Capstone Project Airbnb Nyc 2019 EDA

Instructions:

i) Before running the notebook please Install the below library

- plotnine
- dataprep
- plotly
- matplotlib
- worldclod
- pli

ii) Please try to run on Jupyter Notebook. On google collab dataprep may not work.

iii) Plotly graph may not show on Github please run the code to take effects.

ABOUT THE DATASET

Airbnb has revolutionized the hospitality industry. Prior to 2008, travelers would have likely booked a hotel or hostel for their trip to another town. Nowadays, many of these same people are opting for Airbnb.

The idea behind Airbnb is simple: Find a way for local people to make some extra money renting out their spare home or room to people visiting the area. Hosts using this platform get to advertise their rentals to millions of people worldwide, with the reassurance that a big company will handle payments and offer support when needed. And for guests, Airbnb can offer a homey place to stay that has more character, perhaps even with a kitchen to avoid dining out, often at a lower price than what hotels charge.

This dataset contains 16 features/variables about Airbnb listings within New York City. Below are the features with their descriptions:

Id , id, name , host_id , host_name , neighbourhood_group , neighbourhood , latitude , longitude , room_type , price,minimum_nights , number_of_reviews , last_review , reviews_per_month , calculated_host_listings_count , availability_365

PROBLEM:

What is price according to the neighborhood group and room types?
Where do customers spend most and least numbers of night?
What type of room is preferable by the most of the customers?
Which region is cheaper and costly?
Which region got most and least listings?

SOLUTION:

- In first step I perform data wrangling on the given data set. And clean all the uncomplete features and null values.
- In price analysis I focused over the neighbourhood_group and room_types to know the which is cheaper and costly

- In Region wise room analysis I want to know that the counting of room types in each region.

- Room wise review analysis I focused on which room get maximum average review.
- I calculate the outlier price and filter the nonoutlier prices.
- I establish a relation between review and price to know the choice of price level by the customers.
- I categorized the prices according to the 20% , 40% , 60% , 80% and 100% quantiles. To know that which region's , which types of rooms have highest and lowest price group.
- Through the analysis of name column which consist of description of rooms I draw a word cloud and I got some most frequent words. That can help in searching of rooms.

GRAPHS:

- **SCATTER PLOT:** Relation between latitude and longitude , Relation between price and reviews.
- **HEATMAP**
- **JITTER PLOT USING PLOTNINE:** Relation between room type and availability
- **BAR PLOT:** Relation between Neighborhood group and neighborhood count. Relation between numerical and categorical variable
- **VIOLIN PLOT:** Price Analysis
- **PIE CHART:** Analysis of room types and their percentages.
- **WORD CLOUD:** To check most frequent words

Vinayak Abhinav **vinayakstar.pusa@gmail.com**
All the part of this eda is done by me with the help of AlmaBetter, Kaggle and Youtube.
Github Link: https://github.com/space-star/AlmaBetterAirbnb_CAPSTONE_PROJECT_AIRBNB_DATA_ANALYSIS_EDA

SPECIAL THANKS TO:

- https://www.almabetter.com/
- https://kaggle.com/
- https://www.w3schools.com/