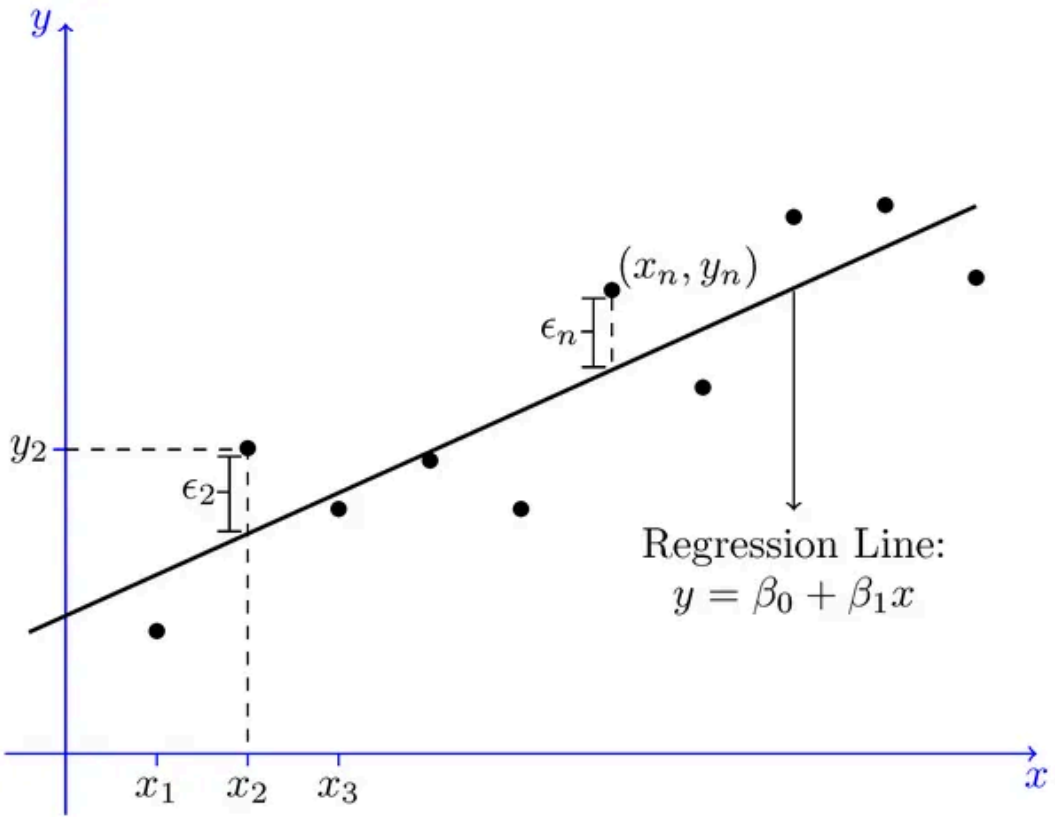
Guwahati House Price Prediction

1. **Objective: The project aims to address the challenge of accurately predicting house prices in Guwahati. Predicting house prices can be difficult for buyers,this project leverages machine learning to provide a more reliable and data-driven way to forecast house prices based on relevant features.By solving this problem, the project helps to make the real estate market more transparent, reducing the risk of overpricing or undervaluation for stakeholders.By integrating historical data and modern prediction techniques, the project contributes to the advancement of data science applications in real estate, while also fostering greater trust and efficiency in the property market.**

2. **Introduction :**
    a. **Machine Learning: Machine Learning (ML) is a subset of artificial intelligence that enables systems to learn from data and make decisions / predictions without being programmed for each task. ML models identify patterns and relationships within data, allowing them to make informed predictions or decisions based on new inputs. In real-world applications, machine learning is used for a wide range of tasks such as image recognition, speech analysis, recommendation systems ,etc based on the historic data.Guwahati's growing real estate sector, using machine learning models to predict house prices offers a data-driven approach to understanding market dynamics and providing more accurate price estimations for both buyers and sellers.**

b. **<u>About the project:</u>** **This project aims to predict house prices in Guwahati using a machine learning model, specifically through a Linear Regression model. Linear regression is a fundamental and widely used algorithm for predictive modeling, particularly in cases where the relationship between the dependent variable (house prices) and independent variables (factors such as size, location, number of rooms, etc.) is assumed to be linear, The choice of linear regression is ideal for this project due to its simplicity and interpretability. In this project, historical housing data from Guwahati is used to train the model. Various features such as property size, number of bedrooms, location, and other amenities are treated as independent variables, while the target variable is the house price. The goal is to identify patterns in these features and estimate the most accurate price for houses in Guwahati .**



3. **<u>Design:</u> Steps involved to create the machine learning model that can predict House price.**
    a. **<u>Data Collection:</u> For this project, the dataset used to train and evaluate the machine learning model was sourced from Kaggle, Kaggle provides a wide range of high-quality datasets for users that are valuable for developing and testing predictive models.**

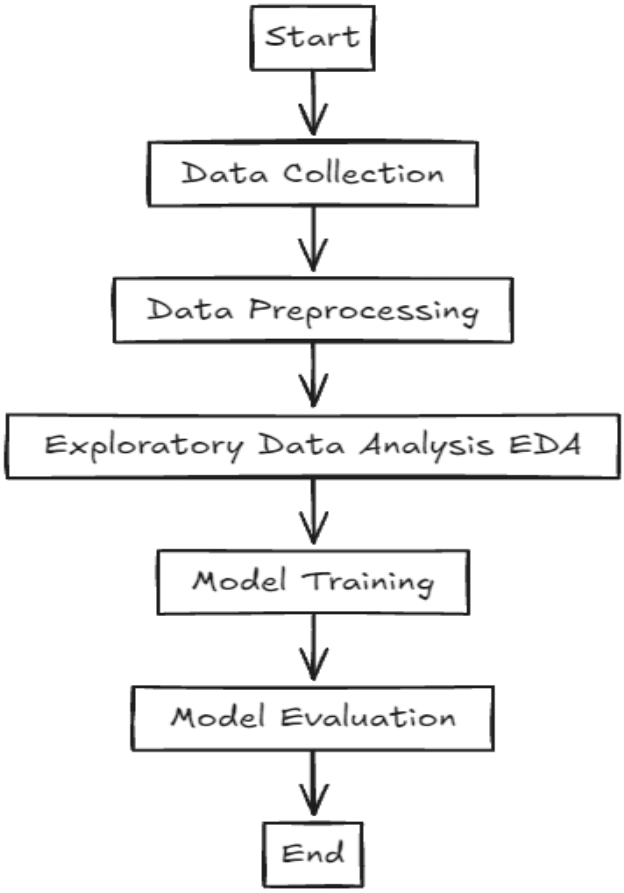
Here is the Dataset . We took this dataset for initial training of our model , as we move forward we scrape data from popular real estate websites such as housing.com, makaan.com etc for batter efficiency of the model performance.

b. Data Preprocessing: Data preprocessing is a crucial step in any machine learning project, as it ensures that the dataset is ready for training ,Here are some steps of data preprocessing-
   i. Handling Missing Values.
   ii. Feature Scaling.
   iii. Encoding Categorical Variables.
   iv. Outlier Detection and Removal.
   v. Splitting the Data.
   vi. Splitting the Data.

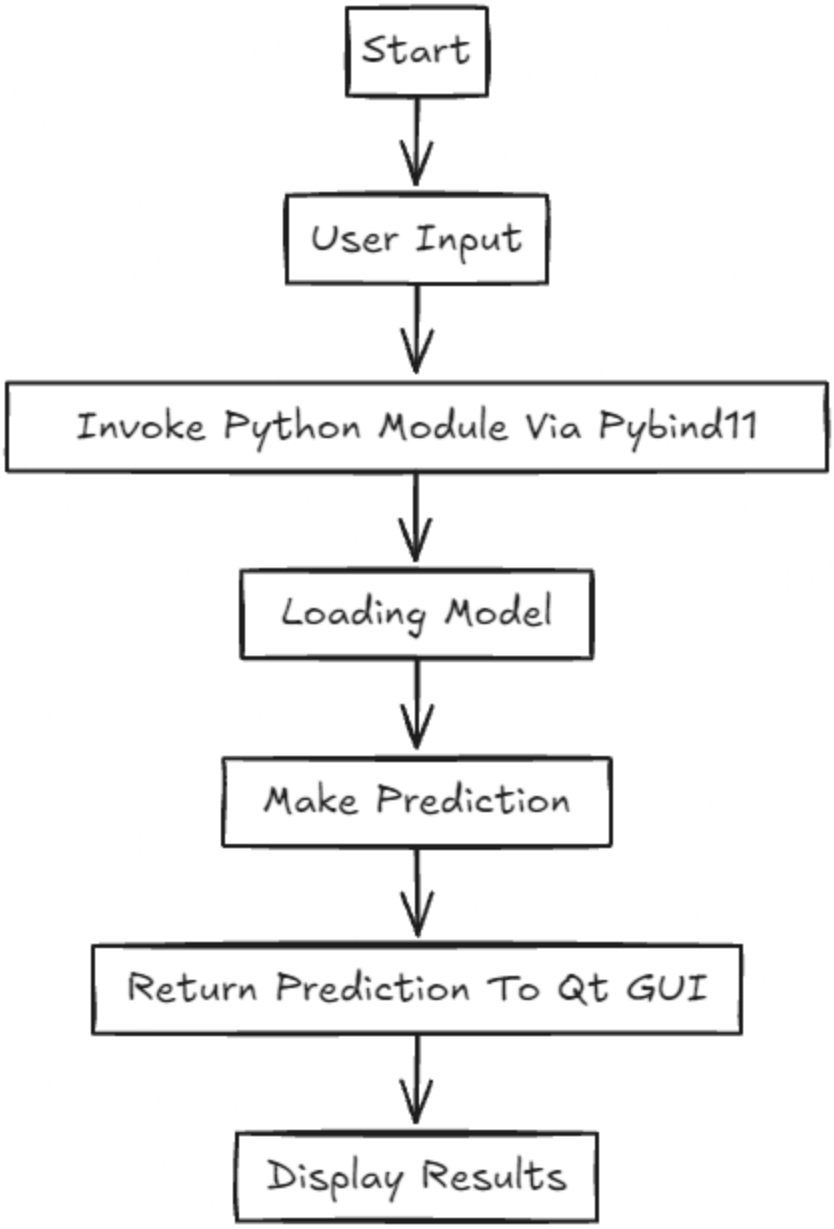
c. Model Training: Once the dataset was preprocessed, the next step was to train the machine learning model. In this project, a Linear Regression model was chosen due to its simplicity and interpretability in predicting continuous numerical values like house prices.
d. Model Evaluation: After training the linear regression model, it was essential to evaluate its performance to ensure it could predict house prices accurately. The evaluation was conducted using the test set, which consists of data unseen by the model during training. This step helps measure the model's capability and how well it will perform when making predictions on new data. After evaluating the model and giving good accuracy, save the model as a .pkl file for future use.
e. Graphical user interface(GUI): To make the model more accessible and user-friendly, a Graphical User Interface (GUI) was developed using Qt. Qt is a popular framework for creating desktop applications with rich, interactive interfaces. The GUI was designed to allow users to input relevant property features, such as the size of the house, number of bedrooms, location, and other attributes. Once

the user inputs the details, the model, loaded from the pickle file, makes predictions and displays the estimated house price.saving the model as a pickle file and using Qt for the GUI ensures that the model is not only functional but also easy to interact with for users who may not be familiar with machine learning or coding.

4. **Data Flow:**
    a. **Data Flow while Training the model:**



b.Data Flow in the whole project:


Start
User Input
Invoke Python Module Via Pybind11
Loading Model
Make Prediction
Return Prediction To Qt GUI
Display Results