PROJECT REPORT – TEAM 39

PROBLEM STATEMENT

In the ever-changing landscape of the New York City Airbnb market, our project aims to analyze data and predict prices, offering valuable insights for potential investors and discerning customers. Our main goal is to discover patterns that reveal areas with the highest number of listings, understand the factors influencing different costs, and grasp the preferences of both hosts and guests. By exploring the complex interactions between neighborhood characteristics, seasonal demand, and pricing dynamics, our research aims to equip new investors with decision-making tools and provide customers with a strategic advantage in selecting listings based on their preferences and budget constraints. This project provides a comprehensive understanding for hosts and guests, offering a valuable resource for strategic decision-making in the dynamic and popular Airbnb market.

Data Source: The dataset was obtained from Kaggle. It comprises a dataset of size [48896*16] with columns including listing names, host details, location coordinates, pricing, and availability metrics.

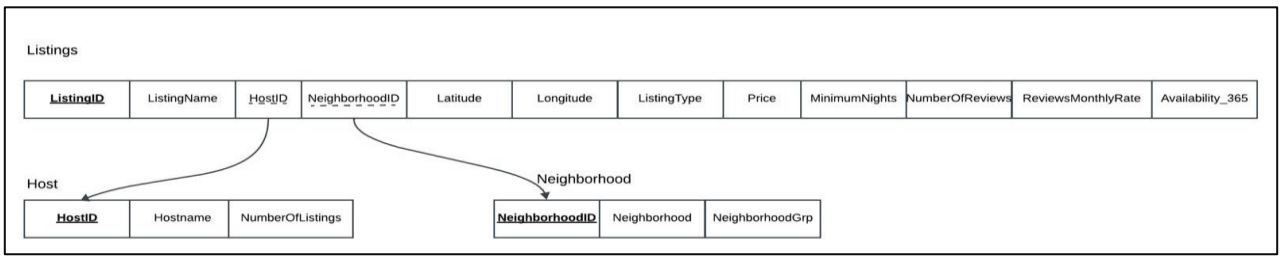
APPROACH & METHODOLOGY

Data Collection and Cleaning:

- Parsed the TSV file to extract the header and listing data.
- Cleaned the data by renaming columns, removing unwanted columns, and handling missing values.

Database Setup and Normalization:

- Created an SQLite database and set up tables for Host, Neighborhood, and Listings.
- Normalized the data to eliminate transitive dependencies, improving data integrity.

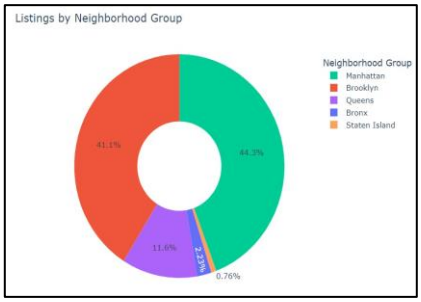


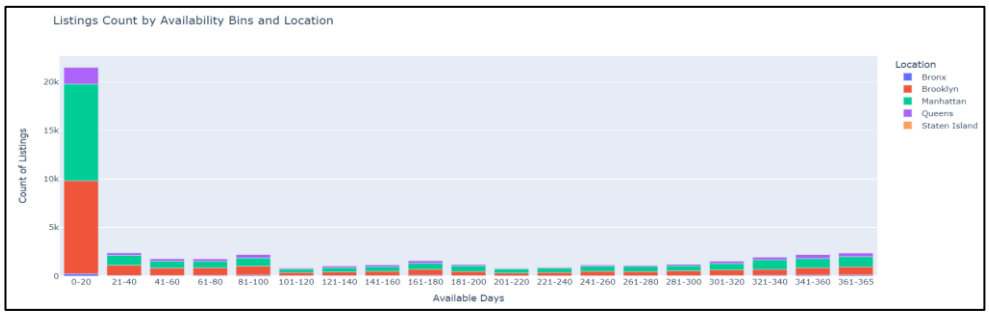
Exploratory Data Analysis (EDA):

- Visualized the distribution of listings by neighborhood group, analyzed the count of listings based on availability and location & explored the distribution of listing types in the dataset.
- Created visualizations for room type vs. price, geographical distribution, average listing prices by neighborhood group, and more during Exploratory Data Analysis (EDA).

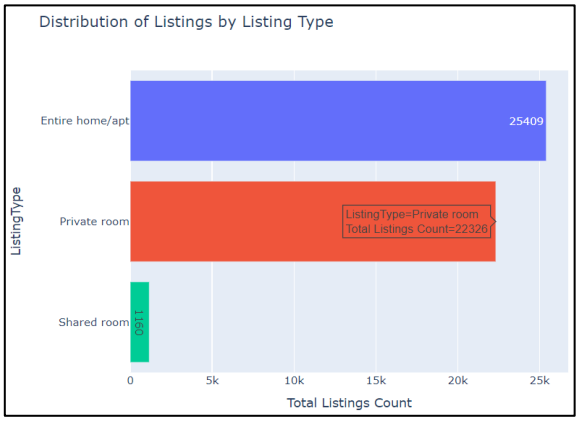
The following graphs portray diverse aspects of the Airbnb Data, facilitating a comprehensive analysis:

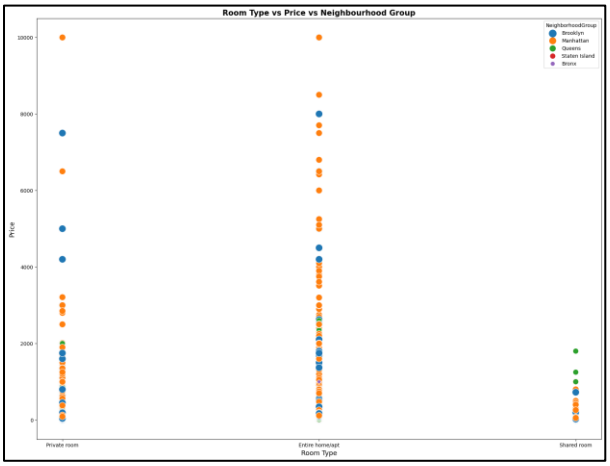
1. **Donut Chart:** Percentage distribution of listings across neighborhood groups. Manhattan & Brooklyn collectively dominate the market, with Manhattan (44.3%), Brooklyn (41.1%) resulting in a combined share of over 80%.
2. **Stacked Column Chart:** Distribution of listing counts based on availability bins and location. Approximately 40% of listings fall within the 0-20days availability bin and Most properties have limited availability, indicating a strong presence of short-term rental options.



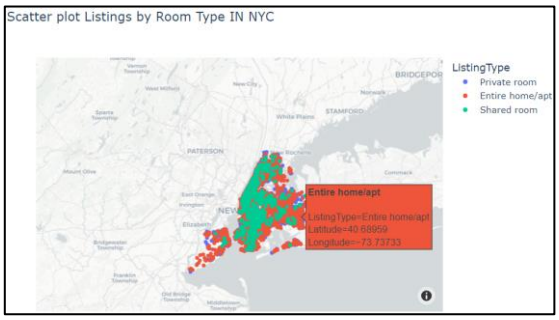


3. **Horizontal Bar Chart:** Illustrates the distribution of listings by listing type, with the majority being of the "Entire home/apt" type, comprising approximately 52% of the total listings in the dataset.
4. **Scatter Plot:** Depicts the relationship between room type, price, and neighborhood group in the dataset, highlighting the distribution of prices across different room types and neighborhood groups.





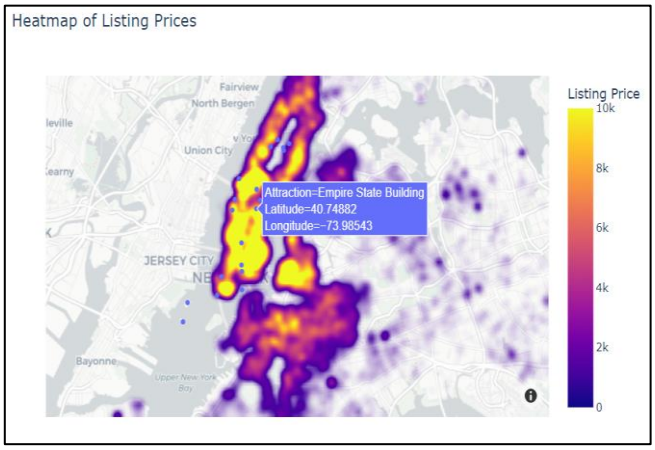
5. **Scatter Map Plot:** Illustrates the geographical distribution of listings in NYC based on latitude and longitude, with color-coded markers representing different room types.
6. **Bar Chart:** Depicts the average listing prices in different neighborhood groups in NYC, highlighting variations in prices across neighborhoods. Manhattan has the highest average listing price of around 197$.

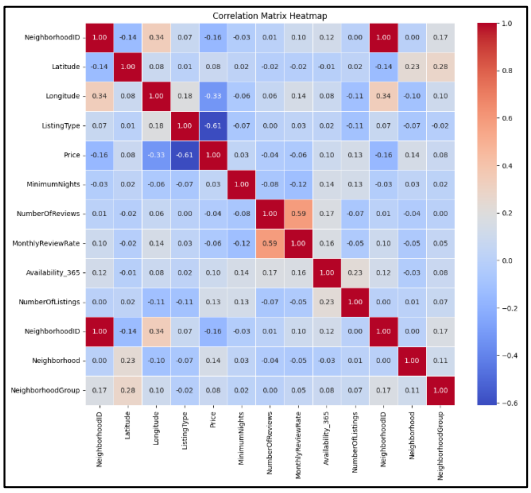


7. **Scatter Plot / Density Heat Map:** It illustrates the concentration and pricing variations of listings across different neighborhoods in New York City. The blue dots highlight locations of key attractions, offering a spatial overview within the diverse NYC landscape.

Combining Data and Correlation Analysis:

- Combined data from normalized database tables.
- Analyzed the correlation matrix to identify relationships between variables.
- Strong Negative Co-relation of Listing Type and Longitude Predictors with Price has been observed.

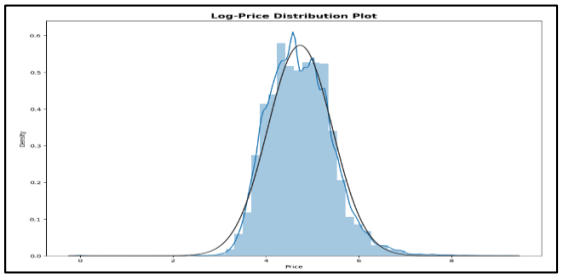




Correcting Skewness in Data:

- In the analysis, we applied a logarithmic transformation to the **price** variable. This was done to address the right-skewness observed in the original distribution, where a few extremely high prices were impacting the overall shape.

The logarithmic transformation of the **price** helps normalize the distribution, making the data more symmetrical and enhancing the reliability of subsequent analyses and model-building **Predictive**

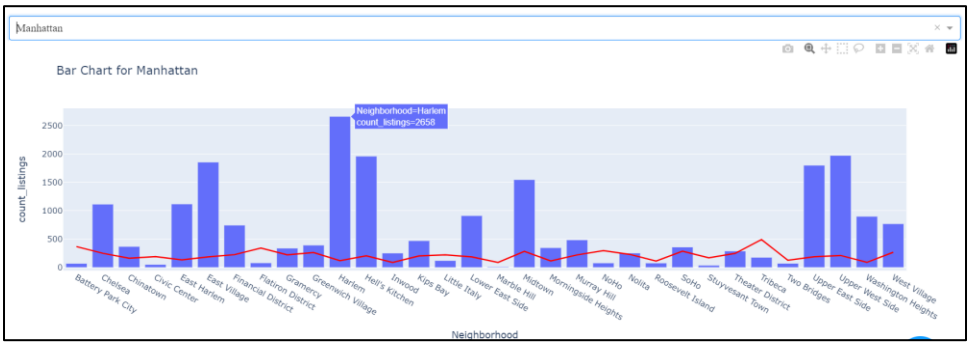


Modeling:

- Predicted listing prices using Linear Regression and Ridge Regression models.
- Evaluated model performance using metrics such as Mean Absolute Error, Mean Squared Error, and R-squared.
- Visualized actual vs. predicted prices to assess model accuracy.

Interactive Dashboard:

- Utilized Dash to create a web-based dashboard with a dropdown menu enabling users to explore and compare the count of listings and average prices across different neighborhoods within a selected neighborhood group in NYC. The dashboard integrates both bar and line charts for a comprehensive analysis.



CONCLUSION

In conclusion, our project conducted an in-depth analysis of the dynamic New York City Airbnb market, revealing significant insights. We meticulously explored room availability, categorized neighborhoods, and pinpointed locations near key tourist attractions and below are our key findings and insights.

KEY FINDINGS:

Geographical Trends:

- Manhattan and Brooklyn are the most dominant areas in terms of both listing count and prices.
- Airbnb listings are densely concentrated around popular tourist spots, including Times Square, the Empire State Building, and Central Park.

Listing Type:

- Entire apartments are the most popular listing type in New York City.

Availability Analysis:

- Most Airbnb listings exhibit limited availability throughout the year.
- The scarcity of short-term rentals in NYC contributes to elevated prices.

Correlation Analysis:

- The correlation matrix suggests that the prices of Airbnb listings in NYC are majorly influenced by the listing type and longitude, suggesting there is a notable relationship between the type of listing (such as entire home/apartment or private room) and the geographical location (longitude) with the pricing of the listings.

TEAM MEMBERS (TEAM 39):

Nihith Nath Kandikattu

Keerthana Allam

Hithesh Kumar Duttuluri

Charan sai pandaraboyina