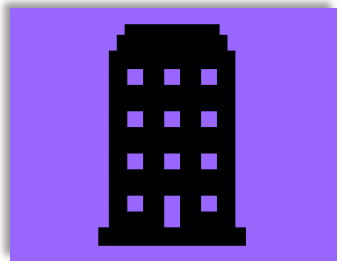


Exploring an international market to understand factors impacting price

Project Objectives

- Source, clean, and analyze data from the web
- Use statistical analyses, including linear regression and cluster analysis, to explore relationships within the data
- Further explore the data through geospatial and time series analysis
- Create a Tableau storyboard to present relevant findings in a digestible format

Data Overview

- Portuguese real estate data sourced from Kaggle
- Historical residential property prices source from the Federal Bank of St. Louis
- Geographic data on Portugals region's sourced from simplemaps

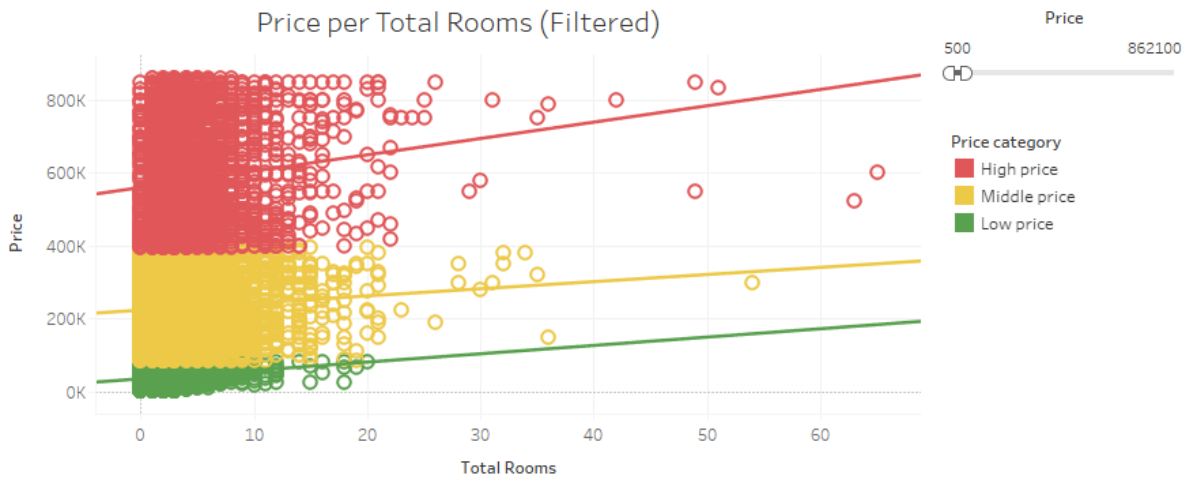
Techniques Applied

- Sourcing data suitable for analysis
- Cleaning data, including the use of machine learning to impute missing values
- Analyzing the relationships between variables within the data to identify the impact variables upon one another
- Forming and answering research questions related to the data
- Presenting relevant findings, including limitations and future possibilities

Tools Used

- Kaggle
  - Data Sourcing
- Anaconda's Jupyter Notebook
  - Data analysis using Python
  - Python libraries including NumPy, Pandas, Seaborn, Matplotlib, and scikit-learn
- Tableau
  - Visualization of analyses
  - Presentation of findings

Visual Highlights



The properties with the most total rooms tend to be in the high price range. However, there is no statistically relevant strength to the relationship between price and total rooms.

Lisboa, the coastal district containing Portugal's capitol, contains more available properties than any other district

Data Limitations and Future Analysis

- Limitations included the presence of null and irregular values in the original data as well as the massive variety of property types and range of variables present in Portuguese real estate.
- Exploring real estate comes with the general limitation that the market is constantly changing and is affected by a near endless number of factors
- Future analyses could dive deeper into the relationship between price and location or price and current economic conditions. Additionally, future analyses could explore the differences between asking prices and selling prices in the market.

Actionable Insights

- Data may not contain obvious relationships between variables. The lack of a strong relationship can be just as meaningful as a relationship with strength.
- The data cleaning process, such as removing outliers or imputing values, is a sensitive process. Comparing an analysis with and without the presence of outliers can be useful.
- Sourcing useable, relevant, available data can be a time-consuming process. The value of data can’t be understated.

Link to GitHub Repository

Link to Tableau Storyboard