

Syracuse University IST 687

# Net Promoter Score (NPS) Analysis for Hyatt Hotel Group

Final Project Report

Ruiyang Chen Mengran Gao Qunfang Wu Xumin Xu Yunxi He

## CONTENTS

Introduction	2
Goals of the report and Target Audience	2
Business Questions	2
Question Analysis	3
Methodology and Data preparation	3
Data Cleaning and Justification	3
Justification of Each Method	7
Models Generated	7
Results and Analysis	7
Descriptive Analysis	7
Purpose of Visit (POV) Trends during the Year	7
Comparison Between Three Different POVs	8
Visitor Ratings, Likelihood to Recommend, and their Relationship with Hotel's type, Scale and Location	11
Types of Hotel, Revenue and Length of Stay	12
Results	13
Interpretation of Results	13
Association Rules	13
Reflections	13
Reference	13
Appendix	14
Codes used to Clean Dataset and Generate Graphs	14
1. Clean Dataset	14

## INTRODUCTION

Our group performed a data analysis for the Hyatt Hotel Group based on a detailed customer survey conducted to collect responses on different hotel services, hotel locations, and customer's overall experience at hotels of the Hyatt Hotel Group within the United States. Our analysis aimed to provide tailored recommendations on how to improve the overall rating and customer experience and evaluations on how these hotels have performed.

The project is initiated with the hope to better facilitate the general operation of Hyatt Hotel Group in the future, and our analysis is based on a series of customer survey responses as well as detailed information of the hotel location and services. To increase the accuracy, feasibility, and validity of our project, our team started with an in-depth understanding of the variables and questions mentioned in our dataset. Our team first filtered out columns that have excessive amount of empty cells and NA values (we are only targeting at cells that we need for our final analysis). By processing our raw data, our time to clean the dataset is significantly reduced which allowed us to focus on the data we need to do the analysis. Our next step was to identify the region that we were to focus on. By zooming in to the region we focus on, we were able to perform a more appropriate analysis for a focused group. Our team then looked at the proportion of different NPS types in our selected months that were used to perform the analysis. Additionally, we generated different models with our selected variables that are corresponding to the customer NPS rating to predict how customer will rate the hotel based on the services they have. Besides, we generated association rules that reveal the relationship between different kinds of service/amenities and the promoter NPS type. Lastly, we built KSVM models and Naive Bayes models to predict NPS\_Type in terms of two different visiting purposes (leisure and business) and four different seasons.

## GOALS OF THE REPORT AND TARGET AUDIENCE

Our report will have potential business value for Hyatt Hotel management team to conduct further business analysis and to facilitate their decision-making process. This well-organized and visualized report will help the management team better understand their current standing in the market and make crucial business decision on how to improve their hotel to accommodate their customers' needs.

People who are interested in investing or working with Hyatt Hotel Group can also use this report as part of their decision-making process and decide if Hyatt Hotel Group is the best match. For example, travel agents or organizations that provide online reservation service for travelers worldwide can refer to this report and design customized trips for customers with different needs and expectations.

Research and education institutes that have a specialized focus on the hotel industry can also use this report as a study material for research purposes. Our test models is authentic, novel, and effective that can be comprehended by people.

This report can be used as part of the annual review for performance of Hyatt hotels. During discussion of how to improve amenities of Hyatt hotels for better NPS performance. Especially

when there are plans for building a new Hyatt hotel, the report provide suggestions that they can refer to in terms of how hotel location, services, and other amenities provided might have impact on customers' level of satisfaction.

## DATA CLEANING AND JUSTIFICATION

The dataset given is excessively comprehensive and detailed, and it has large number of cells that we do not need and thus need to be cleaned. Looking at the whole dataset, it is comprised of data of 12 months, and each month's data has over 500,000 rows to deal with. With such large number of columns and rows, it was urgent that we clean the data set and extract only the relevant data we need.

### Step #1 – Determine the Month

Looking at our dataset, we first decided to use the data in February, May, August, and November. Our logic was to pick one month from each quarter, and each month should have practical and analytical meaning that is worth exploring. We chose February in the first quarter because it is considered the offseason; it is not the time for New Year and it is not the time for spring break. Considering offseason increases the feasibility of the report because hotels should balance their performance during both offseason and busy seasons. We chose May because it is the start of busy season and travelers start to flock to hotels for vacations; it might also be a good start to figure out some potential trends. August is the busiest month of the year as tourists travel domestically and globally for their vacation. This is the time of the year that various ratings of hotel occur, because unexpected situations are more likely to happen during busy season. People are also more likely to travel overseas. Finally, we chose November because of the Thanksgiving break during which people are very likely to go on a trip with their family or their special ones.

### Step #2 – Remove Empty Cells and NAs

First, we decided to delete all rows with empty NPS\_Type columns because the primary purpose of the report is to perform analysis based on NPS rating of each hotel. Since the type of NPS that serves as an indicator of hotel's performance is determined by customers who contributed valid and complete survey responses, our first step was to remove all unnecessary empty cells in NPS\_Type column and the rows that they are in. The logic behind the removal of all empty rows that associated with NPS types is that our ultimate goal is to provide suggestions to improve hotel's NPS score, and thus empty cells of NPS\_Type affect the accuracy of our analysis. Another advantage of removing these rows is that it shrinks our dataset and reduce our processing time.

Second, we decided to divide our processed dataset into two subsets based on purpose of visit, which are leisure and business. Our processed dataset for each month was then made into two subsets and was exported for further analysis.

## Step #3 – Determine the Region

After deleting all rows that has empty NPS type, our group continued to try to focus on a particular region that to perform our analysis. Our team created a bar chart to represent the average likelihood of recommendation ratings (ranged from 1 to 10). Plots were created based on sub-continent and the average ratings in each of the four months, and we have the following plots:

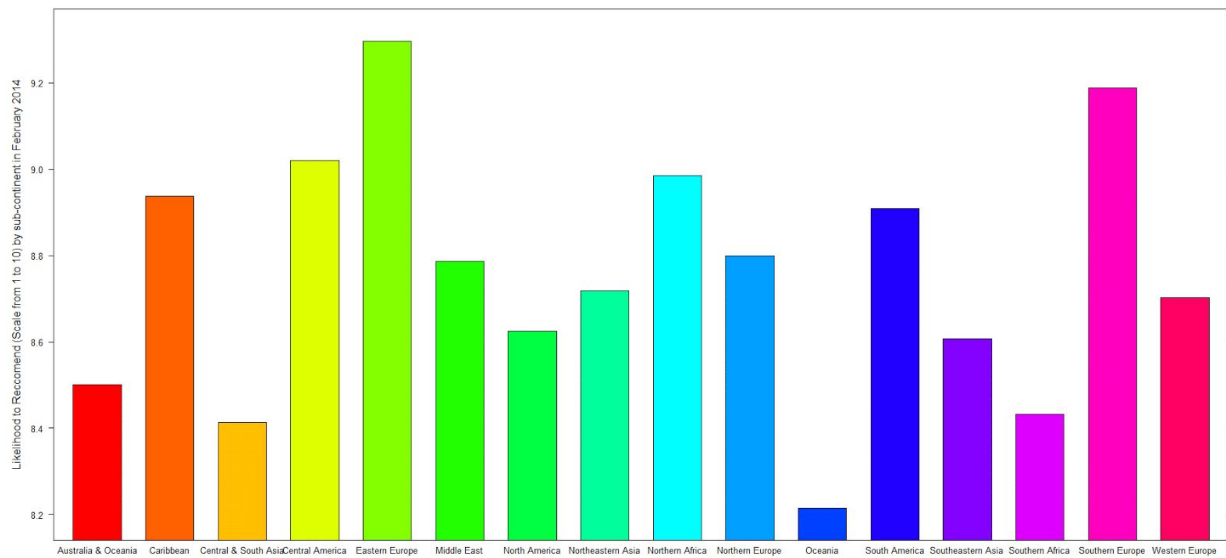


Figure 1 Average Ratings in February

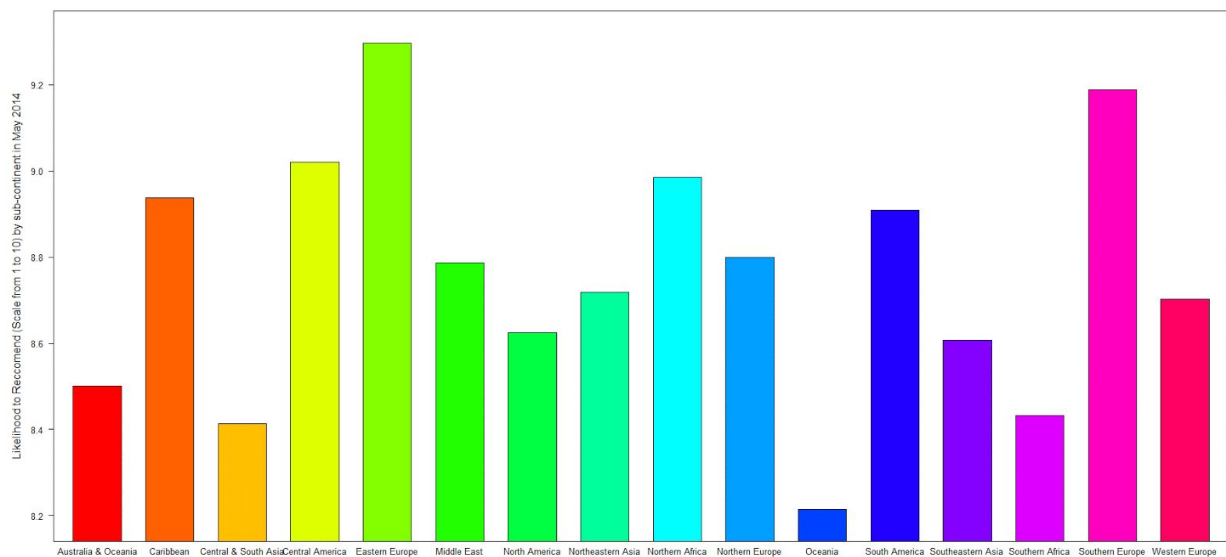


Figure 2 Average Ratings in May

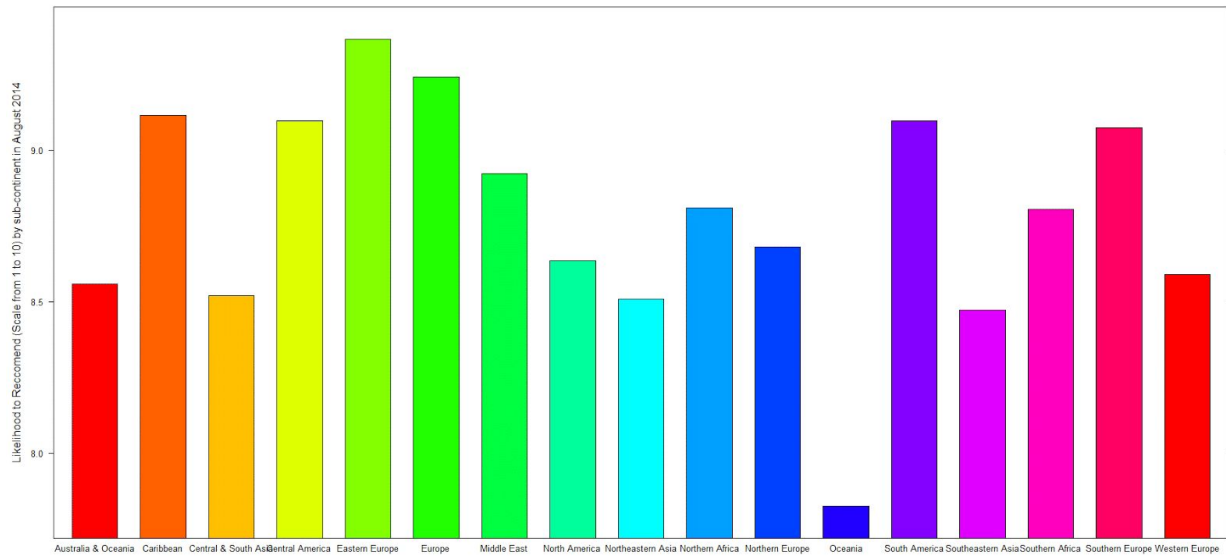


Figure 3 Average Ratings in August

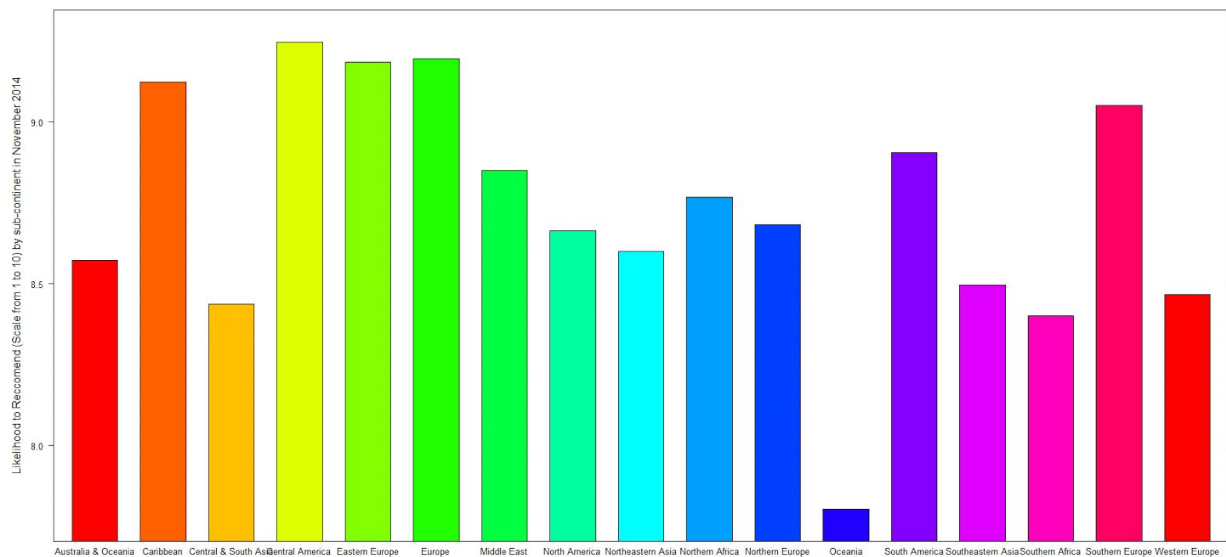
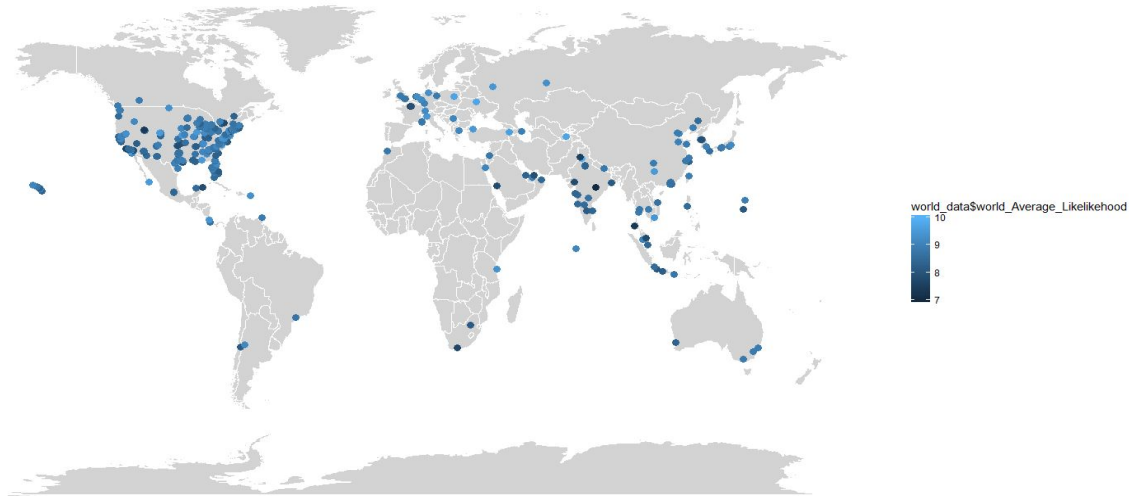


Figure 4 Average Ratings in November

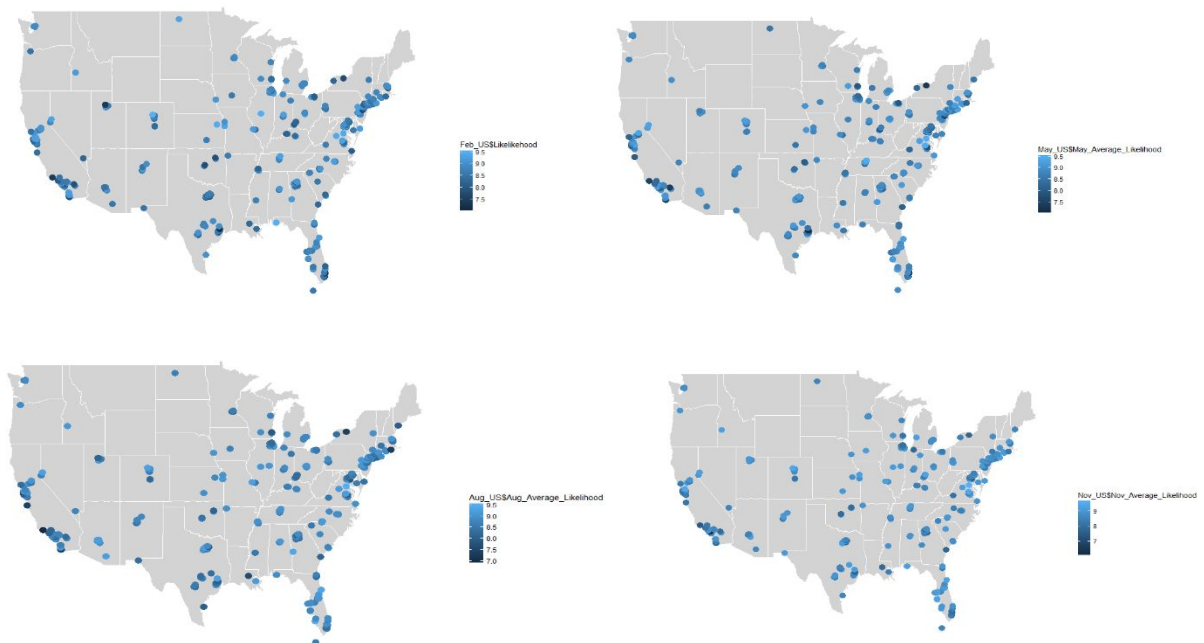
The average ratings of hotels vary a lot based on the sub-region and sub-continent they are located. Thus, based on these bar plots created, we decided to consider cultural difference and value divergence within each sub-continent. In another word, a same rating in different sub-continent might represent very different customer satisfaction situation. Consequently, it is more logical to focus on just one sub-continent and investigate the pattern accurately. While the process of research can be transformed and used to other desirable sub-continent as well.

To have a more visualized and general understanding of the likelihood to recommend, we decided to map out the likelihood to recommend of each hotel worldwide in February only. We

chose that data in February because we only needed to have a general understanding of which part of the region has more customer ratings so that we can focus on one specific regions with enough data amount to continue our analysis. Here is our world map of likelihood to recommend in February.



From the map shown above, it is clear that hotels in the United States are more frequently rated and have more data to analyze. To test our observation and assumption, we took a closer look at each of the four months we selected and the likelihood to recommend for hotels in the United States.



Each point shown on map represents ratings for hotels in a popular travel destination that each Hyatt Hotel is located. The darker the point, the lower the rating. Comparing the four map generated, we found out that ratings of hotels in the United States experienced fluctuations

and varied in likelihood score, and thus we concluded that we would use the data collected for hotels located in the U.S. as our subset data which we would do further analysis on.

## BUSINESS QUESTIONS

Based on the data cleaning process, we have several interesting questions regarding the dataset and its pragmatic meaning.

1. What are the factors affecting NPS for opening hotels in Northern America? And how are factors differ for “Leisure” purpose, “Business” purpose, and “Leisure & Business” purpose?
  - I. Is the ratio of different purpose of visit changes during a year?
  - II. Are the overall ratings different among different purpose of visits?
  - III. Visitors have different purpose of visit, while hotels have different types, is visitors who have different purpose of visit visit different types of hotels?
  - IV. Analyze whether hotel needs to be adjusted by their main type of pov (find out what each pov guest like, and whether the hotel has it)
2. Some interesting facts of Hyatt Hotels in US area.
  - I. Are the types of hotel affecting hotel’s revenue?
  - II. Distribution of hotels in US region.
3. Prediction model for NPS using the factors above? (If the factors are quite different for “leisure” and “business”, we need to build two prediction models.)
  - I. What factors are related to NPS types of promoter the most?
  - II. How to build an accurate predict model?

## QUESTION ANALYSIS

### NPS Promoter Score

First of all, it is imperative to know what the NPS is. NPS, also known as Net Promoter Score, is widely used in the business industry as a measure to evaluate customer experience and to predict future growth. We used the NPS package installed in RStudio to calculate NPS for each hotel in the U.S. in our 4 selected months.

Some ways to use the NPS functions are:

- `nps(x, breaks = list(0:6, 7:8, 9:10))`
  - This function calculates a Net Promoter Score from a vector of Recommend scores



## NPS Analysis for Hyatt Hotel Group

- `nps.se(x, breaks = list(0:6, 7:8, 9:10))`
  - This function calculates the standard error (see below) of a Net Promoter Score
- `nps.var(x, breaks = list(0:6, 7:8, 9:10))`
  - This function calculates the Net Promoter Score variance

1. Our group decided that it is quite necessary for hotels to know whether different NPS Type results from different customers visiting purpose (Business VS Leisure). Our group target at this aspect and further create the sub-dataset based on POV\_H. After that, we use descriptive tools such as curve graphs or bars and charts to compare the proportion of customers for different purpose in each four month selected as well as which types of hotel attract more business-trip customers. Furthermore, we test the association rules to figure out what kind of services business-purpose customers would value most according to the support and confidence value. Based on those results and tests, recommendation for hotels to improve specific for their customers are valid and feasible.

2. To further explore factors closely relate to NPS type, selection is made from hundreds of variables, our group picked several important factors such as Market\_Group\_C, Room\_Type\_Description, Length of Stay to test the correlation with NPS Type. Moreover, we tried to use February data as training data, and May, August November data as test data to build ksvm models and NB models to test how factors (independent variables) influenced NPS\_Type in different seasons. Using the prior data to build models, we can predict the following seasons' or months' NPS\_Type in advance in the same year.

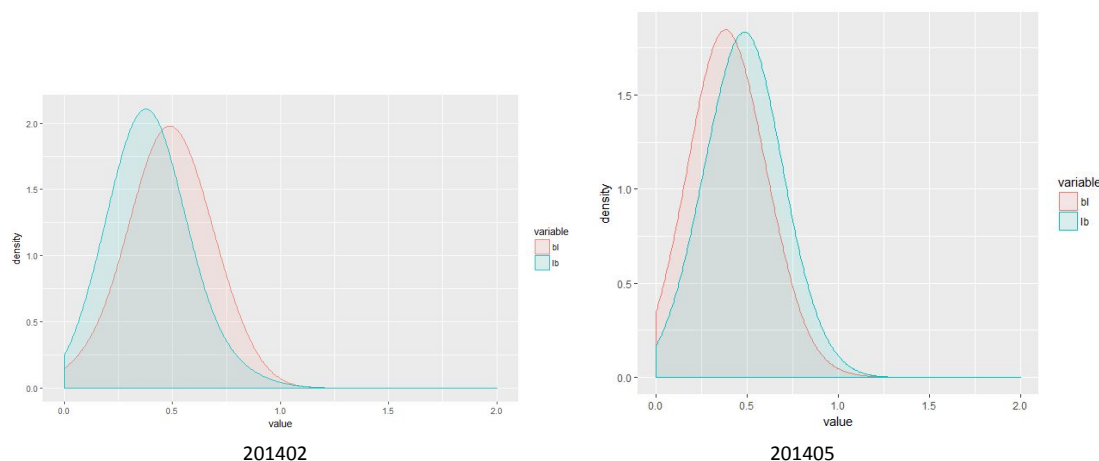
## RESULTS AND ANALYSIS

### PURPOSE OF VACATION (POV) TRENDS AMONG THE YEAR

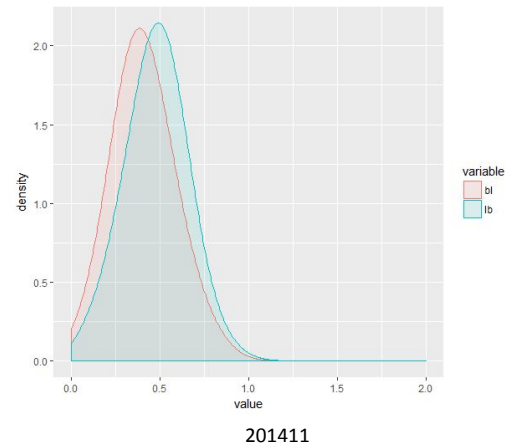
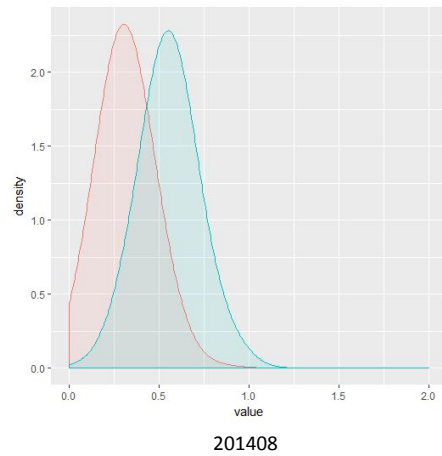
#### 1. Ratio of business visitor/leisure visitor vs. ratio of leisure visitor/business visitor

We took the proportion of business visitors in each of the months we selected divided by the proportion of leisure visitors, and calculated the ratio. We then repeated the process, but this time we used the proportion of leisure visitors divided by the proportion of business visitors.

Red area: Ratio of business visitor/leisure visitor; blue area: ratio of leisure visitor/business visitor

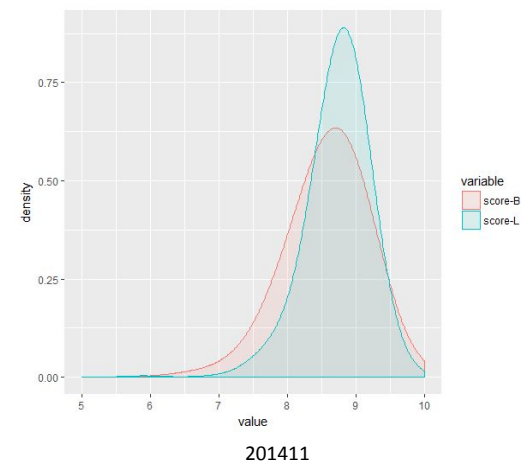
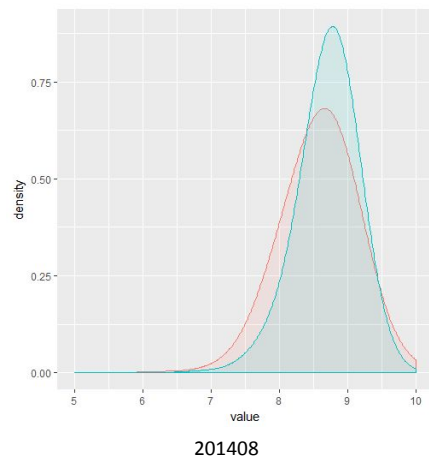
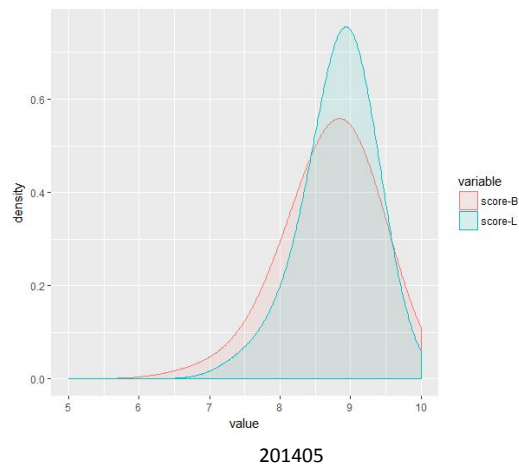
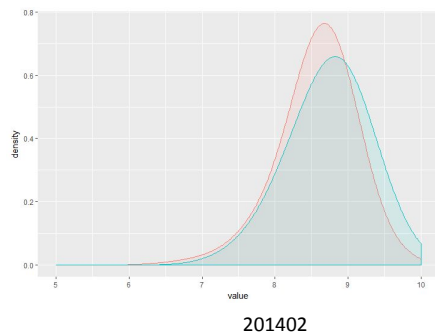


## NPS Analysis for Hyatt Hotel Group



## 2. Business visitor rating vs. leisure visitor rating

Red area: business visitor rating density; blue area: leisure visitor rating density



From the density plots above, we found that business visitor ratio is only more than leisure visitor ratio in February, while leisure visitor ratio is much higher than business visitor ratio in August. This observation meets our expectation that people doing business in the beginning of

the year and travel around for leisure in the summer.

### COMPARISON BETWEEN THREE DIFFERENT POVS

By looking at our variable glossary, we discovered that there are two columns related to visitors' purpose of visit.

Column Name	Definition
POV_H	Guest's purpose of visit
POV_CODE_C	Purpose of visit

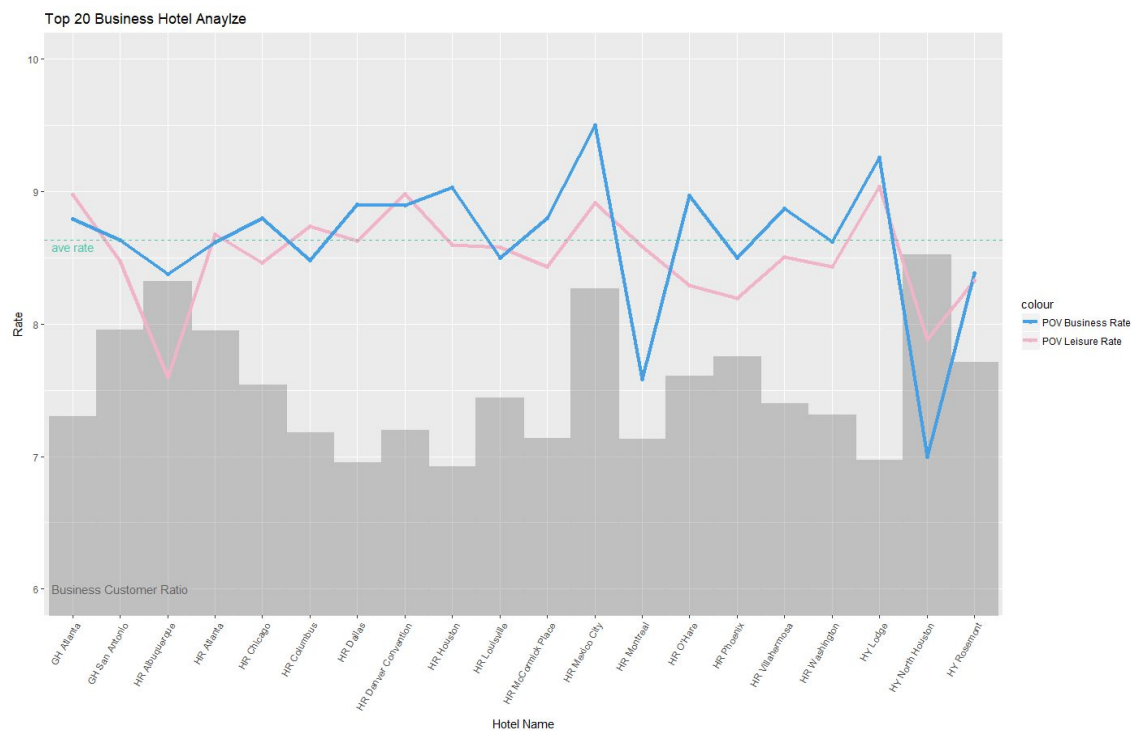
Use the table() function in R, we discovered that customer POV include leisure, business, and combination of leisure and business. Thus, we took a closer look at the relationships between these three different types of POVs as well as their relationships with different hotel brand's focused group.

To continue the comparison between these three different types of POVs, we first defined three hotel type depends on the type of POV its visitors have. The first type of hotel is business hotel, and most of its visitors came for business purposes. Leisure hotel has the most number of travelers visiting for leisure purpose. Leisure and business hotel is a hotel that has an equal (or almost equal) ratio of leisure and business.

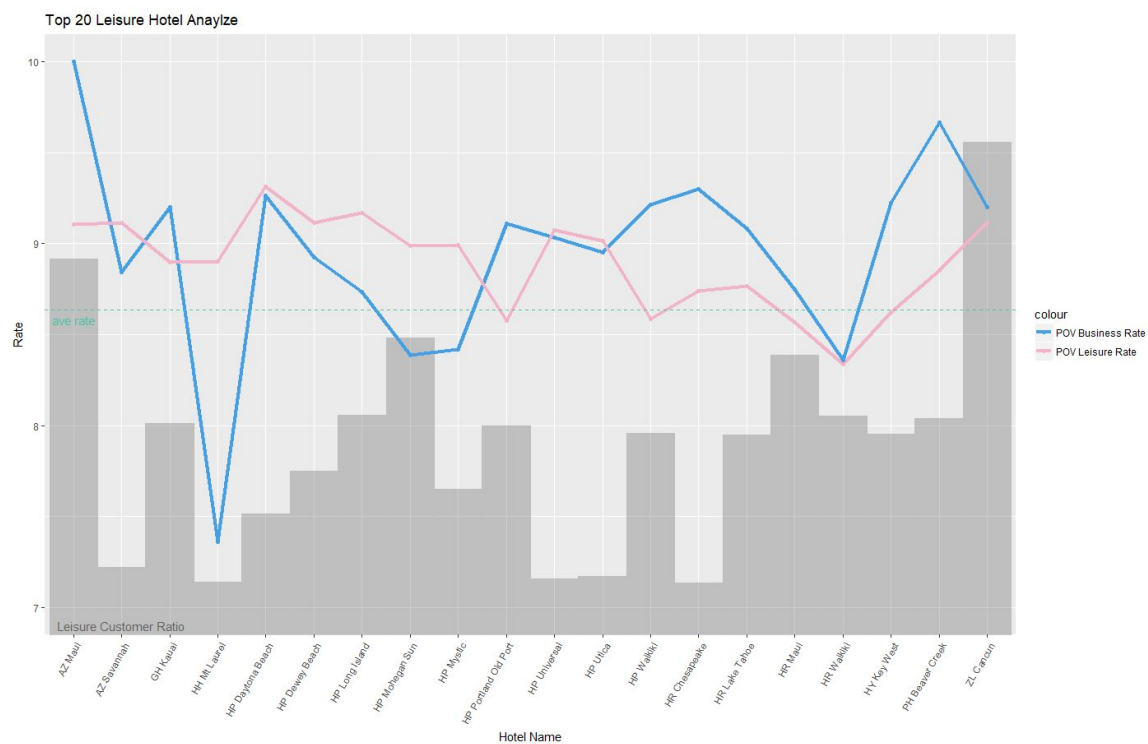
We merged four months' data together and group by different purposes: leisure, business, and leisure & business. We continued to compare different columns among leisure, business and leisure & business, and the expected result is the patterns (i.e., NPS, price, service) among the three purposes.

1. Top 20 hotels that have the highest ratio of business visitors

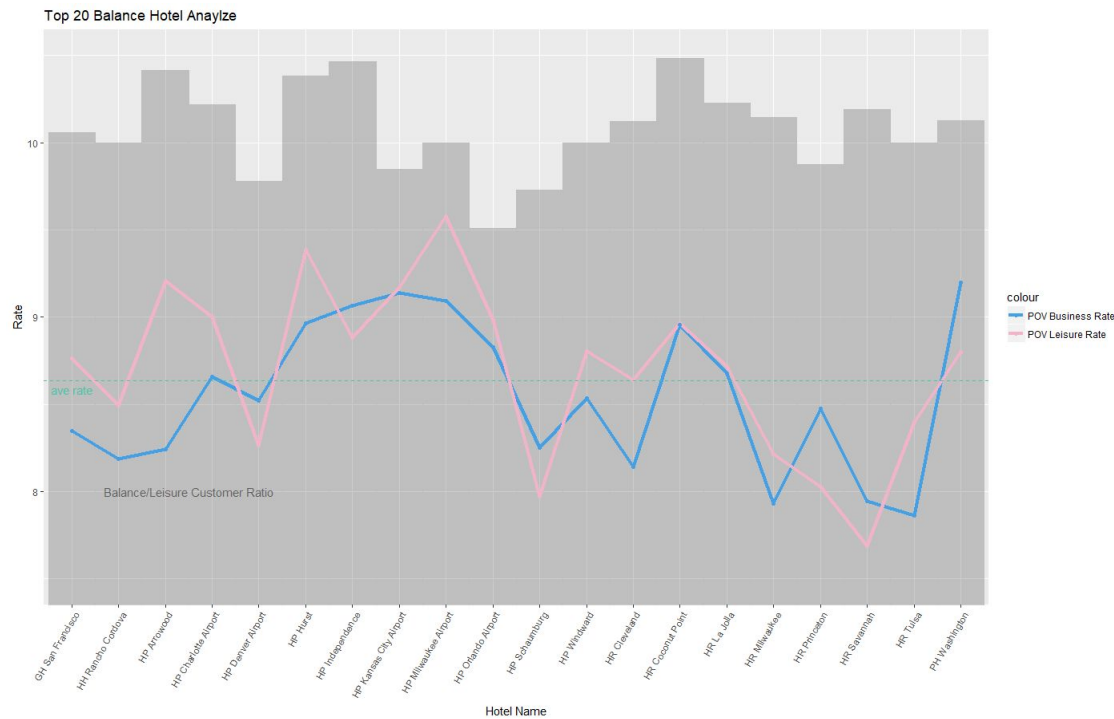
## NPS Analysis for Hyatt Hotel Group



## 2. Top hotels that have the highest ratio of leisure visitors



### 3. Top hotels that have the closest ratio of business and leisure visitors

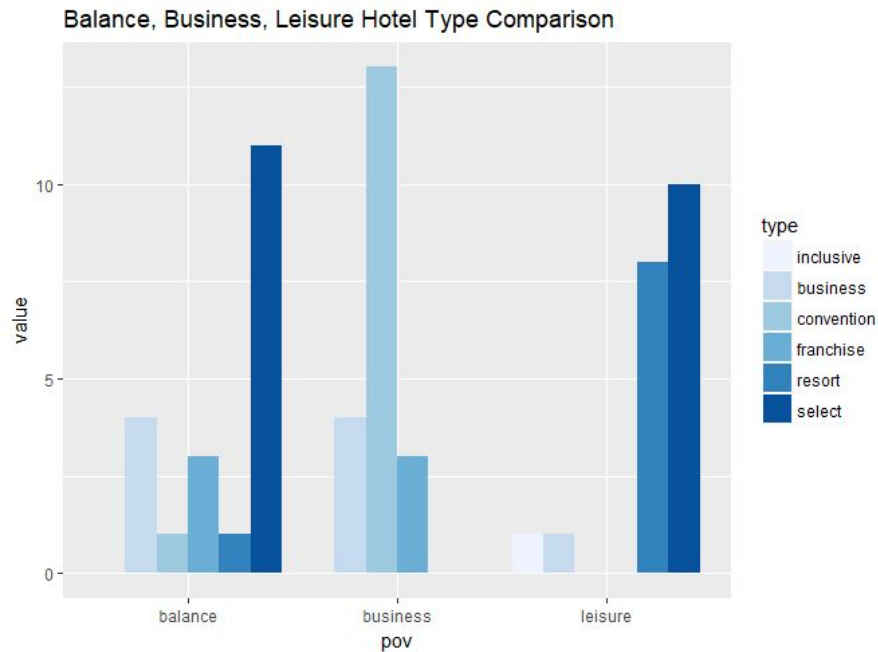


Visitors have different purposes of visit when they visit hotels. In these line charts, it is clear that each hotel has their majority customers who specifically focus on one purpose of visit. Some hotels even have more than 80% of visitors have the same purpose of visit. For the hotels that have almost the same ratio of business visitors and leisure visitors, we label them as balance hotels.

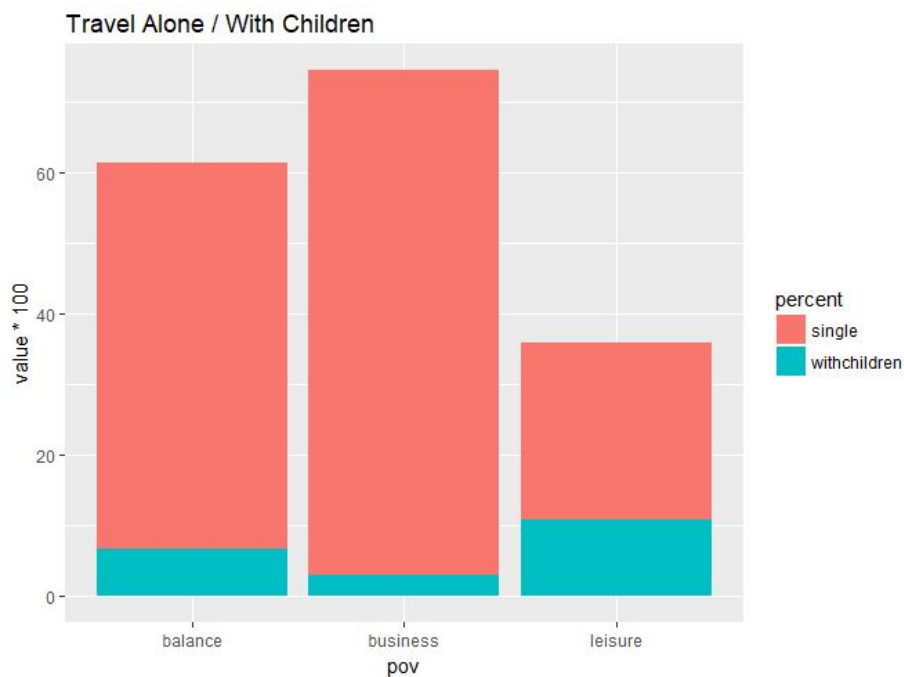
However, from the charts, although some hotels have high ratio of leisure visitors, the ratings from leisure customers is lower than business customers are. On the other hand, some hotels have the same ratio of business or leisure visitors, but ratings from these two type of customers have large gaps. In these cases, we suggest that the hotels might need to consider changing their focused group or adjusting their hotel's amenities to accommodate the majority of visitors and thus to increase their rating and likelihood to recommend.

For example, among the top business hotels, hotel HY North Houston has over 85% of visitors who visited for business purposes. However, the ratings of business visitors who visited HY North Houston is not only much lower than the average rates, but also lower than the rates of leisure visitors. Among the top leisure hotels, visitors who came for leisure purposes and visited hotel HP Waikiki, hotel HR Chesapeake, hotel PH Beaver Creek rated these hotels much lower than the visitors whose purposes were business. We proposed that these hotels needed to adjust their facilities and amenities to meet the majority of visitors' demand to improve their visiting experience.

## DIFFERENT TYPES OF HOTEL IN TERMS OF POVS



From the plot above, we can clearly see that visitors having different POV will visit different types of hotels. Most business visitors visit business, convention and franchise hotel while most leisure visitors visit resort and select service hotels. This observation also agrees with our expectation that business, convention hotels will have most business visitors while resort hotels will have most leisure visitors

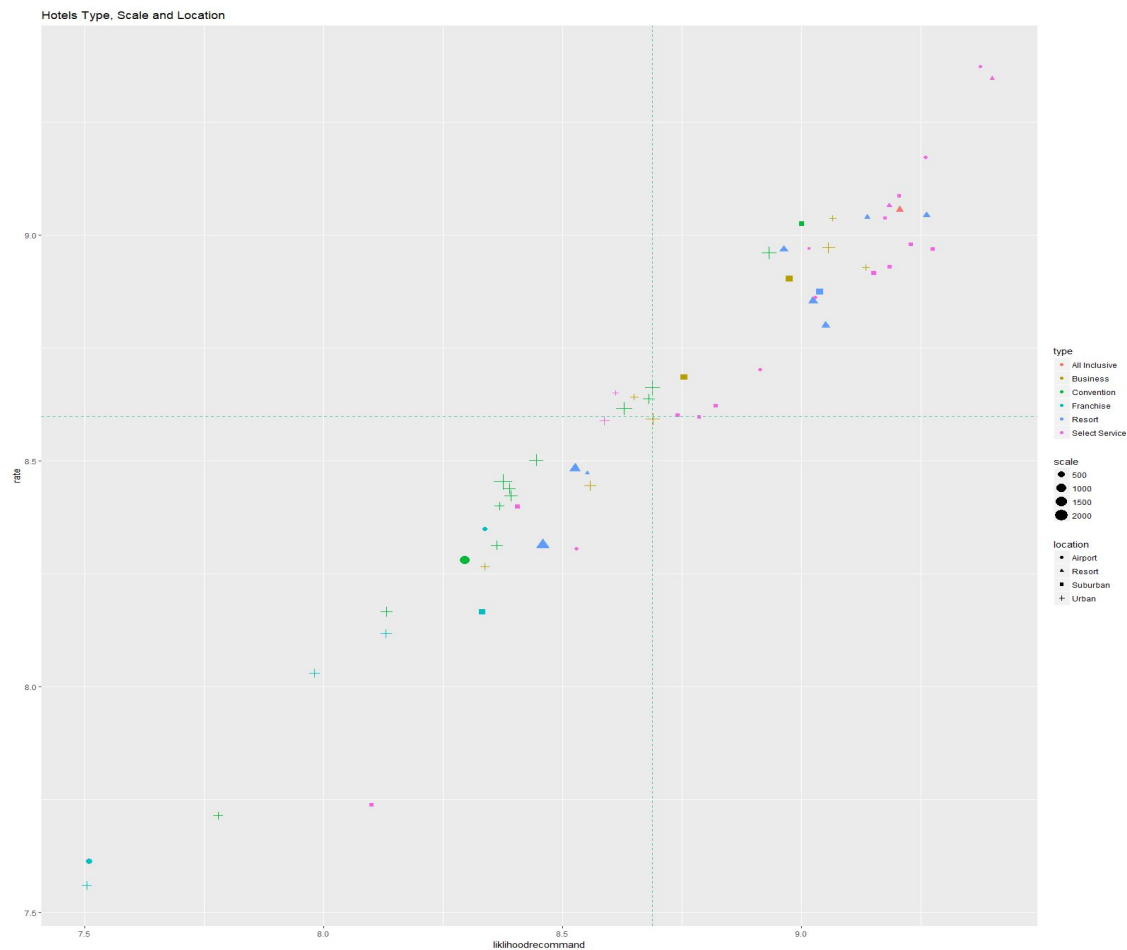


## NPS Analysis for Hyatt Hotel Group

The plot above is the ratio of single visitors and visitors who have children with them on different purpose of visit. It is clear that a lot of visitors are travelling alone. For the hotels that have majority visitors whose purpose is business or the same ratio of business and leisure visitors, there are more than 50% of visitors travelling alone. For these hotels, we think it would be helpful to promote visitors rating by focusing on providing better service for single visitors.

## VISITOR RATINGS, LIKELIHOOD TO RECOMMEND, AND THEIR RELATIONSHIP WITH HOTEL'S TYPE, SCALE AND LOCATION

In our raw data, each hotel has various labels, such as hotels' types, scale, location, and relationship with brands, floors, meeting space, classes, and other facilities. In the following plots, we chose hotel's type, scale and location to identity whether these labels has effects on visitors rating and likelihood to recommend.

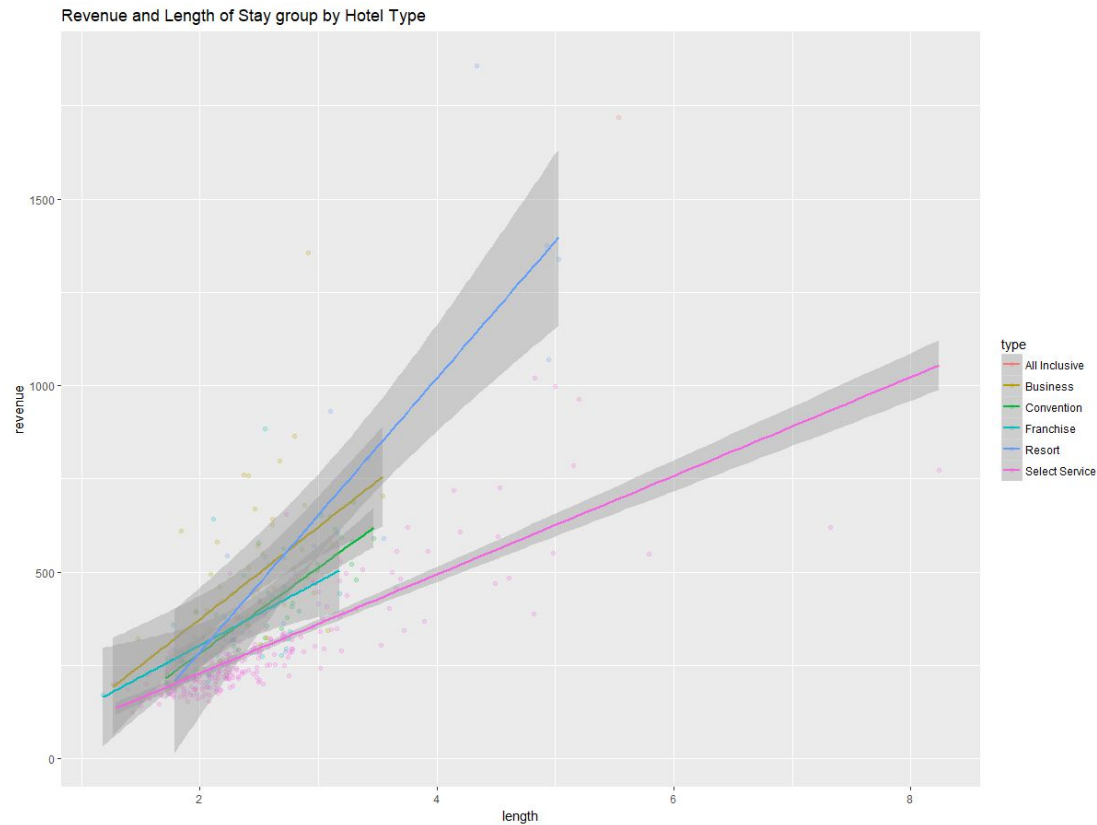


In the plot above, we found that there are a lot of large-scale convention hotels located in urban area that obviously below average rating and likelihood to recommend score. On the other hand, many small-scale hotel only provide selected services but have very high rating and

likelihood to recommend score. This is interesting that although hotels which only provide select service are small, the facilities are not as good as large

## Types of Hotel, Revenue and Length of Stay

### 1. Revenue and Length of Stay

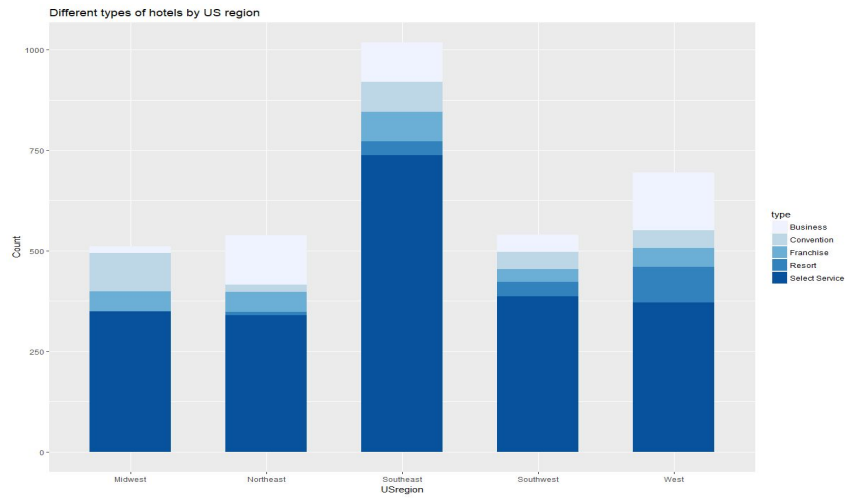


The above plot shows the relationship between hotel revenue per visits and length of stay. The plot aligns our expectation that hotels revenues increasing when visitors stay longer in the hotel. However, the plot tells more than revenue and daily rates. First, most visitors spent less than 3 days in a hotel. Second, selected service hotels earn much less revenue than other types hotels on a daily basis since the linear line of select service hotels in pink is much flatter than other lines, indicating a lower revenue.

### 2. Different types of hotels in US regions

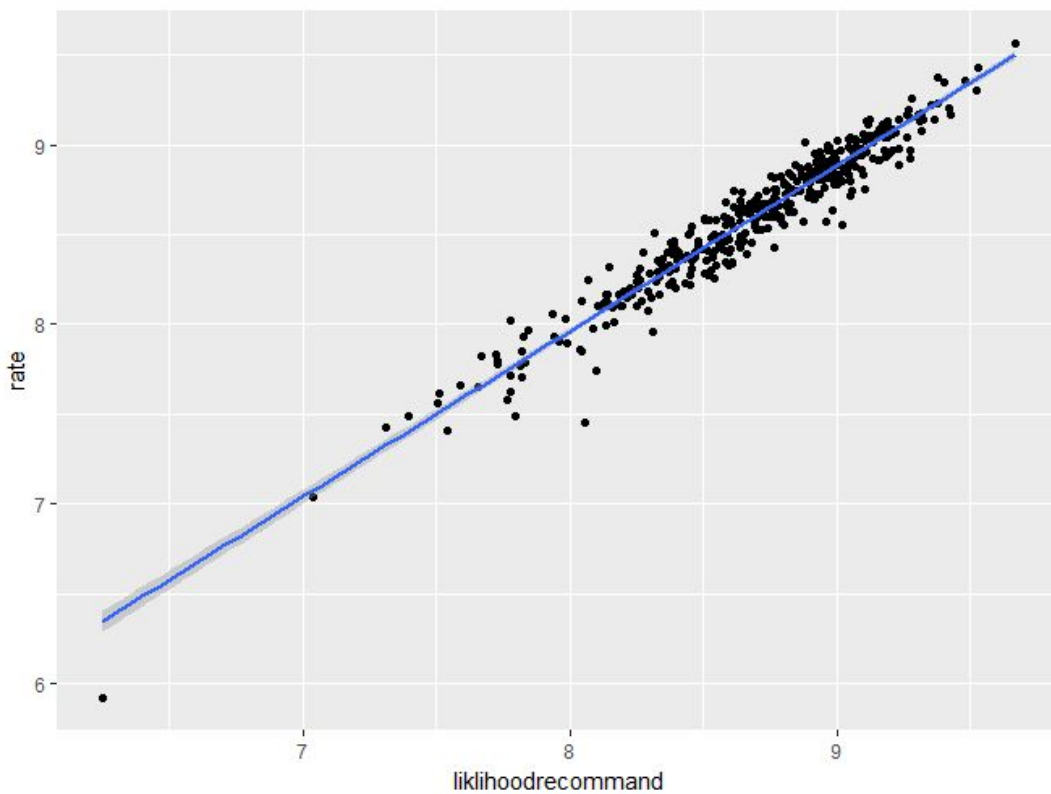


## NPS Analysis for Hyatt Hotel Group

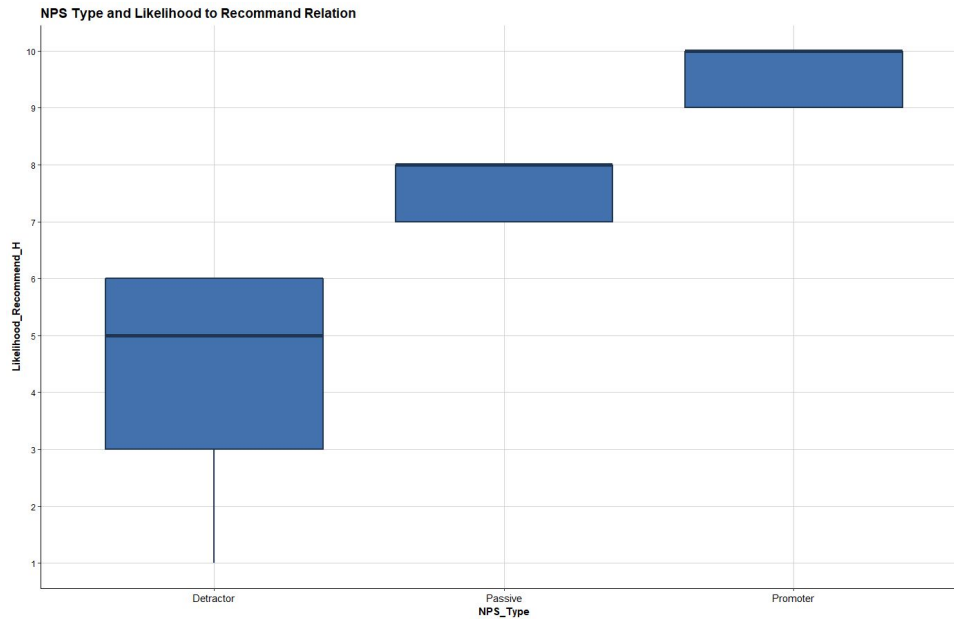


The above plot shows that hotel types are different among different US regions.

## AVERAGE VISITORS LIKELIHOOD TO RECOMMEND AND RATES TO EACH HOTEL



## NPS Analysis for Hyatt Hotel Group



These plot shows how related between visitors likelihood to recommend and their rates to the hotel and the relationship between visitors NPS\_Type and Likelihood to recommend. Likelihood to recommend is almost linear dependence to the rating. From the second plot, we found out the pattern to determine a visitor's NPS types. When visitor's likelihood to recommend is below or equal to six, they are defined as detractor. When visitor's likelihood to recommend between seven and eight, they are defined as passive. When the visitor's likelihood to recommend is higher or equal to 9, they are defined as promoter.

## LM MODEL

```
> summary(model)
```

Call:

```
lm(formula = Likelihood_Recommend_H ~ Guest_Room_H + Tranquility_H +  
    Condition_Hotel_H + Customer_SVC_H + Internet_Sat_H + Check_In_H +  
    F.B_Overall_Experience_H, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.9387	-0.1081	0.0332	0.4248	5.3508

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.120479	0.032436	-65.375	< 2e-16	***
Guest_Room_H	0.299751	0.004115	72.845	< 2e-16	***
Tranquility_H	0.116100	0.002865	40.525	< 2e-16	***
Condition_Hotel_H	0.192779	0.004579	42.098	< 2e-16	***
Customer_SVC_H	0.451690	0.004105	110.042	< 2e-16	***
Internet_Sat_H	0.027413	0.001914	14.325	< 2e-16	***
Check_In_H	0.018499	0.003508	5.273	1.35e-07	***
F.B_Overall_Experience_H	0.107982	0.002612	41.337	< 2e-16	***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.004 on 75339 degrees of freedom  
(190908 observations deleted due to missingness)

Multiple R-squared: 0.6525, Adjusted R-squared: 0.6525  
F-statistic: 2.021e+04 on 7 and 75339 DF, p-value: < 2.2e-16

```
> summary(model2)
```

```
Call:
lm(formula = Likelihood_Recommend_H ~ Guest_Room_H + Customer_SVC_H,
    data = df)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-8.8622 -0.2663  0.1378  0.2598  9.6868
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.747813   0.013571  -55.1   <2e-16 ***
Guest_Room_H   0.534892   0.001500   356.5   <2e-16 ***
Customer_SVC_H  0.526113   0.001709   307.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.112 on 259276 degrees of freedom
(6976 observations deleted due to missingness)
Multiple R-squared:  0.6656,    Adjusted R-squared:  0.6656
F-statistic: 2.581e+05 on 2 and 259276 DF,  p-value: < 2.2e-16
```

Since visitors NPS type is totally dependent on visitors likelihood to recommend, we decided to find out which variables are most related to likelihood to recommend so that we can find out which variables are related to NPS type. We chose guest room quality, tranquility, hotel condition, staff care, internet service, check in quality, and food & beverage quality to make linear model with likelihood to recommend. In the model summary, we find out guest room and staff care are two variables that most related to likelihood to recommend (lowest two t value). Then we make another linear model on likelihood to recommend towards guest room quality and customer service. The model2 r-squared number increased from .6525 to .6656 which mean these guest room quality and customer service are really more dependent to likelihood to recommend.

## ASSOCIATION RULES

Business is a larger portion of the reservations, so we want to look at what are associate with a customer whose pov is business to be a promoter. In order to explore what factors constitute the emergence of promoters, we continued using our subsets of business travelers and leisure travelers. Establishing association rules is the most straightforward way to present the relationships between promoter and hotel services, amenities, and other facilities. They also reveal conditions in which promoters are most likely to occur. Since the ultimate goal of our analysis to improve customers' overall experience of staying at Hyatt hotels, it is imperative to divide our dataset based on POV and to explore each season and comprehensively all seasons. We hope that by establishing such association rules that reveal the relationship between hotel services and promoters, different hotels would develop their own customized business plan to facilitate their customers.

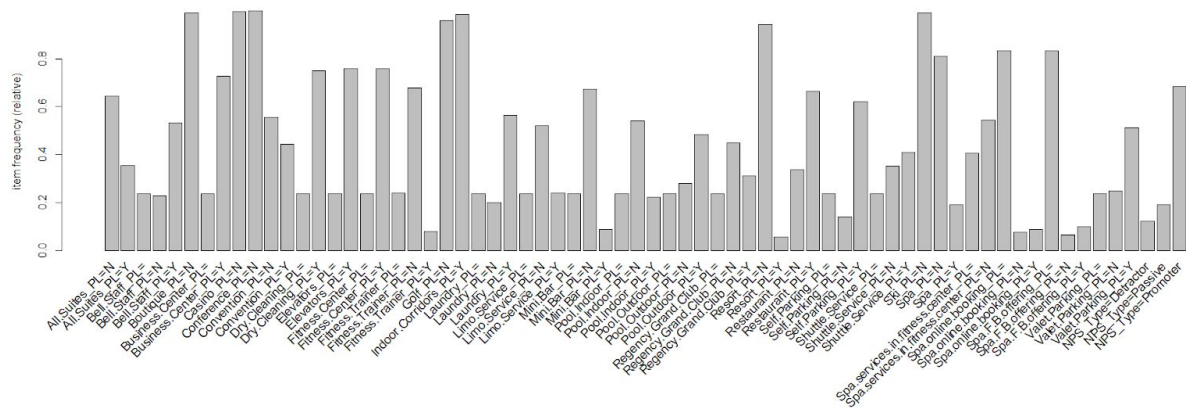
For association rules, we used data only contains POV type of business in four previously selected months of U.S. Hyatt Hotels and tested the association of NPS\_Type as a promoter to 25 services or amenities of the hotel. By using the `table()` in R, we discovered that the number of business visitors was a lot higher than that of leisure visitors.

Based on P-value calculated by chi-square model, we selected variables that are considered relevant to NPS\_Type into association rules analysis. The variables are: Bell.Staff\_PL, Boutique\_PL, Business.Center\_PL, Casino\_PL, Conference\_PL, Convention\_PL, Dry.Cleaning\_PL, Elevators\_PL, Fitness.Center\_PL,

## NPS Analysis for Hyatt Hotel Group

Fitness.Trainer\_PL, Golf\_PL, Indoor.Corridors\_PL, Laundry\_PL, Limo.Service\_PL, Mini.Bar\_PL, Pool.Indoor\_PL, Pool.Outdoor\_PL, Regency.Grand.Club\_PL, Resort\_PL, Restaurant\_PL, Self.Parking\_PL, Shuttle.Service\_PL, Ski\_PL, Spa\_PL, Valet.Parking\_PL.

First we looked at frequency of each value of the variables to have a overlook of the whole data we are analysing. And by looking at 10 rules with highest support



Frequency Table for February

2014 02 business

TOP 10 Support 2014 02 business

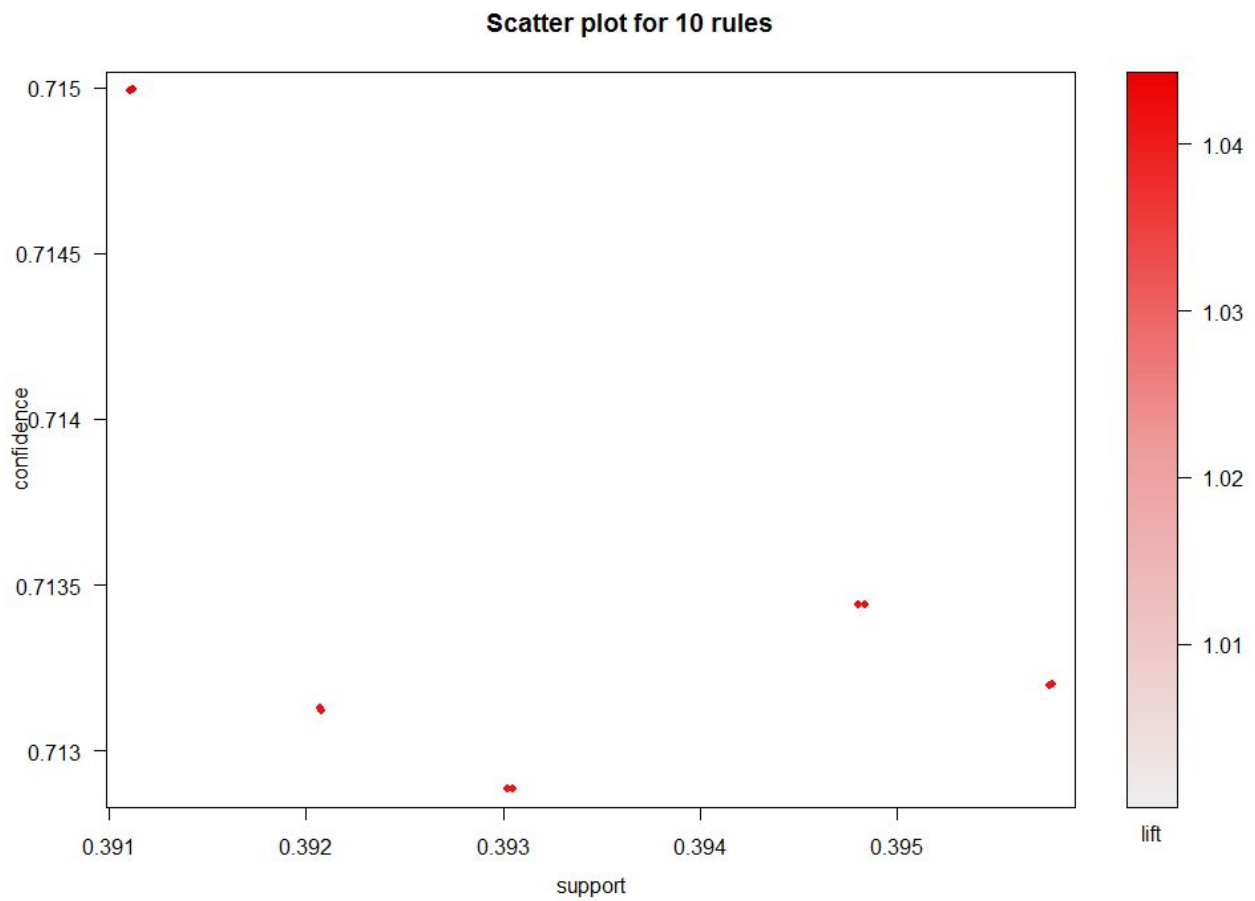
lhs	rhs	support	confidence	lift
[1] {Convention_PL=N}	=> {NPS_Type=Promoter}	0.3957864	0.7131966	1.041446
[2] {Casino_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.3957864	0.7131966	1.041446
[3] {Conference_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.3948199	0.7134430	1.041805
[4] {Casino_PL=N, Conference_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.3948199	0.7134430	1.041805
[5] {Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.3930374	0.7128778	1.040980
[6] {Casino_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.3930374	0.7128778	1.040980
[7] {Conference_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.3920710	0.7131250	1.041341

## NPS Analysis for Hyatt Hotel Group

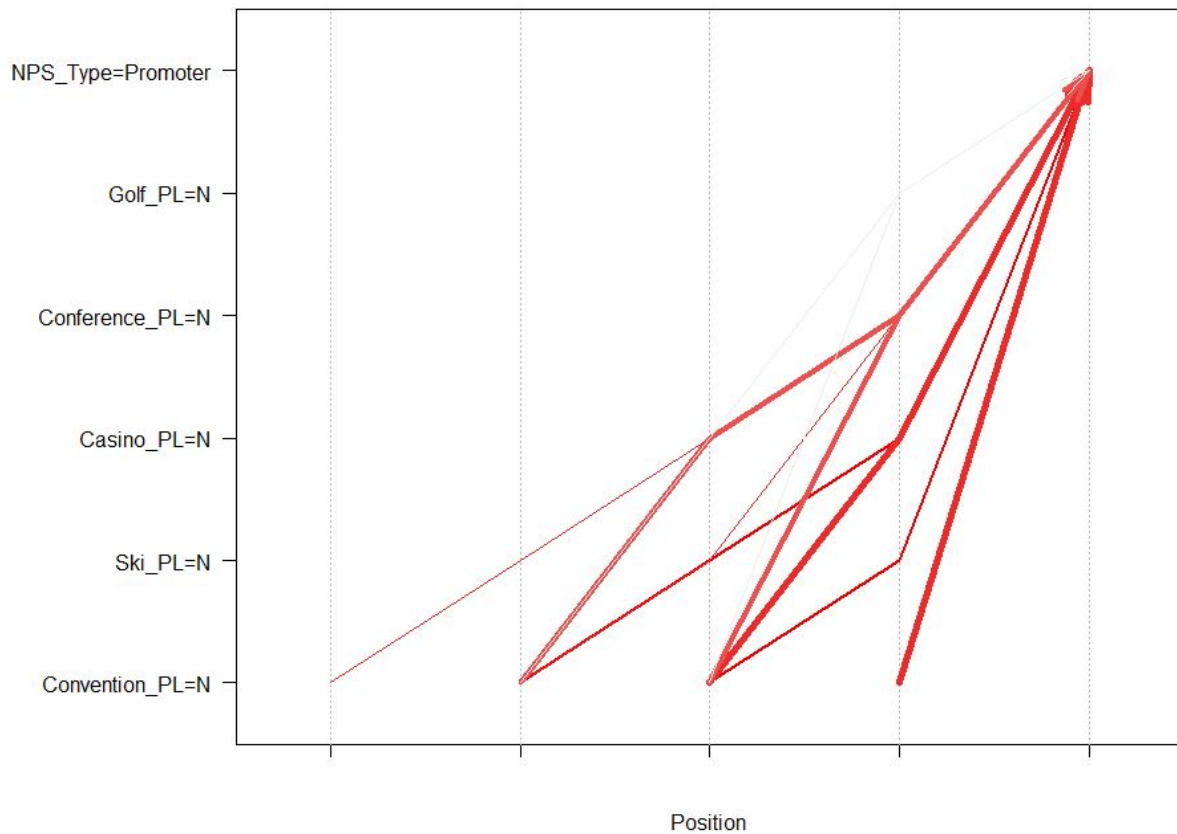
```

[8] {Casino_PL=N,
     Conference_PL=N,
     Convention_PL=N,
     Ski_PL=N}      => {NPS_Type=Promoter} 0.3920710 0.7131250 1.041341
[9] {Convention_PL=N,
     Golf_PL=N}     => {NPS_Type=Promoter} 0.3911045 0.7149980 1.044076
[10] {Casino_PL=N,
      Convention_PL=N,
      Golf_PL=N}    => {NPS_Type=Promoter} 0.3911045 0.7149980 1.044076

```



Parallel coordinates plot for 10 rules



Top10 confidence 14 02 business

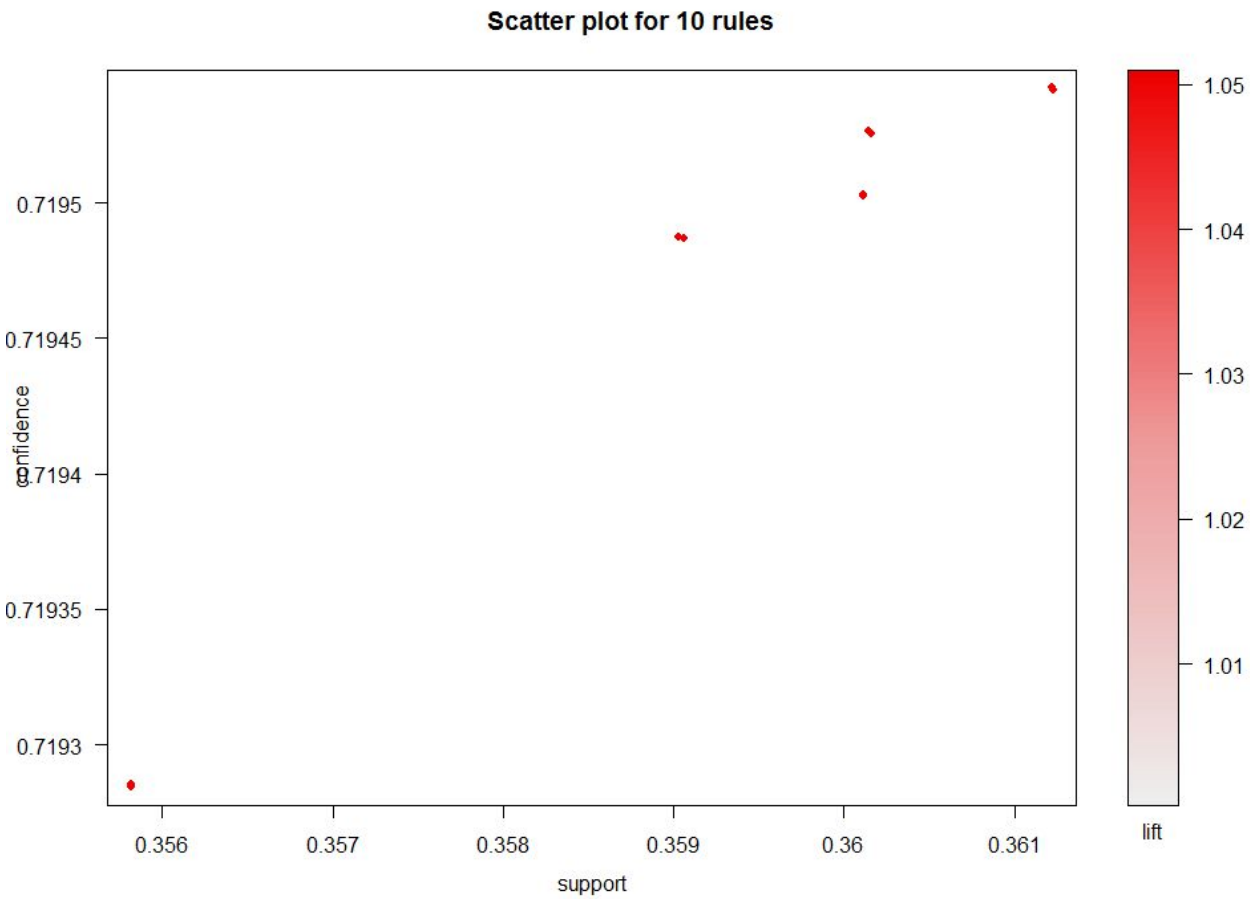
	lhs	rhs	support	confidence	lift
[1]	{Conference_PL=N, Convention_PL=N, Golf_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3612310	0.7195414	1.050711
[2]	{Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3612310	0.7195414	1.050711
[3]	{Conference_PL=N, Convention_PL=N, Golf_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3601572	0.7195263	1.050688
[4]	{Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3601572	0.7195263	1.050688
[5]	{Conference_PL=N, Convention_PL=N,				

## NPS Analysis for Hyatt Hotel Group

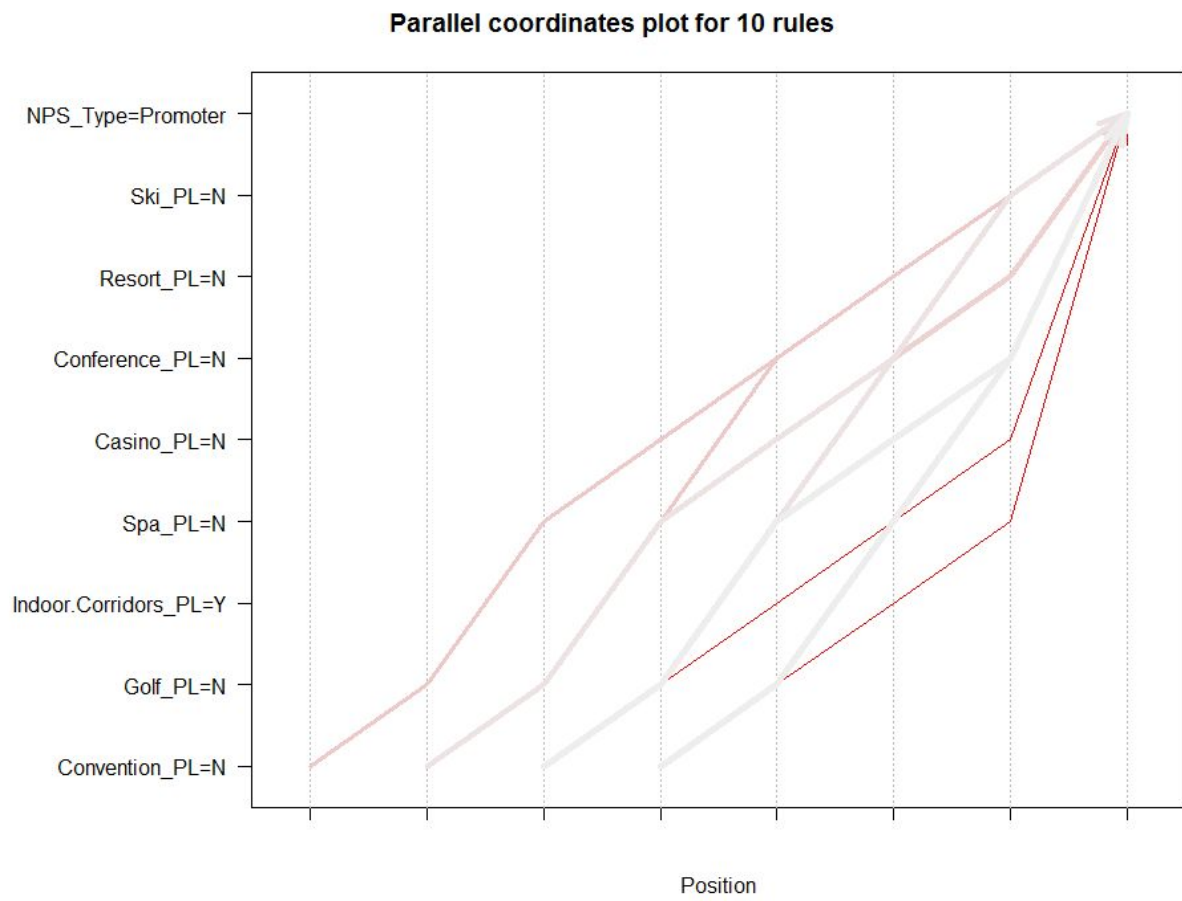
```

      Golf_PL=N,
      Resort_PL=N,
      Spa_PL=N}          => {NPS_Type=Promoter} 0.3601143  0.7195023  1.050653
[6] {Casino_PL=N,
      Conference_PL=N,
      Convention_PL=N,
      Golf_PL=N,
      Resort_PL=N,
      Spa_PL=N}          => {NPS_Type=Promoter} 0.3601143  0.7195023  1.050653
[7] {Conference_PL=N,
      Convention_PL=N,
      Golf_PL=N,
      Resort_PL=N,
      Ski_PL=N,
      Spa_PL=N}          => {NPS_Type=Promoter} 0.3590404  0.7194870  1.050631
[8] {Casino_PL=N,
      Conference_PL=N,
      Convention_PL=N,
      Golf_PL=N,
      Resort_PL=N,
      Ski_PL=N,
      Spa_PL=N}          => {NPS_Type=Promoter} 0.3590404  0.7194870  1.050631
[9] {Convention_PL=N,
      Golf_PL=N,
      Indoor.Corridors_PL=Y,
      Spa_PL=N}          => {NPS_Type=Promoter} 0.3558190  0.7192845  1.050335
[10] {Casino_PL=N,
      Convention_PL=N,
      Golf_PL=N,
      Indoor.Corridors_PL=Y,
      Spa_PL=N}          => {NPS_Type=Promoter} 0.3558190  0.7192845  1.050335

```

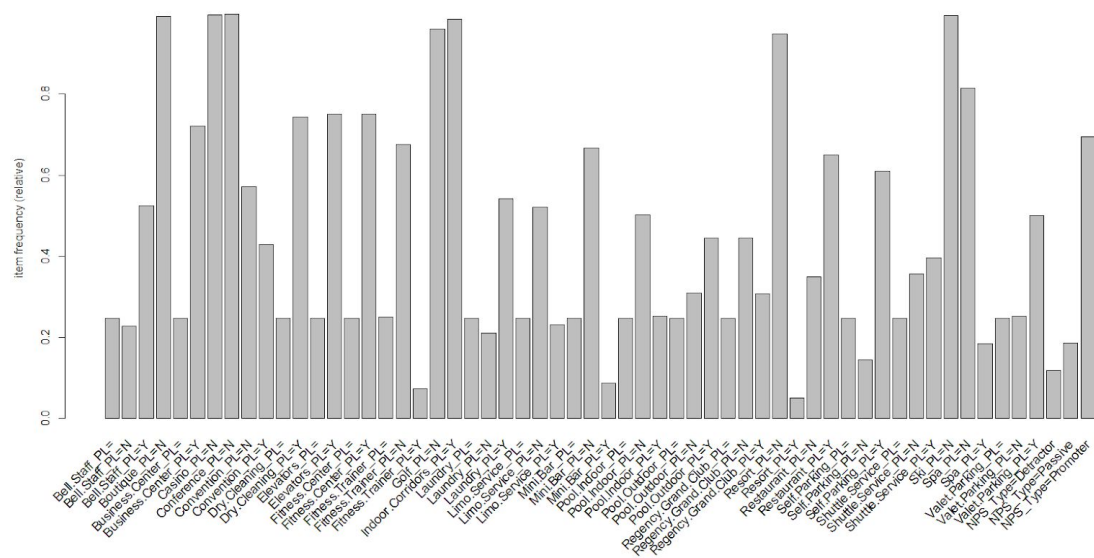






201405 Business

## NPS Analysis for Hyatt Hotel Group



## TOP 10 Support 2014 05 business

lhs	rhs	support	confidence	lift
[1] {Convention_PL=N} => {NPS_Type=Promoter}		0.4101417	0.7188397	1.034107
[2] {Casino_PL=N,	Convention_PL=N} => {NPS_Type=Promoter}	0.4101417	0.7188397	1.034107
[3] {Convention_PL=N,	Ski_PL=N} => {NPS_Type=Promoter}	0.4087594	0.7186391	1.033818
[4] {Casino_PL=N,	Convention_PL=N,			
	Ski_PL=N} => {NPS_Type=Promoter}	0.4087594	0.7186391	1.033818
[5] {Conference_PL=N,	Convention_PL=N} => {NPS_Type=Promoter}	0.4079227	0.7189844	1.034315
[6] {Casino_PL=N,	Conference_PL=N,			
	Convention_PL=N} => {NPS_Type=Promoter}	0.4079227	0.7189844	1.034315
[7] {Convention_PL=N,	Golf_PL=N} => {NPS_Type=Promoter}	0.4070497	0.7205641	1.036587
[8] {Casino_PL=N,	Convention_PL=N,			
	Golf_PL=N} => {NPS_Type=Promoter}	0.4070497	0.7205641	1.036587
[9] {Conference_PL=N,	Convention_PL=N,			

## NPS Analysis for Hyatt Hotel Group

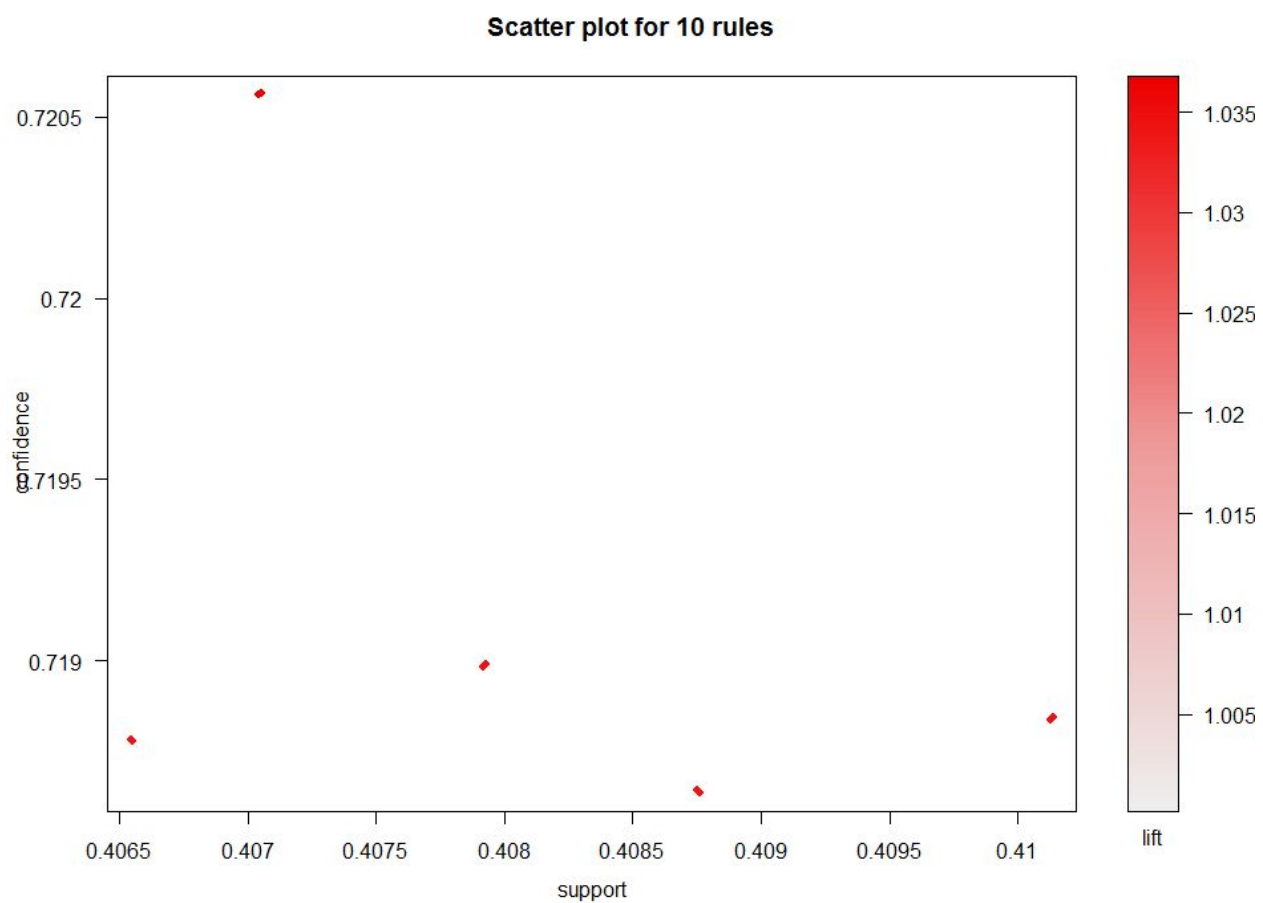
Ski\_PL=N} => {NPS\_Type=Promoter} 0.4065404 0.7187832 1.034025

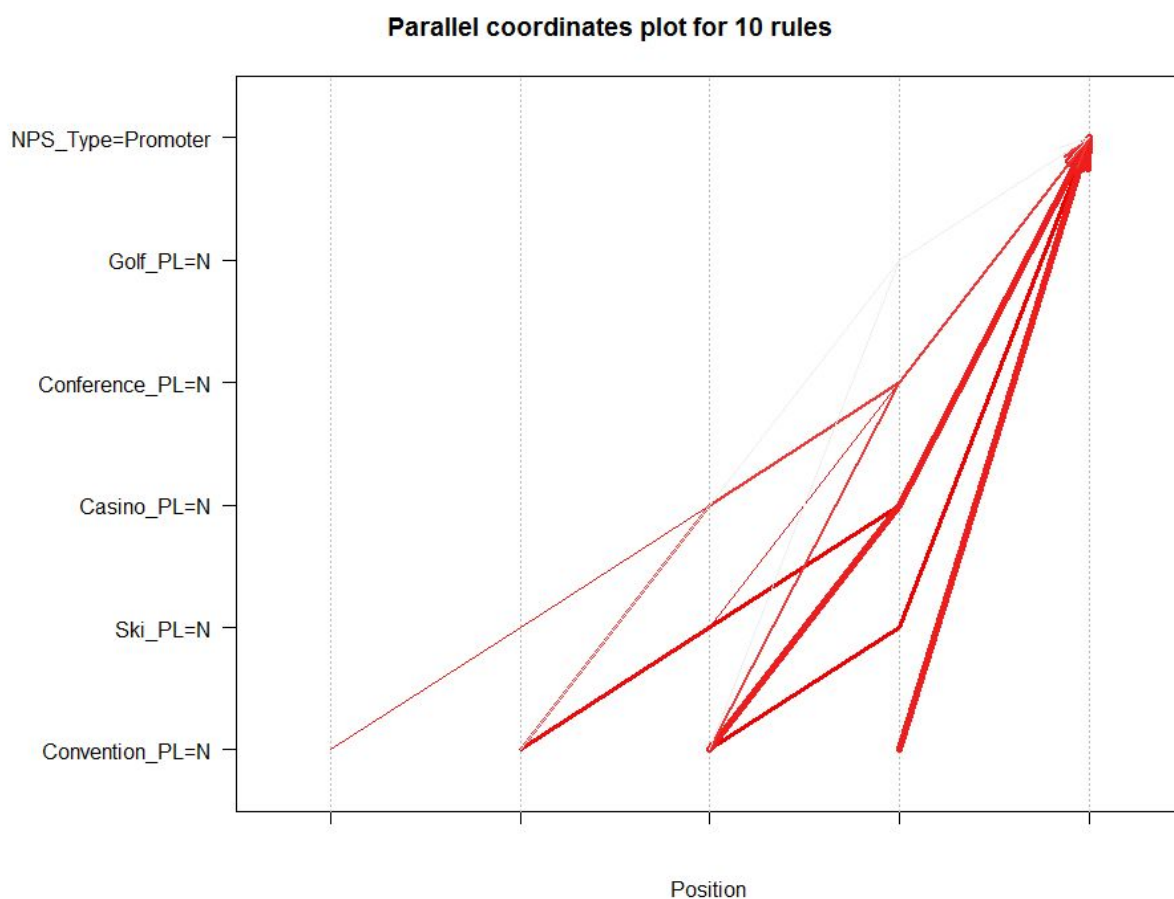
[10] {Casino\_PL=N,

Conference\_PL=N,

Convention\_PL=N,

Ski\_PL=N} => {NPS\_Type=Promoter} 0.4065404 0.7187832 1.034025





Top 10 confidence for 14 05 business

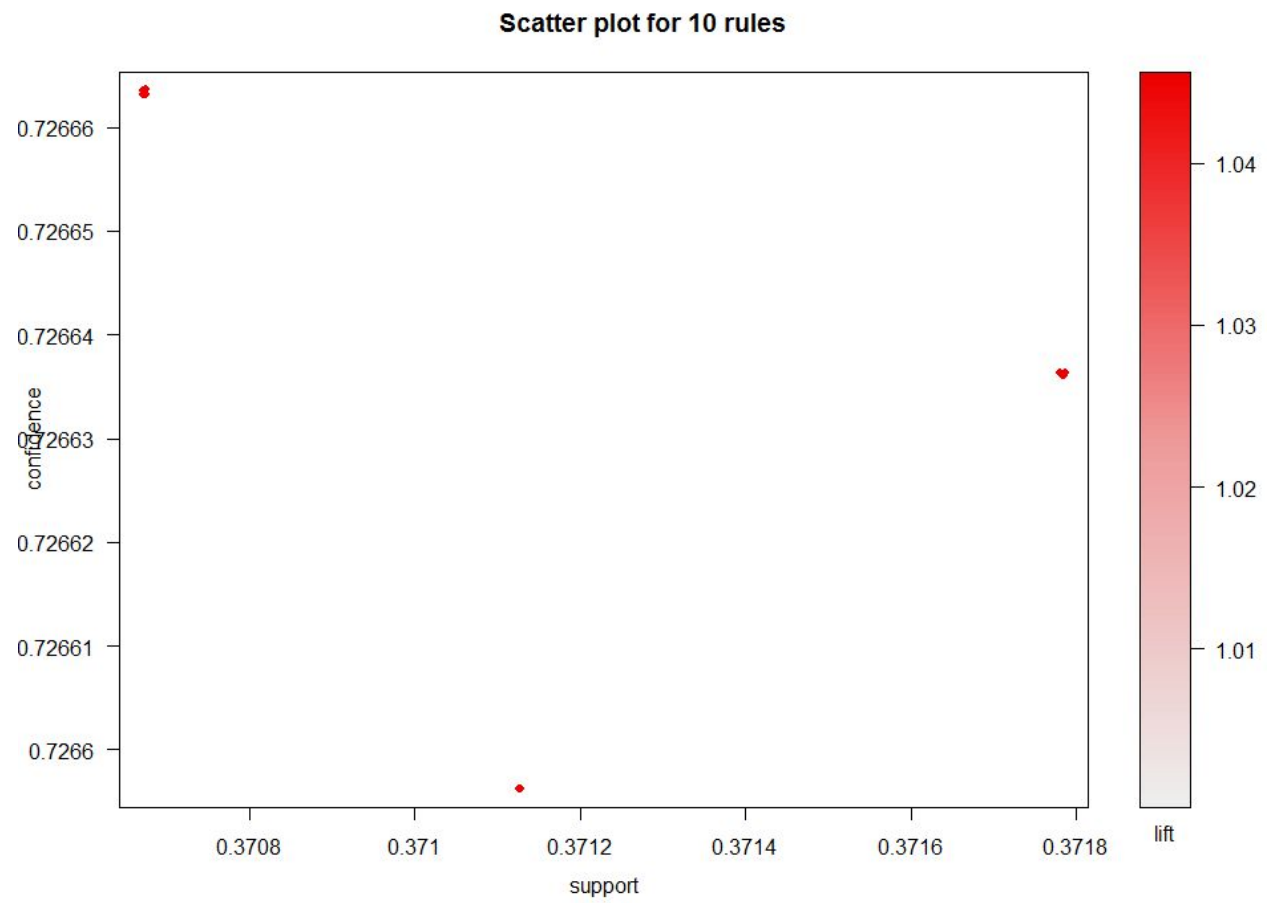
lhs	rhs	support	confidence	lift
[1] {Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3706735	0.7266633	1.045362
[2] {Casino_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3706735	0.7266633	1.045362
[3] {Conference_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3706735	0.7266633	1.045362
[4] {Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N,				

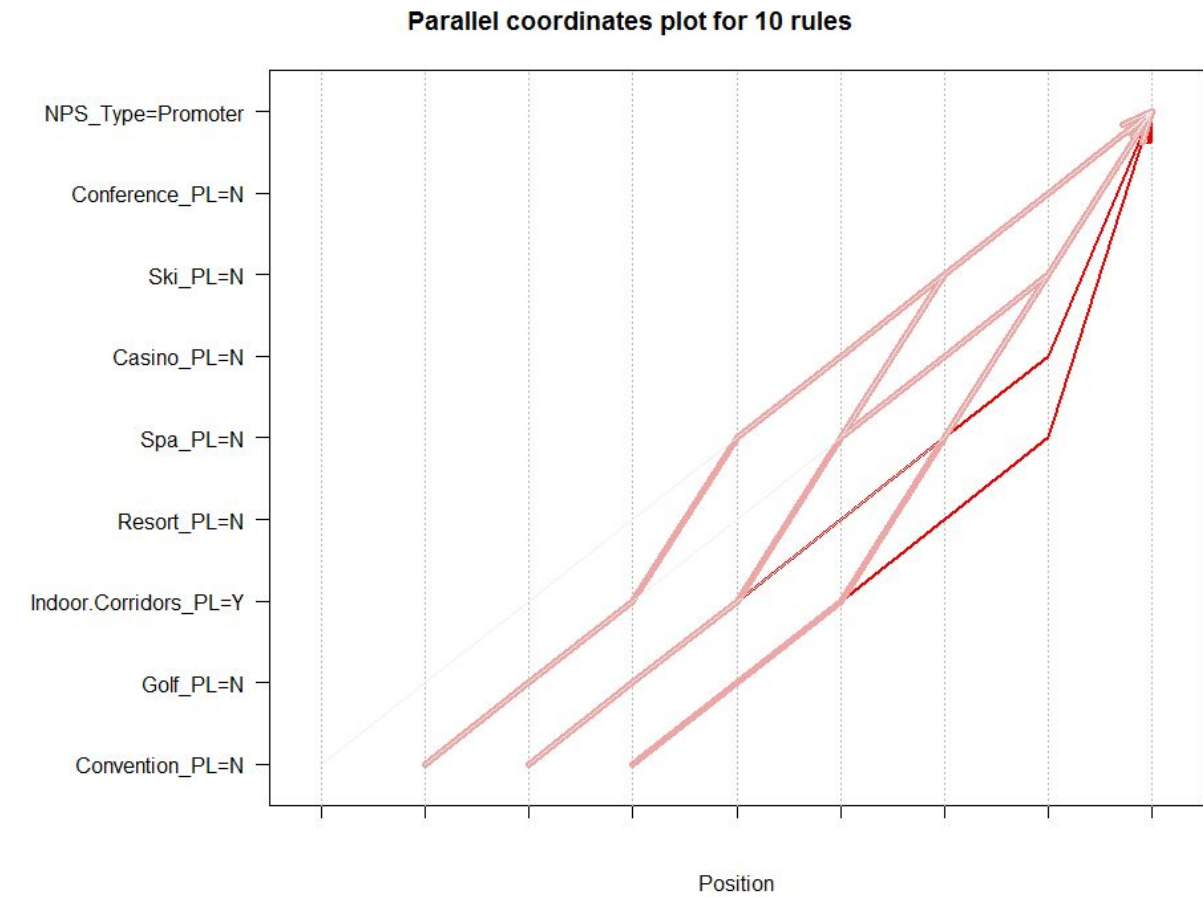
## NPS Analysis for Hyatt Hotel Group

```

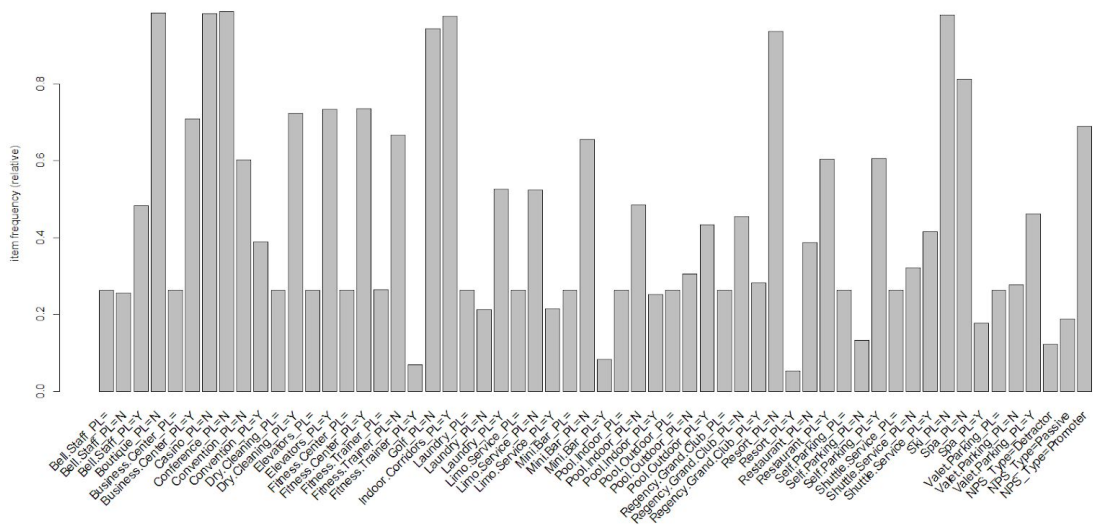
Indoor.Corridors_PL=Y,
Resort_PL=N,
Ski_PL=N,
Spa_PL=N}          => {NPS_Type=Promoter} 0.3706735 0.7266633 1.045362
[5] {Convention_PL=N,
     Golf_PL=N,
     Indoor.Corridors_PL=Y,
     Ski_PL=N,
     Spa_PL=N}      => {NPS_Type=Promoter} 0.3717830 0.7266361 1.045322
[6] {Casino_PL=N,
     Convention_PL=N,
     Golf_PL=N,
     Indoor.Corridors_PL=Y,
     Ski_PL=N,
     Spa_PL=N}      => {NPS_Type=Promoter} 0.3717830 0.7266361 1.045322
[7] {Conference_PL=N,
     Convention_PL=N,
     Golf_PL=N,
     Indoor.Corridors_PL=Y,
     Ski_PL=N,
     Spa_PL=N}      => {NPS_Type=Promoter} 0.3717830 0.7266361 1.045322
[8] {Casino_PL=N,
     Conference_PL=N,
     Convention_PL=N,
     Golf_PL=N,
     Indoor.Corridors_PL=Y,
     Ski_PL=N,
     Spa_PL=N}      => {NPS_Type=Promoter} 0.3717830 0.7266361 1.045322
[9] {Convention_PL=N,
     Golf_PL=N,
     Indoor.Corridors_PL=Y,
     Resort_PL=N,
     Spa_PL=N}      => {NPS_Type=Promoter} 0.3711282 0.7265962 1.045265
[10] {Casino_PL=N,
      Convention_PL=N,
      Golf_PL=N,
      Indoor.Corridors_PL=Y,
      Resort_PL=N,
      Spa_PL=N}      => {NPS_Type=Promoter} 0.3711282 0.7265962 1.045265

```





August 2014 business

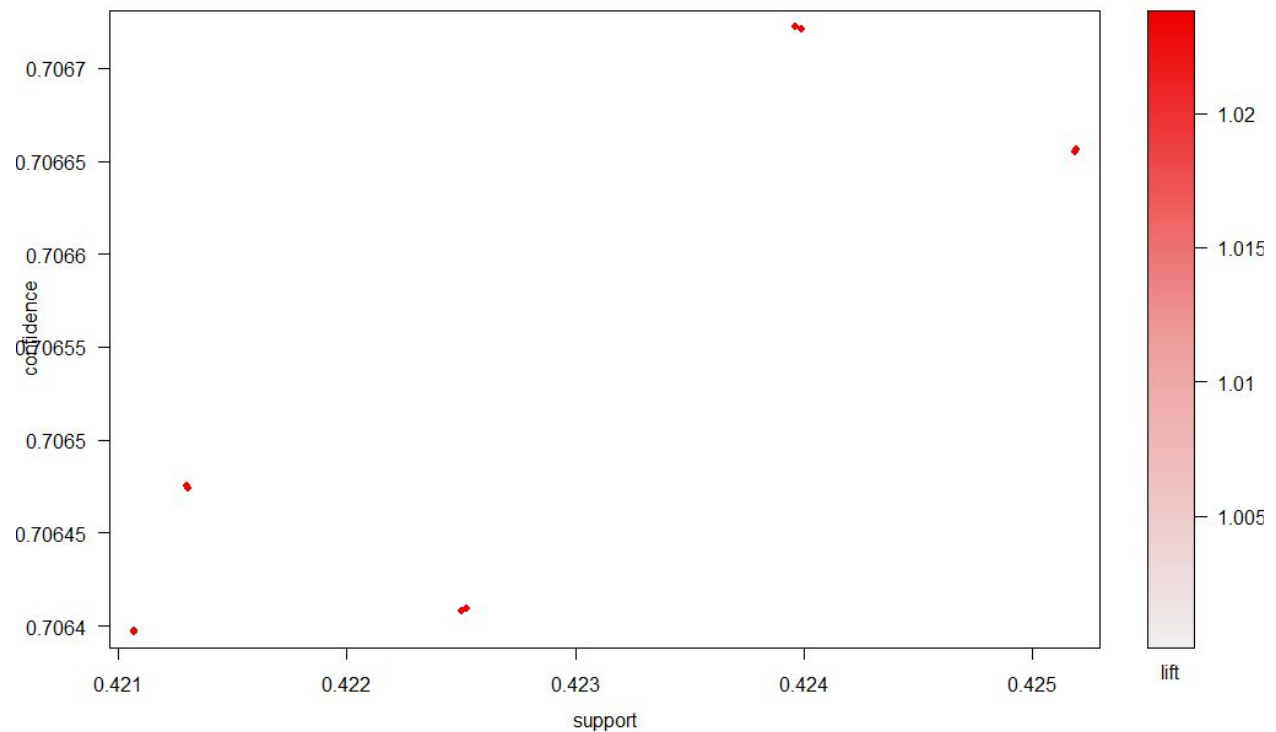


Frequency Table for August

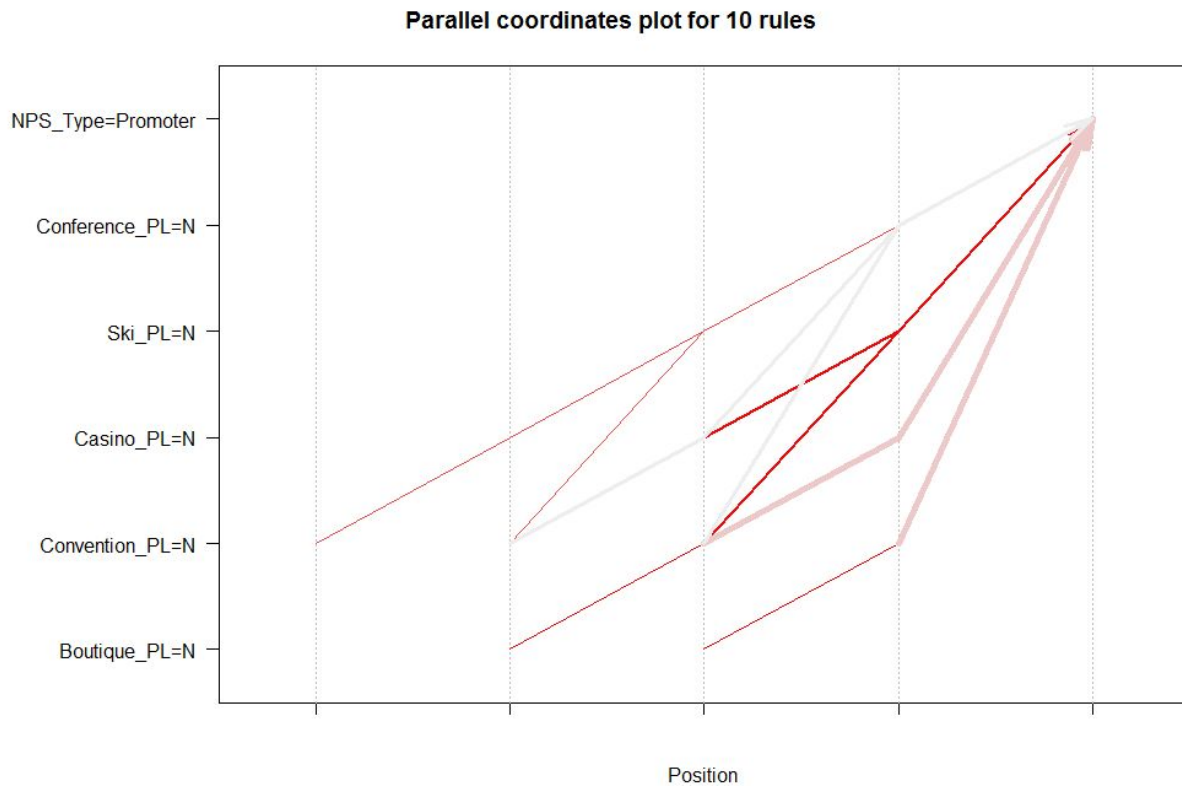
## NPS Analysis for Hyatt Hotel Group

	lhs	rhs	support	confidence	lift
[1]	{Convention_PL=N}	=> {NPS_Type=Promoter}	0.4251875	0.7066553	1.023590
[2]	{Casino_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.4251875	0.7066553	1.023590
[3]	{Conference_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.4239736	0.7067217	1.023686
[4]	{Casino_PL=N, Conference_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.4239736	0.7067217	1.023686
[5]	{Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4225107	0.7064089	1.023233
[6]	{Casino_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4225107	0.7064089	1.023233
[7]	{Conference_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4212967	0.7064749	1.023329
[8]	{Casino_PL=N, Conference_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4212967	0.7064749	1.023329
[9]	{Boutique_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.4210633	0.7063969	1.023216
[10]	{Boutique_PL=N, Casino_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.4210633	0.7063969	1.023216

Scatter plot for 10 rules







Top 10 confidence for 14 08 business

lhs	rhs	support	confidence	lift
[1] {Conference_PL=N, Convention_PL=N, Golf_PL=N, Resort_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3920845	0.7124717	1.032015
[2] {Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Resort_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3920845	0.7124717	1.032015
[3] {Conference_PL=N, Convention_PL=N, Golf_PL=N, Resort_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3930339	0.7124037	1.031917
[4] {Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Resort_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3930339	0.7124037	1.031917
[5] {Convention_PL=N, Golf_PL=N, Resort_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3932985	0.7123809	1.031884
[6] {Casino_PL=N,				

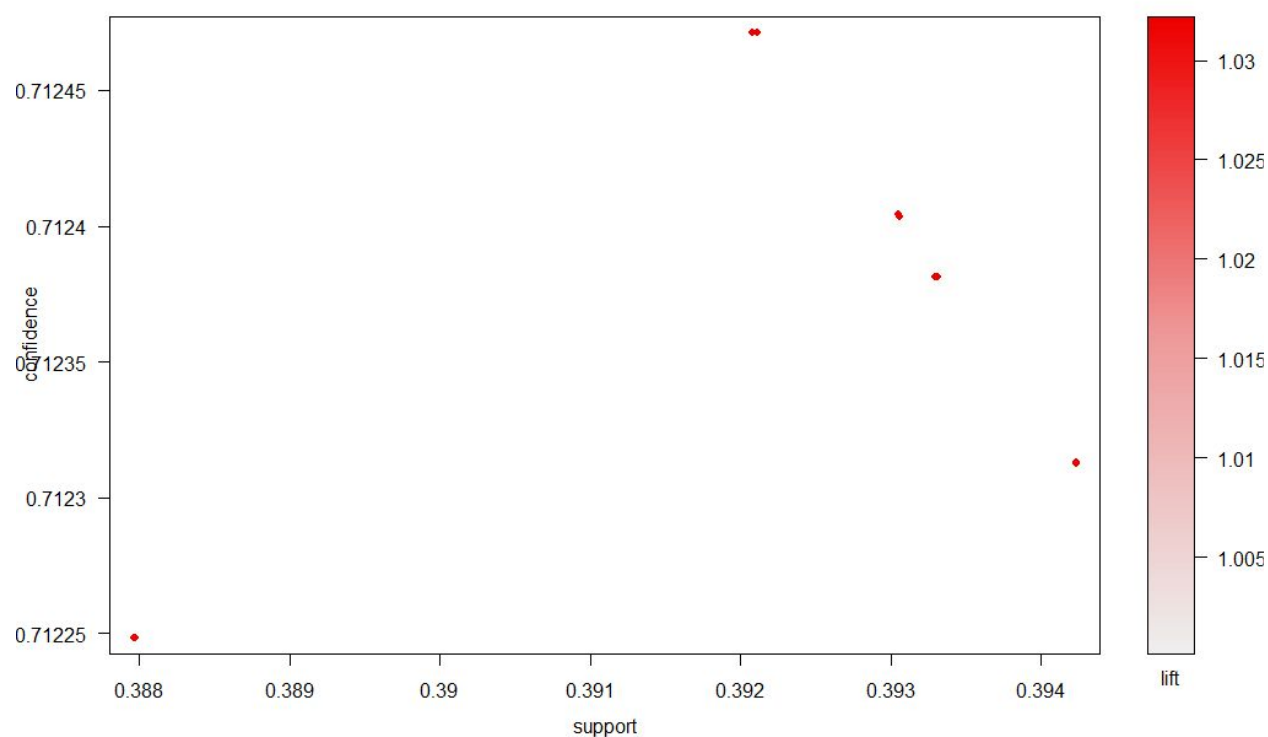
## NPS Analysis for Hyatt Hotel Group

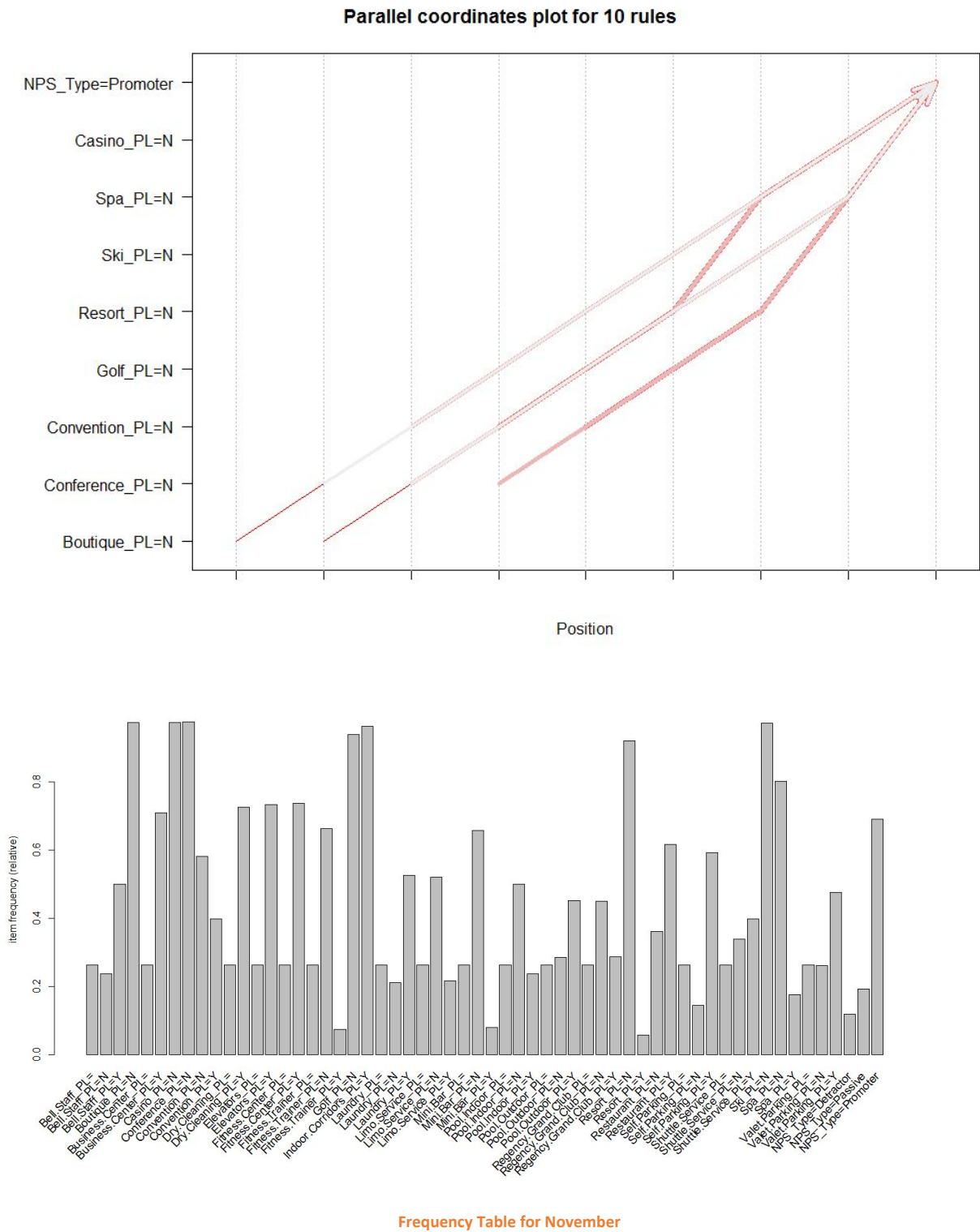
```

Convention_PL=N,
Golf_PL=N,
Resort_PL=N,
Ski_PL=N,
Spa_PL=N}    => {NPS_Type=Promoter} 0.3932985 0.7123809 1.031884
[7] {Convention_PL=N,
Golf_PL=N,
Resort_PL=N,
Spa_PL=N}    => {NPS_Type=Promoter} 0.3942478 0.7123134 1.031786
[8] {Casino_PL=N,
Convention_PL=N,
Golf_PL=N,
Resort_PL=N,
Spa_PL=N}    => {NPS_Type=Promoter} 0.3942478 0.7123134 1.031786
[9] {Boutique_PL=N,
Conference_PL=N,
Convention_PL=N,
Golf_PL=N,
Resort_PL=N,
Ski_PL=N,
Spa_PL=N}    => {NPS_Type=Promoter} 0.3879603 0.7122489 1.031692
[10] {Boutique_PL=N,
Casino_PL=N,
Conference_PL=N,
Convention_PL=N,
Golf_PL=N,
Resort_PL=N,
Ski_PL=N,
Spa_PL=N}    => {NPS_Type=Promoter} 0.3879603 0.7122489 1.031692

```

Scatter plot for 10 rules





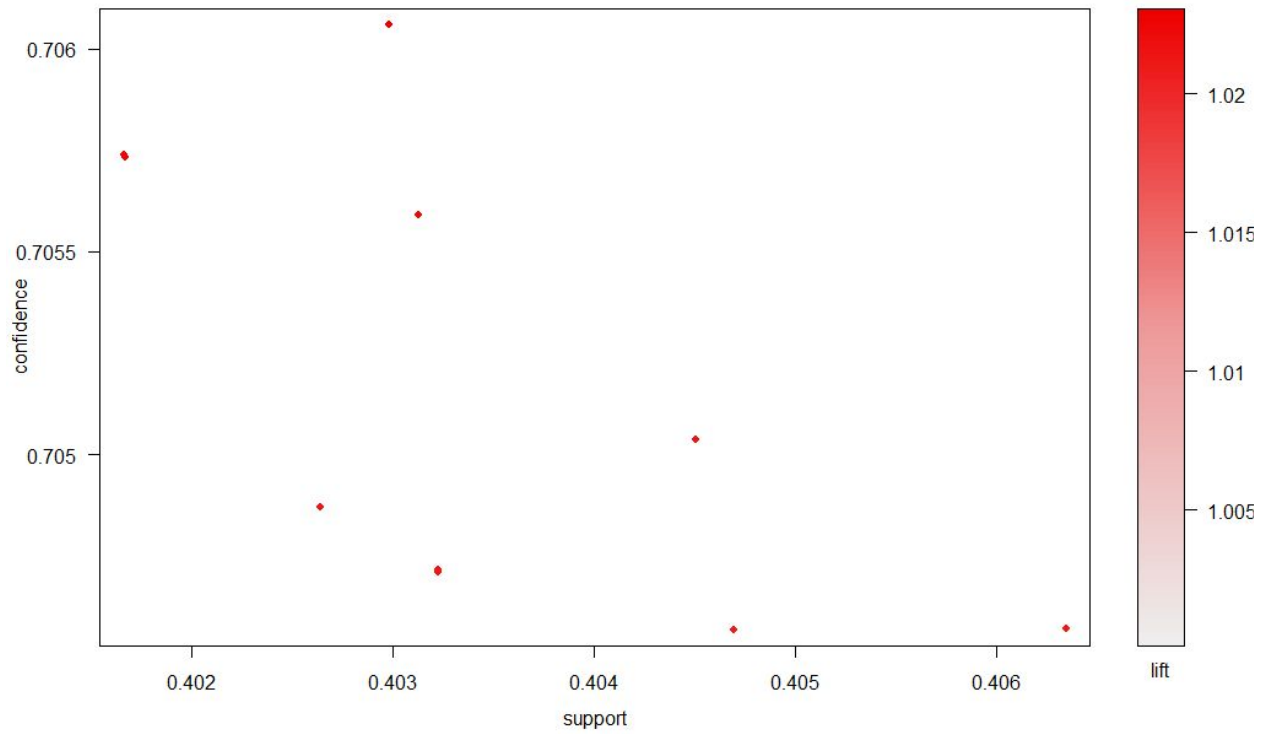
Top 10 support for 14 11 business

lhs	rhs	support	confidence	lift
[1] {Casino_PL=N,				

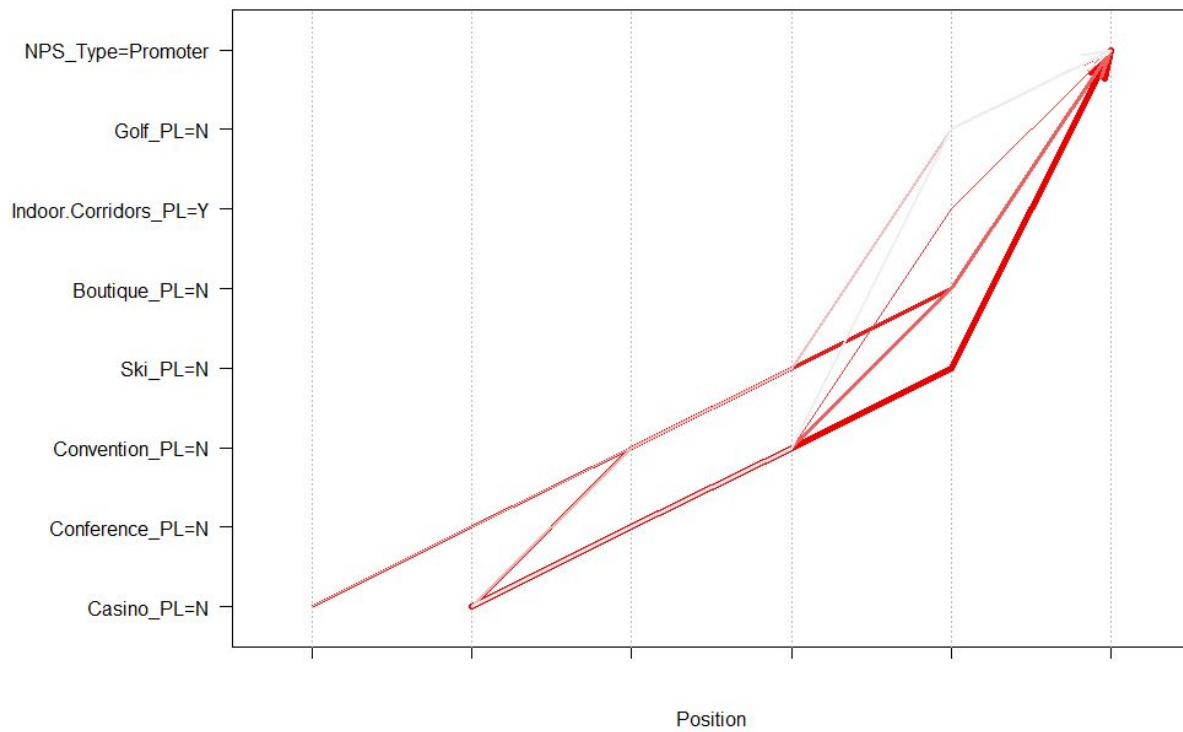
## NPS Analysis for Hyatt Hotel Group

	Conference_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4063307	0.7045668	1.020784
[2]	{Boutique_PL=N, Casino_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4046835	0.7045692	1.020787
[3]	{Boutique_PL=N, Casino_PL=N, Conference_PL=N, Convention_PL=N}	=> {NPS_Type=Promoter}	0.4045150	0.7050341	1.021461
[4]	{Boutique_PL=N, Conference_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4032234	0.7047142	1.020998
[5]	{Boutique_PL=N, Casino_PL=N, Conference_PL=N, Convention_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4032234	0.7047142	1.020998
[6]	{Casino_PL=N, Convention_PL=N, Golf_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4031298	0.7055894	1.022266
[7]	{Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N}	=> {NPS_Type=Promoter}	0.4029613	0.7060579	1.022944
[8]	{Casino_PL=N, Conference_PL=N, Convention_PL=N, Indoor.Corridors_PL=Y}	=> {NPS_Type=Promoter}	0.4026244	0.7048665	1.021218
[9]	{Conference_PL=N, Convention_PL=N, Golf_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4016697	0.7057392	1.022483
[10]	{Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Ski_PL=N}	=> {NPS_Type=Promoter}	0.4016697	0.7057392	1.022483

Scatter plot for 10 rules



Parallel coordinates plot for 10 rules



## NPS Analysis for Hyatt Hotel Group

Top 10 confidence for 14 11 business

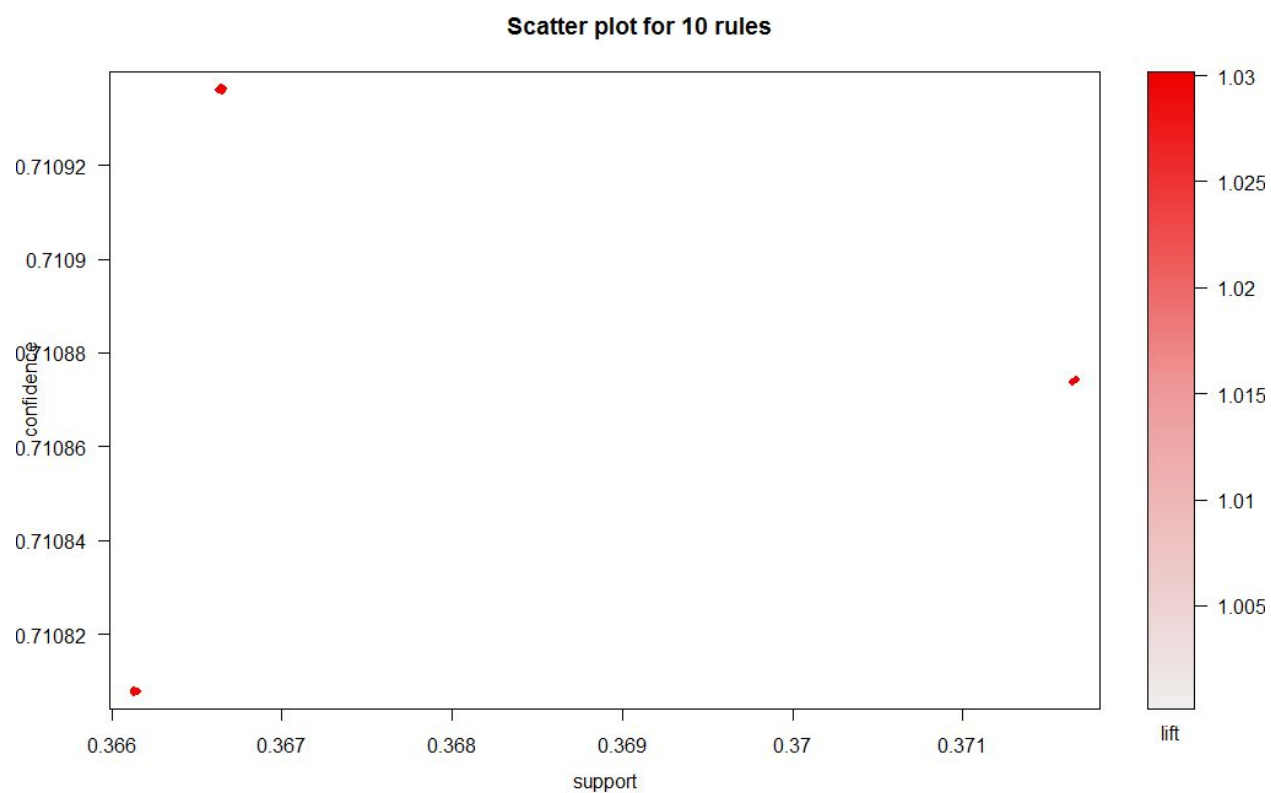
	lhs	rhs	support	confidence	lift
[1]	{Boutique_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3666467	0.7109361	1.030012
[2]	{Boutique_PL=N, Casino_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3666467	0.7109361	1.030012
[3]	{Boutique_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3666467	0.7109361	1.030012
[4]	{Boutique_PL=N, Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3666467	0.7109361	1.030012
[5]	{Boutique_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Resort_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3716446	0.7108740	1.029922
[6]	{Boutique_PL=N, Casino_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Resort_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3716446	0.7108740	1.029922
[7]	{Boutique_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3661413	0.7108075	1.029826
[8]	{Boutique_PL=N, Casino_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y, Resort_PL=N, Ski_PL=N, Spa_PL=N}	=> {NPS_Type=Promoter}	0.3661413	0.7108075	1.029826
[9]	{Boutique_PL=N, Conference_PL=N, Convention_PL=N, Golf_PL=N, Indoor.Corridors_PL=Y,				

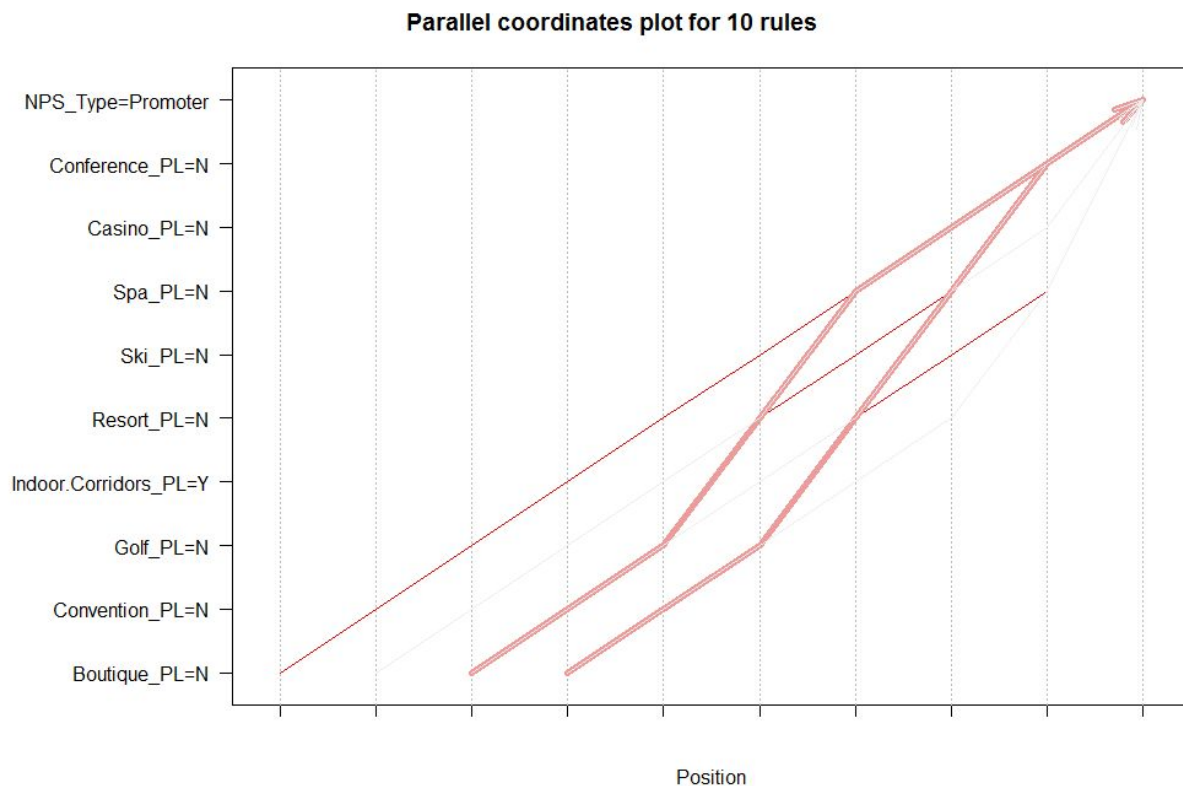
## NPS Analysis for Hyatt Hotel Group

```

Resort_PL=N,
Ski_PL=N,
Spa_PL=N}      => {NPS_Type=Promoter} 0.3661413  0.7108075  1.029826
[10] {Boutique_PL=N,
Casino_PL=N,
Conference_PL=N,
Convention_PL=N,
Golf_PL=N,
Indoor.Corridors_PL=Y,
Resort_PL=N,
Ski_PL=N,
Spa_PL=N}      => {NPS_Type=Promoter} 0.3661413  0.7108075  1.029826

```





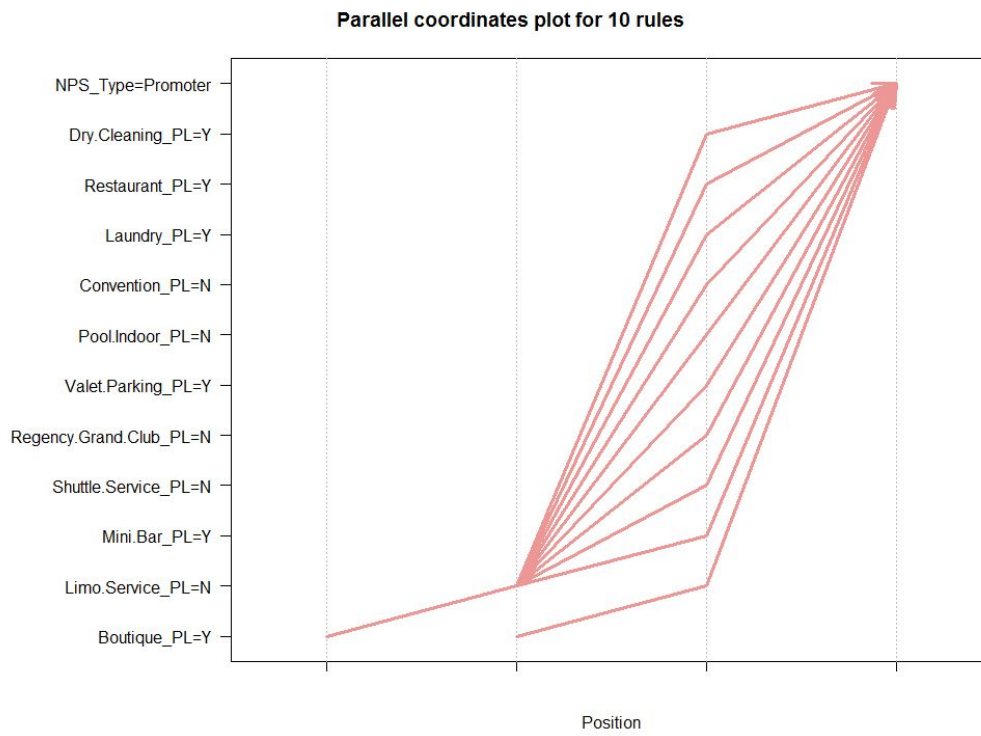
Based on the association rule above with a parameter of  $\text{supp} = 0.35$ ,  $\text{conf} = 0.7$ , we can hardly get any service that are currently associated with a promoter NPS\_Type.

First, it means whether the customer is a promoter or not is not affected too much in many cases by whether there is service. Though, noticeably this is caused by that many hotels don't have the service available. So when we design the association rule with a support of 0.35, most service are labeled N when associated with NPS\_Type promoter. This model is not suitable for the dataset we have in order to gather useful information.

So the next step for association rule analysis is to look high confidence variables that are associated with promoter type. The next step is using  $\text{supp}=0.05$  and look at what rules are have the top 10 confidence.

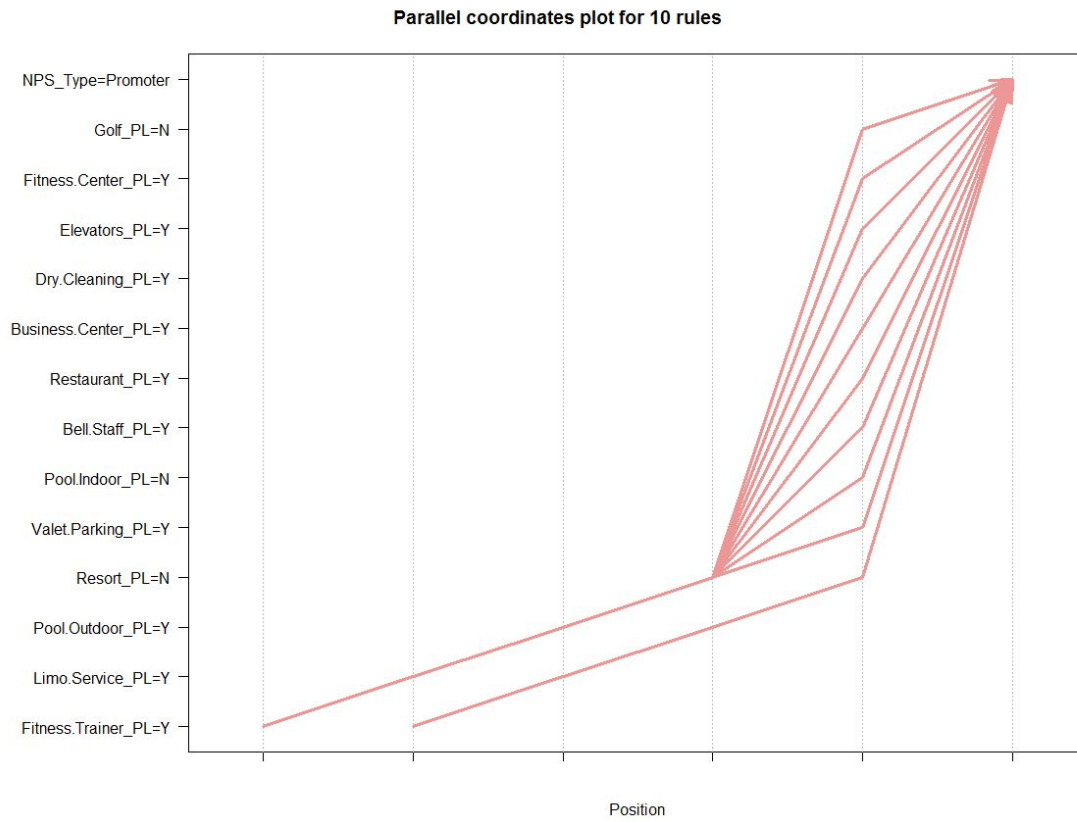
February Top 10 confidence association rules





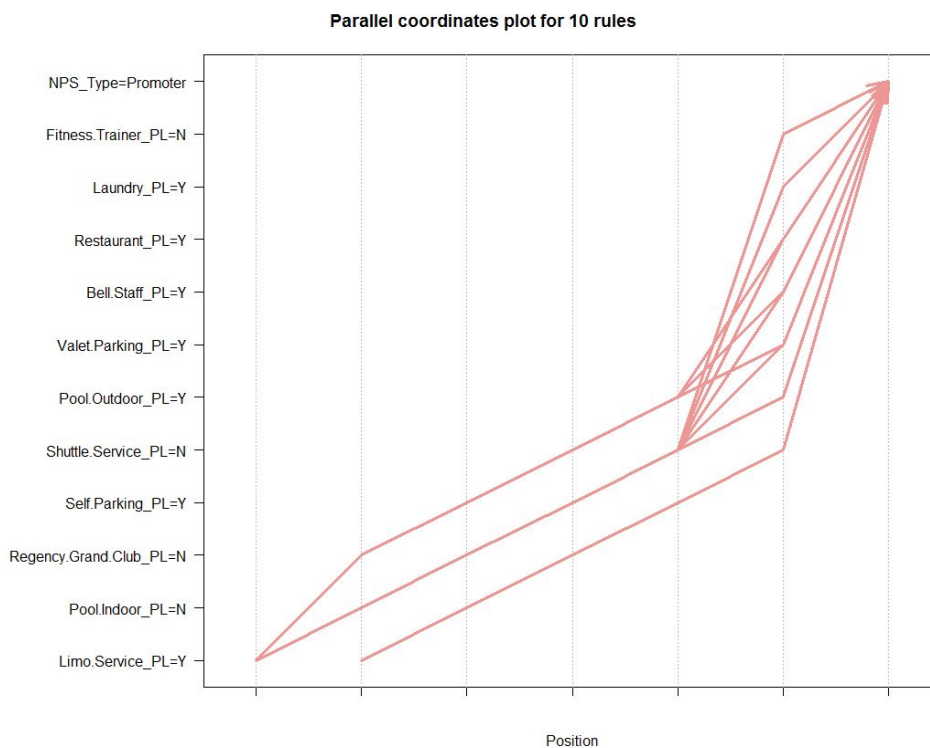
May Top 10 confidence association rules

## NPS Analysis for Hyatt Hotel Group

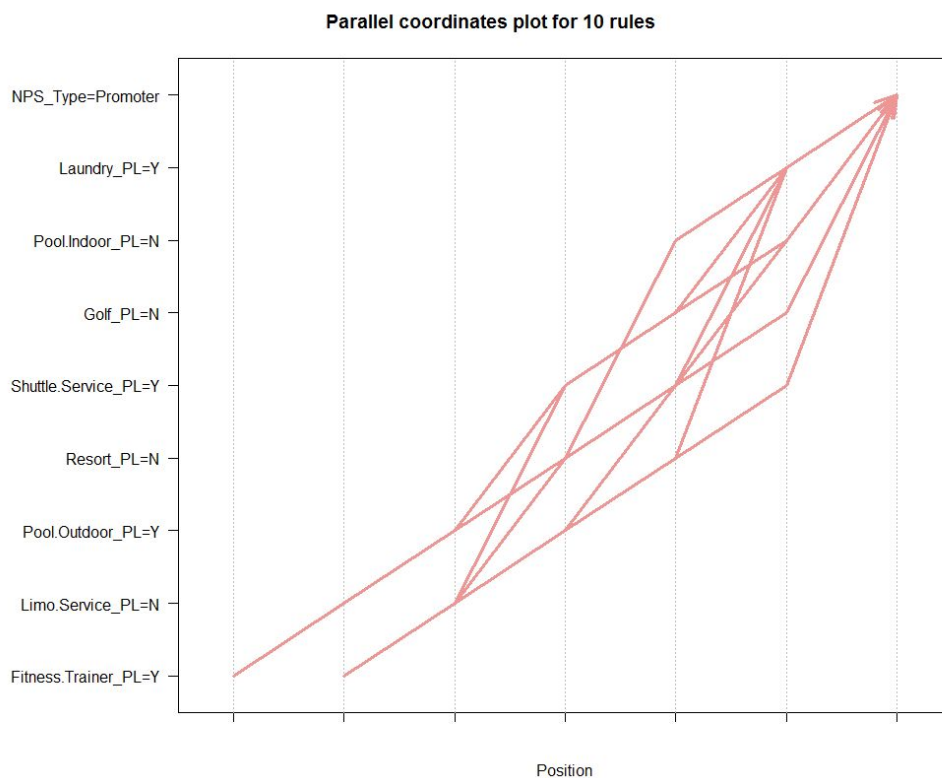


## August Top 10 confidence association rules

## NPS Analysis for Hyatt Hotel Group



## November Top 10 confidence association rules



Month	2	5	8	11	
	Dry Clean	Fitness center	Laundry	Laundry	
	Restaurant	Elevator	Restaurant	Shuttle service	
	Laundry	Dry cleaning	Bell staff	Pool outdoor	
	Valet parking	Business center	Valet parking	Fitness trainer	
	Mini bar	Restaurant	Pool outdoor		
	Boutique	Bell staff	Self parking		
		Valet parking			
		Pool outdoor			
		Limo			
		Fitness trainer			

Appear in 2 month

Appear in 3 month

Based on the association rule between service and NPS\_Type= Promoter. With top 10 confidence and top 10 support from four months from each season, some service are positively associated with a promoter type of NPS(in association rule, the service existence is Yes).

Based on four months of data, restaurant, laundry, outdoor pool are positively associated with promoter type in 3 months.

Dry cleaning, Valet parking, Bell staff, Fitness trainer are associated with promoter type in 2 months.

So these services are closely associated with a NPS\_Type of promoter whose travel reason is business. Possible interpretations from the association rules are:

Hyatt hotels that want to increase the promoter size should work on providing good services in these aspects accordingly to their current situation. And hotels that don't have these services yet but are currently having budget of expanding service can consider these as desirable choices to invest.

## PREDICTION MODELS FOR NPS\_TYPE

### MOTIVATION AND QUESTIONS TO SOLVE

As we mentioned in previous Questions section, there are two purposes of visiting for reservation: leisure and business. Customers with different visiting purposes may have different expectations and preferences which will influence rating NPS\_TYPE, i.e., room types, the length of stay. Besides, we are also wondering whether customers rating NPS\_TYPE behavior will vary among different seasons.

From the two motivations, we built different NPS\_TYPE prediction models (1) for the two different visiting purposes, and (2) for four different seasons.

### INDEPENDENT VARIABLE SELECTION

Two types of independent variables were used to build prediction models, one is about hotel hardware facilities, the other is about customer attributes. In order to select 'good' independent variables, we used correlation analysis between independent variables and NPS\_Type. We used three correlation analysis methods to select independent variables: (1) `chisq.test()` for correlation analysis between categorical variables; (2) `cor.test()` for correlation analysis between numeric variables; (3) `aov()` for correlation analysis between categorical variable and numeric variable. The variable NPS\_Type is categorical. As shown in the screenshot below, for categorical variables (in fact, most of them are categorical variables), we used chi-square test.

```
chisq.test(myOut$MARKET_CODE_C, myOut$NPS_Type) #p-value = 0.4325
chisq.test(myOut$MARKET_GROUP_C, myOut$NPS_Type) #p-value = 0.03644 **
chisq.test(myOut$RATE_PLAN_C, myOut$NPS_Type) #p-value = 0.903
chisq.test(myOut$ROOM_TYPE_CODE_C, myOut$NPS_Type) #p-value = 0.003403 ***
chisq.test(myOut$ROOM_TYPE_DESCRIPTION_C, myOut$NPS_Type) #p-value = 0.006716 ***
chisq.test(myOut$LENGTH_OF_STAY_C, myOut$NPS_Type) #p-value = 0.06735
chisq.test(myOut$POV_CODE_C, myOut$NPS_Type) #p-value = 0.03327 **
chisq.test(myOut$DIRECT_NIGHTS_NUM_R, myOut$NPS_Type) #p-value = 0.05924
chisq.test(myOut$ROOM_TYPE_CODE_R, myOut$NPS_Type) # p-value = 3.362e-07 ***
chisq.test(myOut$ROOM_TYPE_DESCRIPTION_R, myOut$NPS_Type) #p-value = 0.003869 ***
```

Finally, we picked up four hotel-related variables and three customer-related variables as below. The visiting purpose is significantly related with NPS\_Type. But in the models of leisure v.s business, we didn't include this variable.

- About hotel
  - Market group: MARKET\_GROUP\_C
  - Room type: ROOM\_TYPE\_DESCRIPTION\_C & ROOM\_TYPE\_DESCRIPTION\_R
  - Visiting purpose: POV\_CODE\_C
- About customer
  - Length of stay: LENGTH\_OF\_STAY\_CATEGORY\_R

## NPS Analysis for Hyatt Hotel Group

- Individual or group visit: GROUPS\_VS\_FIT\_R
- GP member: GOLDPASSPORT\_FLG\_R

### MODELS FOR LEISURE V.S BUSINESS

We used February data to build two ksvm models for leisure and business respectively. The first  $\frac{2}{3}$  data was used as training data and the rest  $\frac{1}{3}$  was used as test data. For evaluation, we compared the actual NPS\_Type and the predicted results to get the accuracy. The R code for accuracy is:

```
results_201402_l_us <- table(testData_201402_l_us$NPS_Type, pred_201402_l_us)

accuracy_201402_l_us <- (results_201402_l_us[1,1] + results_201402_l_us[2,2] + results_201402_l_us[3,3]) /
length(pred_201402_l_us)

accuracy_201402_l_us #0.675
```

For leisure purpose, the model is:

```
Code: ksvm(NPS_Type ~ MARKET_GROUP_C + ROOM_TYPE_DESCRIPTION_C + ROOM_TYPE_DESCRIPTION_R
+ LENGTH_OF_STAY_CATEGORY_R + GROUPS_VS_FIT_R + GOLDPASSPORT_FLG_R, data =
trainData_201402_l_us, kernel = "rbfdot", kpar = "automatic", C = 5, cross = 3, prob.model = TRUE)
```

For business purpose, the model is:

```
Code: ksvm(NPS_Type ~ MARKET_GROUP_C + ROOM_TYPE_DESCRIPTION_C + ROOM_TYPE_DESCRIPTION_R
+ LENGTH_OF_STAY_CATEGORY_R + GROUPS_VS_FIT_R + GOLDPASSPORT_FLG_R, data =
trainData_201402_b_us, kernel = "rbfdot", kpar = "automatic", C = 5, cross = 3, prob.model = TRUE)
```

The leisure model accuracy is 0.675 and the business model accuracy is 0.685.

### MODELS FOR DIFFERENT SEASONS

We used two algorithm, KSVM and Naive Bayes, to build our models. Given the limited memory space, we used 2/10 February data as training data, and another 1/10 February data, 1/10 May data, 1/10 August data and 1/10 November data as four test data. The evaluation method we used is the same as the last subsection.

Using the same training data in February to predict NPS\_Type in four different months, we can compare the accuracy numbers to see if one season's data could be used to predict other seasons' NPS\_Type.

If the accuracy numbers of May, August and November is almost the same or above the accuracy of February, we can not only get the conclusion that there is no difference in different seasons about how factors (independent variables) influence NPS\_Type but also applied the previous season data to build models and predict the following season's NPS\_Type in the same year.

As shown in the table below, only the NB model for August got the higher accuracy than February. The rest accuracy numbers for May, Aug and Nov are lower than February, especially May.

	Feb	May	Aug	Nov
KSVM	<b>0.678</b>	0.634	0.656	0.644

NB	<b>0.633</b>	0.623	<b>0.634</b>	0.625
----	--------------	-------	--------------	-------

Conclusion: the factors influencing NPS\_Type are varied among different seasons. Only using one month or one season data to build models may not get higher accuracy than using their own data. To improve the accuracy performance of prediction models, we may use each season' or even each month' data as the training data individually. However, in order to get NPS\_Type estimation in advance, we can still try the prior season's or month's data to build prediction models.

## REFLECTIONS

From the project, we first learnt about how to handle large datasets with many variables over 12 months period. We used different seasons and offseason or in season as a subcategory for selecting the months we want to use for analysis. Our criteria is implemented so we can have adequate and representative data for our business questions.

After getting familiar with the data, we learnt to deal with empty cells and NAs in the dataset with different strategies based on analysis we want to perform. In some cases, it is not feasible to just simply remove all empty cells or NAs since it will severely reduce the size of our data and bring selection bias to our analysis.

As our analysis progressed, we learnt the importance of communication between group members of the information we individually gathered. Since the data is large and with so many variables, it is hard for one person to find all detailed useful information and perform every analysis. Communication process is very helpful and improved the productivity of individuals and the group.

Lastly, about finishing the project as a group, we strongly suggest in later semesters if feasible, the size of the group to be reduced to 2 to 3 people maximum. Too many members in group will bring difficulty in efficient communication since it is harder to get everyone together at the same time on the same page with various time schedules. And with less members in the group, it will also be easier to balance individual contributions throughout the semester as well.

## CONCLUSION

Visitor's NPS type is totally dependent on visitor's likelihood to recommend the hotel. From our linear model, we find out that guest room quality and customer service quality are two variables that are dependent on likelihood to recommend. Therefore, we recommend hotels which want to promote their visitor's NPS type can increase guest room quality and customer service quality.

We also recommend every hotel focus on the majority visitors visit of purpose. If the visitors purpose meet the type of hotel, the hotel can provide discount or better service to increase visitor's experiences. If the visitors purpose does not meet the type of hotel, we recommend either the hotels change their types to meet the majority of visitors or the hotels attract more related visitors.

Using one season's or month's data to build models can't predict different seasons' or months' NPS\_Type. Different models in each season or month should be generated to improve accuracy performance. But the

prior data can still be used to predict the following unknown NPS in practice.

## LIMITATIONS

In this project, we only studied the data of hotels within the United States given the limited time and memory space in computer. Although we found that hotels in U.S. were dense in the generated map and there should be different patterns in data for different continents and countries, more data and aspects should be explored in the future to get more comprehensive understanding of improve NPS for hotels. Moreover, we only randomly selected four months to represent four seasons to examine the relationship between factors (independent variables) and NPS which was subjectively. More months should be included to get more accurate results and conclusions.

## REFERENCE

## APPENDIX

### Codes used to Clean Dataset and Generate Graphs

#### 1. CLEAN DATASET

```
#load data201402#####
```

```
file201402.ori <- read.csv("\\\\hd.ad.syr.edu\\01\\877dcb\\Documents\\Desktop\\IST  
687\\Project\\complete_data\\out-201402.csv")
```

```
#subset only north america
```

```
na.open201402 <- file201402.ori[file201402.ori$Sub.Continent_PL == "North America",]
```

```
#subset - opening hotels
```

```
na.open201402 <- na.open201402[na.open201402$Status_PL == "Open",]
```

```
#subset to remove rows that have empty NPS_Type cells
```

```
na.open201402 <- na.open201402[!(is.na(na.open201402$NPS_Type) | na.open201402$NPS_Type == ""), ]
```

```
#export data201402
```

```
write.csv(na.open201402, file = "na.open201402.csv")
```



```
#seperate leisure and business data
```

```
dat.leisure.1402<- na.open201402[na.open201402$POV_CODE_C == "LEISURE",]
```

```
dat.business.1402 <- na.open201402[na.open201402$POV_CODE_C == "BUSINESS",]
```

```
##export data
```

```
write.csv(dat.leisure.1402, file = "dat.leisure.1402.csv")
```

```
write.csv(dat.business.1402, file = "dat.business.1402.csv")
```

```
#load data201405#####
```

```
file201405.ori <- read.csv("\\\\hd.ad.syr.edu\\01\\877dcb\\Documents\\Desktop\\IST  
687\\Project\\complete_data\\out-201405.csv")
```

```
#subset only north america
```

```
na.open201405 <- file201405.ori[file201405.ori$Sub.Continent_PL == "North America",]
```

```
#subset - opening hotels
```

```
na.open201405 <- na.open201405[na.open201405$Status_PL== "Open",]
```

```
#subset to remove rows that have empty NPS_Type cells
```

```
na.open201405 <- na.open201405[!(is.na(na.open201405$NPS_Type) | na.open201405$NPS_Type==""), ]
```

```
#export data201405
```

```
write.csv(na.open201405, file = "na.open201405.csv")
```

```
#seperate leisure and business data
```

```
dat.leisure.1405<- na.open201405[na.open201405$POV_CODE_C == "LEISURE",]
```

## NPS Analysis for Hyatt Hotel Group

```
dat.business.1405 <- na.open201405[na.open201405$POV_CODE_C == "BUSINESS",]
```

```
##export data
```

```
write.csv(dat.leisure.1405, file = "dat.leisure.1405.csv")
```

```
write.csv(dat.business.1405, file = "dat.business.1405.csv")
```

```
#load data201408#####
```

```
file201408.ori <- read.csv("\\\\hd.ad.syr.edu\\01\\877dcb\\Documents\\Desktop\\IST  
687\\Project\\complete_data\\out-201408.csv")
```

```
#subset only north america
```

```
na.open201408 <- file201408.ori[file201408.ori$Sub.Continent_PL == "North America",]
```

```
#subset - opening hotels
```

```
na.open201408 <- na.open201408[na.open201408$Status_PL == "Open",]
```

```
#subset to remove rows that have empty NPS_Type cells
```

```
na.open201408 <- na.open201408[!(is.na(na.open201408$NPS_Type) | na.open201408$NPS_Type == ""), ]
```

```
#export data201408
```

```
write.csv(na.open201408, file = "na.open201408.csv")
```

```
#seperate leisure and business data
```

```
dat.leisure.1408 <- na.open201408[na.open201408$POV_CODE_C == "LEISURE",]
```

```
dat.business.1408 <- na.open201408[na.open201408$POV_CODE_C == "BUSINESS",]
```

```
##export data
```

```
write.csv(dat.leisure.1408, file = "dat.leisure.1408.csv")
```

## NPS Analysis for Hyatt Hotel Group

```
write.csv(dat.business.1408, file = "dat.business.1408.csv")
```

```
#load data201411#####
```

```
file201411.ori <- read.csv("\\\\hd.ad.syr.edu\\01\\877dcb\\Documents\\Desktop\\IST687\\Project\\complete_data\\out-201411.csv")
```

```
#subset only north america
```

```
na.open201411 <- file201411.ori[file201411.ori$Sub.Continent_PL == "North America",]
```

```
#subset - opening hotels
```

```
na.open201411 <- na.open2014011[na.open201411$Status_PL == "Open",]
```

```
#subset to remove rows that have empty NPS_Type cells
```

```
na.open201411 <- na.open201411[!(is.na(na.open201411$NPS_Type) | na.open201411$NPS_Type == ""), ]
```

```
#export data201411
```

```
write.csv(na.open201411, file = "na.open201411.csv")
```

```
#seperate leisure and business data
```

```
dat.leisure.1411 <- na.open201411[na.open201411$POV_CODE_C == "LEISURE",]
```

```
dat.business.1411 <- na.open201411[na.open201411$POV_CODE_C == "BUSINESS",]
```

```
##export data
```

```
write.csv(dat.leisure.1411, file = "dat.leisure.1411.csv")
```

```
write.csv(dat.business.1411, file = "dat.business.1411.csv")
```

## 2. SELECT RELATED VARIABLES FOR NPS\_Type

```
out_201402 <- read.csv("/Users/QunfangWu/Documents/PhD Program/PhD 1-2/IST687 Applied Data
```

## NPS Analysis for Hyatt Hotel Group

```
Science/Test and Project/data/out-201402.csv")
```

```
View(out_201402)
```

```
myOut_201402 <- data.frame(myOut_201402)
```

```
myOut_201402 <- out_201402[, c(1:50, 230:237)]
```

```
View(myOut_201402)
```

```
summary(myOut_201402)
```

```
#chisq.test(myOut_201402$MARKET_CODE_C, myOut_201402$NPS_Type)
```

```
## remove all NAs from the data set
```

```
myOut <- na.omit(myOut_201402)
```

```
## Will test the correlation between NPS_Type and the variables below
```

```
#1-50
```

```
# MARKET_CODE_C
```

```
# MARKET_GROUP_C
```

```
# RATE_PLAN_C
```

```
# ROOM_TYPE_CODE_C
```

```
# ROOM_TYPE_DESCRIPTION_C
```

```
# LENGTH_OF_STAY_C
```

```
# POV_CODE_C
```

```
# DIRECT_NIGHTS_NUM_R
```

```
# ROOM_TYPE_CODE_R
```

```
# ROOM_TYPE_DESCRIPTION_R
```

```
##correlation
```

```
chisq.test(myOut$MARKET_CODE_C, myOut$NPS_Type) #p-value = 0.4325
```

```
chisq.test(myOut$MARKET_GROUP_C, myOut$NPS_Type) #p-value = 0.03644 **
```

```
chisq.test(myOut$RATE_PLAN_C, myOut$NPS_Type) #p-value = 0.903
```

```
chisq.test(myOut$ROOM_TYPE_CODE_C, myOut$NPS_Type) #p-value = 0.003403 ***
```

## NPS Analysis for Hyatt Hotel Group

```
chisq.test(myOut$ROOM_TYPE_DESCRIPTION_C, myOut$NPS_Type) #p-value = 0.006716 ***
chisq.test(myOut$LENGTH_OF_STAY_C, myOut$NPS_Type) #p-value = 0.06735
chisq.test(myOut$POV_CODE_C, myOut$NPS_Type) #p-value = 0.03327 **
chisq.test(myOut$DIRECT_NIGHTS_NUM_R, myOut$NPS_Type) #p-value = 0.05924
chisq.test(myOut$ROOM_TYPE_CODE_R, myOut$NPS_Type) # p-value = 3.362e-07 ***
chisq.test(myOut$ROOM_TYPE_DESCRIPTION_R, myOut$NPS_Type) #p-value = 0.003869 ***
```

### 3. BUILD PREDICTION MODELS

```
#####Prediction Model
```

```
#Read data
```

```
na.open201402 <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687
Project/na.open.nps.dat/na.open201402.csv")
```

```
na.open201405 <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687
Project/na.open.nps.dat/na.open201405.csv")
```

```
na.open201408 <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687
Project/na.open.nps.dat/na.open201408.csv")
```

```
na.open201411 <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687
Project/na.open.nps.dat/na.open201411.csv")
```

```
summary(na.open201402$NPS_Type)
```

```
perc02 <- c(length(na.open201402$NPS_Type[na.open201402$NPS_Type == "Detractor"]),
length(na.open201402$NPS_Type[na.open201402$NPS_Type
=="Passive"]),length(na.open201402$NPS_Type[na.open201402$NPS_Type == "Promoter"]) )
```

```
summary(na.open201405$NPS_Type)
```

```
perc05 <- c(length(na.open201405$NPS_Type[na.open201405$NPS_Type == "Detractor"]),
length(na.open201405$NPS_Type[na.open201405$NPS_Type
=="Passive"]),length(na.open201405$NPS_Type[na.open201405$NPS_Type == "Promoter"]) )
```

```
summary(na.open201408$NPS_Type)
```

```
perc08 <- c(length(na.open201408$NPS_Type[na.open201408$NPS_Type == "Detractor"]),
length(na.open201408$NPS_Type[na.open201408$NPS_Type
```

## NPS Analysis for Hyatt Hotel Group

```

=="Passive"]),length(na.open201408$NPS_Type[na.open201408$NPS_Type=="Promoter"]) )

summary(na.open201411$NPS_Type)

perc11 <- c(length(na.open201411$NPS_Type[na.open201411$NPS_Type=="Detractor"]),
length(na.open201411$NPS_Type[na.open201411$NPS_Type
=="Passive"]),length(na.open201411$NPS_Type[na.open201411$NPS_Type=="Promoter"]) )

labels = c("Detractor", "Passive", "Promoter")

#get the distribution of NPS_Type(s) in different months

P02 <- pie(perc02, labels)

P05 <- pie(perc05, labels)

P08 <- pie(perc08, labels)

P11 <- pie(perc11, labels)

#####build models

library(kernlab)

randIndex <- sample(1:dim(na.open201402_l_us)[1])

cutPoint2_3 <- floor(2 * dim(na.open201402_l_us)[1]/3)

###read 201402 for business v.s leisure

na.open201402_l <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687
Project/Dataset/Leisure/dat.leisure.1402.csv")

na.open201402_l_us <- na.open201402_l[na.open201402_l$Country_PL == "United States",]

na.open201402_b <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687
Project/Dataset/Business/dat.business.1402.csv")

na.open201402_b_us <- na.open201402_b[na.open201402_b$Country_PL == "United States",]

###2/3 in Feb for training data, and the rest 1/3 in Feb for test data

##business

trainData_201402_b_us <-na.open201402_b_us[randIndex[1:cutPoint2_3],]

testData_201402_b_us <- na.open201402_b_us[randIndex[(cutPoint2_3+1):dim(na.open201402_b_us)[1]],]

```

## NPS Analysis for Hyatt Hotel Group

#hotel + customer variables

```
svmOutput_201402_b_us <- ksvm(NPS_Type ~ MARKET_GROUP_C + ROOM_TYPE_DESCRIPTION_C +
ROOM_TYPE_DESCRIPTION_R + LENGTH_OF_STAY_CATEGORY_R + GROUPS_VS_FIT_R +
GOLDPASSPORT_FLG_R, data = trainData_201402_b_us, kernel = "rbfdot", kpar = "automatic", C = 5, cross = 3,
prob.model = TRUE)

pred_201402_b_us <- predict(svmOutput_201402_b_us, testData_201402_b_us)

testData_201402_b_us_na <- na.omit(testData_201402_b_us$NPS_Type)

results_201402_b_us <- table(testData_201402_b_us_na, pred_201402_b_us)

accuracy_201402_b_us <- (results_201402_b_us[1,1] + results_201402_b_us[2,2] +
results_201402_b_us[3,3])/length(pred_201402_b_us)

accuracy_201402_b_us #0.685
```

###leisure

```
trainData_201402_l_us <- na.open201402_l_us[randIndex[1:cutPoint2_3],]

testData_201402_l_us <- na.open201402_l_us[randIndex[(cutPoint2_3+1):dim(na.open201402_l_us)[1]],]

#hotel + customer variables

svmOutput_201402_l_us <- ksvm(NPS_Type ~ MARKET_GROUP_C + ROOM_TYPE_DESCRIPTION_C +
ROOM_TYPE_DESCRIPTION_R + LENGTH_OF_STAY_CATEGORY_R + GROUPS_VS_FIT_R +
GOLDPASSPORT_FLG_R, data = trainData_201402_l_us, kernel = "rbfdot", kpar = "automatic", C = 5, cross = 3,
prob.model = TRUE)

pred_201402_l_us <- predict(svmOutput_201402_l_us, testData_201402_l_us)

results_201402_l_us <- table(testData_201402_l_us$NPS_Type, pred_201402_l_us)

accuracy_201402_l_us <- (results_201402_l_us[1,1] + results_201402_l_us[2,2] +
results_201402_l_us[3,3])/length(pred_201402_l_us)

accuracy_201402_l_us #0.675
```

###business & leisure in different seasons/months

```
na.open201402 <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687 Project/dataset/NA.open
201402-201411/na.open201402.csv")
```

## NPS Analysis for Hyatt Hotel Group

```
na.open201405 <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687 Project/dataset/NA.open
201402-201411/na.open201405.csv")
```

```
na.open201408 <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687 Project/dataset/NA.open
201402-201411/na.open201408.csv")
```

```
na.open201411 <- read.csv("/Users/QunfangWu/Google Drive/PhD Study/IST 687 Project/dataset/NA.open
201402-201411/na.open201411.csv")
```

```
##read data in 4 different months: Feb, May, Aug, and Nov
```

```
##2/10 in Feb for training data, and 1/10 in Feb, May, Aug, and Nov for test data
```

```
#02
```

```
TrainData_201402_us <- na.open201402[na.open201402$Country_PL == "United States",]
```

```
TrainData_201402_us_na <- TrainData_201402_us[TrainData_201402_us$NPS_Type != "",]
```

```
#05
```

```
testData_201405_us <- na.open201405[na.open201405$Country_PL == "United States",]
```

```
testData_201405_us_na <- testData_201405_us[testData_201405_us$NPS_Type != "",]
```

```
randIndex <- sample(1:dim(testData_201405_us_na)[1])
```

```
cutPoint1_10 <- floor(dim(testData_201405_us_na)[1]/10)
```

```
testData_201405_us_na <- testData_201405_us_na[randIndex[1:cutPoint1_10],]
```

```
#08
```

```
testData_201408_us <- na.open201408[na.open201408$Country_PL == "United States",]
```

```
testData_201408_us_na <- testData_201408_us[testData_201408_us$NPS_Type != "",]
```

```
randIndex <- sample(1:dim(testData_201408_us_na)[1])
```

```
cutPoint1_10 <- floor(dim(testData_201408_us_na)[1]/10)
```

```
testData_201408_us_na <- testData_201408_us_na[randIndex[1:cutPoint1_10],]
```

```
#11
```

```
testData_201411_us <- na.open201411[na.open201411$Country_PL == "United States",]
```

```
testData_201411_us_na <- testData_201411_us[testData_201411_us$NPS_Type != "",]
```

```
randIndex <- sample(1:dim(testData_201411_us_na)[1])
```



## NPS Analysis for Hyatt Hotel Group

```

cutPoint1_10 <- floor(dim(testData_201411_us_na)[1]/10)

testData_201411_us_na <- testData_201411_us_na[randIndex[1:cutPoint1_10],]

#ksvm models

library(kernlab)

randIndex <- sample(1:dim(TrainData_201402_us_na)[1])

cutPoint2_10 <- floor(2*dim(TrainData_201402_us_na)[1]/10)

cutPoint3_10 <- floor(3*dim(TrainData_201402_us_na)[1]/10)

trainData_201402_us_na <- TrainData_201402_us_na[randIndex[1:cutPoint2_10],]

testData_201402_us_na <- TrainData_201402_us_na[randIndex[(cutPoint2_10+1):cutPoint3_10],]

svmOutput_201402_us <- ksvm(NPS_Type ~ MARKET_GROUP_C + ROOM_TYPE_DESCRIPTION_C +
ROOM_TYPE_DESCRIPTION_R + POV_CODE_C + LENGTH_OF_STAY_CATEGORY_R + GROUPS_VS_FIT_R +
GOLDPASSPORT_FLG_R, data = trainData_201402_us_na, kernel = "rbfdot", kpar = "automatic", C = 5, cross =
3, prob.model = TRUE)

#02

pred_201402_us <- predict(svmOutput_201402_us, testData_201402_us_na)

results_201402_us <- table(testData_201402_us_na$NPS_Type, pred_201402_us)

accuracy_201402_us <- (results_201402_us[1,1] + results_201402_us[2,2] +
results_201402_us[3,3])/length(pred_201402_us)

accuracy_201402_us #0.678

#05

pred_201405_us <- predict(svmOutput_201402_us, testData_201405_us_na)

results_201405_us <- table(testData_201405_us_na, pred_201405_us)

accuracy_201405_us <- (results_201405_us[1,1] + results_201405_us[2,2] +
results_201405_us[3,3])/length(pred_201405_us)

accuracy_201405_us #0.634

#08

```

## NPS Analysis for Hyatt Hotel Group

```
pred_201408_us <- predict(svmOutput_201402_us, testData_201408_us_na)
```

```
results_201408_us <- table(testData_201408_us, pred_201408_us)
```

```
accuracy_201408_us <- (results_201408_us[1,1] + results_201408_us[2,2] +  
results_201408_us[3,3])/length(pred_201408_us)
```

```
accuracy_201408_us #0.656
```

```
#11
```

```
pred_201411_us <- predict(svmOutput_201402_us, testData_201411_us_na)
```

```
results_201411_us <- table(testData_201411_us, pred_201411_us)
```

```
accuracy_201411_us <- (results_201411_us[1,1] + results_201411_us[2,2] +  
results_201411_us[3,3])/length(pred_201411_us)
```

```
accuracy_201411_us #0.644
```

```
##NB models
```

```
library(e1071)
```

```
nb.model_201402_us <- naiveBayes(NPS_Type ~ MARKET_GROUP_C + ROOM_TYPE_DESCRIPTION_C +  
ROOM_TYPE_DESCRIPTION_R + POV_CODE_C + LENGTH_OF_STAY_CATEGORY_R + GROUPS_VS_FIT_R +  
GOLDPASSPORT_FLG_R, data = trainData_201402_us_na)
```

```
#02
```

```
nb.pred_201402_us <- predict(nb.model_201402_us, testData_201402_us_na)
```

```
results_201402_us_nb <- table(testData_201402_us_na$NPS_Type, nb.pred_201402_us)
```

```
accuracy_201402_us <- (results_201402_us_nb[1,1] + results_201402_us_nb[2,2] +  
results_201402_us_nb[3,3])/length(pred_201402_us)
```

```
accuracy_201402_us #0.633
```

```
#05
```

```
nb.pred_201405_us <- predict(nb.model_201402_us, testData_201405_us_na)
```

```
results_201405_us_nb <- table(testData_201405_us_na$NPS_Type, nb.pred_201402_us)
```

```
accuracy_201405_us <- (results_201405_us_nb[1,1] + results_201405_us_nb[2,2] +  
results_201405_us_nb[3,3])/length(pred_201405_us)
```

## NPS Analysis for Hyatt Hotel Group

```
accuracy_201405_us #0.623
```

```
#08
```

```
nb.pred_201408_us <- predict(nb.model_201402_us, testData_201408_us_na)
```

```
results_201408_us_nb <- table(testData_201408_us_na$NPS_Type, nb.pred_201408_us)
```

```
accuracy_201408_us <- (results_201408_us_nb[1,1] + results_201408_us_nb[2,2] +  
results_201408_us_nb[3,3])/length(pred_201408_us)
```

```
accuracy_201408_us #0.634
```

```
#11
```

```
nb.pred_201411_us <- predict(nb.model_201402_us, testData_201411_us_na)
```

```
results_201411_us_nb <- table(testData_201411_us_na$NPS_Type, nb.pred_201411_us)
```

```
accuracy_201411_us <- (results_201411_us_nb[1,1] + results_201411_us_nb[2,2] +  
results_201411_us_nb[3,3])/length(pred_201411_us)
```

```
accuracy_201411_us #0.625
```

#### 4. US MAP OF FOUR MONTHS SELECTED WITH AVERAGE LIKELIHOOD TO RECOMMEND OF EACH HOTEL

```
install.packages("dplyr")
```

```
library(dplyr)
```

```
install.packages("maps")
```

```
install.packages("mapdata")
```

```
library(mapdata)
```

```
# Feb Data Process
```

```
Feb1 <- na.open201402
```

```
Feb1 <- Feb1[Feb1$Country_PL== "United States",]
```

```
df <- data.frame(Feb1$Property.Latitude_PL, Feb1$Property.Longitude_PL)
```

```
Location <- distinct(df, Feb1$Property.Latitude_PL, Feb1$Property.Longitude_PL)
```

```
Latitude <- Location$`Feb1$Property.Latitude_PL`
```

```
Longitude <- Location$`Feb1$Property.Longitude_PL`
```

## NPS Analysis for Hyatt Hotel Group

Location

```
newdf<- Location[order(Latitude),]
```

```
Feb_Average_Likelihood <- tapply(Feb1$Likelihood_Recommend_H,Feb1$Property.Latitude_PL,mean)
```

```
Latitude <- rownames(Feb_Average_Likelihood)
```

```
meanLikelihood <- data.frame(Latitude,Feb_Average_Likelihood)
```

```
Feb_Likelihood <- data.frame(newdf,meanLikelihood)
```

```
Feb_US <- Feb_Likelihood [,-3]
```

Feb\_US

```
library(ggplot2)
```

```
library(maps)
```

```
library(ggmap)
```

```
#load us map data
```

```
all_states <- map_data("state")
```

```
#plot all states with ggplot
```

```
map1 <- ggplot() + geom_polygon(data=all_states, aes(x=long, y=lat, group = group),fill="light  
grey",color="white") +theme(axis.text = element_blank(),panel.background = element_blank())
```

```
map1
```

```
# May Data Process
```

```
May1 <- na.open201405
```

```
May1 <- May1[May1$Country_PL== "United States",]
```

```
df1 <- data.frame(May1$Property.Latitude_PL,May1$Property.Longitude_PL)
```

```
df1
```

```
Location_May <- distinct(df1,May1$Property.Latitude_PL,May1$Property.Longitude_PL)
```

```
Latitude_May <- Location_May$`May1$Property.Latitude_PL`
```

```
Longitude_May <- Location_May$`May1$Property.Longitude_PL`
```

```

newdf1<- Location_May[order(Latitude_May),]

May_Average_Likelihood <- tapply(May1$Likelihood_Recommend_H,May1$Property.Latitude_PL,mean)

Latitude <- rownames(May_Average_Likelihood)

meanLikelihood <- data.frame(Latitude,May_Average_Likelihood)

May_Likelihood <- data.frame(newdf1,meanLikelihood)

May_US <- May_Likelihood [,-3]

May_US

# Aug Data Process

Aug1 <- na.open201408

Aug1 <- Aug1[Aug1$Country_PL== "United States",]

df2 <- data.frame(Aug1$Property.Latitude_PL,Aug1$Property.Longitude_PL)

Location_Aug <- distinct(df2,Aug1$Property.Latitude_PL,Aug1$Property.Longitude_PL)

Latitude_Aug <- Location_Aug$`Aug1$Property.Latitude_PL`

Longitude_Aug <- Location_Aug$`Aug1$Property.Longitude_PL`

newdf2<- Location_Aug[order(Latitude_Aug),]

Aug_Average_Likelihood <- tapply(Aug1$Likelihood_Recommend_H,Aug1$Property.Latitude_PL,mean)

Latitude <- rownames(Aug_Average_Likelihood)

meanLikelihood <- data.frame(Latitude,Aug_Average_Likelihood)

Aug_Likelihood <- data.frame(newdf2,meanLikelihood)

Aug_US <- Aug_Likelihood [,-3]

Aug_US

# Nov Data Process

Nov1 <- na.open201411

Nov1 <- Nov1[Nov1$Country_PL== "United States",]

df3 <- data.frame(Nov1$Property.Latitude_PL,Nov1$Property.Longitude_PL)

```

## NPS Analysis for Hyatt Hotel Group

```

Location_Nov <- distinct(df3,Nov1$Property.Latitude_PL,Nov1$Property.Longitude_PL)

Latitude_Nov <- Location_Nov$`Nov1$Property.Latitude_PL`

Longitude_Nov <- Location_Nov$`Nov1$Property.Longitude_PL`

newdf3<- Location_Nov[order(Latitude_Nov),]

Nov_Average_Likelihood <- tapply(Nov1$Likelihood_Recommend_H,Nov1$Property.Latitude_PL,mean)

Latitude <- rownames(Nov_Average_Likelihood)

meanLikelihood <- data.frame(Latitude,Nov_Average_Likelihood)

Nov_Likelihood <- data.frame(newdf3,meanLikelihood)

Nov_US <- Nov_Likelihood [,-3]

Nov_US

#point with Likelihood to recomand

map.Feb1 <- map1 + geom_point(data=Feb_US, aes(x=Feb_US$Feb1.Property.Longitude_PL,
y=Feb_US$Feb1.Property.Latitude_PL,color=Feb_US$Likelikehood),size=4)

map.May1 <- map1 + geom_point(data=May_US, aes(x=May_US$May1.Property.Longitude_PL,
y=May_US$May1.Property.Latitude_PL,color=May_US$May_Average_Likelihood),size=4)

map.Aug1 <- map1 + geom_point(data=Aug_US, aes(x=Aug_US$Aug1.Property.Longitude_PL,
y=Aug_US$Aug1.Property.Latitude_PL,color=Aug_US$Aug_Average_Likelihood),size=4)

map.Nov1 <- map1 + geom_point(data=Nov_US, aes(x=Nov_US$Nov1.Property.Longitude_PL,
y=Nov_US$Nov1.Property.Latitude_PL,color=Nov_US$Nov_Average_Likelihood),size=4)


# World Map

# World data

world <- `out.201402.(1)`

world <- world[world$NPS_Type != "",]

str(world)

dfworld <- data.frame(world$Property.Latitude_PL,world$Property.Longitude_PL)

View(dfworld)

```

## NPS Analysis for Hyatt Hotel Group

```

Location.world <- distinct(dfworld,world$Property.Latitude_PL,world$Property.Longitude_PL)

Latitude.world <- Location.world$`world$Property.Latitude_PL`

Longitude <- Location.world$`world$Property.Longitude_PL`

Location.world

newdf.world<- Location.world[order(Latitude.world),]

newdf.world

world_Average_Likelihood <- tapply(world$Likelihood_Recommend_H,world$Property.Latitude_PL,mean)

Latitude.world <- rownames(world_Average_Likelihood)

meanLikelihood <- data.frame(Latitude.world,world_Average_Likelihood)

world_Likelihood <- data.frame(newdf.world,meanLikelihood)

world_data<- world_Likelihood [,3]

world_data

mp <- NULL

mapWorld <- borders("world", colour="white", fill="light grey")

mapWorld

mp <- ggplot() + mapWorld + theme(axis.text = element_blank(),panel.background = element_blank())

#Layer the cities on top

mp <- mp+ geom_point(aes(x=dfworld$world.Property.Longitude_PL,
y=dfworld$world.Property.Latitude_PL),color="blue", size=3)

map.world1 <- mp + geom_point(data=world_data, aes(x=world_data$world.Property.Longitude_PL,
y=world_data$world.Property.Latitude_PL,color=world_data$world_Average_Likelihood),size=3)

```

