Analysing the Factors contributing to the Profit and Loss of a Supermarket Sales.

-Aravindhan Babu
M.sc in Digital and Big data for Value
Essca School of Management

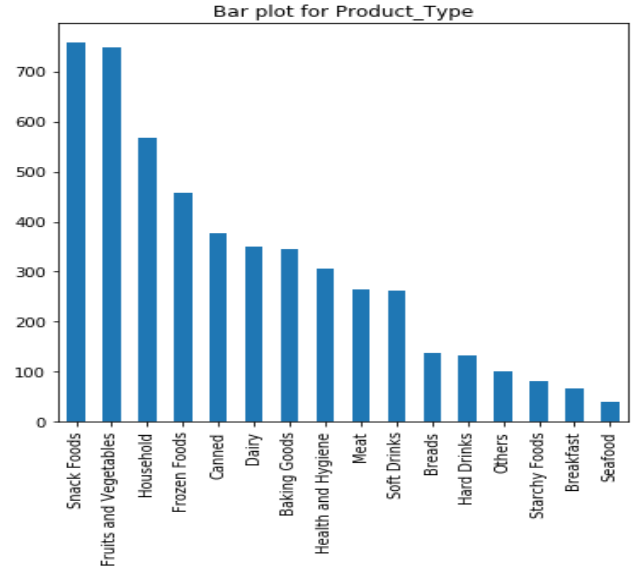
Data analytics takes advantage of the valuable data that customers produce when they purchase products, whether offline or on-line.

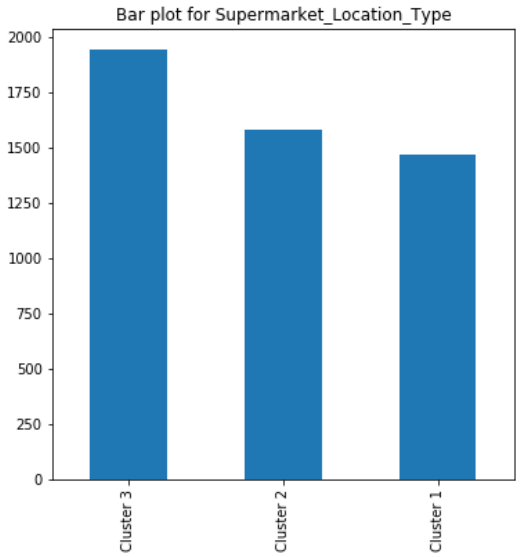
Understanding their purchasing patterns and buying behaviours, helps retailers to develop tailored marketing campaigns and deals that cater to exactly when consumers want, attracting them back to their stores for potential purchases.

The link to the data I have used for analysis is https://github.com/let-me-code-it/Dataset-to-analyse-supermarket-sales/blob/master/Supermarket%20Sales%20Data.csv

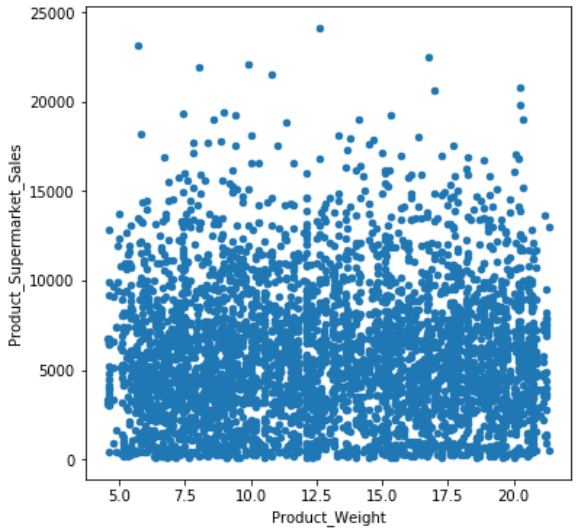
I have taken the columns such as Product weight, Product shelf visibility, Product price, Supermarket opening year, Product supermarket sales, Product Identifier, Product fat content, Supermarket Location Type, Supermarket type, Supermarket Identifier. I have used Python with seaborn library for the analysis.

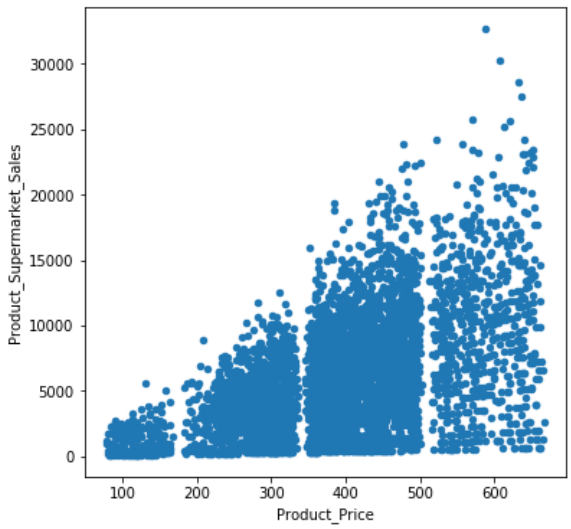
	Product_I	Supermar	Product_Supermarket_Identifier	Product_V	Product_Fat_Content	Product_S	Product_T	Product_P	Supermarket_Open	Supermarket _Size	Supermarket_Location_Type	Supermarket_Type	Pr
2	DRA12	CHUKWUI	DRA12_CHUKWUDI010	11.6	Low Fat	0.068535	Soft Drink	357.54	2005		Cluster 3	Grocery Store	
3	DRA12	CHUKWUI	DRA12_CHUKWUDI013	11.6	Low Fat	0.040912	Soft Drink	355.79	1994	High	Cluster 3	Supermarket Type1	
4	DRA12	CHUKWUI	DRA12_CHUKWUDI017	11.6	Low Fat	0.041178	Soft Drink	350.79	2014		Cluster 2	Supermarket Type1	
5	DRA12	CHUKWUI	DRA12_CHUKWUDI018	11.6	Low Fat	0.041113	Soft Drink	355.04	2016	Medium	Cluster 3	Supermarket Type2	
6	DRA12	CHUKWUI	DRA12_CHUKWUDI035	11.6	Ultra Low fat	0	Soft Drink	354.79	2011	Small	Cluster 2	Supermarket Type1	
7	DRA12	CHUKWUI	DRA12_CHUKWUDI045	11.6	Low Fat	0	Soft Drink	354.04	2009		Cluster 2	Supermarket Type1	
8	DRA24	CHUKWUI	DRA24_CHUKWUDI010	19.35	Normal Fat	0.066832	Soft Drink	409.72	2005		Cluster 3	Grocery Store	
9	DRA24	CHUKWUI	DRA24_CHUKWUDI013	19.35	Normal Fat	0.039895	Soft Drink	406.22	1994	High	Cluster 3	Supermarket Type1	
10	DRA24	CHUKWUI	DRA24_CHUKWUDI017	19.35	Normal Fat	0.040154	Soft Drink	411.72	2014		Cluster 2	Supermarket Type1	





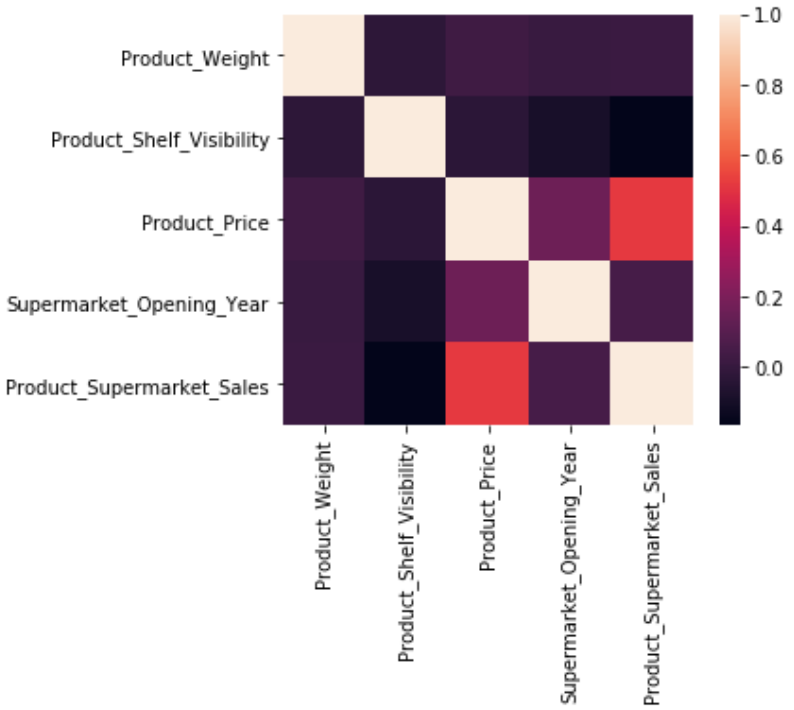
The three clusters of Location includes **Urban ,Rural and Isolated** respectively.





Heat Map :

Correlating the data to find how two different columns correlate with each other using the function pandas .corr( ) and creating a heat map using the function heatmap().



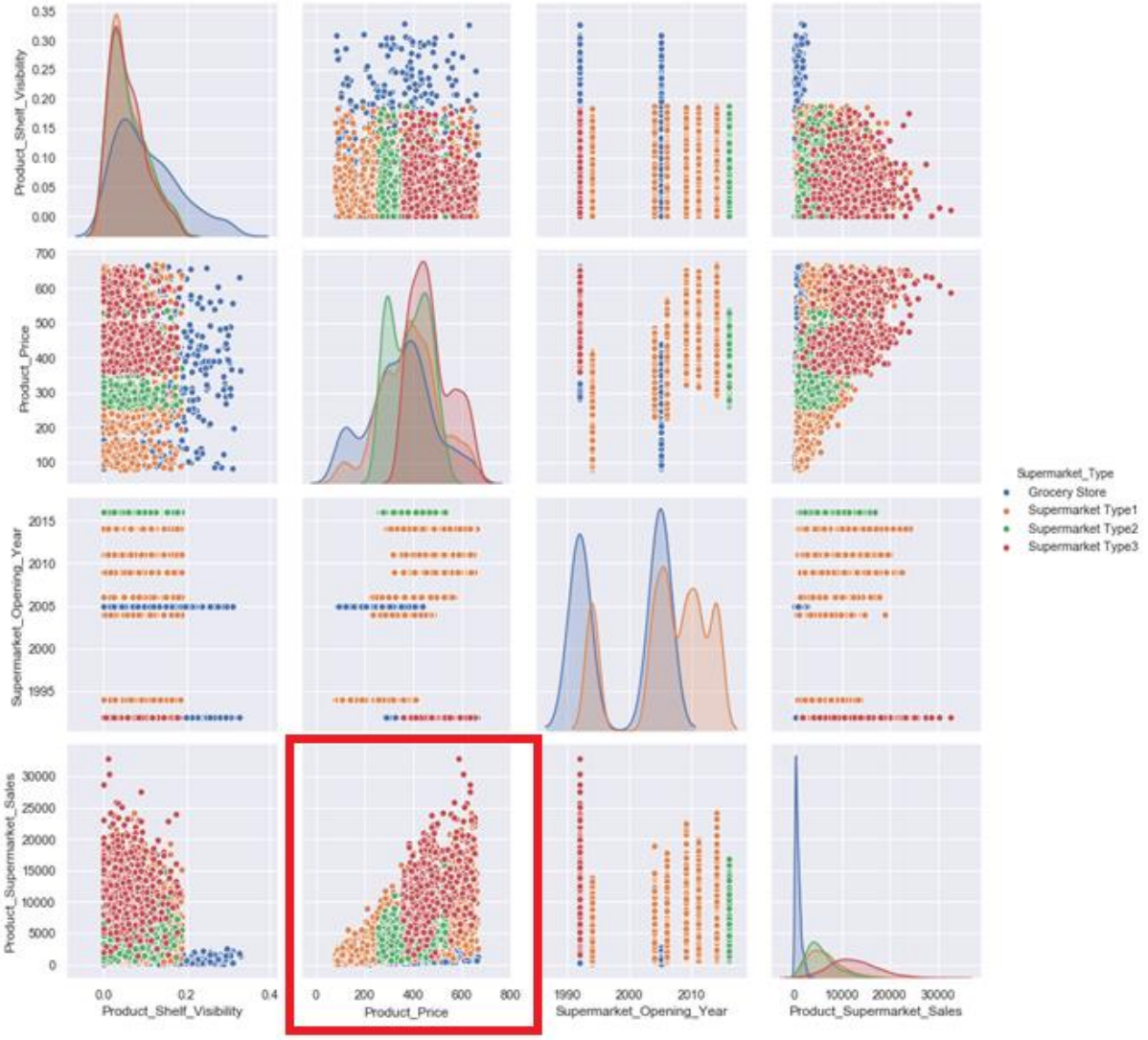
Every square in a heatmap shows how well it correlates with any two features. The Lighter the Square is, higher the correlation is.

Product Price is the most associated function with our Product Supermarket Sales and our hypothesis makes sense that if expensive items or hedonic goods (luxury goods) are sold in higher ratio, then the total product Supermarket Sales will increase and vice versa.

Supermarket Opening Year has also contributed to the price of the products, as an old and a new supermarket having different price ranges and offers.

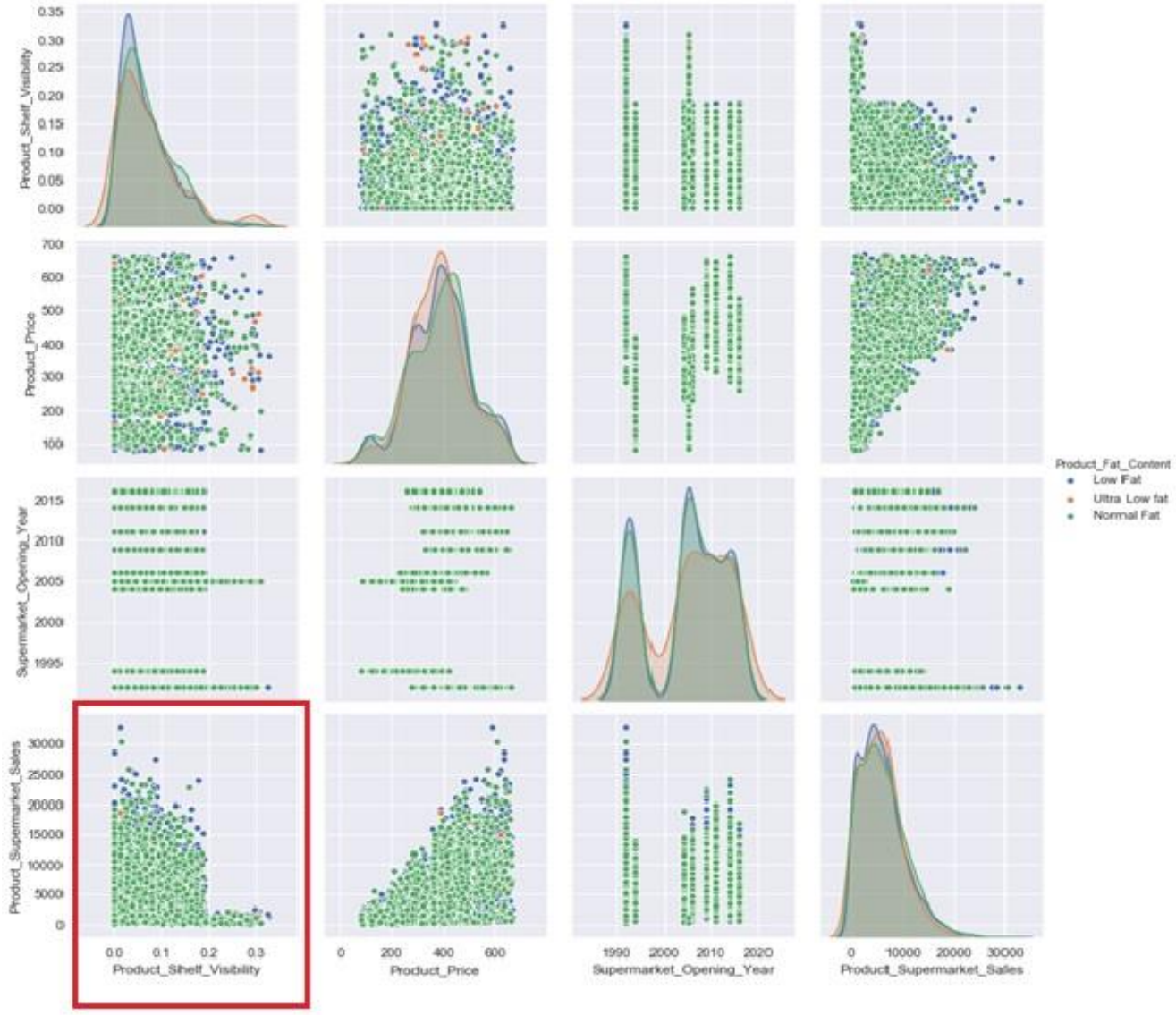
The relationship between different columns in our dataset can be found using Seaborn library.

Plot based on Supermarket Type.



First, we can see that the price of goods in Supermarket Type3 generally starts at around 350 and the price of goods in Supermarket Type1 is approximately between 0–250, while Supermarket Type2 is somewhere between 250 and 350, and finally, Grocery Store takes prices across all ranges .

Plot based on Product Fat Content.

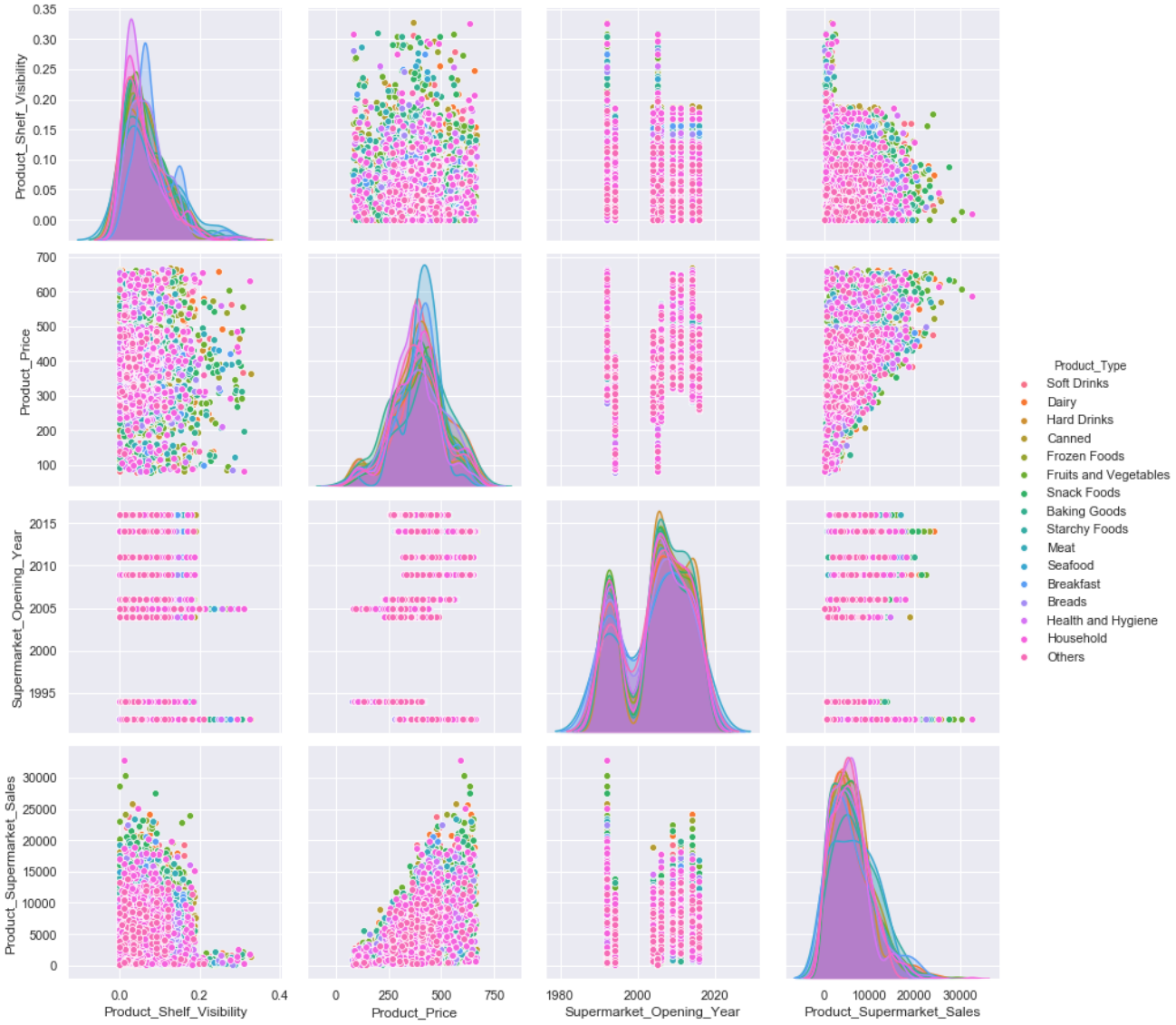


Product Supermarket Sales and the Product Shelf Visibility variable are skewed to the right which clearly determines that the higher the visibility higher the chances of the product being purchased by the customers. Product shelf visibility range (0- higher visibility) (0.3 – lesser visibility).

It is always advisable to place the expensive products or hedonic products (luxury goods) with a higher visibility to increase their sales. Products of Normal Fat has a major contribution to the sales.

So the High and low fat product which contribute less to the sales must be placed in a place with higher visibility.

Plot based on Product Type.



From this plot, we can clearly recognise that the Household, Soft drinks, Health and Hygiene and Others dominate.

Most customers are interested only in these categories.

Based on these plots we can determine which factors and categories must be altered to enhance the sales of a supermarket.