Project Presentation

on

Predicting The Reality Of Mobile Price Range

CHIA-HAO,HSU

Introduction & Background

Introduction(5W1H)

What issue?

How much budget a person should prepare to buy a mobile phone.

Where?

India

When

The resource from 2016

Who(target audience)

People who wants to purchase a new phone based on their own requirements.

Why

Less knowledge or idea about how to measure the budget.

How

The analysis of classification to predict price range.

The price range

Price range label	class	Range	Example
0	Low cost	Up to 8,000	HUAWEI Y9, Asus Zenfone Max Pro M2
1	Medium cost	8,000~15,000	Sony Xperia 10 II, IPhone SE2
2	High cost	15,000~25,000	IPhone 11, Pixel 5, Samsung Galaxy S20 FE 5G
3	Very high cost	At least 25,000	IPhone 12, ASUS ROG Phone 3

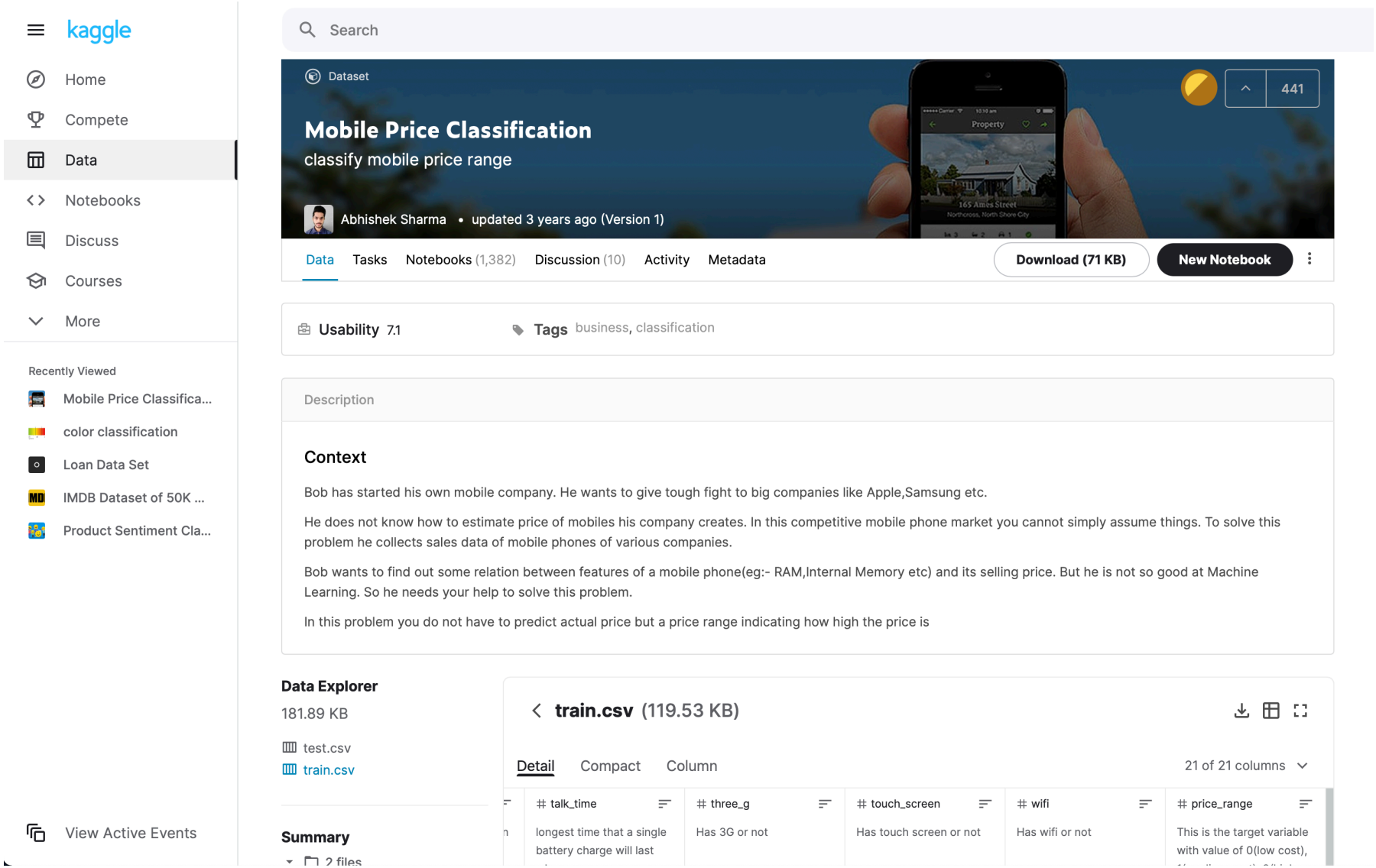
Whole Progresses

1. Introduction to the dataset
2. Data Preparation
3. Data Exploring
4. Regression model
5. Conclusion

Introduction to the Dataset

01

Data Sources



URL:https://www.kaggle.com/iabhishekofficial/mobile-price-classification?select=train.csv

Introduction of dataset(1)

Variable	Type	Unit	Description
battery_power	int64	mAh	The volume of battery
blue	int64		Support Bluetooth (Yes:1, No:0)
clock_speed	float64	GHz	The speed of microprocessors
dual_sim	int64		Support dual Sim card (Yes:1, No:0)
fc	int64	Mega pixels	The pixels of front camera
four_g	int64		Suppprt 4G (Yes:1, No:0)
Int_memory	flost64	Gigabytes	The volume of internal memory

Introduction of dataset(2)

Variable	Type	Unit	Description
m_dep	float64	cm	The depth of mobile phone手機厚度
Mobile_wt	float64	g	The weight of mobile phone
n_cores	int64	Cores	The number of cores in processor
pc	int64	Mega pixels	The pixels of primary camera
px_height	int64	Pixels	The height of pixels
Px_wideth	float64	Pixels	The width of pixels
ram	float64	Mega bytes	The volume of random access memory

Introduction of dataset(3)

Response

Variable	Type	Unit	Description
sc_h	float64	cm	The height of screen
sc_w	float64	cm	The width of screen
talk_time	float64	hour	The largest time of phone calling(without extra charging)
three_g	int64		Support 3G (Yes:1, No:0)
touch_screen	int64		With touch screen (Yes:1, No:0)
wifi	int64		Support WIFI (Yes:1, No:0)
price_range	int64		0(low cost), 1(medium cost), 2(high cost) and 3(very high cost).

Data Preparation

02

Detecting and dealing with Null Values

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   battery_power  2000 non-null   int64
 1   blue           2000 non-null   int64
 2   clock_speed    1957 non-null   float64
 3   dual_sim       2000 non-null   int64
 4   fc             2000 non-null   int64
 5   four_g         2000 non-null   int64
 6   int_memory     1974 non-null   float64
 7   m_dep          1984 non-null   float64
 8   mobile_wt      1972 non-null   float64
 9   n_cores        2000 non-null   int64
 10  pc             2000 non-null   int64
 11  px_height      2000 non-null   int64
 12  px_width       1969 non-null   float64
 13  ram            1978 non-null   float64
 14  sc_h           1982 non-null   float64
 15  sc_w           1982 non-null   float64
 16  talk_time      1986 non-null   float64
 17  three_g        2000 non-null   int64
 18  touch_screen   2000 non-null   int64
 19  wifi           2000 non-null   int64
 20  price_range    2000 non-null   int64
dtypes: float64(9), int64(12)
memory usage: 328.2 KB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   battery_power  2000 non-null   int64
 1   blue           2000 non-null   int64
 2   clock_speed    2000 non-null   float64
 3   dual_sim       2000 non-null   int64
 4   fc             2000 non-null   int64
 5   four_g         2000 non-null   int64
 6   int_memory     2000 non-null   float64
 7   m_dep          2000 non-null   float64
 8   mobile_wt      2000 non-null   float64
 9   n_cores        2000 non-null   int64
 10  pc             2000 non-null   int64
 11  px_height      2000 non-null   int64
 12  px_width       2000 non-null   float64
 13  ram            2000 non-null   float64
 14  sc_h           2000 non-null   float64
 15  sc_w           2000 non-null   float64
 16  talk_time      2000 non-null   float64
 17  three_g        2000 non-null   int64
 18  touch_screen   2000 non-null   int64
 19  wifi           2000 non-null   int64
 20  price_range    2000 non-null   object
dtypes: float64(9), int64(11), object(1)
memory usage: 328.2+ KB
```

Detecting and dealing with Null Values

A	B	C
10	100	1000
20	200	2000
30	300	3000
N/A	400	4000
40	N/A	5000
50	500	N/A
60	600	6000

```
df['A']= df['A'].fillna( ( round( df ['A'].mean() ) ) )
df['B']= df['B'].fillna( ( round( df ['B'].mean() ) ) )
df['C']= df['C'].fillna( ( round( df ['C'].mean() ) ) )
```

When the scale is large, filling null value with mean good at maintaining the distribution in each variable.

A	B	C
10	60	300
20	120	600
30	180	900
35	210	1200
40	190	1500
50	270	1100
60	300	2100

Data Exploring

03

Heat Map Of All Variables

From the correlation coefficient:

r (fc,pc)=0.64

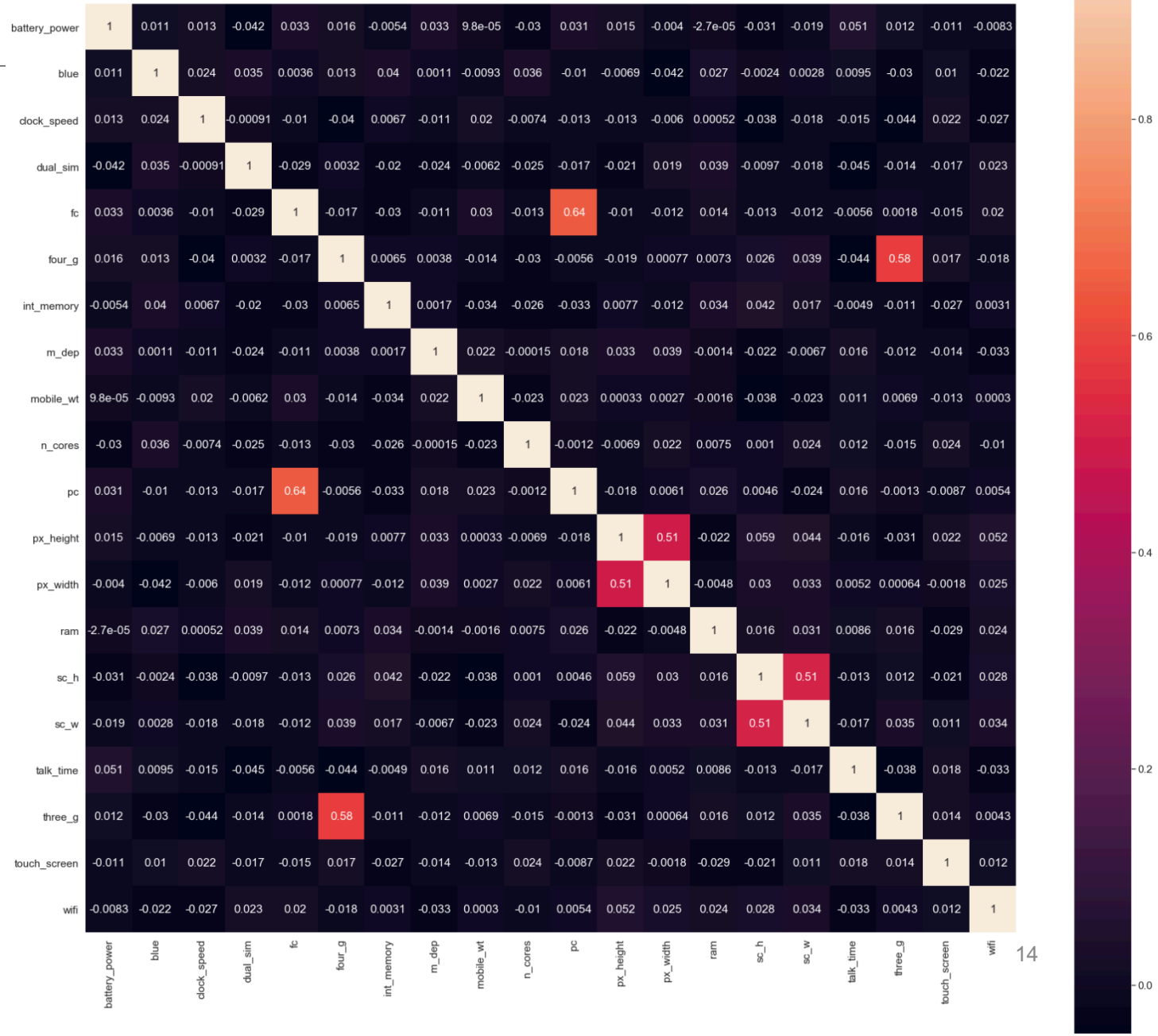
r (three_g, four_g)=0.58

r (px_h, px_width)=0.51

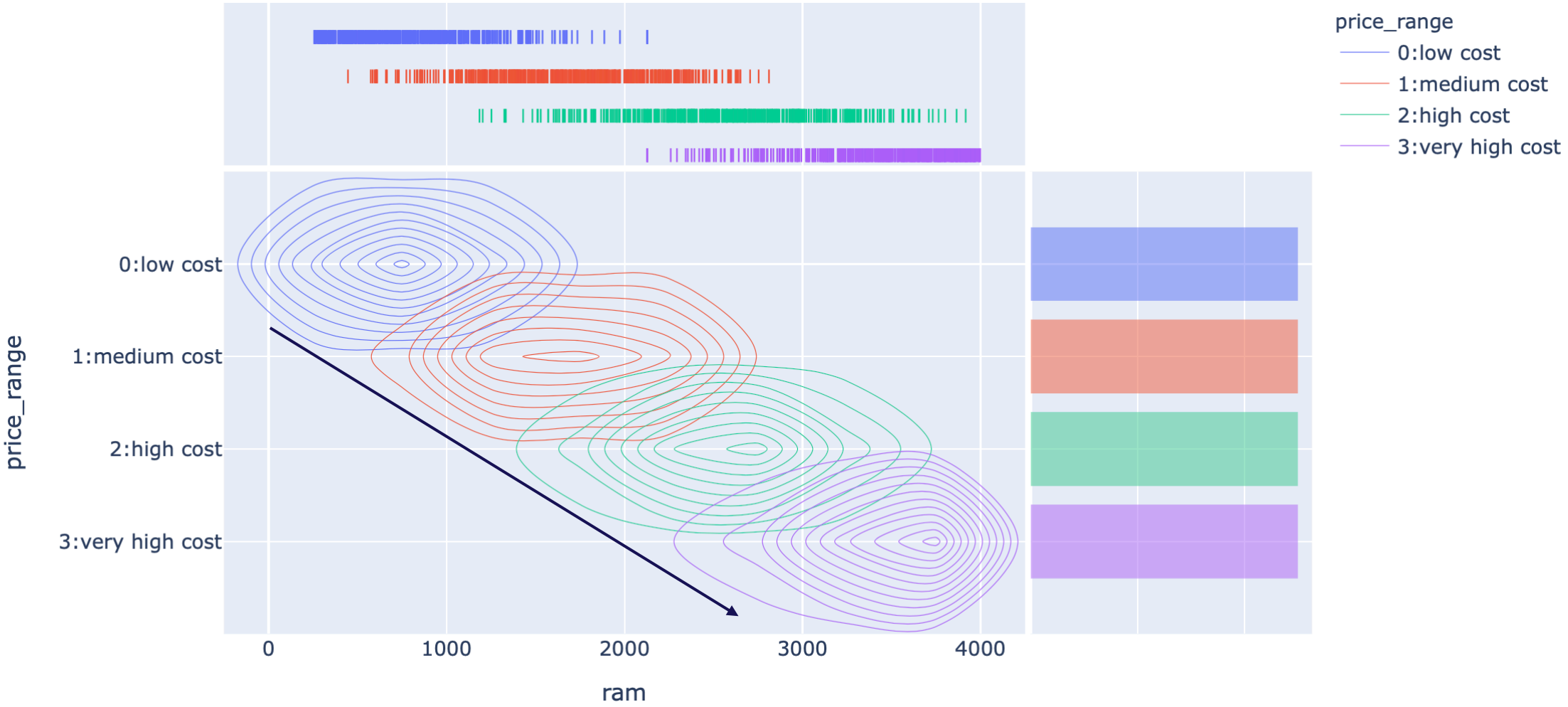
r (sc_w, sc_h)=0.51

If r>0.7, it has collinearity issue,
we must remove one of variable in the combination because both have same feature in the dataset.

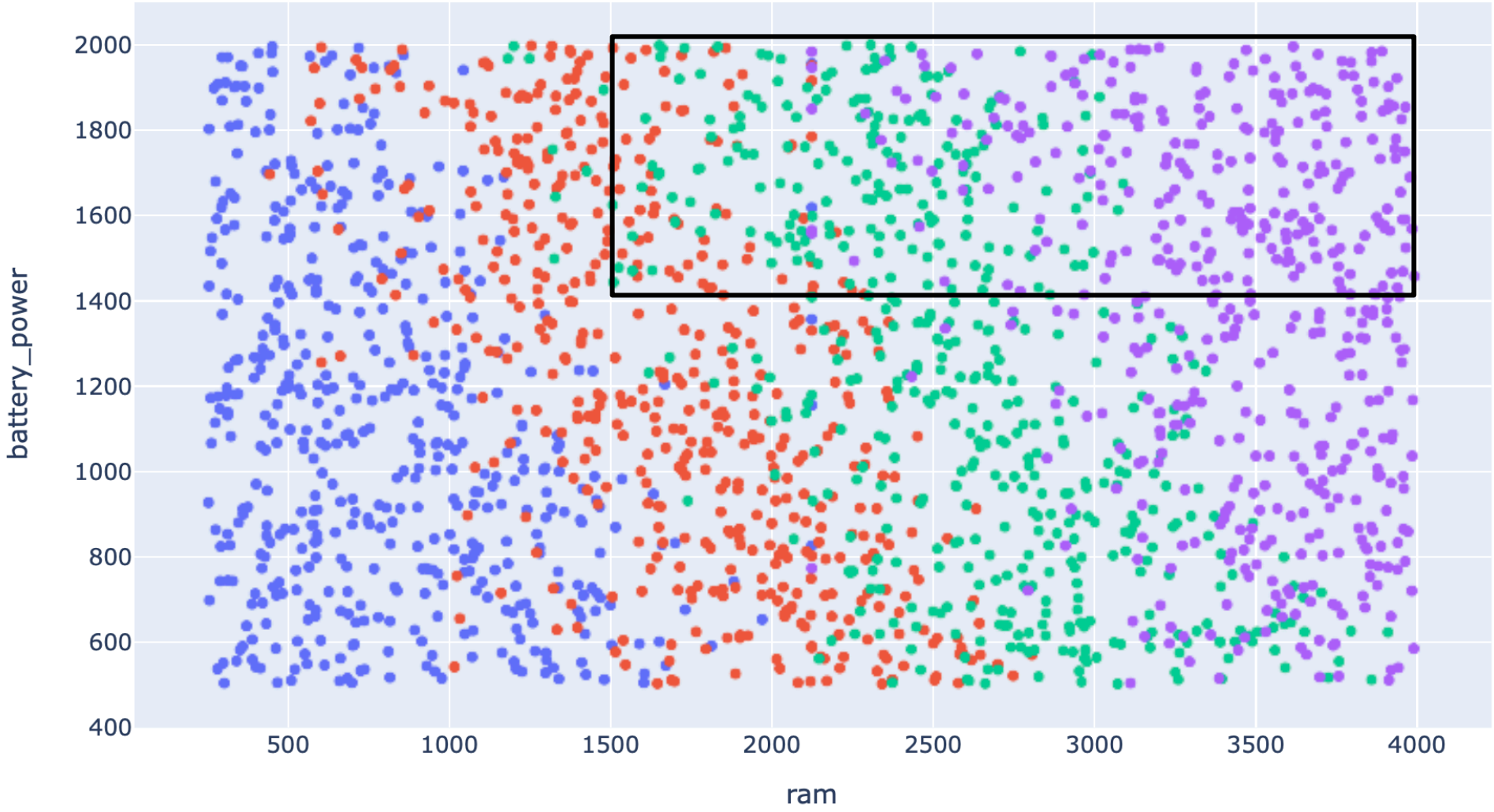
In the heat map,
there is no collinearity issue,
so we **do not need to remove any variable.**



Rule 1-
More volume of ram, the much price that customers should pay.

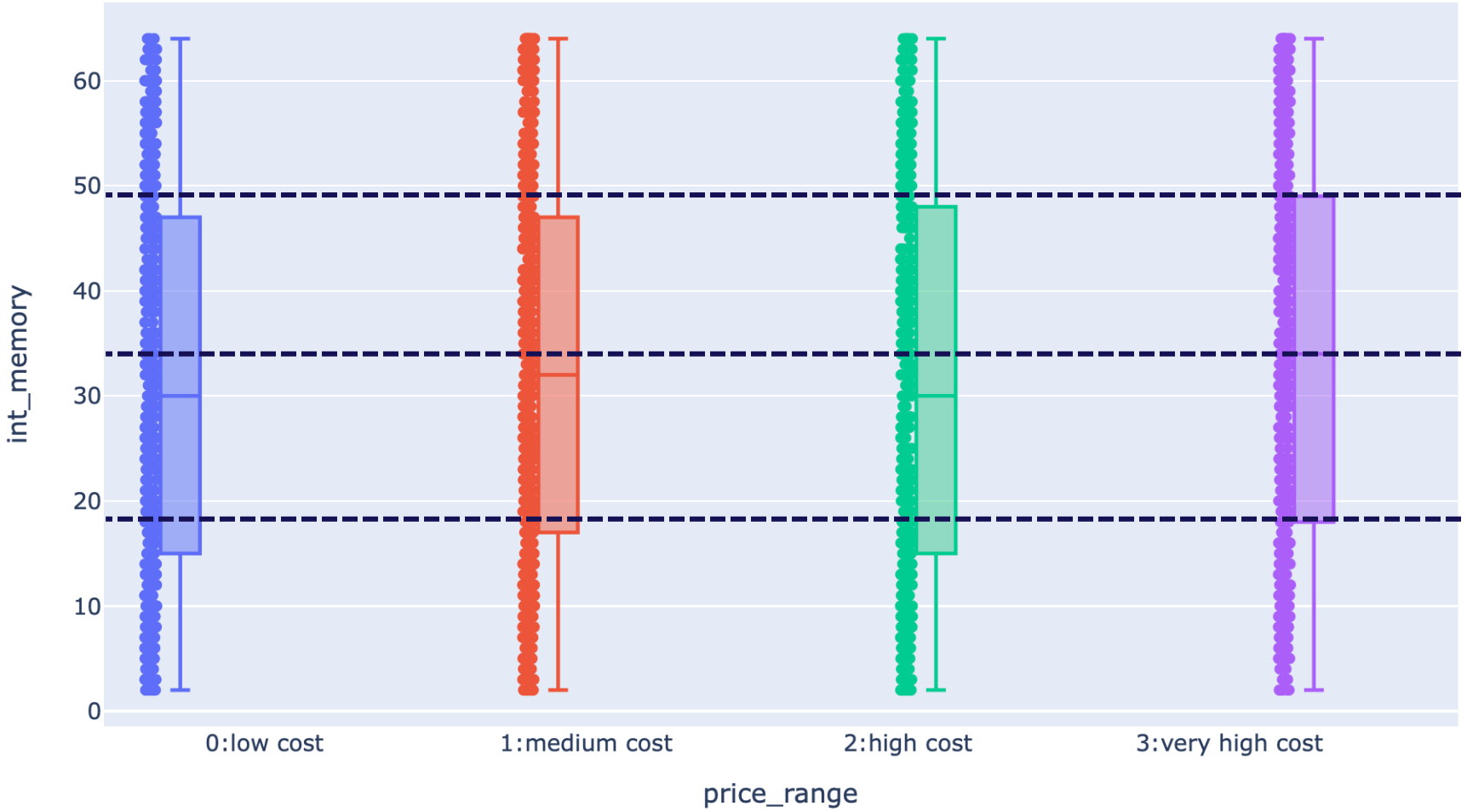


Rule 2-
High battery power & ram spend much money to have.



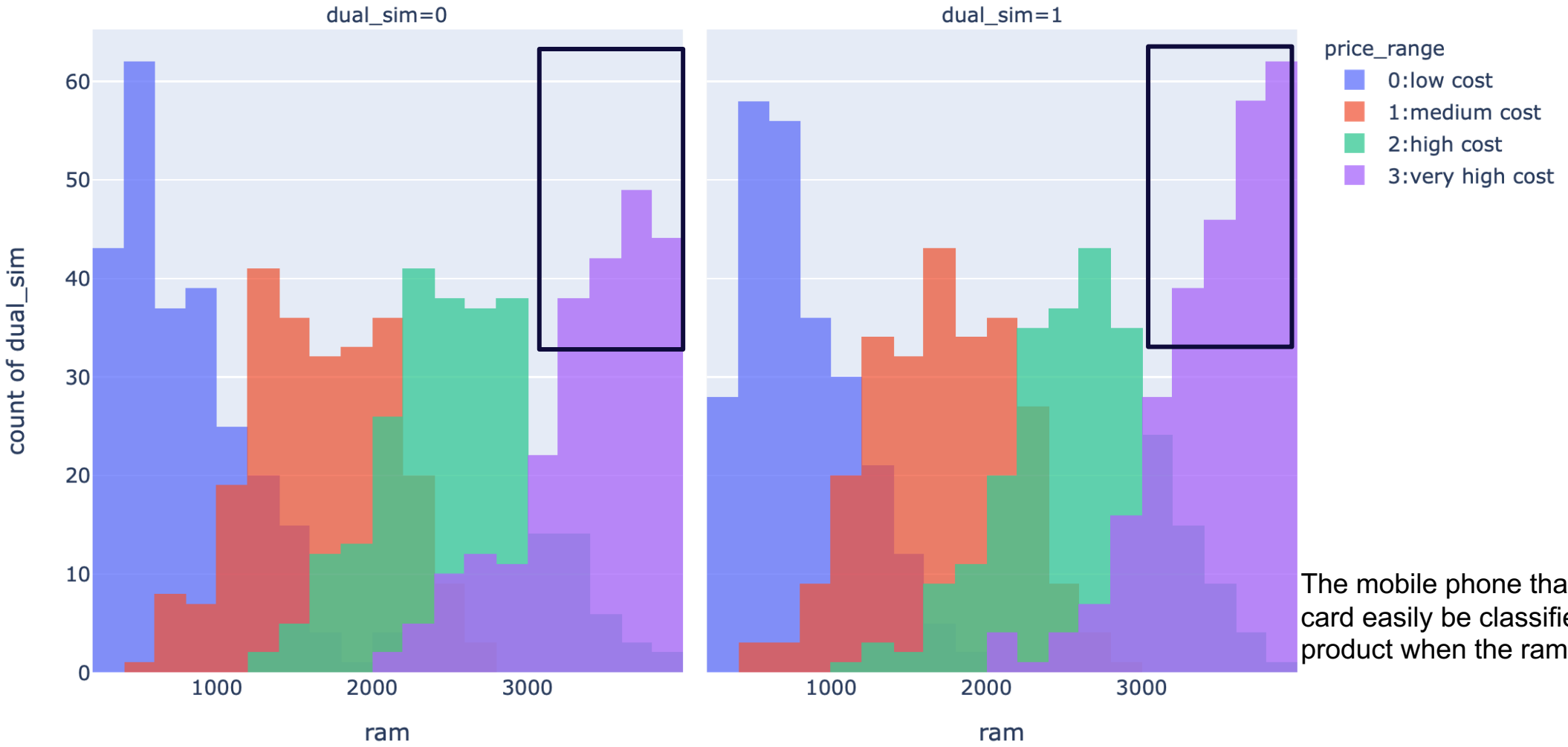
When battery power>1,400 mAh and ram >1500, most product may be classified as high cost or very high-cost mobile phone.

Rule 3-
More volume of memory, much budget to prepare.



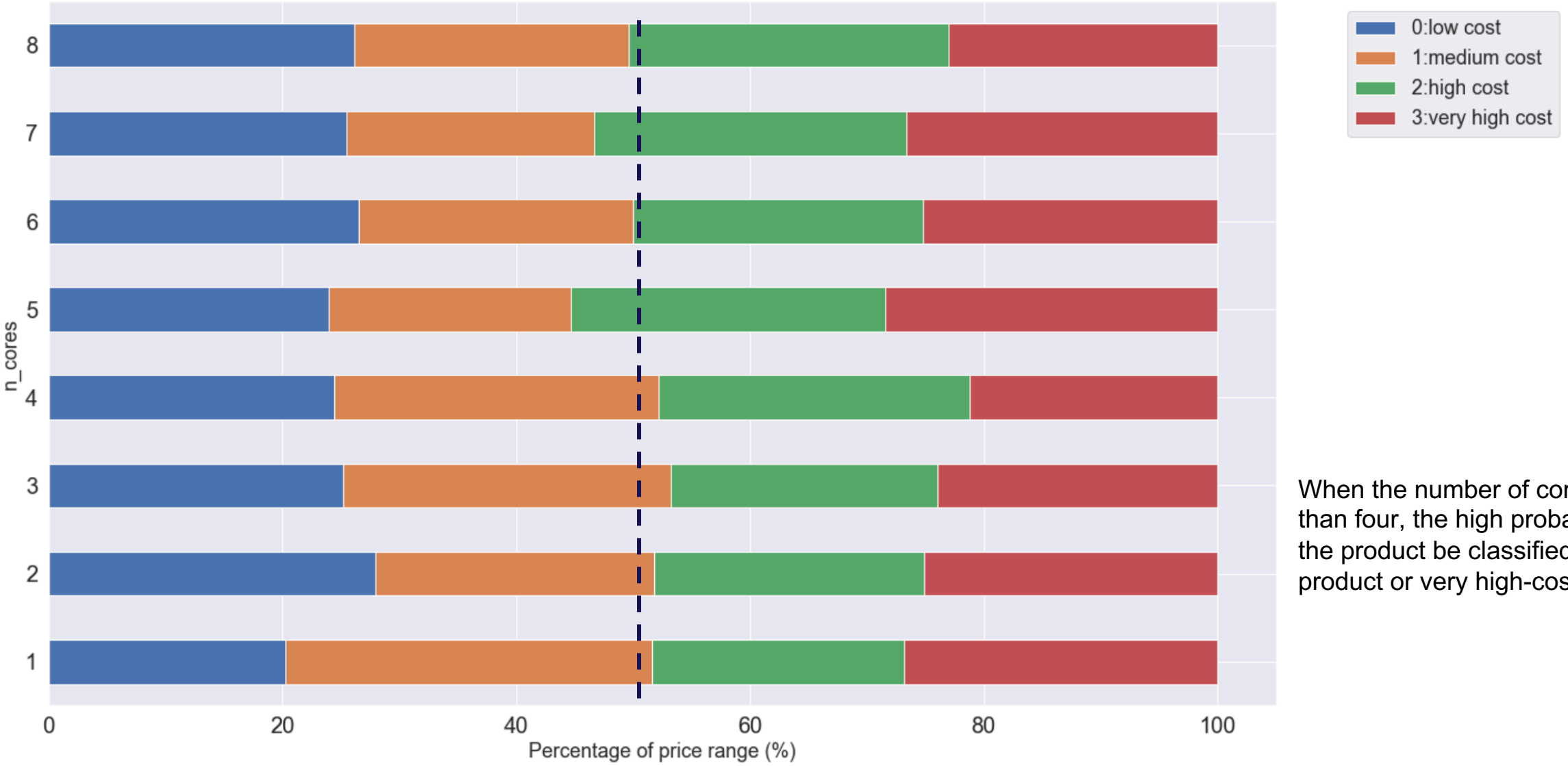
It is clear that if the mobile phone includes more volume of memory, it needs much money to purchase.

Rule 4-
In high volume of ram, supporting dual sim card, the product may increase the price.



The mobile phone that supporting dual sim card easily be classified as very high-cost product when the ram is higher than 3,000.

Rule 5-
More number of cores, the higher price for the mobile phone.



When the number of cores more than four, the high probability that the product be classified as high product or very high-cost label.

Classification Model

04

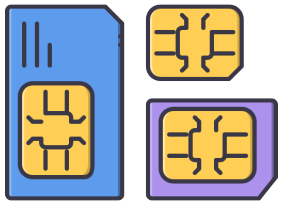
Regression Model

Predictors

Battery Power

Blue tooth

Clock speed



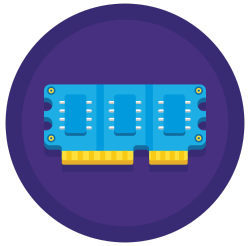
Dual SIM card

Pixels of
Front camera

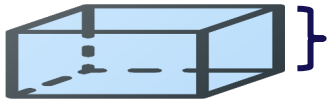
Regression Model

Predictors

4G

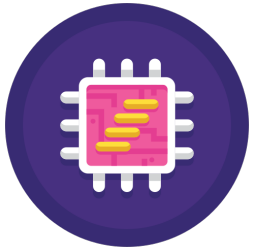


Volume of memory



Mobile depth

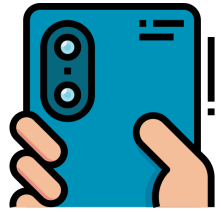
Mobile weight



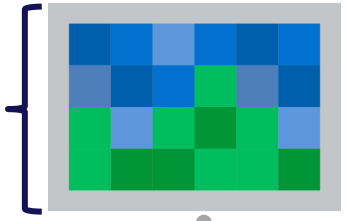
Number of cores

Regression Model

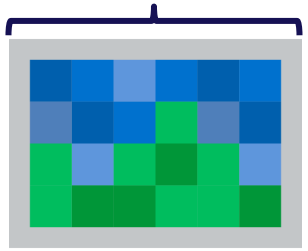
Predictors



Pixels of
Primary camera



Pixels of height



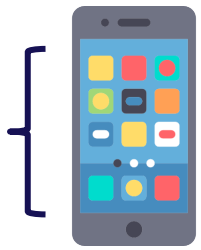
Pixels of width

Volume of RAM

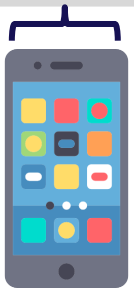
The lonest time for
calling

Regression Model

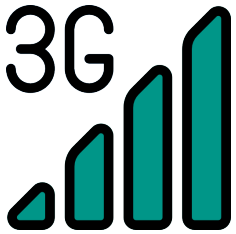
Predictors



Screen height

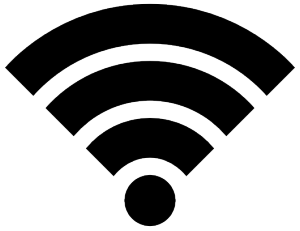


Screen weight



3G

Touch screen



WIFI

Classification Model

Response –Price range

01 DecisionTreeClassifier (Decision Tree)

02 SVC (Support Vector Machines)

03 KNeighborsClassifier (Nearest Neighbors)

04 LogisticRegression

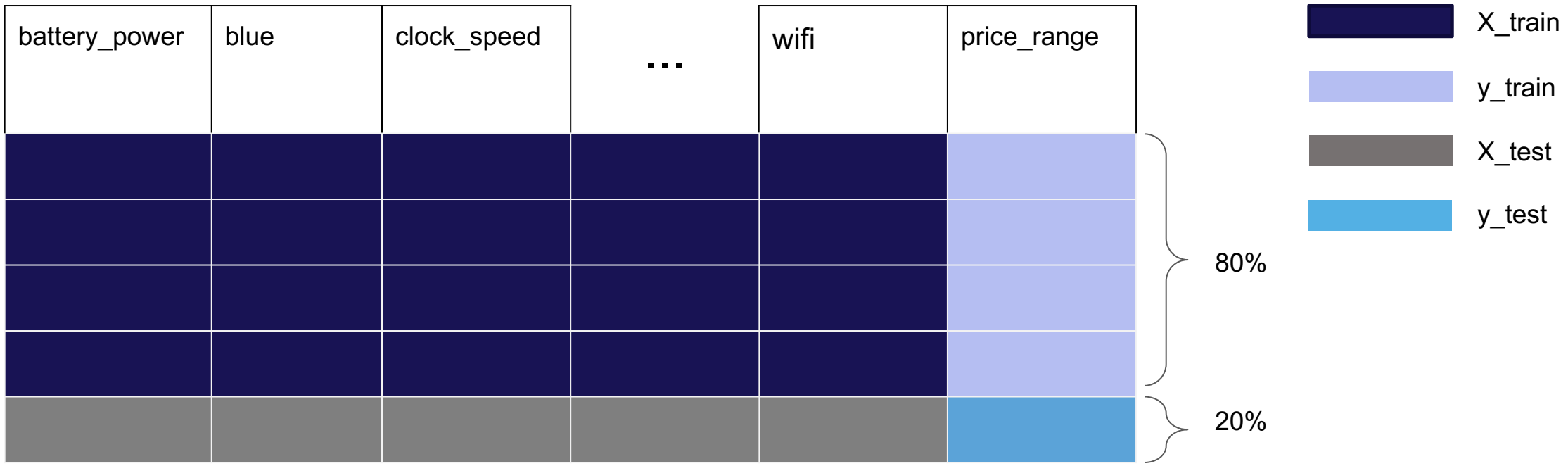
05 RandomForestClassifier (Ensemble Learning)

06 XGBClassifier (Ensemble Learning)

Classification Model

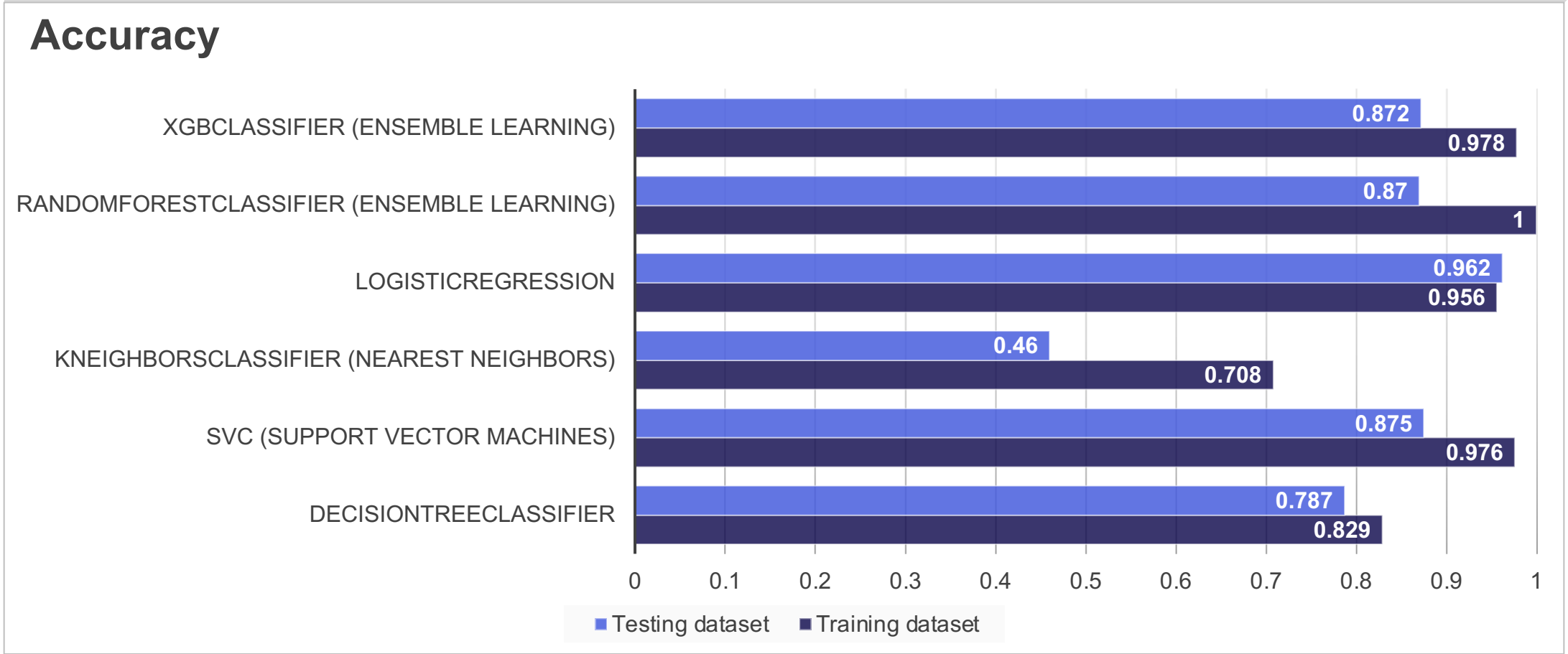
Split the dataset

Training dataset : 80%, Testing dataset : 20%.



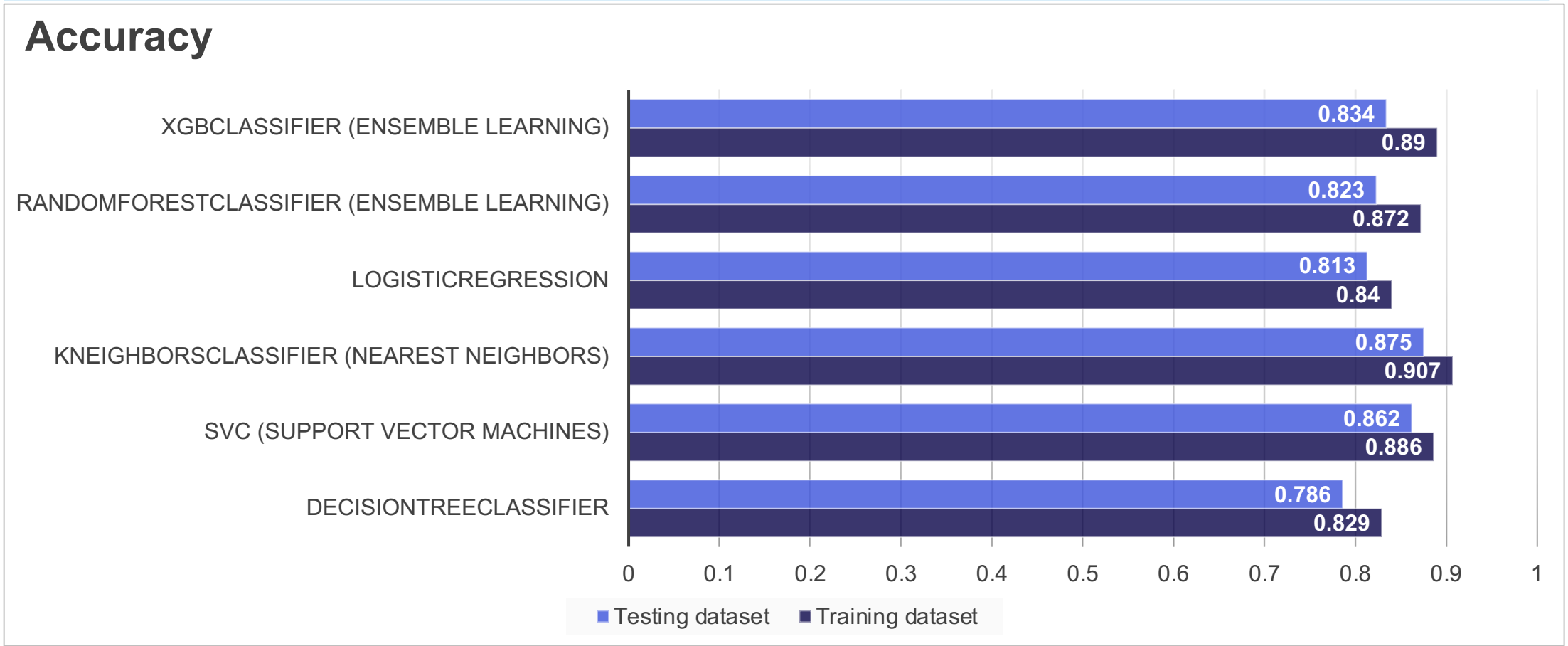
Classification Model

Accuracy



Regression Model-Cross validation

Accuracy



Conclusion

05

Conclusion

Flow Chart

What issue?

We cannot know how much budget we should prepare.

Import recent dataset

Visualizing data

Find some interesting features in our dataset.

Predicting price range

Accuracy 87.5%

We have 87.5% of the observed variation can be explained in this model.

Present on reality

Measuring the budget

How much money to buy a mobile phone.

Reference: https://www.flaticon.com

Thanks for your watching and listening.

CHIA-HAO,HSU