Airbnb in New York City

Team Member

Hicham Bentahila

Suhani Arjin

Wei Zhang

Project Agenda

❖ Project focus

❖ Data Source and Limitation

❖ ETL Process

❖ HTML Production

❖ Visualization - Mapping

❖ Visualization - Data Analysis

❖ Conclusion

Project Focus

❖ How owners choose to listing room type?

❖ How to determine the hot spot in the city? will cost be the main factor?

Data Source

❖   Summary information and metrics for Airbnb listings in New York

City

❖   Data as of June 5th, 2023

❖   Report released quarterly as a CSV

❖   Source: Inside Airbnb Portal : http://insideairbnb.com/get-the-data

Data Limitations

❖ Considering the large amount of data, sometimes the browser fails to display the entire data and returns an error "Out of memory", hence, we have decided to create routes with customized data ready to be used for our visuals.

❖ The accuracy of the data cannot be authenticated as we rely entirely on the data provided by Inside Airbnb in their report released quarterly.

ETL Pipeline

Extract

❖ Data downloaded from the source: "Inside Airbnb Portal" as a CSV

Transform

❖ This project used Pandas to clean the data, remove unnecessary columns and save the cleaned data as a CSV

Load

❖ This project used Pandas and SQLite to create a database, tables and import the CSV file

SQLite Database

Our data is:

❖ Flat and simple dataset
❖ Single table in SQLite that includes all our data from the CSV file

Why SQLite ?

❖ Single file format
❖ Open source and free
❖ Low memory footprint
❖ Serverless architecture

www.quickdatabasediagrams.com

Listings		
id	🔑 INT,
room_type	VARCHAR(50),
price_per_night	INT,
minimum_nights	INT,
rating	FLOAT,
number_of_reviews	INT,
availability	INT,
neighbourhood_group	VARCHAR(50),
neighbourhood	VARCHAR(80),
latitude	FLOAT,
longitude	FLOAT

Python Flask API

Python Flask powered API used to :

❖ deploy the database, and
❖ fetch the necessary data from the database, and
❖ render template

```python
# Use the fucntion render_template to return the html page "index.html" when running the app
@app.route("/")
def index():
    return render_template('airbnb_nyc.html')
```

HTML Production

❖ HTML and CSS: Used to create the structure of the web

application

❖ Javascript with three libraries:

  ➢ Leaflet

  ➢ D3

  ➢ Plotly

  ➢ Highcharts

Database

Full data

Listing location
with ratings

Listing room types

Avalibilty shorter
than 20 days

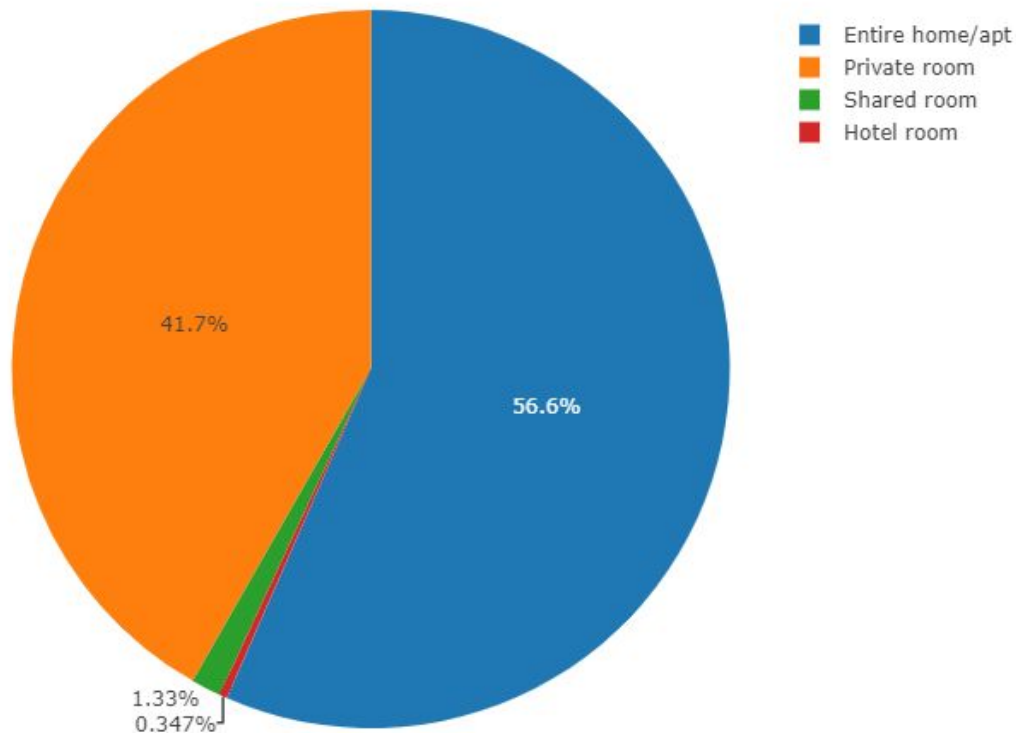Avaliability longer
than 300 days

Avaliablity
comparison

Listings average
price

Data Analysis

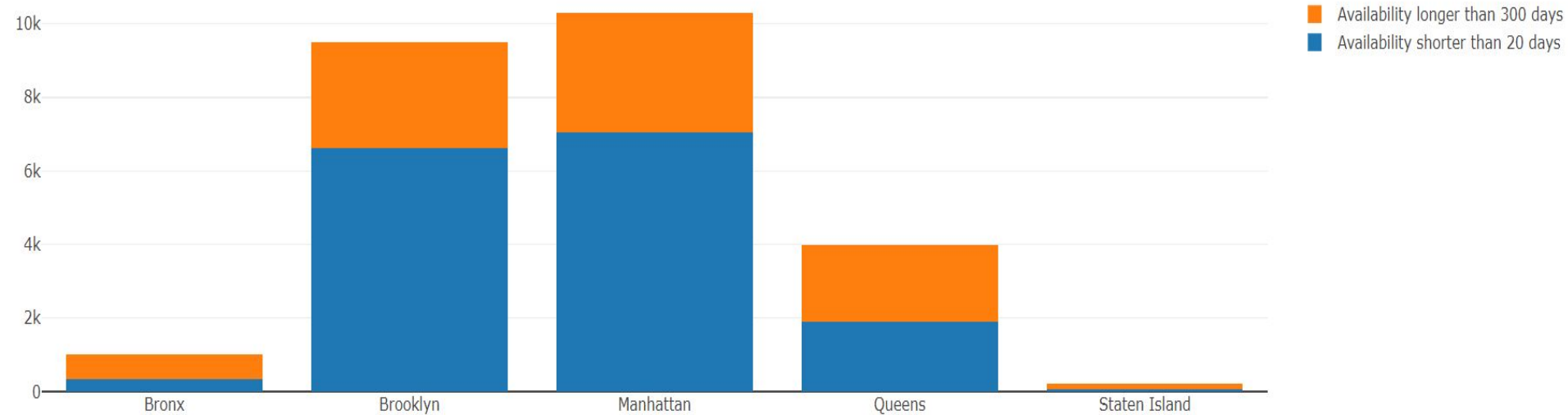There are 43,566 listings in New York City. They have been divided into five broad:

- Entire home/apt
- Private room
- Shared room
- Hotel room

Legend:
- Entire home/apt
- Private room
- Shared room
- Hotel room
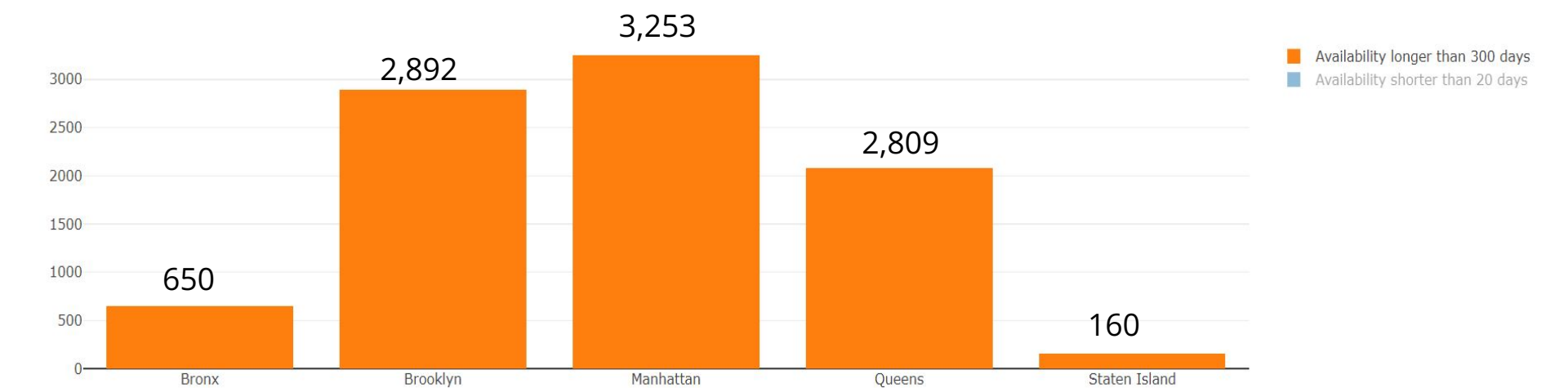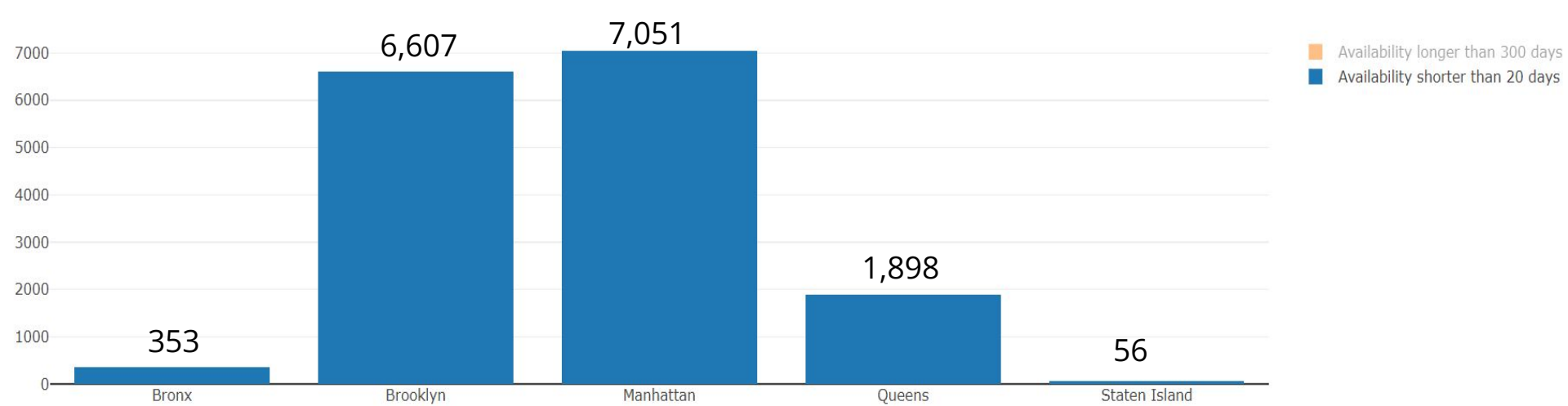
Pie chart values: 56.6%, 41.7%, 1.33%, 0.347%

Popularity Analysis:

Hot Spot - if availability of the listing property is shorter than 20 days

Cold Spot - if availability of the listing property is longer than 300 days

Top chart — Availability shorter than 20 days

Borough	Value
Bronx	353
Brooklyn	6,607
Manhattan	7,051
Queens	1,898
Staten Island	56

Legend: Availability longer than 300 days; Availability shorter than 20 days

Bottom chart — Availability longer than 300 days

Borough	Value
Bronx	650
Brooklyn	2,892
Manhattan	3,253
Queens	2,809
Staten Island	160

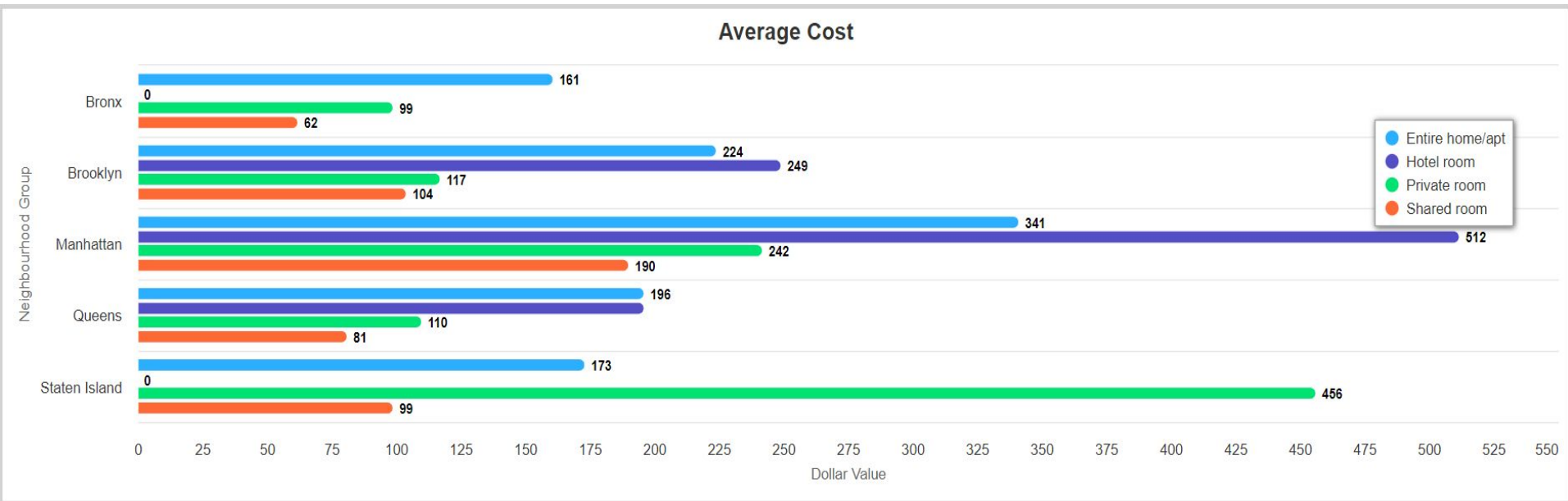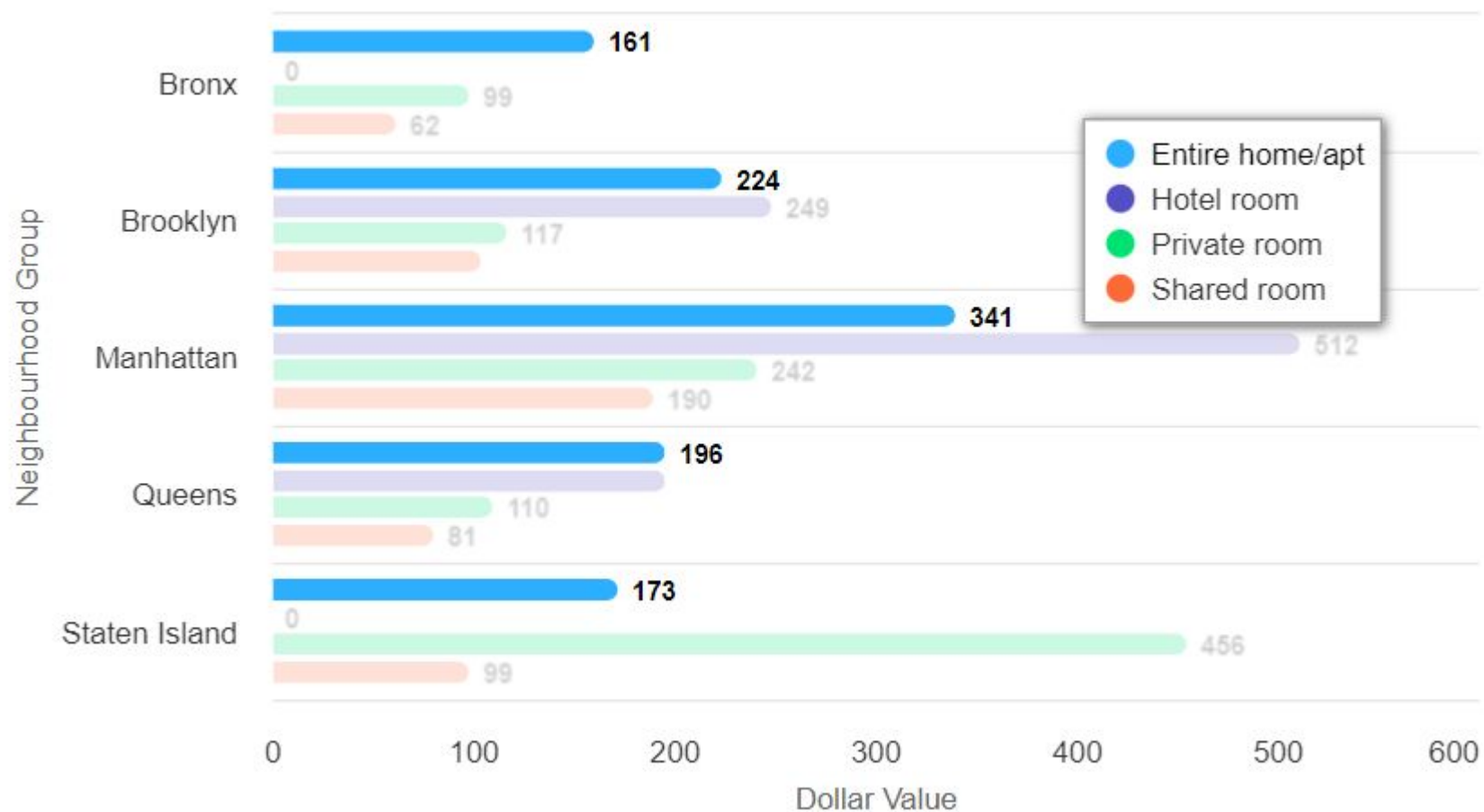Legend: Availability longer than 300 days; Availability shorter than 20 days

For price analysis:

We have the data grouped by different **neighbourhood group** among the different **room types**.

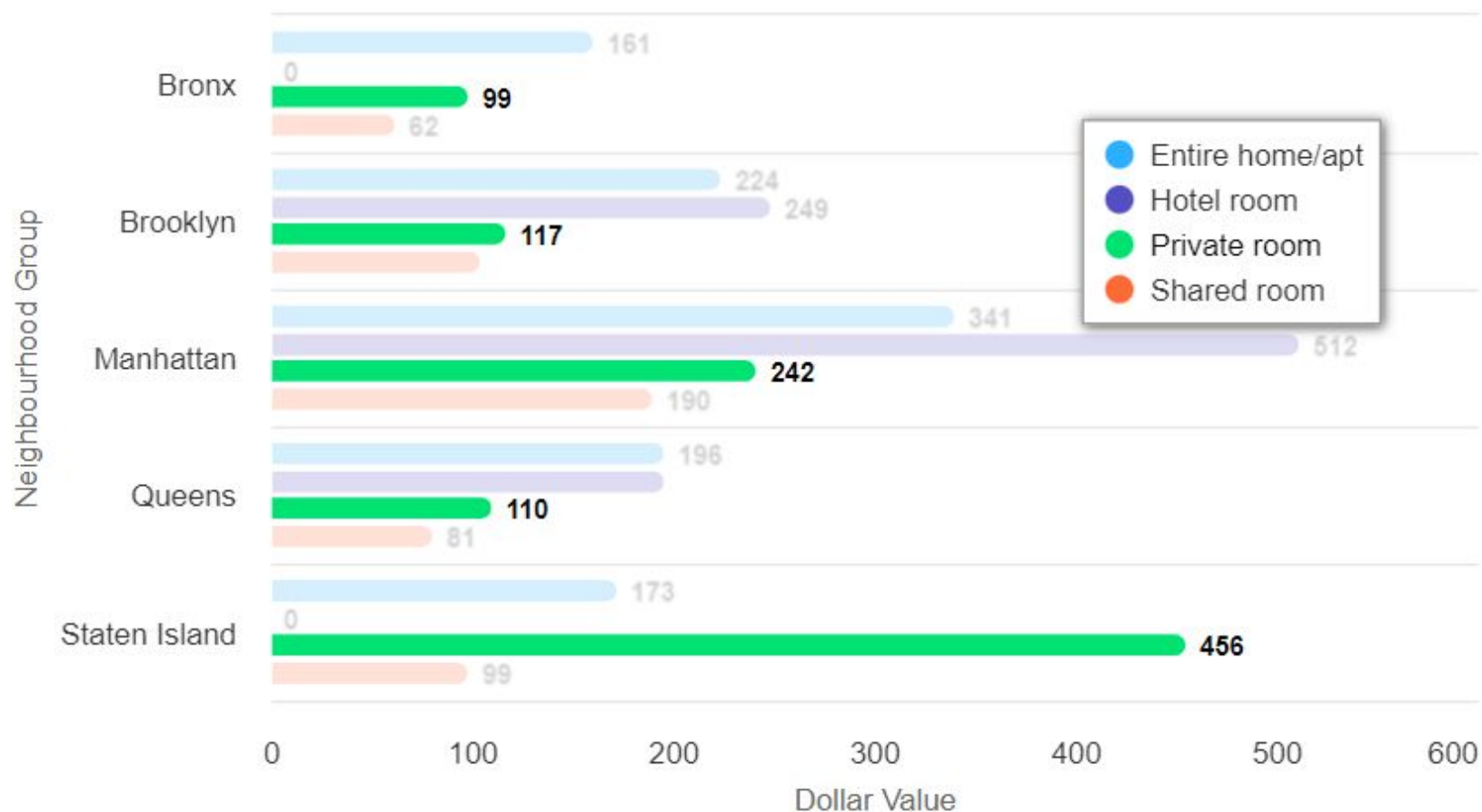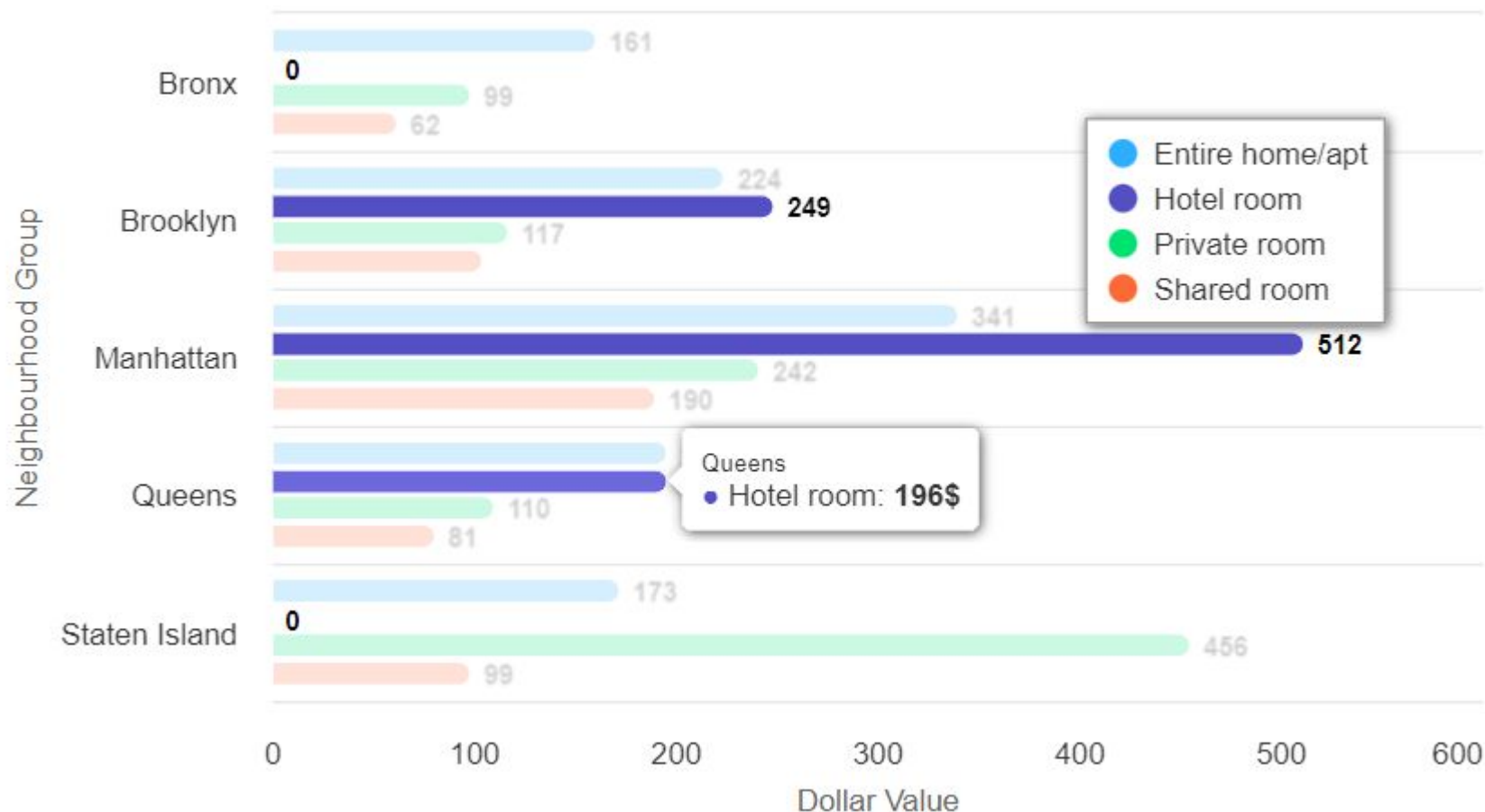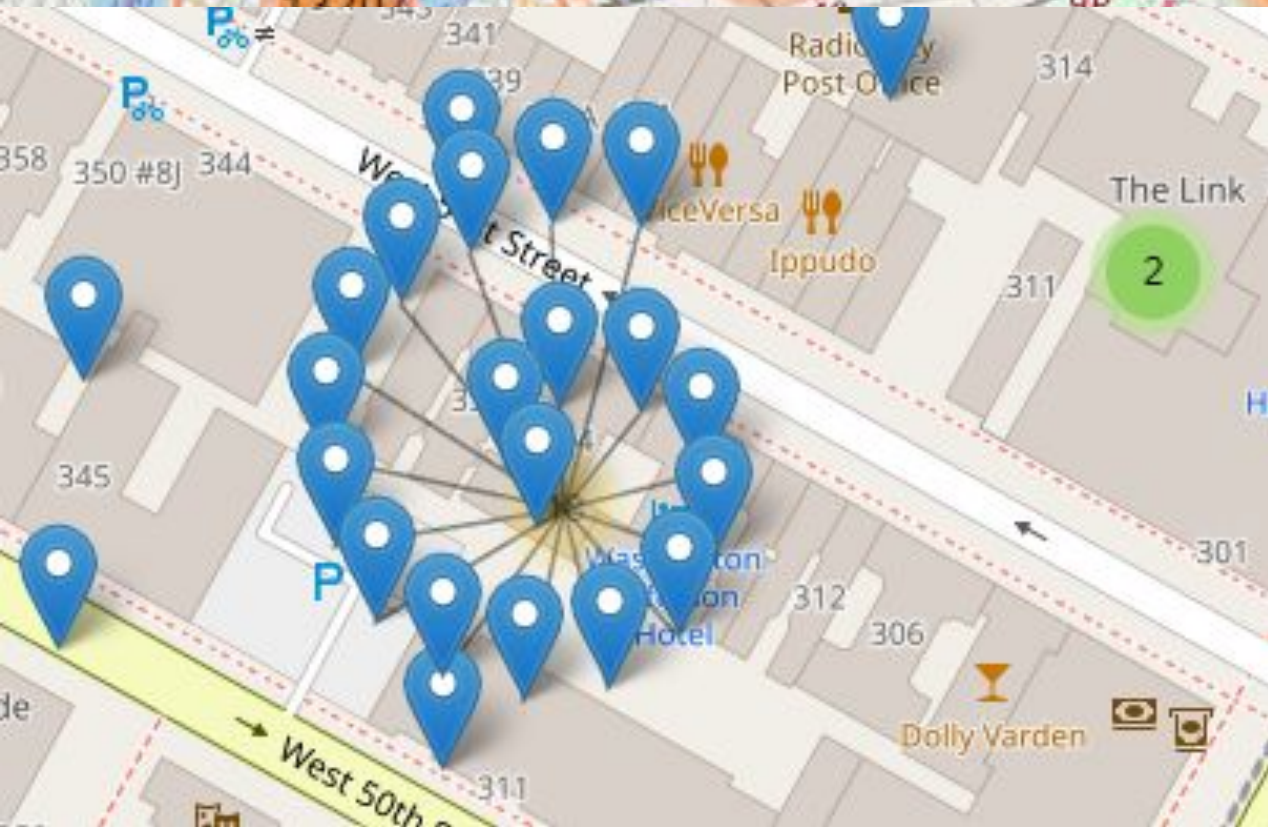The detail graphs show in the behind slides.

Average Cost

Average Cost

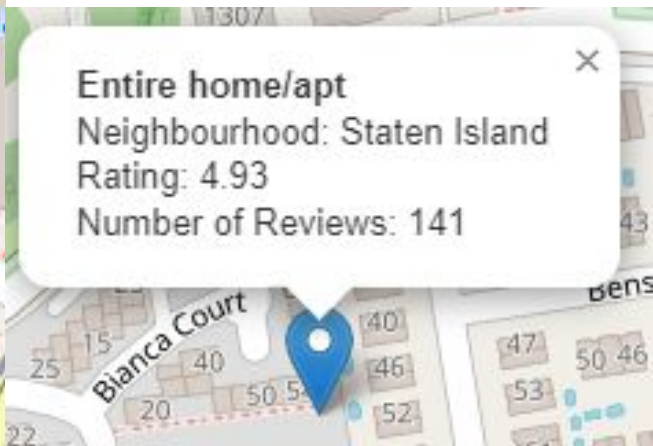| Neighbourhood Group	Entire home/apt	Hotel room	Private room	Shared room
Bronx	161	0	99	62
Brooklyn	224	249	117	
Manhattan	341	512	242	190
Queens	196		110	81
Staten Island	173	0	456	99

Average Cost

Average Cost

Map-demo

Conclusion

- Visitors and hosts can gain valuable insights into the distribution and availability of listings that will help them make better decisions
- Most listed room types are entire room/apt and private room
- People prefer to rent an airbnb in Brooklyn and Manhattan
- Manhattan has the highest average daily cost, and compared to that Bronx seems to have the lowest
- Cost is not the main factor to determine the airbnb hot spot

for listening