

# DATA ANALYSIS PORTFOLIO

**SAMBO MOPELOLA**



# TABLE OF CONTENTS

• Table Of Contents	1
• New York Airbnb Project Description	2
• Data Design	3
• Cleaning the Data	3
• Visualizing the Data	3
• Findings	4-5
• Analysis of the Business problem	6
• Conclusion	7



**EntryLevel**



# New York City Airbnb Project Description

**The Description of this project is based on New York Airbnb. Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world. The business problem is to learn the difference between the hosts and areas, which host is the busiest and differences among the different areas and what could be the reason for it.**

# Data Design

## **Cleaning the data:**

**After sourcing the datasets of the New York Airbnb which includes hosts host names, Neighborhood group, reviews, Room types etc. Having the data sets, the data sets were downloaded and was imported on Excel. The data sets were cleaned, by removing blank spaces, removing duplicates and making sure the Header and the text are consistence and some 'Names' and host 'names' were consistent as others.**

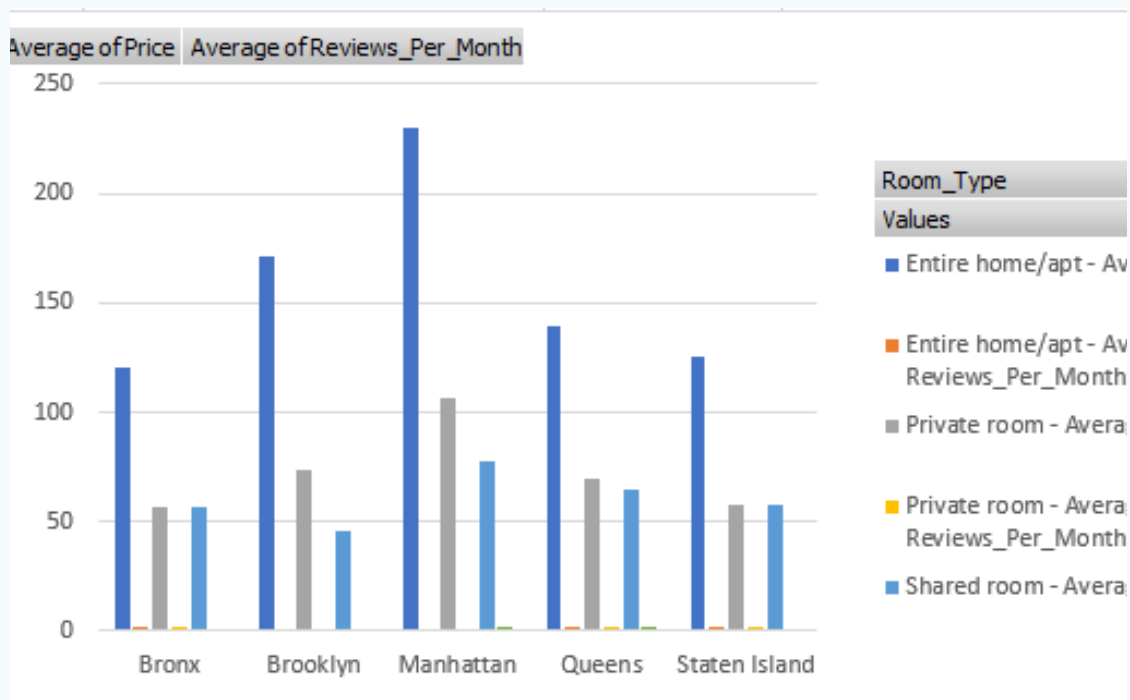
## **Visualizing the data:**

**Pivot tables and charts were created on Excel and visualization was also created on Tableau.**

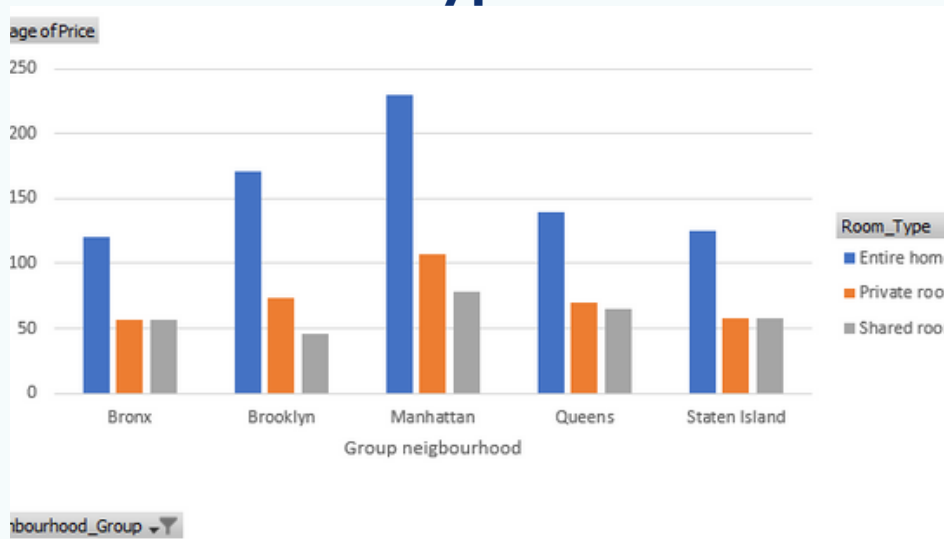
# Findings: 1

## Findings From Pivot table and Charts

**Figure 1: Neighborhood group by Avg price per month.**



**Figure 2: Neighborhood group by Avg price and room type.**



# Findings: 2

## Visualization From Tableau

Figure 3: Availability\_365

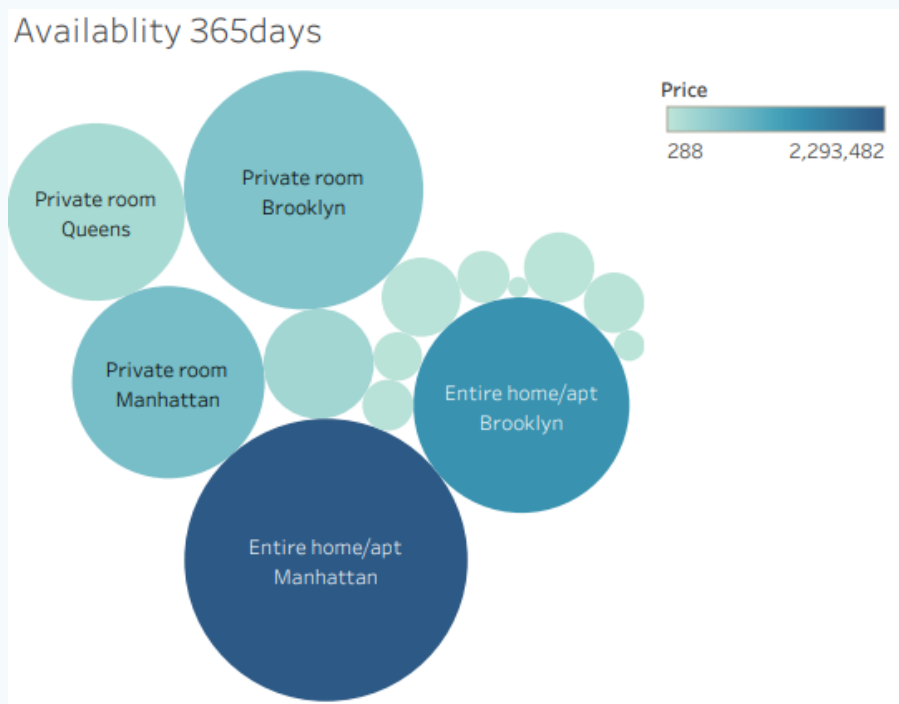
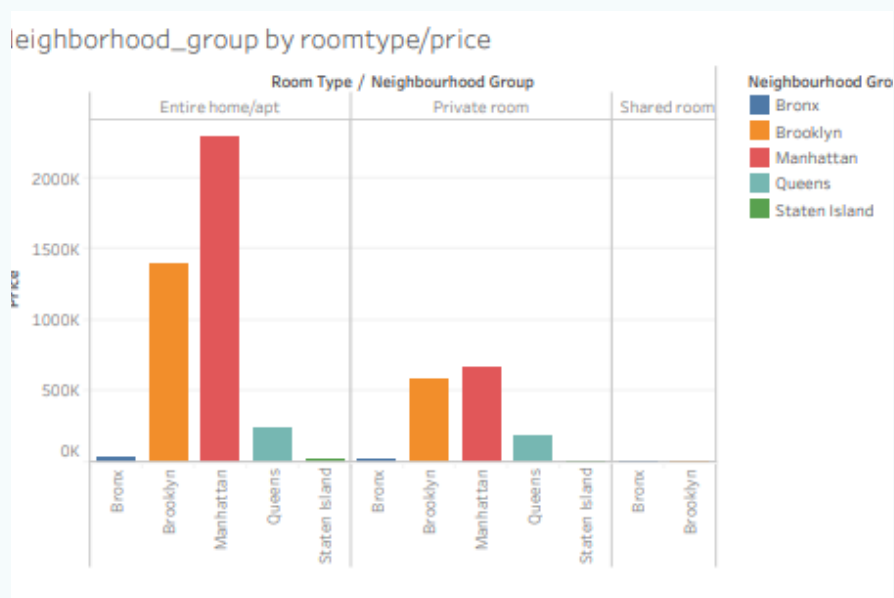


Figure 4: Availability\_365



# Analysis

**Based on the Findings from Above, we find out that of all the neighborhood groups in the New York Airbnb, Manhattan has the highest numbers of Host, with Higher price, Higher numbers of ratings and the highest Availability\_365days.**

**From our root cause analysis, which host is the busiest differences among the different areas and what could be the reason for it, we discover that all hosts in Manhattan are the busiest. The reason for this is being that, of all the five Boroughs in New York, being the neighborhood groups in the Data sets used for this report, despite Manhattan being the smallest geographically it is the most densely populated. Therefore, this report finds that the popularity share is weighted towards Hosts in Manhattan due to its default factor being the most populated city in New York City.**

# Conclusion

**This report concludes that Hosts in Manhattan are the busiest. This report also recommends that there should be more host in the neighborhood group 'Manhattan'. Also, other hosts in the other neighborhood group should do a general survey to find out why and how Manhattan has the busiest hosts, emulates from their services and rebrand on there's to have higher chances of having more guests with higher number of ratings.**