MIS 584 FINAL PROJECT REPORT

Airbnb: Leveraging Business Intelligence for Strategic Insight

BY
Adeniyi Talabi
Ahmad Seayar Sroosh
Dhruvik Patva
Pavani Pingili
Sheroz Shaikh

WPI SPRING '24

Table of Contents

Executive Summary

As one of the biggest online marketplaces for short-term rentals, Airbnb also have lots of trouble storing and utilizing data to support wise decisions and efficient operations. The organization finds it harder to get a cohesive picture of its data due to the volume and complexity of data growing, which results in disjointed insights and ineffective decision-making procedures.

Problem statement:
Due to Airbnb's insufficient data management architecture, there are insufficient data sets, inconsistent data sources, and manual processing. This makes it more difficult to find inefficient regions and automate manual procedures, in addition to impeding the ability to produce insightful analysis. Furthermore, the company's capacity to grow and adapt to shifting market demands is hampered by the absence of a consolidated data repository and defined procedures.

In order to overcome these obstacles, the goal of this project is to create and put into place a complete business intelligence system that will allow Airbnb to use data to propel growth, streamline operations, and surpass consumer expectations in New York City. The following are the project's main goals:

- By developing a single source of truth for data and putting in place a business intelligence platform to make data analysis and visualization easier, to improve data-driven decision making.

- Reduce costs and increase productivity by finding inefficient areas in the business and automating manual tasks.

- Use data to personalize user experiences, expedite guest feedback, and enhance the booking process as a whole to increase customer happiness.

In order to meet these goals, the project will produce the following results:

Create a thorough data warehouse structure that combines and retains information from multiple sources, such as user, transactional, and performance metrics data. This will give the organization a single source of truth for its data, allowing it to get a cohesive picture of it and produce insightful analysis.

Make data-driven decisions possible by implementing interactive dashboards and reports that offer real-time visibility into key performance indicators (KPIs). This will enable the business to swiftly and intelligently make decisions by enabling real-time performance monitoring.

Install a business intelligence solution that makes it easier for users to access and comprehend data by facilitating self-service data analysis, exploration, and visualization. Employees in several areas will be empowered to make data-driven decisions and enhance their workflows as a result.

To guarantee data security, compliance, and quality, define and put into practice data governance rules and standards. As a result, data management and use will be governed by explicit policies and processes, guaranteeing data accuracy, security, and regulatory compliance.

To guarantee that the new business intelligence solution is successfully adopted and used, give important stakeholders support and training. This will guarantee that staff members have the abilities and information needed to operate the new system efficiently and reach its maximum potential.

Note: the data dictionary for the dataset used for this project is attached to appendix A.

Introduction

Airbnb is a peer-to-peer internet marketplace and homestay network that enables people to let tourists and visitors stay in their apartments, homes, or rooms. With listings in more than 191 countries and regions, Airbnb was founded in 2008 and has grown to become one of the biggest online markets for short-term vacation rentals.

Hosts can offer their houses on Airbnb and control the terms of availability and pricing. Based on their travel dates, destination, and other preferences, guests can then look for and reserve lodging. Each booking on Airbnb results in a service fee that is paid by both hosts and guests.
Airbnb is the leading online accommodation rental platform that has disrupted the traditional hotel industry. The company connects people who need a place to stay with people who have extra space to rent, offering a more affordable and flexible alternative for travelers. Airbnb's rapid growth is impressive, with over 900 million arrivals in more than 100,000 cities worldwide. Airbnb's business model is centered around commission fees charged to both hosts and guests, as well as additional services such as insurance and customer support. The company has also expanded its offerings to include "experiences" - activities led by local hosts. Airbnb does not own any properties itself; rather, it serves as a platform that facilitates transactions between hosts and guests. Despite Airbnb's success, the company faces several challenges, including competition from other vacation rental platforms, regulatory hurdles, and the need to optimize its operations and customer experience. To address these challenges, Airbnb can leverage the power of business intelligence (BI) solutions.

Airbnb uses Microsoft Excel for quick data analysis and for storing small amounts of data. They also use Excel as a downloadable format for house owners to download data. They use Spark 3 and Iceberg as their data warehousing tools. Using Metis as a seamless integration for their information systems. Metis is a complete data management software built in-house and used by Airbnb, it integrates the Apache software. It is made up of three core products: Dataportal, Unified Metadata Service (UMS), and Lineage Service. Together, this platform allows Airbnb to manage millions of data assets across many domains. Airbnb currently has a separate BI team called Business Intelligence team and the team uses Apache Superset as their BI and reporting system.

Airbnb currently seats in the “sage” stage of the BI maturity model as depicted below:

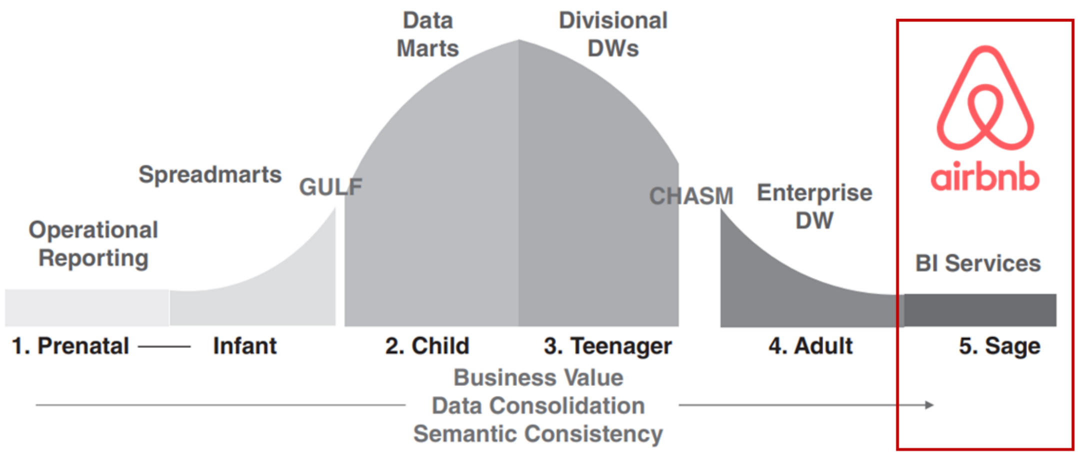

Fig: 01

To provide a guideline for Airbnb's BI implementation and to justify the value of BI solutions, we present two relevant case studies.

Case Study 1: Uber and Dynamic Pricing

Uber, a similar digital marketplace company, has successfully implemented BI to optimize its operations, particularly its dynamic pricing model. Uber uses real-time data and advanced analytics to adjust prices based on factors such as traffic conditions, journey times, driver availability, and customer demand. This data-driven approach has enabled Uber to improve customer satisfaction and operational efficiency.

Case Study 2: Zillow and Real Estate Analytics

Zillow, an online real estate database company, leverages BI to analyze housing market trends and user behavior. Zillow uses BI tools to gather data and provide insights, such as "Zestimates" - estimates of home values. The company also employs AI to enhance real estate transactions and has pioneered advanced BI strategies in the real estate industry.

These case studies highlight the potential benefits of BI for Airbnb, including improved operational efficiency, better decision-making, and enhanced customer experience. By learning from these examples, Airbnb can develop a strategic roadmap for implementing a robust BI system that aligns with its business objectives and helps the company maintain its competitive edge in the dynamic vacation rental market.

The scope of this project is focused on Airbnb's overall business operations and decision-making processes, with a particular emphasis on areas that can benefit from BI, such as dynamic pricing, operational optimization, and customer experience enhancement.

The Proposed Business Intelligence (BI) Solution

Business intelligence is an integrated set of tools to capture, collect, integrate, store, and analyze data with the purpose of generating and presenting information to support business decision making. Business intelligence are practices and systems that are critical in allowing business to access their data and be able to make use of it to address managerial questions. The steps taken to be able to utilize BI are implemented in stages; Firstly, data sources are needed. Vast amounts of data, to be able to gain insights into trends. So, it is gathered in such a way that makes it easier to interpret. The next step would be to store the data properly, in this case either in data marts or data warehouses. This makes sure that the vast quantity of data stored is organized. Lastly, the focus is on visualization and queries that can help us answer managerial questions. So, in other words, BI is not a product, but a framework of the following:

- Collecting and storing operational data
- Aggregating the operational data into decision support data.
- Analyzing decision support data to generate information.
- Presenting such information to the end user to support business decisions.

What are the main components of a BI solution?

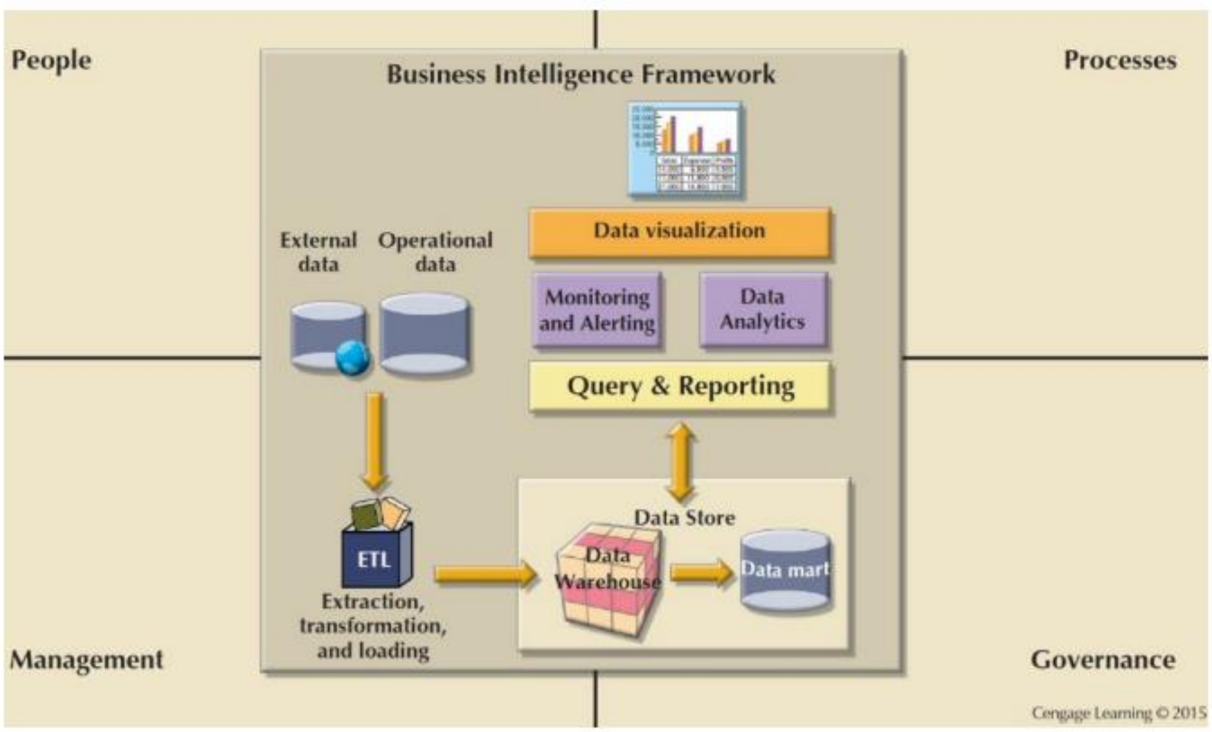

Fig: 02

As depicted in the picture above, the People, Process, Management, and Governance aspects of the Business Intelligence (BI) Framework represent the non-technical but essential components that ensure the BI system operates effectively and aligns with the organization's strategic goals. These four components work in conjunction to create a holistic BI environment where technology is effectively utilized to serve the organization's strategic needs.

1. People - The success of a BI system heavily depends on the individuals involved at every stage, from planning to execution and use. This includes IT professionals who manage and support the BI tools, data analysts who interpret the data, business users who utilize the insights in decision-making, and trainers who educate staff on how to use BI tools effectively. Also, people are responsible for defining the needs, questions, and problems that the BI system should address.

2. Process - Processes refer to the methodologies and procedures used to collect, manage, analyze, and report data. They are the workflows that detail how data moves from one stage to the next within the BI framework, ensuring that data is handled consistently and in a way that supports business objectives. This also includes the process of feedback and iteration, where insights gained from BI are used to refine and improve business processes.

3. Management - Management involves the strategic oversight and administration of the BI system. This includes setting the vision for how BI should be used within the organization, securing funding, managing resources, prioritizing projects, and ensuring that the BI strategy aligns with the company's overall business strategy. Management also has a role in promoting a data-driven culture and driving the adoption of BI tools across the organization.

4. Governance - Governance encompasses the policies, rules, and standards that guide the use of BI within the organization. It includes data governance (ensuring data quality, consistency, and security), as well as governance over the BI processes and technologies themselves. This ensures that the BI system complies with all relevant laws and regulations, such as those related to data privacy and financial reporting, and that the data is used ethically and responsibly.

Understanding the technical aspect of the BI framework:

The technical part outlines the various components and processes involved in transforming raw data into actionable business insights. It captures the entire data life cycle from data acquisition to storage, transformation, integration, presentation, analysis, monitoring, and alerting. These components are briefly discussed below:

1. Data Sources – Finding the right data sources is the first step to a successful BI. Data sources can be either external data or operational data. External data could include market trends, competitor information, economic indicators, etc., while operational data is generated internally from the company's day-to-day operations.

2. ETL (Extraction, Transformation, and Loading) Tools – ETL is a foundational process in data warehousing that involves extracting data from various sources, transforming it to fit operational needs (which can include quality assurance and cleansing), and loading it into a

database for use. This is the stage where raw data is converted into a more structured form suitable for decision making analysis.

3. Data Store (Data Warehouse/Data Mart) - After ETL, the processed data is stored in a centralized repository known as a Data Warehouse. A Data Mart is a subsection of a Data Warehouse and is usually oriented towards a specific business line or team. The warehouse serves as a consolidated source for data analysis and BI processes.

4. Query & Reporting and Data Analytics - This represents the analytics layer where data is queried, and reports are generated. It includes the creation of dashboards, performance reports, and other analytics outputs that help in decision-making. Data Analytics involves more complex processes like predictive modeling, trend analysis, and other advanced statistical techniques.

5. Data Visualization - The results of the data analysis are often visualized using various charts, graphs, and maps to make the data understandable at a glance. The most appropriate presentation format is used in revealing trends, patterns, and outliers within large datasets.

6. Monitoring and Alerting - The framework includes monitoring systems that continuously track business metrics in real-time against predefined thresholds and send alerts when anomalies or critical conditions are detected. This helps businesses respond quickly to changing conditions.

Understanding the aforementioned points, to be able to propose a comprehensive Business Intelligence solution tailored for Airbnb, it's important to understand the dynamics of the short-term rental marketplace and Airbnb's operational distinctions. The goal of our BI solution has many factors, so we aim to deepen insights into guest behaviors, refine listing and pricing strategies, elevate guest experiences, and improve operational and tactical efficiencies.

The foundation of our proposed solution starts with the integration of varied data sources. This makes sure that a robust and holistic understanding of Airbnb's operational and competitive landscape exists. This would include internal data streams such as booking details, cancellations, user platform interactions, host listings, pricing data, and customer service records. To be able to make these data streams useful, external data sources like market trends, economic indicators, tourism statistics and social media sentiment analysis are also vital (Quattrone, G., Kusek, N., & Capra, L., 2022). The overview of all these data sources would provide us with a distinct view of Airbnb's business ecosystem.

Operational Databases:

- Transaction Databases - Contain records of all transactions that guests make, including bookings, cancellations, and payments. This data is crucial for understanding guest booking patterns and revenue flow.

- Customer Relationship Management (CRM) Systems - These databases hold information on guest interactions, service requests, complaints, and feedback. They can be analyzed to understand guests' needs and improve service quality.

- Property Management Systems - Include detailed data on the listings, such as location, amenities, availability, and occupancy rates. This data helps Airbnb analyze the performance of listings and optimize inventory management.

- User Behavior Logs - Track how users interact with the Airbnb platform, including search patterns, click-through rates, and booking funnel progression. Insights from this data can drive personalized marketing strategies.

External Data Sources:

- Market Research Data - Reports and datasets from market research firms can provide insights into travel trends, guest preferences, and competitive benchmarking.

- Economic Indicators - Information such as employment rates, GDP growth, and currency exchange rates can be used to predict travel patterns and demand for accommodations.

- Social media and Online Review Platforms - By monitoring social media and review websites, Airbnb can gauge public sentiment, track guest satisfaction, and respond to emerging trends or issues.

- Government and Tourism Boards - Data from these bodies can provide insights into travel restrictions, tourist influx, popular events, and seasonality, which all impact accommodation demand.

- Competitor Data - Information about competitors' pricing, promotions, and market share can help Airbnb in strategic positioning and pricing decisions.

- Third-Party Analytics Services - Services like Google Analytics can track and analyze web traffic, providing insights into how guests find Airbnb listings and what factors influence their booking decisions.

Regarding the next stage, to be able to accommodate the vast amount of collected data, a data storage strategy employing both a data lake and a data warehouse is recommended. For the data lake, we would employ it as a repository for raw, unstructured data, as stated by Nambiar, A., & Mundra, D. (2022), which offers flexibility for data science and machine learning endeavors. At the same time, our data warehouse would be charged with storing all processed data, making it readily accessible for analysis, reporting, and business intelligence activities; thereby making it easy to cater to the needs of non-technical stakeholders (David, M, n.d).

According to Kimball methodology, the data warehouse structure for Airbnb can be designed around a star schema, which is a type of multidimensional data model. The star schema centers on a fact table surrounded by related dimension tables as described below.

Fact Table:

Bookings: The central fact table that records measurable, quantitative data about each booking. The metrics should include `NumberOfNights`, `BookingRevenue`, `CleaningFees`, `ServiceFees`, and other foreign keys to connect to each dimension table.

Dimension Tables:

Date Dimension:
- Attributes: `DateKey`, `Day`, `Week`, `Month`, `Quarter`, `Year`. This Allows for time-based analysis and trend observation.

Guest Dimension:
- Attributes: `GuestKey`, `GuestName`, `AgeGroup`, `Gender`, `Country`, `Language`. Enables demographic analysis of the customer base.

Host Dimension:
- Attributes: `HostKey`, `HostName`, `HostRating`, `Location`, `DateJoined`. Provides insights into host-related trends and performance.

Property Dimension:
- Attributes: `PropertyKey`, `PropertyName`, `RoomType`, `Amenities`, `MaxOccupancy`, `PetFriendly`. Useful for analyzing listing-specific data.

Location Dimension:
- Attributes: `LocationKey`, `Address`, `Neighborhood`, `City`, `State`, `Country`. Allows for geographic analysis and location-based market trends.

Review Dimension:
- Attributes: `ReviewKey`, `Rating`, `ReviewText`, `SentimentScore`, `ReviewDate`. Captures details about guest reviews for sentiment and text analysis.

Price Dimension:
- Attributes: `PriceKey`, `Date`, `BasePrice`, `DiscountedPrice`, `CleaningFee`, `ExtraPersonFee`. Tracks pricing information over time for dynamic pricing analysis.

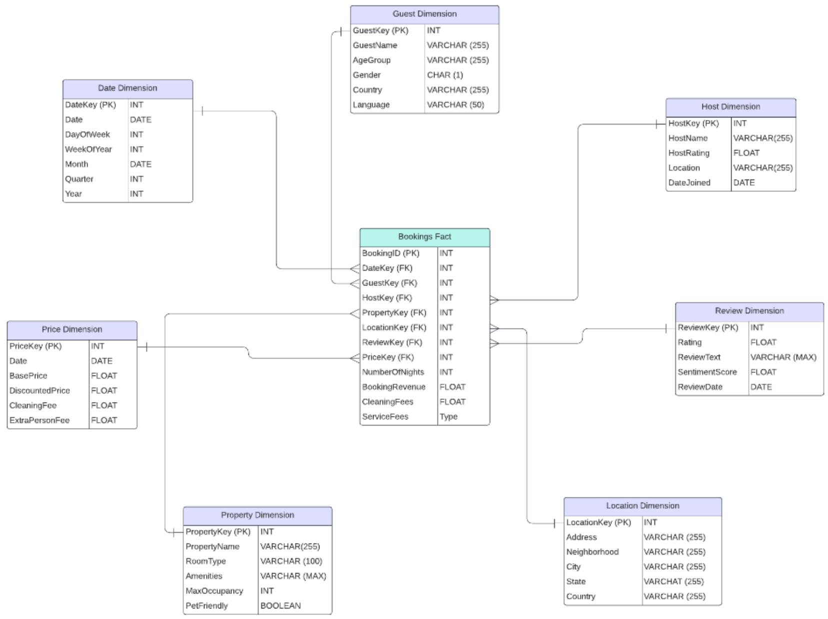

Fig: 03 – Data Structure

At the heart of our BI solution is the analytical and visualization toolkit which is designed to translate complex datasets into actionable insights. Our operational dashboard exemplifies how we can visually represent data to answer critical managerial questions. It provides a comprehensive view of guest behavior, showing metrics like total number of listings, average review scores, occupancy rates, and 5-star rating percentages broken down by room type and neighborhood. This dashboard can identify which neighborhoods have the highest and lowest ratings, correlate these ratings with median pricing, and display occupancy rates to suggest where Airbnb might focus on improving guest experiences or adjust pricing strategies. Furthermore, it gives us an insight into the volume of bookings and the average number of reviews per month, which are essential indicators of market demand and customer engagement. Additionally, when analyzing guest feedback and reviews to be able to monitor service quality and guest satisfaction levels.

A proper monitoring and analytics framework would provide Airbnb with valuable real time insights into a multitude of dimensions of their service. This directly impacts their strategic decision-making and operational effectiveness. For Airbnb, the integration of a sophisticated monitoring and analytics framework, encompassing both Guest and Pricing Analytics dashboards, is fundamental in leveraging data for strategic initiatives. The Guest Analytics dashboard would help with insights regarding guest preferences and behaviors across different neighborhoods, this in turn would inform Airbnb's market approach and enhance host support systems to directly boost guest satisfaction. This empowers Airbnb to advise hosts on creating listings that resonate with guest desires, elevating the attractiveness of their offerings.

Additionally, the Pricing Dashboard provides insights into financial metrics such as pricing trends and forecasts. This would help in assisting hosts in setting prices that balance with market competitiveness. This dashboard is pivotal for Airbnb to monitor its financial well-being, adjusting their commission structures wisely, and understand the fiscal impact of market fluctuations on its revenue.

Combined, these dashboards would help to convert extensive data into intelligible, actionable insights, which enables Airbnb to identify profitable investment areas, understand customer price sensitivity, and fine-tune pricing strategies to improve booking rates. This integrated analytics approach positions Airbnb to tackle operational and financial BI questions, considering the dynamic nature of the market. It aids in optimizing both host and guest experiences and informs financial decisions that underpin growth and service efficiency. Equipped with these data-driven tools, Airbnb is well-positioned to predict market trends and sustain its industry leadership.

However, the effectiveness of a BI solution is not solely dependent on its analytical effectiveness, but also how it adheres to stringent data governance and security protocols; Undoubtedly throughout our process we would also need to ensure data integrity which would involve establishing appropriate data processing pipelines for data cleaning, transformation, and quality assurance. This makes it important to focus on the implementation of a comprehensive data governance framework, filled with rigorous security measures like data encryption and access controls, to safeguard user privacy and ensure regulatory compliance.

Three Use Cases/Prototypes

For this project, we have developed one operational dashboard to draw insights from Guest Analytics and one tactical dashboard for Price Analytics in addition to a predictive regression analysis to identify what factors influence pricing. We will be discussing the dashboards and the data analytics method in-depth in this section.

Case Study 1 – Guest Analytics (Operational Dashboard)

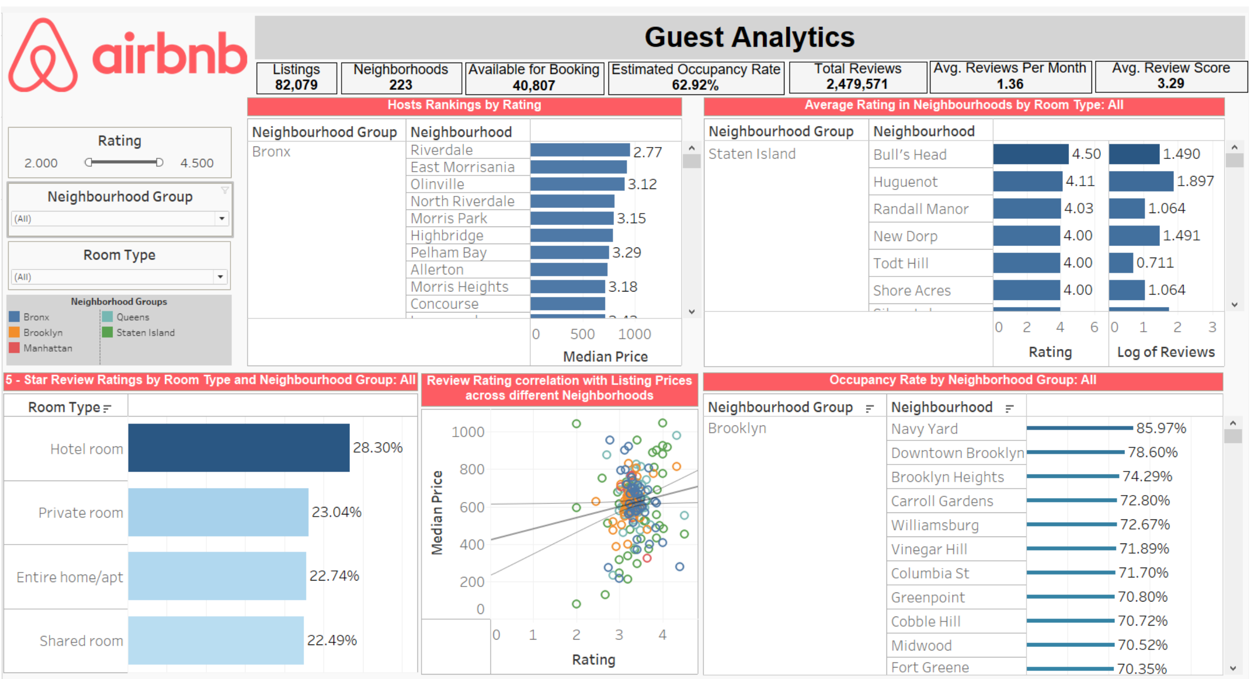

Fig: 04

This operational dashboard is tailored for Airbnb's operational oversight in New York City. The dataset utilized for this project provides multidimensional insights into Airbnb's listings, bookings, and customer feedback across different neighborhoods and room types.

The analytical review of Airbnb's Guest Analytics stresses the intrinsic value of guest satisfaction and its influence on pricing strategies within the New York City market. The dashboard highlights a clear correlation between guest ratings and median listing prices, indicating that higher-rated accommodations command higher prices and are likely to attract more bookings. Additionally, the occupancy rates across various neighborhoods present valuable insights; for instance, areas such as Navy Yard exhibit remarkably high occupancy, signaling strong demand for Airbnb services in these locales. Such analytics empower Airbnb to guide hosts towards improving their offerings, thereby potentially elevating their ratings and justifying higher pricing tiers.

Also, the dashboard facilitates a granular examination of guest preferences, with detailed metrics such as average review scores and review volumes by room type and neighborhood. This tailored analysis is instrumental in pinpointing the precise areas for operational improvements and market opportunities. Armed with this information, Airbnb can implement targeted strategies to bolster high-performing neighborhoods while also addressing and uplifting underperforming areas. By leveraging these guest analytics, Airbnb can craft dynamic pricing models, enrich guest experiences, and optimize the allocation of marketing resources, ultimately shaping a robust, data-driven approach to marketplace competitiveness and revenue growth.

Intended Users - The dashboard is designed for Airbnb managers and analysts who are responsible for monitoring and optimizing the guest experience. It may also serve as an informative tool for hosts who are keen on understanding the competitive landscape and performance within their neighborhood groups.

The filters designed with this dashboard are a range of rating scores, neighborhood groups, and room types. These filters can be used to show different scenarios across all performance metrics. We also have color highlights to differentiate between each neighborhood group on the price correlation metric.

<u>Delving Deeper:</u>

Listings	Neighborhoods	Available for Booking	Estimated Occupancy Rate	Total Reviews	Avg. Reviews Per Month	Avg. Review Score
82,079	**223**	**40,807**	**62.92%**	**2,479,571**	**1.36**	**3.29**

Fig: 05

The top bar of the dashboard displays key performance metrics relevant to the operational management at Airbnb. Each metric contributes to a holistic understanding of Airbnb's operational dynamics in New York City. They allow managers to assess the scale of operations, guest engagement and satisfaction, and market penetration, all of which are essential for informed decision-making.

- Listings (82,079) - This metric represents the total number of properties listed on Airbnb within New York City. It answers the following questions - How many accommodations are available for Airbnb users? And how large is the Airbnb market in New York City?

- Neighborhoods (223) - This indicates the number of distinct neighborhoods in New York City where Airbnb listings are available. It answers the following questions - How extensive is Airbnb's reach within the city? Are there opportunities to expand into more neighborhoods?

- Available for Booking (40,807) -This shows the number of listings available for booking, providing insight into the current inventory open to guests. It answers the following questions - What proportion of the total listings are available for guests? Can customer demand be met with the current availability?

- Estimated Occupancy Rate (62.92%) - This percentage gives an estimate of how frequently listings are occupied, which can be used as a proxy for demand. It answers the question - How often are Airbnb properties occupied? And what is the utilization rate of the listings?

- Total Reviews (2,479,571) – This shows the aggregate number of reviews which provides a quantitative measure of guest interactions and experiences. It answers the following questions - How engaged are guests with the Airbnb community? To what extent are guests providing feedback on their stays?

- Avg. Reviews Per Month (1.36) – This reflects the average frequency of reviews per month, which can indicate the level of booking activity and guest engagement. It answers the following questions - How frequently do guests leave feedback? Is there a correlation between the number of bookings and the reviews received?

- Avg. Review Score (3.29) - This is an average score of the ratings provided by guests, which can be an indicator of overall guest satisfaction. It answers the following questions - How satisfactory are guest experiences on average? What is the quality level of Airbnb accommodation as perceived by guests?

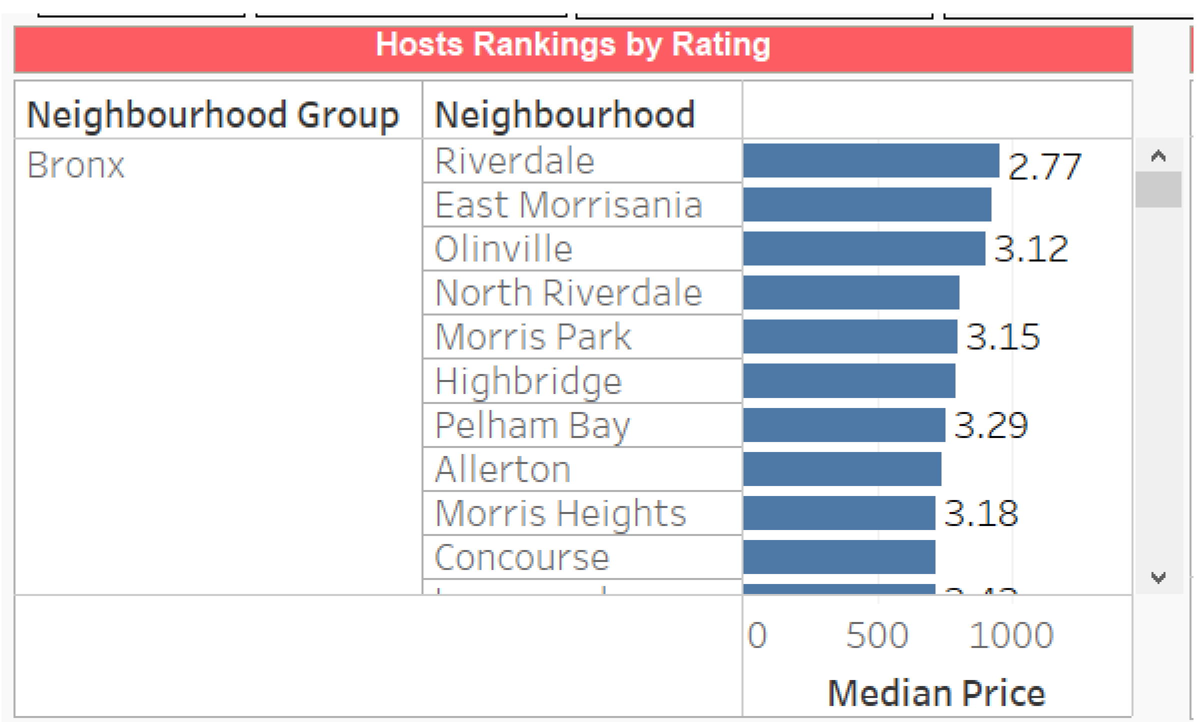

Fig: 06

This performance metric illustrates a bar chart titled "Hosts Rankings by Rating". It shows the neighborhood group, neighborhood, average review score, and the median price of Airbnb listings across each neighborhood.

It answers the following questions:

- Which neighborhoods in each neighborhood group are considered most attractive to guests based on the rating they receive?
- Does a higher rating correlate with a higher median price, indicating that guests may be willing to pay more for higher-rated accommodations?
- Are the prices for listings in certain neighborhoods in line with the perceived quality (as indicated by the rating)?
- Could there be an opportunity to adjust pricing in certain neighborhoods to better reflect the value guests derive from their stay?
- For potential Airbnb hosts or investors, which neighborhoods might offer the most potential for return on investment, considering both the rating and the median price?

How it answers the questions:

- The median price is indicated numerically and represented visually by the length of the bars, making it easy to compare across neighborhoods.
- The rating provides a snapshot of guest satisfaction, enabling a quick assessment of how well hosts in each neighborhood are performing from the guests' perspective.

By applying different filters, users can conduct a customized analysis to compare neighborhoods within different groups, across room types or within score ranges. This targeted approach allows for a more granular understanding of market dynamics. The filters enable managers to isolate data relevant to specific operational questions. This helps in making precise adjustments to strategies related to pricing, marketing, and property improvements.

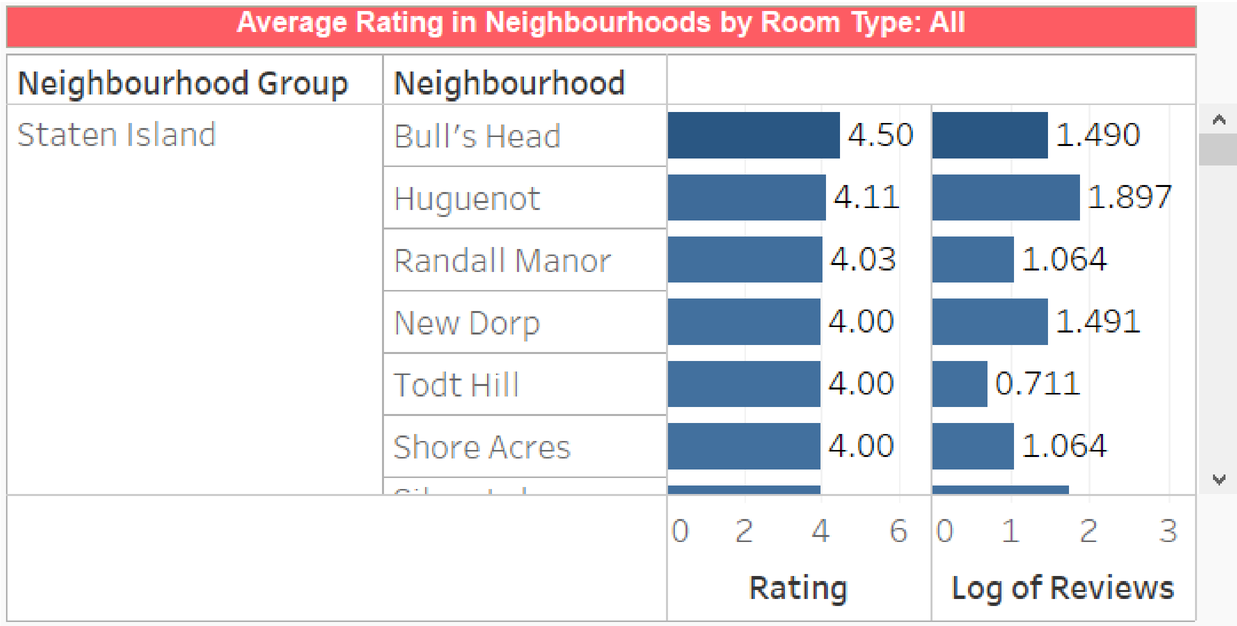

Fig: 07

The performance metric above is a bar chart detailing "Average Rating in Neighborhoods by Room Type". This visualization ranks neighborhoods based on average guest ratings and it provides a logarithmic scale representation of the number of reviews.

Note: we used the logarithmic scale for the number of reviews in this scenario because of the huge disparity between the number of reviews received per neighborhood in comparison to the rating score.

It answers the following questions:

- Which neighborhoods in each neighborhood group are rated highest in comparison to the number of reviews by guests, indicating higher satisfaction levels?
- Is there a significant variation in satisfaction across different neighborhoods?
- How does the volume of guest feedback (number of reviews) correlate with the average ratings?
- Do neighborhoods with higher ratings also have a higher volume of reviews, suggesting a positive guest experience?

How it answers these questions:

- The average rating provides a direct measure of guest satisfaction, and the bars allow for easy comparison between neighborhoods.
- The logarithmic scale for the number of reviews helps to visualize the review volume data across each neighborhood, showing whether a higher volume of feedback correlates with higher ratings.

If a room type filter is applied, it allows users to isolate the data to see how the average ratings differ across various accommodation types within the neighborhoods. For example, users can compare how entire homes/apartments are rated within a particular neighborhood. Also, by selecting different neighborhood groups, users can conduct comparative analysis across various parts of each neighborhood group in New York City.

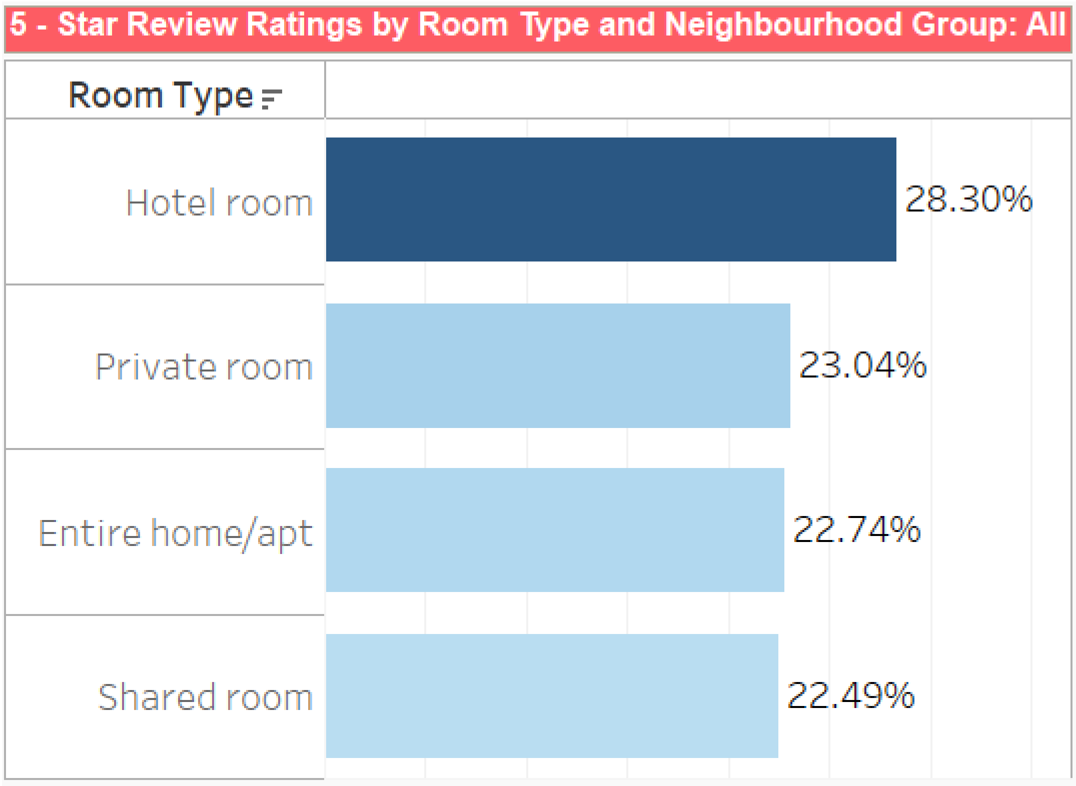

Fig: 08

This performance metric displays a horizontal bar chart detailing "5-Star Review Ratings by Room Type and Neighborhood Group" with a focus on various types of accommodations. This metric is instrumental for Airbnb management in understanding guest satisfaction trends, improving service quality, and refining offerings to match guest preferences.

It answers the following questions:

- Which type of room (Hotel room, Private room, Entire home/apt, Shared room) garners the most 5-star reviews, indicating the highest level of guest satisfaction?
- How does the preference for room types differ, and what does this imply about guest expectations or experiences?
    - How does the quality of experience, as perceived by guests, vary across different room types?

- Are certain types of rooms consistently delivering superior experiences that warrant a 5-star rating across each neighborhood group?

How it answers the questions:

- The percentage of 5-star reviews for each room type is clearly displayed alongside each bar, providing an immediate visual representation of the data.
- The lengths of the bars facilitate a direct comparison of the performance between different room types regarding guest satisfaction.

By filtering for different neighborhood groups, users can explore whether the pattern of 5-star reviews is consistent across different areas. This could highlight geographical variations in guest satisfaction and indicate if certain areas have higher standards or different guest expectations.

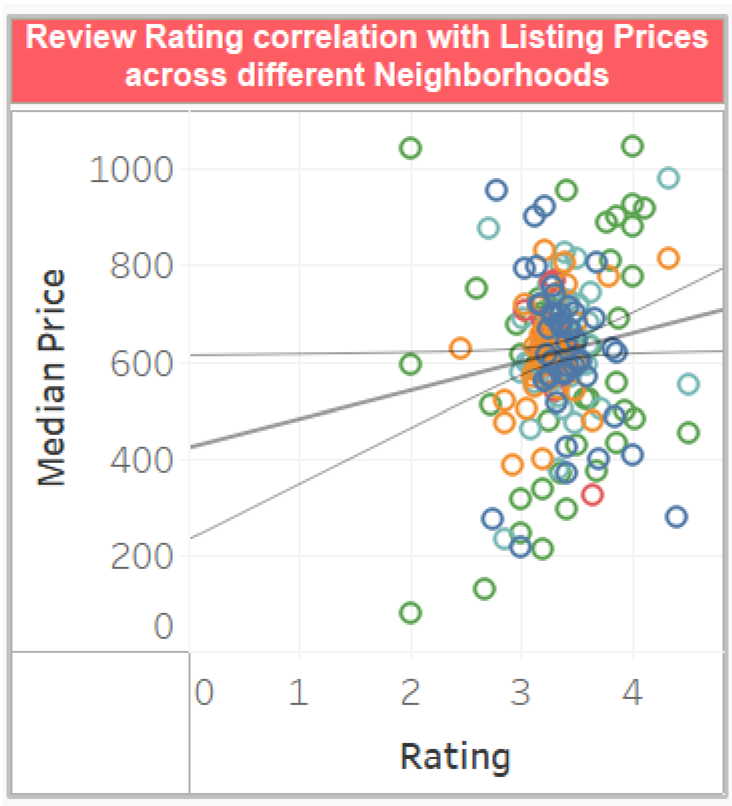

Fig: 09

This performance metric depicted is a scatter plot titled "Review Rating correlation with Listing Prices across different Neighborhoods." It is used to analyze the relationship between guest ratings and the median price of Airbnb listings. This scatter plot serves as a crucial tool for discerning the relationship between guest satisfaction and pricing, with filters enhancing the granularity and relevance of the insights for different market segments.

It answers the following questions:

- Is there a positive correlation between the price of listings and their ratings, suggesting that higher prices are associated with higher guest satisfaction?
- Are there any anomalies or outliers that suggest some listings are overpriced or underpriced relative to their ratings?
- Can higher listing prices be justified by correspondingly higher guest ratings?

How it answers these questions:

- Each point on the scatter plot represents an Airbnb listing, with the position along the x-axis showing the rating and along the y-axis showing the median price.
- The trend line gives a visual indication of the overall relationship (correlation) between the two variables across different neighborhoods.

Users may apply the neighborhood group filter on the scatter plot to see if the correlation holds true across various neighborhoods or to identify neighborhood-specific pricing strategies. This could also reveal if certain neighborhoods are known for better value for money in terms of price-to-satisfaction ratios.
Also, if the filter for room types is used, it can offer insights into whether certain types of rooms (like private rooms or entire homes) have a different price-to-rating correlation. This can be useful for hosts to understand how to price their listings competitively based on room type.

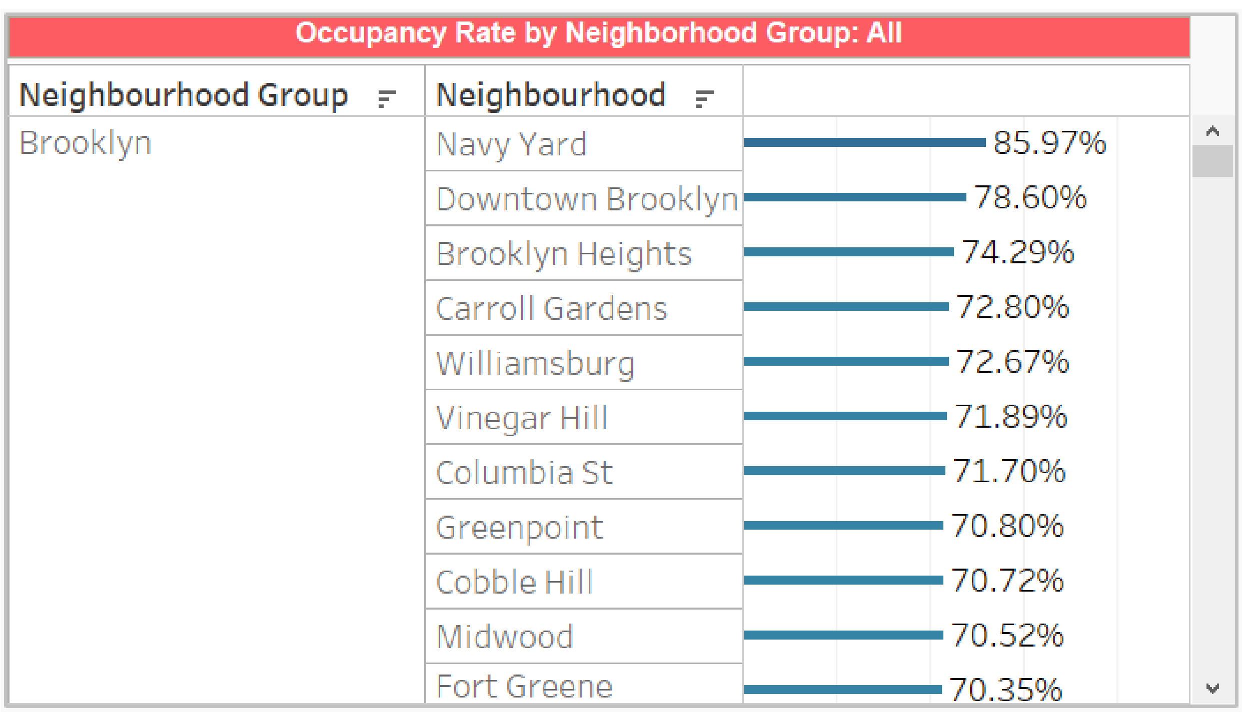

Fig: 10

The performance metric shown is a bar chart detailing the "Occupancy Rate by Neighborhood" within each neighborhood group for Airbnb listings. This metric offers valuable insights into the market demand across different neighborhoods.

It answers the following questions:

- Which neighborhoods in each neighborhood group have the highest occupancy rates, potentially indicating a higher demand for Airbnb listings in those areas?
- How does the popularity of neighborhoods differ in terms of accommodation?
- In which neighborhoods should Airbnb host or the platform itself consider investing more resources to meet higher demand?
- Are there areas with lower occupancy rates that may benefit from marketing efforts to increase their visibility and attractiveness to guests?
- Could the neighborhoods with higher occupancy rates justify higher pricing due to higher demand?
- In neighborhoods with lower occupancy rates, are price reductions or increased amenities needed to attract more guests?

How it answers these questions:

- The occupancy rates are represented by the length of each bar, allowing for quick visual comparison between different neighborhoods within each neighborhood group.
- The numerical value of the occupancy rate next to each bar provides a precise measure of how often listings are booked in each neighborhood.

Adjusting the neighborhood group filter would enable users to view occupancy rates in different neighborhood groups, which could help in comparing demand across various parts of the city. This is useful for understanding broader market dynamics and could influence decisions on where to focus operational efforts.
Filters for room types further refines the data to show how occupancy rates vary not just by location, but also by the type of accommodation.

Case Study 2 – Price Analytics (Tactical Dashboard)

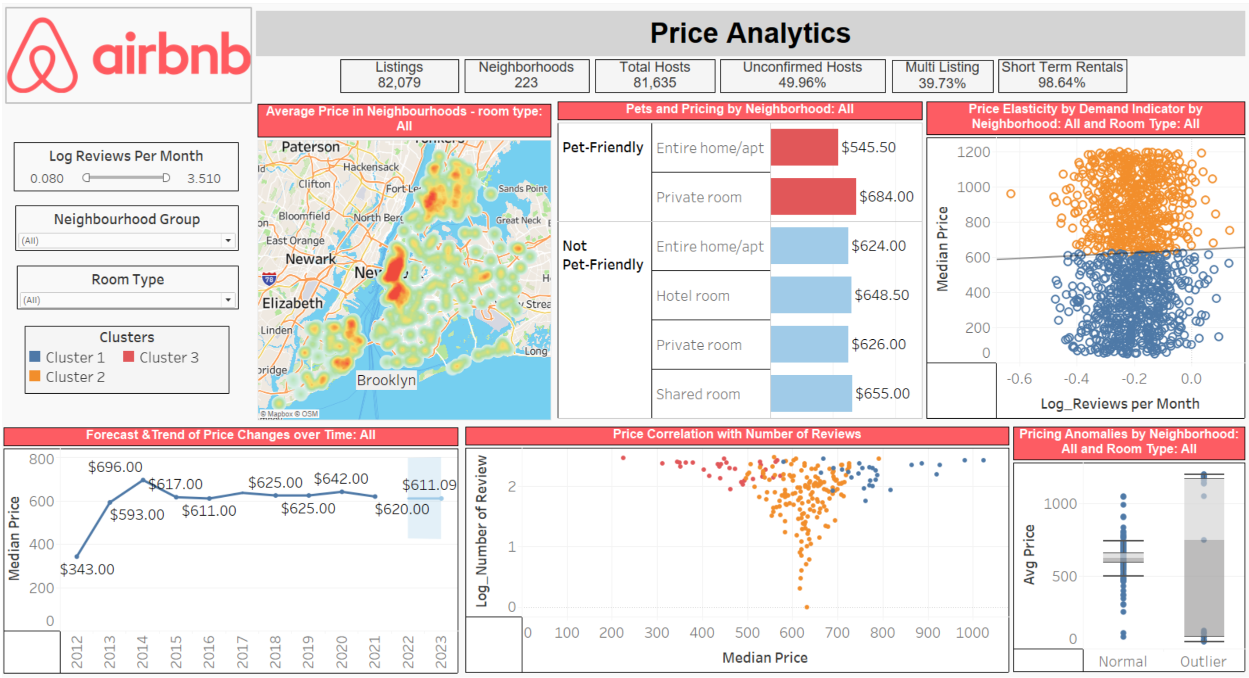

Fig: 11

This tactical dashboard is centered around "Price Analytics" within Airbnb's operations in New York City. The dataset utilized for this project provides multidimensional insights into Airbnb's listings and pricing across different neighborhoods and room types in New York City. In general, the tactical dashboard provides a comprehensive view of Airbnb's pricing dynamics, supporting tactical decision-making through rich visualizations and actionable insights derived from the data.

The dashboard offers an at-a-glance view of critical metrics such as listing counts, neighborhood coverage, host details, and the nature of rentals, which are vital for tactical decision-making. The dashboard's heatmap, displaying average prices across neighborhoods, allows for a quick visual assessment of market hotspots. This geospatial pricing data, alongside the detailed breakdown of pricing by pet-friendliness and room type, equips Airbnb with the tools to adapt their pricing strategies dynamically. The insights gained from pet-friendly accommodations indicate a distinct market segment that could command premium pricing, reflecting the added value guests place on such amenities.

Further depth is provided by the trend analysis of median prices over time, showing how rates have evolved and are projected to continue, which is essential for forecasting and setting future pricing policies. The scatter plots elucidate the relationship between pricing, guest reviews, and demand elasticity, demonstrating the intricate balance between price setting and market response. Such visualizations highlight areas where the pricing could be optimized, based on guest feedback and booking patterns. Meanwhile, the detection of pricing anomalies using a box plot underscores the need for vigilance against pricing inconsistencies, which could either represent untapped opportunities or risks of misalignment with guest expectations. Altogether, this dashboard offers actionable intelligence, aiding Airbnb in refining its competitive positioning and nurturing its growth in the sharing economy.

Intended Users - This dashboard is intended to support mid-level management in making decisions related to pricing strategy, competitive positioning, and market analysis. Managers can utilize this dashboard to make informed decisions on pricing adjustments, and policy implementations (like pet-friendliness) based on current trends and historical data. By examining pricing trends and elasticity, management can better understand how changes in price might affect demand, allowing for more strategic planning around rate setting.

The dashboard also features several filters—Log Reviews Per Month, Neighborhood Group, Room Type, and Clusters—which provide a means to segment and drill down into the data for a more detailed analysis. By adjusting these filters, users can view specific subsets of the data, compare different segments, and identify trends or outliers in pricing strategies.
<u>Delving Deeper:</u>

Fig: 12

These metrics serve as critical data points for assessing Airbnb's market presence in New York City. Each metric is described as follows:

- Listings (82,079) – This indicates the total number of active listings on the platform, and it answers the question - How extensive is Airbnb's inventory in New York City? What is the scale of Airbnb's offerings?

- Neighborhoods (223) – This shows the number of neighborhoods where Airbnb has listings. It answers the question - How widespread is Airbnb's market penetration across different neighborhoods in New York City? Are there areas of untapped market potential?

- Total Hosts (81,635) – This represents the number of hosts on the platform. It answers the question - How many individuals or entities are providing services as hosts? This helps in understanding the size of the hosting community.

- Unconfirmed Hosts (49.96%) – This shows the percentage of hosts who have not yet had their account details confirmed. It answers the question - What proportion of the host community requires further verification? This could be an indicator of potential risk or areas where trust could be improved.

- Multi Listing (39.73%) -This indicates the percentage of hosts that have multiple listings, which may suggest professional or semi-professional hosts versus casual ones. It answers the question - How many hosts are operating on a larger scale with multiple properties? This can impact strategies for host engagement and support.

- Short Term Rentals (98.64%) – This provides the percentage of listings that are available for short-term rental as opposed to longer-term leases. It answers the question - What is the nature of most Airbnb listings in terms of rental duration? This helps in understanding the core business model and customer behavior.

Note: According to Airbnb, short term rentals are a guests stay of 30 days or less.

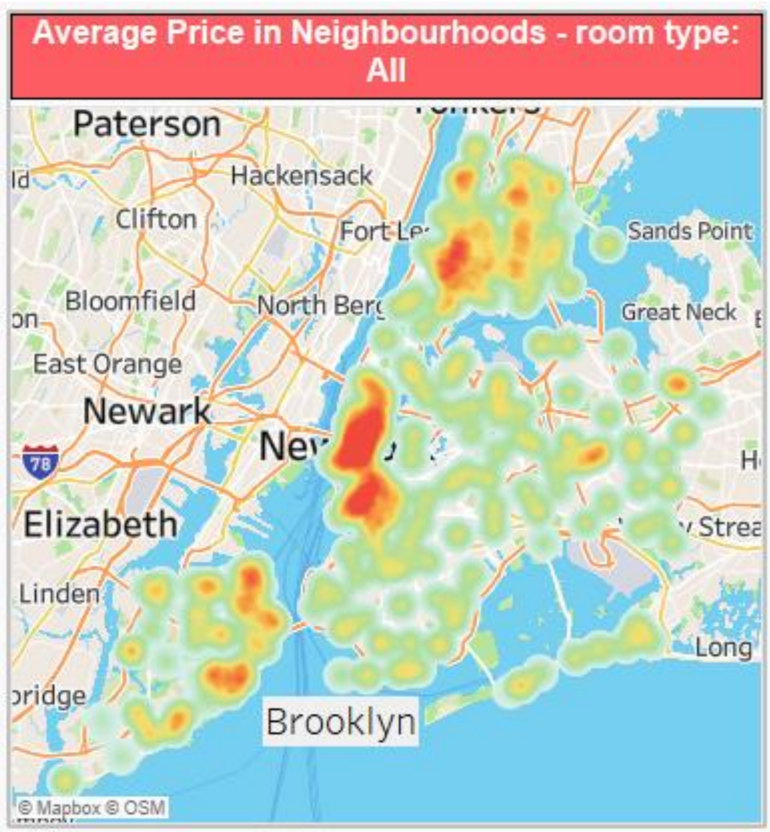

Fig: 13

The heatmap depicted on the dashboard visualizes the "Average Price in Neighborhoods" for Airbnb listings across New York City and can be filtered across various room types. The heatmap serves as a strategic tool for visualizing and analyzing spatial price distribution.

It answers the following questions:

- Which neighborhoods have higher or lower average listing prices? This provides a visual overview of pricing trends across the city, which can be crucial for pricing strategy.
- Where are the hotspots of high pricing within the city?
- Are there neighborhoods with lower average prices that could represent opportunities for new hosts or for market penetration strategies?

How it answers the questions:

- The varying colors on the heatmap provide immediate visual cues on the geographic distribution of prices. Warmer colors typically indicate higher prices, while cooler colors indicate lower prices.
- The intensity of the color can show the concentration of listings at different price points, with deeper shades indicating more pronounced pricing levels.

Adjusting the room type filter would change the data displayed on the heatmap to show average prices for specific types of rooms, such as private rooms, shared rooms, or entire homes/apartments. This allows for a more granular analysis of pricing strategies for different accommodation types.
Also, the neighborhood group filter could provide a way to narrow down the heatmap to a specific neighborhood group, which would be beneficial for localized analysis and strategy.

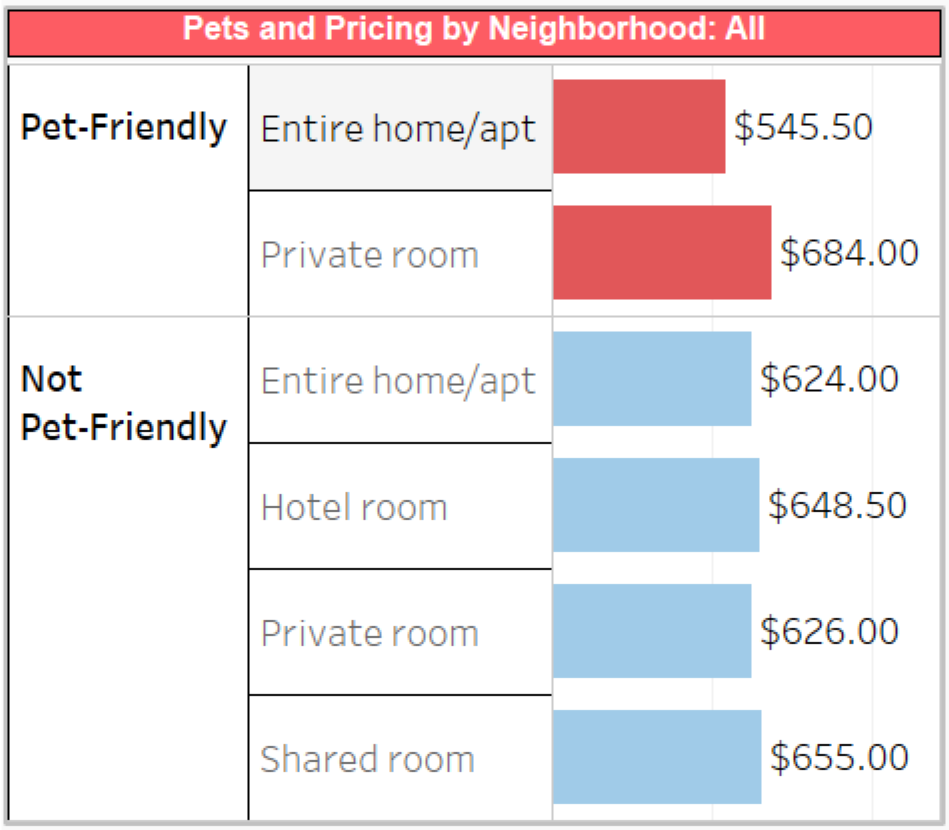

Fig: 14

This bar chart titled "Pets and Pricing by Neighborhood" compares the median prices for Airbnb listings that are pet-friendly against those that are not, across various types of accommodations. This performance metric is key for understanding the value that guests place on pet-friendly accommodations and assists in pricing strategy development for Airbnb listings.
It answers the following questions:

- How does being pet-friendly affect the pricing of Airbnb listings?

- Are guests paying more for the option to bring pets along in different types of accommodations?
- Is there a premium on the price of accommodations that allow pets?
- Can a correlation be inferred between the type of room and the added value of being pet-friendly?

How it answers these questions:

- The bars represent median prices for each type of room and are color-coded to distinguish between pet-friendly and not pet-friendly accommodations.
- The comparison can show if there is a consistent price differential associated with pet-friendly listings, which could suggest a premium for such accommodation.

By filtering the data by neighborhood group, users can see how the pricing for pet-friendly vs. non-pet-friendly accommodations varies across different areas. This might reveal whether some neighborhoods have a higher demand for pet-friendly accommodations, reflected in higher prices. Also, Adjusting the room type filter could isolate the price comparison to just one type of room (e.g., only entire homes/apartments), allowing for a more focused analysis. This might be useful to assess if the pet-friendly price premium is more pronounced in certain types of listings.

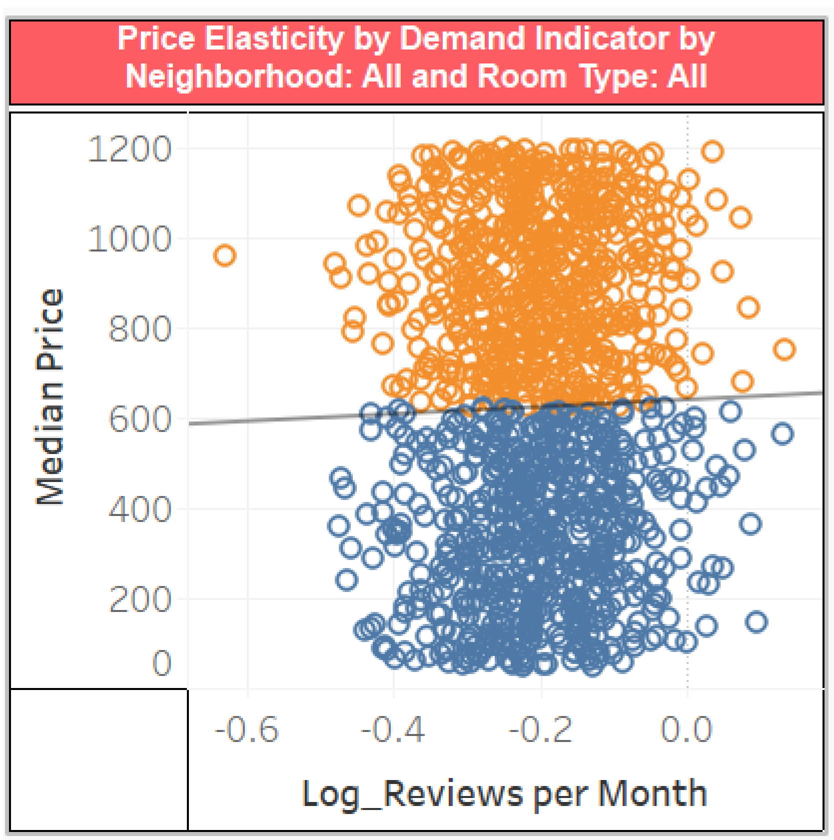

Fig: 15

The scatter plot titled "Price Elasticity by Demand Indicator by Neighborhood: All and Room Type: All" examines the relationship between the number of reviews per month (a proxy for demand) and median listing prices, with data aggregated across different neighborhoods and types of accommodations.

It answers the following questions:

- How sensitive is the demand for Airbnb listings to changes in prices?
- Are there indications that higher or lower prices affect the popularity of listings, as measured by reviews per month?
- What are the pricing levels at which listings seem to achieve a balance between demand (reviews) and profitability (price)?
- Do certain price points lead to a significant increase or decrease in demand?
- Are there any noticeable trends or patterns that suggest a 'sweet spot' for pricing?

How it answers the questions:

- The scatter plot uses the logarithmic value of reviews per month on the x-axis and the median price on the y-axis, with each dot representing a unique data point.
- The distribution of dots can suggest whether there is a correlation between demand and price—dots clustered together indicate a consistent trend, while a more dispersed pattern implies less correlation.

Applying the neighborhood group filter would enable an examination of the price-demand relationship within specific neighborhoods, allowing for local market analysis. Different neighborhood groups may show varying elasticity patterns due to unique local dynamics.

Filtering by room type (e.g., entire home/apt, private room) could reveal differences in price sensitivity between different accommodation types. This helps in understanding which types of rooms might benefit from strategic pricing adjustments to optimize occupancy and revenue.

By understanding elasticity, Airbnb and its hosts can make informed decisions about pricing strategies, such as when to increase prices without significantly affecting demand or when to decrease prices to attract more bookings.

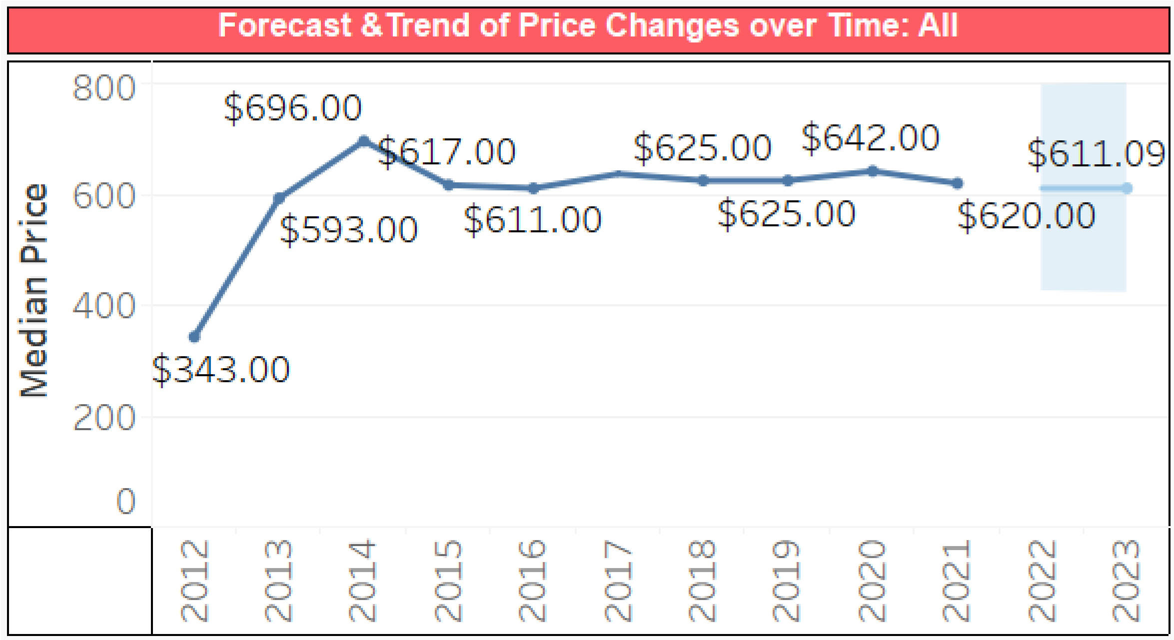

Fig: 16

The line chart titled "Forecast & Trend of Price Changes over Time: All" represents the historical and projected median prices for Airbnb listings over a span of years. The chart also shows the forecast of future prices, which is crucial for planning and strategy development.
It answers the following questions:

- What has been the trend in the median price of Airbnb listings over the past years?
- This metric can identify whether prices have been rising, falling, or remaining stable over time.
- Based on historical data, what is the expected trend in pricing for the near future?

How it answers these questions:

- The chart plots median price points across a timeline, showing the trajectory of price changes. This visual representation makes it easy to grasp the overall trend at a glance.
- Projections into the future (represented by a different line and color) are based on the observed historical trends and can suggest what prices might be expected if current patterns continue.

Using the neighborhood group and room type filters will enable users to view pricing trends for specific areas or types of accommodations, providing more targeted insights for those segments.

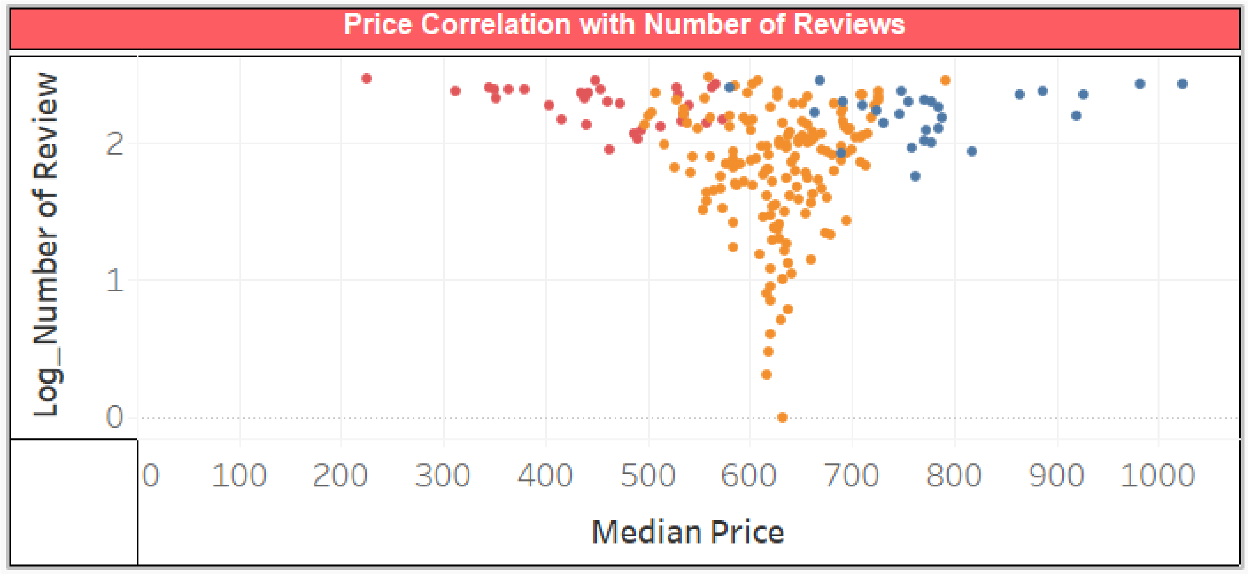

Fig: 17

The scatter plot titled "Price Correlation with Number of Reviews" examines the relationship between the median price of Airbnb listings and the number of reviews they receive, which is a common proxy for demand or popularity. This performance metric helps stakeholders understand how pricing strategies might influence guest engagement.

It answers the following questions:

- Does a higher median price of a listing correlate with more reviews?
- Are more expensive listings less popular, or do they still attract a significant number of reviews?
- Are guests more inclined to leave reviews for higher-priced listings, which might indicate a higher expectation for value?
- Alternatively, are lower-priced listings receiving more reviews due to higher occupancy and turnover?
- Is there an observable price point that maximizes guest engagement, as measured by the number of reviews?
- Where is the concentration of reviews the highest in relation to price, which might indicate an optimal pricing strategy?

How it answers these questions:

- The plot displays individual listings or aggregated data points with their median price on the x-axis and the logarithmic scale of the number of reviews on the y-axis.
- The dispersion of the dots indicates the strength of the correlation; a clear upward or downward trend suggests a strong correlation, while a dispersed cloud of dots suggests a weak or non-existent correlation.

Filtering by neighborhood could reveal how local market conditions affect the price-review relationship.

Room type filters would allow for an analysis of how this correlation varies across different types of accommodations (e.g., entire homes vs. private rooms).

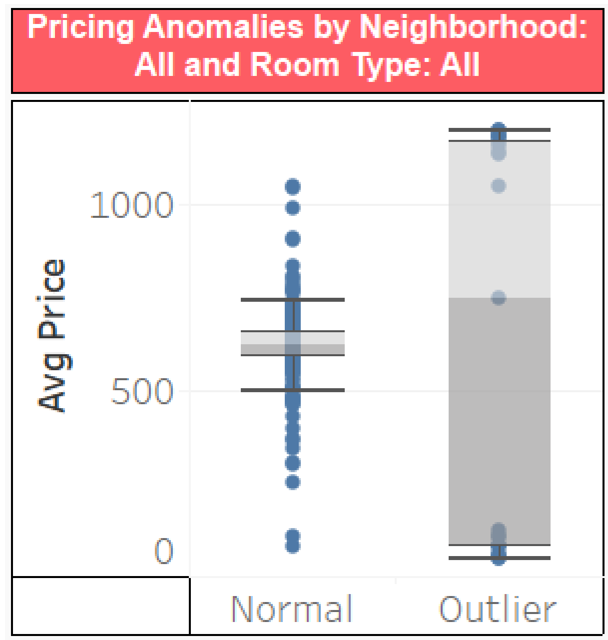

Fig: 18

The box-and-whisker plot is titled "Pricing Anomalies by Neighborhood: All and Room Type: All" alongside a separate set of points categorized as "Outlier." This visualization identifies the distribution of average prices for Airbnb listings and highlights prices that are significantly higher or lower than the norm. It is instrumental in recognizing pricing inconsistencies within the Airbnb market and can signal areas where pricing strategies need to be evaluated or adjusted. It also provides a clear visual means of distinguishing between typical and atypical pricing, guiding tactical pricing decisions.

It answers the following questions:

- Which listings have prices that deviate significantly from the average within specific neighborhoods or room types?
- Are there listings that are priced unusually high or low, indicating potential errors or unique value propositions?
- Is the pricing strategy across different neighborhoods and room types generally consistent, or are there significant anomalies?
- What proportion of listings might be underpriced or overpriced compared to the majority?
- How consistent are listing prices within and across neighborhoods and room types?
- Which listings might be affecting the perceived average market rate due to their anomalous pricing?

How it answers the questions:

- The main section of the plot (labeled "Normal") shows the range of prices where most listings fall (using quartiles and median), indicating the typical pricing structure.
- The "Outlier" section represents listings with prices that fall outside the typical range, which may require further investigation to determine the cause.
- The horizontal line within the "Normal" box indicates the median price, while the top and bottom of the box represent the interquartile range, showing where the middle 50% of the data lies.

Applying the neighborhood group filter would allow users to see pricing anomalies within specific neighborhoods, helping to identify if certain areas have more pronounced pricing issues. Also, the room type filter would segment the data by room type, potentially revealing which types of accommodations are more prone to pricing anomalies.
Understanding pricing anomalies can inform Airbnb management about necessary adjustments to avoid pricing listings too high (which may reduce bookings) or too low (which may undercut potential revenue).

Data Analytics – List Price Prediction Leveraging Machine Learning

Predicting listing prices on Airbnb is pivotal for optimizing dynamic pricing strategies, ensuring hosts maximize their revenue potential while maintaining market competitiveness. It provides foresight for strategic planning and resource allocation, enhances revenue management, and supports informed decision-making regarding promotions and inventory management. Accurate price predictions also contribute to customer satisfaction by ensuring guests receive fair value, fostering trust and loyalty to the platform. Moreover, it helps Airbnb to adapt to market shifts and comply with varying regional regulations, underpinning Airbnb's growth and adaptability within the fast-paced sharing economy.

Leveraging machine learning techniques, we explore the dataset's features and employ regression algorithms to develop predictive models. Our goal is to provide detailed insights into the factors influencing Airbnb listing prices and evaluate the performance of different regression algorithms.

Data Preprocessing:

The analysis began with data preprocessing to ensure the dataset's cleanliness and relevance for model training. Upon loading the dataset, we observe various columns containing missing values and irrelevant information. To address this, we perform thorough data cleaning steps, including dropping unnecessary columns and handling missing values appropriately. Additionally, we remove special characters like '$' and ',' from 'price' and 'service fee' columns and convert them to float data type.

Model Training:

Following data preprocessing, we proceeded to train regression models using three distinct algorithms: Linear Regression, Decision Tree Regression, and Naive Bayes. Before training, we preprocessed the data further by scaling numeric features and encoding categorical features using StandardScaler and OneHotEncoder, respectively. We split the dataset into training and testing sets to evaluate model performance effectively.

Evaluation and Analysis:

The models are evaluated using Mean Squared Error (MSE) as a metric to quantify the difference between predicted and actual prices. This evaluation provides valuable insights into the predictive power of each feature and algorithm combination. Notably, certain features such as 'service fee' and 'room type' exhibit lower MSE, suggesting their significant impact on listing prices. Additionally, we observe differences in performance across algorithms, with Linear Regression and Decision Tree Regression outperforming Naive Bayes in predicting listing prices accurately.

```
Algorithm Used:  Linear Regression
                         Feature            MSE
24                   service fee       1.989963
18                     room type  108746.641971
27                minimum nights  108749.103541
42              availability 365  108752.798123
36            review rate number  109358.773199
39  calculated host listings count  109438.642012
30             number of reviews  109779.022612
12              instant_bookable  110077.083108
21             Construction year  110166.561876
3                  neighbourhood  110252.340302
================================================
Algorithm Used:  Decision Tree
                         Feature            MSE
25                   service fee       1.985475
19                     room type  108739.904704
28                minimum nights  108772.368671
43              availability 365  109137.222445
37            review rate number  109371.429539
40  calculated host listings count  109481.634598
31             number of reviews  110006.253332
13              instant_bookable  110076.544158
22             Construction year  110155.920026
4                  neighbourhood  110252.340302
================================================
Algorithm Used:  Naive Bayes
                         Feature            MSE
26                   service fee       4.135930
41  calculated host listings count  137752.411816
8                            lat  150644.289159
23             Construction year  174004.302252
35             reviews per month  192343.718347
5                  neighbourhood  192826.124180
17           cancellation_policy  202727.144900
29                minimum nights  234001.765298
2            neighbourhood group  239735.031899
38            review rate number  261049.455617
================================================
```

Fig: 19

Discussion:

The analysis uncovers key insights into the determinants of Airbnb listing prices and the effectiveness of regression algorithms in capturing pricing dynamics. Features such as 'service fee' and 'room type' emerge as critical predictors, indicating their influence on pricing decisions. Moreover, the choice of algorithm significantly impacts model performance, emphasizing the importance of selecting appropriate algorithms for predictive tasks.

In conclusion, our comprehensive analysis provides valuable insights into the factors driving Airbnb listing prices and the performance of regression algorithms in predicting them. By leveraging machine learning techniques, stakeholders can make informed decisions to optimize listing prices and enhance profitability. Future research avenues could explore advanced modeling techniques and incorporate additional data sources to further improve predictive accuracy.

PS: The python code for the predictive analytics is attached to the appendix B.

Implementation

Airbnb operates in multiple countries and caters to millions of users worldwide. With its vast network of hosts and guests, Airbnb aims to leverage Business Intelligence (BI) to enhance decision-making, optimize user experience, and sustain growth. This section outlines the BI implementation process tailored for Airbnb, focusing on managerial, technical, and ethical considerations.

In the context of Airbnb's operations, BI implementation holds immense potential to drive efficiency and enhance user satisfaction. One prominent application of BI involves utilizing data analytics for demand forecasting in different markets. By analyzing historical booking data, market trends, and seasonal patterns, BI tools can provide insights into future demand for accommodations. This enables Airbnb to optimize pricing strategies, allocate resources effectively, and tailor its offerings to meet the evolving needs of hosts and guests. Additionally, BI can play a crucial role in optimizing the guest experience by analyzing user feedback, monitoring property performance metrics, and identifying areas for improvement in listing quality and customer service.

Furthermore, BI enables Airbnb to monitor property performance metrics, analyze guest feedback, and identify areas for improvement in listing quality and customer service. Airbnb should implement an organizational change like those mentioned by Kotter's Eight Steps. Currently, there are challenges such as a significant percentage of booking cancellations and negative guest feedback. Our proposed BI system would help identify and implement changes to improve these metrics. The first step is to establish a coalition comprising the Head of Customer Experience and the Head of Product who would spearhead the initiative for creating these changes. This change is aimed at ensuring consistent monitoring of bookings and guest reviews. The next step is to form a committee to oversee the dashboards, involving middle management from both the Customer Experience and Product departments. With middle management onboard, the organizational changes can be effectively communicated to their respective teams. Subsequently, the dashboards would commence their operations, enabling adjustments to increase booking completion rates and address issues leading to negative guest feedback. The dashboards would serve as decision-making tools. The final steps would involve integrating other departments and aspects of the Airbnb platform into the system. These changes will establish a robust monitoring system that drives improvements in customer experience.

BI Implementation Process:

1. **Managerial Implications:**

Kotter's eight-step model for organizational transformation and data warehousing is instrumental for a seamless BI implementation.

Using the Kotter's eight-step model for organizational transformation:

Establish a sense of urgency	A notable percentage of Airbnb bookings, approximately (62%), face booking cancellations issues, while (25%) of guests express dissatisfaction with our products.
Form a powerful guiding coalition	Engage the Head of Customer Experience and the Head of Product at Airbnb to spearhead the change efforts.
Create a Vision	Develop a vision centered around consistent monitoring of guest reviews and booking processes within Airbnb to enhance customer experience.
Communicate the vision	Organize a committee to oversee dashboard monitoring, involving middle management from Customer Experience and Product departments to ensure alignment and commitment.
Empower others to act on the vision	Explain the necessity to teams, highlighting projected outcomes and the importance of addressing booking cancellations and negative feedback.
Plan for and create short-term wins	Aim for short-term victories such as addressing issues related to booking completion rates, address issues leading to negative feedback, and utilize dashboards for decision-making.
Consolidate improvements and produce more changes	Integrate other departments and aspects of the Airbnb platform into the system for comprehensive monitoring.
Institutionalize the new approaches	Enhance overall customer satisfaction, resulting in better average reviews and a strengthened customer experience at Airbnb, solidifying the newly established practices within Airbnb's culture.

Table 1

1.) Create a Sense of Urgency: The adage "If it ain't broke, don't fix it" dominates the culture of many organizations (Lucidchart, 2019). As competitors are swiftly adopting Business Intelligence (BI) solutions, leveraging data-driven insights to gain a competitive advantage, our company faces mounting pressure to enhance operational efficiency, optimize resource allocation, and swiftly adapt to evolving customer preferences. Delaying the implementation of BI could lead to falling behind rivals and missing out on significant growth opportunities in the market. Additionally, prompt adoption of BI is essential to proactively address potential bottlenecks and equipment failures, thereby preventing financial losses. Each day of delay translates to potential revenue loss and decreased production efficiency. One strategy to address this urgency is by implementing data marts for efficient data storage.

2.) Build a Guiding Coalition: Acknowledging the imperative nature of BI integration, pivotal stakeholders spanning top leadership, departmental heads, and influential staff members come together to establish a robust guiding coalition. This coalition embodies a broad spectrum of viewpoints, competencies, and insights, facilitating a thorough grasp of the organization's requirements and obstacles. United in purpose, they pledge to spearhead the BI implementation endeavors and foster widespread backing throughout the organization's hierarchy. "For instance, when Microsoft decided to shift its focus from software to cloud services, CEO Satya Nadella formed a coalition of leaders from various departments to drive this change." (Kotter's 8-Step Change Model: A Comprehensive Overview, n.d.).

3.) Formulate a Strategic Vision: Through collective effort, the guiding coalition formulates a strategic vision outlining the integration of BI. This vision centers on harnessing data to bolster decision-making, elevate product quality, streamline supply chain and manufacturing processes, and attain a competitive edge within the market landscape. Aligned with the company's overarching objectives, the vision underscores the pivotal role of BI in driving transformative outcomes for the organization's long-term prosperity. "For example, when General Electric wanted to shift its focus from manufacturing to digital services, they created a vision of becoming a top digital industrial company. " (Kotter's 8-Step Change Model: A Comprehensive Overview, n.d.).

4.) Communicate the Vision: An extensive communication strategy is implemented to emphasize the urgency and significance of BI adoption across the workforce. Through frequent town hall gatherings, informative newsletters, and digital channels, the BI vision is effectively communicated. Employees gain insight into the beneficial implications of BI for their respective roles and its overarching impact on organizational triumph. This approach fosters a collective comprehension and dedication to advancing the BI initiative throughout the company. "For instance, when Procter & Gamble wanted to change their culture to be more innovative, they used various communication methods such as town hall meetings, newsletters, and videos to reach employees at all levels." (Kotter's 8-Step Change Model: A Comprehensive Overview, n.d.).

5.) Empower Employees to Act: Acknowledging the pivotal role of employees in driving effective BI implementation, the organization commits to providing resources, cutting-edge tools, and comprehensive training initiatives to elevate their data literacy and proficiency in utilizing BI tools. By nurturing a supportive atmosphere, employees are empowered to contribute actively, sharing valuable ideas and insights, thereby cultivating a culture characterized by innovation and collaborative effort. "For example, when IBM wanted to move from traditional products to cloud-based solutions, they provided extensive training and resources for employees to develop new skills." (Kotter's 8-Step Change Model: A Comprehensive Overview, n.d.).

6.) Generate Short-term Wins: The organization strategically selects and prioritizes short-term BI projects aimed at delivering swift benefits such as enhancing inventory management, refining production efficiency, or optimizing distribution channels. Upon successful execution of these initiatives, the achievements are celebrated company-wide, fostering high morale and highlighting the tangible contributions of BI to operational excellence. "For instance, when Airbnb wanted to expand its offerings from just home rentals to experiences, they celebrated their first 1000 bookings in just three months." (Kotter's 8-Step Change Model: A Comprehensive Overview, n.d.).

7.) Consolidate Improvements: Leveraging the momentum from early achievements, the organization expands BI initiatives across diverse departments, integrating data between departments and adopting advanced analytics tools. Persistent challenges and resistance are addressed through continuous communication and guidance from the guiding coalition. The company consolidates improvements by institutionalizing BI practices into standard operating procedures across departments, fostering a culture where data-driven decision-making seamlessly integrates into daily operations. "For example, when Coca-Cola wanted to become a more sustainable company, they continuously evaluated their progress and made changes to their packaging and sourcing practices. " (Kotter's 8-Step Change Model: A Comprehensive Overview, n.d.).

8.) Institutionalize Changes: Embedding BI practices into the organizational DNA, the company implements structural changes, including the establishment of a dedicated data governance department and the realignment of roles to prioritize data-driven decision-making. BI metrics become integral to performance assessments, and a feedback mechanism ensures continuous refinement. By institutionalizing these changes, the organization solidifies its long-term dedication to BI, recognizing it as a fundamental component of its culture and future prosperity. "For example, when Google wanted to create a more open and collaborative culture, they implemented a 20%-time policy where employees could work on personal projects." (Kotter's 8-Step Change Model: A Comprehensive Overview, n.d.).

2. Technical Challenges and Solutions:

Addressing technical challenges is crucial for a successful BI implementation. One of the primary technical hurdles is data quality, encompassing issues such as inconsistent formats, missing values, and inaccuracies.

Data Quality

With guest reviews being a crucial aspect of the data analyzed, ensuring data quality is paramount, particularly in user-submitted content. To address this concern at Airbnb, implementing sentiment analysis could prove beneficial. This entails creating a system where the sentiment of a review is assessed and compared to the rating provided by the guest. If there is alignment between the sentiment analysis and the rating, the data is considered reliable. However, any inconsistencies should be flagged for further investigation. This approach becomes even more relevant if Airbnb intends to expand its review sources to include data scraped from external review websites in addition to its own platform. Implementing sentiment analysis would be a key strategy to ensure the integrity of this combined dataset.

Another data quality challenge encountered relates to the conjoint analysis data. This issue can be addressed by sourcing feedback directly from Airbnb guests. By gathering data directly from actual customers, we can significantly enhance the quantity and accuracy of the data available for conjoint analysis. With a larger and more precise dataset, we can derive more meaningful and statistically significant insights from the conjoint analysis, thereby improving decision-making processes at Airbnb.

Upholding data integrity across BI initiatives amidst evolving business requirements necessitates the establishment of a data governance department. This specialized unit will proactively and reactively address data inaccuracies and null values, employing strategies like regular data audits and the formulation of comprehensive data governance guidelines. By implementing these measures, the organization ensures the reliability and consistency of data utilized in BI projects, enhancing the effectiveness of decision-making processes.

Securing Machine-to-Cloud Connectivity - Transitioning to cloud storage introduces significant cybersecurity considerations. To address this, we propose implementing preprocessing measures on machines to optimize data transmission protocols, thereby minimizing the risk of data loss during transmission. Moreover, establishing redundant local storage serves as a backup mechanism to further mitigate potential data loss. Additionally, deploying redundant routers helps alleviate the risks associated with network connection failures. However, addressing cybersecurity concerns entails financial investment and specialized expertise to ensure regulatory compliance and adequately quantify and manage associated risks, thereby enhancing the overall security of the data transmission process.

3. Ethical Considerations in Business Intelligence (BI) Implementation at Airbnb:

- Privacy Protection - As Airbnb delves into BI implementation, safeguarding user privacy emerges as a paramount ethical concern. In the absence of robust federal privacy regulations, it is imperative for Airbnb to prioritize privacy by obtaining explicit consent from users before data collection. Transparent communication regarding data usage practices and the judicious collection of only essential data to enhance the guest experience are ethical imperatives.

- Data Security Measures - Upholding ethical standards necessitates ensuring the confidentiality and security of collected data at Airbnb. Beyond legal obligations, there exists a moral obligation to safeguard sensitive information from potential data breaches and cyber threats. Airbnb can demonstrate its commitment to ethical data practices by implementing end-to-end encryption protocols and partnering with vendors known for their strong emphasis on data security.

- Transparency and Accountability - Ethical BI implementation entails fostering transparency regarding data utilization practices at Airbnb, enabling users to comprehend the purpose and implications of data usage. Providing users with the option to control the sharing of their information and promptly addressing any breaches of data security are vital aspects of accountability. In the event of data breaches, Airbnb must uphold accountability by promptly informing affected users and taking appropriate remedial measures to mitigate any adverse consequences.

- Addressing Algorithmic Bias - Given the potential for AI-driven BI tools to perpetuate biases, Airbnb must prioritize fairness and equity in algorithmic decision-making processes. Ethical measures include leveraging open-source software, engaging data scientists to detect and mitigate biases, and collaborating with industry peers to develop fair and unbiased analytics tools. By proactively addressing algorithmic bias, Airbnb can uphold its commitment to ethical data practices and ensure fairness in its operations.

4. Other Challenges of Business Intelligence (BI) Implementation at Airbnb

Inadequate Privacy Legislation - The absence of comprehensive federal privacy laws presents a challenge for Airbnb in regulating the collection and utilization of consumer data, thus raising ethical concerns regarding data privacy practices.

Data Governance Complexity - Ethical BI implementation at Airbnb necessitates robust data governance frameworks to oversee data collection and usage. This involves conducting regular audits of data practices to ensure alignment with ethical standards and regulatory requirements.

Ensuring Fairness and Bias Mitigation - BI tools may inadvertently perpetuate biases, potentially resulting in unfair treatment. Addressing biases and ensuring algorithmic fairness pose ongoing challenges for Airbnb, requiring proactive measures such as hiring data scientists and fostering collaborative approaches.

Balancing Organizational Needs with Ethical Considerations - Airbnb faces the challenge of striking a balance between meeting organizational objectives and upholding ethical principles such as data security, transparency, and fairness. Achieving this balance is essential for ethical BI implementation and maintaining trust among users.

Bonus section – Social Media Strategy

The goals of Airbnb's social media strategy are to increase brand recognition and encourage user interaction. Popular social media sites like Facebook, Instagram, TikTok, and Twitter will be used by them to highlight their unique guest experiences and beautiful tourist locations. Airbnb can exhibit the genuineness of their offerings and attract a larger audience by working with influencers and content creators. In order to promote user involvement and foster a feeling of community, the strategy places a strong emphasis on interactive content, such as campaigns and contests. Influencer collaborations will enable them to reach a wider audience and gain the trust of their following, while sentiment analysis will be utilized to comprehend consumer preferences and opinions.

Furthermore, social media platforms with dedicated customer care channels will guarantee fast answers to questions and comments, resolving any problems quickly and enhancing the general consumer experience. Airbnb wants to establish a strong online presence by connecting with tourists through social media, showcasing their varied listings, and eventually encouraging bookings.

Summary and Conclusion

In the quest for enhancing operational efficiency and strategic decision-making, the adoption of Business Intelligence (BI) presents significant opportunities for Airbnb, a global leader in lodging and vacation rentals. Throughout this comprehensive overview, we have delved into the essential components of BI implementation tailored to Airbnb's unique requirements.

Explored BI Implementation Components: Data Sources and Integration: Identified crucial internal data sources such as booking databases, guest feedback, property listings, and external data including market trends and travel insights. The integration of these diverse data sets forms the foundation for comprehensive analysis.

Streamlined Data Warehouse Structure: Proposed a simplified data architecture structured for Airbnb's lodging and rental business, facilitating the consolidation of various data sources. This structured framework enables efficient data querying and analytics, supporting data-driven decision-making.

Data Analytics for Optimization: Emphasized the significance of demand forecasting to optimize pricing strategies, enhance listing performance, and allocate resources effectively. Additionally, real-time analytics for guest feedback and property performance enable continuous improvement in guest satisfaction and property management.

Tactical and Operational Dashboards: Developed strategic dashboards tailored for financial insights, including revenue analysis, expense tracking, and profitability assessment, empowering management in financial planning and decision-making. Simultaneously, operational dashboards focused on monitoring property health and guest engagement ensure enhanced operational efficiency and customer experience.

Recommendations for Strategic Adoption: Cultivating a Culture of Continuous Improvement: Airbnb should embrace BI as an iterative process, fostering a culture of continuous improvement. Regular updates to data models and dashboard functionalities will ensure alignment with evolving business needs.

Employee Training and Adoption: Investing in comprehensive employee training and cultivating a data-driven culture is essential for maximizing the benefits of BI. Empowering employees with the skills and tools to extract actionable insights will amplify BI's impact across all facets of the organization.

Ethical Data Practices and Compliance: Prioritizing ethical data handling and compliance with data privacy regulations is paramount. Establishing robust data security measures and ensuring adherence to privacy guidelines will enhance trust with users and safeguard sensitive information.

Strategic Decision-making with BI Insights: Leadership at Airbnb is encouraged to leverage BI-generated insights in strategic decision-making processes. Utilizing demand forecasts for dynamic pricing strategies and leveraging property performance analytics for proactive maintenance will drive sustained growth and competitiveness in the dynamic hospitality industry.

If Airbnb decides to implement the business intelligence solutions proposed, there should be an overall increase in awareness and monitoring of the guest experience. This will enable informed decision-making to enhance guest satisfaction and identify any issues with listings or any part of the booking process. Initially designed for the accommodation department of Airbnb, the goal is to extend this monitoring system to other customer-facing departments in the future. The plan to establish a data warehouse would facilitate access to booking and review data for relevant stakeholders across the organization. Our implementation plan, aimed at addressing potential ethical and data quality concerns, is crucial in achieving this objective.

We aim to conduct conjoint analysis with Airbnb guests, expanding beyond our initial sample, to derive valuable insights that will inform the marketing department's strategies regarding specific listings, timing, and distribution channels.

In conclusion, the strategic adoption of Business Intelligence aligned with Airbnb's objectives promises transformative benefits. The organization stands poised to harness data-driven insights to optimize operations, enhance guest experiences, and maintain its position as a global leader in the lodging and vacation rental sector.

Appendix A:

Data Dictionary

Airbnb Open Data Dictionary		
Field	Type	Description
id	integer	Airbnb's unique identifier for the listing
scrape_id	bigint	Inside Airbnb "Scrape" this was part of
last_scraped	datetime	UTC. The date and time this listing was "scraped".
source	text	One of "neighbourhood search" or "previous scrape". "neighbourhood search" means that the listing was found by searching the city, while "previous scrape" means that the listing was seen in another scrape performed in the last 65 days, and the listing was confirmed to be still available on the Airbnb site.
name	text	Name of the listing
description	text	Detailed description of the listing
neighborhood_overview	text	Host's description of the neighbourhood
picture_url	text	URL to the Airbnb hosted regular sized image for the listing
host_id	integer	Airbnb's unique identifier for the host/user
host_url	text	The Airbnb page for the host
host_name	text	Name of the host. Usually just the first name(s).
host_since	date	The date the host/user was created. For hosts that are Airbnb guests this could be the date they registered as a guest.
host_location	text	The host's self reported location
host_about	text	Description about the host
host_acceptance_rate		That rate at which a host accepts booking requests.
host_is_superhost	boolean	[t=true; f=false]
host_listings_count	text	The number of listings the host has (per Airbnb calculations)
host_total_listings_count	text	The number of listings the host has (per Airbnb calculations)
host_has_profile_pic	boolean	t=true; f=false
host_identity_verified	boolean	t=true; f=false
neighbourhood_cleansed	text	The neighbourhood as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.
neighbourhood_group_cleansed	text	The neighbourhood group as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.

latitude	numeric	Uses the World Geodetic System (WGS84) projection for latitude and longitude.
longitude	numeric	Uses the World Geodetic System (WGS84) projection for latitude and longitude.
property_type	text	Self selected property type. Hotels and Bed and Breakfasts are described as such by their hosts in this field
room_type	text	[Entire home/apt|Private room|Shared room|Hotel] All homes are grouped into the following three room types: Entire place Private room Shared room Entire place Entire places are best if you're seeking a home away from home. With an entire place, you'll have the whole space to yourself. This usually includes a bedroom, a bathroom, a kitchen, and a separate, dedicated entrance. Hosts should note in the description if they'll be on the property or not (ex: "Host occupies first floor of the home"), and provide further details on the listing. Private rooms Private rooms are great for when you prefer a little privacy, and still value a local connection. When you book a private room, you'll have your own private room for sleeping and may share some spaces with others. You might need to walk through indoor spaces that another host or guest may occupy to get to your room. Shared rooms Shared rooms are for when you don't mind sharing a space with others. When you book a shared room, you'll be sleeping in a space that is shared with others and share the entire space with other people. Shared rooms are popular among flexible travelers looking for new friends and budget-friendly stays.
accommodates	integer	The maximum capacity of the listing
bathrooms	numeric	The number of bathrooms in the listing
bathrooms_text	string	The number of bathrooms in the listing. On the Airbnb web-site, the bathrooms field has evolved from a number to a textual description. For older scrapes, bathrooms is used.
bedrooms	integer	The number of bedrooms
beds	integer	The number of bed(s)
price	currency	daily price in local currency
minimum_nights	integer	minimum number of night stay for the listing (calendar rules may be different)
maximum_nights	integer	maximum number of night stay for the listing (calendar rules may be different)
minimum_minimum_nights	integer	the smallest minimum_night value from the calender (looking 365 nights in the future)

maximum_minimum_nights	integer	the largest minimum_night value from the calender (looking 365 nights in the future)
minimum_maximum_nights	integer	the smallest maximum_night value from the calender (looking 365 nights in the future)
maximum_maximum_nights	integer	the largest maximum_night value from the calender (looking 365 nights in the future)
minimum_nights_avg_ntm	numeric	the average minimum_night value from the calender (looking 365 nights in the future)
maximum_nights_avg_ntm	numeric	the average maximum_night value from the calender (looking 365 nights in the future)
has_availability	boolean	[t=true; f=false]
availability_30	integer	avaliability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
availability_60	integer	avaliability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
availability_90	integer	avaliability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
availability_365	integer	avaliability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
number_of_reviews	integer	The number of reviews the listing has
number_of_reviews_ltm	integer	The number of reviews the listing has (in the last 12 months)
number_of_reviews_l30d	integer	The number of reviews the listing has (in the last 30 days)
first_review	date	The date of the first/oldest review
last_review	date	The date of the last/newest review
license	text	The licence/permit/registration number
instant_bookable	boolean	[t=true; f=false]. Whether the guest can automatically book the listing without the host requiring to accept their booking request. An indicator of a commercial listing.
calculated_host_listings_count	integer	The number of listings the host has in the current scrape, in the city/region geography.

calculated_host_listings_count_entire_homes	integer	The number of Entire home/apt listings the host has in the current scrape, in the city/region geography
calculated_host_listings_count_private_rooms	integer	The number of Private room listings the host has in the current scrape, in the city/region geography
calculated_host_listings_count_shared_rooms	integer	The number of Shared room listings the host has in the current scrape, in the city/region geography
reviews_per_month	numeric	The number of reviews the listing has over the lifetime of the listing

Appendix B:

Python code for predictive analytics:

1.

```
[ ] import pandas as pd
```

```
[ ] # Reading a small set to check if we are not loading a data with empty columns (Unnamed columns)
    airbnb_df=pd.read_csv(filepath_or_buffer='Airbnb_Open_Data.csv',dtype='str',nrows=4)
```

```
[ ] airbnb_df
```

	id	NAME	host id	host_identity_verified	host name	neighbourhood group	neighbourhood	lat	long	country	...	service fee	minimum nights	number of reviews	last review	reviews per month	review rate number	calc li
0	1001254	Clean & quiet apt home by the park	80014485718	unconfirmed	Madaline	Brooklyn	Kensington	40.64749	-73.97237	United States	...	$193	10	9	10/19/2021	0.21	4	
1	1002102	Skylit Midtown Castle	52335172823	verified	Jenna	Manhattan	Midtown	40.75362	-73.98377	United States	...	$28	30	45	5/21/2022	0.38	4	
2	1002403	THE VILLAGE OF HARLEM....NEW YORK !	78829239556	NaN	Elise	Manhattan	Harlem	40.80902	-73.9419	United States	...	$124	3	0	NaN	NaN	5	
3	1002755	NaN	85098326012	unconfirmed	Garry	Brooklyn	Clinton Hill	40.68514	-73.95976	United States	...	$74	30	270	7/5/2019	4.64	4	

4 rows × 26 columns

2.

```
# Review a record to get base idea of data
airbnb_df.head(1).T.to_dict()
```

```
{0: {'id': '1001254',
  'NAME': 'Clean & quiet apt home by the park',
  'host id': '80014485718',
  'host_identity_verified': 'unconfirmed',
  'host name': 'Madaline',
  'neighbourhood group': 'Brooklyn',
  'neighbourhood': 'Kensington',
  'lat': '40.64749',
  'long': '-73.97237',
  'country': 'United States',
  'country code': 'US',
  'instant_bookable': 'FALSE',
  'cancellation_policy': 'strict',
  'room type': 'Private room',
  'Construction year': '2020',
  'price': '$966 ',
  'service fee': '$193 ',
  'minimum nights': '10',
  'number of reviews': '9',
  'last review': '10/19/2021',
  'reviews per month': '0.21',
  'review rate number': '4',
  'calculated host listings count': '6',
  'availability 365': '286',
  'house_rules': "Clean up and treat the home the way you'd like your home to be treated.  No smoking.",
  'license': nan}}
```

3.

```
[ ] airbnb_df.columns
```

```
Index(['id', 'NAME', 'host id', 'host_identity_verified', 'host name',
       'neighbourhood group', 'neighbourhood', 'lat', 'long', 'country',
       'country code', 'instant_bookable', 'cancellation_policy', 'room type',
       'Construction year', 'price', 'service fee', 'minimum nights',
       'number of reviews', 'last review', 'reviews per month',
       'review rate number', 'calculated host listings count',
       'availability 365', 'house_rules', 'license'],
      dtype='object')
```

```
[ ] airbnb_df=pd.read_csv(filepath_or_buffer='Airbnb_Open_Data.csv',dtype='str')
```

```
[ ] airbnb_df.shape
```

```
(102599, 26)
```

4.

```python
# Checking for missing values in the DataFrame
airbnb_df.isna().sum()
```

```
id                                     0
NAME                                 250
host id                                0
host_identity_verified               289
host name                            406
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                        52131
license                           102597
dtype: int64
```

5.

```python
# Printing unique values of each column for the first 50 rows
for c_name in airbnb_df.columns:
  print(f'{c_name} : {airbnb_df[c_name].head(50).unique()[0:10]}')
```

```
id : ['1001254' '1002102' '1002403' '1002755' '1003689' '1004098' '1004650'
 '1005202' '1005754' '1006307']
NAME : ['Clean & quiet apt home by the park' 'Skylit Midtown Castle'
 'THE VILLAGE OF HARLEM....NEW YORK !' nan
 'Entire Apt: Spacious Studio/Loft by central park'
 'Large Cozy 1 BR Apartment In Midtown East' 'BlissArtsSpace!'
 "Large Furnished Room Near B'way" 'Cozy Clean Guest Room - Family Apt'
 'Cute & Cozy Lower East Side 1 bdrm']
host id : ['80014485718' '52335172823' '78829239556' '85098326012' '92037596077'
 '45498551794' '61300605564' '90821839709' '79384379533' '75527839483']
host_identity_verified : ['unconfirmed' 'verified' nan]
host name : ['Madaline' 'Jenna' 'Elise' 'Garry' 'Lyndon' 'Michelle' 'Alberta' 'Emma'
 'Evelyn' 'Carl']
neighbourhood group : ['Brooklyn' 'Manhattan' 'brookln' 'manhatan' 'Queens']
neighbourhood : ['Kensington' 'Midtown' 'Harlem' 'Clinton Hill' 'East Harlem'
 'Murray Hill' 'Bedford-Stuyvesant' "Hell's Kitchen" 'Upper West Side'
 'Chinatown']
lat : ['40.64749' '40.75362' '40.80902' '40.68514' '40.79851' '40.74767'
 '40.68688' '40.76489' '40.80178' '40.71344']
long : ['-73.97237' '-73.98377' '-73.9419' '-73.95976' '-73.94399' '-73.975'
 '-73.95596' '-73.98493' '-73.96723' '-73.99037']
country : ['United States']
country code : ['US' nan]
instant_bookable : ['FALSE' 'TRUE' nan]
cancellation_policy : ['strict' 'moderate' 'flexible']
room type : ['Private room' 'Entire home/apt' 'Shared room']
Construction year : ['2020' '2007' '2005' '2009' '2013' '2015' '2004' '2008' '2010' '2019']
price : ['$966 ' '$142 ' '$620 ' '$368 ' '$204 ' '$577 ' '$71 ' '$1,060 '
 '$1,018 ' '$291 ']
service fee : ['$193 ' '$28 ' '$124 ' '$74 ' '$41 ' '$115 ' '$14 ' '$212 ' '$204 '
 '$58 ']
minimum nights : ['10' '30' '3' '45' '2' '1' '5' '4' '90' '7']
number of reviews : ['9' '45' '0' '270' '74' '49' '430' '118' '160' '53']
last review : ['10/19/2021' '5/21/2022' nan '7/5/2019' '11/19/2018' '6/22/2019'
 '10/5/2017' '6/24/2019' '7/21/2017' '6/9/2019']
reviews per month : ['0.21' '0.38' nan '4.64' '0.1' '0.59' '0.4' '3.47' '0.99' '1.33']
review rate number : ['4' '5' '3' nan '2' '1']
calculated host listings count : ['6' '2' '1' '4' '3']
availability 365 : ['286' '228' '352' '322' '289' '374' '224' '219' '180' '375']
house_rules : ["Clean up and treat the home the way you'd like your home to be treated.  No smoking."
 'Pet friendly but please confirm with me if the pet you are planning on bringing with you is OK. I have a cute and quiet mixed chihuahua. I could accept more guests (for an extra fee) b
 "I encourage you to use my kitchen, cooking and laundry facilities. There is no additional charge to use the washer/dryer in the basement.  No smoking, inside or outside. Come home as l
 nan
```

6.

```
[ ] # Dropping specified columns from the DataFrame
airbnb_df.drop(columns=[
    'license',
    'country',
    'country code',
    'last review',
    'house_rules',
    'host_identity_verified',
    'id',
    'NAME',
    'host id',
    'host name'
    ]
  ,inplace=True)
```

```
# Checking for missing values in the DataFrame after dropping columns
airbnb_df.isna().sum()
```

```
neighbourhood group                 29
neighbourhood                       16
lat                                  8
long                                 8
instant_bookable                   105
cancellation_policy                 76
room type                            0
Construction year                  214
price                              247
service fee                        273
minimum nights                     409
number of reviews                  183
reviews per month                15879
review rate number                 326
calculated host listings count     319
availability 365                   448
dtype: int64
```

```
[ ] # Removing rows with missing values in the 'price' column
airbnb_df=airbnb_df[~airbnb_df['price'].isna()]
airbnb_df.shape
```

```
(102352, 16)
```

7.

```
# Checking for missing values again after removing rows with missing 'price'
airbnb_df.isna().sum()
```

```
neighbourhood group                 28
neighbourhood                       15
lat                                  8
long                                 8
instant_bookable                   100
cancellation_policy                 71
room type                            0
Construction year                  210
price                                0
service fee                        239
minimum nights                     409
number of reviews                  183
reviews per month                15852
review rate number                 326
calculated host listings count     319
availability 365                   448
dtype: int64
```

```
[ ] # Removing '$' and ',' characters from 'price' and 'service fee' columns and converting them to float
airbnb_df['price']=airbnb_df['price'].str.replace('$','').str.replace(',','')
airbnb_df['service fee']=airbnb_df['service fee'].str.replace('$','').str.replace(',','')
```

```
[ ] # Checking the data types of each column in the DataFrame
airbnb_df.dtypes
```

```
neighbourhood group               object
neighbourhood                     object
lat                               object
long                              object
instant_bookable                  object
cancellation_policy               object
room type                         object
Construction year                 object
price                             object
service fee                       object
minimum nights                    object
number of reviews                 object
reviews per month                 object
review rate number                object
calculated host listings count    object
availability 365                  object
dtype: object
```

8.

```python
# Filling missing values with a placeholder
airbnb_df=airbnb_df.fillna(value=-999999)
```

```python
# Converting selected columns to appropriate data types
airbnb_df[
  [
    'lat',
    'long',
    'price',
    'service fee',
    'reviews per month'
    ]
  ]=airbnb_df[
              [
                'lat',
                'long',
                'price',
                'service fee',
                'reviews per month'
                ]
              ].astype('float')
airbnb_df[
  [
    'Construction year',
    'minimum nights',
    'number of reviews',
    'review rate number',
    'calculated host listings count',
    'availability 365'
    ]
  ]=airbnb_df[
              [
                'Construction year',
                'minimum nights',
                'number of reviews',
                'review rate number',
                'calculated host listings count',
                'availability 365'
                ]
              ].astype('int')
```

9.

```python
# Checking the data types of each column in the DataFrame
airbnb_df.dtypes
```

```
neighbourhood group                object
neighbourhood                      object
lat                               float64
long                              float64
instant_bookable                   object
cancellation_policy                object
room type                          object
Construction year                   int64
price                             float64
service fee                       float64
minimum nights                      int64
number of reviews                   int64
reviews per month                 float64
review rate number                  int64
calculated host listings count      int64
availability 365                    int64
dtype: object
```

```python
# Function to generate descriptive statistics for specified columns
def get_descriptive_statistics(df:pd.DataFrame,col_names:list)->pd.DataFrame:
  descp_summary_details:list=[]
  for c_name in col_names:
    temp_dict=df[df[c_name]!=-999999][c_name].describe().to_dict()
    temp_dict['Feature Name']=c_name
    descp_summary_details.append(temp_dict)
    del temp_dict
  return pd.DataFrame(descp_summary_details)
```

10.

```python
# Generating descriptive statistics for selected columns
get_descriptive_statistics(
    df=airbnb_df,
    col_names=[
        'Construction year',
        'price',
        'service fee',
        'minimum nights',
        'number of reviews',
        'reviews per month',
        'review rate number',
        'calculated host listings count',
        'availability 365'
        ],
    )
```

	count	mean	std	min	25%	50%	75%	max	Feature Name
0	102142.0	2012.487038	5.765066	2003.00	2007.00	2012.00	2017.00	2022.0	Construction year
1	102352.0	625.293536	331.671614	50.00	340.00	624.00	913.00	1200.0	price
2	102113.0	125.038399	66.333513	10.00	68.00	125.00	183.00	240.0	service fee
3	101943.0	8.125619	30.556607	-1223.00	2.00	3.00	5.00	5645.0	minimum nights
4	102169.0	27.487878	49.521341	0.00	1.00	7.00	30.00	1024.0	number of reviews
5	86500.0	1.375068	1.747680	0.01	0.22	0.74	2.01	90.0	reviews per month
6	102026.0	3.279282	1.284458	1.00	2.00	3.00	4.00	5.0	review rate number
7	102033.0	7.927200	32.211222	1.00	1.00	1.00	2.00	332.0	calculated host listings count
8	101904.0	141.119897	135.425491	-10.00	3.00	96.00	269.00	3677.0	availability 365

11.

```python
# Filtering out invalid values in 'minimum nights' and 'availability 365' columns
# Based on Above Analysis we can remove some records
airbnb_df=airbnb_df[airbnb_df['minimum nights']>=0]
airbnb_df=airbnb_df[airbnb_df['availability 365']>=0]
airbnb_df=airbnb_df[airbnb_df['minimum nights']<366]
airbnb_df=airbnb_df[airbnb_df['availability 365']<366]
print(airbnb_df.shape)
```

```
(98292, 16)
```

12.

```python
# Generating descriptive statistics again after filtering
get_descriptive_statistics(
    df=airbnb_df,
    col_names=[
        'Construction year',
        'price',
        'service fee',
        'minimum nights',
        'number of reviews',
        'reviews per month',
        'review rate number',
        'calculated host listings count',
        'availability 365'
        ],
    )
```

	count	mean	std	min	25%	50%	75%	max	Feature Name
0	98130.0	2012.489626	5.763392	2003.00	2008.00	2012.00	2017.75	2022.0	Construction year
1	98292.0	625.789851	331.719944	50.00	340.00	626.00	913.00	1200.0	price
2	98062.0	125.141614	66.343466	10.00	68.00	125.00	183.00	240.0	service fee
3	98292.0	7.874507	17.013689	1.00	2.00	3.00	5.00	365.0	minimum nights
4	98166.0	27.154290	48.867710	0.00	1.00	7.00	30.00	1024.0	number of reviews
5	83362.0	1.389219	1.756093	0.01	0.23	0.76	2.03	90.0	reviews per month
6	98008.0	3.287711	1.278845	1.00	2.00	3.00	4.00	5.0	review rate number
7	98002.0	8.048315	32.622870	1.00	1.00	1.00	2.00	332.0	calculated host listings count
8	98292.0	134.458257	129.799474	0.00	2.00	90.00	254.00	365.0	availability 365

13.

Model Training and Summary

```python
import time
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler,OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import mean_squared_error
```

14.

```python
def model_run(df:pd.DataFrame,target_columns:list,algos:list,n:int)->pd.DataFrame:
  """
  Train and evaluate models for regression task using specified algorithms.

  Parameters:
      df (pd.DataFrame): The DataFrame containing the dataset.
      target_columns (list): A list of target columns to predict.
      algos (list): A list of tuples, where each tuple contains an algorithm name and an instance of the algorithm.
      n (int): The number of rows to consider from the dataset.
                This can be used to limit and check the time and resource before running large dataset.

  Returns:
      result: A DataFrame containing the feature name, model name, and Mean Squared Error (MSE) for each combination of feature and algorithm.
  """

  # Identifying numeric and categorical features
  numeric_features=df.select_dtypes(include=['float64','int64']).columns
  categorical_features=df.select_dtypes(include=['object']).columns

  # Initializing transformers
  numeric_transformer=StandardScaler()
  categorical_transformer=OneHotEncoder(handle_unknown='ignore')

  # List to store results
  loop_results:list=[]

  print(time.ctime())

  # Iterating over features
  for feature in df.drop(columns=target_columns).columns:

    temp_df=df[[feature]+target_columns]

    # Filtering the dataset based on specified number of rows
    if n>0:
      temp_df=temp_df.head(n)
```

15.

```python
def display_top_results(df,n,algos)->None:
  """
  Display the top results (features with lowest MSE) for each algorithm.

  Parameters:
      df (pd.DataFrame): The DataFrame containing the model results.
      n (int): The number of top results to display.
      algos (list): A list of tuples, where each tuple contains an algorithm name and an instance of the algorithm.

  Returns:
      None
  """
  for algo_name,algo in algos:
    print('Algorithm Used: ',algo_name)
    print(df[df['Model']==algo_name][['Feature','MSE']].head(n))
    print('='*50)
  del algo_name,algo
  return None
```

```python
# Algorithms to run for our Regression Task
custom_algorithms=[
    ('Linear Regression',LinearRegression()),
    ('Decision Tree',DecisionTreeRegressor(random_state=123)),
    ('Naive Bayes',GaussianNB())
]
```

16.

```python
def display_top_results(df,n,algos)->None:
  """
  Display the top results (features with lowest MSE) for each algorithm.

  Parameters:
      df (pd.DataFrame): The DataFrame containing the model results.
      n (int): The number of top results to display.
      algos (list): A list of tuples, where each tuple contains an algorithm name and an instance of the algorithm.

  Returns:
      None
  """
  for algo_name,algo in algos:
    print('Algorithm Used: ',algo_name)
    print(df[df['Model']==algo_name][['Feature','MSE']].head(n))
    print('='*50)
  del algo_name,algo
  return None
```

```python
# Algorithms to run for our Regression Task
custom_algorithms=[
    ('Linear Regression',LinearRegression()),
    ('Decision Tree',DecisionTreeRegressor(random_state=123)),
    ('Naive Bayes',GaussianNB())
]
```

17.

```python
# Running the model
model_summary=model_run(df=airbnb_df,target_columns=['price'],algos=custom_algorithms,n=0)
print(model_summary.shape)
```

```
Sat Apr 13 10:53:09 2024
Linear Regression neighbourhood group Sat Apr 13 10:53:09 2024
==================================================
Decision Tree neighbourhood group Sat Apr 13 10:53:09 2024
==================================================
Naive Bayes neighbourhood group Sat Apr 13 10:53:13 2024
==================================================
Linear Regression neighbourhood Sat Apr 13 10:53:19 2024
==================================================
Decision Tree neighbourhood Sat Apr 13 10:53:44 2024
==================================================
Naive Bayes neighbourhood Sat Apr 13 10:54:24 2024
==================================================
Linear Regression lat Sat Apr 13 10:54:24 2024
==================================================
Decision Tree lat Sat Apr 13 10:54:25 2024
==================================================
Naive Bayes lat Sat Apr 13 10:54:26 2024
==================================================
Linear Regression long Sat Apr 13 10:54:26 2024
==================================================
Decision Tree long Sat Apr 13 10:54:26 2024
==================================================
Naive Bayes long Sat Apr 13 10:54:27 2024
==================================================
Linear Regression instant_bookable Sat Apr 13 10:54:27 2024
==================================================
Decision Tree instant_bookable Sat Apr 13 10:54:27 2024
==================================================
Naive Bayes instant_bookable Sat Apr 13 10:54:28 2024
==================================================
Linear Regression cancellation_policy Sat Apr 13 10:54:28 2024
==================================================
Decision Tree cancellation_policy Sat Apr 13 10:54:28 2024
==================================================
Naive Bayes cancellation_policy Sat Apr 13 10:54:31 2024
==================================================
Linear Regression room type Sat Apr 13 10:54:31 2024
==================================================
Decision Tree room type Sat Apr 13 10:54:31 2024
```

18.

```
Decision Tree Construction year Sat Apr 13 10:54:35 2024
==================================================
Naive Bayes Construction year Sat Apr 13 10:54:36 2024
==================================================
Linear Regression service fee Sat Apr 13 10:54:36 2024
==================================================
Decision Tree service fee Sat Apr 13 10:54:36 2024
==================================================
Naive Bayes service fee Sat Apr 13 10:54:37 2024
==================================================
Linear Regression minimum nights Sat Apr 13 10:54:37 2024
==================================================
Decision Tree minimum nights Sat Apr 13 10:54:37 2024
==================================================
Naive Bayes minimum nights Sat Apr 13 10:54:38 2024
==================================================
Linear Regression number of reviews Sat Apr 13 10:54:38 2024
==================================================
Decision Tree number of reviews Sat Apr 13 10:54:38 2024
==================================================
Naive Bayes number of reviews Sat Apr 13 10:54:39 2024
==================================================
Linear Regression reviews per month Sat Apr 13 10:54:39 2024
==================================================
Decision Tree reviews per month Sat Apr 13 10:54:39 2024
==================================================
Naive Bayes reviews per month Sat Apr 13 10:54:39 2024
==================================================
Linear Regression review rate number Sat Apr 13 10:54:39 2024
==================================================
Decision Tree review rate number Sat Apr 13 10:54:39 2024
==================================================
Naive Bayes review rate number Sat Apr 13 10:54:40 2024
==================================================
Linear Regression calculated host listings count Sat Apr 13 10:54:40 2024
==================================================
Decision Tree calculated host listings count Sat Apr 13 10:54:40 2024
==================================================
Naive Bayes calculated host listings count Sat Apr 13 10:54:41 2024
==================================================
Linear Regression availability 365 Sat Apr 13 10:54:41 2024
==================================================
Decision Tree availability 365 Sat Apr 13 10:54:41 2024
==================================================
Naive Bayes availability 365 Sat Apr 13 10:54:42 2024
==================================================
(45, 3)
```

19.

```
[ ]  # Displaying top 10 results
     display_top_results(df=model_summary,n=10,algos=custom_algorithms)
```

```
Algorithm Used:  Linear Regression
                        Feature            MSE
24                  service fee       1.989963
18                    room type  108746.641971
27               minimum nights  108749.103541
42             availability 365  108752.798123
36           review rate number  109358.773199
39  calculated host listings count  109438.642012
30            number of reviews  109779.022612
12             instant_bookable  110077.083108
21            Construction year  110166.561876
3                 neighbourhood  110252.340302
==================================================
Algorithm Used:  Decision Tree
                        Feature            MSE
25                  service fee       1.985475
19                    room type  108739.904704
28               minimum nights  108772.368671
43             availability 365  109137.222445
37           review rate number  109371.429539
40  calculated host listings count  109481.634598
31            number of reviews  110006.253332
13             instant_bookable  110076.544158
22            Construction year  110155.920026
4                 neighbourhood  110252.340302
==================================================
Algorithm Used:  Naive Bayes
                        Feature            MSE
26                  service fee       4.135930
41  calculated host listings count  137752.411816
8                           lat  150644.289159
23            Construction year  174004.302252
35            reviews per month  192343.718347
5                 neighbourhood  192826.124180
17          cancellation_policy  202727.144900
29               minimum nights  234001.765298
2           neighbourhood group  239735.031899
38           review rate number  261049.455617
==================================================
```

20.

```
[ ]  # Displaying top 7 results
     display_top_results(df=model_summary,n=7,algos=custom_algorithms)
```

```
Algorithm Used:  Linear Regression
                        Feature            MSE
24                  service fee       1.989963
18                    room type  108746.641971
27               minimum nights  108749.103541
42             availability 365  108752.798123
36           review rate number  109358.773199
39  calculated host listings count  109438.642012
30            number of reviews  109779.022612
==================================================
Algorithm Used:  Decision Tree
                        Feature            MSE
25                  service fee       1.985475
19                    room type  108739.904704
28               minimum nights  108772.368671
43             availability 365  109137.222445
37           review rate number  109371.429539
40  calculated host listings count  109481.634598
31            number of reviews  110006.253332
==================================================
Algorithm Used:  Naive Bayes
                        Feature            MSE
26                  service fee       4.135930
41  calculated host listings count  137752.411816
8                           lat  150644.289159
23            Construction year  174004.302252
35            reviews per month  192343.718347
5                 neighbourhood  192826.124180
17          cancellation_policy  202727.144900
==================================================
```

21.

```
[ ]  # Displaying top 5 results
     display_top_results(df=model_summary,n=5,algos=custom_algorithms)
```

```
Algorithm Used:  Linear Regression
              Feature            MSE
24        service fee       1.989963
18          room type  108746.641971
27     minimum nights  108749.103541
42   availability 365  108752.798123
36 review rate number  109358.773199
==================================================
Algorithm Used:  Decision Tree
              Feature            MSE
25        service fee       1.985475
19          room type  108739.904704
28     minimum nights  108772.368671
43   availability 365  109137.222445
37 review rate number  109371.429539
==================================================
Algorithm Used:  Naive Bayes
                        Feature            MSE
26                  service fee       4.135930
41  calculated host listings count  137752.411816
8                           lat  150644.289159
23            Construction year  174004.302252
35            reviews per month  192343.718347
==================================================
```

References:

Long Beach, CA - Airbnb Help Center. (n.d.). Airbnb. https://www.airbnb.com/help/article/939

Lucidchart. (2019). How to Implement Change with Kotter's 8-Step Change Model | Lucidchart Blog. Lucidchart.com. https://www.lucidchart.com/blog/kotters-8-step-change-model

Christin, S. (n.d.). Kotter's 8-Step Change Model: A Comprehensive Overview. Changechoreographers.com. Retrieved April 18, 2024, from https://www.changechoreographers.com/change-management-models-and-theories-kotter-s-8-step-change-model

Airbnb: Vacation Rentals, Cabins, Beach Houses, Unique Homes & Experiences. (n.d.). Airbnb. https://www.airbnb.com/software-partners

Saxena, P. (2018, November 6). Airbnb's Concept: Business Model & Revenue Source. Appinventiv. https://appinventiv.com/blog/airbnbs-business-model-and-revenue-source/

W, C. (2023, September 1). What's the Airbnb Insights Report? A Guide to Listing Performance. Rankbreeze. https://rankbreeze.com/airbnb-insights-report/

Tiernan, K. (2023, August 28). Airbnb Uses Artificial Intelligence (AI) to Transform their Business, Are You? BDO Digital. https://www.bdodigital.com/insights/analytics/airbnb-artificial-intelligence-transform-business

Airbnb Engineering & Data Science. (n.d.). Airbnb.io. https://airbnb.io/projects/superset/