452 Final Project|Classification of Phone Price Range|Beverly Huang

https://colab.research.google.com/drive/1PuqTA7We_562ZHe1lXoCDGXwWn139EWt

Motivation and Goal

Pricing is the most powerful P&L lever for the retailer. No matter you are creating a new product or starting a new business, you need to think about how to price your product smartly. Especially in a competitive market, pricing is a tool to differentiate yourself from your competitors and get the market share. To price your product, your first step is to understand competitors' pricing strategies. Thus, this project will use mobile phone market as an example to illustrate the method we can use for pricing.

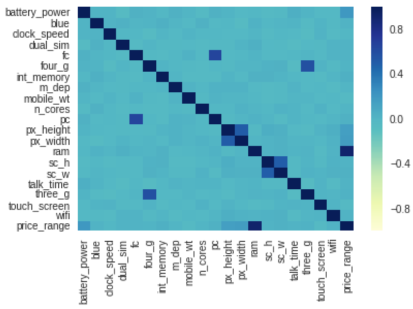
Ingestion: Import Libraries and Load Data

The dataset we have contains data on different mobile phone features such as battery power, internal memory, etc. and its selling price. Data are in the format below.

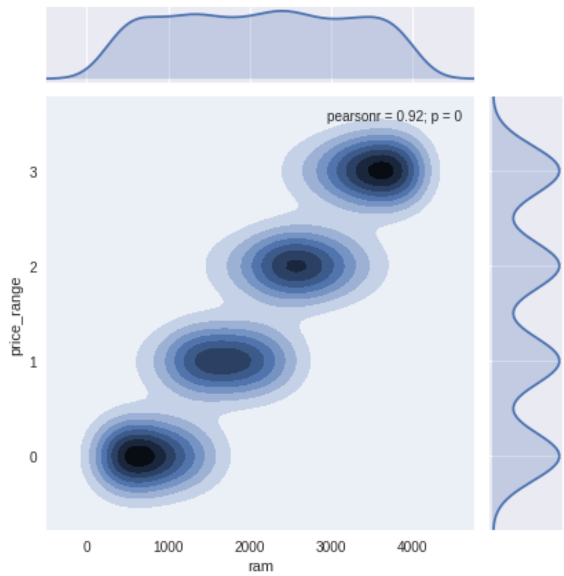
	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	ram	sc_h	sc_w	talk_time	three_g	touch_screen	wifi	price_range
0	842	0	2.2	0	1	0	7	2549	9	7	19	0	0	1	1
1	1021	1	0.5	1	0	1	53	2631	17	3	7	1	1	0	2
2	563	1	0.5	1	2	1	41	2603	11	2	9	1	1	0	2
3	615	1	2.5	0	0	0	10	2769	16	8	11	1	0	0	2
4	1821	1	1.2	0	13	1	44	1411	8	2	15	1	1	0	1

EDA: Exploratory Data Analysis

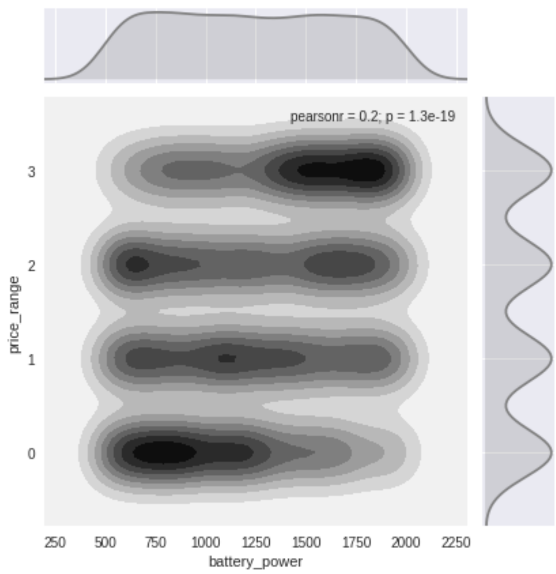
To understand the relationship between the features of a mobile phone and its selling price, we use Seaborn Correlation Heatmap to visualize correlations between all variables. From the Heatmap, we can see that ram, battery power, internal memory and dual sim have higher correlation with selling price than any other features.



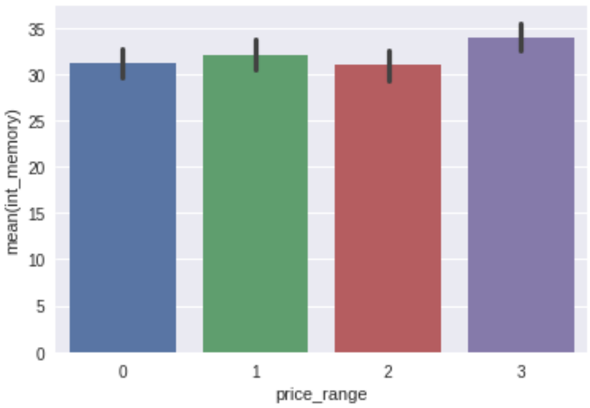
Thus, we take out these four features and use Seaborn plots to look at their relationships with selling price. From the four graphs below, we can see that a mobile phone with more ram, battery power, internal memory and dual sim is more likely to charge higher price.



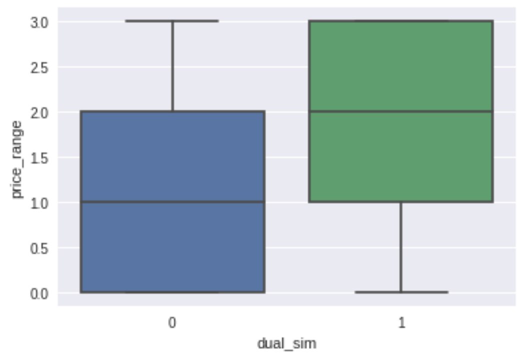
Graph1: RAM vs Price



Graph2: Battery Power vs Price



Graph3: Internal Memory vs Price



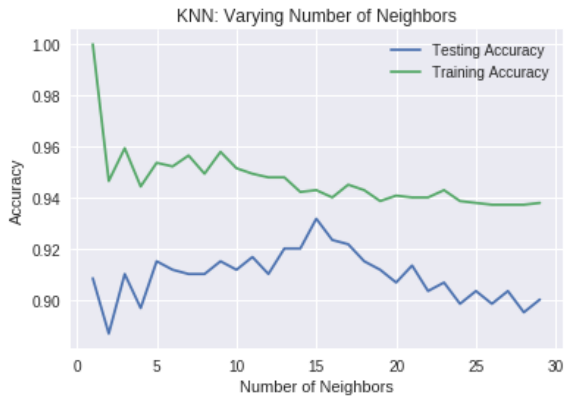
Graph4: Dual Sim vs Price

Modeling: KNN

To capture all relationships, we build a KNN classification model. But the key issue is what K (number of nearest neighbors) should we choose? To solve this problem, we use two methods.

- Method 1

We calculate the accuracy of our model under different K values and generate a plot. From the plot, we can see that when K = 15, our model has the highest accuracy.



- Method 2

Another method we use is Grid Search cross validation. This method will try a bunch of K values, fit all of them separately and choose the best performing one. This method also gives me the same answer: K = 15.

- Use the best K to predict

We apply K = 15 to predict test data, and accuracy is **93.16%**. The classification report is generated below. We can see that both precision and recall are very high. Thus, we are confident that this is a good model.

	precision	recall	f1-score	support
0	0.95	0.97	0.96	150
1	0.92	0.91	0.91	150
2	0.90	0.91	0.91	150
3	0.96	0.94	0.95	150
micro avg	0.93	0.93	0.93	600
macro avg	0.93	0.93	0.93	600
weighted avg	0.93	0.93	0.93	600

Conclusion for Management

This project uses mobile phone market as an example to illustrate how to use competitors' pricing information to build a pricing model for your products. A KNN model is built based on the best K we find from both testing accuracy and GridSearch cross validation result. This model achieves 93.17% accuracy. With this model, the sales team can price mobile phones based on their features easily. This pricing model can also be implemented in other products and industries. It is a pricing model that incorporates both value-based pricing (which considers product value) and competitive pricing (which considers competitors' pricing). Thus, it is a good model for new product pricing.