PARIS HOUSING DATASET

ABDULLAH AL MONAYEM
MD FOYSAL HOSSAIN

INTRODUCTION

"Paris Housing Classification with KNN:

This project explores the fascinating world of Paris housing classification using various algorithms to predict the **REAL ESTATE MARKET** analysis. By analyzing a comprehensive dataset featuring crucial housing attributes, such as location, size, and amenities, we unlock valuable insights to predict property classes. We showcase the potential of KNN in accurately categorizing housing properties, enabling informed decision-making for buyers and investors. Experience the transformative capabilities of machine learning and discover the secrets that lie within the enchanting Parisian real estate market.

WHAT HAVE WE DONE?

We preprocessed a data set and conducted exploratory data analysis such as graphs, plots, Decision Tree model, Linear Regression, K-nearest Neighbour, and K-means Clustering. We will add more analysis methods by learning the following classes.

PRIMARY GOAL

The Paris Housing Classification Dataset offers a wealth of information on housing properties in Paris, including location, size, amenities, and prices. The k-nearest neighbors (KNN) algorithm classifies these properties, revealing valuable patterns and trends. The real estate market and gain actionable insights.

PREPROCESSING AND DATA VISUALIZATION

IN Our Dataset there is no MISSING Values or data.

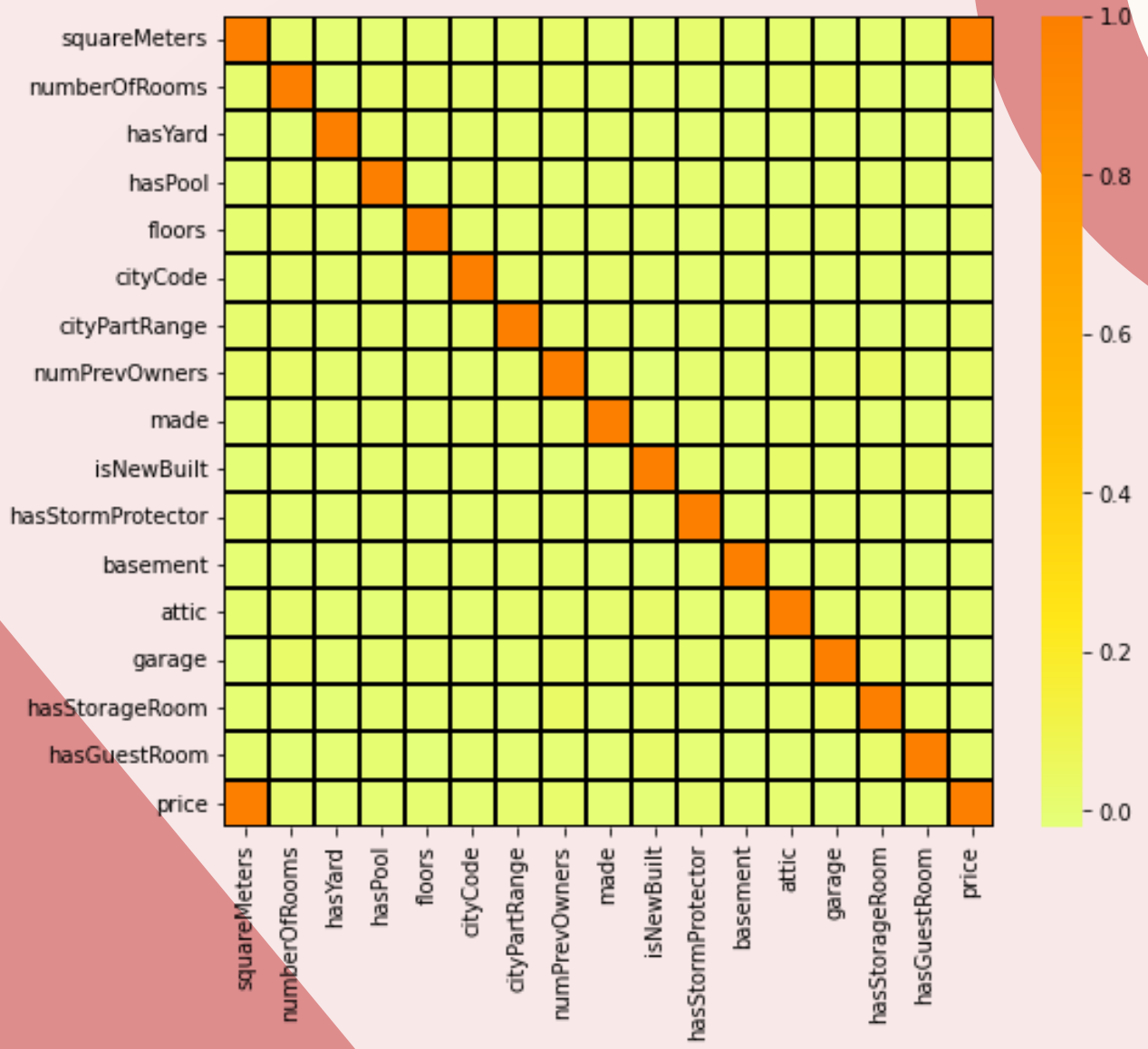
Checking the missing data

```
df.isnull().sum().sort_values(ascending=False)
```

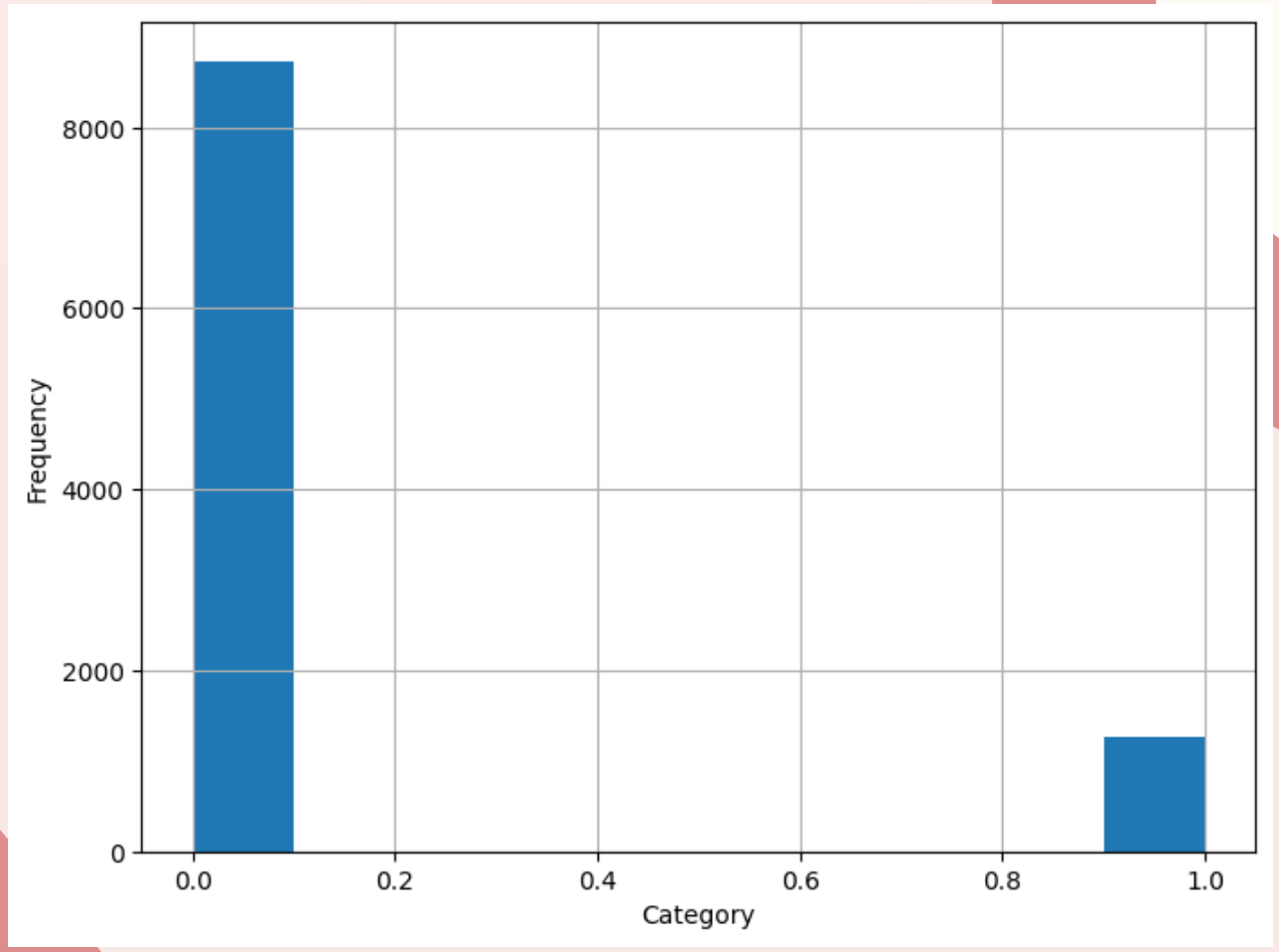
0s

```
squareMeters        0
numberOfRooms       0
price               0
hasGuestRoom        0
hasStorageRoom      0
garage              0
attic               0
basement            0
hasStormProtector   0
isNewBuilt          0
made                0
numPrevOwners       0
cityPartRange       0
cityCode            0
floors              0
hasPool             0
hasYard             0
category            0
dtype: int64
```

CHECKING THE CORRELATION

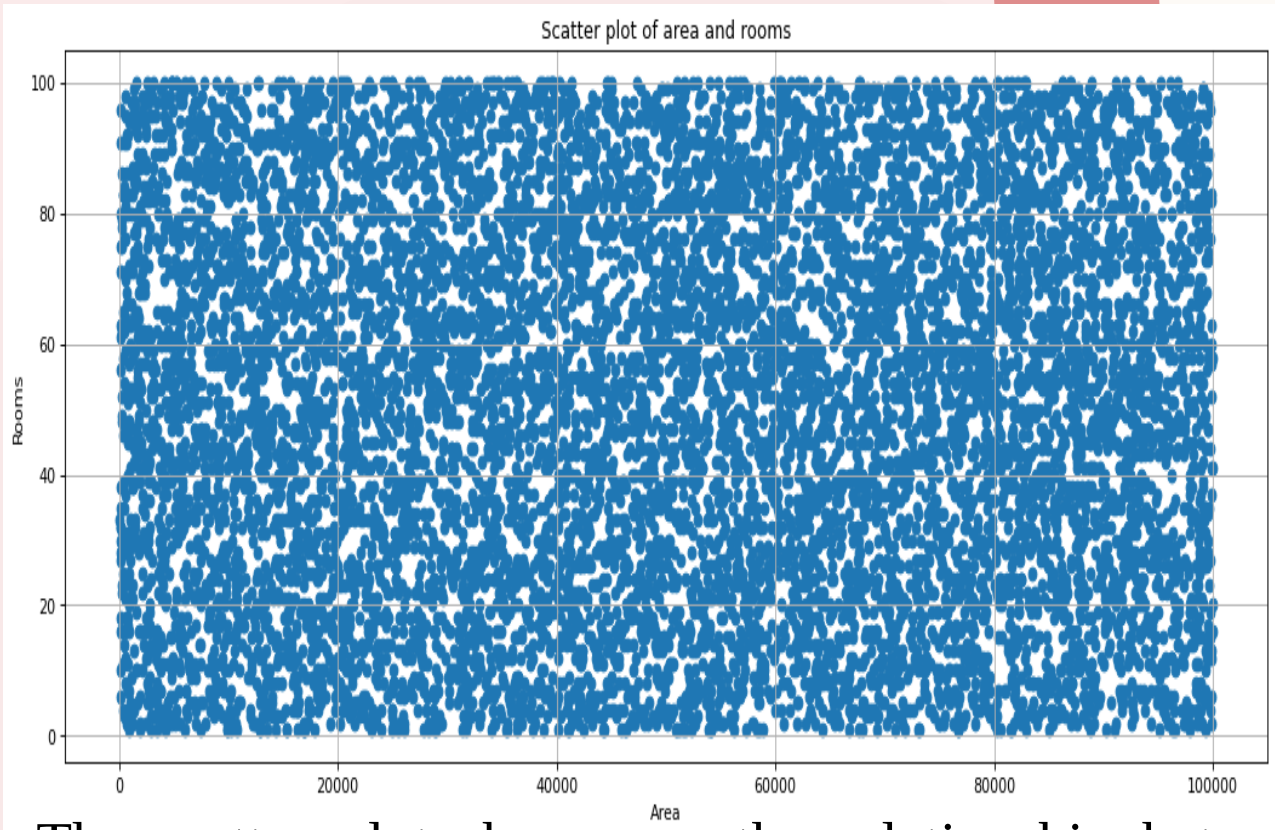


HISTOGRAM



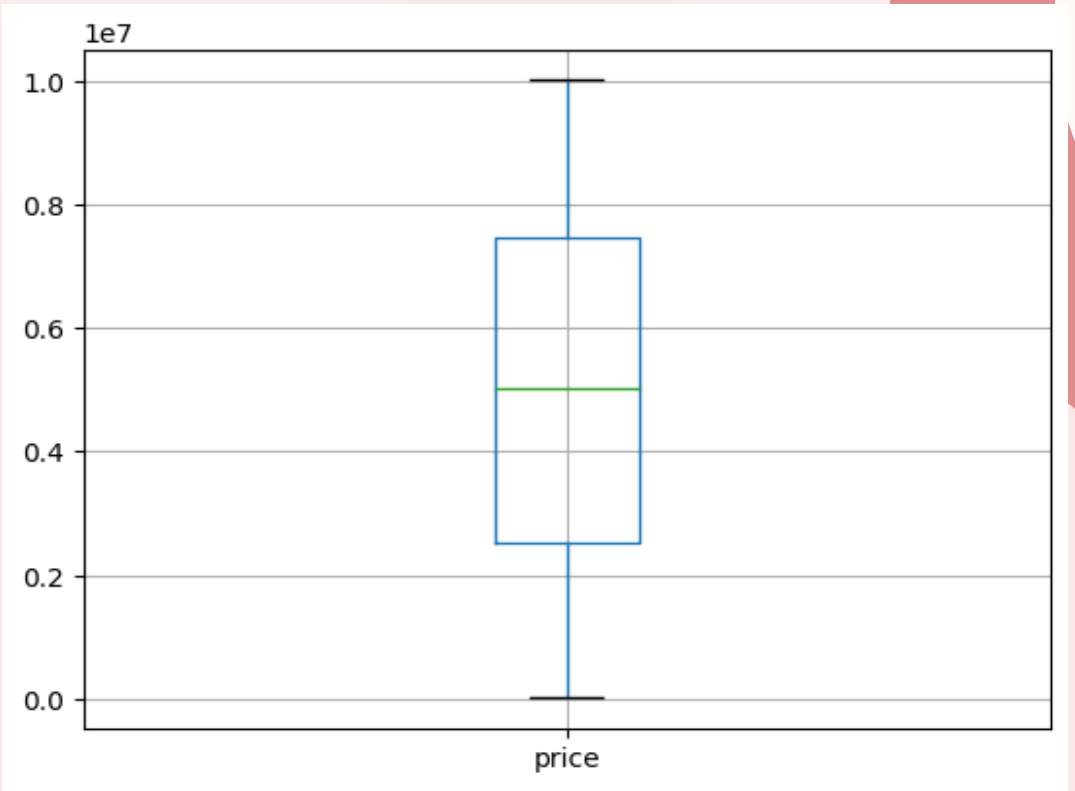
The DataFrame 'df' and display it with the 'Category' label on the x-axis and the 'Frequency' label on the y-axis.

SCATTER PLOT



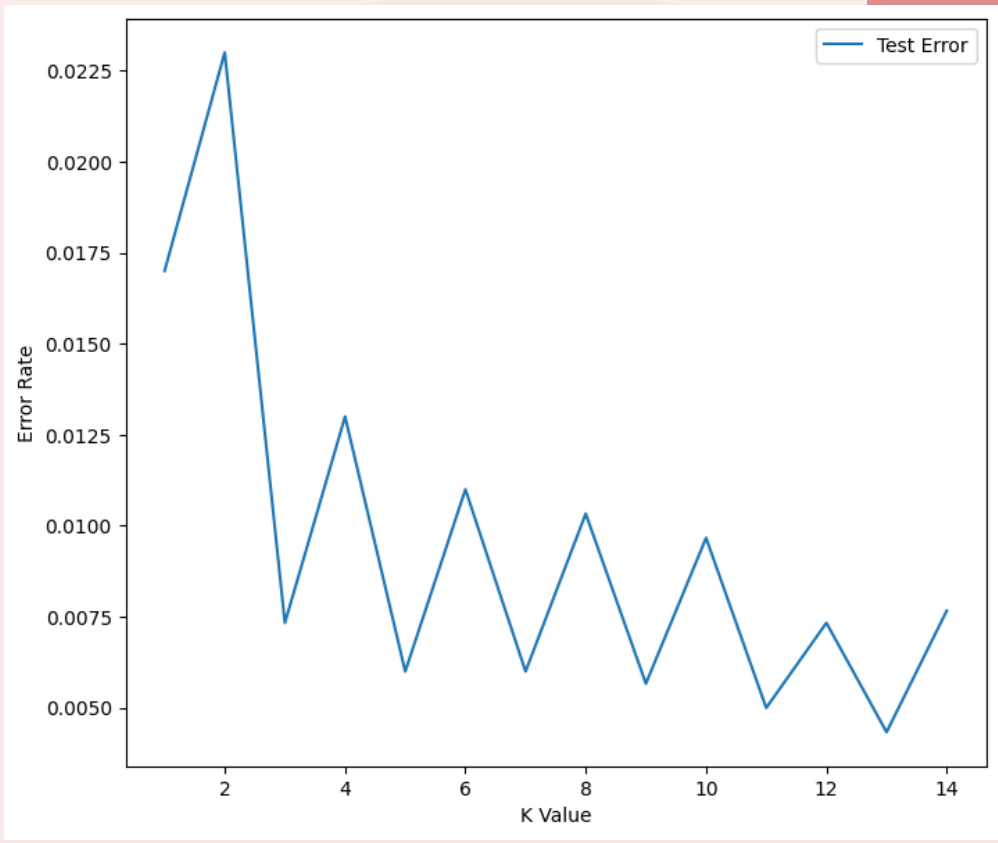
The scatter plot showcases the relationship between the number of rooms and the area (square meters) of housing properties in the Paris housing classification dataset. Each data point represents a unique property, with the x-axis displaying the area and the y-axis representing the number of rooms.

BOX PLOT



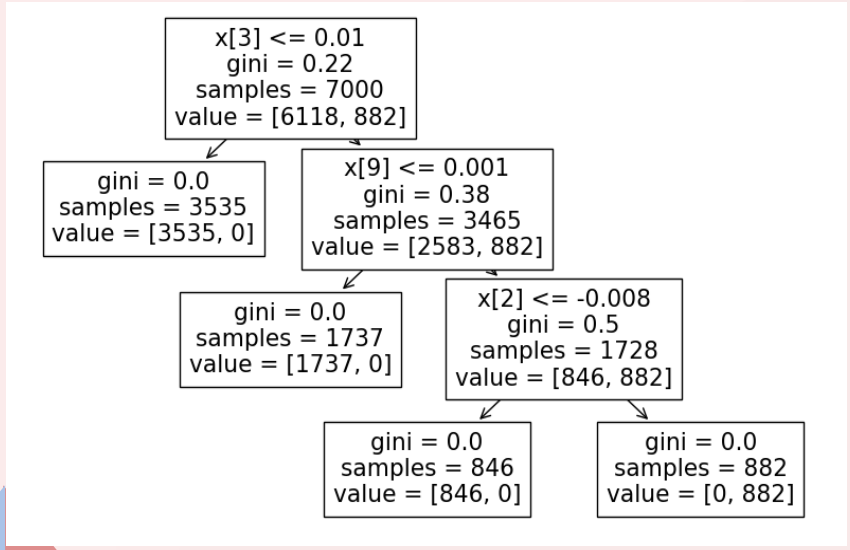
This box plot is utilized for anomaly detection in the 'price' variable of the dataset. It helps identify outliers or anomalies within the distribution of housing prices. The box plot offers a quick assessment of unusual price observations by displaying the central tendencies and potential extreme values. This aids in detecting anomalies and provides valuable insights for data analysis in the context of the Paris housing market. Though we don’t have outliers in this.

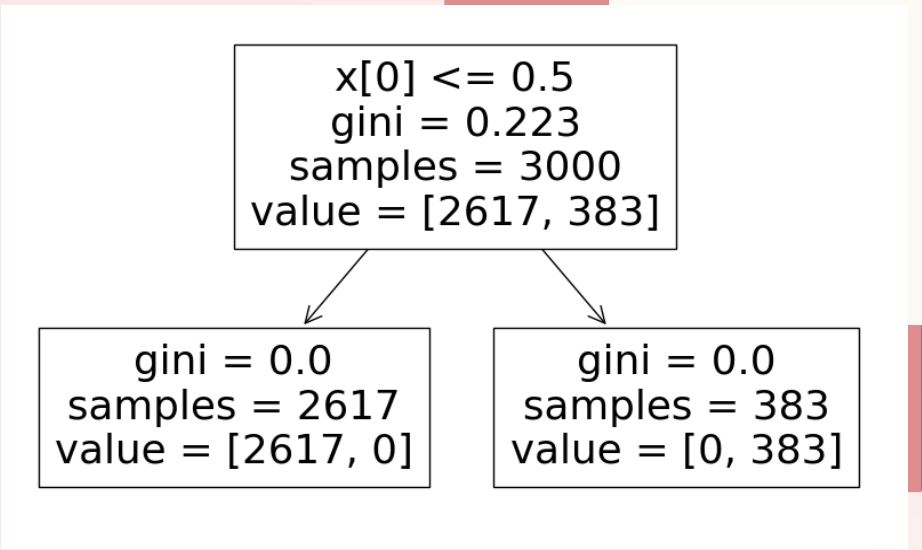
ELBOW FIGURE



The elbow figure depicts the relationship between the error rate and the value of K in the K-nearest neighbors (KNN) algorithm. By varying the value of K and calculating the error rate, we can determine the optimal number of neighbors to consider for classification.

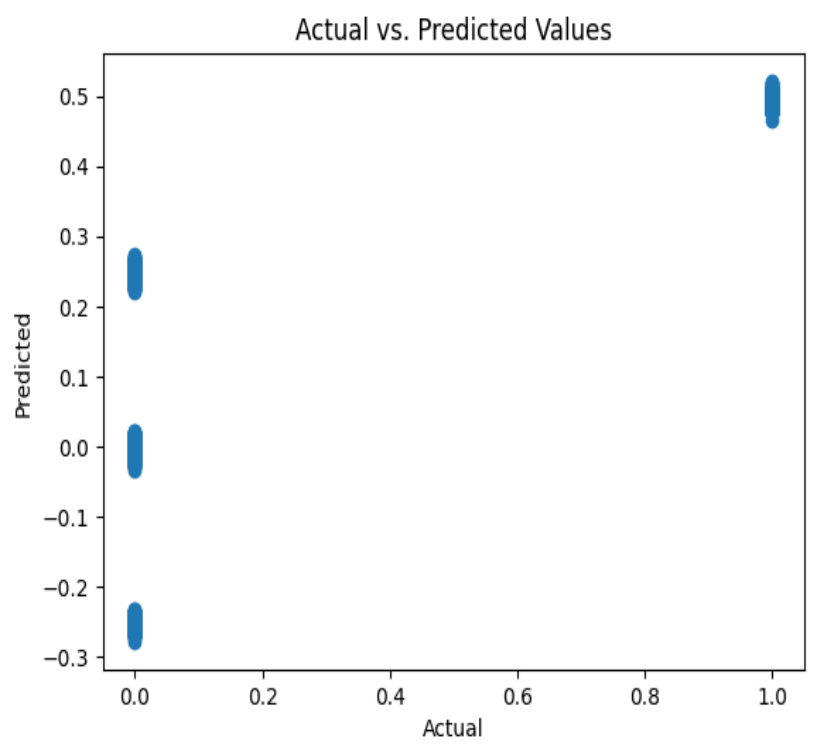
DECISION TREE

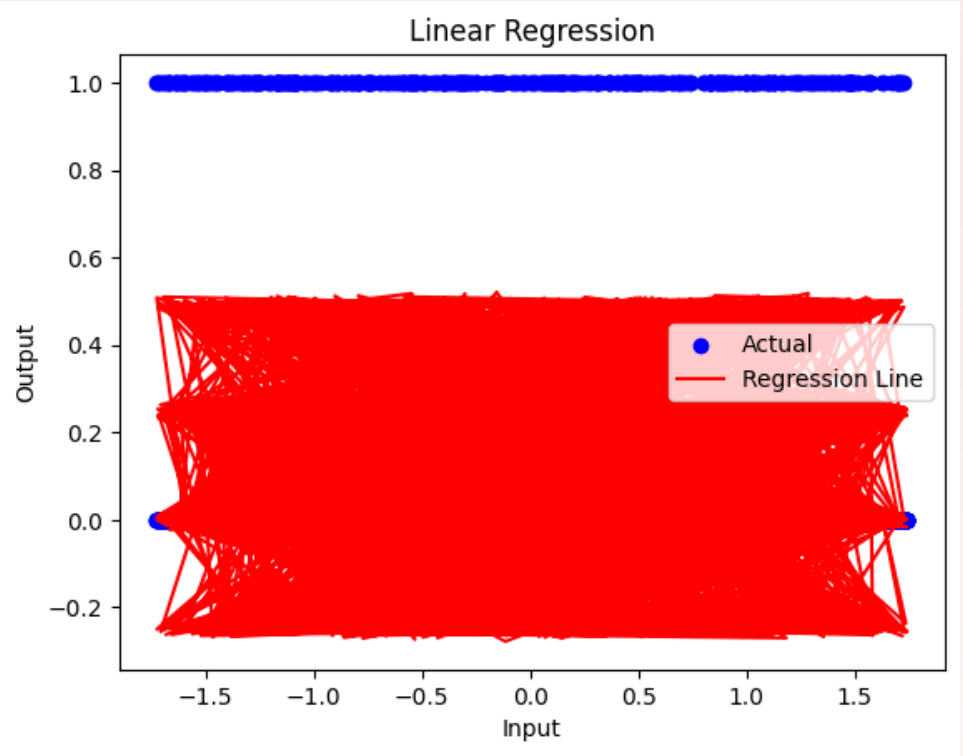




After applying the decision tree algorithm (GINI) we got an accuracy of 1

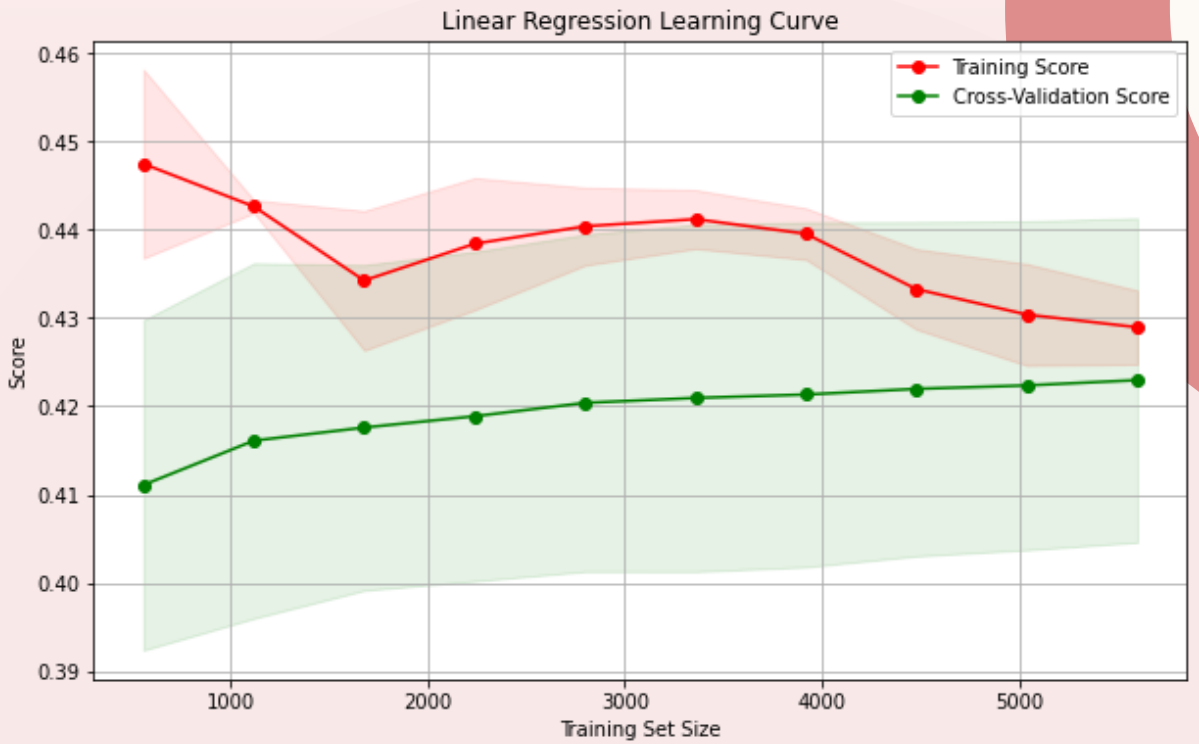
LINEAR REGRESSION





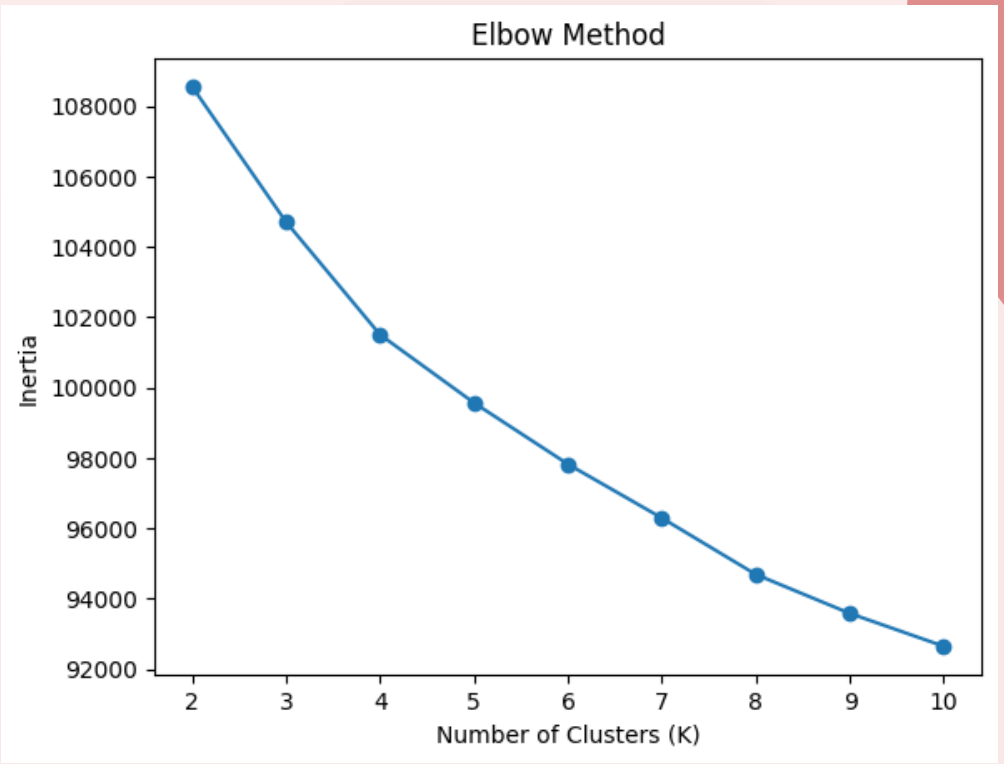
slope value of linear regression Slope (m): 2.6946138286272086 Intercept (c): 0.1259999999999964

LINEAR REGRESSION (CONT.)



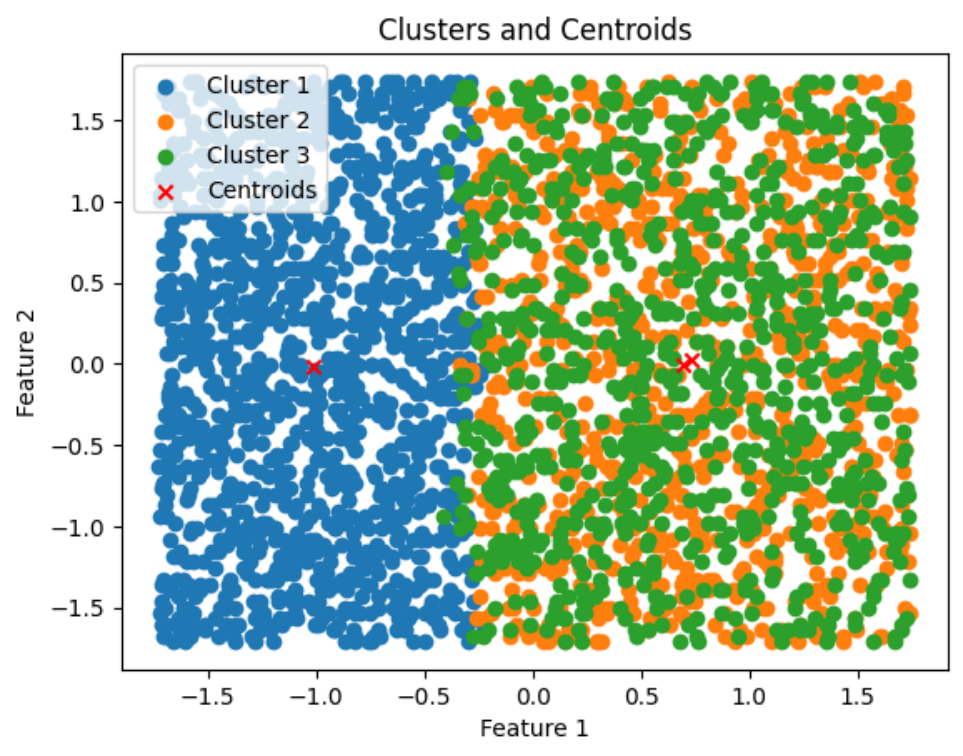
The learning curve of a linear regression model trained on the Paris housing dataset typically shows a decreasing trend, indicating the model's improvement with more training examples. As the number of training examples increases, the model's performance generally improves, resulting in a decrease in both training and validation errors.

ELBOW METHOD



Here in this Elbow method we found as best cluster number is 5.

CLUSTERING



Here we take Two features in cluster.

THANK YOU