2018-1004 IST 687 Applied Data Science

Group A Team Project:
Hyatt Hotels NPS Analysis

Team Members:
Charles Anyanwu, Will Holt, Matt Maloney

Submitted:
December 2018

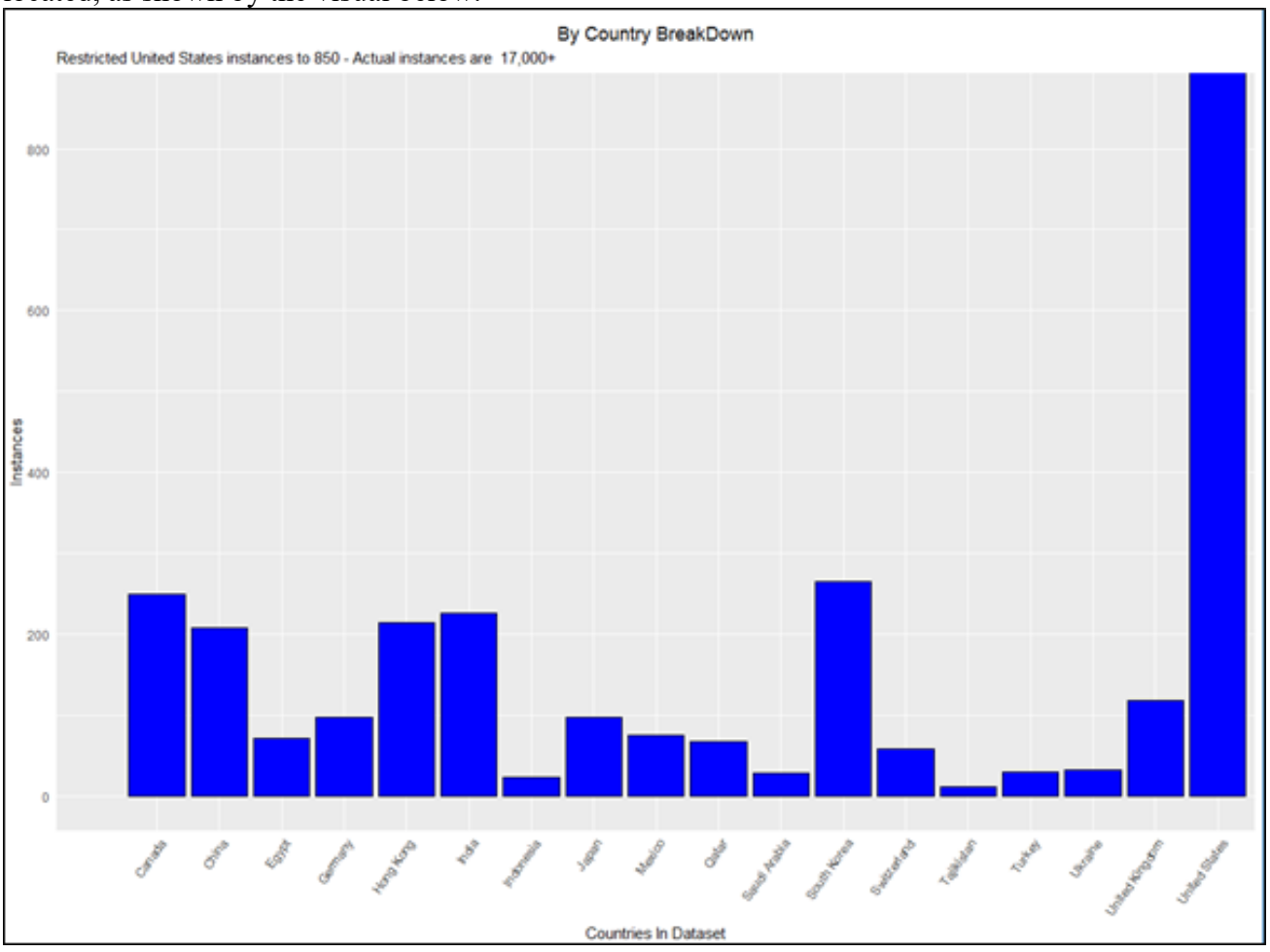
Table of Contents

Background and Understanding

The Hyatt Hotels Corporation (Hyatt) is one of the largest multinational hospitality brands in the world. Headquartered in the United States, Hyatt operates more than 600 hotels, resorts, and vacation properties in more than fifty countries.
Hyatt requested Orange Consulting (Orange) to provide Hyatt with key takeaways after analyzing survey data from guests staying at the company's hotels during February 2014. Driven by a few important business questions, our analysis will deliver actionable insight into the main factors that contribute to Net Promoter Score.

Executive summary

Originally we analyzed data to see if any abnormalities existed. We quickly observed a great degree of skewness when the data was broken out into countries where Hyatt properties are located, as shown by the visual below.



The team consensus was to restrict the dataset to United States.
Based on analysis by using descriptive and different modelling technique such as Association Rules, SVM, etc. Recommendations below were made.

- Services must be packaged and tailored to specific Tiers to keep revenue streams.
- Improvements need to be made in attracting more customer feedback
- Hotel should focus on maintaining a high standard for hotel rooms and hotel condition
- A fair number of revenue and frequency are from business travelers in the age group from age group 36 to 65. This group has to be catered to for continuous patronage.

Business Questions

General Questions

Hyatt would like to explore and understand the following business questions. Orange addresses these questions through the document, but the direct answer is provided in the Business Questions Insights. Please see the Table of Contents for location.

- What is the overall NPS?
    - Provide an overview of the count and percentage for each category of Promoter, Passive and Detractor.
- What is the NPS across different demographics
    - By US Region
    - By State
        - 31 states + Washington D.C.
    - By Hotel (provides same insight as by city)
    - By Gender
    - By Purpose of Visit
- What days of the week are most popular amongst guests?
- What attributes are best at predicting if a guest will be a promoter?

NPS Contributors

- What are the key drivers for Promoters / Detractors / Passives

Data load

The dataset that was used by the team had 19,342 observations (nrows) and 58 variables (ncol). Using str() on ProjectSurvey, below is the output:

```
> str(ProjectSurvey)
'data.frame':    19342 obs. of  58 variables:
 $ MARKET_GROUP_C          : Factor w/ 7 levels "CONVENTION","DISCOUNT",..: 3 3 3 2 3 3 3 3 2 3 ...
 $ ROOM_TYPE_CODE_C        : Factor w/ 164 levels "","1B2A","1BCN",..: 100 100 20 130 20 65 20 129 100 129 ...
 $ WALK_IN_FLG_C           : Factor w/ 2 levels "N","Y": 1 1 1 1 1 1 1 1 1 1 ...
 $ CHECK_IN_DATE_C         : Factor w/ 37 levels "1/11/2014","1/26/2014",..: 5 32 24 6 19 20 4 10 16 16 ...
 $ CHECK_OUT_DATE_C        : Factor w/ 72 levels "1/16/2014","1/28/2014",..: 15 10 37 55 23 51 30 17 43 22 ...
 $ LENGTH_OF_STAY_C        : int  10 7 14 22 5 7 7 7 10 7 ...
 $ NUMBER_OF_ROOMS_C       : int  1 1 1 1 1 1 1 1 1 1 ...
 $ ADULT_NUM_C             : int  2 2 2 2 2 2 2 2 2 2 ...
 $ CHILDREN_NUM_C          : int  NA NA NA NA NA NA NA NA NA NA ...
 $ POV_CODE_C              : Factor w/ 2 levels "BUSINESS","LEISURE": 1 1 1 2 1 1 1 1 1 1 ...
 $ QUOTED_RATE_C           : num  225 158 158 110 201 ...
 $ RESERVATION_DATE_R      : Factor w/ 295 levels "1/1/2014 0:00",..: 204 207 257 40 66 81 123 97 104 110 ...
 $ ENTRY_TIME_R            : Factor w/ 16580 levels "0:00:04","0:00:12",..: 13673 15568 11299 14066 15687 13616 195 190 13939 186 ...
 $ ENTRY_HOTEL_CODE_R      : Factor w/ 210 levels "ABEZB","ABQZA",..: 195 195 195 110 195 195 195 195 110 195 ...
 $ LAST_CHANGE_DATE_R      : Factor w/ 258 levels "1/1/2014 0:00",..: 180 183 223 124 64 79 121 95 105 108 ...
 $ ROOM_TYPE_CODE_R        : Factor w/ 144 levels "1B2A","1BCN",..: 56 88 88 88 17 17 17 17 56 56 ...
 $ STATE_R                 : Factor w/ 112 levels "","AB","AC","AE",..: 1 1 1 1 1 25 25 1 12 1 ...
 $ GUEST_COUNTRY_R         : Factor w/ 92 levels "","$CO","AC",..: 27 27 87 87 33 88 88 87 33 27 ...
 $ PACE_CATEGORY_R         : Factor w/ 9 levels "0-3 Days","15-21 Days",..: 8 8 8 8 8 8 8 4 6 6 4 ...
 $ PACE_R                  : Factor w/ 284 levels "-1","0","1","10",..: 138 134 96 24 9 276 237 247 245 240 ...
 $ REVENUE_USD_R           : num  2250 1108 2217 2429 1406 ...
 $ Guest_Country_H         : Factor w/ 86 levels "","American Samoa",..: 81 81 81 81 53 81 81 81 27 81 ...
 $ Gender_H                : Factor w/ 4 levels "","Female","Male",..: 2 3 2 3 3 3 3 3 3 2 ...
 $ Age_Range_H             : Factor w/ 8 levels "","18-25","26-35",..: 7 5 7 7 7 6 2 7 6 6 ...
 $ Language_H              : Factor w/ 13 levels "Arabic","Chinese Simplified",..: 4 4 4 4 4 4 4 4 6 4 ...
 $ Likelihood_Recommend_H  : int  10 10 10 10 10 9 10 10 10 10 ...
 $ Overall_Sat_H           : int  9 9 10 10 9 9 8 10 10 10 ...
 $ Guest_Room_H            : int  10 10 10 10 10 9 9 10 8 10 ...
 $ Tranquility_H           : int  10 9 10 NA 10 9 NA NA 10 10 ...
 $ Condition_Hotel_H       : int  9 10 10 9 10 9 10 10 9 10 ...
 $ Customer_SVC_H          : int  10 10 10 10 10 10 10 10 9 10 ...
 $ Staff_Cared_H           : int  10 10 10 NA 10 10 NA NA 10 10 ...
 $ Internet_Sat_H          : int  10 7 NA NA NA 6 NA NA 9 9 ...
 $ Check_In_H              : int  10 9 10 NA 9 10 NA NA 8 10 ...
 $ F.B_FREQ_H              : int  NA 1 NA NA NA 2 NA NA 3 3 ...
 $ F.B_Overall_Experience_H: num  NA 9 NA NA NA ...
 $ Property_ID_PL          : int  3275 3275 3275 3275 3275 3275 3275 3275 3275 3275 ...
 $ Hotel.Name.Long_PL      : Factor w/ 195 levels "Andaz Liverpool Street, London",..: 185 185 185 185 185 185 185 185 185 185 ...
 $ Hotel.Name.Short_PL     : Factor w/ 195 levels "AZ Liverpool Street",..: 175 175 175 175 175 175 175 175 175 175 ...
 $ Award.Category_PL       : int  2 2 2 2 2 2 2 2 2 2 ...
 $ City_PL                 : Factor w/ 153 levels "Addison","Ahmedabad",..: 138 138 138 138 138 138 138 138 138 138 ...
 $ Country_PL              : Factor w/ 18 levels "Canada","China",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Ops.Region_PL           : Factor w/ 4 levels "Americas","ASPAC",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Property.Latitude_PL    : num  28 28 28 28 28 ...
 $ Property.Longitude_PL   : num  34.4 34.4 34.4 34.4 34.4 ...
 $ Currency_PL             : Factor w/ 18 levels "CAD","CHF","CNY",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ Dom.Int.1_PL            : Factor w/ 2 levels "Domestic","International": 2 2 2 2 2 2 2 2 2 2 ...
 $ Brand_PL                : Factor w/ 7 levels "Andaz","Grand Hyatt",..: 6 6 6 6 6 6 6 6 6 6 ...
 $ Club.Type_PL            : Factor w/ 4 levels "","Club","Grand Club",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ Region_PL               : Factor w/ 4 levels "Americas","Asia Pacific",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ Category_PL             : Factor w/ 2 levels "Full Service",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Type_PL                 : Factor w/ 5 levels "Business","Convention",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ Class_PL                : Factor w/ 3 levels "Luxury Class",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Location_PL             : Factor w/ 4 levels "Airport","Resort",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Relationship_PL         : Factor w/ 3 levels "Franchised","Managed",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ GP_Tier                 : Factor w/ 11 levels "","CARD","Courtesy",..: 7 1 7 5 1 1 1 1 11 7 ...
 $ Booking_Channel         : Factor w/ 3 levels "Electronic Distribution",..: 3 3 3 2 3 3 3 3 2 3 ...
 $ NPS_Type                : Factor w/ 3 levels "Detractor","Passive",..: 3 3 3 3 3 3 3 3 3 3 ...
```

After restricting dataset to Hyatt operations in the United States, we reduced the data set 17,439 observations(nrows), but our team added additional variables. We also changed the names of the variables in order to more easily identify them throughout the analysis. While we did not explore all of the variables venues, we feel that our variable selection fairly represents the data set and offers important insight. Below is the updated and formatted data set used for analysis.

```
> str(ProjectSurveyUSA)
'data.frame':    17439 obs. of  66 variables:
 $ MarketGroup       : Factor w/ 7 levels "CONVENTION","DISCOUNT",..: 4 4 5 4 1 1 2 1 4 7 ...
 $ RTC_CheckOut      : Factor w/ 164 levels "","1B2A","1BCN",..: 161 162 160 162 161 160 162 162 160 161 ...
 $ WalkIn            : Factor w/ 2 levels "N","Y": 1 1 1 1 1 1 1 1 1 1 ...
 $ CheckInDate       : Date, format: "2014-02-28" "2014-02-14" "2014-02-03" "2014-02-15" ...
 $ CheckOut          : Date, format: "2014-03-04" "2014-02-15" "2014-02-06" "2014-02-17" ...
 $ LOS               : num  4 1 3 2 3 2 1 5 2 5 ...
 $ NumRooms          : num  1 1 1 1 1 1 1 1 1 1 ...
 $ Adult             : num  1 2 1 1 1 1 2 2 2 2 ...
 $ Children          : num  1.65 1.65 1.65 1.65 1.65 ...
 $ PurposeOfVisit    : Factor w/ 2 levels "BUSINESS","LEISURE": 1 2 1 2 1 1 1 1 1 1 ...
 $ QuotedRate        : num  229 340 183 209 209 ...
 $ ResvDate          : Date, format: "2014-01-28" "2014-01-28" "2014-01-28" "2014-01-29" ...
 $ ResvTime          : Factor w/ 16580 levels "0:00:04","0:00:12",..: 11302 11666 13044 742 1483 1837 4607 4694 350 14938 ...
 $ EntryHotelCode    : Factor w/ 210 levels "ABEZB","ABQZA",..: 165 165 165 164 165 165 150 165 165 202 ...
 $ ResvDateChange    : Date, format: "2014-01-28" "2014-01-31" "2014-01-28" "2014-01-30" ...
 $ RTC_Book          : Factor w/ 144 levels "1B2A","1BCN",..: 139 142 139 140 139 140 140 140 140 139 ...
 $ State             : Factor w/ 112 levels "","AB","AC","AE",..: 52 25 95 78 76 76 104 71 25 25 ...
 $ GuestCountryOrigin: Factor w/ 92 levels "","$CO","AC",..: 88 88 88 14 88 88 88 88 88 88 ...
 $ FutureResvDaysCat : Factor w/ 9 levels "0-3 Days","15-21 Days",..: 4 2 5 2 2 2 7 2 1 3 ...
 $ FutureResvDays    : num  183 81 241 81 103 114 26 103 113 167 ...
 $ Revenue           : num  916 340 550 418 627 ...
 $ GuestCountryHome  : Factor w/ 86 levels "","American Samoa",..: 83 83 83 12 83 83 83 83 83 83 ...
 $ Gender            : Factor w/ 4 levels "","Female","Male",..: 2 3 3 3 3 2 2 3 3 3 ...
 $ AgeRange          : Factor w/ 8 levels "","18-25","26-35",..: 5 5 5 6 5 4 5 6 5 4 ...
 $ Language          : Factor w/ 13 levels "Arabic","Chinese Simplified",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ LR_Score          : int  10 9 5 9 10 10 10 9 10 10 ...
 $ OS_Score          : num  10 8 6 9 10 10 10 9 9 9 ...
 $ GR_Score          : num  10 7 10 10 10 10 10 10 9 10 ...
 $ T_Score           : num  8.77 8.77 8.77 10 8.77 ...
 $ HC_Score          : num  10 9 8 10 9 10 10 9 9 10 ...
 $ CS_Score          : num  10 8 3 10 10 ...
 $ SC_Score          : num  9.04 9.04 9.04 9 9.04 ...
 $ IS_Score          : num  8.75 8.75 8.75 9 8.75 ...
 $ CI_Score          : num  9.28 9.28 9.28 10 9.28 ...
 $ FB_Frequency      : num  1.47 1.47 1.47 2 1.47 ...
 $ FBE_Score         : num  8.63 8.63 8.63 9.5 8.63 ...
 $ PropertyID        : int  3328 3328 3328 3328 3328 3328 3328 3328 3328 3328 ...
 $ HotelName         : Factor w/ 195 levels "Andaz Liverpool Street, London",..: 176 176 176 176 176 176 176 176 176 176 ...
 $ HotelAbbrv        : Factor w/ 195 levels "AZ Liverpool Street",..: 167 167 167 167 167 167 167 167 167 167 ...
 $ AwardCat          : int  5 5 5 5 5 5 5 5 5 5 ...
 $ HotelCity         : Factor w/ 153 levels "Addison","Ahmedabad",..: 113 113 113 113 113 113 113 113 113 113 ...
 $ HotelCountry      : Factor w/ 18 levels "Canada","China",..: 18 18 18 18 18 18 18 18 18 18 ...
 $ OperatingRegion   : Factor w/ 4 levels "Americas","ASPAC",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Latitude          : num  28.4 28.4 28.4 28.4 28.4 ...
 $ Longitude         : num  -81.5 -81.5 -81.5 -81.5 -81.5 ...
 $ Currency          : Factor w/ 18 levels "CAD","CHF","CNY",..: 18 18 18 18 18 18 18 18 18 18 ...
 $ ValidValues       : Factor w/ 2 levels "Domestic","International": 1 1 1 1 1 1 1 1 1 1 ...
 $ HotelBrand        : Factor w/ 7 levels "Andaz","Grand Hyatt",..: 6 6 6 6 6 6 6 6 6 6 ...
 $ Club              : Factor w/ 4 levels "","Club","Grand Club",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ Region            : Factor w/ 4 levels "Americas","Asia Pacific",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Category          : Factor w/ 2 levels "Full Service",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Type              : Factor w/ 5 levels "Business","Convention",..: 4 4 4 4 4 4 4 4 4 4 ...
 $ Class             : Factor w/ 3 levels "Luxury Class",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Location          : Factor w/ 4 levels "Airport","Resort",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Relationship      : Factor w/ 3 levels "Franchised","Managed",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ GPTier            : Factor w/ 11 levels "","CARD","Courtesy",..: 7 1 11 1 7 7 11 7 6 7 ...
 $ Channel           : Factor w/ 3 levels "Electronic Distribution",..: 1 1 1 2 1 1 2 1 1 3 ...
 $ NPS               : Factor w/ 3 levels "Detractor","Passive",..: 3 3 1 3 3 3 3 3 3 3 ...
 $ CheckInDay        : chr  "Friday" "Friday" "Monday" "Saturday" ...
 $ CheckOutDay       : chr  "Tuesday" "Saturday" "Thursday" "Monday" ...
 $ Coords            : num  28.4 28.4 28.4 28.4 28.4 ...
 $ stateName         : chr  "florida" "florida" "florida" "florida" ...
 $ USRegion          : chr  "Southeast" "Southeast" "Southeast" "Southeast" ...
 $ Promoter          : num  1 1 0 1 1 1 1 1 1 1 ...
 $ Detractor         : num  0 0 1 0 0 0 0 0 0 0 ...
 $ Passive           : num  0 0 0 0 0 0 0 0 0 0 ...
```

Description of the data set

The data set contains survey data from guests staying at Hyatt hotels during February 2014. One of the most important data points across all observations is called Net Promoter Score (NPS), which measures customer experience and can help predict business growth. It provides a measurement of customer satisfaction.
Survey respondents are grouped as:

- 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 the brand and impede growth through negative word-of-mouth.

§ Sources - https://www.netpromoter.com/know

While many of the variables will be used in our analysis, the following changes were made prior to beginning:

- Columns with a large number of NAs were given various treatments, including converting to the mean
- "Children_Num_C" were discarded from the beginning based on the numbers of NAs (15,874) within its column of data or Internet_Sat_H which had 9,649 NAs
- "Check_in_H" – 6,423
- F.B_FREQ_H - 8,605
- F.B_Overall_Exprience – 8,605
- Tranquility – 6,497
- Staff_cared - 6,422

The additional columns added helped in the data manipulation and explanation of the NPS score

Data Cleansing

In order to run some descriptive analysis, the team had to address the NAs by converting factors to numeric values, factors to data, and factors to characters where needed. Some of these methods were achieved by creating functions and data manipulations.

Before →

```
summary(ProjectSurvey)
   MARKET_GROUP_C ROOM_TYPE_CODE_C WALK_IN_FLG_C  CHECK_IN_DATE_C   CHECK_OUT_DATE_C LENGTH_OF_STAY_C NUMBER_OF_ROOMS_C  ADULT_NUM_C    CHILDREN_NUM_C     POV_CODE_C   QUOTED_RATE_C          RESERVATION_DATE_R
 CONVENTION:3847   KING   :8105     N:19329       2/14/2014: 1355   2/23/2014: 1481   Min.   : 1.000   Min.   :1        Min.   :1.000   Min.   :1.000   BUSINESS:15678   Min.   :      0   2/3/2014 0:00 :  617
 DISCOUNT  :4519   DDBL   :4011     Y:   13       2/21/2014: 1133   2/16/2014: 1401   1st Qu.: 1.000   1st Qu.:1        1st Qu.:1.000   1st Qu.:1.000   LEISURE : 3664   1st Qu.:     92   2/4/2014 0:00 :  587
 OTHER     : 936   1BKN   : 986                2/28/2014: 1083   2/9/2014 : 1086   Median : 2.000   Median :1        Median :1.000   Median :2.000                    Median :    119   2/10/2014 0:00:  582
 RACK      :5134   QNQN   : 856                2/15/2014: 1005   2/21/2014:  831   Mean   : 2.408   Mean   :1        Mean   :1.432   Mean   :1.622                    Mean   :   5340   2/6/2014 0:00 :  580
 SPECIAL   :1402   QUEN   : 314                2/22/2014:  919   2/28/2014:  827   3rd Qu.: 3.000   3rd Qu.:1        3rd Qu.:2.000   3rd Qu.:2.000                    3rd Qu.:    169   2/11/2014 0:00:  546
 VOLUME    :2903   DLXX   : 312                2/7/2014 :  895   2/15/2014:  771   Max.   :97.000   Max.   :1        Max.   :8.000   Max.   :4.000                    Max.   :2213984   2/5/2014 0:00 :  522
 WHOLESALE : 601   (Other):4758                (Other)  :12952   (Other)  :12945   NA's   :97       NA's   :51       NA's   :7       NA's   :17654                                      (Other)       :15908
  ENTRY_TIME_R   ENTRY_HOTEL_CODE_R      LAST_CHANGE_DATE_R ROOM_TYPE_CODE_R    STATE_R          GUEST_COUNTRY_R    PACE_CATEGORY_R     PACE_R      REVENUE_USD_R          Guest_Country_H
 11:18:16:    5   OMGDS  :11569      2/3/2014 0:00 :  634   KING   :9145          : 3342   UNITED STATES :15448   0-3 Days  :5084   0      : 1886   Min.   :    0.0   USA           :17221
 11:21:42:    5   OMARO  : 3418      2/10/2014 0:00:  626   DDBL   :3895   TX     : 1837                 : 1964   8-14 Days :2758   1      : 1381   1st Qu.:  115.0   Canada        :  359
 14:47:25:    5   MARRO  : 1079      2/6/2014 0:00 :  622   1BKN   :1053   CA     : 1203   CANADA        :  404   31-60 Days:2726   2      :  941   Median :  210.0   Korea         :  246
 10:18:58:    4   BUSPH  :  175      2/5/2014 0:00 :  605   QNQN   : 854   FL     : 1058   SOUTH KOREA   :  233   4-7 Days  :2672   3      :  876   Mean   :  339.7   United Kingdom:  178
 10:19:06:    4   EURRC  :  135      2/4/2014 0:00 :  587   QUEN   : 486   IL     :  946   JAPAN         :  146   15-21 Days:2141   4      :  808   3rd Qu.:  398.0   Germany       :  150
 10:25:50:    4   INDRO  :   93      2/12/2014 0:00:  585   DLXN   : 357   NY     :  679   UNITED KINGDOM:  141   22-30 Days:2063   5      :  683   Max.   :18415.0   Japan         :  148
 (Other) :19315   (Other): 2873      (Other)       :15683   (Other):3552   (Other):10277   (Other)       : 1006   (Other)   :1898   (Other):12767                      (Other)       : 1040
          Gender_H      Age_Range_H              Language_H    Likelihood_Recommend_H Overall_Sat_H    Guest_Room_H    Tranquility_H   Condition_Hotel_H Customer_SVC_H  Staff_Cared_H   Internet_Sat_H
                :  213   46-55 :5783   English          :18302   Min.   : 1.0          Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.00   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
 Female         : 8511   36-45 :4405   Korean           :  238   1st Qu.: 8.0          1st Qu.: 8.000   1st Qu.: 8.000   1st Qu.: 8.000   1st Qu.: 8.00   1st Qu.: 9.000   1st Qu.: 9.000   1st Qu.: 8.000
 Male           :10271   56-65 :4027   German           :  208   Median : 9.0          Median : 9.000   Median : 9.000   Median : 9.000   Median : 9.00   Median :10.000   Median :10.000   Median : 9.000
 Prefer not to answer:  347   26-35 :2714   Japanese ja_JP   :  152   Mean   : 8.7          Mean   : 8.666   Mean   : 8.792   Mean   : 8.777   Mean   : 8.94   Mean   : 9.098   Mean   : 9.025   Mean   : 8.681
                      66-75 :1262   Spanish (Americas):  124   3rd Qu.:10.0          3rd Qu.:10.000   3rd Qu.:10.000   3rd Qu.:10.000   3rd Qu.:10.00   3rd Qu.:10.000   3rd Qu.:10.000   3rd Qu.:10.000
                      18-25 : 560   Chinese Simplified:   95   Max.   :10.0          Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.00   Max.   :10.000   Max.   :10.000   Max.   :10.000
                      (Other): 591   (Other)          :  223                         NA's   :21       NA's   :188      NA's   :7182     NA's   :212     NA's   :276      NA's   :7108     NA's   :10633
  Check_In_H      F.B_FREQ_H    F.B_Overall_Experience_H Property_ID_PL                                Hotel.Name.Long_PL      Hotel.Name.Short_PL Award.Category_PL        City_PL                Country_PL
 Min.   : 1.000   Min.   :1.000   Min.   : 0.000          Min.   : 3010   Hyatt Regency Jacksonville Riverfront   :  579   HR Jacksonville :  579   Min.   :1.000   Washington   : 1088   United States:17467
 1st Qu.: 9.000   1st Qu.:1.000   1st Qu.: 8.000          1st Qu.: 3496   Hyatt Regency O'Hare                   :  579   HR O'Hare       :  579   1st Qu.:1.000   Jacksonville:  667   South Korea  :  265
 Median :10.000   Median :1.000   Median : 9.000          Median : 3583   Grand Hyatt Washington                 :  484   GH Washington   :  484   Median :2.000   Orlando      :  667   Canada       :  250
 Mean   : 9.252   Mean   :1.498   Mean   : 8.637          Mean   : 5067   Hyatt Regency Washington on Capitol Hill:  425   HR Washington   :  425   Mean   :2.227   Rosemont     :  579   India        :  226
 3rd Qu.:10.000   3rd Qu.:2.000   3rd Qu.:10.000          3rd Qu.: 3694   Hyatt Regency Grand Cypress            :  414   HR Grand Cypress:  414   3rd Qu.:3.000   Arlington    :  473   Hong Kong    :  215
 Max.   :10.000   Max.   :4.000   Max.   :10.000          Max.   :39001   Hyatt Regency Santa Clara              :  351   HR Santa Clara  :  351   Max.   :9.000   Atlanta      :  440   China        :  208
 NA's   :7112     NA's   :9640    NA's   :9640                            (Other)                                :16510   (Other)         :16510                   (Other)      :15428   (Other)      :  711
 Ops.Region_PL   Property.Latitude_PL Property.Longitude_PL  Currency_PL          Dom.Int.l_PL          Brand_PL          Club.Type_PL          Region_PL          Category_PL          Type_PL
 Americas:17859   Min.   :-6.909       Min.   :-123.12       USD    :17467   Domestic     :17467   Andaz       :  118              :14818   Americas          :17792   Full Service  : 6490   Business      : 2207
 ASPAC   : 1035   1st Qu.:32.751       1st Qu.: -96.77       KRW    :  265   International: 1875   Grand Hyatt :  929   Club       :   94   Asia Pacific      : 1035   Select Service:12852   Convention    : 2512
 EAME    :  330   Median :35.829       Median : -84.36       CAD    :  250                         Hyatt       :  743   Grand Club :  681   Europe            :  348                          Franchise     :  753
 Europe  :  118   Mean   :35.781       Mean   : -75.49       INR    :  226                         Hyatt House : 2206   Regency Club: 3749   Middle East & Africa:  167                          Resort        : 1018
                  3rd Qu.:39.940       3rd Qu.: -77.56       HKD    :  215                         Hyatt Place :10646                                                                                Select Service:12852
                  Max.   :51.517       Max.   : 139.69       CNY    :  208                         Hyatt Regency: 4099
                                                             (Other):  711                         Park Hyatt  :  601
          Class_PL       Location_PL      Relationship_PL     GP_Tier                Booking_Channel      NPS_Type
 Luxury Class      : 1648   Airport : 3963   Franchised:8514   Gold    :7157   Electronic Distribution:11166   Detractor: 2169
 Upper Upscale Class: 4842   Resort  : 1115   Managed   :6555           :5955   Global Contact Center  : 5457   Passive  : 3603
 Upscale Class     :12852   Suburban:10537   Owned     :4273   GOLD    :3061   Hotel                  : 2719   Promoter :13570
                            Urban   : 3727                    Platinum:2084
                                                              Diamond : 759
                                                              PLAT    : 244
                                                              (Other) :  82
```

After →

```
> summary(ProjectSurveyUSA)
    MarketGroup    RTC_CheckOut  WalkIn      CheckInDate           CheckOut               LOS            NumRooms     Adult           Children       PurposeOfVisit   QuotedRate       ResvDate               ResvTime
 CONVENTION:3720   KING    :7540  N:17426  Min.   :2014-02-01  Min.   :2014-02-01  Min.   : 1.000  Min.   :1   Min.   :1.000  Min.   :1.000  BUSINESS:14086  Min.   :  0.0  Min.   :2012-08-21  11:21:42:     5
 DISCOUNT  :4353   DDBL    :3932  Y:   13  1st Qu.:2014-02-09  1st Qu.:2014-02-11  1st Qu.: 1.000  1st Qu.:1   1st Qu.:1.000  1st Qu.:1.649  LEISURE : 3353  1st Qu.: 89.0  1st Qu.:2014-01-15  14:47:25:     5
 OTHER     : 901   1KKN    : 984           Median :2014-02-15  Median :2014-02-17  Median : 2.000  Median :1   Median :1.000  Median :1.649                  Median :113.0  Median :2014-02-01  10:18:58:     4
 RACK      :4514   QNQN    : 855           Mean   :2014-02-15  Mean   :2014-02-17  Mean   : 2.366  Mean   :1   Mean   :1.423  Mean   :1.650                  Mean   :125.2  Mean   :2014-01-22  10:19:06:     4
 SPECIAL   :1322   DLXN    : 297           3rd Qu.:2014-02-22  3rd Qu.:2014-02-24  3rd Qu.: 3.000  3rd Qu.:1   3rd Qu.:2.000  3rd Qu.:1.649                  3rd Qu.:152.2  3rd Qu.:2014-02-11  10:25:50:     4
 VOLUME    :2389   QUEN    : 267           Max.   :2014-02-28  Max.   :2014-05-14  Max.   :97.000  Max.   :1   Max.   :8.000  Max.   :4.000                  Max.   :989.0  Max.   :2014-02-28  11:04:28:     4
 WHOLESALE : 240   (Other):3564                                                                                                                                                                (Other) :17413
 EntryHotelCode  ResvDateChange          RTC_Book        State          GuestCountryOrigin  FutureResvDaysCat FutureResvDays     Revenue                GuestCountryHome                   Gender              AgeRange
 OMGDS   :10743  Min.   :2013-03-05   KING    :8215          :2055   UNITED STATES :15099   0-3 Days   :4660   Min.   :  1.0   Min.   :    0.0   USA              :16899                          : 196   46-55   :5295
 OMARO   : 3346  1st Qu.:2014-01-21   DDBL    :3824   TX     :1815                 : 1707   31-60 Days:2477   1st Qu.: 26.0   1st Qu.:  109.7   Canada           :  198   Female               :7950   36-45   :3826
 MARRO   : 1053  Median :2014-02-04   1KKN    :1050   CA     :1157   CANADA        :  278   8-14 Days :2430   Median :143.0   Median :  198.0                    :   79   Male                 :8968   56-65   :3763
 CHIRO   :   52  Mean   :2014-01-27   QNQN    : 853   FL     :1033   UNITED KINGDOM:   49   4-7 Days  :2367   Mean   :132.6   Mean   :  311.7   United Kingdom:   34   Prefer not to answer: 325   26-35   :2304
 ORLZA   :   49  3rd Qu.:2014-02-13   QUEN    : 427   IL     : 911   GERMANY       :   33   15-21 Days:1917   3rd Qu.:221.0   3rd Qu.:  368.0   Germany          :   28                                66-75   :1190
 FLLZA   :   42  Max.   :2014-03-19   DLXN    : 357   MI     : 657   MEXICO        :   32   22-30 Days:1861   Max.   :284.0   Max.   :18415.0   Mexico           :   25                                18-25   : 501
 (Other): 2154                        (Other):2713   (Other):9811   (Other)       :  241   (Other)    :1727                                     (Other)          :  176                                (Other): 560
            Language       LR_Score         OS_Score         GR_Score         T_Score          HC_Score          CS_Score          SC_Score          IS_Score          CI_Score        FB_Frequency     FBE_Score
 English           :17249  Min.   : 1.000  Min.   : 1.00   Min.   : 1.00   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   :1.000   Min.   : 0.00
 Spanish (Americas):    53  1st Qu.: 8.000  1st Qu.: 8.00   1st Qu.: 8.00   1st Qu.: 8.766   1st Qu.: 8.000   1st Qu.: 9.000   1st Qu.: 9.000   1st Qu.: 8.747   1st Qu.: 9.281   1st Qu.:1.000   1st Qu.: 8.63
 German            :    47  Median : 9.000  Median : 9.00   Median : 9.00   Median : 8.766   Median : 9.000   Median :10.000   Median : 9.035   Median : 8.747   Median : 9.281   Median :1.475   Median : 8.63
 French            :    15  Mean   : 8.688  Mean   : 8.66   Mean   : 8.79   Mean   : 8.766   Mean   : 8.936   Mean   : 9.112   Mean   : 9.035   Mean   : 8.747   Mean   : 9.281   Mean   :1.475   Mean   : 8.63
 Japanese ja_JP    :    14  3rd Qu.:10.000  3rd Qu.:10.00   3rd Qu.:10.00   3rd Qu.:10.000   3rd Qu.:10.000   3rd Qu.:10.000   3rd Qu.:10.000   3rd Qu.: 9.000   3rd Qu.:10.000   3rd Qu.:1.475   3rd Qu.: 9.00
 Korean            :    12  Max.   :10.000  Max.   :10.00   Max.   :10.00   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :4.000   Max.   :10.00
 (Other)           :    49
   PropertyID                                 HotelName              HotelAbbrv           AwardCat             HotelCity               HotelCountry      OperatingRegion     Latitude        Longitude          Currency
 Min.   : 3010   Hyatt Regency Jacksonville Riverfront   :  579   HR Jacksonville :  579   Min.   :1.000   Washington   : 1086   United States:17439   Americas:17439   Min.   :25.78   Min.   :-122.27   USD    :17439
 1st Qu.: 3498   Hyatt Regency O'Hare                    :  579   HR O'Hare       :  579   1st Qu.:1.000   Jacksonville:  667   Canada       :     0   ASPAC   :    0   1st Qu.:32.88   1st Qu.: -96.82   CAD    :     0
 Median : 3577   Grand Hyatt Washington                  :  483   GH Washington   :  483   Median :2.000   Orlando      :  667   China        :     0   EAME    :    0   Median :36.05   Median : -84.45   CHF    :     0
 Mean   : 4604   Hyatt Regency Washington on Capitol Hill:  424   HR Washington   :  424   Mean   :2.038   Rosemont     :  579   Egypt        :     0   Europe  :    0   Mean   :35.97   Mean   : -88.67   CNY    :     0
 3rd Qu.: 3674   Hyatt Regency Grand Cypress             :  414   HR Grand Cypress:  414   3rd Qu.:2.000   Arlington    :  473   Germany      :     0                    3rd Qu.:39.92   3rd Qu.: -80.16   EGP    :     0
 Max.   :34001   Hyatt Regency Santa Clara               :  351   HR Santa Clara  :  351   Max.   :8.000   Atlanta      :  440   Hong Kong    :     0                    Max.   :44.86   Max.   : -71.11   EUR    :     0
                 (Other)                                 :14609   (Other)         :14609                   (Other)      :13527   (Other)      :     0                                                        (Other):     0
      ValidValues          HotelBrand               Club                      Region                Category                    Type                         Class                  Location          Relationship
 Domestic     :17439   Andaz         :    0                :13989   Americas              :17439   Full Service  : 4611   Business         : 1005   Luxury Class         :  910   Airport :3958   Franchised:8494
 International:     0   Grand Hyatt   :  731   Club        :    0   Asia Pacific          :    0   Select Service:12828   Convention       : 2162   Upper Upscale Class: 3701   Resort  : 976   Managed   :5101
                        Hyatt         :  507   Grand Club  :  483   Europe                :    0                          Franchise        :  751   Upscale Class        :12828   Suburban:9940   Owned     :3844
                        Hyatt House   : 2198   Regency Club: 2967   Middle East & Africa:    0                          Resort           :  693                                 Urban   :2565
                        Hyatt Place   :10630                                                                             Select Service:12828
                        Hyatt Regency: 3194
                        Park Hyatt    :  179
     GPTier                    Channel                  NPS            CheckInDay           CheckOutDay             Coords          stateName            USRegion              Promoter         Detractor          Passive
 Gold    :6741   Electronic Distribution:10211   Detractor: 2006   Length:17439         Length:17439         Min.   :25.78   Length:17439         Length:17439         Min.   :0.000   Min.   :0.000   Min.   :0.000
         :5441   Global Contact Center  : 5073   Passive  : 3208   Class :character   Class :character   1st Qu.:32.88   Class :character   Class :character   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
 GOLD    :2517   Hotel                  : 2155   Promoter :12225   Mode  :character   Mode  :character   Median :36.05   Mode  :character   Mode  :character   Median :1.000   Median :0.000   Median :0.000
 Platinum:1888                                                                                             Mean   :35.97                                            Mean   :0.701   Mean   :0.115   Mean   :0.184
 Diamond : 603                                                                                             3rd Qu.:39.92                                            3rd Qu.:1.000   3rd Qu.:0.000   3rd Qu.:0.000
 PLAT    : 196                                                                                             Max.   :44.86                                            Max.   :1.000   Max.   :1.000   Max.   :1.000
 (Other) :  53
```

Data Quality Checks

To double check data integrity for NAs, we used the below command to find which columns had NA's: names(which(sapply(ProjectSurveyUSA, anyNA)))
We then created a text editor to validate the definition of variable while working to clean and check the quality of the data with :

- #Creating a Data Frame Using the R Data Editor
- Glossary<- data.frame(Column = character(), Defintion = character(), NewName = character())
- Glossary<- edit(Glossary)
- View(Glossary)

	Column	Defintion	NewName
1	MARKET_GROUP_C	Sub-group of major markets	MarketGroup
2	ROOM_TYPE_CODE_C	Hyatt standard room type code of the guest's room upo...	RTC_CheckOut
3	WALK_IN_FLG_C	Flag indicating a walk-in	WalkIn
4	CHECK_IN_DATE_C	Check in date; for WALK status adjusted to the first in-h...	CheckInDate
5	CHECK_OUT_DATE_C	Check out date	CheckOut
6	LENGTH_OF_STAY_C	Length of stay	LOS
7	NUMBER_OF_ROOMS_C	Number of rooms occupied. Valid values are 0 or 1; for c...	NumRooms
8	ADULT_NUM_C	Number of adults on the last day of the stay	Adult
9	CHILDREN_NUM_C	Check in length	Children
10	POV_CODE_C	Purpose of visit	POVCode
11	QUOTED_RATE_C	Quoted rate derived from Reservation table	QuotedRate
12	RESERVATION_DATE_R	Date when the booking was made	ResvDate
13	ENTRY_TIME_R	Time the reservation was made	ResvTime
14	ENTRY_HOTEL_CODE_R	Entry hotel code	EntryHotelCode
15	LAST_CHANGE_DATE_R	Date of last change to the reservation	ResvDateChange
16	ROOM_TYPE_CODE_R	Hyatt standard room type code of the guest's room as p...	RTC_Book
17	STATE_R	State of the party making the reservation	State
18	GUEST_COUNTRY_R	The country to which the actual guest belongs to. Differ...	GuestCountryOrigin
19	PACE_CATEGORY_R	Categorizes the number of nights in the future the reser...	FutureResvDaysCat
20	PACE_R	Pace identifies the number of nights in the future the re...	FutureResvDays
21	REVENUE_USD_R	Total revenue from the reservation in USD	Revenue

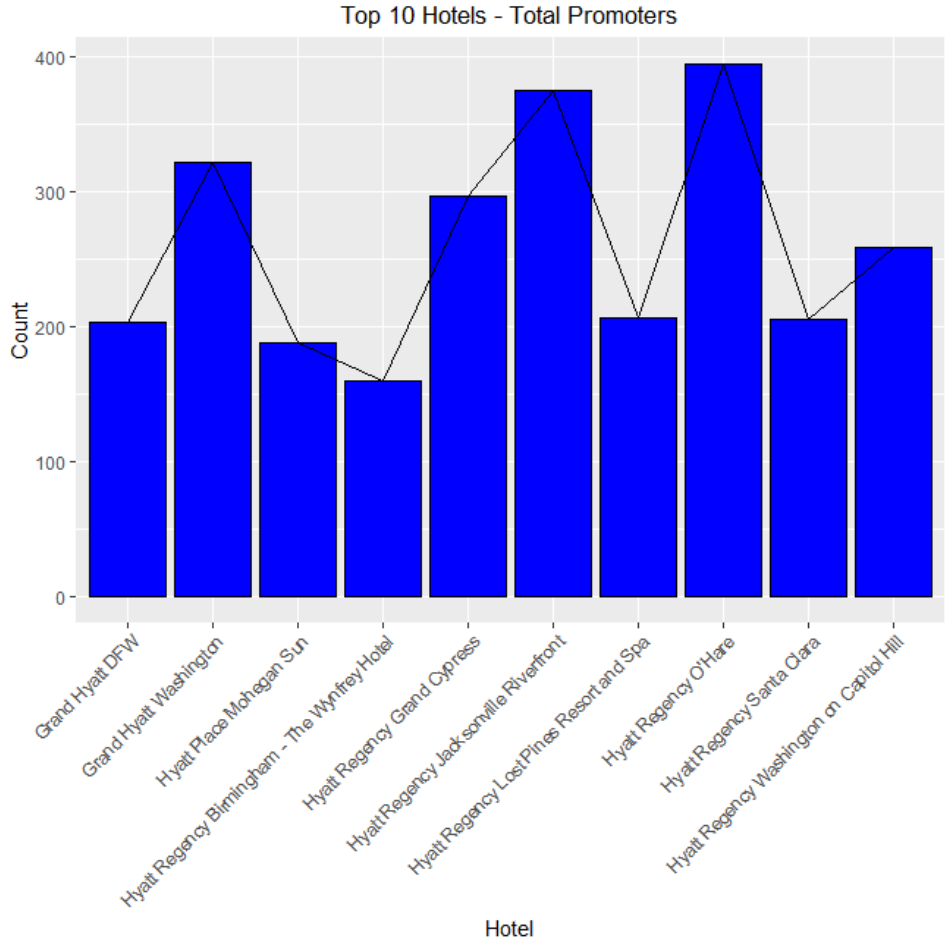
Descriptive Statistics

After the data had been cleansed and checked for quality, we were able to gather some initial descriptive statistics on the data set. Before diving into specifics around NPS, we wanted to look at the ten hotels with the:

a. most promoters
b. fewest promoters
c. most detractors/passives
d. highest percentage of promoters among survey takers
e. lowest percentage of promoters among survey takers

Ten hotels with the most promoters:

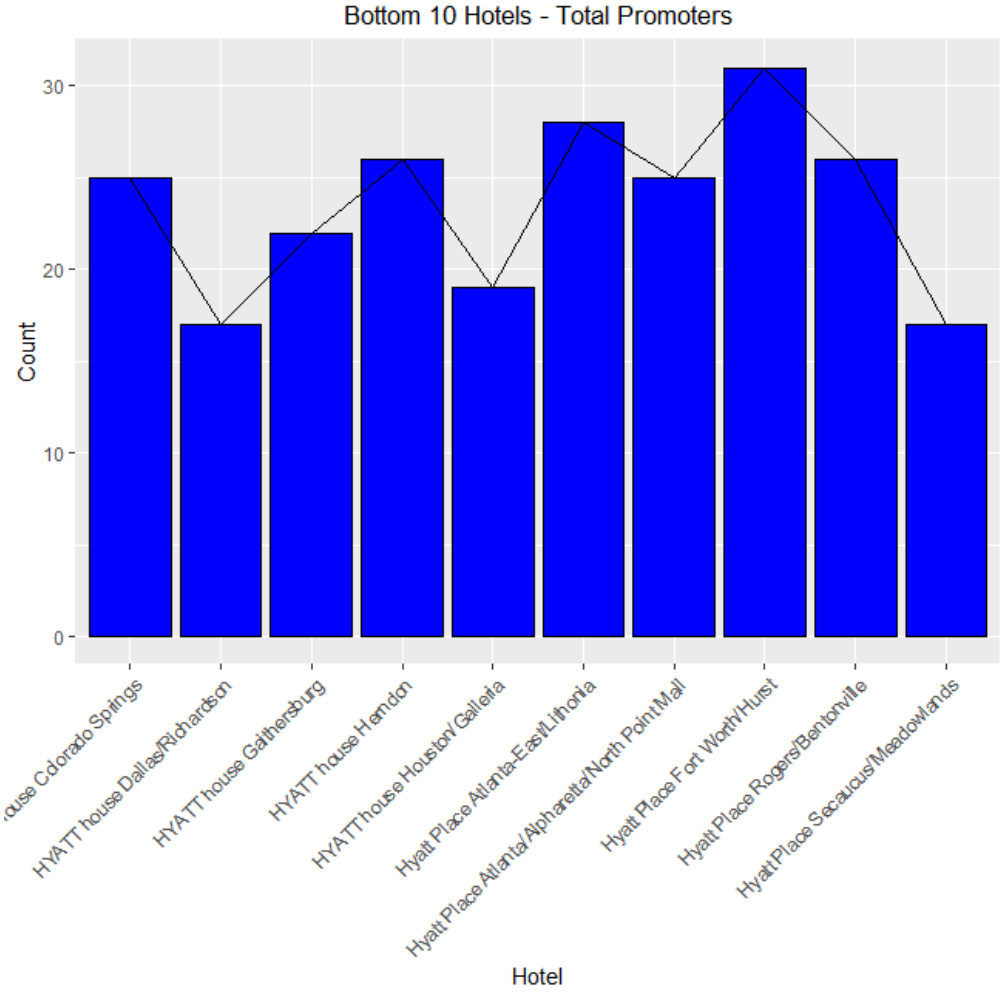
	HotelName	Promoter	n
1	Hyatt Regency O'Hare	1	394
2	Hyatt Regency Jacksonville Riverfront	1	375
3	Grand Hyatt Washington	1	321
4	Hyatt Regency Grand Cypress	1	296
5	Hyatt Regency Washington on Capitol Hill	1	258
6	Hyatt Regency Lost Pines Resort and Spa	1	206
7	Hyatt Regency Santa Clara	1	205
8	Grand Hyatt DFW	1	203
9	Hyatt Place Mohegan Sun	1	188
10	Hyatt Regency Birmingham - The Wynfrey Hotel	1	160



At this point, it is worthwhile to remember that we are looking at a group of 17,000+ surveys across 170 hotels. With this in mind, we see that each of the top ten hotels have a significant sample size (all above 160). The Hyatt Regency O'Hare led the way with 394 surveys resulting in promoters.

Ten hotels with the fewest promoters:

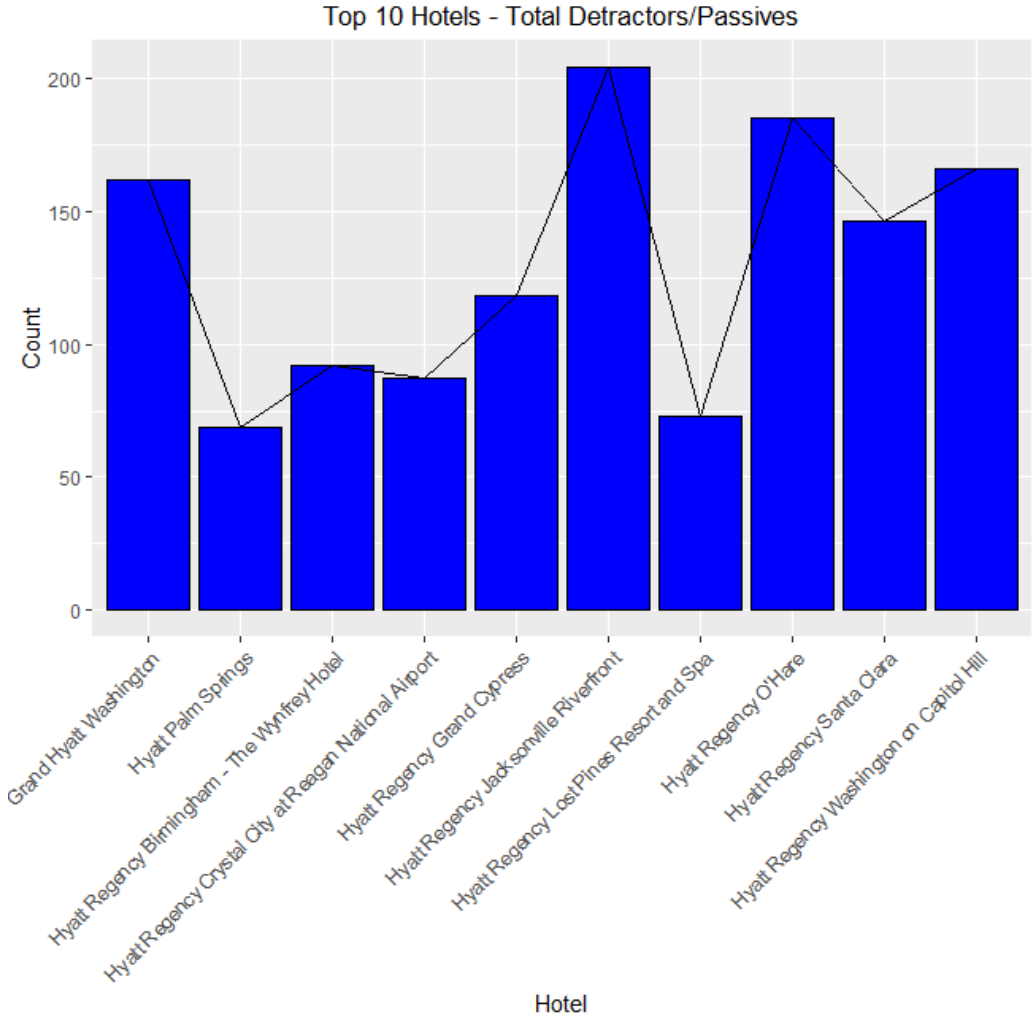
	HotelName	Promoter	n
161	Hyatt Place Fort Worth/Hurst	1	31
162	Hyatt Place Atlanta-East/Lithonia	1	28
163	HYATT house Herndon	1	26
164	Hyatt Place Rogers/Bentonville	1	26
165	HYATT house Colorado Springs	1	25
166	Hyatt Place Atlanta/Alpharetta/North Point Mall	1	25
167	HYATT house Gaithersburg	1	22
168	HYATT house Houston/Galleria	1	19
169	HYATT house Dallas/Richardson	1	17
170	Hyatt Place Secaucus/Meadowlands	1	17



When we compare the surveys resulting in promoters from the hotels in spots 161-170 on this same list, we see that the sample sizes were significantly lacking. More than anything, this tells us that the hotels could do a better job of promoting and/or incentivizing survey completion among guests.

Ten hotels with the most detractors/passives:

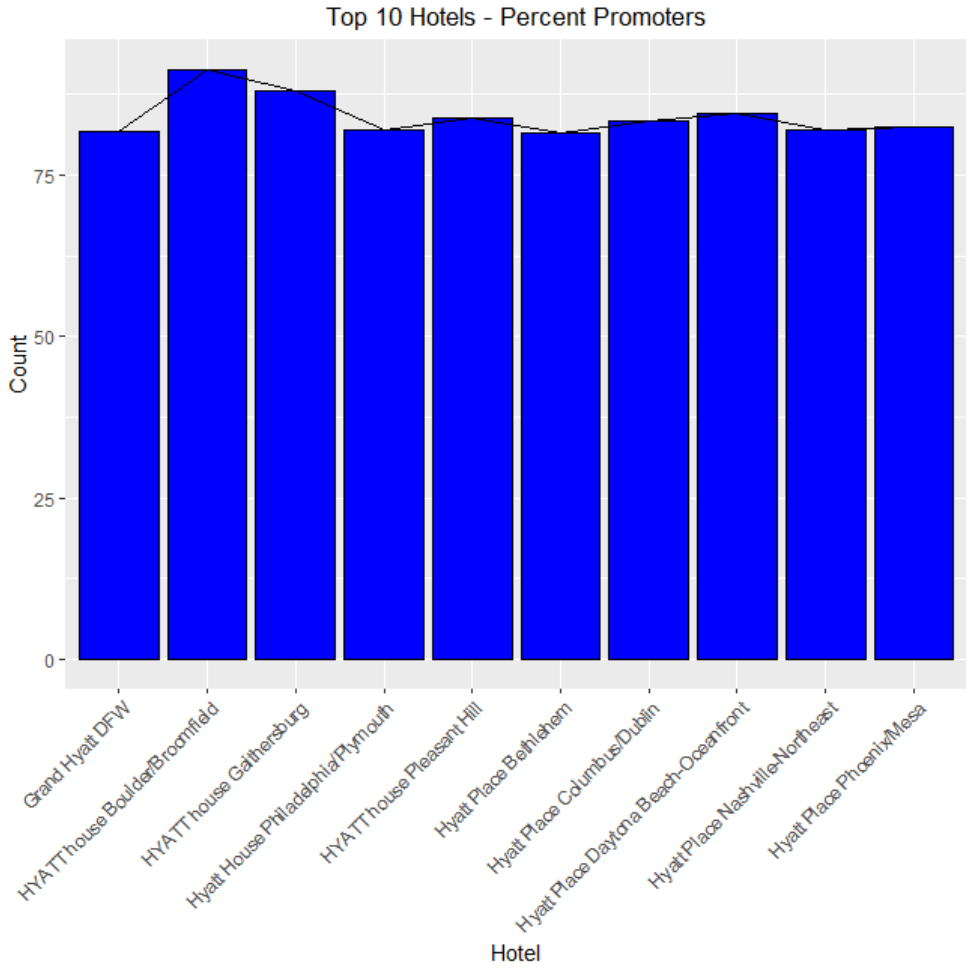
	HotelName	Promoter	n
1	Hyatt Regency Jacksonville Riverfront	0	204
2	Hyatt Regency O'Hare	0	185
3	Hyatt Regency Washington on Capitol Hill	0	166
4	Grand Hyatt Washington	0	162
5	Hyatt Regency Santa Clara	0	146
6	Hyatt Regency Grand Cypress	0	118
7	Hyatt Regency Birmingham - The Wynfrey Hotel	0	92
8	Hyatt Regency Crystal City at Reagan National Airport	0	87
9	Hyatt Regency Lost Pines Resort and Spa	0	73
10	Hyatt Palm Springs	0	69



As stated above, looking at the total number of promoters (or non-promoters in this case) is not the most useful metric, simply because the sample sizes vary so greatly between hotels. The hotels above have the most non-promoters, but this may be just a function of their surveys being completed by a wider audience.

Ten hotels with the highest percentage of promoters among survey takers:

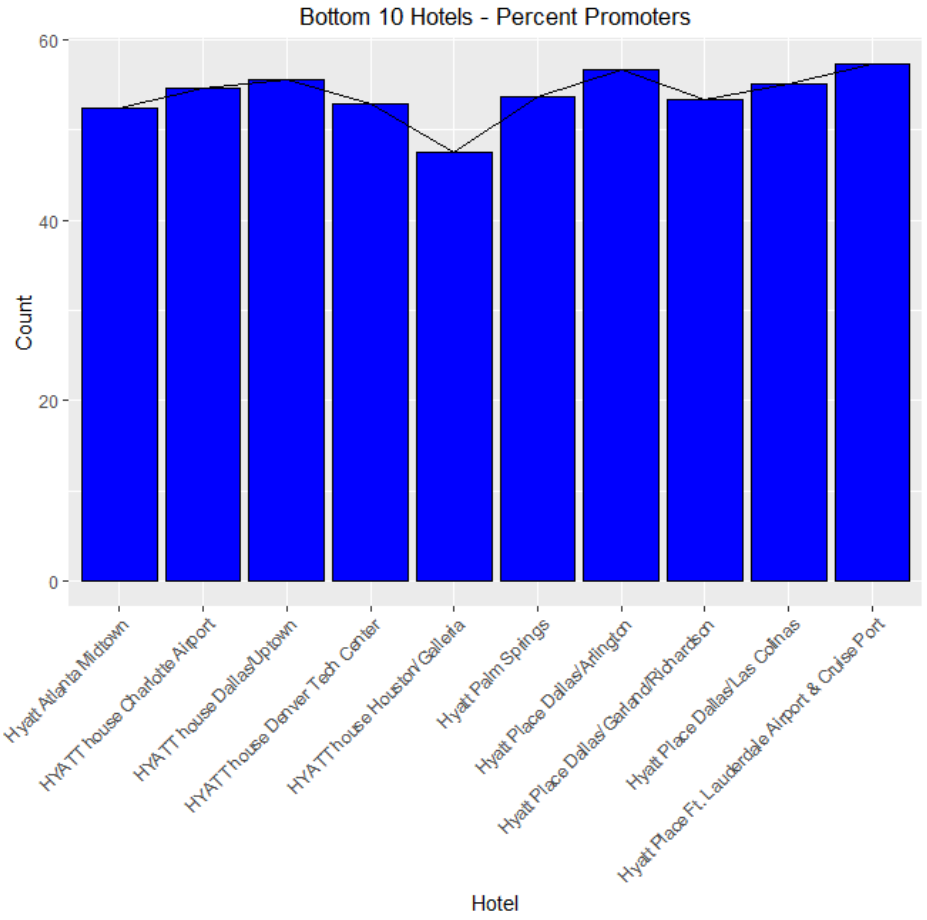
	HotelName	Promoter	n	p
93	HYATT house Boulder/Broomfield	1	53	91.37931
167	HYATT house Gaithersburg	1	22	88.00000
17	Hyatt Place Daytona Beach-Oceanfront	1	109	84.49612
123	HYATT house Pleasant Hill	1	47	83.92857
53	Hyatt Place Columbus/Dublin	1	70	83.33333
31	Hyatt Place Phoenix/Mesa	1	90	82.56881
86	Hyatt House Philadelphia/Plymouth	1	55	82.08955
88	Hyatt Place Nashville-Northeast	1	55	82.08955
8	Grand Hyatt DFW	1	203	81.85484
94	Hyatt Place Bethlehem	1	53	81.53846



Here we find a more useful view of the top ten hotels: viewing them by "promoter percentage" - the percentage of surveys that resulted in promoters. While the data does include some hotels with a very small sample size (Hyatt House Gaithersburg is the most extreme example with 22), it gives us a better idea of the overall service that a hotel provides.

Ten hotels with the lowest percentage of promoters among survey takers:

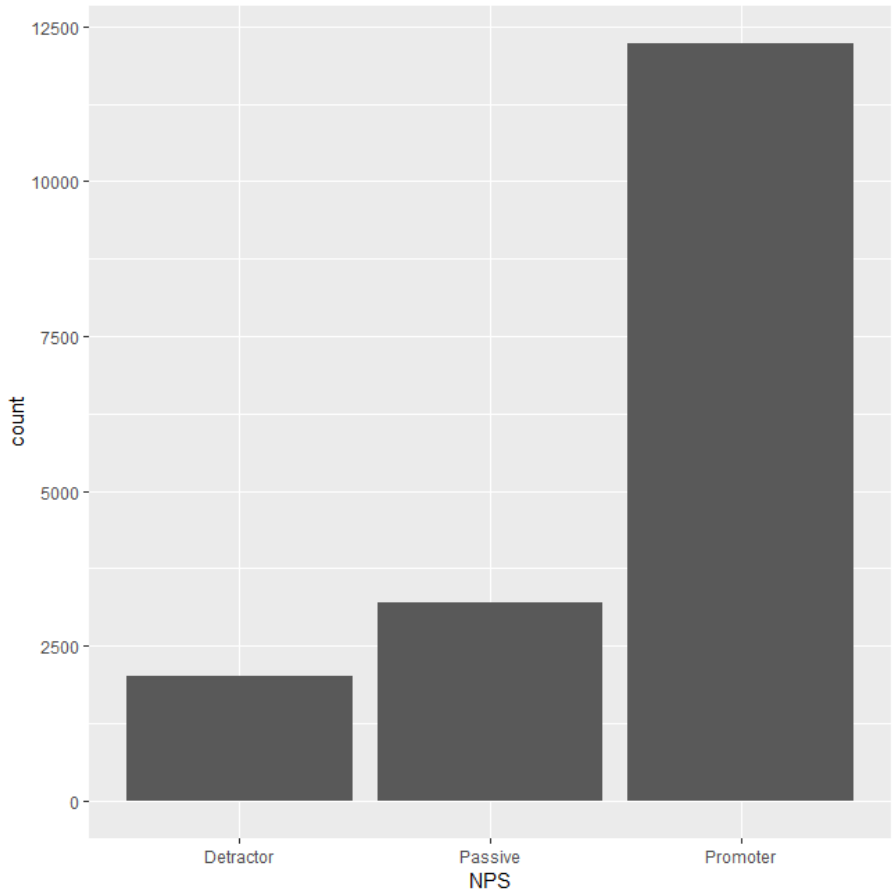
	HotelName	Promoter	n	p
87	Hyatt Place Ft. Lauderdale Airport & Cruise Port	1	55	57.29167
146	Hyatt Place Dallas/Arlington	1	38	56.71642
112	HYATT house Dallas/Uptown	1	50	55.55556
147	Hyatt Place Dallas/Las Colinas	1	38	55.07246
136	HYATT house Charlotte Airport	1	41	54.66667
41	Hyatt Palm Springs	1	80	53.69128
139	Hyatt Place Dallas/Garland/Richardson	1	40	53.33333
148	HYATT house Denver Tech Center	1	36	52.94118
131	Hyatt Atlanta Midtown	1	43	52.43902
168	HYATT house Houston/Galleria	1	19	47.50000



Lastly, we have a snapshot that looks at the lowest promoter rates among the 170 hotel data set. Again, sample size issues affect this group (in fact, the number of surveys here is significantly lower than those at the top). That said, we can still glean that there may be pervasive issues at some of these hotels that leave customers dissatisfied with their stays.

NPS type analysis

Since our aim is to actually discover which factors drive NPS, we thought it best to next dive into some analysis of the NPS types within this data set.
Looking at our 17,439 observations, we are able to see that far and away, Hyatt's guests are promoters. Once that had been established, we needed to drill down further to see if NPS type changed widely across our other variables.

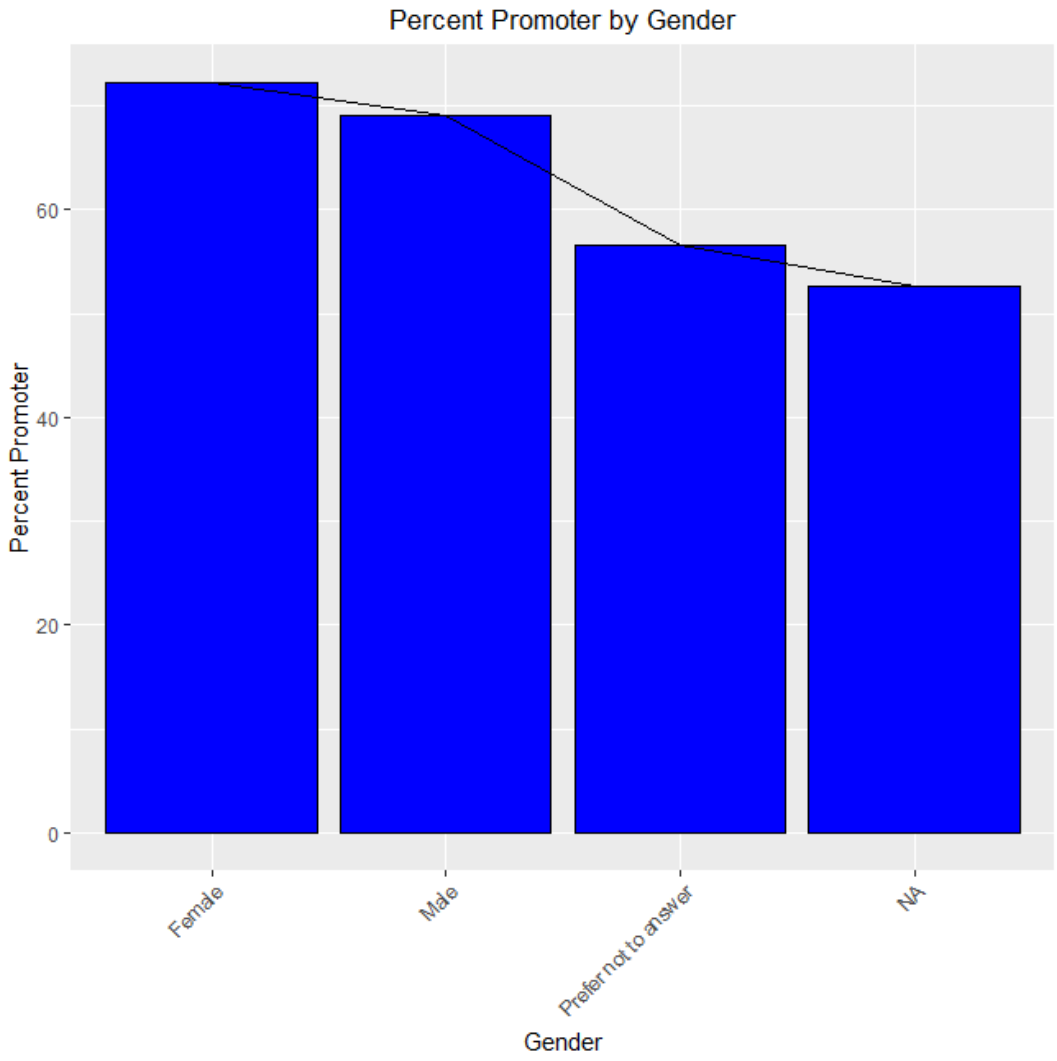


In addition to this high-level analysis of NPS types, we also analyzed how the following criteria affected NPS type:

a. **Gender**
b. **Age**
c. **Tier status**
d. **Purpose of visit**
e. **Hotel state**
f. **Hotel brand**
g. **Hotel ownership type**

Gender breakdown of NPS promoters:

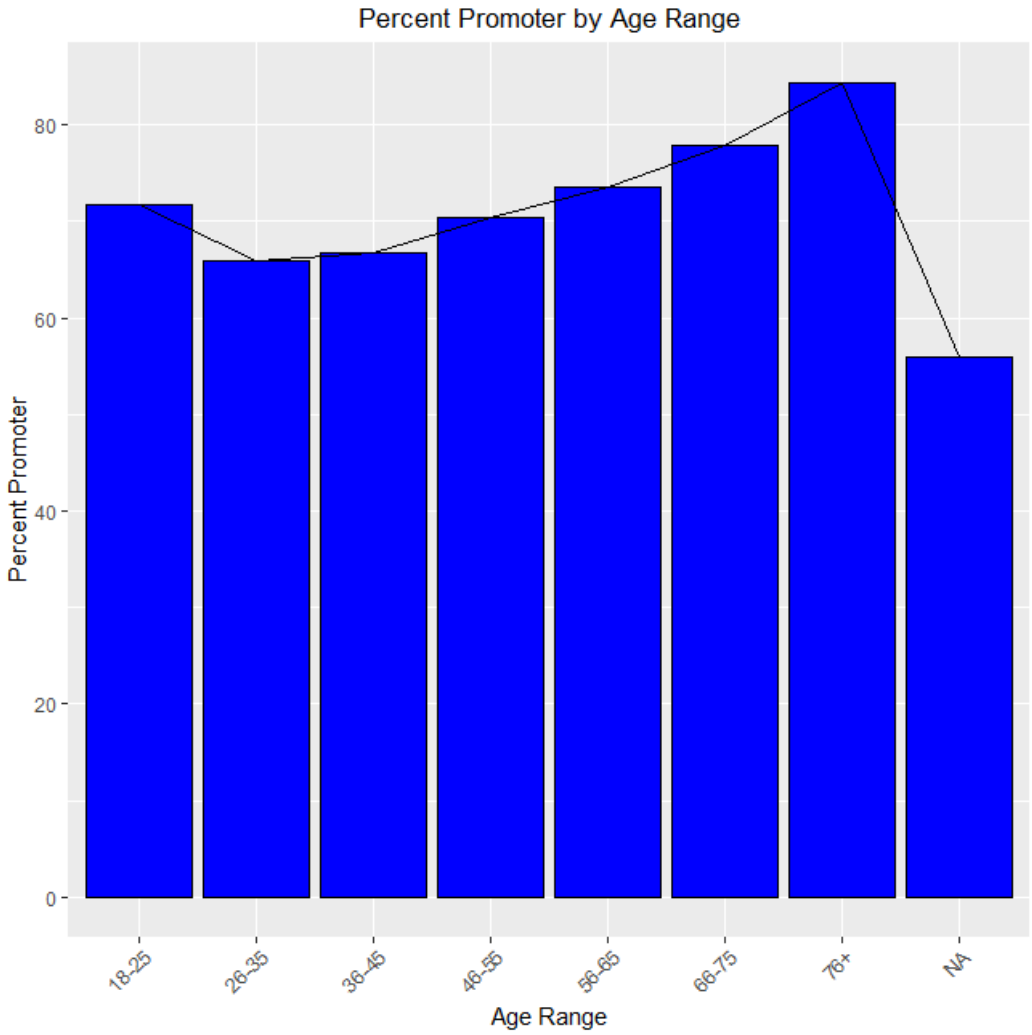
	Gender	Promoter	n	tn	p
2	Female	1	5740	7950	72.20126
1	Male	1	6198	8968	69.11240
3	Prefer not to answer	1	184	325	56.61538
4	<NA>	1	103	196	52.55102



Here we see that the majority of men (6198 of 8968, or 69%), women (5740 of 7950, or 72%), and those that preferred not to answer (184 of 325, or 57%) were classified as promoters. One of the largest takeaways here (one that will be echoed throughout our analysis) is that Hyatt's data collection could be more effective. By making this question mandatory, Hyatt could have prevented the 196 <NA> values that appear in this chart and graph.

Age breakdown of NPS promoters:

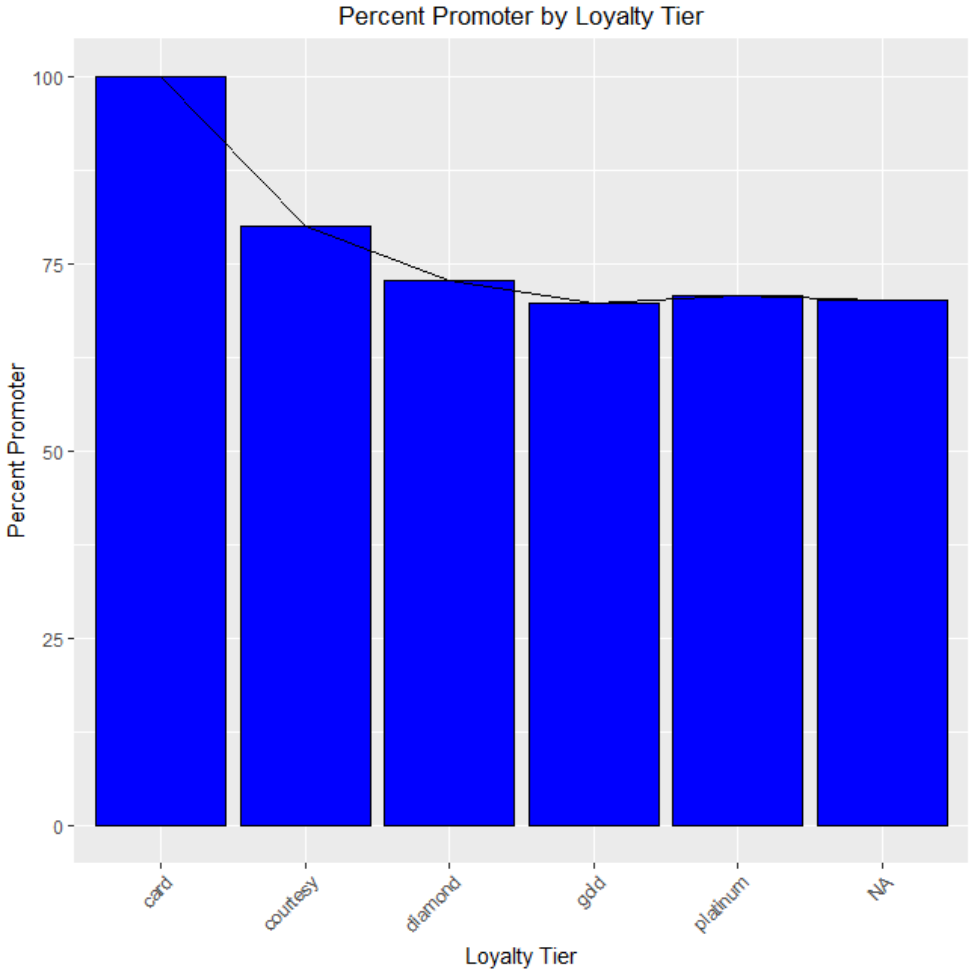
	AgeRange	Promoter	n	tn	p
8	76+	1	166	197	84.26396
5	66-75	1	927	1190	77.89916
2	56-65	1	2769	3763	73.58491
6	18-25	1	359	501	71.65669
1	46-55	1	3731	5295	70.46270
3	36-45	1	2552	3826	66.70152
4	26-35	1	1518	2304	65.88542
7	<NA>	1	203	363	55.92287



This shows us that the age group with the highest rate of promoters is the 76+ group (84%), and that the lowest defined group was the 26-35 group (66%). The group with the most promoters overall was the 46-55 block, and, despite having the highest promoter rate, the 76+ bracket had the smallest number of promoters. If the 18-25 age group is dismissed as an anomaly produced by relatively small sample size, we can see that there is a correlation between guests' age and their overall satisfaction with their stay.

Tier Status breakdown of NPS promoters:

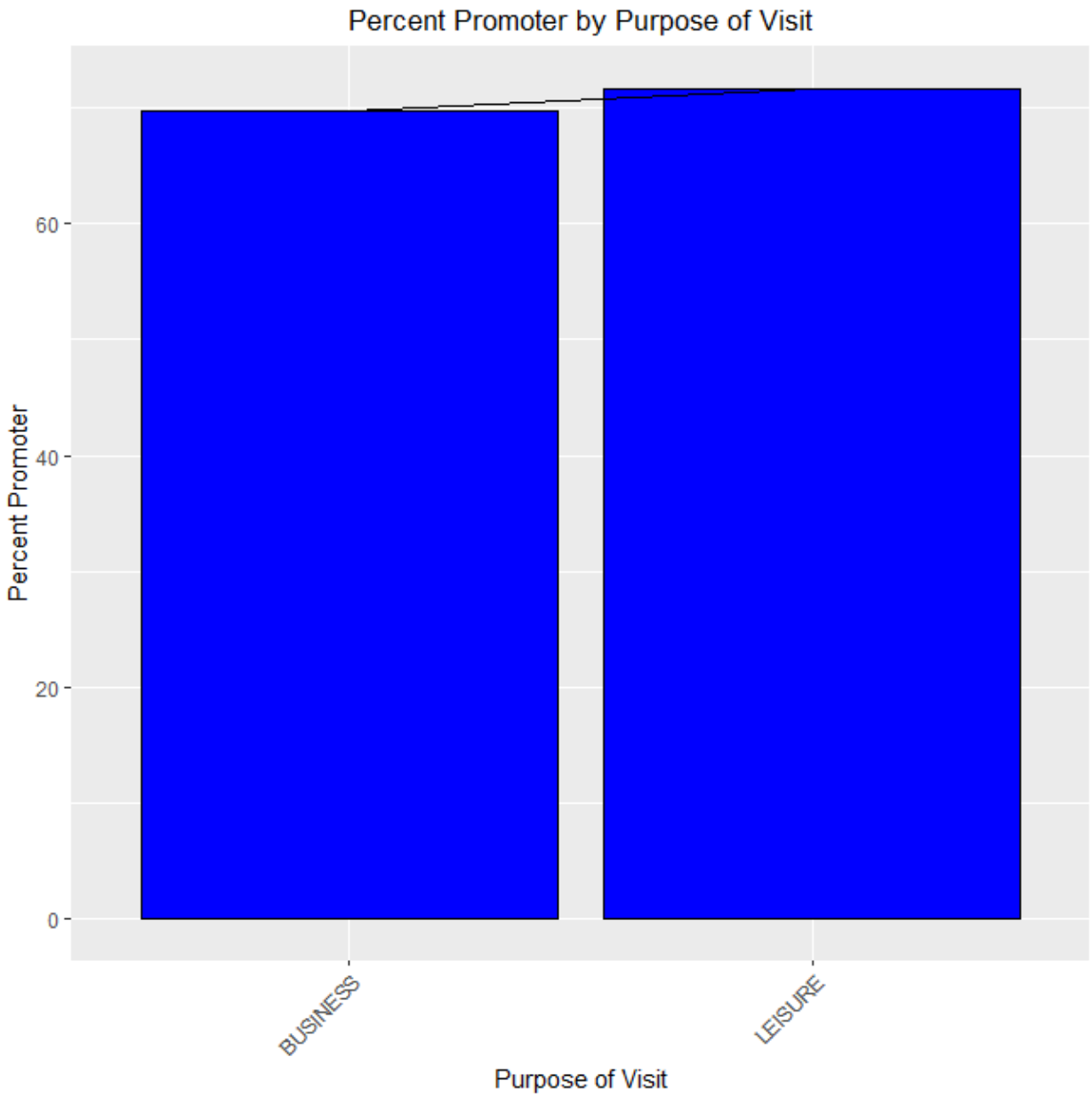
	GPTier	Promoter	n	tn	p
6	card	1	1	1	100.00000
5	courtesy	1	8	10	80.00000
4	diamond	1	469	645	72.71318
3	platinum	1	1477	2084	70.87332
2	<NA>	1	3818	5441	70.17092
1	gold	1	6452	9258	69.69108



By conducting analysis on Hyatt's loyalty program (as it relates to NPS), we see that NPS types across the three status tiers represented are relatively similar. Additionally, the promoter rate among survey-takers who did not participate in the loyalty program was similar as well. That said, Hyatt is missing opportunities to enroll guests in its loyalty program. We see that the higher a guest's status, the more satisfied they are with their experience; why not enroll as many guests as possible? Additionally, if "card" represents a guest who holds a Hyatt-branded credit card, Hyatt can do much better in its promotion of the card, given that there is only one observation in the entire dataset.

Purpose of visit breakdown of NPS promoters:

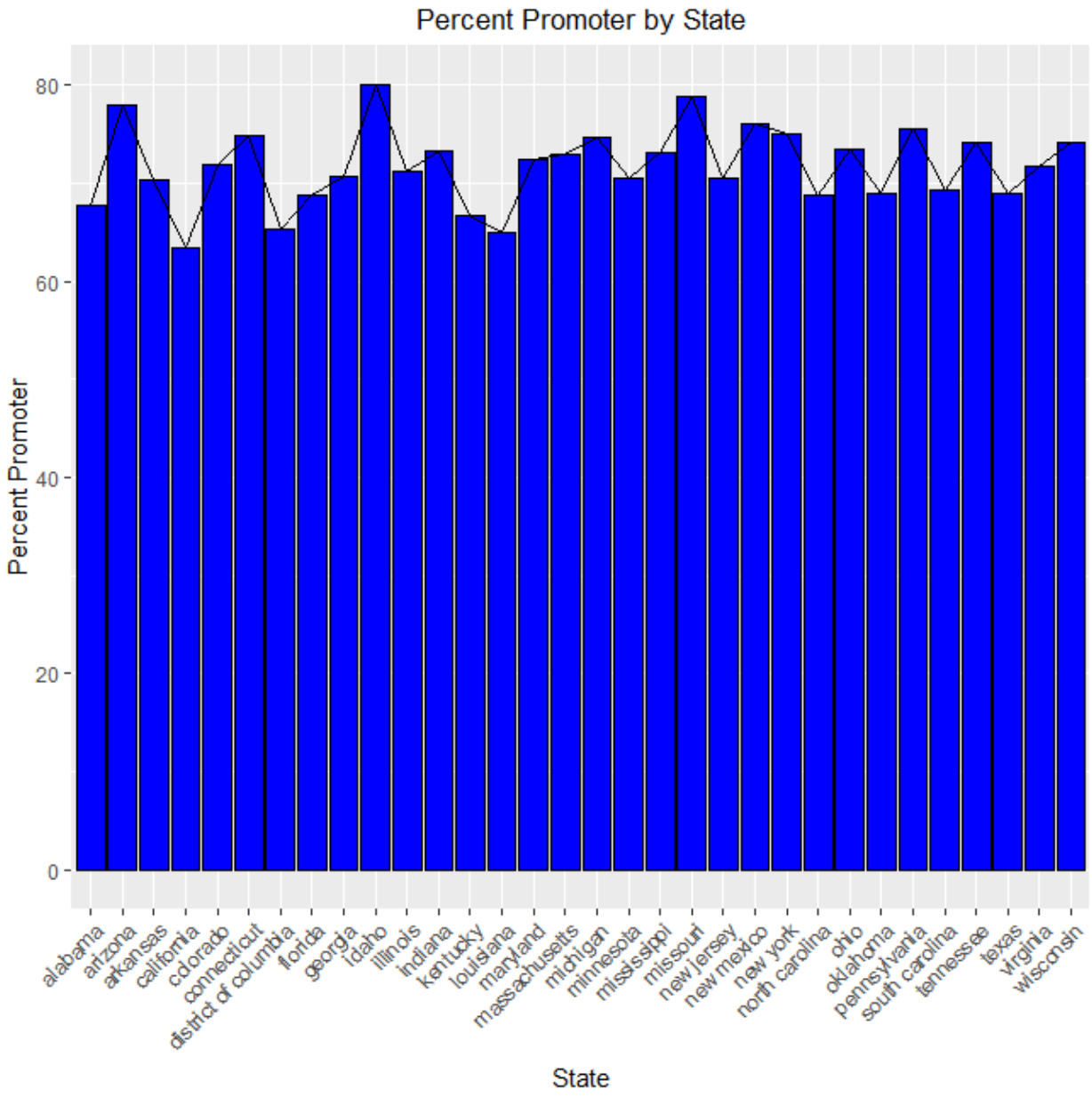
	PurposeOfVisit	Promoter	n	tn	p
2	LEISURE	1	2404	3353	71.69699
1	BUSINESS	1	9821	14086	69.72171



In looking at the number of promoters and rate of promoters among business and leisure travelers, we see that there is little difference between the two. Nonetheless, the business travelers promoted at a rate that was about two percentage points lower than leisure travelers. Notwithstanding, the sample size of the business travelers was more than four times that of the pool of leisure travelers.

Hotel state breakdown as it relates to NPS promoters:

	stateName	Promoter	n	tn	p
28	idaho	1	92	115	80.00000
29	missouri	1	78	99	78.78788
20	arizona	1	163	209	77.99043
17	new mexico	1	212	279	75.98566
18	pennsylvania	1	206	273	75.45788
25	new york	1	99	132	75.00000
13	connecticut	1	321	429	74.82517
7	michigan	1	505	677	74.59380
11	tennessee	1	390	526	74.14449
22	wisconsin	1	126	170	74.11765
12	ohio	1	380	517	73.50097
19	indiana	1	176	240	73.33333
26	mississippi	1	95	130	73.07692
21	massachusetts	1	146	200	73.00000
24	maryland	1	108	149	72.48322
8	colorado	1	429	597	71.85930
9	virginia	1	421	587	71.72061
4	illinois	1	896	1259	71.16759
5	georgia	1	750	1060	70.75472
14	new jersey	1	294	417	70.50360
27	minnesota	1	93	132	70.45455
32	arkansas	1	26	37	70.27027
16	south carolina	1	214	309	69.25566
23	oklahoma	1	118	171	69.00585
1	texas	1	1926	2795	68.90877
2	florida	1	1508	2191	68.82702
10	north carolina	1	392	570	68.77193
15	alabama	1	283	418	67.70335
30	kentucky	1	76	114	66.66667
3	district of columbia	1	975	1492	65.34853
31	louisiana	1	52	80	65.00000
6	california	1	675	1065	63.38028



After grouping the hotels into the states where they are located and honing in on their promoter rates, there appears to be a great deal of disparity between the top states and those further down the list. While Idaho's 80% promoter rate can be explained away by its relatively small sample size, other states with considerable sample sizes still performed much higher than others. For instance, Ohio, Tennessee, Michigan, and Connecticut all had over 400 survey takers yet still cleared the 73.5% mark for promoter rate. Meanwhile, California's promoter rate was around 63% over a sample size of 1,065.

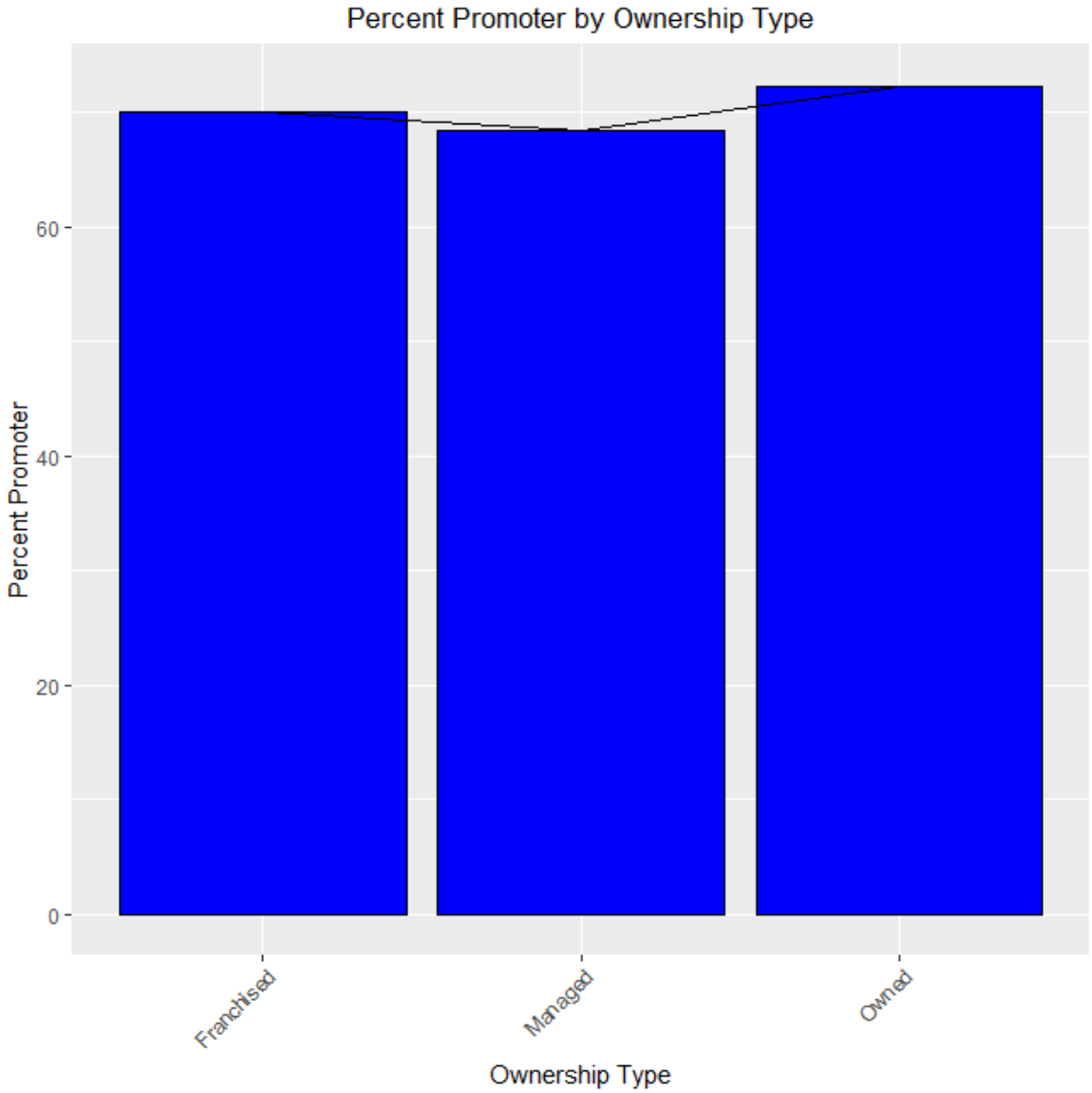
Hotel brand breakdown for NPS promoters:

	HotelBrand	Promoter	n	tn	p
6	Park Hyatt	1	145	179	81.00559
4	Grand Hyatt	1	524	731	71.68263
1	Hyatt Place	1	7603	10630	71.52399
3	Hyatt House	1	1556	2198	70.79163
2	Hyatt Regency	1	2091	3194	65.46650
5	Hyatt	1	306	507	60.35503

Here we see that the two brands with the smallest sample sizes (Hyatt and Park Hyatt) fell well outside of the other brands' promoter rates.That said, the three brands with large sample sizes (Hyatt Regency, Hyatt House, and Hyatt Place) did experience some disparity, with 65%, 71%, and 72% promoter rates, respectively.

Hotel ownership type breakdown for NPS promoters:

	Relationship	Promoter	n	tn	p
3	Owned	1	2781	3844	72.34651
1	Franchised	1	5952	8494	70.07299
2	Managed	1	3492	5101	68.45717

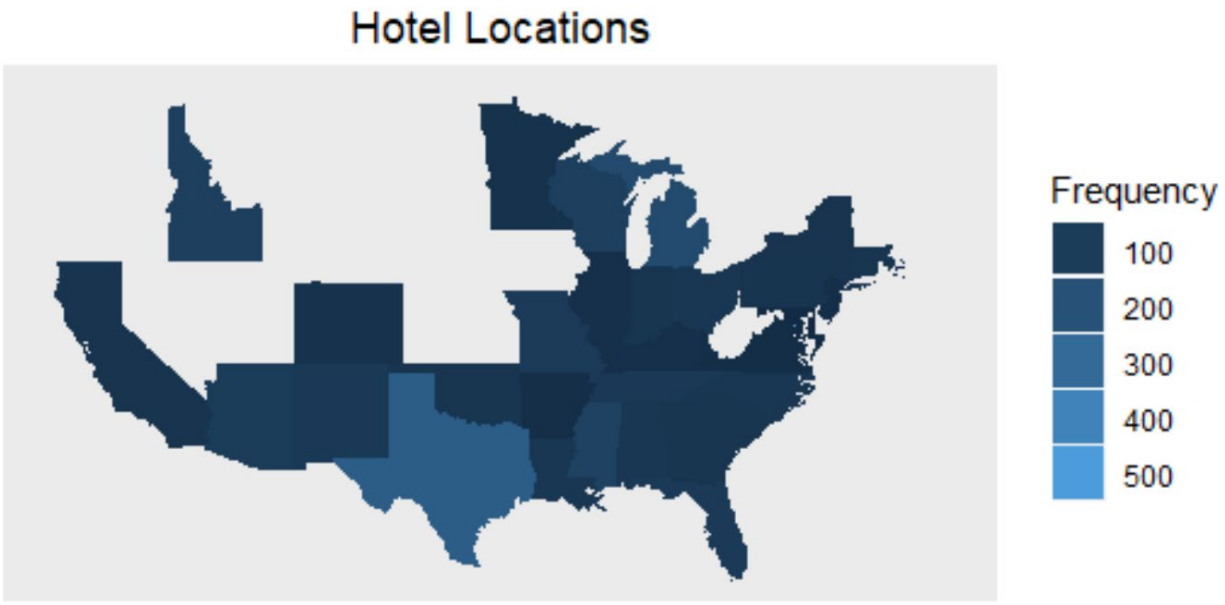


The above chart seems to suggest that the Hyatt brand's stewardship of hotels tends to be more effective than independent ownership. The hotels with the lowest scores tended to be those that were managed by third party companies.

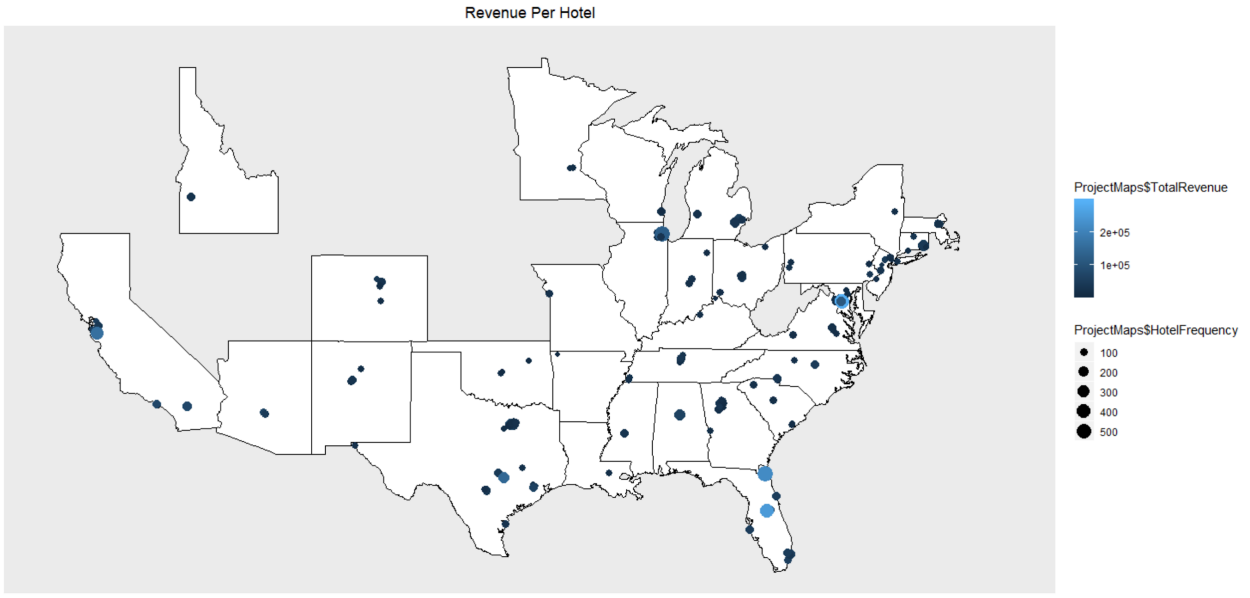
Maps

Orange plotted the 170 unique hotels to better understand the location, revenue and frequency of stay for each hotel.

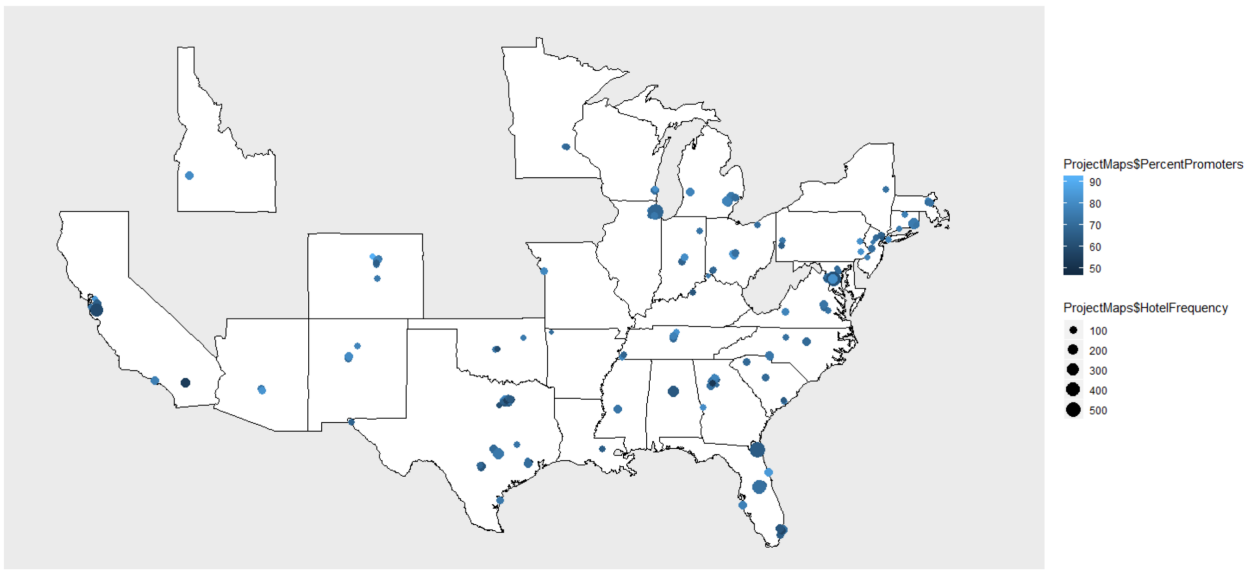
We are able to see that Texas has the most frequently visited hotels. No major patterns begin to emerge with this view in terms of frequency, but we are able to see a large swath of the country (in the Midwest and West) that is not represented in the data set. This information can be key for planning future hotels.



The image below allows us to understand the relationship between frequency and revenue generated for each hotel. The visual suggests what we expected: More visits to a hotel will result in greater revenue as seen by the bigger and brighter circles. The only exception is in California and Illinois. Due to the data being collected in February it makes sense to see Chicago not producing much revenue despite a relatively large circle if we make the assumption that the Chicago hotels were discounted during the colder months. The same assumption would likely not hold true for hotels in a warmer climate such as California.



The following map allows us to see the relationship between Promoter Percentage (Percentage of Promote vs. Non Promote) and the frequency at which guests visit the hotels. There appears to be a strong correlation here due to the larger circles being brighter. This may suggest that guests are providing a Promote ranking and returning as a repeat guest.



Top and Bottom 10 Profiles

Orange wanted to determine the differentiating factors between the Top 10 and Bottom 10 hotels. We ranked hotels based on the percentage of guests that submitted a Promoter survey, and leveraged numerical and factorial variables to gain insights.

The following code and function can be used as surveys are populated so that Hyatt can have live updates of the Top and Bottom 10 Hotels.

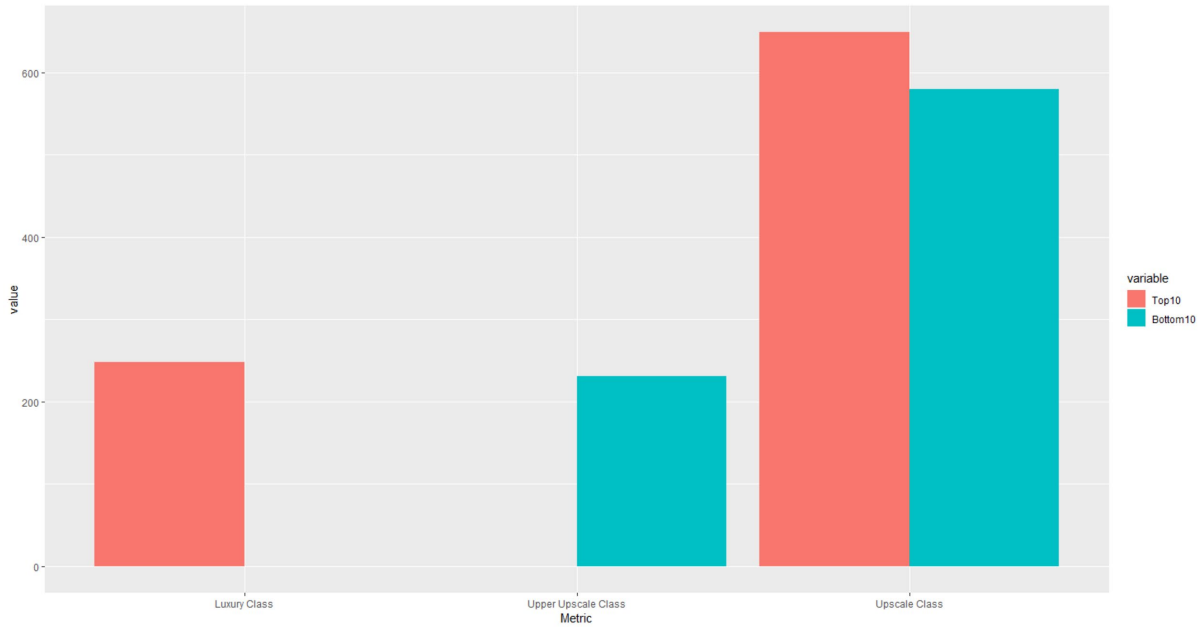
```
HotelPromotePercent <- function()
{
  hotelPromoter <- ProjectSurveyUSA %>% mutate(promoter = ifelse(NPS == "Promoter", 1, 0)) %>%
    group_by(HotelAbbrv = fct_inorder(HotelAbbrv)) %>%
    summarise(Percent = percent(mean(promoter)))
  return(hotelPromoter[order(desc(hotelPromoter$Percent)),])
}

HotelRankings <- HotelPromotePercent()

Top10 <- as.data.frame(HotelRankings[1:10,])
Top10 <- merge(Top10, ProjectSurveyUSA, by = "HotelAbbrv")

Bottom10 <- as.data.frame(HotelRankings[(dim(HotelRankings)[1]-9):dim(HotelRankings)[1],])
Bottom10 <- merge(Bottom10, ProjectSurveyUSA, by = "HotelAbbrv")
```

In order to compare the two categories, charts and visuals were generated.

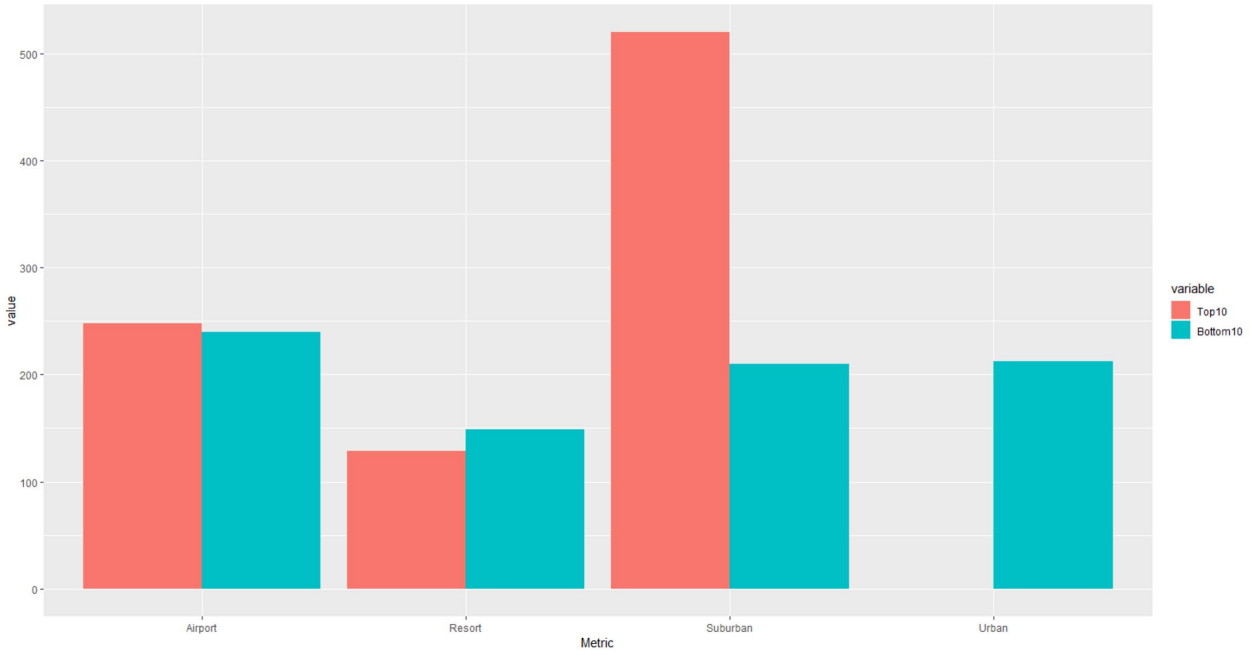
As a result of this comparison we are able to see that guest scores for the Top 10 hotels are higher in nearly every category. Perhaps the most interesting piece is that Quoted Rates and Revenue are higher with the Top 10 hotels. This suggests that guests are more concerned about the experience during their stay than they are with saving money. This insight is critical as Hyatt attempts to convert Non Promoters to Promoters.

	Metric	Top10	Bottom10
1	AwardCat	2.1917503	1.9038224
2	CI_Score	9.3717343	9.0000275
3	CS_Score	9.3740826	8.5392483
4	FBE_Score	8.8602026	8.4486094
5	FB_Frequency	1.4891112	1.4782775
6	FutureResvDays	124.7190635	133.8249075
7	GR_Score	9.2922595	8.1971542
8	HC_Score	9.4344979	8.1405404
9	IS_Score	8.8932721	8.7039651
10	LOS	2.7775648	2.6626004
11	LR_Score	9.2307692	7.8051788
12	OS_Score	9.1036789	7.8865312
13	Promoter	0.8338907	0.5425401
14	QuotedRate	145.6821962	131.6670530
15	Revenue	378.2024080	346.0555857
16	SC_Score	9.1474880	8.7163978
17	T_Score	9.0359993	8.3930850

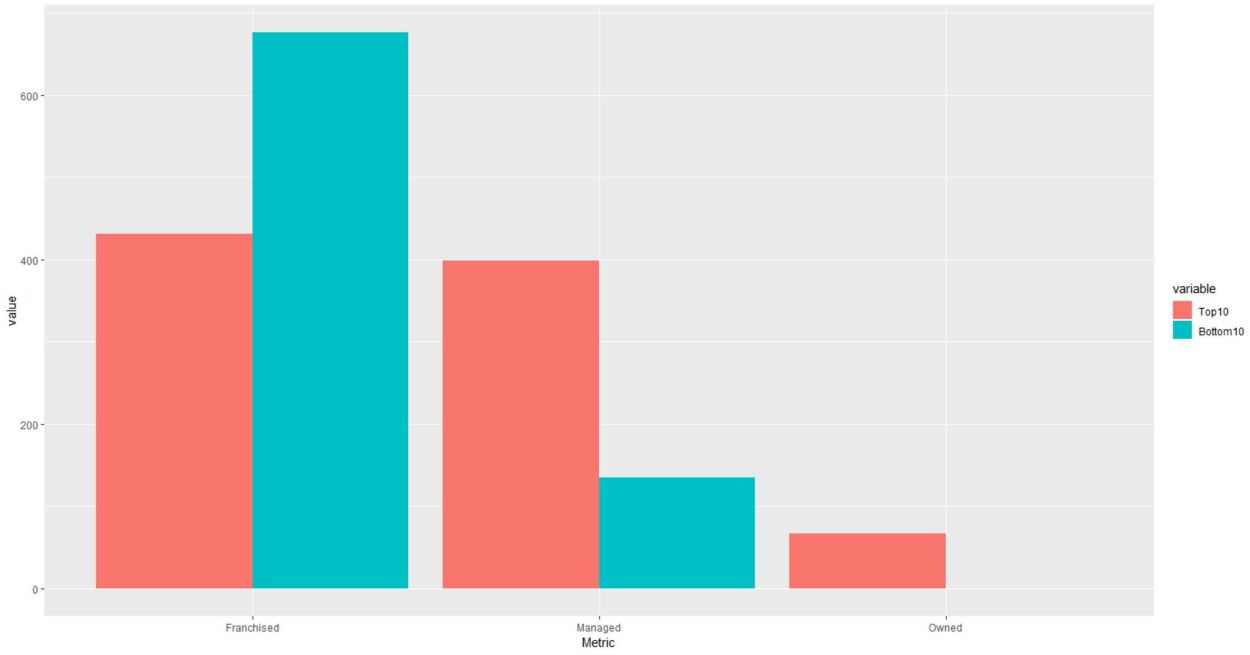
Our observations about spending seem to hold true given the image below. Luxury hotels are absent from the Bottom 10, but formed an important part of the Top 10. While the Bottom 10 does focus include Upper Upscale properties, the surveys resulting in promoters still fall short of the Top 10's Luxury totals. A shift in focus and marketing might better serve the Bottom 10.



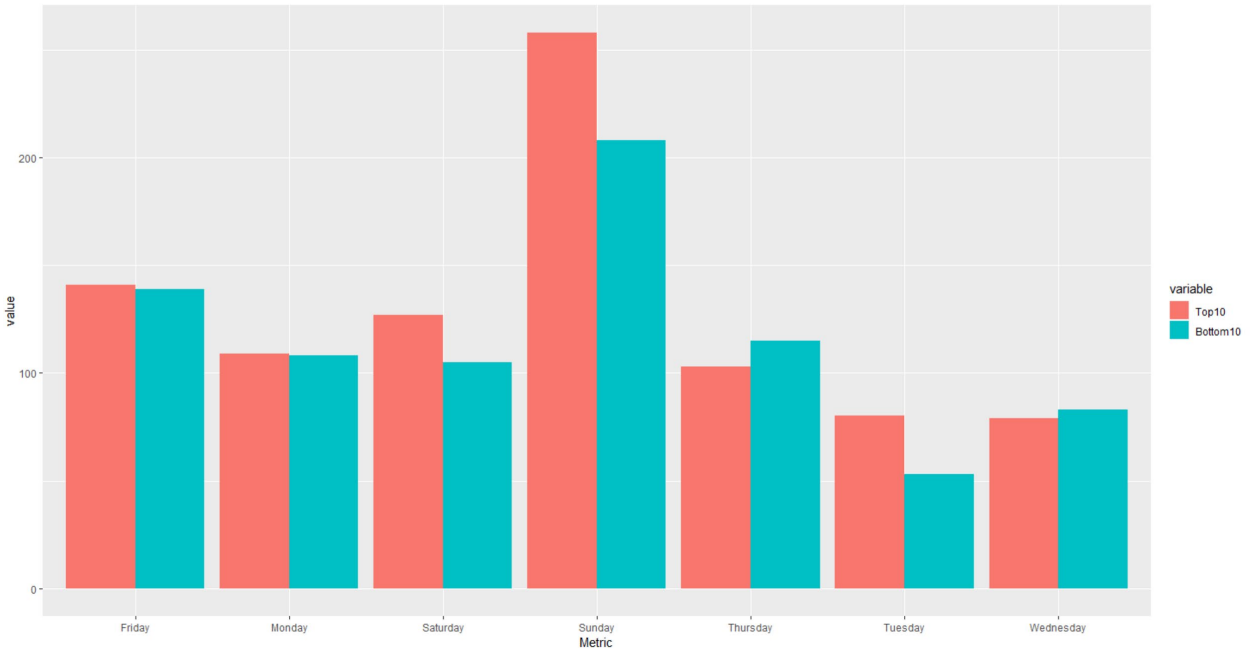
If Hyatt does consider expansion to the Midwest and West it may want to consider building hotels in the suburbs, as the Top 10 hotels capitalized on this location when compared to the Bottom 10. Of note: Urban locations did not contribute to the success of the Top 10 hotels.



Hyatt should also consider the types of owners operating the hotels as franchisees are connected more to the Bottom 10 while Hyatt-owned properties are associated exclusively with the Top 10.



Orange also recognized a trend on check out day for the Top and Bottom 10. The Top 10 hotels appeared to have more guests checkout on the weekends when compared to the Bottom 10. Perhaps the Bottom 10 hotels can offer promotions that incentivize guests to stay over for the weekend, which will hopefully result in more Promoter surveys for the Bottom 10.



Models

In order to determine the factors that are the most significant contributors to the NPS, we utilized our knowledge and understanding from the previous sections.

Linear Model

The following code was used to predict Detractor, Passive or Promoter based on a training data set and then tested against the test data set.

```
syrModel <- polr(NPS~CS_Score + GR_Score + HC_Score + OS_Score + QuotedRate + Revenue + SC_Score + Relationsh
ip + CheckOutDay + T_Score + Type + Location , data = trainData, Hess = TRUE)
```

This resulted in a model with 89.97% accuracy as seen in the matrix below.

```
testData...59. Detractor Passive Promoter
Detractor         511      109       20
Passive            60      768      199
Promoter           13      182     3951

Accuracy <- round((511+768+3951)/(511+60+13+109+768+182+20+199+3951)*100,2)
Accuracy
```

Association Rules Mining

The purpose is to find the rules which help the team predict the occurrence of specific characteristics based on the other variables in the data set

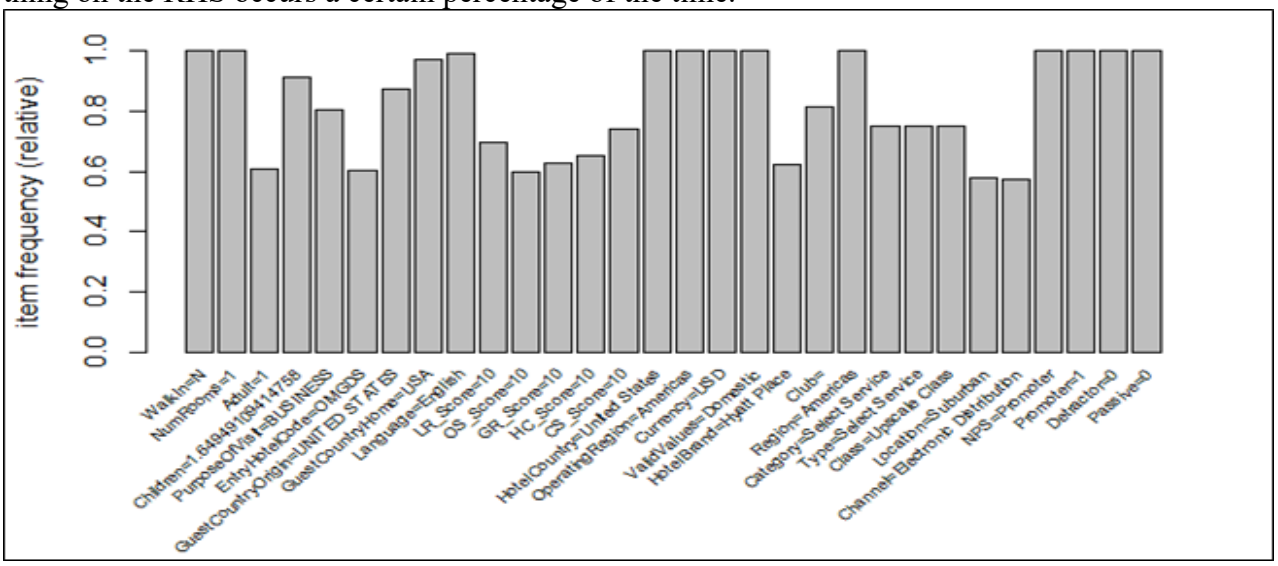
We used the criteria support and confidence to identify the most important relationships.

The support is an indication of how frequently the items appear in the data.

We originally looked at the overall data set. To do this, we converted the data-frame to transaction data by converting numbers or strings to factors in the data set, then convert ultimately the data-frame to a transaction data set. We also restricted the data to exclusively look at promoters.

At the first level we have one-item sets (commonly found). The next level is the two-items sets - they need to have the property that each of their subsets must be frequent enough to include (sufficient support).

Each rule states that when the thing or things on the left hand side of the equation occurs, the thing on the RHS occurs a certain percentage of the time.



Using a generic support of 2% and a confidence of 50%, the set of rules is 198, which is a lot of rules to examine manually. Hence, it was plotted using the arulesViz package.

The lift is a measure of the performance of association rule at predicting with respect to the population.

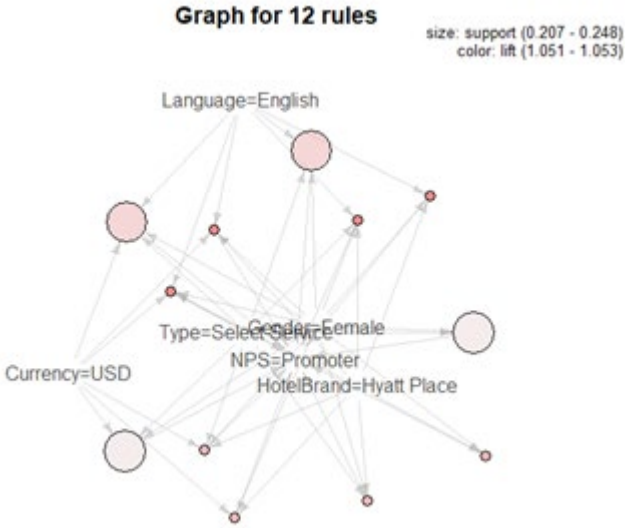
Lift serves as a measure of interestingness (from textbook) and the higher the lift the more attention it gets.

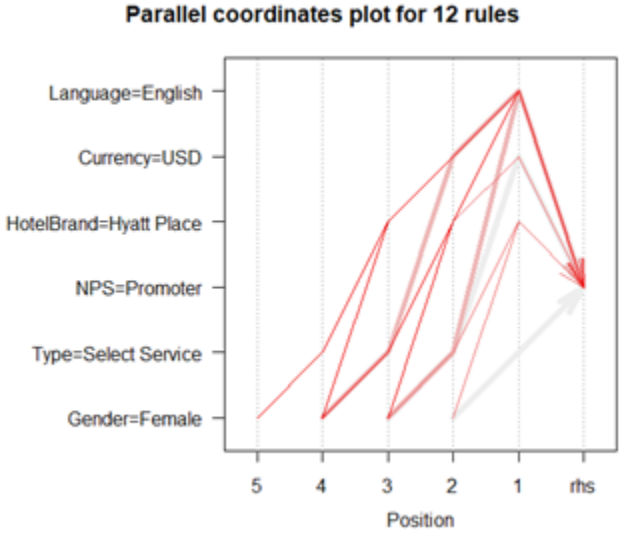
- More favorable are rules that have high support and high confidence, Some rules with high lifts seem to have low support, Support is an indication of how frequently the items appear in the data.

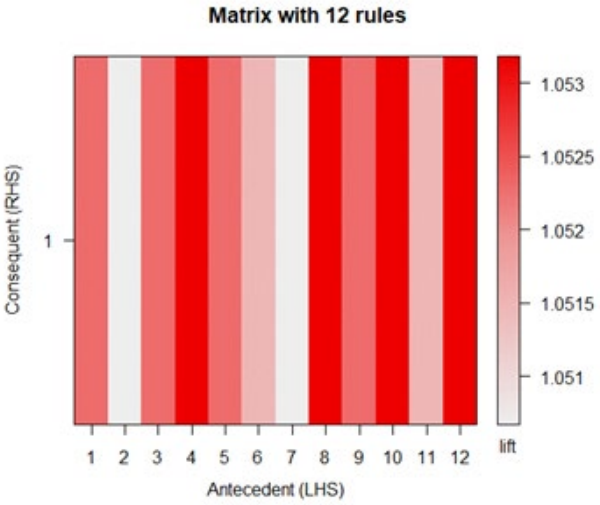
```
> # creating rules with respect to NPS_type
> Rules.NPS <- apriori(NPS.df, parameter = list(support = 0.2, confidence =0.5),
+                      appearance = list(rhs=c("NPS=Promoter", "NPS=Detractor",
+                                              "NPS=Passive"), default="lhs"))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5     0.2      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 3487

set item appearances ...[3 item(s)] done [0.00s].
set transactions ...[61 item(s), 17439 transaction(s)] done [0.01s].
sorting and recoding items ... [20 item(s)] done [0.00s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 3 4 5 6 7 done [0.01s].
writing ... [198 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

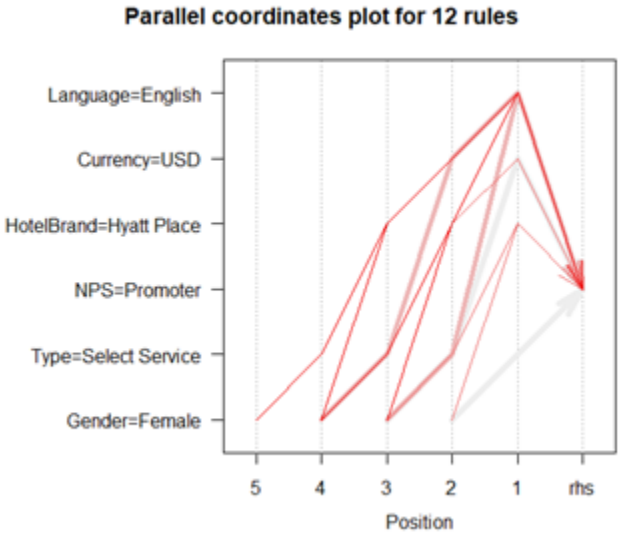
We re-ran the good rule with a lift of 1.05 and got rules with the highest lift; this amounted to twelve rules.

```
set of 12 rules
> inspect(Good.Rules)
     lhs                                                                                                rhs             support   confidence lift     count
[1]  {HotelBrand=Hyatt Place,Gender=Female}                                                          => {NPS=Promoter} 0.2073513 0.7376581  1.052272 3616
[2]  {Type=Select Service,Gender=Female}                                                             => {NPS=Promoter} 0.2476633 0.7365280  1.050659 4319
[3]  {HotelBrand=Hyatt Place,Type=Select Service,Gender=Female}                                      => {NPS=Promoter} 0.2073513 0.7376581  1.052272 3616
[4]  {Language=English,HotelBrand=Hyatt Place,Gender=Female}                                         => {NPS=Promoter} 0.2068926 0.7382852  1.053166 3608
[5]  {Currency=USD,HotelBrand=Hyatt Place,Gender=Female}                                             => {NPS=Promoter} 0.2073513 0.7376581  1.052272 3616
[6]  {Language=English,Type=Select Service,Gender=Female}                                            => {NPS=Promoter} 0.2466311 0.7371037  1.051481 4301
[7]  {Currency=USD,Type=Select Service,Gender=Female}                                                => {NPS=Promoter} 0.2476633 0.7365280  1.050659 4319
[8]  {Language=English,HotelBrand=Hyatt Place,Type=Select Service,Gender=Female}                     => {NPS=Promoter} 0.2068926 0.7382852  1.053166 3608
[9]  {Currency=USD,HotelBrand=Hyatt Place,Type=Select Service,Gender=Female}                         => {NPS=Promoter} 0.2073513 0.7376581  1.052272 3616
[10] {Language=English,Currency=USD,HotelBrand=Hyatt Place,Gender=Female}                             => {NPS=Promoter} 0.2068926 0.7382852  1.053166 3608
[11] {Language=English,Currency=USD,Type=Select Service,Gender=Female}                               => {NPS=Promoter} 0.2466311 0.7371037  1.051481 4301
[12] {Language=English,Currency=USD,HotelBrand=Hyatt Place,Type=Select Service,Gender=Female}        => {NPS=Promoter} 0.2068926 0.7382852  1.053166 3608
>
```









Support Vector Machine

Our Support Vector Machine utilized several more variables and produced an accuracy of 99.9%. This accuracy was higher, due in part, to the Passive and Detractors aggregated into a NonPromoter bucket. We took this approach as Hyatt's goal should be to convert all non-promoters to become promoters.

```
> syr <- read.csv("ProjectSurveyUSA_12.10.18v2.csv")
>
> library(kernlab)
> table(syr$NPS)

NonPromoter    Promoter
       5214       12225
> randindex <- sample(1:dim(syr)[1])
> summary(randindex)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1    4360    8720    8720   13080   17439
> length(randindex)
[1] 17439
> cutPoint2_3 <- floor((2/3)*dim(syr)[1])
>
> #Training set
> trainData <- syr[randindex[1:cutPoint2_3],]
>
> #Test set
> testData <- syr[randindex[(cutPoint2_3+1):dim(syr)[1]],]
>
>
> #Build model based on Training
> svmOutput <- ksvm(NPS ~., data = trainData, kernel = 'rbfdot',
+                   kpar = 'automatic', C=50, cross = 3, prob.model = TRUE)
>
> #Make a prediction
> svmPred <- predict(svmOutput, testData, type = 'votes')
>
> compTable <- data.frame(testData[,35], svmPred[1,])
> table(compTable)
                                  svmPred.1...
                    testData...35.    0    1
                       NonPromoter    2 1711
                       Promoter    4094    6
```

Best Model

The Ordered Logistic model provides more detail with its prediction but the Support Vector Machine achieves the most important goal for Hyatt: Identify Promoters and Non-Promoters. The Non-Promoters should remain Hyatt's biggest focus with the assumption that Promoters will remain in this category. If all hotels only improve then it should be reasonable to assume that Promoters will only become more satisfied and Non-Promoters will slowly convert. Yes, there may be some Promoters that do not like some of the changes, but the net gain will be positive for Promoters.

Business Question Insights

- What is the overall NPS?
  - Provide an overview of the count and percentage for each category of Promoter, Passive and Detractor

```
> #Calculate the total and percentage of Promoter, Passive and Detractor
> NPSsummary <- as.data.frame(table(ProjectSurveyUSA$NPS))
> colnames(NPSsummary) <- c("Category", "Freq")
> NPSsummary$Percent <- percent((NPSsummary$Freq / sum(NPSsummary$Freq)))
> NPSsummary
   Category  Freq Percent
1 Detractor  2006   11.5%
2   Passive  3208   18.4%
3  Promoter 12225   70.1%
```

- What is the NPS amongst different demographics
  - By US Region

```
> #Calculate the NPS by Region
> ProjectSurveyUSA$stateRegion <- ifelse(ProjectSurveyUSA$stateName %in% c("washington", "oregon", "california", "id
aho", "nevada", "montana", "wymoing", "utah", "arizona", "colorado"), "West",
+                                ifelse(ProjectSurveyUSA$stateName %in% c("new mexico", "oklahoma", "texas",
 "arkansas", "louisiana", "mississippi", "alabama"), "South",
+                                       ifelse(ProjectSurveyUSA$stateName %in% c("north dakota", "south dako
ta", "nebraska", "kansas", "minnesota", "iowa", "missouri", "wisconsin", "illinois", "michigan", "indiana", "kentuck
y", "ohio"), "Midwest",
+                                              ifelse(ProjectSurveyUSA$stateName %in% c("maine", "new hampsh
ire", "vermont", "massachusetts", "rhode island", "connecticut", "new york", "new jersey", "pennsylvania", "delaware
", "maryland", "district of columbia"), "Northeast",
+                                                     ifelse(ProjectSurveyUSA$stateName %in% c("west virgini
a", "virginia", "tennessee", "north carolina", "south carolina", "georgia", "florida"),"Southeast",
+                                                            "ERROR")))))
```

```
> #View the values populated in stateRegion. ERROR should not be included
> unique(ProjectSurveyUSA$stateRegion)
[1] "Southeast" "Northeast" "South"     "Midwest"   "West"
> #If ERROR appears from the previous line then run the code below to see the row where the error is occuring. If th
ere is not an error you should get integer(0) as output
> which(grepl("ERROR", ProjectSurveyUSA$stateRegion))
integer(0)
>
> #Get final view of the dataset to make sure the column was added
> View(ProjectSurveyUSA)
>
> #Create a Numerator and Denominator for each region and divide
> regionPromote <- function(region)
+ {
+   n <- length(which(ProjectSurveyUSA$NPS == "Promoter" & ProjectSurveyUSA$stateRegion == region))
+   d <- length(which(ProjectSurveyUSA$stateRegion == region))
+   return(percent(n/d))
+ }
>
> NEpromote <- regionPromote("Northeast")
> SEpromote <- regionPromote("Southeast")
> MWpromote <- regionPromote("Midwest")
> Spromote <- regionPromote("South")
> Wpromote <- regionPromote("West")
>
> rp <- data.frame(NEpromote, SEpromote, MWpromote, Spromote, Wpromote)
>
> colnames(rp) <- c("NorthEastPromote", "SouthEastPromote", "MidWestPromote", "SouthPromote", "WestPromote")
>
> rp
  NorthEastPromote SouthEastPromote MidWestPromote SouthPromote WestPromote
1            69.5%            70.1%          72.6%        69.4%       68.4%
```

- By State
  - 31 states + Washington D.C.

```
> StatePromotePercent <- function()
+ {
+ StatePromo <- ProjectSurveyUSA %>% mutate(promoter = ifelse(NPS == "Promoter", 1, 0)) %>%
+   group_by(stateName = fct_inorder(stateName)) %>%
+   summarise(Percent = percent(mean(promoter)))
+ StatePromo <- as.data.frame(StatePromo)
+ return(StatePromo)
+ }
> StatePromotePercent()
```

	stateName	Percent
1	florida	68.8%
2	district of columbia	65.3%
3	new mexico	76.0%
4	georgia	70.8%
5	texas	68.9%
6	tennessee	74.1%
7	massachusetts	73.0%
8	louisiana	65.0%
9	south carolina	69.3%
10	illinois	71.2%
11	kentucky	66.7%
12	ohio	73.5%
13	north carolina	68.8%
14	colorado	71.9%
15	new jersey	70.5%
16	virginia	71.7%
17	california	63.4%
18	missouri	78.8%
19	wisconsin	74.1%
20	minnesota	70.5%
21	oklahoma	69.0%
22	pennsylvania	75.5%
23	arizona	78.0%
24	arkansas	70.3%
25	maryland	72.5%
26	new york	75.0%
27	connecticut	74.8%
28	alabama	67.7%
29	idaho	80.0%
30	michigan	74.6%
31	indiana	73.3%
32	mississippi	73.1%

- By Hotel (provides same insight as by city)

```
> HotelPromotePercent <- function()
+ {
+   hotelPromoter <- ProjectSurveyUSA %>% mutate(promoter = ifelse(NPS == "Promoter", 1, 0)) %>%
+     group_by(HotelAbbrv = fct_inorder(HotelAbbrv)) %>%
+     summarise(Percent = percent(mean(promoter)))
+   hotelPromoter <- as.data.frame(hotelPromoter)
+   return(hotelPromoter)
+ }
> HotelPromotePercent()
                HotelAbbrv Percent
1          HR Grand Cypress   71.5%
2              HY Arlington   62.0%
3             GH Washington   66.5%
4             PH Washington   81.0%
5           HR Crystal City   61.7%
6             HR Washington   60.8%
7    HP Albuquerque Airport   69.3%
8     HP Albuquerque Uptown   80.0%
9               HP Windward   78.0%
10            HP Johns Creek   80.3%
```

- By Gender

```
> GenderPromotePercent <- function()
+ {
+   genderPromoter <- ProjectSurveyUSA %>% mutate(promoter = ifelse(NPS == "Promoter", 1, 0)) %>%
+     group_by(Gender = fct_inorder(Gender)) %>%
+     summarise(Percent = percent(mean(promoter)))
+   genderPromoter <- as.data.frame(genderPromoter)
+   return(genderPromoter)
+ }
> GenderPromotePercent()
               Gender Percent
1              Female   72.2%
2                Male   69.1%
3                       52.6%
4 Prefer not to answer   56.6%
```

- By Purpose of Visit

```
> PurposePromotePercent <- function()
+ {
+   purposePromoter <- ProjectSurveyUSA %>% mutate(promoter = ifelse(NPS == "Promoter", 1, 0)) %>%
+     group_by(PurposeOfVisit = fct_inorder(PurposeOfVisit)) %>%
+     summarise(Percent = percent(mean(promoter)))
+   purposePromoter <- as.data.frame(purposePromoter)
+   return(purposePromoter)
+ }
> PurposePromotePercent()
  PurposeOfVisit Percent
1       BUSINESS   69.7%
2        LEISURE   71.7%
```

- What days of the week are most popular amongst guests?
  - Check-in Day

```
> CheckInPromotePercent <- function()
+   {
+   CheckInPromoter <- ProjectSurveyUSA %>% mutate(promoter = ifelse(NPS == "Promoter", 1, 0)) %>%
+         group_by(CheckInDay = fct_inorder(CheckInDay)) %>%
+         summarise(Percent = percent(mean(promoter)))
+         CheckInPromoter <- as.data.frame(CheckInPromoter)
+         return(CheckInPromoter[order(desc(CheckInPromoter$Percent)),])
+   }
> CheckInPromotePercent()
  CheckInDay Percent
3   Saturday   72.4%
5   Thursday   71.2%
1     Friday   70.4%
7  Wednesday   70.1%
2     Monday   69.5%
4    Tuesday   68.1%
6     Sunday   67.6%
```

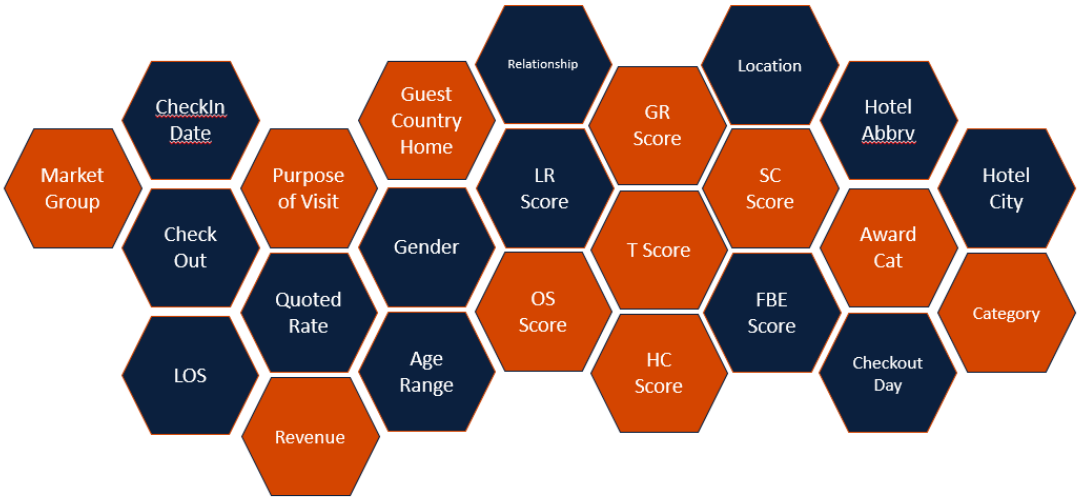
- Check-out Day

```
> CheckOutPromotePercent <- function()
+ {
+   CheckOutPromoter <- ProjectSurveyUSA %>% mutate(promoter = ifelse(NPS == "Promoter", 1, 0)) %>%
+     group_by(CheckOutDay = fct_inorder(CheckOutDay)) %>%
+     summarise(Percent = percent(mean(promoter)))
+   CheckOutPromoter <- as.data.frame(CheckOutPromoter)
+   return(CheckOutPromoter[order(desc(CheckOutPromoter$Percent)),])
+ }
> CheckOutPromotePercent()
  CheckOutDay Percent
2    Saturday   72.0%
1     Tuesday   71.7%
6      Sunday   71.5%
4      Monday   70.9%
5      Friday   68.5%
3    Thursday   67.9%
7   Wednesday   67.1%
```

- What attributes are best at predicting if a guest will be a promoter?

MarketGroup
CheckInDate
CheckOut
LOS
PurposeOfVisit
QuotedRate
Revenue
GuestCountryHome
Gender
AgeRange
LR_Score
OS_Score
GR_Score

T_Score
HC_Score
CS_Score
SC_Score
IS_Score
CI_Score
FB_Frequency
FBE_Score
HotelAbbrv
AwardCat
HotelCity
HotelBrand
Club
Category
Type
Class
Location
Relationship
GPTier
Channel
NPS
CheckInDay
CheckOutDay



Recommendations

Orange recommends Hyatt gather more data and places a stronger emphasis on making fields required to avoid the large number of NAs that were present in the data. This can be achieved by offering the surveys online. While this may reduce the number of surveys it will at least provide better data, which will ultimately lead to better insights and decision making.
If Hyatt is interested in expanding its reach, Orange recommends expansions to the Midwest and West with an understanding and objective to place hotels in suburban settings. Franchising these hotels should be avoided as it appears franchised hotels have an association to the Bottom 10 hotels.