD_AP1.Staff_Cared_H=high,
    CombD_AP1.F.B_Overall_Experience_H=high,
    CombD_AP1.Internet_Sat_H=high,
    CombD_AP1.Fitness.Center_PL=Y}          => {CombD_AP1.NPS_Type=Promoter} 0.5624505  0.8800849
1.332782 14928
[16] {CombD_AP1.Guest_Room_H=high,
    CombD_AP1.Tranquility_H=high,
    CombD_AP1.Condition_Hotel_H=high,
    CombD_AP1.Customer_SVC_H=high,
    CombD_AP1.Staff_Cared_H=high,
    CombD_AP1.F.B_Overall_Experience_H=high,
    CombD_AP1.Business.Center_PL=Y,
    CombD_AP1.Fitness.Center_PL=Y,
    CombD_AP1.Convention_PL=N}              => {CombD_AP1.NPS_Type=Promoter} 0.3708602  0.8828594
1.336983  9843

[17] {CombD_AP1.Guest_Room_H=high,
    CombD_AP1.Tranquility_H=high,
    CombD_AP1.Condition_Hotel_H=high,
    CombD_AP1.Customer_SVC_H=high,
    CombD_AP1.Staff_Cared_H=high,
    CombD_AP1.F.B_Overall_Experience_H=high,
    CombD_AP1.Internet_Sat_H=high,
    CombD_AP1.Business.Center_PL=Y,
    CombD_AP1.Convention_PL=N}              => {CombD_AP1.NPS_Type=Promoter} 0.3708602  0.8828594
1.336983  9843
[18] {CombD_AP1.Guest_Room_H=high,
    CombD_AP1.Tranquility_H=high,
    CombD_AP1.Condition_Hotel_H=high,
    CombD_AP1.Customer_SVC_H=high,
    CombD_AP1.Staff_Cared_H=high,
    CombD_AP1.F.B_Overall_Experience_H=high,
    CombD_AP1.Internet_Sat_H=high,
    CombD_AP1.Fitness.Center_PL=Y,
    CombD_AP1.Convention_PL=N}              => {CombD_AP1.NPS_Type=Promoter} 0.4015674  0.8841878
1.338995 10658
[19] {CombD_AP1.Guest_Room_H=high,
    CombD_AP1.Tranquility_H=high,
    CombD_AP1.Condition_Hotel_H=high,
    CombD_AP1.Customer_SVC_H=high,
    CombD_AP1.Staff_Cared_H=high,
    CombD_AP1.F.B_Overall_Experience_H=high,
    CombD_AP1.Internet_Sat_H=high,
    CombD_AP1.Business.Center_PL=Y,
    CombD_AP1.Fitness.Center_PL=Y}          => {CombD_AP1.NPS_Type=Promoter} 0.5317433  0.8789313
1.331035 14113
```

The visualization for Apriori algorithm implemented in our project is as follows:

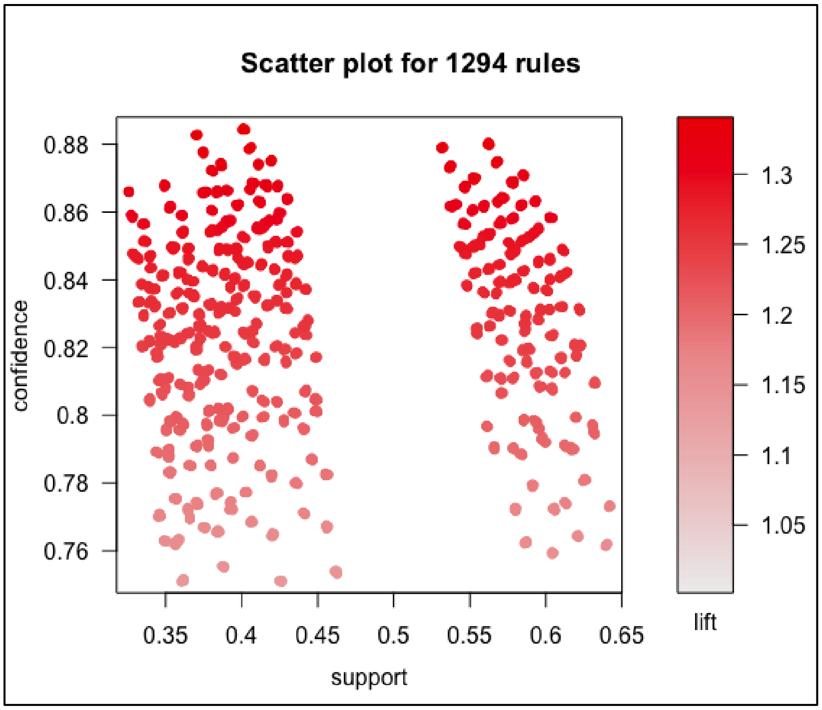

Fig.: Scatter Plot for Association Rules

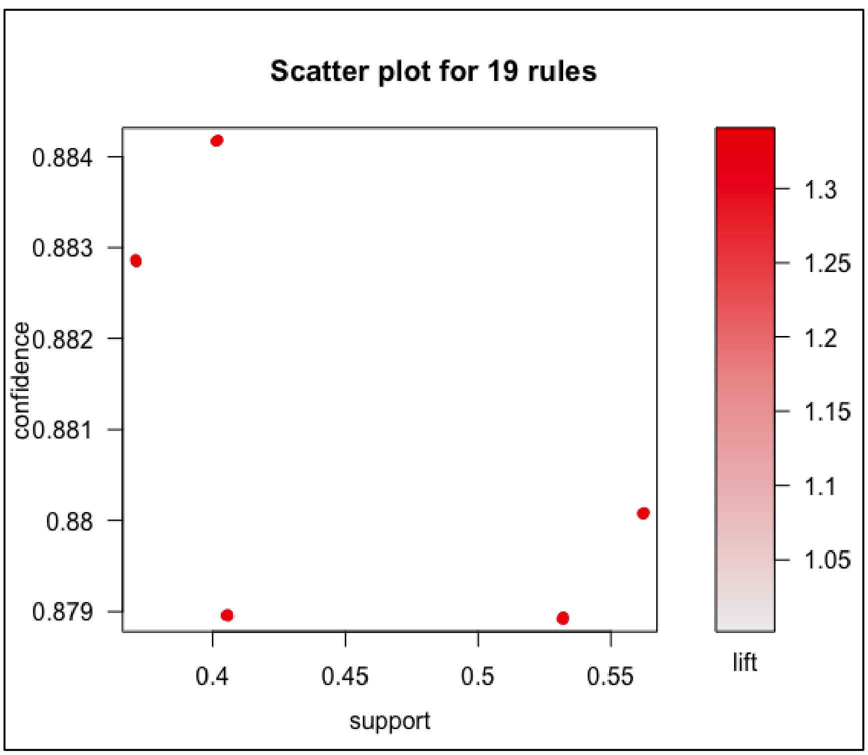

Fig.: Scatter Plot for Strongest Association Rules

All 19 associative rules thus observed are for the promoters, with a Support of 0.3, a Confidence of 0.65, and a Lift factor greater than 1.329. As we plot the rules shown above, we obtain a scatter plot showing the strongest associations among the rules obtained.

7.2 KSVM

Support Vector Machines are an excellent tool for classification, novelty detection, and regression.

The code for KSVM algorithm implemented in our project is as follows:

```
################ KSVM ######################
CombD1 <- read.csv(file = "CombD.csv")
View(CombD1)
str(CombD1)
nrow(CombD1)
CombD1[CombD1==""] <- NA
CombD1 <- na.omit(CombD1)
View(CombD1)
nrow(CombD1)
is.na(CombD1)
#Divide the data into separate rows
random_index1 <- sample(1:nrow(CombD1))
cutPnt1 <- floor(nrow(CombD1)/3*2)
#Create the test and train datasets for future computation
CombD1.train <- CombD1[random_index1[1:cutPnt1],]
CombD1.train
CombD1.test <- CombD1[random_index1[(cutPnt1+1):nrow(CombD1)],]
CombD1.test
nrow(CombD1.test)
nrow(CombD1)
View(CombD1.test)
#1) Build a model (using the 'ksvm' function, trying to predict onzone).
#Install and load packages e1071 and kernlab to build model
install.packages("e1071")
library(e1071)
install.packages("kernlab")
library(kernlab)
#Build the model
KSVMmodel2 <- ksvm(NPS_Type ~ Guest_Room_H + Tranquility_H + Condition_Hotel_H + Customer_SVC_H +
Staff_Cared_H + Check_In_H + F.B_FREQ_H + F.B_Overall_Experience_H , data=CombD1.train, kernel="rbfdot",
kpar="automatic", C=10, cross=10, prob.model=TRUE)
KSVMmodel2
KSVMmodel1 <- ksvm(NPS_Type ~ Guest_Room_H + Tranquility_H + Condition_Hotel_H + Customer_SVC_H +
Staff_Cared_H + Check_In_H + F.B_FREQ_H + F.B_Overall_Experience_H , data=CombD1.test, kernel="rbfdot",
kpar="automatic", C=10, cross=10, prob.model=TRUE)
KSVMmodel1
predictY1 <- predict(KSVMmodel1, CombD1.test)
predictY1
table(predictY1)
nrow(predictY1)
variance <- CombD1.test$NPS_Type
variance
table(variance)
nrow(variance)
df <- data.frame(predictY1,variance)
```

```
df<- na.omit(df)
nrow(df)
str(df)
df$predictY1 <- as.numeric(df$predictY1)
df$variance <- as.numeric(df$variance)
diff <- df$variance - df$predictY1
diff
plot_ksvm <- ggplot(CombD1.test,aes(x= CombD1.test$NPS_Type, y= (CombD1.test$Guest_Room_H)+(CombD1.test$Tranquility_H)+(CombD1.test$Condition_Hotel_H)+(CombD1.test$Customer_SVC_H)+(CombD1.test$Staff_Cared_H)+(CombD1.test$Check_In_H)+(CombD1.test$F.B_FREQ_H)+(CombD1.test$F.B_Overall_Experience_H) )) + geom_point(aes(colour= diff,size= diff)) + ylab("Factors affecting NPS") + xlab("NPS Type")
plot_ksvm
ctable <- data.frame(CombD1.test$NPS_Type,df$predictY1)
table(ctable)
################### End of KSVM ###############
```

The output for KSVM algorithm implemented in our project is as follows:

Training Data Output

```
Support Vector Machine object of class "ksvm"
SV type: C-svc  (classification)
 parameter : cost C = 10
Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  0.221943549467345
Number of Support Vectors : 1400
Objective Function Value : -1808.339 -1027.433 -6147.842
Training error : 0.102185
Cross validation error : 0.203089
Probability model included.
```

Test Data Output

```
Support Vector Machine object of class "ksvm"
SV type: C-svc  (classification)
 parameter : cost C = 10
Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  0.242143590632303
Number of Support Vectors : 712
Objective Function Value : -544.7166 -332.8583 -2541.368
Training error : 0.068766
Cross validation error : 0.209549
Probability model included.
```

Matrix

```
> ctable <- data.frame(CombD1.test$NPS_Type,df$predictY1)
> table(ctable)
> table(ctable)
                 df.predictY1
CombD1.test.NPS_Type    1    2    3
           Detractor  144    4    9
           Passive      0  228   77
           Promoter     3   14 1077
```

The visualization for KSVM algorithm implemented in our project is as follows:

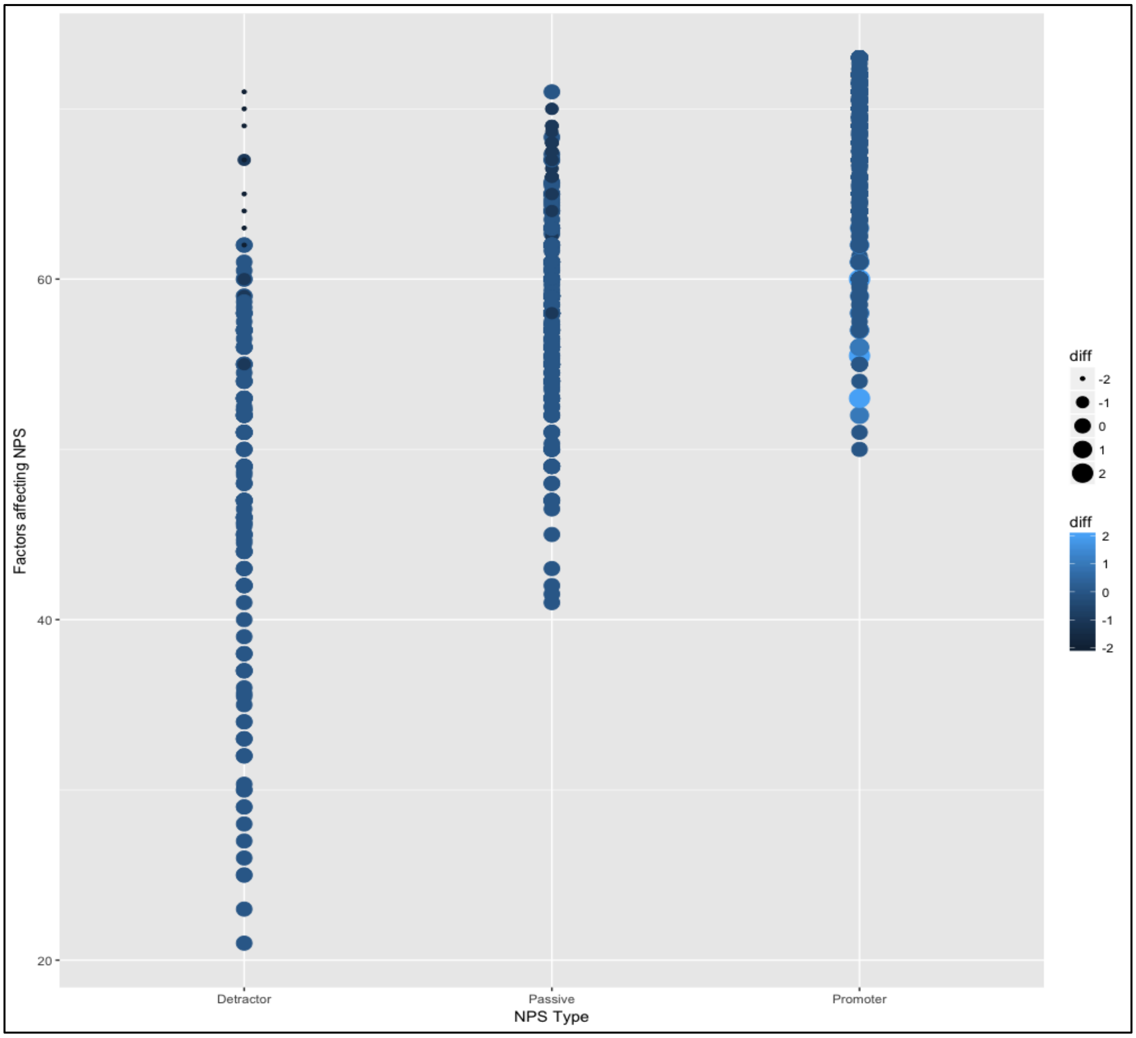

Fig.: Scatter Intensity Plot for KSVM

We have used KSVM to firstly, divide the data for four months, namely, October, November, December 2014, and January 2015, into essentially two parts: test data (Combd1.test) and train data (Combd1.train). The proportion of division was 2:1, two parts train data and one part test data. We then subsequently ran KSVM on test and train data, and plotted the difference between outputs of each the corresponding KSVM models.

Linear Modeling

Linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables denoted X.

The code for Linear Modeling algorithm implemented in our project is as follows:

```
################################## Linear Modeling ###############
CombD12 <- read.csv(file = "CombD.csv")
View(CombD12)
str(CombD1)
nrow(CombD1)
for(i in 1:ncol(CombD1))
{
  CombD1[,i] <- as.numeric(CombD1[,i])
}
View(CombD1)
Lm_All <- lm(formula = NPS_Type ~ Guest_Room_H + Tranquility_H + Condition_Hotel_H + Customer_SVC_H + Staff_Cared_H + Check_In_H + F.B_FREQ_H + F.B_Overall_Experience_H + Overall_Sat_H + Internet_Sat_H, data = CombD1)
Lm_All
summary(Lm_All)
Lm_All1 <- lm(formula = NPS_Type ~ Guest_Room_H + Tranquility_H + Condition_Hotel_H + Customer_SVC_H + Staff_Cared_H + Check_In_H + F.B_FREQ_H + F.B_Overall_Experience_H + Internet_Sat_H, data = CombD1)
Lm_All1
summary(Lm_All1)
Lm_All2 <- lm(formula = NPS_Type ~ Guest_Room_H + Tranquility_H + Condition_Hotel_H + Customer_SVC_H + Staff_Cared_H + F.B_FREQ_H + F.B_Overall_Experience_H + Internet_Sat_H, data = CombD1)
Lm_All2
summary(Lm_All2)
Lm_All3 <- lm(formula = NPS_Type ~ Guest_Room_H + Tranquility_H + Condition_Hotel_H + Customer_SVC_H + Staff_Cared_H + F.B_Overall_Experience_H + Internet_Sat_H, data = CombD1)
Lm_All3
summary(Lm_All3)
# NPS_Type - 0.5889
```

The visualization for Linear Modeling algorithm implemented in our project is as follows:

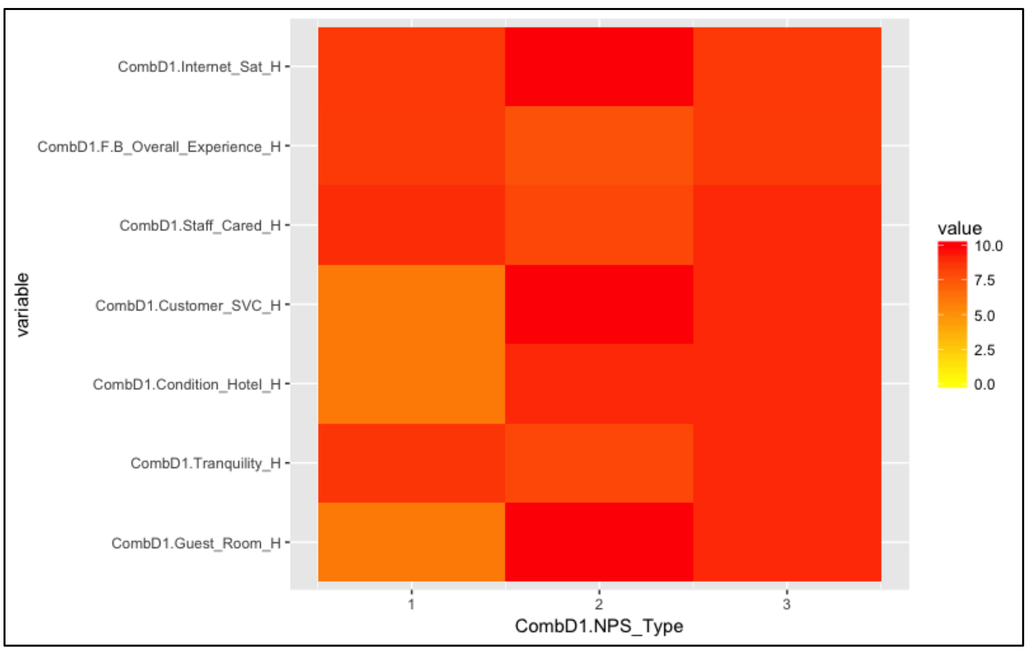

Fig.: Heat Map for Linear Modeling

The above graph represents an intensity heat map for factors affecting NPS bifurcated according to the NPS_Type.

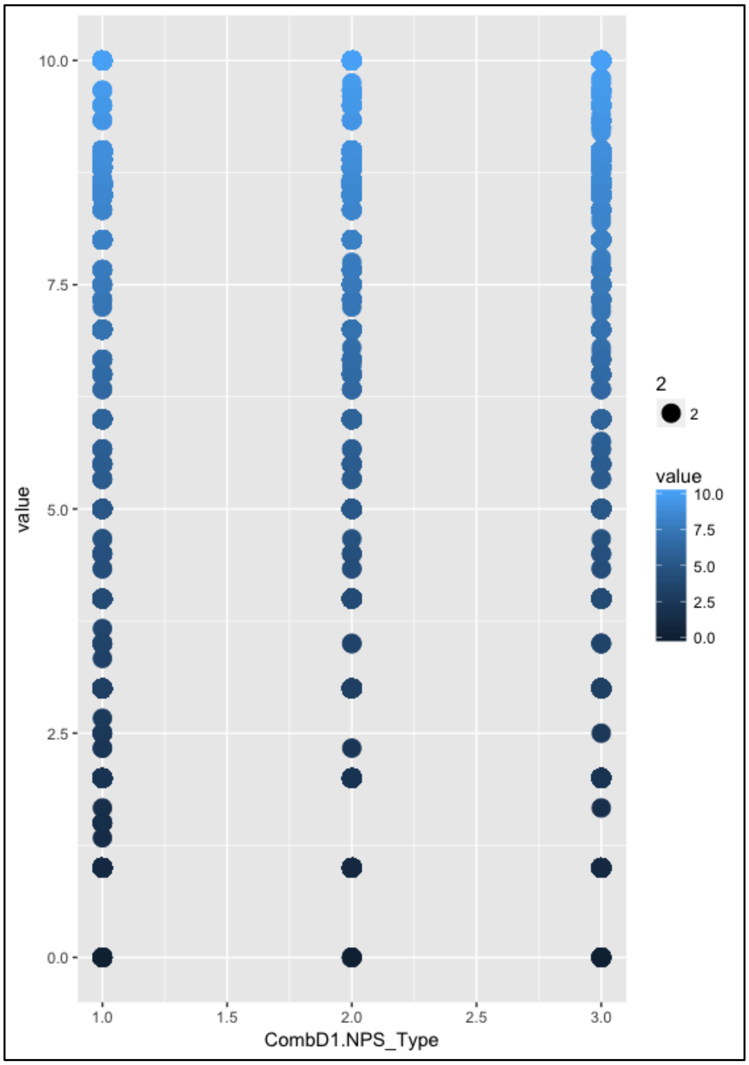

Fig.: Scatter Plot for Linear Modeling

The above graph represents the scatter plot for the values of recommendations arranged according to the NPS_Type, and the variations in shade of the data points depicts the strength of the recommendations.

Through the means of linear modeling, we have established interdependencies and linkages amidst the variable NPS_Type as mapped with different sets of variables namely Tranquility_H, Overall_Sat_H, Condition_Hotel_H, Customer_SVC_H, Staff_Cared_H, Guest_Room_H, Check_In_H, and F.B_Overall_Experience_H. Meticulously maneuvered in varying combinations, showcased using four models.

8. Business Questions

Feedback from every customer provides information that lead to the formulation of actionable insights. Analyzing this feedback would help the Hyatt Corporation improve their business practices and services, which in turn would result in higher revenues.

The data set consists of various attributes like tranquility, overall satisfaction, likelihood to recommend, etc., which have been put through algorithms for descriptive analytics and explored using data mining techniques.

Following are the business questions that have been identified and answered through the project:

1. What is the likelihood to recommend a specific hotel as per a particular state in the North-East region of the United States?
2. Out of the various brands of hotels under the Hyatt Corporation, which brand is better based on the likelihood to recommend?
3. What is the number of Promoters and Detractors in the categories of age, sex, country, purpose of visit?
4. What are the attributes that influence the likelihood of a customer to recommend a particular hotel?
5. What are the facilities and services provided at the hotel that need to be focused on in order to increase overall satisfaction of customers?

9. Descriptive Analysis

The various kinds of descriptive analysis that has been performed in the project are:

- Calculation of NPS for each state in the North-East region of the United States. The visualization for that is as follows:

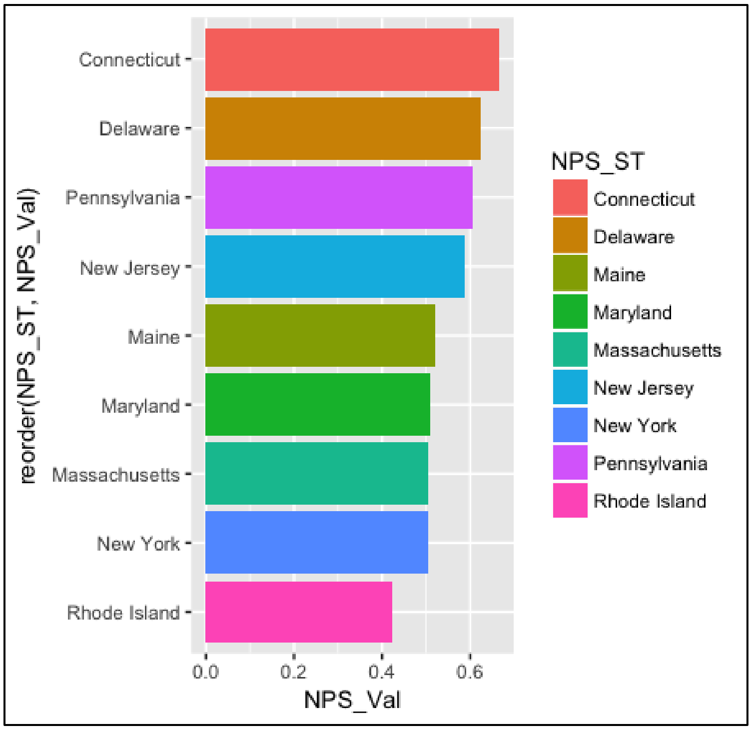

Fig.: NPS for States in North-East

- The finding from the above analysis is that Connecticut has the highest NPS while Rhode Island has the lowest. However, there exists only one record towards the calculation of score for Rhode Island, we decided to consider the second lowest NPS on the graph which is for New York State.
- Considering the various levels of NPS (high, medium and low), we further analyzed the states of Connecticut, New Jersey and New York.

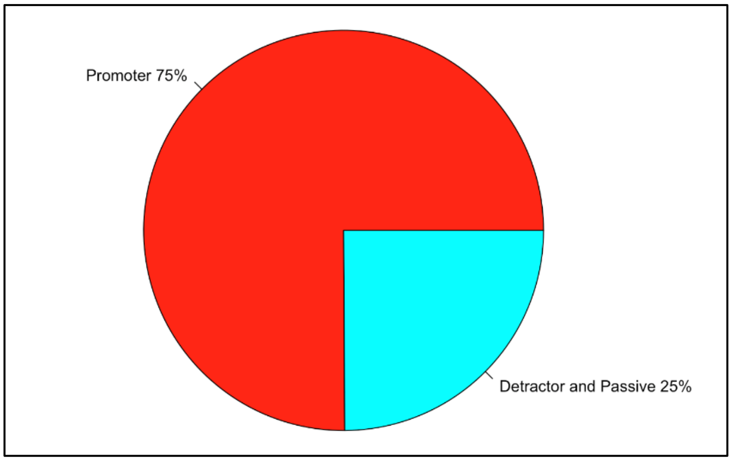

Fig.: Promoter, Detractors and Passives for Connecticut

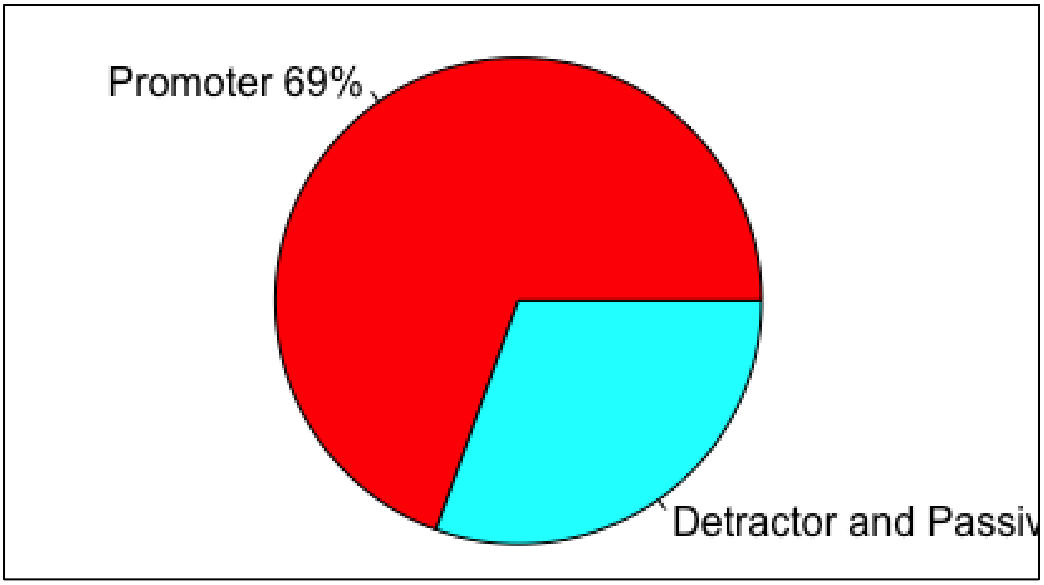

Fig.: Promoter, Detractors and Passives for New Jersey

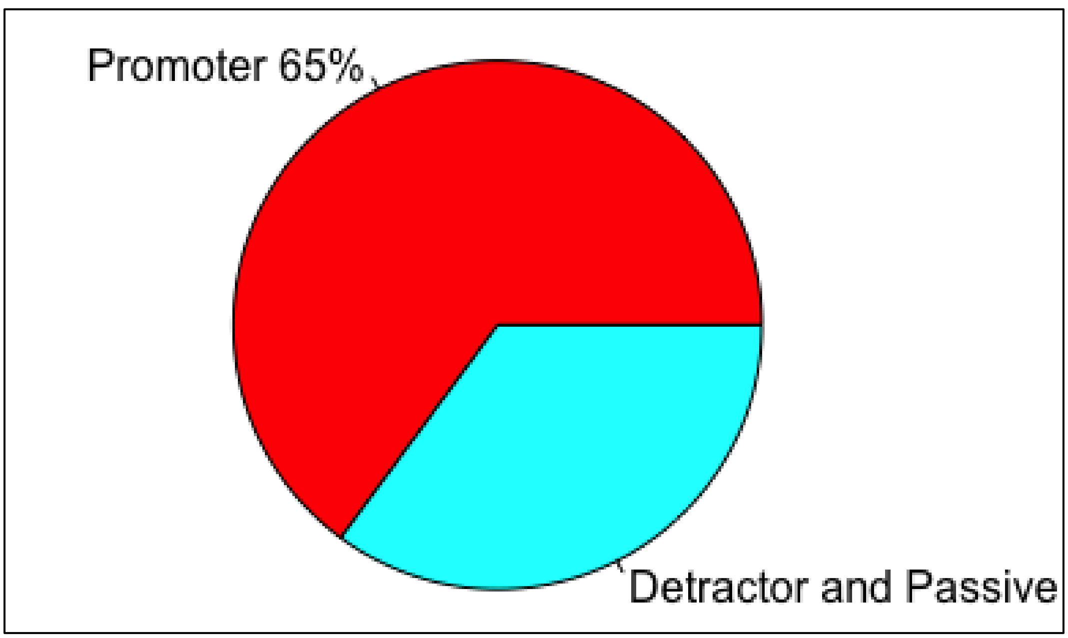

Fig.: Promoter, Detractors and Passives for New York

- Calculating the percentage of Promoters and Detractors on the basis of various brands under the Hyatt Corporation, in the North-East region.

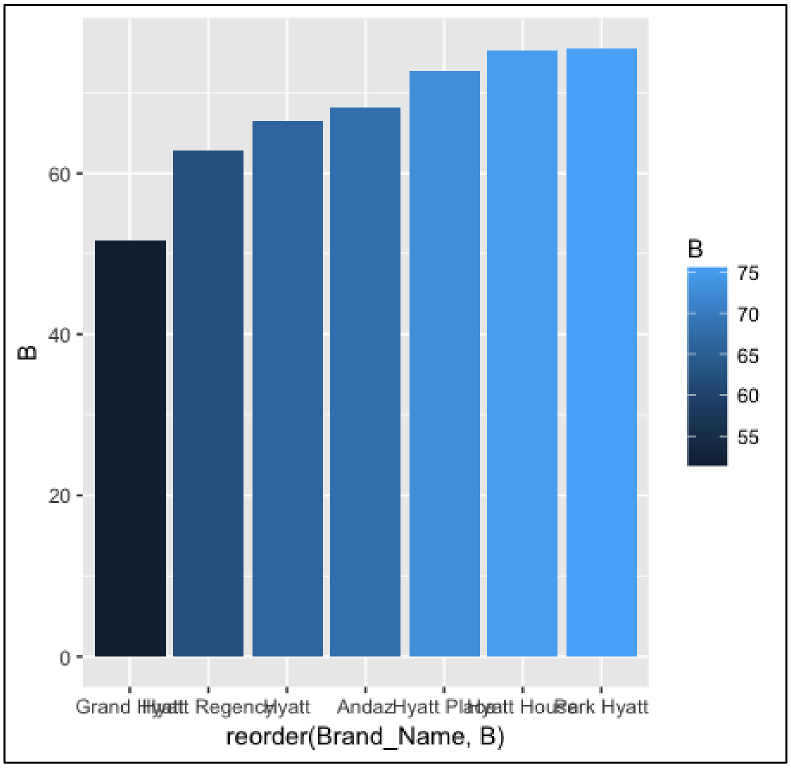

Fig.: Promoters for Hyatt Brands

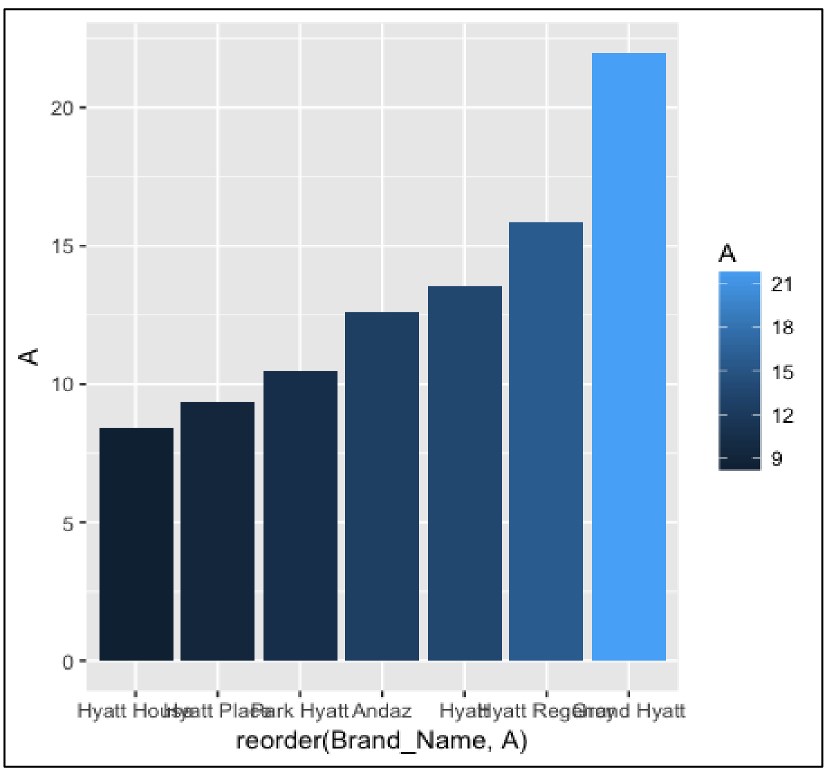

Fig.: Detractors for Hyatt Brands

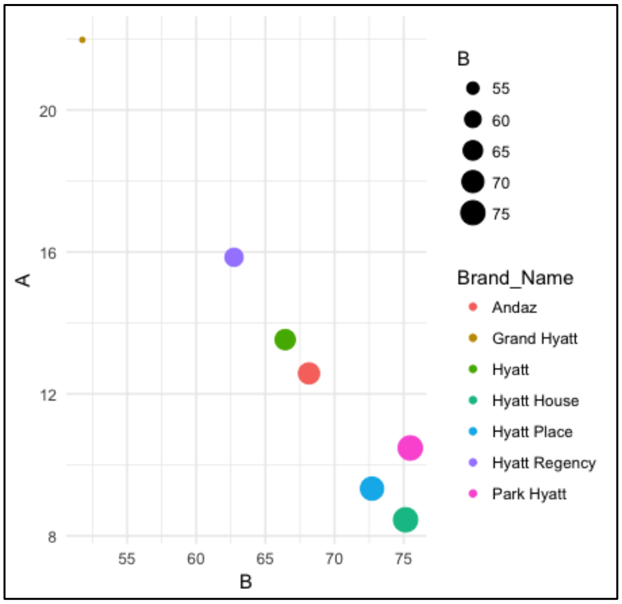

Fig.: Scatter Plot for Promoter and Detractors for Hyatt Brands

- Based on the above graphs, it is clearly visible that Hyatt House has proportionately the highest percentage of Promoters, and Grand Hyatt has the least. However, this may not have been the case in every state of the North-East region.
- Further, we looked into the best and worst brands of hotels in each of the above state that we have selected for the purpose of analysis, namely, Connecticut, New Jersey and New York. The following are the representations for the same:

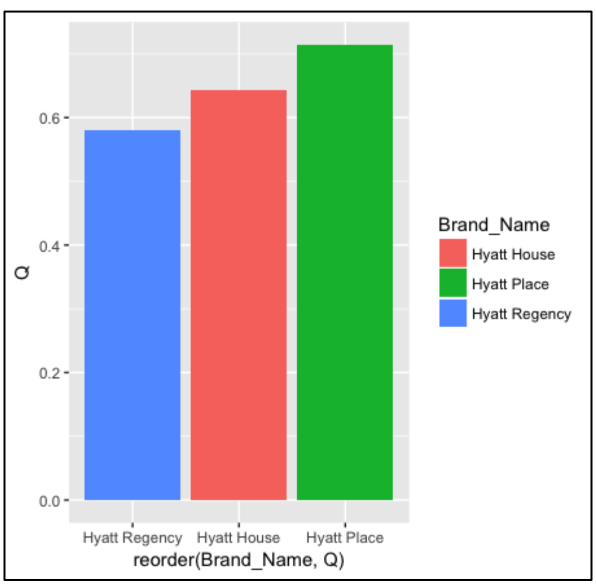

Fig.: Brand-wise Ranking for Connecticut State

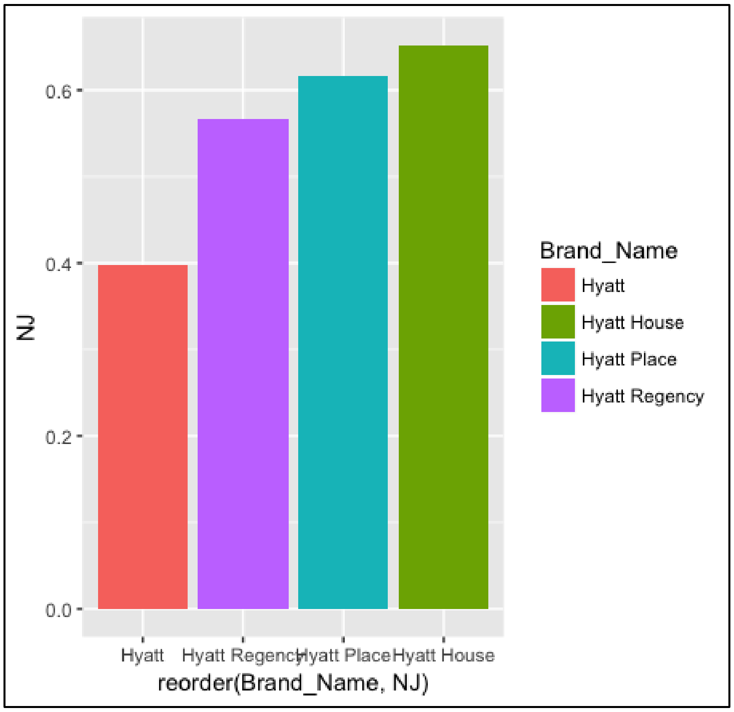

Fig.: Brand-wise Ranking for New Jersey State

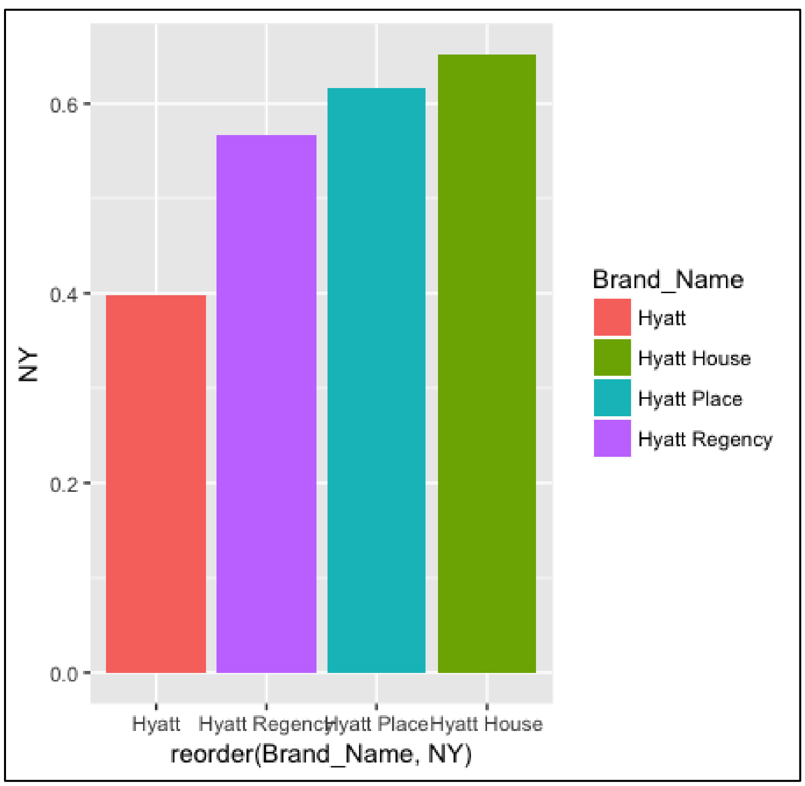

Fig.: Brand-wise Ranking for New York State

- Next, we explored three parameters, purpose of visit, gender and age of visitor, against the frequency of likelihood to recommend. These demographics gave us a clear understanding of the target customers for Hyatt Corporation.

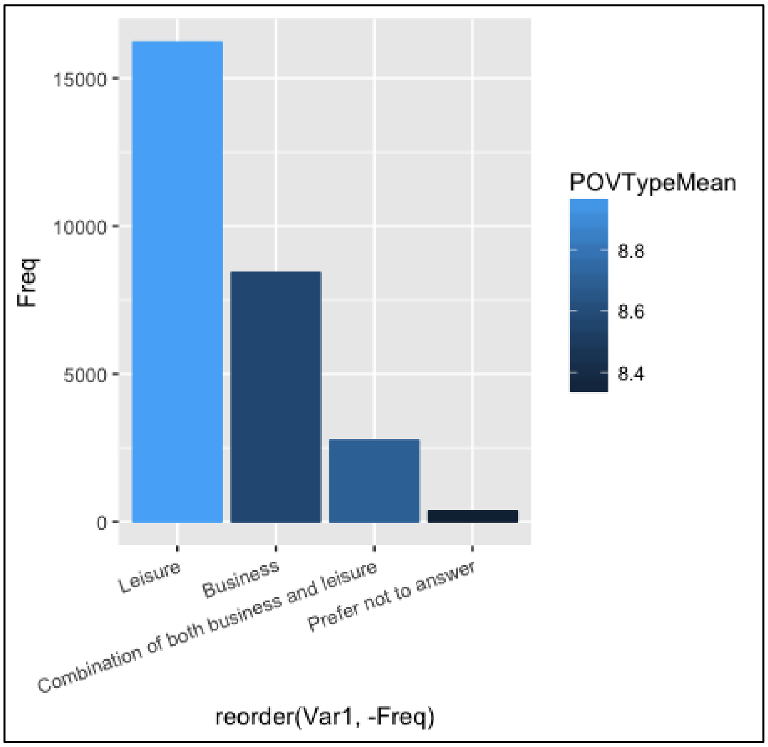

Fig.: Purpose of Visit

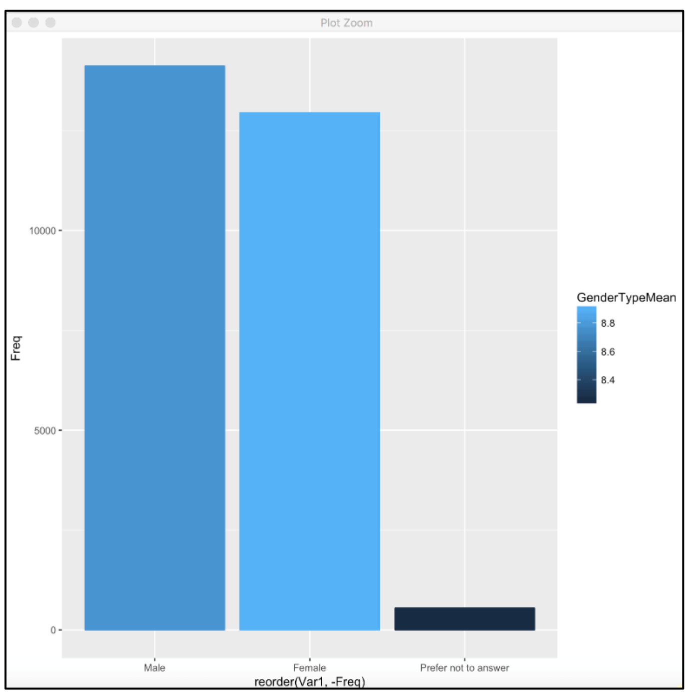

Fig.: Gender of Visitors

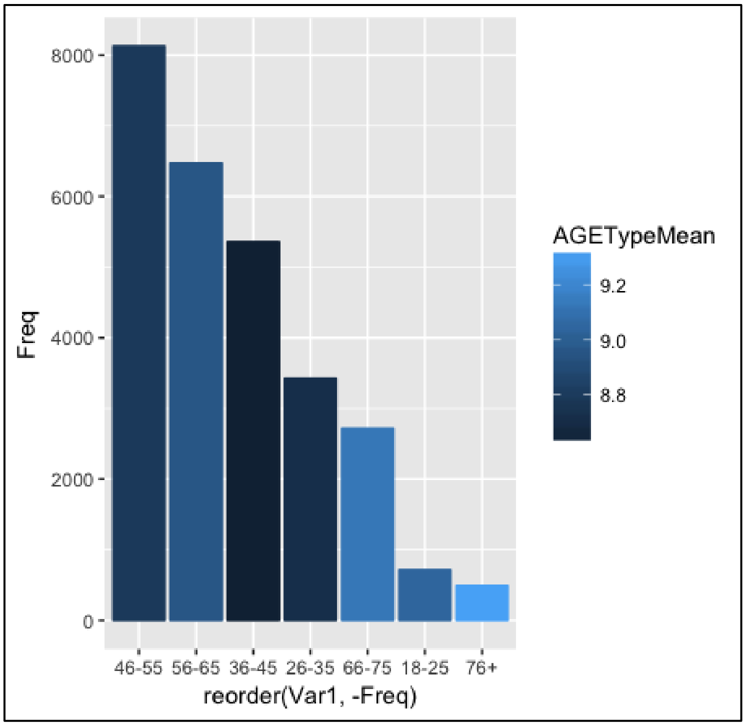

Fig.: Age Group of Visitors

Data Visualizations

Data visualizations have been done to graphically map all factors except the NPS separately on the North-East region of the United States in order to get a better understanding of the proportionality of these individual factors. This provides actionable insights about the factors that contribute the most towards a greater NPS Score.

[P. T. O.]

Following are the various maps that have been plotted:

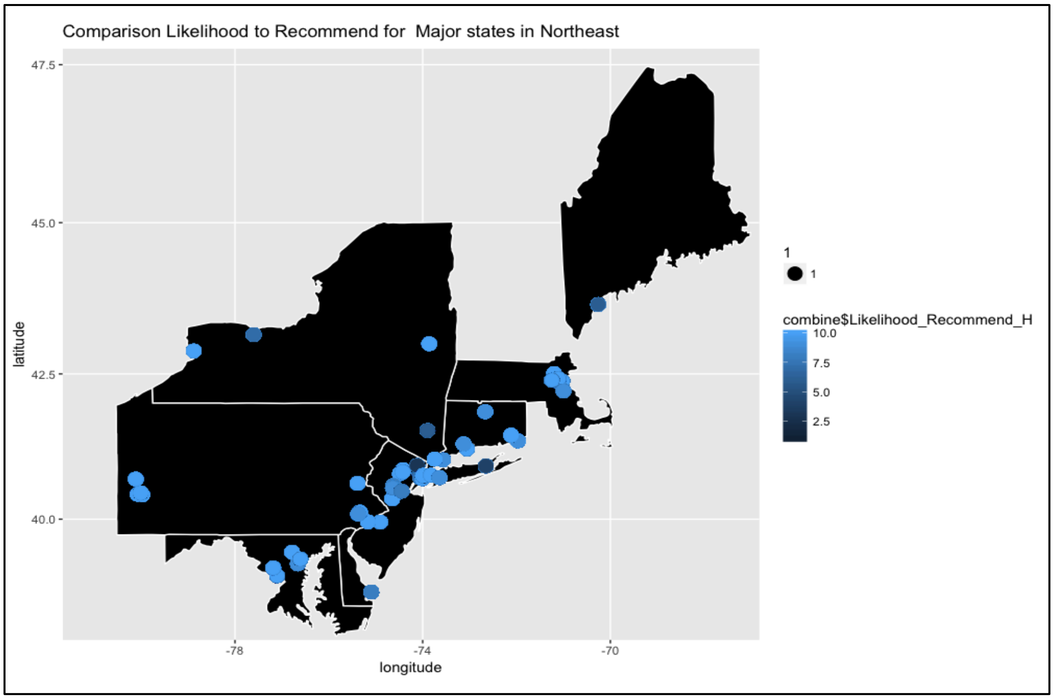

Fig.: Likelihood to Recommend for Major States in North-East

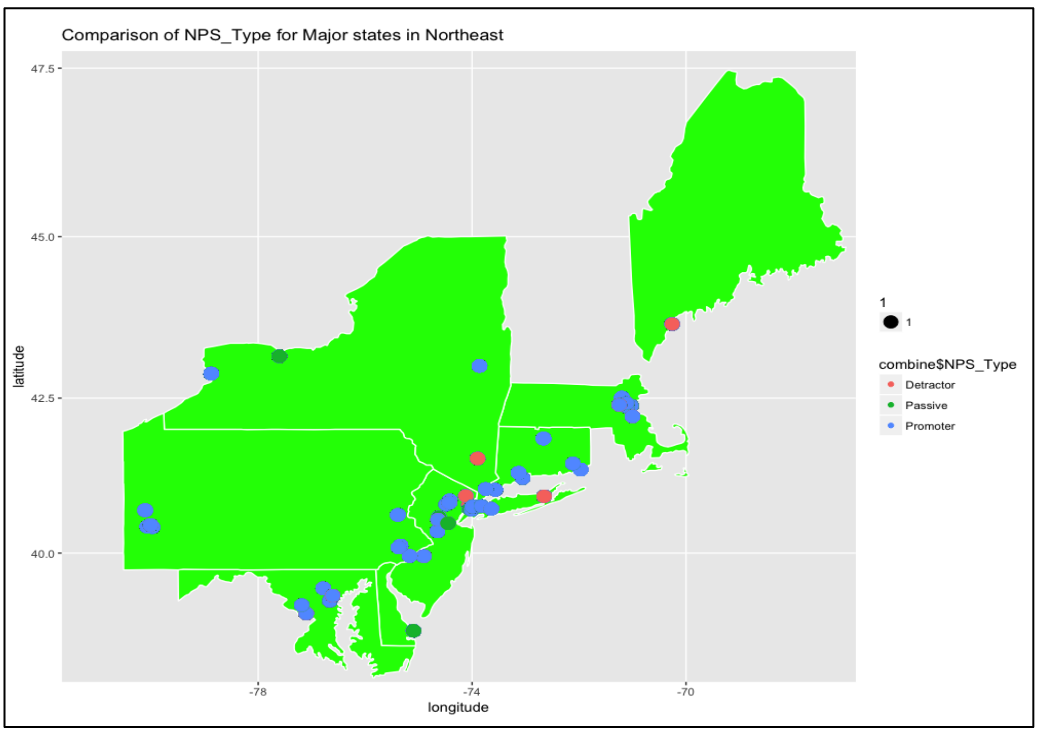

Fig.: NPS_Type for Major States in North-East

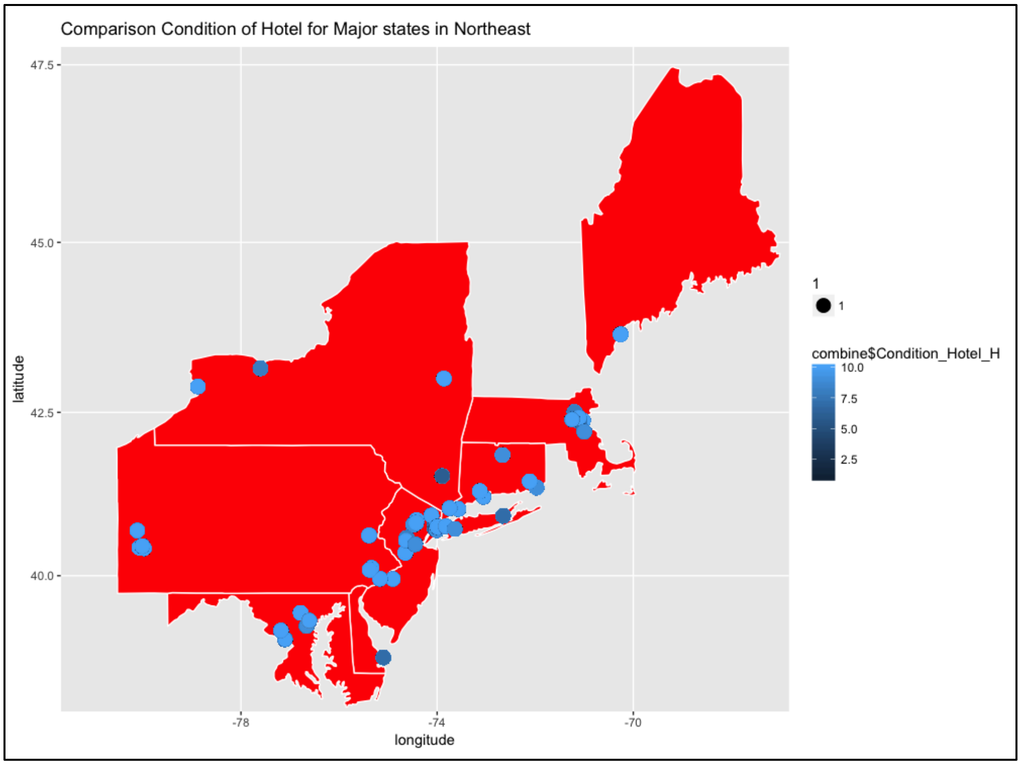

Fig.: Condition of Hotel for Major States in North-East

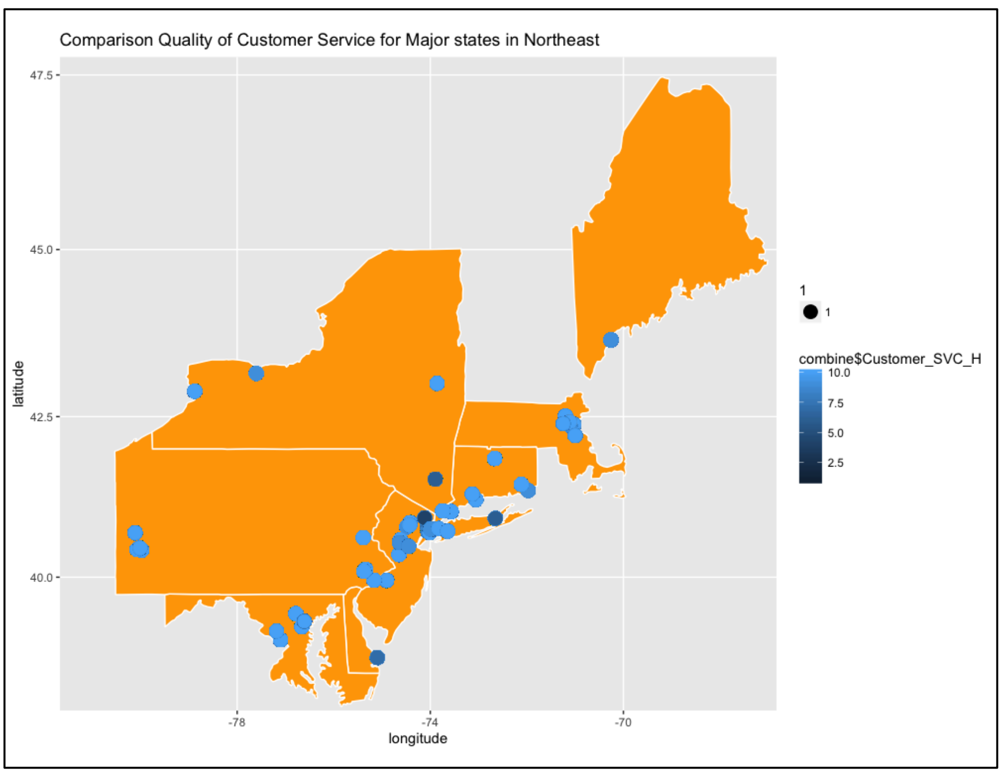

Fig.: Quality of Customer Service for Major States in North-East

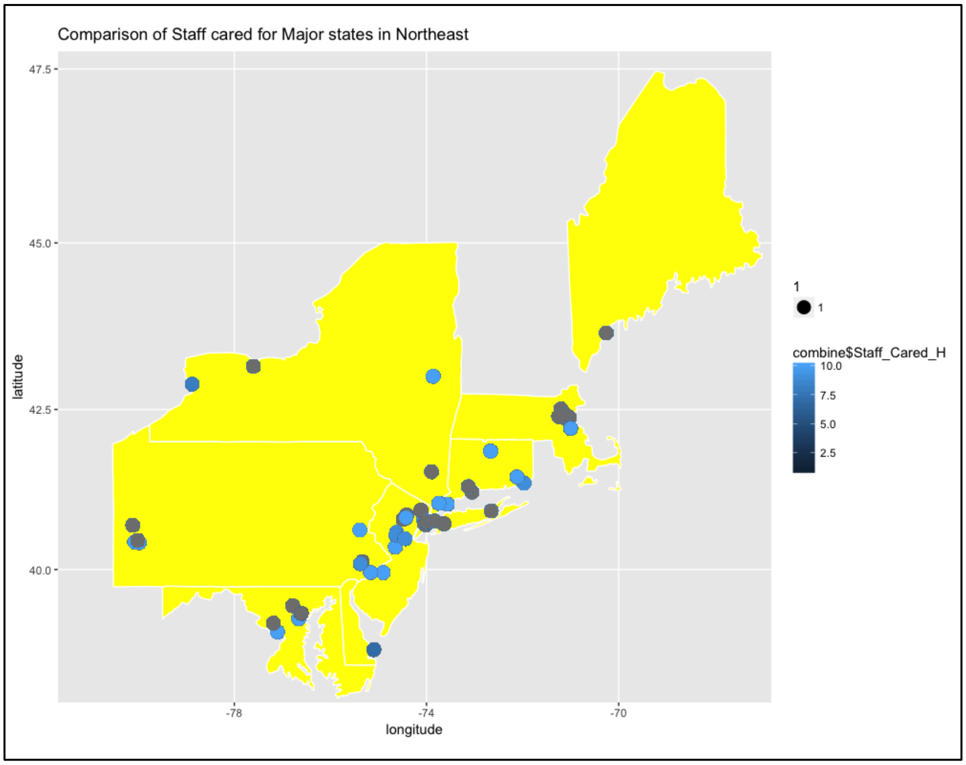

Fig.: Staff Hospitality Rating for Major States in North-East

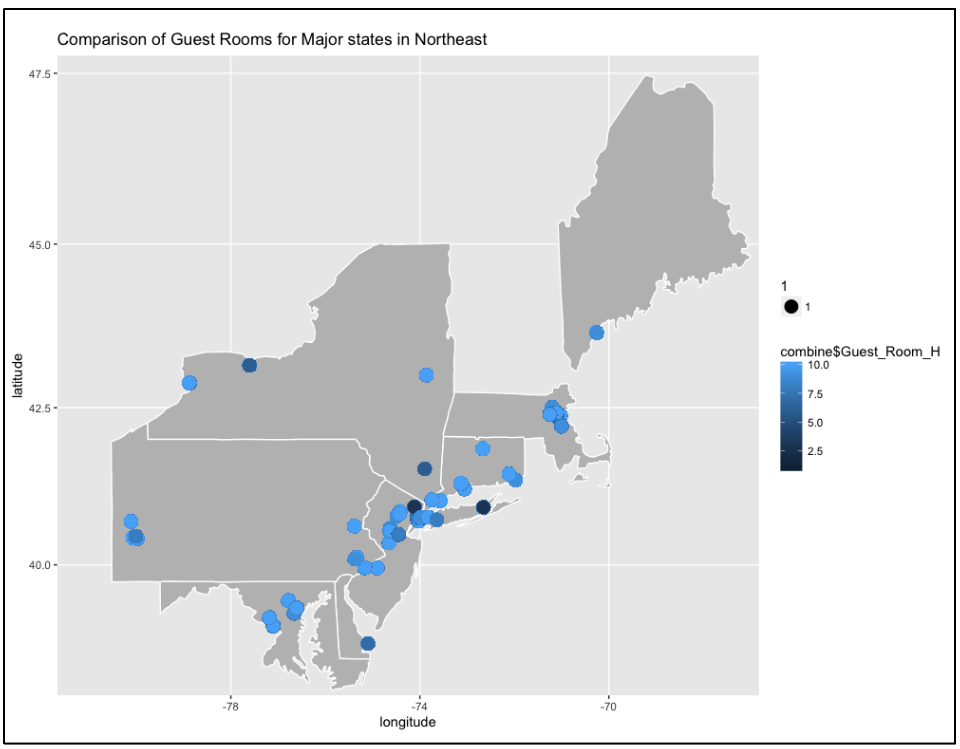

Fig.: Comparison of Guest Rooms for Major States in North-East

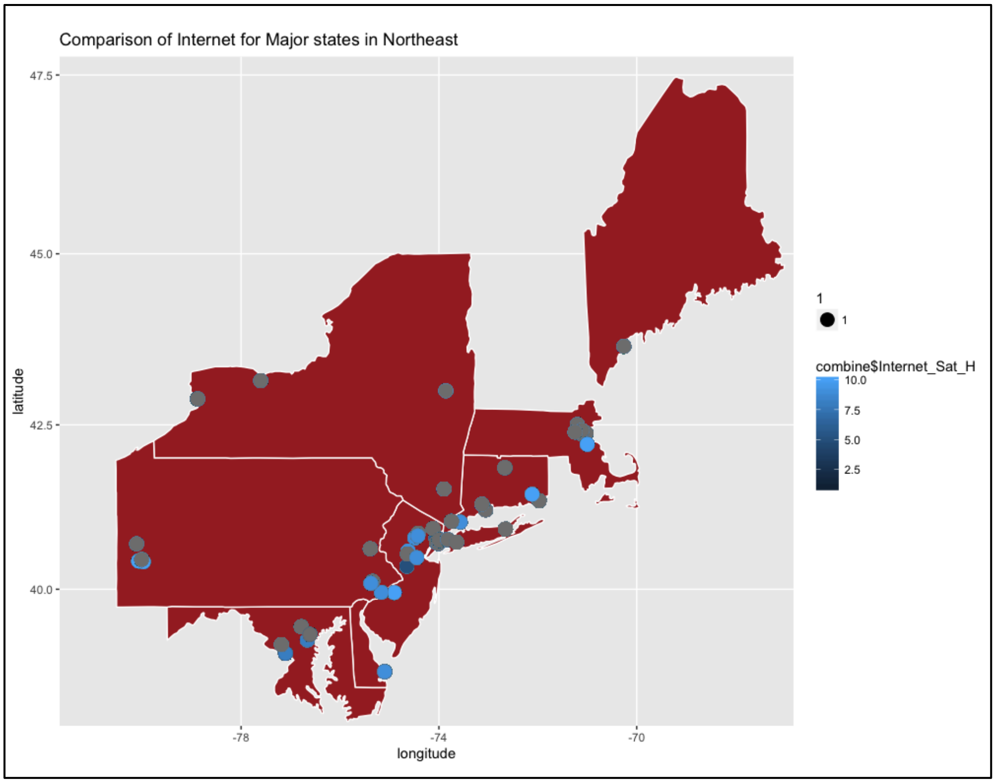

Fig.: Internet Service Ratings for Major States in North-East

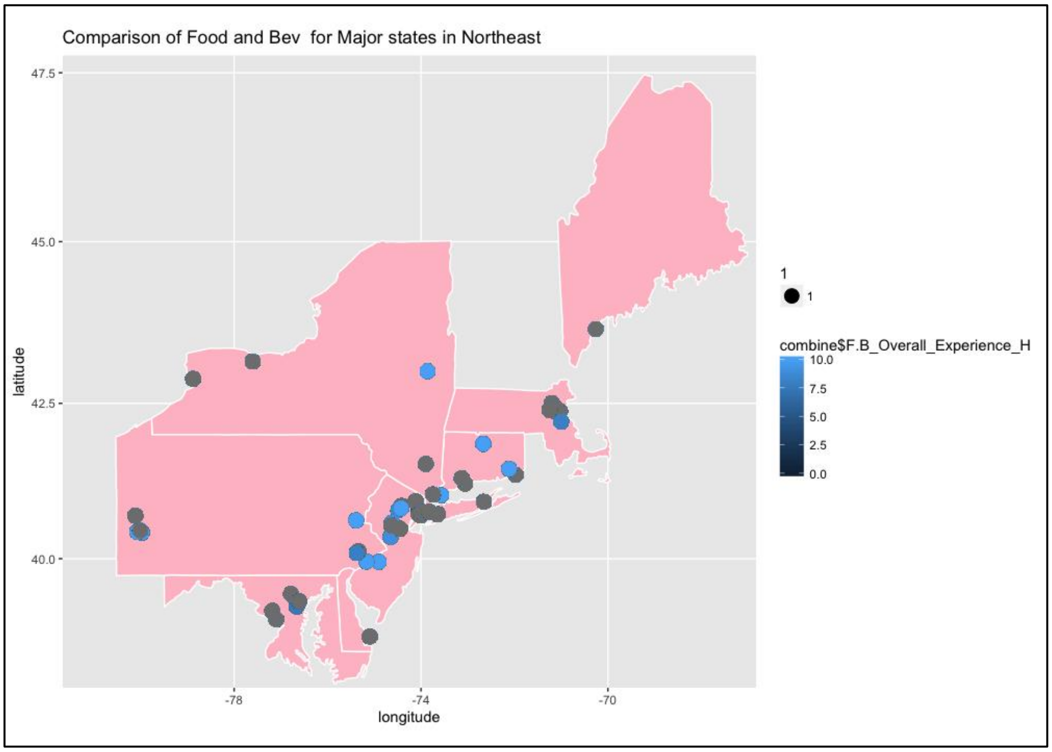

Fig.: Comparison of Food and Beverages for Major States in North-East

11. Summary of Data Visualizations

From the above graphs, we formulated an accordance of attributes that support each other. The following attributes were found to be in accordance with each other:

- Likelihood to Recommend
- NPS_Type
- Condition of Hotel
- Quality of Customer Service
- Comparison of Guest Rooms

This basically means that for attributes such as condition of hotel and quality of service are high for a particular location, the likelihood of the customer to recommend that Hotel is high, consequently increasing the NPS for that hotel.

Similarly, when attributes like quality of food and beverages, internet service ratings, and staff hospitality ratings are low, likelihood of the customer to recommend that hotel becomes low, reducing the NPS for that hotel.

12. Results

It has been observed that the most amount of customers visit Hyatt Hotels for primarily two reasons, which are:

- Leisure
- Business

In order to improve NPS, presence of a fitness center satisfies the needs of customers that visit for leisure, and presence of a business center satisfy the needs of customers that visit for business.

Prioritize on customer service and hospitality ratings of staff members, as customer satisfaction is of utmost importance to whether or not the customers are likely to recommend the hotel.

Our suggestion is two-fold:

- Training staff members for increasing their hospitality towards customers.

- Catering to specific customer behavior and needs, for example, stocking the customers' favorite edibles and products before their arrival.

In order to increase NPS Score in particular areas of the North-East region of the United States, heightened emphasis is required for the following factors:

- Staff Hospitality Ratings
- Quality of Internet Service
- Quality of Food and Beverages provided

This can be done by improving the above attributes by comparing them with successful regions and implementing better strategies.

13. Conclusion

The analysis that has been done through the medium of this project provides actionable insights into various issues with the services provided by hotels of the Hyatt Corporation, especially in the North-East region of the United States. Hotels with the most number of detractors could use this information to improve their services and increase customer satisfaction.