ALI MUSA

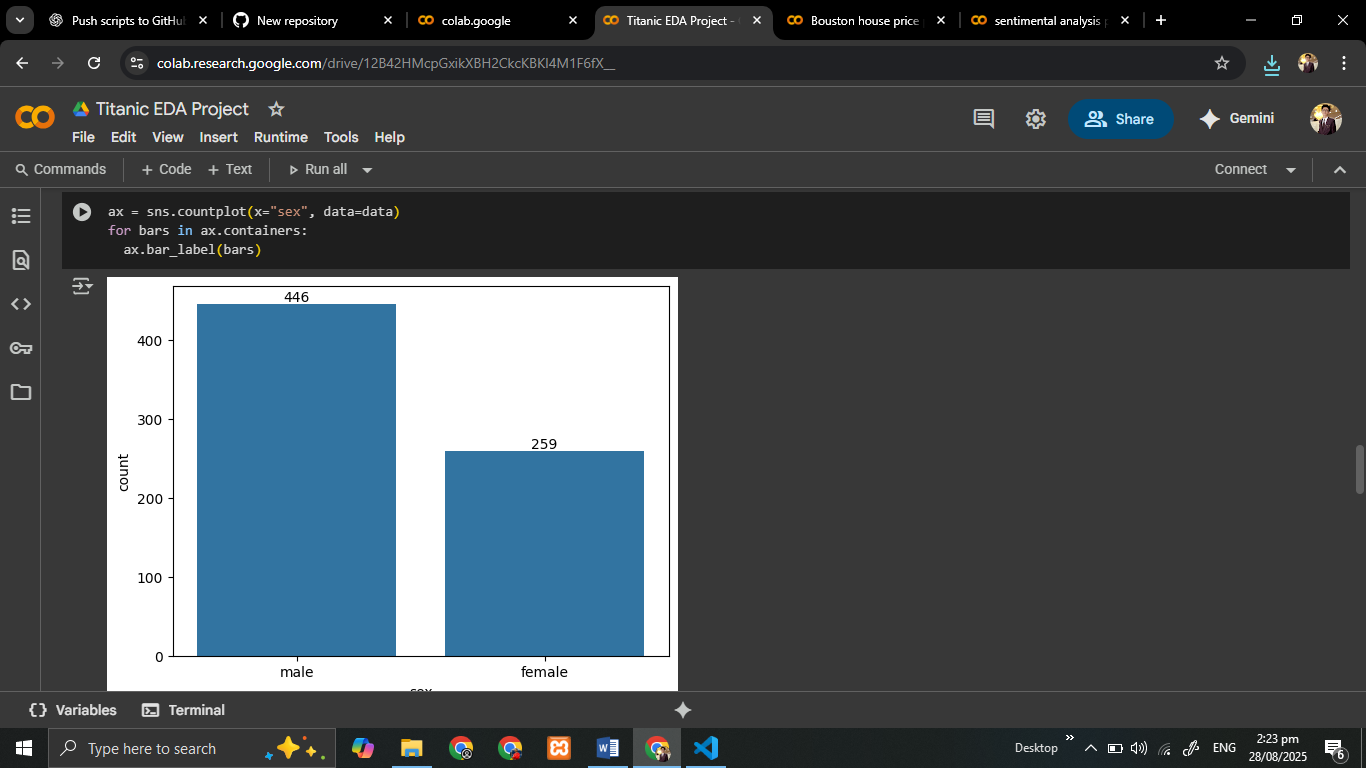
# 📌 Project Overview

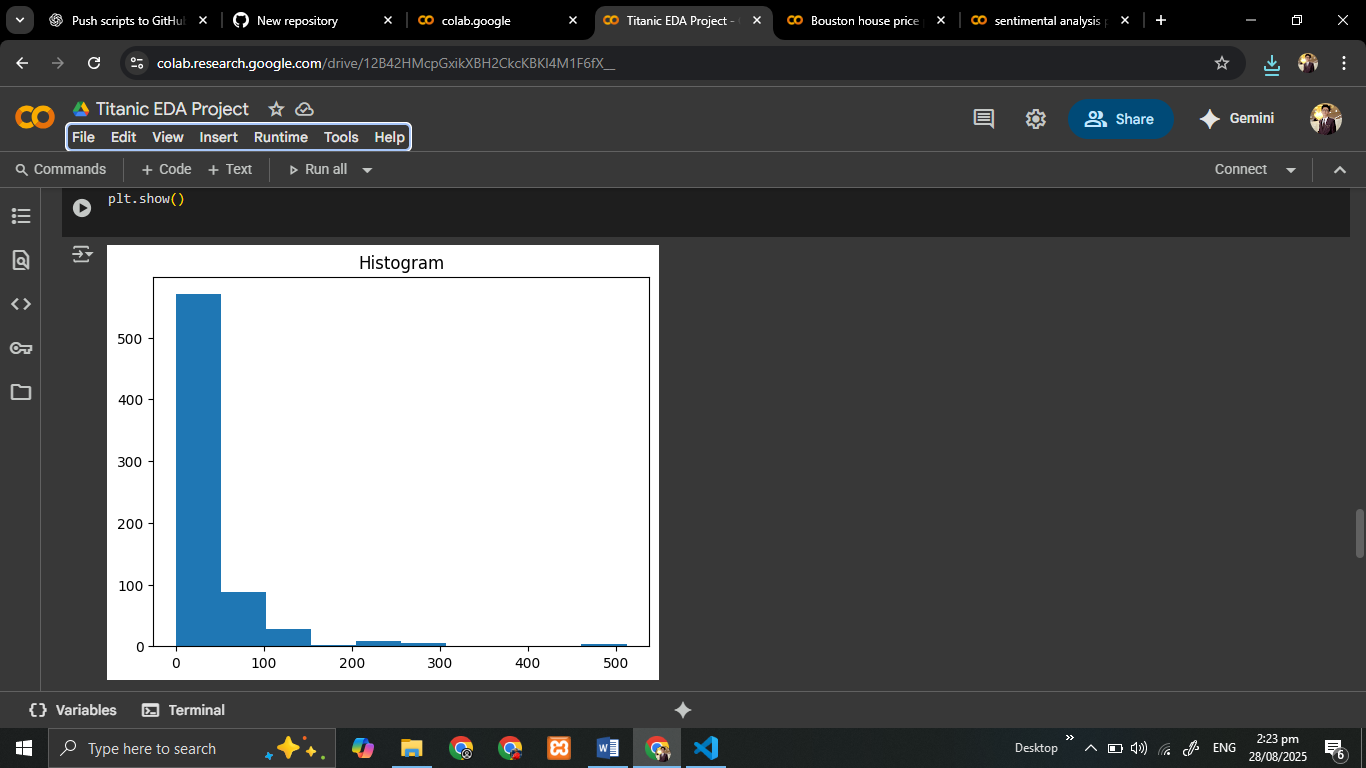
This repository contains three main tasks that demonstrate data analysis and machine learning concepts. The tasks cover Exploratory Data Analysis (EDA), Sentiment Analysis, and Regression modeling.

# 📝 Task 1: Exploratory Data Analysis (Titanic Dataset)

In this task, the Titanic dataset was used to perform exploratory data analysis (EDA). The goal was to understand the data, clean missing values, and visualize patterns related to passengers’ survival.  
  
Key steps performed:  
- Data Cleaning: Handling missing values in 'Age', 'Embarked', and 'Cabin'.  
- Data Visualization: Distribution of survival rate by gender, class, and age groups.  
- Feature Analysis: Understanding how passenger class, gender, and age influenced survival.

**VISUALS :**





# 📝

OBSERVATIONS :

Socio-economic status mattered → Wealthier, higher-class passengers survived more.

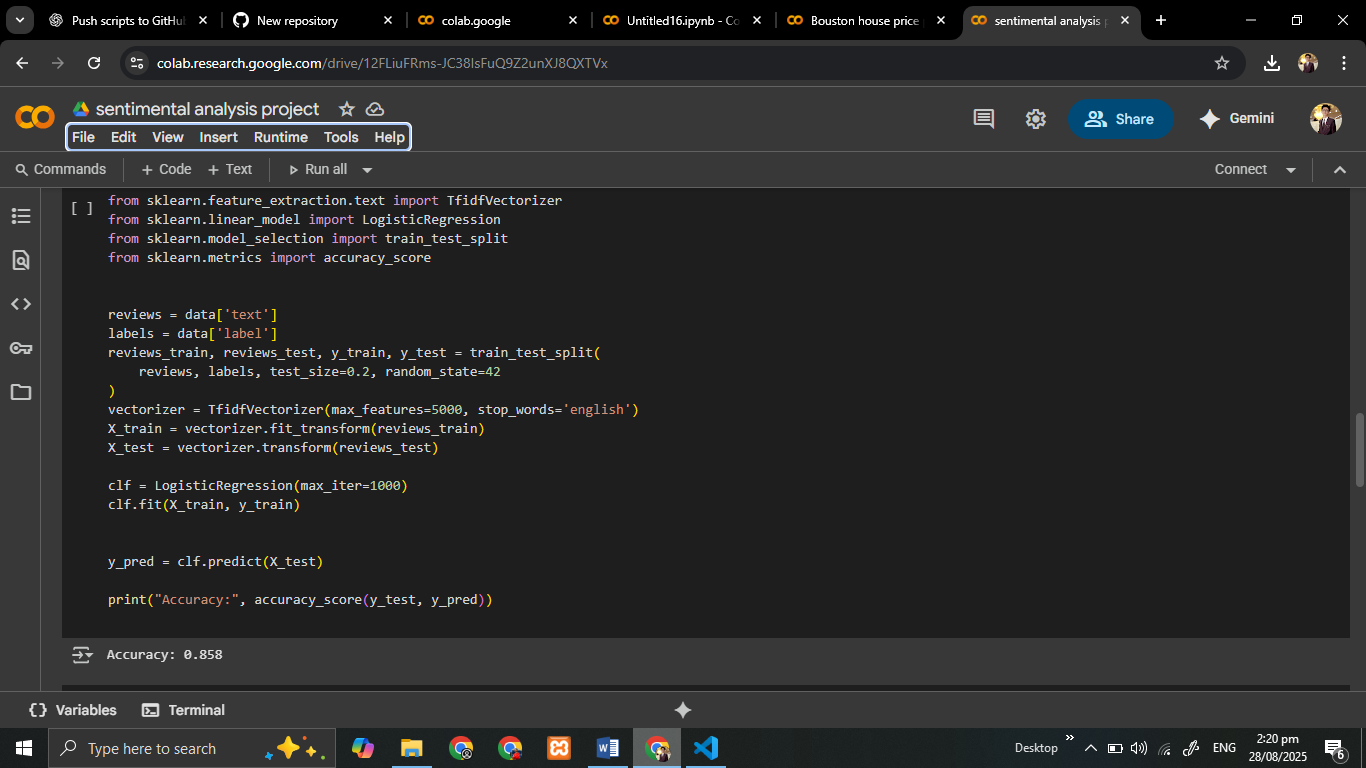
Fare & Pclass strongly explain survival patterns.

Age had a small effect → younger had slightly better survival.

# Task 2: Sentiment Analysis (Movie Reviews)

In this task, a sentiment analysis model was built using a dataset of movie reviews. The goal was to classify reviews as positive or negative based on textual content.  
  
Key steps performed:  
- Text Preprocessing: Lowercasing, removing stopwords, punctuation, and special characters.  
- Feature Extraction: Using Bag of Words (BoW) or TF-IDF.  
- Model Training: Logistic Regression / Naive Bayes for binary classification.  
- Evaluation: Accuracy, Precision, Recall, and F1 Score were calculated.

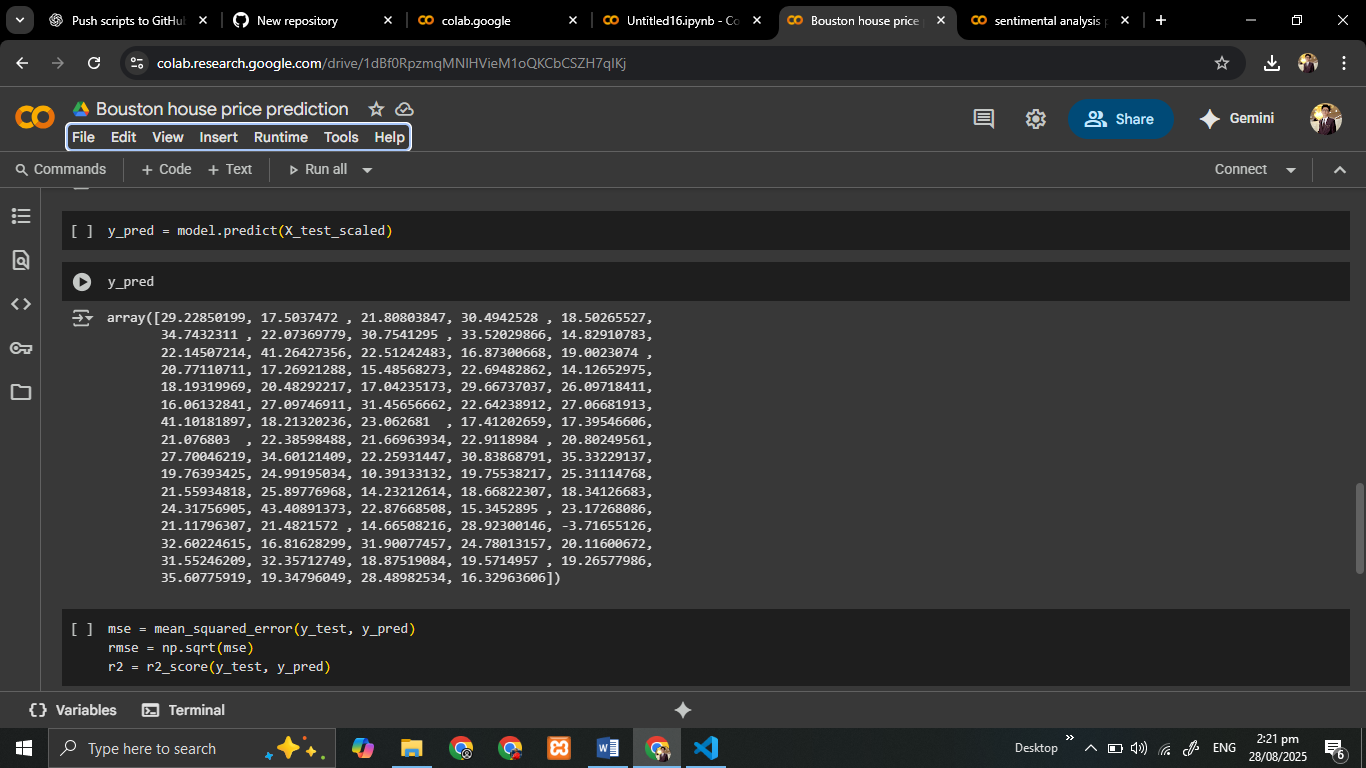
**VISUALS :**



# 📝 Task 3: Boston House Price Prediction

This task involved predicting house prices using the Boston Housing dataset. The objective was to apply regression techniques to estimate the median value of houses.  
  
Key steps performed:  
- Data Exploration: Checking correlations between features such as RM (average rooms), LSTAT (lower status population), and MEDV (house value).  
- Data Preprocessing: Normalization and feature selection.  
- Model Training: Linear Regression was applied to predict housing prices.  
- Evaluation: Model performance was measured using RMSE (Root Mean Square Error).

**VISUALS :**



# 

# Requirements

- Python 3.12  
- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, nltk  
- Jupyter Notebook (for analysis and visualization)

# 👨‍💻 Author

Ali Musa