Mexico Real Estate Price Analysis Project Report

Objective: The main goal of this project was to analyze a dataset of 21,000 property listings in Mexico, sourced from the real estate platform Properati.com. The aim was to investigate whether property sale prices are more influenced by property size or location.


Key Learnings:

1. Organizing and cleaning data using Python data structures and the `pandas` library.
2. Importing and combining multiple CSV files.
3. Cleaning and transforming data (e.g., removing currency symbols, handling missing values).
4. Creating visualizations including scatter plots and box plots to analyze patterns.
5. Using correlation coefficients to quantify relationships between variables.


Data Preparation:

- Three CSV files were loaded and combined (`df1`, `df2`, `df3`).
- Null values and irrelevant columns (e.g., `Unnamed: 0`) were removed.
- The `price_usd` column was cleaned by stripping the "$" symbol and commas.

Exploratory Data Analysis:

- Distribution plots and histograms were used to understand the spread of `price_usd` and `area_m2`.

Mean House Price by State
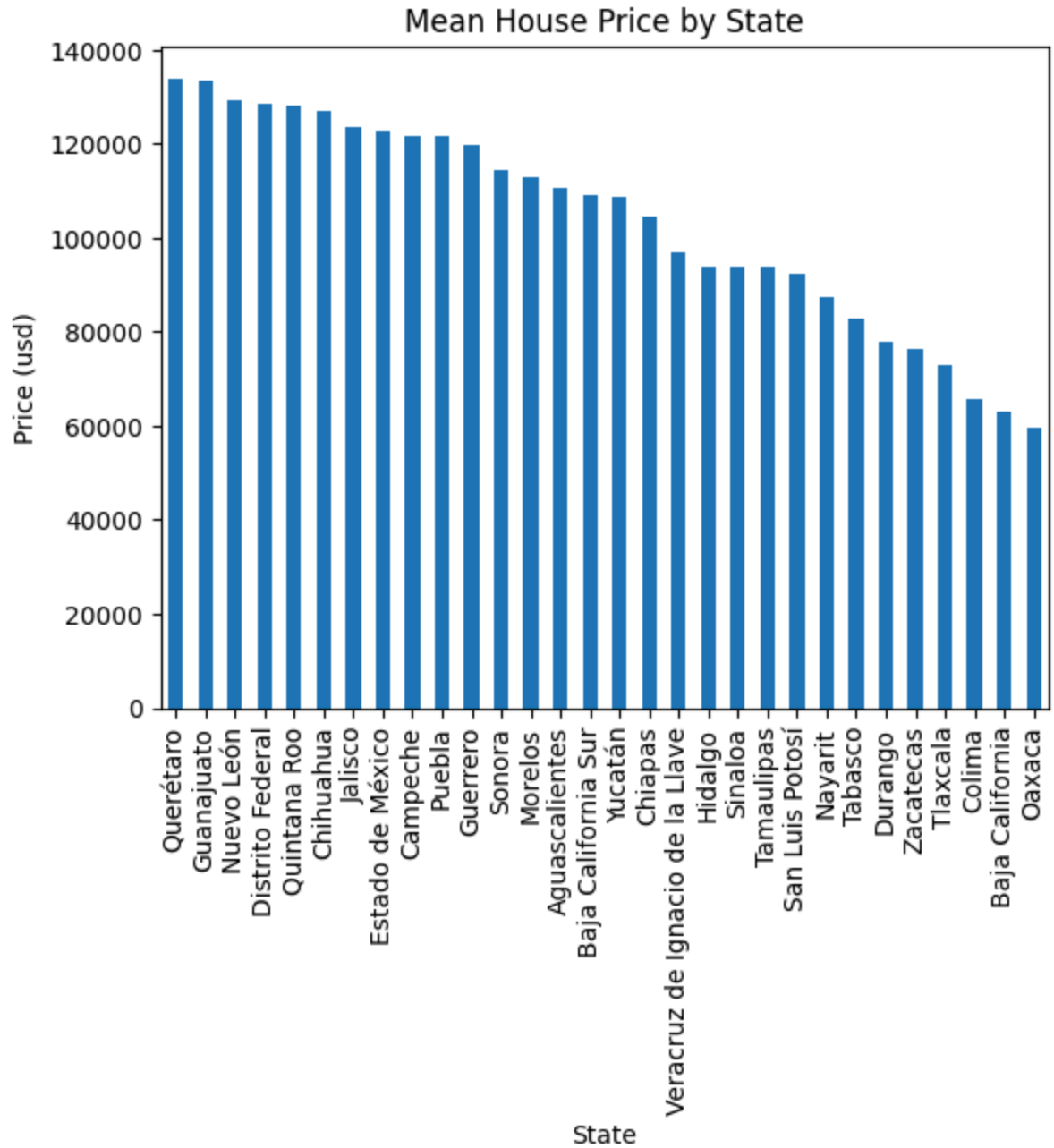
- Box plots were used to examine price outliers.

Distribution of Home Sizes

Area (sq meters)
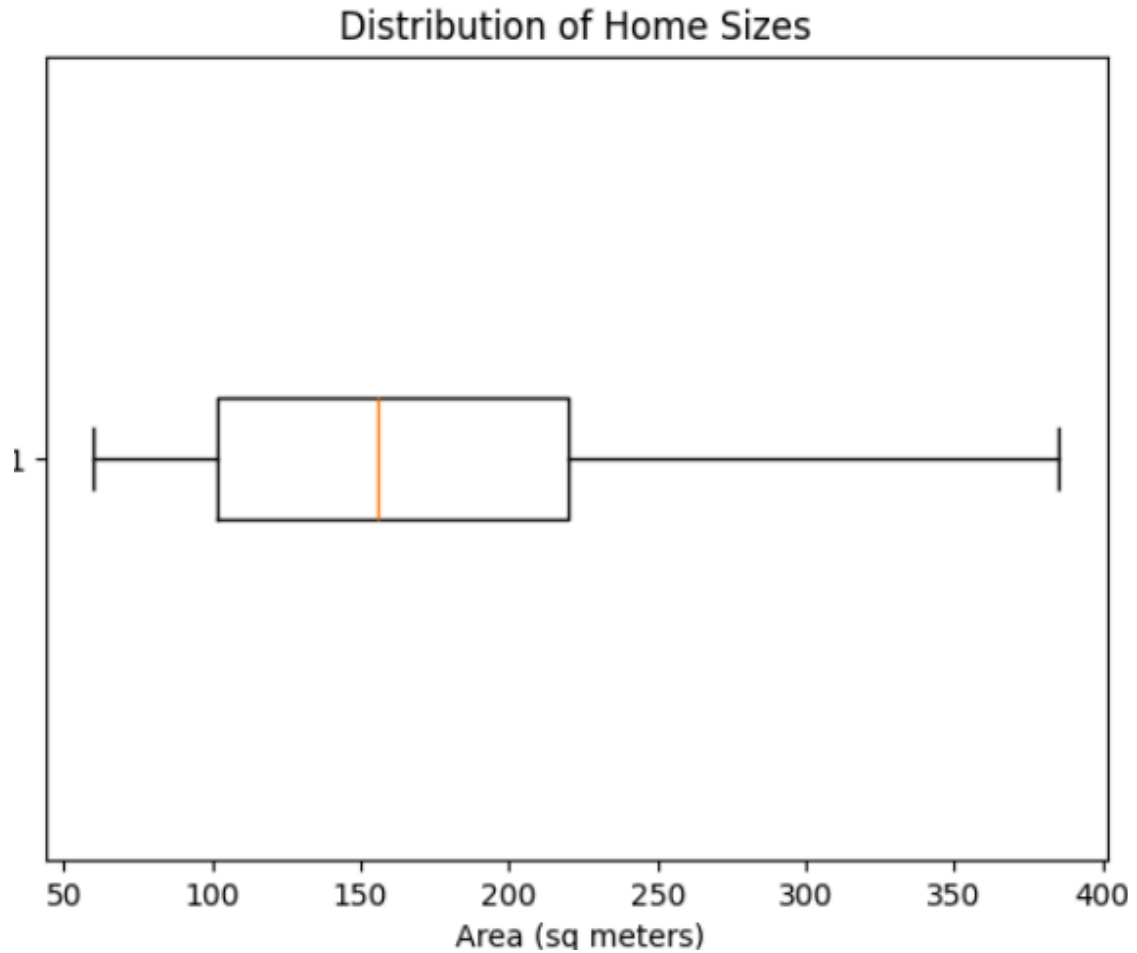
- Value counts and group-by aggregations were used to count the number of listings per state and region.

Correlation Analysis: The correlation between property size (`area_m2`) and price (`price_usd`) was calculated for different states.

For the **Southern states of Mexico**, the correlation coefficients were:

{'Paraná': 0.521, 'Santa Catarina': 0.548, 'Rio Grande do Sul': 0.497}

These values suggest a **moderate positive correlation** between property size and price in these states.

Interpretation of Results:

- Property size does have a positive correlation with price, meaning larger properties tend to cost more.
- However, the strength of this relationship varies by location.
- Location also has a significant impact on price, as seen in regional comparisons and boxplots.
- Some high-demand states show stronger correlation, suggesting both location and size play important roles.

Conclusion: Both **property size** and **location** significantly affect the price of real estate in Mexico. However, their influence varies by region. By quantifying these relationships, real estate professionals and buyers can make more informed decisions.

Tools Used:

- Python (Pandas, Matplotlib, Seaborn)
- Jupyter Notebook(Google Collab)

Check out project in [Github](#)

This project demonstrates a full data science pipeline from data cleaning to interpretation, using real-world data.