**Phase-1**

1. Define Problem statement ?

You are given a dataset containing various attributes of houses in Boston, Massachusetts, and their corresponding median values (prices). Your task is to build a predictive model that can estimate the median value of owner-occupied homes (in thousands of dollars) based on these attributes. In other words, you need to create a regression model to predict house prices based on the available features.

2. Create a project plan and product backlog ?

**Project Plan:**

**Project Initiation**

* + Define project objectives and scope.
  + Identify stakeholders and their expectations.

**Data Collection and Preprocessing**

* + Obtain the Boston Housing dataset.
  + Explore and clean the data.
  + Handle missing values.
  + Normalize or scale features if needed.

**Data Analysis**

* + Perform exploratory data analysis.
  + Visualize data distribution and correlations.
  + Identify potential feature selection or engineering.

**Model Development**

* + Split the data into training and testing sets.
  + Implement and train various regression models (e.g., Linear Regression, Decision Trees, Random Forest).
  + Evaluate model performance using appropriate metrics.
  + Fine-tune hyperparameters.

**Model Validation**

* + Validate the model's performance on unseen data.
  + Cross-validation to assess model generalization.
  + Address any overfitting or underfitting issues.

**Documentation**

* + Create documentation for the project, including the dataset, data preprocessing steps, and model details.
  + Prepare a report or presentation summarizing the findings and the model's performance.

**Deployment and Integration**

* + Deploy the model to a relevant platform or system.
  + Integrate the model with any necessary applications or interfaces.

**Testing and Quality Assurance**

* + Thoroughly test the deployed model to ensure it functions as expected.
  + Address any issues or bugs.

**Maintenance and Monitoring (Ongoing)**

* + Establish a system for monitoring model performance in production.
  + Plan for model updates and maintenance

**Product backlog:**

**Task: Data Collection**

* + Subtask: Obtain the Boston Housing dataset.
  + Subtask: Verify data integrity and sources.

**Task: Data Preprocessing**

* + Subtask: Data cleaning and handling missing values.
  + Subtask: Normalize or scale features.

**Task: Data Analysis**

* + Subtask: EDA - Visualize data distribution.
  + Subtask: EDA - Identify correlations between features.
  + Subtask: Feature selection and engineering.

**Task: Model Development**

* + Subtask: Implement Linear Regression model.
  + Subtask: Implement Decision Trees and Random Forest models.
  + Subtask: Train and evaluate models.

**Task: Model Validation**

* + Subtask: Cross-validation for model generalization.
  + Subtask: Address model overfitting/underfitting if required.

**Task: Documentation**

* + Subtask: Create project documentation.
  + Subtask: Prepare a report or presentation.

**Task: Deployment and Integration**

* + Subtask: Deploy the model to a platform.
  + Subtask: Integration with an application or system.

**Task: Testing and Quality Assurance**

* + Subtask: Thoroughly test the deployed model.
  + Subtask: Address any issues or bugs.

**Task: Maintenance and Monitoring**

* + Subtask: Set up a monitoring system for the model in production.
  + Subtask: Plan for model updates and maintenance.