Week 3 - Project

Case Study

A superstore retail business is a large, multi-department store that sells various products, including groceries, electronics, home goods, clothing, and more. These stores are often designed to be a one-stop shop for customers, offering a wide range of products and services under one roof. Superstores are typically larger than traditional retail stores and may have a larger product selection. Superstores are often part of a larger chain and have multiple locations in a region or country.

A new store manager needs your help to better understand his/her Data Operations Team. You are provided with part of the sales data that a Business Intelligence Analyst encounters daily. Design the dashboard to analyze and interpret the data to help provide valuable insights to the store manager.

Dataset

A Superstore dataset typically includes information about the products, customers, and sales associated with a retail store. It may include the following columns:

Table Description:

Sl.No	Column Name	Column Description
1	Order ID	Unique Identifier For the Order
2	Order Date	Date of the order placed
3	Ship Date	Date of the order shipped
4	Ship Mode	Priority Mode of Shipping (Same Day, First Class, Second Class, Standard Class)
5	Customer ID	Customer Unique Identifier
6	Customer Name	Name of the customer
7	Segment	Customer Segment (Consumer, Corporate, Home Office)
11	Postal Code	Address from the order was placed
12	Region	Name of the Region
13	Product ID	Unique Product Identifier
14	Category	Product Category
15	Sub-Category	Product Sub - Category
16	Product Name	Name of the Product

18	Quantity	Quantity of the product ordered
19	Discount	Discount % on the product
20	Buy Price	Buying price for each item
21	Price Per Each	Selling price for each item

Problem Statement:

Data Extraction, Cleaning,Loading and Transformation

Q1. Desk representatives at the stores are not tech savvy hence they directly share the data in the single excel file. As a Power BI Developer, read the data directly from the excel file.

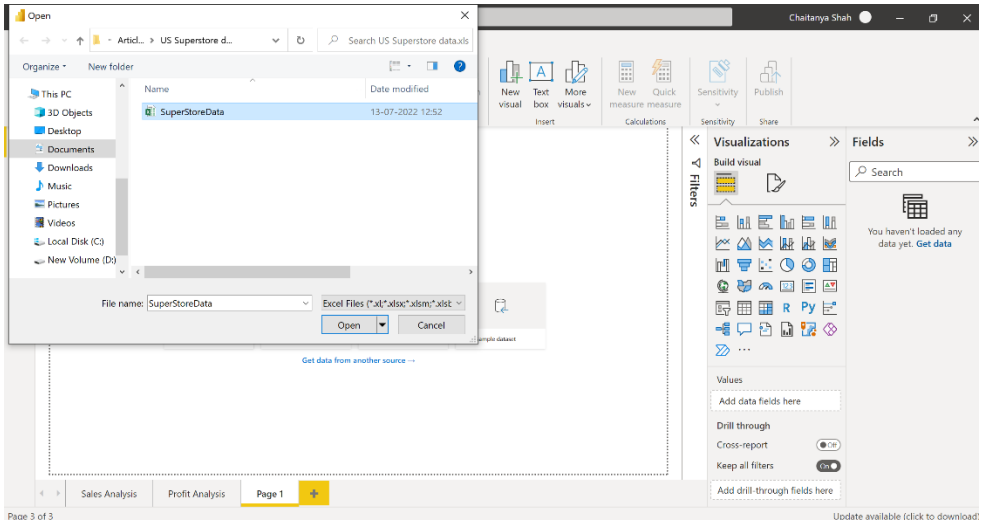
- **Connect to Data Sources in Power BI**

The get data option is used for loading data for visualization and analysis. Developers can either load the data first or perform a data transformation before loading the data. The get data option is used to connect to *different data sources* and is present in the Home tab on the Power BI desktop. Users can import data from Azure Synapse Analytics SQL, Excel, Text/CSV, Web, Amazon RedShift, Oracle, MySQL, Snowflake, SAP databases, Google BigQuery, MariaDB, SharePoint List, etc., in Power BI.

Importing Data in Power BI

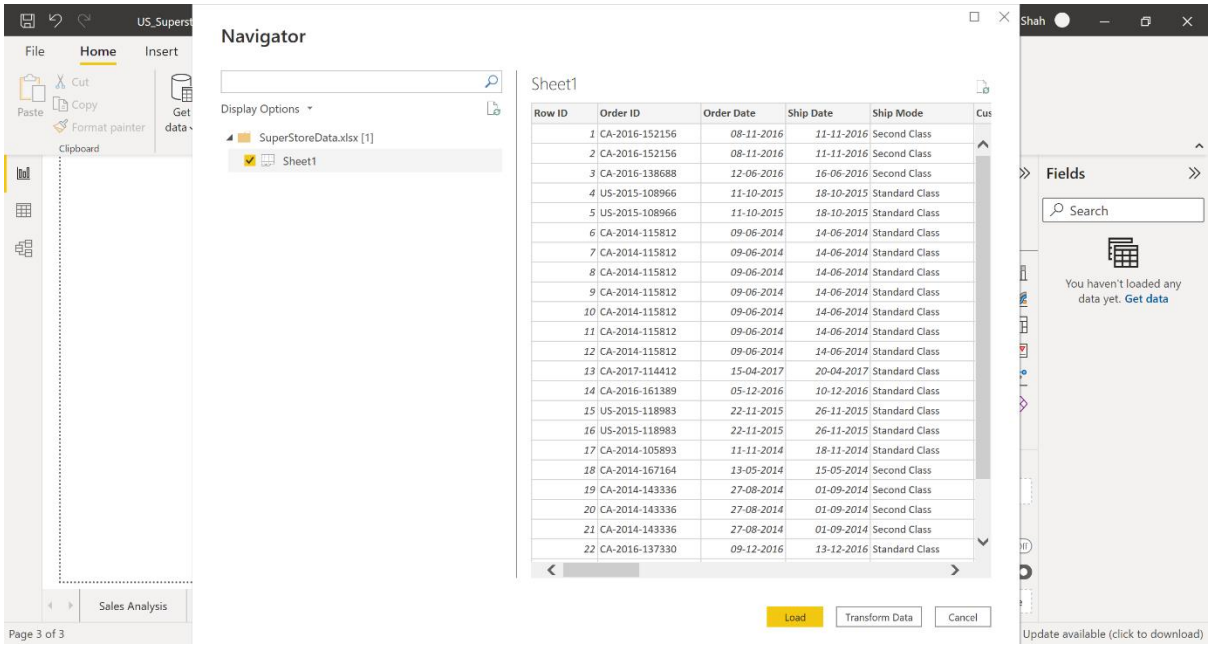
The get data option is used for loading data for visualization and analysis.

Step 1: Click on ***Get data in the Home tab***, then select ***Excel workbook,*** now choose the downloaded excel file and ***Open*** it.



2. The data coming from the source is in raw form in the flat file; hence clean and prepare (transform) the dataset for efficient use. Delete the empty columns and rows, change the fields to appropriate data types and split the fields and rename the columns appropriately.

Step 2: Select ***Sheet1-> Transform Data***.



Performing Data Transformation

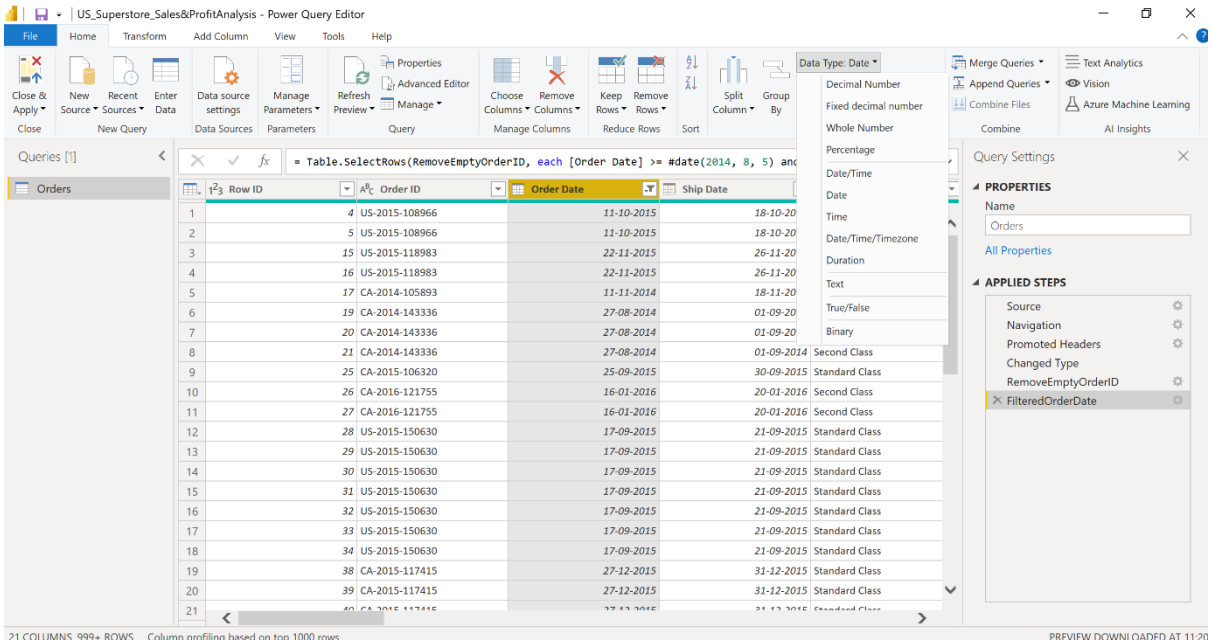
A. Shaping & Cleaning the Data

1. Right Click ***Sheet1 -> Rename.*** Rename it to ***Orders***.

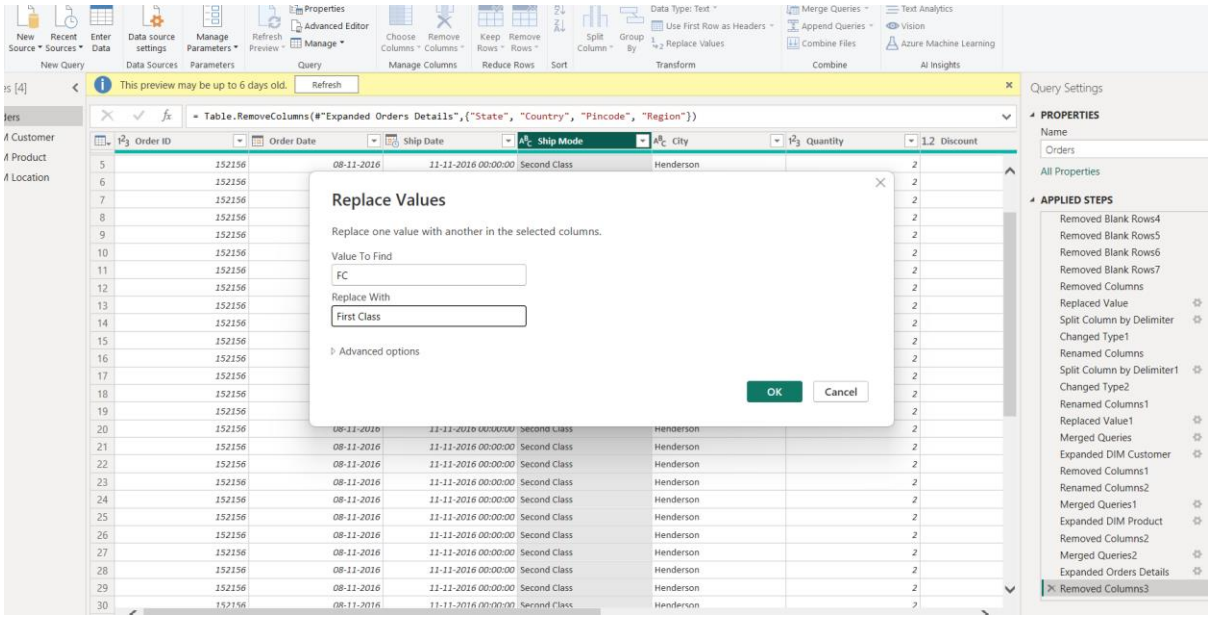
Evaluating and changing column data types

Now, click on every column and check if the Data Type is correctly selected or not.

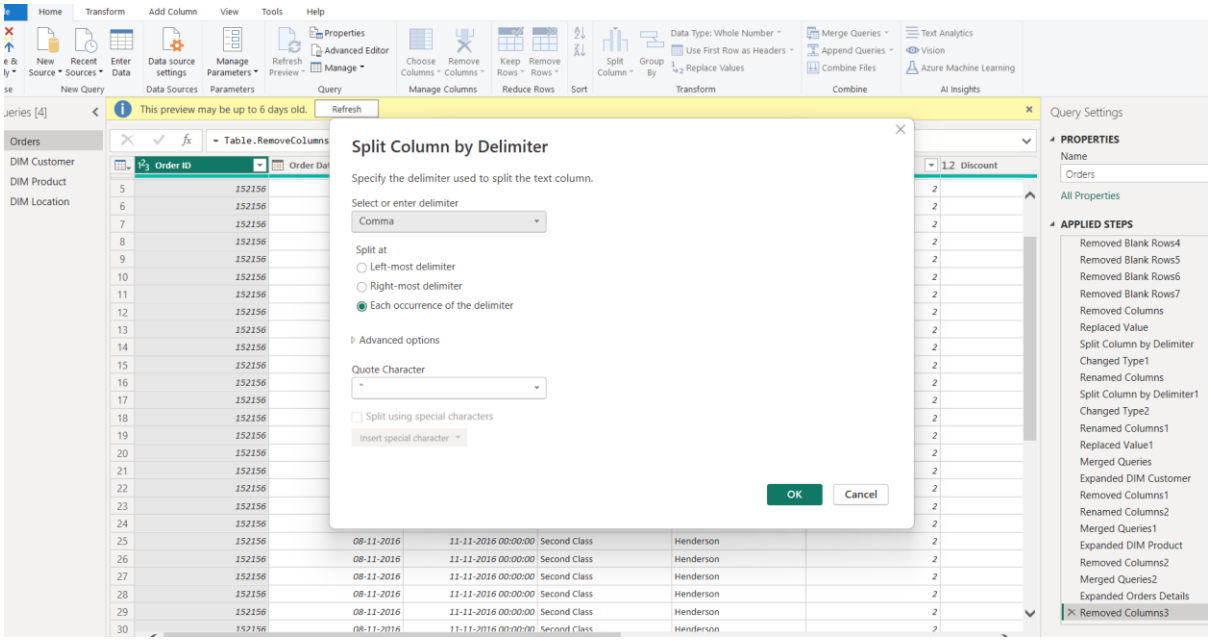
If the ***Data Type*** is not correctly selected, change the Data Type from the dropdown. *For example*, the *Data Type of Order Date & Ship Date column* must be ***Date***.



3. Standardize the values in the column Ship mode [Hint: Replace FC with first class]



4. Split the address column to City, State, Country and Pincode



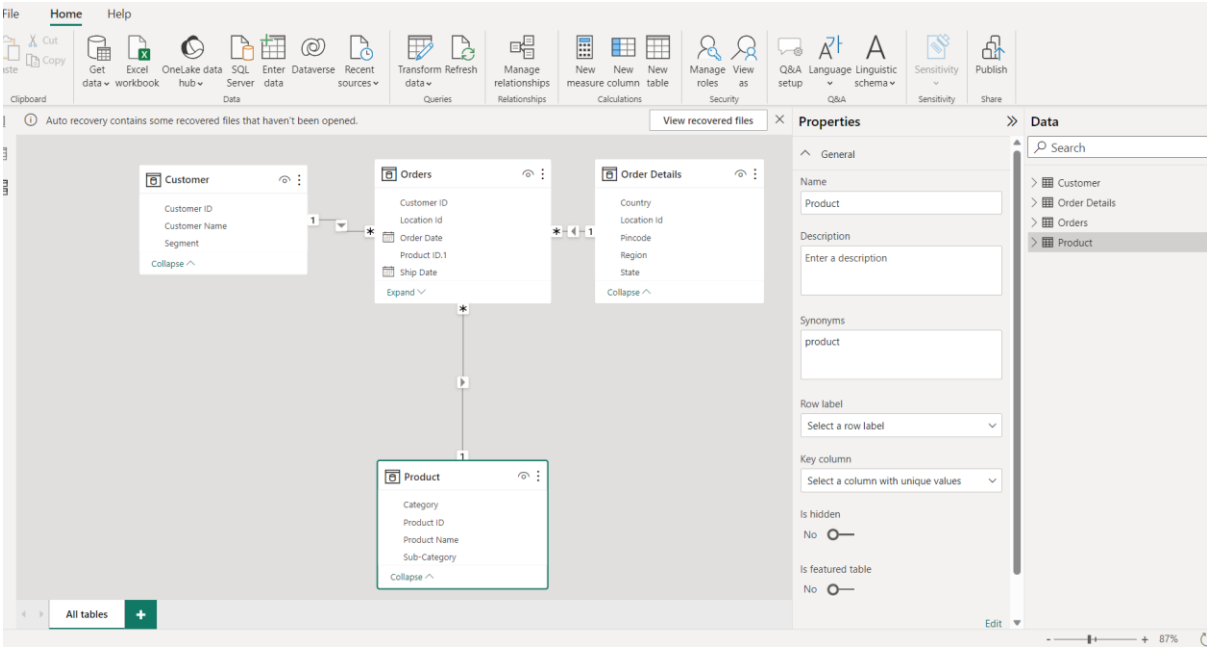
Data Modeling:

1. Tracking sales in the retail business is a weekly task; hence setting up the data model will be crucial for this. Convert a flat file into STAR schema for better performance of the analysis. The schema shall have a central Fact table, ‘Orders’ and three dimension tables,’Order details’, ‘Customer’ and ‘Product’. Refer to the table below for creating appropriate columns in each table.

   Power BI data modeling is the process of creating a relationship between common columns of multiple tables. If the column headings are the same across tables, then Power BI auto-detects the relationship between tables. Using these columns, we can merge the tables as well.

   Microsoft Power BI Desktop is not only a visualization tool, but it also has the capabilities for data modeling. If the data is not modeled correctly, it will cause redundancy, consume more storage, and make data retrieval difficult. Using a Star Schema is the mature modeling approach that overcomes most poor modeling problems.

   Star Schema:
   A star schema is a database modeling technique commonly used in data warehousing and business intelligence projects. It is a schema design used to organize and store data in a way that simplifies query processing and reporting in a relational data model. Its basic form consists of a single **fact table** surrounded by multiple **dimension tables**, all connected by joins between the primary key and foreign keys (often called surrogate keys).



2. Remove duplicate rows from the newly created dimension tables, and ensure there are no empty rows

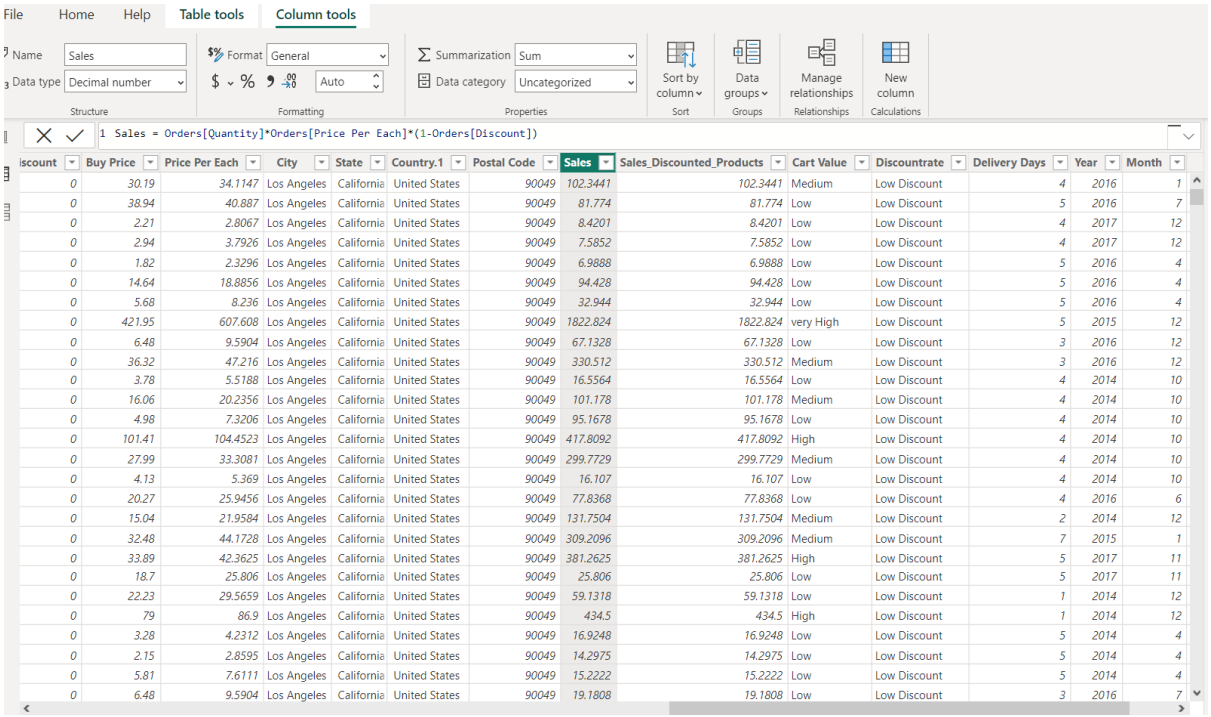
3. Once the tables are created, ensure one- to - many relationships are created between dimensions and fact table.

Column Name	**Type**
Ship Mode	Dimension(Order Details)
Postal Code	Dimension(Order Details)
Region	Dimension(Order Details)
Customer ID	Dimension (Customer)
Customer Name	Dimension (Customer)
Segment	Dimension (Customer)
Product ID	Dimension (Product)
Category	Dimension (Product)
Sub-Category	Dimension (Product)
Product Name	Dimension (Product)

Data Analysis:

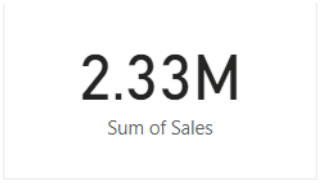
Furthermore, to grow the footholds in the store and achieve an ambitious sales target, the store manager wants to track and visualize the following metrics;

1. Create a new column ‘Sales’ or ‘Order value’ [Hint: Sales = Qty*price per qty*(1- % Discount)].



iscount	Buy Price	Price Per Each	City	State	Country.1	Postal Code	Sales	Sales_Discounted_Products	Cart Value	Discountrate	Delivery Days	Year	Month
0	30.19	34.1147	Los Angeles	California	United States	90049	102.3441	102.3441	Medium	Low Discount	4	2016	1
0	38.94	40.887	Los Angeles	California	United States	90049	81.774	81.774	Low	Low Discount	5	2016	7
0	2.21	2.8067	Los Angeles	California	United States	90049	8.4201	8.4201	Low	Low Discount	4	2017	12
0	2.94	3.7926	Los Angeles	California	United States	90049	7.5852	7.5852	Low	Low Discount	4	2017	12
0	1.82	2.3296	Los Angeles	California	United States	90049	6.9888	6.9888	Low	Low Discount	5	2016	4
0	14.64	18.8856	Los Angeles	California	United States	90049	94.428	94.428	Low	Low Discount	5	2016	4
0	5.68	8.236	Los Angeles	California	United States	90049	32.944	32.944	Low	Low Discount	5	2016	4
0	421.95	607.608	Los Angeles	California	United States	90049	1822.824	1822.824	very High	Low Discount	5	2015	12
0	6.48	9.5904	Los Angeles	California	United States	90049	67.1328	67.1328	Low	Low Discount	3	2016	12
0	36.32	47.216	Los Angeles	California	United States	90049	330.512	330.512	Medium	Low Discount	3	2016	12
0	3.78	5.5188	Los Angeles	California	United States	90049	16.5564	16.5564	Low	Low Discount	4	2014	10
0	16.06	20.2356	Los Angeles	California	United States	90049	101.178	101.178	Medium	Low Discount	4	2014	10
0	4.98	7.3206	Los Angeles	California	United States	90049	95.1678	95.1678	Low	Low Discount	4	2014	10
0	101.41	104.4523	Los Angeles	California	United States	90049	417.8092	417.8092	High	Low Discount	4	2014	10
0	27.99	33.3081	Los Angeles	California	United States	90049	299.7729	299.7729	Medium	Low Discount	4	2014	10
0	4.13	5.369	Los Angeles	California	United States	90049	16.107	16.107	Low	Low Discount	4	2014	10
0	20.27	25.9456	Los Angeles	California	United States	90049	77.8368	77.8368	Low	Low Discount	4	2016	6
0	15.04	21.9584	Los Angeles	California	United States	90049	131.7504	131.7504	Medium	Low Discount	2	2014	12
0	32.48	44.1728	Los Angeles	California	United States	90049	309.2096	309.2096	Medium	Low Discount	7	2015	1
0	33.89	42.3625	Los Angeles	California	United States	90049	381.2625	381.2625	High	Low Discount	5	2017	11
0	18.7	25.806	Los Angeles	California	United States	90049	25.806	25.806	Low	Low Discount	5	2017	11
0	22.23	29.5659	Los Angeles	California	United States	90049	59.1318	59.1318	Low	Low Discount	1	2014	12
0	79	86.9	Los Angeles	California	United States	90049	434.5	434.5	High	Low Discount	1	2014	12
0	3.28	4.2312	Los Angeles	California	United States	90049	16.9248	16.9248	Low	Low Discount	5	2014	4
0	2.15	2.8595	Los Angeles	California	United States	90049	14.2975	14.2975	Low	Low Discount	5	2014	4
0	5.81	7.6111	Los Angeles	California	United States	90049	15.2222	15.2222	Low	Low Discount	5	2014	4
0	6.48	9.5904	Los Angeles	California	United States	90049	19.1808	19.1808	Low	Low Discount	3	2016	7

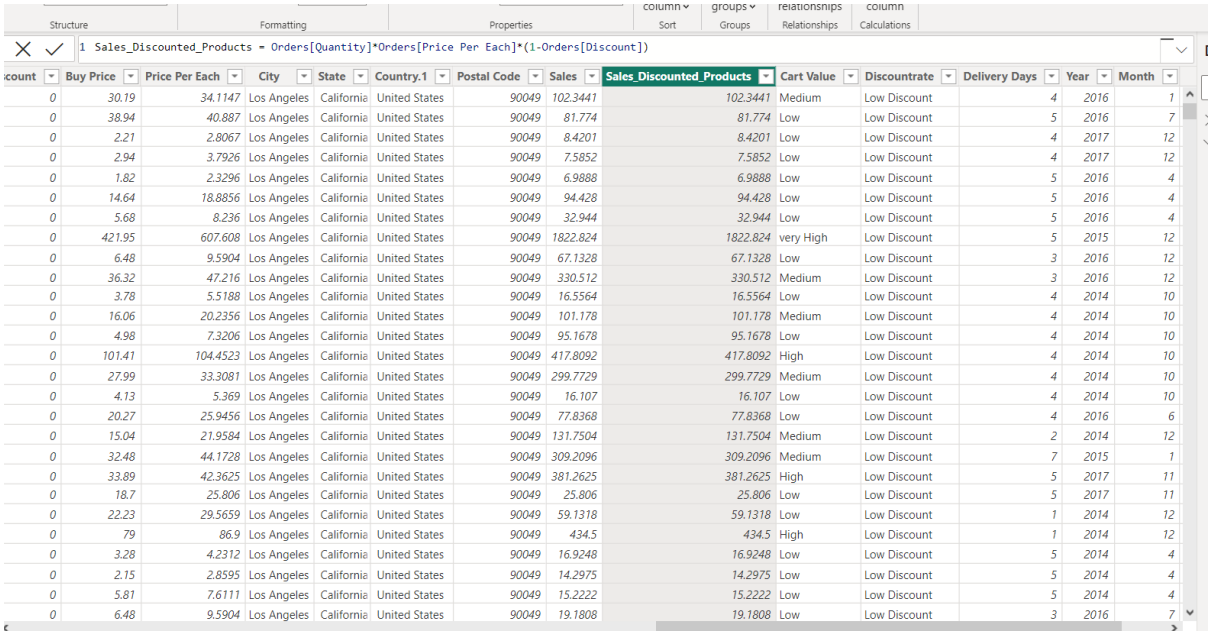
Create a card visual displaying the total sales.



-

2. Similarly calculate the Sales from discounted products and display the total sales from discounted products.

- First Filter the rows by filter options where discount is not equal to Zero.
- Then calculate Sales of discounted products

- 
Structure Formatting Properties column Sort groups Groups relationships Relationships column Calculations

```
1 Sales_Discounted_Products = Orders[Quantity]*Orders[Price Per Each]*(1-Orders[Discount])
```

count	Buy Price	Price Per Each	City	State	Country.1	Postal Code	Sales	Sales_Discounted_Products	Cart Value	Discountrate	Delivery Days	Year	Month
0	30.19	34.1147	Los Angeles	California	United States	90049	102.3441	102.3441	Medium	Low Discount	4	2016	1
0	38.94	40.887	Los Angeles	California	United States	90049	81.774	81.774	Low	Low Discount	5	2016	7
0	2.21	2.8067	Los Angeles	California	United States	90049	8.4201	8.4201	Low	Low Discount	4	2017	12
0	2.94	3.7926	Los Angeles	California	United States	90049	7.5852	7.5852	Low	Low Discount	4	2017	12
0	1.82	2.3296	Los Angeles	California	United States	90049	6.9888	6.9888	Low	Low Discount	5	2016	4
0	14.64	18.8856	Los Angeles	California	United States	90049	94.428	94.428	Low	Low Discount	5	2016	4
0	5.68	8.236	Los Angeles	California	United States	90049	32.944	32.944	Low	Low Discount	5	2016	4
0	421.95	607.608	Los Angeles	California	United States	90049	1822.824	1822.824	very High	Low Discount	5	2015	12
0	6.48	9.5904	Los Angeles	California	United States	90049	67.1328	67.1328	Low	Low Discount	3	2016	12
0	36.32	47.216	Los Angeles	California	United States	90049	330.512	330.512	Medium	Low Discount	3	2016	12
0	3.78	5.5188	Los Angeles	California	United States	90049	16.5564	16.5564	Low	Low Discount	4	2014	10
0	16.06	20.2356	Los Angeles	California	United States	90049	101.178	101.178	Medium	Low Discount	4	2014	10
0	4.98	7.3206	Los Angeles	California	United States	90049	95.1678	95.1678	Low	Low Discount	4	2014	10
0	101.41	104.4523	Los Angeles	California	United States	90049	417.8092	417.8092	High	Low Discount	4	2014	10
0	27.99	33.3081	Los Angeles	California	United States	90049	299.7729	299.7729	Medium	Low Discount	4	2014	10
0	4.13	5.369	Los Angeles	California	United States	90049	16.107	16.107	Low	Low Discount	4	2014	10
0	20.27	25.9456	Los Angeles	California	United States	90049	77.8368	77.8368	Low	Low Discount	4	2016	6
0	15.04	21.9584	Los Angeles	California	United States	90049	131.7504	131.7504	Medium	Low Discount	2	2014	12
0	32.48	44.1728	Los Angeles	California	United States	90049	309.2096	309.2096	Medium	Low Discount	7	2015	1
0	33.89	42.3625	Los Angeles	California	United States	90049	381.2625	381.2625	High	Low Discount	5	2017	11
0	18.7	25.806	Los Angeles	California	United States	90049	25.806	25.806	Low	Low Discount	5	2017	11
0	22.23	29.5659	Los Angeles	California	United States	90049	59.1318	59.1318	Low	Low Discount	1	2014	12
0	79	86.9	Los Angeles	California	United States	90049	434.5	434.5	High	Low Discount	1	2014	12
0	3.28	4.2312	Los Angeles	California	United States	90049	16.9248	16.9248	Low	Low Discount	5	2014	4
0	2.15	2.8595	Los Angeles	California	United States	90049	14.2975	14.2975	Low	Low Discount	5	2014	4
0	5.81	7.6111	Los Angeles	California	United States	90049	15.2222	15.2222	Low	Low Discount	5	2014	4
0	6.48	9.5904	Los Angeles	California	United States	90049	19.1808	19.1808	Low	Low Discount	3	2016	7

3. Since supermarkets sell bulk items, store managers want to know each order's cart value. Create a column “Cart Value” that categorizes the order value/sales as Low, medium, high or very high.

Cart Value,

< 1000: Low

<3500: Medium

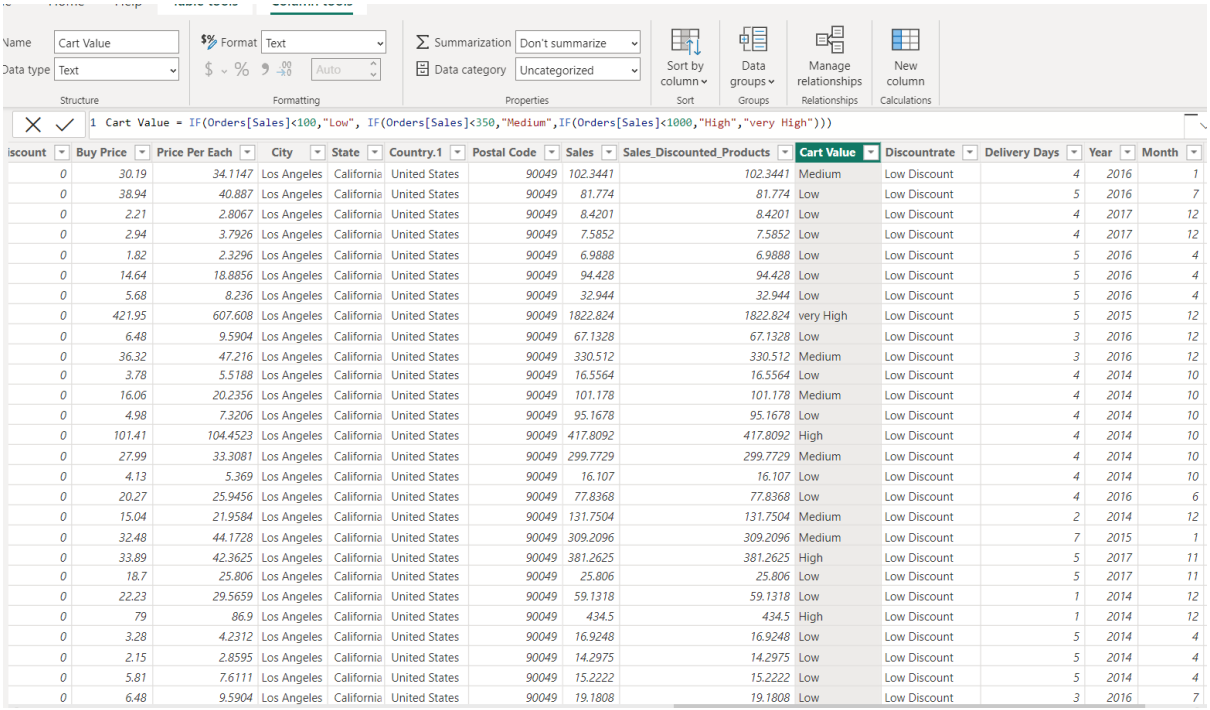
< 10000: High

> 10000: Very High

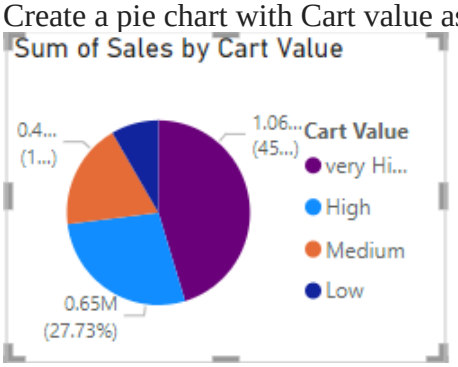
[Hint : From Power Query editor->Add column-> Conditional column OR

DAX formula using nested IF function]

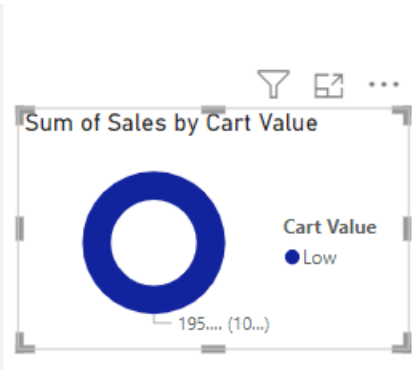
- Cart Value = IF(Orders[Sales]<1000,"Low", IF(Orders[Sales]<3500,"Medium",IF(Orders[Sales]<10000,"High","very High")))
-



iscount	Buy Price	Price Per Each	City	State	Country.1	Postal Code	Sales	Sales_Discounted_Products	Cart Value	Discountrate	Delivery Days	Year	Month
0	30.19	34.1147	Los Angeles	California	United States	90049	102.3441	102.3441	Medium	Low Discount	4	2016	1
0	38.94	40.887	Los Angeles	California	United States	90049	81.774	81.774	Low	Low Discount	5	2016	7
0	2.21	2.8067	Los Angeles	California	United States	90049	8.4201	8.4201	Low	Low Discount	4	2017	12
0	2.94	3.7926	Los Angeles	California	United States	90049	7.5852	7.5852	Low	Low Discount	4	2017	12
0	1.82	2.3296	Los Angeles	California	United States	90049	6.9888	6.9888	Low	Low Discount	5	2016	4
0	14.64	18.8856	Los Angeles	California	United States	90049	94.428	94.428	Low	Low Discount	5	2016	4
0	5.68	8.236	Los Angeles	California	United States	90049	32.944	32.944	Low	Low Discount	5	2016	4
0	421.95	607.608	Los Angeles	California	United States	90049	1822.824	1822.824	very High	Low Discount	5	2015	12
0	6.48	9.5904	Los Angeles	California	United States	90049	67.1328	67.1328	Low	Low Discount	3	2016	12
0	36.32	47.216	Los Angeles	California	United States	90049	330.512	330.512	Medium	Low Discount	3	2016	12
0	3.78	5.5188	Los Angeles	California	United States	90049	16.5564	16.5564	Low	Low Discount	4	2014	10
0	16.06	20.2356	Los Angeles	California	United States	90049	101.178	101.178	Medium	Low Discount	4	2014	10
0	4.98	7.3206	Los Angeles	California	United States	90049	95.1678	95.1678	Low	Low Discount	4	2014	10
0	101.41	104.4523	Los Angeles	California	United States	90049	417.8092	417.8092	High	Low Discount	4	2014	10
0	27.99	33.3081	Los Angeles	California	United States	90049	299.7729	299.7729	Medium	Low Discount	4	2014	10
0	4.13	5.369	Los Angeles	California	United States	90049	16.107	16.107	Low	Low Discount	4	2014	10
0	20.27	25.9456	Los Angeles	California	United States	90049	77.8368	77.8368	Low	Low Discount	4	2016	6
0	15.04	21.9584	Los Angeles	California	United States	90049	131.7504	131.7504	Medium	Low Discount	2	2014	12
0	32.48	44.1728	Los Angeles	California	United States	90049	309.2096	309.2096	Medium	Low Discount	7	2015	1
0	33.89	42.3625	Los Angeles	California	United States	90049	381.2625	381.2625	High	Low Discount	5	2017	11
0	18.7	25.806	Los Angeles	California	United States	90049	25.806	25.806	Low	Low Discount	5	2017	11
0	22.23	29.5659	Los Angeles	California	United States	90049	59.1318	59.1318	Low	Low Discount	1	2014	12
0	79	86.9	Los Angeles	California	United States	90049	434.5	434.5	High	Low Discount	1	2014	12
0	3.28	4.2312	Los Angeles	California	United States	90049	16.9248	16.9248	Low	Low Discount	5	2014	4
0	2.15	2.8595	Los Angeles	California	United States	90049	14.2975	14.2975	Low	Low Discount	5	2014	4
0	5.81	7.6111	Los Angeles	California	United States	90049	15.2222	15.2222	Low	Low Discount	5	2014	4
0	6.48	9.5904	Los Angeles	California	United States	90049	19.1808	19.1808	Low	Low Discount	3	2016	7

- Create a pie chart with Cart value as legend and Order value in Values field.
- 

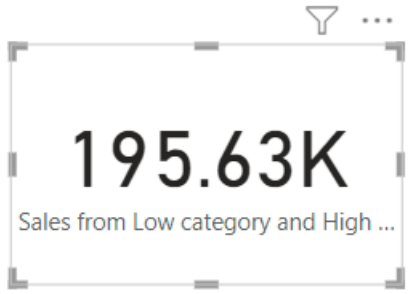
4. Separately visualize the total sales just from the low cart value category (as mentioned, any value below 1000 can be considered as low value category).



5. Using card visual, track the total sales coming from the low cart category and discount more than or equal to 50% to find out the contribution and cause.

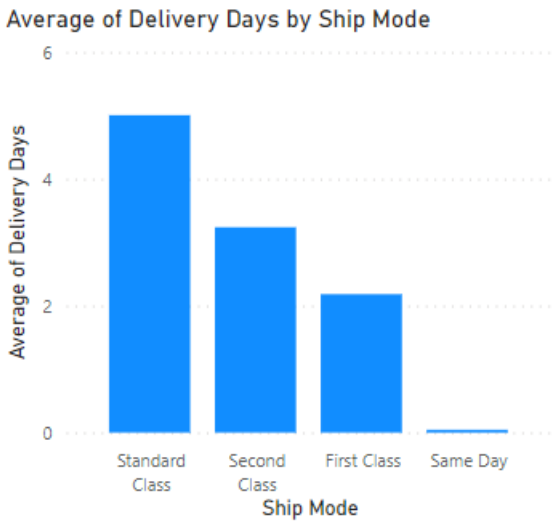
[Hint: Create a new measure, sales using calculate() and sum() function with the two filters, Low cart category and discount>=50% OR calculate sales using sumx() with if() and and()]

Sales from Low category and High Discount = CALCULATE(SUM(Orders[Sales]))

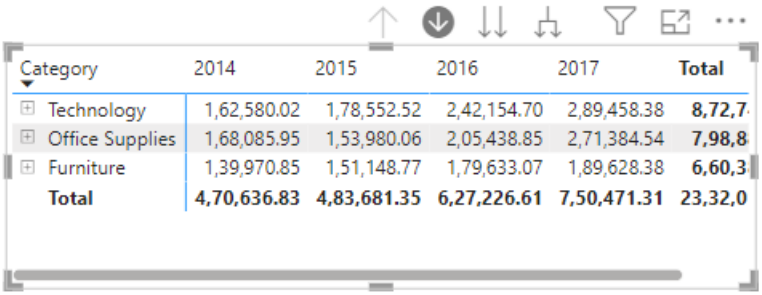
- 

6. Find out the number of days it takes to deliver for each shipment type (refer ship mode) so that delivery issues can be looked at on priority [Hint: No of days to deliver can be calculated from the difference between order_date and shipping_date]. Create a column chart that shows the average number of days it takes to deliver for each shipment type.

- Delivery Days = DATEDIFF(Orders[Order Date],Orders[Ship Date],DAY)



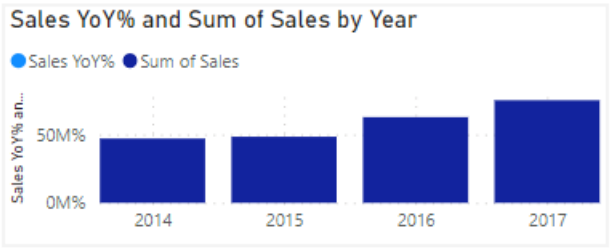
7. So far the store manager has managed to see the current snapshot of the sales based on various criteria. In the Retail business, do we see a spike in sales on special occasions like festivals? To achieve this, create a matrix visualization that displays order date as hierarchy , sales and sales year to date.



Category	2014	2015	2016	2017	Total
Technology	1,62,580.02	1,78,552.52	2,42,154.70	2,89,458.38	**8,72,7**
Office Supplies	1,68,085.95	1,53,980.06	2,05,438.85	2,71,384.54	**7,98,8**
Furniture	1,39,970.85	1,51,148.77	1,79,633.07	1,89,628.38	**6,60,3**
Total	**4,70,636.83**	**4,83,681.35**	**6,27,226.61**	**7,50,471.31**	**23,32,0**

-

8. Visualize the cumulative sales for each month for all the years to calculate Year on Year Sales Growth. Calculate YoY growth.

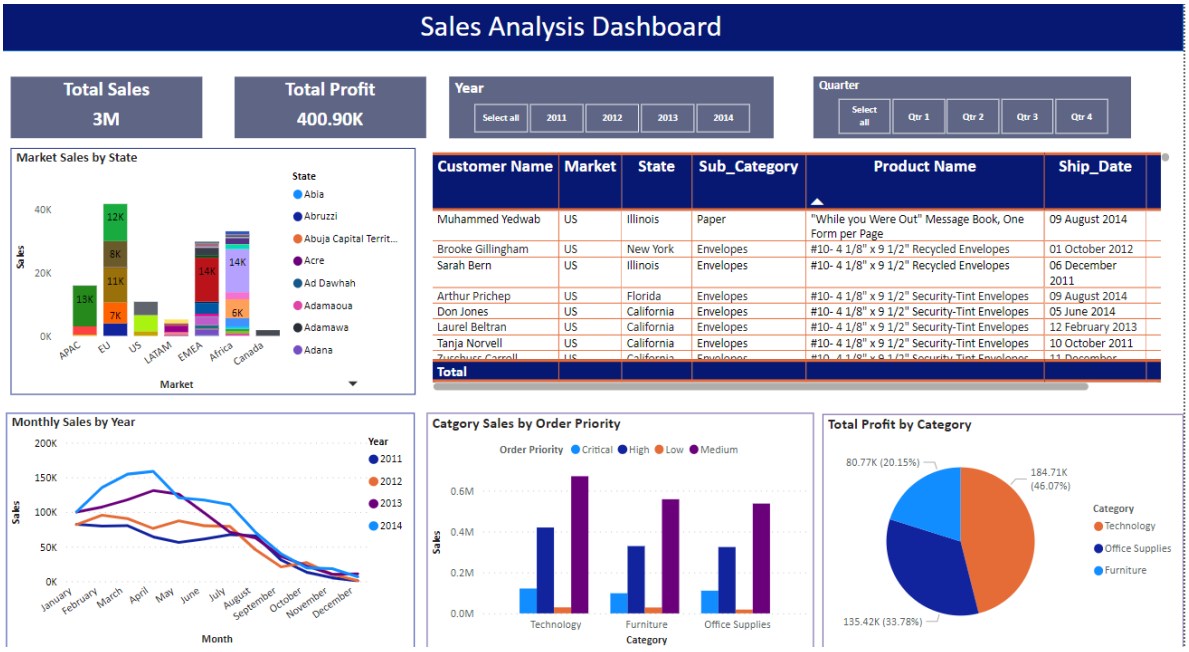


-

.

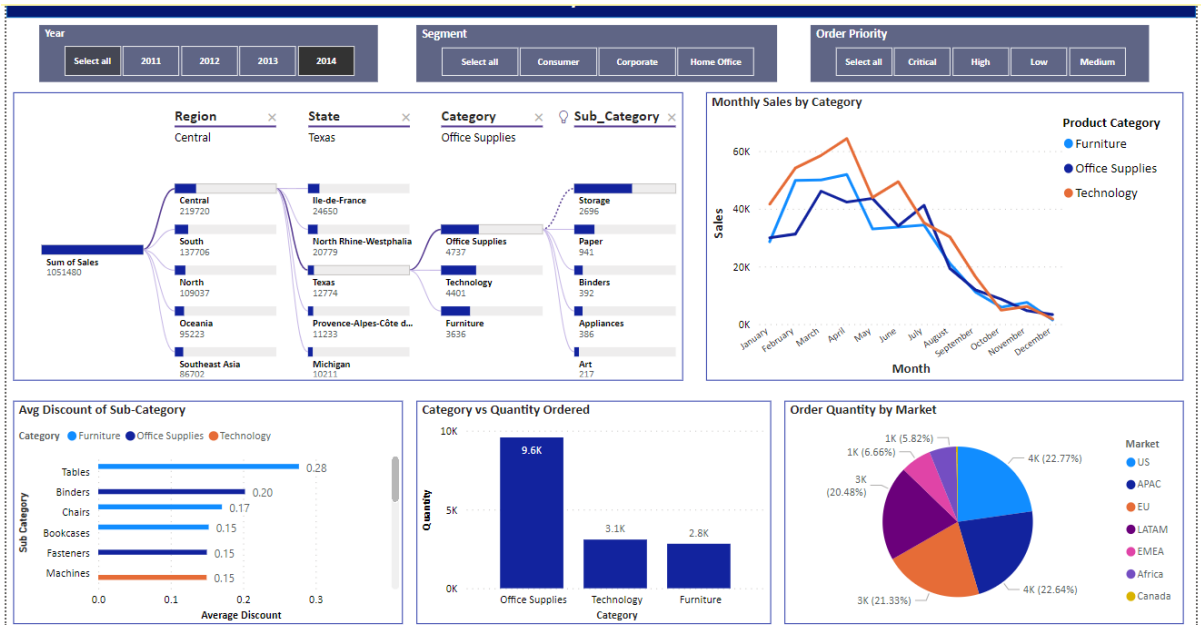
Sales Analysis Dashboard:

- Sales department can analyze sales and profit through this dashboard. Filter through years and quarterly sales.
- The company has made a Total Sales of 3 million and Total Profit of 400.9K between 2011- 2014.
- 1 million worth of sales and 119.11K profit was made in 2014.
- 46% of profit came from product category technology.
- Year 2014 March and April has seen high sales revenue and December being the holiday Season has the lowest revenue sales.



Product Analysis Dashboard:

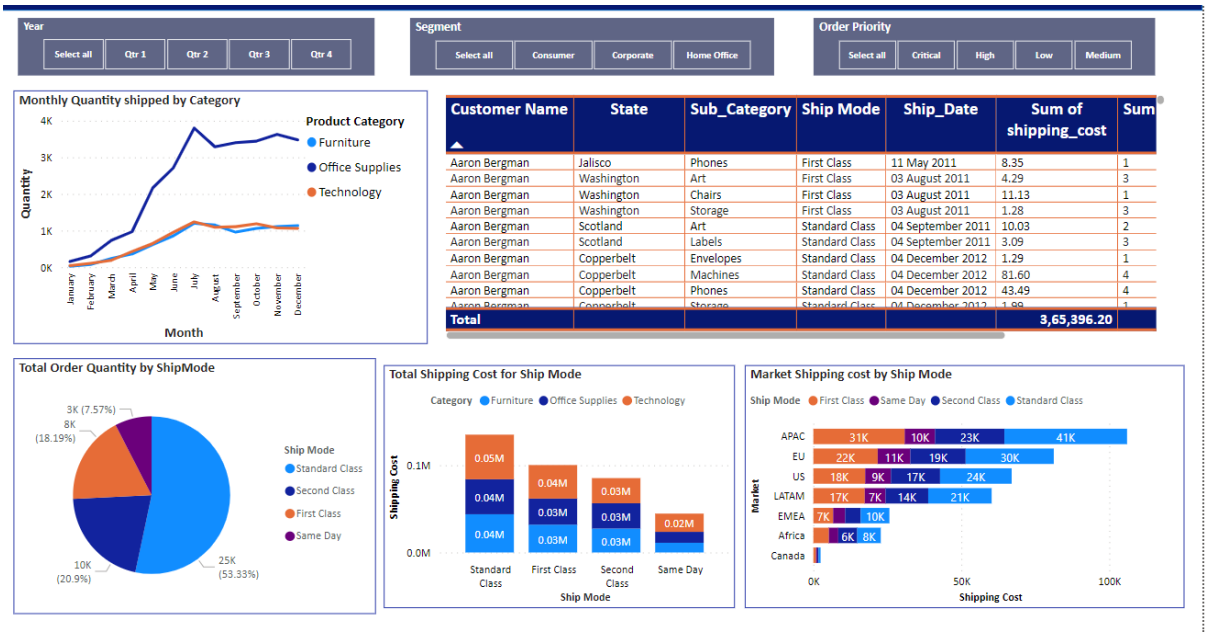
- This dashboard dive deep into product categories, quantity ordered, product-subcategory
- Highest amount of office supplies have been ordered in terms of quantity.
- Tables have been sold at an average discount of 28% in 2014



Shipping Analysis Dashboard:

- 53% of the orders have Standard shipping mode and lowest preferred is same day shipping.

- 41% of Orders with critical order priority have been ordered by first class ship mode.
- 100 % of the orders with low priority have been ordered through standard class.
- APAC markets have highest shipping cost, followed by EU and US.



Customer Name	State	Sub_Category	Ship Mode	Ship_Date	Sum of shipping_cost	Sum
Aaron Bergman	Jalisco	Phones	First Class	11 May 2011	8.35	1
Aaron Bergman	Washington	Art	First Class	03 August 2011	4.29	3
Aaron Bergman	Washington	Chairs	First Class	03 August 2011	11.13	1
Aaron Bergman	Washington	Storage	First Class	03 August 2011	1.28	3
Aaron Bergman	Scotland	Art	Standard Class	04 September 2011	10.03	2
Aaron Bergman	Scotland	Labels	Standard Class	04 September 2011	3.09	3
Aaron Bergman	Copperbelt	Envelopes	Standard Class	04 December 2012	1.29	1
Aaron Bergman	Copperbelt	Machines	Standard Class	04 December 2012	81.60	4
Aaron Bergman	Copperbelt	Phones	Standard Class	04 December 2012	43.49	4
Total					3,65,396.20	

Conclusion

We have seen how to build and publish a superstore sales and profit report in Power BI using the US superstore retail transaction data to get valuable insights about sales and profits. We understood how reports developed using its support of multiple data sources, are secure, and are easily scalable. We learned how we could build it for a real-time scenario. Below are the major takeaways:

- We understood what Power BI is and how we can get meaningful insights from data using its report.
- We have learned about the building blocks, such as visualizations, datasets, reports, dashboards, and tiles.
- We understood how to apply desktop data transformations using Power Query Editor.
- We have also seen how we can install its Desktop.

- We got a good understanding of how we can get data from multiple data sources in Power BI Desktop.
- We have also developed a report containing 3 pages based on superstore retail transaction data.
- ·We have learned how to shape data, drop unnecessary columns, filter, and sort columns in a real-world scenario.
- ·Apart from this, we saw how we could develop reports using built-in visuals such as cards, slicers, stacked bar charts, waterfall charts, maps, etc.