### **Mobile Price Classification**



## Background of The Company and The Problem:

Bob has started his own mobile company. He wants to give tough fights to big companies like Apple, Samsung etc. He does not know how to estimate the price of mobiles his company creates. In this competitive mobile phone market you cannot simply assume things. To solve this problem he collects sales data of mobile phones of various companies. Bob wants to find out some relation between features of a mobile phone(eg:- RAM, Internal Memory etc) and its selling price.

Dataset Website: <a href="https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification/data">https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification/data</a>



# Data Wrangling

By displaying the data information, we can see that there are too many individual units and too many features in the data. We have done some data cleaning to make the features in our data look clearer, thus clarifying our judgment.

<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	int64
7	m_dep	2000 non-null	float64
8	mobile_wt	2000 non-null	int64
9	n_cores	2000 non-null	int64
10	рс	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
dtyp	es: float64(2),	int64(19)	

memory usage: 328.2 KB

# Data Cleaning

After we did the data cleaning we found 0 NaN values in the dataframe. Very good, the data looks much clearer, which is very helpful for the EDA process.

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
рс	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	

## **EDA**

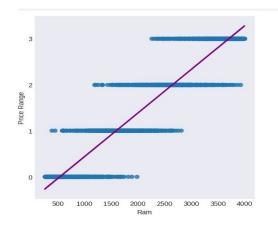
battery_power		0.011	0.011	-0.042	0.033	0.016	-0.004	0.034	0.0018	-0.03	0.031	0.015	-0.0084	-0.00065	-0.03	-0.021	0.053	0.012	-0.011	-0.0083	0.2
blue	0.011		0.021	0.035	0.0036	0.013	0.041	0.004	-0.0086	0.036	-0.01	-0.0069	-0.042	0.026	-0.003	0.00061	0.014	-0.03	0.01	-0.022	0.021
dock_speed	0.011	0.021		-0.0013	-0.00043	-0.043	0.0065	0.014	0.012	-0.0057	-0.0052	-0.015	-0.0095	0.0034	-0.029	-0.0074	-0.011	-0.046	0.02	-0.024	-0.0066
dual_sim	0.042	0.035	-0.0013		-0.029	0.0032	0.016	0.022	0.009	0.025	-0.017	0.021	0.014	0.041	0.012	0.017	0.039	0.014	0.017	0.023	0.017
fc	0.033	0.0036	-0.00043	-0.029		-0.017	-0.029	-0.0018	0.024	-0.013	0.64	-0.01	-0.0052	0.015	-0.011	-0.012	-0.0068	0.0018	-0.015	0.02	0.022
four_g	0.016	0.013	0.043	0.0032	0.017		0.0087	0.0018	0.017	0.03	0.0056	0.019	0.0074	0.0073	0.027	0.037	0.047		0.017	0.018	0.015
int_memory	-0.004	0.041	0.0065	-0.016	-0.029	0.0087		0.0069	-0.034	-0.028	-0.033	0.01	-0.0083	0.033	0.038	0.012	-0.0028	0.0094	-0.027	0.007	0.044
m_dep	0.034	0.004	-0.014	-0.022	-0.0018	-0.0018	0.0069		0.022	-0.0035	0.026	0.025	0.024	-0.0094	-0.025	-0.018	0.017	-0.012	-0.0026	-0.028	0.00085
mobile_wt	0.0018	-0.0086	0.012	-0.009	0.024	-0.017	-0.034	0.022		-0.019	0.019	0.00094	9e-05	-0.0026	-0.034	-0.021	0.0062	0.0016	-0.014	-0.00041	-0.03
n_cores	-0.03	0.036	-0.0057	-0.025	-0.013	-0.03	-0.028	-0.0035	-0.019		-0.0012	-0.0069	0.024	0.0049	-0.00031	0.026	0.013	-0.015	0.024	-0.01	0.0044
pc	0.031	-0.01	-0.0052	0.017	0.64	0.0056	0.033	0.026	0.019	0.0012		0.018	0.0042	0.029	0.0049	-0.024	0.015	0.0013	-0.0087	0.0054	0.034
px_height	0.015	-0.0069	-0.015	-0.021	-0.01	-0.019	0.01	0.025	0.00094	-0.0069	-0.018		0.51	-0.02	0.06	0.043	-0.011	-0.031	0.022	0.052	0.15
px_width	-0.0084	-0.042	-0.0095	0.014	-0.0052	0.0074	-0.0083	0.024	9e-05	0.024	0.0042	0.51		0.0041	0.022	0.035	0.0087	0.00035	-0.0016	0.03	0.17
ram	-0.00065	0.026	0.0034	0.041	0.015	0.0073	0.033	-0.0094	-0.0026	0.0049	0.029	-0.02	0.0041		0.016	0.036	0.011	0.016	-0.03	0.023	0.92
sc_h	-0.03	-0.003	-0.029	-0.012	-0.011	0.027	0.038	-0.025	-0.034	-0.00031	0.0049	0.06	0.022	0.016			-0.017	0.012	-0.02	0.026	0.023
sc_w	-0.021	0.00061	-0.0074	-0.017	-0.012	0.037	0.012	-0.018	-0.021	0.026	-0.024	0.043	0.035	0.036	0.51		-0.023	0.031	0.013	0.035	0.039
talk_time	0.053	0.014	-0.011	-0.039	-0.0068	-0.047	-0.0028	0.017	0.0062	0.013	0.015	-0.011	0.0067	0.011	-0.017	-0.023		-0.043	0.017	-0.03	0.022
three_g	0.012	-0.03	0.046	0.014	0.0018		0.0094	0.012	0.0016	-0.015	0.0013	0.031	0.00035	0.016	0.012	0.031	0.043		0.014	0.0043	0.024
touch_screen	-0.011	0.01	0.02	0.017	-0.015	0.017	0.027	-0.0026	0.014	0.024	0.0087	0.022	-0.0016	-0.03	0.02	0.013	0.017	0.014		0.012	-0.03
vafi	-0.0083	-0.022	-0.024	0.023	0.02	-0.018	0.007	-0.028	-0.00041	-0.01	0.0054	0.052	0.03	0.023	0.026	0.035	-0.03	0.0043	0.012		0.019
price_range	0.2	0.021	-0.0066	0.017	0.022	0.015	0.044	0.00085	-0.03	0.0044	0.034	0.15	0.17	0.92	0.023	0.039	0.022	0.024	-0.03	0.019	
	battery_power	pine	dock_speed	dual_sim	Ψ	lour_g	int memory	дар ш	mobile_wt	sauco u	8	px_height	px_width	ram	8	8	talk time	three_g	touch screen	Me	price_range

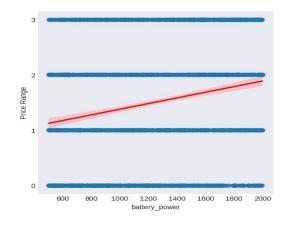
ram	0.917046	
battery_power	0.200723	
px_width	0.165818	
px_height	0.148858	
int_memory	0.044435	
sc_w	0.038711	
рс	0.033599	
touch_screen	0.030411	
mobile_wt	0.030302	
three_g	0.023611	
sc_h	0.022986	
fc	0.021998	
talk_time	0.021859	
blue	0.020573	
wifi	0.018785	
dual_sim	0.017444	
four_g	0.014772	
clock_speed	0.006606	
n_cores	0.004399	
m_dep	0.000853	
price_range	NaN	
Name: price_ran	ge, dtype:	float64

#### **EDA**

Used a heat map to show the highest correlation between Price\_range and other features in the dataset. We found that the correlation between "price\_range" and "ram" is very high, which means that in future modeling we may need to use the important feature of "ram" to predict the price range of mobile phones. Moreover, battery power also has a high correlation with our target variable.

#### **EDA** and Some Data Visualization

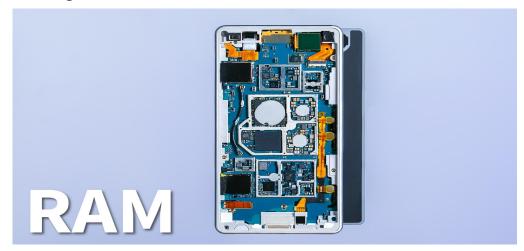




After we identified the two features that most affect the price range (battery power and ram), we made some comparisons using bar charts and found that the increasing trend of ram is stronger than that of battery power.

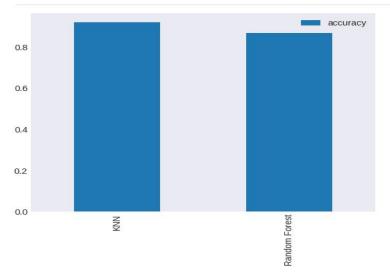
#### **EDA**

There is an increasing relationship between RAM and mobile phone prices. Let's also see some relations between price range and battery power. We can see that although there is an increasing relationship between battery power and price range, it does not have as big an impact as ram on the price range. This strengthens our reason for choosing the ram feature for modeling.



## Pre-Processing and Training Data

Two modeling methods, KNN and random forest regression, are established and compared. After comparing the data, we found that the KNN modeling method has more accuracy, so KNN is a more suitable modeling method.



## Modeling and Hyperparameter Method

Model accuracy comparison cannot produce a model with a final high score. We used the Hyperparameter Method to ensure that KNN is the final model we want to choose.

#### 1. Hyperparemeter tuning for Random Forest:

#### 2. Hyperparemeter tuning KNN:

## Modeling and Hyperparameter Method

The outlook from last silde pretty much says that the KNN model is correct to use the best model with the best hyperparameters. After we determined the best model, we did the model evaluation. We used two model evaluation methods: classification report and Cross Validation.

	precision	recall	f1-score	support	
0	0.96	0.98	0.97	178	Cross Validation Scores: [0.94  0.935  0.9425  0.935  0.9175]
1	0.91	0.90	0.90	163	
2	0.87	0.88	0.88	161	1 011/10 01111 0 70 1 1 07/
3	0.95	0.93	0.94	158	print(f'Cross Validation Score (Mean): ' + str(np.mean(cross_val_score(KNN, X, y, cv=5)))
accuracy			0.92	660	Cross Validation Score (Mean): 0.93399999999999
macro avg	0.92	0.92	0.92	660	
weighted avg	0.92	0.92	0.92	660	

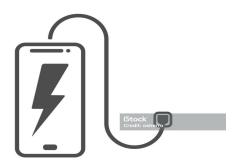
Through the cross validation method, it is not difficult to see that similar scores can be maintained by performing data segmentation in time and even increasing some accuracy.

#### Conclusion

The important factors that affect the price range of mobile phones are RAM and battery power. The important factors that affect the price range of mobile phones are RAM and battery power. The company can use these two pieces of information to make changes in the production of mobile phones to compete with Major mobile phone brands.



**VS** 



### Advice on business:

- 1. While maintaining almost the same price range and the same phone conditions, appropriately increase the phone RAM. Because in our opinion, mobile phone RAM is an important factor that affects the price range, and companies need to pay attention to it, so that they can compete with big-brand mobile phone companies.
- 2. If the first method does not achieve good results, we can also try to improve the battery performance while also increasing the RAM, because battery performance, as a factor second only to the mobile phone RAM, will also affect pricing.

