



Airbnb NYC Storytelling Assignment

PRESENTED BY:

Anand Pratap
Abhijit
Sukanya Sarkar





Agenda

- Problem Statement
- Background
- Data preparation methodology
- Key areas
- Annexture:
 - Data Source
 - Data Methodology

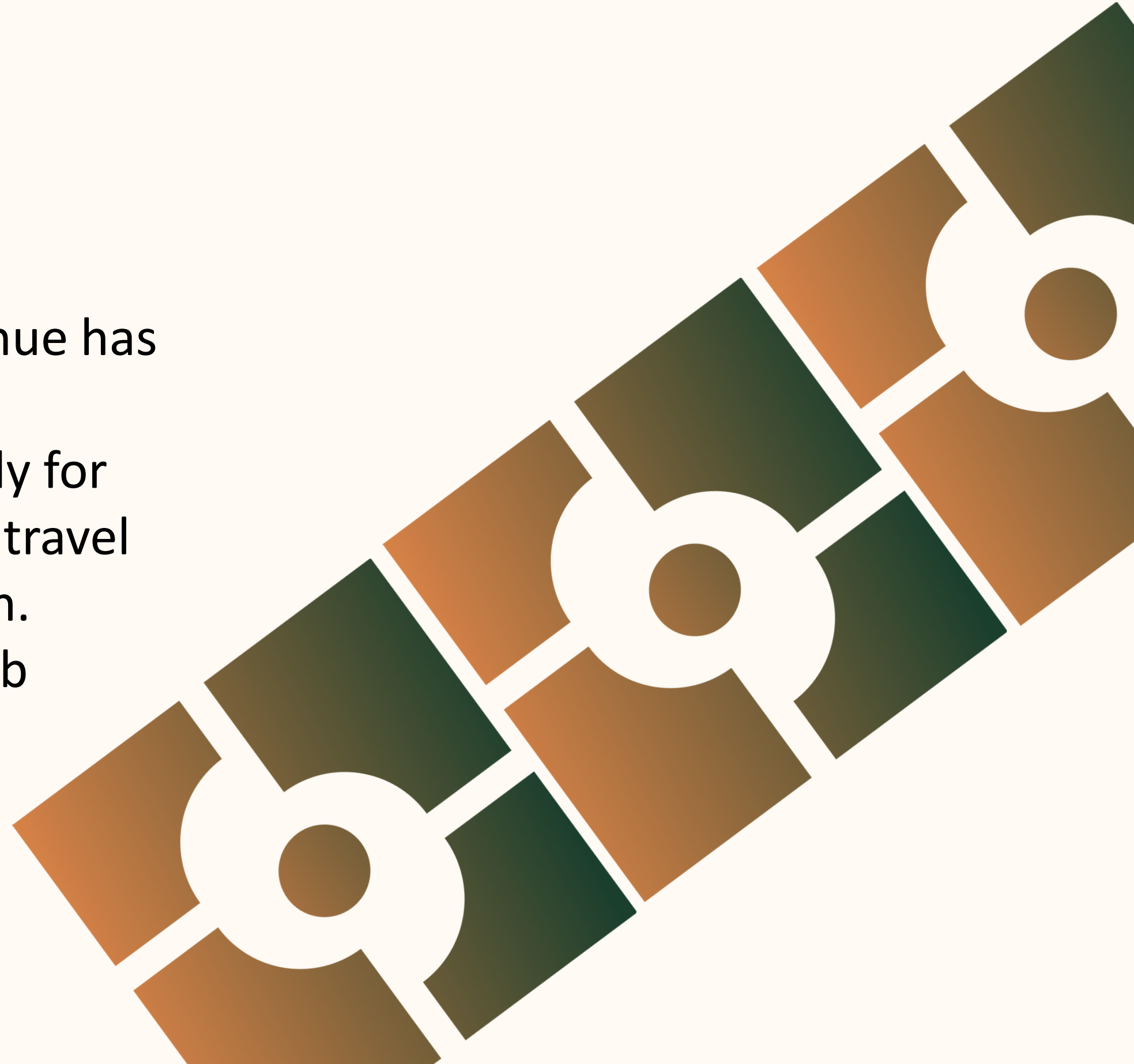
Problem Statement

- People can rent out their extra lodging by using the internet platform Airbnb.
- Airbnb lost a significant amount of money during the COVID-19 pandemic.
- Now that people are starting to travel again, Airbnb wants to revive its company and be prepared to serve clients.



Background

- Over the last few months, Airbnb's revenue has significantly decreased.
- Airbnb needs to make sure that it is ready for this shift, as more people are starting to travel and the constraints are starting to loosen.
- Thus, a dataset comprising several Airbnb listings in New York has been analyzed.



Data Preparation Methodology

- Data cleansing to find any duplicates, outliers, or missing values
- Removing unnecessary columns
- Created a separate Excel sheet with the cleaned data with the given data source.
- Using cleansed data files, creating visual charts through bivariate and univariate analysis and more to help comprehend the data



Data cleansing and upload

```
1 bnb_df = pd.read_csv('AB_NYC_2019.csv')
2 bnb_df.head()
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK!	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	2
4	5022	Entire Apt. Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	

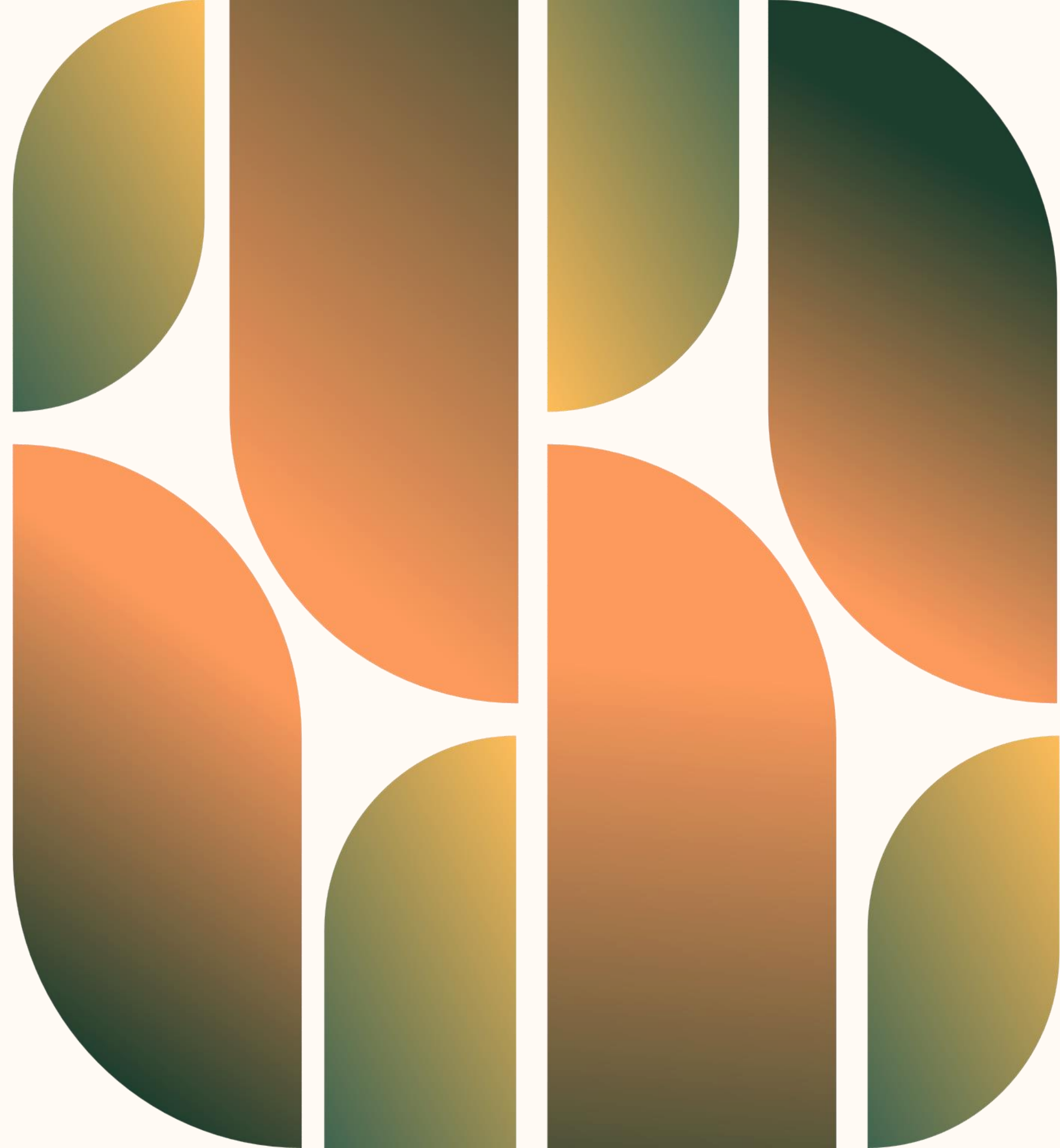
```
1 # Checking null values
2
3 bnb_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 20 columns):
#   Column                                          Non-Null Count  Dtype
---  -
0   id                                              48895 non-null  int64
1   name                                           48879 non-null  object
2   host_id                                         48895 non-null  int64
3   host_name                                       48874 non-null  object
4   neighbourhood_group                           48895 non-null  object
5   neighbourhood                                   48895 non-null  object
6   latitude                                       48895 non-null  float64
7   longitude                                       48895 non-null  float64
8   room_type                                       48895 non-null  object
9   price                                           48895 non-null  int64
10  minimum_nights                                48895 non-null  int64
11  number_of_reviews                             48895 non-null  int64
12  last_review                                    38843 non-null  object
13  reviews_per_month                             38843 non-null  float64
14  calculated_host_listings_count                48895 non-null  int64
15  availability_365                               48895 non-null  int64
16  availability_365_categories                   48895 non-null  object
17  minimum_night_categories                      48895 non-null  object
18  number_of_reviews_categories                 48895 non-null  object
19  price_categories                              48895 non-null  object
dtypes: float64(3), int64(7), object(10)
memory usage: 7.5+ MB
```



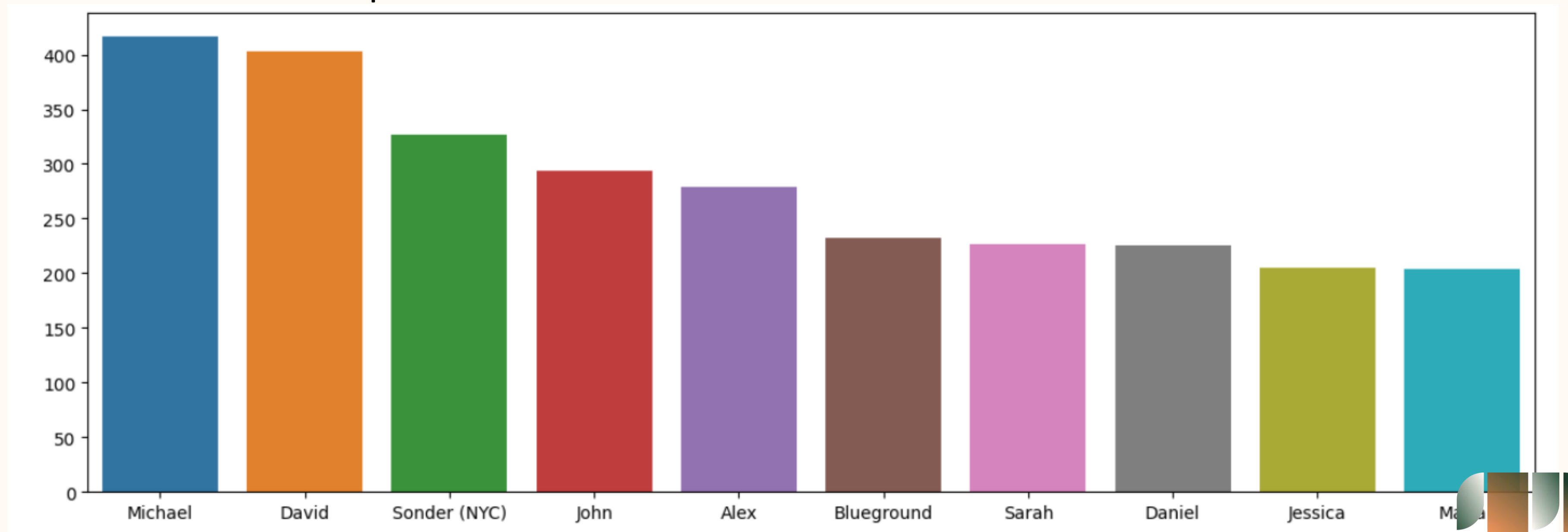
Key Areas

1. The top ten hosts
2. Well-liked neighborhood association
3. Client reservations with regard to minimum nights
4. The accessibility of Airbnb in various neighborhood groupings concerning Minimum number of nights
5. Availability versus Neighbourhood regarding minimum nights
6. The respective neighborhood groups' median prices
7. Favorite room type in relation to neighborhood group



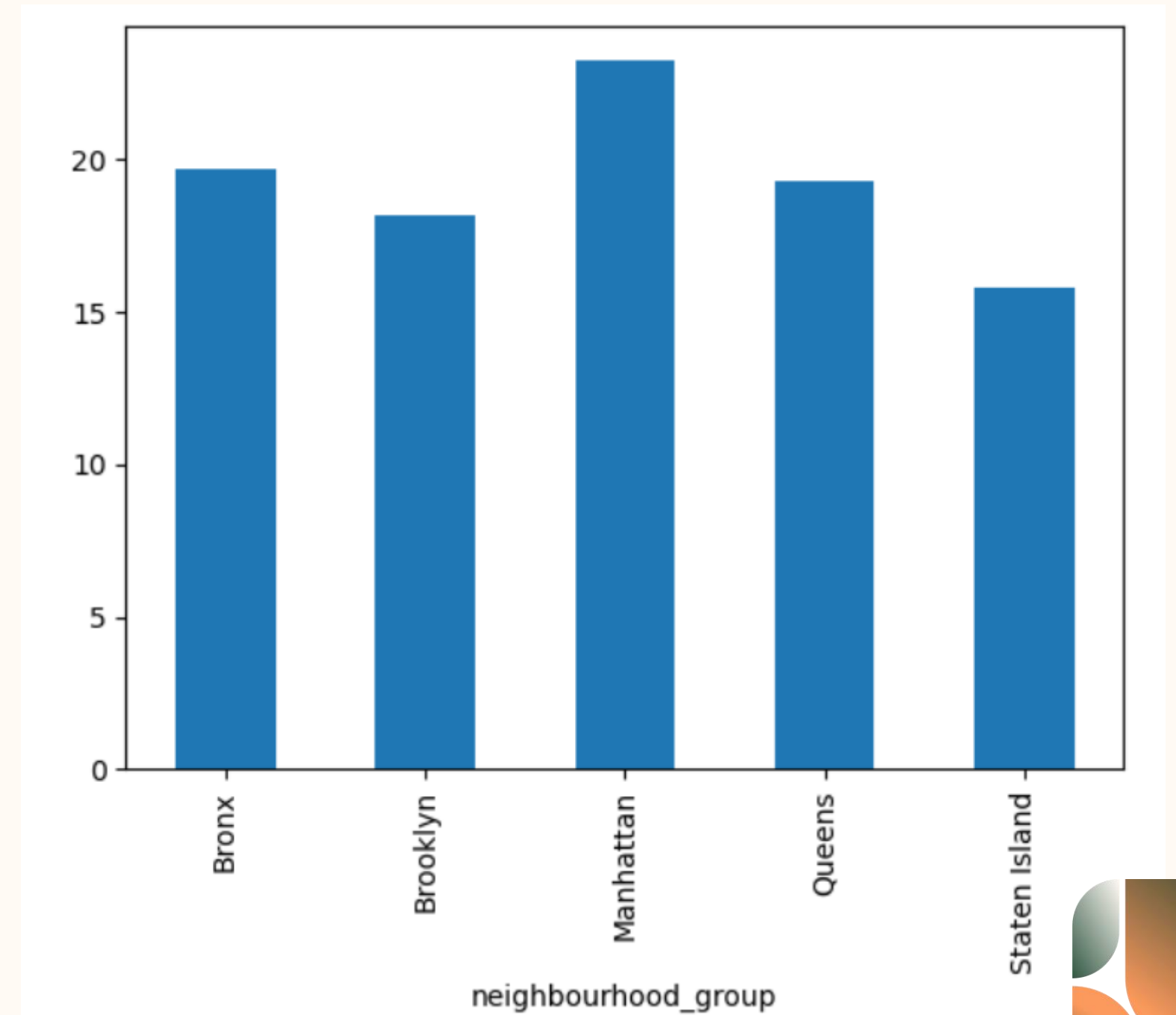
The Top 10 hosts

Among the 10 list of hosts for Airbnb seems Michael being the top most followed by David and to the end at 10 position its Maria



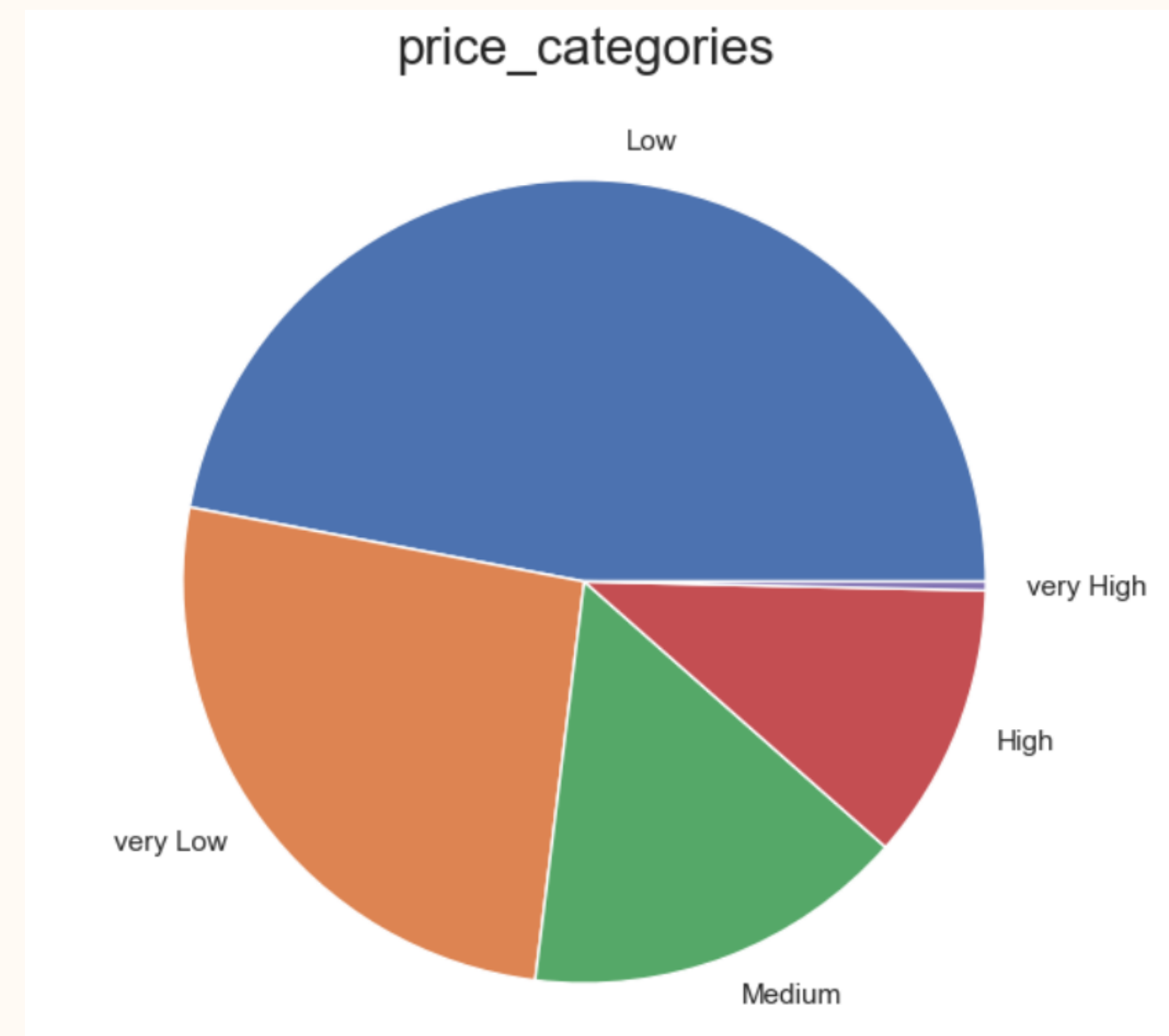
Well-liked neighborhood

- a) In New York, the largest percentage of Airbnb listings are found in Manhattan and Brooklyn.
- b) The island with the fewest listings is Staten Island.
- c) Queens is the third most popular Airbnb location, with a third preference in Bronx



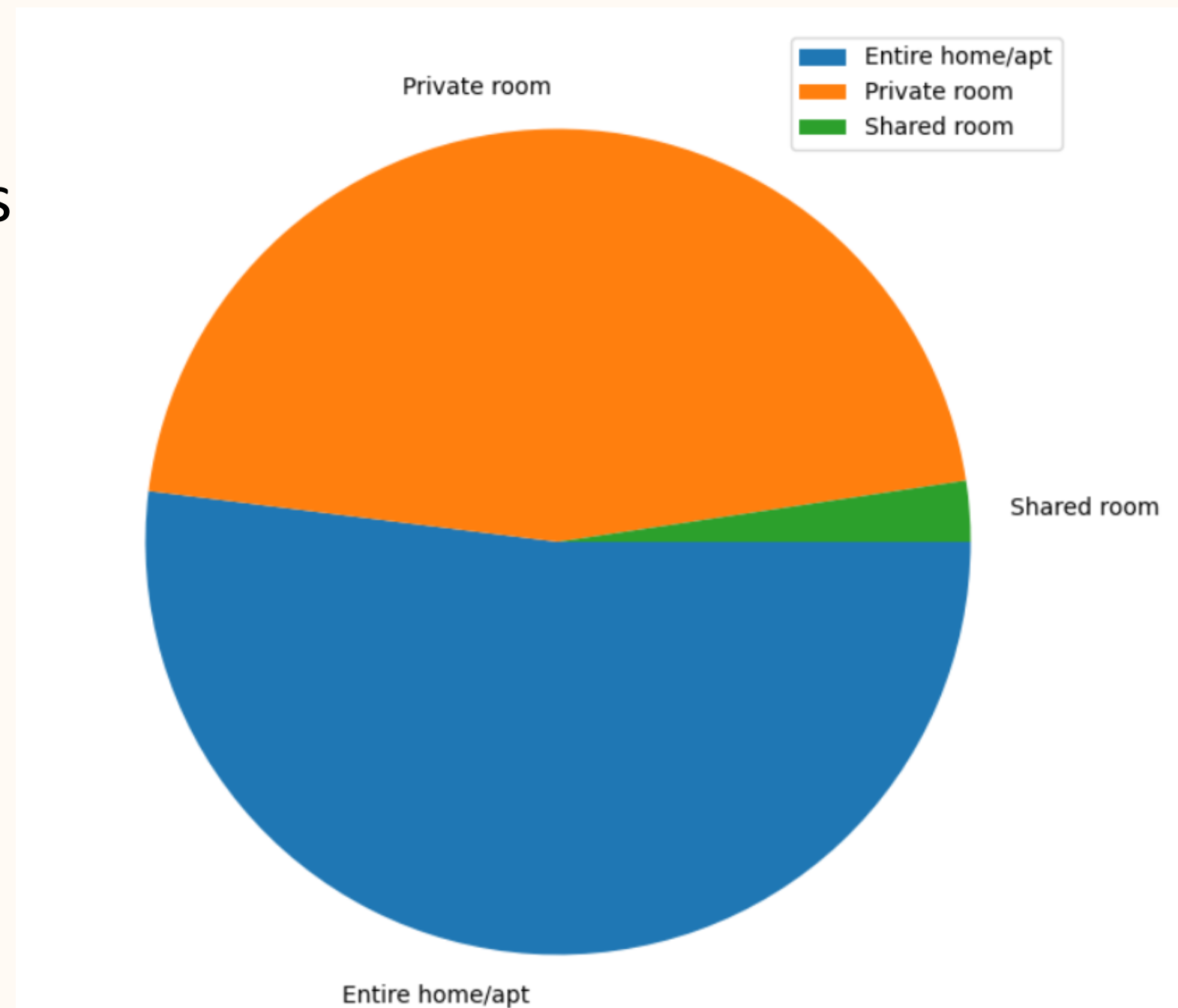
Median price of the different Neighborhood Groups

- a) According to the data, Manhattan's median price is greater than that of the other neighborhood groups.
- b) The median price is highest in Brooklyn.
- c) The least is the Bronx.



Preferred room type w.r.t neighborhood group

Preferred room types and their categories defines the status of bookings done



Annexure: Data Sources

- **Dataset overview:** Index(['id', 'name', 'host_id', 'host_name', 'neighbourhood_group', 'neighbourhood', 'latitude', 'longitude', 'room_type', 'price', 'minimum_nights', 'number_of_reviews', 'last_review', 'reviews_per_month', 'calculated_host_listings_count', 'availability_365', 'availability_365_categories', 'minimum_night_categories', 'number_of_reviews_categories', 'price_categories'], dtype='object')
- **Column:** 'id', 'name', 'host_name', 'neighbourhood_group', 'neighbourhood', 'room_type', 'availability_365_categories', 'minimum_night_categories', 'number_of_reviews_categories', 'price_categories'], dtype='object'

	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
0	149	1	9	0.21	6	365
1	225	1	45	0.38	2	355
2	150	3	0	NaN	1	365
3	89	1	270	4.64	1	194
4	80	10	9	0.10	1	0

Thank you

