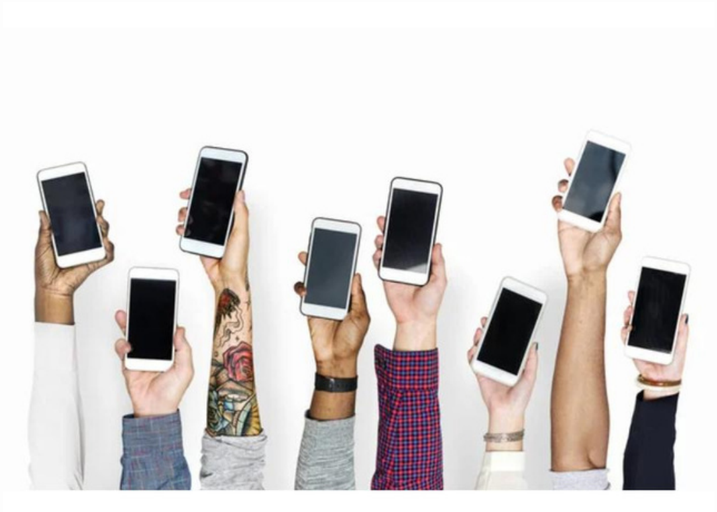


MACHINE LEARNING PROJECT

DATA SCIENCE BOOTCAMP| BIG BLUE DATA ACADEMY

TABLE OF CONTENTS

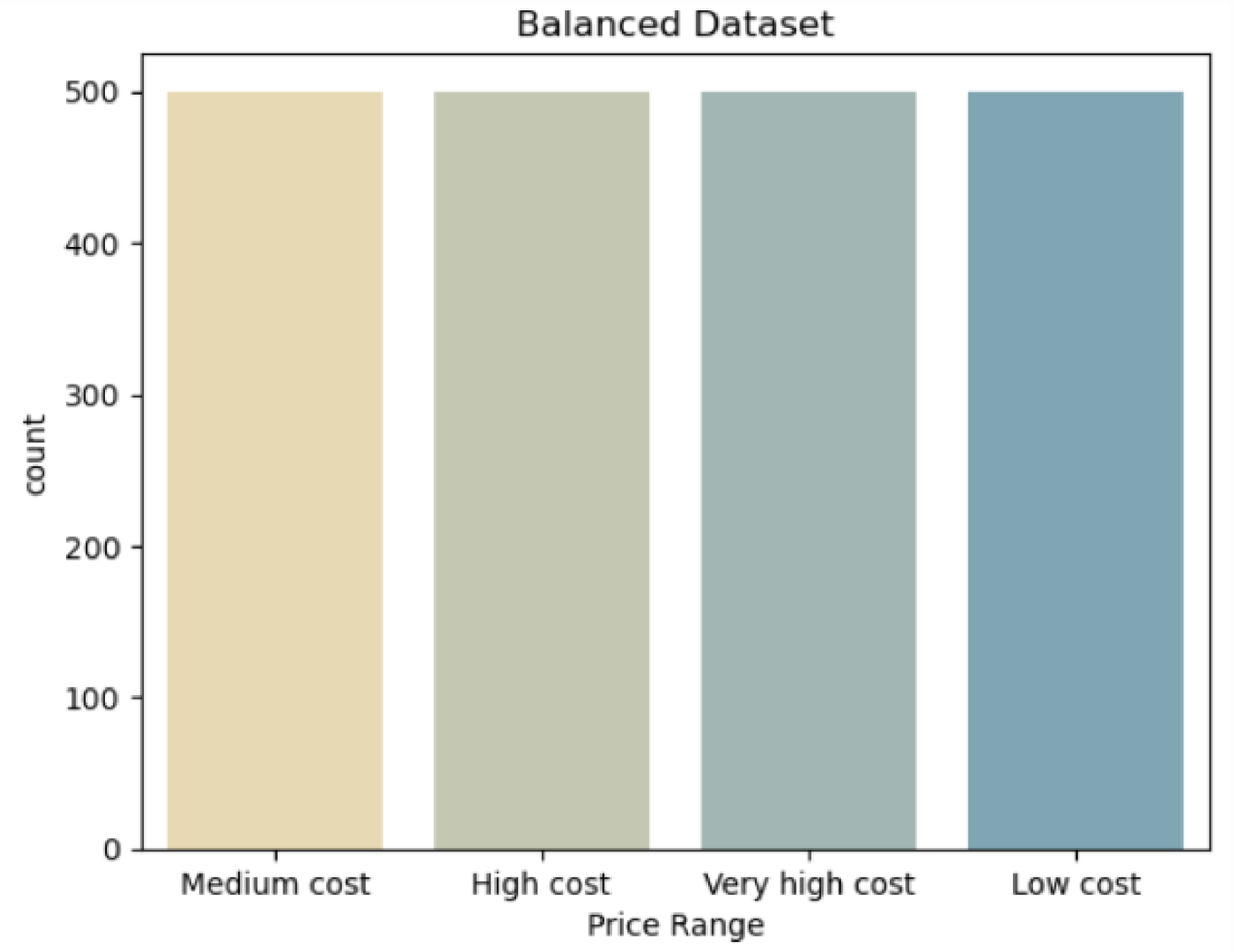
ABOUT US

We are a consulting company and we work for a client that has just started his own mobile company.

Our wants to give tough fight to big companies like Apple,Samsung etc.

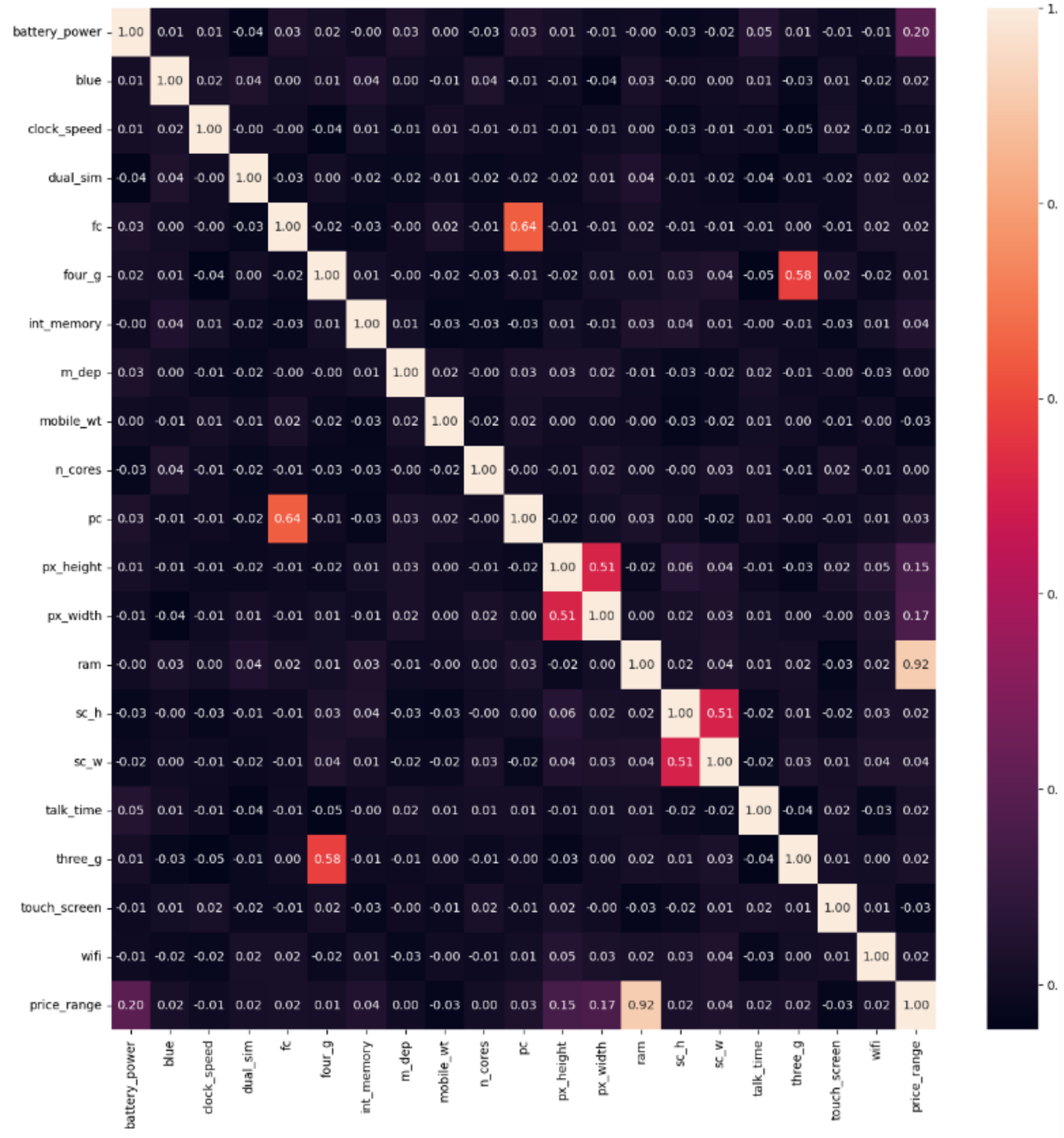
So,our goal is to identify a relationship between different features of a mobile phone.

Our objective is not to predict the actual price of a mobile phone, but to determine a price range that indicates how high the price is likely to be.



The dataset that we used had four classes for the predicted value, reffering to the range of mobile prices and as we observe it is fully balanced.

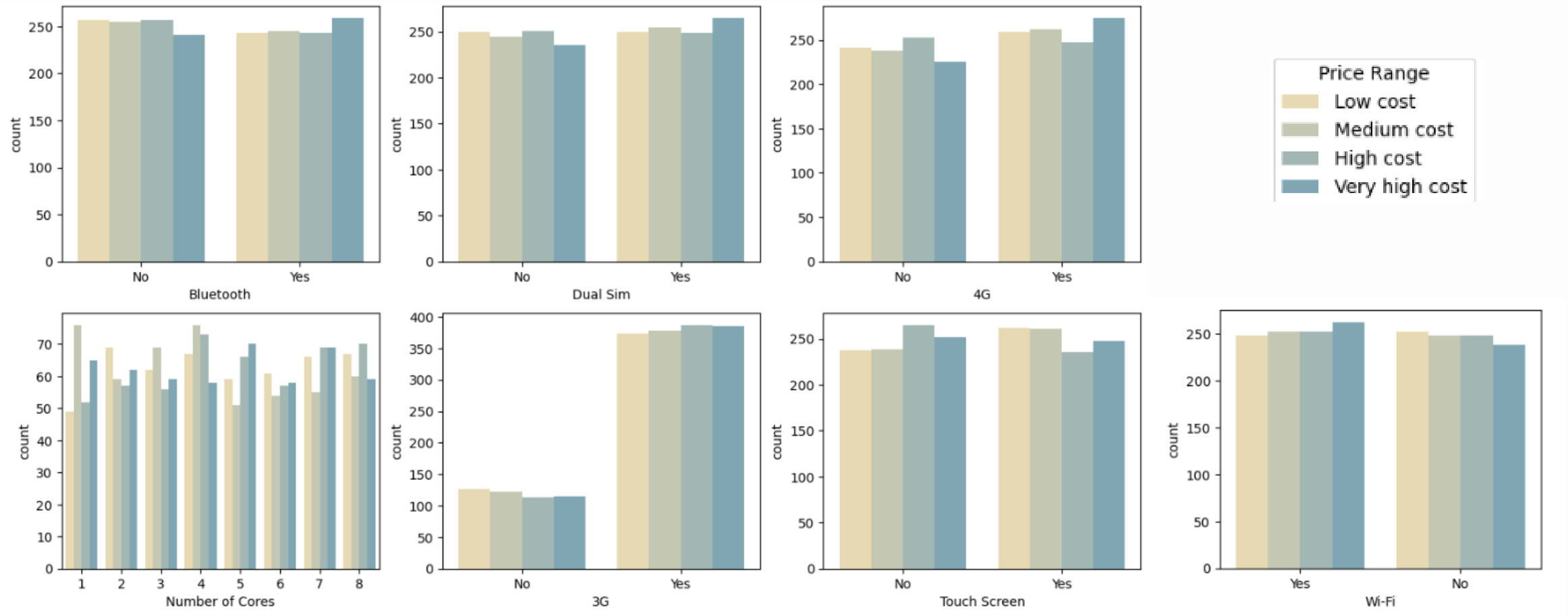
Correlation matxix



There are not any highly correlated feautures except from ram which we consider it as rational and not as a leakage clue.

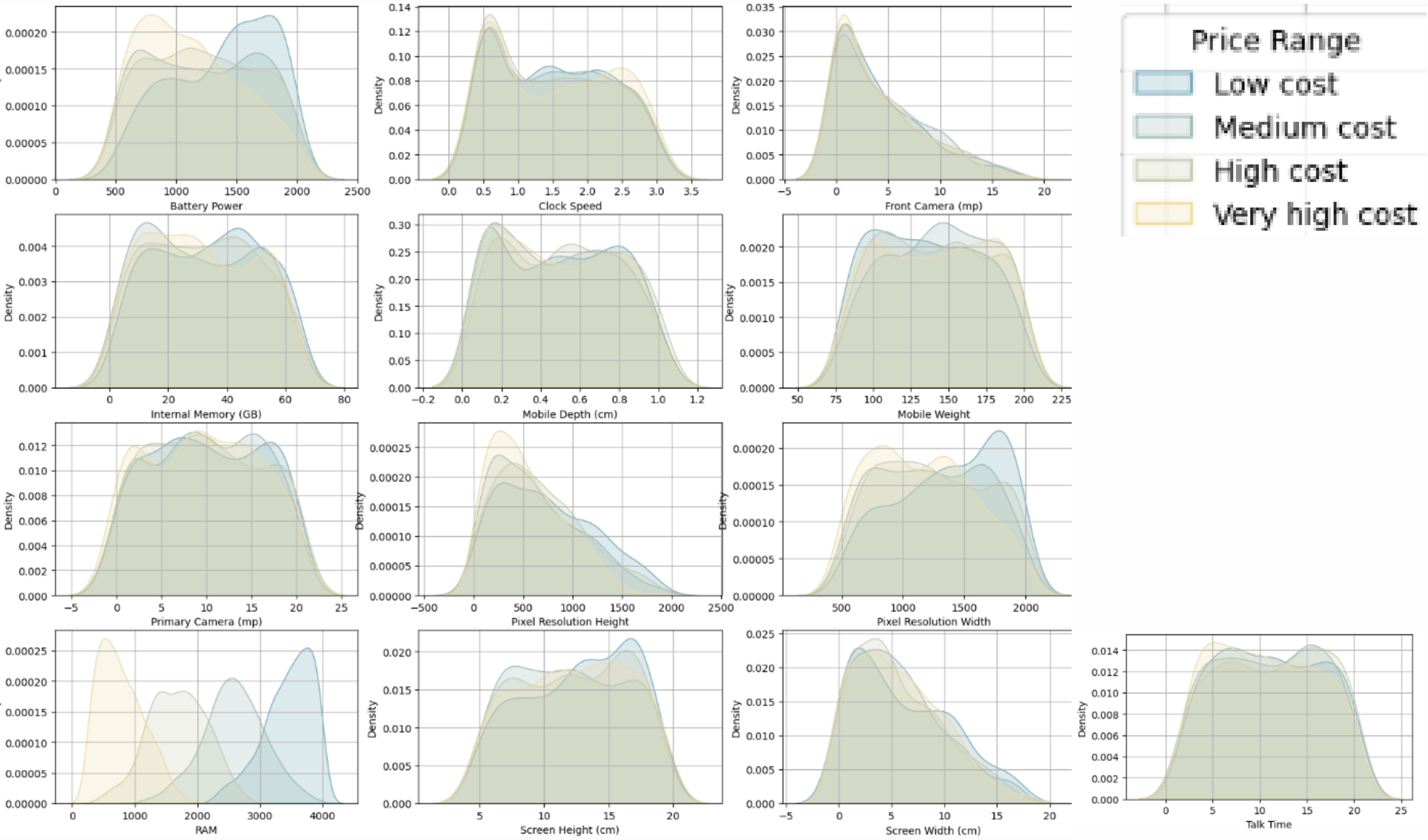
We had also to handle some zero values in screen width column and beacuase it is 50% correlated with sceen height, we used the average value from the sc_height to fill the zeros.

ANALYSIS OF CATEGORICAL FEATURES



- Most of the features are balanced
- There is an imbalance of phones equipped with 3g

ANALYSIS OF NUMERICAL FEATURES



The target seems changes according to the battery, RAM and Pixel Resolution width features

MODEL SELECTION

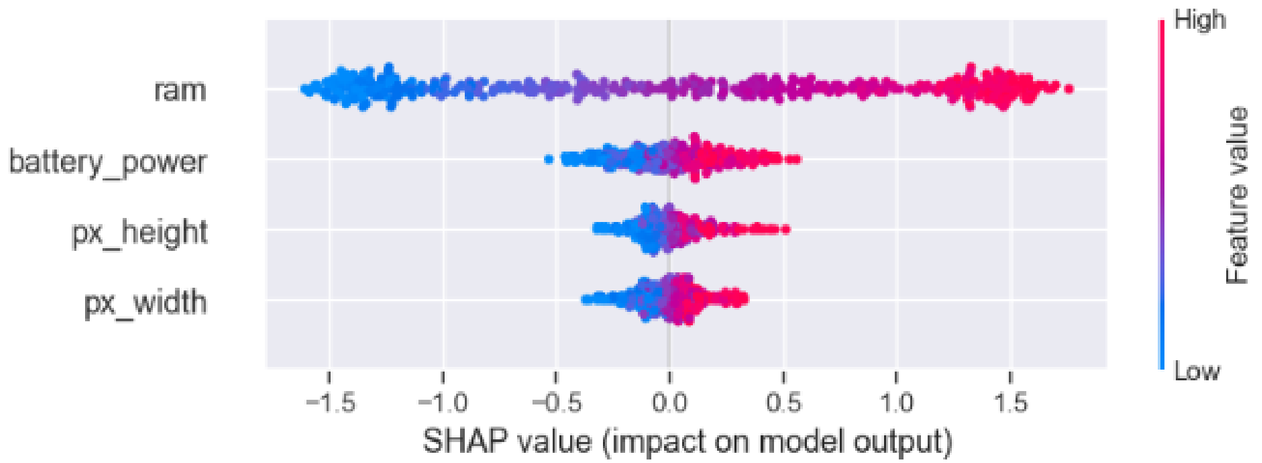
	Scaler	Classifier	Accuracy Score	F1-Score	T-Run
0	RobustScaler()	SVC(random_state=42)	0.9075	0.905198	0.177977
1	RobustScaler()	(DecisionTreeClassifier(max_features='sqrt', r...	0.8850	0.881654	2.362449
2	RobustScaler()	KNeighborsClassifier()	0.5075	0.506186	0.318610
3	MinMaxScaler()	SVC(random_state=42)	0.8625	0.858246	0.246982
4	MinMaxScaler()	(DecisionTreeClassifier(max_features='sqrt', r...	0.8875	0.884350	2.234623
5	MinMaxScaler()	KNeighborsClassifier()	0.4175	0.416180	0.031443
6	StandardScaler()	SVC(random_state=42)	0.8925	0.888857	0.175755
7	StandardScaler()	(DecisionTreeClassifier(max_features='sqrt', r...	0.8875	0.884350	2.337297
8	StandardScaler()	KNeighborsClassifier()	0.5300	0.528238	0.029875
9	None	SVC(random_state=42)	0.9650	0.964391	0.076395
10	None	(DecisionTreeClassifier(max_features='sqrt', r...	0.8850	0.881654	2.357937
11	None	KNeighborsClassifier()	0.9425	0.940759	0.022143

FEATURE SELECTION

	Specs	Score
13	ram	931267.519053
11	px_height	17363.569536
0	battery_power	14129.866576
12	px_width	9810.586750
8	mobile_wt	95.972863
6	int_memory	89.839124
15	sc_w	16.480319
16	talk_time	13.236400
4	fc	10.135166
14	sc_h	9.614878

Final Features for better score

- RAM
- Pixel Resolution Height
- Battery power
- Pixel Resolution Width



ACCURACY SUMMARY

SVC PERFORMANCE SUMMARY ON TEST DATA

	SVM
Accuracy	96.75%
Macro Precision	96.61%
Macro Recall	96.83%
Macro F1-score	96.68%

Here we see the final information about the performance of our selected model.

THANK YOU!

Alexandra Britsa
Athina Gerolymou
Evridiki Kourachani