**Choosing the Best Neighborhood to Rent an Airbnb in NYC**

1. **Introduction**

   **1.1. Background**

More and more people choose to stay in Airbnbs when they decide to stay in a foreign city, be it for tourism or work. When they do, there are a number of factors to consider when choosing where to stay. Airbnb provides users with maps of the different Airbnbs at their disposal, but especially when entering a new city, there is little a person knows about a given neighborhood. Of course, distance to the city center and price are important, but it is just as important to many people that the type of neighborhood they stay in fits their requirements. That type of knowledge, however, is much harder to retrieve on Airbnb's website. This is where this project enters the stage.

   **1.2 Problem**

With this project, I want to add a layer to the Airbnb search that is typically harder to include: the types of venues and amenities surrounding the neighborhood the Airbnb is located in.

   **1.3 Interest**

The project is interesting due to the ubiquity of Airbnb in travel and the clear benefit it can provide to travelers.

2. **Data acquisition and cleaning**

   **2.1 Data sources**

I use two different data sources in my project. On the one hand, I use the Foursquare.com Places API to identify venues around NYC. This data encompasses the types of venues in the New York City area, the name of the establishment, the location of the venue, and other aspects that we will not use for this particular project.  I also use a public data set provided by Airbnb and posted on Kaggle.com for Airbnb listings in New York City in 2019. This dataset provides the name of the listing, the location, the neighborhood it is located in, the price, number of reviews and other aspects to an Airbnb listing.[1]

---

[1] https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data

**2.2 Data cleaning**

I first create a dataset of venues in New York City based on the neighborhood they are located in. To do this, I parse through each neighborhood and find recommended venues in that neighborhood. The airbnb data does not have the exact same classification of neighborhoods as the neighborhood data I use to identify the longitude and latitude of each neighborhood. I therefore merge the two neighborhood lists to account for this.

I will use this data in the following way:

- Use Foursquare Data to cluster neighborhoods based on venue.
- Use airbnb data in New York City to look at these clusters and compare the airbnbs based on quality and price.

**3. Exploratory Data Analysis**

**4. Modeling**

**5. Results**

**6. Discussion**

**7. Conclusion**