

Hotel Data Analysis

Prof Jeff Saltz

Prof Erik Anderson

Jeni Adenshia

Cordell Hudson

Omkar Mutreja

Romil Shroff

IST 687 | December 6, 2017

Table of Contents

[Introduction 2](#_Toc500423205)

[Industry Analysis 2](#_Toc500423206)

[Overview of Hyatt 3](#_Toc500423207)

[Business Questions 4](#_Toc500423208)

[Data Identification 4](#_Toc500423209)

[Data Preprocessing 4](#_Toc500423210)

[Descriptive Statistics 12](#_Toc500423211)

[Recommendation Analysis 15](#_Toc500423212)

[Interactive Data Visualization 25](#_Toc500423213)

[Conclusion: 27](#_Toc500423214)

# Introduction

Project Objective:

The objective of the data analysis project is to analyze data collected through the customer feedback of the Hyatt hotels throughout the world. The focus is to identify patterns and provide substantial recommendations to any hotel or hotels to improve the facilities and amenities offered. This will in turn help increase the hotels profit and customer satisfaction.

Scope:

The scope of the project is limited to a few hotels in the United States. The survey data for the months of November 2014, December 2014 and January 2015 have been analyzed, pre-processed and analyzed again.

Deliverables:

* Identification of patterns based on the performance of the hotels in US
* Visualization of different hotels based on performance and feedback received on basis of the services offered
* Recommendation of the services offered to improve the business performance and revenue

# Industry Analysis

Hotel industry:

In the past decade, the industry has thrived, as both leisure and business travel have ascended alongside GDP. The hospitality industry's backbone is comprised of customer service, a concept shared by all segments of the industry. This business may focus on one or all facets of hospitality. How accomplished you and your staff are at serving others will determine your business' level of success. However, though costs and challenges will increase, owning or managing several facets of hospitality can provide you with many more opportunities to generate success. The U.S. travel and tourism industry generated over $1.5 trillion in economic output in 2016. In 2016 travelers spent more than $293 billion on traveler accommodations. This sector supports more than 1.4 million U.S. jobs.

The Hotel Industry Customer Experience Benchmarks have been identified which are as follows:

* Ease of reservations
* Check-In process
* Room cleanliness and comfort
* Food services
* In-Room entertainment (TV, movies)
* Amenities
* In-Room internet service
* Staff courtesy and helpfulness
* Loyalty programs
* Call center
* Website

NPS trends:

Net Promoter Score, or NPS, measures customer experience and predicts business growth. This proven metric transformed the business world and now provides the core measurement for customer experience management programs the world round.

NPS is calculated using the answer to a key question, using a 0-10 scale: How likely is it that you would recommend [brand] to a friend or colleague? Subtracting the percentage of Detractors (score 0-6) from the percentage of Promoters (score 9-10) yields the Net Promoter Score, which can range from a low of -100 (if every customer is a Detractor) to a high of 100 (if every customer is a Promoter). Promoters (score 9-10) are loyal enthusiasts who will keep buying and refer others, fueling growth. Passives (score 7-8) are satisfied but unenthusiastic customers who are vulnerable to competitive offerings. Detractors (score 0-6) are unhappy customers who can damage your brand and impede growth through negative word-of-mouth.

# Overview of Hyatt

Hyatt Hotels Corporation is an American multinational owner, operator, and franchisor of hotels, resorts, and vacation properties. Hyatt was founded by Jay Pritzker in 1957. The Hyatt Corporation came into being upon purchase of the Hyatt House, at Los Angeles International Airport, on September 27, 1957. On December 31, 2004, all the hospitality assets owned by Pritzker family business interests were consolidated under a single entity now Hyatt Hotels Corporation. In 1968, Hyatt International was formed and subsequently became a separate public company. 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.

# Business Questions

The focus is to improve the services provided by the Hyatt hotels to facilitate the customer recommendations and reviews. Customer recommendations and reviews are important factors for any hotel industry. We have performed descriptive analysis and data mining to analyze the survey data set to answer the following questions:

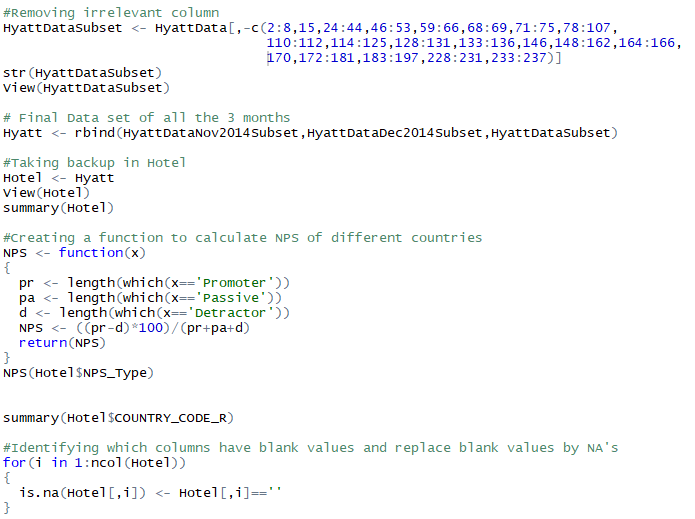
* How should HYATT hotels reduce the friction in the customer experience?
* Which country has the highest detractor ratio?
* Which state within that country is showcasing poor customer feedback?
* Which category of hotels is leading the poor customer feedback trend within the specific state?
* What facilities and services have a positive and negative impact on the NPS count of the hotel?
* Which services should be given priority for improvement?
* Are there any trends to categorize hotels based on customer feedback?
* What services and facilities are driving transformation of customer experience from detractors, passive to promotors?

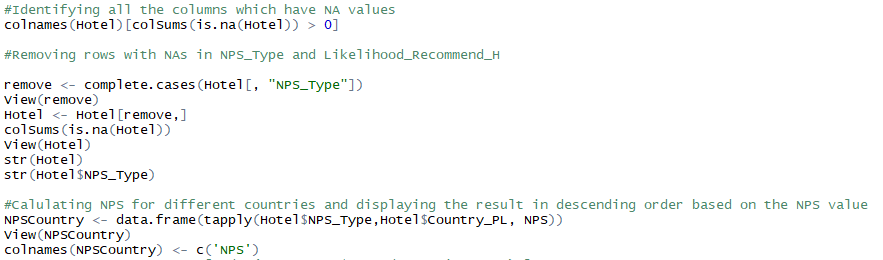
# Data Identification

The survey data of November 2014, December 2014 and January 2015 have **1151316** customer data with around 78 relevant attributes of the hotel.

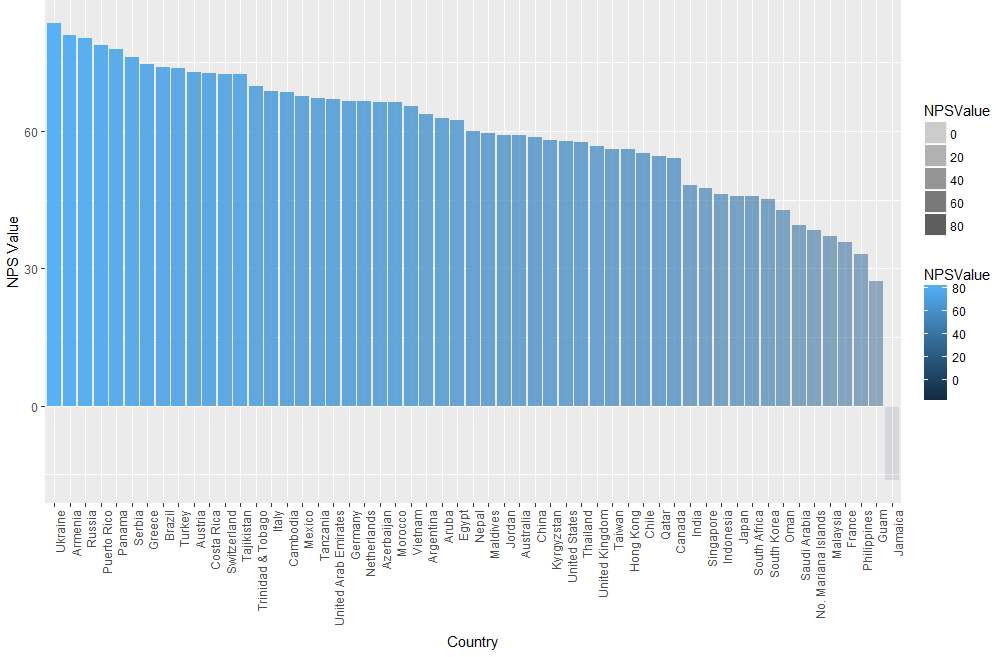
# Data Preprocessing

This survey data has null values which will possibly create a hindrance to obtain accurate trends. Omitting the NA values is not advisable as we may lose out on important trends. Initially, all the null values are replaced with blank values and then replaced with the string ‘NA’. But the best idea is to remove the NA values ‘Likelihood to Recommend’ and ‘NPS\_Type’ as they are the key factors for this recommendation analysis. By doing this we drill down to **185091** customer data instances.

Here is a code snippet for removing the redundant, irrelevant and empty values:  


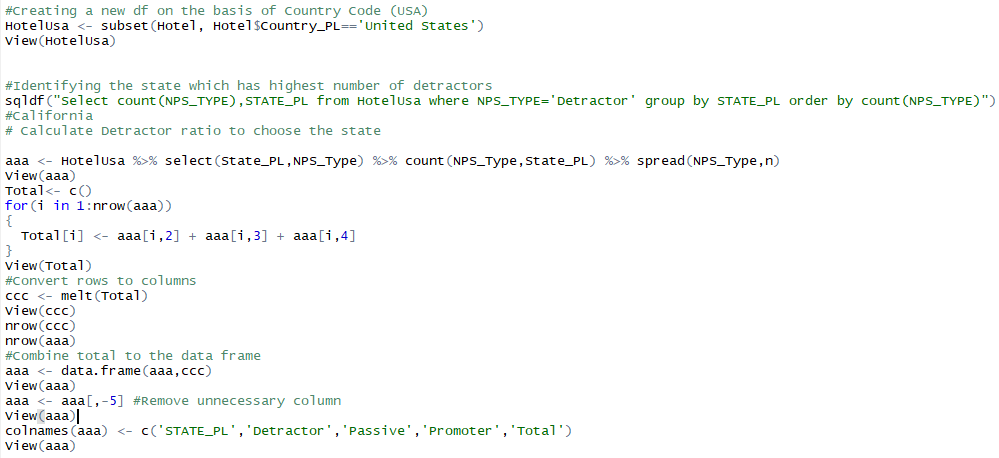


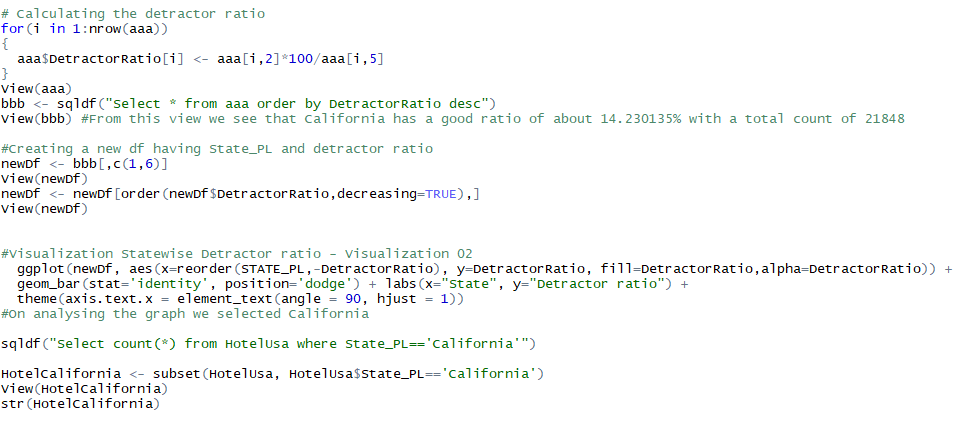
Now that the data has been cleaned, the data trends for different countries was analyzed according to the NPS value.

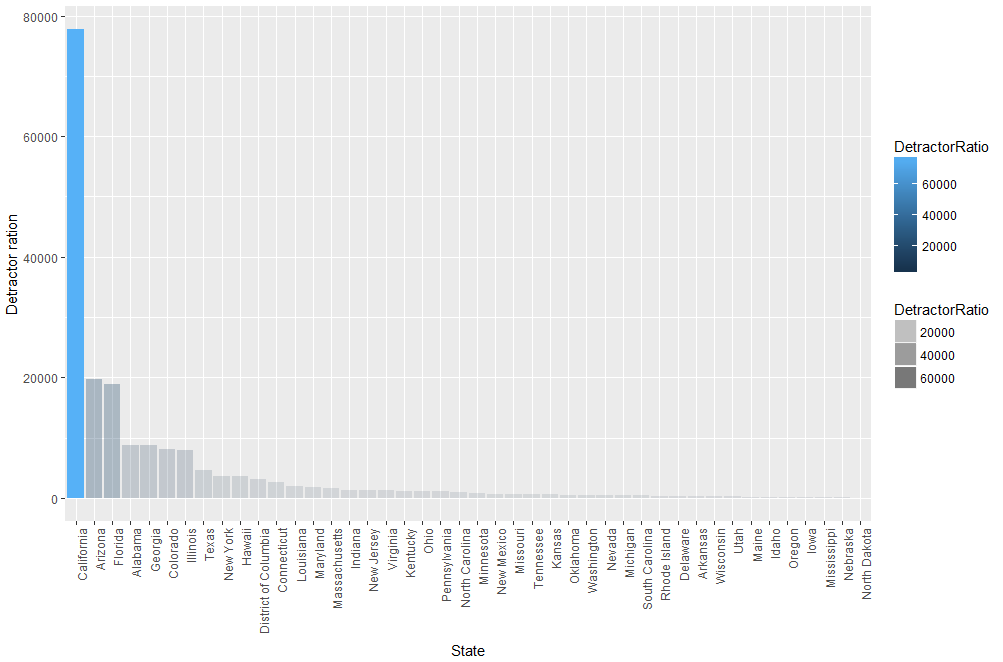


Most of the top countries did not have sufficient number of instances despite having a bad NPS value. But the United States has a total count of 143363 instances with a NPS value of 57.83. The focus of this analysis to identify why the specific Hyatt Hotels detractors instead of promotors.

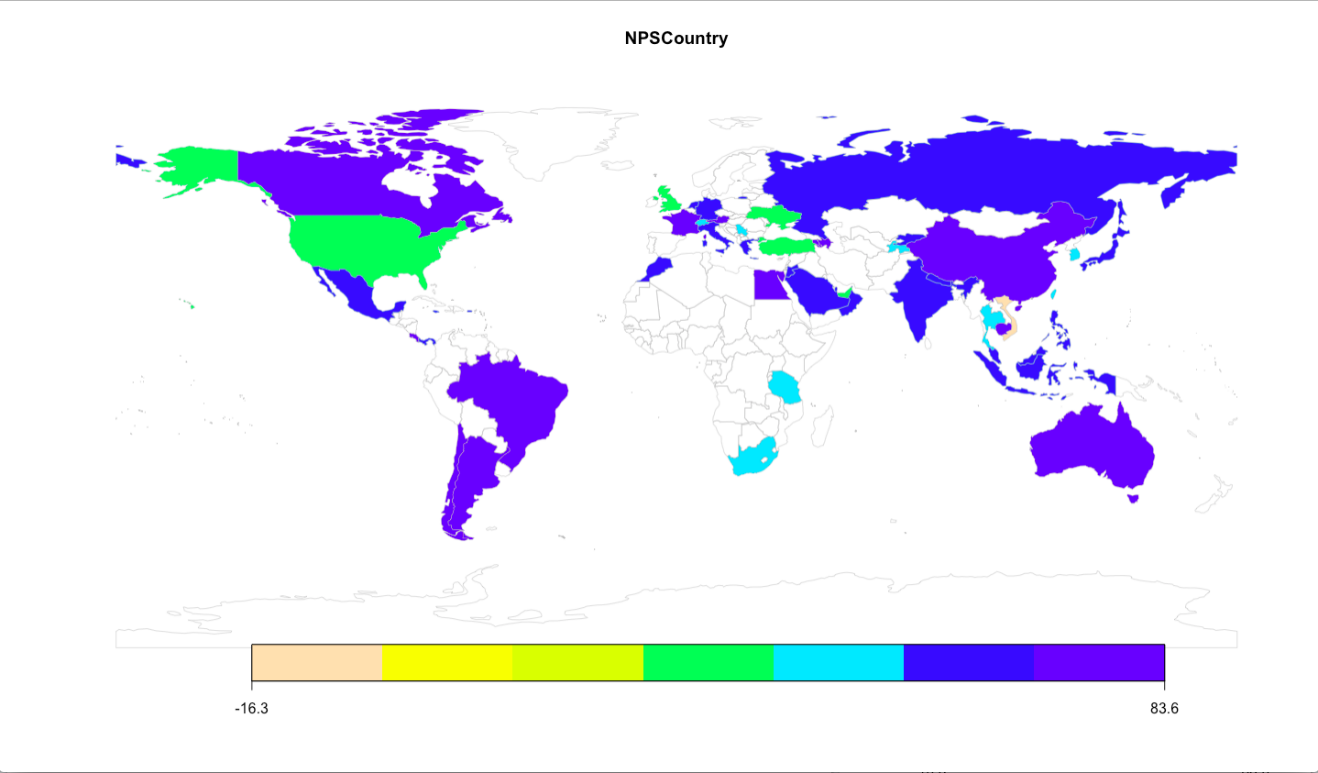
Next, the trends for recommendation was analyzed for all the states in the US.



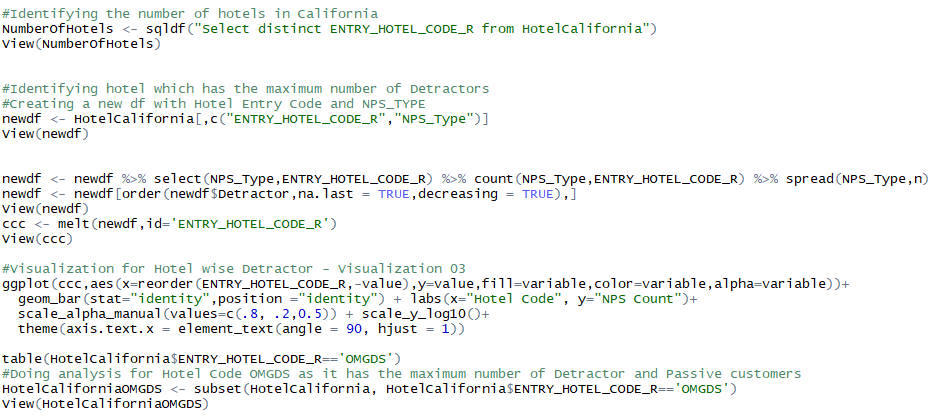


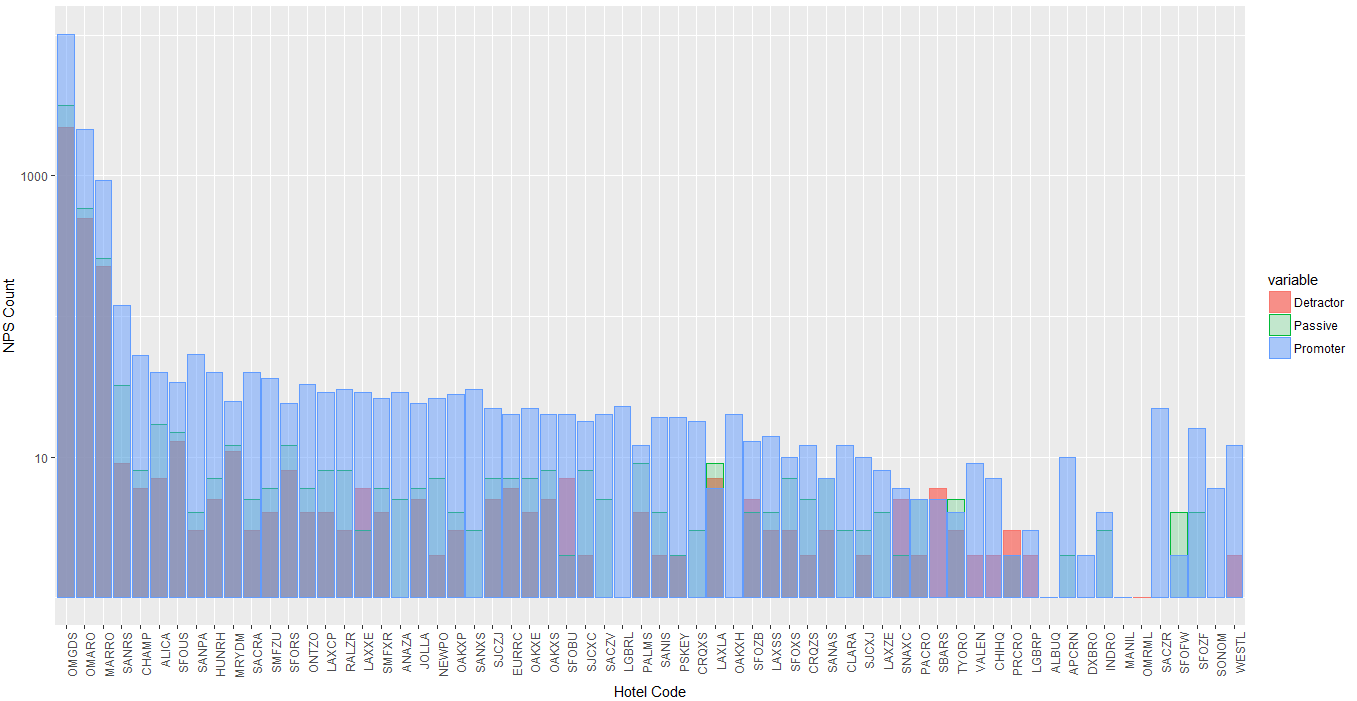


The top 5 states according to the detractor ratio are California, Arizona, Florida, Alabama and Georgia. We have selected California which has the most number of detractors with about 21848 instances.



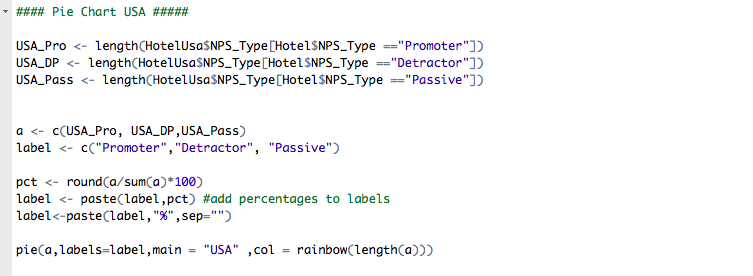
As the focus is to improve the services provided by the hotels to increase the promotor value, the hotel with highest detractor ratio has been selected to perform further analysis.

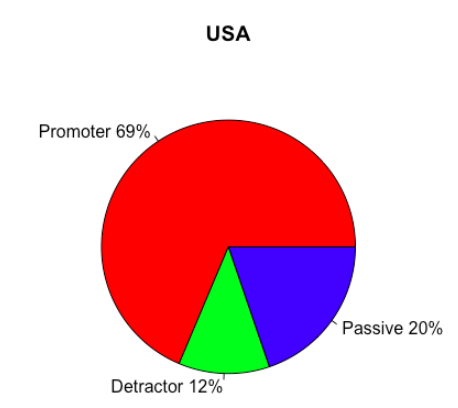


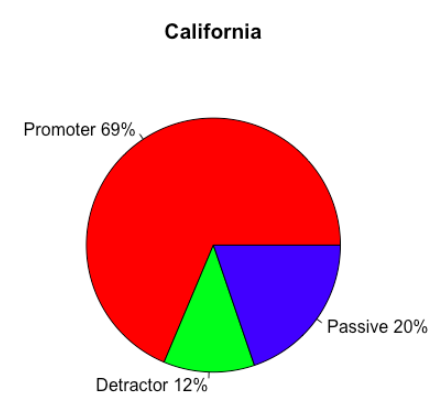


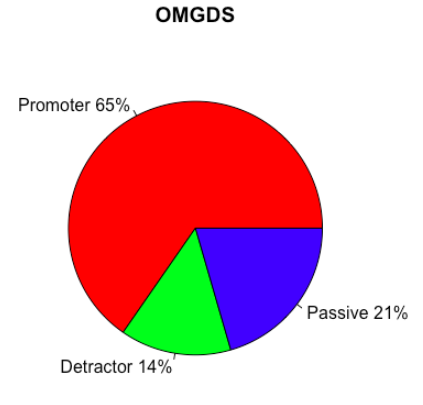
Hotel with the code OMGDS which is in El Segundo city of California has been further analyzed which has 15482 instances. Now we dive into deep analysis for providing appropriate recommendations backed up with statistical proof.

To summarize why we narrowed down to the respective geographic location – the US, California and specific hotel in California and visually we have created a chart which shows the NPS (Promoter, Detractor and Passive) ratios for the same.









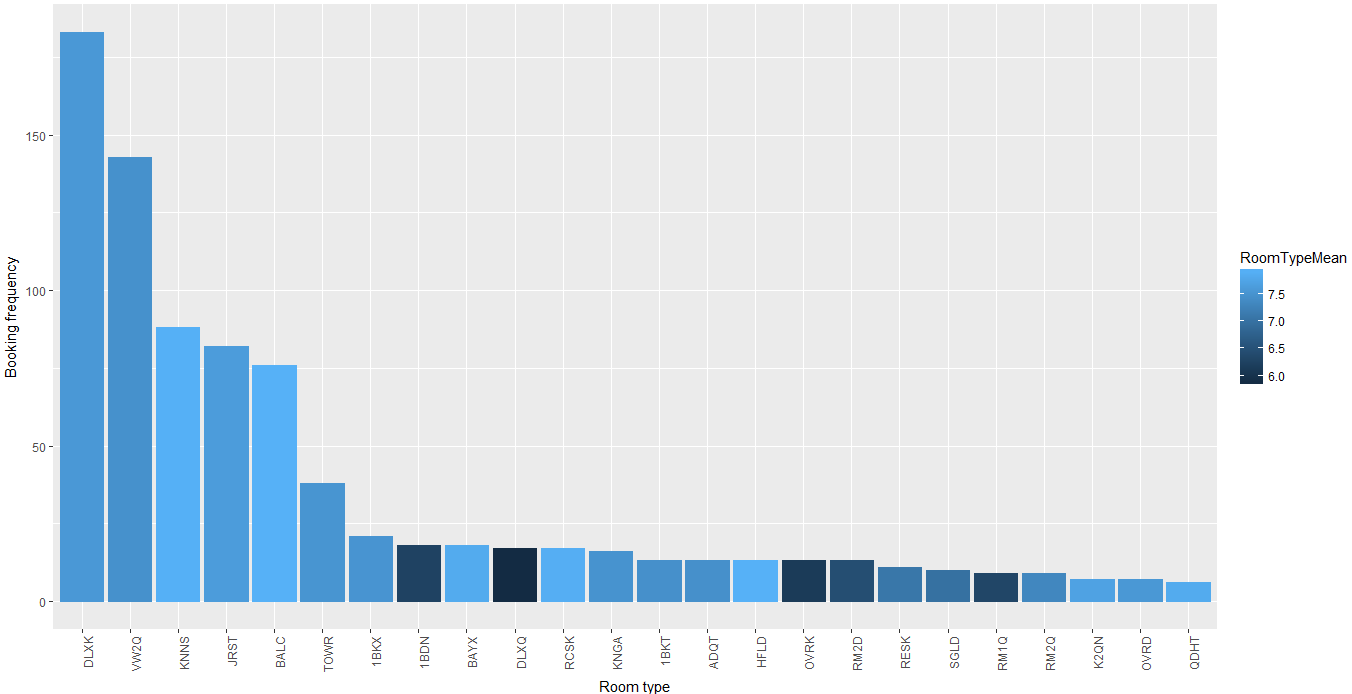
# Descriptive Statistics

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

The descriptive statistics analysis includes the following:

Room Type Analysis:

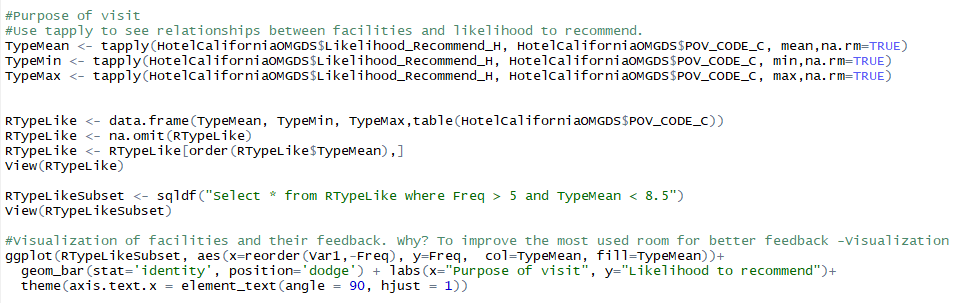
The feedback received for room types which are frequently booked has be analyzed as shown below:

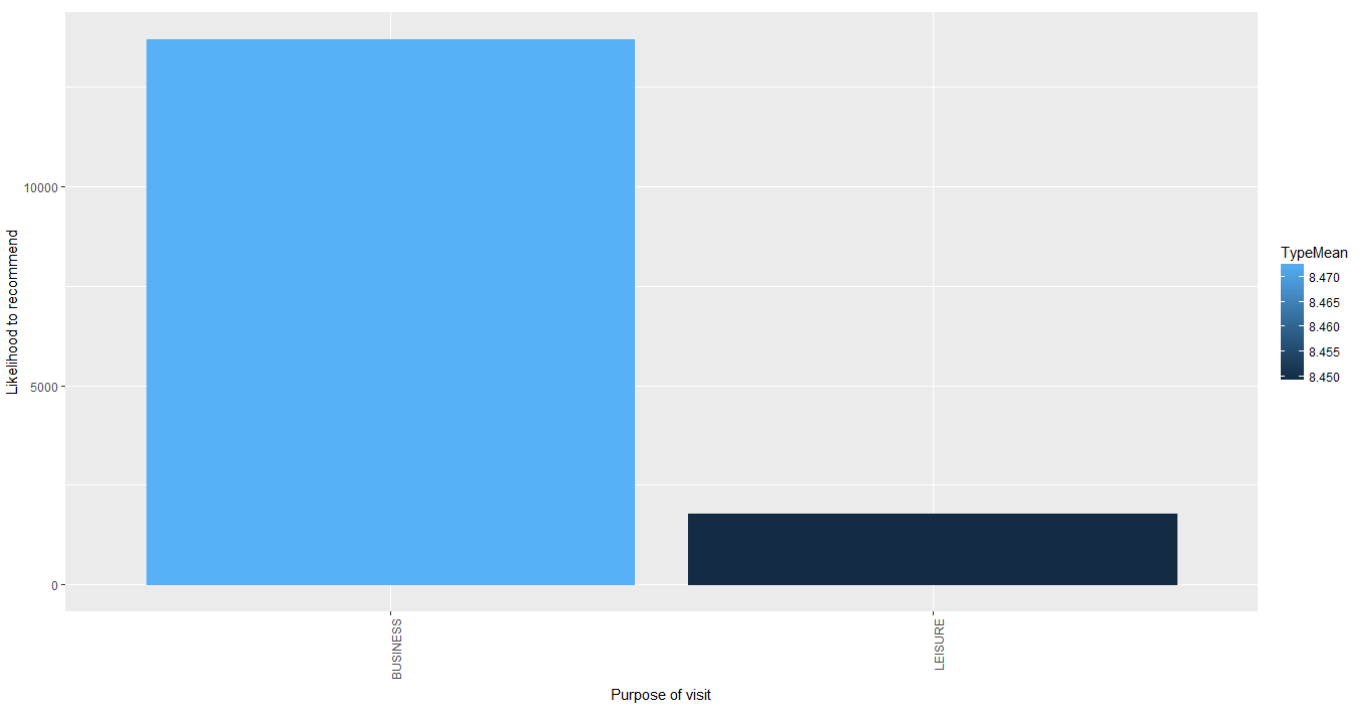


As we can see Room type DLXK, VW2Q, KNNS are frequently booked by the customers but they do not have good NPS rating. The management should focus on improving the condition of these room types to ensure increase in promoters.

Purpose of Visit:

Purpose of Visit (POV) is either for Business or Leisure.

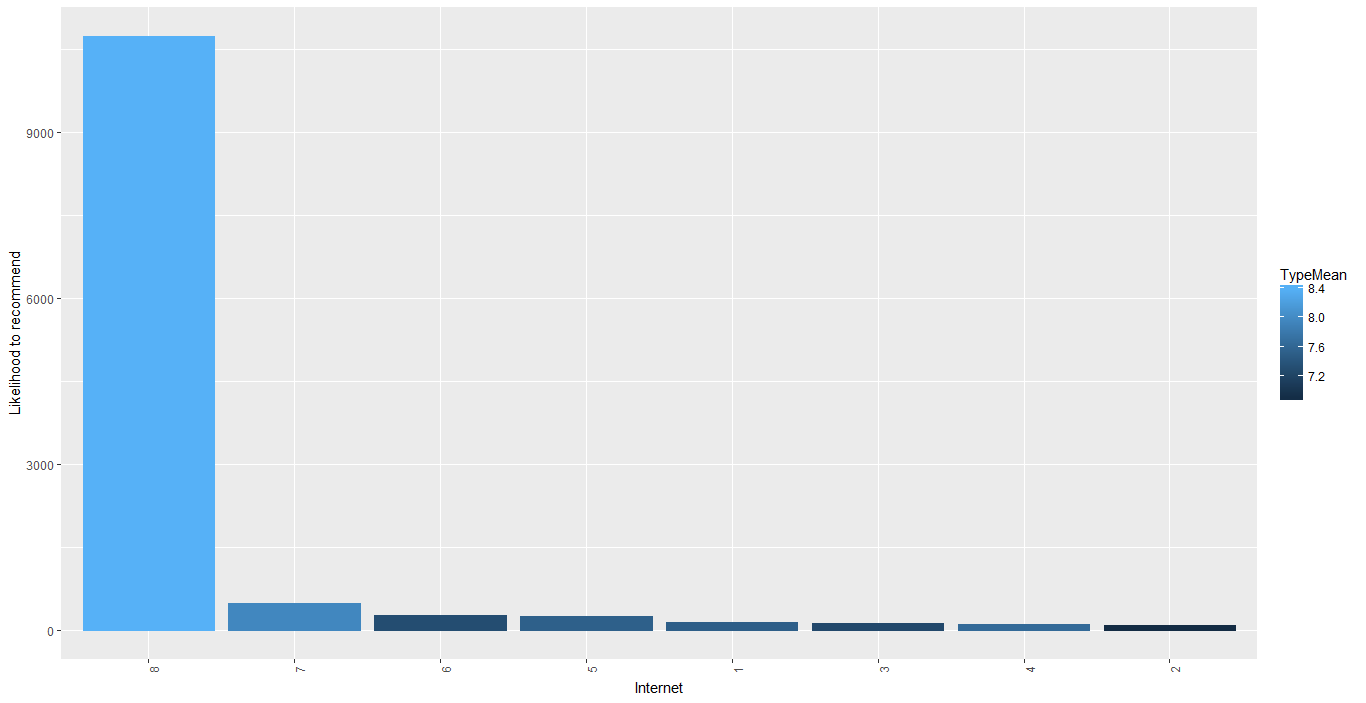




Most of the customers for Hyatt hotels come for Business trips. We should focus more on improving the facilities provided for the same purpose to ensure wide range of happy customers. But at the same time, we see that the Leisure based customers are not happy with their facilities. Happy customers, glowing recommendations!

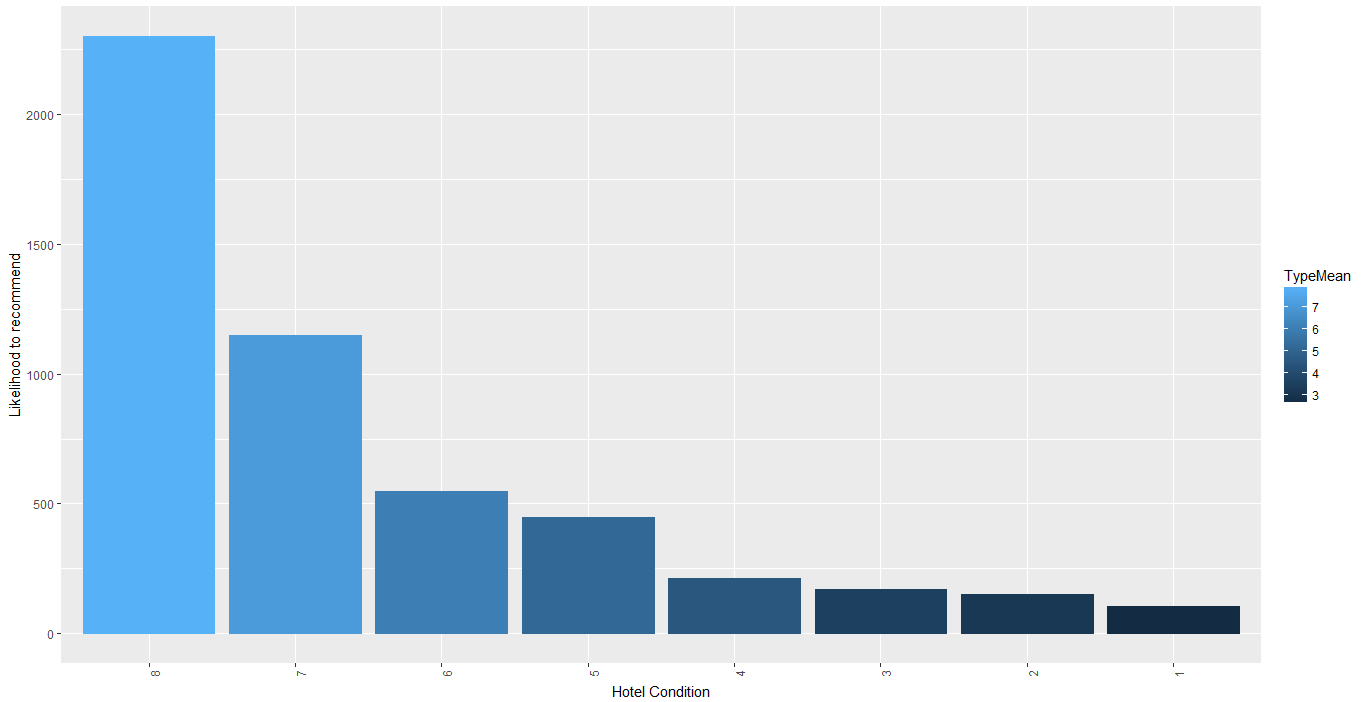
Internet:

The facilities provided by the Hotel are important as well. For example, as shown below good and reliable Internet services keep the customers happy in this tech savy and dependent world.



Condition of the Hotel:

Condition of the hotel the customers are living in should be good and also has a great impact on the likelihood to recommend atribute.



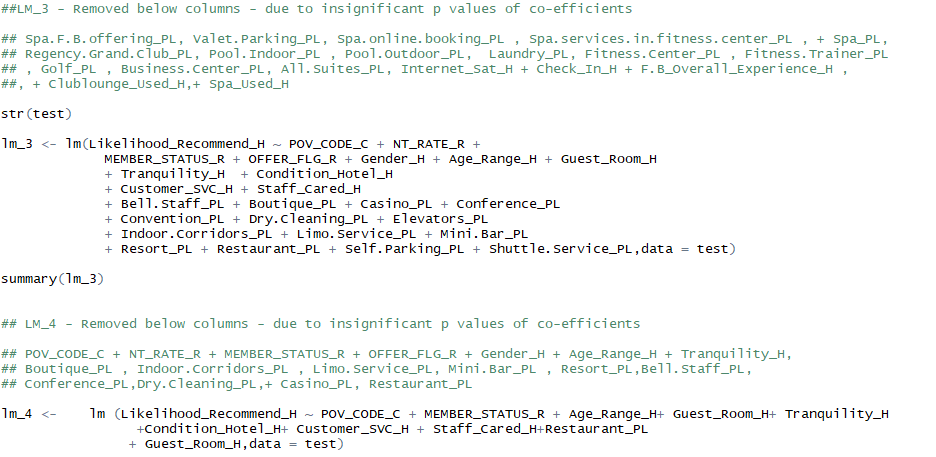
# Recommendation Analysis

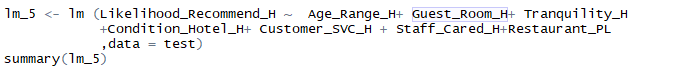
Modeling Techniques:

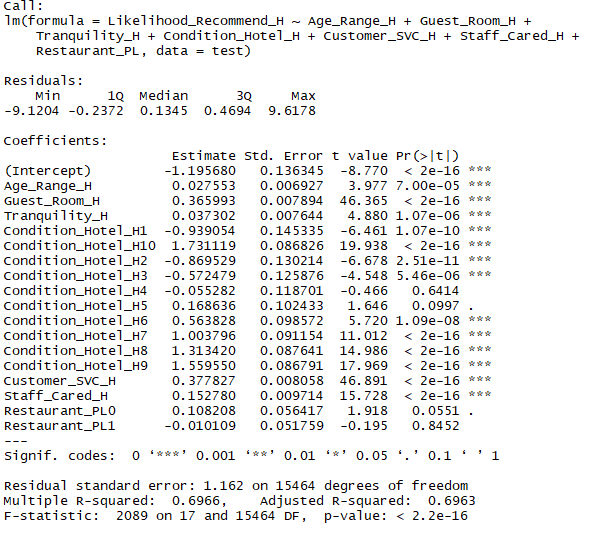
Instead of query based approach, we focused on model based approach and we started with linear regression model. We performed linear modeling with dependent variables Likelihood\_Recommend\_H and secondly NPS\_Type.

Linear Modeling to understand and predict the behavior of customers based ‘Likelihood to Recommend’. Initially we chose a few relevant variables according to the descriptive analysis and kept performing linear modeling until we obtained statistically relevant independent variables.



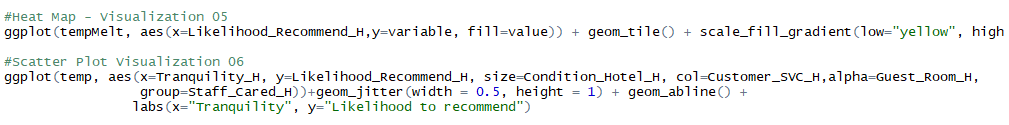


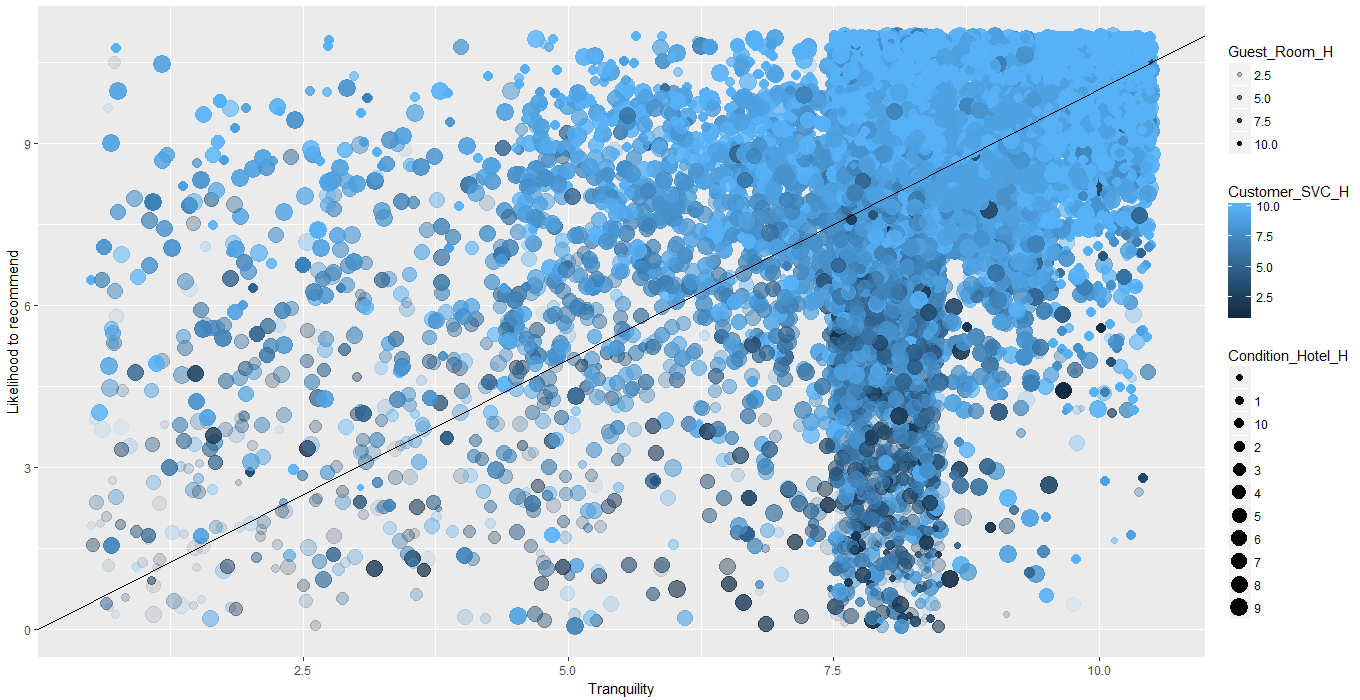


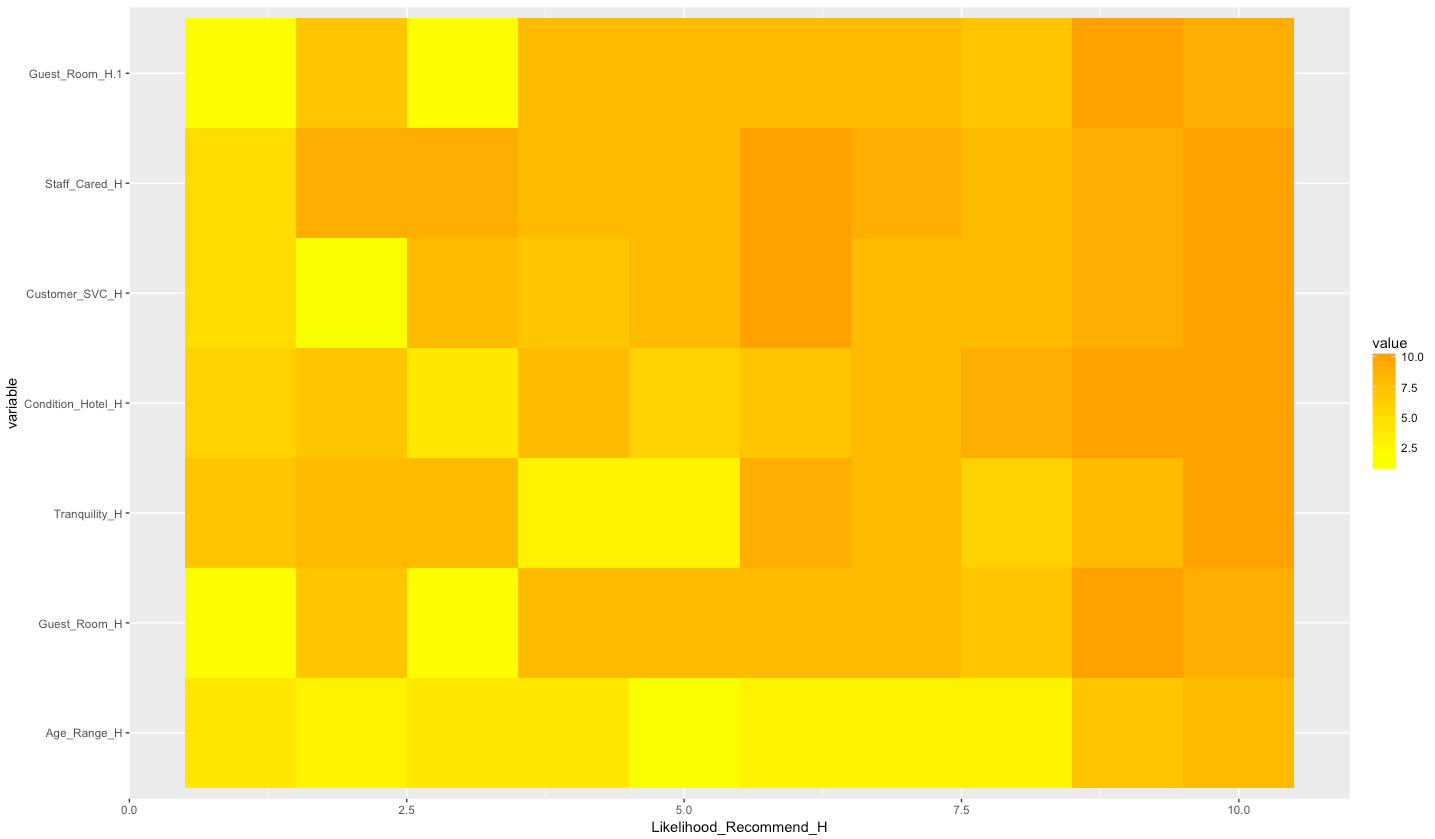


As we can see in the above summary, the R squared value of about **0.69** for Likelihood to Recommend is dependent on Age Range, Tranquility, Condition Hotel, Customer SVC, Staff Cared and Guest Room. The correlation between Guest Room and likelihood to recommend is high as we can see it in the summary above.

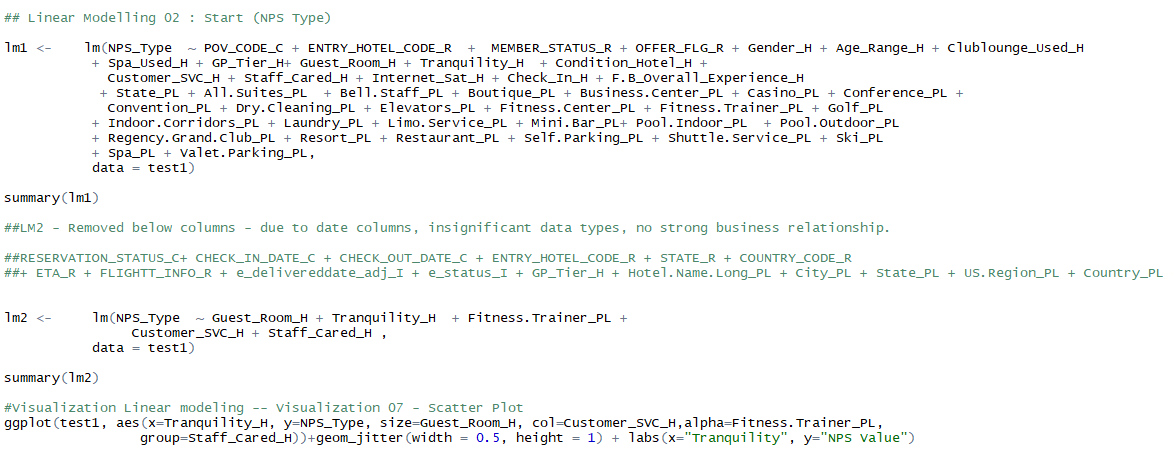
Heat map and scatter plot has been created for the same

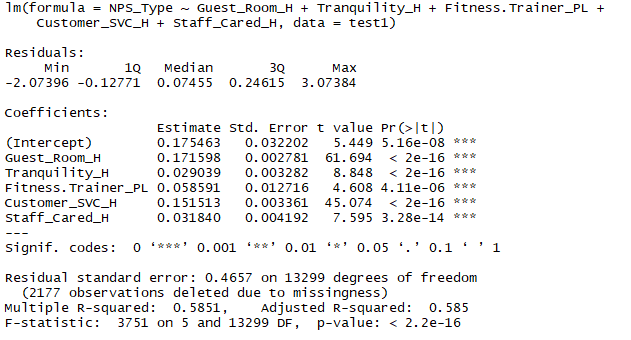






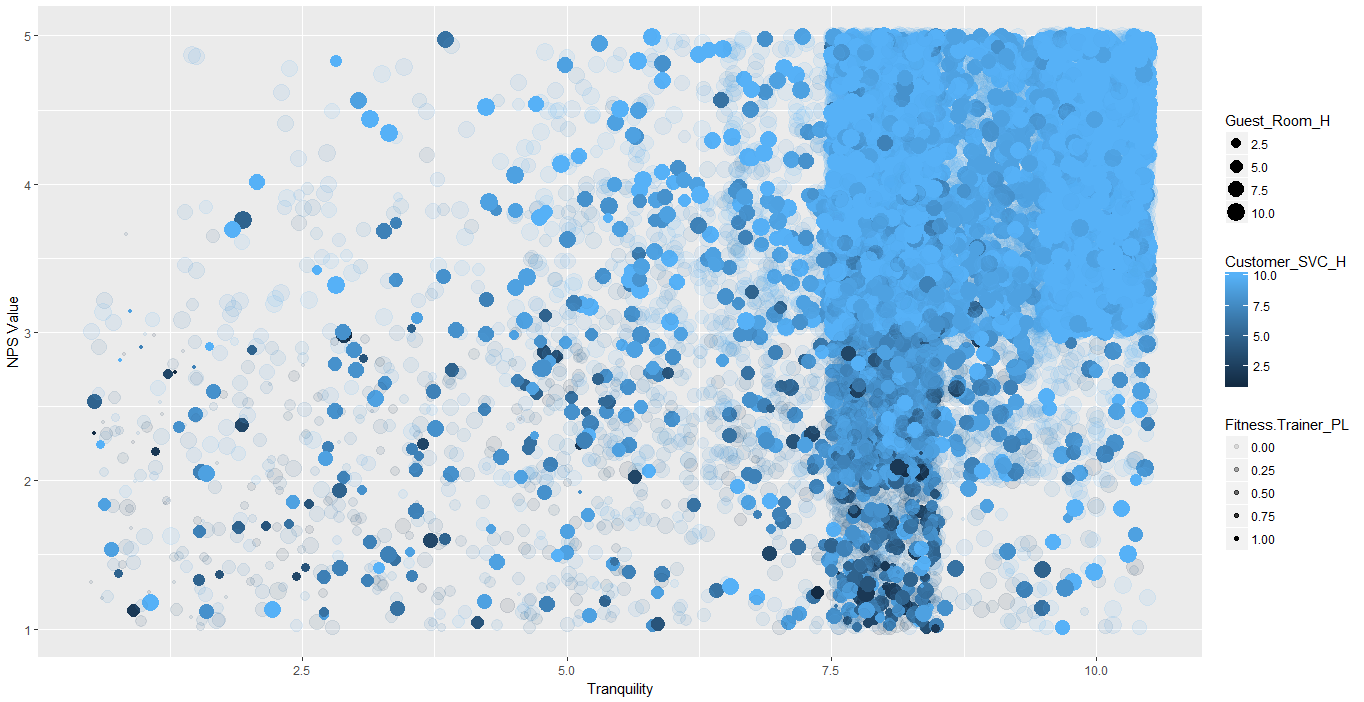
Next, we performed linear modeling to understand and predict the behavior of customers based ‘NPS Type’



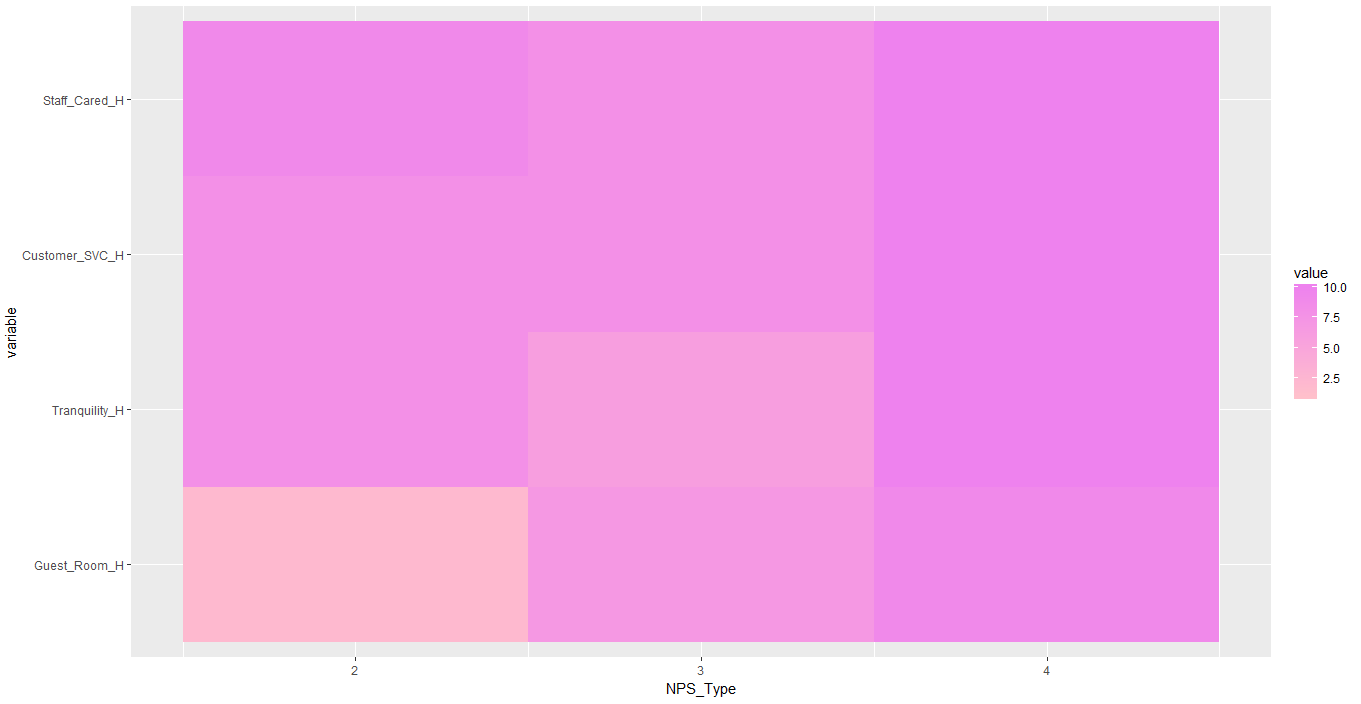


As we can see in the above summary, the R squared value of about **0.585** for NPS\_Type which is dependent on Guest Room, Tranquility, Customer SVC, Staff Cared and Fitness Trainer.

Heat map and scatter plot has been created for the same which provides an easier visual appeal for the business analysis.



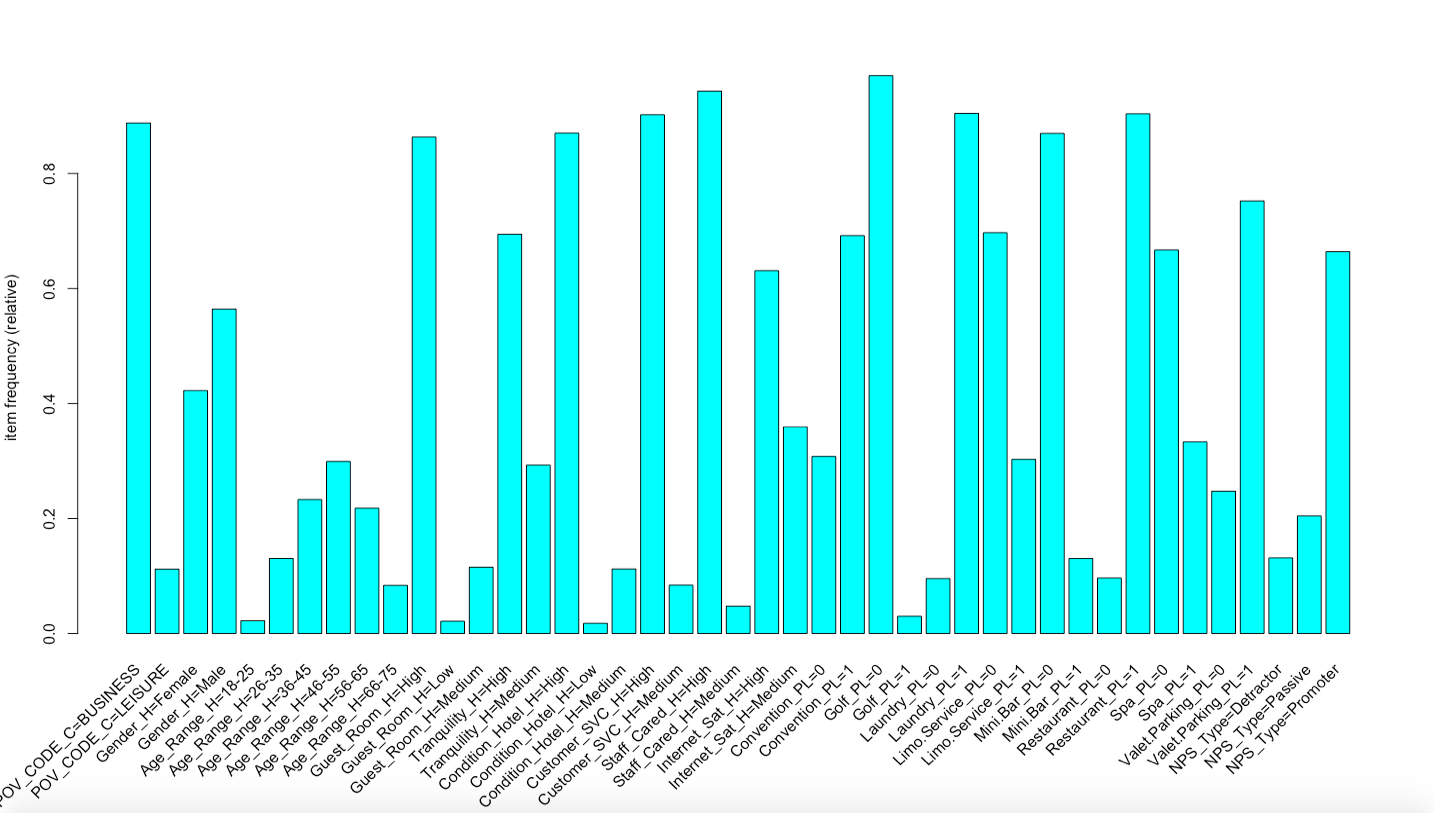
As shown in the scatter plot, the rightmost upper corner of Guest room, Customer satisfaction and Fitness training program services and feedback received for the same has the better chances of improving the business.

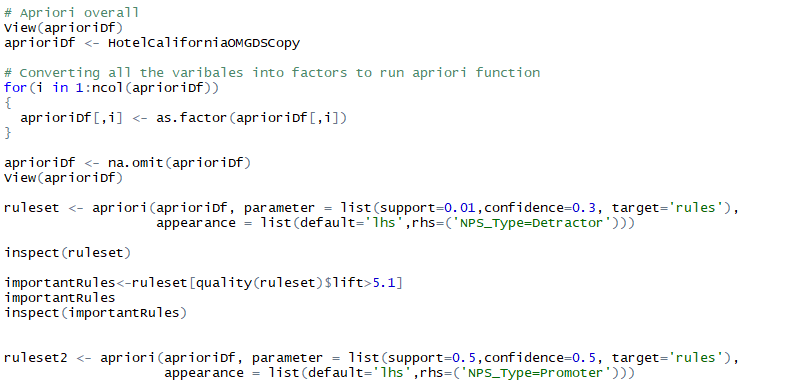


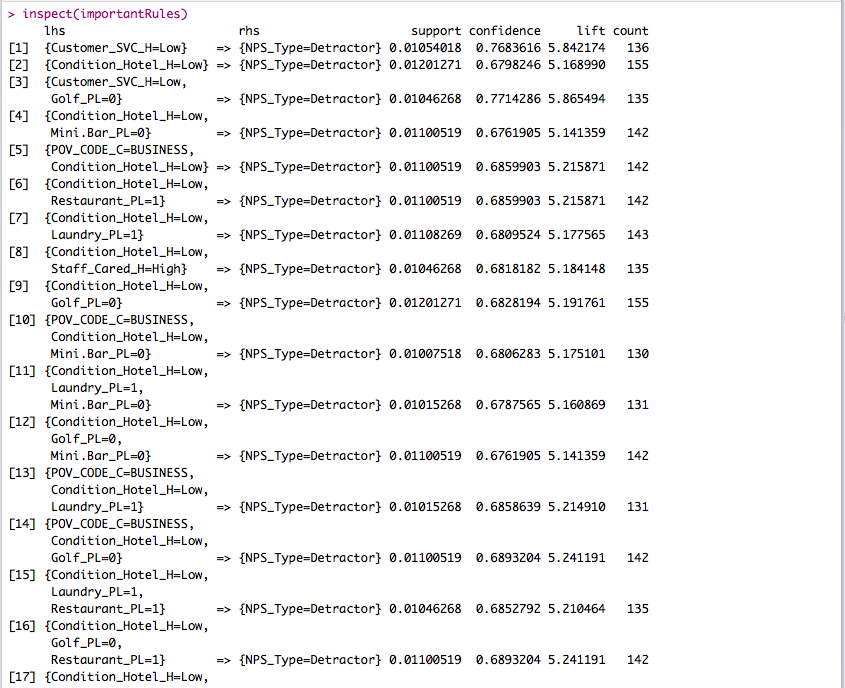
Apriori:

Apriori Algorithm is a data mining algorithm for mining frequent combination of services and facilities which specifically caters to customer preferences.

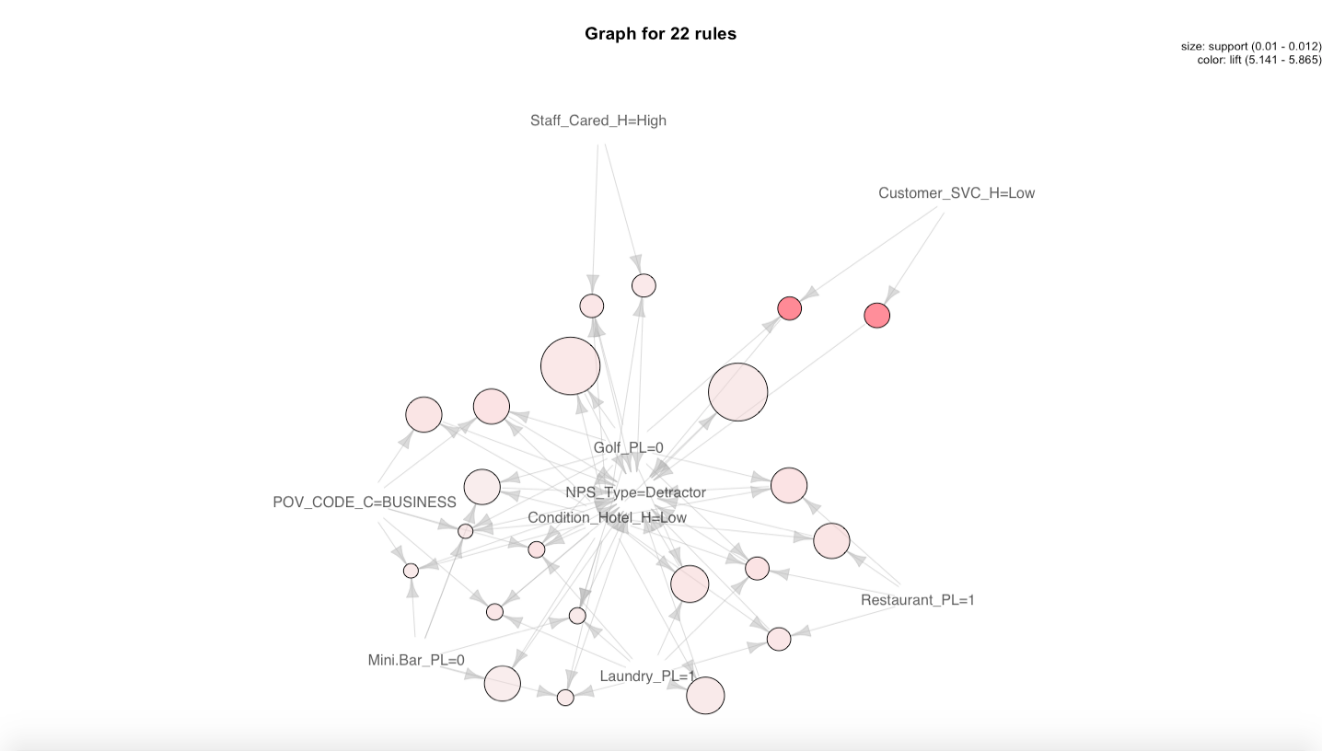
Item Frequency Plot to determine the frequency of occurrence of services





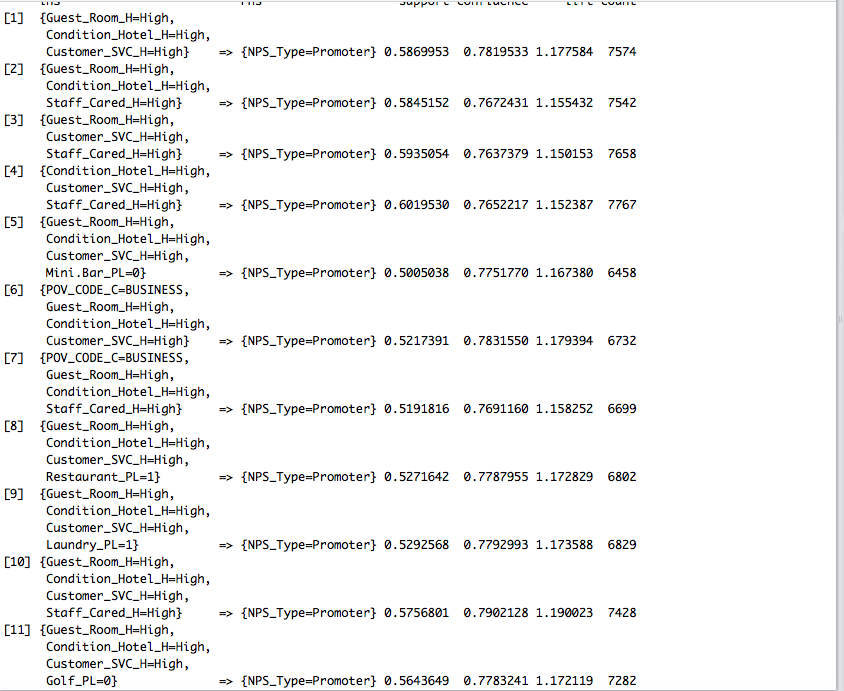


When we perform analysis on the NPS Type Detractor, we obtain good rules stating the association between the NPS Type Detractor and Customer SVC, Condition Hotel, Purpose of Visit, Mini Bar, Laundry and a few more with a support of about **0.01**, confidence as **0.3** and Lift above **5.1**.



With this analysis we recommend that we should improve the variables mentioned above to make sure we gain more number of promoters.

If we analysis the variables ensuring more number of promoters using the association rules as shown below:



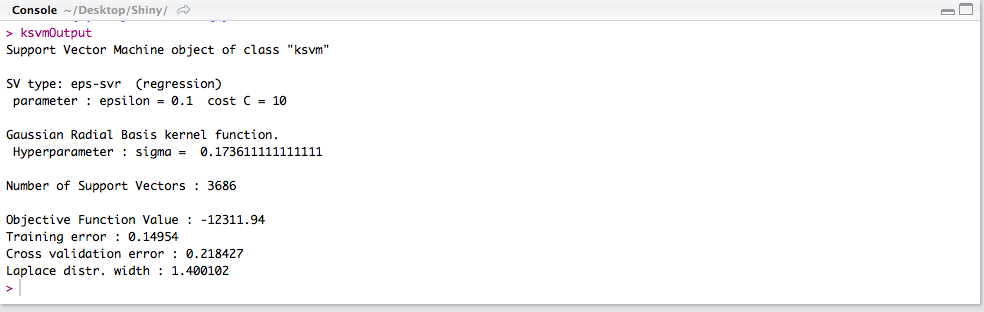
As we observed in linear modeling, Guest room, Hotel Condition, Customer SVC, Purpose of Visit are highly associated with gaining good promoters for the Hotel amongst the customer who come for Business trips.

SVM:

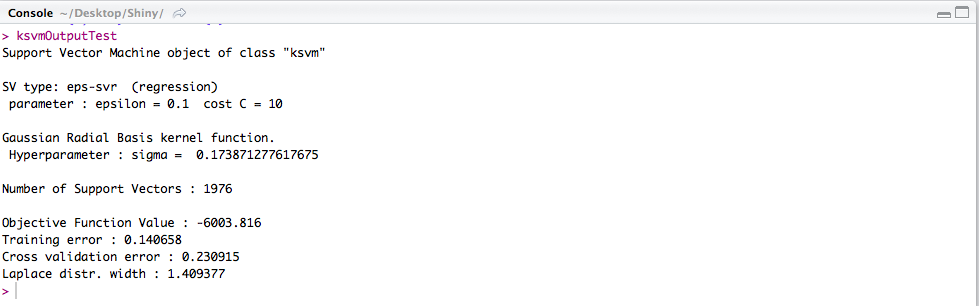
Machine learning algorithms like Support Vector Machines (SVM) are used to implement supervised learning of data for classification and prediction. SVM facilities the identification of a combination of services and facilities which will create a positive, neutral, and negative customer experience.

Guest Rooms, Tranquility, Hotel Condition, Customer Satisfaction, Staff cared, and Internet Usage have been identified as crucial parameters to influence customer favorability towards the hotel

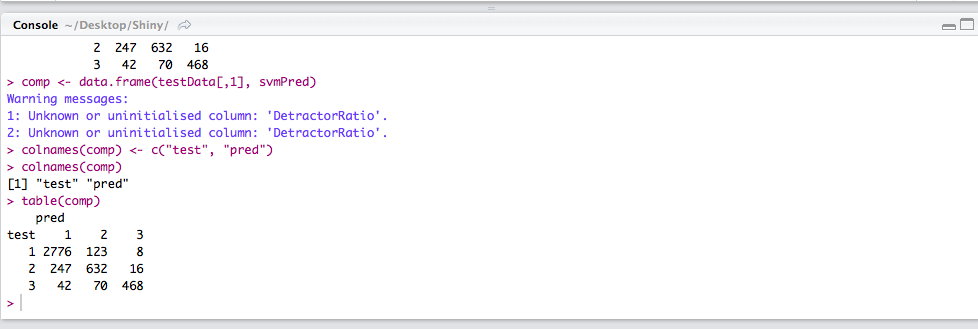
Training Data



Testing Data

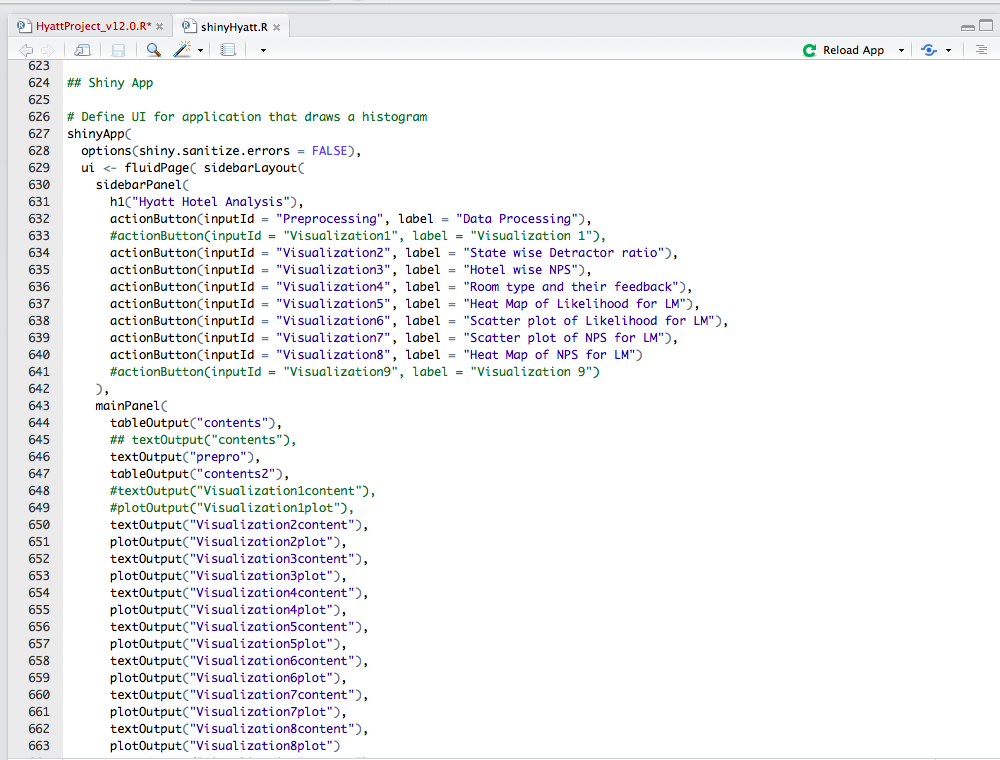


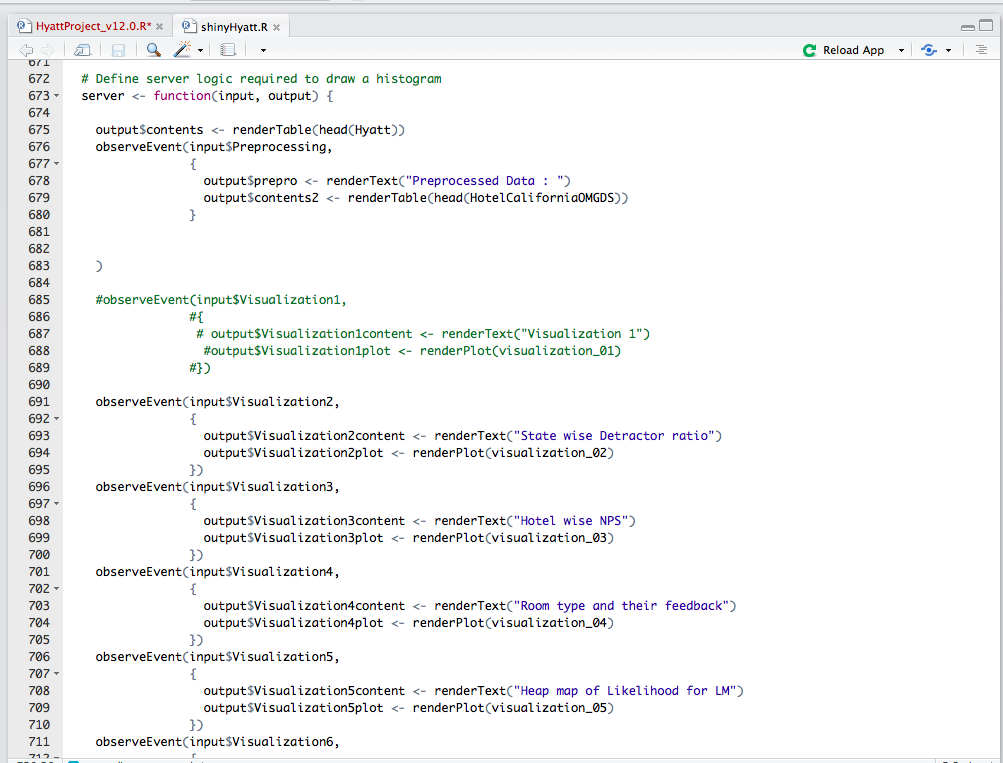
As shown below, Confusion Matrix from the SVM data model clearly indicates an accurate model for classification with elements in the diagonal showcasing precise classifications.



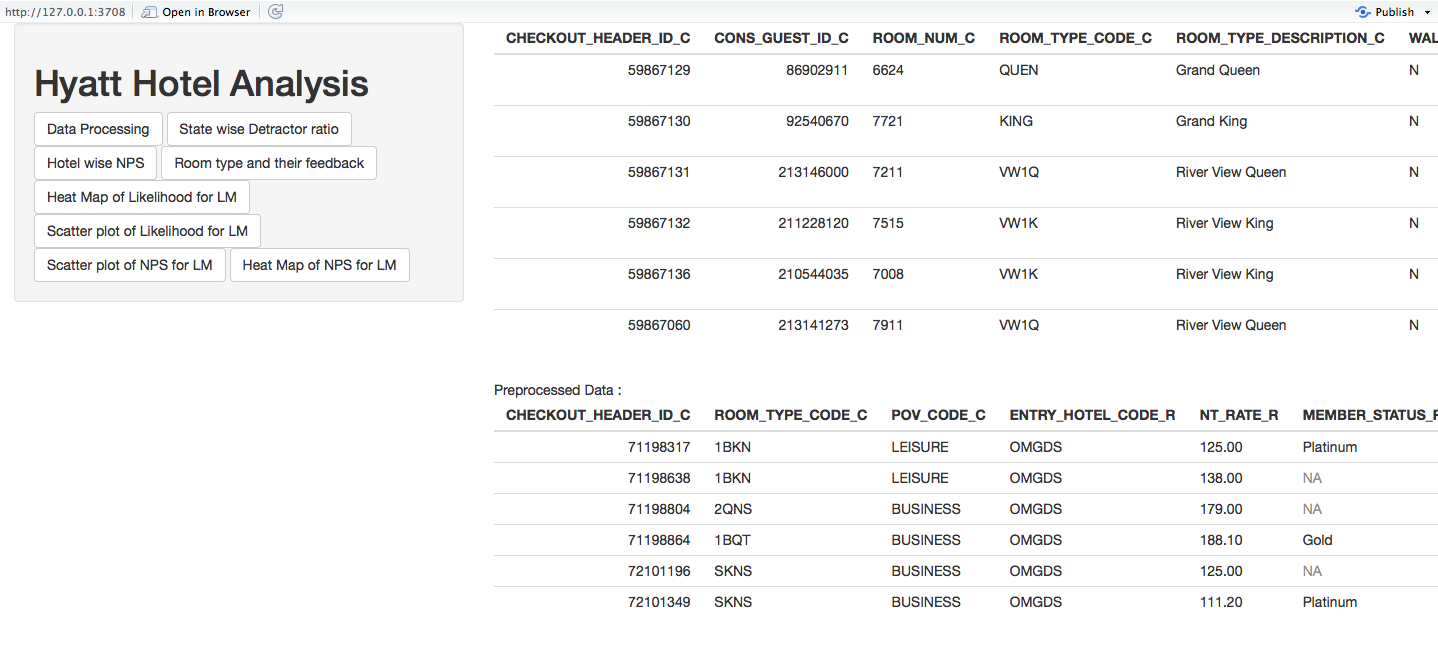
# Interactive Data Visualization

The graphs and charts created in R is not interactive hence we implemented Shiny apps for the same.









# Conclusion:

* To improve customer feedback from detractors, the hotel must focus on improving the services and facilities which have been proven to significantly influence customer feedback
* Services must be packaged and promoted in a combination of Guest Rooms, Tranquility, Hotel Condition, Customer Satisfaction, Staff cared, and Internet Usage tailored to specific class and category of customers
* Hyatt Hotel in El Segundo City of California should make significant improvements by providing better staff services to customers
* The Hotel should focus on maintaining a high standard for Hotel Rooms and Hotel Conditions
* Room upgrades, premium internet services, and personal staff could be provided to frequent customers with detractor and passive ratings
* Restaurants and cafes significantly influence customer feedback towards the hotel; Survey forms and feedback questionnaire could be filled up to customize food and bar menu
* Special rules must be set-up to ensure noise free environment and tranquility of customers