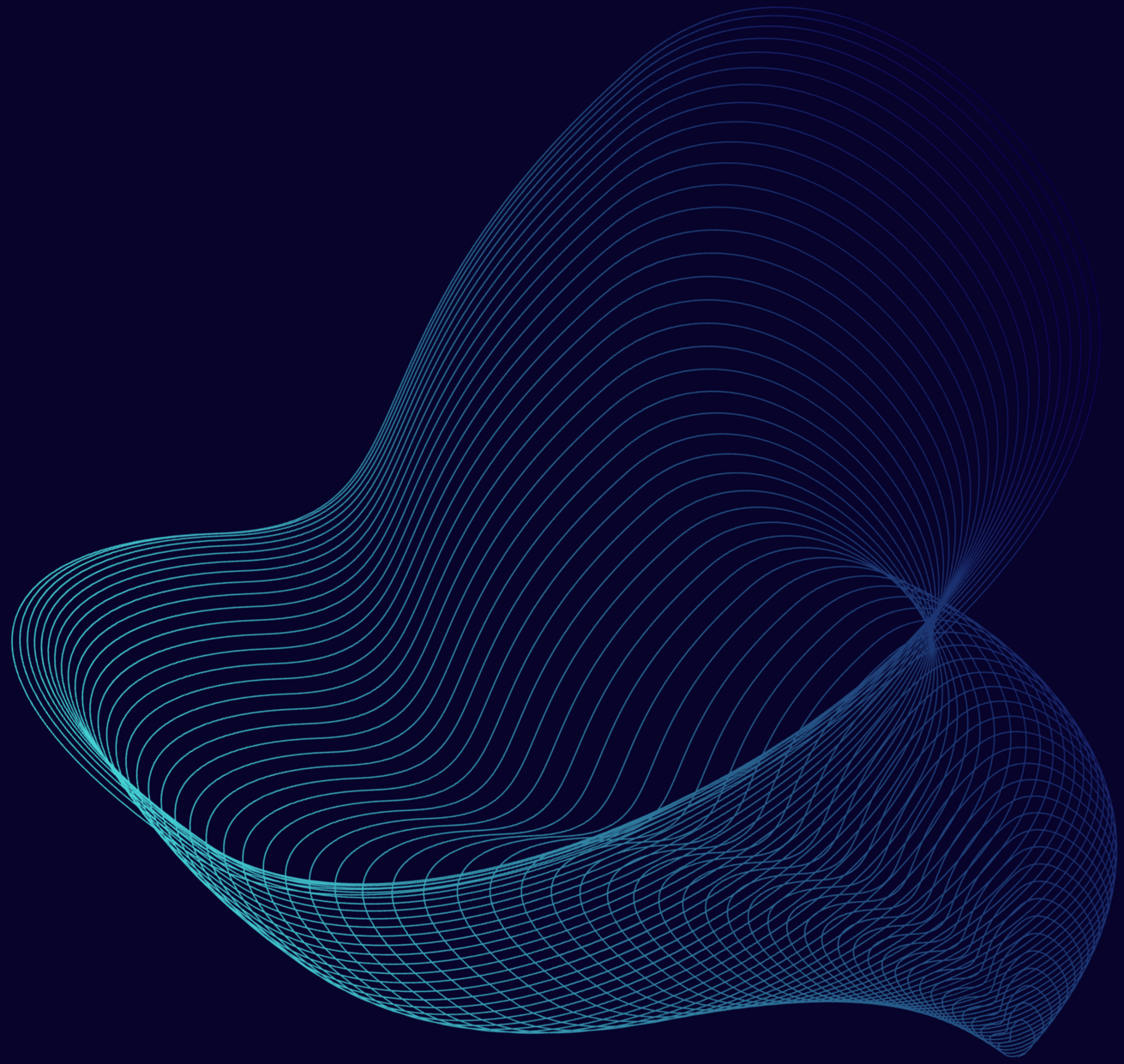
NEWYORK OPEN DATA ANALYSIS

Team:
Ramanjul Reddy Kotlo
koundinya (kp)
Sumayya Ahmed

AGENDA

Objectives
Datasets
Data preprocessing
Data visualizations
Results
Conclusion
Challenges and Solutions
Future Scope
References



OBJECTIVE

- Exploring data to discover key trends and insights
- Understanding distribution of Newyork wifi hotspot locations
- Recognising pattern in crimes relted to praking violation data for trend analysis.

DATA SETS

- Recognising pattern in crimes related to parking violation data for trend analysis. - violation code, violation, description.
- https://data.cityofnewyork.us/City-Government/Open-Parking-and-Camera-Violations/nc67-uf89/about_data

- Newyork wifi hotspot locations- Borough,Type,Provider,Name,Location,Latitude
- https://data.cityofnewyork.us/City-Government/NYC-Wi-Fi-Hotspot-Locations/yjub-udmw/about_data

DATA PRE PROCESSING

- Data is preprocessed using pandas

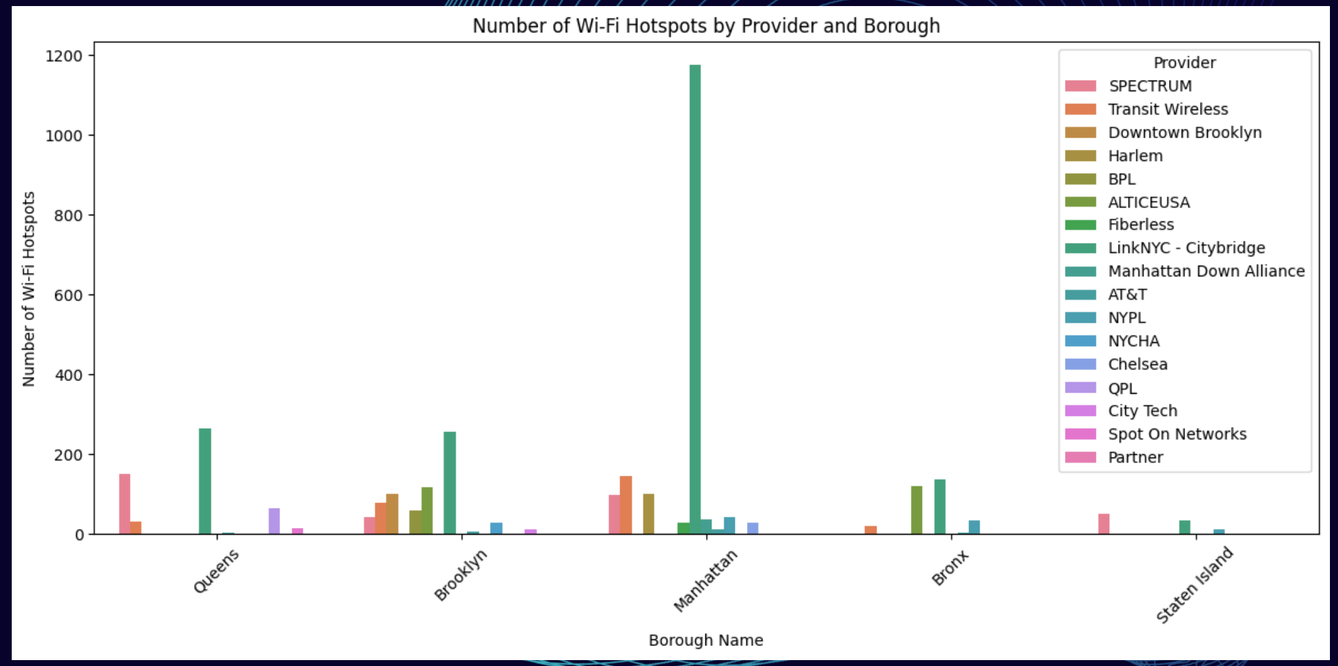
- Pre-processing involved tasks such as filtering out irrelevant data, handling missing values, and converting data types
- Encoding categorical variables

```
# Install necessary libraries
%pip install pandas
%pip install matplotlib
%pip install seaborn
%pip install folium
# Import necessary libraries for data manipulation and visualization
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import folium

# Import warnings module to suppress warnings
import warnings

# Ignore warnings of the category FutureWarning
# FutureWarning is a category of warning that indicates that a certain feature or beha
warnings.simplefilter(action='ignore', category=FutureWarning)
```

DATA VISUALIZATION

NYC DATA HOSTSPOT LOCATION

- It appears that LinkNYC - Citybridge is the dominant provider across most boroughs, with the highest number of hotspots in all except Staten Island.
- Transit Wireless has a significant presence in Queens, Manhattan, and Staten Island.
- Other providers like AT&T, NYPL seem to have a lower overall presence compared to Spectrum and Transit Wireless.

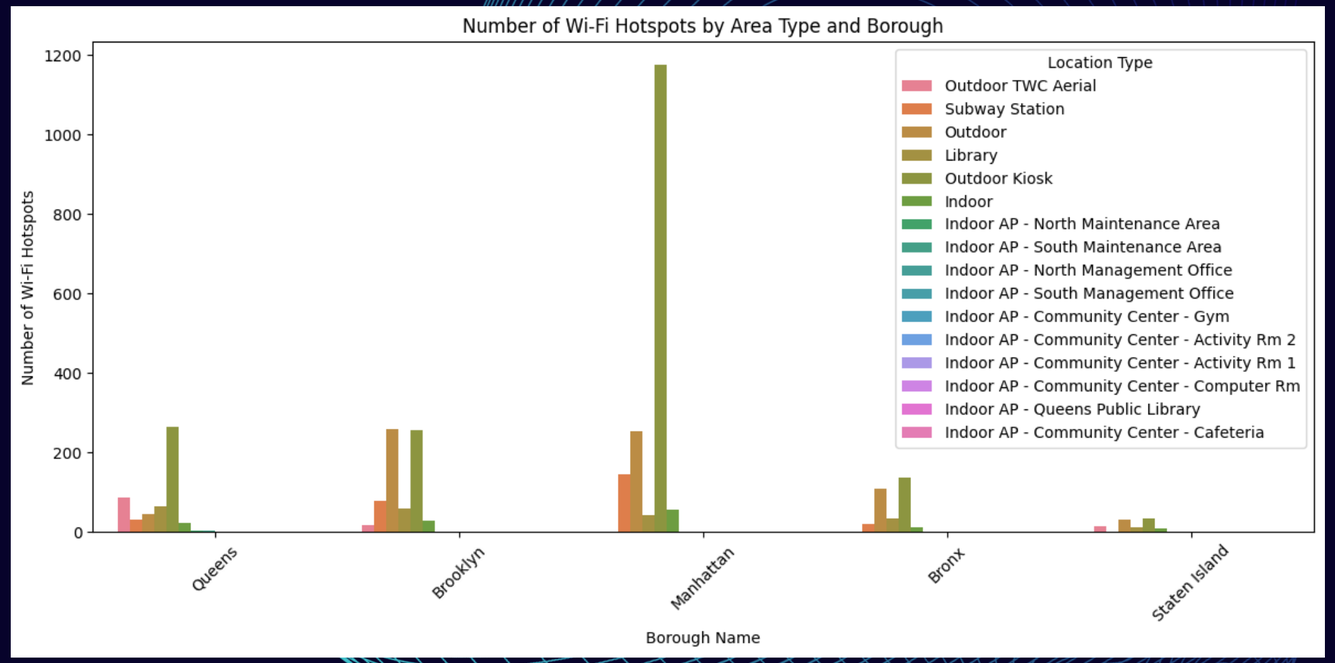


DATA VISUALIZATION

NYC DATA HOSTSPOT LOCATION

- It appears that Indoor appears to be the dominant location type across most boroughs, likely consisting of hotspots in buildings, libraries, or community centres.

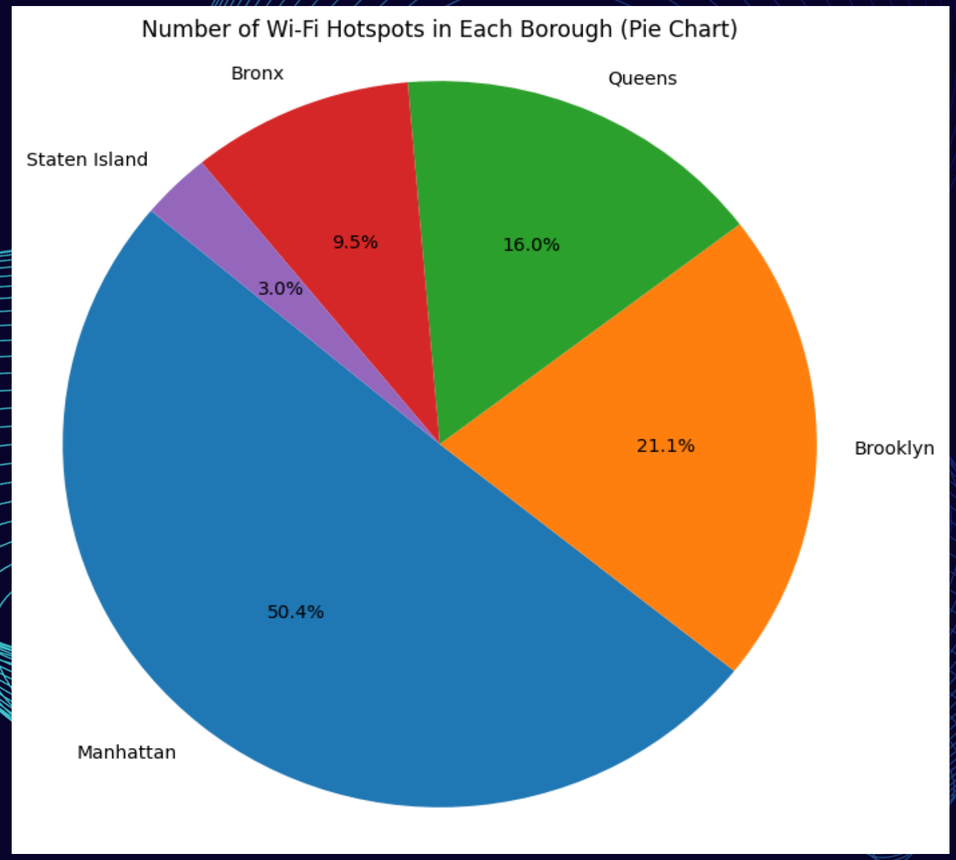
- Outdoor hotspots, including those on streetlights, kiosks, or aerial installations, also have a noticeable presence across all boroughs.



DATA VISUALIZATION

NYC DATA HOSTSPOT LOCATION

- Manhattan appears to have the most Wi-Fi hotspots, followed by Brooklyn and Queens.
- Bronx and Staten Island seem to have the least number of hotspots compared to the other boroughs.
- The percentages provide a clearer view of the distribution compared to a bar chart, highlighting the dominance of Bronx in terms of hotspot count.



DATA VISUALIZATION

```python
# Create a Folium map centered around New York City
nyc_map_least_number = folium.Map(location=[40.7128, -74.0060], zoom_start=11)

# Define a list of boroughs
boroughs = ['Manhattan', 'Bronx', 'Brooklyn', 'Queens', 'Staten Island']

# Iterate over each borough
for borough in boroughs:
    # Filter data for the current borough
    filtered_df = df[df['Borough Name'] == borough]

    # If the filtered DataFrame is not empty
    if not filtered_df.empty:
        # Group data by provider and count unique Wi-Fi providers
        provider_counts = filtered_df['Provider'].value_counts()

        # Find the least common Wi-Fi provider
        least_common_provider = provider_counts.idxmin()

        # Filter data for the least common provider
        filtered_df = filtered_df[filtered_df['Provider'] == least_common_provider]

        # Add markers for each Wi-Fi hotspot location in the filtered DataFrame
        for index, row in filtered_df.iterrows():
            folium.Marker([row['Latitude'], row['Longitude']], popup=row['Provider'])

# Save the map as an HTML file
nyc_map_least_number.save("boroughs_least_wifi_providers_locations.html")

nyc_map_least_number
```

DATA VISUALIZATION

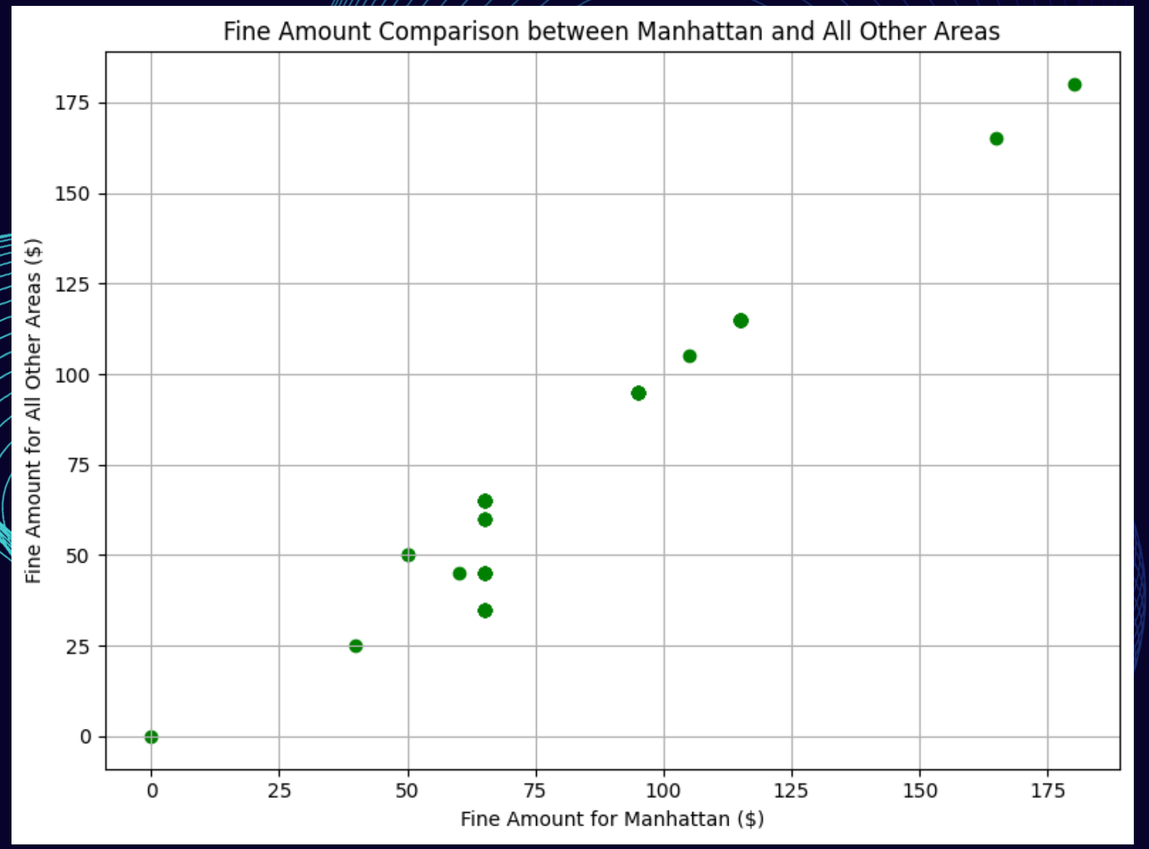
PARKING VIOALTION

Observations:

- Manhattan Distribution:
- The data points for Manhattan fines are clustered around specific values.
- There seems to be a concentration of fines at certain levels.
- This could indicate consistent enforcement or specific violation patterns in Manhattan.

Other Areas Distribution:

- The fines in other areas exhibit more variability.
- The data points are spread out across a wider range of fine amounts.
- Some areas have significantly higher fines, while others have lower fines.



CONCLUSION

- Manhattan has the highest number of Wi-Fi hotspots overall, with LinkNYC - Citybridge being the dominant provider and also leads in the number of hotspots, with the majority being outdoor hotspots.
- The Bronx and Staten Island have the least, indicating a potential area for infrastructural development
- For most points, the fine amount for all other areas is lower than for Manhattan, which suggests that Manhattan has a higher fine rate for parking violations.

CHALLENGES AND SOLUTIONS

- To understand what columns meant in the data Solution: We overcame this challenge by going through data dictionary files on the website and explored relevant blogs to gain further insight.

- Working with GitHub Solution: We faced authentication issues while accessing GitHub to collaborate as a team. However, we overcame this challenge by establishing an SSH key connection..

- Working with huge data sets and cleaning and visualizing them. Solution: Specifically, when implementing the parking violation dataset, we encountered issues with column names causing errors during cleanup. To address this, we updated the column names to rectify the errors.

FUTURE SCOPE

- Investigate the reasons behind the fine distribution in Manhattan and Evaluate the impact of fines on compliance and public behavior.

- Explore other relationships and trends between variables in the data Implement real-time data visualizations to dynamically represent and keep up with the evolving data

REFERENCES

https://data.cityofnewyork.us/City-Government/Open-Parking-and-Camera-Violations/nc67-uf89/about_data

https://data.cityofnewyork.us/City-Government/NYC-Wi-Fi-Hotspot-Locations/yjub-udmw/about_data


End of presentation
THANK YOU