**PROJECTS: SELECT ONE (1) PROJECT OUT OF THE THREE PROVIDED.**

**Timeline:**

* Week 3-5: Data Exploration and Preprocessing
* Week 6-7: Model Building and Evaluation
* Week 8: Hyperparameter Tuning and Model Interpretation
* Week 9: Reporting and Presentation Preparation
* Week 10: Submission and Presentation

**Evaluation Criteria:**

* Understanding of data and preprocessing techniques
* Correct implementation of machine learning models
* Proper evaluation and comparison of models
* Effective hyperparameter tuning
* Clarity and thoroughness of the report
* Quality of the presentation

**Tools and Libraries:**

* Python
* Pandas, NumPy for data manipulation
* Matplotlib, Seaborn for visualization
* Scikit-learn for machine learning models
* Jupyter Notebook for code development

**Additional Resources:**

* Kaggle Tutorials
* Scikit-learn Documentation
* Pandas Documentation

**Project Assignment 3: Predictive Modelling for House Prices**

**Objective:**

The goal of this project is to build a predictive model that accurately estimates the prices of houses based on various features such as location, size, number of rooms, and other relevant factors. This project will help you apply various machine learning techniques and evaluate their performance.

**Dataset:**

Use the **Ames Housing Dataset**, which can be downloaded from Kaggle. This dataset contains information on 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa.

**Project Requirements:**

1. **Data Exploration and Preprocessing:**
   1. Load the dataset and understand its structure.
   2. Perform exploratory data analysis (EDA) to identify trends and patterns.
   3. Handle missing values appropriately.
   4. Encode categorical variables.
   5. Normalize/standardize the data if necessary.
   6. Feature selection: Identify and select features that are most predictive of house prices.
2. **Model Building:**
   1. Split the data into training and test sets (e.g., 80% training, 20% test).
   2. Implement the following regression models:
      * Linear Regression
      * Decision Tree Regression
      * Random Forest Regression
      * Gradient Boosting Regression
      * Any other model of your choice (e.g., XGBoost, Lasso Regression)
3. **Model Evaluation:**
   1. Evaluate the performance of each model using appropriate metrics (e.g., RMSE, MAE, R^2 score).
   2. Perform cross-validation to ensure the robustness of your model.
   3. Compare the results of different models and discuss which model performed the best and why.
4. **Hyperparameter Tuning:**
   1. Use techniques such as Grid Search or Random Search to find the best hyperparameters for your models.
5. **Model Interpretation:**
   1. Analyze and interpret the results of your best-performing model.
   2. Identify the most important features influencing house prices.
   3. Provide visualizations to support your findings.
6. **Reporting:**
   1. Write a comprehensive report summarizing your findings and methodologies. Your report should include:
      * Introduction
      * Data Exploration and Preprocessing
      * Model Building and Evaluation
      * Hyperparameter Tuning
      * Model Interpretation
      * Conclusion
7. **Presentation:**
   1. Prepare a presentation to showcase your work. Highlight key findings, methodologies, and the performance of your models.
   2. Be ready to answer questions regarding your approach and the results obtained.

**Deliverables:**

1. **Code:** Submit a well-documented Jupyter notebook or Python script containing all your code.
2. **Report:** A detailed report in PDF format.
3. **Presentation:** Slides in PowerPoint or PDF format.