Airbnb Project Report

Computational Thinking Sp23

Ben Onderick and Evan Kale

**Problem:**

The problem that we chose to tackle deals with Airbnb data. When you use Airbnb, you can’t view average prices specific to certain neighborhoods, or compare them easily side by side. This can be a problem because you could be overpaying for something that isn’t worth it. You could get a better place in a better neighborhood for a cheaper price. You also can not see the number of average reviews that are left on a place. We chose to create a program that could do this.

**Data:**

The dataset that we chose was an Airbnb dataset that we found on Kaggle. The data has over 40,000 different entries of Airbnb locations and listings that have been on the site in New York City. The different parameters of each location include various pieces of location information, information about reviews, and pricing information for each location.

<https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data>

**Function Implementation:**

We have created 6 different functions to aid us in creating this program.

load\_data – This function is to just simply load the data into python so we can analyze it.

get\_top\_neighborhoods – this function uses a groupby and sort command to sort all the neighborhoods in order of the most expensive neighborhoods by nightly price.

get\_low\_neighborhoods - This function also uses a grouby and sort command, only it sorts the lowest neighborhoods by nightly price.

plot\_price\_histogram – This function creates a histogram that shows the price on the x-axis and the frequency of stay on the y-axis.

calculate\_review\_stats – This function shows summary statistics, in particular it shows the number of review, the mean and median number of reviews and the minimum and maximum of reviews made.

display\_avg\_price\_per\_room – This function shows the average price per room by room type. The room types include a whole house or apartment, private room, or shared room.