

Exploring the Relationship between Airbnb Listings and Crime in New York

Akash Venugopal

Ruthu Gandal Shankare Gowda

Ruthwik Reddy Bommana

Problem Statement

- Airbnb has rapidly grown to become a popular alternative accommodation option for tourists visiting New York. While this has brought many benefits to the city's tourism industry, concerns have been raised about the potential negative impacts of Airbnb on local communities. One such concern is that the increase in Airbnb listings may lead to an increase in crime in the neighborhoods where these listings are located. The purpose of this project is to explore the relationship between Airbnb listings and crime in New York City, and to provide insights that can help stakeholders better understand the potential impacts of Airbnb on public safety and the overall well-being of local communities.

Datasets

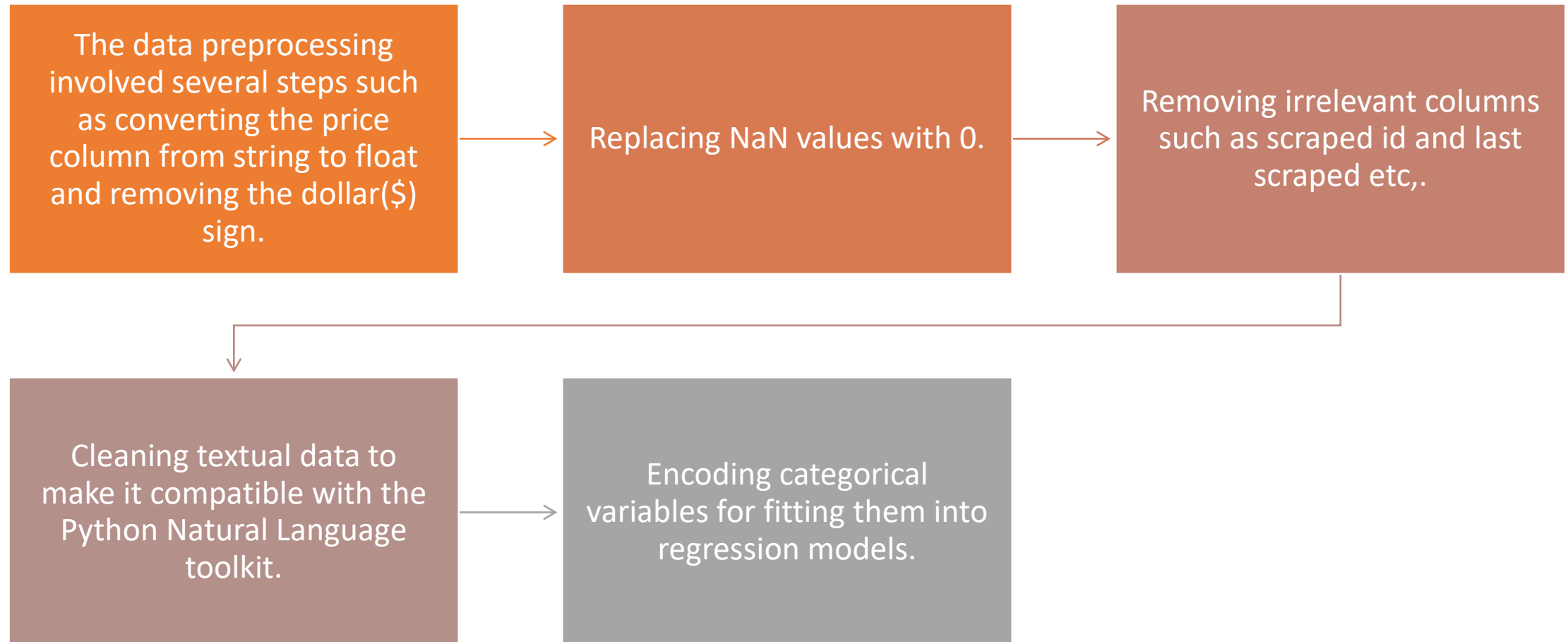
- NYC city dataset - Inside Airbnb link

<http://insideairbnb.com/get-the-data/>

- Complementary dataset - NYPD shooting incident data - NYC Open Data

<https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Year-To-Date-/5ucz-vwe8>

Data Preparation



Data Visualization and Business Intelligence Tool



Tableau is a data visualization tool that allows users to create interactive and dynamic dashboards, reports, and charts from various data sources.



It provides a user-friendly interface and drag-and-drop functionality that allows users to quickly and easily create visualizations without the need for advanced programming or coding skills.



It is also a powerful tool that helps in data cleaning enables users to analyze and visualize large amounts of data easily.

Data Visualization and Business Intelligence Tool

Primary Components of Tableau are

Data Sources → Tableau connects to various data sources such as Excel, CSV, SQL Server, Google Analytics.

Worksheets → are where the visualizations are created.

Dashboards → allows users to combine multiple worksheets and visualizations onto a single page, providing a high-level overview of data insights.

Filters → allows users to dynamically change the data being displayed in the visualization, filter out unwanted data.

Formatting Options → Tableau provides various formatting options to customize the visual appearance of the charts and graphs.

Data Implementation

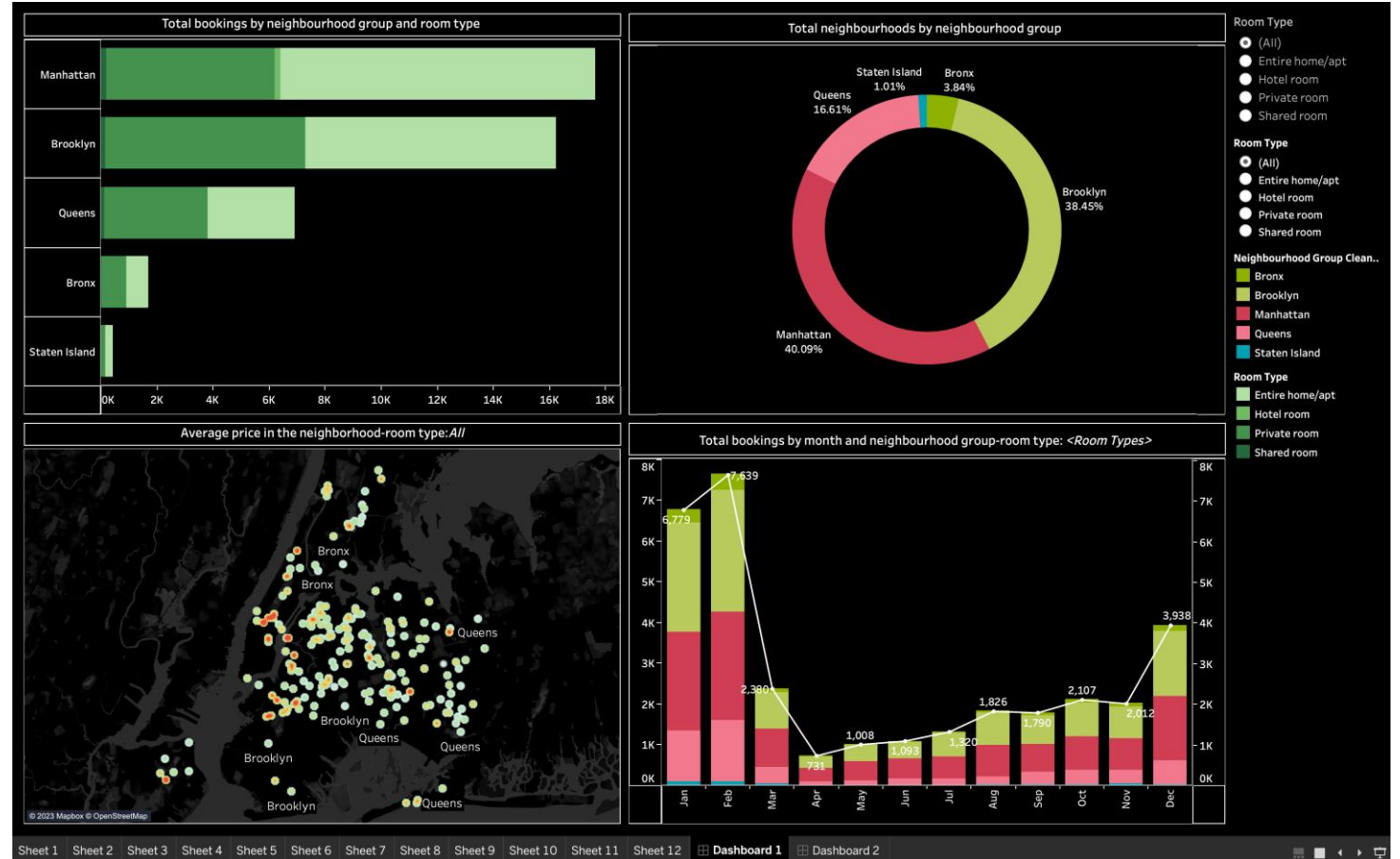
Two primary tools were used for implementing data analysis and visualization - Tableau and Jupyter Notebook with Python. Several libraries such as Matplotlib, pandas, scikit-learn and tree interpreter were also utilized.

Utilized Tableau for data visualization and sharing insights with others
Developed interactive dashboards for users to explore and analyze the data in detail

Python was utilized for data analysis and visualization, with the assistance of libraries such as Matplotlib and Tree Interpreter, for developing regression models.

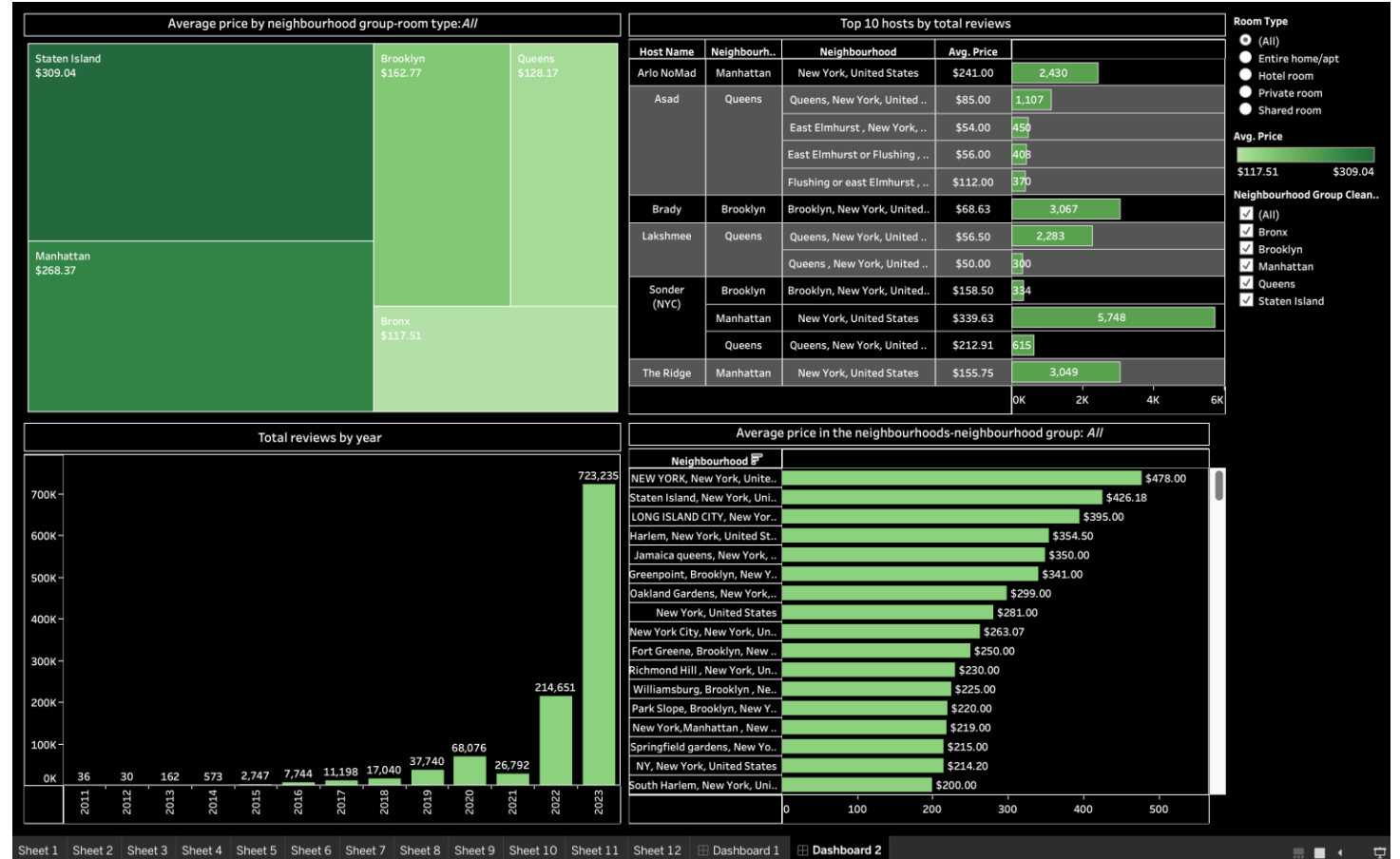
Exploring Data through Analysis and Visualization Technique

The created Tableau sheets show important insights from the data analysis, such as the total bookings and neighborhoods by neighborhood group and room type, average prices, and bookings by month and neighborhood group-room type. These visualizations provide a comprehensive understanding of the data and can help in making informed decisions regarding the Airbnb listings in New York.



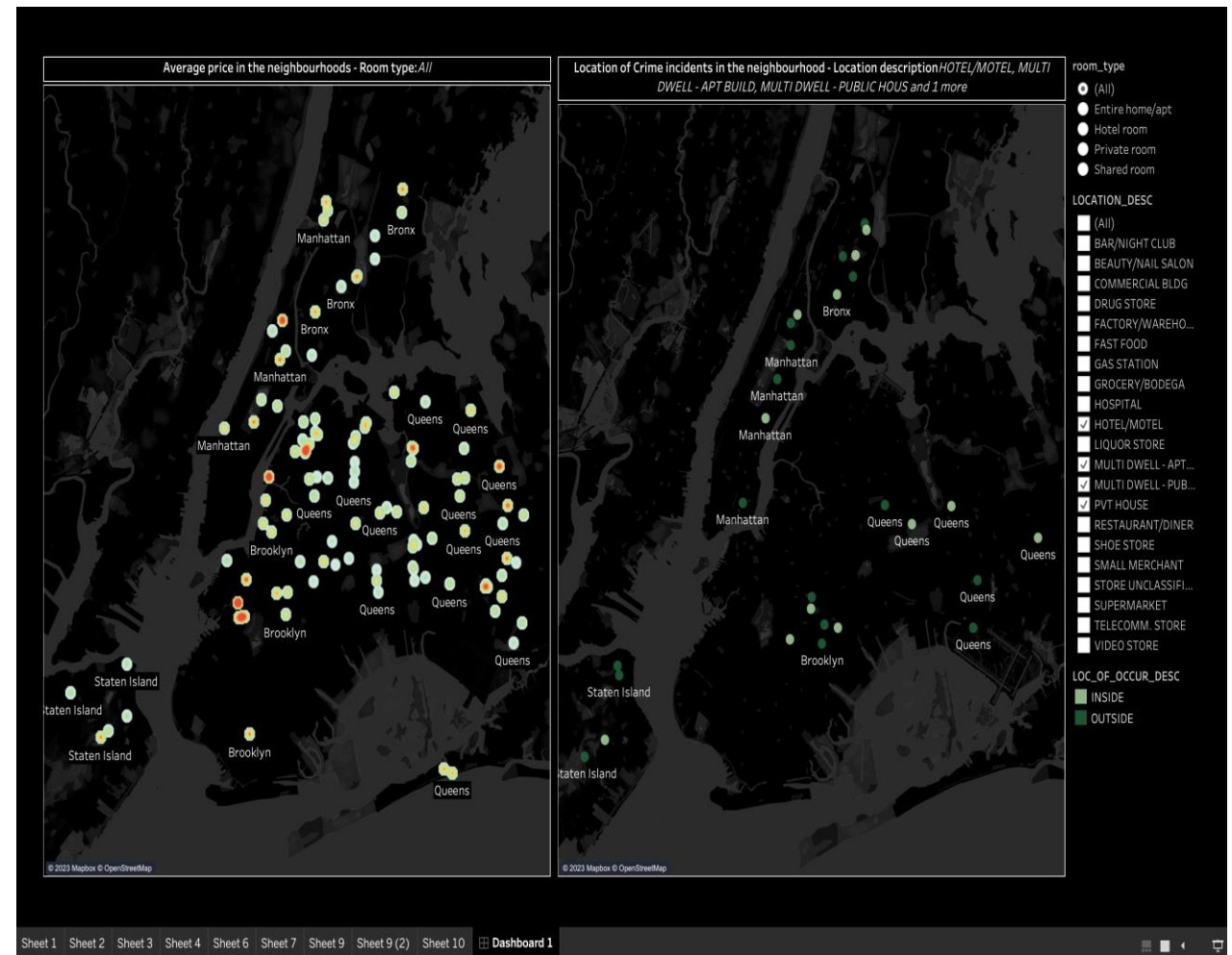
Exploring Data through Analysis and Visualization Technique

We created four different Tableau sheets for the analysis of Airbnb data. The first sheet shows the average price of listings by neighborhood group and room type. The second sheet displays the top 10 hosts by total reviews. The third sheet shows the total reviews by year. Finally, the fourth sheet displays the average price by neighborhood group and neighborhood.



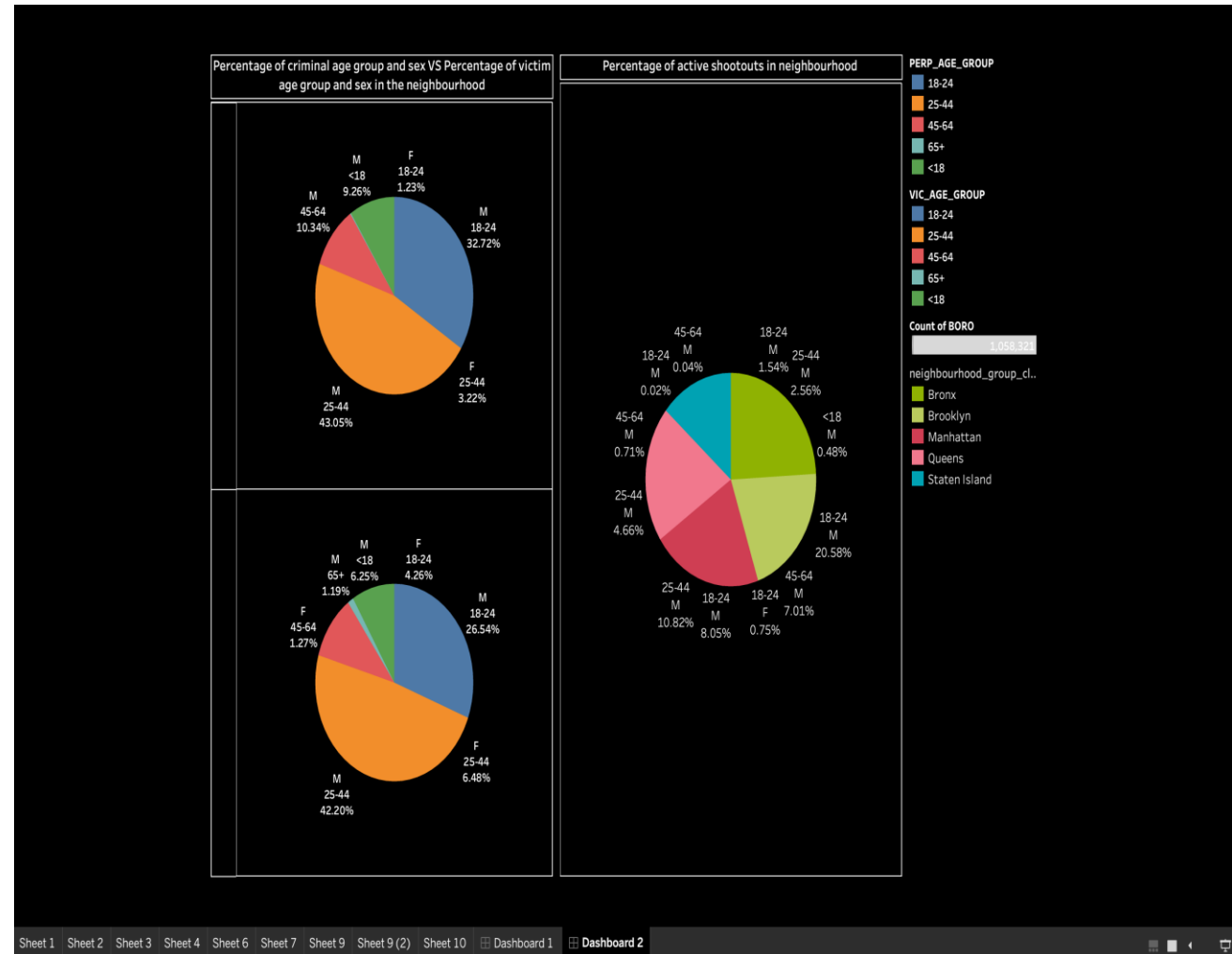
Analysis and Visualization of Airbnb and Crime Data.

- Location of crime incidents in the neighborhood can help us understand the safety of the neighborhoods we plan to visit or stay in. We used color coding to indicate the crime rate and locations with higher crime rates are shown in darker colors.
- We used the location description of crime to identify patterns of crime in different types of locations. This information can be used to make informed decisions about where to stay during travel or which neighborhoods to avoid.



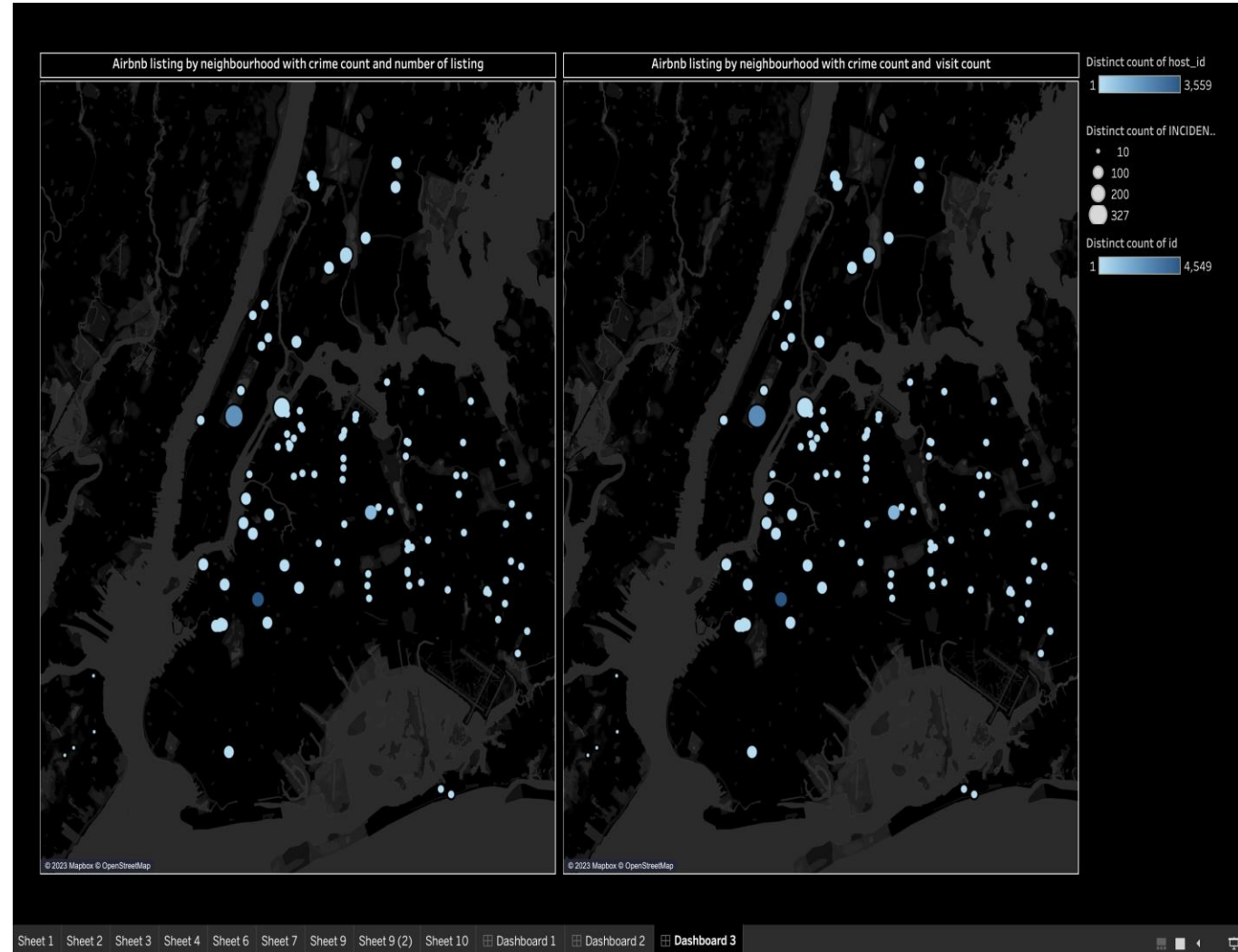
Analysis and Visualization of Airbnb and Crime Data.

- Pie charts were used to visualize the percentages of criminal age group and sex, as well as victim age group and sex, in the neighborhood. This helped us gain insight into the demographics of crimes and victims in the area.
- We also created a pie chart to show the percentage of active shootouts in the neighborhood, providing a quick overview of the incidence of such violent crimes.



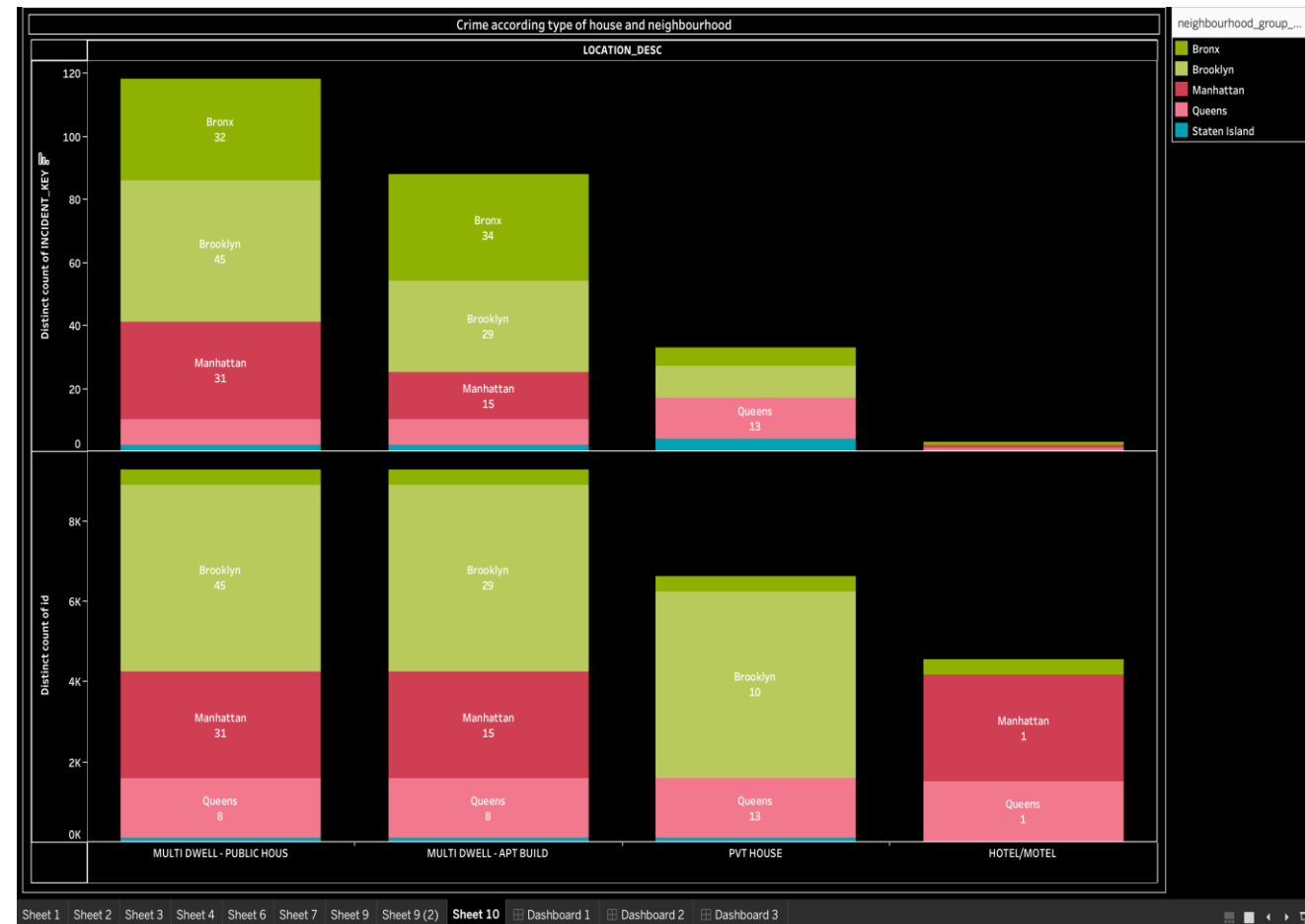
Analysis and Visualization of Airbnb and Crime Data.

- We created interactive maps using Tableau that show the correlation between crime incidents and Airbnb listings in different neighborhoods in New York. The maps highlight the number of listings, crime incidents, and visits to the neighborhoods, allowing us to identify any trends or patterns between the variables.



Analysis and Visualization of Airbnb and Crime Data.

- We used a bar graph to visualize crime according to the type of house and neighborhood. This helped us understand the correlation between crime and the type of house/neighborhood and enabled us to identify the areas with the highest crime rates.



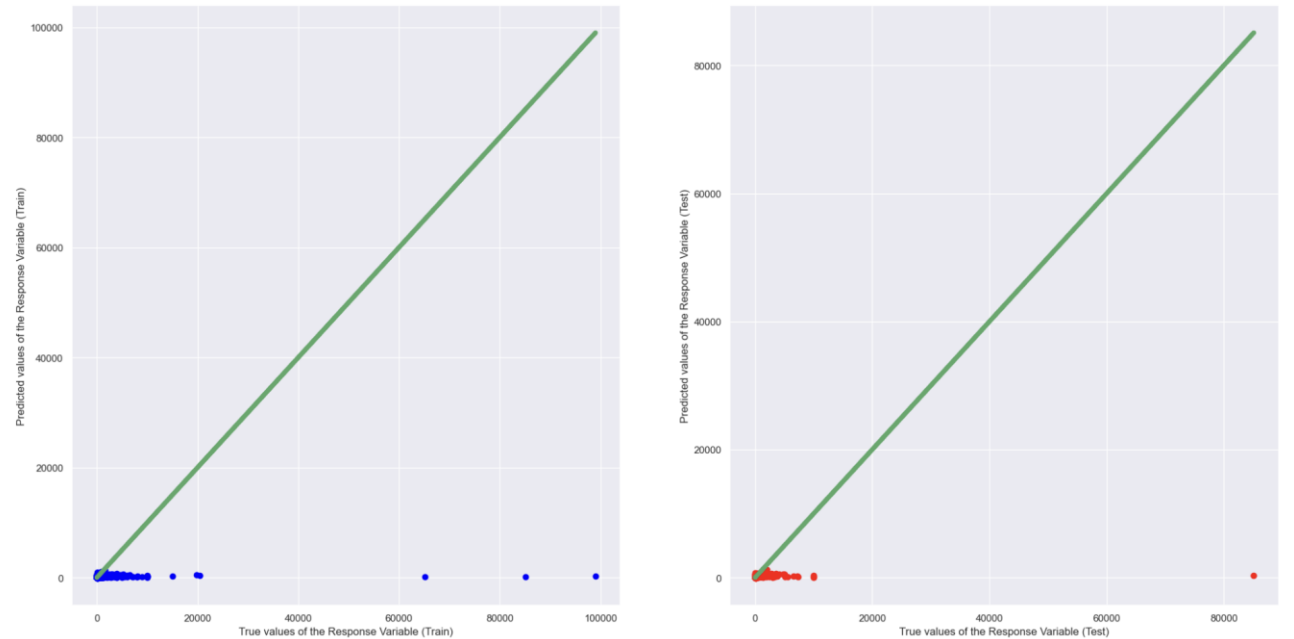
Regression Models

- Regression models are statistical tools that enable the prediction of a target variable based on one or more independent variables. They are commonly used to explore and quantify the relationships between variables, and to make predictions or forecasts based on those relationships.
- Regression models helps in Relationship Identification, Prediction and forecasting, Data exploration, Model Validation and more.

Linear Regression

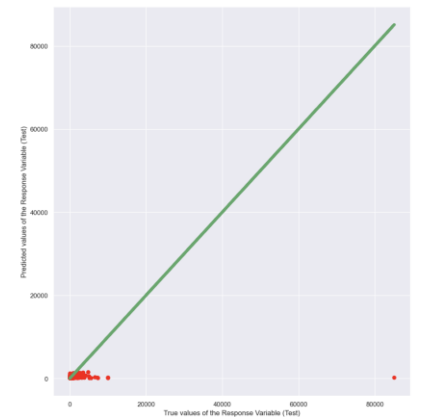
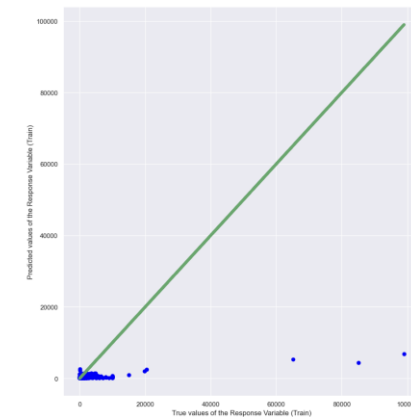
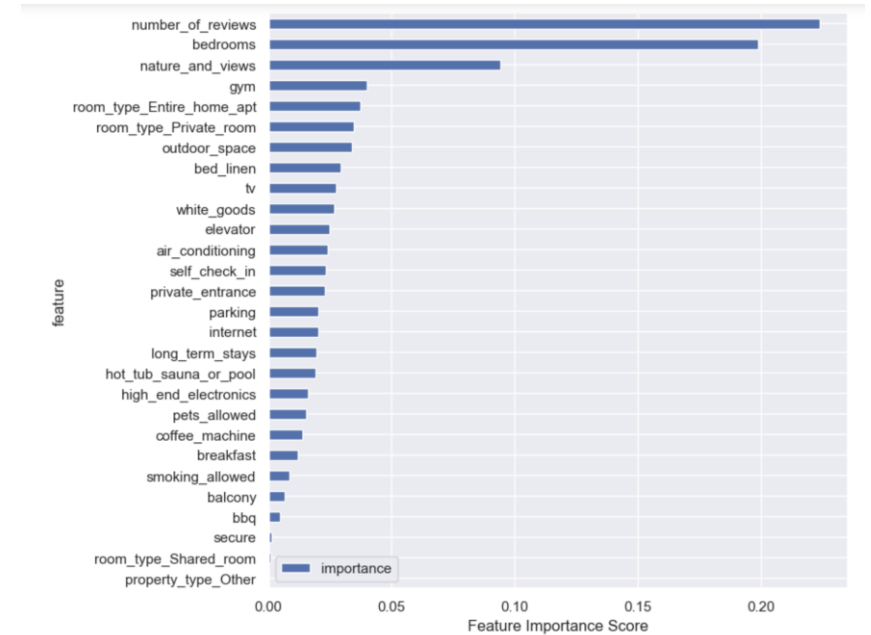
- Linear regression models the relationship between a dependent variable and independent variables, fitting a linear equation to the data to predict the dependent variable. It is used in many fields for trend analysis and prediction.
- The regression task used to predict a dependent variable value (in this case, price) based on given independent variables (in this case, "bedrooms", "breakfast", "tv", "white_goods", "smoking_allowed", "property_type_Other").

Out[21]: Text(0, 0.5, 'Predicted values of the Response Variable (Test)')



Random Forest Regression

- Random Forest Regression is a machine learning algorithm that utilizes a combination of decision trees to create an ensemble model. This model makes predictions by aggregating the predictions of multiple decision trees. Random forest regression can handle both numerical and categorical data and can be used for both regression and classification tasks.
- Benefits of this model are accurate predictions, feature selection, Scalability and more.



Sentiment Analysis

- Sentiment analysis, known as an emotion AI, is a technique that involves the use of natural language processing (NLP) and machine learning to identify, extract and analyze subjective information from text data such as social media posts, customer feedback, product reviews, and more.
- The primary goal of sentiment analysis is to determine whether the text data expresses a positive, negative, or neutral sentiment.
- Sentiment analysis is important for businesses and organizations as it helps them to gain insights into customer opinions, emotions, and attitudes towards their products, or services.

Sentiment Analysis

Jupyter Sentimental analysis Last Checkpoint: Yesterday at 12:11 AM (autosaved) Python 3 (ipykernel) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted

```
In [1]: import pandas as pd
        from textblob import TextBlob

        # Load the CSV file containing the Airbnb reviews
        df = pd.read_csv('reviews.csv')

        # Define a function to determine the sentiment of a comment
        def get_sentiment(comment):
            blob = TextBlob(str(comment))
            sentiment = blob.sentiment.polarity
            if sentiment > 0:
                return 'Positive'
            elif sentiment < 0:
                return 'Negative'
            else:
                return 'Neutral'

        # Apply the get_sentiment function to the 'comments' column
        df['Sentiment'] = df['comments'].apply(get_sentiment)

        # Group the data by listing ID and calculate the percentage of positive, negative, and neutral reviews
        grouped = df.groupby('listing_id')['Sentiment'].value_counts(normalize=True).unstack(fill_value=0) * 100

        # Print the results
        print(grouped)
```

Sentiment	Negative	Neutral	Positive
listing_id			
2595	0.000000	10.204082	89.795918
5121	0.000000	4.000000	96.000000
5136	0.000000	0.000000	100.000000
5178	3.130435	14.086957	82.782609
5203	0.000000	1.694915	98.305085
...
835703681287484641	0.000000	0.000000	100.000000
837075395576130377	0.000000	0.000000	100.000000
838131275214287871	0.000000	0.000000	100.000000
838548659003355204	0.000000	0.000000	100.000000
839119919829213270	0.000000	0.000000	100.000000

[32627 rows x 3 columns]

OUTCOMES

- The analysis conducted provided insights on the following.
- Neighborhoods with a high concentration of both Airbnb listings and reported crimes are mainly situated in the Manhattan and Brooklyn boroughs of New York City.
- Compared to private houses and hotel/motels, the concentration of Airbnb listings and reported crimes is higher in public housing and apartment buildings.
- Population density and income levels, may have a greater influence on crime.
- Crime rates tend to be higher in areas with high Airbnb activity during peak tourist season, such as summer months.



Thank you
