

## *A Sneak Peak into Airbnb Dataset at Seattle*



### *Introduction*

**Seattle** is a seaport city on the West Coast of United States and the most populous city in the state of Washington. It's one of the most famous tourist destinations. Many people come to visit this place and Airbnb is the best option for tourists.

**Airbnb** is an online marketplace for short term and long term stays and experiences. This dataset describes the listing activity of homestays in Seattle, WA.

Everyone wants to have a pleasant and comfortable stay. This analysis tells what is the best time to visit Seattle, which neighbourhoods have the best vibes to stay and how the price variations correlate to various numeric variables from the customers point of view and from the Airbnb point of view they can take a look at how to increase their occupancy rates during off season based on the above view points .

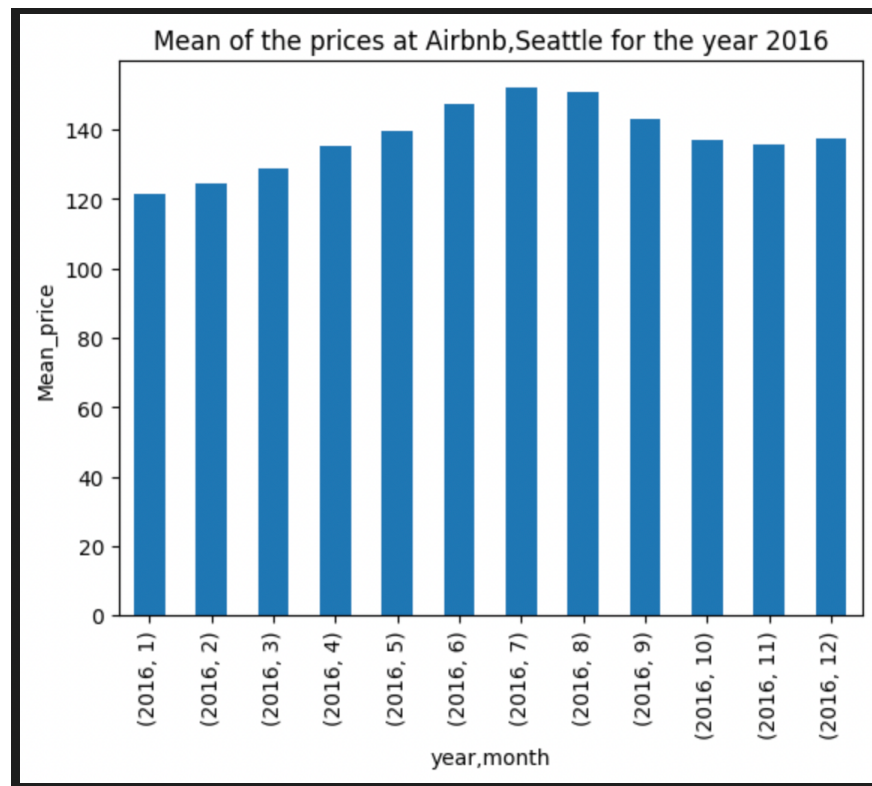
To find out, I have taken the data from **Kaggle** and answered some interesting questions and gave some insights into the business.

1. What are the busiest times of the year to visit Seattle?
2. What are the price variations based on neighborhood ?
3. What is the vibe of each Seattle neighborhood using listing descriptions?
4. What are the numerical variables which have high correlation to price?

. . .

### *1. What are the busiest times of the year to visit Seattle?*

Firstly, I wanted to know what is the highest occupancy time based on mean price of every month in the year 2016.



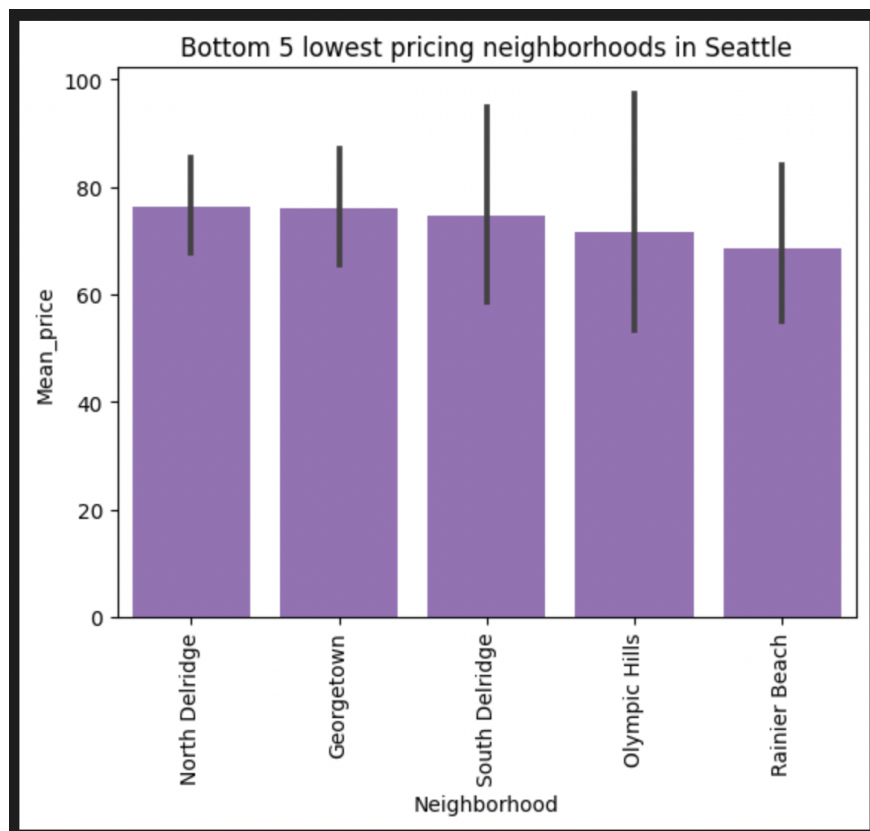
From the above bar plot, we can conclude that **max occupancy** at Airbnb in Seattle is recorded in the months of **June to August** with peak at **July**. The least occupancy rates are in first two months rapidly increasing from April reaching the peak in July and slowly decreasing till October and thereafter maintaining stability towards the end of the year.

. . .

## 2. What are the price variations based on neighborhood ?

Next, we shall take a look at top5 and bottom5 highest and lowest pricing neighbourhoods in Seattle.

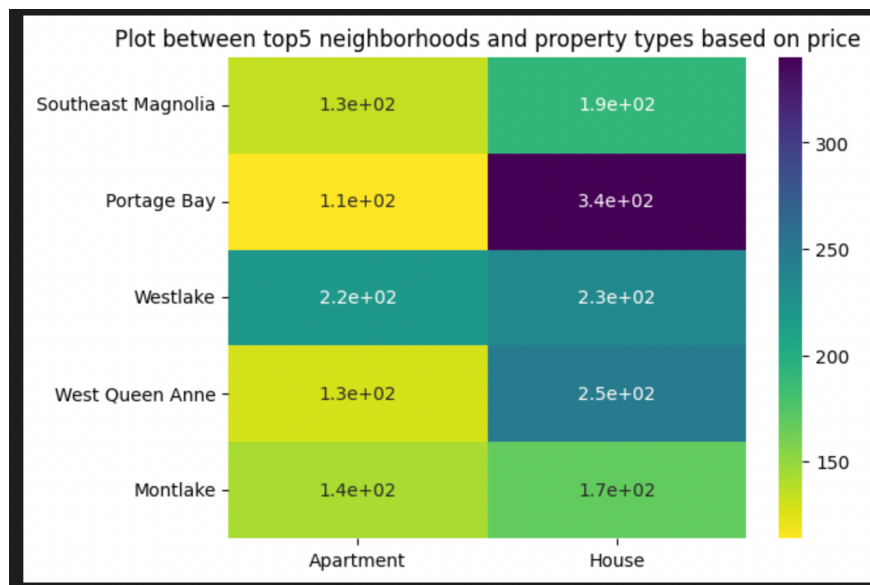




**Southeast Magnolia** is the highest pricing neighborhood with mean price of \$231.7 followed by **Portage Bay** with \$227.85 and the least pricing neighborhood is **Rainier beach** with \$68.5.

I wanted to further extend my analysis to property type in top 5 highest pricing neighbourhoods.

Almost 90% of property types are House and Apartments, so plot a heatmap between these two property types and top5 neighbourhoods .



From the above heatmap, considering the two major property types i.e., House and Apartment, we can conclude that

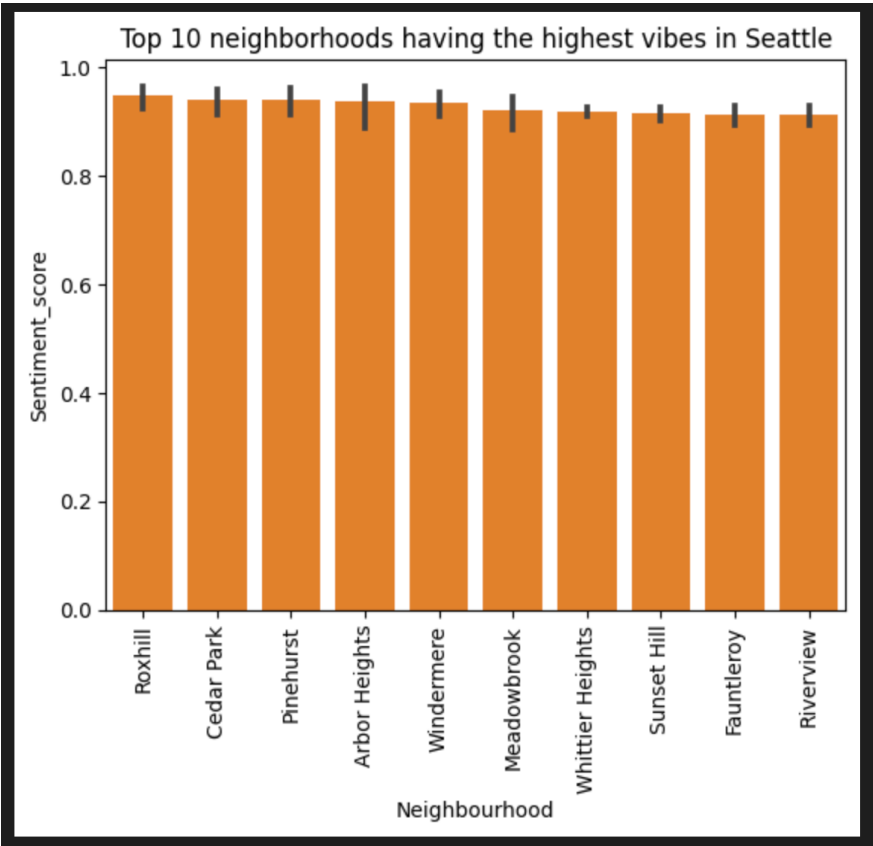
1. **House in Portage Bay** is the highest pricing neighborhood with \$340 .
2. **Apartment in Portage Bay** is the lowest pricing neighborhood with \$113.8 in the top5 neighborhoods.

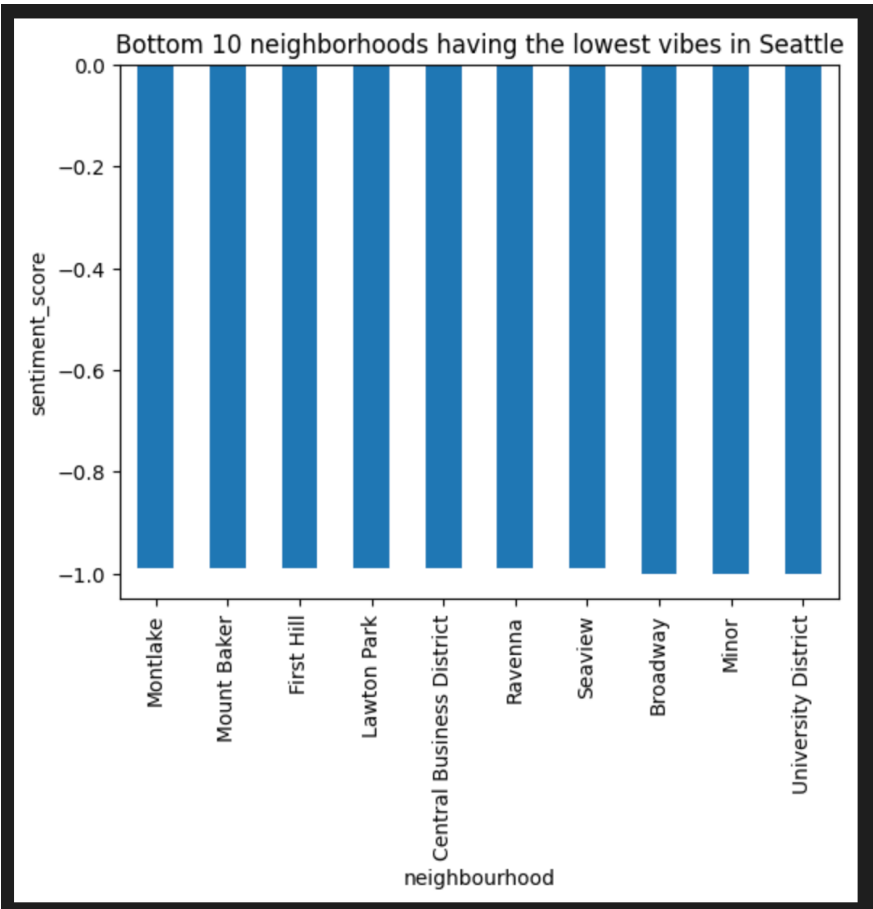
. . .

### 3. *What is the vibe of each Seattle neighborhood using listing descriptions?*

To find the vibe of Seattle neighborhoods i.e., analyse the positive and negative reviews we have to make use of the concept '**NLTK Sentiment Analyzer**'.

Polarity score lies between (-1 to 1) where -1 identifies the most negative words such as '*disgusting*', '*awful*', '*pathetic*', and 1 identifies the most positive words like '*excellent*', '*best*'.





Based on polarity/sentiment scores,the above bar charts are plotted and we can conclude **Roxhill** is the neighbourhood having the highest positive reviews followed by Cedar Park and Pinehurst and **Broadway,Minor and University District** have the most negative reviews.

. . .

4. What are the numerical variables which have high correlation to price?

Finally, I wanted to model the date using Linear Regression and evaluate the coefficients to find which numeric variables have the highest impact on price.

The r2\_score for training data set is 0.52 and for test data set is 0.51 .

	est_int	coefs	abs_coefs
16	bathrooms	26.232495	26.232495
17	bedrooms	24.443734	24.443734
48	Buzzer/Wireless Intercom	17.489005	17.489005
55	Hot Tub	15.838931	15.838931
15	accommodates	15.434493	15.434493
41	Air Conditioning	15.195588	15.195588
56	Pets live on this property	-14.442436	14.442436
47	Free Parking on Premises	-9.988123	9.988123
45	Washer	-9.279528	9.279528
52	Breakfast	-8.976874	8.976874

Based on the above coefficient analysis, the features that have the most impact on price are bathrooms, bedrooms and some of the amenities like Wireless Intercom, Hot Tub, Air Conditioning etc.

To see more about this analysis, see the link to my Github available [here](#).