MACHINE LEARNING PROJECT

Housing Market Value Estimation

PROJECT INSTRUCTOR : Dr. U. Surya Kameswari

SPEAKER'S NAME :

Ponnada Kiran   (Y22MC20049)

INTRODUCTION

- This project aims at getting a good insight in the data for the House Price Prediction.
- It focuses mainly on preparing the data for modeling.
- The project can be classified into these following main sections:
  - Visual inspection of data
  - Defining the metadata
  - Descriptive statistics
  - Handling imbalanced classes
  - Checking Data quality
  - Exploratory data visualization
  - Feature engineering
  - Feature selection
  - Feature scaling

DATA Description

- Description of the housing dataset used
- Overview of key features and target variable (market value)
- Data preprocessingsteps (e.g., handling missing vaalues, data cleaning)

DATA Management

For better data management, the features are categorized on multiple grounds. This helps to select specific variables for tasks like analysis, visualization, modeling etc. Following are the grounds and categories used :

- role : input, ID, target
- level : nominal, interval, ordinal, binary
- keep : True, False
- data_type : integer, float

Handling Imbalanced Classes

- Explanation of imbalanced classes in market value estimation
- Techniques to address class imbalance (e.g., oversampling, undersampling, SMOTE)

Handling Data Quality

Following are the measures taken to improve data quality and handle missing values :

- Addressing outliers and noisy data
- Dealing with missing values (imputation strategies)
- Quality assessment and data validation techniques.

# Exploratory Data Visualization

Heatmap of interval features

For checking the correlations between interval features, heatmap is a good way to visualize the correlation between features.

- Visualization techniques for understanding feature distributions and relationships
- Scatter plots, bar charts, heatmaps, and other relevant visualizations
- Uncovering patterns and trends in the data

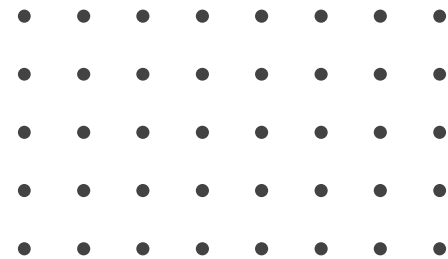# Using pair plots to visualize relationship between features
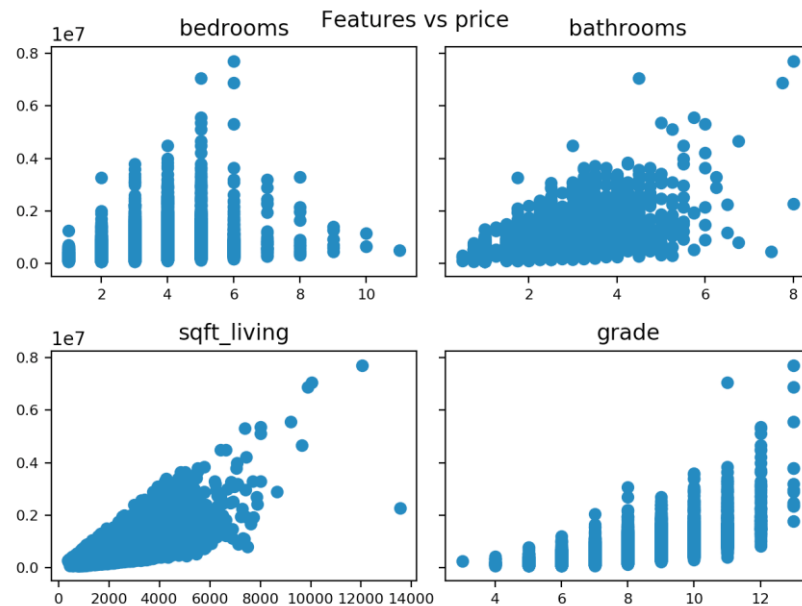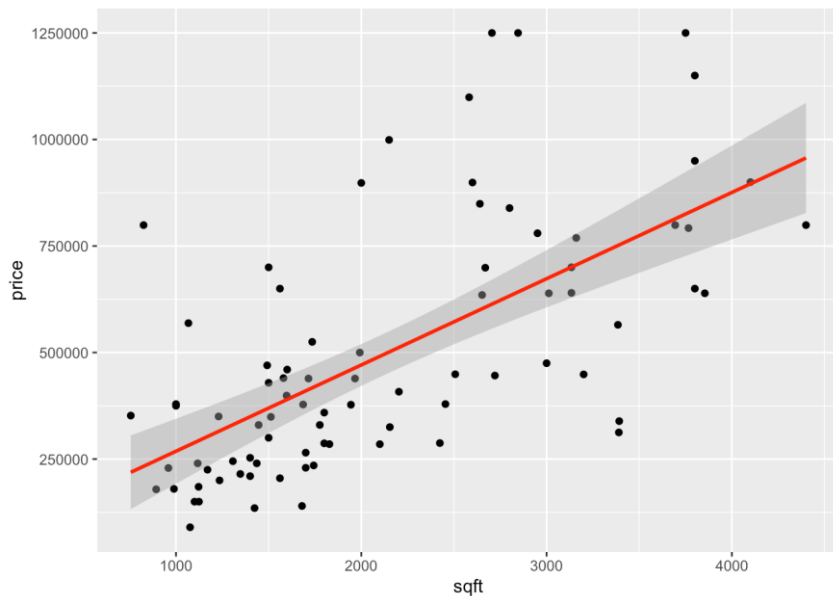
Feature Selection

- Methods for selecting relevant features for market value estimation
- Statistical techniques (e.g., correlation analysis, feature importance)
- Domain knowledge and expert input for feature selection

Feature Scaling

- Importance of feature scaling in machine learning models
- Techniques such as standardization and normalization
- Ensuring consistent scales for accurate model training and prediction

THANKS!