New York Airbnb Data Project

One-sentence summary: This site allows travellers, Airbnb host and people who work in the tourism industry to choose a satisfying housing and better understand the Airbnb housing market in New York.

Summary of the dataset: We use the dataset “New York City Airbnb Open Data: Airbnb Listings and Metrics in NYC, NY, USA (2019)” from Kaggle.com. Guests and hosts use Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, NY for 2019. This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions. More specifically, the dataset contains `host names`, `neighbourhood location`, `room location`, `room type`, `price`, `minimum number of nights`, `number of reviews`, `last review`, `reviews per month` and etc. The data is uploaded by Dgomonov, who acknowledged Airbnb’s public release of the data. We also didn’t find any restrictions on the terms of use for this dataset. We downloaded version 3 of the dataset in csv format from Kaggle.com on October 4th, 2019.

Column name	Column Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	price in dollars
minimum_nights	amount of nights minimum

number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Audience and goals:

- Incoming landlords and current hosts
    - Figuring out the popular locations by neighborhoods
    - Getting knowledge of reasonable prices within current neighborhoods
    - Popular styles, e.g. number of rooms and requirements for renters
- Tourists interested in big data of airbnb in New York
    - Reviewing the popular housing and availability to figure out where to stay
    - Comparing and Contrasting among various neighborhoods
    - Knowing reasonable price in certain neighborhood
- Investigators and researchers interested in the entire industry and economic situation in New York City
    - Knowing trends of pricing according to neighborhood variation
    - Research in what price of housing relates to
- Tourism business owner, e.g restaurant owner
    - Knowing average price of neighborhoods to determine the location of their target customers

List of requirements:

- Functional:
    - Provide range of price of apartments and housing in given area
    - Provide availability of all housing in given location
    - Provide rank of popularity of given area housing
        - Use number of reviews as a proxy for popularity
    - Provide housing capacity of given area
    - Show the host a word cloud of the reviews
    - Recommend what price to set/adjust to
        - Compute the average airbnb price in the host's region

        - Show whether the host is more expensive or less expensive than the average
    - Use maps to display data related to location
- Non-Functional:
    - Present data to be understandable and straightforward
    - Assurance of the accuracy and consistency of the data
    - Remain functional without excessive maintenance or repair
    - Designed in a way that is usable for people with disabilities
        - Pass WAVE test (especially for the maps)
        - Able to navigate the website using Tab

Team collaboration: Since the project is mainly separated into three parts: frontend, mediator, backend. We roughly distributed them to members according to their strength in different languages:

- Rosa: mediator. Since the backend deals with the database and the frontend both displays and gets user inputs, Rosa will be translating the data between them. Rosa has experience with Javascript and jQuery. She will also learn more specific techniques needed for this project.
- Emma: backend. Backend deals with the dataset in .csv format. Ideally, we will use SQL to filter data and crop data according to queries given by the frontend. Emma is familiar with SQL which will be helpful if she takes charge of backend.
- Shiyue: frontend. Since we will be working on a lot of data visualization which is a domain that we are unfamiliar with, Shiyue will be in charge of searching related libraries that we can use. Shiyue is also familiar with Flask, so she will take a lead when we use this library.

We will remain in constant email contact over the course of this research and respond with 3 hours or remain in contact through social media. We will meet at least 3 times each week where all team members will be present. We will make all important decisions together such as the design of the framework and the main content. We will make sure that everyone agrees are what we are working on. Emma will be in charge of shared documents and files. Rosa will be responsible for the overall project structure, such as functions that we have implemented, methods that need improvement and what each sprint would consist of. Shiyue will be responsible for checking the deadlines and making sure that everything is on track.