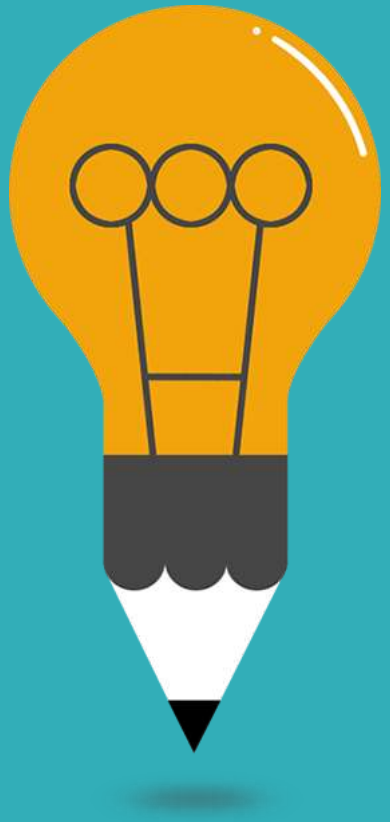
LogoType



NYC Airbnb Host Analysis

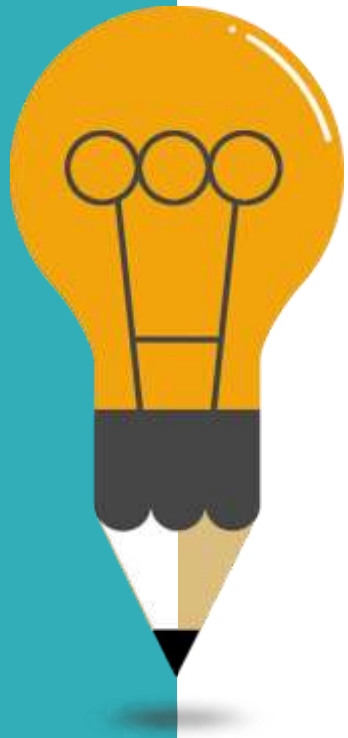
Presented by Siyi Mo

Project Purpose

This project is aimed to provide business insights of Airbnb business status in NYC over the years of 2008 to 2016 by the Airbnb dataset containing 30478 records.

During the process of the project, different visualizations will be built to interpret demographic distributions, in four different districts of NYC which are Manhattan, Brooklyn, Queens, Bronx and Staten Island regarding room type, number of hosts, and average price per night. Meanwhile, growth trend of new hosts over the years, forecast of future district average price and correlation between host price and review rating and booked records will also be studied. All the analyzed results will be included in an interactive dashboard.

Flow Structure



01

Database connection

You can simply impress your audience and add a unique zing.

02

Data analysis

You can simply impress your audience and add a unique zing.

03

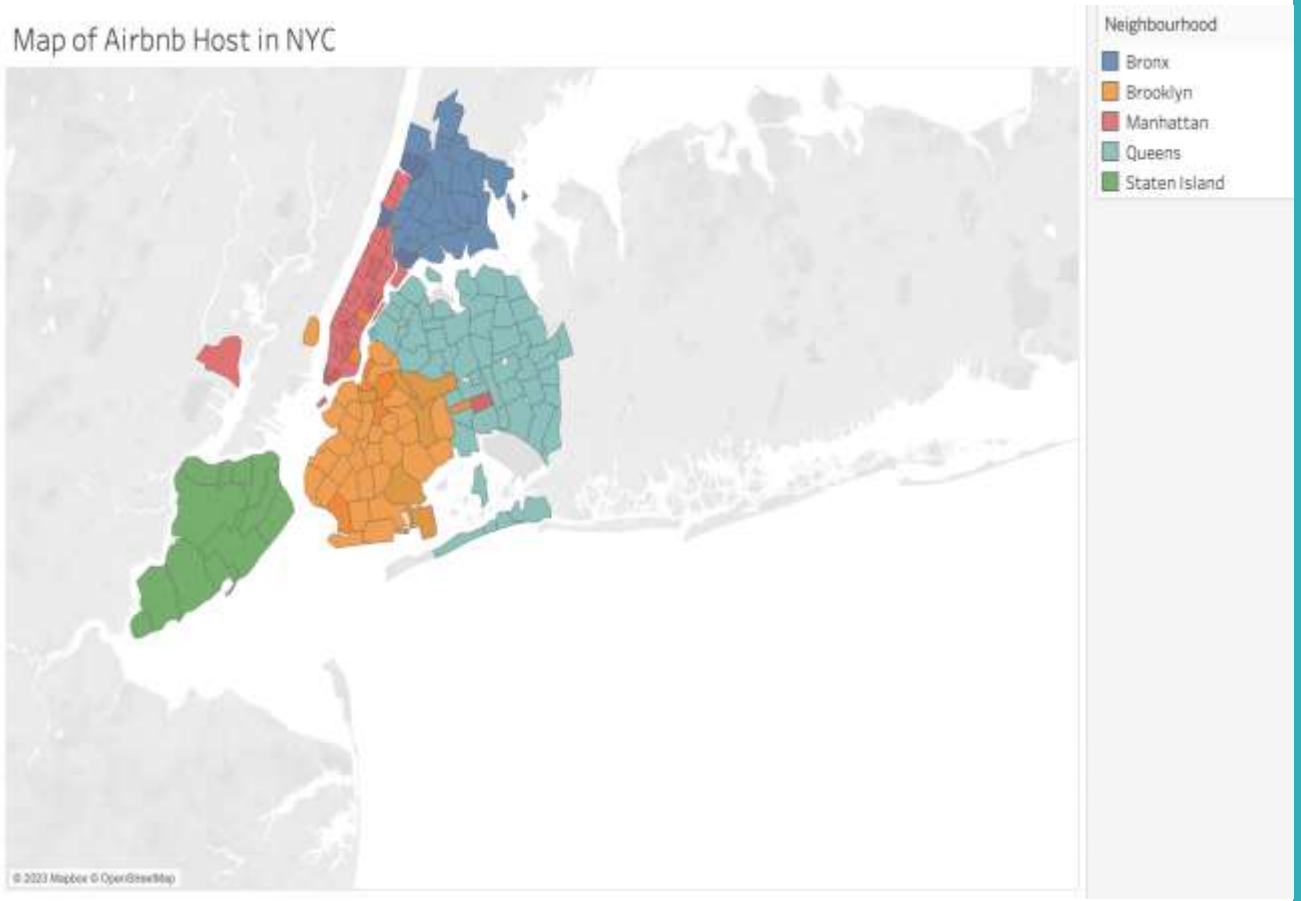
Challenges

You can simply impress your audience and add a unique zing.

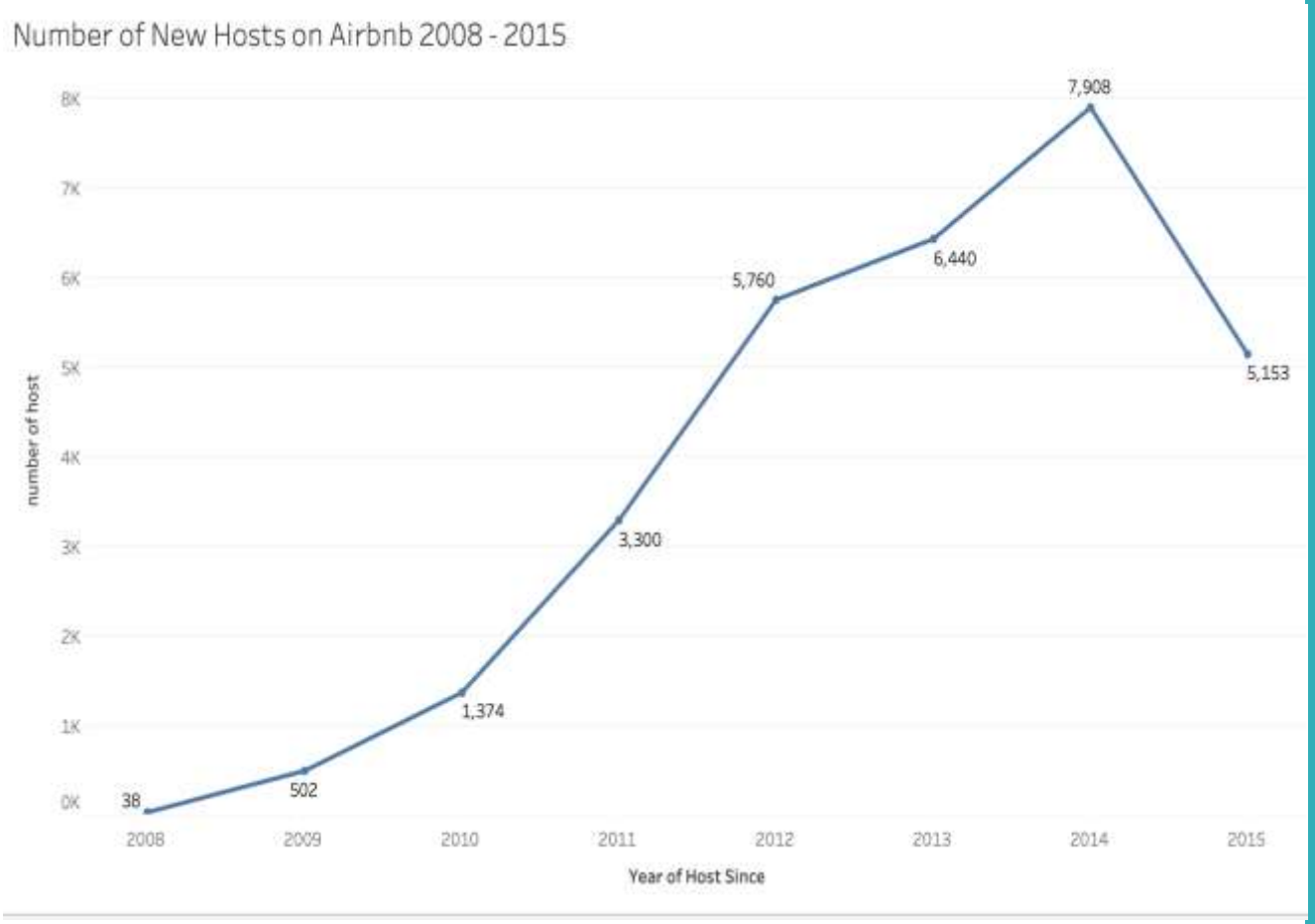
04

Further exploring

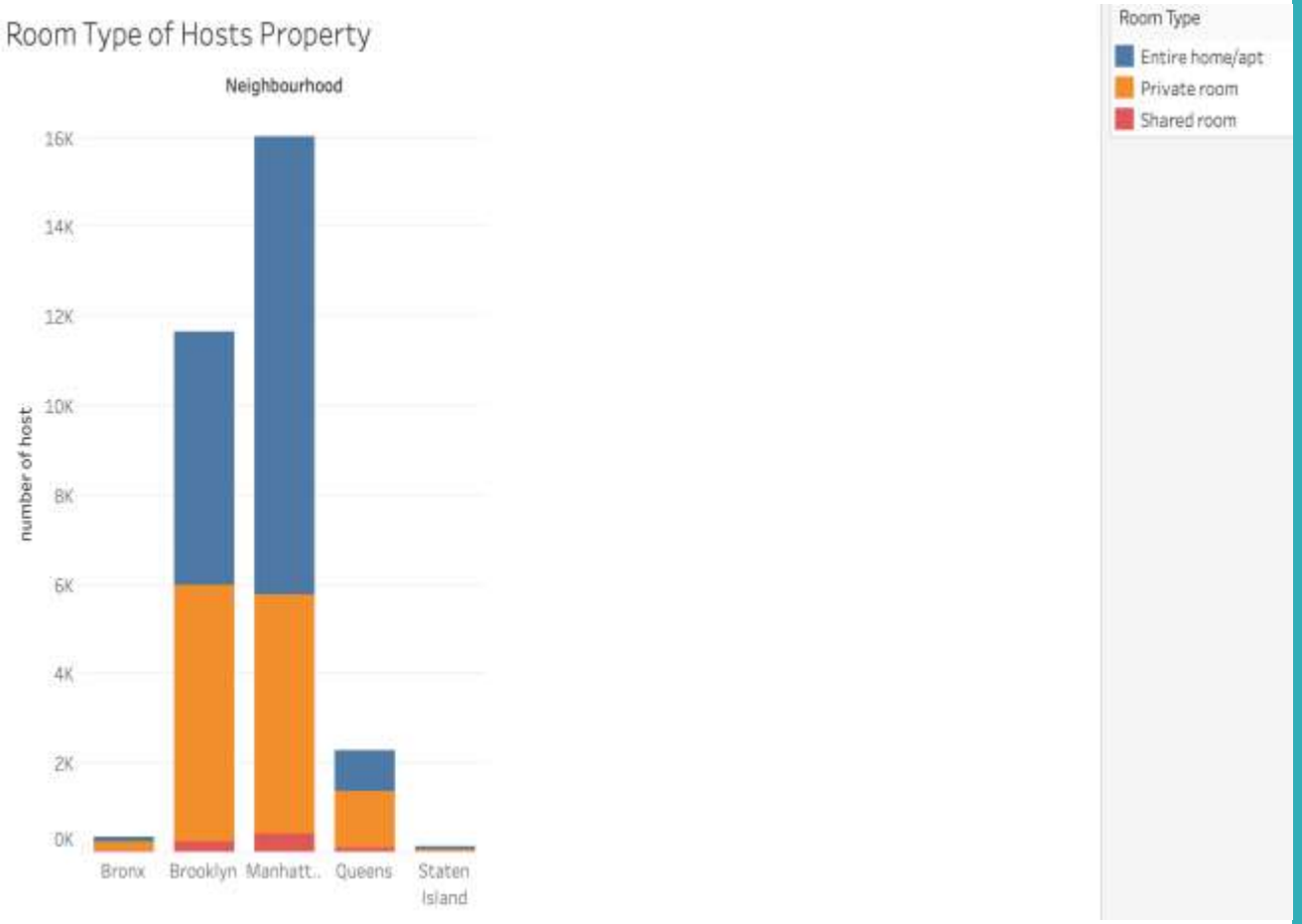
You can simply impress your audience and add a unique zing.



Map of NYC Host

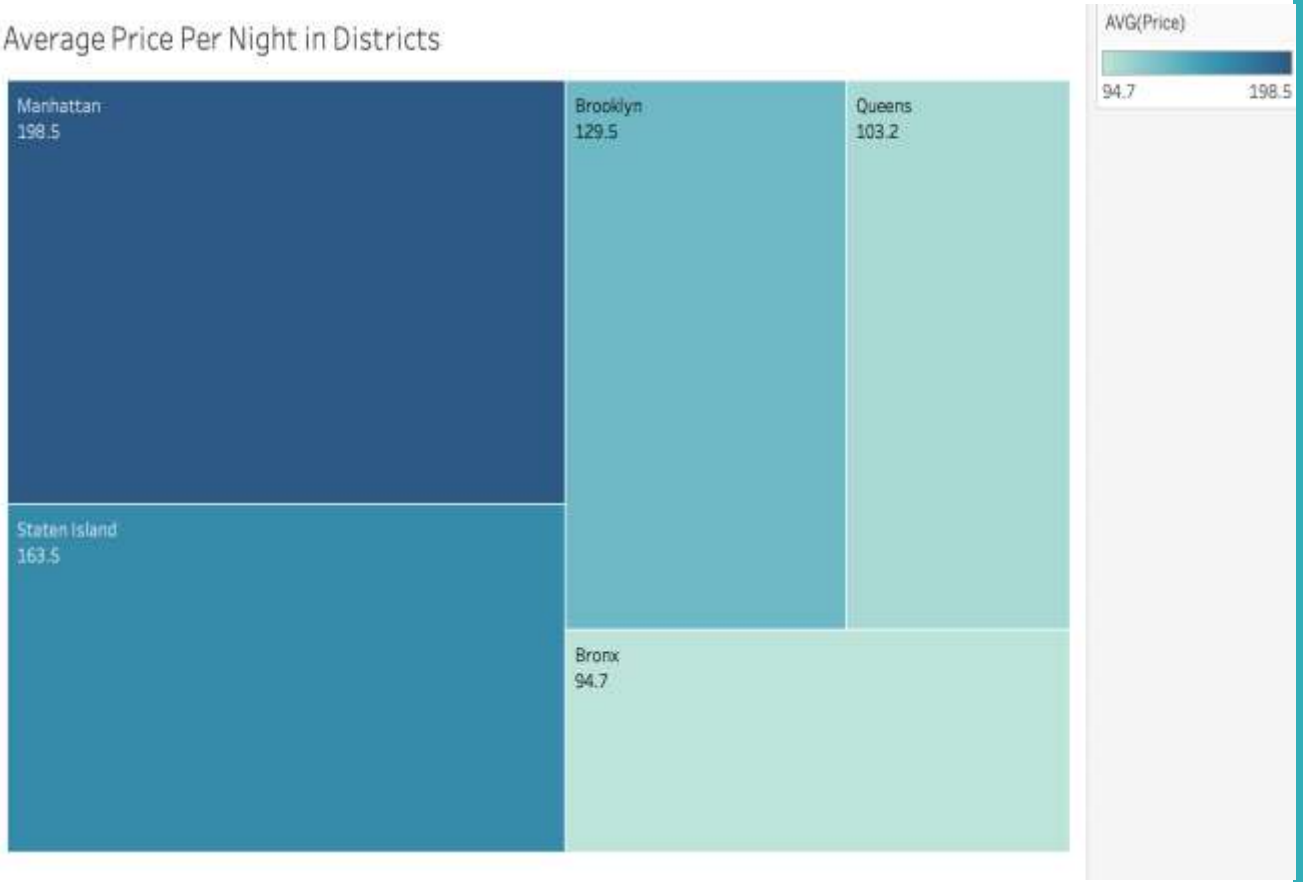


Growth trend of new host per year

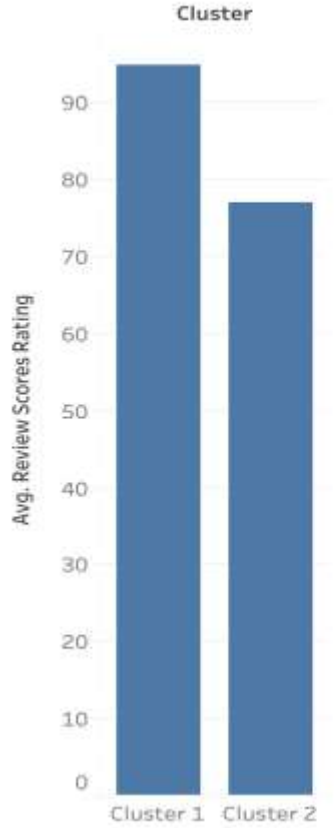

Room Type of Hosts Property
Neighbourhood
number of host
16K
14K
12K
10K
8K
6K
4K
2K
0K
Bronx
Brooklyn
Manhatt..
Queens
Staten Island
Room Type
Entire home/apt
Private room
Shared room

Room type

Average Price



Average review scores in 2 clusters



Centers

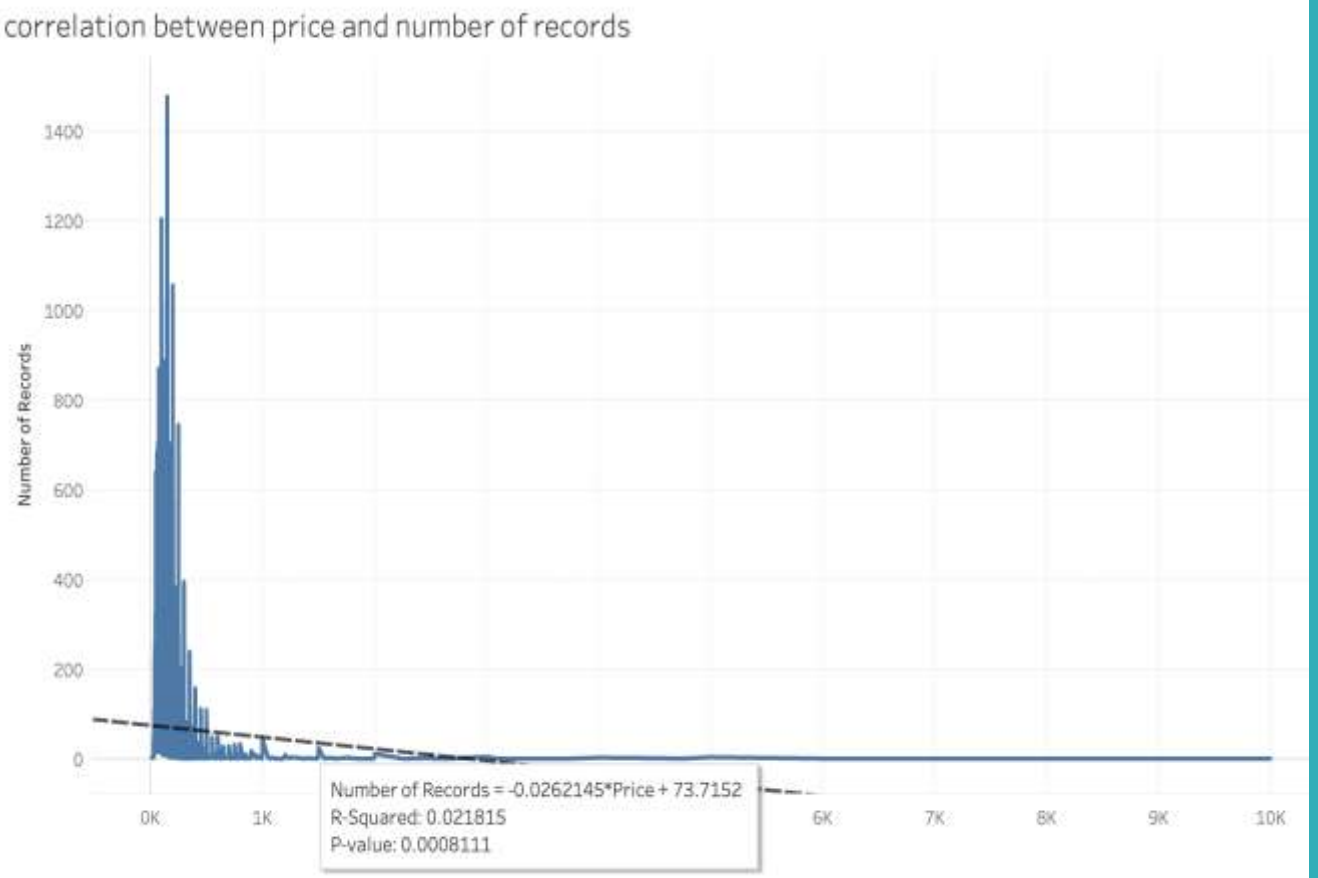
Clusters	Number of Items	Sum of Price	Sum of Review Scores Rating	Sum of Number Of Reviews
Cluster 1	18479	157.99	94.983	17.889
Cluster 2	3676	138.68	76.965	9.5528
Not Clustered	0			

Analysis of Variance:

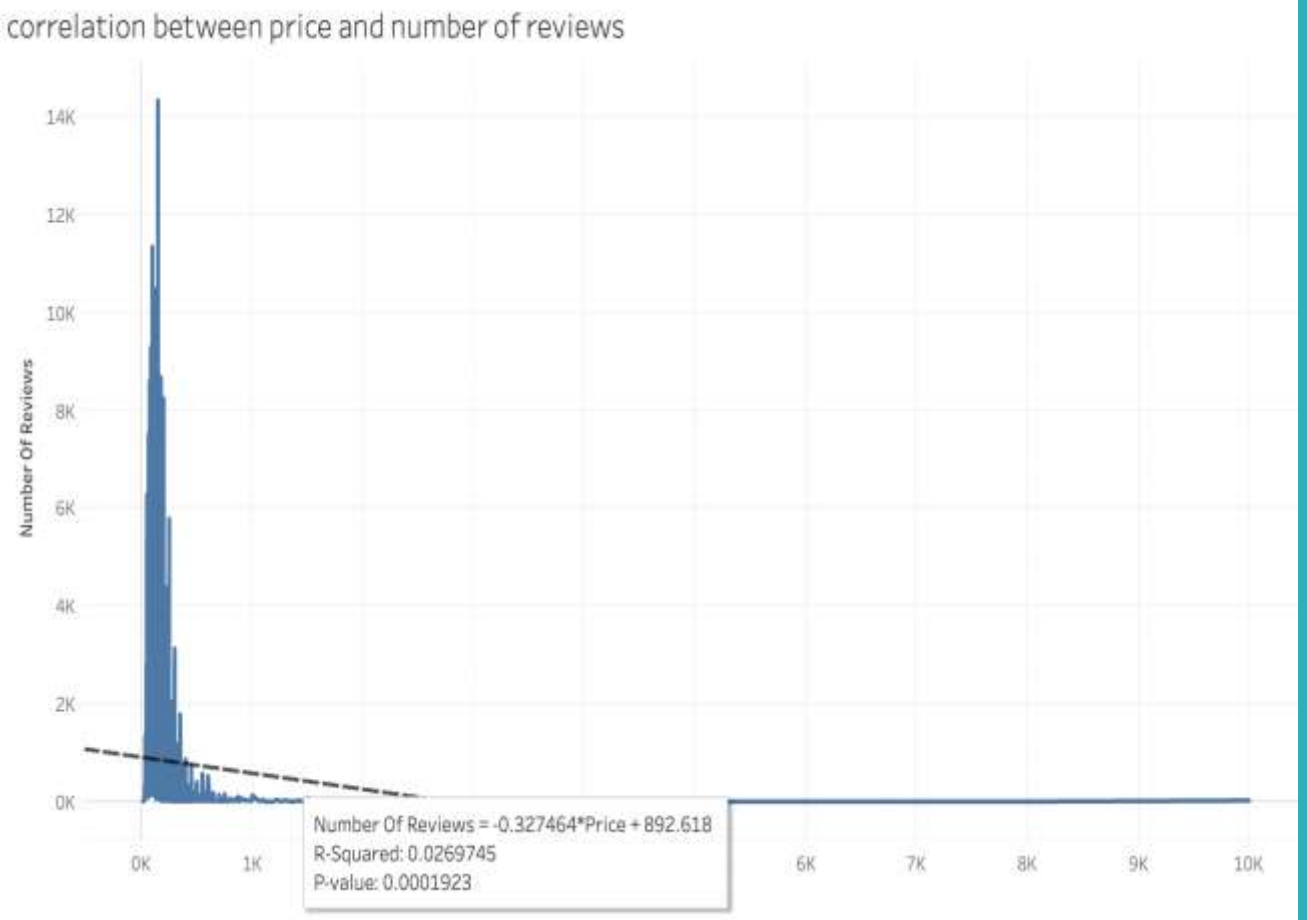
			Model		Error	
Variable	F-statistic	p-value	Sum of Squares	DF	Sum of Squares	DF
Sum of Review Scores Rating	1.271e+04	0.0	155.5	1	271.1	22153
Sum of Number Of Reviews	360.5	0.0	3.251	1	199.7	22153
Sum of Price	51.61	6.998e-13	0.01146	1	4.917	22153

Patterns between price and review scores & numbers

Host Price Forecast



Two Columns Designed



Two Columns Designed

Challenges

1.hard to find variables for better clustering since the n umeric paymasters are not many.

2. R2 of linear regression model is small which means data is not well explained by model.

Future Exploratory

1.To improve the linear regression model, outlier values need to be predicted and missing values need to be identified.
2.Transform the feature data before put into a regression model.

3.multiple linear regression model can be conducted for better fit of the data.