



Analysis on AIRBNB NYC Listings

Presentation to Data Analysis Managers
and Lead Data Analysts

Agenda

- Objective
- Background
- Data Profile
- Data Preparation
- Key Insights
- Assumptions Made
- Methodology followed

Objective



- Study key insights from the Airbnb NYC data of property listings
- Understand customer preference and activities in different property types and areas
- Catalyze management decisions in acquisitions and customer experience, by providing key indicators of customer behavior

Background

- Travel restrictions around the world due to global pandemic has affected AirBNB's business heavily
- Now that restrictions have lifted, different leaders at AirBNB want to understand the strategy to increase revenue
- Dataset of Airbnb listings in New York for the past 8 years have been used to analyse and understand customer preference and behaviour

Data Profile and Preparation

- Property and Host Listings of Airbnb in NYC are has been provided for the period from 2011 to 2019
- Customer ratings provided on properties is used as the key factor indicating customer preference
- Other details of property listings such as neighborhood group, room type, price, minimum nights of stay and availability around the year have been used to analyze and derive insights



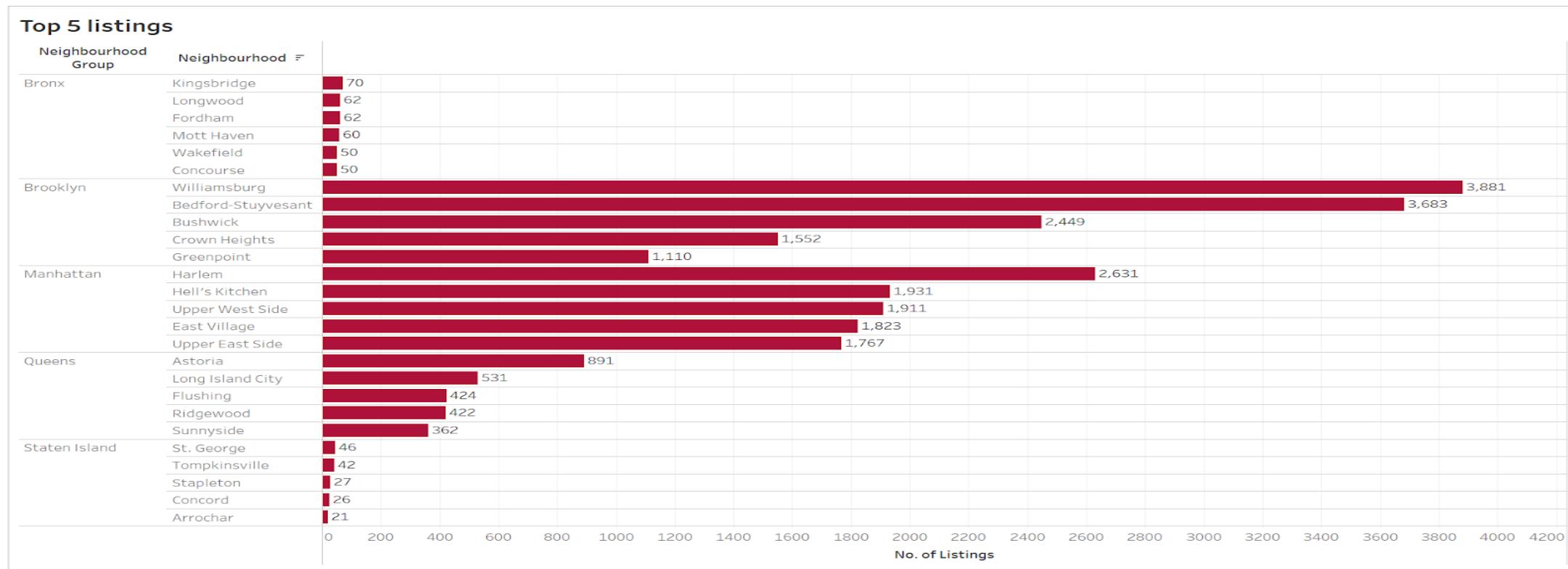
- The data was cleaned for any missing and null values
- We used Pandas library to explore and identify outliers in all numerical columns
- The outliers were treated and 1.6% of the dataset was removed to avoid any bias when further exploring the dataset
- Further analysis of the data was performed using Tableau



KEY INSIGHTS

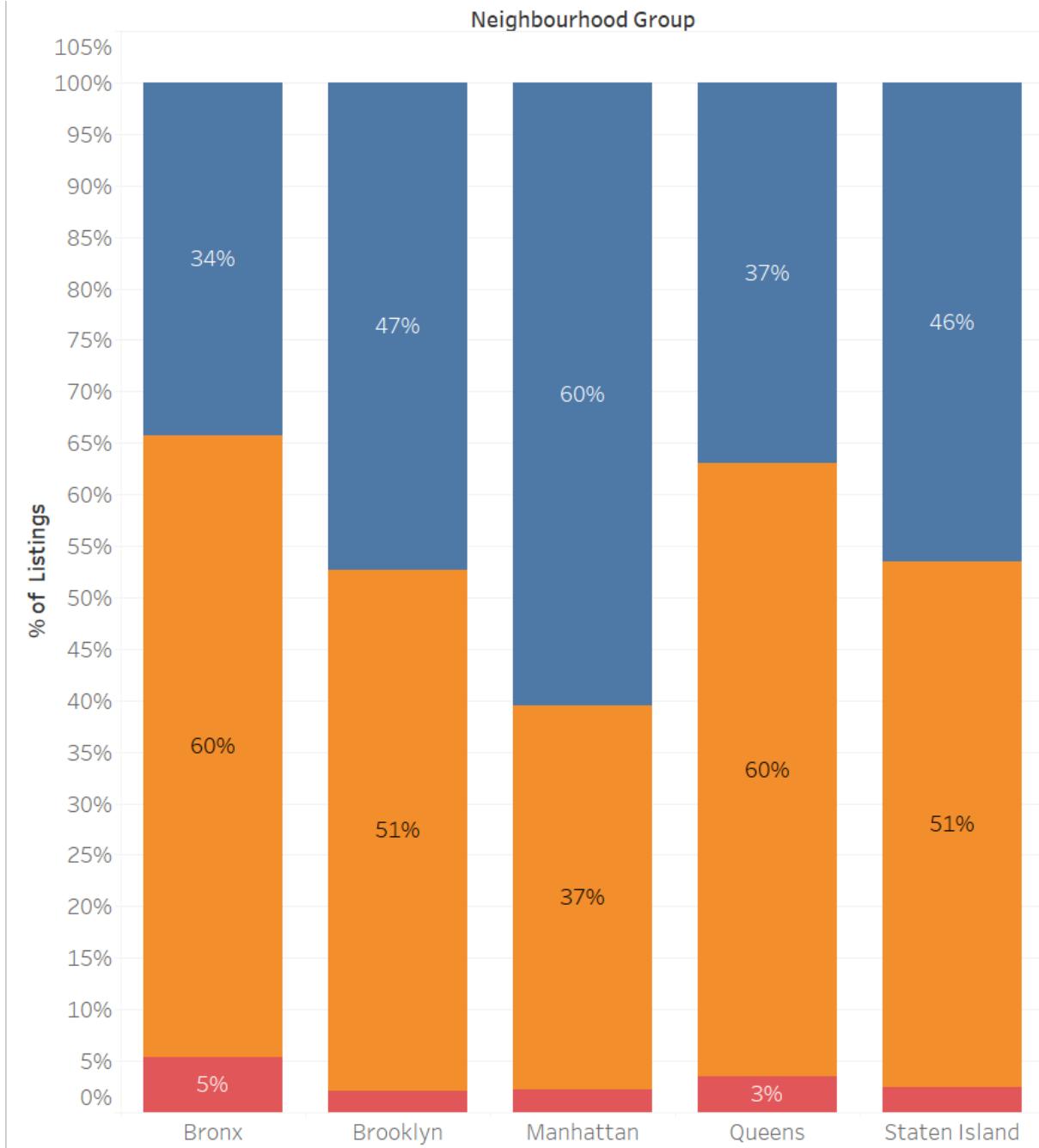
Neighborhood with highest number of listings

- Brooklyn and Manhattan are the most popular neighborhoods groups, with the highest number of listings
- Around 25% of the property listings are concentrated in the neighborhoods of Williamsburgh, Bedford – Stuyvesant, Bushwick and Harlem
- Staten Island has the least number of bookings among the 5 neighborhood groups



Most prevailing room type

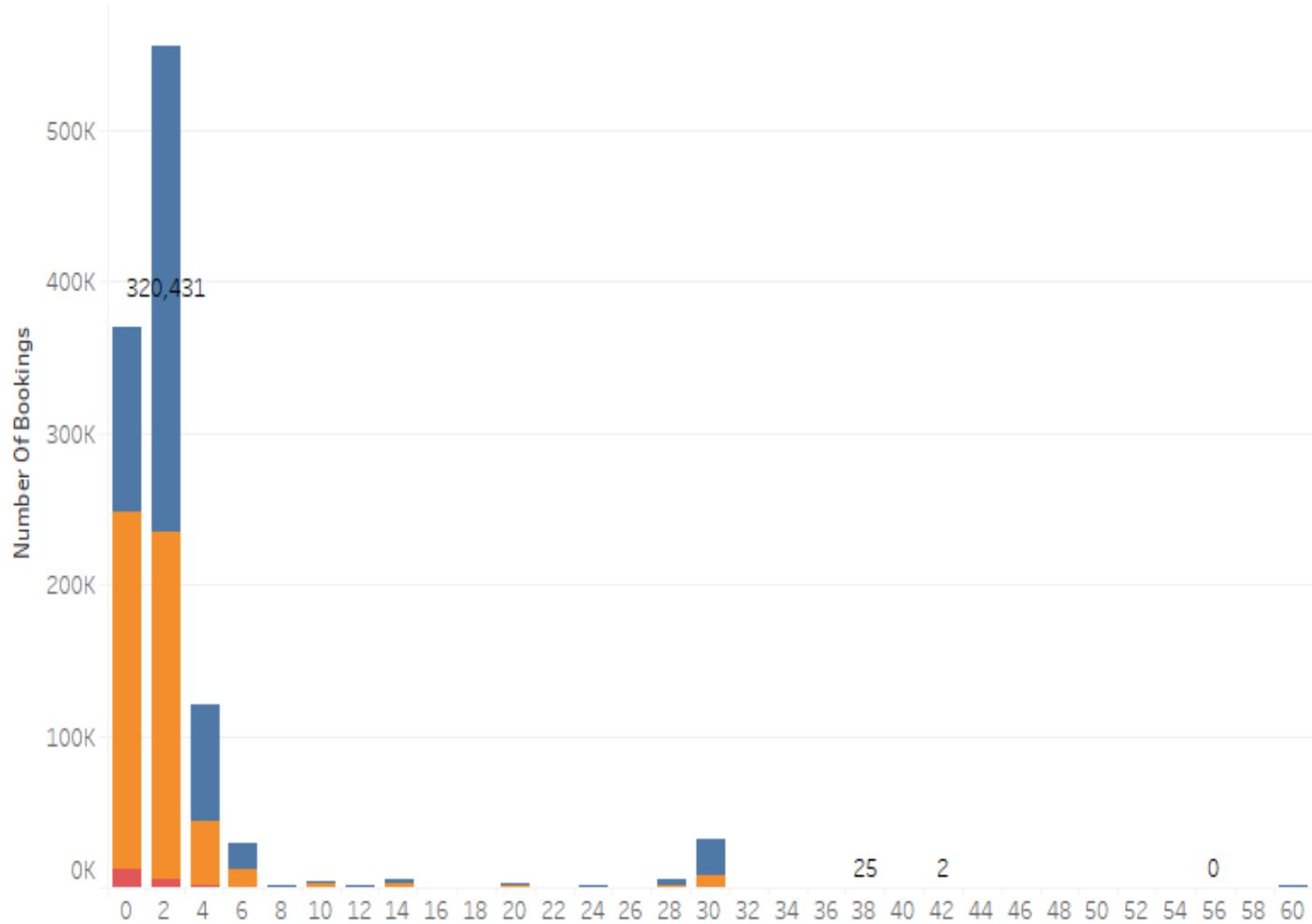
- Private rooms are the most prevailing listing across all neighborhoods, covering 52% of the total listings
- The second most prevailing room type is Entire homes/apartments. Shared room are the least preferred
- Manhattan has an exception where entire home type listings are more available than private rooms



Bookings as per minimum stay term

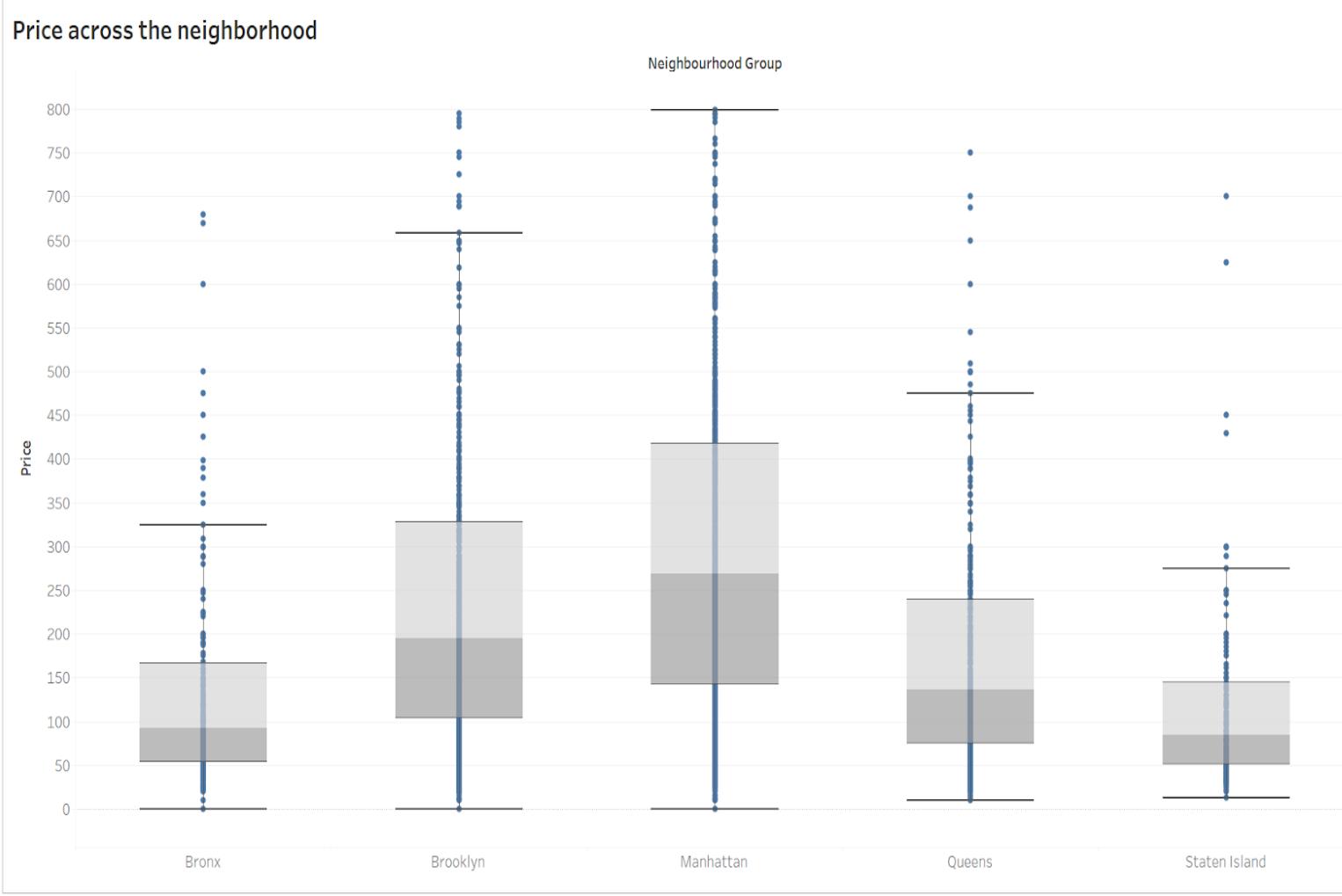
- High number of bookings are noticed in listings with minimum stay of 2 days
- Private room customers prefer to have minimum stay of less than 2 days
- Sudden spike in bookings on 30 days minimum stay is noticed reasonably for entire homes/apartments and private rooms

How many nights do you want to stay?



Price across Neighborhood

- Manhattan has the highest median price of \$269, followed by Brooklyn at \$196
- Manhattan also has the widest range of listings as per price
- Staten Island is the most affordable neighborhood, with very few premium listings



Most Preferred Neighborhood and Room type

- Staten Island's entire homes have the highest average bookings/listing out of all neighborhoods
- Generally we can see that average bookings of Entire home type listings are higher than other types, except for Manhattan
- Even though Brooklyn and Manhattan have high no of listings, their average bookings are lower as compared to other neighborhoods

Bookings in the neighborhood



Assumptions made during analysis

- There is no column that shows the number of bookings, hence it is assumed that number of reviews against a listing is the number of bookings made.
- It was also assumed that outliers in the dataset was an anomaly and that such prices or minimum number of nights is impossible and hence those data were removed in the data cleaning process
- No. of listings has been created based on whether the listing has a name or not and those which did not have a name were assumed that the listing is not available any more.

Methodology used:

- The given data was explored using pandas library in Python and all different numerical columns were explored to find any outliers
- The outliers were treated and 1.6% of the dataset was removed to avoid any bias when further exploring the dataset
- The new dataset was then analyzed and visualized in Tableau
- We also followed different story-telling approach for different audience:
 - Data Managers & Lead Data Analysts : We presented them key insights of the past data and the procedures followed in cleaning the data and analyzing it.
 - Heads of Acquisitions & operations and User experience: We presented them suggestive insights in business language, and refrained from using any technical details. Our presentation was structured to provide major feedback on customer experience and suggest further acquisitions.