1. Objective

The goal of this project is to build a robust and explainable machine learning pipeline to give insights on private housing in Singapore. This involves end-to-end implementation, from data preprocessing and feature engineering to model building, explainability, and insight extraction.

2. Data Overview

The dataset used for this project contains detailed transaction records of private residential property sales in Singapore from a specific time period.

Field	Description
Project Name	Name of the property development
Transaction Price ($SGD)	Final sale price of the unit in Singapore dollars
Area (SQFT)	Floor area of the unit in square feet
Sale Date	Date the transaction was completed
Street Name	Location of the property
Type of Sale	Indicates if the sale was a new sale or resale
Type of Area	Specifies whether the floor area is strata or land area
Area (SQM)	Floor area in square meters
Property Type	Type of housing
Number of Units	Number of units transacted in the deal
Tenure	Lease type of the property
Postal District	Numerical code representing the geographical zone
Market Segment	Broad area classification
Floor Level	General indication of which level the unit is on

3. Task

The primary task of this project is to build a robust and explainable machine learning pipeline to uncover insights from private housing transaction data in Singapore. While prediction of transaction prices is part of

the pipeline, the main focus is on using the model outputs and explainability techniques to answer the following questions:

Required Analysis Questions:

1. **Top 3 Key Factors**: State the top 3 key factors that influence private property prices
2. **Patterns & Trends**: Identify patterns and trends across locations, property types, and time
3. **Buyer Insights**: Support property buyers in making informed decisions with 3 most important findings

---

4. Deliverables

1. Jupyter Notebook (.ipynb)

- Well-documented code implementing the complete machine learning pipeline
- Clear explanations for each step: data preprocessing, feature engineering, modelling, and insight generation
- Must include clear sections addressing the analysis questions

2. HTML Export of Notebook (.html)

- Export of the complete Jupyter notebook in HTML format

3. Prediction Results (predictions.csv)

- Model predictions from test dataset in CSV format

4. Written Responses Document

Answer the following questions with **less than 100 words each** (can reference specific sections in Jupyter notebook):

Question 1: Top 3 Key Factors

- Provide the top 3 factors influencing private property prices
- Include brief explanation of each factor's impact
- Reference: Point to relevant section in Jupyter notebook

Question 2: Patterns & Trends

- Describe key patterns across locations, property types, and time
- Highlight most significant trends discovered
- Reference: Point to relevant visualizations/analysis in notebook

Question 3: Buyer Recommendations

- Present 3 most important findings for property buyers
- Provide actionable insights for decision-making
- Reference: Point to supporting evidence in notebook

---

5. Grading Criteria (Total: 100%)

Component	Weight	Description
Model Performance	25%	Mean Squared Error (MSE) score on test dataset

Notebook Quality	50%	Clear, concise, and well-structured Jupyter notebook that comprehensively answers the analysis questions
Written Responses	25%	Quality and insight of answers in the written responses document

Evaluation Details:

- **MSE Score (25%)**: Lower MSE indicates better model performance
- **Notebook Structure (50%)**:
  - Code clarity and documentation
  - Logical flow of analysis
  - Visualization quality
  - Comprehensive coverage of required questions
- **Written Answers (25%)**:
  - Clarity and conciseness (within 100-word limit)
  - Insight quality and practical relevance
  - Proper references to notebook sections

6. Submission Instructions

1. **File Naming**:

   - Zip all files and name as: `Assignment_1_A1234567B_JohnDoe.zip`
   - Where A1234567B is your student number and JohnDoe is your name

2. **Upload Location**:

   - Submit to "Assignment 1" in Canvas → Assignments
   - Ensure submission is completed before the assignment due date

3. **Required Files in Zip**:

   - Jupyter notebook (.ipynb)
   - HTML export (.html)
   - Predictions CSV file
   - Written responses document (PDF or Word)