**Airbnb – Amsterdam**

**About Airbnb**

Airbnb, Inc. is an online platform for arranging or offering lodging, homestays or tourism experiences. The company acts as a broker, receiving commissions from each booking, without owning any of the real estate listings itself.

**About the Data Set**

We’ll be focusing on the listings for Amsterdam over the months, primarity exploring the number of bookings in different areas. The data set provides a description of the host, neighborhood, room types, prices, date of bookings and number of reviews, which can be used to find correlations between different attributes.

Since Amsterdam hosts numerous festivals over the year, the number of listings can be interrelated with the event, aiding us in exploring whether the festivities increase the chances of renting a room for a higher price, lower number of reviews etc.

**Possible Research Questions**

Although, we haven’t decided a primary question for our research, some of the possible ones are listed below:

* Difference in rents in the neighborhoods
* Frequency of bookings per neighborhood
* Frequency of bookings per season in a particular area
* Role of ratings in the bookings
* Effect of profile verification on Airbnb booking and ratings
* Correlation of seasons with the Airbnb size that’s booked
* Effects of festivals on the bookings
* Effect on rent near major holiday seasons
* Correlation of rent with the number of bookings
* Correlation of rent per area with the number of bookings it gets

**Data Cleaning**

* Removing outliers
* Replacing missing values and anomalies

**Exploratory Data Analysis**

* Box plots to identify and remove outliers
* Bar graphs to identify the trends of different variables – grouped by area, number of reviews, months and price.
* Heatmap and scatterplots between different variables to determine the correlation – whether its positive, negative or doesn’t exist.
* Bar and line plots to determine if the attributes are skewed and correlating them with other variables to see if the two are related.
* Line plot for average rent – to determine the relation of rent with the number of bookings, grouped by months, reviews and neighborhood.

**Research Conclusion**

Answering the research question and testing if our initial hypothesis was true.

We haven’t decided a definitive hypothesis yet, but it might pertain to an initial assumption about the relationship between different attributes of the dataset.