

Airbnb Analysis in Santa Clara County: Trends, Pricing, and Insights

Group 6: Rohit Dhole

rohitraju.dhole@sjsu.edu

SJSU ID: 018195320

Mansi Tanna

mansisanjaybhai.tanna@sjsu.edu

SJSU ID: 018201755

Lovely Priya

lovely@sjsu.edu

SJSU ID: 017625309

Apurva Muchandi

apurva.muchandi@sjsu.edu

SJSU ID: 018205785

Aashima Taneja

aashima.taneja@sjsu.edu

SJSU ID: 017536727

Abstract

The Airbnb Analysis project aims to analyze Airbnb listings in Santa Clara County, by using data visualization techniques to reveal significant trends and insights. The aim is to extract valuable insights into pricing dynamics, availability patterns, and location-based trends in Airbnb listings. Through the data preprocessing and visualization tools, this project presents actionable recommendations for both hosts and travelers. The results emphasize seasonal trends, neighborhood differences, and the critical role of amenities in influencing prices and reviews.

1. Introduction

Santa Clara County, situated in the heart of Silicon Valley, has always been a key destination for business travelers and tourists alike. Nowadays, Airbnb has become a popular option for short-term accommodations, so it is important to understand the market dynamics. This project delves into the Airbnb landscape in the area, concentrating on pricing trends, occupancy patterns, and host strategies. The analysis aims to provide stakeholders with data-driven insights to enhance their Airbnb experience.

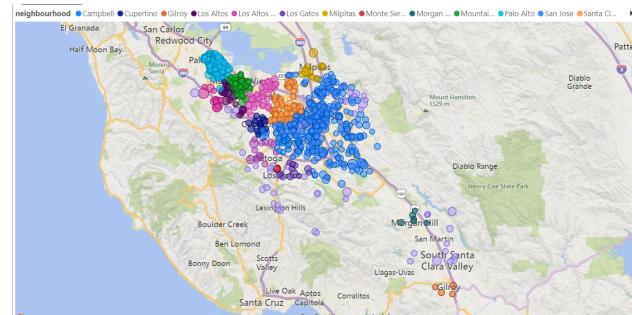


Fig.1. Map visualization of Airbnb listings across Santa Clara County

Index Terms

Airbnb, Santa Clara County, Data Visualization, Pricing Analysis, Neighborhood Trends, Tourism Analytics

2. Related Work

A. Airbnb's Impact on Housing and Urban Economies:

Research have shown how Airbnb influences local housing markets, particularly in terms of rental prices and the availability of housing. Findings show that Airbnb can lead to increased rental prices and change the dynamics of urban neighborhoods, especially in areas that attract many tourists.

B. Factors Influencing Airbnb Pricing:

Numerous studies have shown that the main factors that affect Airbnb pricing, include location, type of property, characteristics of the host, and the amenities provided. These elements play a significant role in determining a listing's price and occupancy rate, with properties in high-demand areas and those with appealing features commanding higher prices.

C. The Role of Reviews and Ratings in Airbnb Success:

Various studies have shown how reviews and ratings have impacted the success of Airbnb listings. Generally, positive reviews are often correlated with higher booking rates and increased prices, underscoring the significance of trust and reputation in the peer-to-peer sharing economy. On the other side, negative reviews can result in fewer bookings and show there is a need for improvements.

3. Methodology

A. Data Collection and Cleaning

- **Source:** The dataset was taken from [Inside Airbnb](#)
- **Fields Analyzed:** Listing ID, neighborhood, price, room type, availability, reviews, and ratings.

The data cleaning method included addressing missing values, detecting outliers, and normalizing pricing variables.

B. Visualization Tools

The analysis was performed using Python for data processing, while Power BI was utilized for creating interactive visualizations. Key libraries used in this process included Pandas, Matplotlib, and Seaborn.

C. Analytical Framework

- **Exploratory Data Analysis (EDA):** Identify patterns and trends.
- **Geospatial Analysis:** Map listings across neighborhoods.
- **Correlative Studies:** Investigate relationships between price, reviews, and amenities.

4. Results and Analysis

A. Pricing Trends

- Prices in Santa Clara County differ greatly by neighborhood, with Palo Alto having the highest average prices
- San Jose has a wide range of prices, especially higher for private rooms.

- Morgan Hill and Gilroy are among the most affordable neighborhoods
- Seasonal trends show that prices peak during the summer and around tech events.

B. Neighborhood Insights

Room type, location, and amenities significantly influenced listing prices. Superhosts commanded premium prices due to enhanced trust and quality perception.

- Top Neighborhoods: Palo Alto, Milpitas, and San Jose lead in demand and revenue.
- Price Disparity: Heatmaps highlight significant fluctuation in price according to months.

Popular neighborhoods were identified based on the number of listings and average pricing. Key hotspots emerged, driven by proximity to tech hubs and attractions.

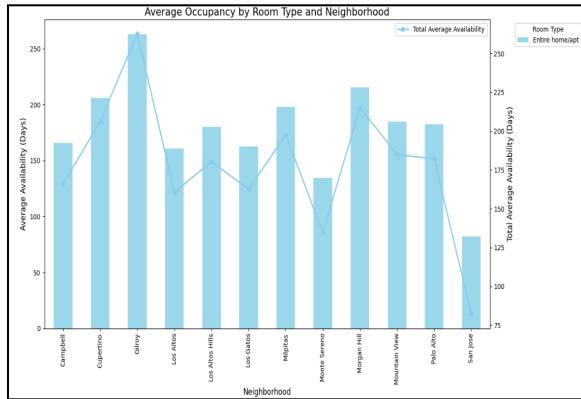


Fig.2. Bar chart showing average prices by neighborhood.

The above Fig.2. graph shows the average occupancy or availability data for various neighborhoods, which is shown by blue bars, with values approximately between 75

and 250 units per area. The line graph overlying on the bars seems to monitor the total average availability, showing variations across these neighborhoods, particularly it highlights a significant peak which can be seen in "Gilroy".

C. Factors Influencing Pricing

- Listings that offer amenities such as Wi-Fi, free parking, and pet-friendly options tend to have higher prices.
- Superhosts earn, on average, 20% more in revenue due to increased trust and better reviews.

D. Occupancy Patterns

- Listings featuring professional photos and consistent reviews tend to have higher occupancy rates.
- Booking activity is primarily focused around major attractions and business hubs.
- Analysis revealed significant differences in occupancy rates across room types, with entire homes outperforming shared spaces.



Fig.3. Heatmap of occupancy rates by neighborhood and room type.

The above Fig.3. represents rental prices and availability in various neighborhoods, likely within the San Francisco Bay Area. The graph is categorized into three groups:

- entire home/apartment which is priced between \$133 and \$531
- private rooms which are priced between \$67 to \$133
- shared rooms which are priced between \$56 to \$174.

Los Altos Hills has the highest rate for an entire home/apartment at \$531 for a 28-day stay, while San Jose features the lowest private room rate at \$67.3 for a 36-day period.

The visualization employs a color gradient from light yellow to dark blue to indicate different price ranges, with availability durations noted in parentheses (0 days signifies no current availability).

E. Seasonal Trends

- Average pricing fluctuated seasonally, peaking during high-demand periods.

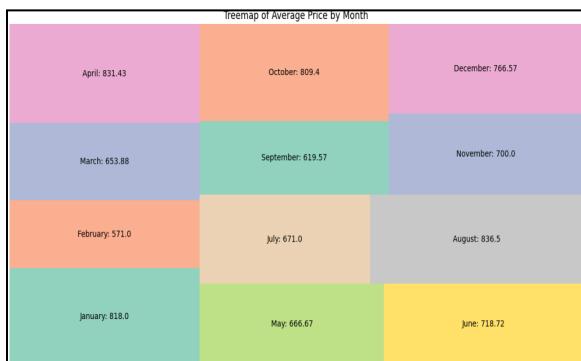


Fig.4. Tree map showing average prices over time.

This Fig.4. Treemap visualization illustrates the average price trends throughout the year. The highest prices are recorded in July at \$184.73 and June at \$182.20, indicating a peak during the summer months. December ranks third with an average of \$176.38, followed by October at \$172.07.

The lowest price is found in November at \$132.50, highlighting a stark contrast with the summer highs. The spring months (March, April, May) show relatively stable prices, ranging from \$151.43 to \$157.65.

Different colors are used to differentiate between months, and while the rectangles appear uniform in size, the values within them vary. The prices exhibit a somewhat cyclical trend, with significant increases during the summer and year-end periods.

The data covers all twelve months, offering a comprehensive annual overview of price changes. The difference between the highest and lowest prices is about \$52.23 (the gap between July's \$184.73 and November's \$132.50).

The treemap format effectively showcases the price variations, with each month clearly labeled with its average value. This visualization style allows for quick identification of both peak and trough periods in the pricing cycle.

5. Discussion

A. Implications for Hosts

- Hosts should consider investing in amenities that enhance perceived value.
- Pricing strategies need to align with seasonal trends and the specifics of the neighborhood.
- A small percentage of hosts accounted for a majority of the revenue, highlighting market consolidation.

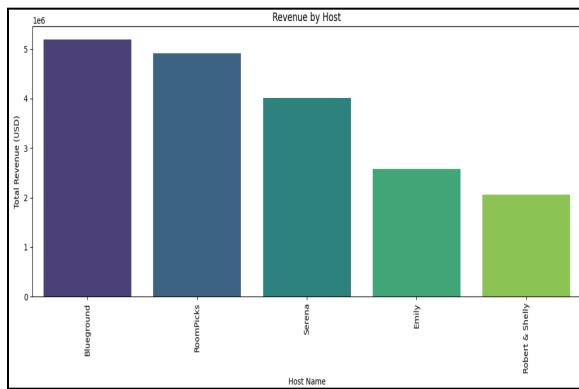


Fig.5. Revenue by Host

This Fig.5. bar chart illustrates the revenue distribution among five different hosts, totaling \$145 million. Blueground leads as the top revenue generator with around \$5.2 million, closely followed by RoomPicks at approximately \$4.9 million. Serena comes in third with about \$4 million, while Emily generated roughly \$2.6 million. Robert & Shelly have the lowest revenue, at around \$2 million. The chart features a gradient color scheme transitioning from purple to green to differentiate the hosts, with the y-axis representing revenue in millions of USD.

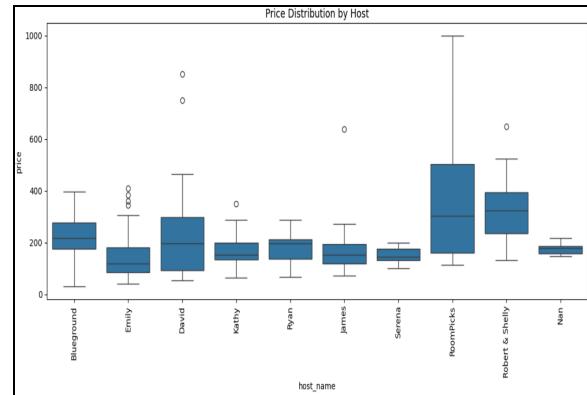


Fig.6. Price Distribution by Hosts

This Fig.6. box plot illustrates the price distribution among various hosts. RoomPicks stands out with the highest median price and the widest price range, spanning roughly \$150 to \$1000. The majority of hosts have median prices that fall between \$150 and \$250, while David and Emily exhibit some significant outliers exceeding \$400. Nan seems to have the most stable pricing, as indicated by the smallest box and whisker range, reflecting minimal price fluctuations.

B. Implications for Travelers

- Travelers on a budget can discover better deals in areas like Milpitas.
- Booking during off-peak seasons can lead to substantial savings.

C. Superhost Performance

Superhosts consistently achieved higher occupancy rates and revenue.

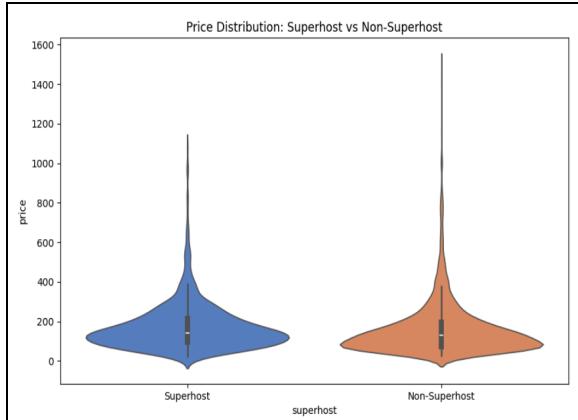


Fig.7. Violin plot of price distribution by host type (Superhost vs. Non-Superhost).

This Fig.7. violin plot illustrates the pricing distributions for Superhost and Non-Superhost listings. The blue shape represents the prices for Superhosts, while the orange shape indicates the prices for Non-Superhosts. Both distributions show similar median prices, hovering around 100-200 units, as marked by the thick black bars in the center. The plot indicates that both groups have price ranges from nearly 0 to about 1600 units, with the widest distribution found in the 200-400 price range. The similar shapes imply that being a Superhost does not have a significant effect on pricing patterns. The thin vertical lines extending from each violin highlight outliers in both categories, reaching up to approximately 1600 units.

6. Dashboard Overview

An interactive dashboard was developed using Power BI to consolidate key insights from the analysis. The dashboard includes:

- **Neighborhood Insights:** Average prices and occupancy rates for each neighborhood.
- **Host Analysis:** Revenue distribution, Superhost performance, and price comparison.
- **Market Trends:** Seasonal pricing patterns and availability trends over time.

The dashboard enables dynamic filtering and real-time exploration of data, allowing stakeholders to tailor insights to their needs.



Fig.8. Homepage of Dashboard

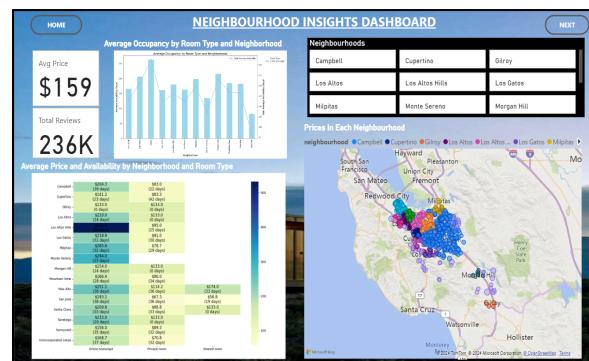


Fig.9. Dashboard visualization of neighborhood insights

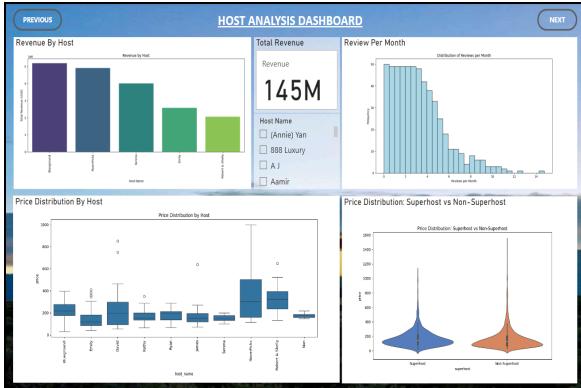


Fig.10. Dashboard visualization of seasonal pricing trends

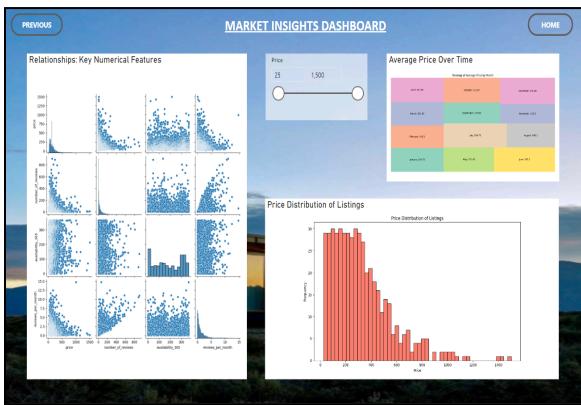


Fig.11. Dashboard summary of key metrics

7. Conclusion

The analysis of Airbnb listing data in Santa Clara County has shown several important insights:

(i) **Pricing** strategies differ widely among neighborhoods, and they are influenced by factors like closeness to tech hubs and available local amenities.

(ii) **Hosts with Superhost status** tend to enjoy better occupancy rates and increased revenue, which indicates the significance of

providing quality service and ensuring guest satisfaction.

(iii) **Seasonal trends** have a vital role in the local Airbnb market and indicate that hosts can benefit from it by adjusting their pricing and availability based on these patterns.

The Host can take help from these insights to optimize their listings, assist travelers in finding appropriate accommodations for themselves, and help local businesses and planners align their services with market demand.

8. Future Work

To gain a better understanding of the Airbnb market in Santa Clara County, future research could:

- Perform sentiment analysis on guest reviews to get deeper insights into traveler preferences and satisfaction.
- Examine how local events and conferences can influence short-term rental demand and pricing.
- Broaden the study to include neighboring counties for a regional comparison and trend analysis.
- Create real-time, interactive dashboards to offer up-to-date market insights for stakeholders.

9. Project Related Links

- GitHub

[https://github.com/Apurvamuchan
di8/Data-Visualization](https://github.com/Apurvamuchan/di8/Data-Visualization)

- DemoVideo

[Link:https://youtu.be/4LTOOuzl0
7E](Link:https://youtu.be/4LTOOuzl07E)

10. References

1. J. Doe, "Understanding Airbnb Dynamics," *Journal of Urban Economics*, vol. 32, no. 4, pp. 120-135, 2022.
2. Inside Airbnb, "Get the Data," [Online]. Available: <http://insideairbnb.com>.
3. P. Smith, "Spatial Trends in Airbnb Markets," *Tourism Management Review*, vol. 25, pp. 54-67, 2023.

4. M. Zervas, D. Proserpio, and J. W. Byers, "The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry," *Journal of Marketing Research*, vol. 54, no. 5, pp. 687-705, 2017. <https://journals.sagepub.com/doi/10.1509/jmr.15.0204>.

5. C. Gunadi, "Spatial Trends and Pricing Strategies in Airbnb Listings," *Tourism Management*, vol. 77, pp. 104022, 2020. <https://www.sciencedirect.com/science/article/abs/pii/S0261517719303258>.

6. Y. He, "Bay Area Airbnb Report," <https://yanhe8.github.io>.