

zenius

Kampus  
Merdeka  
INDONESIA JAYA

# Final Project Presentation

Nomor Kelompok: 5

Nama Mentor: Diajeng Ciptaning Ayu

Nama:

- Arnold Julian
- Kayliev Nabil Auza

Machine Learning Class

Program Studi Independen Bersertifikat  
Zenius Bersama Kampus Merdeka



# Petunjuk

- Waktu presentasi adalah 5 menit (tentatif, tergantung dari banyaknya kelompok yang mendaftarkan diri)
- Waktu tanya jawab adalah 5 menit
- Silakan menambahkan gambar/visualisasi pada slide presentasi
- Upayakan agar tetap dalam format poin-poin (ingat, ini presentasi, bukan esai)
- Jangan masukkan *code* ke dalam slide presentasi (tidak usah memasukan screenshot jupyter notebook)

1. Latar Belakang
2. Explorasi Data dan Visualisasi
3. Modelling
4. Kesimpulan

# Latar Belakang

# Latar Belakang Project

Sumber Data:

<https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>

Problem: **Classification**

Tujuan:

- Memprediksi pengaruh harga smartphone terhadap spesifikasi
- Mencari tahu model algoritma untuk menentukan akurasi terhadap harga smartphone
- Menganalisis Range harga dari spesifikasi smartphone

# Explorasi Data dan Visualisasi

# Business Understanding

Smartphone sudah menjadi bagian penting dalam hidup, selain sebagai alat komunikasi bisa juga membantu kegiatan sehari hari.

Sebuah pengusaha ingin membuat perusahaan smartphone nya sendiri. Pengusaha tersebut tidak tau bagaimana caranya membuat estimasi harganya. Karena itu, dia perlu mengumpulkan data-data spesifikasi dan mengelompokan nya. Dia ingin mengetahui apakah ada korelasi dari macam-macam spesifikasi dan jangkauan harganya. Harga smartphone juga dipengaruhi oleh spesifikasinya.



# Data Cleansing

```
battery_power    0
blue             0
clock_speed      0
dual_sim         0
fc              0
four_g          0
int_memory       0
m_dep           0
mobile_wt        0
n_cores          0
pc               0
px_height        0
px_width         0
ram              0
sc_h             0
sc_w             0
talk_time        0
three_g          0
touch_screen     0
wifi             0
price_range      0
dtype: int64
```

Data tersebut tidak perlu dibersihkan karena dari awal data sudah bersih, sehingga tidak ada baris atau kolom yang missing/null

```
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   int64
7   m_dep                2000 non-null   float64
8   mobile_wt            2000 non-null   int64
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   int64
13  ram                  2000 non-null   int64
14  sc_h                 2000 non-null   int64
15  sc_w                 2000 non-null   int64
16  talk_time            2000 non-null   int64
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(2), int64(19)
memory usage: 328.2 KB
```



# Exploratory Data Analysis

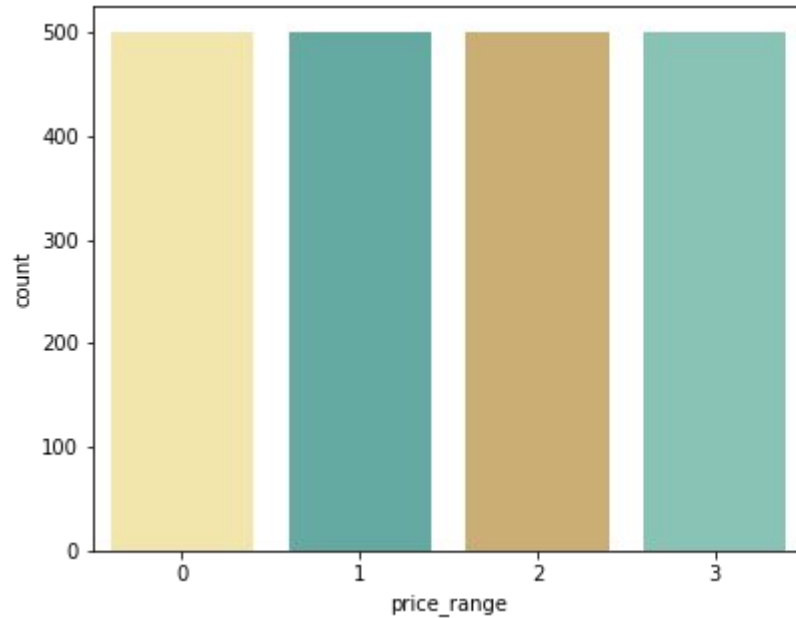
	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	...	px_height	px_width	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi	price_range
0	842	0	2.2	0	1	0	7	0.6	188	2	...	20	756	2549	9	7	19	0	0	1	1
1	1021	1	0.5	1	0	1	53	0.7	136	3	...	905	1988	2631	17	3	7	1	1	0	2
2	563	1	0.5	1	2	1	41	0.9	145	5	...	1263	1716	2603	11	2	9	1	1	0	2
3	615	1	2.5	0	0	0	10	0.8	131	6	...	1216	1786	2769	16	8	11	1	0	0	2
4	1821	1	1.2	0	13	1	44	0.6	141	2	...	1208	1212	1411	8	2	15	1	1	0	1

5 rows x 21 columns

```
Index(['battery_power', 'blue', 'clock_speed', 'dual_sim', 'fc', 'four_g',  
      'int_memory', 'm_dep', 'mobile_wt', 'n_cores', 'pc', 'px_height',  
      'px_width', 'ram', 'sc_h', 'sc_w', 'talk_time', 'three_g',  
      'touch_screen', 'wifi', 'price_range'],  
      dtype='object')
```

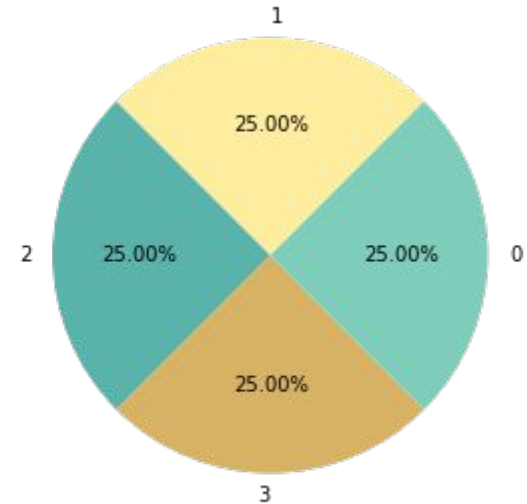
```
(2000, 21)
```

# Price Range Column

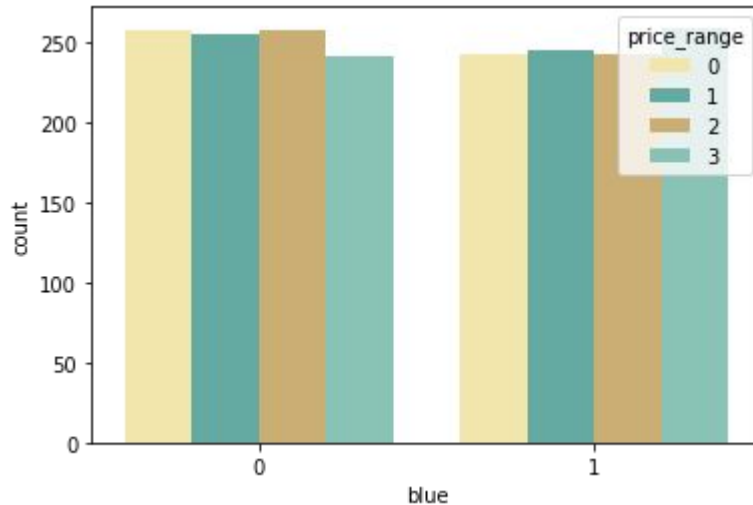


Keterangan :

- 0 = Low cost
- 1 = Medium cost
- 2 = High cost
- 3 = Very High cost



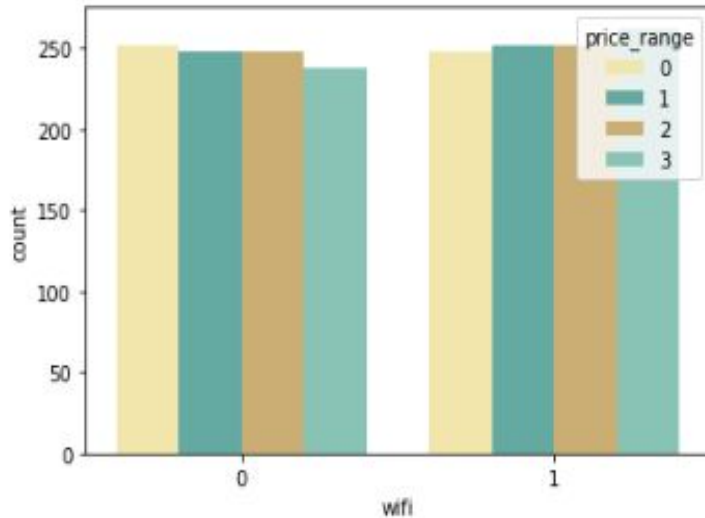
# Bluetooth



price_range	0	1	2	3
blue				
0	257	255	257	241
1	243	245	243	259

Terdapat perbandingan yang tidak jauh antara yang tidak memiliki bluetooth dan yang memiliki

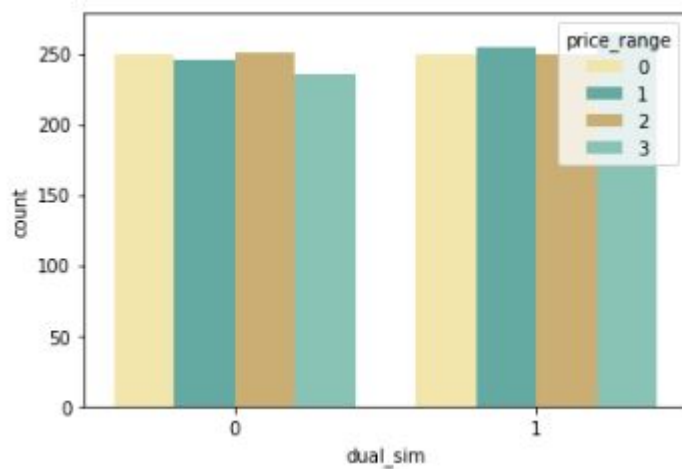
# WiFi



price_range	0	1	2	3
wifi				
0	252	248	248	238
1	248	252	252	262

Terdapat perbandingan yang tidak jauh antara yang memiliki wifi dan tidak

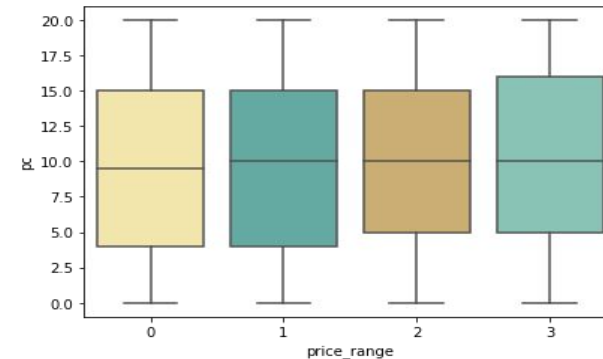
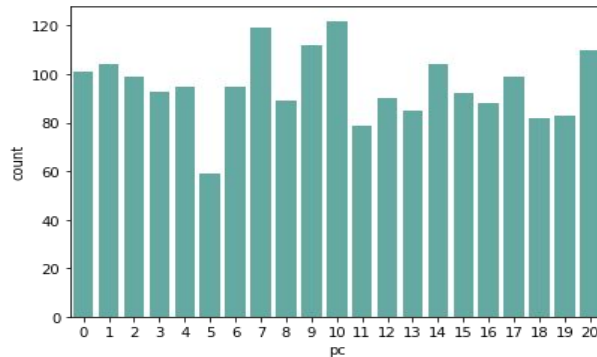
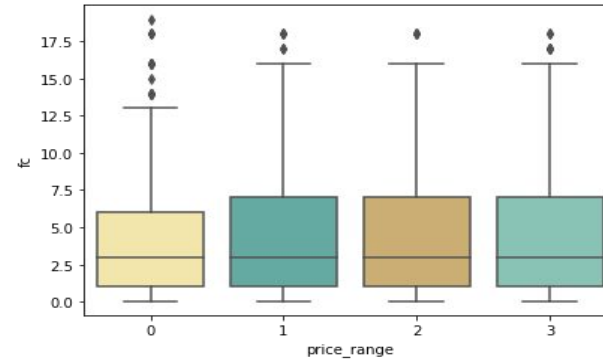
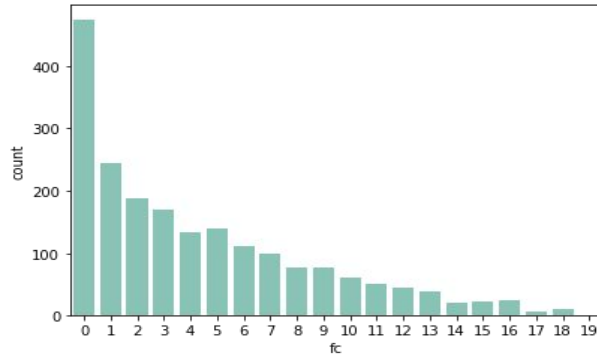
# Dual Sim



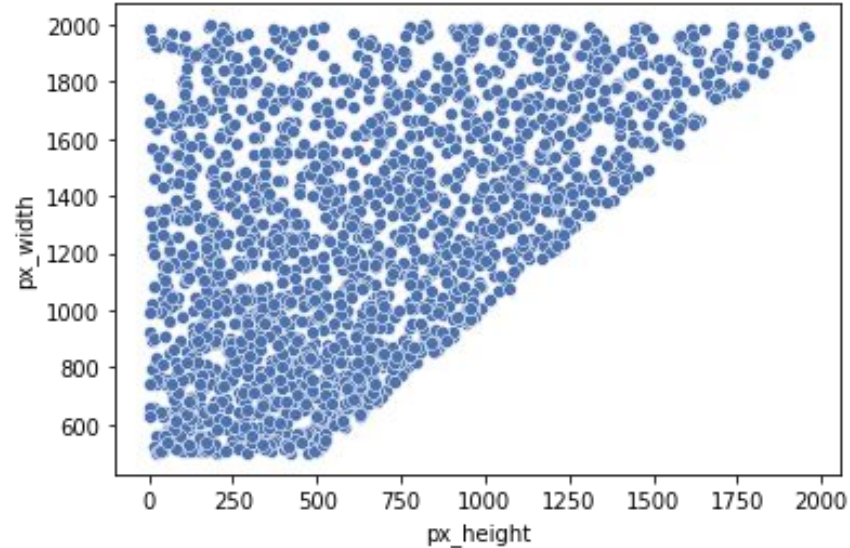
price_range	0	1	2	3
dual_sim				
0	250	245	251	235
1	250	255	249	265

Terdapat perbandingan yang tidak jauh antara yang memiliki dual sim dan tidak

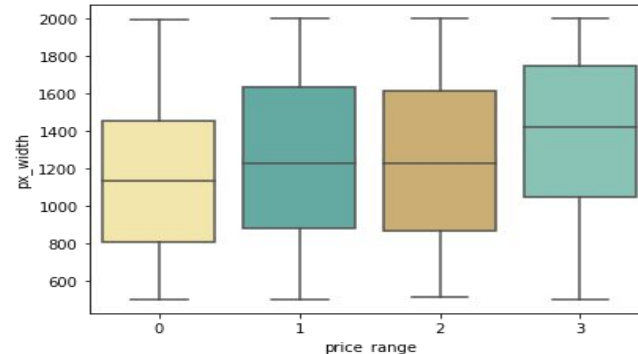
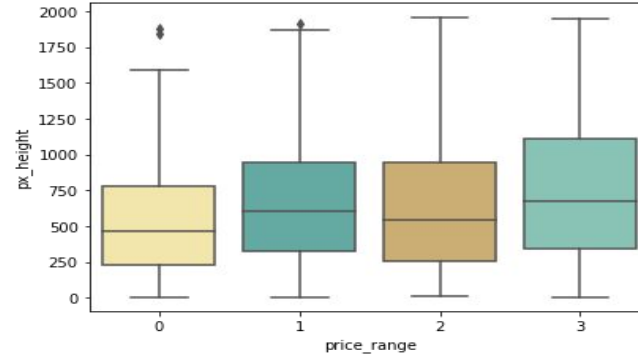
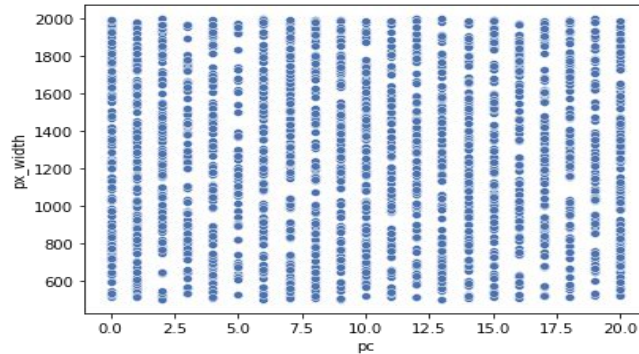
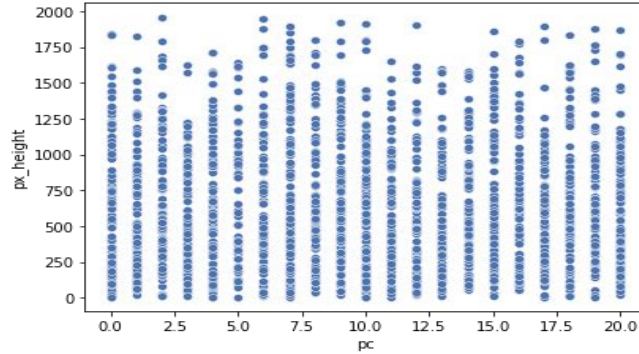
# Front and Primary Camera Column



# Px Height and Px Width

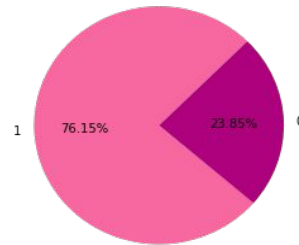
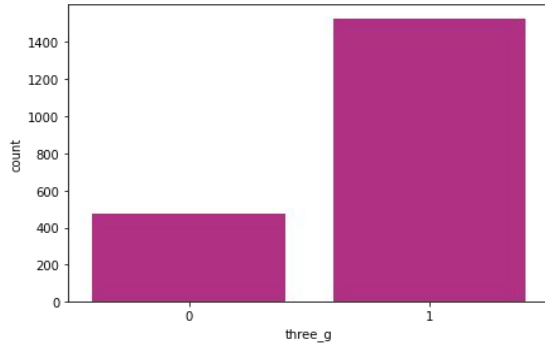


# Px Height and Px Width

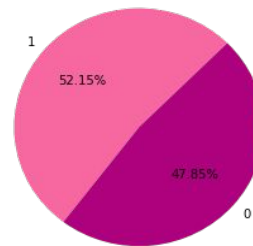
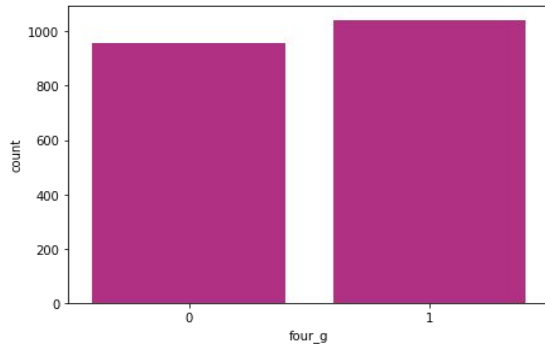




# 3G and 4G Column

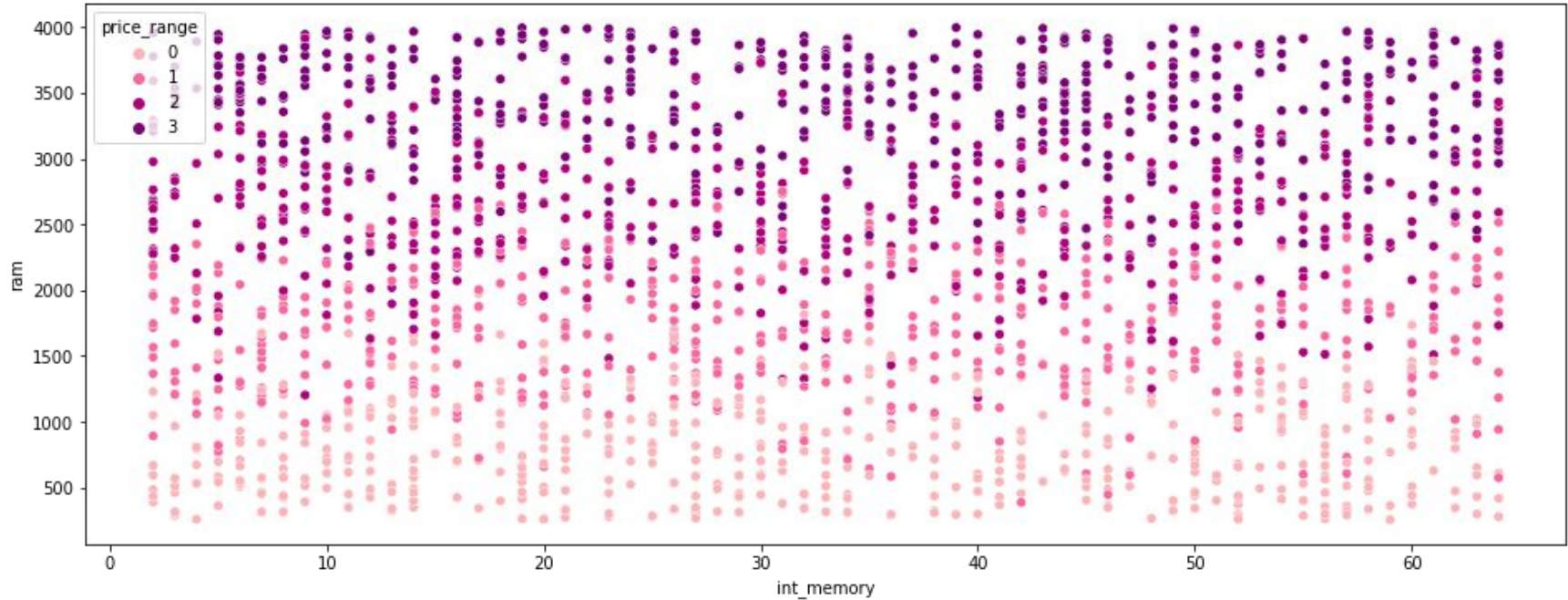


price_range	0	1	2	3
three_g				
0	127	122	113	115
1	373	378	387	385

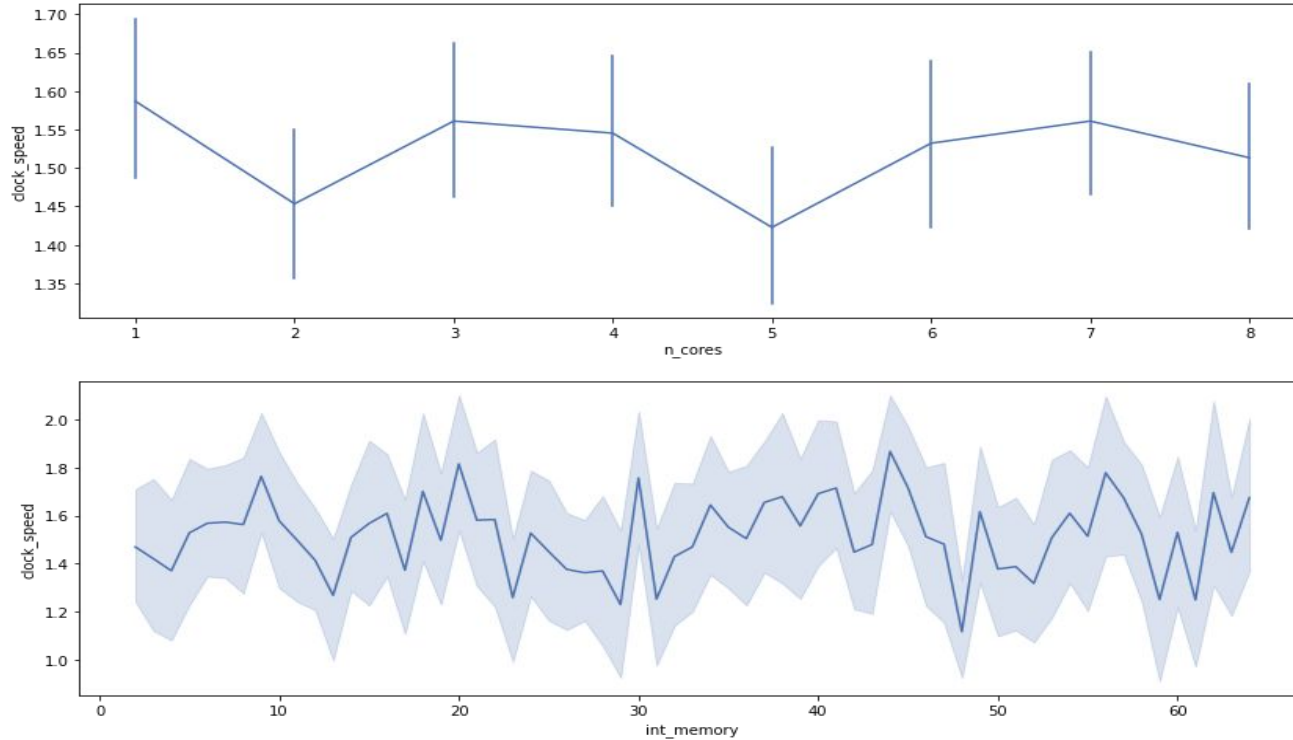


price_range	0	1	2	3
four_g				
0	241	238	253	225
1	259	262	247	275

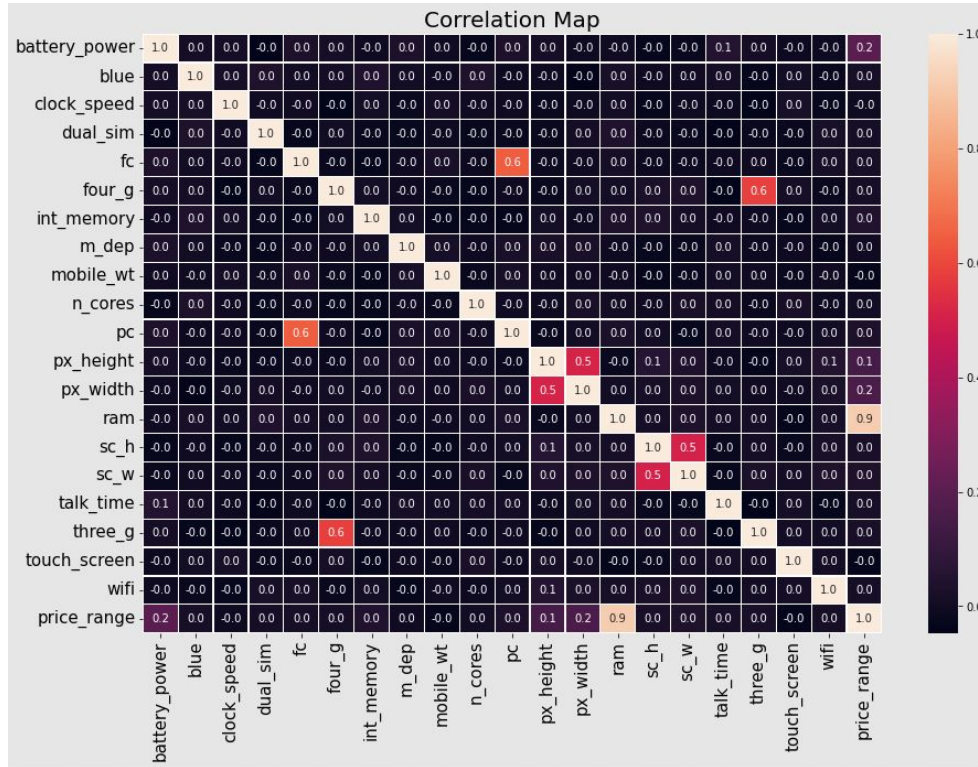
# Memory and RAM



# Clock Speed



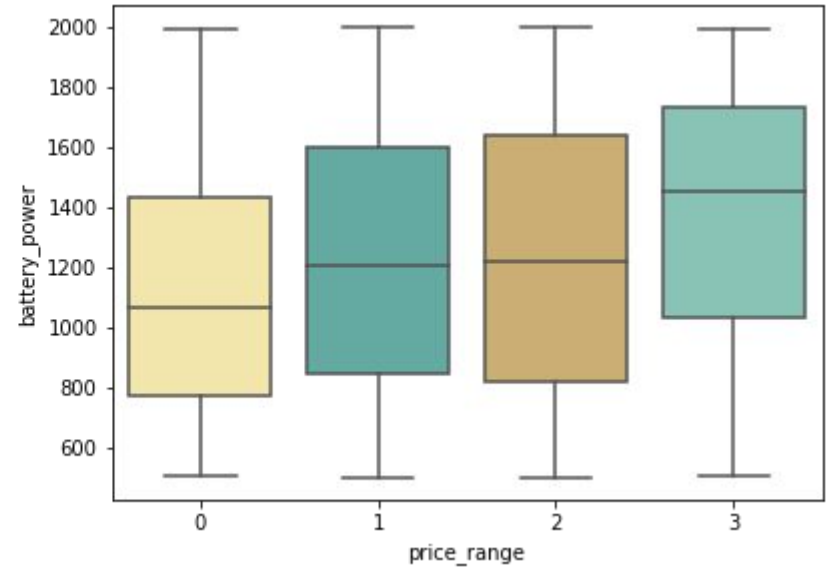
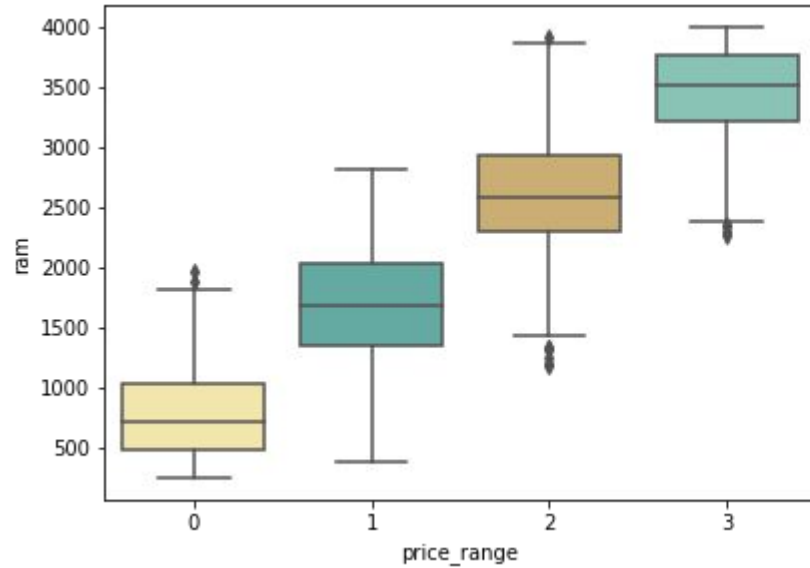
# Exploratory Data Analysis



## Heatmap

Terdapat korelasi paling kuat untuk price range dengan ram. Namun ada korelasi yang kurang kuat untuk baterai, px height dan width

# Price Range



# Modelling

# Preparation

- Membuat Variable x dan y dimana x = semua kolom kecuali price range dan y = price range
- Membagi variabel x dan y menjadi 4 ( x train, x test, y train, y test)
- Untuk pembagian datanya menggunakan train test split dengan data test 30%
- Mengecek variabel dari x dan y

# Model yang digunakan

- Logistic Regression
- Decision Tree
- Random Forest
- Random Forest (Tuning)
- K Nearest Neighbor



# Logistic Regression Model

- Membuat variabel  $l_r$  sebagai logistic regression
- Melakukan fit untuk variabel  $x$  train dan  $y$  train
- Melakukan tes pada model dan didapatkan hasil 0.6283

# Decision Tree Model

- Membuat variabel dtree untuk decision tree classifier
- Melakukan fit untuk variabel x train dan y train
- Melakukan tes pada model dan didapatkan hasil 0.86

# Random Forest Model

- Membuat variabel rfc sebagai Random Forest Classifier
- Melakukan fit untuk variabel x train dan y train
- Melakukan tes pada model dan didapatkan hasil 0.873

# K Nearest Neighbor

- Membuat variabel knn sebagai K Neighbor Classifier
- Melakukan fit untuk variabel x train dan y train
- Melakukan tes pada model dan didapatkan hasil 0.92

# Hyperparameter Tuning

- Membuat variabel rf sebagai random forest classifier tuning
- Menggunakan grid search cv untuk menaikkan model
- Membuat variabel grid search untuk membuat metode tuningnya dengan estimator 200
- Membuat parameter dengan max depth, min sample leaf dan n estimator
- Melakukan fit untuk variabel x train dan y train dan didapatkan score 0.859

# Evaluation

- Accuracy Score

```
Akurasi untuk Logistic Regression: 0.63
Akurasi untuk Decision Tree 0.86
Akurasi untuk Random Forest 0.87
Akurasi untuk Random Forest (Tuning) 0.84
Akurasi untuk KNeighbors 0.92
```

- Precision Score

```
Precision score untuk Logistic Regression: 0.63
Precision score untuk Decision Tree: 0.87
Precision score untuk Random Forest: 0.87
Precision score untuk Random Forest(Tuning): 0.85
Precision score untuk KNeighbors: 0.92
```

- Recall Score

```
Recall score untuk Logistic Regression: 0.64
Recall score untuk Decision Tree: 0.86
Recall score untuk Random Forest: 0.88
Recall score untuk Random Forest (Tuning): 0.85
Recall score untuk KNeighbors: 0.92
```

Berdasarkan hasil tes dari akurasi yang dilakukan, k neighbors memiliki akurasi yang paling besar yaitu 0.92

# Conclusion

# Kesimpulan

Dari Analisis tersebut, dapat disimpulkan bahwa ram memiliki faktor penting ketika melakukan pembelian untuk smartphone. Maka dari itu, hal yang perlu kita lihat untuk pembelian smartphone yaitu ram nya.

Tapi kita juga perlu melihat aspek dari baterai dan juga pixel kamera dari heatmap masih ada korelasi dengan price range.

Pastikan ketika ingin membeli smartphone, fitur smartphone nya sudah memiliki bluetooth, wifi, dan memiliki dual sim sehingga mudah untuk kita memiliki dua nomor serta memiliki penyimpanan yang banyak karena dapat menguntungkan pengguna ketika memiliki banyak aplikasi



# Terima kasih!

Ada pertanyaan?

zenius



Kampus  
Merdeka  
INDONESIA JAYA