

# Real Estate Market Analysis: Mexico and Brazil

## 1. Project Overview

This project involved a comparative **exploratory data analysis (EDA)** of real estate markets in **Mexico and Brazil**. The main objective was to uncover pricing trends, property characteristics, and regional differences that could help **investors make more informed decisions**.

---

## 2. Data Overview

- **Source:** Publicly available datasets from **Kaggle**
  - **Attributes Included:**
    - **Numerical features:** price, surface area (area\_m2)
    - **Categorical features:** property type, operation type, region/state
    - **Location data:** latitude, longitude (for mapping)
- 

## 3. Project Workflow

### 3.1 Data Collection & Understanding

- Sourced real estate data for both countries.
- Data fields included price, property type, surface area, number of rooms, and location coordinates.

### 3.2 Data Cleaning & Preparation

- Handled missing values and filtered irrelevant records.
  - Standardized numerical formats (prices and area measurements).
  - Concatenated and merged cleaned datasets into a unified dataframe.
- 

## 4. Exploratory Data Analysis (EDA)

### 4.1 Geographic Distribution

#### LOCATION DATA(lat and lon)

- I analysed location data using "lat" and "lon" Columns to get a sense of where the properties are located

- Used latitude and longitude columns to visualize property locations with `scatter_mapbox` (via Plotly)
  - The houses in our dataset distributed evenly throughout the country with high property concentration around **Mexico City**, likely due to urban development and economic activity.
- 

- **4.2 Categorical Analysis: Top States by Listings**
  - Used `.value_counts()` on the state column to identify the **top 10 states** with the most listings:
    - **Top 3:** Distrito Federal (CDMX), Estado de México, Yucatán
  - **Distrito Federal was the most prevalent state with** the highest number of listings, indicating a highly active urban real estate market.
  - States like **Yucatán** and **Morelos**, which are not the largest economically, also showed strong activity, possibly due to lifestyle or tourism-related demand.
  - Noted high activity in non-economic hubs like **Yucatán** and **Morelos**, possibly due to lifestyle or tourism factors.
- 

#### 4.3 Numerical Analysis: Area and Price Distributions

- Analyzed `area_m2` and `price_usd` using descriptive statistics and histograms.
  - Found **right-skewed distributions** in both variables.
  - Price distribution was more skewed, with outliers pulling the mean significantly higher than the median.
- 

### 5. Research Questions & Key Findings

#### Research Question 1: Which State Has the Most Expensive Real Estate?

- Initially computed average price by state:
- **Querétaro** appeared most expensive by average price, but this conflicted with economic data.
- Introduced a better metric: **price per square meter**

**Finding:** Mexico City (Distrito Federal) emerged as the **most expensive market** based on price per m<sup>2</sup>, aligning with GDP rankings and economic influence.

---

## Research Question 2: Is There a Relationship Between Home Size and Price?

- Created scatter plots between area\_m2 and price\_usd.
- Found a **moderate positive correlation** nationally (~0.5), indicating that larger homes tend to cost more.

### State-Level Correlation Analysis

- **Morelos**: Strong correlation between home size and price.
  - **Mexico City**: Weak correlation, suggesting other factors (e.g., location, amenities, urban constraints) influence price more heavily in economic centers.
- 

## 6. Interpretation & Insights

- **Price and size are generally related**, but the strength of that relationship **varies by region**.
  - In urban centers like **Mexico City**, price is influenced more by **location and demand** than by size alone.
  - Using **price per m<sup>2</sup>** gives a more accurate measure of real estate value across regions.
- 

## 7. Conclusion

By the end of the analysis, I was able to deliver **actionable, data-driven insights**:

- Identified which regions are more **affordable** or **expensive**
- Clarified which **property types** are most common
- Showed how **price relates to home size and location**

These findings are valuable to **real estate investors, developers, and analysts** seeking to understand and make decisions in the **Latin American property market**.

---

## Key Insights Delivered

By the end of the analysis, I was able to deliver the following actionable, data-driven insights:

- **Most Expensive Regions:**  
Mexico City (Distrito Federal) was identified as the most expensive real estate market when using price per square meter as the metric.

- **Most Affordable Regions**  
States such as San Luis Potosí and Veracruz were among the most affordable based on average price per square meter.
  - **Common Property Types:**  
Apartments and houses were the most frequently listed property types across both countries, with variation by region.
  - **Size Price Relationship:**  
There is a moderate positive correlation between home size and price nationwide, however, this relationship weakens in major urban centers like Mexico City where location has a stronger influence
  - **Regional Distribution Patterns:**  
Property listings are heavily concentrated in urban centers such as Mexico City, while some lifestyle or tourism driven states like Yucatán and Morelos also showed unexpectedly high listing volumes
-