# NEXT HIKES IT SOLUTION PROJECT 4

## IFTISUM RAHAMAN

# INTRODUCTION

- In this project, we have worked with a dataset that contains detailed information about various mobile phones, including their model,color,memory,RAM,battery capacity,rear camera,front camera,AI lens,mobile height,processor and most importantly price. Our primary goal is to develop a predictive model for mobile phone prizes.

- We will perform a feature extraction analysis to identify the most influential features.
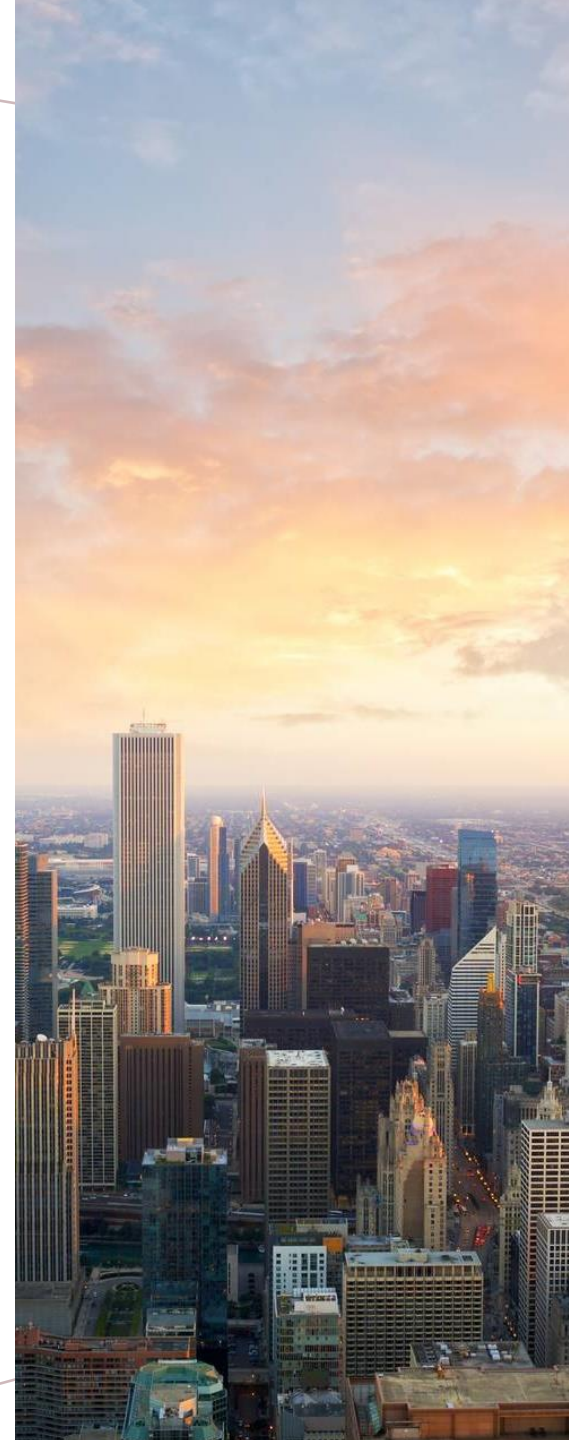
# OBJECTIVE

- OUR OBJECT IS TO BUILD A PREDICTIVE MODEL THAT CAN ACCURATELY ESTIMATE THE PRICE OF A MOBILE PHONE BASED ON ITS FEATURES.

- TO ACHIEVE THIS, WE WILL PERFORM A FEATURE EXTRACTION ANALYSIS TO IDENTIFY THE MOST INFLUENTIAL FEATURES.

- THE ORGANIZATION IS KEEN TO ENHANCE ITS PRICING STRATEGY BY GAINING A DEEPER UNDERSTANDING OF THE KEY FEATURES THAT INFLUENCE THE PRICES OF MOBILE PHONES IN TODAY'S HIGHLY COMPETITIVE MARKET.

# DATASET DESCRIPTION

- NUMBER OF ENTRIES: 541

- NUMBER OF FEATURES: 12

-  KEY FEATURES: MODEL, COLOUR, MEMORY, RAM, BATTERY, REAR CAMERA, FRONT CAMERA, AI LENS MOBILE HEIGHT, PROCESSOR, PRIZE.

# DATA PREPROCESSING

- DATA CLEANING: In this data cleaning we will find about the missing values, duplicate values. Solving all the complications which will effect the preprocessing of the data. Dropping the unnecessary column, finding the data type and all the information (shape, duplicate, describe, info, isnull) value.

- FEATURE ENCODING: One-hot encoding is applied to categorical features such as Model, Colour, Rear Camera, Front Camera and Processor to convert them into a format suitable for machine learning algorithms.

# EXPLORATORY DATA ANALYSIS

- DISTRIBUTION OF FEATURES: Analyzing the distribution of numerical features like Memory, RAM, Battery and Price to understand their range and central tendencies.

- CORRELATION ANALYSIS: Evaluate the correlation between different numerical features and the target variable to identify potential predictors.

- CATEGORY FREQUENCY: Analyzing the frequency of different categories in categorical features like Model, Colour, and Processor to understand their prevalence in the dataset.
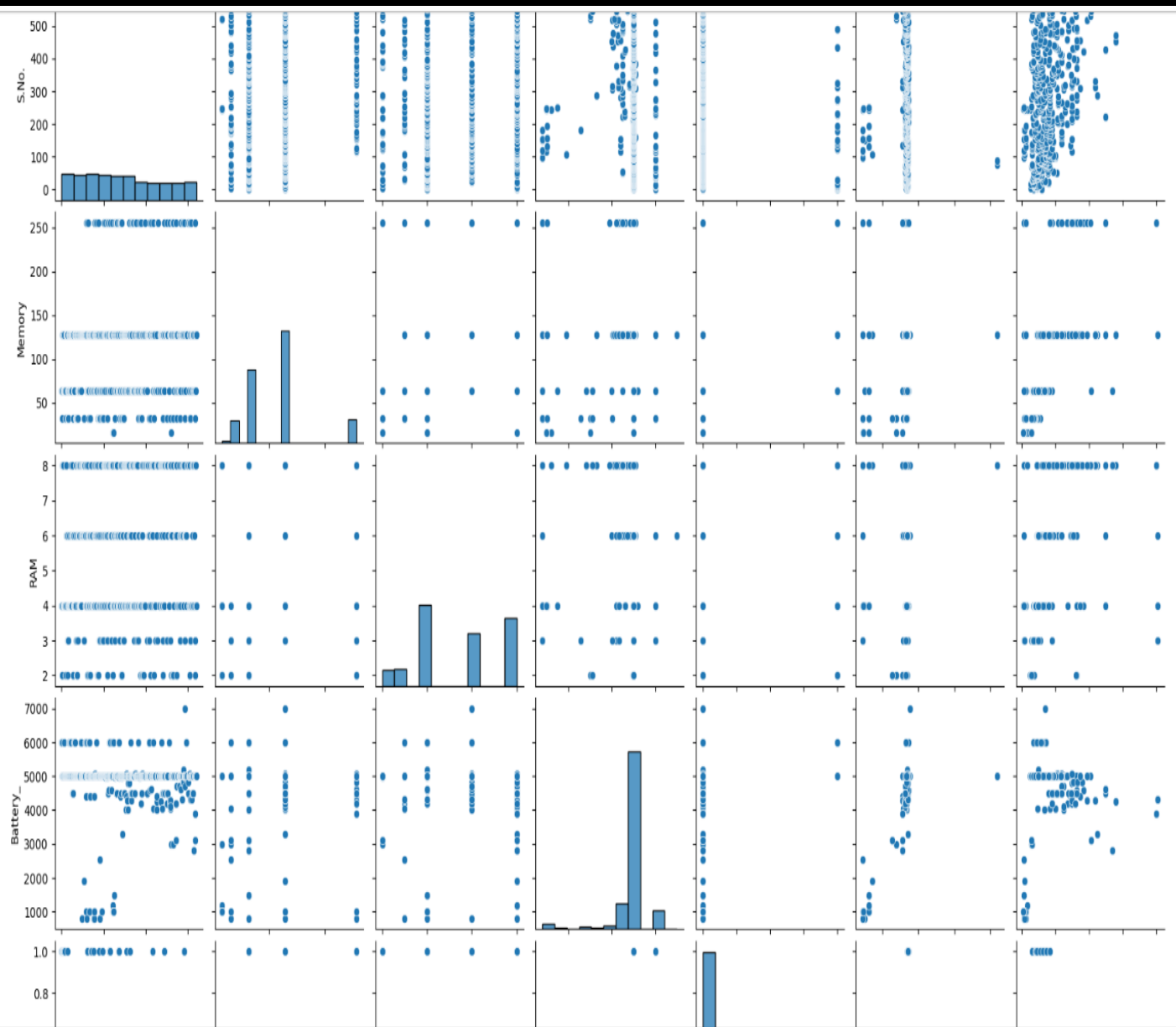
# VISUALIZATIONS
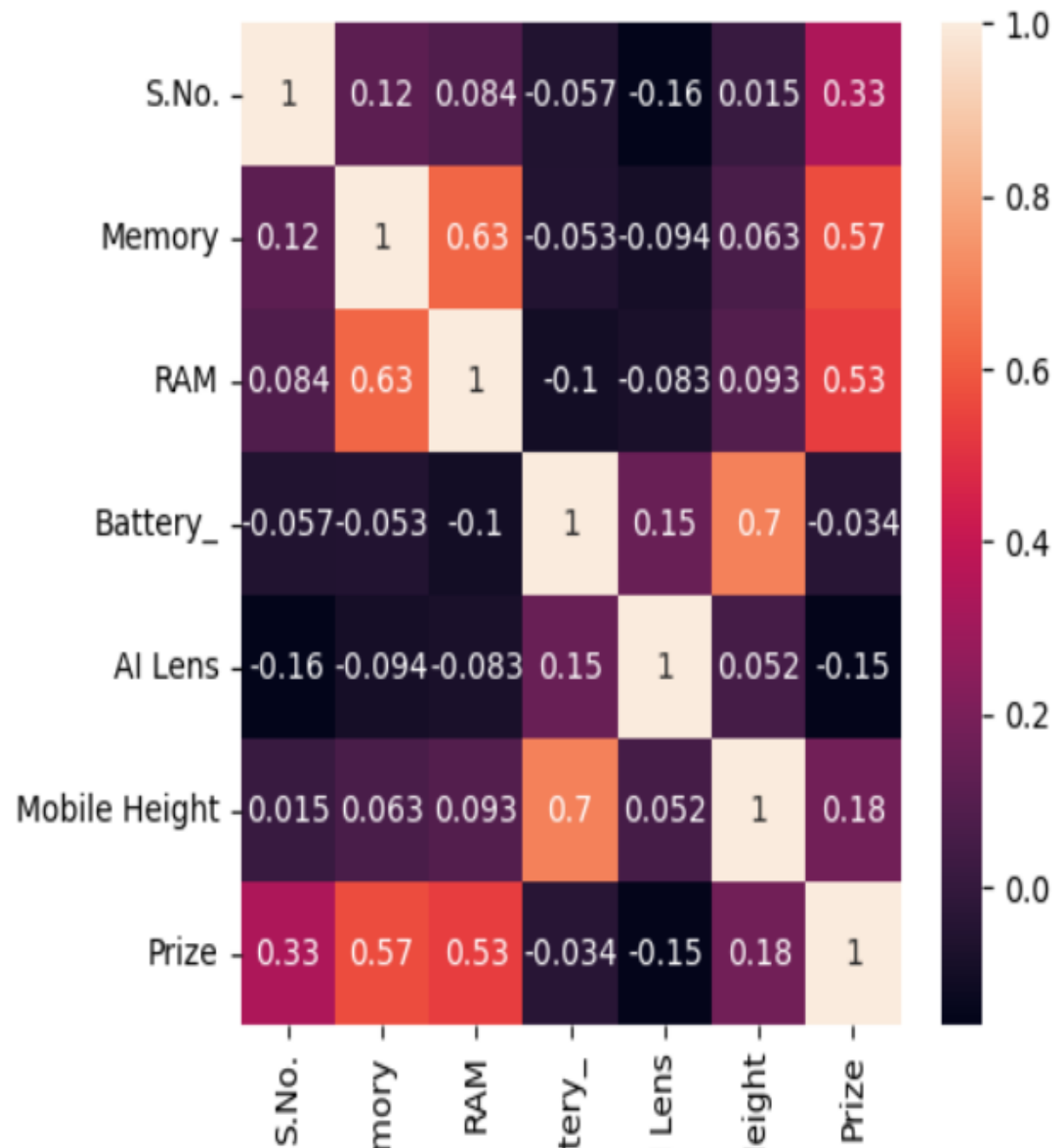
HEATMAP
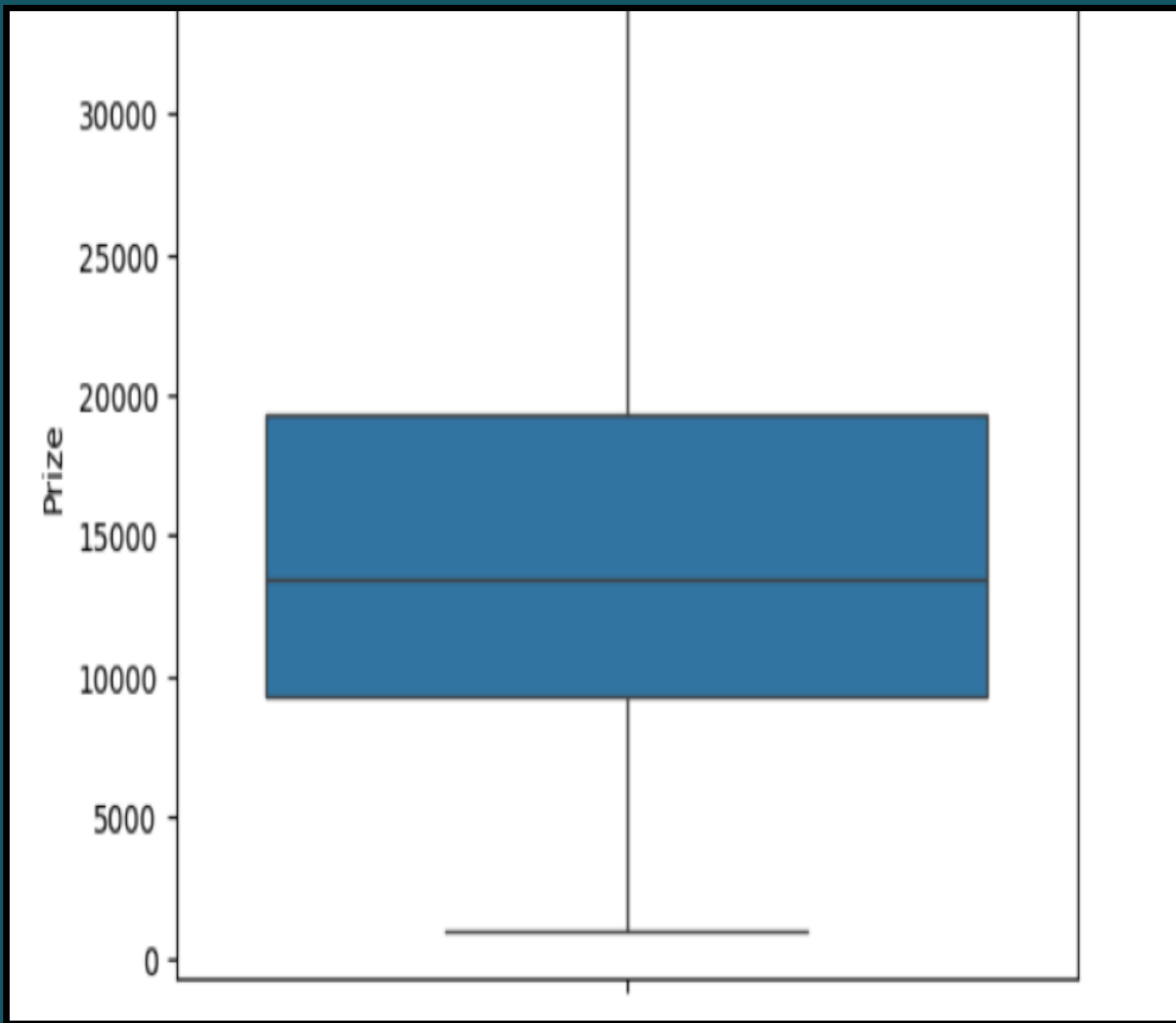
BOXPLOT

PAIRPLOT

# CONCLUSION

- Best Performers: The XGB Regressor and CatBoosting Regressor demonstrated the best performance on both the training and test sets, with R2 scores of 9.496581e-01 and 9.445088e-01 respectively.

- Ridge Regression also performed well with R2 score of 9.424866e-01 on the test set, making it a viable option.

- Random Forest Regression and Decision Tree Regression also showed strong performance with R2 score of 9.406551e-01 and 9.039107e-01 respectively, indicating their effectiveness for this task.

- Lasso Regression and Linear Regression did not perform well, indicating that they are not suitable for this dataset without significant modifications or improvements.

# THANK YOU

IFTISUM RAHAMAN