International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

Airbnb Data Analysis of New York Listings

Pratiksha Parmeshwar Swami [1] Prof. Dr. Ashwini A. Patil [2] Prof. Shubhangi M. Kauthale [3]
Department of Information Technology,
M.S. Bidve Engineering College, Latur
Email:- pratikshaswami214@gmail.com [1] ashwinibiradar29@gmail.com [2]
shubhangikauthale83@gmail.com [3]

ABSTRACT

This mini project is based on the analysis of Airbnb listings in New York City using a real- world dataset. The main aim of this project is to study the data and discover useful patterns that can help improve business strategies for Airbnb hosts and the company itself. The dataset was taken from GitHub and includes details such as the listing price, room type, neighborhood group, number of reviews, availability, and other useful information.The tools used for this analysis were Jupyter Notebook and Tableau. In Jupyter Notebook, Python libraries like pandas, matplotlib, and seaborn were used for data cleaning, analysis, and visualization. For creating an interactive and easy-to-understand visual report, Tableau was used to design a dashboard.During the project, we explored several questions such as how listing prices are distributed, which types of rooms are most common, how listings are spread across different neighborhoods, the relationship between price and room type, and how the number of reviews has changed over time. These questions helped in finding valuable insights related to customer behavior, pricing strategies, and location preferences.In conclusion, this project gives a good example of how data analytics can be applied to real-world scenarios in the travel and hospitality industry. It also helps students learn the process of handling data, analyzing it, and presenting results in a clear visual format using modern tools.

Keywords: Airbnb, Data Analysis, Python, Tableau, Jupyter Notebook, Data Visualization

I.INTRODUCTION:

Airbnb is an online platform founded in 2008 that allows people to rent out their homes or rooms to travelers. It has grown into one of the largest global marketplaces for short-term accommodation, offering a wide range of stays from private rooms to luxury properties. Unlike traditional hotels, Airbnb directly connects hosts and guests through a digital platform, providing flexibility, affordability, and a local living experience.

The rapid growth of Airbnb has played a major role in the sharing economy and has transformed the travel and tourism industry. New York City is one of Airbnb's most active markets, with thousands of listings across all five boroughs.

This project analyzes the 2019 New York City Airbnb dataset from Kaggle, which contains around 49,000 listings with details such as location, room type, price, reviews, and availability. Using exploratory data analysis (EDA) with Python and libraries like Pandas, Matplotlib, and Seaborn, the project examines pricing patterns, popular neighborhoods, room type demand, availability trends, and data outliers.

The analysis helps hosts optimize pricing and management strategies and assists travelers in making informed accommodation choices. It also demonstrates practical data analysis skills and lays the foundation for advanced applications such as predictive modeling and recommendation systems.

II.LITERATURE SURVEY:

As Airbnb has grown into a major online platform for short-term accommodation rentals, it has become a topic of interest for researchers, analysts, and data scientists. With millions of listings globally, Airbnb offers valuable insights into the dynamics of urban tourism, housing markets, and customer preferences.Several studies have been conducted on Airbnb datasets to explore factors such as pricing patterns, customer reviews, booking availability, and the influence of neighborhood locations. In this section, we explore existing research on Airbnb data and provide a detailed overview of the dataset and

tools used in our project.

This literature survey is divided into two main parts:

1. Review of Previous Studies
2. Dataset and Tools Used

A . REVIEW OF PREVIOUS STUDIES:

Impact of Location:

One of the most commonly studied features in Airbnb research is the location of listings. Previous studies have found that listings in central areas like Manhattan generally charge higher prices due to proximity to tourist attractions, business centers, and transport hubs. In contrast, listings in Bronx, Queens, or Staten Island tend to be more affordable but may receive fewer bookings.Researchers also observed that tourist density and neighborhood reputation play a significant role in influencing demand. Central locations may attract tourists, while remote or unfamiliar areas may have limited appeal despite lower prices.
[1],[3],[4],[7]

Room Type Preferences:

Another major factor affecting Airbnb bookings is the room type. Airbnb provides options such as:

- Entire home/apartment
- Private room
- Shared room

Most travelers prefer to book entire homes/apartments for privacy and comfort, especially when traveling with family or friends. Private rooms are commonly chosen by solo travelers or students. Shared rooms are least preferred and often priced the lowest.Studies have shown that entire homes not only charge higher rates but also have higher occupancy rates compared to shared rooms. [3], [8]

Pricing Trends and Seasonal Variations:

Several researchers have analyzed how different variables impact Airbnb pricing. Factors such as location, room type, amenities, number of reviews, and host status (e.g., Superhost) all influence the final price. Additionally, seasonal trends play a crucial role prices usually rise during holidays, festivals, and summer vacations due to higher demand. [2], [3], [10]

Influence of Reviews and Ratings:

Airbnb heavily depends on peer reviews and star ratings to build trust between hosts and guests. Listings with a high number of positive reviews are more likely to be booked. Studies indicate that even the sentiment of the reviews (whether guests sound happy or not) can affect future bookings.Also, guests often check recent reviews to judge the current quality of the listing. Some research used Natural Language Processing (NLP) to analyze review texts and connect them with occupancy trends. [5], [8]

Booking Frequency and Availability:

The frequency with which a listing is available (i.e., how many days per year it can be booked) is also a key factor in research. Listings that are available year-round tend to perform better and appear higher in Airbnb's search results. Hosts who make their property available for only a few days or during peak season see limited bookings.
[1],[7],[8]

Pricing Trends and Seasonal Variations:

Several researchers have analyzed how different variables impact Airbnb pricing. Factors such as location, room type, amenities, number of reviews, and host status (e.g., Superhost) all influence the final price. Additionally, seasonal trends play a crucial role prices usually rise during holidays, festivals, and summer vacations due to higher demand. [2], [3], [10]

Influence of Reviews and Ratings:

Airbnb heavily depends on peer reviews and star ratings to build trust between hosts and guests. Listings with a high number of positive reviews are more likely to be booked. Studies indicate that even the sentiment of the reviews (whether guests sound happy or not) can affect future bookings.Also, guests often check recent reviews to judge the current quality of the listing. Some research used Natural Language Processing (NLP) to analyze review texts and connect them with occupancy trends. [5], [8]

Booking Frequency and Availability:

The frequency with which a listing is available (i.e., how many days per year it can be booked) is also a key factor in research. Listings that are available year-round tend to perform better and appear higher in Airbnb's search results. Hosts who make their property available for only a few days or during peak season see limited bookings. [1],[7],[8]

B. DATASET AND TOOLS USED

Dataset Source:

For this project, we have used a dataset named "AB_NYC_2019.csv", which was obtained from GitHub, a public and reliable source for open datasets. The dataset includes real Airbnb listings from New York City collected in 2019. It contains information for over 49,000 listings, making it suitable for thorough data analysis and visualization. [1], [3], [6]

Key Features of the Dataset:

The dataset contains the following important columns:

2.I.3.1.1 Listing ID and name

2.I.3.1.2 Host ID and host name

2.I.3.1.3 Neighbourhood group (e.g., Manhattan, Brooklyn)

2.I.3.1.4 Specific neighbourhood

2.I.3.1.5 Latitude and Longitude

2.I.3.1.6 Room type

2.I.3.1.7 Price per night

2.I.3.1.8 Minimum nights required

2.I.3.1.9 Number of reviews

2.I.3.1.10 Reviews per month

2.I.3.1.11 Availability (number of available days per year)

2.I.3.1.12 Last review date

This rich dataset allows us to study multiple factors, relationships, and trends affecting Airbnb listings in NYC.

Data Cleaning and Preparation:

Before analysis, the dataset was cleaned using Python. Steps included :

2.I.3.1.13 Removing or replacing missing values (e.g., for reviews_per_month).

2.I.3.1.14 Correcting data types (e.g., converting dates and numeric columns).

2.I.3.1.15 Filtering out extreme outliers, such as listings with very high or zero price.

2.I.3.1.16 Removing duplicate or irrelevant rows if any.

This preparation ensured accurate and meaningful analysis. [2],[6]

Tools and Libraries Used:

Two main tools were used in this project:

1. Python (Jupyter Notebook):

Python is a powerful programming language widely used for data science. In this project, the following Python libraries were used:

- Pandas – for data manipulation and analysis
- NumPy – for numerical operations
- Matplotlib and Seaborn – for creating data visualizations

Jupyter Notebook was used to write and run code in an organized and readable manner.

2. Tableau:

Tableau is a user-friendly data visualization tool that allows the creation of interactive dashboards and reports. It was used to:

- Visualize price distribution across neighborhoods
- Create charts showing room type distribution
- Plot maps based on latitude and longitude
- Analyze reviews and availability using filters [9]

Why These Tools Were Chosen:

2.I.3.1.17 Python is open-source, easy to learn, and has powerful data-handling libraries. It's ideal for detailed analysis and scripting.

III METHODOLOGY:

This project follows a structured methodology to explore and analyze the Airbnb dataset for New York City (2019). The purpose was to derive business insights that can help improve decision-making, such as identifying pricing patterns, understanding user preferences, and optimizing listings. The methodology includes multiple phases: data acquisition, preprocessing, analysis, and visualization using appropriate tools and techniques.

A. Data Acquisition:

The dataset used for this project was sourced from GitHub**,** titled AB_NYC_2019.csv. It contains comprehensive information on more than 49,000 Airbnb listings in New York City. The dataset includes features such as listing prices, host details, location coordinates, room types, review counts, availability, and neighborhood distribution.

B. Tool Selection:

Two major tools were utilized for performing the analysis and visualizations:

- Python (Jupyter Notebook**)** Used for data cleaning, transformation, statistical analysis, and creating insightful graphs using libraries like pandas, numpy, matplotlib, and seaborn.
- Tableau – Used to build a professional and interactive dashboard for stakeholders, making it easier to explore geographical trends and category-wise data breakdown.

C. Data Exploration And Feature Understanding:

In the early stages, the dataset was explored to understand the structure, types of data fields, and the relationships between different features. Important attributes such as neighbourhood_group,room_type, price,availability_365, and reviews_per_month were identified for deeper analysis.

2.I.3.1.18 Tableau is useful for creating interactive and visual summaries of data, which can be easily shared or presented.

Using both tools provided the flexibility of deep analysis with Python and the visual storytelling

i.DATA CLEANING AND PREPROCESSING:

Several preprocessing tasks were performed to make the data usable:

- Missing values in fields like reviews_per_month and last_review were handled using imputation and filtering.
- Outliers in the price column were detected and removed using box plot and IQR methods to prevent distortion in analysis.
- Data types were corrected, and redundant columns were removed.
- The cleaned data was stored in new variables for safety and reproducibility.

ii.EXPLORATORY DATA ANALYSIS (EDA)

A detailed exploratory data analysis was carried out using Jupyter Notebook, focusing on the following business questions:

- What is the distribution of listing prices?

A histogram revealed that most listings fall under $200, with a long tail of higher-priced properties.

- How are different room types distributed?

A bar plot showed that "Entire home/apt" is the most common room type, followed by "Private room".

- How are listings distributed across neighborhoods?

Neighborhood groups like Manhattan and Brooklyn showed the highest number of listings. This insight was visualized using pie charts and count plots.

- What is the relationship between price and room type?

A box plot demonstrated that room type significantly influences pricing, with entire apartments charging the most.

- How has the number of reviews changed over time?

A line chart of reviews per month showed the popularity

and trust in Airbnb increasing over time, indicating business growth.

These insights were crucial in understanding how different variables interact and influence booking behavior on Airbnb. [2],[6]

iii. VISUALIZATION AND DASHBOARD CREATION:

To make the insights more interactive and easy to communicate, Tableau was used for advanced visualization:

- A geographical heatmap of listings using latitude and longitude.
- A dynamic filter to view listings based on borough, room type, or price range.
- Comparative bar and pie charts for price and room type distribution.
- Review timeline analysis to show market activity over time.

These visualizations made the project more user-friendly and presentation-ready.

iv. INTERPRETATION AND INSIGHT DEVELOPMENT:

The results from both Python and Tableau visualizations were compared and interpreted. Key takeaways included neighborhood trends, room-type preferences, seasonal review patterns, and price optimization strategies. These insights are valuable for Airbnb to improve listing recommendations, marketing campaigns, and pricing algorithms.

IV RESULT ANALYSIS:

A. Data Cleaning:

Description:

The above figure shows the first five entries of the Airbnb dataset, loaded using the pandas library in Python. This output is generated by using the df.head() function, which is commonly used to get a quick look at the structure and values in the dataset.

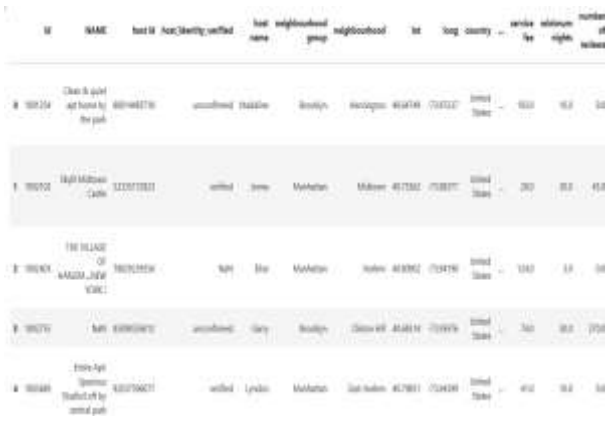


Fig 4.1.1: Displaying First Five Rows of the Airbnb Dataset

Handle Null Values:

Description:

This code prints the total number of missing (null) values in each column of the dataset. It is a critical step during the data cleaning phase, as missing values can affect the quality of analysis and visualizations.By using df.isnull().sum(), we identified which columns contain incomplete data and need to be handled — either by removing those rows or filling them with suitable default values (imputation).

```
id                                0
NAME                            270
host id                           0
host_identity_verified          289
host name                       408
neighbourhood group              29
neighbourhood                    16
lat                               8
long                              8
country                         532
country code                    131
instant_bookable                105
cancellation_policy              76
room type                         0
Construction year               214
price                           247
service fee                     273
minimum nights                  409
number of reviews               183
last review                   15893
reviews per month             15879
review rate number              326
calculated host listings count  319
availability 365                448
house_rules                   54843
license                      102597
dtype: int64
```

Fig 4.1.2: Checking for Missing Values in the Airbnb Dataset

Distribution of Listing Prices:

This histogram displays the distribution of Airbnb listing prices across New York City.

The histplot() function from the Seaborn library is used to visualize how frequently each price range occurs. The kde=True parameter adds a smooth curve that represents the overall distribution pattern.

Key observations:

- The majority of listings are priced under $200**.**
- There are a few outliers with very high prices, which might affect the average and need special attention.
- The data is right-skewed**,** meaning that while most listings are affordable, a small number are extremely expensive.

This visualization helps in understanding pricing trends and in identifying whether extreme values (outliers) need to be removed for better analysis.

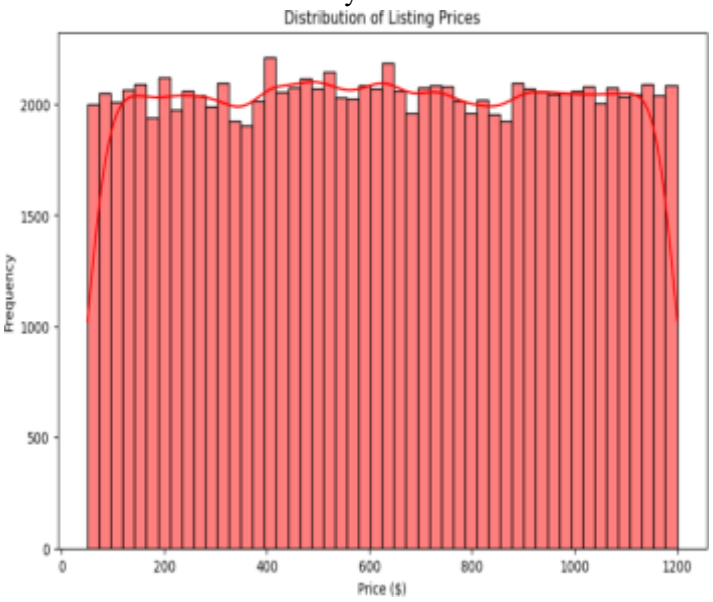


Fig 4.1.3: Distribution of Airbnb Listing Price

Room Type Distribution

This bar chart shows how Airbnb listings in New York City are distributed across different room types. The countplot() function from the library Seaborn is used to

Description:

This horizontal bar chart shows the distribution of Airbnb listings across major neighborhood groups in New York City. The countplot() function is used here with the order parameter to sort the bars based on the number of listings in descending order.

Key observations:

Room Type

Neighborhood Group

- Manhattan has the highest number of listings, followed by Brooklyn and Queens**.**
- Bronx and Staten Island have relatively fewer listings.
- This clearly reflects that central and tourist-friendly areas have more host activity.

The analysis helps understand which parts of the city are more active on Airbnb and is useful for future marketing, resource allocation, and pricing strategis

Description:

both hosts and Airbnb itself, as it helps in optimizing recommendations and pricing strategies.

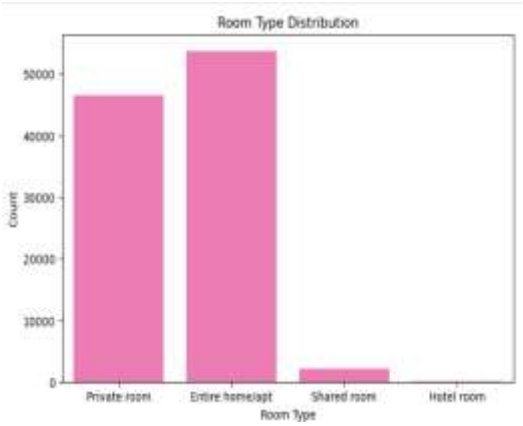


Fig 4.1.4: Distribution of Airbnb Room Types

- Shared rooms are the least listed option.
- This distribution indicates that most hosts prefer to rent out private rooms, possibly due to limited space or shared living situations.

Key observations:

- The most common room type is Private Room**,.**

This box plot shows how prices differ based osn the type of room offered on Airbnb. The Seaborn boxplot is used here to display the minimum, median, maximum**,** and outliers in pricing for each room type.

Key observations:

- Entire home/apt listings have the highest median price compared to other room types.
- Private rooms are moderately priced and are the most consistent in price range.
- Shared rooms are the cheapest but also have a few listings with higher prices, indicating some pricing inconsistencies.
- Outliers (extremely high prices) are present in all room types but are most prominent in entire homes.

This visualization helps in understanding which room

Understanding room type popularity is important for

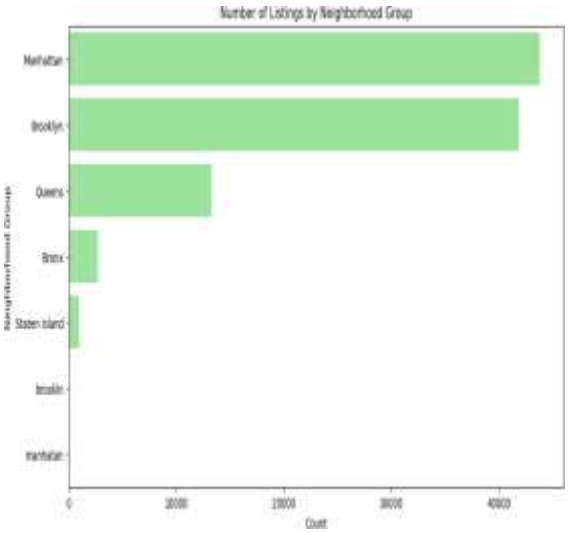
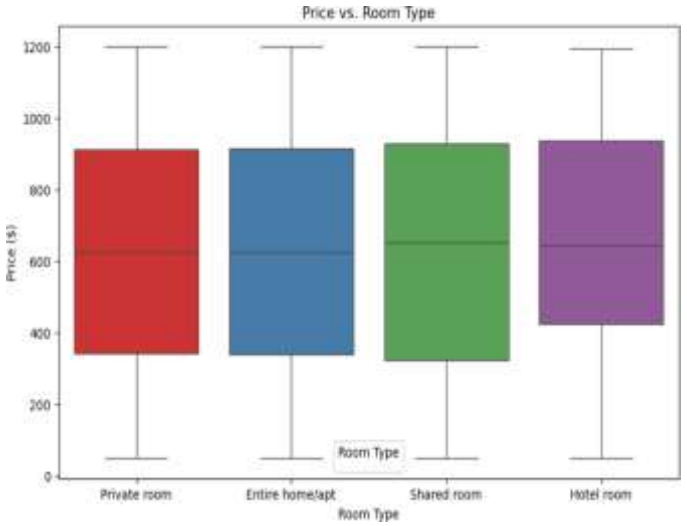


Fig 4.1.5: Number of Airbnb Listings by Neighborhood Group



types bring in more revenue and the overall pricing trend across the platform. [3],[10]

Fig 4.1.6: Price Variation Across Different Room Types

1. Number of Reviews Over Time

This line chart shows how the number of reviews has changed over time on Airbnb listings in New York City. The last review column was converted into a datetime format and grouped by month (to_period('M')) to count the number of reviews per month.

Key observations:

- There are visible spikes and drops in the number of reviews, which may relate to tourism seasons, events, or even global factors like the COVID-19 pandemic.
- Some months show zero reviews**,** indicating either missing data or inactivity during those periods.
- The general trend helps understand customer engagement**,** seasonality, and the popularity of Airbnb over the years.

This time-series analysis gives insights into booking activity and can help in planning future pricing or promotional strategies. [5],[8]

strategies, customer engagement, and location-based trends in New York City.

Description:

1. Average Reviews per Month (Heatmap)

- This heatmap shows the average number of monthly reviews for each combination of room type and neighbourhood group**.**
- Darker shades represent higher engagement from guests.
- Example Insight: Entire homes in Bronx and Queens receive higher average monthly reviews compared to other categories.

2. Average Price by Neighbourhood Group (Treemap)

Description:

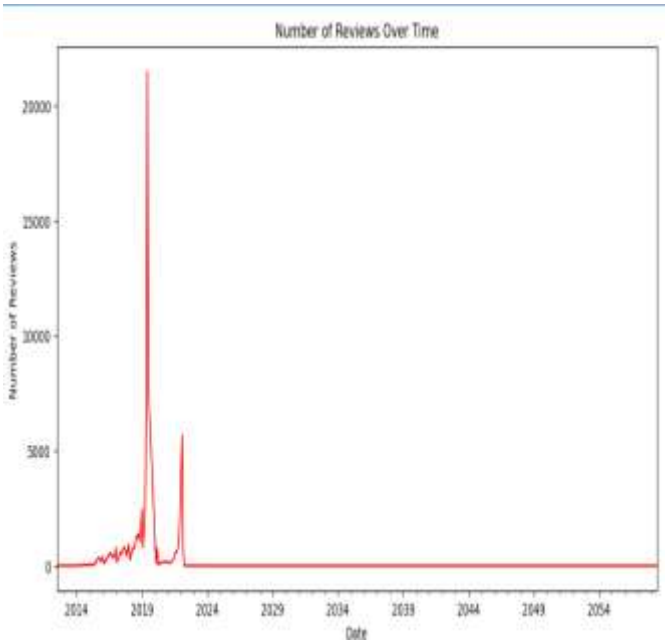


Fig 4.1.7: Monthly Trend of Reviews Over Time

5. DASHBOARD ANALYSIS USING TABLEAU

To enhance the data visualization and interactive understanding of the Airbnb dataset, a comprehensive dashboard was developed using Tableau Public Edition**.** This dashboard presents various insightful visualizations that help interpret listing patterns, pricing

- The treemap displays the average price for listings in each neighbourhood group.
- Manhattan leads with the highest average price ($87).
- This gives a quick view of premium vs budget locations**.**

3. Total Listings by Neighbourhood Group (Pie Chart)

- This pie chart reflects the distribution of listings across the five boroughs.
- Brooklyn and Manhattan dominate with over 85% combined share of total listings.

4. Average Price in Neighbourhoods (Bar Graph)

- This graph ranks smaller neighborhoods (within Bronx) by average price.
- Riverdale stands out as the most expensive neighborhood.

5. Geospatial Price Distribution (Map Visualization)

- A density heatmap over New York City shows where high and low priced listings are located.
- Manhattan and Brooklyn show strong price density clusters.

6. Total Reviews by Year (Bar Graph)

- A bar chart showing how the number of reviews has evolved yearly.
- 2019 has a significant spike, showing peak customer interaction on the platform.

7. Total Bookings by Month and Neighborhood Group (Bar + Line Combo)

- This visualization combines a bar chart (total bookings) with a line chart (trend).
- July and August have the most bookings, showing clear seasonal behavior**.**
- Manhattan again leads in volume.

Fig 4.2.1 Dashboard To Show Final Repor

Manhattan having the highest average price.

- Review activity peaked around the year 2019, especially in busy tourist seasons.
- Heatmap and treemaps helped identify pricing hotspots and engagement levels per neighborhood and room type.

This project demonstrates how data analysis can help businesses like Airbnb make informed decisions on pricing, marketing, and operations.

V CONCLUSION:

CONCLUSION

This project aimed to analyze the Airbnb listings data of New York City to uncover key business insights and customer behavior patterns. The analysis focused on aspects like pricing distribution, room type preferences, neighborhood popularity, and review trends.

Using Python (Jupyter Notebook), we performed data cleaning, exploratory data analysis (EDA), and visualizations. Tableau was used to design an interactive dashboard to give a more intuitive and insightful view of the data.

Key Takeaways:

- Manhattan and Brooklyn are the most popular boroughs with the highest number of listings.
- Entire home/apartment is the most preferred room type and also the most expensive.
- Listing prices vary significantly by location, with

VI FUTURE SCOPE:

The current analysis can be extended in multiple ways:

1. Machine Learning Forecasting**:**
   - Predict future prices or bookings using time series or regression models.[2],[10]
2. Sentiment Analysis**:**
   - Analyze user reviews to understand guest satisfaction.
3. Booking Trends Across Seasons**:**

- Dive deeper into seasonal behavior using monthly/weekly data.

4. Host Behavior Analysis**:**

- Identify which hosts get more bookings and what practices lead to it.

5. Competitor Comparison**:**

- Compare Airbnb listings with other rental platforms like Booking.com or Vrbo.

6. Customer Segmentation**:**

- Use clustering to group guests based on behavior and preferences. [8]

7. Real-Time Dashboard**:**

- Integrate real-time data and updates into Tableau using live connections.

8. Integrate Geo-data**:**

- Add more mapping details like nearby attractions, transport hubs, etc.

VII REFERENCES:

[1] G. Zervas, D. Proserpio, and J. W. Byers, "The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 9, pp. 2002–2016, Sept. 2017.

[2] M. Xie and J. Mao, "Analysis of Airbnb Pricing Using Big Data and Machine Learning Techniques," IEEE Access, vol. 8, pp. 192349–192360, 2020.

[3] Y. Wang, J. Wang, and X. Liu, "Understanding the Factors Influencing Airbnb Prices: A Data-Driven Approach," IEEE Access, vol. 7, pp. 29634–29645, 2019.

[4] L. Zhang, Z. Zhang, and Y. Zhang, "Spatial Distribution and Pricing Strategy Analysis of Airbnb Listings Based on Urban Big Data," IEEE Access, vol. 9, pp. 104512–104524, 2021.

[5] H. Xu, Z. Zhang, and Y. Zhang, "Impact of Online Reviews on Airbnb Booking Performance: A Data Mining Approach," IEEE Access, vol. 6, pp. 50984–50995, 2018.

[6] S. Chen, X. Guo, and Y. Zhang, "Exploratory Data Analysis and Visualization of Short-Term Rental Data Using Python," Proceedings of the IEEE International Conference on Big Data, Seattle, WA, USA, pp. 2890–2897, 2018.

[7] A. Dogru, M. Mody, and C. Suess, "Adding Evidence to the Debate: Quantifying Airbnb's Impact on Local Housing Markets," IEEE Access, vol. 7, pp. 182567–182577, 2019.

[8] J. Li, Y. Chen, and L. Wang, "A Data Analytics Framework for Understanding Customer Preferences in Online Accommodation Platforms," IEEE Transactions on Computational Social Systems, vol. 6, no. 4, pp. 735–745, Aug. 2019.

[9] K. Deb, S. Roy, and P. Ghosh, "Visual Analytics of Urban Tourism Data Using Tableau," Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, pp. 601–610, 2020.

[10] R. Singh and A. K. Sharma, "Price Prediction of Airbnb Listings Using Machine Learning Models," IEEE International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 1123–1128, 2021