GUATEMALA REAL ESTATE
ING. DANIEL PEDRO PABLO ALVAREZ VIELMAN

REAL ESTATE PRICE ANALYSIS REPORT

GUATEMALA PROPERTY MARKET OVERVIEW 2024

Content

Introduction

The "Predicting House and Apartment Prices" project offers a detailed analysis on real estate prices within specific Guatemala's municipalities context. Drawing from data extracted from a real estate website (*www.encuentra24.com*), this study focuses on properties within the price range between Q400,000 to Q2,000,000, examining both houses and apartments. The project serves as a practical data analysis skills demonstration, aiming to equip potential buyers, investors, and industry professionals with valuable market insights and the ability to predict property prices.

Initially, the project involved meticulous data extraction and cleaning phases. The data was filtered and normalized to address variabilities in listings, ensuring only properties within the specified price range were included. Through comprehensive analysis, municipality and zone were accurately determined from listing titles.

In the exploratory phase, we delve into price distributions and feature correlations, utilizing statistical methods such as correlation matrices and ANOVA tests. These analyses highlight the most influential factors on property pricing, emphasizing significant geographical location role.

Guatemala's diverse municipal landscape informs this study, showcasing variations in real estate market dynamics across different areas. Ultimately, the project will culminate in a predictive model for property prices and visualizations on a dashboard.

Objectives

1. Data Extraction and Cleaning:

- Extract real estate data from an online platform, focusing on properties within the price range between Q400,000 to Q2,000,000.
- Clean and normalize the dataset, ensuring accuracy in representing property details.

2. Geographical Analysis:

- Determine municipality and zone for each property listing, leveraging title information and improving locality insights.

3. Descriptive Analysis:

- Explore price distributions and identify key features: square meters, rooms, parking, and baths number that influence property prices.

4. Feature Correlation and Statistical Testing:

- Perform correlation analysis and ANOVA tests to assess different municipalities and zones impact on property prices.

5. Predictive Modeling:

- Use different models to accurately predict property prices, using the insights gained from the exploratory analysis phase.

6. Dashboard Visualizations:

- Create a Dashboard on Power BI to provide an interactive real estate visualization analysis.

7. Market Insights:

- Provide valuable insights into the Guatemalan real estate market dynamics to guide buyers, investors, and industry professionals.

Context

This project focuses on the real estate market within Guatemala's municipalities. Using data sourced from a real estate website, we explore property listings filtered to reflect the price range between Q400,000 to Q2,000,000. The dataset includes both houses and apartments, offering a market dynamics snapshot across diverse regions.

Guatemala's municipalities exhibit varied characteristics, influencing property valuations significantly. By understanding these regional differences, we aim to shed light on the factors driving real estate pricing and equip stakeholders with actionable insights.

The analysis considers properties from key municipalities, including Guatemala City, Mixco, Villa Nueva, Santa Catarina Pinula, Fraijanes and others, providing an in-depth view how these areas compare in property affordability and demand terms. This contextual backdrop sets the stage for developing a predictive price model, aimed at enhancing strategic decision-making in the property market.

Methodology

This project employs a comprehensive methodology that integrates various tools and techniques across different analysis phases:

1. Web Scraping:

Data was extracted from a real estate website using **Selenium** for browser automation and **BeautifulSoup** for HTML parsing. Key libraries used include:

- **Selenium.webdriver**: For automating web browser interaction and managing Firefox drivers.
- **Webdriver_manager**: To streamline driver setup.
- **BeautifulSoup**: For parsing HTML content and extracting data.

2. Data Cleaning: The dataset was cleaned and prepared using **Pandas** for data manipulation, with regular expressions (**re**) and text normalization (**unidecode**) to ensure data consistency.

3. Geographical and Property Type Determination:

- **Fuzzywuzzy:** was utilized to accurately determine municipality and zone from property titles and to categorize property types as house or apartment.

4. Descriptive Analysis:

- Mathematical and statistical analyses were carried out using **Pandas**, **Numpy**, and **Scipy** for insights into numerical data.
- Data visualization was performed with **Matplotlib** and **Seaborn**, aiding in price distributions examination, correlation analysis, and regional price variations.

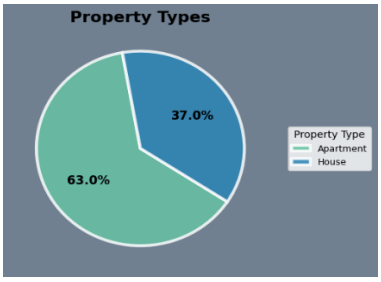
5. Predictive Modeling:

- Planned for future implementation, predictive modeling will involve utilizing regression techniques and other algorithms to forecast property prices based on identified key features.

These methodologies collectively facilitate a thorough real estate market exploration and analysis, supporting informed decision-making through robust data insights.

Results and Interpretations

1. **Property Type Distribution:** The dataset is composed of 37% houses and 63% apartments. This distribution suggests a larger availability or market demand for apartments within specified price range.



2. **Descriptive Statistics:** Houses tend to be larger with more rooms and parking spaces compared to apartments, and both property types exhibit similar price ranges. Houses average 183.03 m² and 3.30 rooms, while apartments average 84.34 m² and 2.10 rooms. These differences likely influence price predictions and suggest key factors to consider in future model improvements.

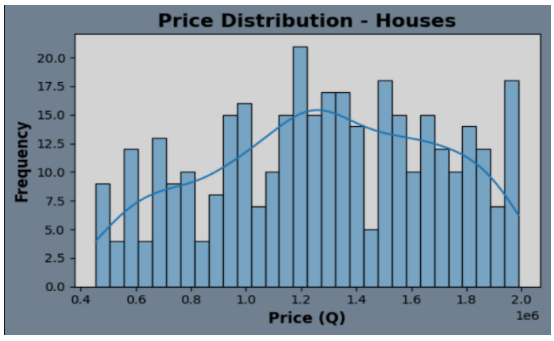
House

	mean	std	min	25%	50%	75%	max
Price (Q.)	1,282,560.38	411,895.80	455,000.00	975,000.00	1,300,000.00	1,618,700.00	1,989,000.00
Square meters	183.03	86.09	37	120	175.5	225.25	544
Rooms	3.3	1.3	1	3	3	3	13
Parking	2.23	1.01	0	2	2	2	6
Baths	2.3	0.95	1	2	2	3	7

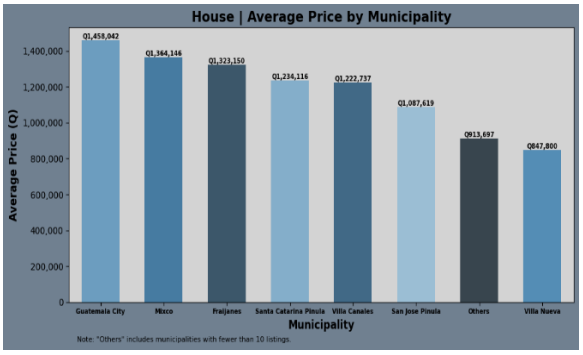
Apartment

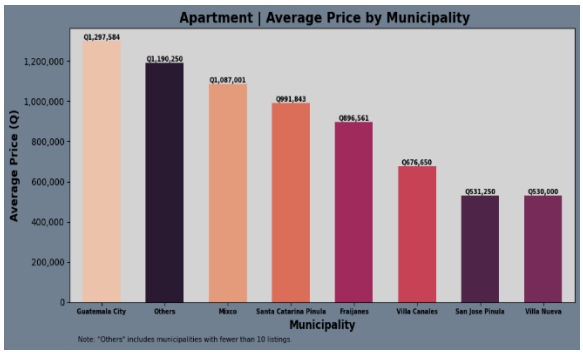
	mean	std	min	25%	50%	75%	max
Price (Q.)	1,259,094.10	373,304.17	415,000.00	975,000.00	1,250,000.00	1,550,250.00	2,012,400.00
Square meters	84.34	34.21	19	62.25	79	100	400
Rooms	2.1	0.8	1	2	2	3	10
Parking	1.61	0.58	0	1	2	2	3
Baths	1.6	0.6	0	1	2	2	6

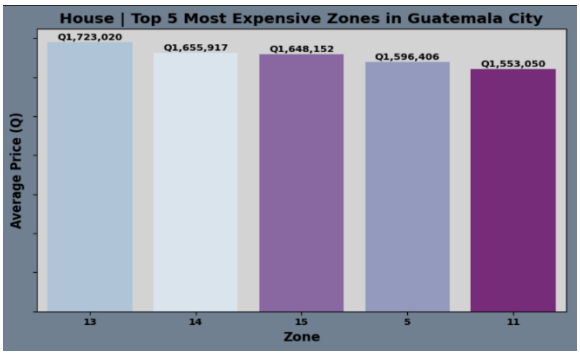
3. **Price Distribution:** Price distribution for houses shows slight bimodality, while apartments display a more symmetric distribution.

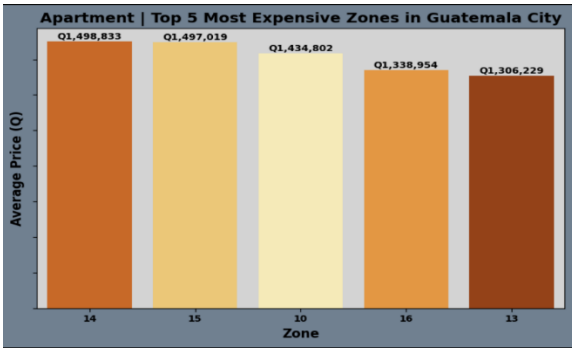


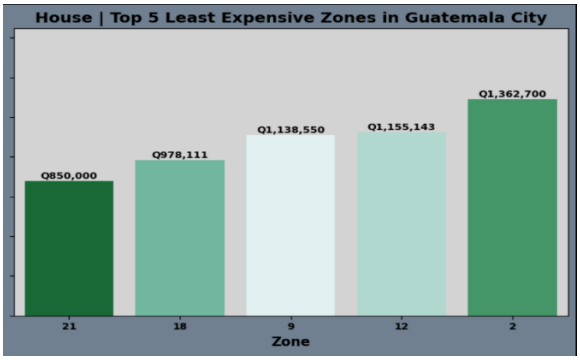
4. **Price Analysis by Municipality and Zone:** Guatemala City shows the highest average prices for both houses and apartments, highlighting significant location role in property valuation.

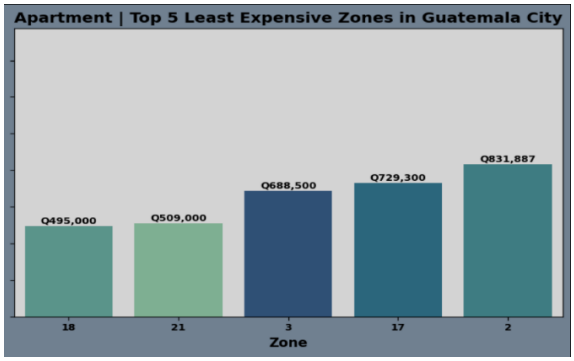




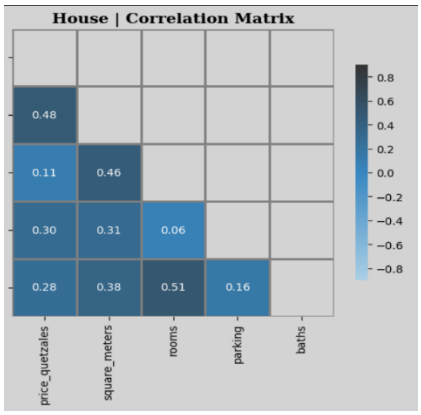


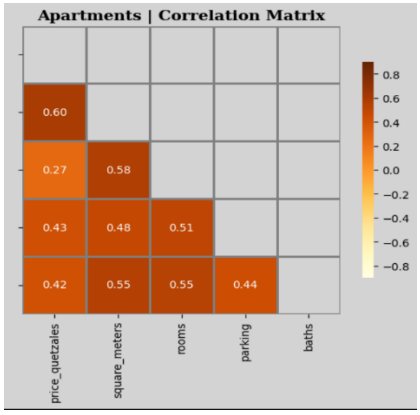






5. **Correlation Insights:** Strong correlations were observed between price and attributes like square meters and rooms number, indicating these factors are essential property prices determinants.

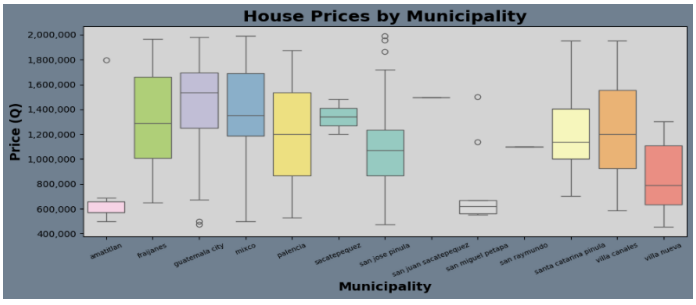


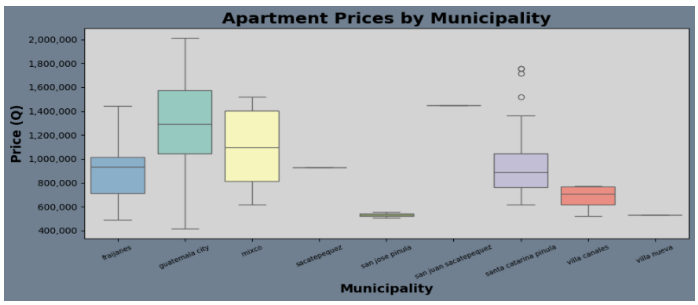


6. ANOVA Analysis: Both zones and municipalities significantly impact property prices, with zones showing a particularly strong effect on apartment valuations.

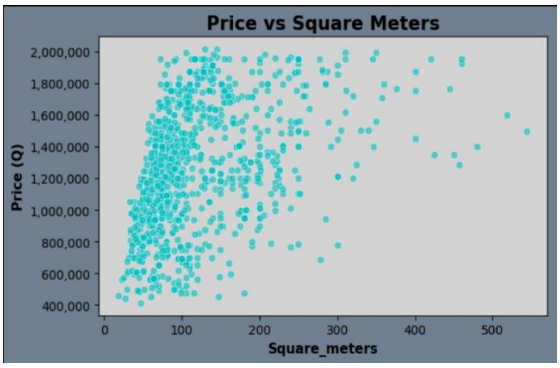
Type	Category	F-Statistic	p-Value
House	Zone	4.558079	2.81E-09
	Municipality	7.797261	7.36E-13
Apartment	Zone	18.826713	3.53E-45
	Municipality	8.409038	7.40E-11

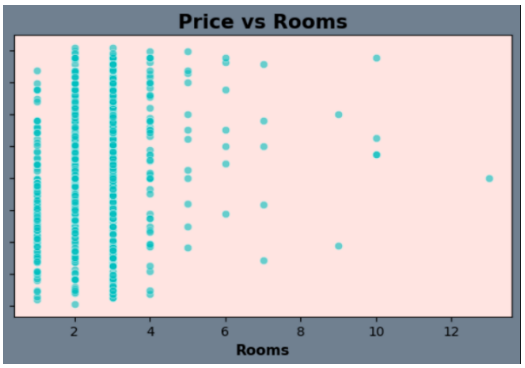
7. Boxplots by Municipality: Boxplots highlight price distributions across municipalities, revealing variability and outliers in pricing.

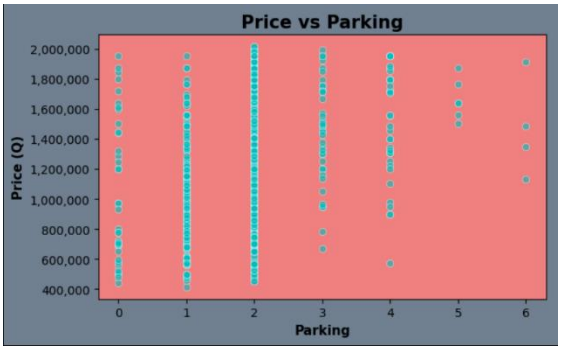


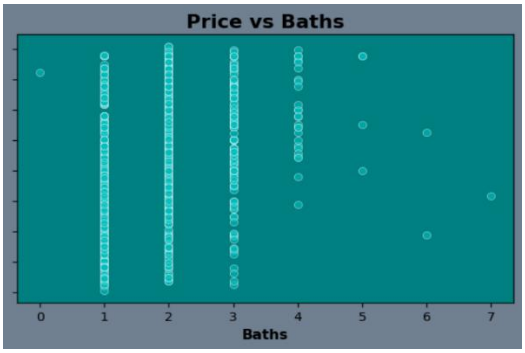


8. **Price vs Key Features:** Scatter plots indicate a positive correlation between price and features like square meters and rooms number. However, variability suggests additional influencing factors.









9. Predictive Modeling Results: We applied multiple regression models to predict property prices for both houses and apartments. The results are summarized as follows:

- **Houses:** Ensemble models like Random Forest and Gradient Boosting showed superior performance, capturing complex patterns better than basic linear models.
- **Apartments:** Linear models performed well, but ensemble methods offered slightly better accuracy, with Gradient Boosting leading.
- The smaller houses dataset may have constrained model performance, highlighting potential benefit of acquiring more data.
- Apartment prices are more predictable with the current features, indicating a more straightforward relationship between variables.

Model	Houses MSE	Houses R2	Apartments MSE	Apartments R2
Linear Regression	1.14E+11	**0.308819**	3.76E+10	**0.676422**
Lasso Regression	1.14E+11	**0.308812**	3.76E+10	**0.67642**
Decision Tree	1.14E+11	**0.312328**	5.47E+10	**0.52928**
Random Forest	9.05E+10	**0.453939**	3.82E+10	**0.671704**
Gradient Boosting	9.35E+10	**0.4358**	3.64E+10	**0.68664**

10. Power BI: The Power BI dashboard complements the analytical results by offering a visual interaction with the data. Key insights are represented through various visualizations:

1. **Property Type Distribution:** Donut chart illustrating houses and apartments proportion.

2. **Price Analysis by Type:** Stacked bar chart showing average price comparisons between houses and apartments.

3. **Geographical Price Trends:** Combined column and line chart revealing price differences and price per square meter across municipalities. Area chart showing average prices by zone, highlighting regional variations.

4. **Feature Overviews:** Gauges providing quick key property features snapshots such as square meters, rooms, parking, and baths.

These visual tools support findings from exploratory data analysis, reinforcing conclusions drawn from statistical methods used throughout the project.

Conclusions

- **Property Type Distribution:** The dataset is predominantly composed of apartments (63%), which may indicate higher availability or demand within analyzed price range.
- **Descriptive Statistics:** Houses are generally larger and more expensive than apartments. However, both property types exhibit a similar prices range, suggesting variability in other influencing factors.
- **Price Analysis by Municipality and Zone:** Guatemala City consistently showcases the highest average prices for both houses and apartments. Significant price disparities are observed across different zones and municipalities, emphasizing location role in property valuation.
- **Correlation Insights:** Strong correlations between certain features (e.g., price and square meters) suggest that physical attributes are crucial property prices determinants, particularly for apartments.
- **ANOVA Findings:** Both zones and municipalities influence property prices significantly, with a notably stronger effect observed in apartments. This underscores need to incorporate geographical factors in price prediction models.
- **Overall Market Dynamics:** The analysis demonstrates that while physical attributes and location heavily influence property prices, there is substantial variability, indicating potential role on other factors not captured within this dataset.
- **Data Insights:** The analysis emphasizes crucial location role and physical property attributes in determining real estate prices.
- **Model Efficacy:** Ensemble models provided better predictions for both houses and apartments, indicating their capability to capture complex relationships within the data.
- **Challenges and Limitations:** Limited houses data affected model accuracy. A larger dataset could improve findings.
- **Future Directions:** Enhanced feature engineering and additional data collection are recommended to further refine model predictions and gain deeper insights into the real estate market.