**Task no 1**

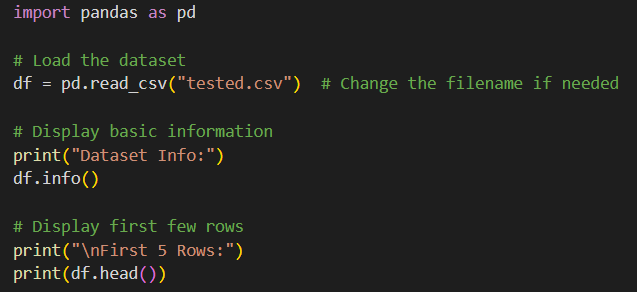
**EDA and Visualization of a Real-World Dataset**

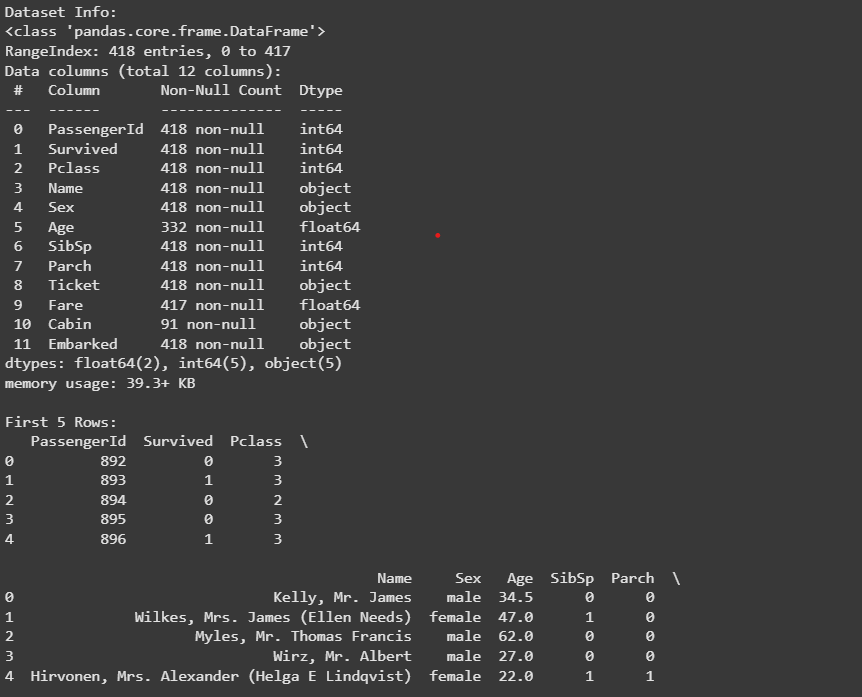
Step by Step performance of EDA and Visualization on real world Dataset taking Titanic Dataset as example.

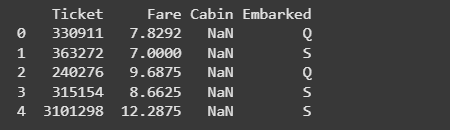
**Step 1:**

* **Load the Dataset: Use Pandas to load and explore the dataset**

INPUT:

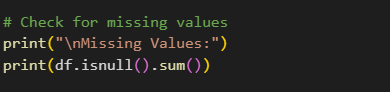


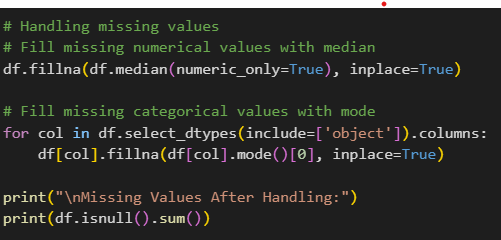
OUTPUT: 



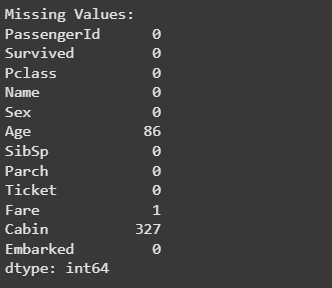
* **Data Cleaning: Handle missing values using imputation techniques or removal**

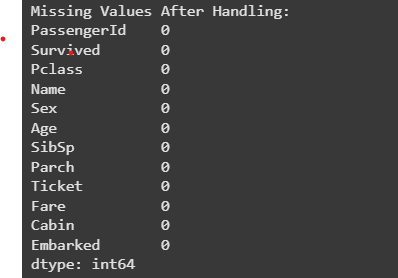
INPUT:



****

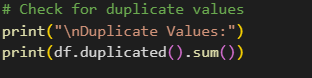
OUTPUT:

****

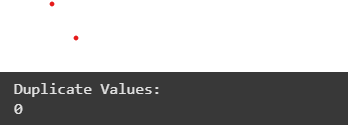


* **Remove duplicates.**

INPUT:



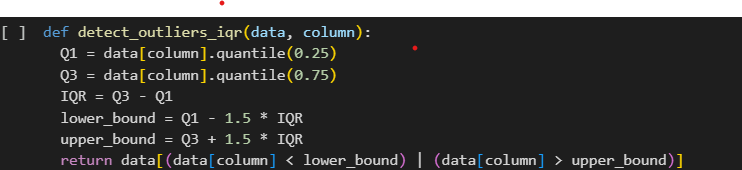
OUTPUT:



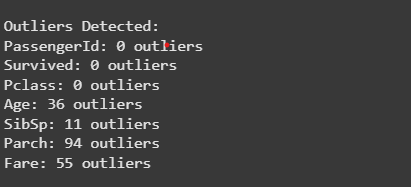
* **Identify and manage outliers using statistical methods or visualization.**

INPUT:

Identifying the outliers

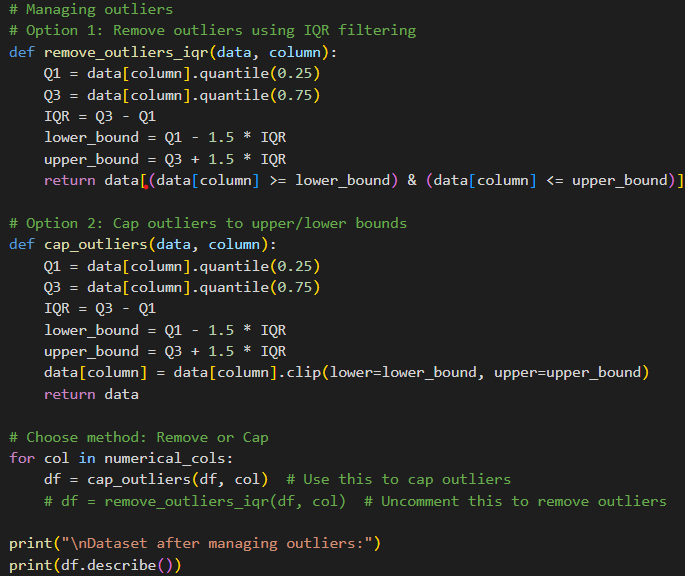


OUTPUT:

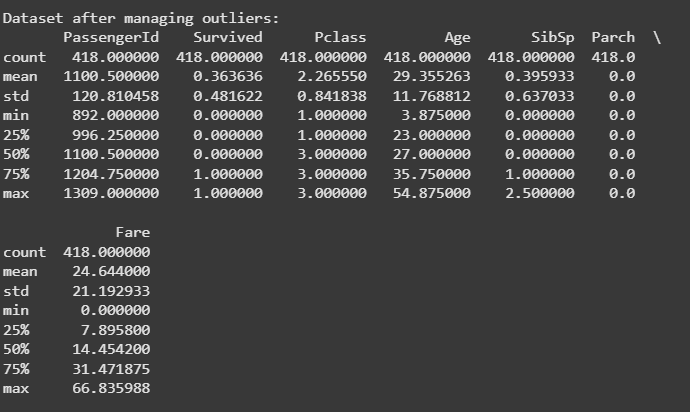


**Managing the outliers**

INPUT:



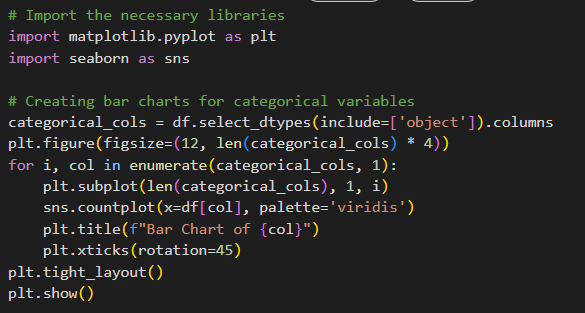
OUTPUT:



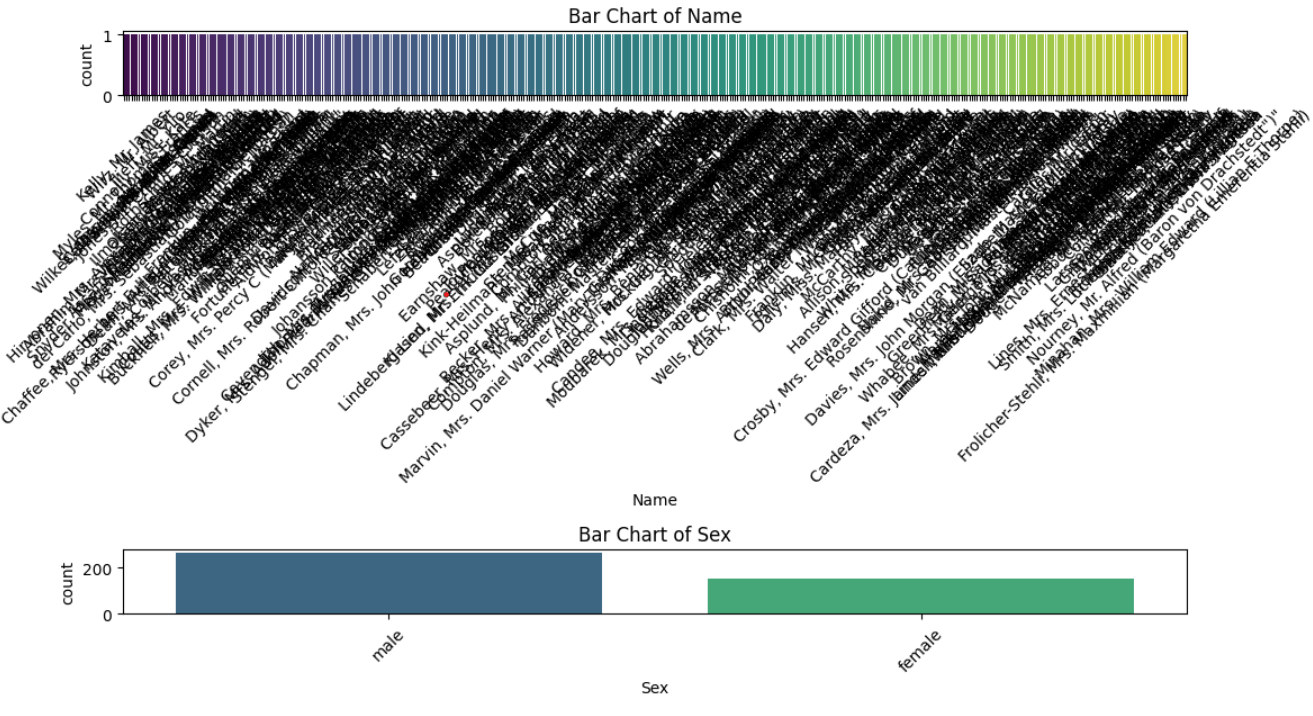
**Visualizations:**

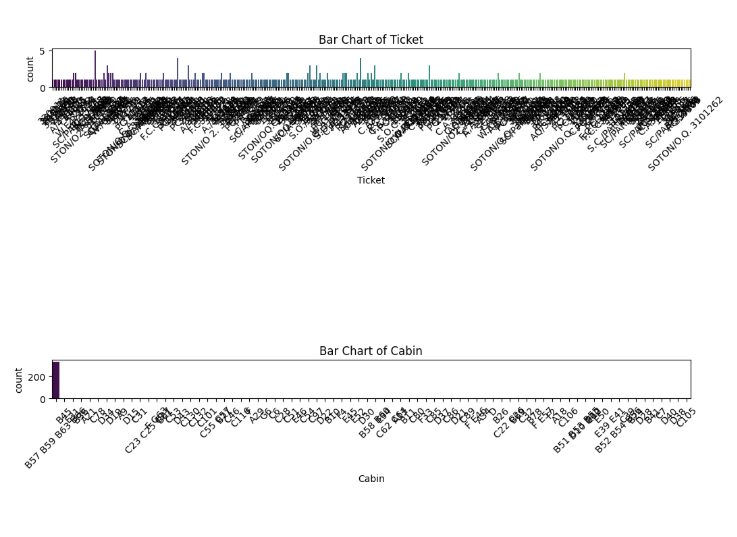
* + **Create bar charts for categorical variables.**

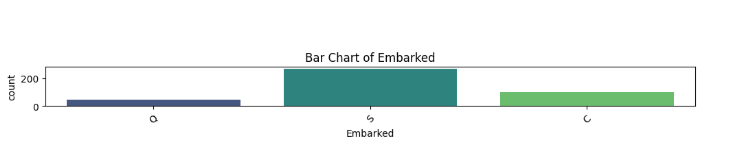
INPUT:



OUTPUT:

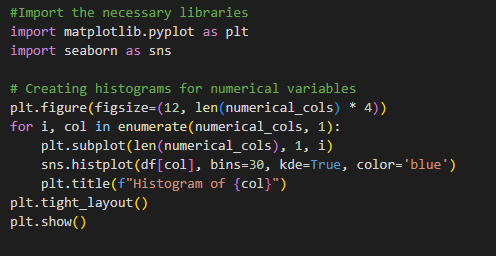




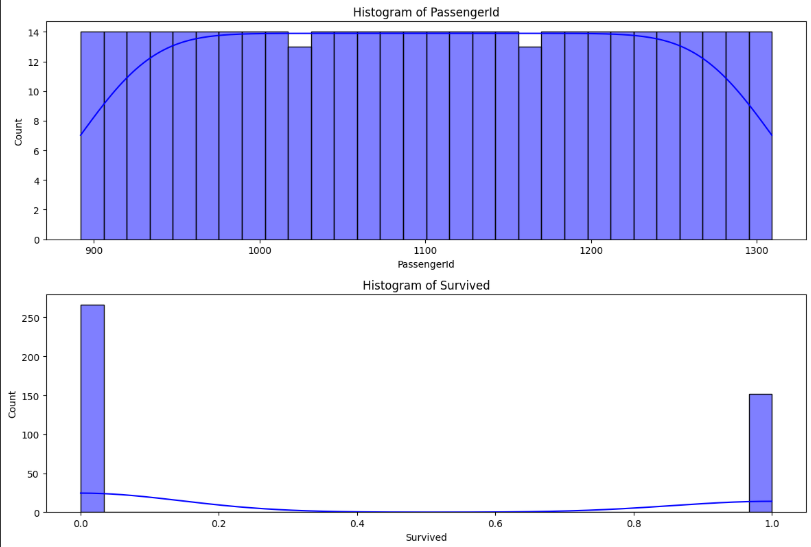


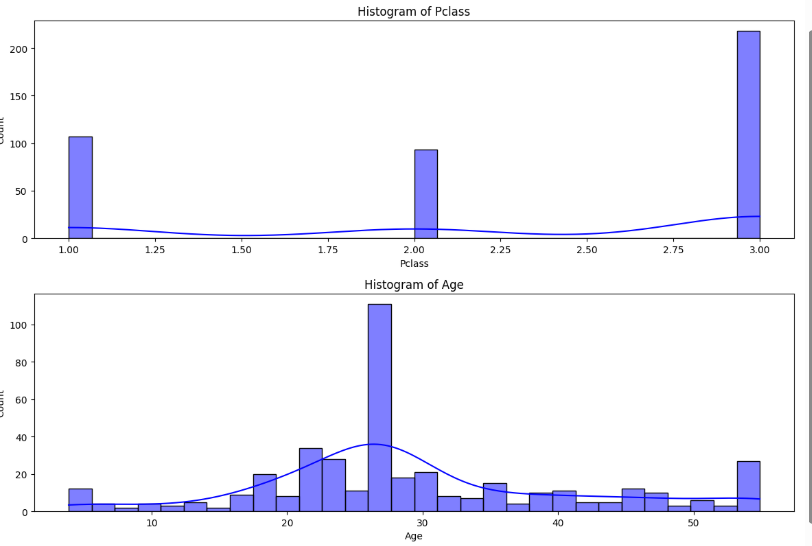
* **PLOT HISTOGRAMS FOR NUMERIC DISTRIBUTIONS**

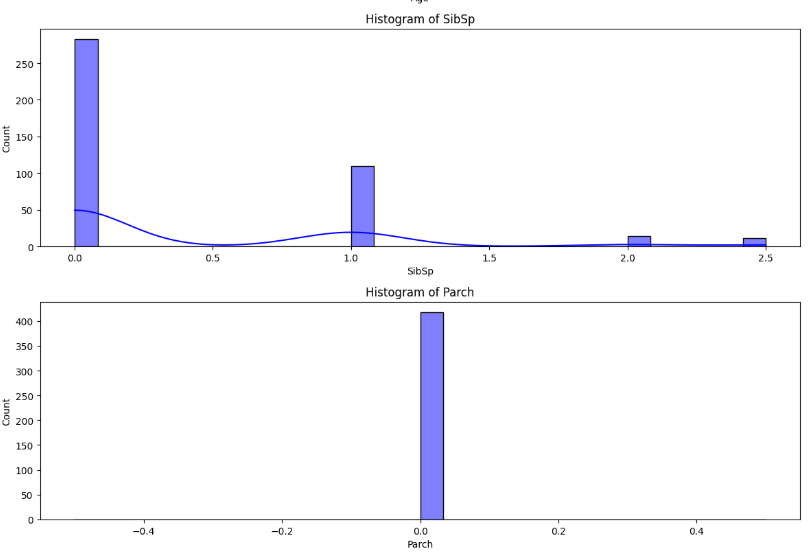
INPUT:

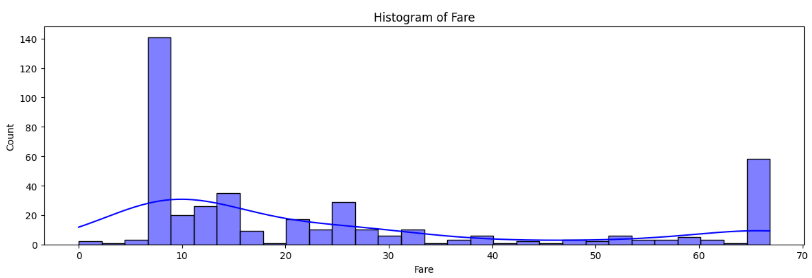


OUTPUT:



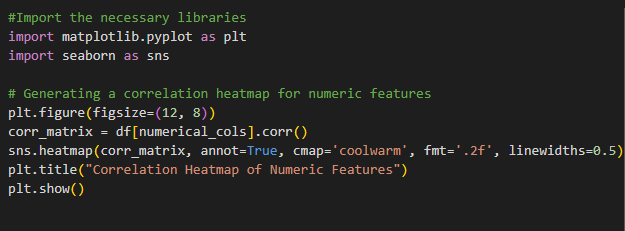






* **Generate a correlation heatmap for numeric features**

INPUT:



OUTPUT:



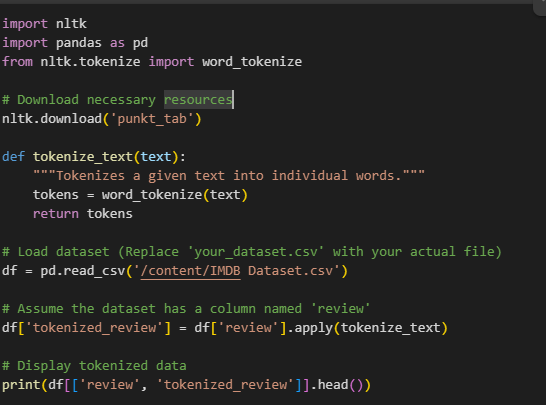
**Task No 2**

**Build a sentiment analysis model using a dataset such as IMDB Reviews**

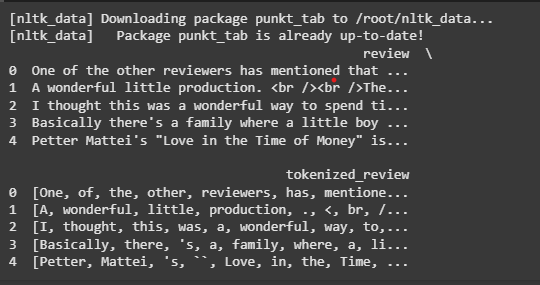
**Step 1:**

* Text Preprocessing: Tokenize Text into individual words.

INPUT:

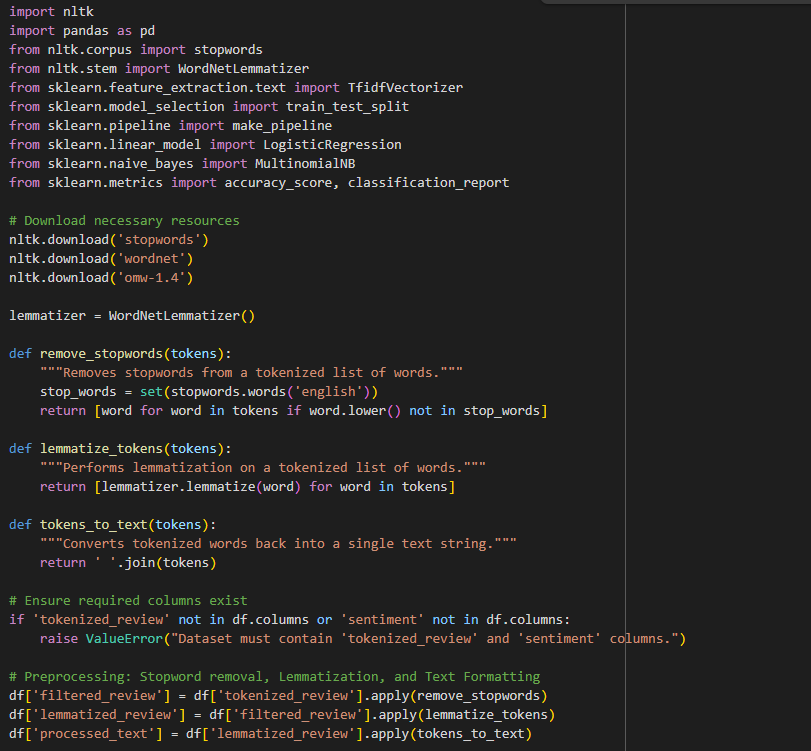


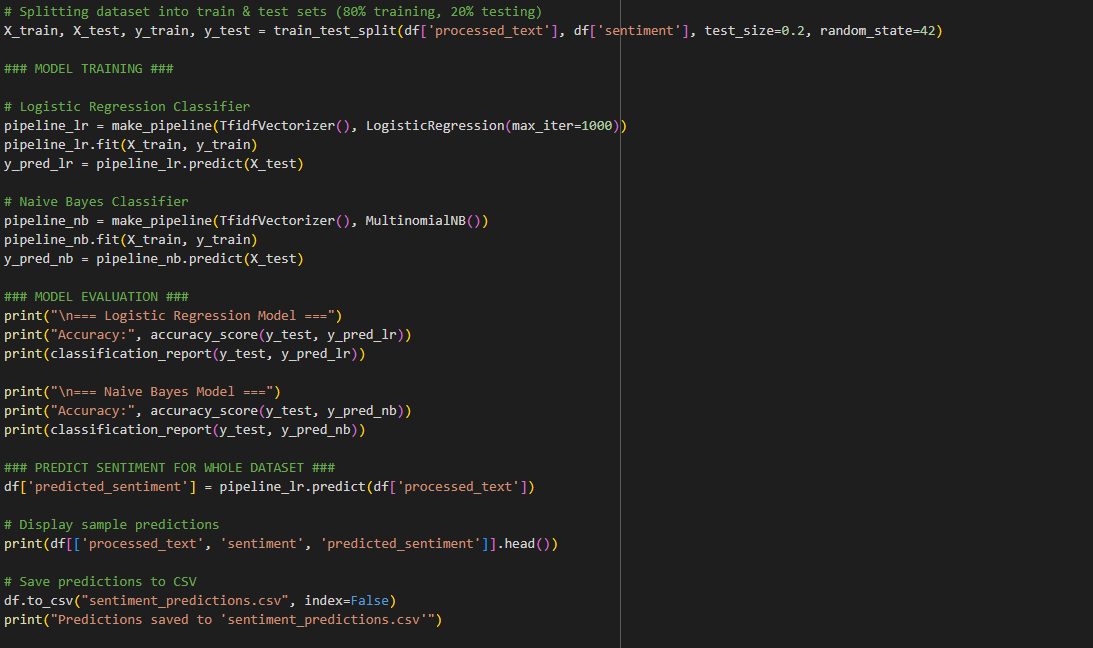
OUTPUT:



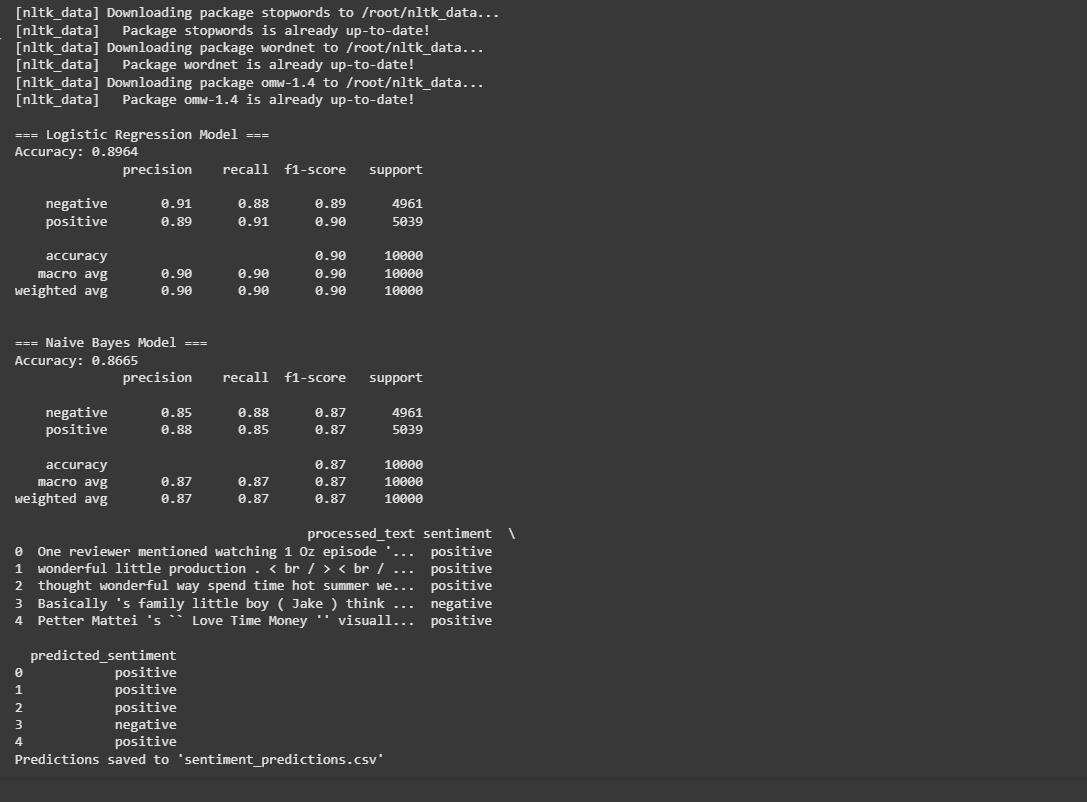
* Remove stopwords.
* Perform lemmatization for text normalization.
* Convert text data into numerical format using TF-IDF or word embeddings
* Train a classifier such as Logistic Regression or Naïve Bayes to predict sentiment
* Evaluate the model’s performance using metrices like precision, recall, and F1-score.

INPUT:





INPUT:



**Task No 3**

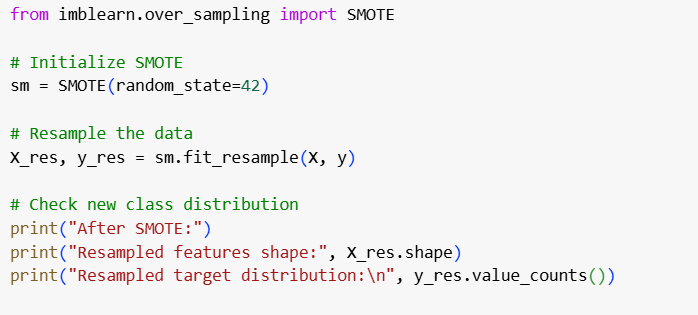
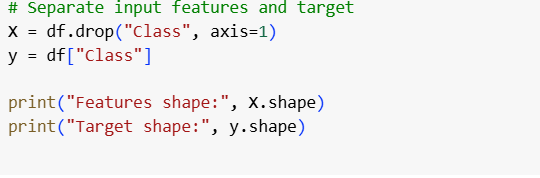
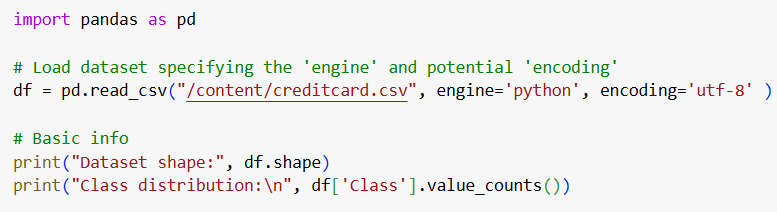
**Fraud Detection System**

Description:  
 Develop a fraud detection system using a dataset like the Credit Card Fraud Dataset.

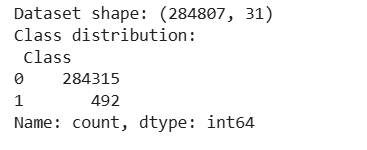
Steps:

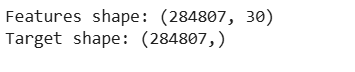
1. **Data Preprocessing:**
   * Handle imbalanced data using techniques like SMOTE or undersampling.

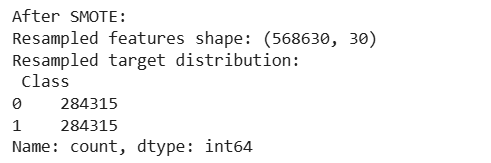
INPUT:

****

OUTPUT:



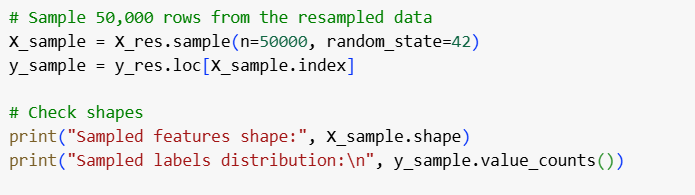


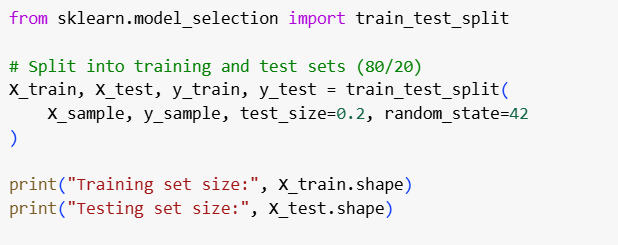


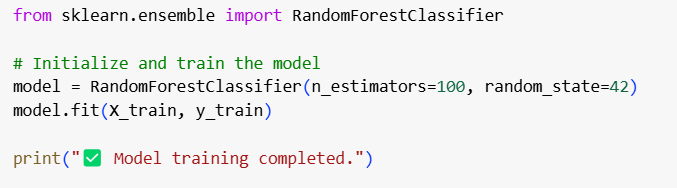
1. **Model Training:**

* Train a Random Forest or Gradient Boosting model to detect Fraudulent transactions.

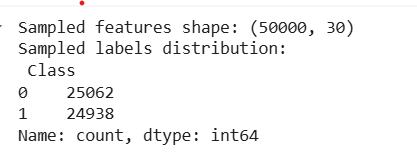
INPUT:

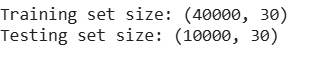






OUTPUT:



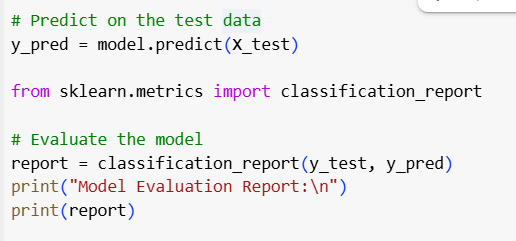


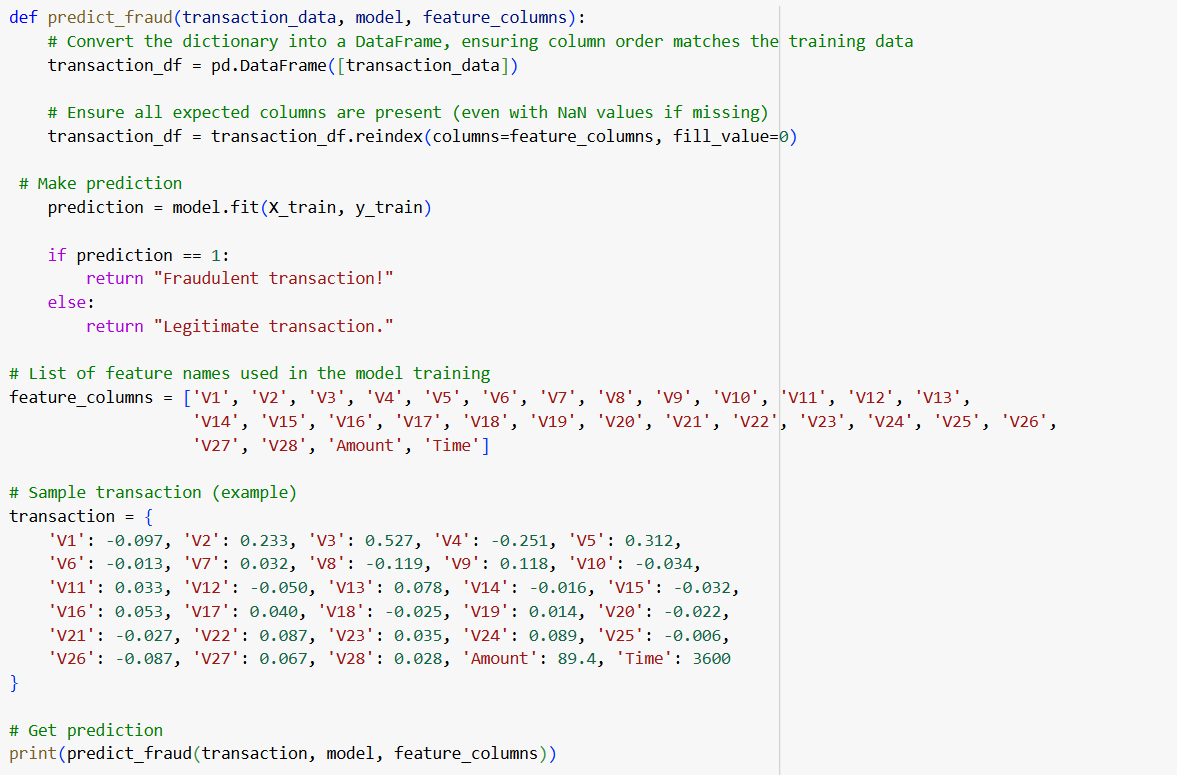


1. **Testing Interface:**

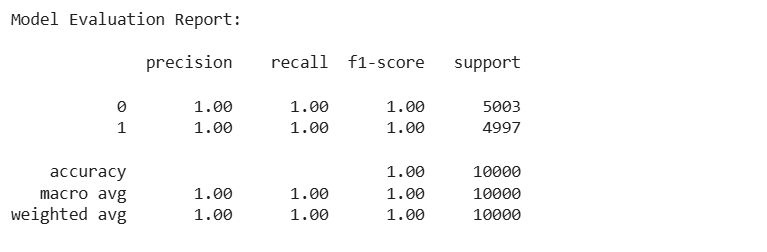
* Create a simple interface (e.g., a command-line input) to test the fraud detection system.

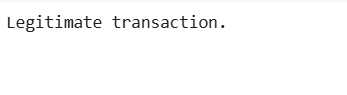
INPUT:





OUTPUT:



 **Task No 4**

**Predicting House Prices Using the Boston Housing Dataset**

Description:  
  Build a regression model from scratch to predict house prices using the Boston Housing Dataset.

Steps:

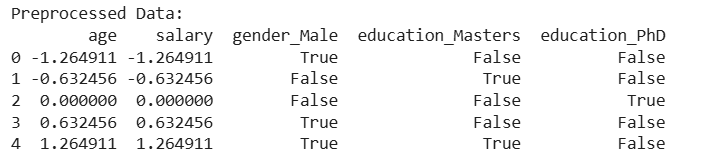
1. **Data Preprocessing:**

Normalize numerical features and preprocess categorical variables

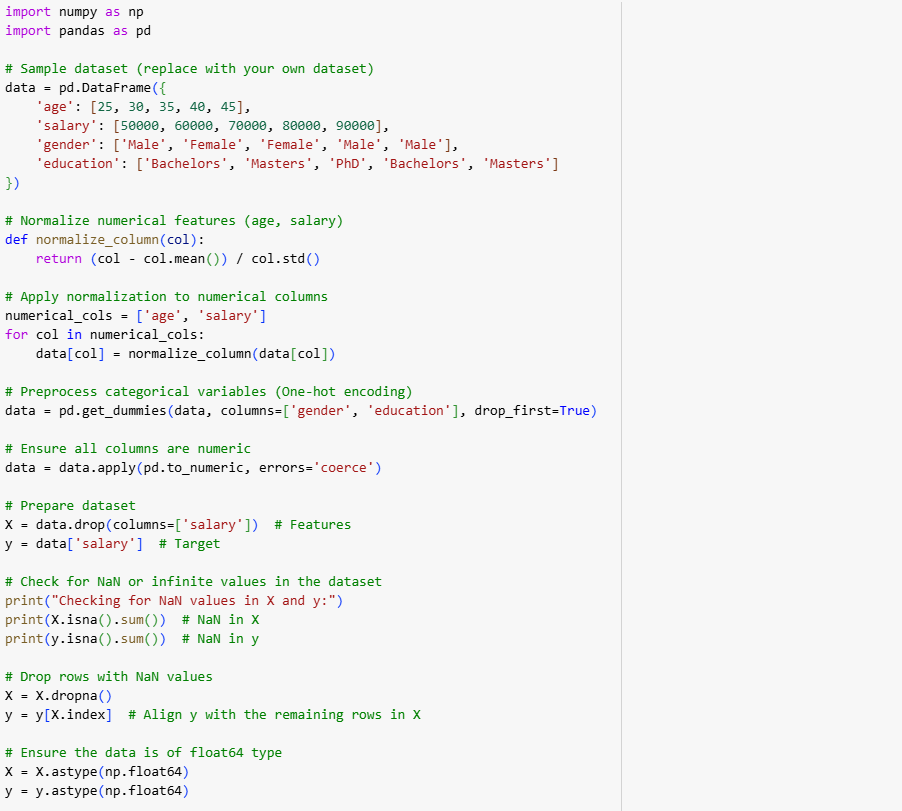
INPUT:

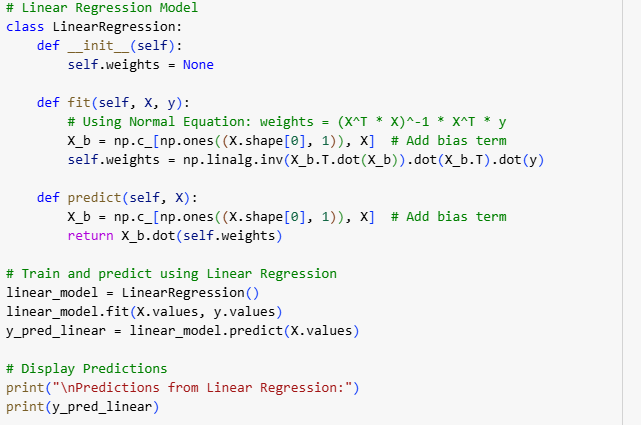


OUTPUT:

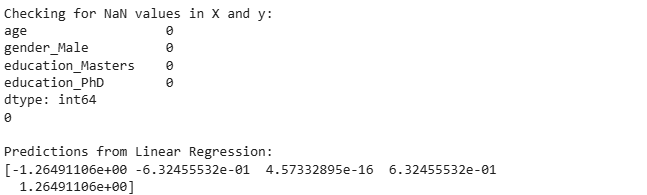


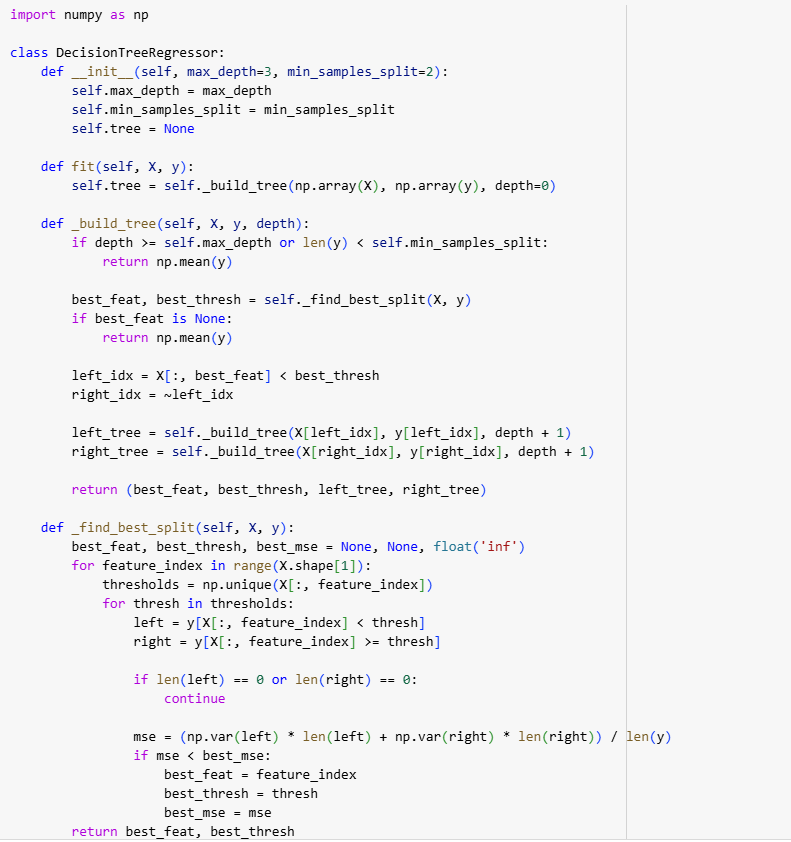
* **Model Implementation:**
* Implement Linear Regression, Random Forest, and XGBoost models from scratch (avoid using built-in libraries like sklearn.linear\_model).
* INPUT

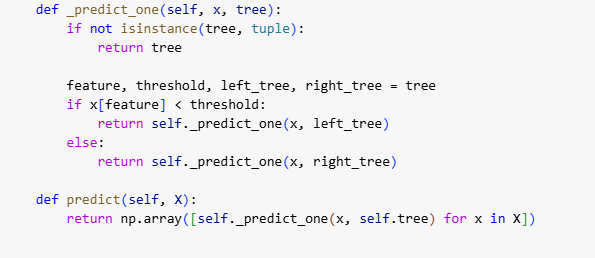


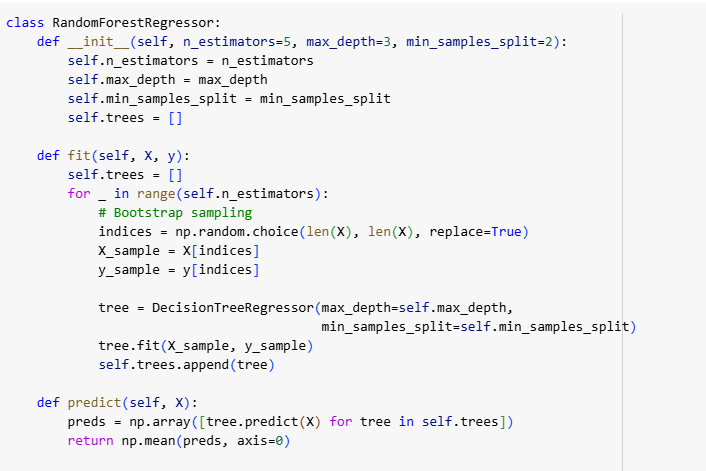


OUTPUT:

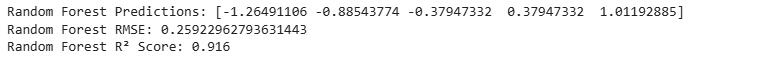






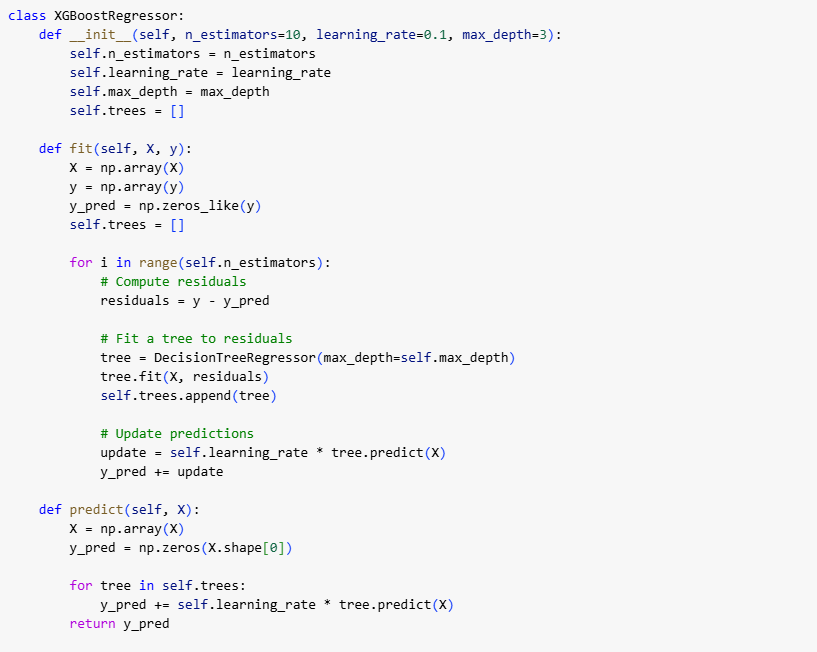


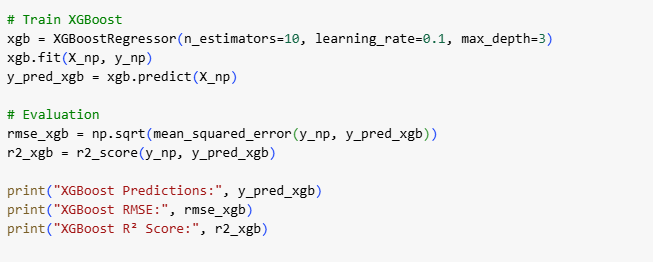
OUTPUT

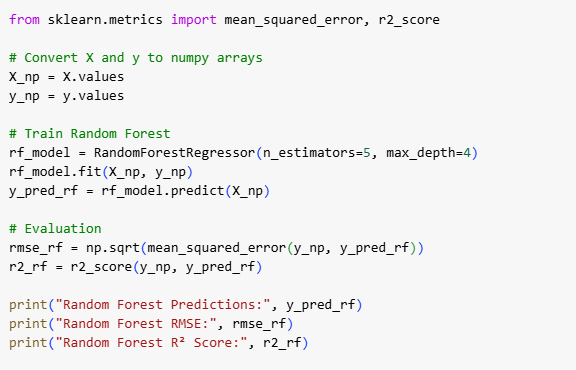


1. **Performance Comparison:**
   * Compare the models using RMSE and R² metrics

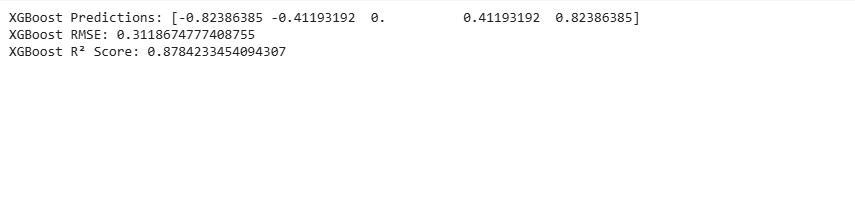
INPUT:







INPUT:



1. **Feature Importance:**
   * Visualize feature importance for tree-based models.

INPUT:



OUTPUT:

