Problem Solving

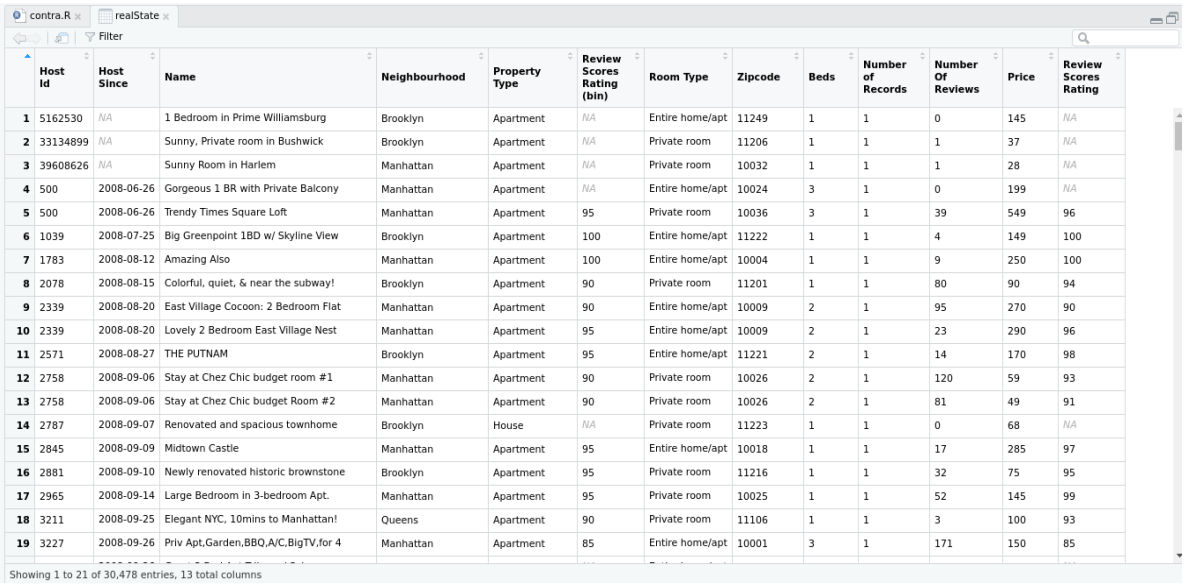
(NAME)
(REGISTRATION NUMBER)

1.0 Introduction

For the data analysis purpose there are number of popular tools and application that can be used. Considering the performance ratings and user experiences R-Studio is effective and poplar tool among the community of the data science. R Studio contains more attractive and effective features when we compare with the competitors. Considering the programming languages like python and other analyzing languages R programming stand at the top of the performance list because of the incredible performance. Specialty with the R programming is users can simply run the commands line by line and do no have to compile the entire document for the outputs. That saves a lot of time and effort in the programming manner.

Apart from all that R Studio allow users to power up the R programming platform and in that way users can freely manage the way of using the visualization tools and other additional features with a minimum time period. Analyzing features in the R studio tool is very reliable and put the flags on the precise locations in the data sets that are questionable. Along with that R programming and R studio consist of vector operations which can be helpful in a considerably larger operations. Which means users do not have to so many loops and repetitive operations to go through the document. Well supported community and so many packages are to make sure that the users are not left alone in the data science.

I have selected a data set from a real state business in USA. This data set represents all the apartments that have been owned by the business and the customer feedback as well as the description. The data set has imported to the R Studio and rest of the analyzing portions are based on this data set.



contra.R | realState

Filter

	Host Id	Host Since	Name	Neighbourhood	Property Type	Review Scores Rating (bin)	Room Type	Zipcode	Beds	Number of Records	Number Of Reviews	Price	Review Scores Rating
1	5162530	NA	1 Bedroom in Prime Williamsburg	Brooklyn	Apartment	NA	Entire home/apt	11249	1	1	0	145	NA
2	33134899	NA	Sunny, Private room in Bushwick	Brooklyn	Apartment	NA	Private room	11206	1	1	1	37	NA
3	39608626	NA	Sunny Room in Harlem	Manhattan	Apartment	NA	Private room	10032	1	1	1	28	NA
4	500	2008-06-26	Gorgeous 1 BR with Private Balcony	Manhattan	Apartment	NA	Entire home/apt	10024	3	1	0	199	NA
5	500	2008-06-26	Trendy Times Square Loft	Manhattan	Apartment	95	Private room	10036	3	1	39	549	96
6	1039	2008-07-25	Big Greenpoint 1BD w/ Skyline View	Brooklyn	Apartment	100	Entire home/apt	11222	1	1	4	149	100
7	1783	2008-08-12	Amazing Also	Manhattan	Apartment	100	Entire home/apt	10004	1	1	9	250	100
8	2078	2008-08-15	Colorful, quiet, & near the subway!	Brooklyn	Apartment	90	Private room	11201	1	1	80	90	94
9	2339	2008-08-20	East Village Cocoon: 2 Bedroom Flat	Manhattan	Apartment	90	Entire home/apt	10009	2	1	95	270	90
10	2339	2008-08-20	Lovely 2 Bedroom East Village Nest	Manhattan	Apartment	95	Entire home/apt	10009	2	1	23	290	96
11	2571	2008-08-27	THE PUTNAM	Brooklyn	Apartment	95	Entire home/apt	11221	2	1	14	170	98
12	2758	2008-09-06	Stay at Chez Chic budget room #1	Manhattan	Apartment	90	Private room	10026	2	1	120	59	93
13	2758	2008-09-06	Stay at Chez Chic budget Room #2	Manhattan	Apartment	90	Private room	10026	2	1	81	49	91
14	2787	2008-09-07	Renovated and spacious townhome	Brooklyn	House	NA	Private room	11223	1	1	0	68	NA
15	2845	2008-09-09	Midtown Castle	Manhattan	Apartment	95	Entire home/apt	10018	1	1	17	285	97
16	2881	2008-09-10	Newly renovated historic brownstone	Brooklyn	Apartment	95	Private room	11216	1	1	32	75	95
17	2965	2008-09-14	Large Bedroom in 3-bedroom Apt.	Manhattan	Apartment	95	Private room	10025	1	1	52	145	99
18	3211	2008-09-25	Elegant NYC, 10mins to Manhattan!	Queens	Apartment	90	Private room	11106	1	1	3	100	93
19	3227	2008-09-26	Priv Apt,Garden,BBQ,A/C,BigTV,for 4	Manhattan	Apartment	85	Entire home/apt	10001	3	1	171	150	85

Showing 1 to 21 of 30,478 entries, 13 total columns

Image 1: Data Set for the Analyzing

2.0 Visualizations of the data-set

2.1 Mean

In this data-set we have to look in to the pricing and rating of the Apartments. There are more than 30,000 inputs and the mean of the categories are calculated as follows.

```
> mean(realState$Price)
[1] 163.5897
```

Image 2: Mean of the data set

Here I have exercised the mean syntax in R programming in R Studio. Simply it gives the mean of the column that I have selected in the syntax in this case the prices of the apartments. There are limited number of objects that the visualization tools can be applied because some of the columns contain unfilled rows and lines which can leads to a false analytical summary.

2.2 Standard Deviation

Standard deviation is a comparison measurement that can be used to analysis the behavior of the data set in the future and mostly used in forecasting.

```
> sd(realState$Price)
[1] 197.7855
```

Image 3: Standard Deviation

This data set has a standard deviation of 197.7855 in the price section and I used sd syntax for the calculation.

2.3 Histogram

Histogram is a really good method of data visualization and in this scenario we have used several data columns for the representation purpose. Considering the amount of information given in the data set we can see that the variation between the reviews on the properties and the price of the properties are proportional to each other. So that if the business organization could more focus on the customer satisfaction and their feedback on the properties then they can achieve a better outcome in their business.

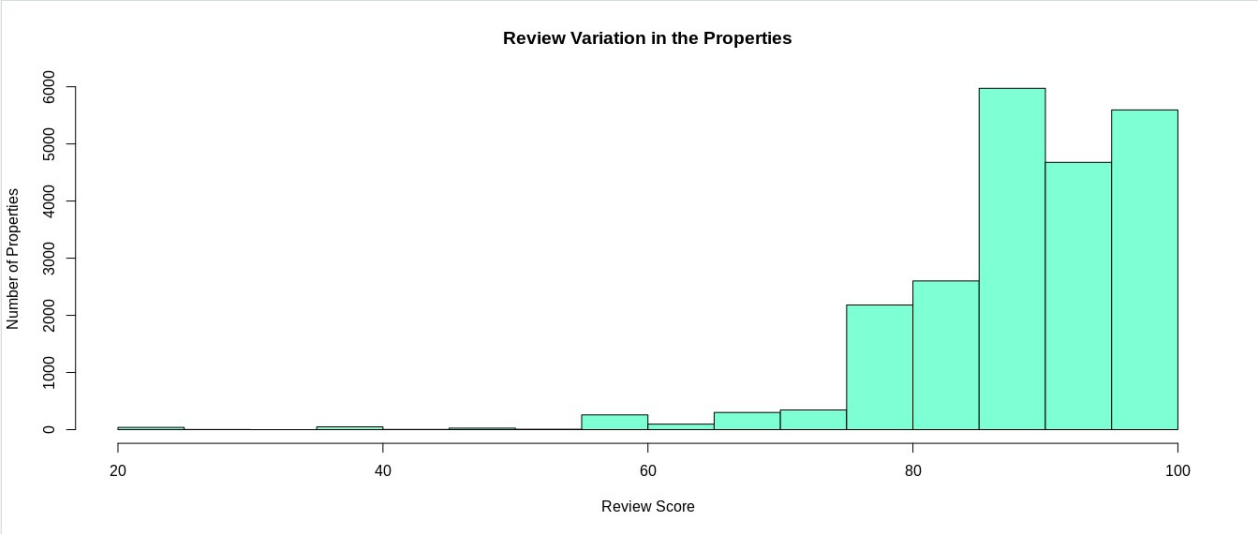


Image 4: Histogram - Review Score Vs Price

As you can see in this representation all the low reviewed places of the company owns are at a low level of the pricing scale. When we explore the data set there are considerably higher amount of customers are in that potion of the customers are in the low reviewed are of properties and because of the low demand on that segment organization cannot charge the higher amount of pay rates on those properties.

2.3 Scatter Diagram

Scatter diagram extracted from the R Studio by using the ggplot2 package and the visualizing environment. Same as in the histogram we can see the variation of the data along with the review and rating mechanism.

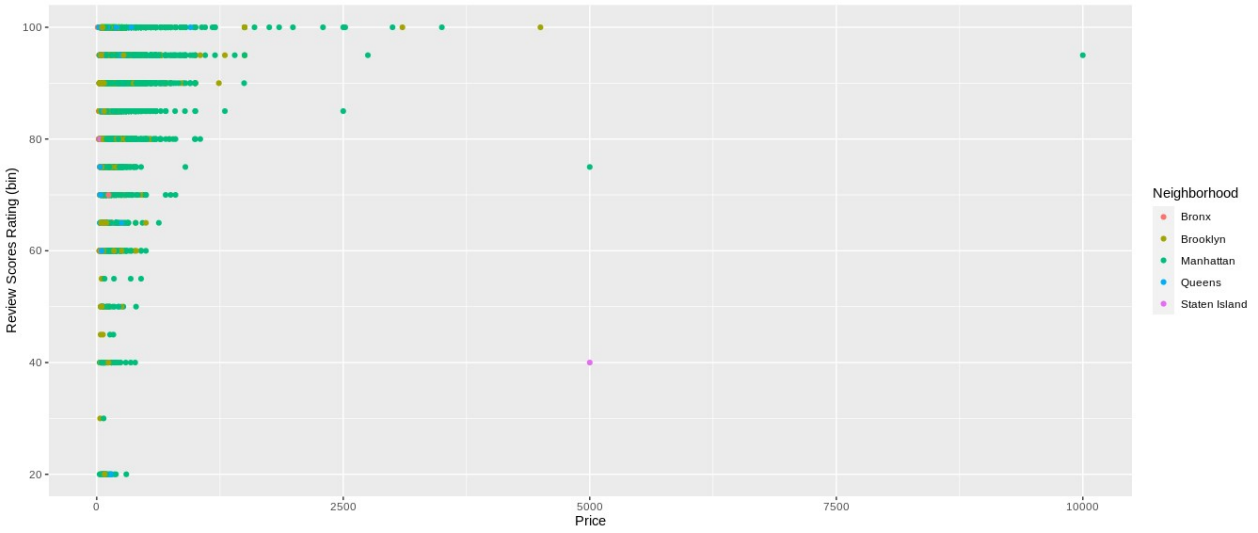


Image 5: Scatter Diagram

2.4 Box-plot

The box plot is very effective way of visualizing data according to the categorical factors. Mostly in the common and basic data visualizations we use non categorical data which has a controlled flow in the data paths and easy represent in the outputs. But categorical data should br converted as the non categorical data using the factor key word in the R programming.

In this scenario the location of the properties are counted as categorical data and a box plot is used represent them in a understandable manner to the receivers by not using the factoring feature.

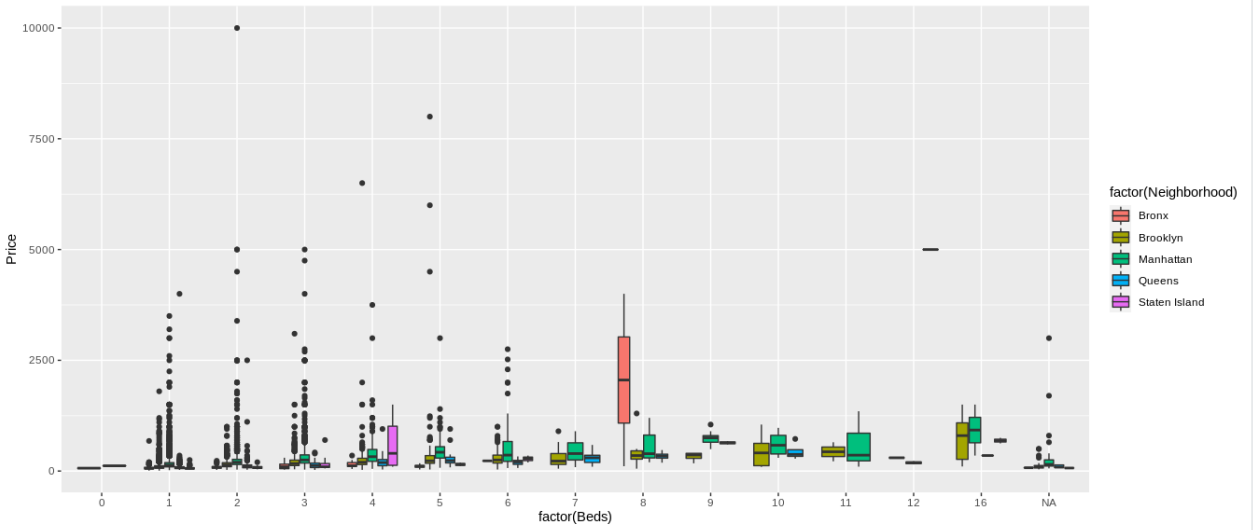


Image 6: Box Plot - Price Vs No of Beds Vs Location

In this figure we have obtained the box plot between price variation according to the size of the apartment and also we have considered the location of the apartment. The positive side of this representation is we can have a clear picture on our data set because we have used more than two variables and that gives a wider angle on the data set. At the same time there are some significant amount of outliers in the data set and using the box plot we can understand them in a more effective way.

Considering the outliers we can determine the customer choice according to the respective city. So that we can see when the size of the apartment is smaller than the average then the price variation is going higher in urban arias.

2.5 Polygon

Polygon is created using the histogram itself in a more delicate manner. In this case the it is created considering the amount of money people have to pay for a apartment. To make this information more useful the locations of the apartments has been added to the polygon.

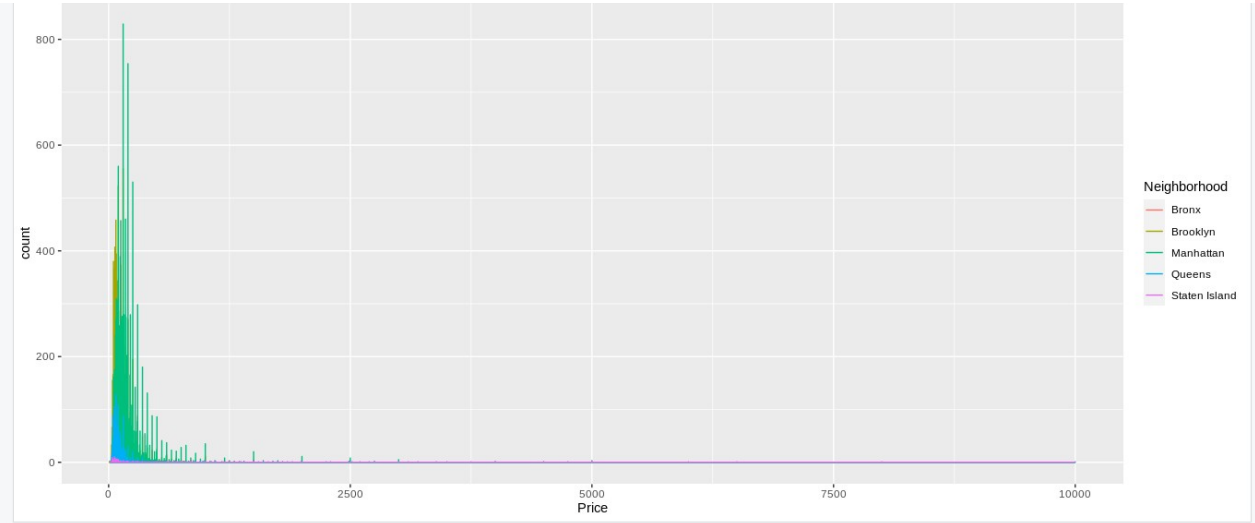


Image 7: Polygon - Price Vs Count and the Location

As we can see in the figure there are very high amount of demand for the apartments which are in the price range of the first 500 units. In that segment most of the places are located in the Manhattan Island or Queens.

3.0 Discussion

3.1 Problem Identification

This organization is based on urban real state business and there are some certain limitations that the company already faced. In this scenario there is some push back in the low price rated apartments. Surprisingly there are some considerable amount of drawbacks in the customer ratings in that segment of business. Therefore ultimately we can see there is a problem to be fixed in the organizational point of view in this segment.

So that the organization should see through a way to get customer attraction on the lower scale apartments and their facilities. Both the demand and the review going down indicates the low user satisfaction in the smaller apartments. But in general there is some good solid demand for the small sized apartments in the city and urban areas. As in the histogram and scatter diagram indicates there is a possible market for the features but there are not enough good conditioned apartments available for the customers. So that the organization should focus on funding the development sites more.

3.2 Solution

There are two reasons that the small scale apartment demand going down. One is the lack of infrastructure and conditions in the apartments. The next thing is the bad review and reputation on small scale apartments in the area.

So that the first thing the organization can focus on is getting the review average up. In order to do that the organization should concern on giving a good marketing strategy on the small scale apartments. Good marketing campaign can always change the way of the work line in an organization. In this way the organization can ensure a good crowd will come together for their products and services. The next thing is paying attention to the bad reviews. The bad reviews are made for a reason. If the organization can figure out the reason for the reviews then it will help through the recovery procedure. As an example one very possible reason for the bad reviews can be the bad conditions in the apartments. Low conditioned bathroom, bad air conditioning system are two most common failures in the apartments. The surrounding also matters to the customer when they are going to rent out a place for living. If the neighborhood is not that safer and calmer as the customers expected then the bad review effect can happen and the organization should think about the surrounding before they invest on the properties.

Understanding the situations and provide solution at the proper pot will save the organization from making losses. So ensure that company should focus on the marketing campaign and the review management mechanism more the less.

Talking about solutions company should consider the future market behaviour as well. There might be some significant scenarios that could be a effect on the business and company should forecast the statistical values and take the appropriate management decision at the appropriate movements. So that when the organization considering the solutions for these issues they should the consider the upcoming situations as well.

4.0 Conclusion

Considering all the factors and features in the given data set of the organization statistically real state business has a good rate of success and the profitability ratio of the organization also got to a good high point. This data set interprets a good variation between the variables and that variation makes the business more interesting and important at the same time. If the company manages to keep them out of the issues then they can make a positive impact on the market.

In future they have to forecast the statistics according to the variable behavior and there are some extraordinary movements In the data set. Organization should consider those data sets and make sure these kinds of data will not make any trouble in the reasonable future.

The next big thing in the forecasting is the factors. There are seasonal factors, circular factors and so on. Therefor organization should consider the behavior of the factors in the future forecasting.