PHASE I  
**House Price Prediction Project Plan**

**SREE VAASINI S**

**2021503562**

**ARTIFICIAL INTELLIGENCE**

**NAAN MUDHALAN**

**Problem Definition**

The problem at hand is to predict house prices using machine learning techniques. The objective is to develop a model that accurately predicts the prices of houses based on a set of features such as location, square footage, number of bedrooms and bathrooms, and other relevant factors. The project involves several key steps including data preprocessing, feature engineering, model selection, training, and evaluation.

**Design Thinking**

1. **Data Source**

For this project, we will utilize the dataset "USA Housing" available on Kaggle. The dataset contains information about houses, including features like location, square footage, bedrooms, bathrooms, and price. The dataset will serve as the foundation for our predictive modeling.

Dataset Link: [USA Housing Dataset](https://www.kaggle.com/datasets/vedavyasv/usa-housing)

1. **Data Preprocessing**

In this step, we will clean and preprocess the data to ensure it is suitable for training machine learning models. This involves handling missing values, converting categorical features into numerical representations, and addressing any inconsistencies in the dataset.

1. **Feature Selection**

We will select the most relevant features from the dataset that will aid in predicting house prices accurately. Factors such as location, square footage, number of bedrooms and bathrooms, and other relevant attributes will be considered during this process.

1. **Model Selection**

For the regression task of predicting house prices, we will choose a suitable regression algorithm. Potential options include Linear Regression, Random Forest Regressor, and others. The choice of model will be based on its performance and suitability for the given task.

1. **Model Training**

Once the model is selected, we will proceed to train it using the preprocessed data. This involves feeding the selected features and target variable (house prices) into the chosen regression algorithm and allowing it to learn the underlying patterns.

1. **Evaluation**

To evaluate the performance of the trained model, we will use appropriate evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared. These metrics will provide insights into how well the model predicts house prices and help us fine-tune the model if needed.

**Conclusion:**

Summarize the results, discuss the model's performance, and suggest potential future improvements or extensions to the project.