



Data Storytelling Case Study Presentation – I

Data Analysis Team

By

Mr. Justin Dias

Mr. Mandar Dhumal

Cohort - DS-37

Agenda

- Objective
- Background
- Key findings
- Recommendations
- Appendix:
 - Data sources
 - Data methodology
 - Data assumptions

Objective

- Acquisition of the best properties in New York
- Optimal price negotiation, and negotiating the services
- Optimising the order of property listings in certain neighbourhoods and cities

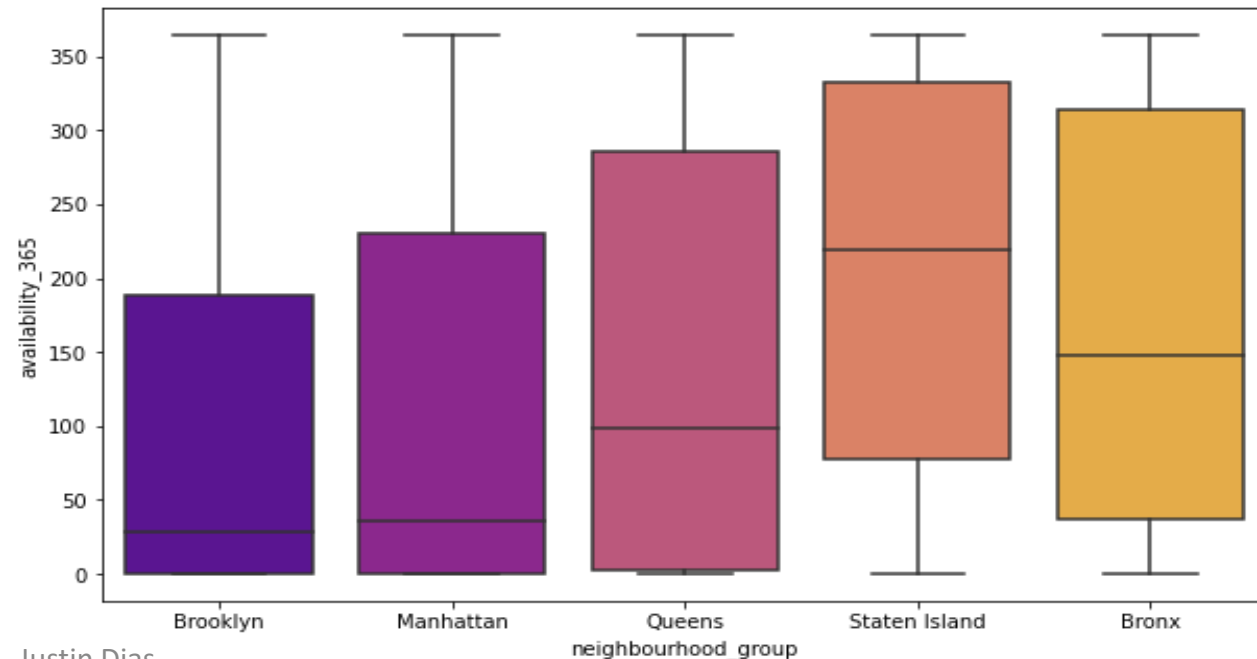
Background

- For the past few months, Airbnb has seen a major decline in revenue. due to Covid pandemic
- Recently most of countries have lifted Covid restrictions and people have started to travel more
- Airbnb wants to make sure that it is fully prepared for this change.

Staten Island has the highest availability on days and Brooklyn has rooms with 0 Price



- Staten Island has a very high median room availability appr. of 230 days. Hence business is low
- Brooklyn and Manhattan have less than 30 days of availability. Hence business is more
- Data has few properties with Price as 0 which needs to be checked/corrected at the data collection stage

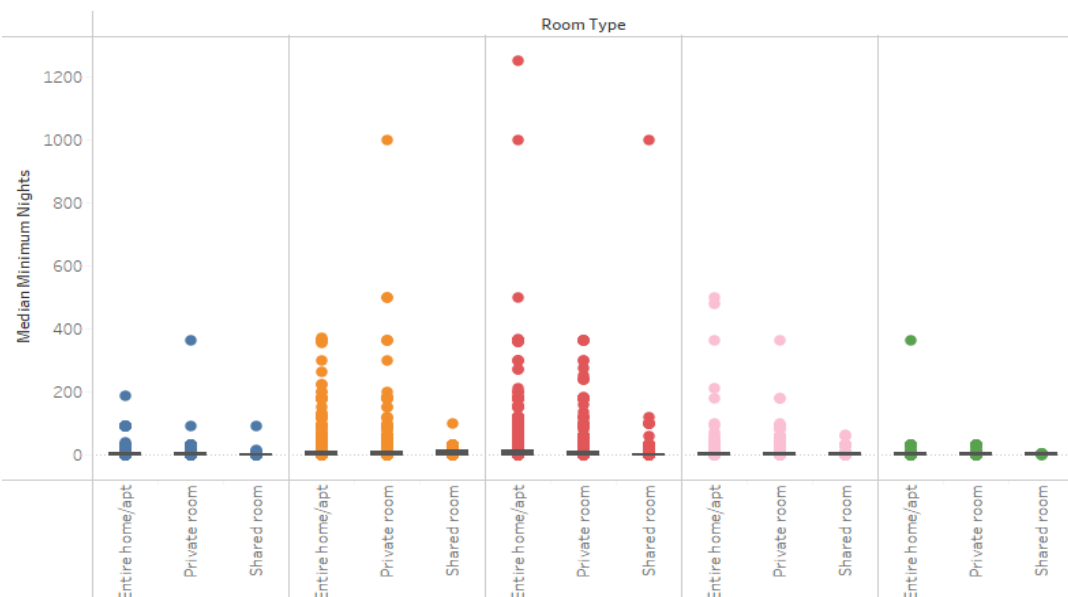


Properties with Price = 0			
Neighbourhoo..	Neighbourhood	Id	
Bronx	East Morrisania	20333471	0
		18750597	0
Brooklyn	Bedford-Stuyvesant	20639628	0
		20639792	0
		20639914	0
		20523843	0
	Bushwick	21291569	0
		21304320	0
		20608117	0
	Greenpoint	20624541	0
	Williamsburg	20933849	0
	Murray Hill		

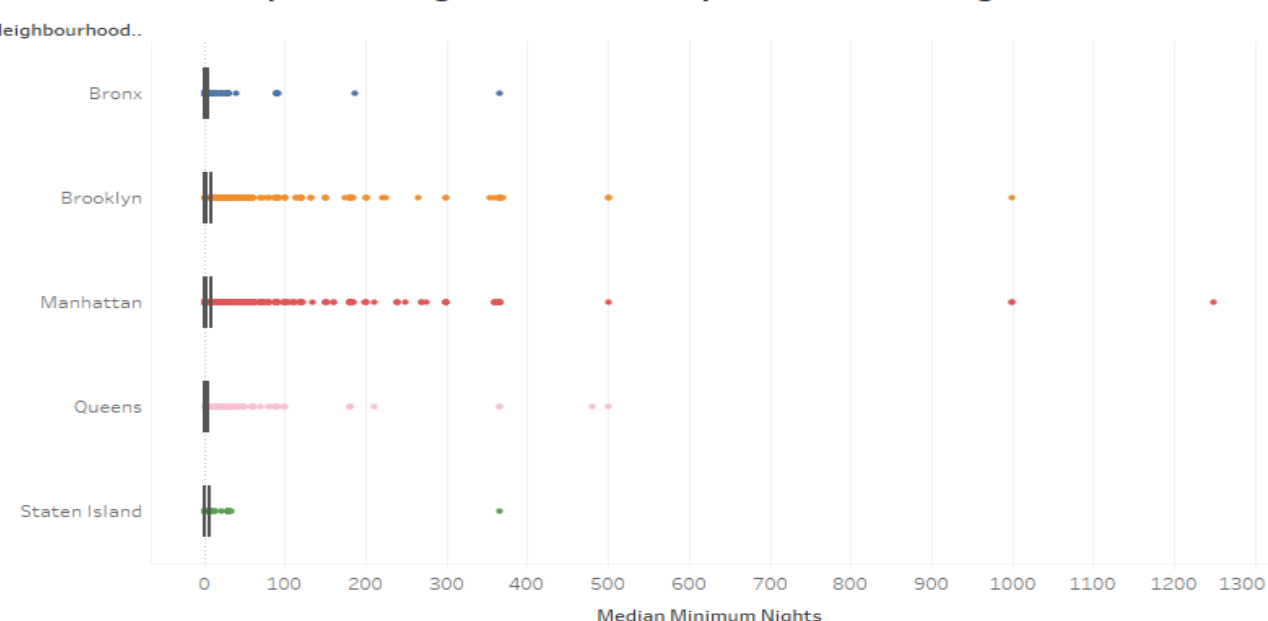
Outliers observed in the data

- We found that there were outliers in the data which may affect the bookings of some properties
- There are many properties with minimum nights of more than the median night of around 6. More minimum nights can hamper the booking of that property
- Majority of such outliers can be observed in Entire Home/Apt

Boxplot of Neighbourhood Group and Minimum Nights



Boxplot of Neighbourhood Group and Minimum Nights



Recommendations

- Concentrate on reducing the availability days in Staten Island and the Bronx boroughs by increasing the listings
- Data collection stage should check for outliers as some of them are beyond reason/logic (like price = 0 and high minimum nights for booking)
- Efforts to get clean data (as null values and outliers were observed) should be made. This will help in having insightful understanding of the customer preferences

Appendix – Data Sources

- We were provided with a New York Airbnbs Dataset
 - The data dictionary consists

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Appendix – Data Methodology

- We have done an in-depth analysis of the data sets. The process included:
 - Cleaning the data by removing the null values of some columns with median values
 - Outliers were not dealt with as that did not affect our analysis (as during visualizations we have taken median where outliers were found)
 - The cleaning was conducted in Python and the visualisations were done in Tableau

Appendix – Assumptions

- There were certain gaps in the data due to data unavailability hence, we assumed:
 - That number of reviews shows the preference of the customers
 - Not all customers send reviews
 - Sum of Price is assumed to the revenue generated for the getting valuable insights

Thank You