Airbnb in New York City

DATA VISUALIZATION AND TABLEAU DASHBOARDS PROJECT

Anna Karlova

September, 2023

INTRODUCTION

This presentation focuses on the Airbnb listings dataset in New York City and aims to show exploratory data analysis to understand the relationships and trends in the development of Airbnb's market from 2008 to 2015.

The visualizations and dashboards provide a comprehensive view of the data to support decision-making and understanding of the market dynamics.

PROJECT EXECUTION

Connect to data

Step 1

Loading the data from Airbnb's New York City public data from 2008 to 2015.

Explore and clean data

Step 2

Detecting different data types in the dataset and cleaning data (removing duplicates, outliers).

Data visualization

Step 3

Building charts, graphs, maps for understanding data relationships, and forecast analysis.

Data investigation

Step 4

Studying different trends and patterns, detecting meaningful key points, formatting questions.

Creating dashboards

Step 5

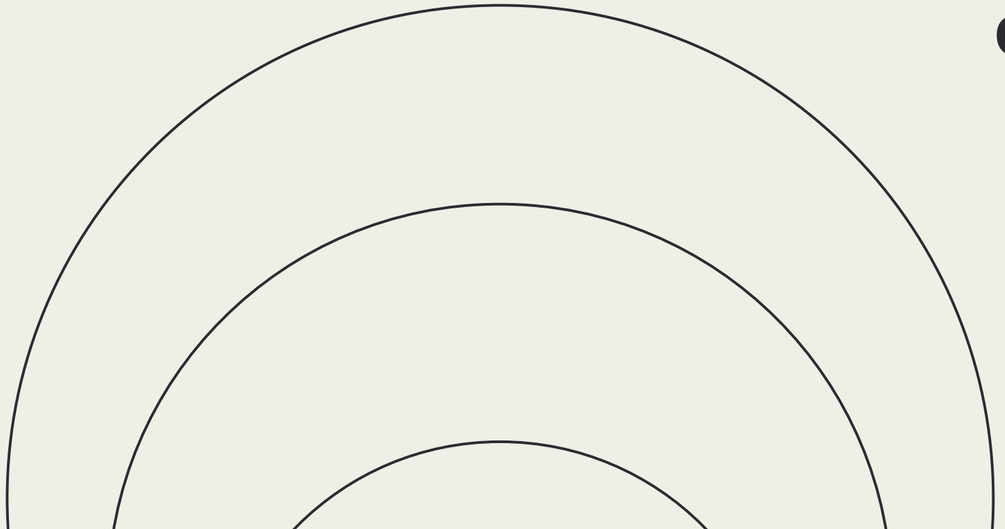
Design the layout, effectively representing the data, and adding filters, dropdowns, and other controls.

DATASET DESCRIPTION

The dataset contains 13 variables:

- 7 numeric, 1 date, and 5 categorical fields (zipcode was changed from numerical to ZIP Code type)

These variables describe the hosts that began operating on the Airbnb market from 2008 to 2015, their ID, date of registration, listing description, property types, neighbourhood, zipcode location, room types, number of beds, records, prices, ratings, and reviews.



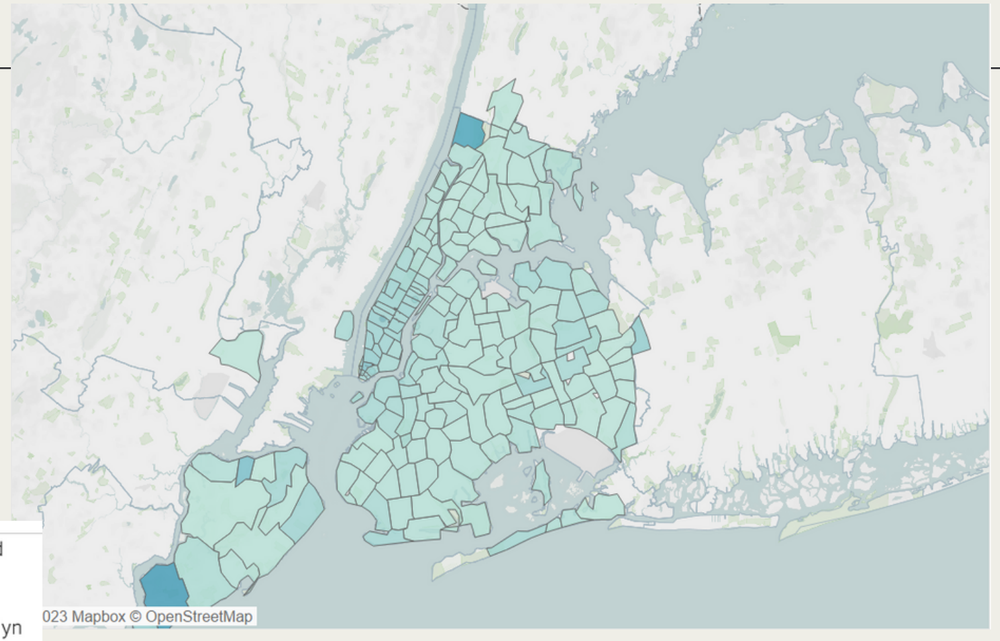
Cleaning data:

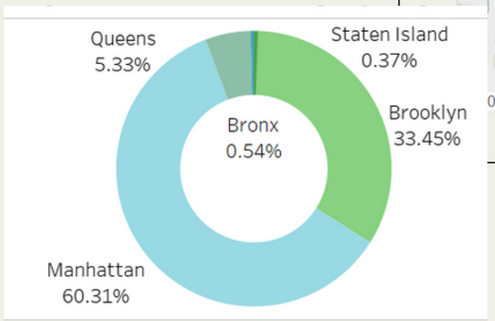
- removed 17 rows of duplicates,
- excluded 4 rows that contain zip codes of other states or cities,
- after data cleaning, there are 30457 observations.

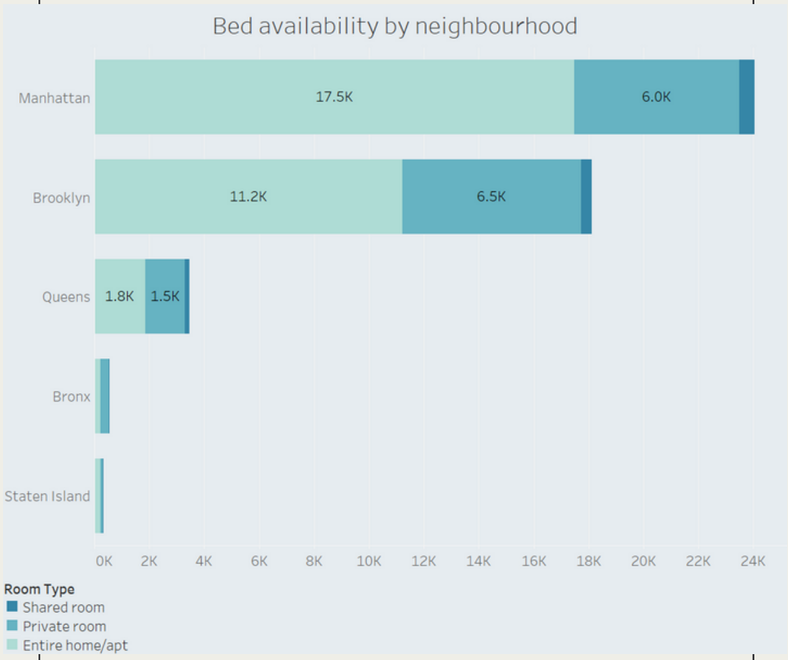
VISUALIZATION

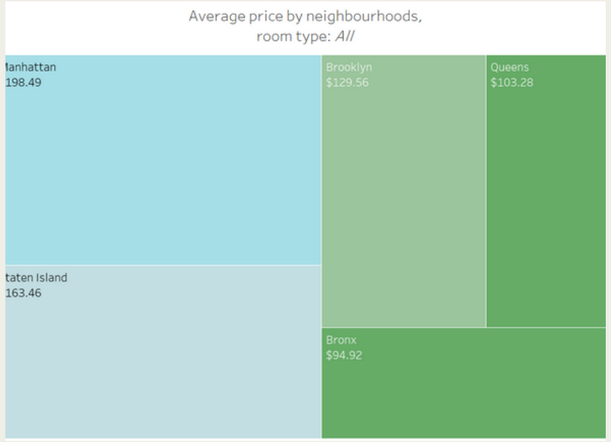
To visualize and analyze data effectively, the common types of charts were used:

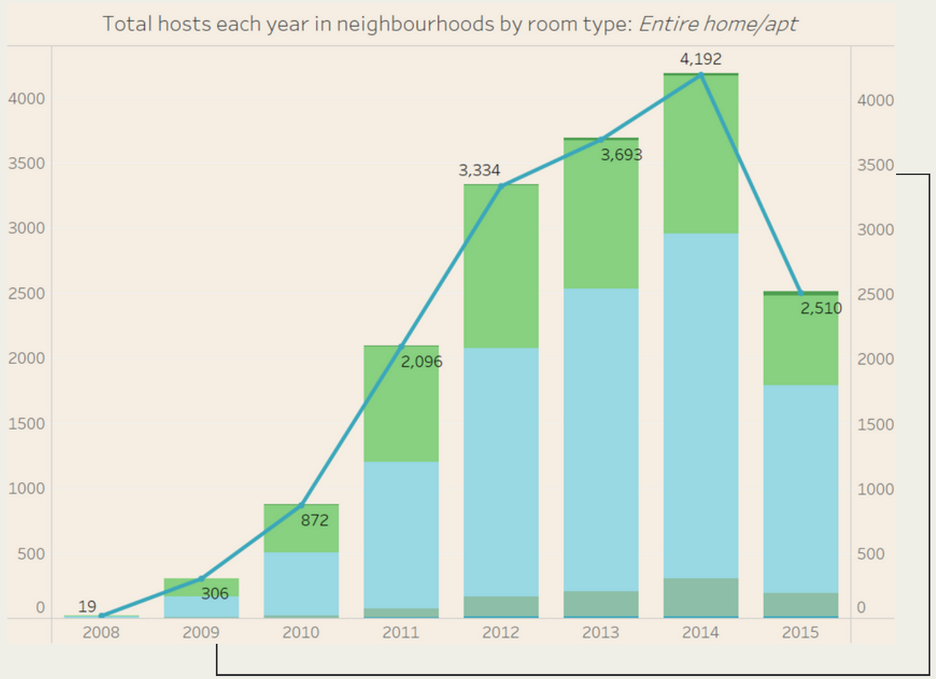
- maps, bar and line charts, pie charts, and tree maps.





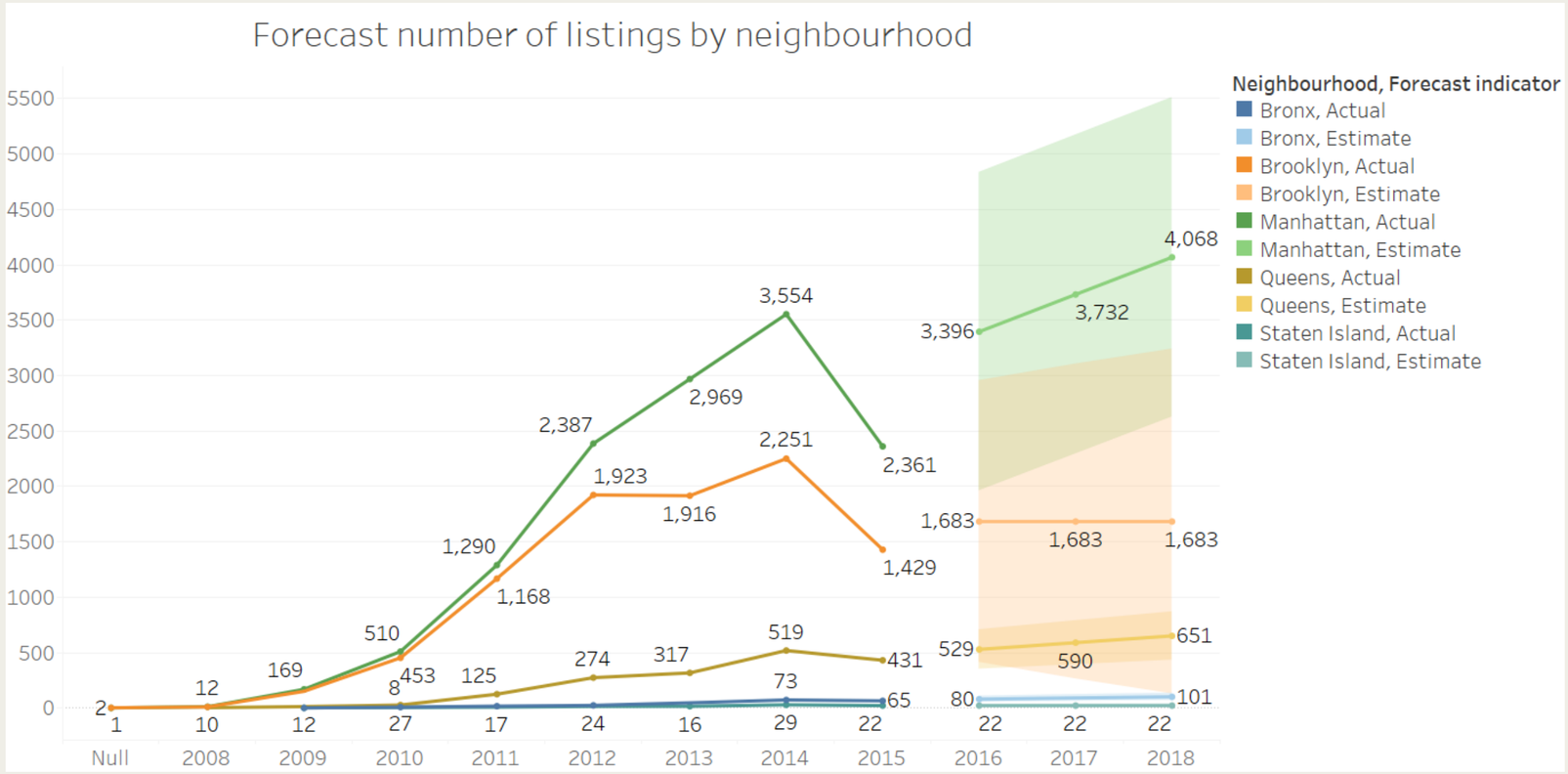






FORECAST

- The number of newly listed hosts increased from 2008 to 2014 and experienced a sharp decline in 2015.
- Manhattan and Brooklyn saw the most number of increases in listings.
- With forecast analysis, Manhattan and Queens are both projected to increase the number of listings by ~10% from 2016 to 2018



Additional research:

Airbnb removed a significant amount of listings in 2015. General reports that 72% of Airbnb listings violated New York City regulations. In 2015, Airbnb published its data from 2014 to 2015 to the public, but 1,500 listings were removed from the dataset, which makes one question the validity of the released data and, consequently, the forecast results.
The impact of regulatory changes and other factors on the Airbnb market in New York City continued to be a subject of fluctuation in the years following 2015.

INVESTIGATION

- Manhattan and Brooklyn have the highest number of listings compared to other neighborhoods, likely owing to their strong appeal to tourists.
- Bronx, Queens, and Staten Island exhibit a higher prevalence of listings for private rooms as opposed to entire homes or shared rooms.
- The average bed price tends to rise with the size of the room provided, particularly for entire homes/apartments.
- Across all neighborhoods, entire homes/apartments consistently have the highest average bed prices.
- Manhattan stands out as the neighborhood with the most expensive room types on average.
- In general, the average rating scores per zipcode are quite positive, with most zipcodes boasting ratings above 90%.
- There is a consistent pattern of increasing average listing prices in every NYC neighborhood until the year 2011. However, in 2012, there was a decline in the average price for each neighborhood on an annual basis.
- This abrupt decrease in average listing prices may be attributed to a significant surge in the number of listings.
- Between 2014 and 2015, there was another decline in average listing prices, possibly stemming from a combination of factors such as increased competition due to the rising number of listings and stricter regulations imposed on short-term housing rentals in NYC.

DASHBOARDS

Several essential steps were involved in creating dashboards:

- design the layouts, add worksheets that contain visualizations, use actions to allow users to interact with the dashboard and dynamically update the displayed data, use floating elements, add filters specific to the dashboards, and create calculated fields.

The created dashboards enables to:

- Track changes in average prices over the years by zipcode, room type, and neighbourhood.
- Display average ratings of listings based on location as well as bed availability.
- Illustrate the distribution of listings across neighbourhoods.
- Present the total number of hosts and reviews categorized by room type and location.

Additionally, the dashboards explore and answer various questions related to pricing, ratings, listing distribution, and host demographics.

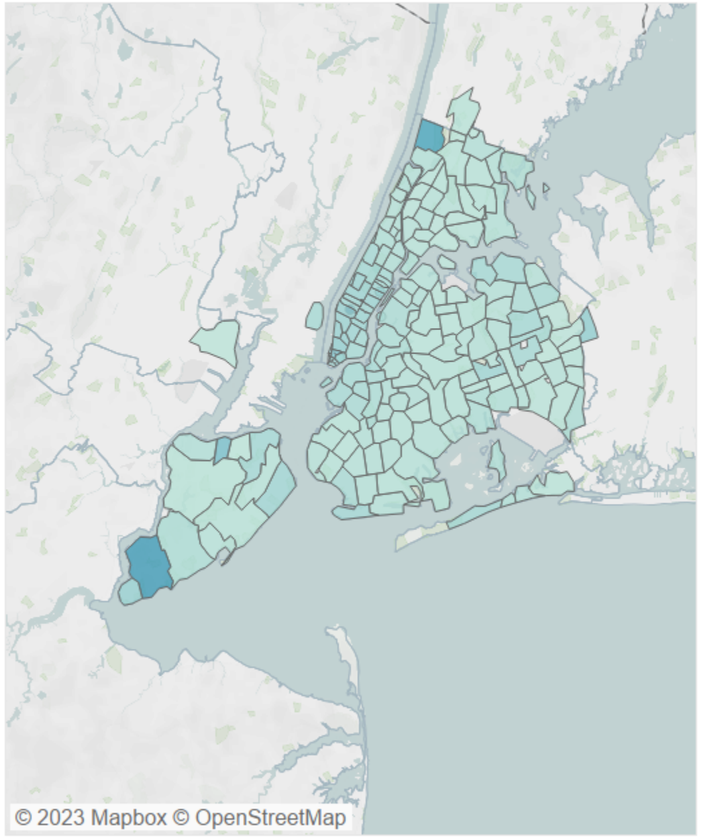
Airbnb Dashboard 1

Select Year
All

Select Room Type
All

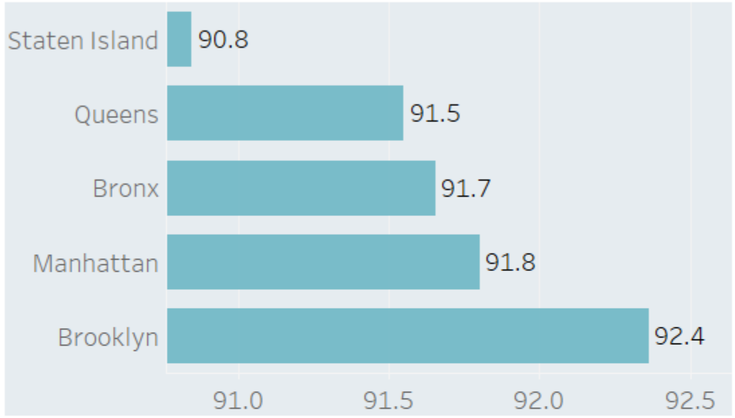
Select Property Type
All

Average price by zipcode

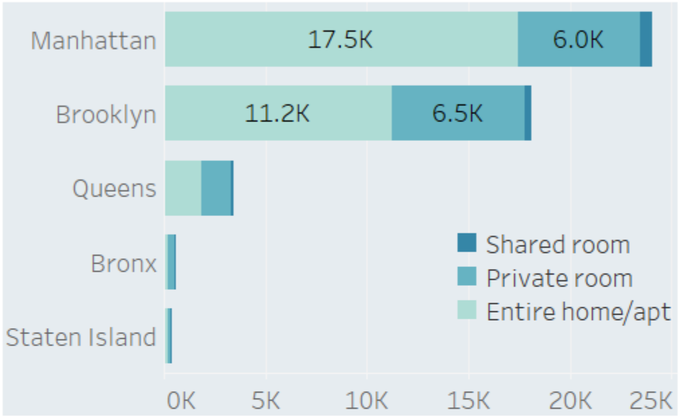


Avg. Price
44 1,500

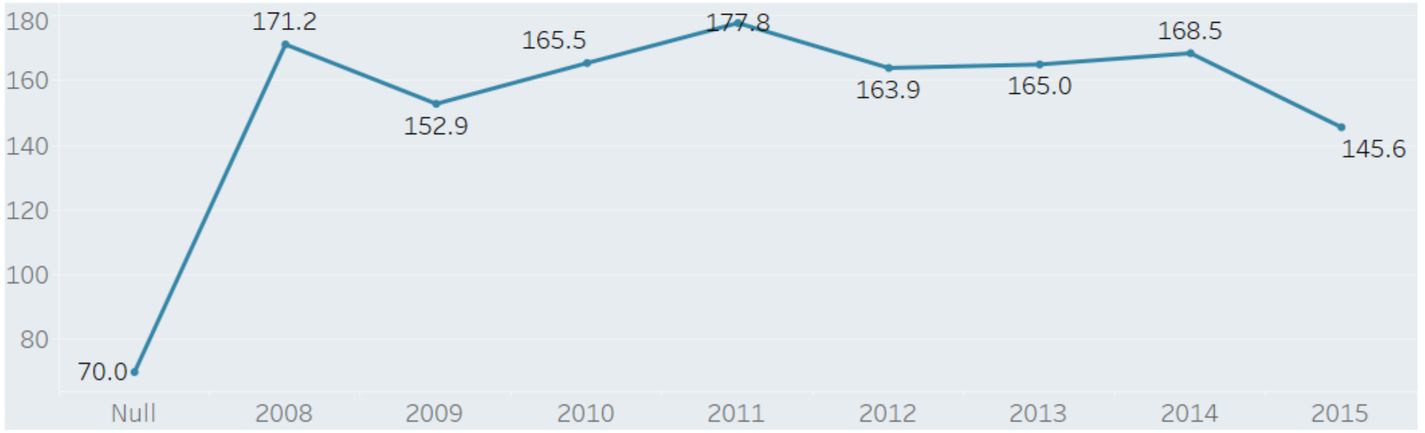
Average rating of neighbourhood



Bed availability by neighbourhood



Average price per year



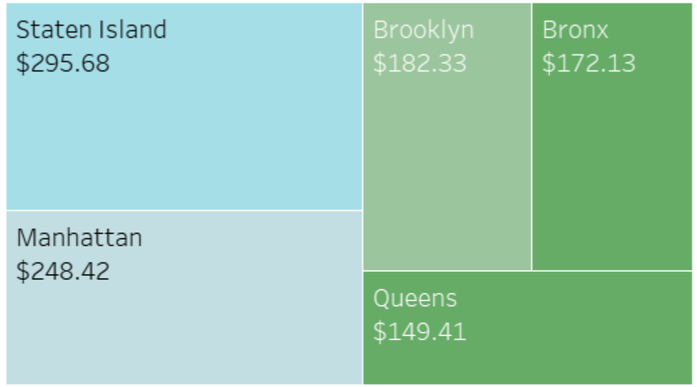
Airbnb Dashboard 2

Select Room Type
Entire home/apt

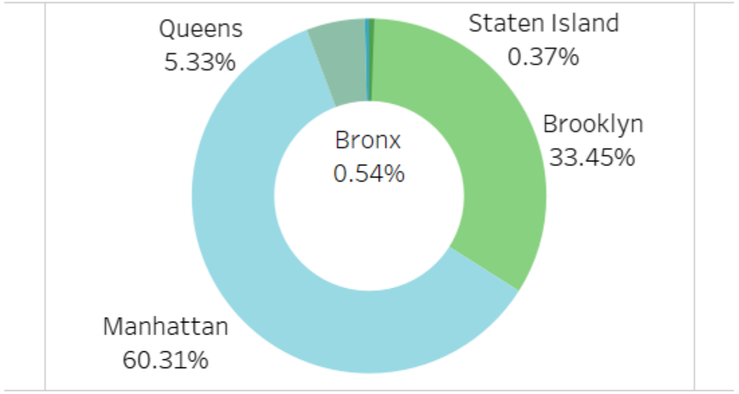
Total reviews
366,304

Total hosts
24,421

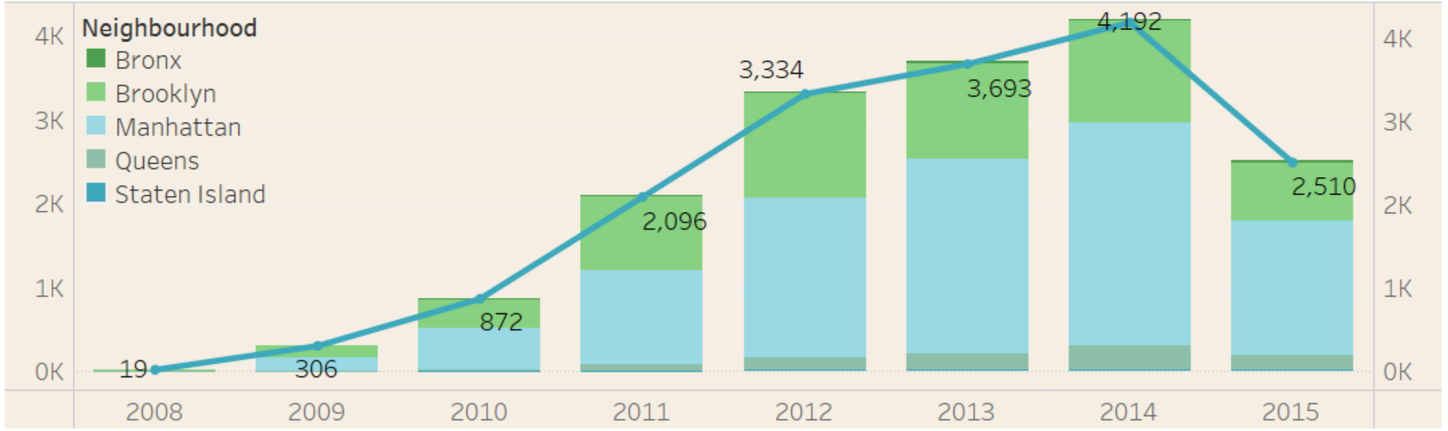
Average price by neighbourhoods, room type: *Entire home/apt*



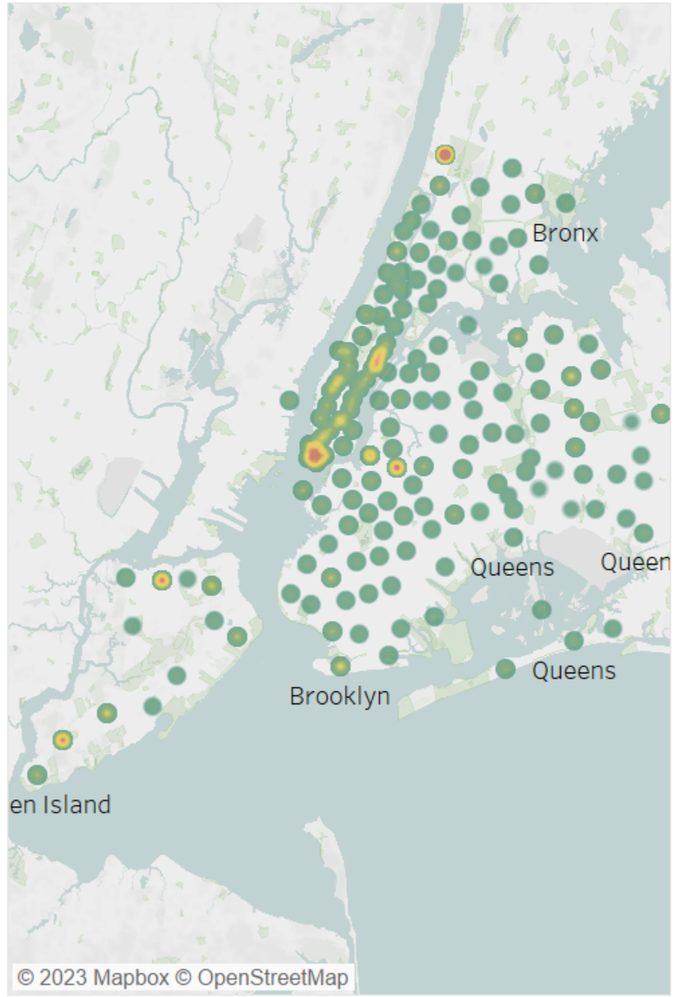
Percentage distribution of listings by neighbourhoods



Total hosts each year in neighbourhoods by room type: *Entire home/apt*



Average price in neighbourhoods by room type: *Entire home/apt*



CHALLENGES

- Generating creative chart ideas and selecting appropriate chart types.
- Creating a clean and efficient dashboard demanded significant time and effort.
- Encountered some warnings when trying to load Name due to the number of values.

FUTURE GOALS

- Further cleaning of data because a significant part of fields/columns contains null values.
- Get updated data for subsequent years to compare the results of the forecast and the real market situation.
- Get data of points of interest for further investigation of how geographical proximity to attractions impacts on price and popularity of listings.

Thank you!