Capstone Project Submission

Airbnb Booking Analysis Summary

GitHub Link: - https://github.com/ajinkyamorade/Airbnb-Bookings-Analysis

Airbnb was conceived in 2008 by two roommates who rented out an air mattress in their living room. This San-Francisco-based start-up offers you someone's home as a place to stay instead of a hotel. Airbnb does not own properties. It acts as an intermediary between those who want to rent out space and those who are looking for space to rent.

We have provided with host data in NYC with about 49000 rows and 16 columns. Starting with raw data, we performed data wrangling over it. In addition, we divided this project into different sections, such as Data Exploration and Variable Identification, Understanding Data, Handling NaN Values, Exploring and Visualizing Data, Single Variable Analysis, Bi-Variable Analysis. For managing NaN values, we mainly focused on important elements like neighborhoods, neighborhood groups, host ids, counting, etc. Several features were dropped from the analysis because they were of no value to us. Depending on the feature, we replace all null values with '0', 'unknown', etc.

Then We found out correlation between various features using heat map. During the exploration and visualization of data, we performed single and bi-variable analysis according to the requirements. Through exploration we get to that Manhattan has the highest number of listed properties and also all top 5 costliest properties in entire NYC. We also found out that guest's preferred living in an entire home/apt followed. We inferred from analysis that there are very less shared rooms throughout NYC as compared to private and Entire home/apt.

We also analyzed that Private room has received the most no of reviews/month where Manhattan has received highest number of reviews for private room type. Manhattan & Queens got the most no of reviews for Entire home/apt room type. There were less reviews received from shared rooms as compared to other room types and it was from Staten Island followed by Bronx.

Almost 95% of the listings on Airbnb are either Private room or Entire home/ apt. We also found out that Bronx and Staten Island has listings which are mostly available throughout the year (which may be the case as they are not much costlier as compared to other neighbourhood groups.

At the end we examined the Price feature where we detected and removed outliers from it by using the quantile method. Also, we found out cheapest and costliest listings in entire NYC, average prices of rooms in each neighbourhood group.

Team Member's Name, Email and Contribution:

Nitesh Gajakosh |niteshgajakosh1998@gmail.com|

Contribution Role:

- Data Wrangling
    1) Acquiring Data – Airbnb NYC 2019
    2) Variable Identification
- Data Analysis
    1) Co-relation matrix
    2) Single-variable Analysis
        a. Neighbourhood group Vs Number of listings
        b. Top 10 reviewed hosts on basis of review per month
    3) Bi-Variable Analysis
        a. Count of each room type in neighbourhood group in entire NYC
        b. Monthly review variations with room types in each neighbourhood groups
    4) Price Feature
        a. Room types Vs Price in different neighbourhood groups
        b. The costliest and cheapest listing & their respective hosts in entire NYC

Ajinkya Morade| ajinkya.morade1998@gmail.com|

Contribution Role:

- Data Wrangling
    1) Acquiring Data – Airbnb NYC 2019
    2) Variable Identification
- Data Analysis
    1) Handling Null Values
    2) Single-variable Analysis
        a. Top 10 neighbourhood entire NYC on the basic of count of listings
        b. Minimum stays in different room types
        c. Top 10 host (id) with the greatest number of listings
    3) Bi-Variable Analysis
        a. room type and their relationship with availability and with different neighbourhood groups
    4) Price Feature
        a. Detecting outliers and removing them
        b. Top neighbourhood groups in NYC with respect to average price/day of Airbnb listings