House Prices in Argentina

JonPaul Ferzacca, Joey Musholt

Project Description

In the Argentine real estate landscape, our dataset offers a robust .csv compilation of property listings including over 1 million data point. Through this data, we endeavor to discern key trends such as price prediction based on property attributes like bedroom count and size, the correlation between listing duration and price, and the interplay between price and location. This exploration can give insights for local governance and real estate investors, guiding data-informed housing decisions.

Prior Work

- Housing Price Predictor
  - Mapping of Data
  - Measuring Feature Importance
  - https://www.kaggle.com/code/msorondo/a-housing-price-predictor-for-buenos-aires-c

Data Source

- 2 million rows
- 25 potential features
- Accessed via Kaggle:

  https://www.kaggle.com/datasets/msorondo/argentina-venta-de-propiedades/data

- Originally sourced from Properati (property listing platform):

  https://www.properati.com.ar/

- Downloaded on both PCs

Proposed Work

Data Cleaning

- Cleaning of missing data point (Longitude and Latitude)
- Cleaning of Ad Type
- Potential drop of Uruguay
- Checking for outliers

Data Preprocessing

- Converting to uniform currency
- Feature Engineering
    - Date Extraction
    - Convert categorical variables

Data Integration

- Map Integration
- Currency conversion rates(Change with interest)
- With location data, distance to important landmarks or city centers.

Tools

- Python
- NumPy
- Pandas
- Sklearn
- MatPlotLib
- Seaborn
- Github

If needed:

- SQL Databases (like PostgreSQL, MySQL, MS SQL Server)
- Tableau (for visualizations)

Evaluation

- Comparing predicted price to the labelled price feature
- Overall model accuracy, precision, recall, F1-score
- Accuracy by subsets of data (ex: models have better fit by structure type, location)