DATA ANALYSIS ON HYATT HOTELS IN UNITED STATES

Identification and Recommendations for Infrastructure Improvement

Group-4:
Charan Kumar Durairaj
Harsh Avlani
Jash Mehta
Priyanka Srivastava
Shrutik Katchhi

Prof. Jeff Saltz
Prof. Erik Anderson
IST 687: Applied Data Science

Project Objective:

The objective of the project is to analyze the data from the surveys filled by the customers of Hyatt Hotels and provide recommendations to any particular hotel or hotels in a particular region, in order to improve their facilities and thereby increase the profit and customer satisfaction.

Scope:

The scope of the project is limited to all the Hyatt hotels located only within United States. The survey data is for the month of January-2016 and analysis had been done based on the processed data (explained later).

Deliverables:

- Identification of hidden patterns over the performance of the hotels within United States
- Visualization of the performance of hotels based on the hotel facilities and services
- Recommendation of facilities and services to be improved in a hotel to improve its performance and increase its revenue

Data Requisition:

Initially, our team requested the customer survey data for one month and 80 attributes (columns). However, after a thorough analysis and understanding of the dataset, we required three more attributes for our analysis. The final raw dataset that we requested had 1.7 million instances and 83 columns.

Data Preprocessing:

The dataset for January, 2016 has 1.7 million instances of customer data and 83 columns mentioning about different attributes of the hotels. However, there were too many instances, where the 'Likelihood to Recommend' column had no value in it. Since it would be a best idea to work on the data, which is real instead of making an assumption and replacing the blank values with mean/median, we ignored all the instances which had blank values for 'Likelihood to Recommend' column. Now, the dataset had only around 38,000 instances and again as per the scope of the project, we filtered the instances for 'United States' and had around 27,000 instances. We are only focusing on the top 5 cities with most number of detractors. We found out that the top 5 cities with most number of detractors are New York, San Francisco, San Antonio, Atlanta and San Diego.

Business question to be answered

Every customer feedback provides insight about the hotels in which they reside.

Analyzing feedback would help Hyatt identify inappropriate practices and stop them before they become major issues.

The dataset consists of attributes like likelihood to recommend, overall satisfaction, tranquility, purpose of visit, postal code etc.

We have applied descriptive statistics and data mining techniques to analyze the dataset.

Following are the Questions which will be addressed:

- Which are the top 5 cities in USA with most number of detractors?
- Which is the best Hyatt hotel (best NPS) in a particular postal code in the top 5 cities?
- In which areas (overall satisfaction, tranquility, Spa etc.) other Hyatt hotels lag with respect to the best Hyatt hotels?
- Which hotel in each city receives the maximum number of detractors?
- What are the common issues associated with each hotel?
- Number of cases of Consumer Dispute
- Which are the most common amenities used by customers?

Initial phase:

Our team decided to go ahead on the project by projecting 'Likelihood to Recommend' column as the main attribute, which can contribute the hotel to improve its business. Therefore, we calculated the Net Promotor Score (NPS) for 'Likelihood to Recommend'.

The entire analysis is done considering the following assumptions:

- Rating of 9 and 10: Promotors
- Rating of 7 and 8: Passive
- Rating of 1-6: Detractors

During the initial phase of the project, our team was working on identifying the possible patterns on the processed dataset through querying in R Studio.

We tried identifying different parameters for the possible reasons of customers being detractor. We divided customers into groups and investigated for basic requirements of these groups, whether it was provided in the hotel or not.

The below code will give a pie chart that will compare count of detractor whose purpose of visit was leisure and hotels were not having a spa in it to the count of detractor whose purpose of visit was leisure and hotels were having spa.

```
spa_n <- length(df$Likelihood_Recommend_H[((df$Likelihood_Recommend_H) < 8 ) &
((df$POV_CODE_C)=="LEISURE") & ((df$Spa_PL)=="N")])
spa_y <-length(df$Likelihood_Recommend_H[((df$Likelihood_Recommend_H) < 8 ) &
((df$POV_CODE_C)=="LEISURE") & ((df$Spa_PL)=="Y")])

pies <- c(spa_n,spa_y)
lbls <- c("N","Y")
par(mar = rep(2, 4))
pie(pies, lbls)
```

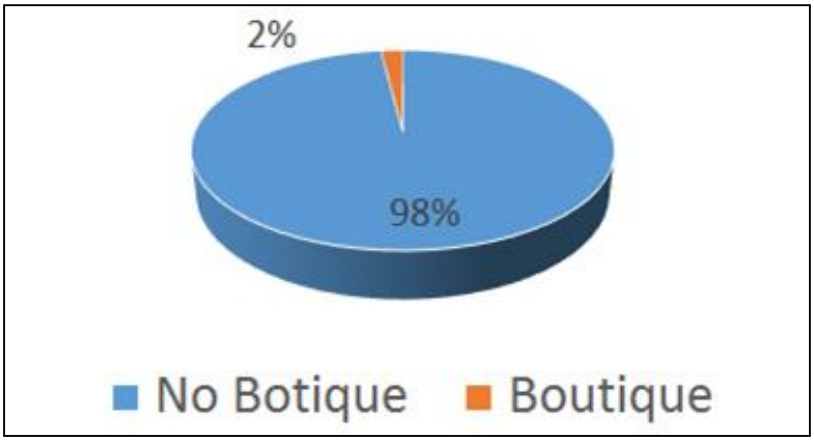
Similarly, we identified some more patterns

- Comparing count of detractor who were females and purpose of visit was leisure and hotels were not having any boutique within the premises to the hotels with boutiques
- Comparing count of detractors in San Diego, whose purpose of visit was Business were having guest room rating as 1-7 to the guest room ratings as 8 and more
- Comparing count of detractors in San Francisco, whose purpose of visit was Business were having condition of hotel rating as 1-7 to the condition of hotel rating as 8 and more

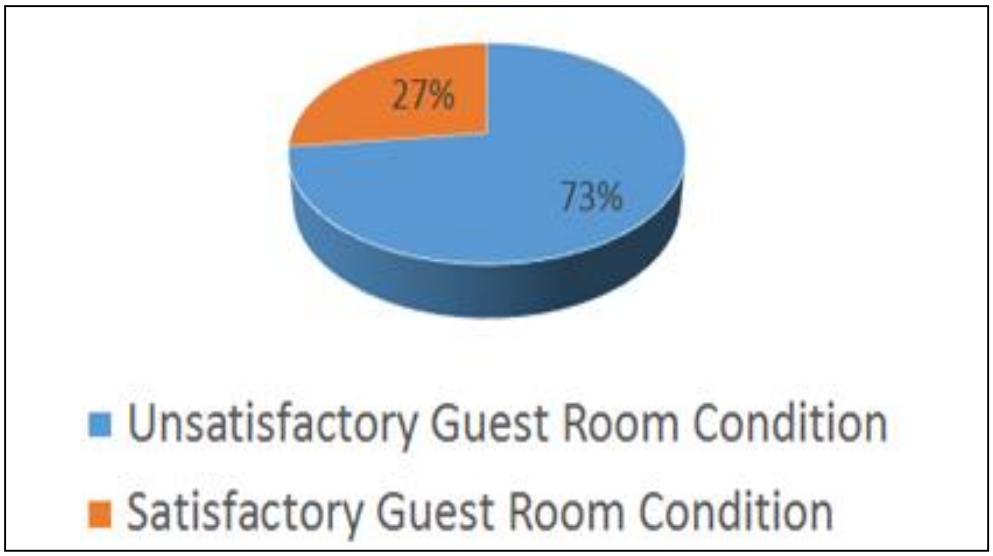
Below are the pie charts generated for identifying possible reasons of customers being detractor.

- 80% of the customers (couples) whose purpose of visit was Leisure could be detractors because the hotel had no Spa Service

- 98% of the Female customers whose purpose of visit was Leisure might be detractors because the hotel had no Boutique



- 73% of the customers in San Diego whose purpose of visit was Business were possibly detractors because of poor guest rooms



- 70% of the customers in San Francisco whose purpose of visit was Business were possibly detractors because of poor condition of hotel

Modelling:

However, the team decided to move on to a 'model based approach' instead of query-based, trial and error approach. The first model we tried was 'linear regression', however with the given dataset which is processed, we didn't get fruitful results based on linear modeling and the R^2 value for the model was below 0.4 for all the models.

The next model, we tried is the J48 classification algorithm. However, since we didn't have an original target variable in the dataset, we didn't continue that approach.

Since, we didn't get effective results through modeling techniques, we decided to develop a function to model the data based on the following rules, in order to identify the parameters that a hotel has to concentrate to improve its business.

The model is based on the following set of rules and conditions:

1) We find out the best hotel in the city by using the table function. This function helps to visualize the number of detractors and promoters. So, the hotel having the best promoter ratio is taken into consideration for the following function. In the below example for Atlanta, we found out that "Hyatt" brand hotel in postal code 30319 is the best hotel. The logic behind taking the best promoter ratio is because that hotel will have the highest NPS. Therefore, we are taking the hotel with best NPS into consideration.

2) Next, we take into consideration all the "Hyatt" hotels in the neighboring postal codes. We input postal code and the data frame to compare with the best hotel.

The below function compares the following attributes of the Best hotel with the attributes of hotel which we input in the function:

- Overall Satisfaction
- Tranquility
- Hotel Condition

- Customer Service
- Staff Cared
- SPA
- Casino
- Boutique
- Business center
- Conference
- Convention
- Dry Cleaning
- Elevators
- Fitness Center
- Fitness Trainer
- Golf
- Indoor Corridors
- Laundry
- Limo Service
- Minibar
- Pool Indoor
- Pool Outdoor
- Regency Grand Club
- Resort
- Restaurant
- Self-Parking
- Shuttle Service
- Ski

Below is the R code that implements the above said function and recommends the hotel about the facilities and services that it has to improve to attract customers and increase the profit.

```
atlantaData = x[(x$City_PL=="Atlanta"),]
fun <- function(df,postalCode)
{
  # Overall Satisfaction:
  cs1 = mean(df$Overall_Sat_H[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")],
na.rm=T)
  print(paste("Desired Overall Satisfaction: ",cs1))

  cs2 = mean(df$Customer_SVC_H[(df$Postal.Code_PL==postalCode) &
(df$Brand_PL=="Hyatt Regency")], na.rm=T)
  print(paste("Current Overall Satisfaction: ",cs2))
  # Tranquility:
```

```
 cs1 = mean(df$Tranquility_H[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")],
na.rm=T)
 print(paste("Desired Tranquility: ",cs1))

 cs2 = mean(df$Tranquility_H[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")], na.rm=T)
 print(paste("Current Tranquility: ",cs2))

 # Hotel Condition:
 cs1 = mean(df$Condition_Hotel_H[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")],
na.rm=T)
 print(paste("Desired Hotel Condition: ",cs1))

 cs2 = mean(df$Condition_Hotel_H[(df$Postal.Code_PL==postalCode) &
(df$Brand_PL=="Hyatt Regency")], na.rm=T)
 print(paste("Current Hotel Condition: ",cs2))


 # Customer Service:
 cs1 = mean(df$Customer_SVC_H[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")],
na.rm=T)
 print(paste("Desired Customer Service: ",cs1))

 cs2 = mean(df$Customer_SVC_H[(df$Postal.Code_PL==postalCode) &
(df$Brand_PL=="Hyatt Regency")], na.rm=T)
 print(paste("Current Customer Service: ",cs2))

 # Staff Cared:
 cs1 = mean(df$Staff_Cared_H[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")],
na.rm=T)
 print(paste("Desired Staff Cared: ",cs1))

 cs2 = mean(df$Staff_Cared_H[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")], na.rm=T)
 print(paste("Current Staff Cared: ",cs2))

 # Spa Service:

 a1=  (df$Spa_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(a1[a1=="Y"])>length(a1[a1=="N"]) )
 {a1<-c("Y")}
 if(length(a1[a1=="Y"])<length(a1[a1=="N"]))
 {a1<-c("N")}
 if(length(a1[a1=="Y"])==length(a1[a1=="N"]))
```

```
{a1<-c("Data Not Available")}

a2=  (df$Spa_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt Regency")])
if( length(a2[a2=="Y"])>length(a2[a2=="N"]) )
{a2<-c("Y")}
if(length(a2[a2=="Y"])<length(a2[a2=="N"]))
{a2<-c("N")}
if(length(a2[a2=="Y"])==length(a2[a2=="N"]))
{a2<-c("Data Not Available")}

if(a1!=a2)
{
  print(paste("Desired Spa Service: ",a1))
  print(paste("Current Spa Service: ",a2))
}
#Casino
b1=  (df$Casino_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
{b1<-c("Y")}
if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
{b1<-c("N")}
if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
{b1<-c("Data Not Available")}


b2=  (df$Casino_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt Regency")])
if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
{b2<-c("Y")}
if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
{b2<-c("N")}
if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
{b2<-c("Data Not Available")}


if(b1!=b2)
{
  print(paste("Desired Casino: ",b1))
  print(paste("Current Casino: ",b2))
}

# Boutique_PL

b1=  (df$Boutique_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
```

```
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Boutique_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Boutique: ",b1))
  print(paste("Current Boutique: ",b2))
 }







 #,df$Business.Center_PL


 b1=  (df$Business.Center_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}
 b2=  (df$Business.Center_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
```

```
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Business.Center_PL: ",b1))
  print(paste("Current Business.Center_PL: ",b2))
 }




 #,df$Conference_PL
 #
 b1=  (df$Conference_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}

 b2=  (df$Conference_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Conference_PL: ",b1))
  print(paste("Current Conference_PL: ",b2))
 }

 #,df$Convention_PL
 #
 b1=  (df$Convention_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
```

```
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Convention_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Convention_PL: ",b1))
  print(paste("Current Convention_PL: ",b2))
 }

 #,df$Dry.Cleaning_PL
 #
 #
 b1=  (df$Dry.Cleaning_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Dry.Cleaning_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
```

```
 {b2<-c("Data Not Available")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Dry.Cleaning_PL: ",b1))
  print(paste("Current Dry.Cleaning_PL: ",b2))
 }


 #,df$Elevators_PL

 b1=  (df$Elevators_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Elevators_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Elevators_PL: ",b1))
  print(paste("Current Elevators_PL: ",b2))
 }


 #,df$Fitness.Center_PL

 b1=  (df$Fitness.Center_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
```

```
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Fitness.Center_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}

 if(b1!=b2)
 {
  print(paste("Desired Fitness.Center_PL: ",b1))
  print(paste("Current Fitness.Center_PL: ",b2))
 }




 #,df$Fitness.Trainer_PL


 b1=  (df$Fitness.Trainer_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Fitness.Trainer_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
```

```
{b2<-c("Data Not Available")}


if(b1!=b2)
{
  print(paste("Desired Fitness.Trainer_PL: ",b1))
  print(paste("Current Fitness.Trainer_PL: ",b2))
}



#,df$Golf_PL

b1=  (df$Golf_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
{b1<-c("Y")}
if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
{b1<-c("N")}
if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
{b1<-c("Data Not Available")}



b2=  (df$Golf_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt Regency")])
if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
{b2<-c("Y")}
if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
{b2<-c("N")}
if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
{b2<-c("Data Not Available")}


if(b1!=b2)
{
  print(paste("Desired Golf_PL: ",b1))
  print(paste("Current Golf_PL: ",b2))
}



#,df$Indoor.Corridors_PL

b1=  (df$Indoor.Corridors_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
```

```
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Indoor.Corridors_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Indoor.Corridors_PL: ",b1))
  print(paste("Current Indoor.Corridors_PL: ",b2))
 }



 #,df$Laundry_PL

 b1=  (df$Laundry_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Laundry_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}
```

```
 if(b1!=b2)
 {
  print(paste("Desired Laundry_PL: ",b1))
  print(paste("Current Laundry_PL: ",b2))
 }


 #,df$Limo.Service_PL

 b1=  (df$Limo.Service_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Limo.Service_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}

 if(b1!=b2)
 {
  print(paste("Desired Limo.Service_PL: ",b1))
  print(paste("Current Limo.Service_PL: ",b2))
 }




 #,df$Mini.Bar_PL

 b1=  (df$Mini.Bar_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
```

```
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}



 b2=  (df$Mini.Bar_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Mini.Bar_PL: ",b1))
  print(paste("Current Mini.Bar_PL: ",b2))
 }



 #,df$Pool.Indoor_PL

 b1=  (df$Pool.Indoor_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}


 b2=  (df$Pool.Indoor_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}
```

```
 if(b1!=b2)
 {
  print(paste("Desired Pool.Indoor_PL: ",b1))
  print(paste("Current Pool.Indoor_PL: ",b2))
 }



 #,df$Pool.Outdoor_PL

 b1=  (df$Pool.Outdoor_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}



 b2=  (df$Pool.Outdoor_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Pool.Outdoor_PL: ",b1))
  print(paste("Current Pool.Outdoor_PL: ",b2))
 }



 #,df$Regency.Grand.Club_PL

 b1=  (df$Regency.Grand.Club_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
```

```
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}



 b2=  (df$Regency.Grand.Club_PL[(df$Postal.Code_PL==postalCode) &
(df$Brand_PL=="Hyatt Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Regency.Grand.Club_PL: ",b1))
  print(paste("Current Regency.Grand.Club_PL: ",b2))
 }



 #,df$Resort_PL


 b1=  (df$Resort_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}



 b2=  (df$Resort_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}
```

```
 if(b1!=b2)
 {
  print(paste("Desired Resort_PL: ",b1))
  print(paste("Current Resort_PL: ",b2))
 }


 #,df$Restaurant_PL

 b1=  (df$Restaurant_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}



 b2=  (df$Restaurant_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Restaurant_PL: ",b1))
  print(paste("Current Restaurant_PL: ",b2))
 }


 #,df$Self.Parking_PL


 b1=  (df$Self.Parking_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
```

```
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}
 b2=  (df$Self.Parking_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")
 if(b1!=b2)
 {
  print(paste("Desired Self.Parking_PL: ",b1))
  print(paste("Current Self.Parking_PL: ",b2))
 }
 #,df$Shuttle.Service_PL
 b1=  (df$Shuttle.Service_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
 if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
 {b1<-c("N")}
 if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
 {b1<-c("Data Not Available")}
 b2=  (df$Shuttle.Service_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
 if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
 {b2<-c("Y")}
 if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
 {b2<-c("N")}
 if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
 {b2<-c("Data Not Available")}


 if(b1!=b2)
 {
  print(paste("Desired Shuttle.Service_PL: ",b1))
  print(paste("Current Shuttle.Service_PL: ",b2))
 }
 #,df$Ski_PL
 b1=  (df$Ski_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
 if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
 {b1<-c("Y")}
```

```
  if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
  {b1<-c("N")}
  if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
  {b1<-c("Data Not Available")}
b2=  (df$Ski_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt Regency")])
  if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
  {b2<-c("Y")}
  if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
  {b2<-c("N")}
  if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
  {b2<-c("Data Not Available")}
if(b1!=b2)
  {
    print(paste("Desired Ski_PL: ",b1))
    print(paste("Current Ski_PL: ",b2))
  }
  b1=  (df$Valet.Parking_PL[(df$Postal.Code_PL==30319) & (df$Brand_PL=="Hyatt")])
  if( length(b1[b1=="Y"])>length(b1[b1=="N"]) )
  {b1<-c("Y")}
  if(length(b1[b1=="Y"])<length(b1[b1=="N"]))
  {b1<-c("N")}
  if(length(b1[b1=="Y"])==length(b1[b1=="N"]))
  {b1<-c("Data Not Available")
  b2=  (df$Valet.Parking_PL[(df$Postal.Code_PL==postalCode) & (df$Brand_PL=="Hyatt
Regency")])
  if( length(b2[b2=="Y"])>length(b2[b2=="N"]) )
  {b2<-c("Y")}
  if(length(b2[b2=="Y"])<length(b2[b2=="N"]))
  {b2<-c("N")}
  if(length(b2[b2=="Y"])==length(b2[b2=="N"]))
  {b2<-c("Data Not Available")}
if(b1!=b2)
  {
    print(paste("Desired Valet.Parking_PL: ",b1))
    print(paste("Current Valet.Parking_PL: ",b2))
  }
}
table(atlantaData$Postal.Code_PL,atlantaData$NPS)
# Best: 30319
# worst: 30303
fun(atlantaData,30303)
colnames(atlantadf) <- c("Rating","Desired","Current")
melteddf = melt(atlantadf, id = "Rating")
ggplot(melteddf, aes(Rating, value)) +
```

```
geom_bar(aes(fill = variable), position = "dodge", stat="identity") +
labs(title = "Atlanta")
```

Demo of the model through visualization:

The function requires two arguments in order to calculate the effectiveness to attract customers and identify the areas of concern for a hotel in a particular zip code. They are the following:

- Dataframe
- Zip code of the hotel

Sample Output:

fun(sfData,94133)

[1] "Desired Overall Satisfaction: 8.59375"

[1] "Current Overall Satisfaction: 7.80392156862745"

[1] "Desired Tranquility: 8.45454545454546"

[1] "Current Tranquility: 6.8"

[1] "Desired Hotel Condition: 8.904"

[1] "Current Hotel Condition: 6.52941176470588"

[1] "Desired Customer Service: 8.992"

[1] "Current Customer Service: 7.80392156862745"

[1] "Desired Staff Cared: 8.72058823529412"

[1] "Current Staff Cared: 7.80952380952381“

[1] "Desired Convention_PL: Y"

[1] "Current Convention_PL: N"

[1] "Desired Pool.Outdoor_PL: N"

[1] "Current Pool.Outdoor_PL: Y"

[1] "Desired Regency.Grand.Club_PL: Y"

[1] "Current Regency.Grand.Club_PL: N"

[1] "Desired Self.Parking_PL: Y"

[1] "Current Self.Parking_PL: N"

This is the tool(function) which would be our deliverable to the client.

If the client enters the data frame (customer survey data) along with the desired postal code, then the recommendations like above should be the final output. This way the client can determine which areas of the hotels need to be fixed in order to increase the NPS to reach to the desired level.

For example, in the above output:

Desired Overall Satisfaction: 8.59375 [Result from the best hotel]

Current Overall Satisfaction: 7.80392156862745 [Hotel in the 30303]

Desired Tranquility: 8.45454545454546 [Result from the best hotel]

Current Tranquility: 6.8 [Hotel in the 30303]

Desired Hotel Condition: 8.904 [Result from the best hotel]

Current Hotel Condition: 6.52941176470588 [Hotel in the 30303]

Desired Customer Service: 8.992 [Result from the best hotel]

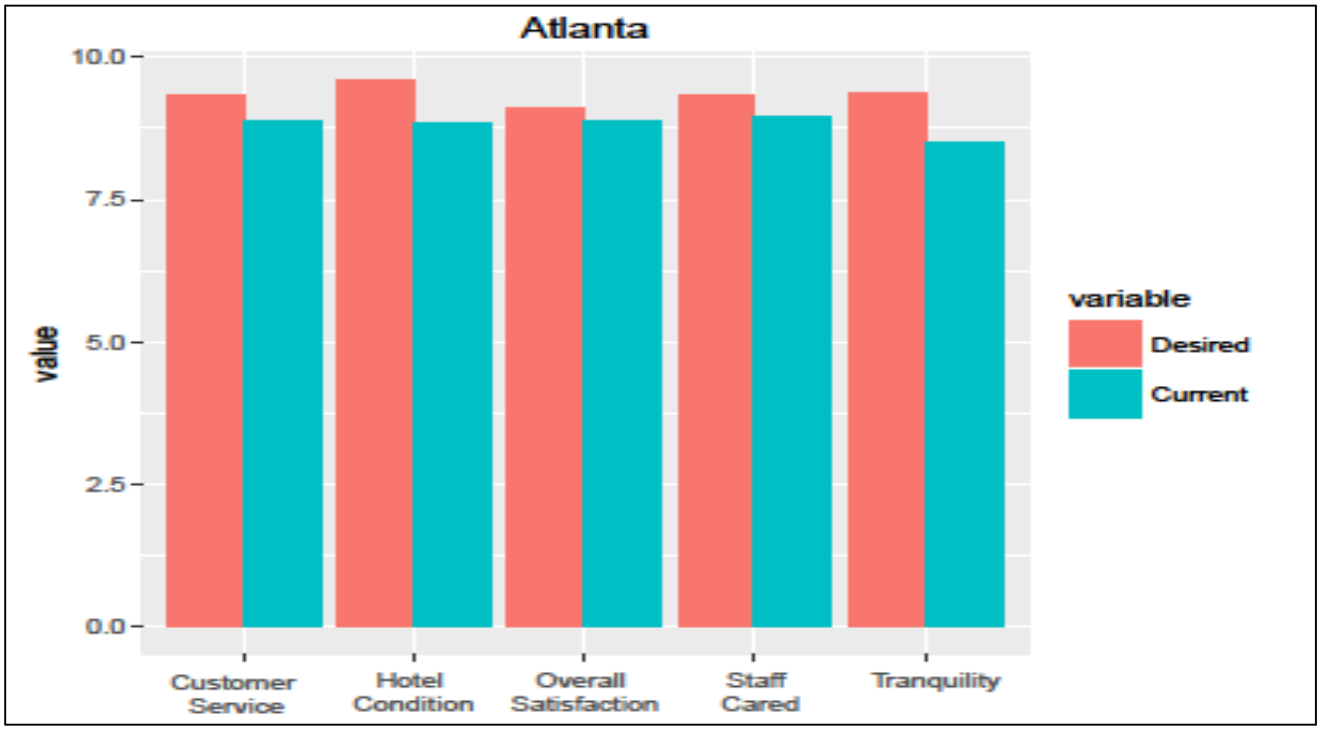
Current Customer Service: 7.80392156862745 [Hotel in the 30303]

Desired Staff Cared: 8.72058823529412 [Result from the best hotel]

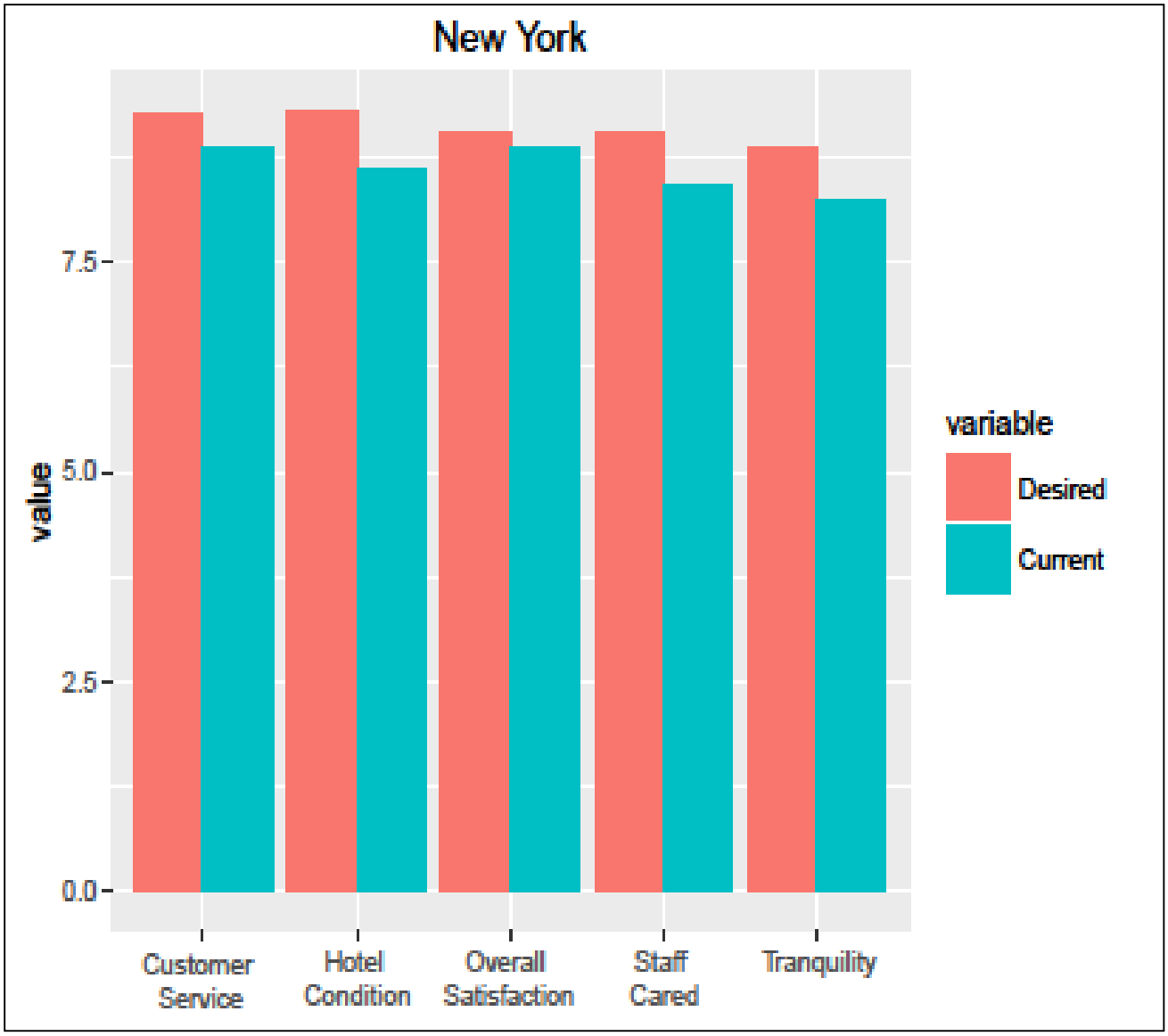
Current Staff Cared: 7.80952380952381[Hotel in the 30303]

So, clearly the hotel in 30303 needs to look into the best hotels and implement in their hotel too so that they can reach the desired results. The hotel in 30303 postal code must also include facilities such as outdoor pool, self-parking, convention, regency grand club.

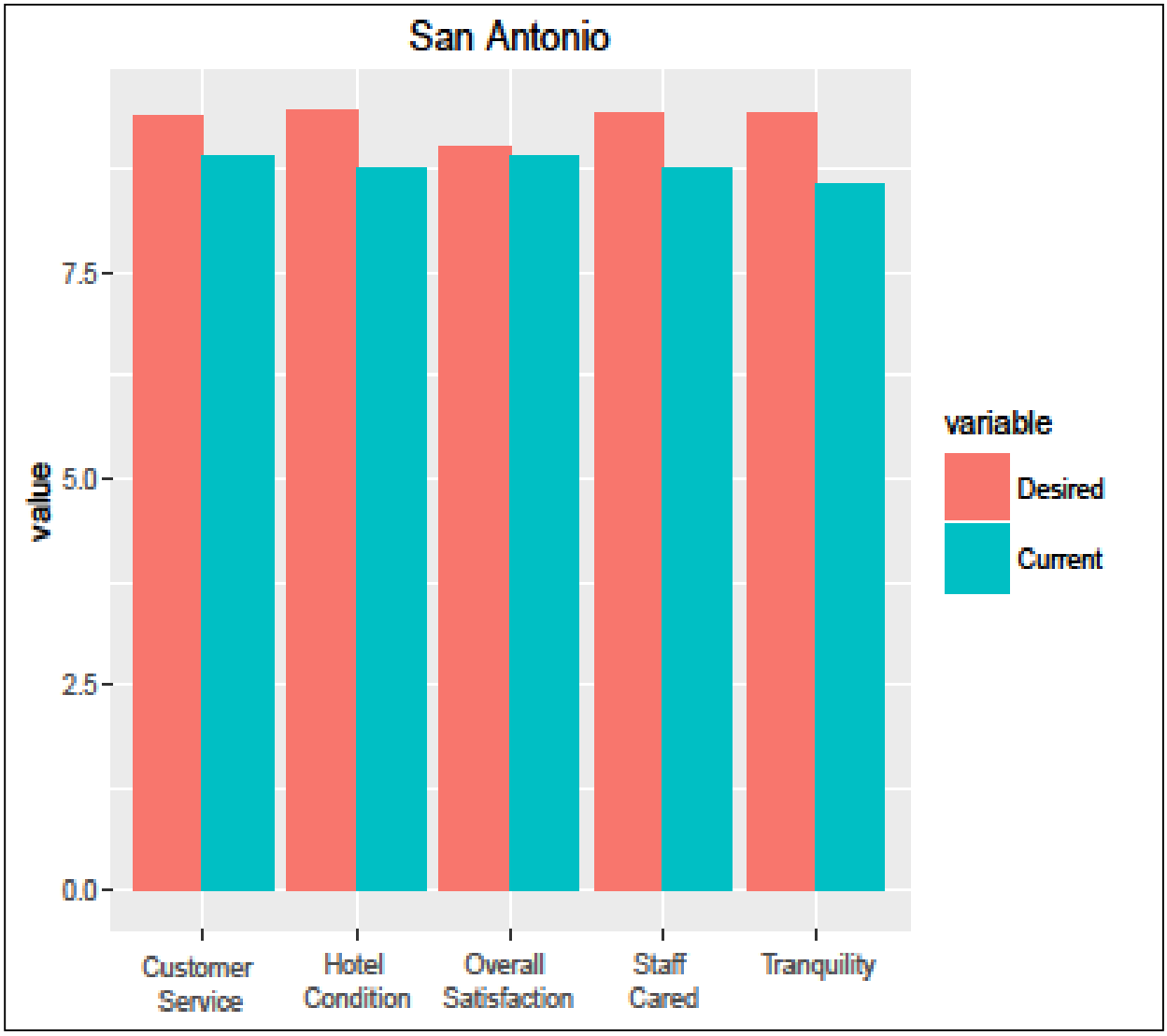
Below is the bar chart represents the hotels in the zip code: 30303 at Atlanta. The red bar represents the service of the best hotel in the region and the blue bar represents the service of the hotels in the zip code: 30303. The hotels in the area of 30303 has to fill the gap between red and blue bars, in order to be more profitable.



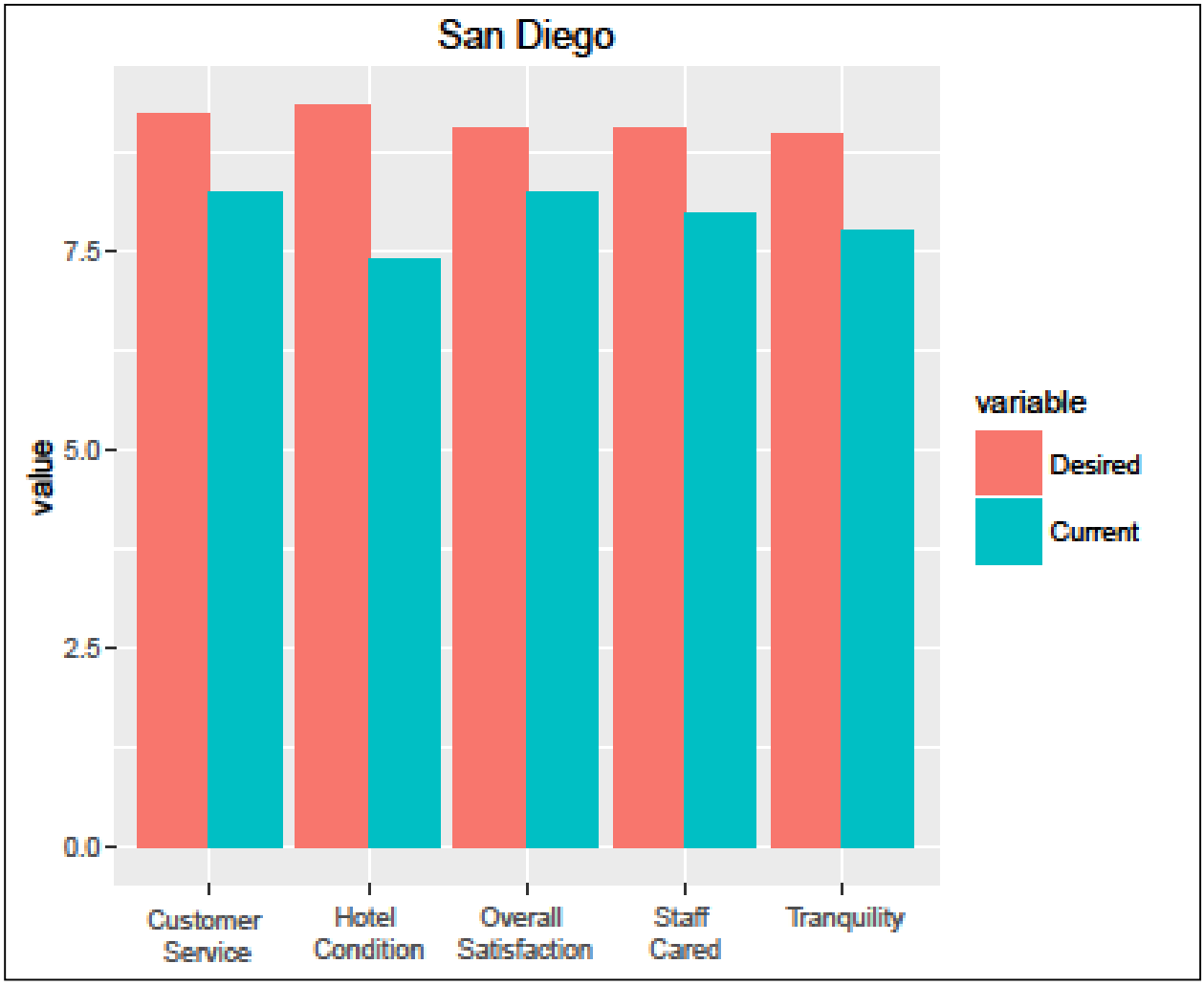
Below is the bar chart represents the Hyatt hotels in the zip code: 10017 at New York. The red bar represents the service of the best hotel in the region and the blue bar represents the service of the hotels in the zip code: 10017. The hotels in the area of 10017 has to fill the gap between red and blue bars, in order to be more profitable.


New York
value
7.5
5.0
2.5
0.0
Customer Service
Hotel Condition
Overall Satisfaction
Staff Cared
Tranquility
variable
Desired
Current

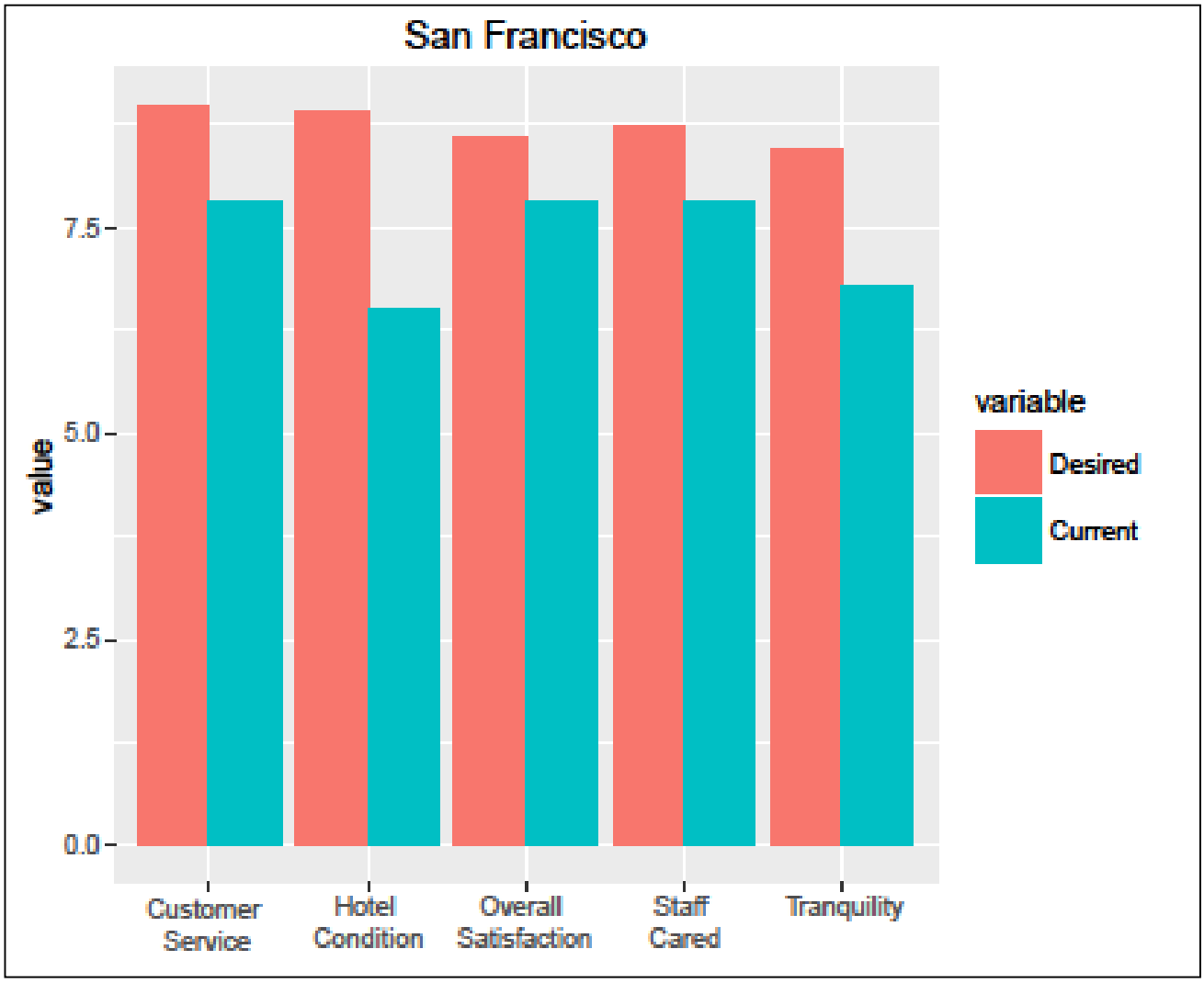
Below is the bar chart represents the hotels in the zip code: 78205 at San Antonio. The red bar represents the service of the best hotel in the region and the blue bar represents the service of the hotels in the zip code: 78205. The hotels in the area of 78205 has to fill the gap between red and blue bars, in order to be more profitable.



Below is the bar chart represents the hotels in the zip code: 92109 at San Diego. The red bar represents the service of the best hotel in the region and the blue bar represents the service of the hotels in the zip code: 92109. The hotels in the area of 92109 has to fill the gap between red and blue bars, in order to be more profitable.



Below is the bar chart represents the hotels in the zip code: 94133 at San Francisco. The red bar represents the service of the best hotel in the region and the blue bar represents the service of the hotels in the zip code: 94133. The hotels in the area of 94133 has to fill the gap between red and blue bars, in order to be more profitable



Conclusion

The analysis that we performed on the data from the surveys filled by the customers helped the Hotels improve facilities and increase their NPS. Hotels with most number of detractors could use this data and hence improve the customer ratings.