

Project Report

“Superstore Sale Analysis”

CONTENTS

1. <u>Introduction</u>	
1.1. Superstore	02
1.2. Data Analysis	03
1.3. Power BI.....	03
2. <u>Literature Survey</u>	
2.1. Why is Data Analysis Important	04
2.2. Power BI used for	05
2.3. Scope of Project.....	06
3. <u>System Analysis</u>	
3.1. Import the Dataset	06
3.2. Dataset Cleaning.....	07
3.3. Cleaning Steps	07
4. <u>Feasibility Study</u>	
4.1. Data Transformation.....	08
4.2. Power Query Editor.....	09
5. <u>Visualization Charts</u>	
5.1. Stacked Column Chart.....	09
5.2. Pie Chart.....	11
5.3. Line Chart	12
5.4. Table	13
5.5. Map	14
5.6. Donut Chart	16
5.7. Stacked Area Chart	17
5.8. Tree Map	19
5.9. Funnel Chart	20

6. Implementation Phase

6.1 Analysis	22
6.2 Screen Layouts	23
6.3 Dashboard	29

7. <u>Conclusion and discussion</u>	31
---	----

1. Introduction

Superstore



A **superstore** is a large store traditionally aimed at meeting consumers' total needs for routinely purchased food products, non-food items, and services.

Superstores are large-scale retail establishments designed to meet consumers' comprehensive needs under a single roof.

They go beyond the traditional supermarket model by offering an extensive range of products and services. With their expansive layouts and diverse offerings, superstores aim to provide convenience and efficiency to shoppers.

Understanding the Superstore

Superstores have revolutionized the retail industry by providing consumers with a one-stop shopping experience. They are mammoth-sized retail establishments designed to cater to the diverse needs of consumers. These retail giants take the concept of convenience to new heights by offering an extensive array of products and services, all housed within a single expansive space.

1.2 Advantages of Superstore

Superstores have redefined how consumers shop by providing a comprehensive and convenient retail experience.



Convenience

Superstores are synonymous with convenience. They eliminate the need for customers to visit multiple stores to fulfil their shopping needs. By providing a comprehensive range of products and services, superstores save time and effort for consumers, who can find everything they require in one place. This aspect mainly benefits busy individuals and families seeking a streamlined shopping experience.

Competitive Pricing

The combination of grocery and non-food offerings in superstores often leads to cost savings for consumers. Supercenters, for instance, offer discounted non-food products alongside competitively priced groceries, making them an attractive option for budget-conscious shoppers. Moreover, category killers frequently leverage their specialized market dominance to negotiate favorable prices with suppliers, passing on the savings to customers.

Extensive Product Selection

Superstores excel in providing an unparalleled selection of products. Their vast floor spaces allow for a broad assortment, enabling customers to choose from numerous brands, styles, and variants. Whether it's groceries, clothing, electronics, or household goods, superstores offer an extensive range of choices to cater to diverse consumer preferences.

Enhanced Customer Experience

Category killers set themselves apart by employing knowledgeable staff members well-versed in their respective product domains. These experts can guide customers through the available options, provide recommendations, and address queries effectively. Such personalized assistance enhances the overall customer experience, fostering loyalty and trust.

1.3 Data Analysis

Data analysis is the process of cleaning, changing, and processing raw data and extracting actionable, relevant information that helps businesses make informed decisions. The procedure helps reduce the risks inherent in decision-making by providing useful insights and statistics, often presented in charts, images, tables, and graphs.



1.4 Power BI

Power BI is a powerful data visualization and analytics tool that can help you quickly make sense of your data by extracting it from different data sources.

Power BI is a data visualization platform used primarily for business intelligence purposes. Designed to be used by business professionals with varying levels of data knowledge, Power BI's dashboard is capable of reporting and visualizing data in a wide range of different styles, including graphs, maps, charts, scatter plots, and more.

2. LITRATURE SURVEY

2.1 Why is Data Analysis Important



Better Customer Targeting: You don't want to waste your business's precious time, resources, and money putting together advertising campaigns targeted at demographic groups that have little to no interest in the goods and services you offer. Data analysis helps you see where you should be focusing your advertising and marketing efforts.

- **You Will Know Your Target Customers Better:** Data analysis tracks how well your products and campaigns are performing within your target demographic. Through data analysis, your business can get a better idea of your target audience's spending habits, disposable income, and most likely areas of interest. This data helps businesses set prices, determine the length of ad campaigns, and even help project the number of goods needed.
- **Reduce Operational Costs:** Data analysis shows you which areas in your business need more resources and money, and which areas are not producing and thus should be scaled back or eliminated outright.
- **Better Problem-Solving Methods:** Informed decisions are more likely to be successful decisions. Data provides businesses with information. You can see where this progression is leading. Data analysis helps businesses make the right choices and avoid costly pitfalls.

- **You Get More Accurate Data:** If you want to make informed decisions, you need data, but there's more to it. The data in question must be accurate. Data analysis helps businesses acquire relevant, accurate information, suitable for developing future marketing strategies, business plans, and realigning the company's vision or mission

2.2 Power BI used for

Whether you're a data pro or are just entering the business world, Power BI is designed to empower you with data-driven insights. Some of the most common uses for the platform include:

- Creating reports and dashboards that present data sets in multiple ways using visuals
- Connecting various data sources, such as Excel sheets, onsite data warehouses, and cloud-based data storage, and then transforming them into business insights
- Turning data into a wide range of different visuals, including pie charts, decomposition trees, gauge charts, KPIs, combo charts, bar and column charts, and ribbon charts – among many other options
- Providing company-wide access to data, data visualization tools, and insights in order to create a data-driven work culture

2.3 Scope of Project

1. **Geographic Scope:** Define the specific regions or countries where the superstore project will be implemented. This could involve opening new branches in certain locations or establishing an online presence that caters to a global audience.
2. **Product and Service Offerings:** Specify the types of products or services the superstore will provide. This could range from groceries and household items to electronics, clothing, and other retail categories.

3. **Technology and Infrastructure:** Describe the technology and infrastructure requirements for the project. This might involve setting up IT systems, POS (Point of Sale) systems, online platforms, and any other technology necessary for smooth operations.
4. **Timeline:** Provide an estimated timeline for the project, including key milestones and deadlines. This helps in tracking progress and managing expectations.
5. **Budget:** Outline the financial resources allocated to the superstore project. This includes initial investments, operational costs, marketing expenses, and any other relevant financial considerations.
6. **Regulatory Compliance:** Address any legal or regulatory requirements that the project must adhere to. This could involve obtaining permits, meeting safety standards, and complying with local business regulations.

Stakeholders:

Identify the key stakeholders involved in the project. This may include investors, project managers, employees, customers, and any other parties with a significant interest in the project.

Risks and Mitigation Strategies:

Identify potential risks that could impact the success of the project and propose strategies for mitigating or managing these risks.

Success Criteria:

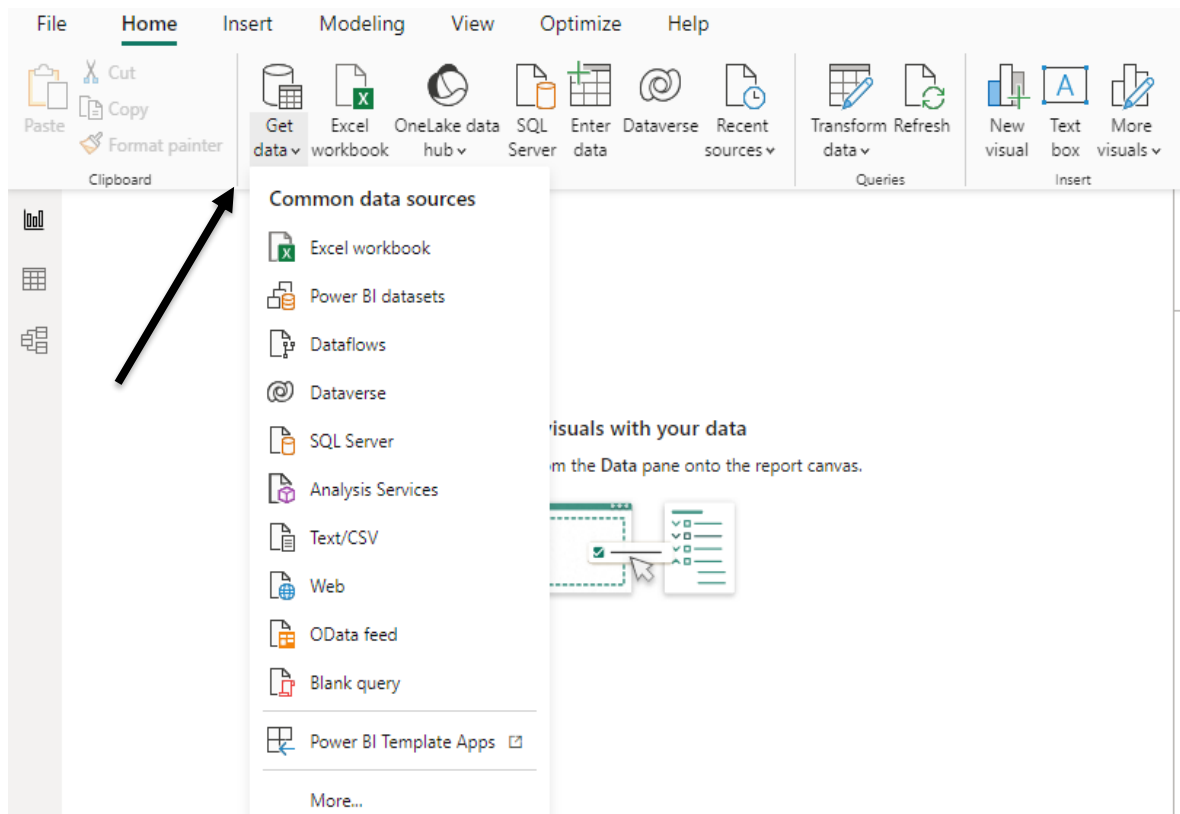
Define the criteria that will be used to measure the success of the superstore project. This could include financial metrics, customer satisfaction levels, and other key performance indicators.

3. SYSTEM ANALYSIS

3.1 Import the Dataset

Initially, we identify the type of data we are working with, such as Excel data, SQL Server

data, web data, or Power BI datasets. Once determined, we proceed to load and transform the data, initiating the cleaning process as the first step.



3.2 Dataset Cleaning

There are some steps we use to clean data in Power BI, which are helpful in creating meaningful insights.

1. Use the first row as a header.
2. Remove duplicate rows.
3. Remove duplicate and null columns.

4. Remove blank rows.

5. Check for null and blank values; if they exist, delete them.

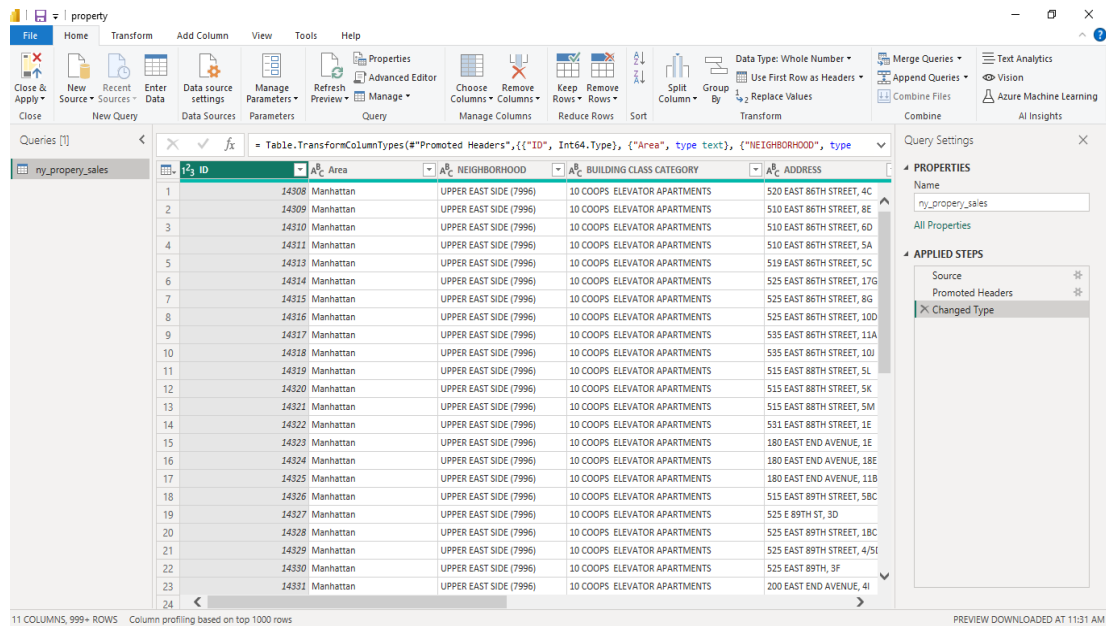
6. Check data types; always convert numerical values into decimal or fixed decimal values. 7. Check date data types and convert them into the US and UK time zones using their respective locales

3.3 Cleaning Steps

Here, I commence the cleaning process for H1B visa data. The initial steps include:

1. Removing duplicate and blank rows.
2. Checking for null and blank values and deleting them if they exist.
3. Converting the data types of each column; numerical values to decimals & fixed decimal, and dates are adjusted to their local zones, such as US and UK.

Following these steps, I close and apply the transformed dataset, where I proceed to conduct insightful analysis.



4. Feasibility

4.1 Data Transformation:

Power BI Desktop has three views:

Report view – You can use queries that you create to build compelling visualizations, arranged as you want them to appear, and with multiple pages, that you can share with others.

Data view – See the data in your report in data model format, where you can add measures, create new columns, and manage relationships.

Model view – Get a graphical representation of the relationships that are established in your data model, and manage or modify them as needed.

Access these views by selecting one of the three icons along the left side of Power BI Desktop. In the following image, Report view is selected, indicated by the yellow band beside the icon.

Power BI Desktop also comes with Power Query Editor. Use Power Query Editor to connect to one or many data sources, shape and transform the data to meet your needs, then load that model into Power BI Desktop.

4.2 Power Query Editor

To get to Power Query Editor, select Transform data from the Home tab of Power BI Desktop.

With no data connections, Power Query Editor appears as a blank pane, ready for data.

After a query is loaded, Power Query Editor view becomes more interesting. If you connect to a Web data source using the New Source button in the top left, Power Query Editor loads information about the data, which you can then begin to shape.

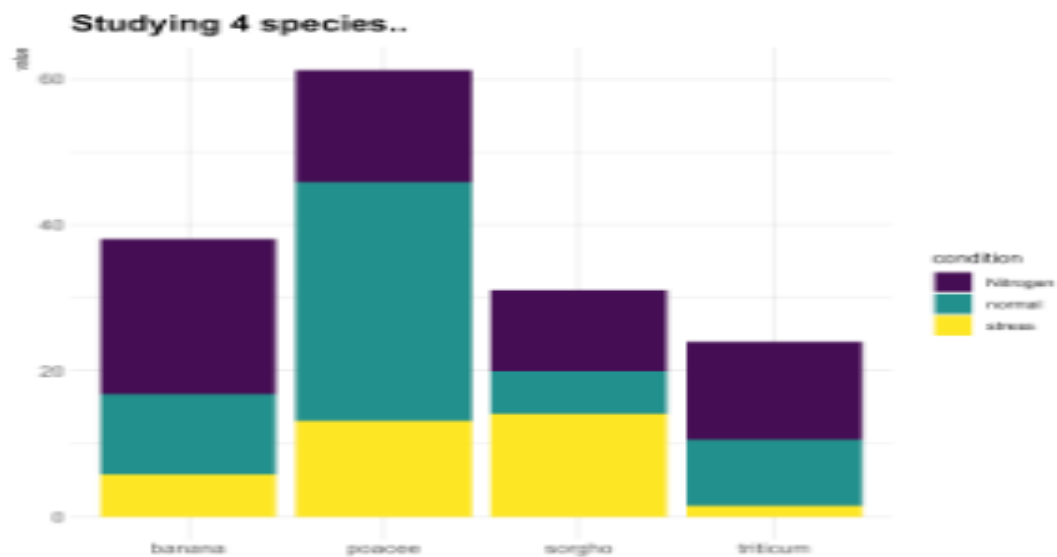
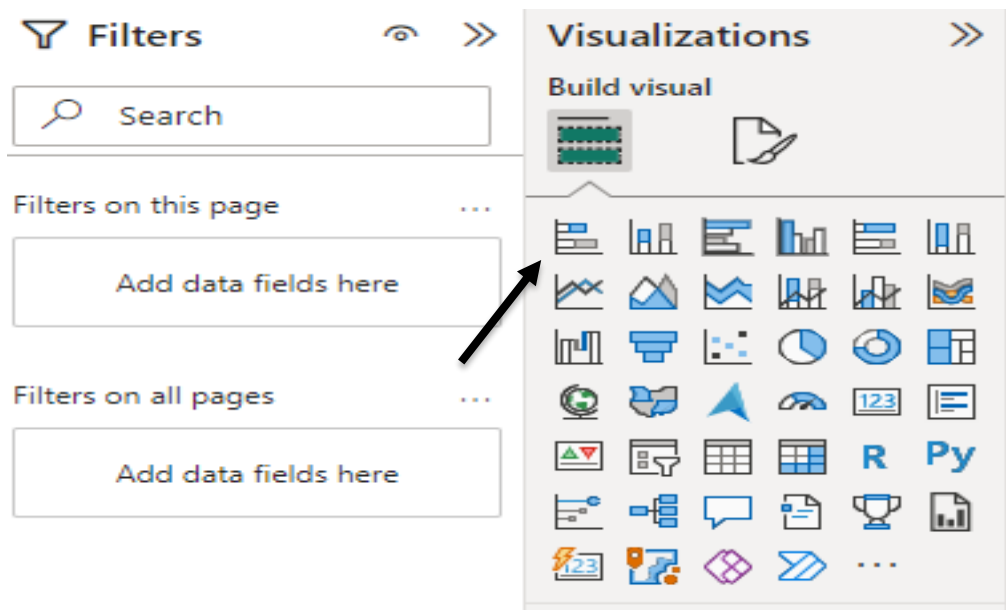
To connect to data and begin the query building process, select New Source. A menu appears, providing the most common data sources.

Advanced Editor

The Advanced Editor lets you see the code that Power Query Editor is creating with each step. It also lets you create your own code in the Power Query M formula language. To launch the advanced editor, select View from the ribbon, then select Advanced Editor. A window appears, showing the code generated for the selected query.

5. Visualizations

5.1 Stacked Column Chart & Stacked Bar Chart



Both are most usable visuals in Power BI. Stacked Column Chart is useful to compare multiple dimensions against a single measure. In a stacked column chart, the vertical axis represents the numerical values of the data, while the horizontal axis displays the categories or time periods.

Stacked column charts are useful for comparing the total values of different groups, as well as for identifying the contribution of each group to the overall total.

How to formatting Stacked bar Chart?

Visual Tab:

X-axis : Mange the X-axis Value font size, color, Title and Display Unit etc.

Y-axis: Mange the Y-axis Value font size, color and Title.

Legend: Specified the Legend Text color, font size, Position & Title.

Bars: Change the color of chart & maintain the spacing between bars.

Data Labels: Enable the data labels on chart, manage the display unit of labels.

Total Labels: Enable Total labels on chart.

General Tab:

Properties: In this section you can manage the chart height, width, horizontal & vertical position.

Title: Under general tab you can see the title section> Here you can set the below properties-

Title- Specified the title for chart, & manage the font size, color, background for chart.

Subtitle- Specified the subtitle for chart, & manage the font size, color, background for chart.

Divider- Enable the line between Title & chart.

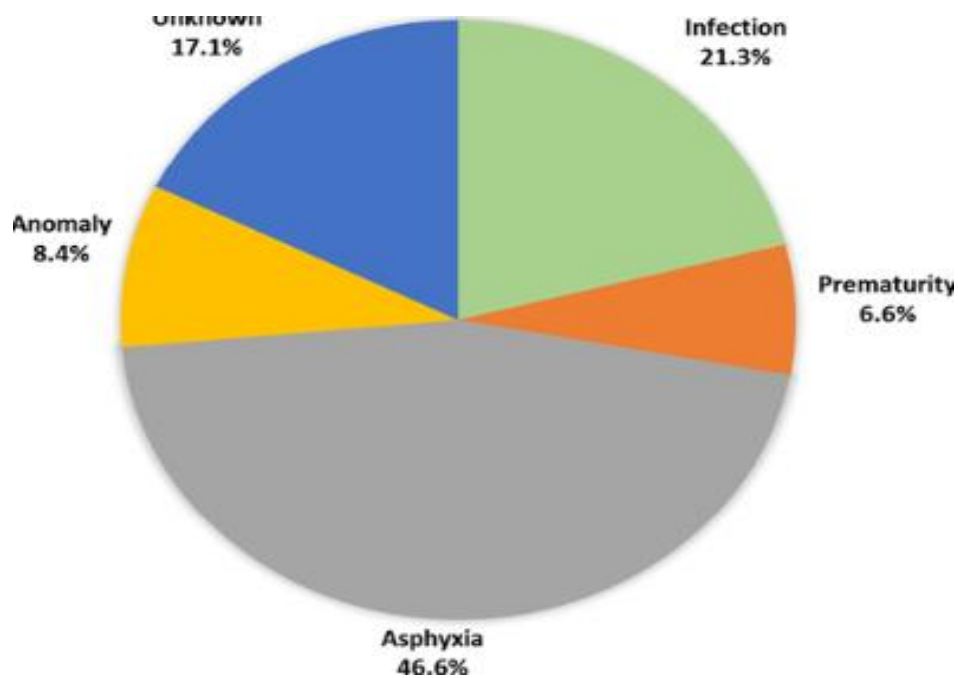
Spacing- Manage the space between title, Subtitle & chart area.

Key features of a Stacked Bar Chart in Power BI include:

- **Multiple Categories:** Each bar on the chart represents a specific category, and the individual segments within the bar represent the subcategories or components.
- **Stacked Segments:** Subcategories are stacked on top of each other within a single bar, illustrating the contribution of each subcategory to the total of that category.
- **Comparison and Composition:** Stacked bar charts allow you to compare both the overall distribution of categories and the distribution of subcategories within each category.
- **Color Coding:** Different subcategories are typically assigned distinct colors to aid in differentiation.
- **Legend:** The legend helps identify the subcategories and their colors.
- **Tooltip:** Hovering over a segment provides detailed information about the value it represents

5.2 Pie Chart in Power BI

Pie Chart in Power BI is a built-in visualization chart available with all versions of Power BI. The pie chart is a round-shaped circle chart where each category data set is shown in a pie shape based on the value of each data label. In addition, the pie chart of each category is adjusted against the overall portion of the data labels.



Legend: This is nothing based on what column values we need to see the pie chart. In our example, we need to see the “Region-wise” pie chart. Our legend will be the “Region” column from “Pie Table.”

Details: If you want to show any further partition of the data you can add here, we will come back to this in the second example.

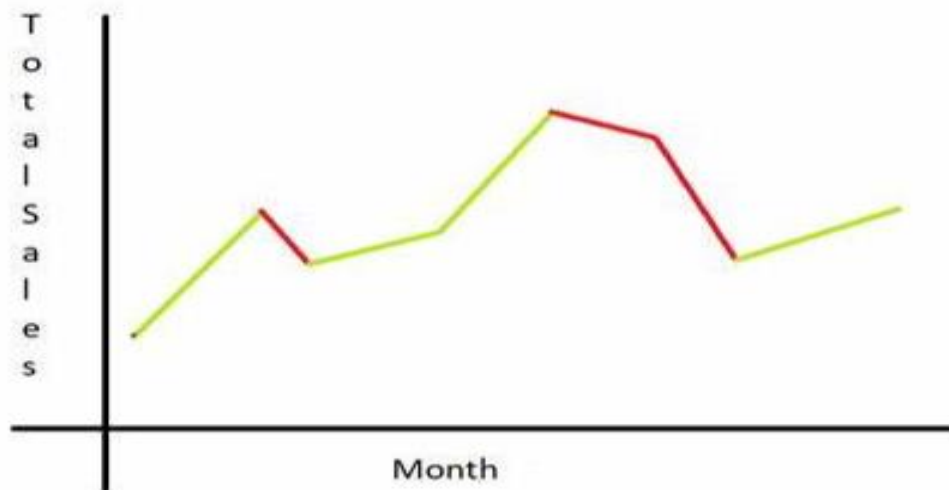
Values: This is nothing but what numerical value we need to show in the pie circle.

Tooltips: This is nothing but adding an extra element to the pie chart when we hover on any of the slices of the pie.

5.3 Line Chart in Power BI

A line chart is a sequence of data points defined by dots and joined by straight lines. A line chart may include one or multiple lines. It is one of the visuals to display the data in the Power BI report in a visually immersive and interactive manner.

It is used to define continuous data sets and the line charts maintain an X and a Y-axis in Power BI.



Multiple lines on a Line chart in Power BI

Let us see how to display multiple lines plotted on a Line Chart visual in Power BI,

Yes, it is possible to display the line chart with multiple lines plotted on it.

To achieve this, follow the below steps:

- Initially, make sure that you have loaded the data source to the Power BI desktop.
- Once the data is loaded you can confirm the data under the Fields section and the fields pane contains the column data presented in the source data.
- Now under the Visualizations pane, select the Line Chart option, you can see that the line chart is added to the report canvas.

5.4 Tables in Power BI reports

EMPLOYER_NAME	Average of PREVAILING_WAGE
OMPI OF AMERICA INC.	₹ 212,987,840
UROLOGY TODAY, LLC	₹ 173,139,200
INNOBOX SYSTEMS	₹ 172,363,360
BAALAAJEE INC.	₹ 143,440,766
DEPARTMENT OF VETERANS AFFAIRS/NMVAHCS	₹ 134,702,089.5
BAYSHORE GLOBAL MANAGEMENT LLC	₹ 126,235,824
MANHATTAN SOFTWARE, INC.	₹ 123,907,680
GYM-PACT CORP	₹ 123,389,760
BOOKRENTER.COM, INC.	₹ 123,032,124.8
CATANIA, MAHON, MILLIGRAM & RIDER, PLLC	₹ 122,697,120
AGROPLASMA, INC	₹ 122,133,440
RAMONA MEDICAL CLINIC	₹ 121,846,712
AQUA LUNG AMERICA INC.	₹ 118,110,720
BRONX LIGHTHOUSE CHARTER SCHOOL	₹ 112,424,000
E AND D MEDIA INC.	₹ 109,824,032.6667
SHELBY HOSPITAL LCC	₹ 109,250,419
TAMPA BAY DIGITAL DESIGN EXPERTS, LLC	₹ 107,043,518.4
NUVIEW SYSTEMS, INC.	₹ 102,823,250.5
XANTREX TECHNOLOGY, INC.	₹ 101,957,554.5
PETRO MANAGEMENT INC.	₹ 101,920,000
OUTPOST RESEARCH AND DEVELOPMENT LLC	₹ 94,641,799
Total	₹ 146,998.4319

A table is a grid that contains related data in a logical series of rows and columns. A table can also contain headers and a row for totals. Tables work well with quantitative comparisons where you're looking at many values for a single category. In the following example, the table

displays five different measures for the Category items, including average prices, year over year sales, and sales goals.

Power BI helps you create tables in reports and cross-highlight elements within the table with other visuals on the same report page. You can select rows, columns, and even individual cells, and then cross-highlight the values. You can also copy and paste individual cells and multiple cell selections into other applications.

When to use a table

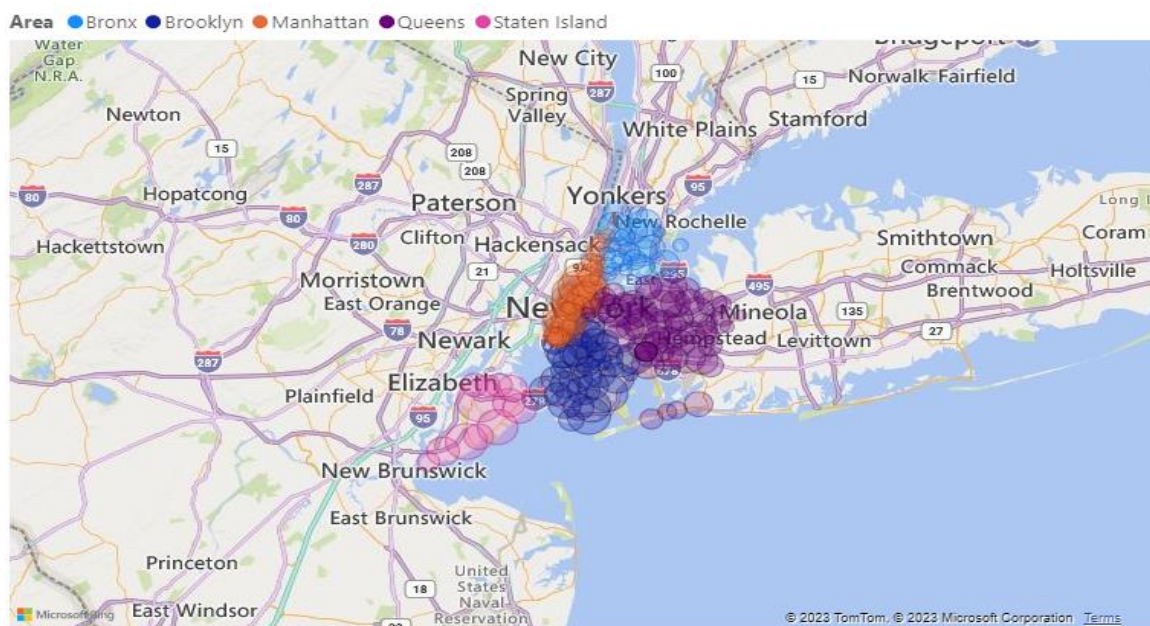
Tables are a great choice for several scenarios:

Represent numerical data by category with multiple measures.

Display data as a matrix or in a tabular format with rows and columns.

Review and compare detailed data and exact values rather than visual representations.

5.5 Map visualization



The map visualization in Power BI is a useful tool for analyzing spatial data. It allows you to create a map that represents your data in a way that is easy to understand and visually compelling. The map visualization works by integrating with Bing Maps, which provides a rich set of map styles and geographic data.

The map visualization gives you the ability to add multiple layers and data points to your map, customize the map's appearance, and create drill-down hierarchies for deeper analysis.

One of the key features of the map visualization in Power BI is the ability to use custom map layers.

This means that you can import your own geographic data. Additionally, the map visualization supports a wide range of data types, including latitude and longitude coordinates, addresses, and even postal codes, making it a versatile tool for spatial analysis.

1. Categorize geographic fields

In Power BI Desktop, you can ensure fields are correctly geo-coded by setting the Data Category on the data fields. In Data view, select the desired column. From the ribbon, select the Column tools tab and then set the Data Category to Address, City, Continent, Country, County, Postal Code, State, or Province.

2. Use more than one location column

Sometimes, even setting the data categories for mapping isn't enough for Bing to correctly guess your intent. Some designations are ambiguous because the location exists in multiple countries/regions. For example, there's a Southampton in England, Pennsylvania, and New York.

3. Use specific Latitude and Longitude

Add latitude and longitude values to your semantic model. This data removes any ambiguity and returns results more quickly. Latitude and Longitude fields must be in Decimal Number format, which you can set in the data model.

4. Use Place category for columns with full location information

While we encourage you to use geo-hierarchies in your maps, if you must use a single location column with full geographical information, you can set the data categorization to Place. For example, if the data in your column is full addresses, such as 1 Microsoft Way, Redmond Washington 98052, this generalized data category works best with Bing.

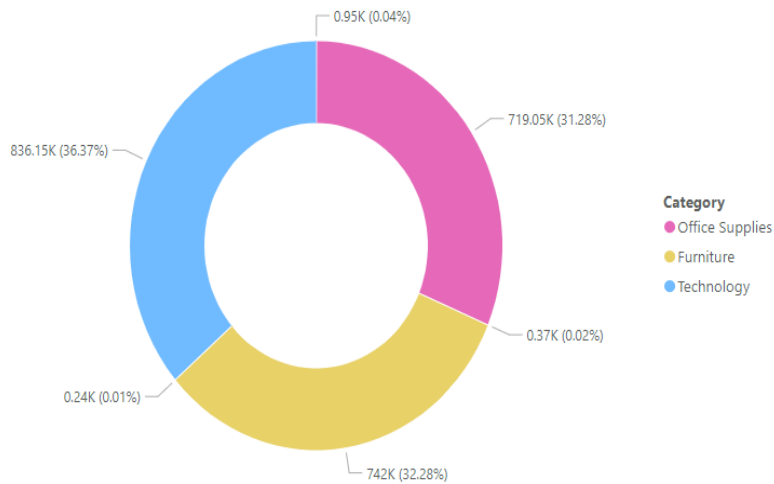
In Power BI: tips to get better results when using map visualizations

1. Use latitude and longitude fields (if they exist)

In Power BI, if the semantic model you are using has fields for longitude and latitude--use them! Power BI has special buckets to help make the map data unambiguous. Just drag the field that contains your latitude data into the Visualizations > Latitude area. And do the same for your longitude data. When you add this data, you also need to fill the Location field when creating your visualizations. Otherwise, the data is aggregated by default, so for example, the latitude and longitude would be paired at the state level, not the city level.

5.6 Donut Chart

A donut chart is a circular chart, which could present values of a dataset in the form of slices of a donut. The donut chart is exactly the same as a pie chart, the only difference is pie chart has a circle, but a donut chart has a hole in the circle. We have various options to format donut charts, we can change the value of the legends, rotation, detail labels, etc. In this article, we will learn how to format a donut chart in Power BI and explore its various options.



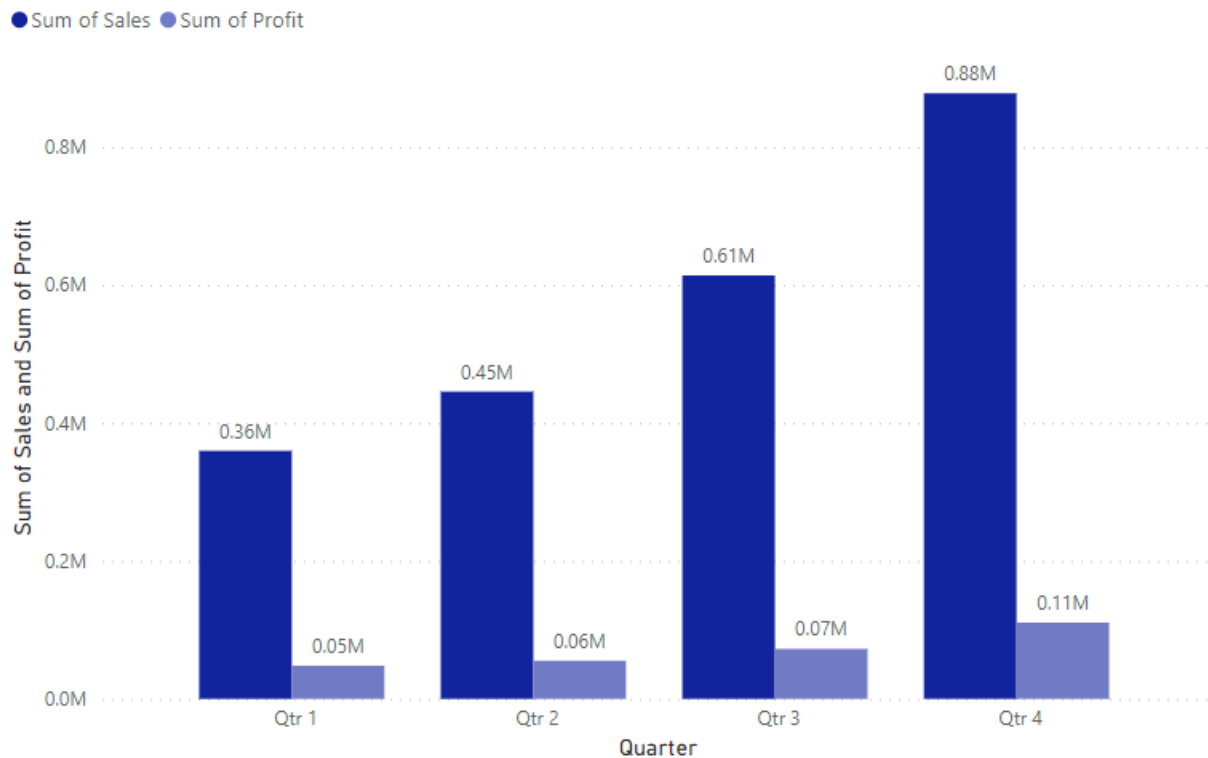
Formatting a Donut Chart In Power BI

After the successful, creation of a donut chart in Power BI, we have multiple options to format it. For example, adding the title to the chart, changing the color, and position of the chart, and adding tooltips, slicer colors, and detail labels to the chart.

6. Implementation Phase:

6.1 Analysis:

1. What are the total profits and total sales per quarter?

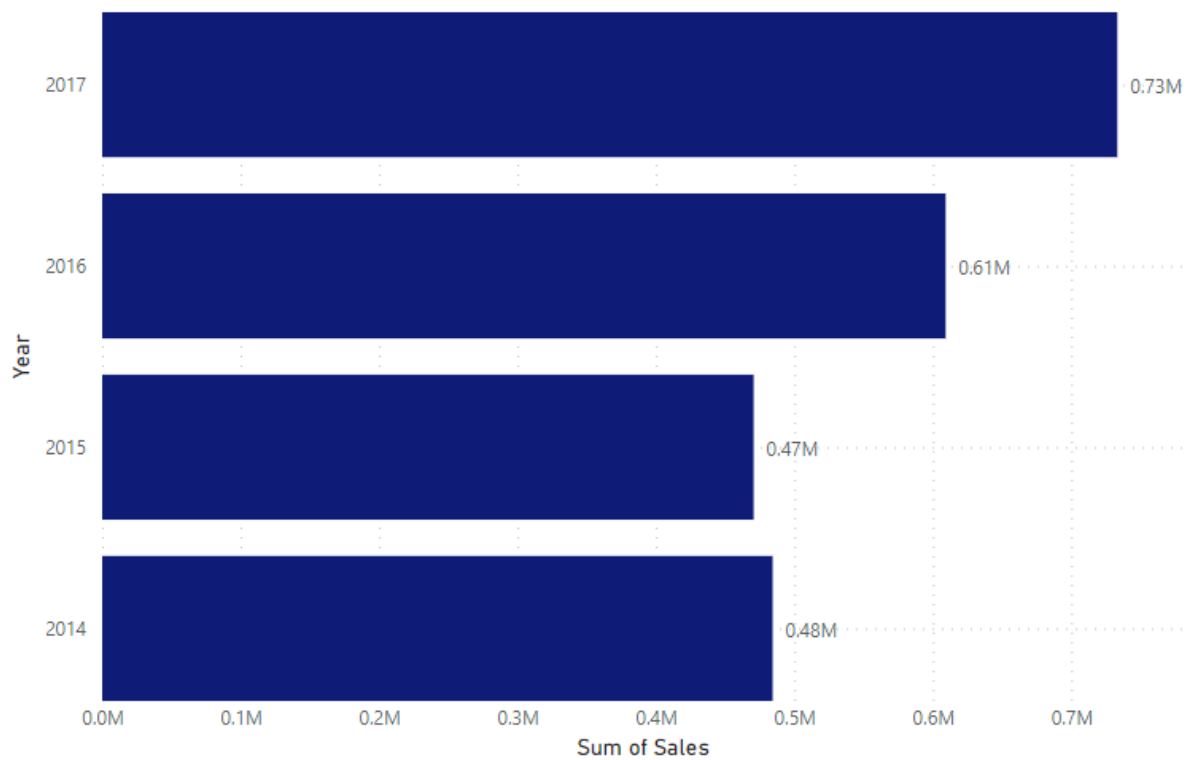


Analysis:

Firstly, input the sales and profits by quarter. Finally, observe the results to identify the sum of total profits and sum of total profit per quarter.

Hence, the result allows us to identify the total profits and total profit per quarter.

2. What are total sales and total profits of each year?

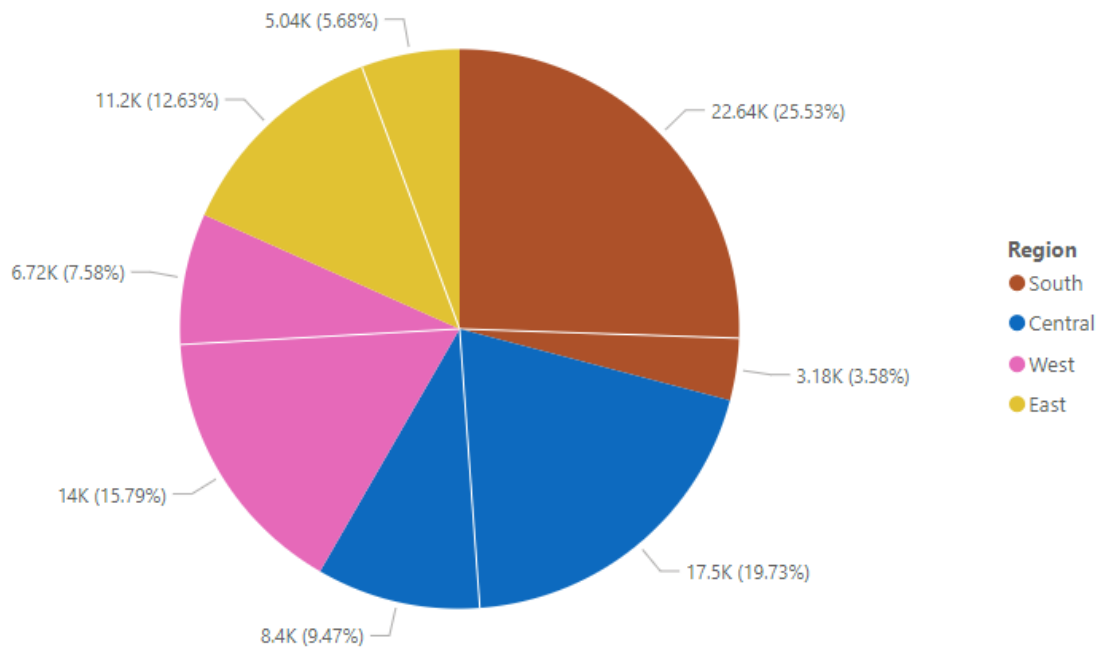


Analysis:

Here, input the sales and profits by year. Finally, observe the results to identify the sum of total profits and sum of total profit by year.

Hence, the result allows us to identify the total profits and total profit per year.

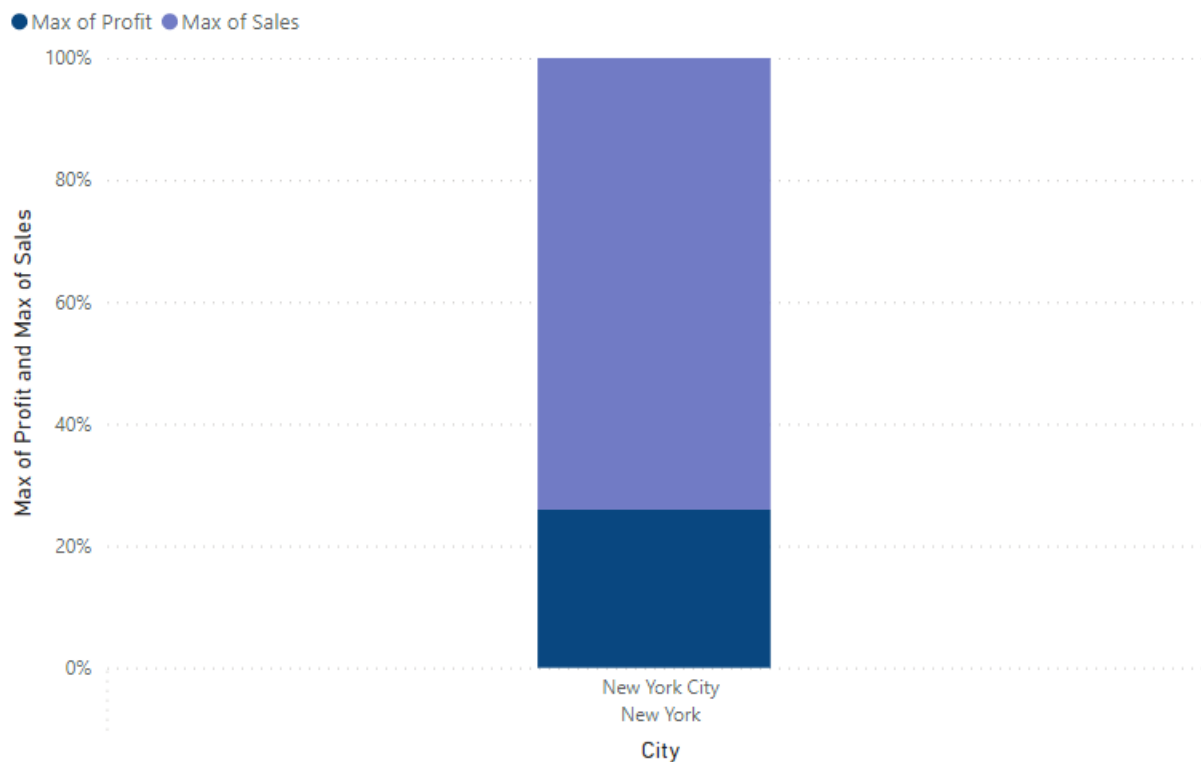
3. What region generates the highest sales and profits?



Analysis:

I used a stacked column chart with a column for the region, sales and identified the sum of sales and sum of profits by region and I find what region generate the highest sales and profit.

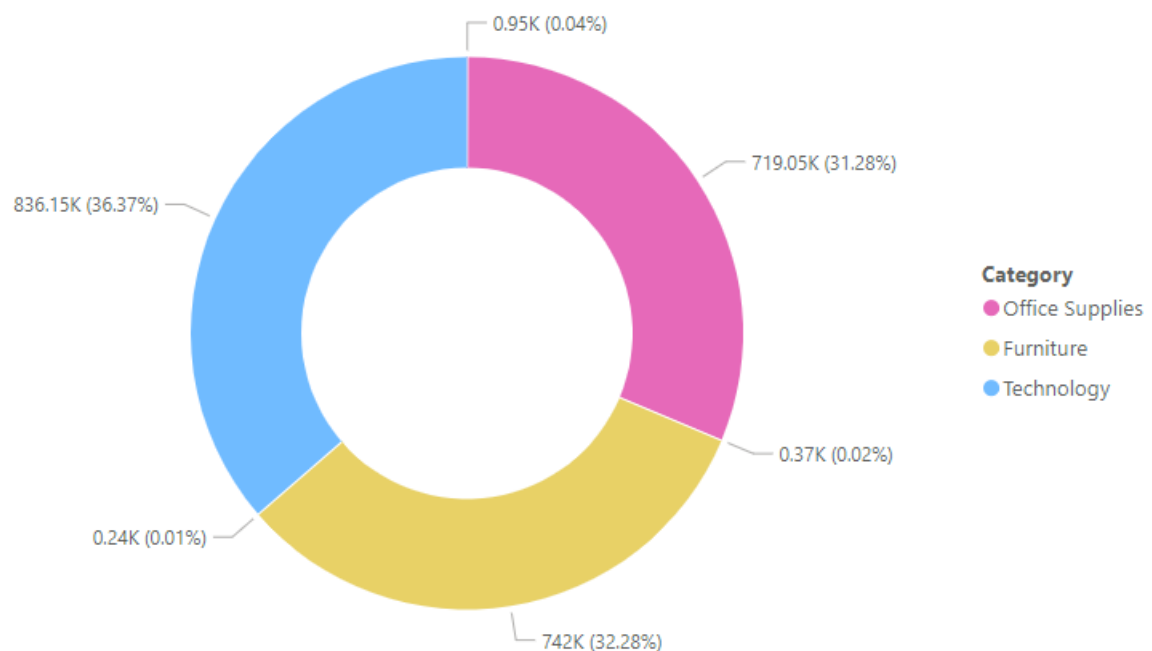
4. What state and city brings in the highest sales and profits?



Analysis:

I used a Map with a column for city and states. I analyzed the sum of sales and sum of profit by state and city.

5. The relationship between discount and sales and the total discount per category



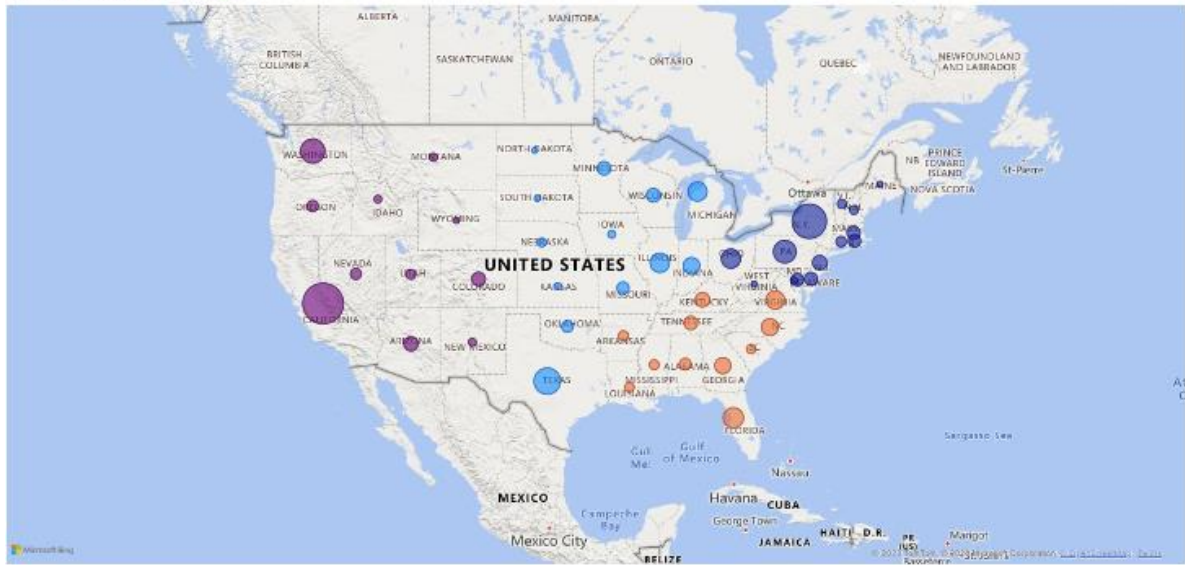
Analysis:

Here, we have a column for sales, category and the total discount. We identified the relationship between discount and sales and the total discount per category.

6. What category generates the highest sales and profits in each region and state?

Sum of Sales and Sum of Profit by State and Region

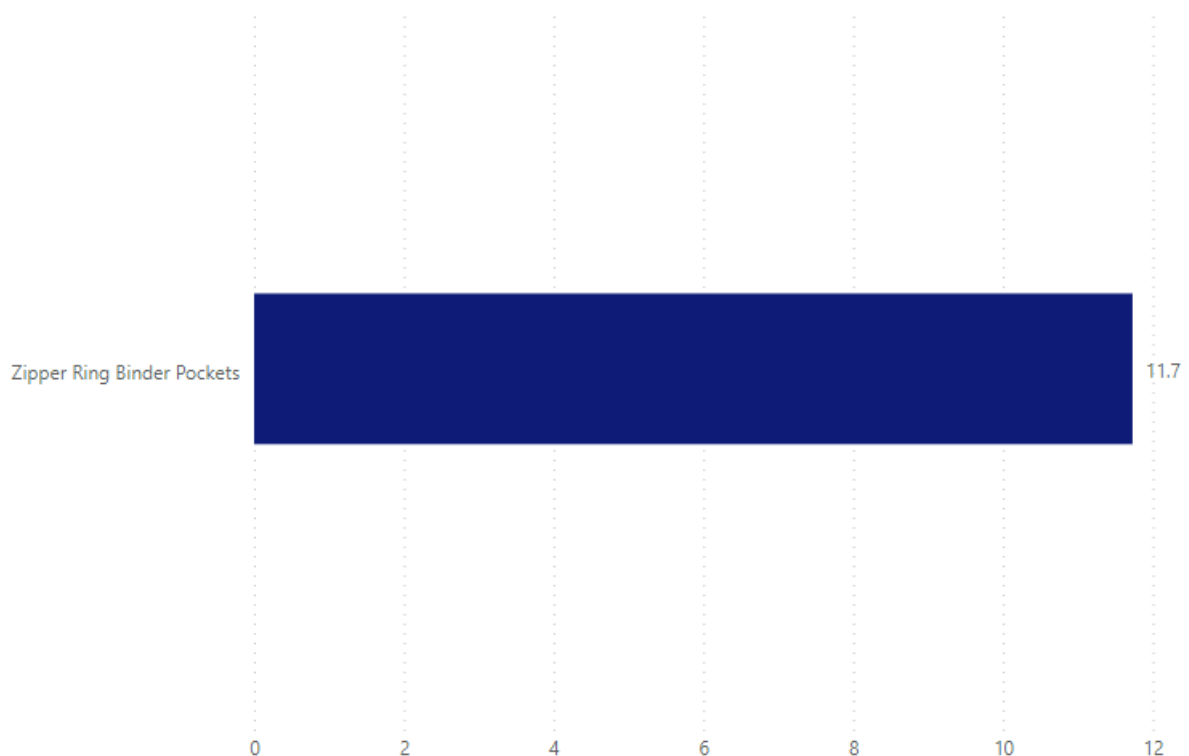
Region ● Central ● East ● South ● West



Analysis:

Here, I have columns for category, sales, profits and region, state and I identified the category generates the highest sales and profits in each region and state.

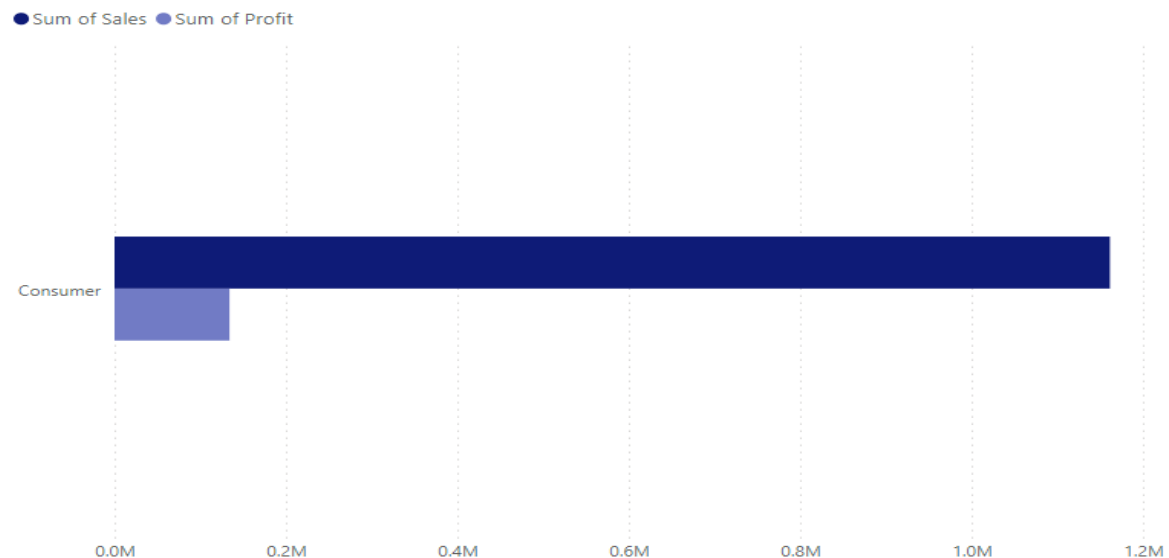
8. What are the names of the products that are the most and least profitable to us?



Analysis:

I used a clustered column chart with a column for products and profit. I analyzed the names of the products are least profitable.

8. What segment makes the most of our profits and sales?

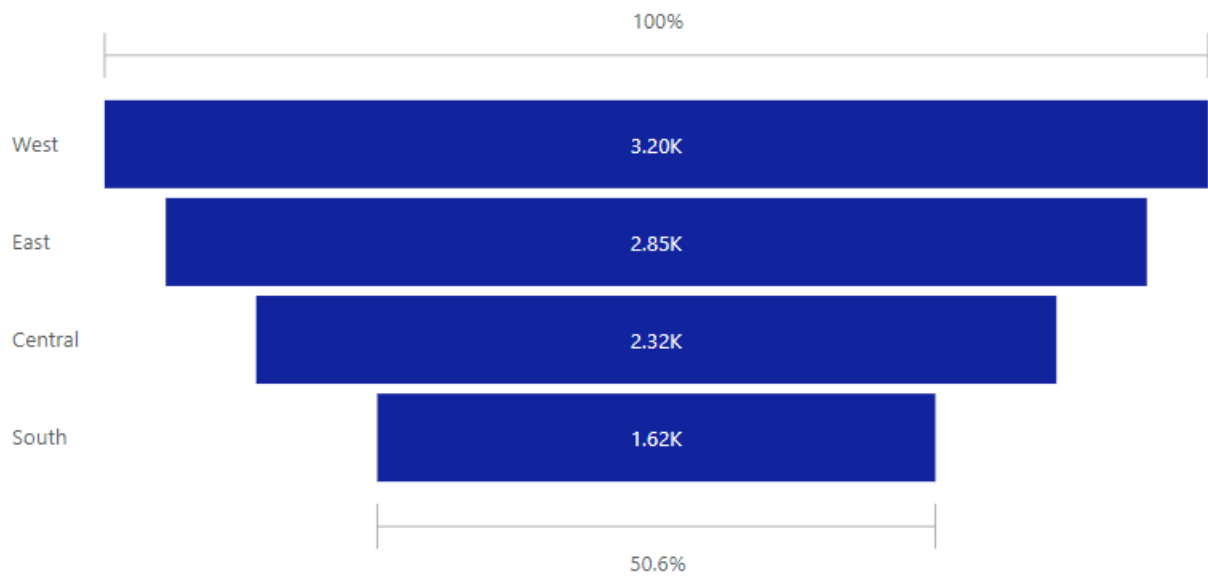


Segment	Sum of Sales	Sum of Profit
Consumer	1,161,401.34	134,119.33
Corporate	706,146.44	91,979.45
Home Office	429,653.29	60,299.01
Total	2,297,201.07	286,397.79

Analysis:

Firstly, input the sum of sales or profit and the segment. Next, apply basic filtering by selecting the "Top N" option and entering the value 1. Then, drag the sum of profit into the designated area. Finally, observe the results to identify the segment is most of our sales and profit.

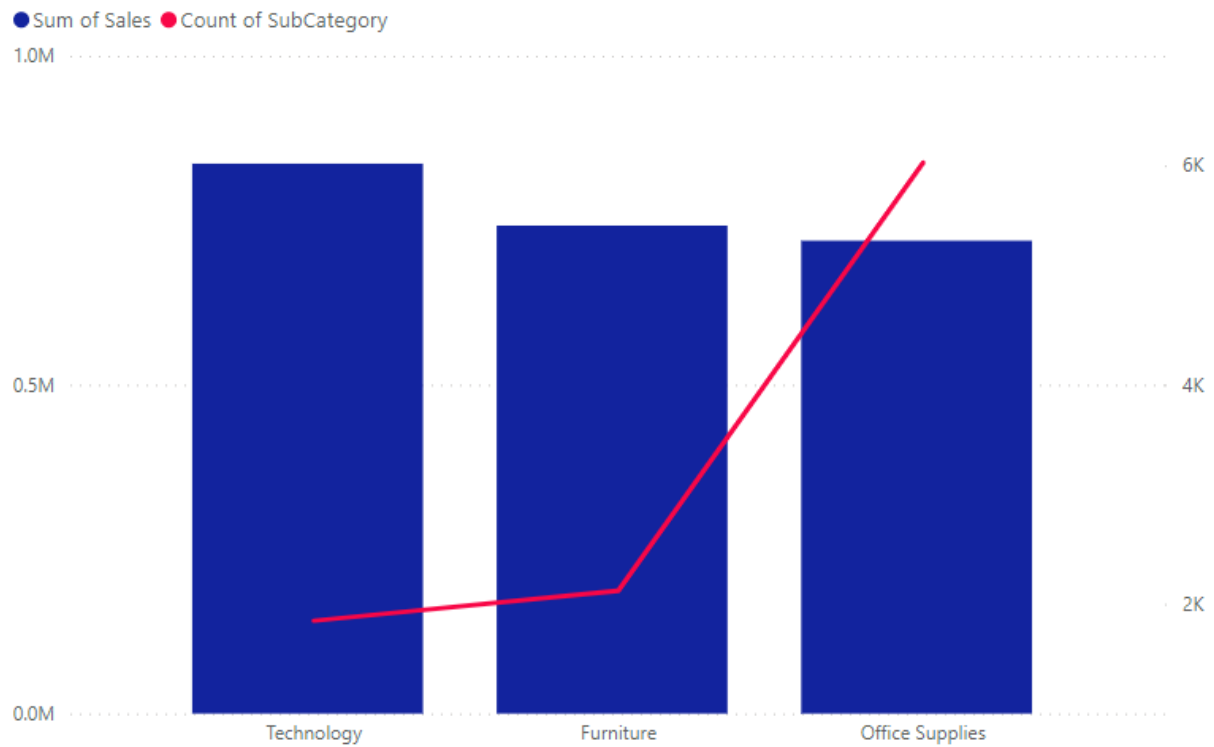
9. How many customers do we have (unique customer IDs) in total and how much per region and state?



Analysis:

Here, initially capturing data on states, regions. Within the available columns, I sought to identify the unique customer IDs where the majority of unique customers IDs in per states and also its regions.

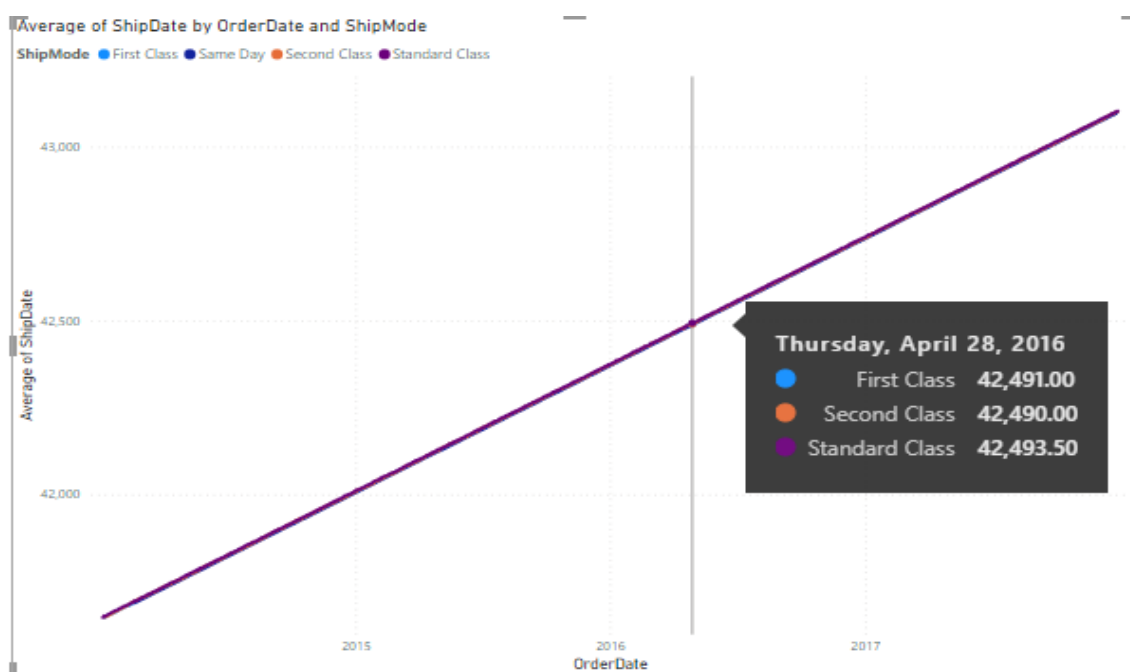
10.Customer rewards program



Analysis:

Here, initially capturing data on category, subcategory and total sales. Within the available columns, I sought to identify the customers are rewarded based on the occurrence of the highest sales in their respective category.

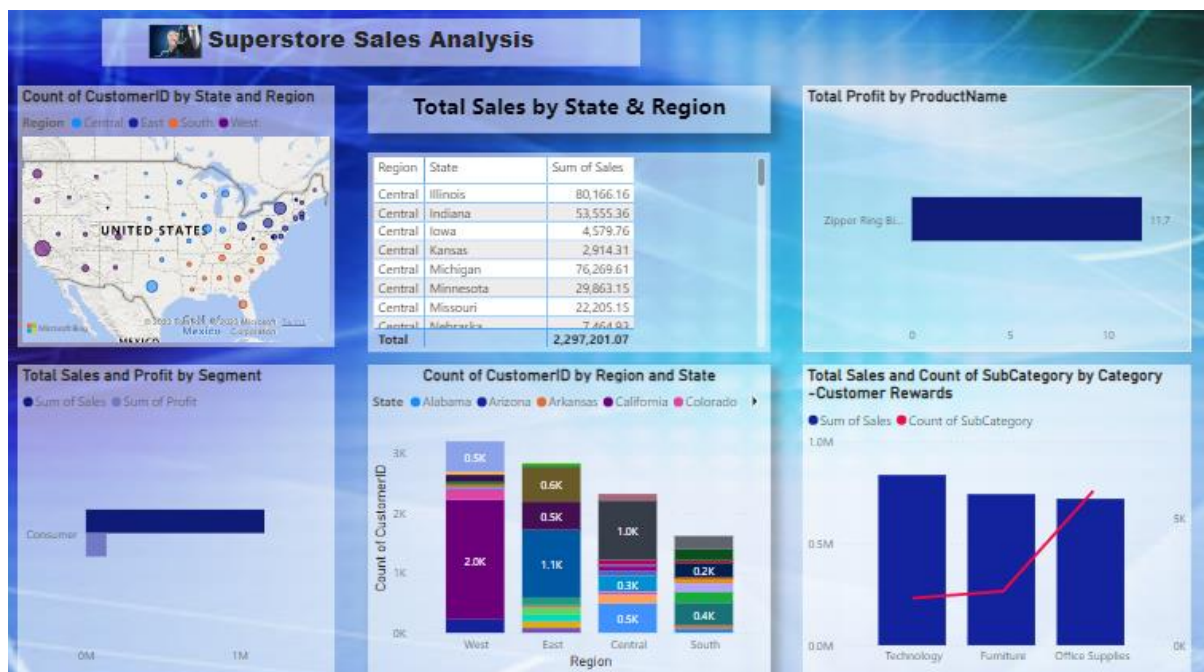
11.Average shipping time per class and in total



Analysis:

Here, Initially, I calculated the average duration between the order date and ship date for each ship mode. Among the available columns, I aimed to determine the average shipping time per class, and I presented it as part of the analysis.

Dashboard



8. Conclusion

Trends in Total Sales and Profits:

Identify trends in total sales and profits over the years to understand the overall business performance. Finally, observe the results to identify the sum of total profits and the sum of total profit by year.

Quarterly Performance:

Analyze total profits and sales per quarter to identify any seasonality or specific periods of high/low performance. Finally, observe the results to identify the sum of total profits and the sum of total profit per quarter.

Regional and State Analysis:

I identified the sum of sales and the sum of profits by region. This analysis provides insights into the regional and state-wise performance, helping in strategic decision-making and resource allocation.

Discount and Sales Relationship:

Identify the relationship between discount and sales and the total discount per category. This analysis contributes to optimizing discount strategies for different product categories.

Category Performance:

Analyze the names of the most and least profitable products. This insight aids in optimizing the product portfolio by understanding which products contribute significantly to profits and which ones may need further evaluation.

Product Profitability:

Finally, observe the results to identify the segment that contributes most to our sales and profit. This analysis provides a focus on the product segments driving the majority of the business's sales and profits.

Customer Metrics:

Sought to identify the unique customer IDs where the majority of unique customer IDs are per states and regions. This customer-centric analysis assists in tailoring marketing efforts and enhancing customer relationship management strategies.