# Data Science & Analyst Internship (POWERBI PROJECT WORK) _ Spark Foundation

# Name: Vishnupriya.K

**PROBLEM STATEMENT:** To Perform the Exploratory Data Analysis on Retail store data to identify the weak areas to work to make more profitable business.

## 1. Objective:

To identify the actionable insights to help outlet increase sales, identifying weak areas, customer buying choices and improve overall business efficiency.

## 2. Data Cleaning and preparation:

The given data set was clean with no blanks/empty rows and can be considered for further analysis.

The data in csv. file need to load into power bi, then check the data type of each column in a table and change to respective data type.

The data consist of following columns:

- Ship mode – it represents mode of shipment
- Segment – it represents type of customers
- Contry,city,state, pincode, region – Details of location
- Category – It represents type of product which it belongs
- Sub-category - – It represents the product name
- Qty – qty of product purchased
- Sales – purchases of product
- Discount – Discount given to product on total price after purchase

## 3. Data Analysis:

The Exploratory Data Analysis was performed on the given data set & all insights were categorized into 3 types.

- EDA is an approach to analyze the datasets to summarize their main characteristics in form of visual methods.
- EDA helps us to find errors, discovering data, mapping out data structure, finding out anomalies (outliers).

The Analysis was divided into following 3 categories:

**i) Overview:**

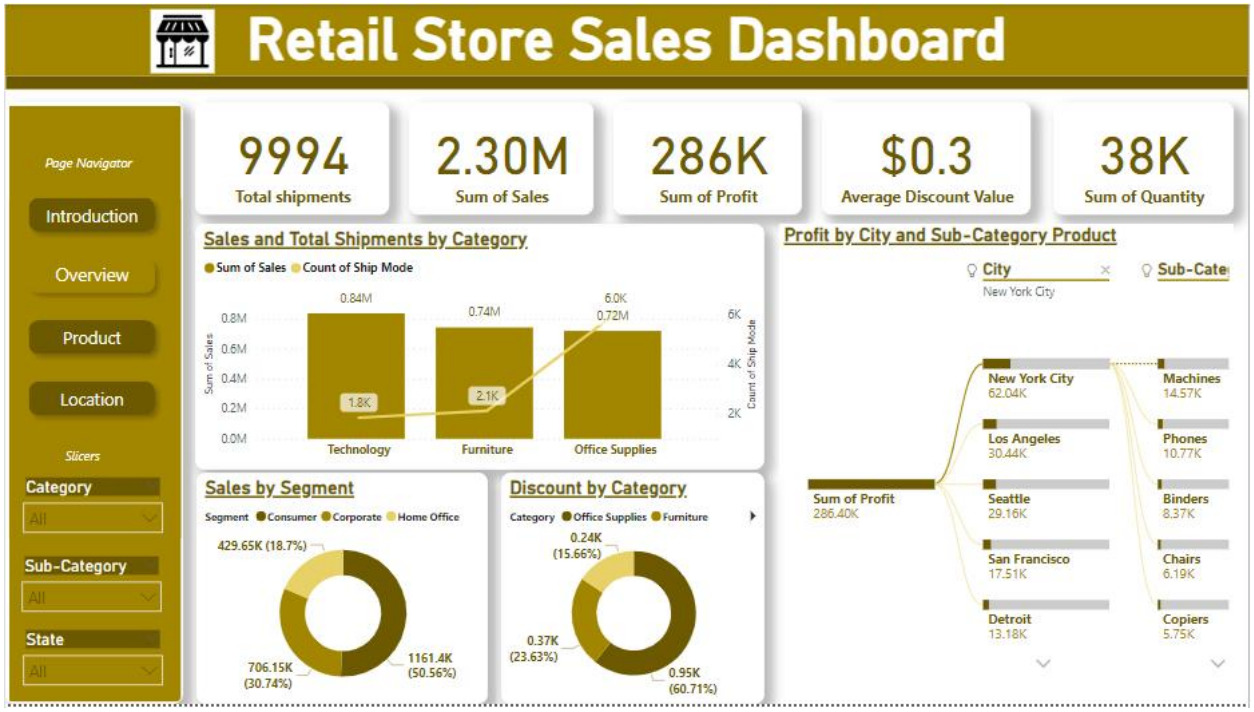
The following are the dax (Data Analysis Expressions) functions used to calculate overall analysis of the data:

# Data Science & Analyst Internship (POWERBI PROJECT WORK) _ Spark Foundation

## Name: Vishnupriya.K

| Column/Measure name | Dax formula used |
|---|---|
| Total shipments | COUNT(SampleSuperstore[Ship Mode]) |
| Sum of sales | Sum(SampleSuperstore[Sales]) |
| Sum of profit | Sum(SampleSuperstore[profit]) |
| Sum of discount | Sum(SampleSuperstore[discount]) |
| Sum of Quantity | Sum(SampleSuperstore[Quantity]) |

Based on above formulas, the following highlights on dashboard was created.




**a) Sales & total shipments by Category trend:**

- The retail store was made around 10,000 shipments (time period was not available in data) with 2.3 million sales, 286 k profits for a quantity of 38k products.
- In terms of sales, the technology has highest sales (0.84 M) followed by furniture (0.74 M), office suppliers (0.72 M).

**b) Discount by Category Trend:**

- The discount was given high for office suppliers (60%) followed by furniture(~24%) & technology(~16%) category.

**c) Sales by Segment Trend:**

- In terms of segment, the consumer segment holds highest sales of ~ 1.2m sales, highest shipments (5.2k) followed by corporate 0.71 m sales, 3 k shipments and home office 0.43 m sales, 1.8 k shipments.

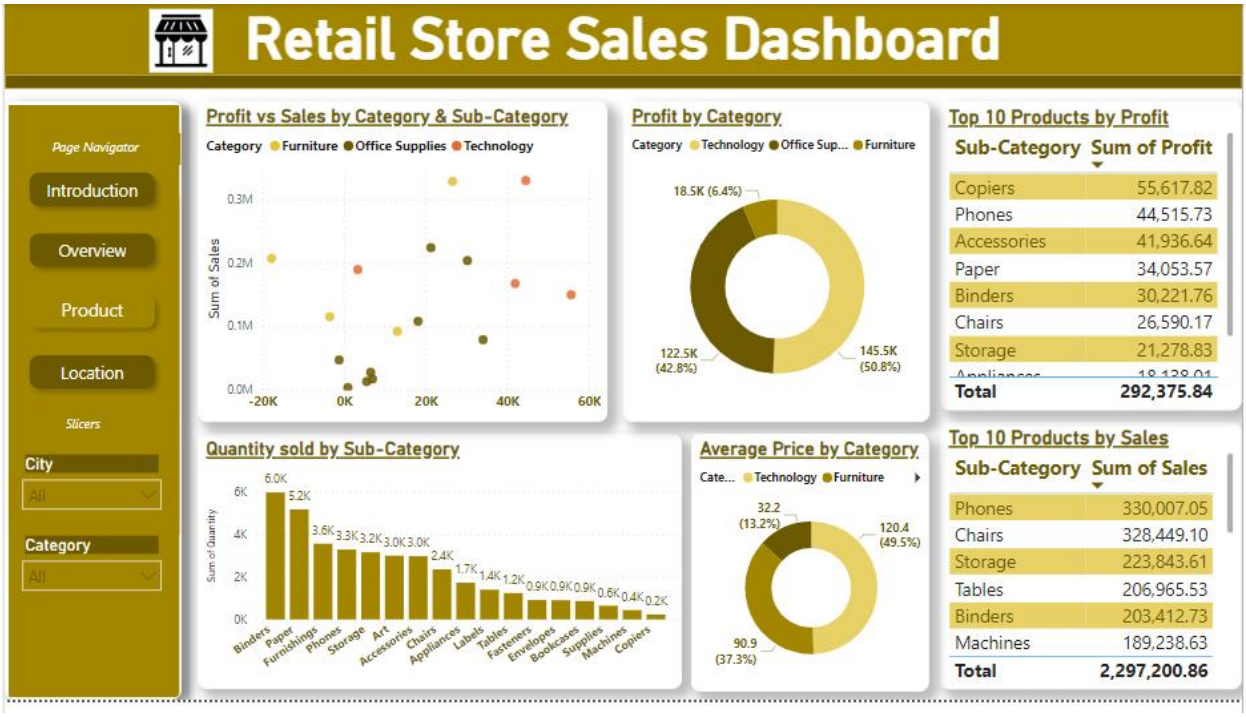
**d) Profit by city & sub Category Analysis:**

- Total profit was 286k which was around 12% of profit margin with respective to revenue.
- New York city was most profitable city with machines, phones products has highest profits followed by losangeles where copies has highest profits in seattle, sanfrancisco cities.
- In the technology category (145k profit), machines, phones, copiers, accessories has highest profits from new york city which was 36k & remaining cities doesn't seems much profits from technology category.
- Here discount was offered less for technology items, still it has highest sales & profit means there is no correlation for discount & sales & people interested to buy items at any price.
- In Office suppliers category(122k profit), binders, applications, storage items has highest profits from new york city(20.62k) followed by los Angeles (13.46k), seattle, sanfrancisco hold (~9 k) profits & majority are from binders, applications, storage items these three items.
- The discount was higher for office suppliers, it incurred more profits & sales. So, there is a correlation b/w discounts & sales for this category items & people interested to buy according to discounts.
- In the furniture category (18.45k profit), tables , furnishings have higher profits from Seattle city, followed by new York city, surprisingly tables (-3.54k) shows negative profits and chairs shows higher profits, & also in Los Angeles, tables (-0.06k) shows negative profits.
- Furniture category has least profits although sales are good & discount also offered less, which means need to revise/recheck the original prices of this product to reduce negative profits.
- It represents that every city has their own demand products & this analysis helps to predict further demand and maintain inventory according to the city requirement to reduce losses.
- To understand further, lets deep down into individual product for further analysis.

# Data Science & Analyst Internship (POWERBI PROJECT WORK) _ Spark Foundation
# Name: Vishnupriya.K

**ii) Product Analysis:**

- The following was the dashboard created for in depth profit & sales analysis of product.




- From the above dashboard, we clearly see that phones,machines (technology), storage, binders (office suppliers), chairs, tables (furniture) holds maximum sales.
- In profits, binders, storage, paper (office suppliers), Phones, copiers, accessories, (technology), chairs (furniture) has more profits.
- If we see profit vs sales, there is no correlation b/w these two because some products also incurred losses even though sales are higher.
- In terms of profits, technology (50%) & office suppliers (43%) holds over all profit & least was furniture (6%).
- In Technology all sub category items (4 nos) gave profits except machines (<5k) although sales are better, in office suppliers ( 8 nos) out of 8 items, 6 items gave (>5k) profits & remaining 2 items fasteners (<1k) & supplies (-1.1 k) gave negative returns sales also for thes items are less, so better to maintain inventory less according to demand.

## Data Science & Analyst Internship (POWERBI PROJECT WORK) _ Spark Foundation

## Name: Vishnupriya.K

- In furniture out of 4 products, bookcases (-3.4 k) and tables (- 17k) gave huge losses in overall products.
- There is a need to focus & revise the prices of tables & bookcases to reduce losses & understand other factors for causing losses in particular items.
- Majority of the qty sold by binders, followed by papers, furnishings, phones etc.
- To understand location where business is huge/low need to deep down the data.

**iii) Location wise analysis:**


- New york city was the location which is more profitable, higher sales & qty sold.
- Most of the sales are at west, south & central regions & few in east.
- In terms of mode of shipment, all locations are preferring standard class delivery followed by second class, first class & very few in same day.
- Majority of transactions are with customers followed by corporate orders & few home office orders.

# Data Science & Analyst Internship (POWERBI PROJECT WORK) _ Spark Foundation

# Name: Vishnupriya.K

## Insights & Recommendations:

- In terms of sales, the technology has highest sales (0.84 M) followed by furniture (0.74 M), office suppliers (0.72 M).
- The discount was higher for office suppliers, it incurred more profits & sales. So, there is a correlation b/w discounts & sales for this category items & people interested to buy according to discounts.
- The discount was offered less for technology items, still it has highest sales & profit means there is no correlation for discount & sales and people interested to buy items at any price.
- If we see profit vs sales, there is no correlation b/w these two because some products also incurred losses even though sales are higher.
- In terms of profits, technology (50%) & office suppliers (43%) holds over all profit & least was furniture (6%).
- In Technology we need to focus on machines where ales are good, but profits are less, need to revise the prices because the demand of the product.
- In office suppliers , 2 items fasteners (<1k) & supplies (-1.1 k) these 2 items are very weak in this category so better to reduce the inventory of these items to reduce losses.
- In furniture out of 4 products, bookcases (-3.4 k) and tables (- 17k) gave huge losses in overall products.
- In the furniture category (18.45k profit), tables , furnishings have higher profits from Seattle city, followed by new York city, surprisingly tables (-3.54k) shows negative profits and chairs shows higher profits, & also in Los Angeles, tables (-0.06k) shows negative profits.
- Furniture category has least profits although sales are good & discount also offered less, which means need to revise/recheck the original prices of this product to reduce negative profits.
- It represents that every city has their own demand products & this analysis helps to predict further demand and maintain inventory according to the city requirement to reduce losses.