

Midterm Project

Team Name - **team-deep-thinkers**

Topic

Real Estate Price Analysis

Team Members:

- Miyar Karthik Kamath
- Mohith Girigowdara Girish
- Ahmed Zaytoun

Objective

The availability of land suitable for residential construction and variations in the affordability index of the population around the country, especially in states like Hawaii, New York, and California, have led to the housing market's enormous growth over the years. The goal here is to look at aspects like profitability and competitiveness to analyze a particular investment and find the several elements that might affect housing affordability. Among the main techniques that will be used to assess the correctness of the model are regression using a Muller loop, classification, and fractal clustering to find golden clusters. It is possible to identify latent variables that could improve a location outcome.

Business Purpose:

Goal:

The goal is to use machine learning to analyze a particular investment and help investors make data-driven decisions when purchasing investment properties by identifying possible factors impacting the housing deficit and analyzing real estate prices by looking at aspects like profitability and competitiveness to maximize profitability and minimize risk.

Metrics:

Metrics are measured by:

- **Profitability Analysis:** This involves analyzing factors like rent income versus expenses such as mortgage, insurance, HOA fees, property taxes, and utility fees.
- **Affordability Index** such as house price to income ratio.
- **Model Accuracy** is based on Gini score, precision, recall, and accuracy.

Service:

- **Refinements:** Clean data by removing unwanted symbols such as \$ or sqft, assuming the missing values by taking the mean value, and dropping unnecessary columns.
- **Amalgamation:** Data from three datasets will be amalgamated based on median income, location, age, and housing value.
- **Enrichment:** Adding latent variables such as flood or wildfire zones to enhance its predictive power.
- **Application/Microservices:** Develop a monolith-based architecture for preprocessing data with microservices for other functionalities like model training.
- **User Experience (U/X):** Design a user-friendly dashboard that allows users to visualize results.
- **Business Logic:** Identifying ideal properties based on rent income and purchase price.

Data Narrative:

1. What is the data telling me?
The data shows trends and patterns revealing real estate prices from different regions and periods.
2. What can the data tell me?
The data is capable of predicting future real estate prices and identifying various factors that influence price fluctuations.
3. Patterns you see in the data
We can observe a pattern in which prices in urban areas increase compared to rural regions, and seasonal variations impact overall trends.
4. What are you looking for in the data?

While examining the datasets, we are focusing on key factors that play important roles in property prices, such as location, property size, and market conditions.

5. Hypothesis:

Hypothesis 1:

Location-specific factors such as proximity to urban centers, crime rates, and quality of schools play a significant role in real estate prices.

Hypothesis 2:

Interest rates and unemployment rates contribute heavily to real estate price fluctuations over time.

Articles:

- K-Means Clustering on Neighborhoods and Real estates:

<https://medium.com/@tenzkunsang648/k-means-clustering-of-homes-and-neighborhoods-in-minneapolis-6b4fb02bc308>

- ML-driven house price prediction:

<https://www.akkio.com/post/house-price-prediction-using-machine-learning>

- Enhancing Your Sharpe Ratio: A Quantitative Approach:

<https://medium.com/@serdarilarlan/enhancing-your-sharpe-ratio-a-quantitative-approach-7f633407dc5a>

Experiments:

Predict what properties and where I can buy.

We are using ML to match properties in certain areas to budget constraints.

Predict how we minimize the purchase cost.

We use clustering and regression to identify properties with the lowest cost for the best property value.

Predict which properties will have the highest future value.

Using regression, we can establish relationships between property features and future prices.

Key Performance Indicators (KPIs):

1. Accuracy:

Definition: Measures the percentage of correctly classified properties out of the total number of properties.

Example: The accuracy of classifying properties into categories like 'High-Value,' 'Medium-Value,' and 'Low-Value.'

Associated Method: Classification

2. Precision:

Definition: Measures the proportion of valid positive classifications among all positive classifications.

Example: The precision of identifying 'High-Value' properties in the market from the predictions.

Associated Method: Classification

3. Mean Absolute Error (MAE):

Definition: Measures the average magnitude of errors in predicting property prices without considering the direction.

Example: Predicting property prices based on historical data and market trends.

Associated Method: Regression

4. R-Squared (Coefficient of Determination):

Definition: Indicates the proportion of the variance in property prices that can be predicted from independent variables such as location, size, and amenities.

Example: The R-squared value of a regression model predicting future property prices based on features like proximity to amenities.

Associated Method: Regression

5. Silhouette Score:

Definition: Measures how similar a property is to its cluster compared to other clusters, indicating the quality of the clustering.

Example: The Silhouette Score for clustering properties based on features like price, location, and size, ensuring distinct groups are well-separated.

Associated Method: Clustering

Datasets:

- California Housing Prices:

<https://www.kaggle.com/datasets/camnugent/california-housing-prices>

- Housing Prices SoCal:

<https://kaggle.com/datasets/ted8080/house-prices-and-images-socal>

- USA Real Estate dataset:

<https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset>

Machine Learning Experiments:

- **Clustering:** Using the K-means clustering algorithm to identify properties with positive rent income.
- **Regression:** Using regression models to predict future property prices or rent values.
- **Classification:** Classify properties based on investment risk, profitability, or other significant factors.