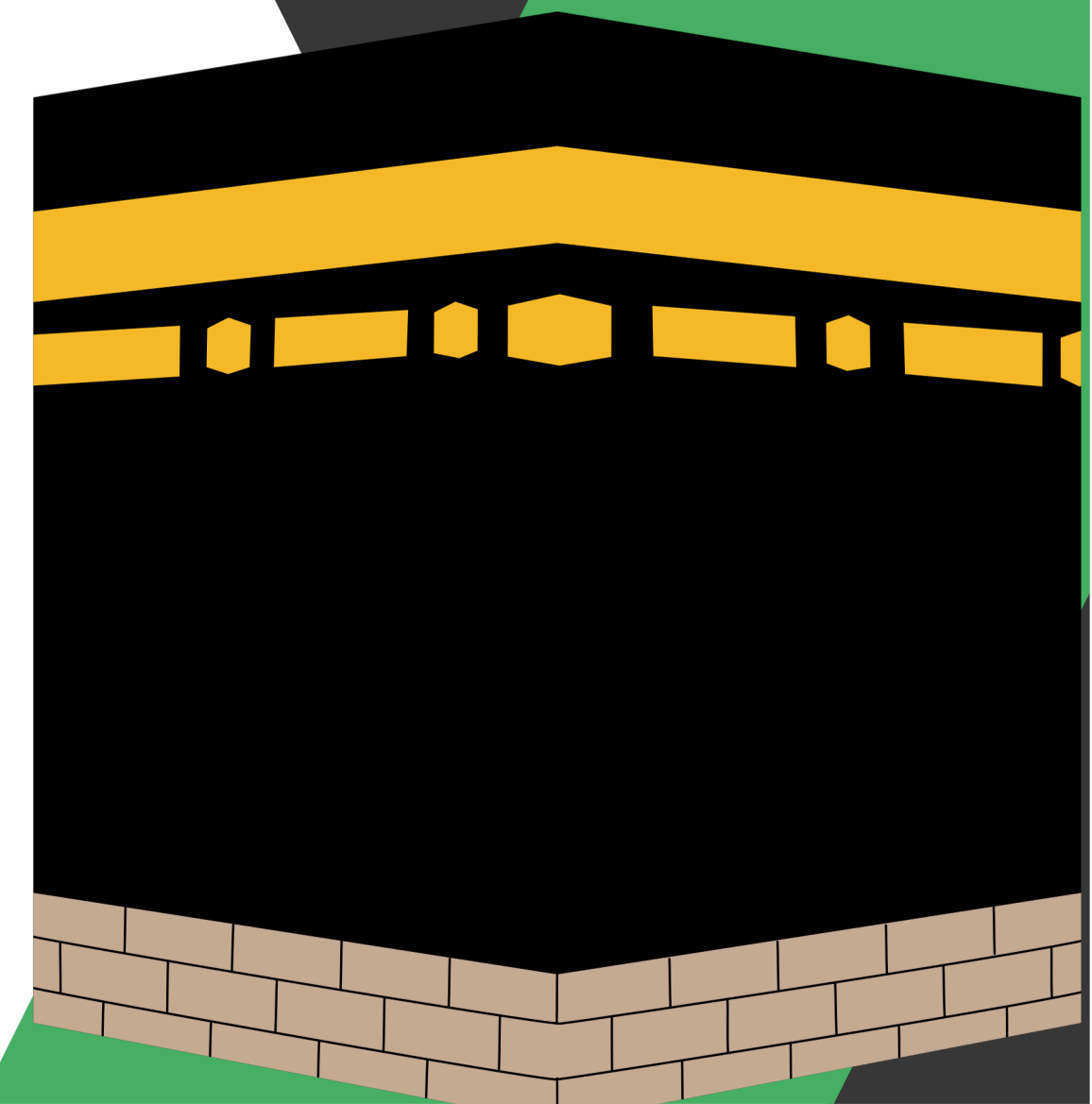# Predicting and Analyzing the Number of Umrah Pilgrims

Data Science and Analytics

Agenda

- Introduction
- Dataset
- Statistical Analysis
- Data preprocess
- Methodology
- Results
- Conclusion

Introduction

Umrah Overview
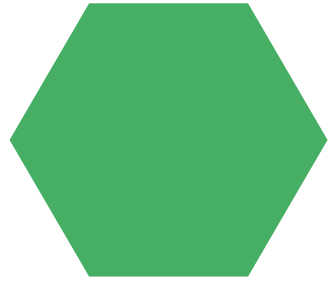- A continuous pilgrimage in Makkah, attracting millions of participants annually.

Technological Enhancements
- Use of AI for security and crowd management to improve response times and safety.

Future of Data Science
- Using Machine Learning to predict pilgrim numbers and optimize logistics.

Dataset

Obtained from

King Abdullah Petroleum Studies and Research Center (KAPSARC) Data Portal provided by the General Authority for Statistics
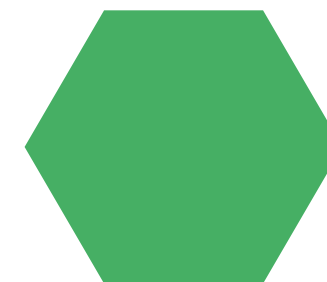
Number of Samples

266

Number of Features
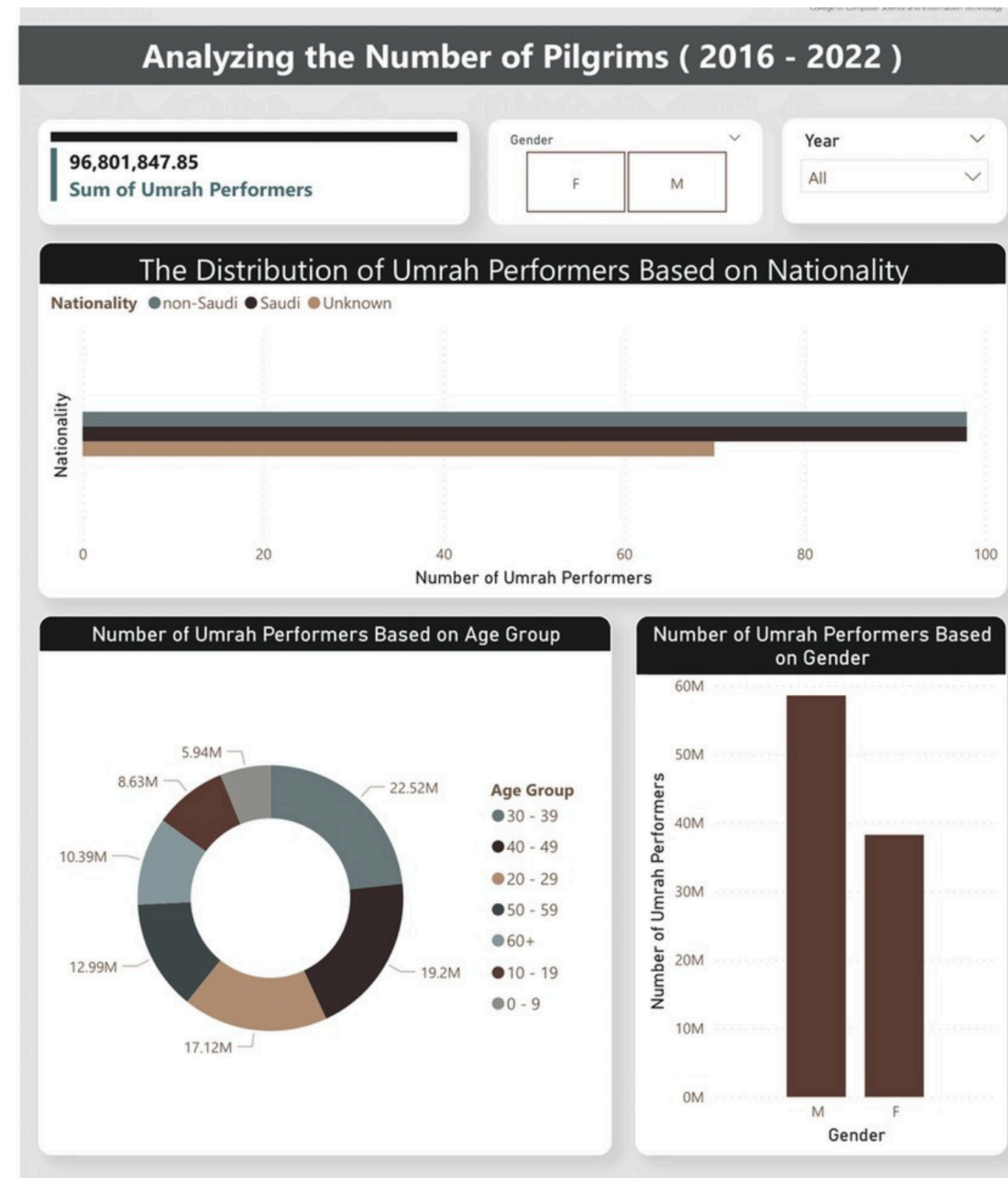
6
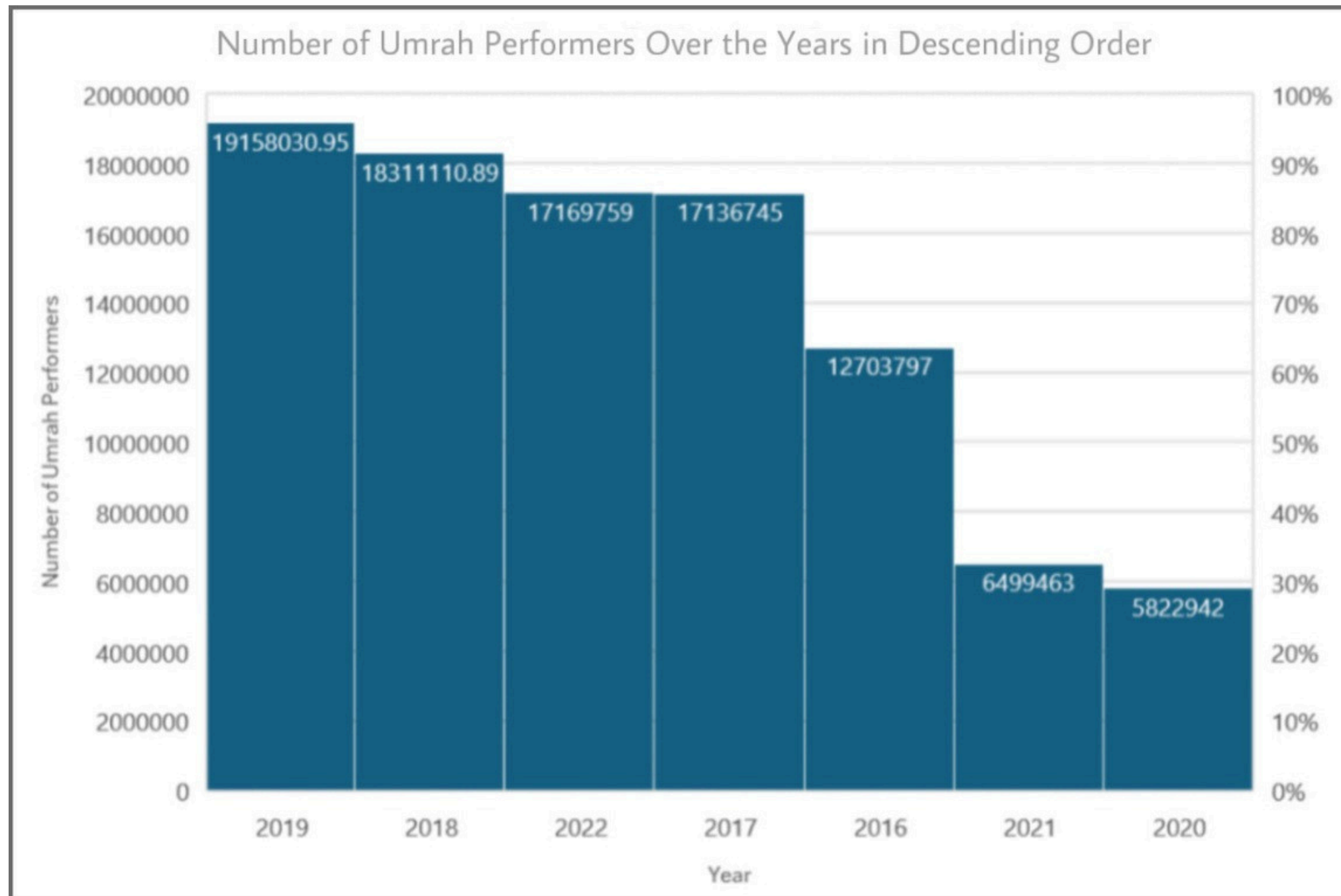
Dataset: Features

Gender

Nationality (Saudi/non-Saudi)

Year

Internal/external

Age Group

Number of Umrah Performers

Statistical Analysis

Statistical Analysis

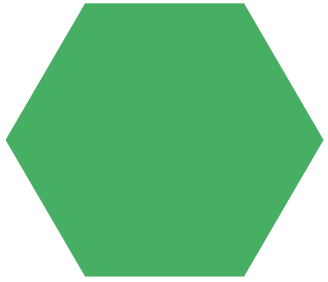Number of Umrah Performers Over the Years in Descending Order

Data preprocess

- Correct Spelling Inconsistency
- Remove Irrelevant Column
- Assess Missing Values
- Handle Missing Values
- List and Count Unique Values
- Transform Categorical Variables
- Rearrange Data Frame

Methodology

1. **Random Forest**

2. **Extreme Gradient Boosting**

3. **Gradient Boosting**
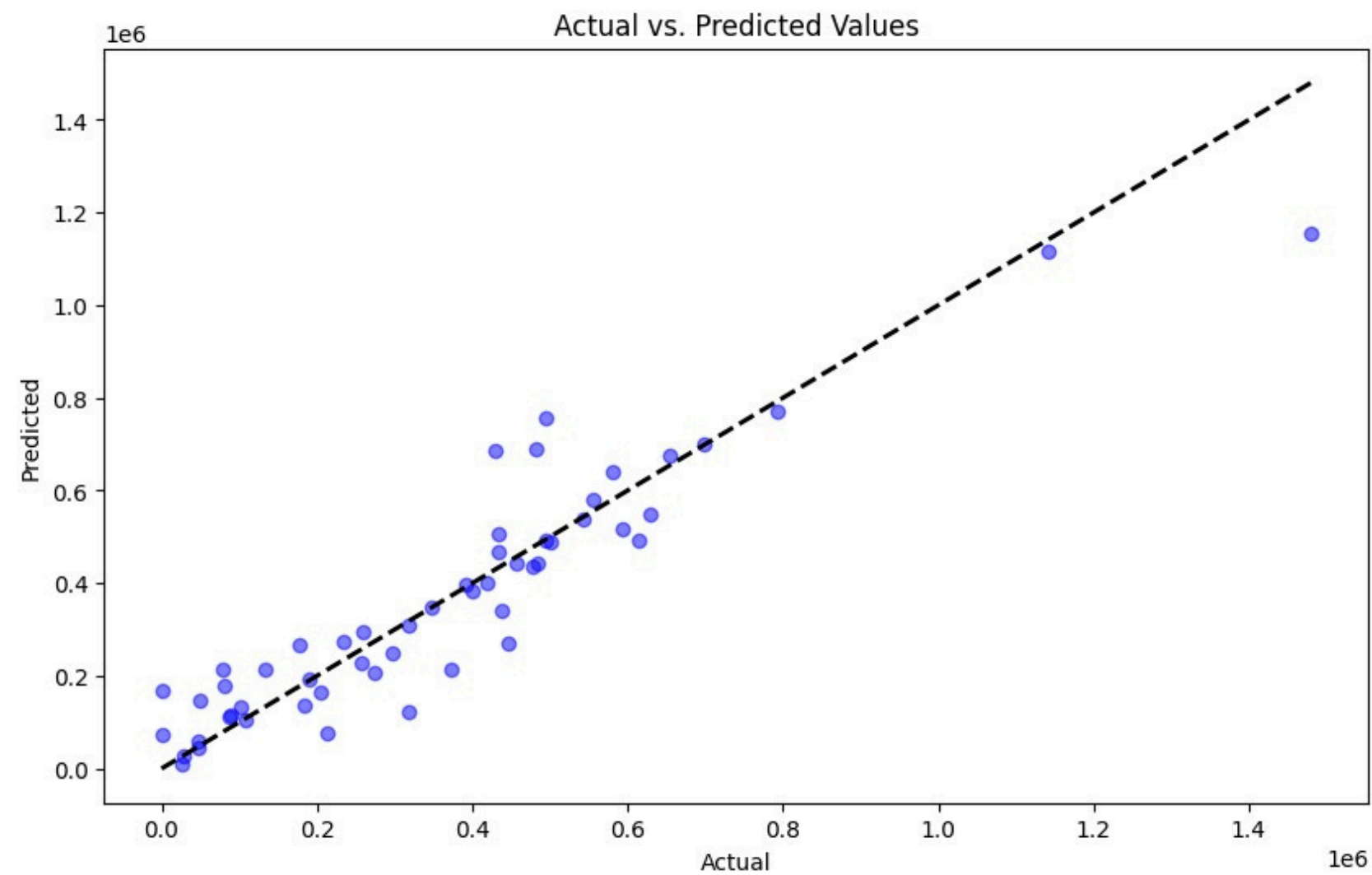
4. **Linear Regression**

5. **K Neighbors Regressor**

Results

Classifier	R-squared	Mean Squared Error (MSE)
XGBoost Regressor	0.8674	10,214,327,021.96
Gradient Boosting Regressor	0.8299	13,103,788,222.44
Random Forest Regressor	0.8227	10,954,178,629.40
K-Neighbors Regressor	0.7058	22,667,497,379.18
Linear Regression	0.3141	55,586,674,022.77

Results

XGBoosting

Actual vs. Predicted Values

Linear Regression

Actual vs. Predicted Values