R Notebook

1. Cover Page

Data Mining and Predictive Analytics

Project Title : Exploratory and Predictive Analytics of Airbnb listings to yield high booking rate

Market assigned to team : Las Vegas

"We, the undersigned, certify that the report submitted is our original work; all authors participated in the work in a substantive way; all authors have seen and approved the report as submitted; the text, images, illustrations, and other items included in the manuscript do not carry any infringement / plagiarism issue upon any existing copyrighted materials."

	Names of the signed team members
Contact Member	`Suchit Sanghvi`
Team Member 2	`Kaivan Shah`
Team Member 3	`Divya Vakil`
Team Member 4	`Aniket Datar`
Team Member 5	`Ronit Motiwale`
Team Member 6	`Prachit Puranik`

---

2. Executive Summary

The aim is to help an investor make a data driven decision to invest in property to list it as an airbnb that would yield high booking rates in Las Vegas. We conducted a market research registering the pros and cons of Las Vegas as an airbnb hub. The first priority was to analyze factors that make it a favorable property such as location, type and size. This analysis was a blend of market driven and data driven effort to narrow down the options. Once the type of property was finalized, the focus was to help the investor decide on the management style, which would alleviate the customer engagement This analysis is completely based on the kind of user experience the customers are expecting when they visit Vegas. We looked into the prudent charecteristics of the host such as customer interaction on the platform and the kind of amenities offered. This was done by inspecting and consolidating the most frequently asked for services. On understanding the important variables, we built a predictive model which would help envision a good investment. We would be analysing in detail on which is the most favorable configuration of a property.

This process was carefully crafted taking multiple variable combinations and iterations into consideration which would be in the investor's best interest. This two part analysis can be used as a tool to understand importance of underlying elements in the Las Vegas airbnb market and predict the possibility of an investment being a successful one.

---

3. Research Questions

Las Vegas is the Entertainment capital of the World. It has lot of attractions which are situated in and around the Strip which is stretched over 4.2 miles. This lane is the main hub of entertainment and lot of tourists that are visiting Vegas essentially would want to stay as close to the Strip as possible. From business point of view, it was crucial to identify the localities which would be profitable and attractive so as to generate maximum traction. From our initial analysis we found that not every locality was generating similar traction which is why we came to our first Research Question

1. Which will be a favorable location for buying the property?

It is very important to invest in the right property because in the long run the locality of the property and its proximity to the popular areas will play a part in determining if the investment succeeded. Which is why it was a crucial question. Taking this area of research further, it was important to consider which type of properties make a sensible investment from an ROI point of view. After diving further into the visitor demographics, we found out that the average age of visitors in Las Vegas has been around the 45 years mark and there has been a trend of huge proportion of visitors being married and having an income of more than $40k. Also, from our market research we found out that over the past few years, the percentage of people visiting Vegas in groups had been as low as 15%-19%, which indicated that there was a huge market from comparatively smaller properties. This led us to our subsequent two sub questions:

i. What type of property should the investor buy?
ii. What should be the property layout?

After working on investing in the perfect property, we delved further into the Customer Satisfaction domain, since a wonderful place to stay coupled with an even better host could be the key towards generating large revenue. When booking an Airbnb, customers are looking for a great place to stay and feel welcomed at the perfect price. So, it was important for the host to be at a certain standard because in Las Vegas, the law demands owner to stay on the property. This would also mean that the host should look reliable. This led us to our Second main question

2. What should be the ideal host benchmarks?

We looked at the data we had and agreed upon working on certain aspects such as the host's response time, their verified status, the number of listing they have, if the host is a Superhost and the overall rating. Apart from all these variables, Amenities was also a very important factor in determining which property had a high booking rate and there were certain amenities that appeared a lot more frequently than others. Also, a lenient cancellation policy ensures calmness amongst the customers. We thought that it might be a critical aspect of Customer Satisfaction. That was how we ended up on our next two sub questions:

i. What are the essential amenities to be provided?
ii. What should be the optimal reservation policy?

All these questions essentially enveloped the purpose of our Business Case along with certain assumptions to keep in mind regarding the laws towards AirBnb and short term booking in Las Vegas, and some other functionalities such as Transit and finding the Best time to invest which were solved by Domain Knowledge and Data to a certain extent.

---

4. Methodology

Importing Data

The data (dfTrain) consists of approximately 150k Airbnb listings in the United States as of January, 2020 with 66 variables. Below we extract the data for Las Vegas Market through the use of Random Control Variable.

marketTrain is our Las Vegas data with 6296 listings and 66 variables.

Quick inspection of Las Vegas data shows us the data conversion, cleaning and pre-processing needed. The following section and code will go through some data cleaning steps.

Data Exploration and Pre-processing

We will start with selecting features based on our domain knowledge from the data that will help serve our objective and answer our research questions. The choice of these variables was also influenced by our work on the Kaggle Competition. Our learning of universal importance of some variables influenced us to add them to our market specific model as well.

Below is the list of the variables that we selected:

Accommodates, amenities, bathrooms, bedrooms, beds, neighborhood_overview, property_type, room_type, hostSince, hostVerifications, host_listings_count, host_response_time, host_is_superhost, host_has_profile_pic, host_identity_verified, maximum_nights, minimum_nights, price, cleaning fee, require_guest_phone_verification, security_deposit, cancellation policy, transit, instant_bookable

Cleaning Step: Converting character to numeric and numeric to factor

Variables Cleaned: price, security_deposit, cleaning_fee, high_booking_rate

Sample Code:

```
marketTrain$price <- as.numeric(gsub('\\$|,', '', marketTrain$price))
```

Cleaning Step: Dealing with missing values on the basis of domain knowledge(for example: missing values in cleaning fee variables were filled with zero assuming host did not take any addition fees)

Variables Cleaned: cleaning_fee, bathrooms, bedrooms, security_deposit, host_listings_count, host_is_superhost, beds

Sample Code:

```
marketTrain$host_identity_verified[is.na(marketTrain$host_identity_verified)]
= 0
```

Cleaning Step: Creating dummy variables for categorical fields

Variables Cleaned:

- amenities(for significant ones found through kaggle part) - Dyer, Hair Dryer, Laptop friendly workspace, Hot water, Heating
- Transit - dummyTransit will be ‘1’ if the variable is not empty and ‘0’ otherwise
- host_Verifications - dummyHostVerificationsReviews, dummyHostVerificationsFB and dummyHostVerificationsReviews created for ‘reviews’, ‘FB’, ‘KBA’ in host_verification variable

Sample Code:

```
marketTrain <-  marketTrain %>% mutate(dummyTransit = ifelse(is.na(transit),
0, 1))

dummyHairDryer  <-  as.numeric(sapply("\"Hair dryer\"", grepl,
marketTrain$amenities))
marketTrain <-  marketTrain %>% mutate(dummyHairDryer)
```

Cleaning Step: Convertting character variables into categorical variables and grouping infrequent categories to ‘other’ categories.

Variables Cleaned:

- Cancellation Ploicy - grouping categories with less than ten percent data to ‘other’ category.
- Property Type - grouping categories with less than three percent data to ‘other’ category.
- Host Response Time - grouping categories with less than ten percent data or ‘N/A’ values to ‘other’ category.

Sample Code:

```
marketTrain$cancellation_policy <-
as.character(marketTrain$cancellation_policy)

othervar <-  as.list(marketTrain %>% group_by(cancellation_policy) %>%
tally() %>%
  mutate(pct = n/sum(n)*100) %>% arrange(desc(pct)) %>% filter(pct<10) )[1]

marketTrain <-  marketTrain %>% mutate(cancellation_policy =
if_else(cancellation_policy %in% unlist(othervar), "Other",
cancellation_policy))
```

Cleaning Step: Converting variables to numerical variables.

Variables Cleaned:

- Amenities - amenitiesCount will have a count of total number of amenities
- Host_Since - hostDays will have a count of total number of days, the host has been on Airbnb

Sample Code:

```
# amenitiesCount

marketTrain <- marketTrain %>% mutate(amenitiesCount =
as.numeric(lapply(marketTrain$amenities, function(x) (str_count(x, ",") +
1))))

# hostDays
marketTrain <- marketTrain %>% mutate(hostDays = as.numeric(Sys.Date() -
host_since))

# replacing na values by the average value of hostDays
marketTrain$hostDays[is.na(marketTrain$hostDays)] = 1394
```

Exploratory Analysis

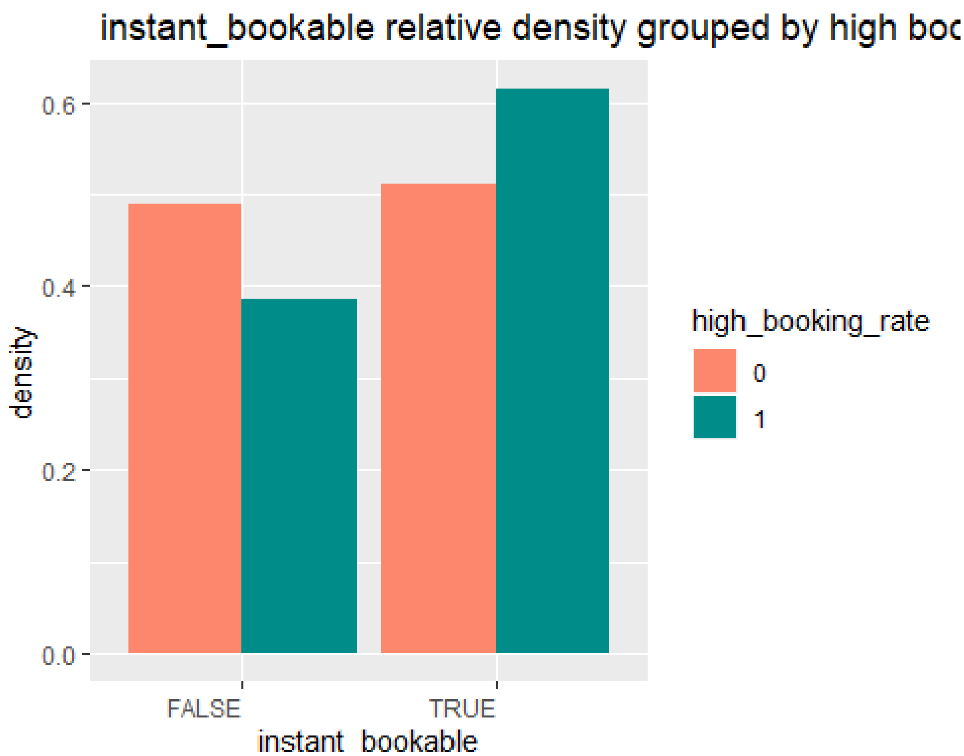
We will now visualise distribution of some relevant variables that we have selected, breaking them down by popularity.We will also provide the interpretation of these variables based on the plot.

1 - Discrete Variable (instant_bookable, dummyTranist, host_is_superhost, host_has_profile_pic, host_identity_verified, require_guest_phone_verification)

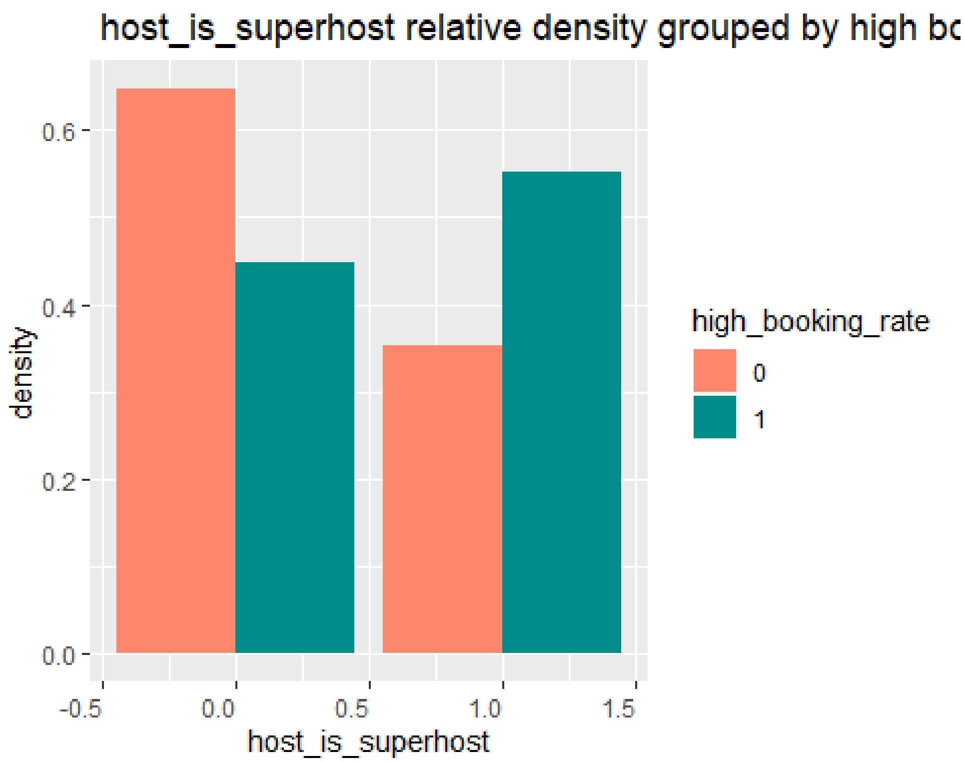
Variable Visualised

- Instant Bookable
- host_is_superhost
- host_identity_verified

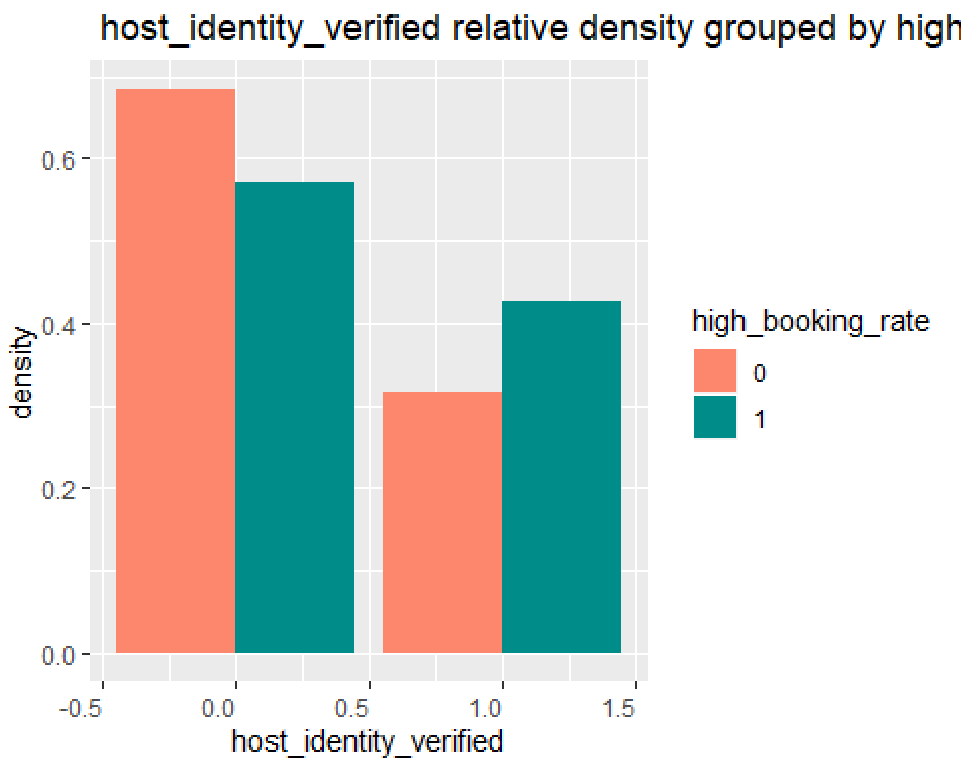
```
 #  "instant_bookable"

    temp <- subset(marketTrain, high_booking_rate == 1)
    temp <- temp %>% group_by(high_booking_rate , instant_bookable) %>%
      summarise(density = n()/nrow(.))

    temp1 <- subset(marketTrain, high_booking_rate == 0)
    temp1 <- temp1 %>% group_by(high_booking_rate, instant_bookable) %>%
      summarise(density = n()/nrow(.))

    temp2 <- rbind(temp, temp1)

    plotHT <- ggplot(data=temp2, aes(x=instant_bookable, y=density,
fill=as.factor(high_booking_rate))) + geom_bar(position = 'dodge',
stat='identity') + labs(fill = "high_booking_rate", x = "instant_bookable",
title = paste(" instant_bookable relative density grouped by high booking
rate")) +           theme(axis.text.x = element_text(angle = 0, hjust = 1)) +
scale_fill_manual(values= c("#FD876C","#008D8A"))

    print(plotHT)
```




```r
    #  "host_is_superhost"

    temp <- subset(marketTrain, high_booking_rate == 1)
    temp <- temp %>% group_by(high_booking_rate , host_is_superhost) %>%
      summarise(density = n()/nrow(.))

    temp1 <- subset(marketTrain, high_booking_rate == 0)
    temp1 <- temp1 %>% group_by(high_booking_rate, host_is_superhost) %>%
      summarise(density = n()/nrow(.))

    temp2 <- rbind(temp, temp1)

    plotHT <- ggplot(data=temp2, aes(x=host_is_superhost, y=density,
fill=as.factor(high_booking_rate))) + geom_bar(position = 'dodge',
stat='identity') + labs(fill = "high_booking_rate", x = "host_is_superhost",
title = paste(" host_is_superhost relative density grouped by high booking
rate")) +           theme(axis.text.x = element_text(angle = 0, hjust = 1)) +
scale_fill_manual(values= c("#FD876C","#008D8A"))

    print(plotHT)
```




```
    #  "host_identity_verified"

    temp <- subset(marketTrain, high_booking_rate == 1)
    temp <- temp %>% group_by(high_booking_rate , host_identity_verified) %>%
      summarise(density = n()/nrow(.))

    temp1 <- subset(marketTrain, high_booking_rate == 0)
    temp1 <- temp1 %>% group_by(high_booking_rate, host_identity_verified)
%>%
      summarise(density = n()/nrow(.))

    temp2 <- rbind(temp, temp1)

    plotHT <- ggplot(data=temp2, aes(x=host_identity_verified, y=density,
fill=as.factor(high_booking_rate))) + geom_bar(position = 'dodge',
stat='identity') + labs(fill = "high_booking_rate", x =
"host_identity_verified", title = paste(" host_identity_verified relative
density grouped by high booking rate")) +           theme(axis.text.x =
element_text(angle = 0, hjust = 1)) + scale_fill_manual(values=
c("#FD876C","#008D8A"))

    print(plotHT)
```



Interpretation:

instant_bookable: When Properties are instantly bookable they have higher booking rate

host_is_superhost : When host is super host, then properties have higher probability of getting booked

host_identity_verified : If host identity is verified then it has higher probability of getting booked

---

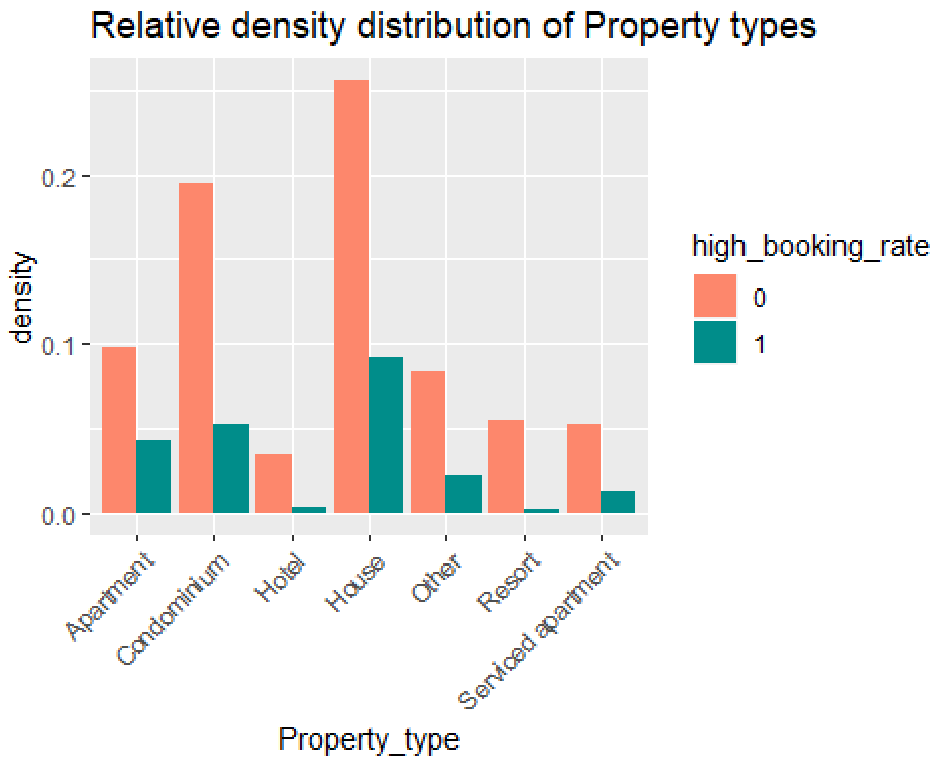
2 - Categorical Variable (property_type, cancellation_policy, host_response_time, room_type, host_verifications)

Variable Visualised

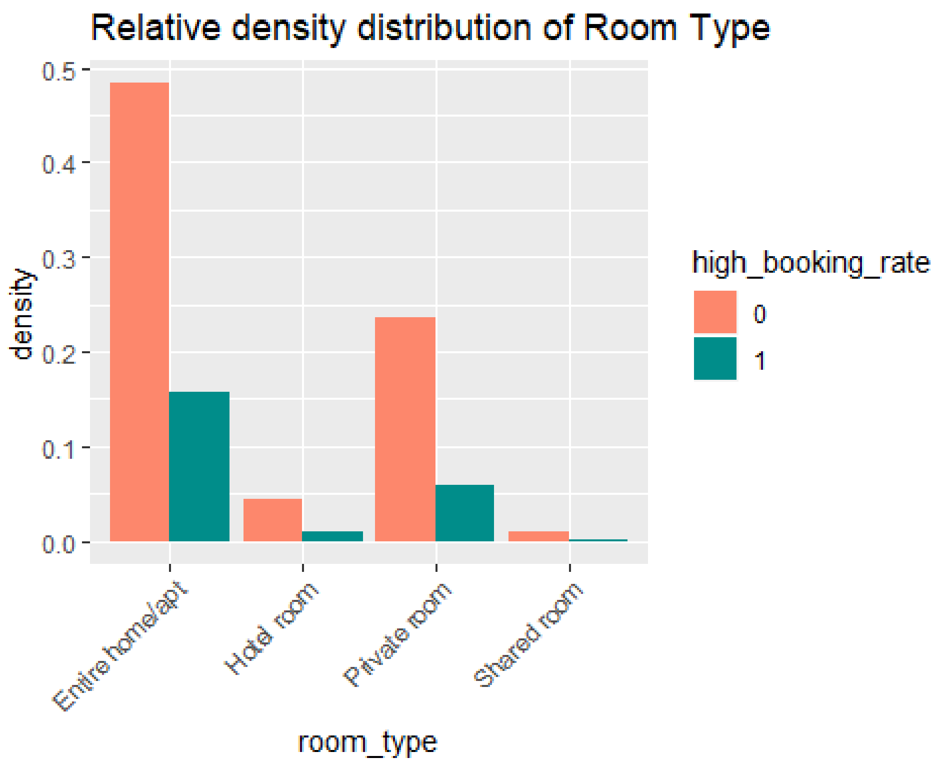
- property_type
- room_type
- host_response_time
- cancellation_policy

```
#"property_type"


   marketTrain %>%  group_by(high_booking_rate,property_type ) %>%
summarise(density = n()/nrow(.)) %>%
  ggplot(aes(x=as.factor(property_type), y=density,
```

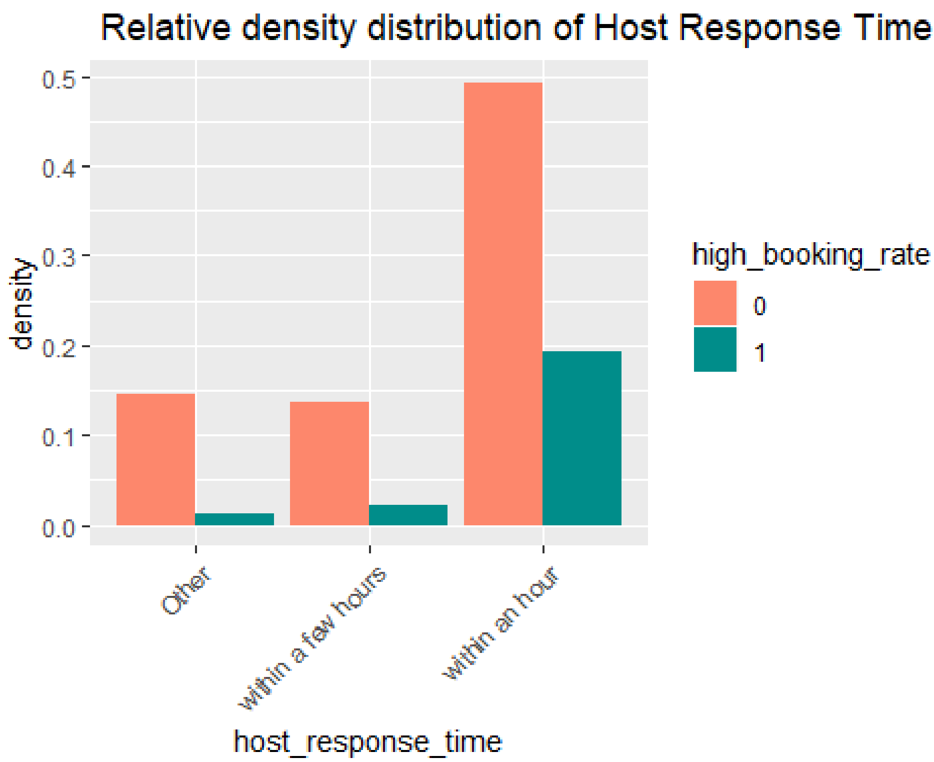
```
fill=as.factor(high_booking_rate))) +
      geom_bar(position = 'dodge', stat='identity') + labs(fill =
"high_booking_rate", x = "Property_type",
            title = paste("Relative density distribution of Property types"))
+scale_fill_manual(values=c( "#FD876C","#008D8A"))+
            theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



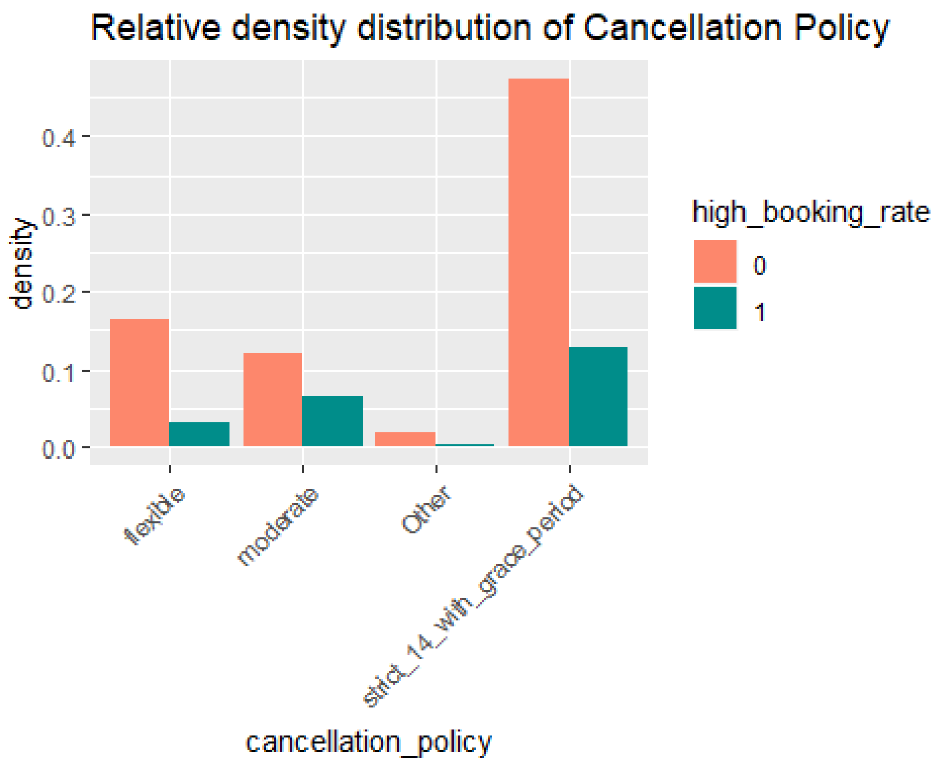

```
 #"room_type"

    marketTrain %>%  group_by(high_booking_rate,room_type ) %>%
summarise(density = n()/nrow(.)) %>% ggplot(aes(x=as.factor(room_type),
y=density, fill=as.factor(high_booking_rate))) + geom_bar(position = 'dodge',
stat='identity') + labs(fill = "high_booking_rate", x = "room_type",
            title = paste("Relative density distribution of Room Type "))
+scale_fill_manual(values=c( "#FD876C","#008D8A"))+
            theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
#  "host_response_time"

    marketTrain %>%  group_by(high_booking_rate,host_response_time ) %>%
summarise(density = n()/nrow(.)) %>%
ggplot(aes(x=as.factor(host_response_time), y=density,
fill=as.factor(high_booking_rate))) + geom_bar(position = 'dodge',
stat='identity') + labs(fill = "high_booking_rate", x = "host_response_time",
            title = paste(" Relative density distribution of Host Response
Time")) +scale_fill_manual(values=c( "#FD876C","#008D8A"))+
            theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
    # 'cancellation_policy'

    marketTrain %>%  group_by(high_booking_rate,cancellation_policy ) %>%
summarise(density = n()/nrow(.)) %>%
ggplot(aes(x=as.factor(cancellation_policy), y=density,
fill=as.factor(high_booking_rate))) + geom_bar(position = 'dodge',
stat='identity') + labs(fill = "high_booking_rate", x =
"cancellation_policy",              title = paste("Relative density
distribution of Cancellation Policy")) +scale_fill_manual(values=c(
"#FD876C","#008D8A"))+
            theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Interpretation:

property_type: High Booking Rate is represented in House, Condominium and Apartment and very small ratio of service apartment.

room_type: Entire home/apt and private room has the most High Booking Rate which is evident from the plot. Also shared room has absolutely zero high booking rate.

host_response_time: Host response time being within an hour shows High Booking Rate

cancellation_policy: strict_14_with_grace_period and moderate categories of cancellation policy has High Booking Rate as compared to other categories

---

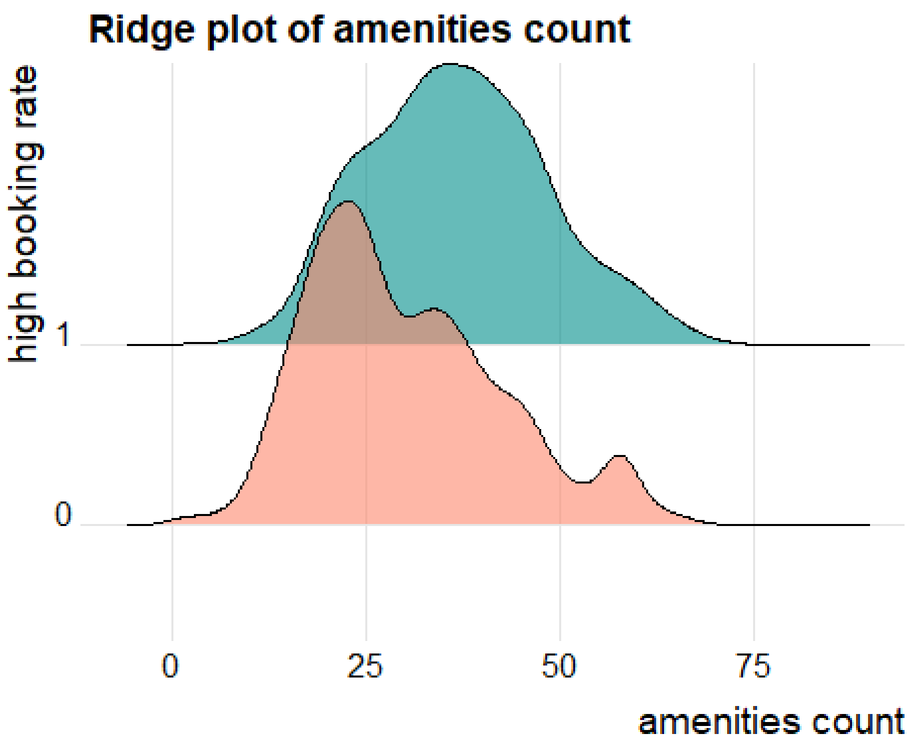
3 - Continuous Variable (amenitiesCount, accommodates, bedrooms, bathrooms, beds, hostSince, price, security_deposit, cleaning_fee)

Variable Visualised:

- amenitiesCount
- price

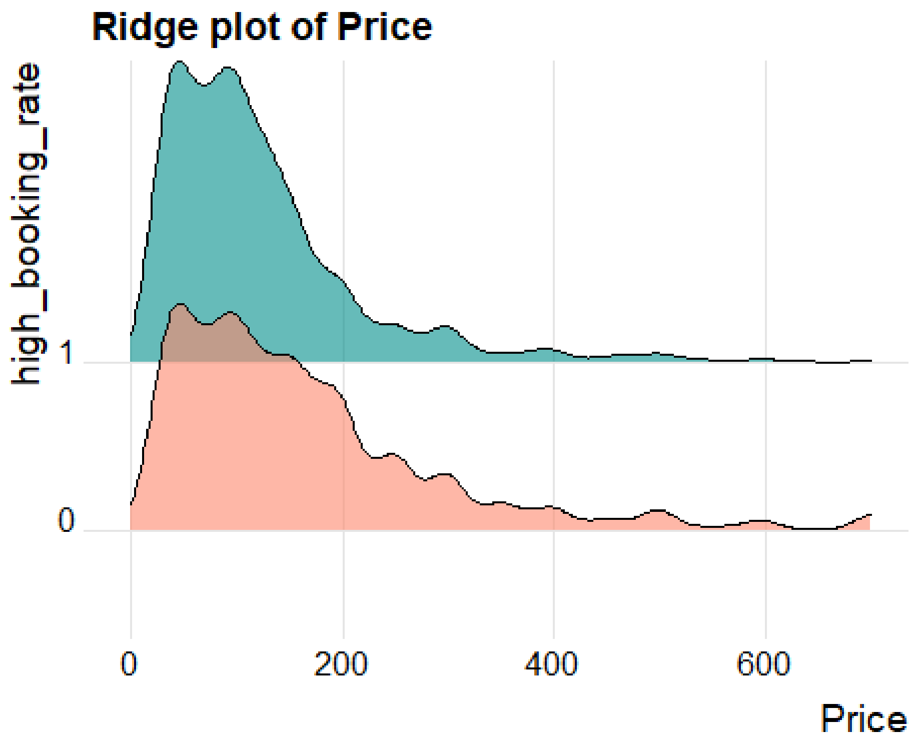
```
# Amenities Count

amenityCount_plot <- ggplot(data= marketTrain , aes(x=amenitiesCount,
y=as.factor(high_booking_rate),
fill=as.factor(high_booking_rate)))+geom_density_ridges(alpha=0.6,
```

```
bins=20)+theme_ridges()+theme(legend.position = "none")+labs(fill =
"high_booking_rate", x = "amenities count",y="high booking rate", title =
paste("Ridge plot of amenities count"))+scale_fill_manual(values=c(
"#FD876C","#008D8A"))

## Warning: Ignoring unknown parameters: bins

amenityCount_plot

## Picking joint bandwidth of 2.27
```




```
# Price

price_plot <- ggplot(data= marketTrain , aes(x=price,
y=as.factor(high_booking_rate),
fill=as.factor(high_booking_rate)))+geom_density_ridges(alpha=0.6,bins =
0.5)+theme_ridges()+theme(legend.position = "center")+labs(fill =
"high_booking_rate", title = paste("Ridge plot of Price"))+ xlab("Price") +
ylab("high_booking_rate") + scale_fill_manual(values=c(
"#FD876C","#008D8A"))+scale_x_continuous(limits=c(0, 700))

## Warning: Ignoring unknown parameters: bins

price_plot

## Picking joint bandwidth of 15.4
```

```
## Warning: Removed 249 rows containing non-finite values
(stat_density_ridges).
```



Interpretation:

Amenities: Properties with higher amenities are more likely to get booked

Price : It shows that properties which are booked more have low prices.

Predictive Analysis

Model 1 - Logistic Model

We will now build a logistic regression model that will take our selected variables as input parameters and will give a probability of a listing getting a high booking rate. For this purpose, we will divide our data into train set(80%) and test set(20%). We will then build our model on the train set and run it on test set to measure accuracy.

```r
# Splitting the data into train and test set

set.seed(123)
lvTrain <- marketTrain %>% sample_frac(size = 0.8)
lvTest <- dplyr::setdiff(marketTrain, lvTrain)
```

```r
# Building Logistic Regression Model

lvfitGLM <-  glm(formula =   high_booking_rate ~   cleaning_fee + bathrooms +
bedrooms + instant_bookable + dummyHostVerificationsFB  +
dummyHostVerificationsKBA + dummyHostVerificationsReviews +
host_listings_count +  dummyHairDryer + dummyTransit  + amenitiesCount  +
room_type + security_deposit + dummyCable + host_response_time  +
cancellation_policy + price + accommodates +  host_is_superhost +
minimum_nights + maximum_nights + dummyDryer + dummyHotWater + property_type
+ host_has_profile_pic + host_identity_verified , family = 'binomial', data =
lvTrain)
```

Model 1 - Random Forest Model

Now we will build a Random Forest Model. Random Forest is an algorithm that consists of building a large number of decision trees together. Each decision tree gives an output and the class with the most votes will become our models prediction for a particular airbnb listing. Random Forest will help us get more accurately predict whether or not a certain property will give us a high booking rate in future.

After converting variables to make them suitable to run random forest, we build the model as shown below.

```r
lvfitRF <-  train(high_booking_rate ~ cleaning_fee + bathrooms + bedrooms +
instant_bookable + dummyHostVerificationsFB  + dummyHostVerificationsKBA +
dummyHostVerificationsReviews + host_listings_count +  dummyHairDryer +
dummyTransit  + amenitiesCount  + room_type + security_deposit + dummyCable +
host_response_time  + cancellation_policy + price + accommodates +
host_is_superhost + minimum_nights + maximum_nights + dummyDryer +
dummyHotWater + property_type + host_has_profile_pic +
host_identity_verified, data = lvTrainRF, method='rf')
```

---

5. Results and Findings

Our first objective was to select the right property to acheive a high booking rate in future. To serve this objective we needed to answer the type of property that the investor would need to buy and the layout of the property in terms of bedrooms, bathrooms, etc.

Our logistic model will help us understand how different features that we added in the model relates with the probability of getting a high booking rate.

Coefficients of variables such as property_type and room_type from the logistic regression will tell us what property type or room type increases or decreases the probability of getting a high booking rate. In a similar way, our model will help us answer our questions to buy the

right property and manage it effectively by telling us about how different featutres affect our booking rate.

Model Performance Evaluation

Now, we need measures that will tell us about how well our model predicts the different types of classes. That is, how accurately the model classifies the airbnb listing as high booking listing.

Before investing, we need to be sure about the property that we are placing our bet on. For this purpose, we are selecting Specificity as our primary performance metric. We choose Specificity to make sure that most of our decisions are correct and we don't wrongly identify a property as a potential highly booked listing.

On testing our models at various cutoffs, we selected 0.5 as our final cutoff values based on the results and our objective. We wanted to make sure that our model rarely classifies a low booked property as highly booked.

Logistic Model Performance - Confusion Matrix

```
resultslvGLM %>%
  xtabs(~predictedClass+high_booking_rate, .) %>%
  confusionMatrix(positive = '1')
```

```
## Confusion Matrix and Statistics
##
##               high_booking_rate
## predictedClass   0   1
##              0 900 175
##              1  79 105
##
##                Accuracy : 0.7983
##                  95% CI : (0.775, 0.8201)
##     No Information Rate : 0.7776
##     P-Value [Acc > NIR] : 0.04081
##
##                   Kappa : 0.3354
##
##  Mcnemar's Test P-Value : 2.51e-09
##
##             Sensitivity : 0.3750
##             Specificity : 0.9193
##          Pos Pred Value : 0.5707
##          Neg Pred Value : 0.8372
##              Prevalence : 0.2224
##          Detection Rate : 0.0834
##    Detection Prevalence : 0.1461
```

```
##       Balanced Accuracy : 0.6472
##
##        'Positive' Class : 1
##
```

Logistic Model Performance - AUC Score

```
resultslvGLM %>%
 roc_auc(truth = high_booking_rate, predictedProb)

## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 roc_auc binary         0.810
```

As seen above, out Logistic Model gives us Specificity of 0.92531. There are only 77 listings that had low booking rates however model classified as highly booked. We used AUC metric to set cutoff at the most optimal value of sensitivity and specificity. Final AUC value of our model was 0.8101342.

Random Forest Model Performance - Confusion Matrix

```
resultslvRF %>%
  xtabs(~predictedClass+high_booking_rate, .) %>%
  confusionMatrix(positive = '1')

## Confusion Matrix and Statistics
##
##               high_booking_rate
## predictedClass   0   1
##              0 903 138
##              1  76 142
##
##                Accuracy : 0.83
##                  95% CI : (0.8081, 0.8504)
##     No Information Rate : 0.7776
##     P-Value [Acc > NIR] : 2.412e-06
##
##                   Kappa : 0.4664
##
##  Mcnemar's Test P-Value : 3.048e-05
##
##             Sensitivity : 0.5071
##             Specificity : 0.9224
##          Pos Pred Value : 0.6514
##          Neg Pred Value : 0.8674
##              Prevalence : 0.2224
##          Detection Rate : 0.1128
##    Detection Prevalence : 0.1732
##       Balanced Accuracy : 0.7148
##
```

```
##        'Positive' Class : 1
##
```

Random Forest Model Performance - AUC Score

```
resultslvRF %>%
 roc_auc(truth = high_booking_rate, high_booking_rate_new)

## # A tibble: 1 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 roc_auc binary         0.876
```

While there isn't a significant difference, our Random Forest Model helps us capture better Accuracy, AUC Score and slightly better Specificity as seen from the above output.

---

6. Conclusion

Consolidating the discoveries from market research, exploratory analysis and data modeling, we can summarize

1. Location : After a brief analysis on data according to the location, there weren't any significant results. However, according to market research, recommendation for the investor would be to purchase a property near prime locations in the city such as the Strip, Downtown and Fremont Street.

2. Property Characteristics : By observing the property characteristics, it tends to be exhorted that an investor should plan to go for property types such as Houses, Apartments and Condos with 2-5 bedrooms that acommodate 3-8 guests.

3. Amenities : Regarding the amenities provided, it would be good for the proprietor to provide around 20 to 40 amenities. The proprietor should also make sure that amenities like AC, Heating, Dryer, Hot Water, Cable and Wifi are made accessible for a higher booking rate.

4. Transit : It is prudent for the property to be in localities from where the availability of public transit is good. Likewise, there are Uber and Lyft rides easily available to get to the Strip.

5. Best time to invest: The best time to invest from a cost perspective would be in January. The cost is the lowest in January. However, the options for buying property are relevantly less. The month of September sees a great deal of alternatives opening up. With the options, the prices are higher than the month of January. Subsequently, if the investor is eager to spend liberally, then September would offer an extraordinary number of property alternatives.

6. Management : In terms of customer satisfaction, correspondence with the client is vital for the host as and when required. Furthermore, factors like instant booking and a quick response time are helpful.

Future research

it is prudent to utilize the logistic model to get insights on factors that are being thought of and their relation to getting higher booking rates. The ensemble model ought to be utilized for getting precision to determine the booking rate of the property alternatives. The primary drivers of the Las Vegas economy are Tourism, Gaming and Conventions, which in turn feed the Retail and Restaurant industries. In this way, the travel industry segment may be a good factor to be considered later for investigation as it can radically influence the outcomes. The logistic model can be utilized on informational purpose for an alternate city as well.

Limitations

Laws in Las Vegas such as short term stay and owners staying in the property added a few restrictions on the research. Factors such as political, neighborhood features are bound to change which can affect where a property can be listed as an Airbnb. The ensemble (random forest) model is ideal for Vegas but not as useful for a different. The reason to this can be attributed to the high accuracy of the ensemble method and ought to be checked for overfitting similar influenced features in the new location where the model is utilized for prediction.

---

7. References:

1. https://listwithclever.com/real-estate-blog/best-time-to-buy-a-house-in-las-vegas/
2. https://www.mashvisor.com/blog/airbnb-las-vegas/
3. https://files.lasvegasnevada.gov/business-licensing/Short-Term-Residential-Rental-Business-General-Information-Handout-BL255.pdf
4. https://matadornetwork.com/destinations/north-america/united-states/las-vegas/7-coolest-neighborhoods-las-vegas-stay-eat-party/
5. https://www.investopedia.com/articles/personal-finance/031815/hotels-vs-airbnb-vegas-visitors.asp
6. https://www.mashvisor.com/blog/guide-airbnb-investment-property/
7. http://www.vegascasinoinfo.com/vegas-busy-months-and-slower-months/
8. https://www.mashvisor.com/blog/13-tips-high-airbnb-occupancy-rate/
9. https://assets.simpleviewcms.com/simpleview/image/upload/v1/clients/lasvegas/2018_Las_Vegas_Visitors_Profile_Study_94443c1d-334f-4d0b-b997-5c8800f990b0.pdf
10. https://www.lvcva.com/stats-and-facts/