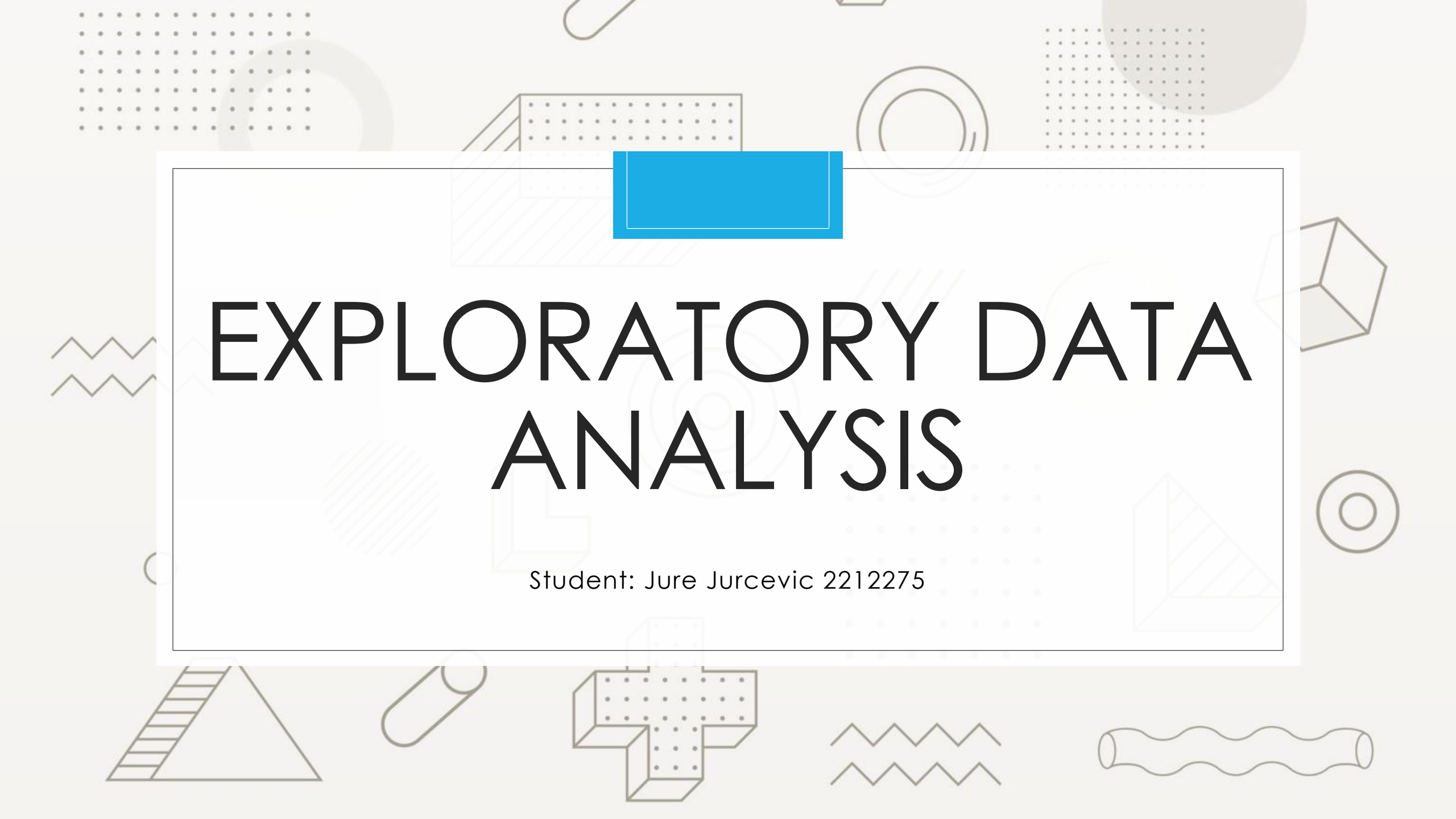


The background is a light gray with various faint, stylized geometric shapes and patterns. These include a grid of dots in the top left, a large circle in the top right, a paperclip in the top center, a 3D cube in the middle right, a triangle with a staircase pattern in the bottom left, a 3D cross with dots in the bottom center, and a wavy line in the bottom right. A solid blue rectangle is positioned at the top center, partially overlapping the white title box.

EXPLORATORY DATA ANALYSIS

Student: Jure Jurcevic 2212275

AGENDA



Data Overview



Univariate Analysis



Multivariate Analysis



Data Quality & Cleaning



Executive Summary & Recommendations

DATASET DESCRIPTION

- Source & Scope

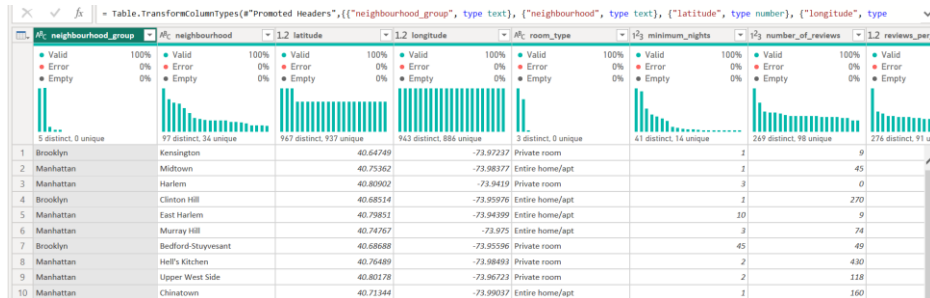
For this project I am analysing Airbnb's New York City (NYC) data of 2019. Main objective is to find out the key metrics that influence the listing of properties on the platform. For this task I will explore and visualise the dataset using basic exploratory data analysis (EDA) techniques [1] [2].

- 49 000 observations with various types of field in dataset

- List of field:

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_listing_count, availability_365

Data Preparation and Quality Checks



- The dataset (AB_NYC_2019.csv) was first imported into Power BI Power Query for inspection.
- Column profiling revealed that all key attributes (borough, room type, latitude, longitude, price, etc.) were 100 % valid, with no empty or error values.
- This validation confirmed data completeness and helped ensure that later transformations (e.g., cleaning duplicates, removing outliers) were based on reliable inputs.
- Visual indicators (green bars) in Power BI confirm no missing or corrupted entries across columns [3].

KEY MATRICS

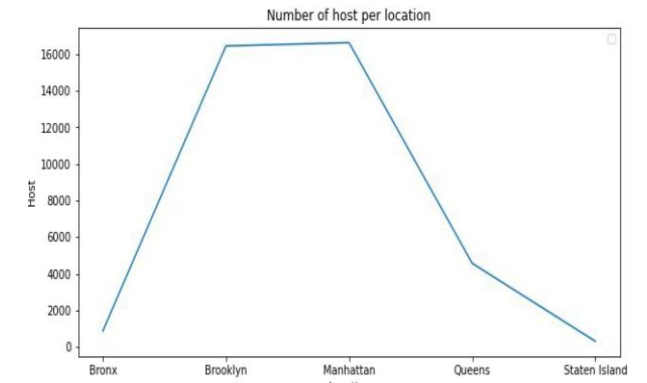
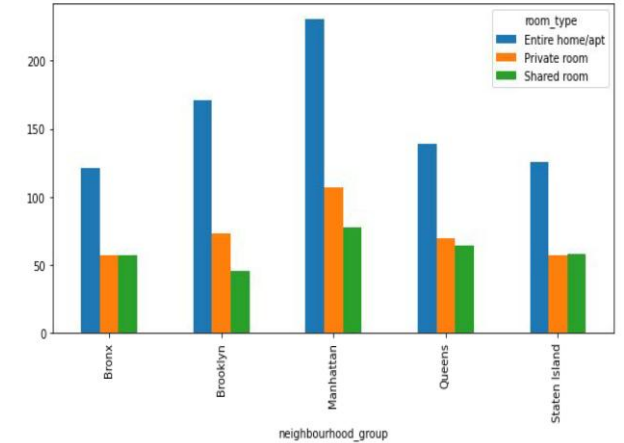
Category	Metric	Purpose
Supply	Number of listings, active hosts, availability	Measures market size & saturation
Demand	Number of reviews, occupancy, min nights	Indicates popularity & guest behavior
Pricing	Average price per borough/room type, turnover	Shows profitability & pricing power
Product	Room type distribution	Reveals product mix & customer preference
Location	Top neighbourhoods, borough analysis	Maps spatial concertation of listings
Engagement	Text keywords, review counts	Reflects marketing trends & host strategies [4]

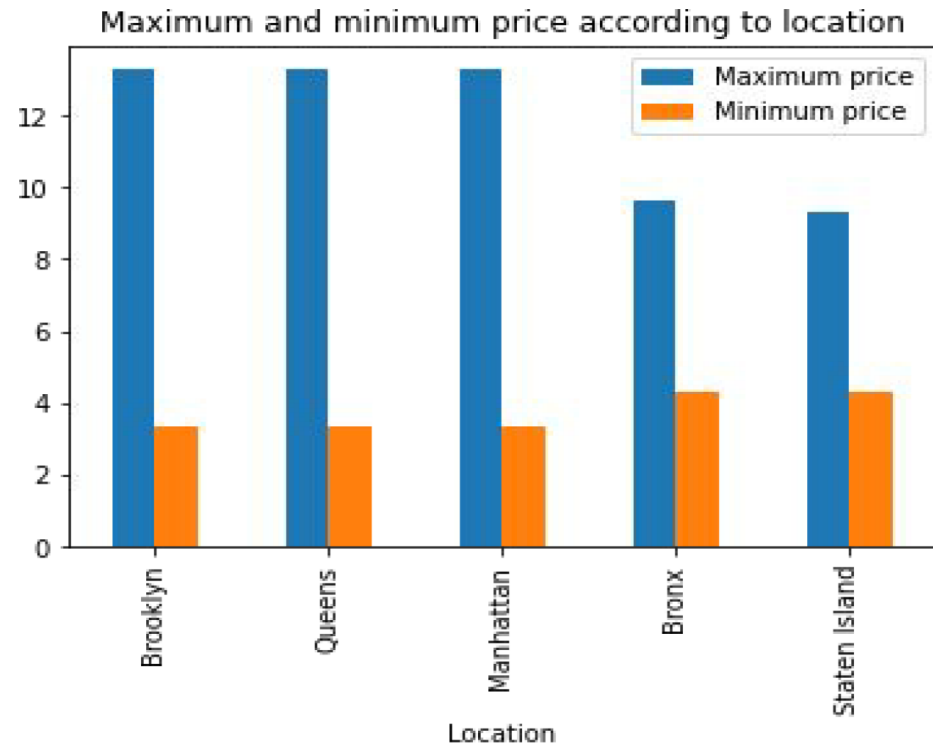
Airbnb Pricing and Host Distribution Across New York City

Manhattan and Brooklyn dominate both pricing and hosting activity.

Shared and private rooms remain most affordable in outer boroughs.

Queens shows mid-range pricing and average host presence.

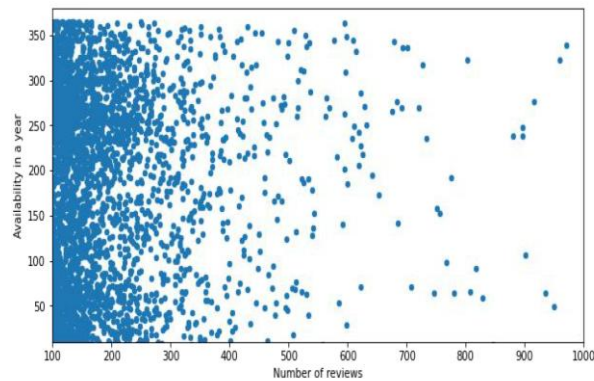




Highest and Lowest Airbnb Prices by Location

- Brooklyn, Queens, and Manhattan show almost the same maximum price levels, indicating strong demand and premium listings in these boroughs.
- Minimum prices are also similar across these three areas, showing a consistent lower threshold for budget stays.
- Log transformation was applied to visualize wide differences between minimum and maximum price values on the same scale.
- Overall, price extremes are concentrated in major boroughs, reflecting high pricing variability driven by property type and location.

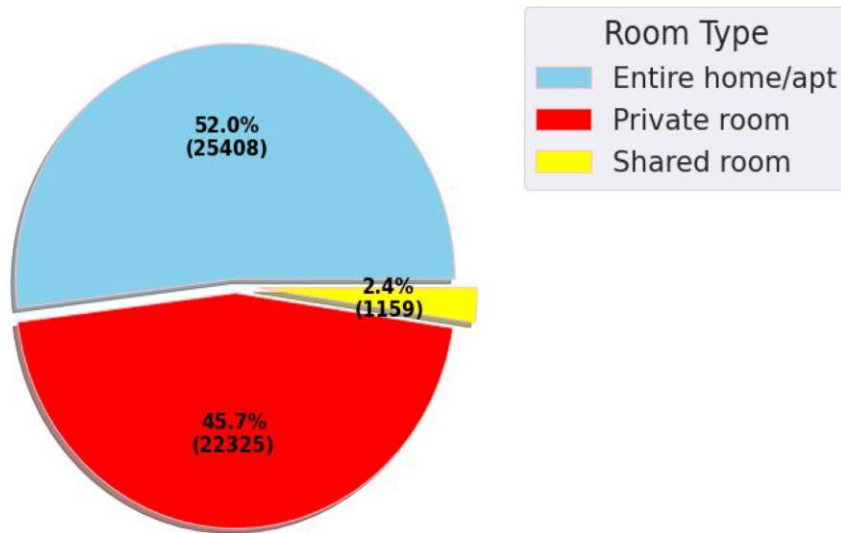
	host_id	host_name	number_of_reviews	availability_365
12	22959695	Gurpreet Singh	1157	0.0
41	99392252	Michael	732	0.0
47	121391142	Deloris	693	0.0
123	792159	Wanda	480	0.0
125	37818581	Sofia	479	0.0



Host Popularity vs Availability in New York City

- Hosts with high review counts (e.g., Gurpreet Singh, Michael, Deloris) indicate strong customer demand and high booking frequency.
- Most top hosts show low availability (0 days), suggesting their listings are fully booked or consistently occupied.
- The scatter plot reveals a dense cluster of moderately reviewed listings with varied availability, while a few outliers dominate total reviews — showing a power-law effect where a small number of hosts capture most of the activity.

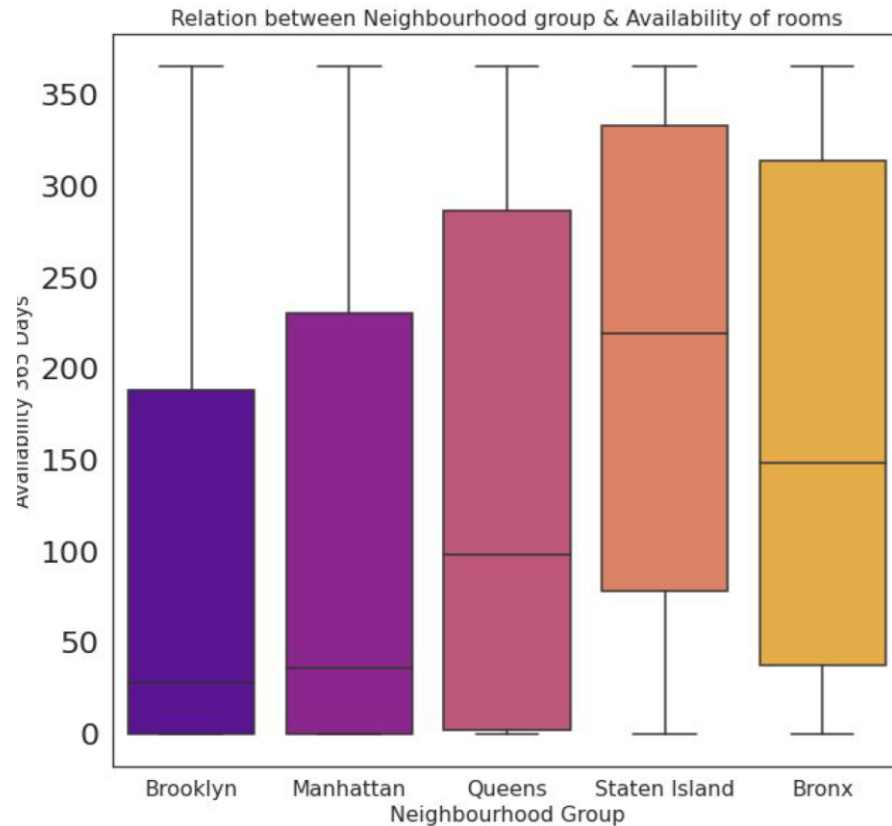
Count of Listed Rooms



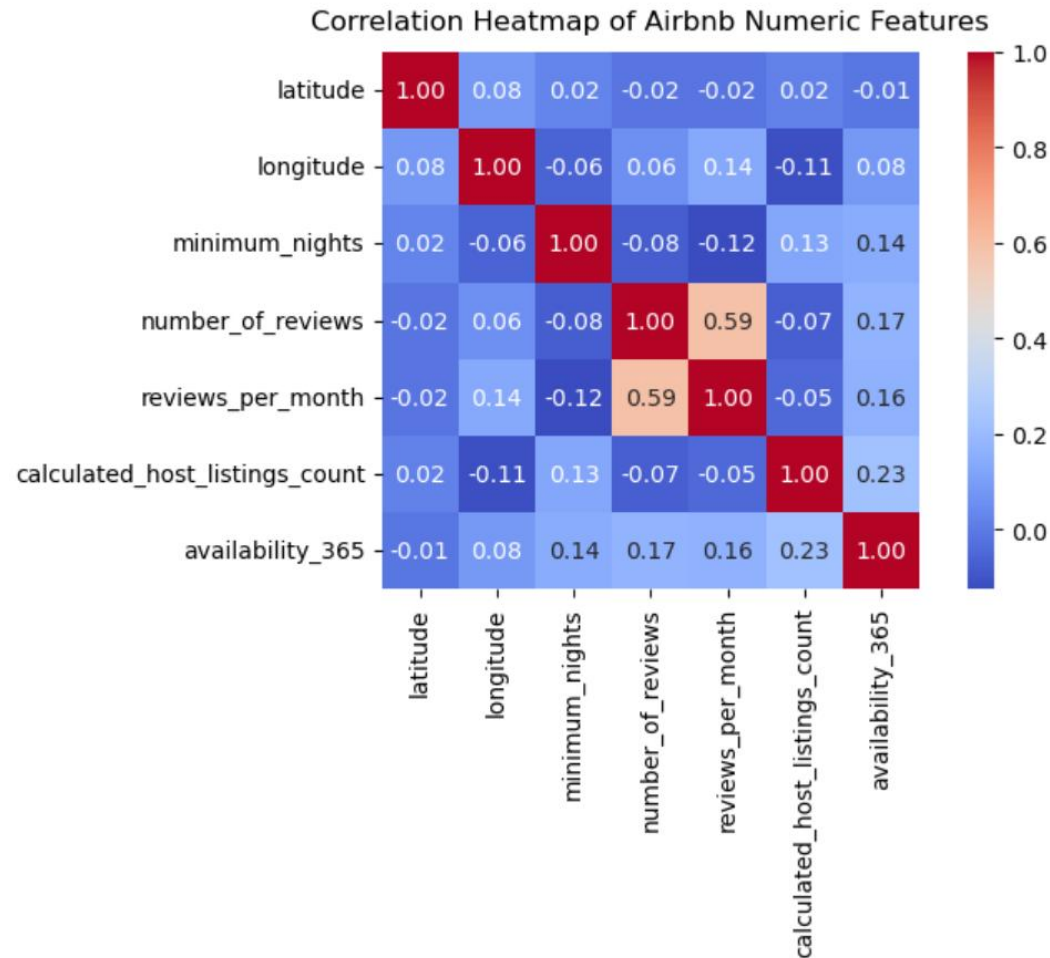
Room Type Composition of Airbnb Listings in New York City

- Entire homes/apartments make up the largest share (52%) of all listings — showing guests prefer full-property rentals.
- Private rooms represent about 46% of listings, catering to budget-conscious travellers.
- Shared rooms are rare (only 2.4%), indicating low demand for shared accommodations.
- Overall, Airbnb's NYC market is dominated by private and entire-home stays, highlighting a strong preference for privacy and comfort.

Room Availability by Borough and Room Type in New York City



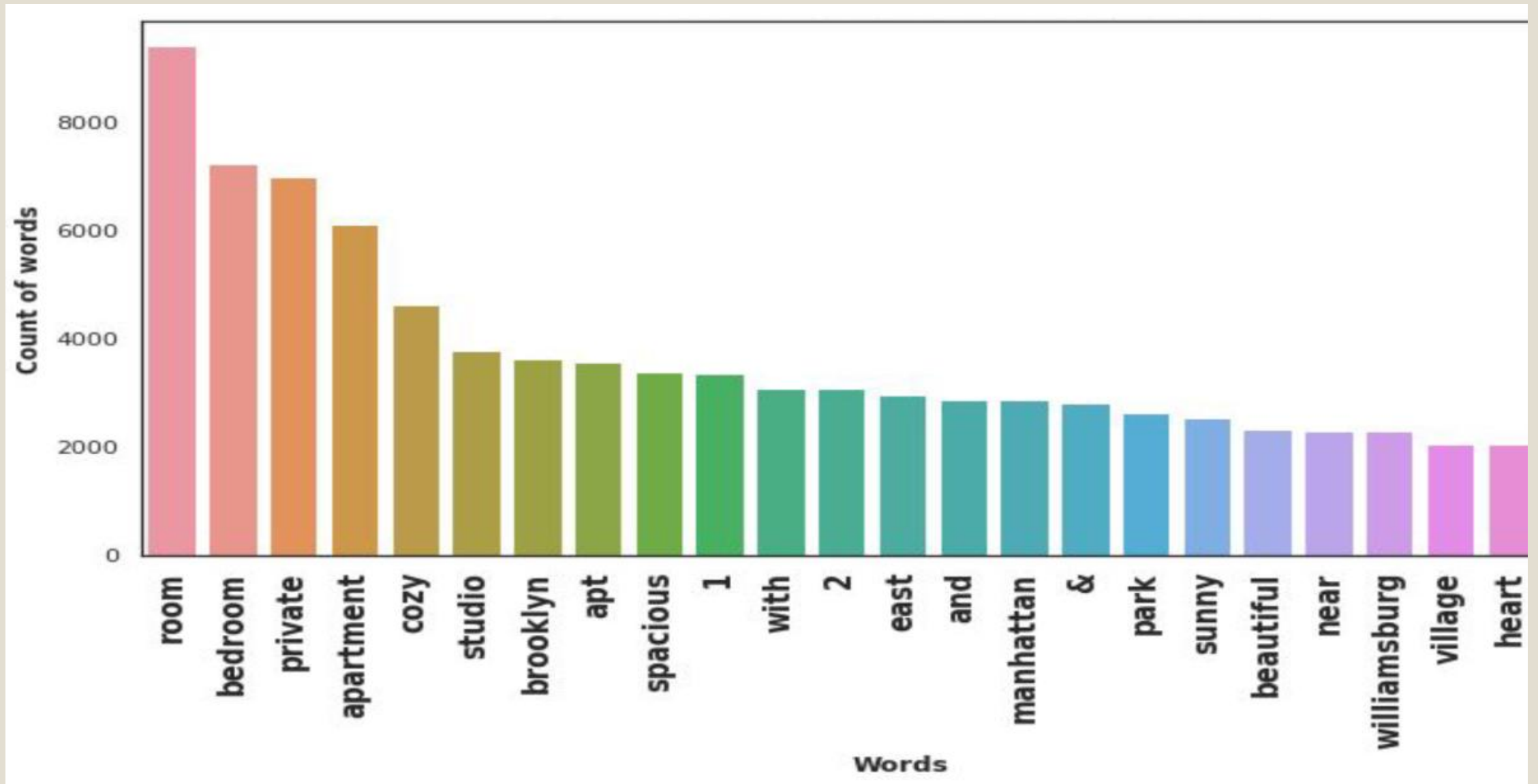
- Brooklyn and Manhattan have the lowest average availability, with many listings showing 0 days available - indicating high occupancy or full bookings.
- Bronx and Staten Island show the highest average availability (around 300 days/year), reflecting lower demand or longer idle periods.
- This pattern suggests longer stays or higher turnover in central boroughs (Manhattan, Brooklyn) and underutilised listings in outer areas.



Relationships Between Key Numerical Features

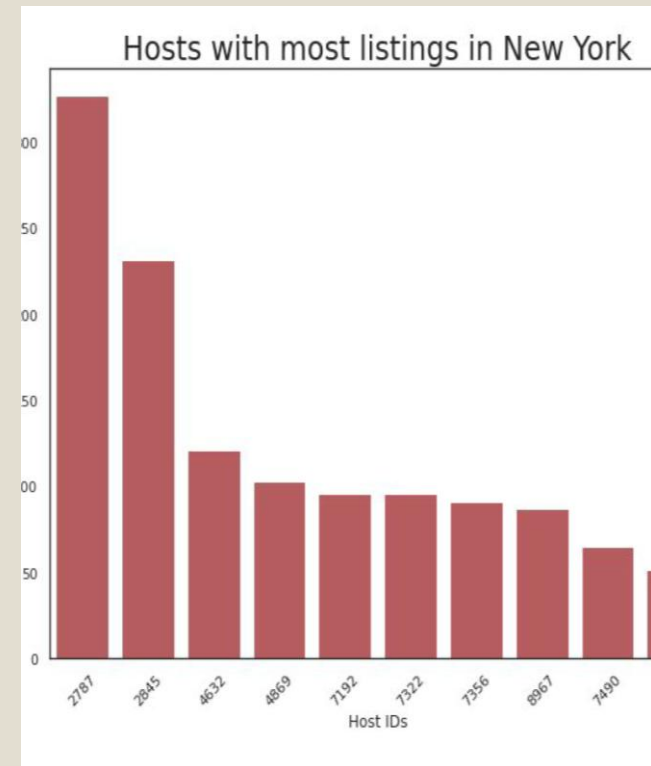
- Most variables show weak or no correlation, indicating that Airbnb metrics behave largely independently of one another.
- The strongest relationship ($r = 0.59$) occurs between `number_of_reviews` and `reviews_per_month`, meaning listings with more total reviews tend to receive consistent monthly feedback.
- `Availability_365` and `calculated_host_listings_count` show a slight positive correlation ($r = 0.23$) — hosts managing multiple properties tend to have more days open for booking.
- `Minimum_nights` and `reviews_per_month` display a small negative correlation, suggesting that listings requiring longer stays receive fewer monthly reviews.
- Overall, location and room type (categorical features) likely influence Airbnb performance more than numerical interactions.

Top 25 Most Common Words Used in listing names



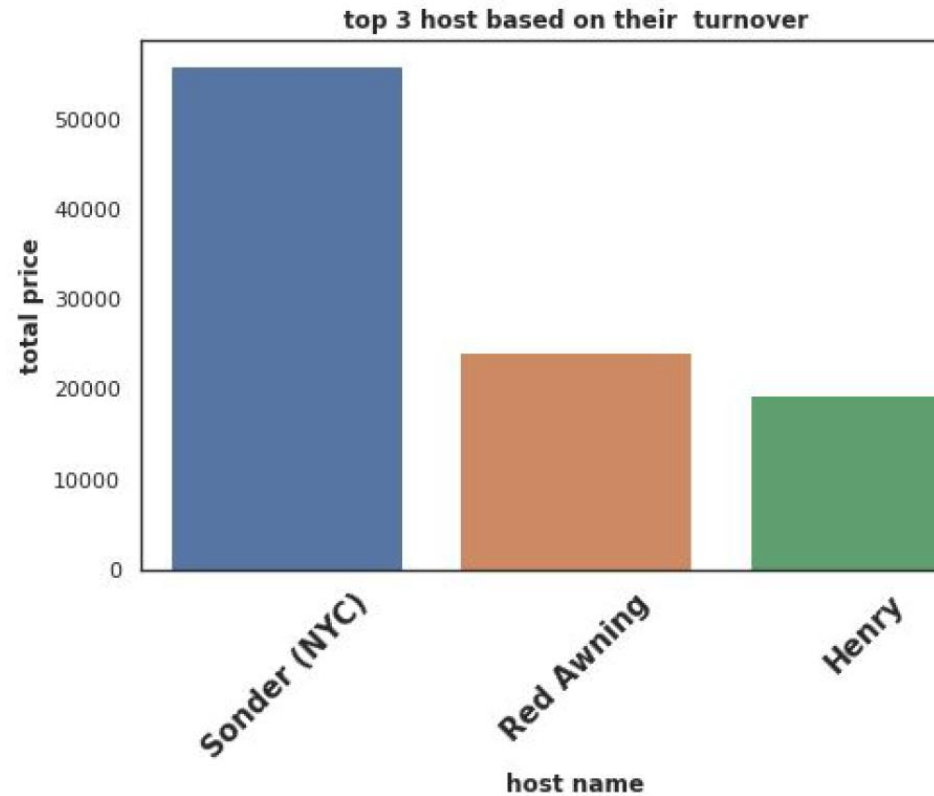
Top 10 Airbnb Hosts by Number of Listings in New York City

- The top 10 hosts together account for roughly 2.5% of all listings (about 1,270 out of 49,000+), showing high market concentration among a few large hosts.
- One host alone manages over 300 listings, suggesting the presence of professional or corporate property managers in the Airbnb NYC market.
- This indicates that while most hosts manage one or two listings, a small elite group controls a significant portion of total supply.



Top 3 Airbnb Hosts by Total Turnover in New York City

- Sonder (NYC) ranks as the highest-earning host, with total turnover exceeding \$50,000.
- Red Awning and Henry follow in second and third positions, generating significantly lower revenue.
- The sharp gap between Sonder and others suggests a strong market advantage and possibly a larger property portfolio or higher pricing strategy.



Challenges Encountered During Data Analysis

01

Understanding the dataset and interpreting column meanings required in-depth exploration of Airbnb's data structure.

02

Managing missing values, duplicates, and data inconsistencies was essential for accurate insights [5].

03

Translating the business model of Airbnb into measurable data relationships took additional research [6].

04

Creating clear and meaningful visualizations to communicate findings effectively was time-intensive but crucial for storytelling.

Key Findings

Manhattan and Brooklyn are the most active markets, attracting both the highest number of hosts and the highest customer spending.

Entire homes/apartments dominate listings (52%), showing strong guest preference for private stays, while shared rooms remain minimal (2.4%).

Sonder (NYC), Red Awning, and Henry lead in host turnover, indicating professional management dominance.

Longer stays are common in private rooms within central boroughs like Manhattan and Brooklyn.

Recommendations



Hosts: Focus on *entire home* and *private room* offerings to align with guest demand and maximize revenue.



Airbnb: Encourage more listings in outer boroughs (Bronx, Staten Island) to balance supply and meet potential untapped demand.



Pricing Strategy: Maintain competitive pricing in Manhattan and Brooklyn; consider promotional pricing in less-booked areas.



Platform Optimization: Highlight high-performing hosts and create tools for new hosts to replicate their success strategies.

References:

- [1] Dgomonov, "New York City Airbnb Open Data," [www.kaggle.com](https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data), 2019.
<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>
- [2] Suresh Kumar Mukhiya and U. Ahmed, *Hands-on exploratory data analysis with Python : perform EDA techniques to understand, summarize, and investigate your data*. Birmingham, Uk: Packt Publishing, 2020.
- [3] Microsoft, "Power BI - Data Visualization | Microsoft Power Platform," Microsoft, 2025.
<https://www.microsoft.com/en-us/power-platform/products/power-bi>
- [4] E. Alizadeh, "A Guide to Metrics in Exploratory Data Analysis | Towards Data Science," *Towards Data Science*, Dec. 15, 2020. <https://towardsdatascience.com/a-guide-to-metrics-in-exploratory-data-analysis-250b33f72297/>
- [5] M. Afkanpour, E. Hosseinzadeh, and H. Tabesh, "Identify the most appropriate imputation method for handling missing values in clinical structured datasets: a systematic review," *BMC Medical Research Methodology*, vol. 24, no. 1, Aug. 2024, doi: <https://doi.org/10.1186/s12874-024-02310-6>.
- [6] A. Boswijk, "Transforming Business Value through Digitalized Networks: A Case Study on the Value Drivers of Airbnb," *Journal of Creating Value*, vol. 3, no. 1, pp. 104–114, Apr. 2017, doi: <https://doi.org/10.1177/2394964317697736>.

The background of the image is a dense, overlapping collage of numerous small, rectangular sticky notes. These notes are in various colors, including shades of blue, green, purple, and yellow. Each sticky note features a large, bold, black question mark. The notes are scattered across the entire frame, creating a textured and busy visual effect.

Thank you, questions welcome.