Optimizing Revenue for Airbnb Hosts

Team members: Kanchi Tank, Shubham Gujar, Chaitanya Priya Kakkehalli Jayaram

Introduction

Airbnb is an excellent way for people to rent out their homes or rooms to travelers. It's been around for a while and has changed how we travel and stay in new places. But, as someone new to Airbnb hosting, making the most money from it can take time and effort. That's why analyzing how to make my Airbnb listing the best and earn more money is essential.

For many new Airbnb hosts, optimizing revenue can seem daunting. Understanding how to set the right price for their listing, attract guests, and maximize earnings requires a grasp of various strategies and tools. A wide array of approaches are available, from utilizing technology to analyze market trends and adjust pricing accordingly to seeking assistance from seasoned professionals in the field.

Exploring these strategies and tools marks the beginning of an exciting journey for hosts aiming to enhance their Airbnb listings. By delving into the intricacies of revenue optimization, hosts can uncover valuable insights that make their listings stand out and attract more guests. Through this exploration, hosts may even stumble upon tips and tricks that benefit themselves and fellow newcomers to the world of Airbnb hosting.

Problem statement

New Airbnb hosts need help to juggle various factors to maximize their earnings. It's a lot to handle, from adjusting prices to matching demand and tracking when their space is available to meet travelers' needs. Doing all this manually wastes precious time and often means they miss opportunities to earn more money.

One big issue is that hosts often set one price and need to remember about it, even when they could charge more during busy times like holidays or events. Managing their information across different platforms is a headache; mistakes can cost them bookings and cash.

Another problem is hosts need to know what's happening in the market. With insights into what guests want and when they want, hosts can make intelligent decisions about their listings. This puts them at risk of losing potential earnings and falling behind competitors.

To overcome these challenges, hosts need tools that simplify tasks, help them set competitive prices, and provide market insights. By addressing these issues, hosts can uncover new ways to boost their earnings and thrive in the competitive world of Airbnb hosting.

Dataset

1) Airbnb: We've obtained the latest Airbnb data from the Inside Airbnb website, which includes detailed listing data, calendar data, reviews data, and neighborhood data. For this project, we've obtained data covering the period from 3 December 2022 to 3 December 2023.
2) Crime Data: We obtained data on crime in Los Angeles from the data.gov LAPD website. This data has all the related features for a crime, including type of crime, address, and zip code. We have used this project's latest available data from 10 February 2020 to 8 March 2024. We have further extracted the number of crimes per zip code and integrated it with our dataset.
3) Tourist Attractions: Prior review suggested that the availability of tourist spots around Airbnb strongly correlates with its price and number of bookings. We extracted the total number of tourist attractions around each listing using Google API and then integrated it with our dataset.

Methodology or proposed solution

1) **New Hosts**

a) **Analysis**

We performed Pearson correlation over the number of bookings and price for all the features, aiming to uncover any underlying patterns or dependencies.

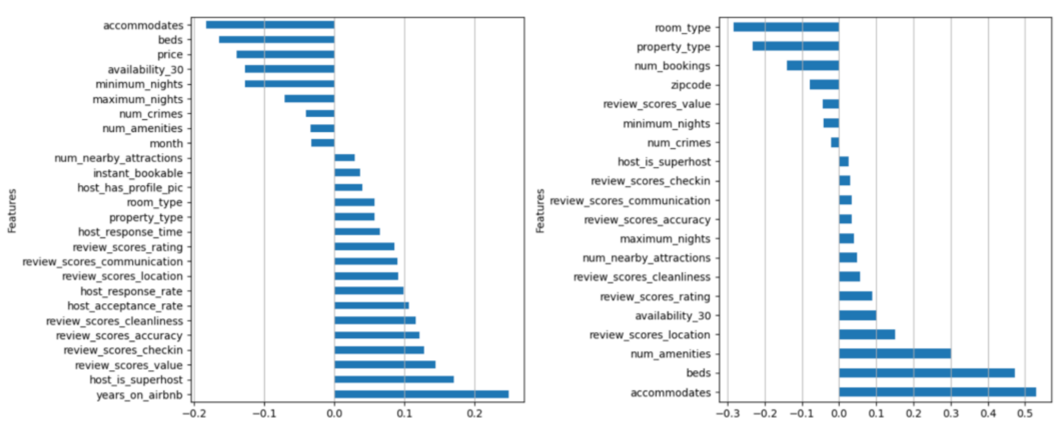

Fig 1. Graph representing Pearson correlation between the number of bookings and price of all the other features.

However, only some of these features are available when a new host registers his place on Airbnb. We selected a list of foundational features that the user will be aware of when listing his property on Airbnb for the first time. These features are 'minimum_nights,' 'maximum_nights,' 'property_type,' 'room_type,' 'accommodates,' 'beds,' 'price,' 'instant_bookable,' 'EncodedZip,' 'num_crimes,' 'num_nearby_attractions,' 'num_amenities.'

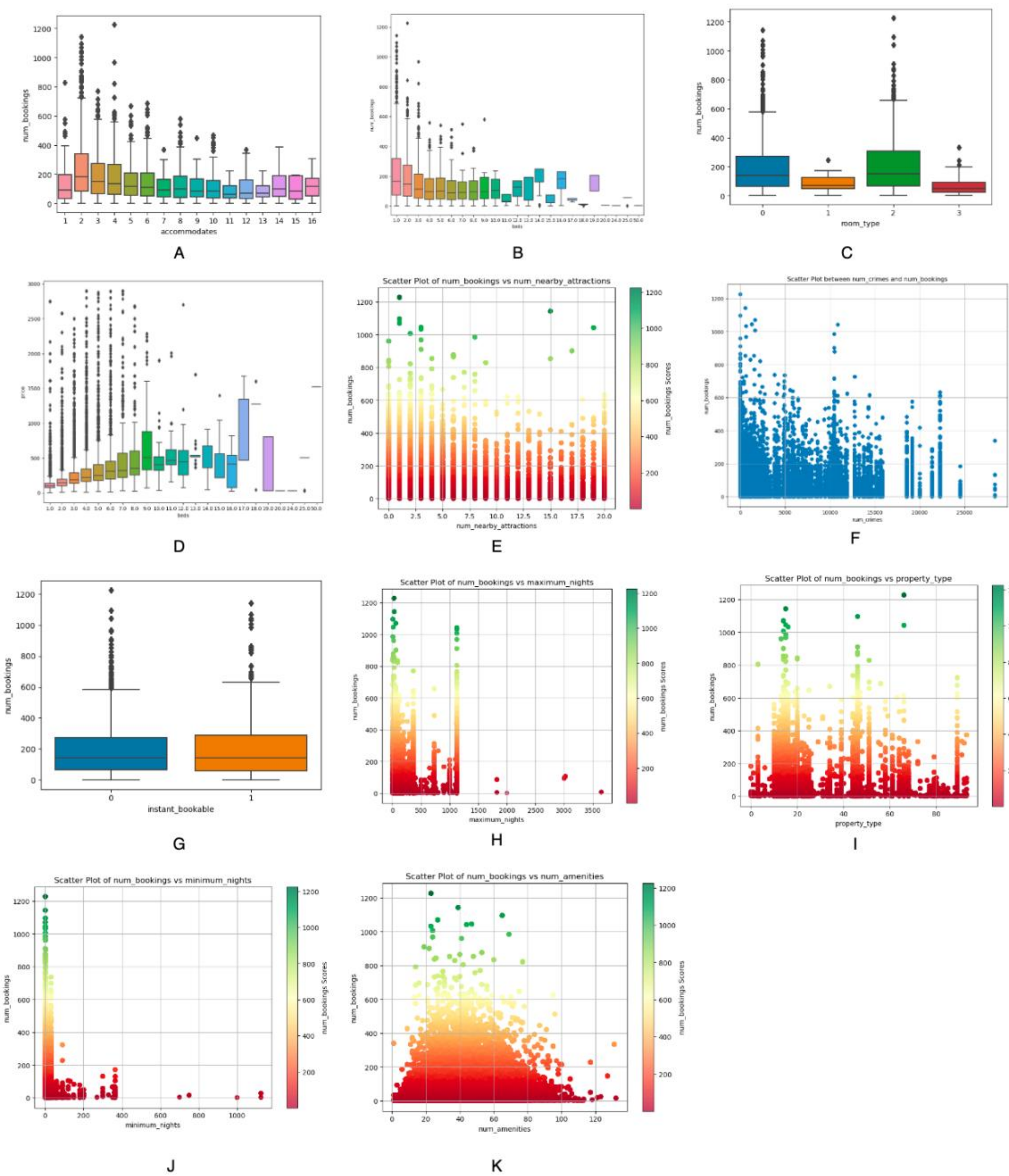

Fig 2. Plots between the number of bookings and the features.

Figure 2 helps us visually analyze how new host features are related to the number of bookings. Accommodations have the highest correlation with the number of bookings. The bar plot (A) between them suggests that the number of bookings is more when the property can accommodate more than one person.

The bar plot (B) between the number of beds and the number of bookings suggests that the number of beds between 3 and 10 attracts similar bookings. In contrast, properties with single or double beds were able to attract more bookings. The barplot (D) extends to (B), analyzing price vs beds. This leads to a more obvious scenario, where the Price increase in the number of beds increases.

The bar plot (C) between room type and number of bookings explains a significant increase in the number of bookings for property entire home/apt and private room compared to hotel rooms and shared rooms. Individually, the whole home/apt has a similar booking trend with a private room; similarly, a hotel room has a similar trend with a shared room.

The scatterplot (E) does not showcase a specific trend. It is more scattered, as seen in the Pearson correlation. Scatterplot (F) talks about the number of crimes; it is seen that as the number of crimes increases, the number of bookings drops. Thus, this is an essential feature when developing a prediction model. Nothing specific can be comprehended from the boxplot (G). Scatterplot (H) and (J) displays maximum_nights and minimum_nights. As the minimum nights increase, the number of bookings drops. This is also partially true for maximum nights. More bookings have been seen when the maximum number of nights is less than 50.

b) Strategy

After analyzing these features, we preprocessed the data. We performed label encoding and normalization. We aim to predict the number of bookings a new Airbnb host will get using the property-based features. Since this is a regression problem, we further explored Machine Learning models. We eventually trained our data on RandomForest, XGBoost, CatBoost, and Linear regression. We applied GridSearchCV to obtain the optimal hyperparameters for each of the models. We then applied a K-fold cross-validation on the method with K=5. We used RMSE (Root Means Square Error) and R-2 value as the evaluation metrics. Once all the models were trained, we also obtained the feature importance of each model and further analyzed it. CatBoost is the best model, and the crucial hyperparameters were interactions = 1000, learning rate = 0.1, loss_function = RMSE, and depth = 7. Although we have performed all the steps for Machine Learning analysis, we found that our models have high RMSE and low R-2 scores since the model cannot converge on the data. This was also one of the reasons we tried to extract more data or data from different modalities, such as the number of crimes and tourist attractions. However, a significant correlation between them and the target variable must be observed. Thus, we could only observe a performance improvement. We also concluded that the categorical features, such as property type and property in each zip code, could be more evenly distributed. Performing this analysis on a larger dataset version will help converge the ML models and thus obtain better performance.

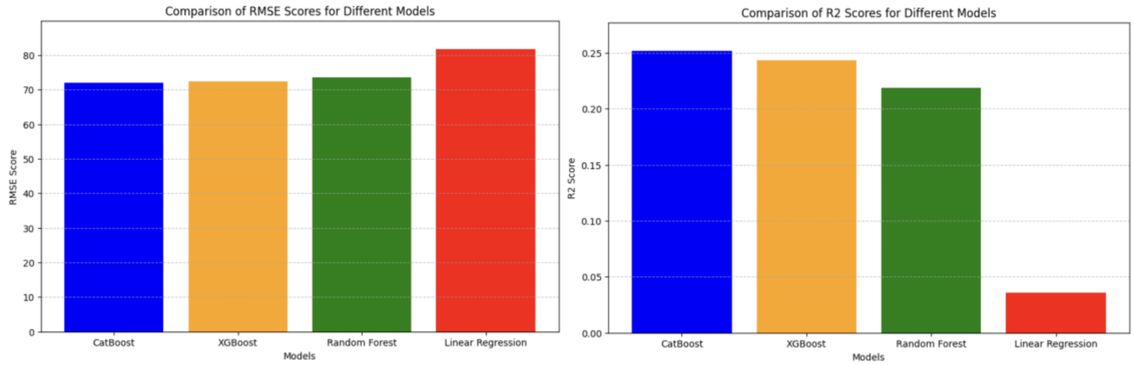

Fig 3. Comparison between ML models using RMSE and R2 metrics.

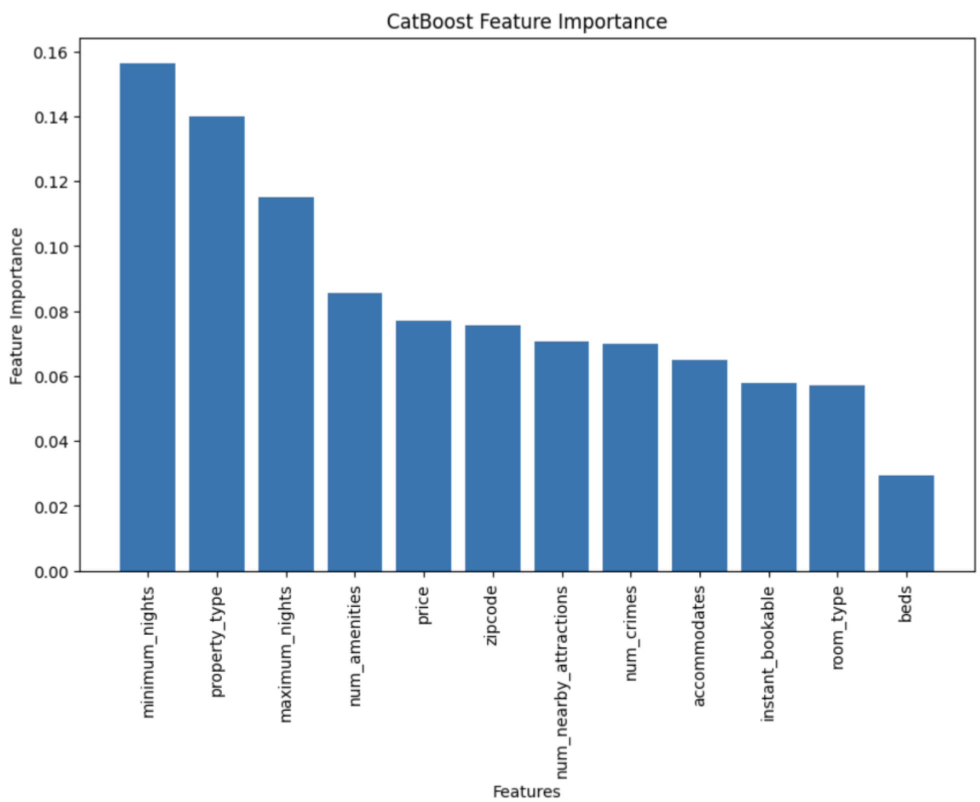

Fig 4. Feature performance of the best ML model- CatBoost suggests the support each feature needs for the final prediction.

2) Existing hosts:

a) Analysis

The heatmap provides a visual representation of the correlation coefficients between different variables. It reveals that "host_is_superhost" has the strongest positive correlation with the number of bookings, followed by "host_since," "host_acceptance_rate," and "host_response_rate." Review-related features also show positive correlations with bookings, while "minimum_nights" and "price" have negative correlations.

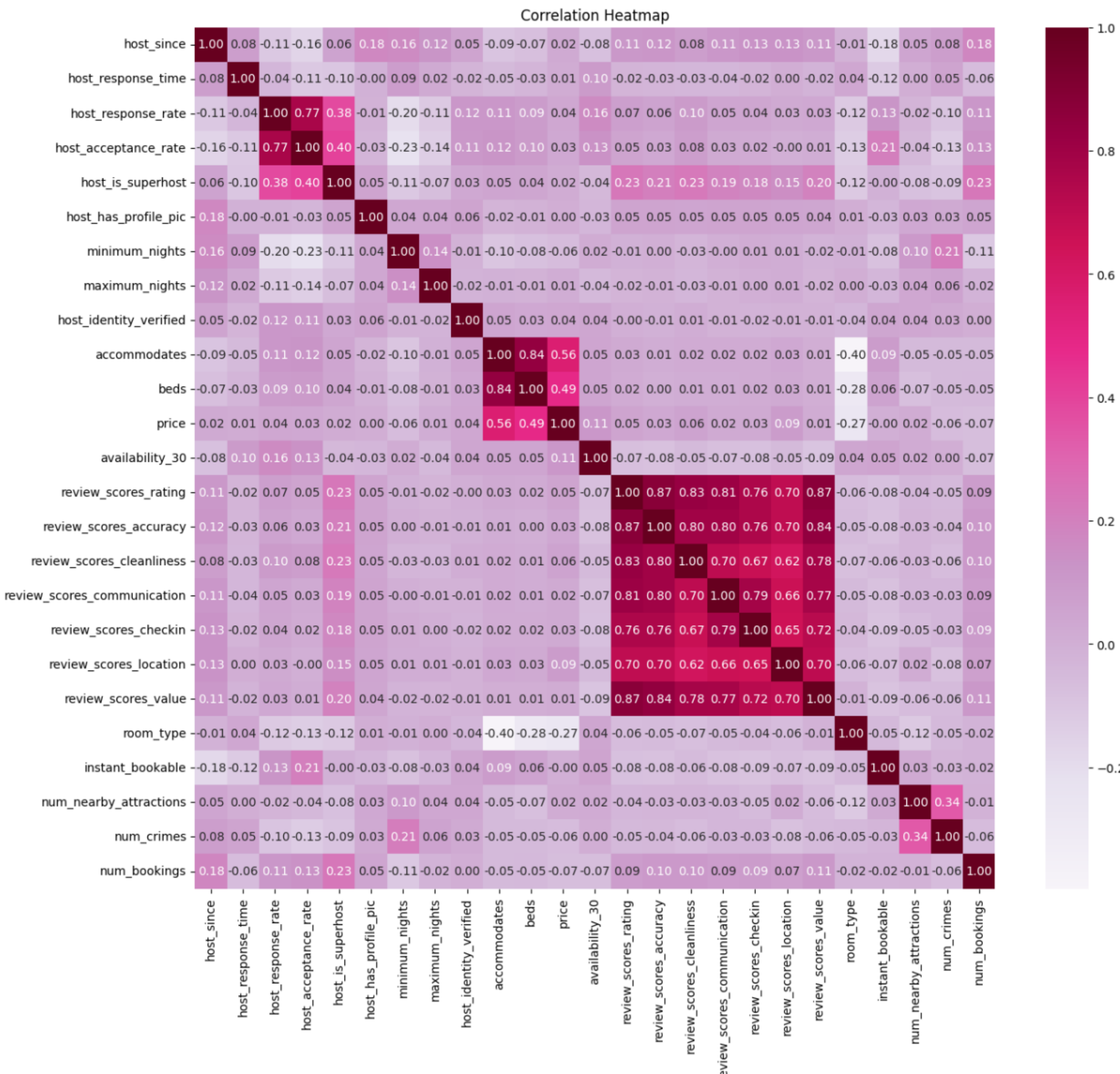

Fig 4. Heatmap representing correlation coefficients between different variables.

The "host_since vs. Num Bookings" plot suggests a positive correlation, indicating that more experienced hosts tend to have more bookings.

The "host_acceptance_rate vs. Num Bookings" plot also shows a positive correlation, suggesting that hosts who accept more booking requests tend to have more bookings.

The "review_scores_value vs. Num Bookings" plot indicates that listings with higher value scores tend to have more bookings.

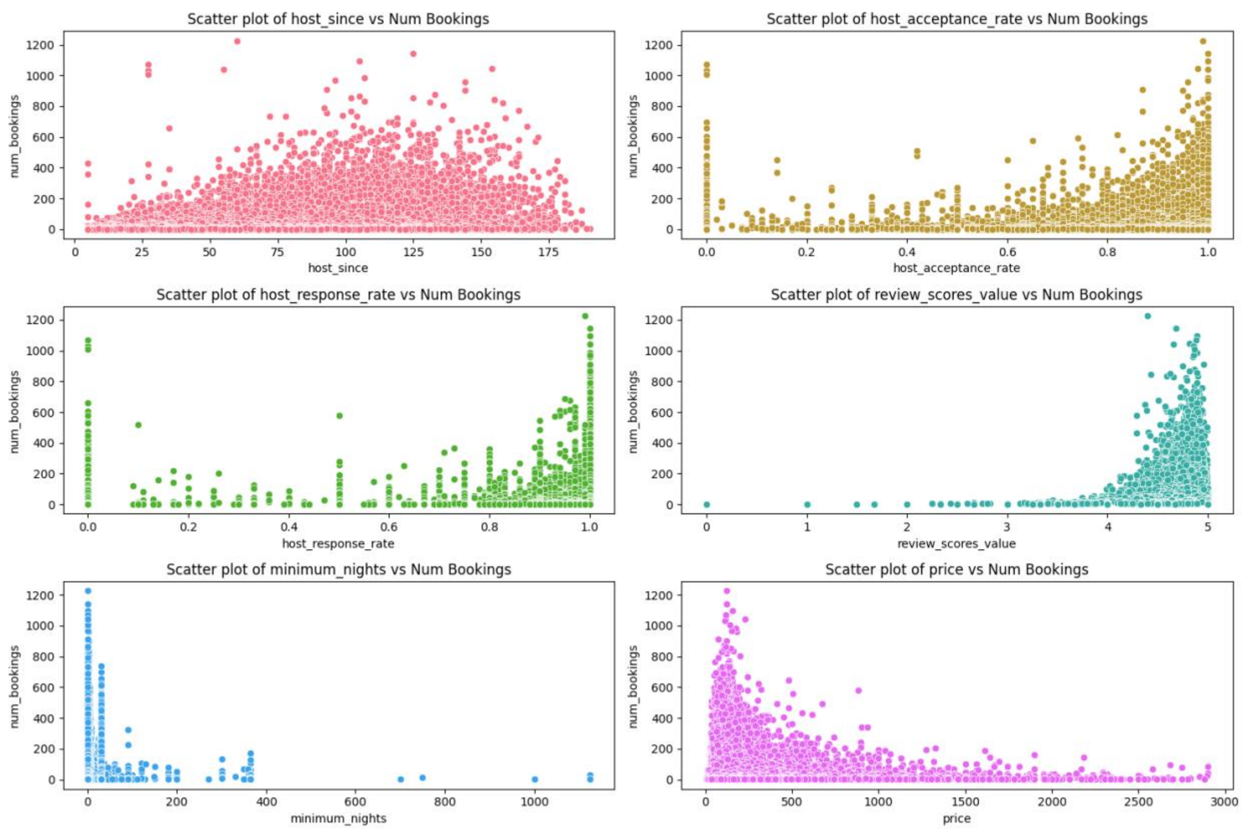

Fig 5. Scatter plots explore the relationships between different features and the number of bookings

The bar chart compares the percentage values for low and high performers across various attributes. It highlights the attributes where high performers significantly outperform low performers, such as "num_bookings," "host_is_superhost," and various review score metrics. Conversely, it shows attributes where low performers have higher values, like "minimum_nights" and "price."

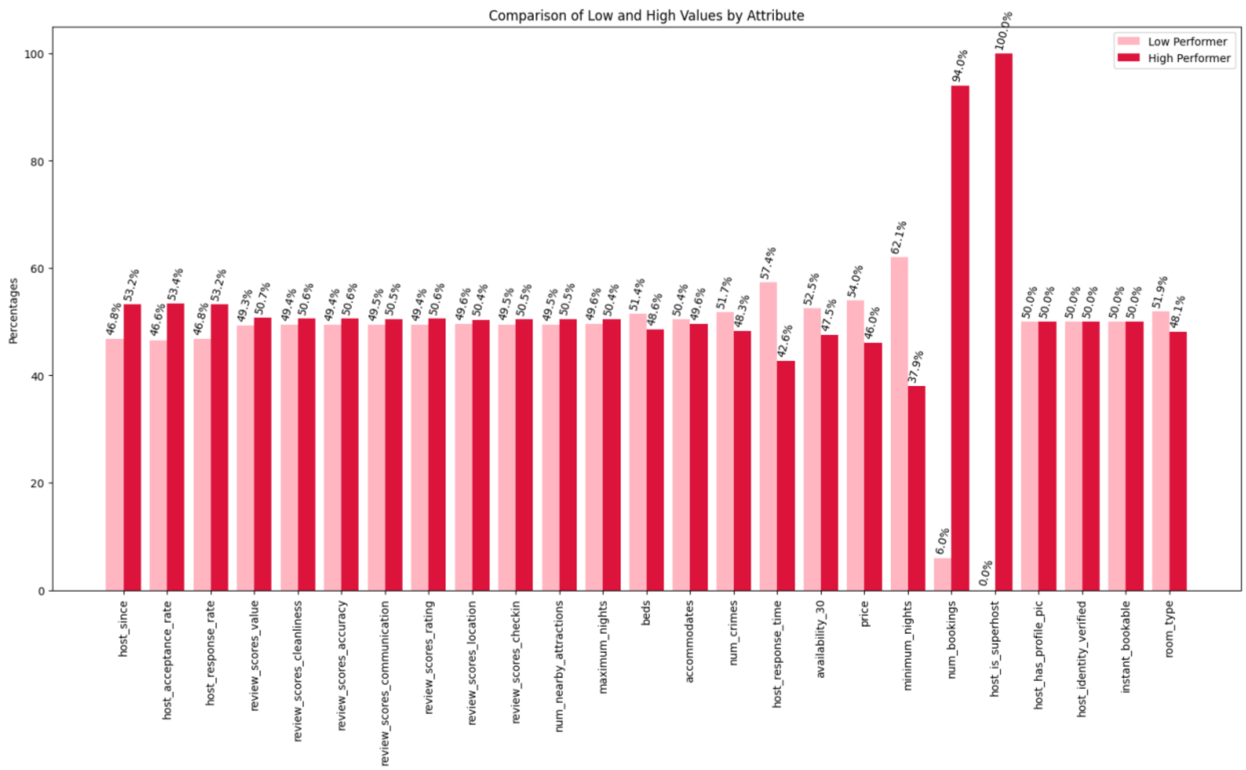

Fig 6. Comparison of Low and High Values by Attribute

b) Strategy

This analysis used location- or region-based clustering to group Airbnb listings. We applied k-means clustering to form clusters and determined the optimal number of clusters (k) using the elbow method. The optimal value for k was 35, resulting in 35 distinct location-based clusters. Below are the details of the methodology and the recommendations for hosts to increase bookings.

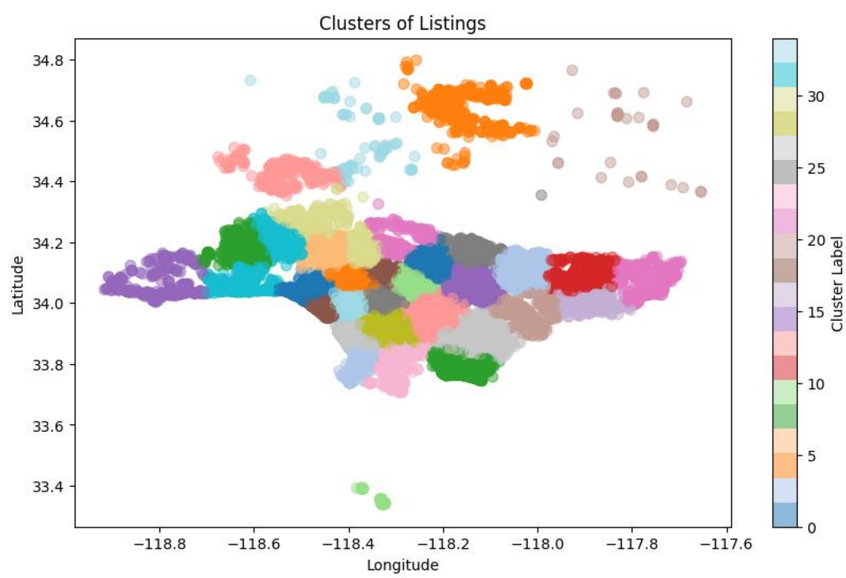

Fig 7. Distinct location-based clusters

Cluster Formation: Each Airbnb listing was assigned to a cluster based on its geographic location using k-means clustering. This approach creates groups of listings that are geographically close to one another.
Elbow Method: We used the elbow method to determine the optimal number of clusters. This method involves plotting the within-cluster sum of squares against the number of clusters and identifying the "elbow" point where the rate of decrease flattens. The optimal k was identified as 35.
Cluster-Based Analysis: For a given Airbnb listing, we assigned it to one of the 35 clusters based on its location. We then divided each cluster into two groups: high and low performers. This division was made by calculating the median number of bookings within each cluster. Listings with bookings below the median were categorized as low performers, while those above the median were categorized as high performers.

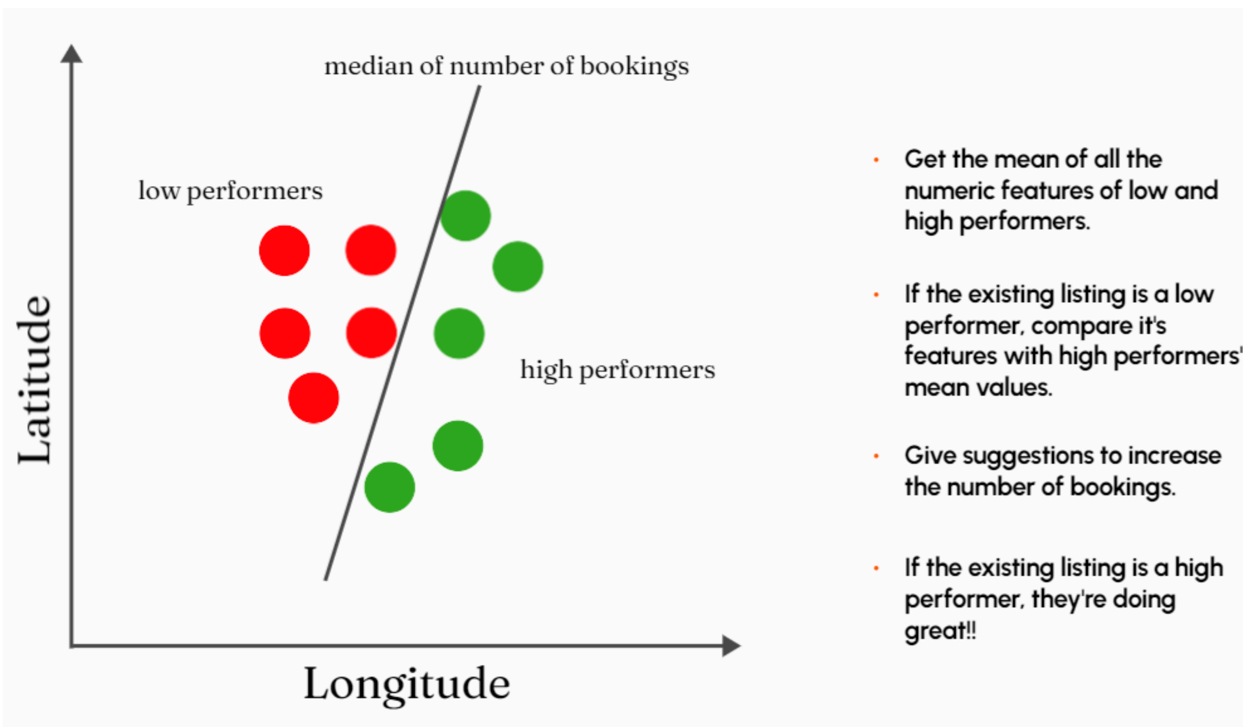

Fig 8. Cluster-Based Analysis

Recommendations for Hosts:
For Low-Performing Hosts: If an Airbnb listing falls into the low-performing category, we recommend comparing its features with the mean values of the high performers within the same cluster.

For High-Performing Hosts: If an Airbnb listing is in the high-performing category, congratulations! Keep up the great work. To maintain your success, continue monitoring guest feedback and ensure your listing remains competitive.

Application

In the culmination of our efforts, we have developed a Streamlit application that caters to novice and experienced hosts in the realm of Airbnb management. Tailored to address the diverse needs of users, our application serves as a comprehensive tool for preliminary analysis and ongoing optimization of Airbnb listings.

Our application offers insights into potential investment opportunities for novice hosts embarking on their Airbnb journey. By leveraging the Pandas library, we facilitate preliminary analysis of key statistics within specific zip codes or enable direct comparison between two areas. Critical factors such as pricing dynamics, listing availability, amenities, property types, and seasonal booking trends are examined to empower hosts to make informed decisions. Moreover, our application goes beyond statistical analysis by providing actionable recommendations, including identifying optimal locations for Airbnb ventures and highlighting nearby attractions to enhance guest experiences.

For hosts evaluating the potential profitability of existing properties, the application integrates predictive modeling using the XGBoost algorithm. By forecasting the approximate number of bookings a property could garner within a year, hosts gain valuable insights into potential returns on investment. This predictive analysis enables hosts to assess the viability of their listings and make informed decisions regarding pricing strategies and resource allocation.

Furthermore, the application offers a comprehensive analysis tool for seasoned hosts seeking to optimize their current listings. By inputting their listing_id, hosts receive a detailed overview of their listing's performance, including areas of strength and opportunities for enhancement. Leveraging KMeans clustering, the application identifies critical features associated with high-performing listings and provides personalized recommendations for optimization.

Implemented on the Streamlit framework and seamlessly integrated with Python, the application prioritizes usability and accessibility. Leveraging the capabilities of the Pandas library and Matplotlib for data manipulation and visualization, the tool provides users with a user-friendly interface and robust functionality.

Project outcomes

Hosts need to set the correct prices for their rentals to make more money. Hosts can charge correctly by checking what others are charging, seeing when people want to book, and knowing the market trends. This can help them get more bookings and cash, making their Airbnb business more successful. Using tools that do things automatically can save hosts time and money. These tools include talking to guests, scheduling cleaning, and managing bookings. Hosts can spend less time on tedious tasks and focus on other important stuff. It makes running an Airbnb smoother and can boost profits.

In the busy world of Airbnb, hosts need to stand out to get more bookings. Analyzing data can help hosts understand what guests like and what's happening in the market. With this info, hosts can make their listings more appealing, beat the competition, and get more bookings. This can help hosts make more money and fill up their rentals more often when the host wants their place empty. To make sure their rentals are always booked, hosts need to make them attractive and available. They can do this by adjusting prices based on demand, offering deals during slow times, and making their listings look great. Hosts can earn more money by keeping their places in demand and making their Airbnb business more profitable.

Making guests happy is super essential for hosts. By knowing what guests want and giving them a great experience, hosts can get good reviews, more repeat bookings, and attract new guests. From providing tips on local hotspots to making check-ins easy, hosts can use data to make their guests feel special. This can lead to more bookings and more money in the long run. Using data wisely can help hosts stay ahead of the game. By looking at past trends and what's happening now, hosts can spot problems or chances to make more money early on. They can change prices or update their listings to match what guests want. This smart way of working can help hosts avoid problems and strengthen their business. Growing an Airbnb business means managing lots of listings without sacrificing quality. With tools that do things automatically and keep all the info in one place, hosts can handle many rentals easily. This means they can expand their business, make more money, and do it all without making guests unhappy. It's a great way for hosts to succeed in the competitive world of vacation rentals.

The code

Link to the code:
https://drive.google.com/drive/folders/1oEhjMGmdB5KLW6EBmzc5YB0hMz7sXkGt?usp=sharing
Link to the presentation: https://pitch.com/v/revenue-optimization-for-airbnb-hosts-h8bu2u

The core of the application lies in the app.py file. It is the primary driver of all functionalities, with the essential code required to power the entire system. This pivotal file seamlessly integrates various features by orchestrating the execution of functions stored in files such as compare.py, analyze.py, improve.py, and make_suggestion.py. These files contain the specific code implementations that enable the range of functionalities within the application. The data folder contains all the data files we have used for the application, to make analysis, and also for the predictive modeling.

Installing all necessary dependencies outlined in the requirement file is essential to initiate the application.

To install all dependencies:

```
pip install -r requirements.txt
```

Following the installation process, the application can be executed using the command:

```
streamlit run app.py
```

This command triggers the activation of the application, displaying different features we have discussed.

Learning experiences

- This project provided hands-on experience in data analysis, including data manipulation, statistical analysis, and machine learning techniques. Working with real-world datasets enabled the application of theoretical knowledge to practical scenarios.
- Using machine learning models like XGBoost and KMeans provided insights into model training, evaluation, and deployment, which are essential aspects of machine learning.
- Building a Streamlit application integrated with Python helped us gain proficiency in structuring and modularizing code, managing dependencies, and creating user-friendly interfaces.

Appendix

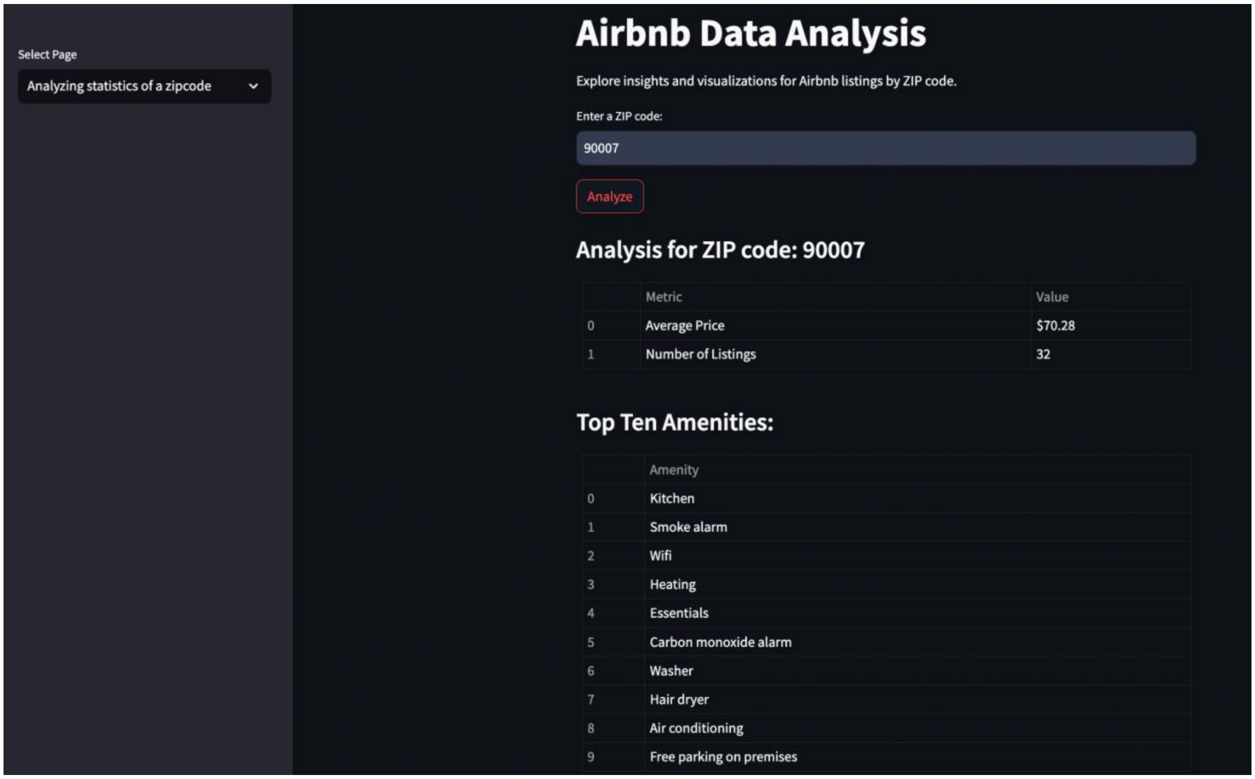

Fig 9. Analyzing a particular zip code

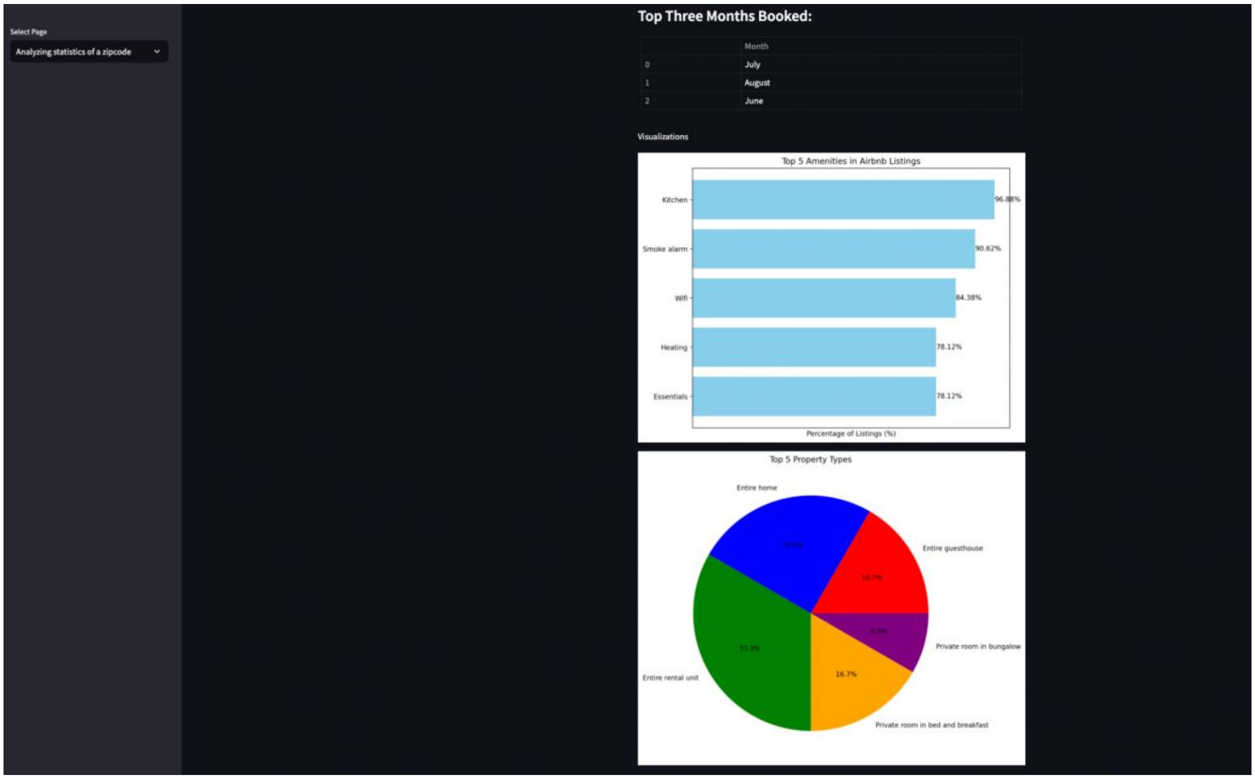

Fig 10. Analyzing a particular zip code with more features

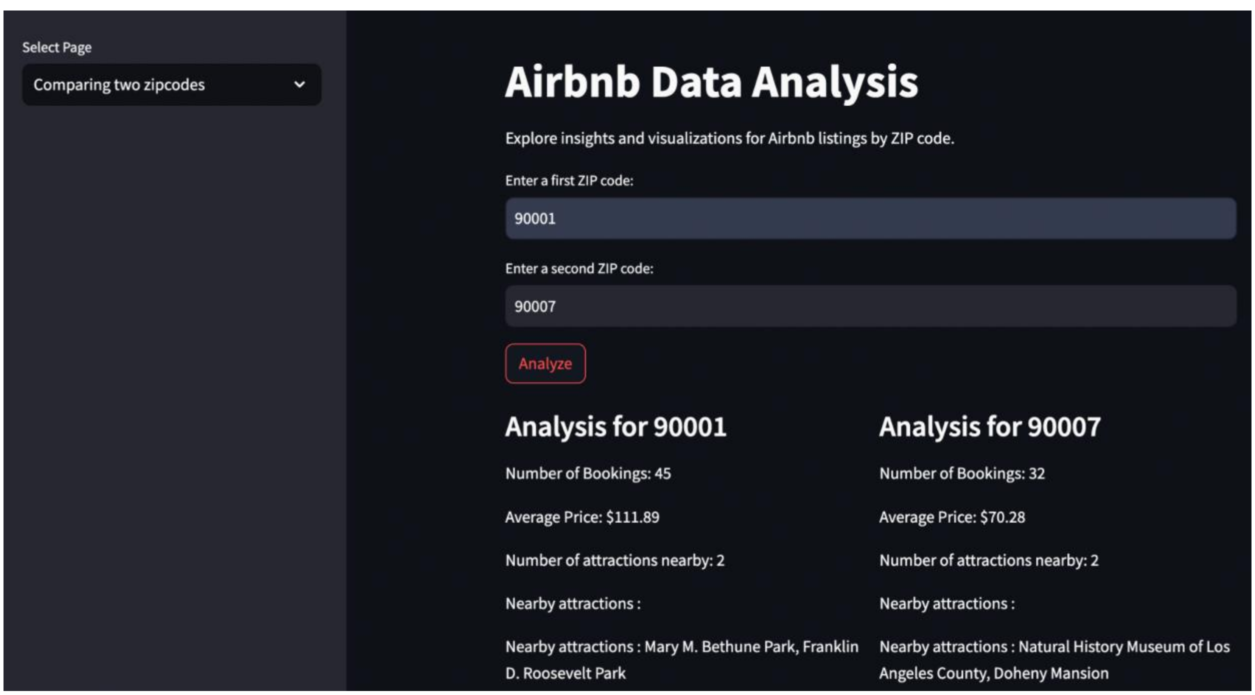

Fig 11. Comparison of two different area codes

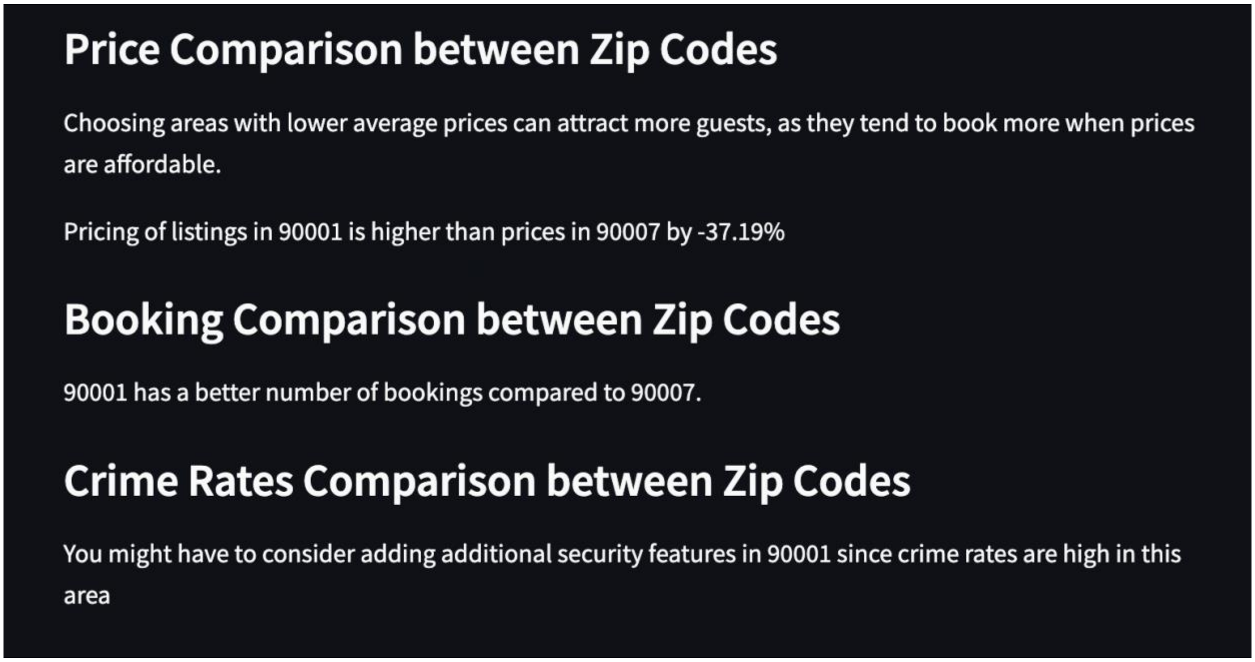

Fig 12. Comparison of features and recommendations of two different area codes

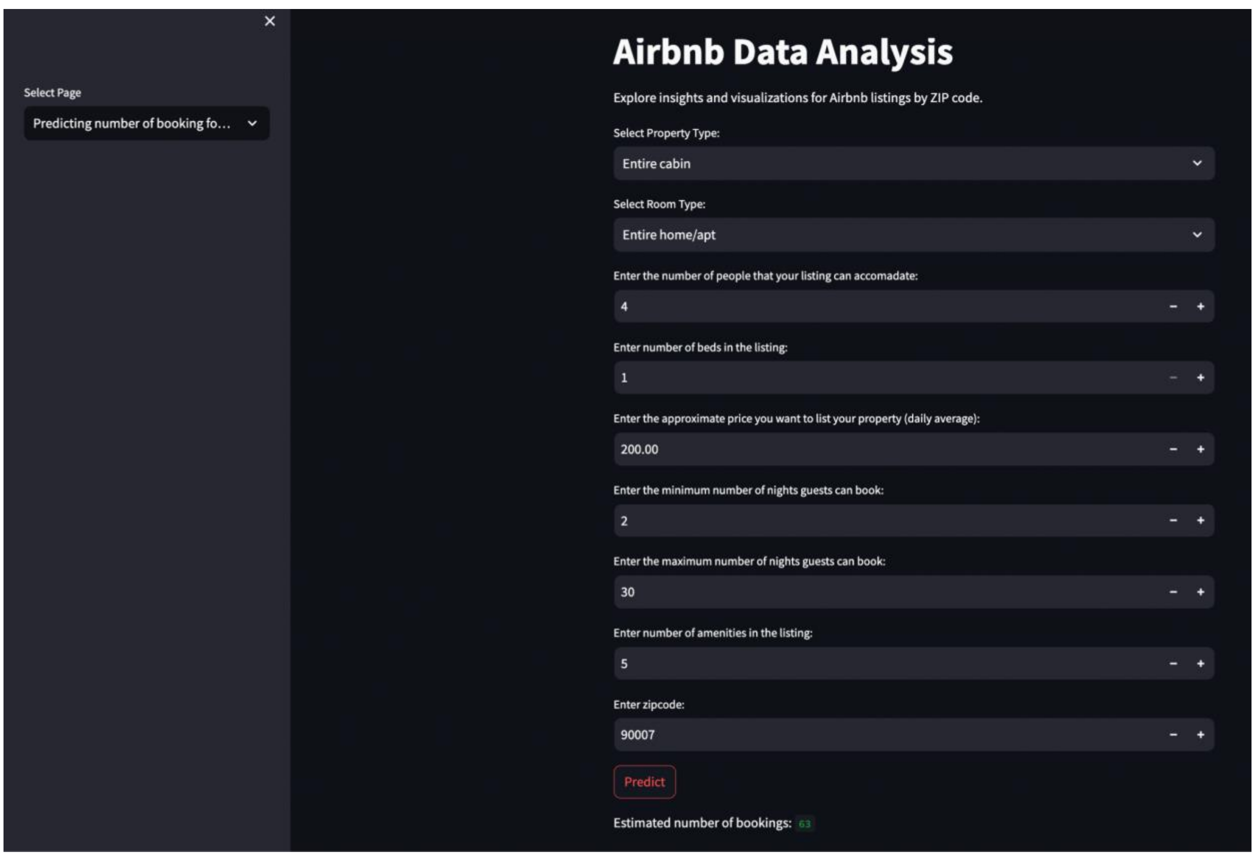

Fig 13. Predicting the number of bookings for new hosts

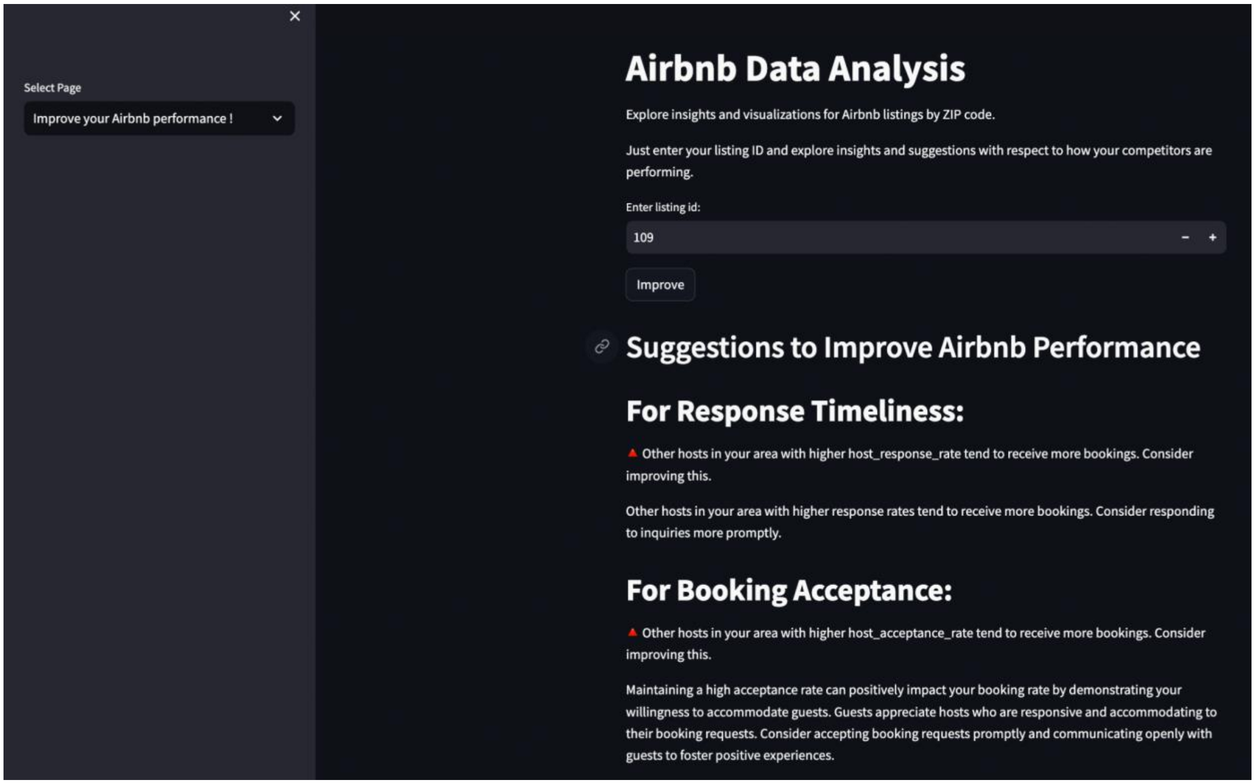

Fig 14. Recommendations to seasoned hosts on how to improve their current listings