Smartphone Price Prediction using ML

**Abstract-**

Whenever we plan to purchase any mobile phone ,we consider many factors like Camera, Battery life, Memory etc. But the most important factor we should consider is whether the mobile phone we are going to purchase is worth the cost or not . As we don’t have any resources to cross check the price, we tend to take wrong decision sometimes

Through this project, we will be able to solve this problem by taking the historical data pertaining to the key features of smartphones along with its cost and develop a model that will predict the approximate price of the new smartphone with a reasonable accuracy

**Requirements-**

1)Jupyter Notebook

2)Numpy for Linear Algebra

3)Sklearn for Statistical modelling and Machine Learning

4) Pandas for Data Processing

5) Matplotlib for Data visualization

6) Seaborn for Data visualization that are more statistically sophisticated

**Approach-**

**1)INTRODUCTION:**

The Dataset required for the project has been taken from Kaggle

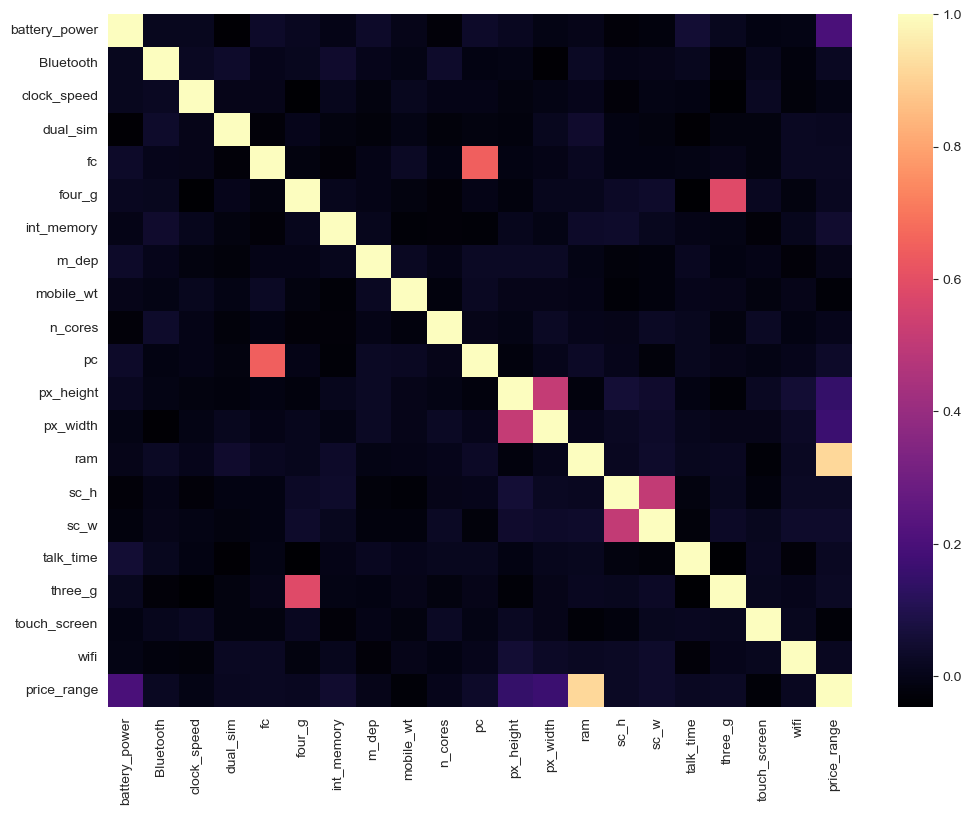
The data contains information regarding features, specification, and price range of 2000 Smartphones. In this Project, I am going to explore and analyse a dataset and try to predict optimum price ranges for a list of Smartphones in the Market

**2)Data Analysis**

CORRELATION MATRIX-

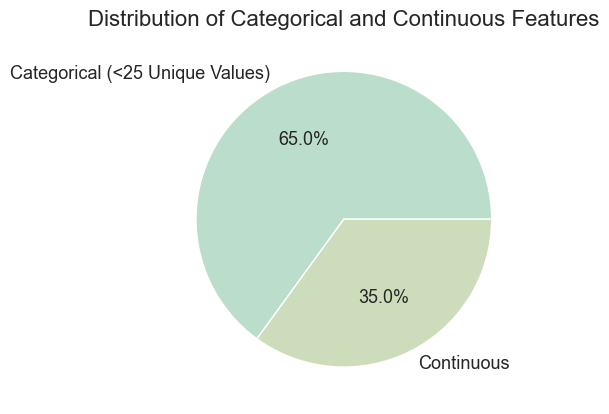
plt.figure(figsize=(12,9))

sns.heatmap(df.corr(),cmap='magma')



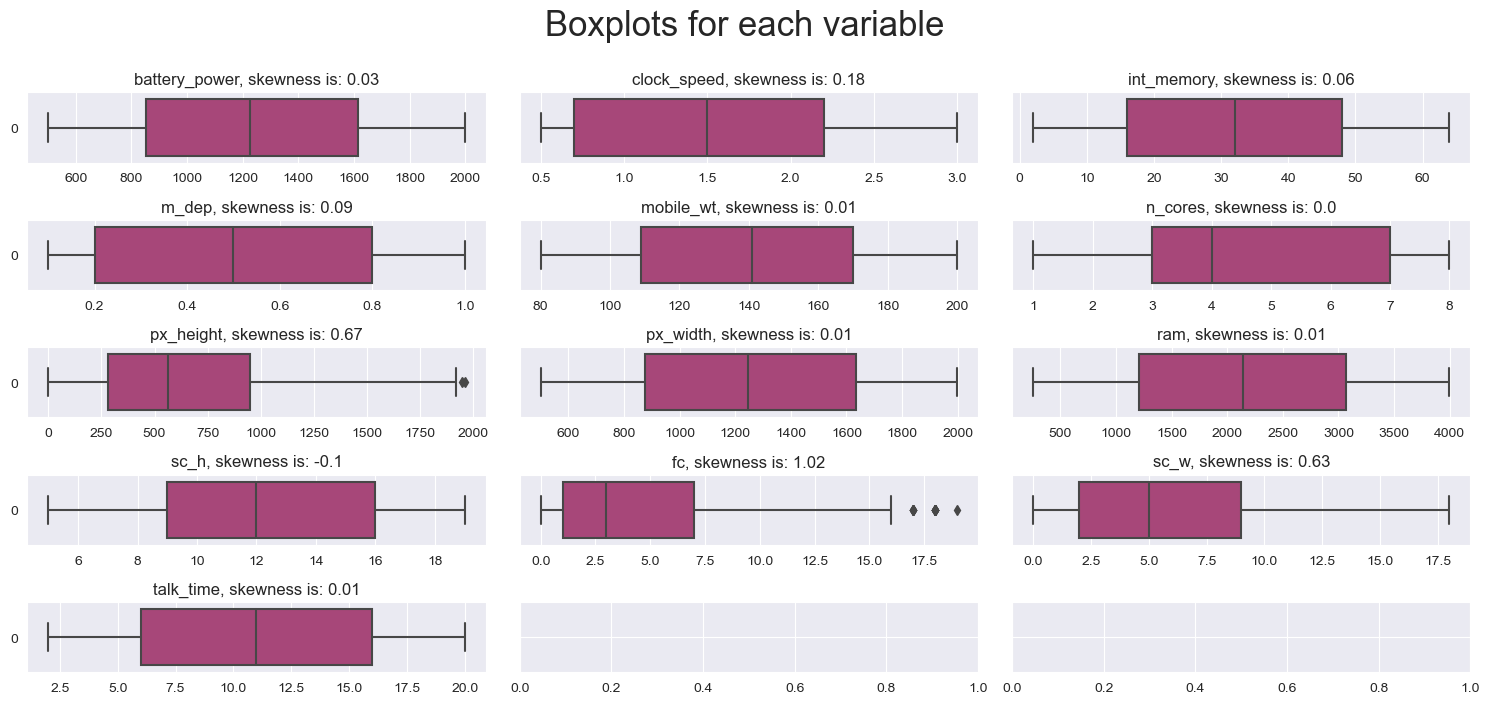
The Matrix shows that ram and price range are highly correlated

FEATURE DISTRIBUTION-



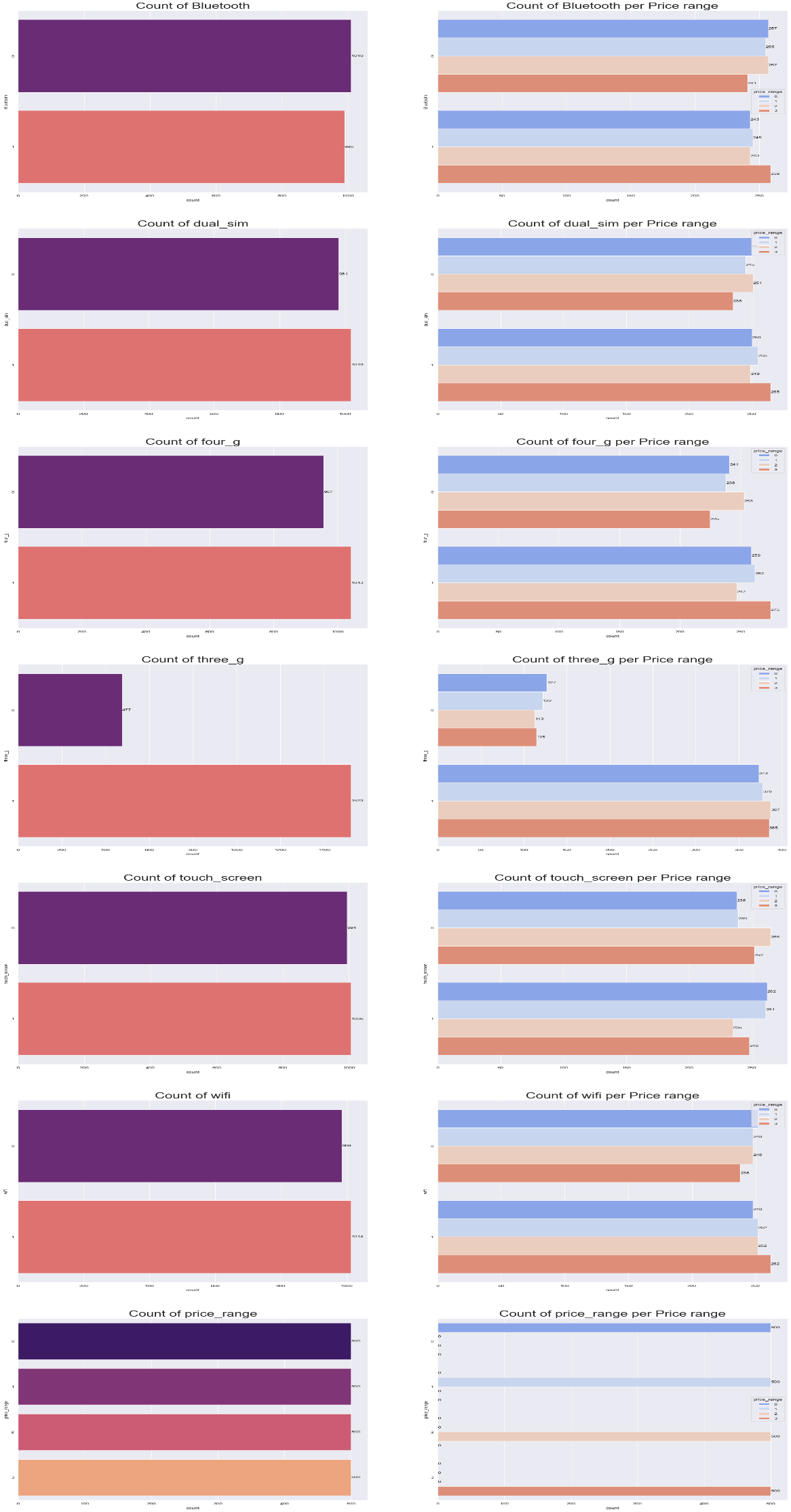
The pie chart shows that 65% contributes to Categorical Data whereas 35% contributes to Continuous Data

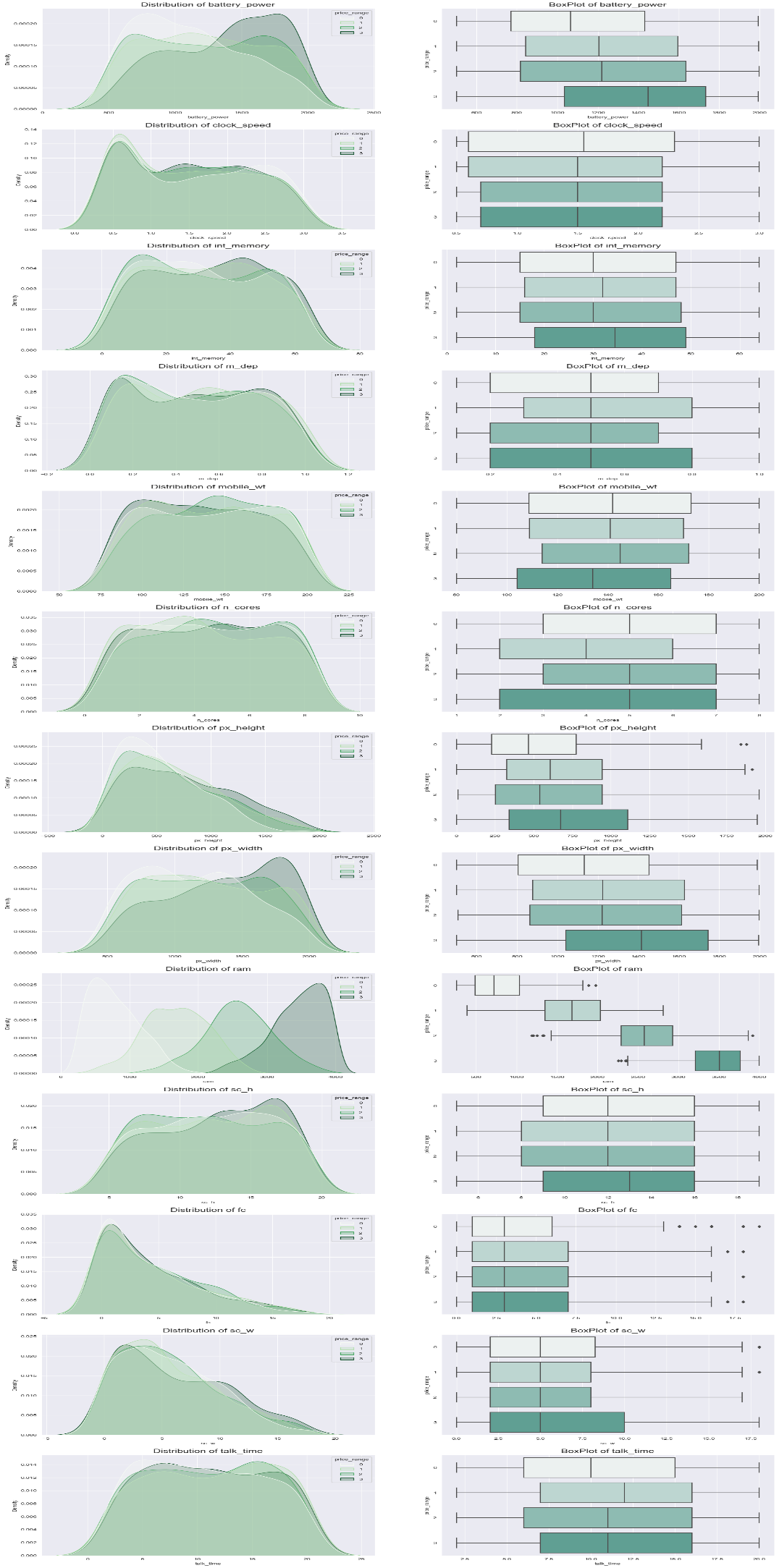
CHECK FOR OUTLIERS-



Maximum number of Outliers lie in px\_height and fc

EXPLORING CATEGORICAL FEATURES AND NUMERICAL FEATURES





2)MACHINE LEARNING MODELS-

In this Project I’ve used 3 different models such as Logistic Regression, Decision Tree, Random Forest, KNN and selected the model which gives best Accuracy out of those

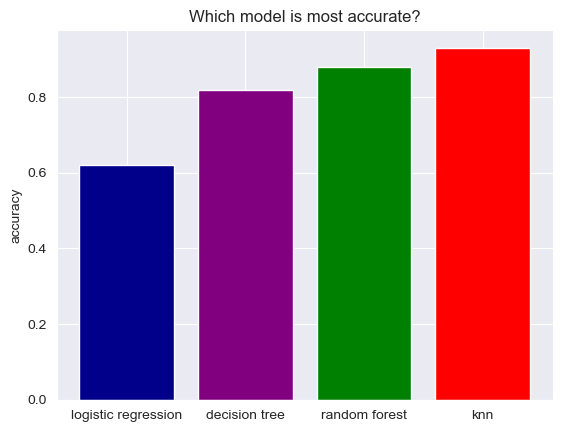
Accuracy of :

Logistic Regression-62%

Decision Tree-82%

Random Forest-88%

K-Nearest Neighbors-93%



Here, we can conclude by saying that KNN is best model for this dataset as it achieved highest accuracy

Then I’ve run KNN model to predict target values on the test dataset