**NAME OF THE COURSE : ARTIFICIAL INTELLIGENCE**

**NAME OF THE PROJECT : PREDICITING HOUSE PRICES USING MACHINE LEARNING**

**PHASE - 1**

**PROBLEM STATEMENT:**

 The problem 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. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

**WHAT I UNDERSTAND :**

1. Data Collection: Gathering relevant data is the first step. This dataset should contain information about houses, such as their location, size, number of bedrooms and bathrooms, and any other relevant attributes. Additionally, historical sales data can be useful.
2. Data Preprocessing: Raw data usually needs cleaning and formatting. This involves handling missing values, removing duplicates, and encoding categorical variables. Data outliers may also need to be addressed.
3. Feature Engineering: Creating new features or transforming existing ones can enhance the model's predictive power. For example, you can calculate the price per square foot, create a feature indicating the age of the house, or consider proximity to amenities like schools and parks.
4. Data Splitting: Split the data into two sets: a training set used to train the model and a test set used to evaluate its performance. A common split is 80% for training and 20% for testing.
5. Model Selection: Choose an appropriate machine learning algorithm. Regression models like Linear Regression, Decision Trees, Random Forests, or Gradient Boosting are commonly used for predicting house prices.
6. Model Training: Train the selected model using the training data. The model learns the relationship between the features and the target variable (house prices) during this phase.
7. Model Evaluation: Evaluate the model's performance on the test data using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE). These metrics quantify how well the model's predictions align with the actual prices.
8. Hyperparameter Tuning: Adjust the model's hyperparameters to optimize its performance. Techniques like cross-validation can help fine-tune the model.
9. Deployment: Once satisfied with the model's performance, it can be deployed for making real predictions. Users can input house features, and the model will estimate the price.
10. Monitoring and Maintenance: Continuously monitor the model's performance in a production environment and retrain it periodically with new data to ensure it remains accurate over time.

**SOURCE CODE :**

# Import necessary libraries

import pandas as pd

import numpy as np

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression # Example algorithm

# Step 1: Data Collection

# Load the dataset containing information about houses

data = pd.read\_csv('house\_data.csv')

# Step 2: Data Preprocessing

# Handle missing values, duplicates, and encode categorical variables

# Address outliers if necessary

# Step 3: Feature Engineering

# Create new features or transform existing ones

# Step 4: Data Splitting

# Split the data into training and testing sets

X = data[['Location', 'SquareFootage', 'Bedrooms', 'Bathrooms', 'OtherFeatures']] # Features

y = data['Price'] # Target variable

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Step 5: Model Selection

# Choose a machine learning algorithm (e.g., Linear Regression)

model = LinearRegression()

# Step 6: Model Training

model.fit(X\_train, y\_train)

# Step 7: Model Evaluation

# Evaluate the model's performance on the test data

y\_pred = model.predict(X\_test)

# Calculate metrics like MAE, MSE, RMSE, etc.

# Step 8: Hyperparameter Tuning (if needed)

# Fine-tune the model's hyperparameters using techniques like cross-validation

# Step 9: Deployment (Optional)

# Deploy the model for making real predictions in a production environment

# Step 10: Monitoring and Maintenance (Continuous)

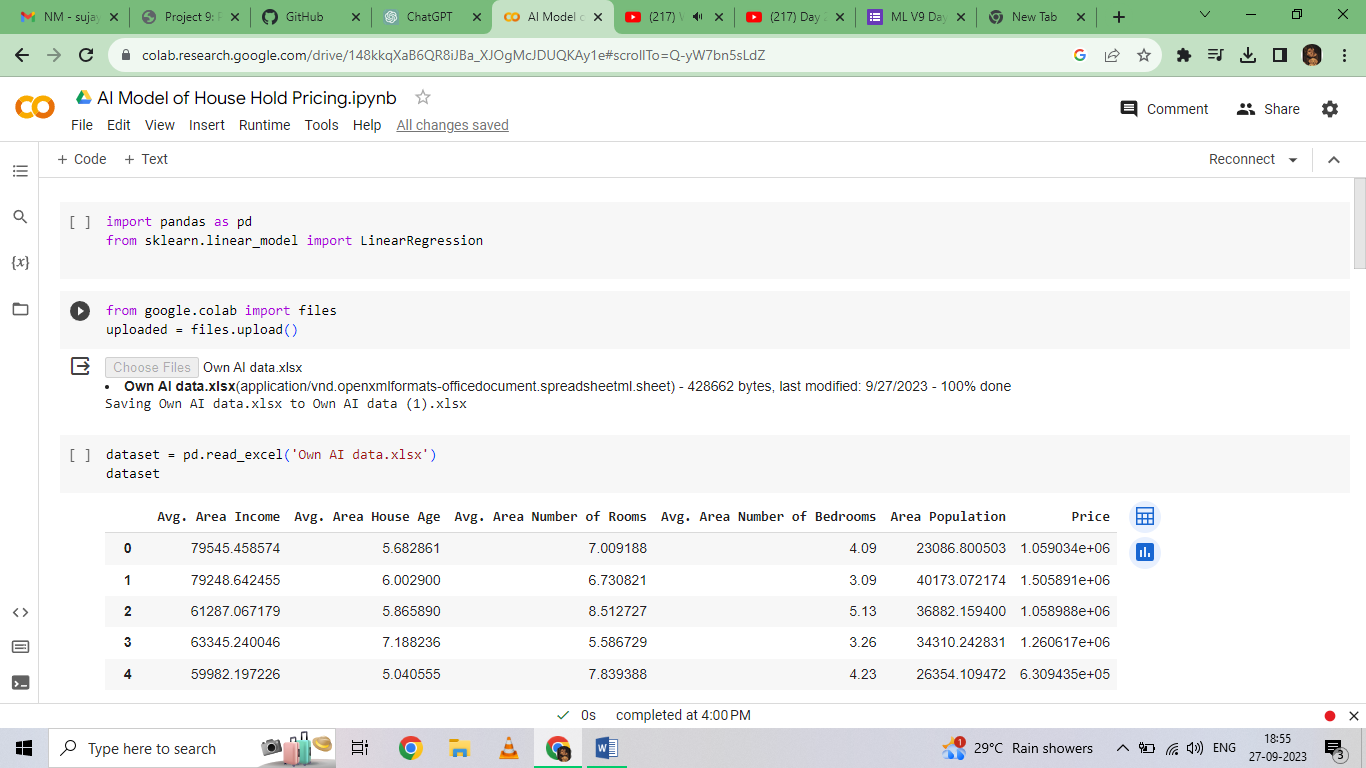
# Continuously monitor the model's performance and retrain it with new data

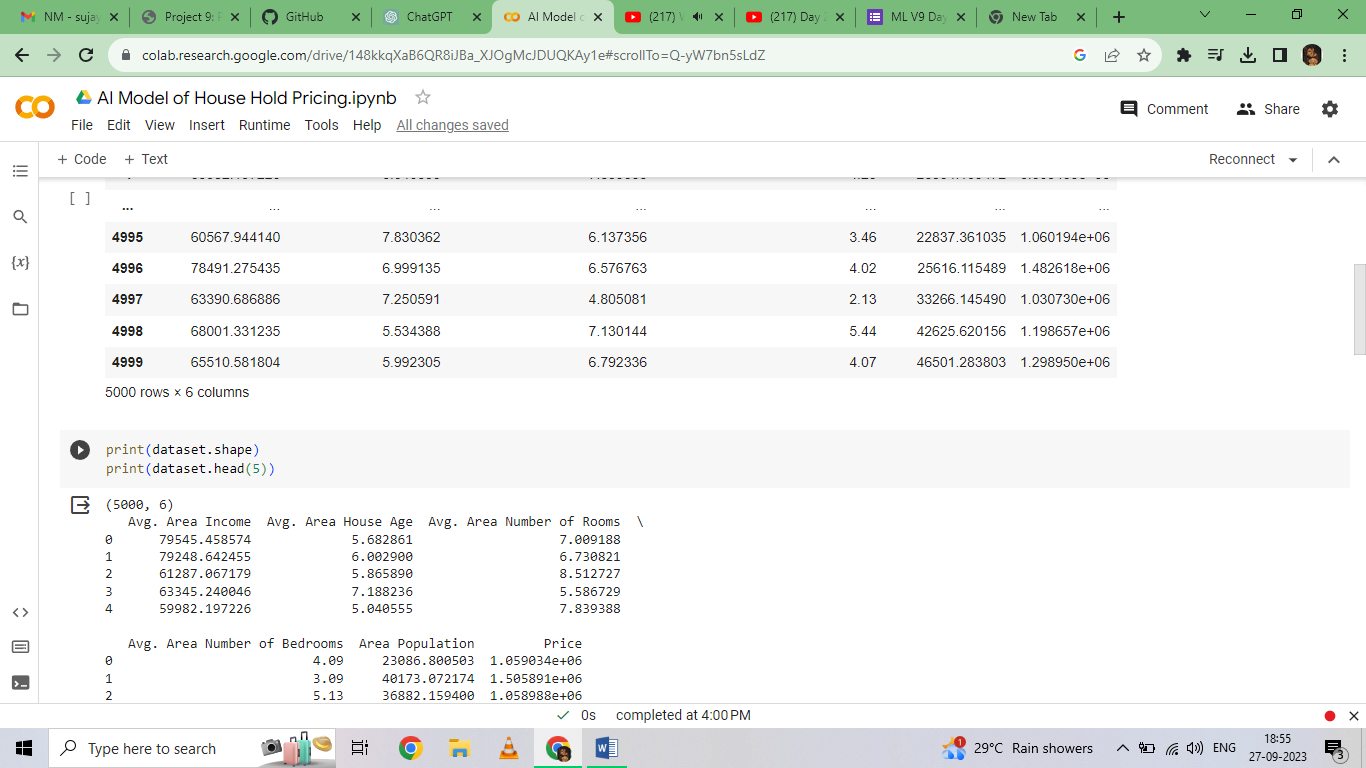
# Predicting house prices for new input data

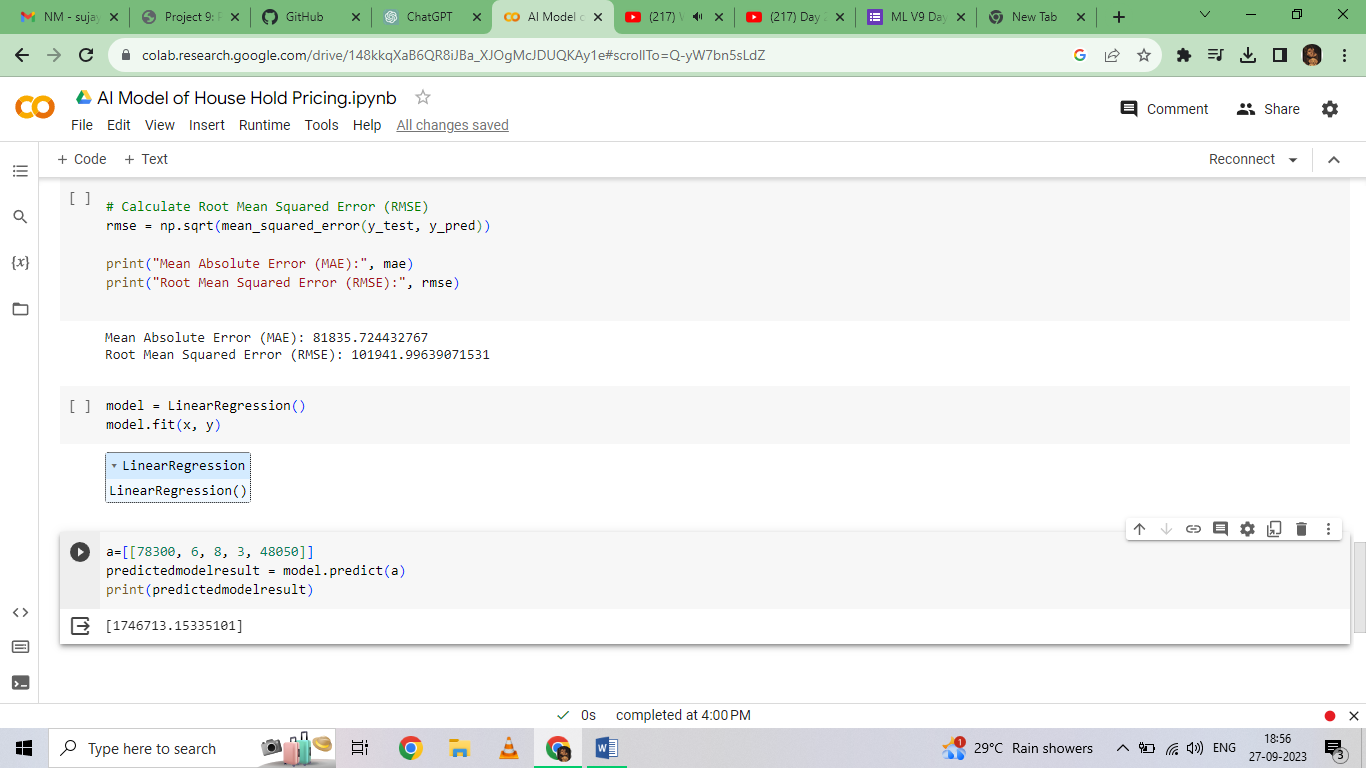
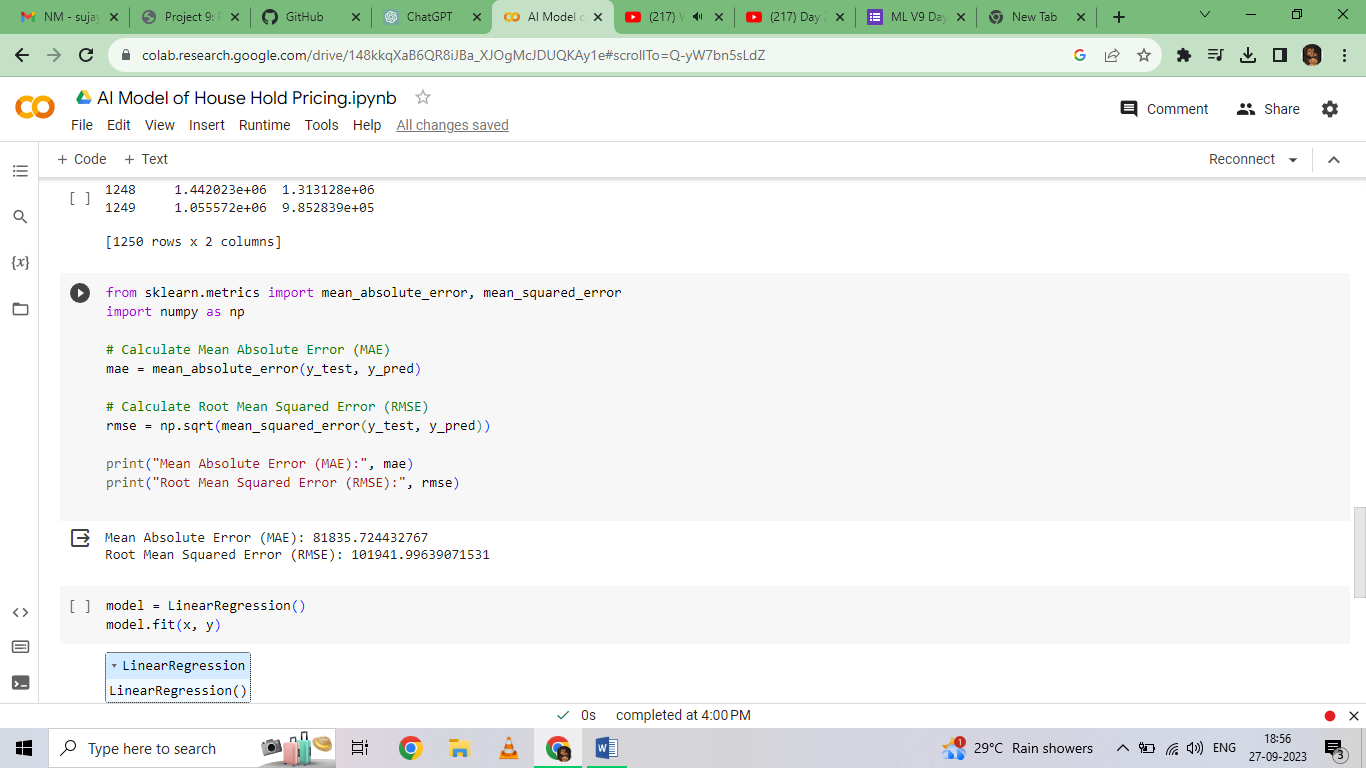
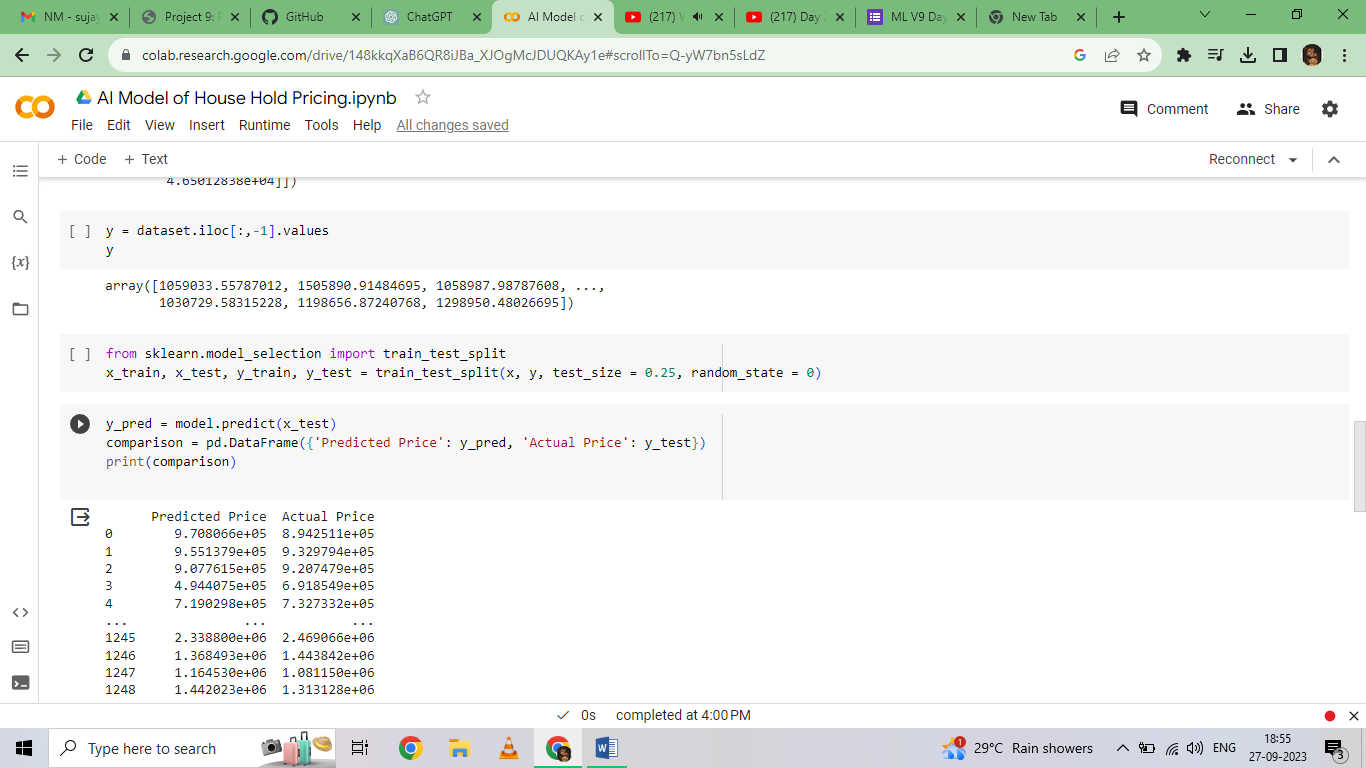
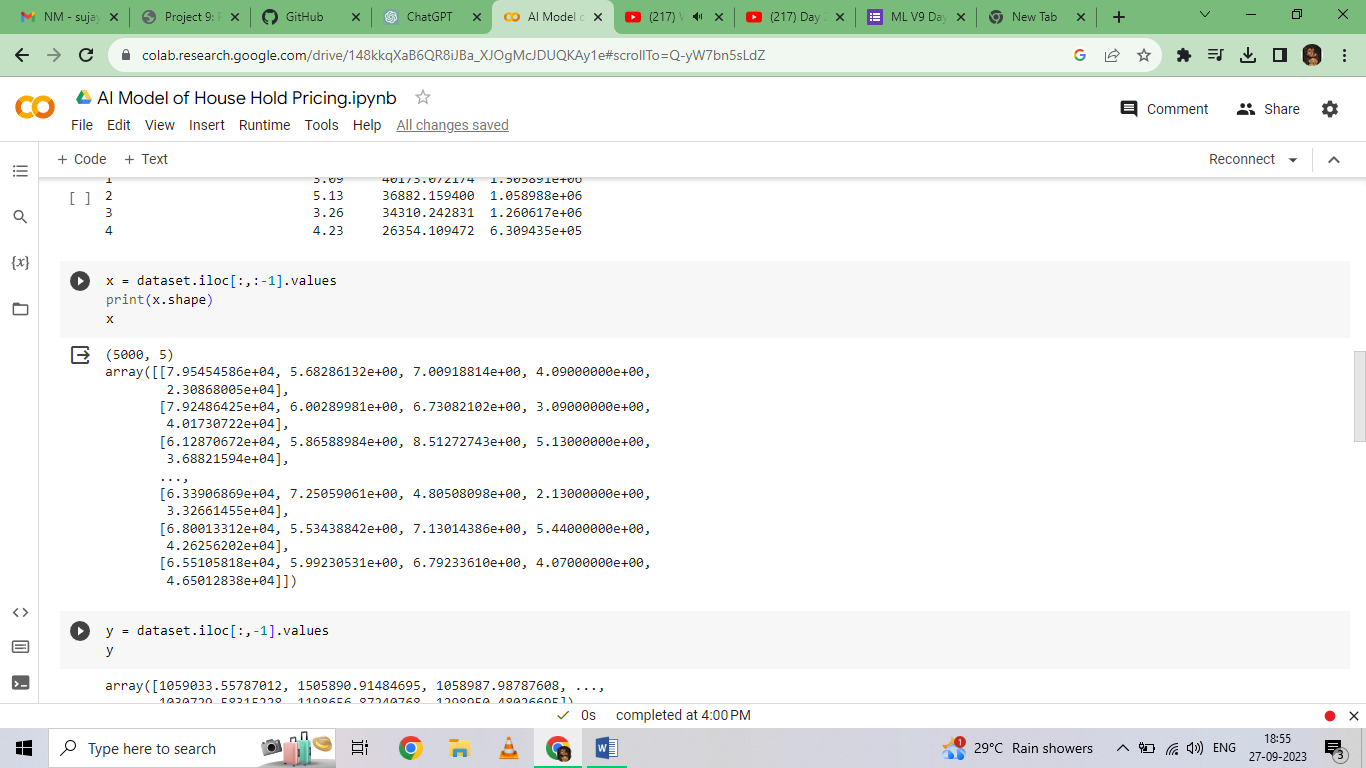
new\_data = pd.DataFrame({'Location': ['New\_Location'], 'SquareFootage': [2500], 'Bedrooms': [3], 'Bathrooms': [2], 'OtherFeatures': ['Some Features']})

predicted\_price = model.predict(new\_data)

print("Predicted House Price:", predicted\_price)

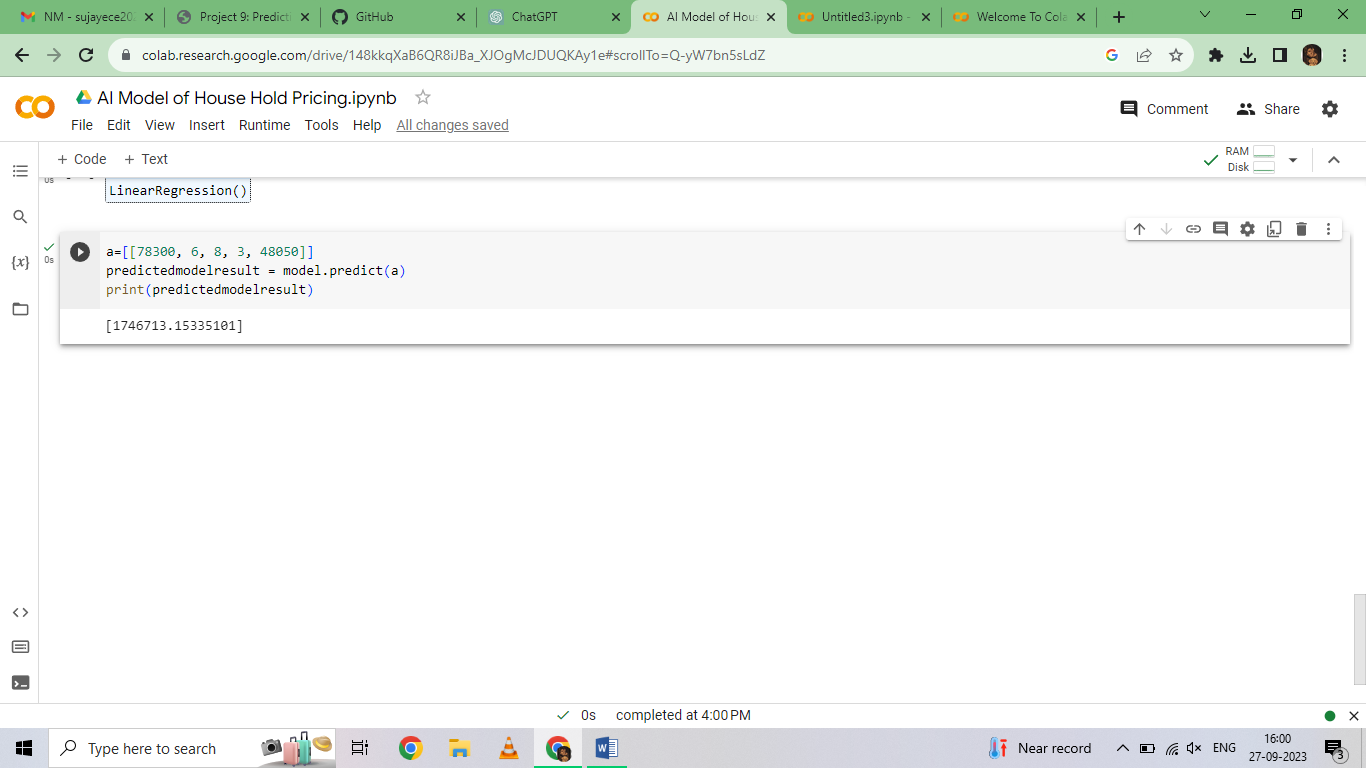
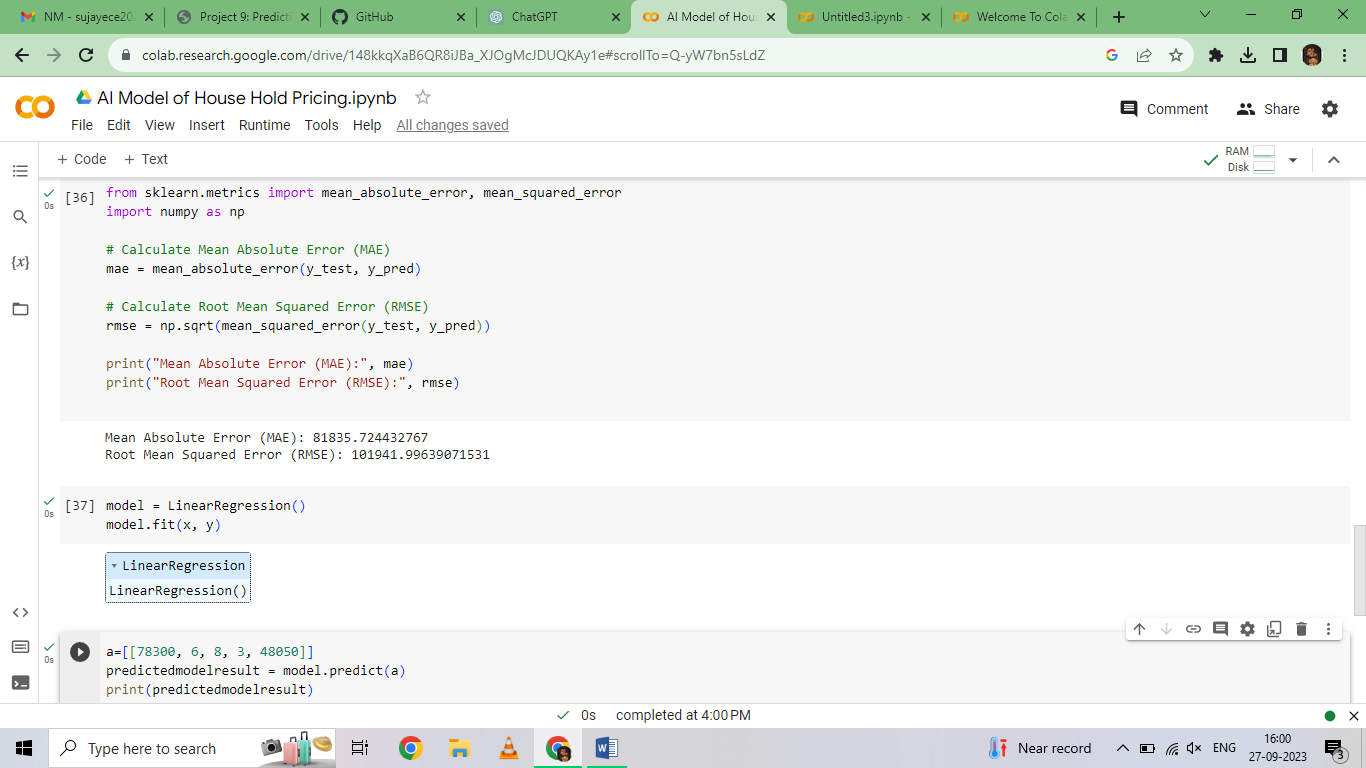
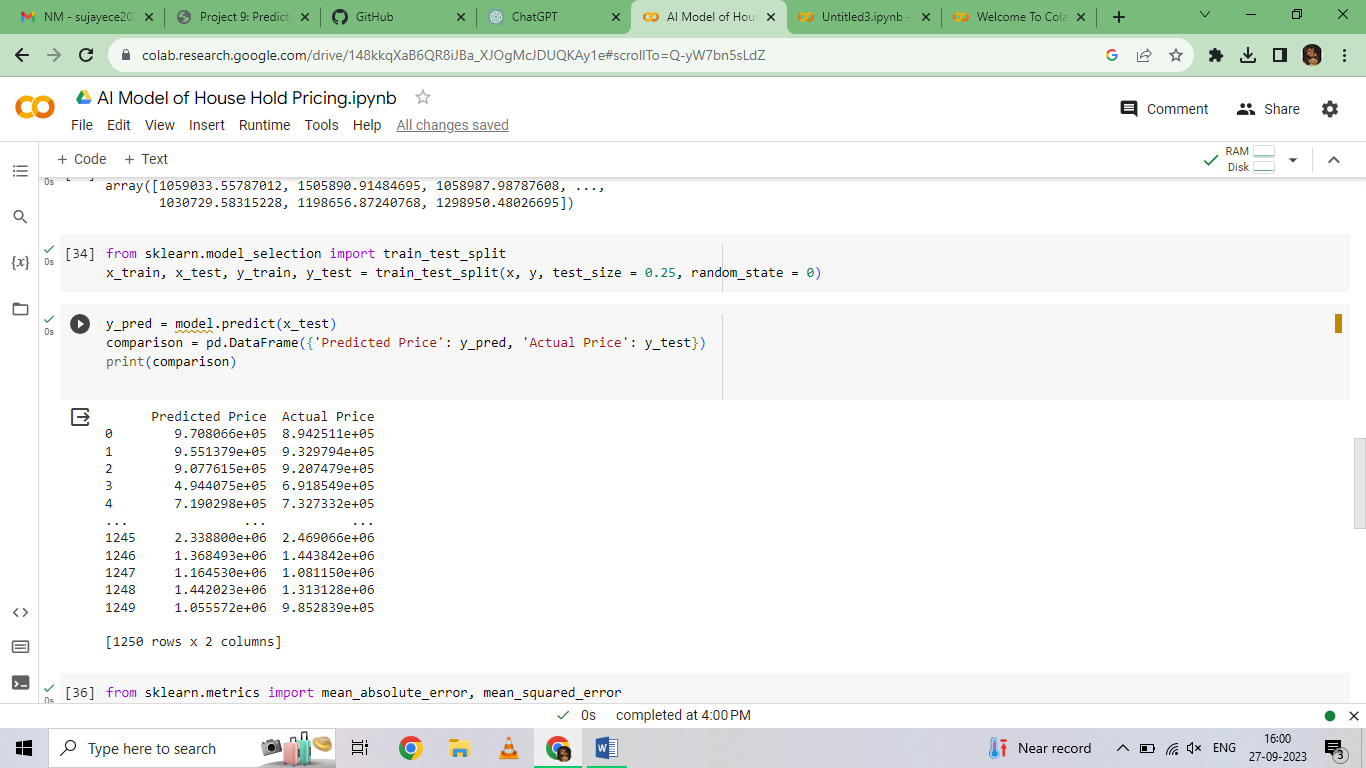
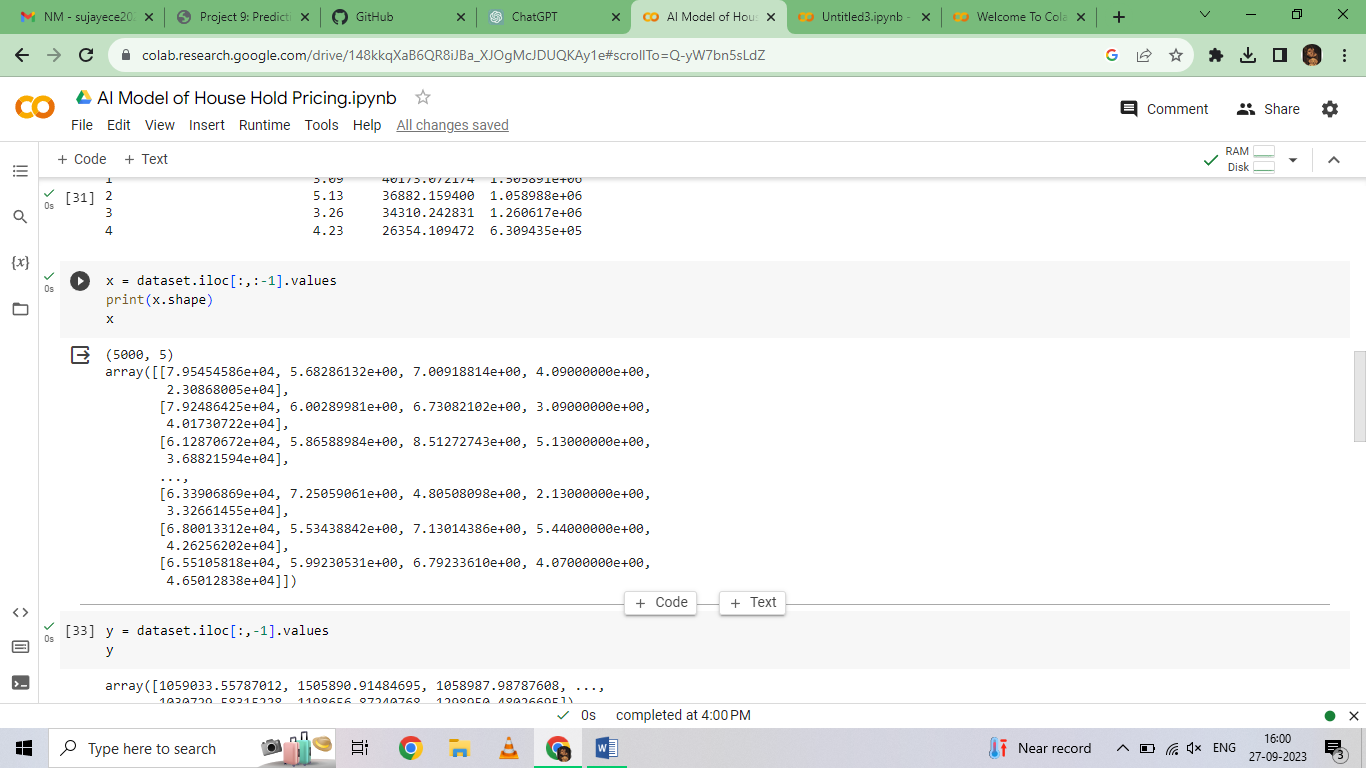
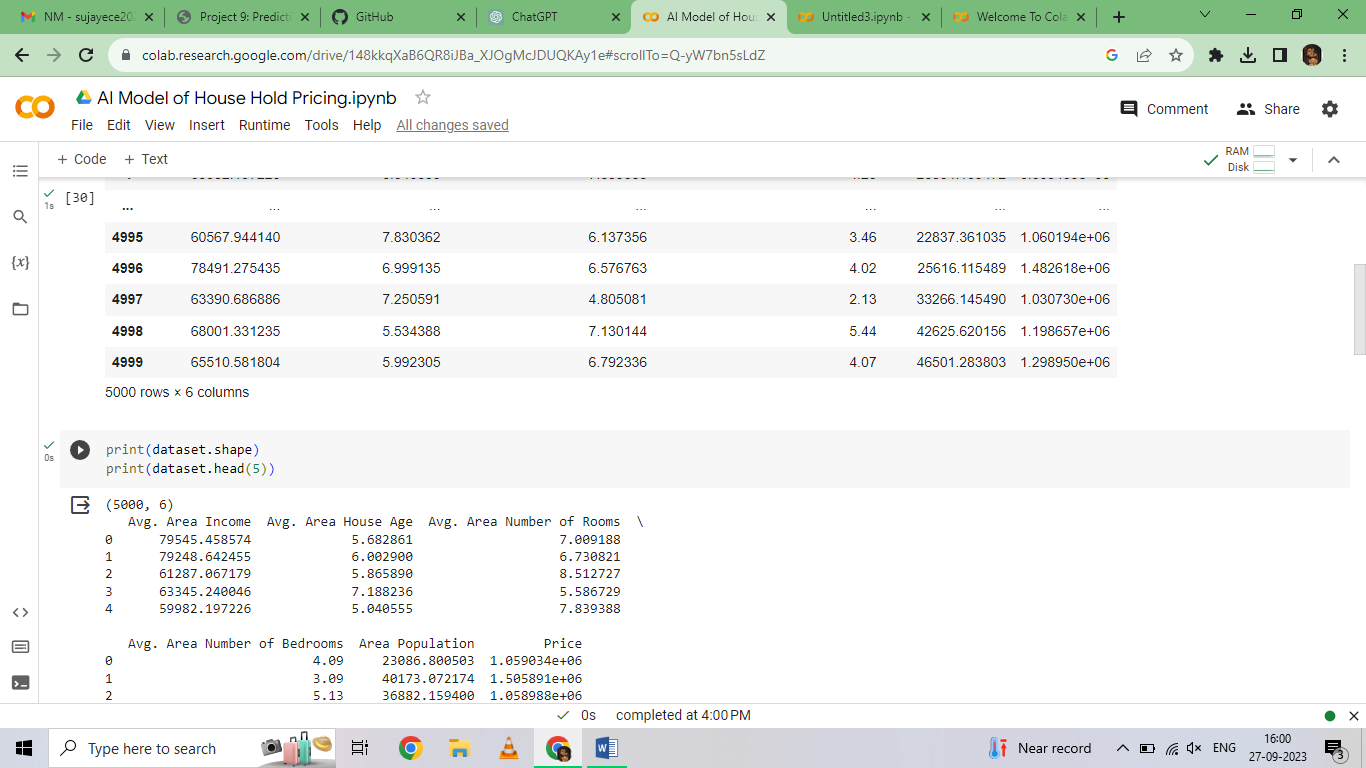
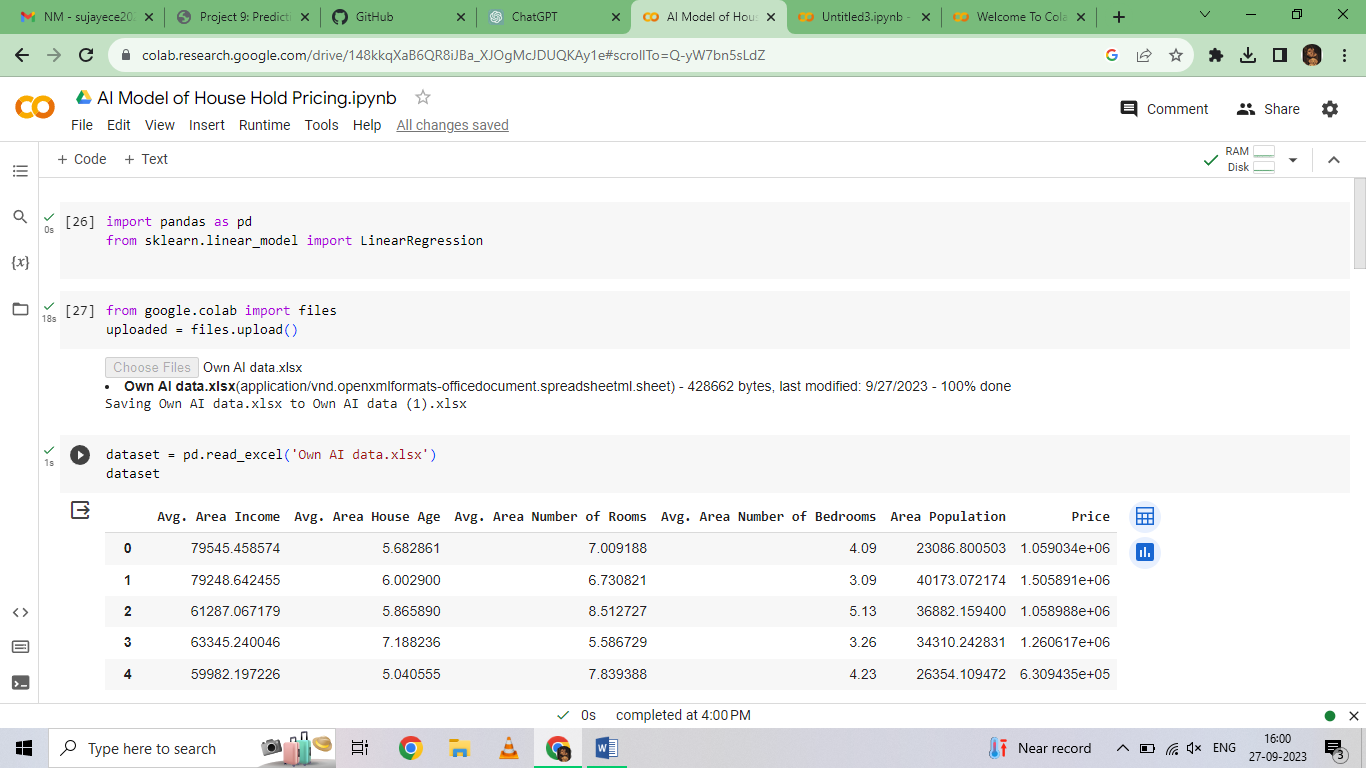
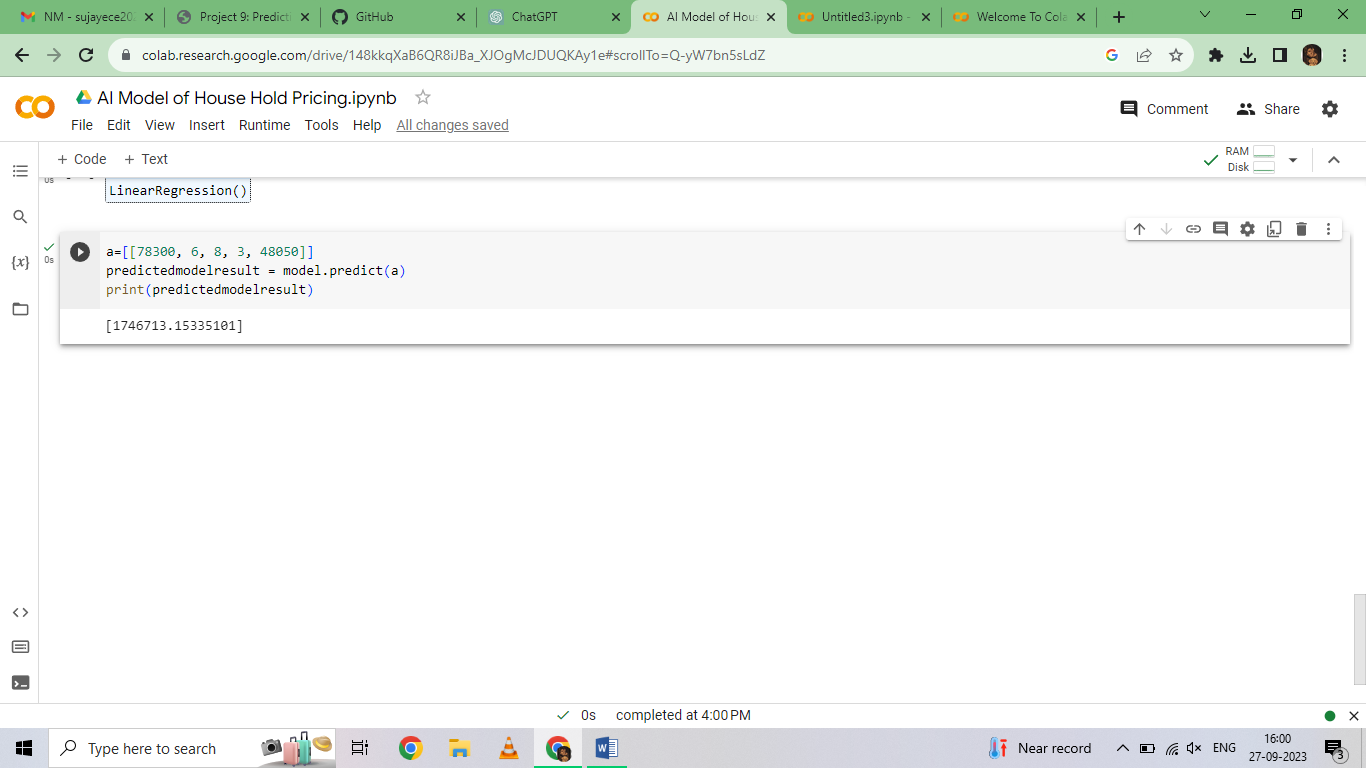
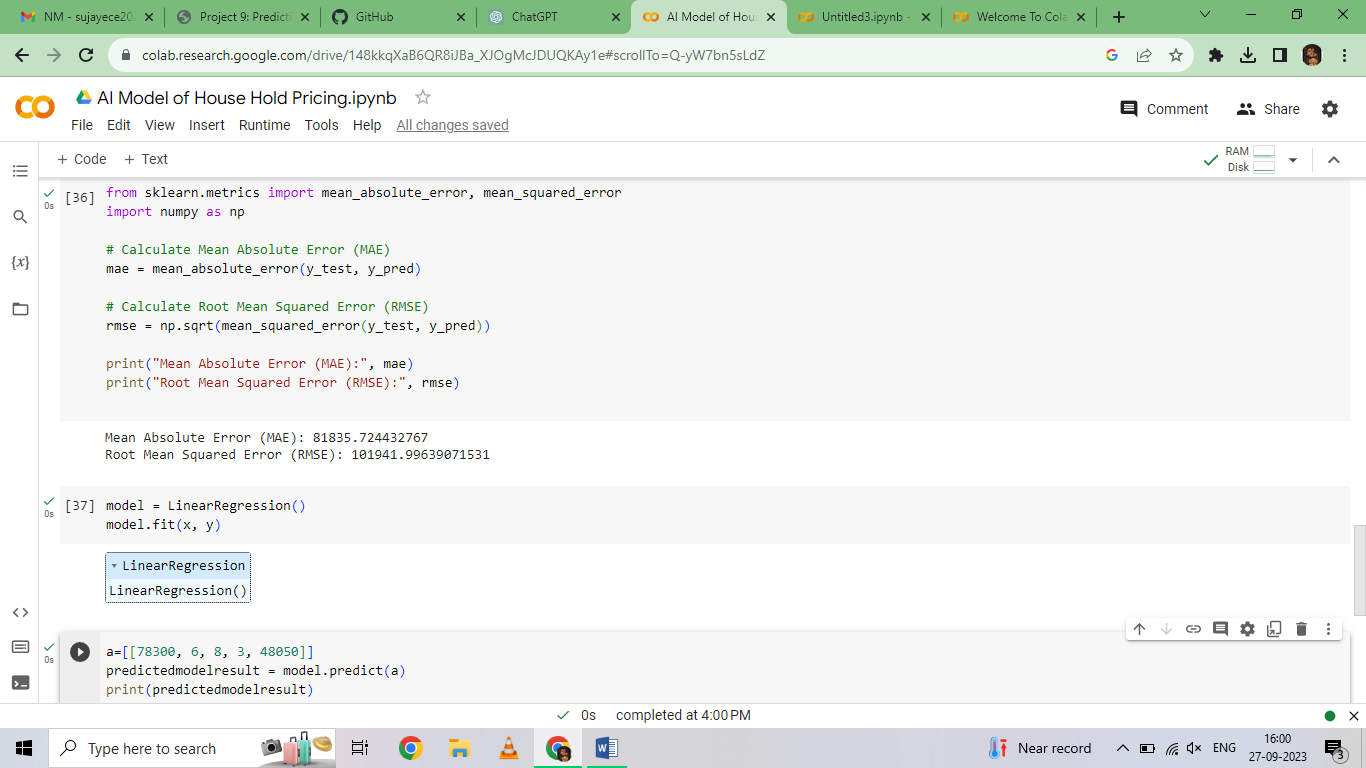
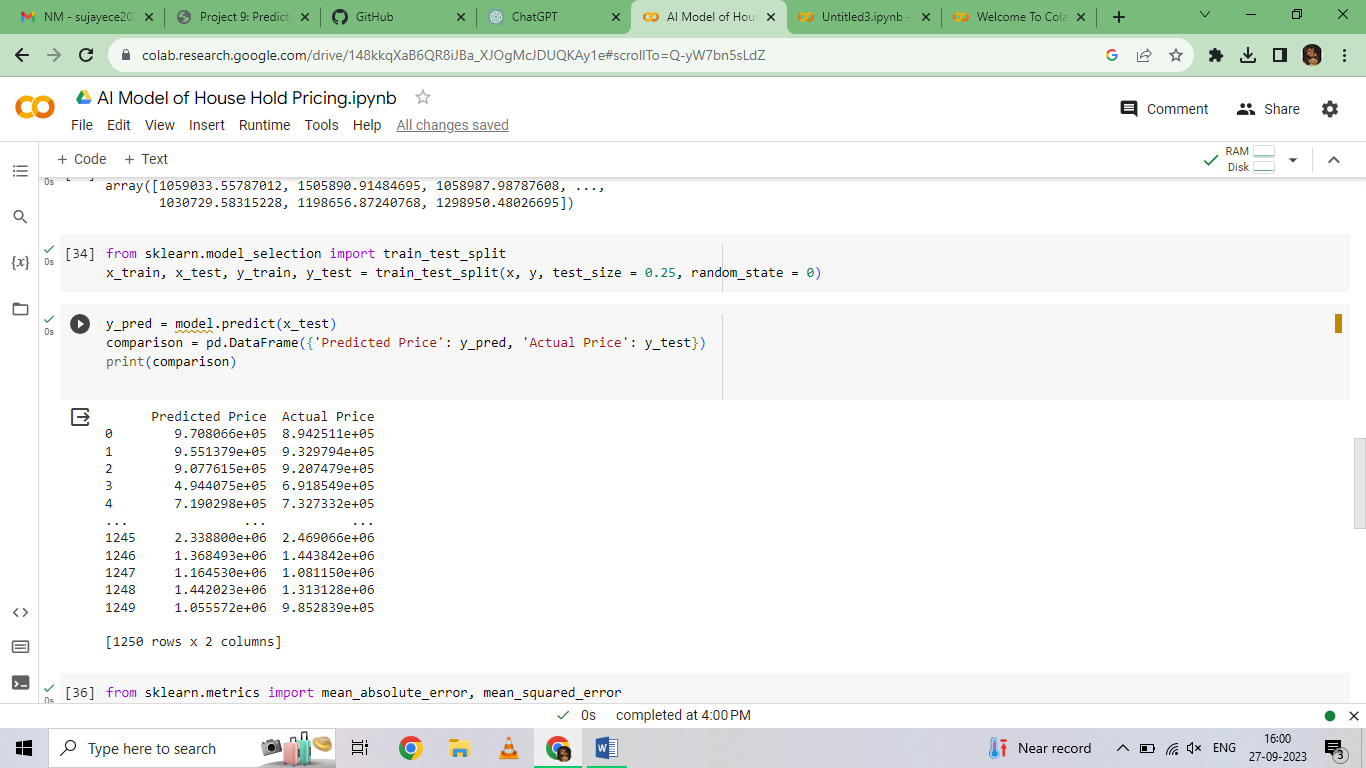
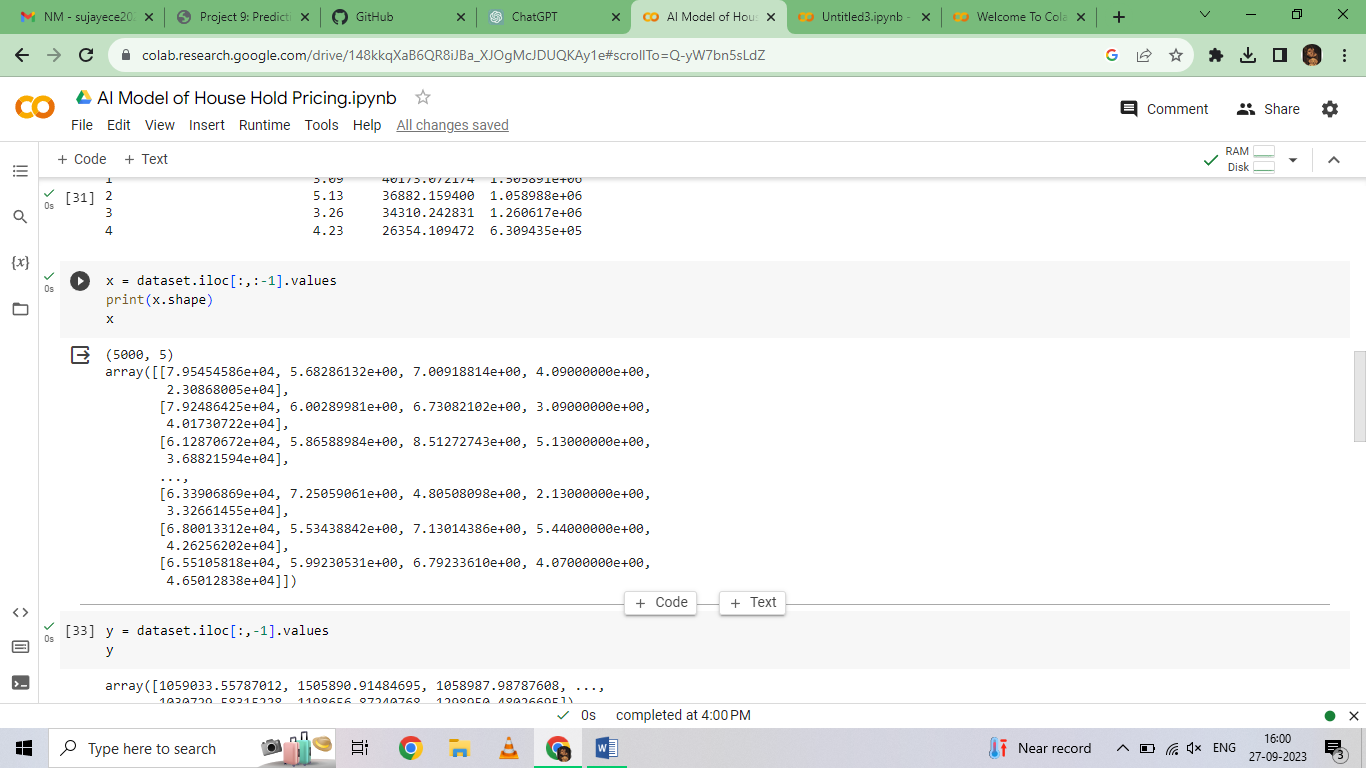
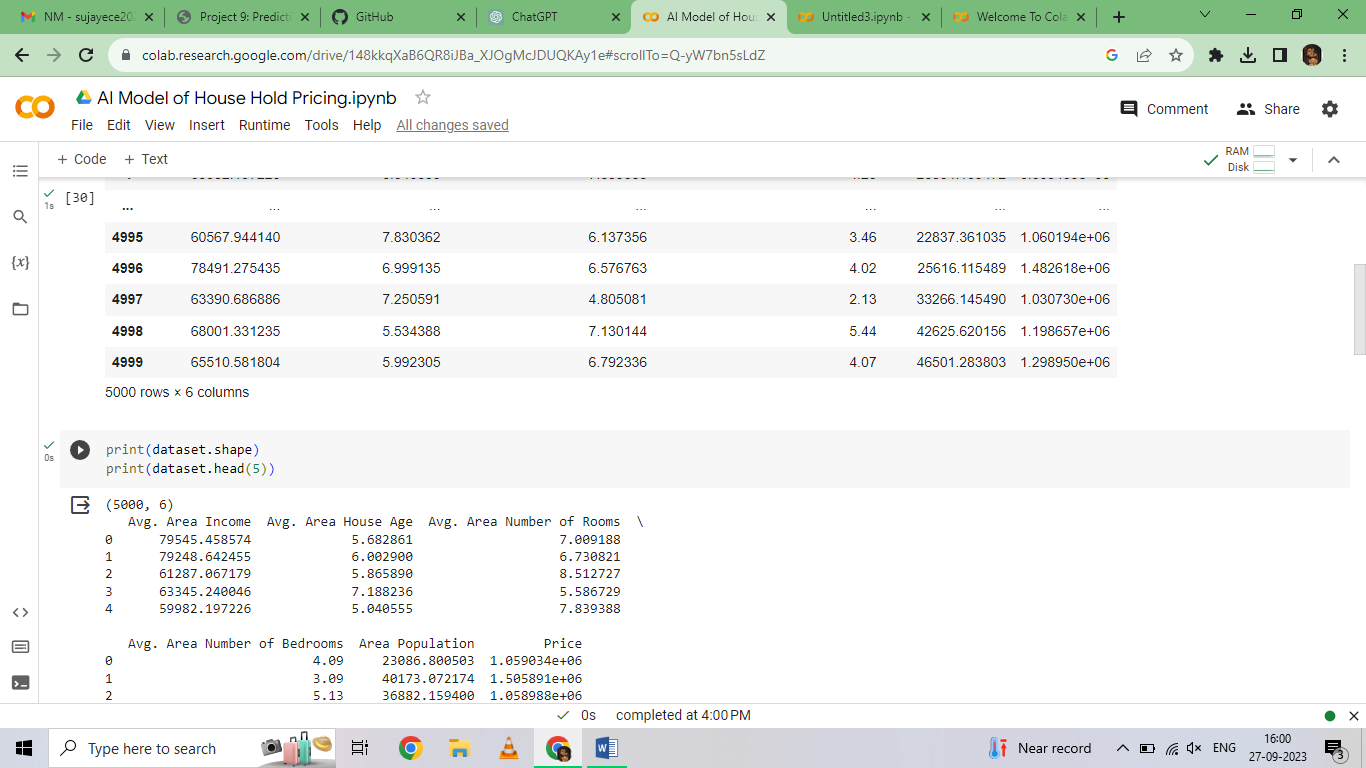
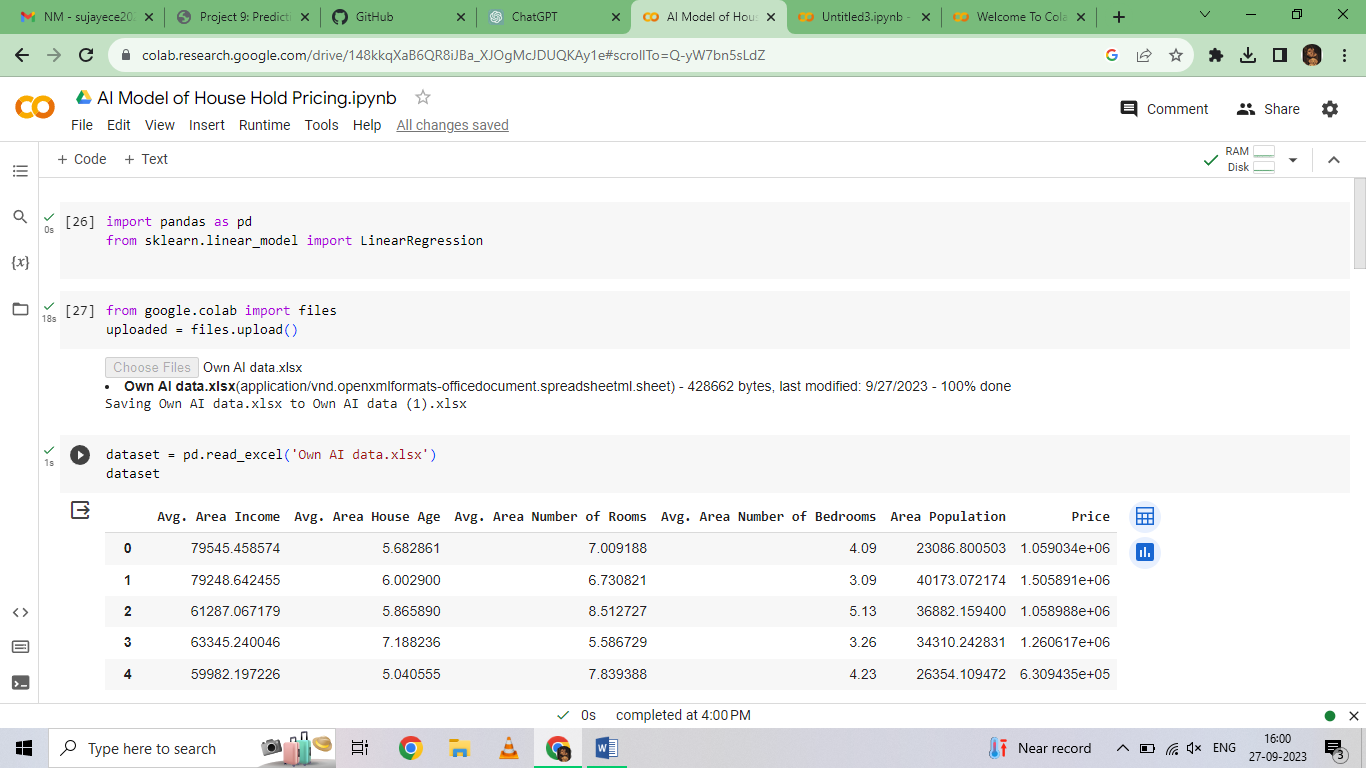






NAME OF THE STUDENT : SUJAY RAVICHANDRAN

DEPARTMENT : BE-ECE / 3RD YEAR

COLLEGE : M.P.NACHIMUTHU.M.JAGANADHAN ENGINEERING COLLEGETop of Form