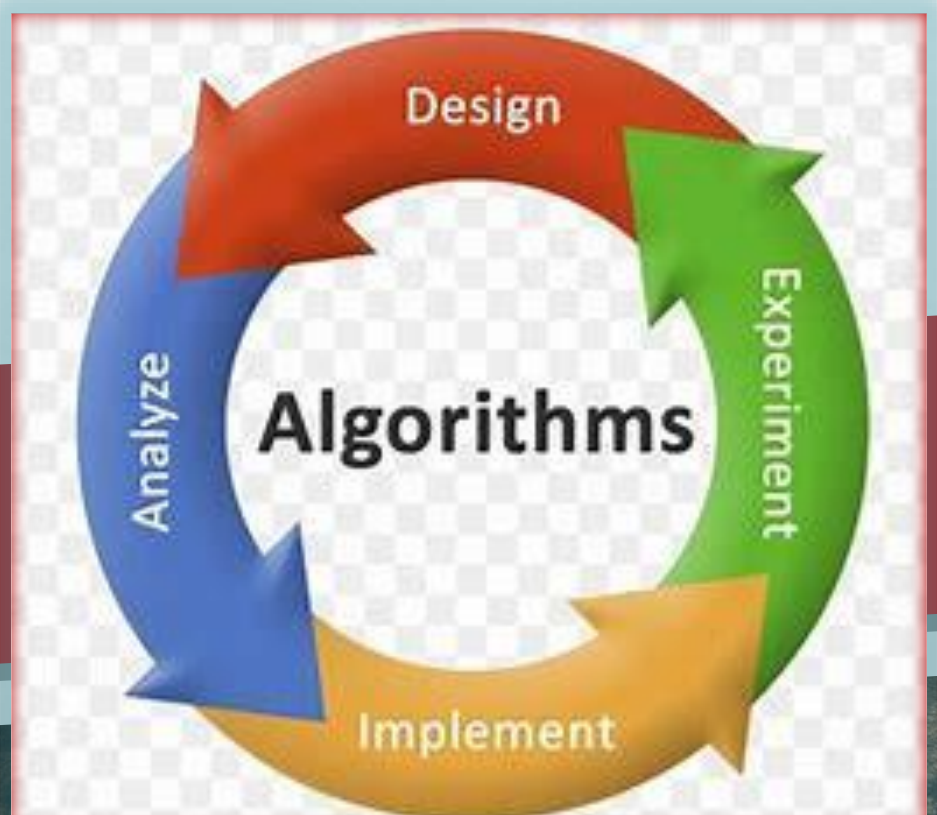
PROJECT – 4
FEATURE EXTRACTION AND PRICE PREDICTION FOR MOBILE PHONES

"A JOURNEY THROUGH THE WORLD OF MOBILE DATA WITH ADVANCED ANALYTICS TECHNIQUES"



PROJECT DESCRIPTION:

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

DATA PREPROCESSING:

- COVERTED OBJECT COLUMN INTO INTEGER LIKE:

  PRICE WAS 7,299

  NOW IT IS 7299

  IT HAS BEEN CONVERTED BY STRING FUNCTION

- NO MISSING VALUES ARE THERE
- ORDINAL ENCODING:

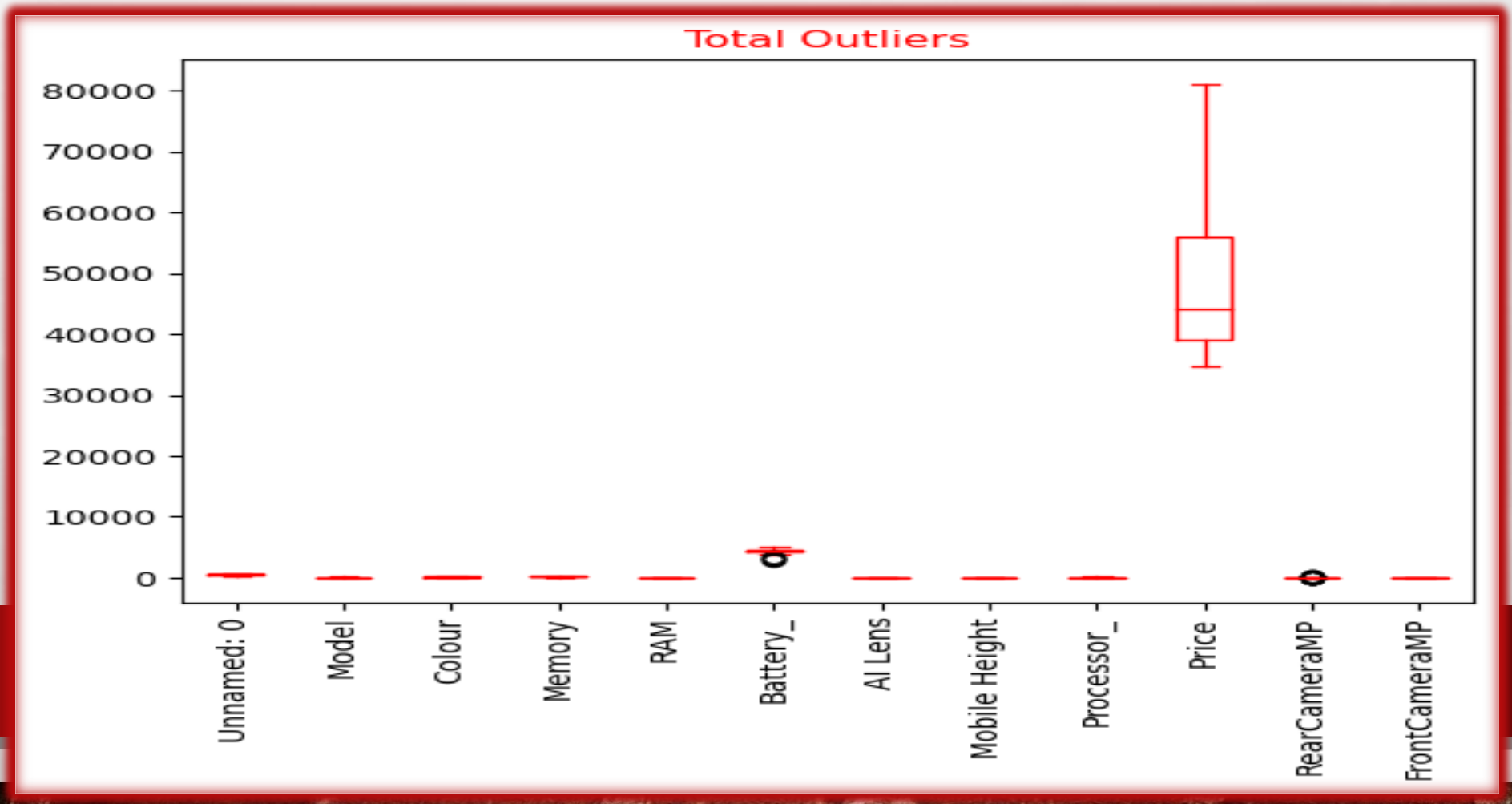
  THIS ENCODING HAS BEEN USED TO CONVERT CATEGORICAL VALUES INTO NUMERICAL FORMAT FOR THE COLUMNS LIKE:

  - MODEL
  - COLOUR
  - PROCESSOR

- CONVERTED REAR CAMERA AND FRONT CAMERA OBJECT INTO INTEGER TO REMOVE MP FROM NUMERICAL VALUE, BY USING LAMBDA FUNCTION

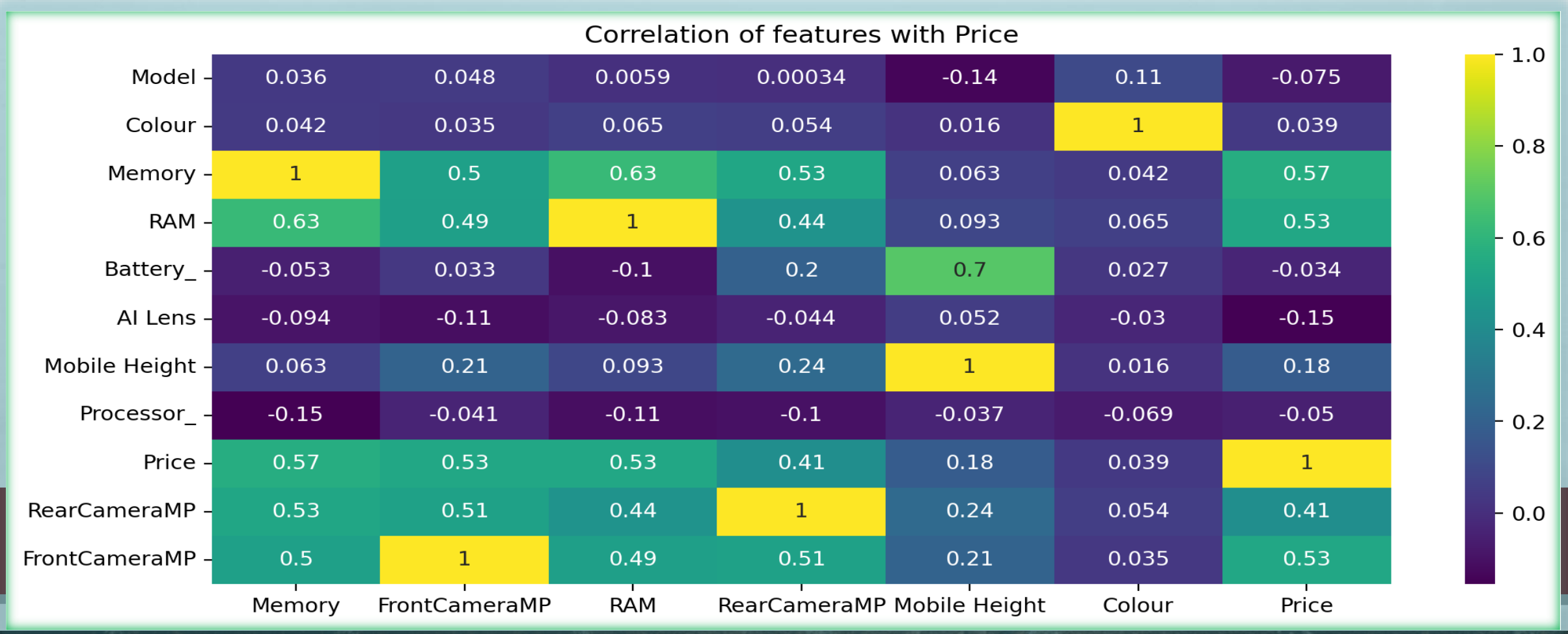
DATA PREPROCESSING:

- OUTLIERS HAVE BEEN CALCULATED BY:
  - MOST AFFECTING FEATURES METHOD
  - QUANTILE METHOD



FEATURE EXTRACTION:

- USED CORRELATION ANALYSIS TO IDENTIFY THE MOST RELEVANT FEATURES THAT STRONGLY AFFECT THE PRICE
- MEMORY, FRONT CAMERA, RAM, REAR CAMERA ARE THE MOST IMPORTANT FEATURES ACCORDING TO THEIR CORRELATION WITH PRICE



FEATURE EXTRACTION:

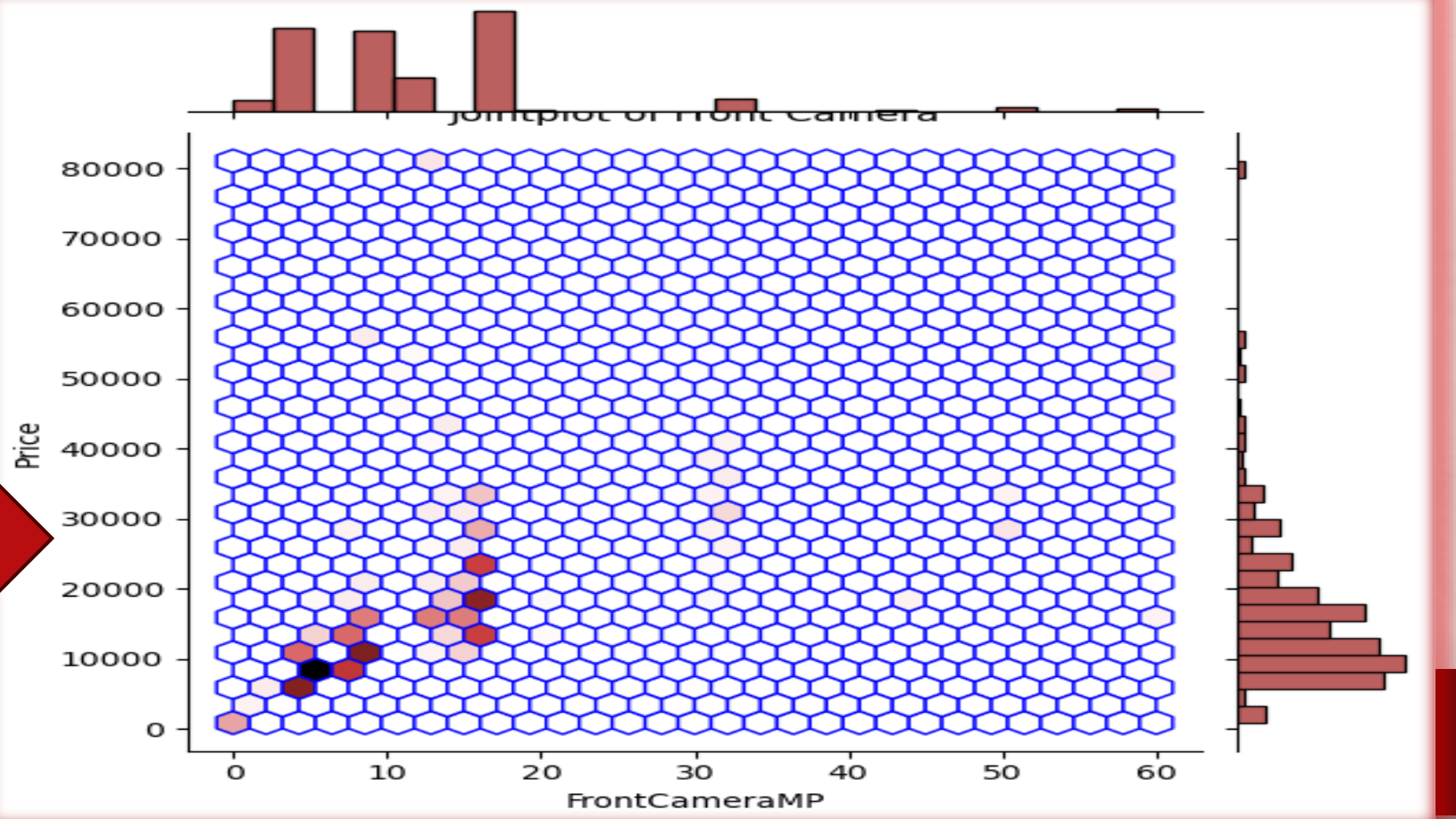
- DIMENSIONALITY REDUCTION

PCA METHOD HAS BEEN USED FOR THIS AND THE ACCURACY IS 1.0

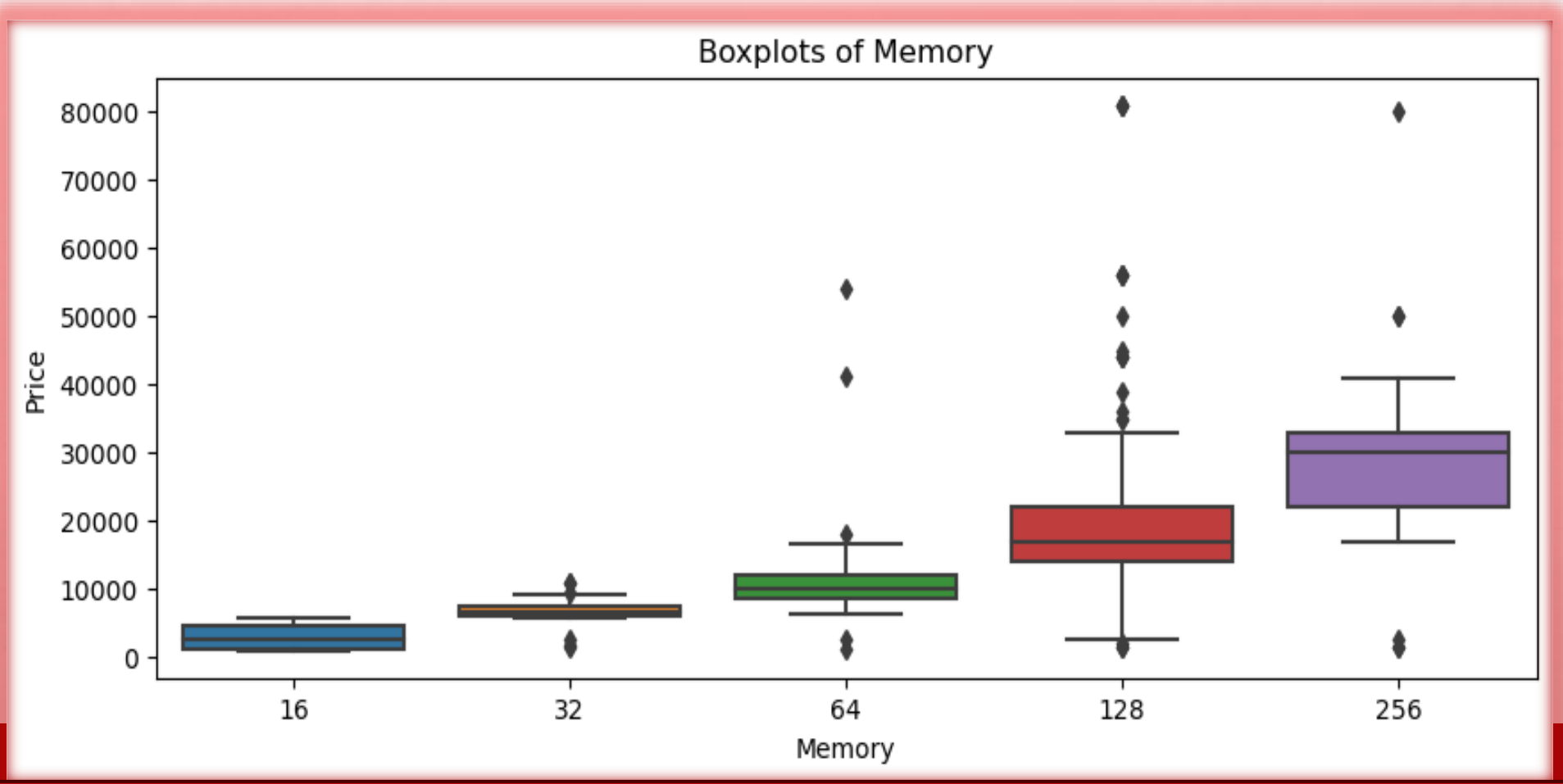
VISUALIZATION

JOINTPLOT, COUNTPLOT,
BOXPLOT,
SCATTERPLOT
AND MANY OTHER
PLOTS
HAVE BEEN USED

JOINT PLOT OF FRONT CAMERA



BOXPLOT OF MEMORY



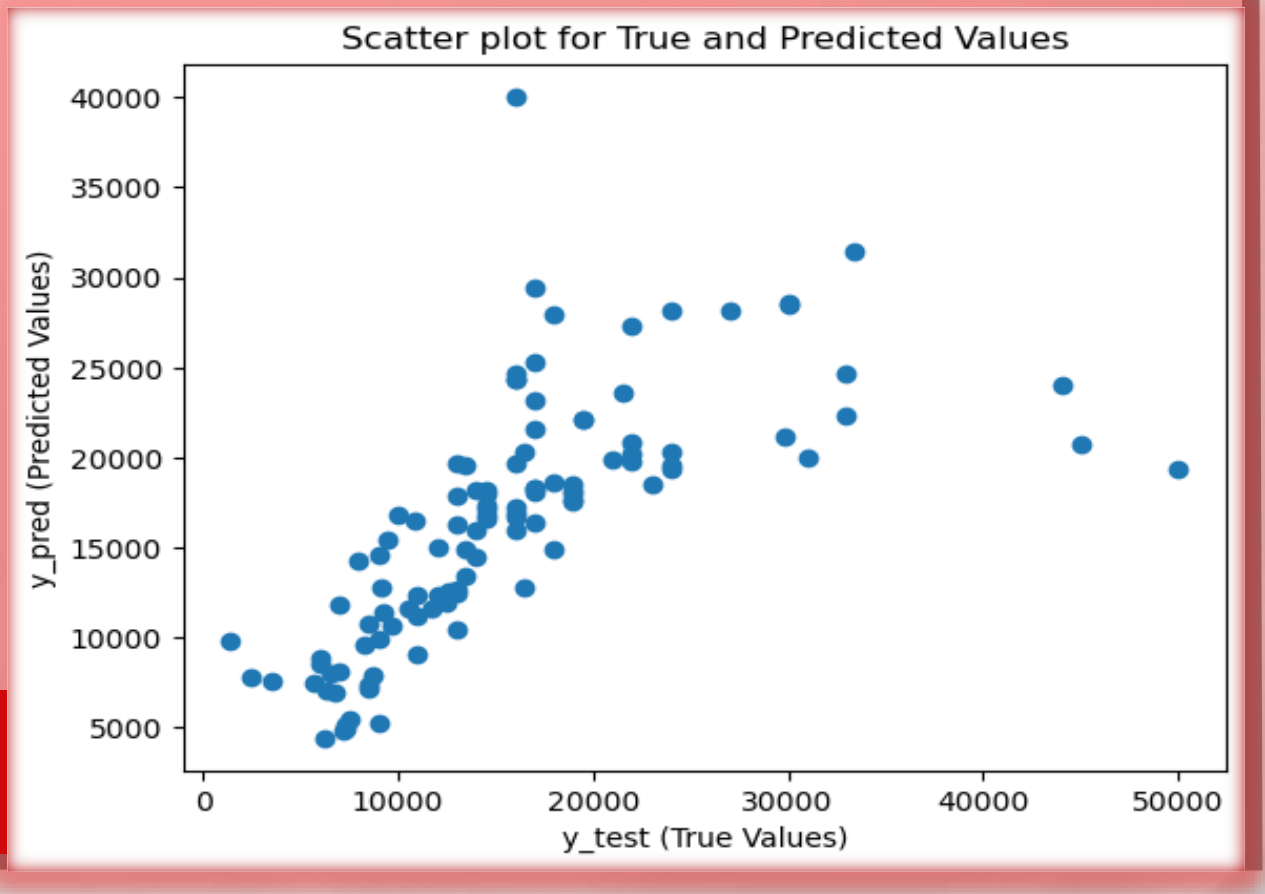
Most relevant feature which affects the Price

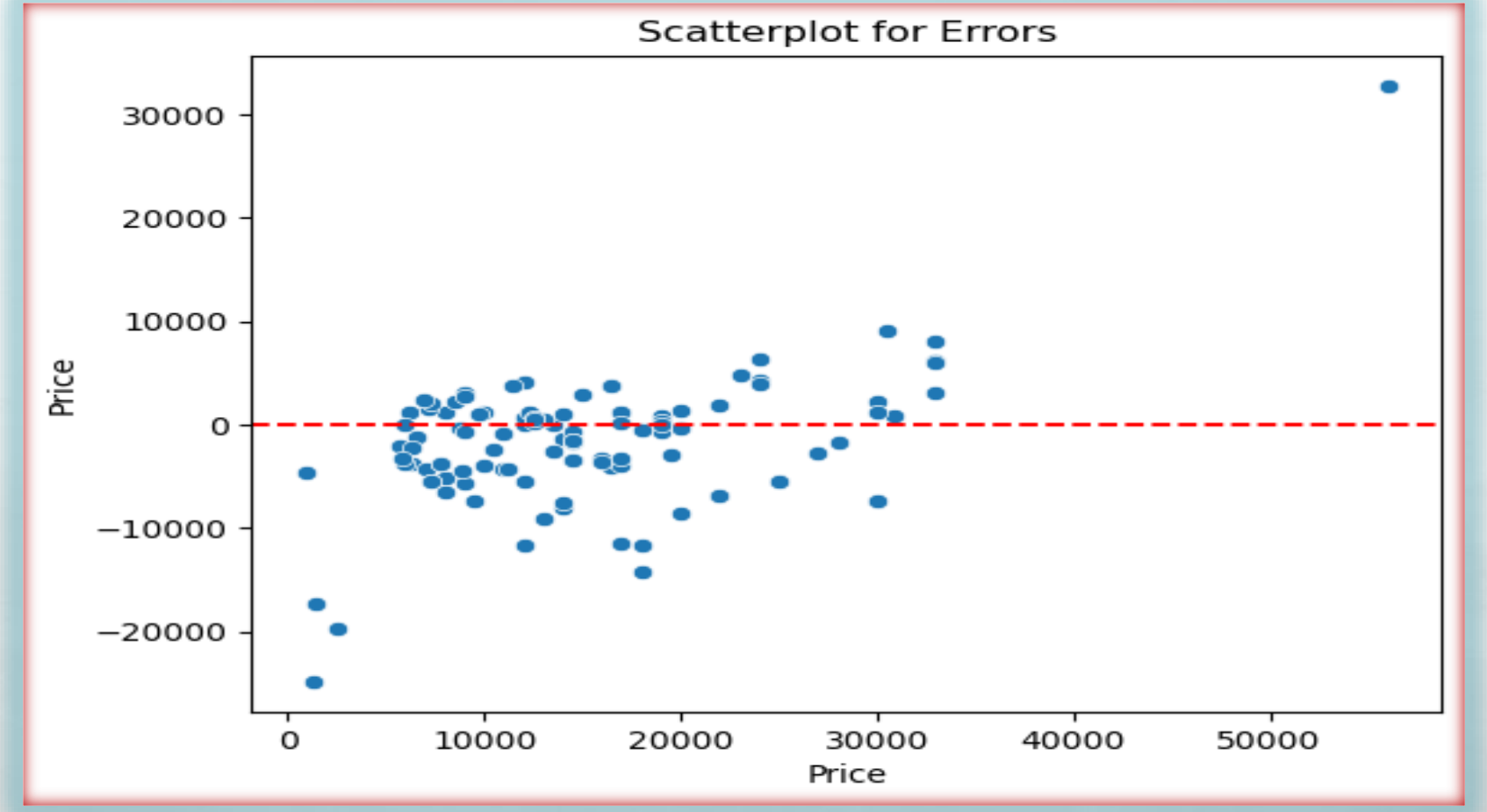
MODEL BUILDING:

- LINEAR REGRESSIO
- DECISION TREE
- RANDOM FOREST
- GRADIENT BOOSTING

LINEAR REGRESSION

SCORE = 0.44



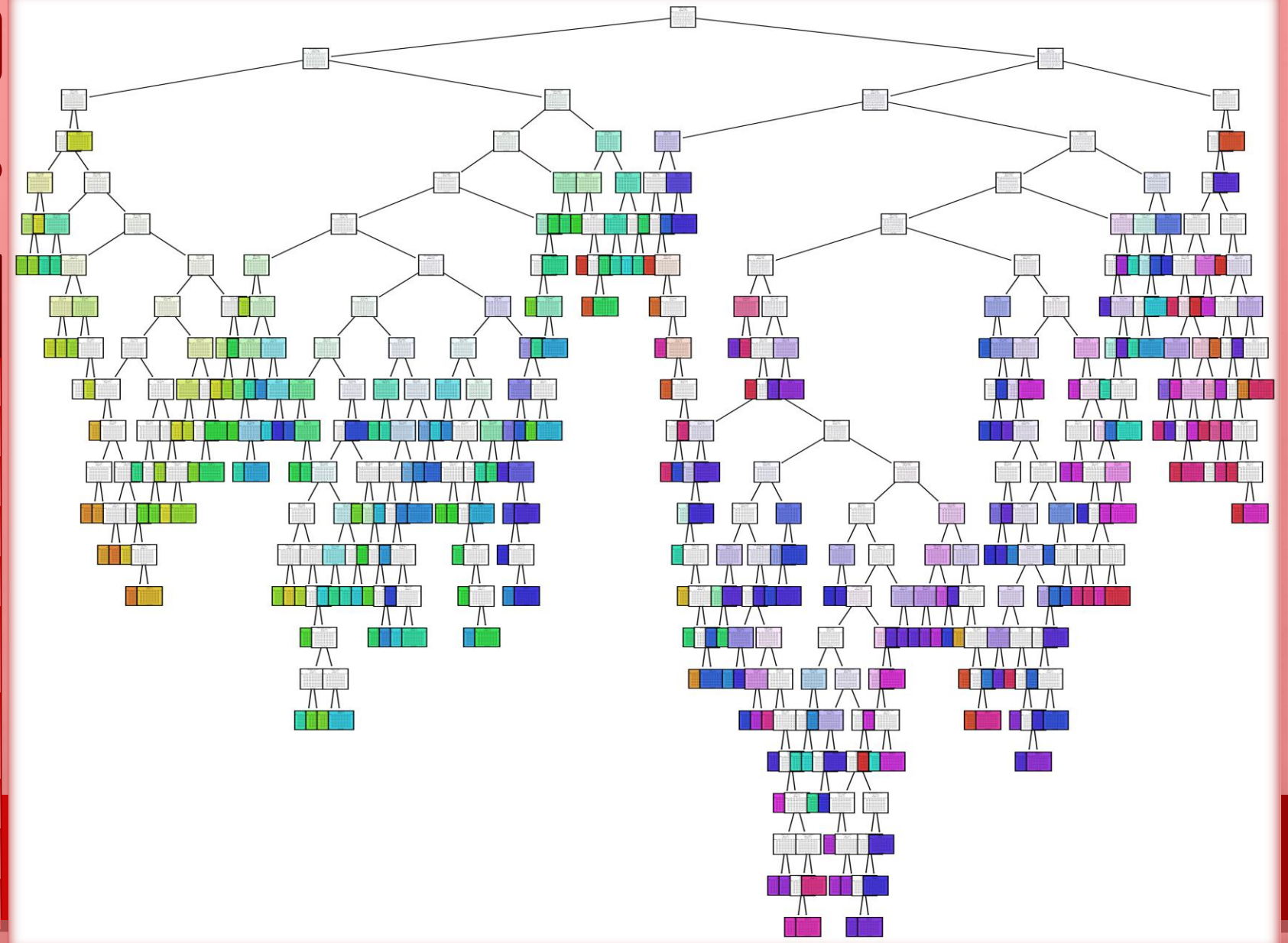


NO CLEAR LINE OR CURVE
SO WE SHOULD NOT USE LINEAR REGRESSION

DECISION TREE

SCORE = 0.61

	FEATURE IMPORTANCE
MODEL	0.219095
PROCESSOR_	0.184691
RAM	0.119651
MOBILE HEIGHT	0.116859
COLOUR	0.114248
FRONTCAMERAMP	0.081010
REARCAMERAMP	0.069034
BATTERY_	0.042313
MEMORY	0.040941
AI LENS	0.012159

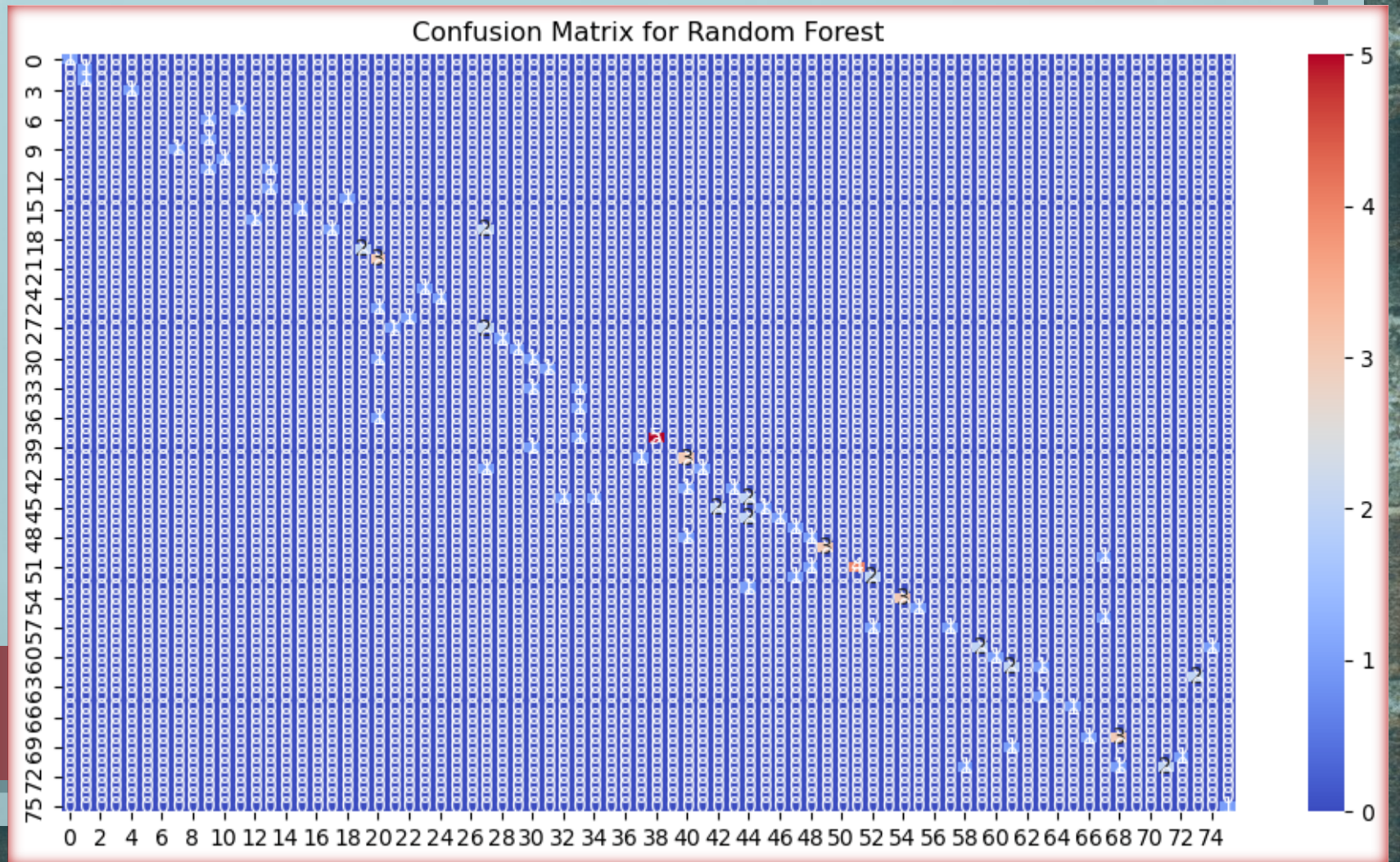


RANDOM FOREST

CLASSIFICATION REPORT
SCORE = 0.53

SCORE = 0.57

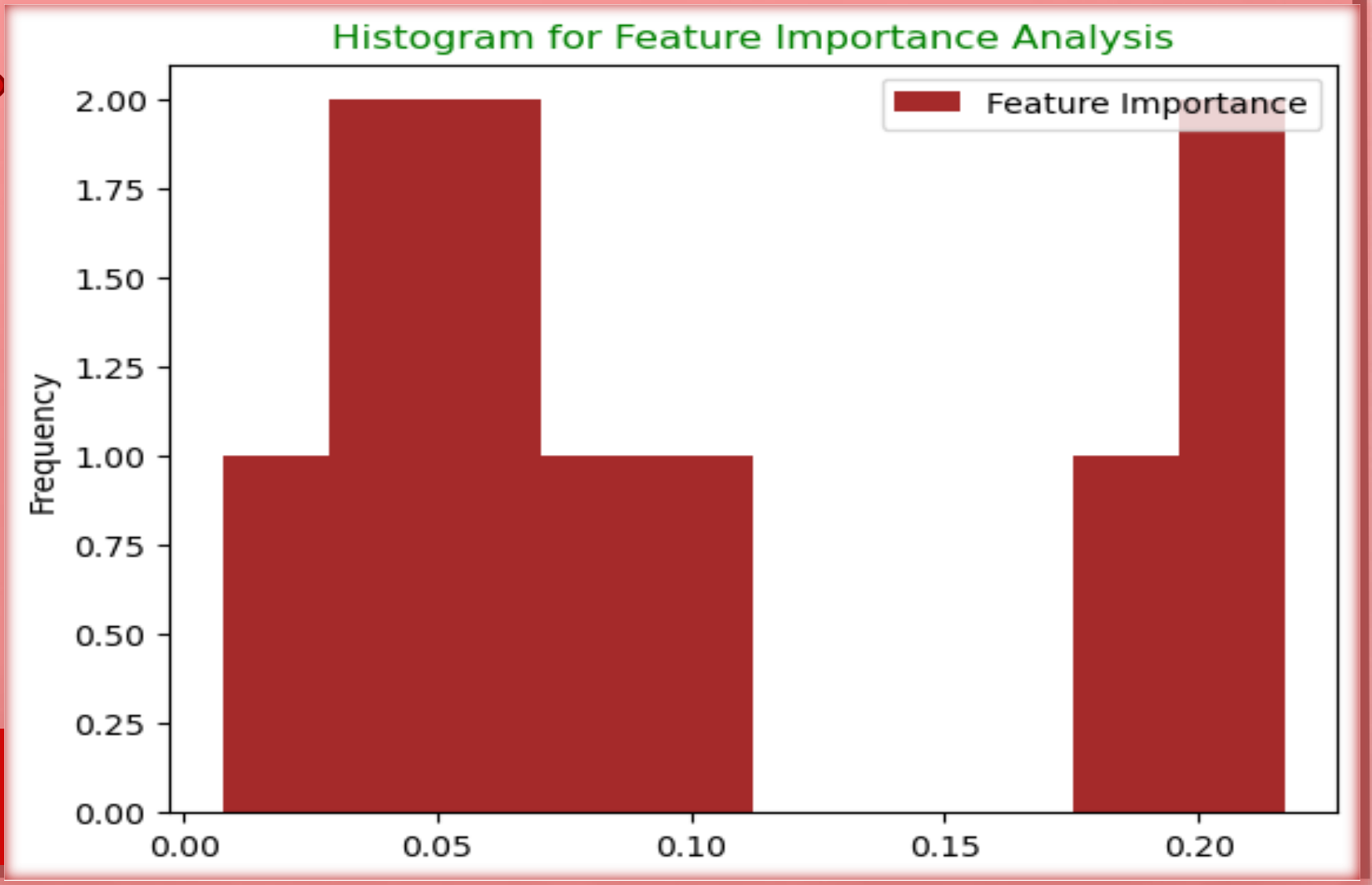
	FEATURE IMPORTANCE
MODEL	0.168056
COLOUR	0.163538
PROCESSOR_	0.160473
MOBILE HEIGHT	0.126542
RAM	0.100919
FRONTCAMERAMP	0.071675
REARCAMERAMP	0.071483
MEMORY	0.069416
BATTERY_	0.057153
AILENS	0.010744



GRADIENT BOOSTING CLASSIFICATION

SCORE = 0.61

	FEATURE IMPORTANCE
MODEL	0.216958
COLOUR	0.210234
PROCESSOR_	0.194729
RAM	0.110268
MOBILE HEIGHT	0.074244
MEMORY	0.068250
REARCAMERAMP	0.049575
FRONTCAMERAMP	0.039309
BATTERY_	0.028763
AI LENS	0.007671

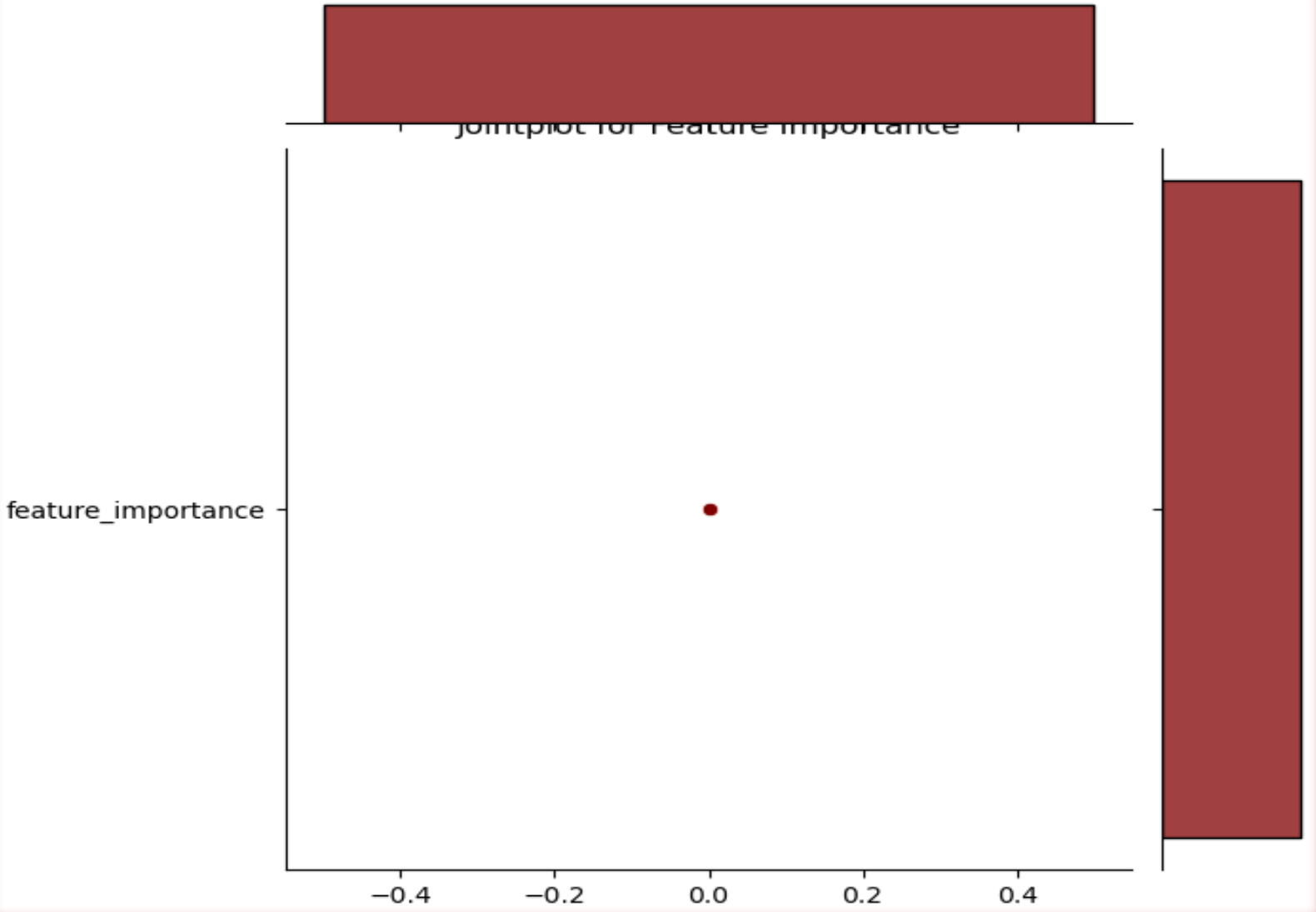


GRADIENT BOOSTING REGRESSION

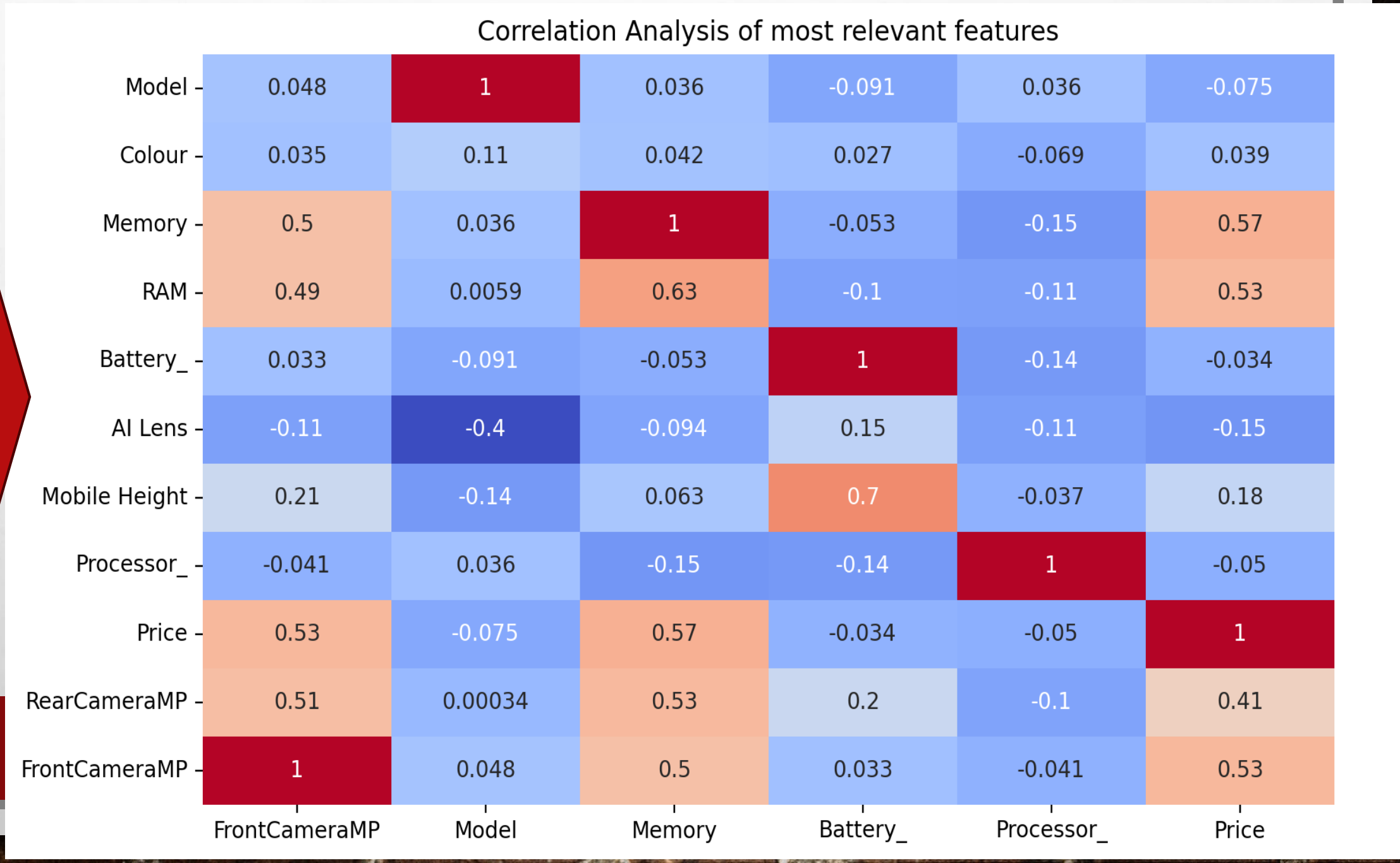
SCORE = 0.90

	FEATURE IMPORTANCE
FRONTCAMERAMP	0.460586
MODEL	0.197077
MEMORY	0.140548
BATTERY_	0.077638
REARCAMERAMP	0.028453
PROCESSOR_	0.026109
RAM	0.024242
MOBILE HEIGHT	0.024029
COLOUR	0.021075
AI LENS	0.000244

JOINTPLOT FOR FEATURE IMPORTANCE



Gradient Boosting Regression is the best algorithm for this data



RECOMMENDATIONS: FRONT CAMERA, MODEL AND MEMORY HAVE THE MOST SIGNIFICANT INFLUENCE ON MOBILE PHONE PRICES. THIS INFORMATION CAN INFORM PRICING STRATEGIES AND MARKETING DECISIONS.