Airbnb Investment Opportunity Analysis (Mexico) Report

Project Overview

The project uses official Airbnb datasets and modern data analytics tools to turn raw data into clear insights that support business and investment planning.

Tools: Python, SQL, Tableau

Purpose of the Analysis

The analysis was carried out to understand Airbnb market performance across neighborhoods in Mexico and to identify locations with strong demand, stable occupancy, and revenue potential. The goal was to support data-driven decision-making for investors, hosts, and stakeholders by answering key business questions related to demand concentration, pricing strength, market competition, and overall listing performance.

By combining Python, SQL, and Tableau, the project aimed to move beyond raw data exploration and provide clear, actionable insights that can guide Airbnb investment strategies and hosting decisions.

Tool 1 Python:

Data Loading and Validation

The first step involved loading all datasets into Python using the Pandas library. Once loaded, basic validation checks were performed to confirm the number of records, data types, and completeness of each dataset.

These checks ensured that:

- All datasets were successfully imported
- Column data types were appropriate for analysis
- No missing or corrupted values were present

This step confirmed that the data was clean and ready for analysis.

Data Understanding and Cleaning

To fully understand the structure of the data, exploratory checks were conducted to examine column meanings, data ranges, and distributions.

The date columns doesn't have the right datatype so it was converted to the right datatype using pandas. In the calendar dataset, availability was represented using:

- `t` (true) means the property is available for booking on that date
- `f` (false) means the property is already booked on that date

For analysis purposes, this text-based representation was converted into a numeric format:

- `1` = booked (false)
- `0` = available (true)

This conversion made it easier to calculate booking frequency and occupancy trends using Python.

Key Business Questions

1. Which areas in Mexico show the highest Airbnb demand?

Demand Analysis

Demand was analyzed using guest review activity as a proxy. Reviews were linked to listings and neighborhoods, and the total number of reviews per neighborhood was calculated.

Insight: More reviews means higher demand
From the result Roma Norte had the highest reviews, which means it is the neighborhood with th e highest demand.

2. Where do properties achieve strong occupancy and pricing performance?

Occupancy and pricing performance were evaluated using average occupancy rates, nightly prices, and estimated monthly revenue.
The analysis shows that Polanco, Roma Norte, and Condesa achieve the strongest performance, combining high occupancy with higher pricing levels, which results in higher estimated monthly revenue compared to other neighborhoods.

3. Which locations have lower competition compared to demand?

Demand and competition were compared using the number of reviews per listing.
The results show that Polanco, Roma Norte, and Condesa have the highest demand relative to competition, indicating that listings in these areas receive more demand per property.
Neighborhoods such as Narvarte and Juárez show lower demand-to-competition ratios, suggesting a more competitive market relative to demand.

4. Which areas offer the best investment opportunities for new Airbnb listings?

Investment opportunities were evaluated using a composite ROI score that combines estimated monthly revenue, occupancy rate, and pricing balance.
The analysis shows that Roma Norte, Narvarte, and Coyoacán offer the strongest investment opportunities, as they record the highest average ROI scores alongside solid occupancy levels and competitive revenue performance.
While areas such as Condesa and Polanco also perform well, their slightly lower ROI scores suggest marginally higher competitive or pricing pressure compared to the top-ranked locations.

Tool 2: SQL

Data Importation

A database was created named (Airbnb), the csv files was then imported into the created database.

The tables was confirmed using select query in sql.

Data Cleaning

The date column was cleaned using sql queries to change it to normal sql date format.

Key Performance Indicators using SQL

1. **Occupancy rate**

Insight: This measures how often listens are booked and high values indicates stronger booking demand. From the queries average occupancy rate was calculated and Rome Norte had highest average occupancy rate which means it's the neighborhood with the highest demand.

2. **Average nightly price per neighborhood**

Insight: This shows pricing strength by location and helps identify premium vs budget areas. From the result Condesa had the highest average average nightly price although the differences between the average nightly prices for each neighborhood isn't much.

3. **Estimated monthly revenue by neighborhood**

The average monthly revenue was calculated for each neighborhood, it's a strong indicator of financial performance.

4. **Number of listings per neighborhood (competition level)**

Insight: This measures market saturation and higher count is equal to higher competition.

5. **Total Reviews per Neighborhood (Basic Demand)**

Insight: It shows neighborhoods with the most guest activity, from the result Roma Norte had the highest reviews, which means it is the neighborhood with the highest demand.

SQL was used to compute key performance indicators at the neighborhood level, including occupancy rate, average nightly price, estimated monthly revenue, and the number of active listings. In addition, the reviews and listings tables were joined to measure demand using review activity as a proxy for booking behavior.

Tool 3: Tableau

Market & Performance Dashboard Interpretation

Tableau was used to visually present and validate insights derived from Python and SQL analysis. The interactive dashboard provides a clear overview of Airbnb performance across neighborhoods, focusing on demand, occupancy, pricing, and revenue.

The dashboard shows an average occupancy rate of approximately 49%, indicating steady booking activity across the market. Total reviews **(**328,600**)** reflect strong overall guest engagement.

Roma Norte stands out with the highest demand and occupancy levels, while Narvarte and Polanco also perform strongly. Revenue analysis shows that Narvarte, Condesa, and Roma Norte generate the highest average monthly revenue. Pricing analysis confirms that entire homes/apartments command significantly higher nightly prices than private rooms.

Conclusion

The analysis shows that Airbnb performance varies by neighborhood, with Roma Norte, Narvarte, and Polanco demonstrating the strongest demand, occupancy, and revenue potential. These findings highlight clear location-based opportunities for Airbnb hosting and investment.

Recommendations

Investors and hosts should prioritize Roma Norte, Narvarte, and Polanco and focus on entire home/apartment listings to maximize pricing and occupancy performance.