

PROBLEM STATEMENT

Pillow Palooza is a start-up focused on disrupting the traditional real estate market by offering short-term rental options for travellers. The company has seen a significant increase in demand for short-term rentals in New York City but wants to better understand the market to optimize their business strategy.

Pillow Palooza has collected data from various sources on Airbnb listings in New York City to gain insight into the market. The company now plans to analyse this data using its internal teams, to provide recommendations on how to maximize revenue and occupancy rates for their short-term rental properties.

To accomplish this, the Head of Data has requested that I should build a project that collects Airbnb listing data from multiple sources, clean, and analyse it. The analysis will aim to uncover key trends in popular neighbourhoods, rental prices, property types, length of stay, and demand over time.

The insights gained from the analysis had the potential to provide valuable information to the start-up on how to make informed business decisions. For example, Pillow Palooza could use the insights to identify which neighbourhoods to invest in, which property types to focus on, and how to price their rentals to remain competitive in the market.

By providing these insights, the project aimed to help Pillow Palooza make strategic decisions that could lead to growth and success in the short-term rental market in New York City.

DATA DESCRIPTION

The data contains three different tables, Airbnb prices, Reviews and Room types for several listings within New York city.

The first task was to validate the data and carry out data wrangling and cleaning using python.

Airbnb_price.csv

Listing_id: Unique identifier of listing

Price: Nightly listing price in USD

Neighbourhood name : Name of borough and neighbourhood where listing is located

Airbnb_room_type.xlsx

This is an Excel file containing data on Airbnb listing descriptions and room types.

Listing_id: Unique identifier of listing

Description: Listing description

Room type: Airbnb has three types of rooms: shared rooms, private rooms, and entire homes/apartments

Airbnb_last_review.tsv

This is a TSV file containing data on Airbnb host names and review dates.

Listing_id: Unique identifier of listing

Host name: Name of listing host

Last_review: Date when the listing was last reviewed

The clean data was then downloaded from python to excel and then into Tableau workbook to generate charts and visualizations. A total of five charts were created, grouped into a dashboard.

1. Listing map: A mapbox was used to show the number of listings per borough within New York City.

2. Length of stay per neighbourhood: A bar chart showing median minimum nights and median bookings per neighbourhood.

product category.

3. Revenue generated: A bar chart showing the total revenue generated and the total listing counts for each room type per borough.

indicating the states. It has a dropdown filter option for selecting either top or bottom five.

4. Most Booked Neighbourhoods: A horizontal bar chart was used to show the occupancy rate by using stacked bar to show the availability 365, booked days, and listings for each neighbourhood grouped by borough.

INSIGHTS

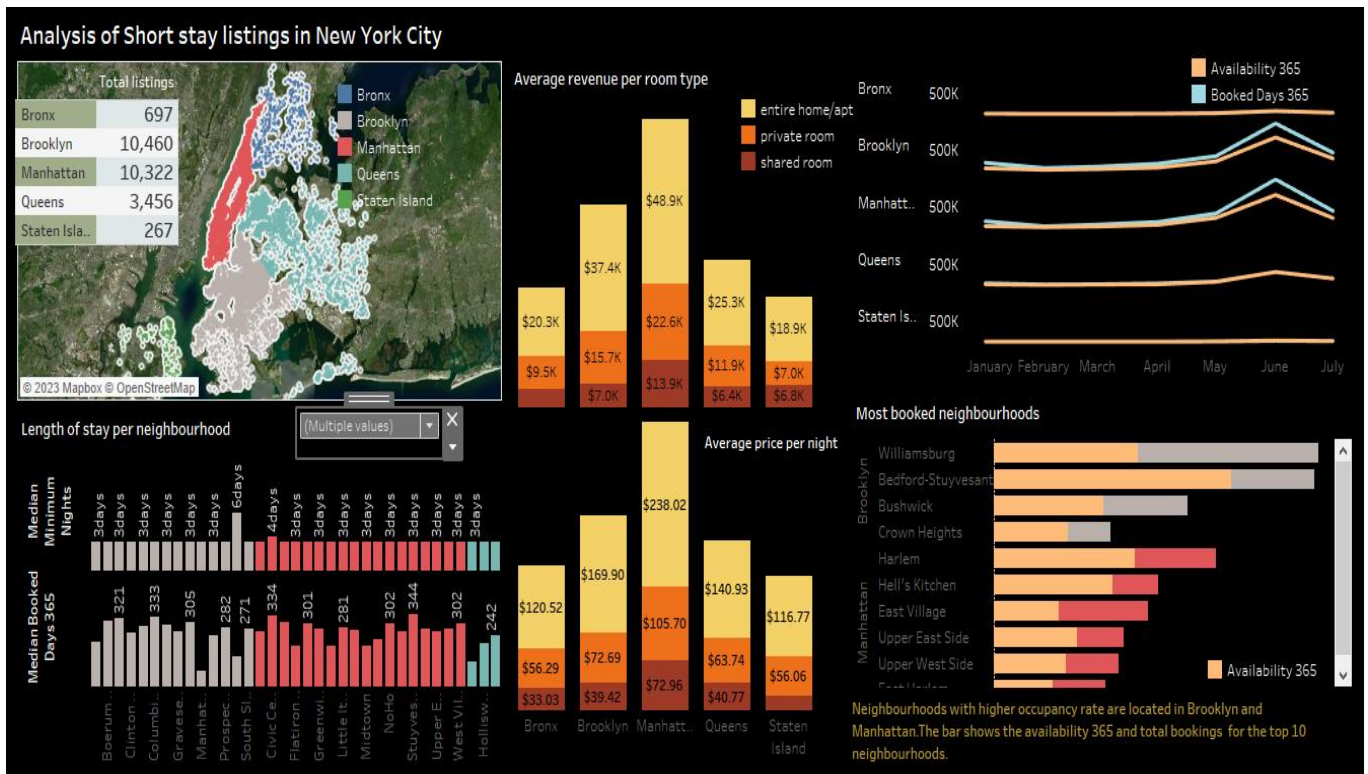
1. There is a higher concentration of listings in Brooklyn and Manhattan compared to other boroughs. Both boroughs have over 80% of all the listings in New York City (Brooklyn - 41.5%, Manhattan - 41%).

2. Entire room apartment in Manhattan have the highest overall revenue for all the short-term Airbnb listings in NYC. Also, for every Borough entire room recorded higher returns than other room types.

3. Entire room type in Manhattan is the most expensive room followed by Brooklyn, Queens, Bronx and then Staten Island.

4. Sea gate in Brooklyn and civic centre in Manhattan are the neighbourhoods with the most length of stay. Half of all bookings in sea gate neighbourhood stayed for at least 6 nights while civic centre was about 4 nights.

5. Williamsburg, Bedford-Stuyvesant and Harlem are the most booked neighbourhoods with the first two located in Brooklyn while the third in Manhattan.



RECOMMENDATIONS

1. Following the huge concentration of listings, bookings and revenue in Brooklyn and Manhattan, I would recommend that the business should focus growth on both boroughs.
2. The neighbourhoods with higher occupancy rate should also be considered, these are Williamsburg, Bedford-Stuyvesant, and Harlem all located in the two busiest boroughs. costs and grow profit ratio.
3. More investment should be channelled towards entire rooms apartment due to the high booking rate and higher revenue.

APPENDIX

```
import numpy as np
import pandas as pd
prices = pd.read_csv("data/airbnb_price.csv")
print(prices.head())
np.random.seed(0)
print(prices.sample(5))
print(prices.shape)
xls = pd.read_excel("data/airbnb_room_type.xlsx")
print(xls.head())
xls = pd.ExcelFile("data/airbnb_room_type.xlsx")

# Assuming the first sheet is the one we want
room_types = xls.parse(0)
reviews = pd.read_csv("data/airbnb_last_review.tsv", sep='\t')
print(reviews)
prices["price"] = prices["price"].str.replace(" dollars", " ")

prices["price"] = pd.to_numeric(prices["price"])
prices.describe()

#Subsetting prices for listings costing $0, free_listing
free_listing = prices["price"] == 0
max_listing = prices["price"] >= 7500

#Update prices by removing all free listings
```

```
prices = prices.loc[~free_listing]
prices = prices.loc[~max_listing]
#Average the price column in the prices dataframe
avg_price = round(prices["price"].mean(),2)
print(avg_price)
#Add a new column, price_per_month, to the prices DataFrame
prices["price_per_month"] = prices["price"]* 365/12
print(prices)
average_price_per_month = round(prices["price_per_month"].mean(),2)
print(average_price_per_month)
difference = round(average_price_per_month - 3100, 2)
print(difference)
#Change all values in the room_type column to lowercase.
room_types["room_type"] = room_types["room_type"].str.lower()
print(room_types)
#Convert the room_type column to a dtype.
room_types["room_type"] = room_types["room_type"].astype("category")
#Store the count of values for room_type as room_frequencies.
room_frequencies = room_types["room_type"].value_counts()
print(room_frequencies)
import datetime as dt
# Change the data type of the last_review column to datetime
reviews["last_review"] = pd.to_datetime(reviews["last_review"])
# Create first_reviewed, the earliest review date
first_reviewed = reviews["last_review"].dt.date.min()
# Create last_reviewed, the earliest review date
last_reviewed = reviews["last_review"].dt.date.max()
print(first_reviewed)
print(last_reviewed)
```

#Joining the DataFrames

Merge prices and room_types to create rooms_and_prices

```
rooms_and_prices = prices.merge(room_types, how="outer", on="listing_id")
```

Merge rooms_and_prices with the reviews DataFrame to create airbnb_merged

```
airbnb_merged = rooms_and_prices.merge(reviews, how="outer", on="listing_id")
```

#Create a new column in airbnb_merged called borough by using the str.partition() method on airbnb_merged["nbhood_full"] and indexing the first value using [0]

```
airbnb_merged['borough'] = airbnb_merged['nbhood_full'].str.partition(',')[0]
```