Amazon Sales Analysis

Analysed the Amazon sales dataset I have use libraries for data analysis and visualisation that I have used in this Project Pandas, Numpy, Matplotlib, Seaborn, and plotly.

Data Preparation and Cleaning

*Cleaning and filtering and formatting data using Excel
*Load file using Pandas
* Handing Missing values
*show some information about the data.

Inferences and Conclusion

I start With 100 sample with 18 columns (about 1800 of total sample size) missing values was 0 found and dataset.

Analysis-Univarate

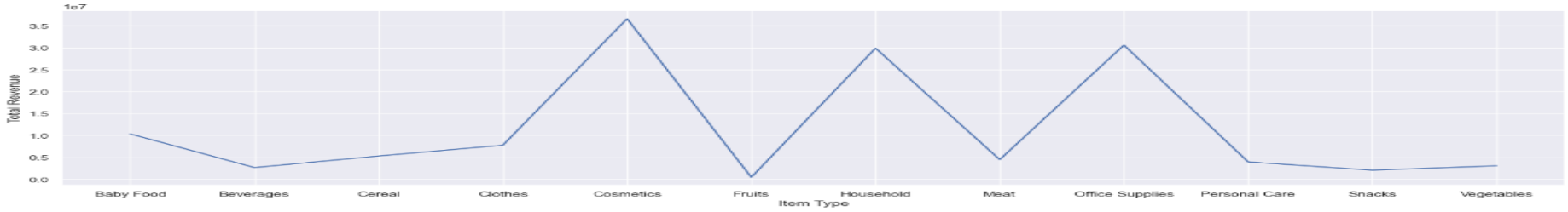
Region - Count 7 of Region name top region contribution to dataset is Sub-Saharan Africa, Europe, Asia and Middle East North Africa.

Order ID - Region wise segregation Count of order Id and maximum order ID in Sub-Saharan Africa, Europe and Middle East and North Africa and minimum order ID views is North America, Central America and the Caribbean and Middle East and North Africa There are increasing Offline Channel

Asia	11
Australia and Oceania	11
Central America and the Caribbean	07
Europe	22
Middle East and North Africa	10
North America	03
Sub-Saharan Africa	36

Country - Region in count of 76. Units Sold (512871) Total Profit: (44168198). We discuss about top 20 countries in Maximum Units Sold: and Total Profit Djibouti, Azerbaijan, Brunei, Cameroon and Fiji and Minimum Albania, Costa Rica, Federated States of Micronesia, Democratic Republic of the Congo and Comoros.

Item Type – Total of 12 Items are available in database Maximum Profitable are Cosmetics, Household, Office supply and Clothes And Minimum are Beverages, Snacks, meat and Fruits.
Total revenue Items wise maximum revenues are Cosmetics, Office Supplies, Household and Baby Food and minimum revenues are Vegetables, Beverages, Snacks and Fruits



Sales Channel – Type of 2 sales channel are Offline and Online. Offline channel are more profitable then online channel.
Offline Total Profit:- 24920727 and Online Total Profit is 19247472.
Customers are mostly trust offline more than online in purchasing products

Order Priority - Type of 4 Order Priority Maximum Units sold is H, L and Minimum C, M according Total Profit Views. And units wise Sold Units with revenue And Profit

Order Priority	Units Sold	Total Revenue	Total Profit
H	154212	48749543	16891601
L	146876	36628129	10858727
C	116951	18855064	6748328
M	94832	33116030	9669543

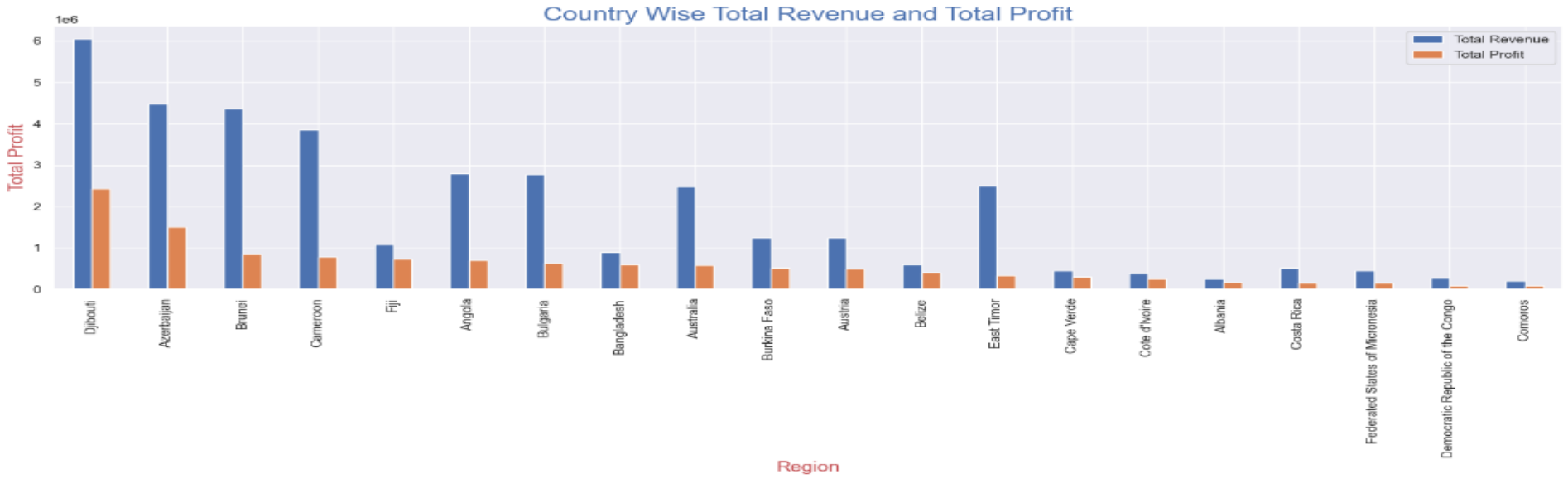
Year - Dataset represent data for 8 years from 2010 to 2017. Year wise Maximum Profitable Years: 2012, 2013, and 2010.
And Minimum Profitable Year: 2011, 2015 and 2017.

Visualisation of Data

All of 7 Region wise total profit maximum Sub-Saharan Africa, Europe and minimum Australia and Oceania, Central America and the Caribbean, North America there in increase a sales channel communication and delivery soon as soon fast providers Items.

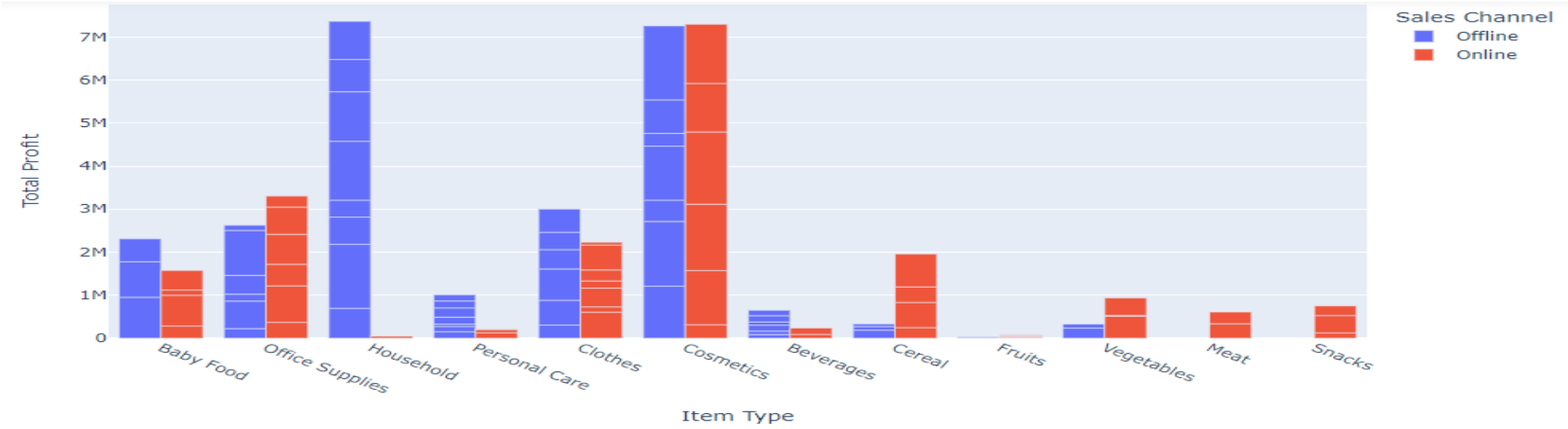
Country wise total revenue and total profit

Region in count of 76 countries in top 20 Total Profit: (44168198). With item per revenues is best viewing in each top 20 countries but revenues according more increase delivery channels and items quality and best services provide for more increase profit.



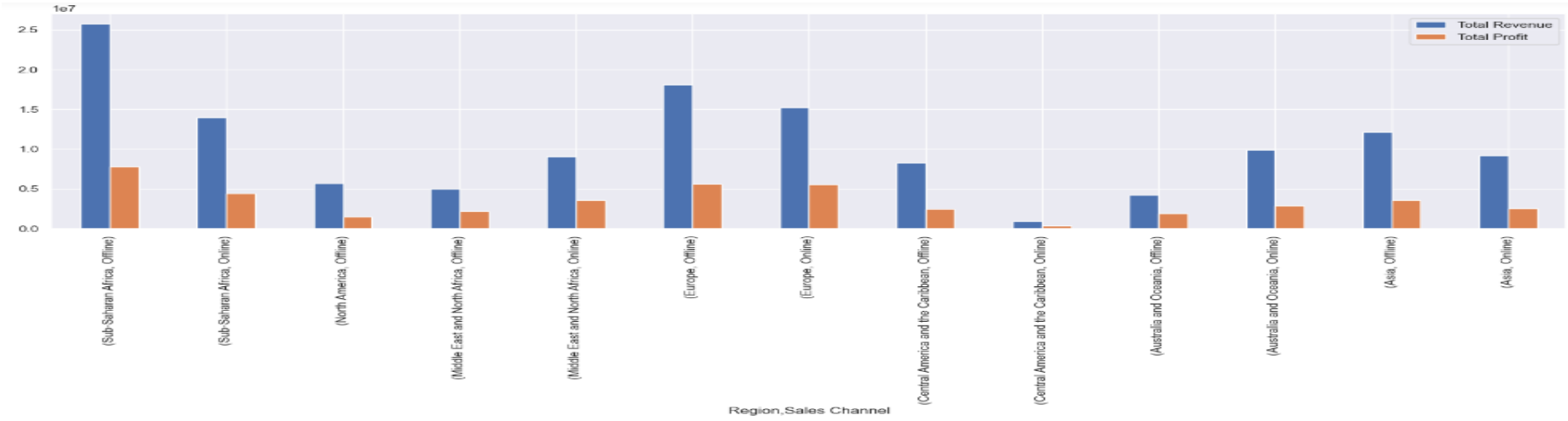
Sales Channel wise profit Total Profit.

Type of 2 sales channel are Offline and Online both channels are good performance as per items types maximum are Cosmetics, household, Office supply best Profit view and best services by sales channels and Minimum are fruits , meat, Vegetables, Snacks and Beverages are low profit view. That why most is food related items spoil quickly. Therefor minimum sales channels use we m ore fast delivery for food items.



Region by revenue and profit

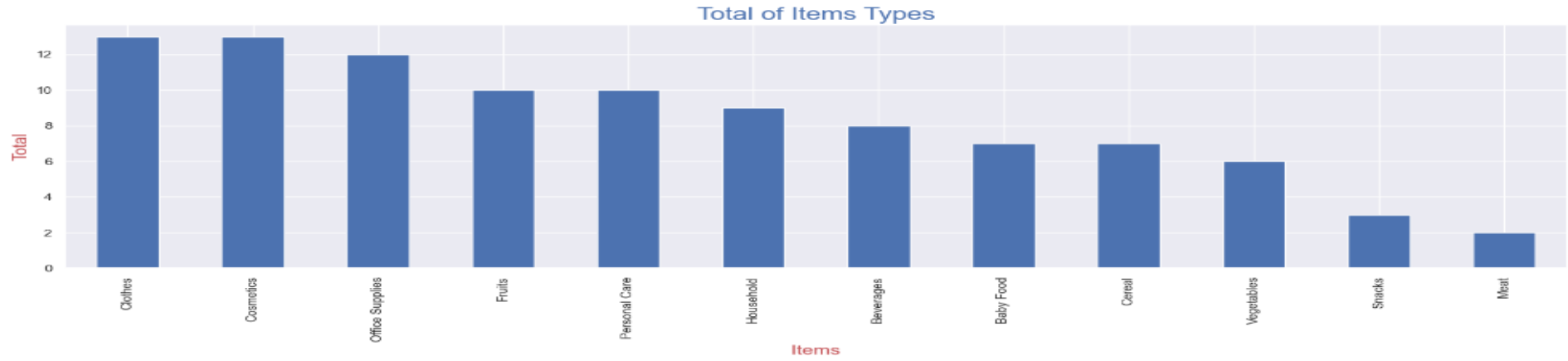
Total of 13 revenues by total revenues and total profit showing highest there best sales communication sale channels services provider and low views revenues and profit there are not possible to fast delivery items according customers required



Total of Items Types

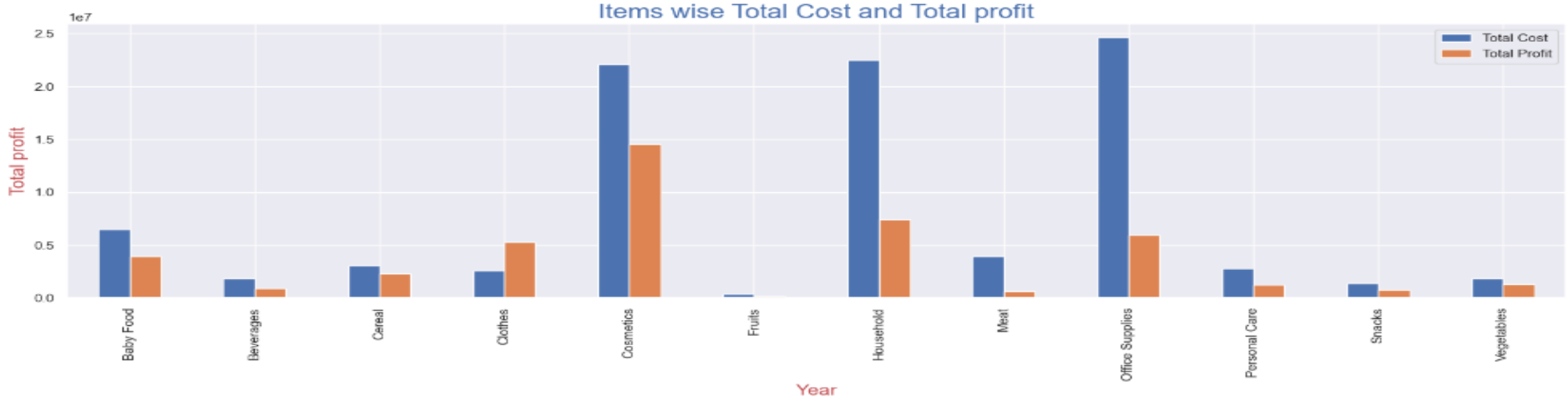
Total of 12 Items are available in database Maximum Profitable are Cosmetics, Household, Office supply and Clothes And Minimum are Beverages, Snacks, meat and Fruits.

Food related items are stored in according environment therefor log time use and no easy to spoiled for log delivery.



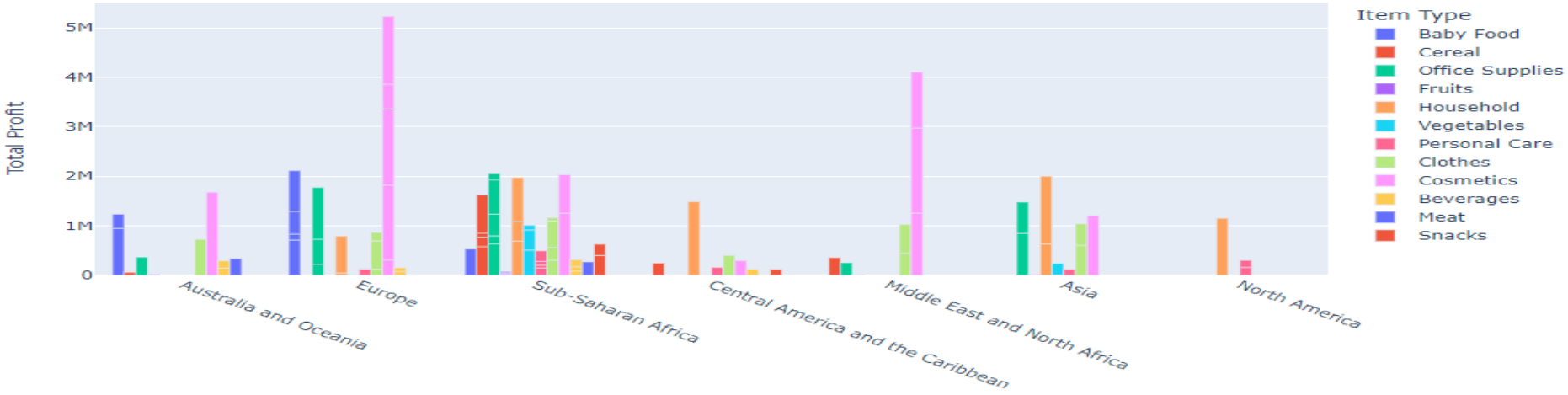
Item type wise cost and profit

Total of 12 Items are available in database Maximum Profitable are Cosmetics, Household, Office supply and Clothes And Minimum are Beverages, Snacks, meat and Fruits. And cost according profit are down side indicated so we want unnecessary cost will be remove from items.



Item type wise Total Profit by Region

Total of 7 Region in sales provides by amazon and total of 12 types of item available in each region as per customers required. Highest profitable and all types category sold are Sub-Saharan, Europe and Central America and the Caribbean and down side profitable and some category sold are North America, Middle East and North Africa and Australia and Oceania.

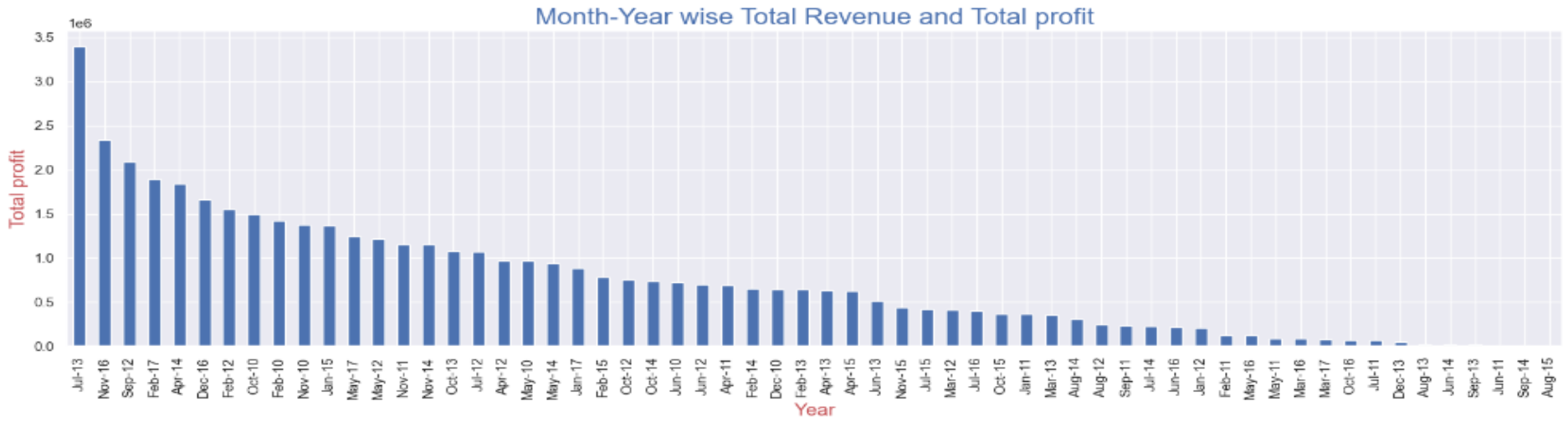


Year wise Total profit find out.

Dataset represent data for 8 years from 2010 to 2017. Year wise Maximum Profitable Years: 2012, 2013, and 2010. And Minimum Profitable Year: 2011, 2015 and 2017

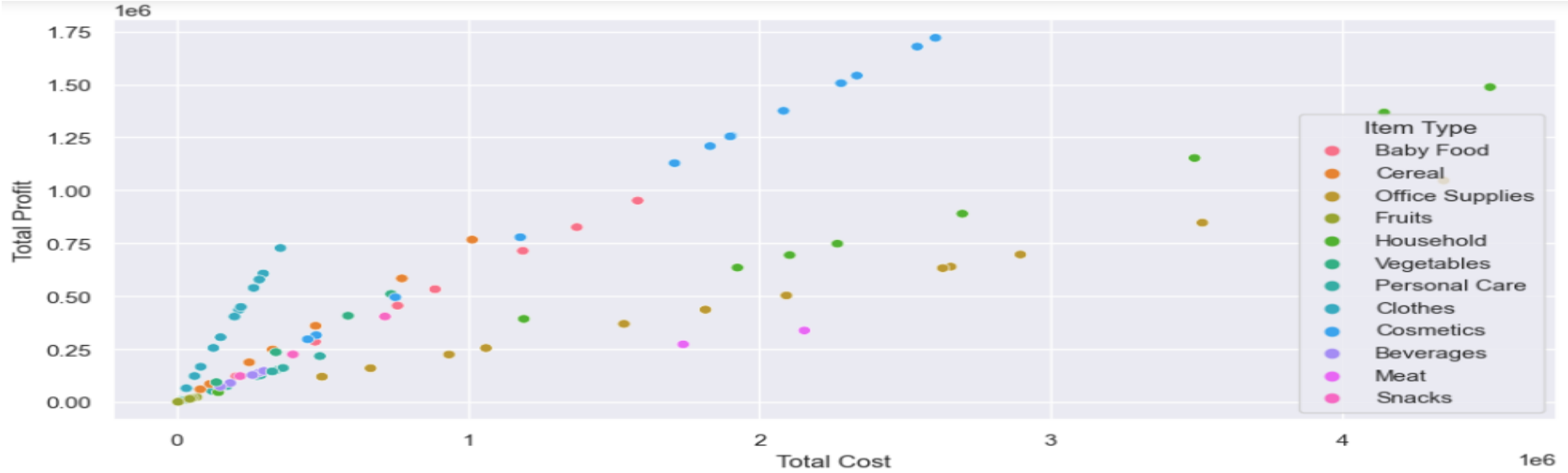
Month-Year wise Total profit find out

Months Year wise up to down side trend views each bar we can justify years according month by profit represents profit in bar graph.



Correlation Item Types wise Total Cost between Total Profit

It is correlation Item type of total cost between total profit this scatter graph is indicate represent items wise cost and profit and food items at giving high cost paying because month by month increase per items cost.
Food related items are stored in according environment therefor log time use and no easy to spoiled for log delivery that why increase cost.



Correlation with all columns

The heatmap () function takes a data matrix as input and plots that matrix as a heatmap. Each cell in the heatmap corresponds to a data point in the matrix, and the colour of the cell represents the value of that data point.
Correlation from 1 to -1. A positive 1 mean that the tow variables are positively coordinated to a degree of one decimal points.
A negative -1 mean that the variables are negatively coordinated.
all of dataset in 1 represent are positively view and some values are will going to indicates -1 its mean may be loss profit and cost values is coordination not good with profit and dark colour indicate to negatively view.

Thank You