**Housing Price Prediction**

**Team: 16**

Yash Zaveri & Zeel Patel

**INTRODUCTION:**

**Background:**

The housing price prediction project is a cutting-edge initiative aimed at leveraging advanced data analysis and machine learning techniques to forecast real estate values based on a comprehensive set of factors. By incorporating features such as the property's location, historical housing prices in the area, the number of bedrooms and bathrooms, and the overall area of the house, the project seeks to create a predictive model that can offer valuable insights into future property values. This approach represents a paradigm shift in real estate decision-making, moving beyond traditional methods by harnessing the power of extensive datasets to identify intricate patterns and correlations. The project not only addresses the inherent challenges of market volatility and regulatory changes but also provides a dynamic, adaptable model that can adjust to evolving market conditions.

**Motivation:**

Our motivation for this housing price prediction project is grounded in the need to provide valuable insights for participants in the real estate market. Similar to the cryptocurrency market, where price fluctuations and market sentiment are crucial, the housing market also faces challenges influenced by factors like location, historical pricing, and property attributes. By harnessing advanced data analysis and machine learning techniques, our project aims to develop a predictive model tailored to assist individuals in understanding and forecasting housing price dynamics.

**Goal:**

Our project objectives extend to analysing historical data in the real estate domain, aiming to identify patterns, trends, and influential factors for predictive modelling. Recognizing that some houses may lack individual past data, we will employ surrounding housing prices and neighbourhood trends as crucial indicators. By scrutinizing neighbouring property values and broader housing market dynamics, our goal is to develop a robust predictive model for housing prices, enabling informed decision-making even in cases where specific past data is unavailable. This approach seeks to enhance the predictability of property values by considering the broader context of the surrounding housing landscape and neighbourhoods trends.

**METHODOLOGY:**

**Data Pre-processing:**

This involves handling missing values, addressing outliers, standardizing and normalizing numerical features, and encoding categorical variables. Additionally, feature engineering, dealing with duplicates, and splitting the dataset into training and testing sets are key steps. Th e process ensures a clean and standardized dataset, laying the groundwork for effective machine learning model development.

**Data Analysis:**

This involves analysing summary statistics like mean, median, and standard deviation for each property. In addition to numerical and categorical exploration of the data, one can conduct correlation analyses between variables, investigate sentiment scores derived from property descriptions, and visualize property prices. Geospatial insights can be gained if latitude and longitude information is available. Overall, a comprehensive data analysis can unveil trends, patterns, and relationships within the property dataset, facilitating informed decision-making for various stakeholders

**Feature Engineering:**

In housing price prediction, feature engineering is crucial for transforming raw data into informative model input. Key features include historical property values, square footage trends, and neighbourhoods characteristics. Lag features, moving averages, and sentiment indicators capture historical dynamics, while normalization ensures consistency. Custom features may reflect specific events or regulatory changes in the real estate market. Dummy variables, representing categorical variables like neighbourhood or property type, play a crucial role in encoding qualitative information for machine learning models.

**Model Selection:**

In our approach to housing price prediction, we adopt a diverse set of machine learning models, including Linear Regression, Decision Trees and Support Vector Machines. These models, trained on a comprehensive dataset comprising historical housing prices and property attributes, collectively aim to forecast property values and discern trends in the dynamic real estate market.

**DATASET:**

**Description:**

This dataset provides details about various real estate listings, encompassing information such as the region, price, property type (e.g., "apartment" or "condo"), square footage, number of bedrooms and bathrooms, as well as amenities such as whether cats or dogs are allowed, smoking policies, wheelchair access, and availability of electric vehicle charging. Additionally, the dataset includes a sentiment score derived from the textual descriptions of each property. The descriptions capture key features and characteristics of the listings, offering a comprehensive view for potential renters or buyers. This dataset is valuable for tasks such as price prediction, trend analysis, and understanding the impact of property descriptions on sentiment.

**SHAPE AND SIZE:**

(384978, 22) & 110.8 MB

<https://www.kaggle.com/datasets/austinreese/usa-housing-listings>