

Assignment1 : Exploratory Data Analysis and Feature Engineering with the Melbourne Housing Dataset

Objective:

The objective of this assignment is to conduct an exploratory data analysis (EDA) and apply feature engineering techniques to enhance the predictive modelling capabilities using the Melbourne Housing dataset.

Dataset:

You can obtain the Melbourne Housing dataset from Kaggle

(<https://www.kaggle.com/datasets/ronikmalhotra/melbourne-housing-dataset/data>).

Assignment Steps:

Data Loading and Initial Exploration (20 points):

- Provide information on the dataset, including the number of rows and columns.
- Briefly describe the target variable (e.g., 'Price') and its distribution.
- Display summary statistics and data types of the features.
- Identify any missing values and outline a plan to handle them.

Exploratory Data Analysis (EDA) (30 points):

- Visualize the distribution of numeric variables using histograms and box plots.
- Explore relationships between features and the target variable using scatter plots and correlation matrices.
- Examine categorical variables with bar plots and frequency tables.
- Identify potential outliers and discuss their impact on the dataset.

Feature Engineering (40 points):

- Apply at least five feature engineering techniques to improve the dataset for modeling purposes. Some ideas include:
- Handling missing data (e.g., imputation methods).
- Encoding categorical variables (e.g., one-hot encoding or label encoding).
- Creating interaction features or polynomial features.
- Scaling or normalizing numeric features.
- Handling outliers (e.g., winsorizing or removing extreme values).
- Provide clear explanations and justifications for each feature engineering step.

Conclusion and Recommendations (10 points):

- Summarize the key findings from the EDA and feature engineering processes.

Submission Guidelines: You should submit a Jupyter Notebook or a report documenting their **analysis, code, visualizations, and explanations**. **Do not** just submit a Notebook with blocks of codes. The most important thing I want to see in your assignment is your thinking process and your rationale for each movement, so detailed Markdown is required.