TRIBHUVAN UNIVERSITY

A Practical Report

On

Data Analysis and

Visualization

(CACS455)

Submitted by   *Submitted to*

Name: Krishna Prasad Lamichhane   Er. Suvash Chandra Gautam

Symbol Number:92102067

TABLE OF CONTENTS

Lab1: Bhatvateni Stores aims to create an annual sales report for 2023 to gain a better understanding of their customers and develop strategies to increase sales in 2025. To achieve this, the report should address the following standardized questions.

1. What are the trends in sales revenue and order volume over 2023, and how are they distributed?

Answer: The following charts shows the comparison between sales and orders with monthly data to see the trends in sales revenue and order volume over 2023.
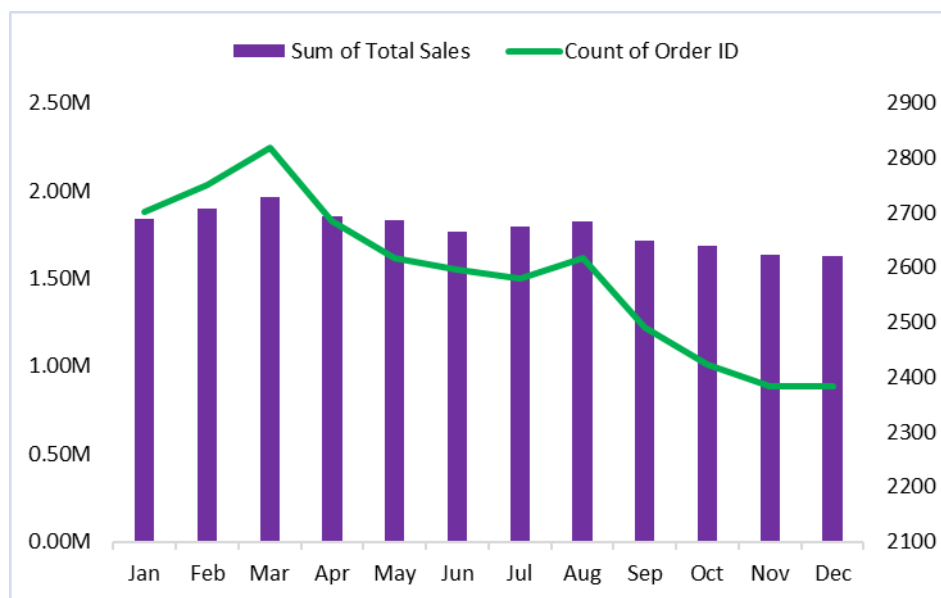

Fig: chart of sales revenue and order volume

2. Which month recorded the highest sales revenue and had the highest order volume?

Answer: March was a standout month, leading both in terms of revenue and order volume.

Highest Sales Revenue: March, with a total of 1,928,066

Highest Order Volume: Also in March, with 2,843 items sold

Month	Highest Sales Revenue	Highest Order Volume
March	**1928066**	**2843**

3. What is the distribution of purchases between male and female customers in 2023?

Answer: we can see that female (Women) customers have a higher total purchase amount then male (Men) customers. below pivot table and in pie chart shows the visual data of purchase distribution between male and female

Pivot table:

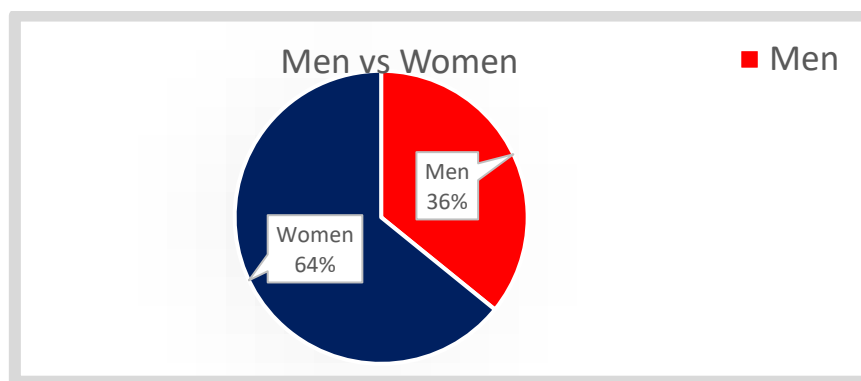| Row Labels	Sum of Total Salaes
Men	7684958
Women	13756251
Grand Total	**21441209**

Fig: Pie Chart of purchase distribution between male and female.

4. What are the different order statuses recorded in 2023, and how are they distributed?

Answer: The majority of orders (92%) were successfully delivered, with a small percentage being returned, refunded, or canceled. We can see that in the pivot table and pie chart below clearly

Pivot Table:

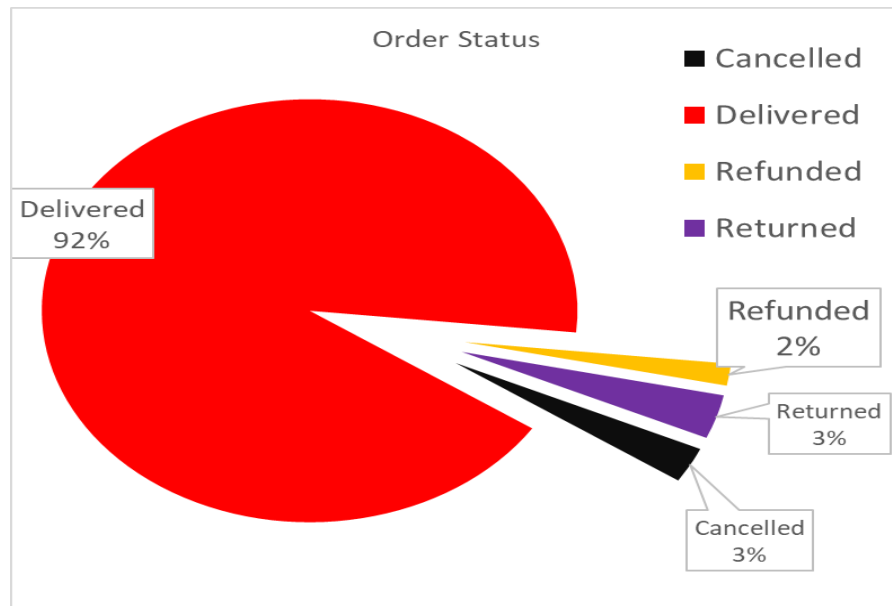| Row Labels	Count of Order ID
Cancelled	844
Delivered	28641
Refunded	517
Returned	1045
Grand Total	**31047**

Fig: Pie Chart of order statuses recorded in 2023

5. Which are the top 10 states contributing the most to sales revenue?

Answer: We can see that Maharashtra is the state contributing the most to sales revenue, followed by Karnataka, Uttar Pradesh, Andhra Pradesh, Delhi, Haryana, Kerala, Tamil Nadu, Telangana, and West Bengal. This is clearly demonstrated in the pivot table and chart below.

Pivot Table:

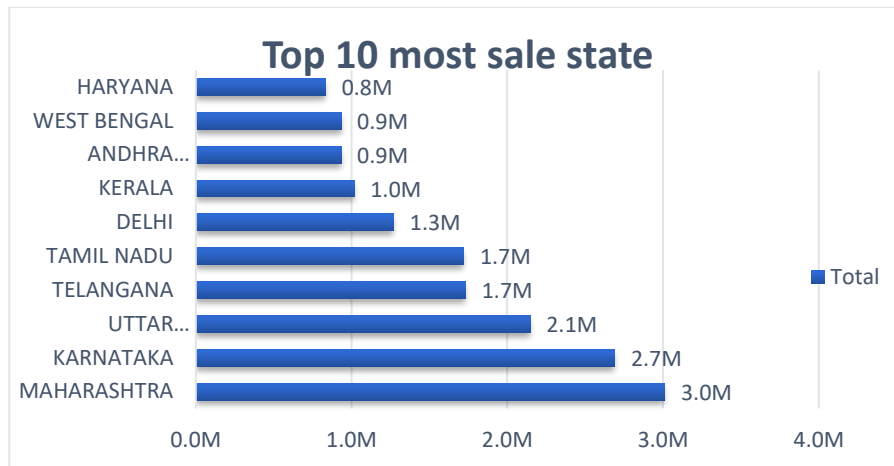| Row Labels	Sum of Total Salaes
MAHARASHTRA	3011677
KARNATAKA	2690050
UTTAR PRADESH	2147265
TELANGANA	1731869
TAMIL NADU	1720783
DELHI	1270412
KERALA	1017282
ANDHRA PRADESH	933191
WEST BENGAL	933102
HARYANA	833048

Fig: Chart of top 10 states contributing the most to sales revenue

6. What is the relationship between customer age and gender in terms of the number of orders placed?

Answer: The dataset reveals the connection between customer age, gender, and the number of orders placed. The below pivot table and chart shows the relationship between customer age and gender in terms of the number of orders placed

Pivot Table:

Count of Order ID	Column Labels	
Row Labels	Men	Women
Adult	7.82%	17.80%
Senior	5.91%	13.70%
Teenager	1.38%	3.33%
Young	15.47%	34.59%

Relation between age and gender

■ Men ■ Women

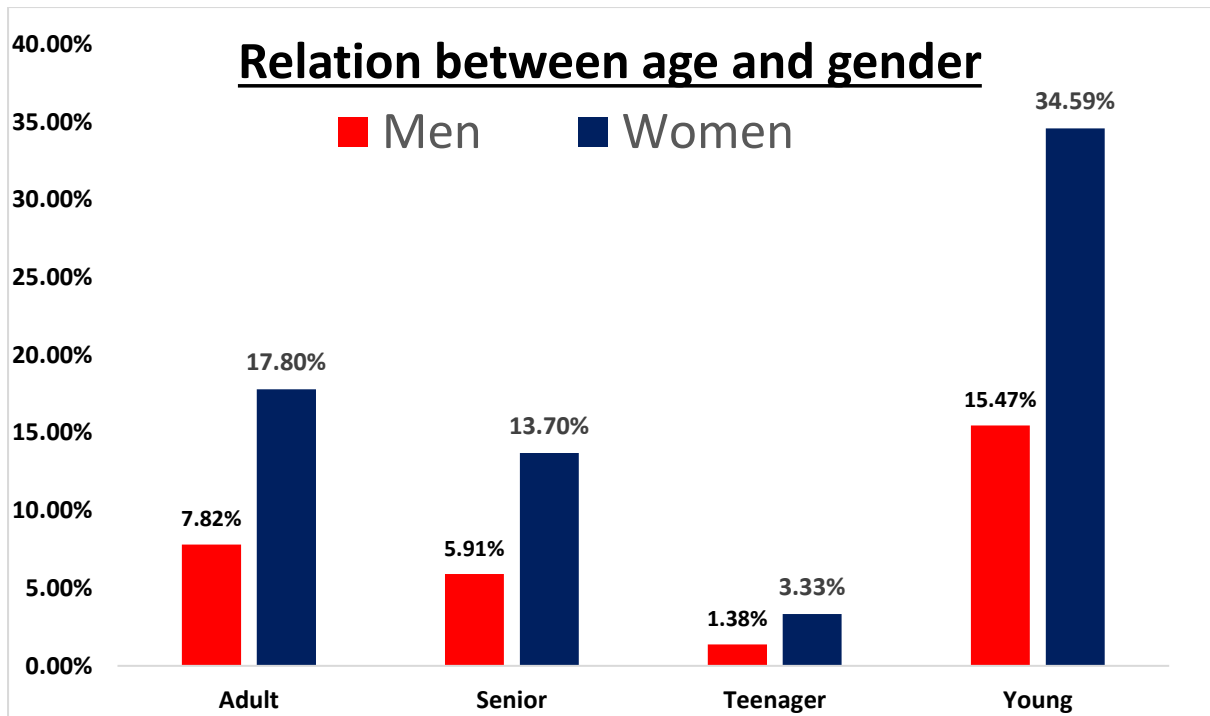|	Adult	Senior	Teenager	Young
Men	7.82%	5.91%	1.38%	15.47%
Women	17.80%	13.70%	3.33%	34.59%

Fig: Bar Chart of number of orders placed by gender and age

7. Which sales channel (e.g., online, in-store) contributed the most to overall sales revenue?

Answer: The sales channels contributing the most to overall sales revenue are as follows

Pivot Table

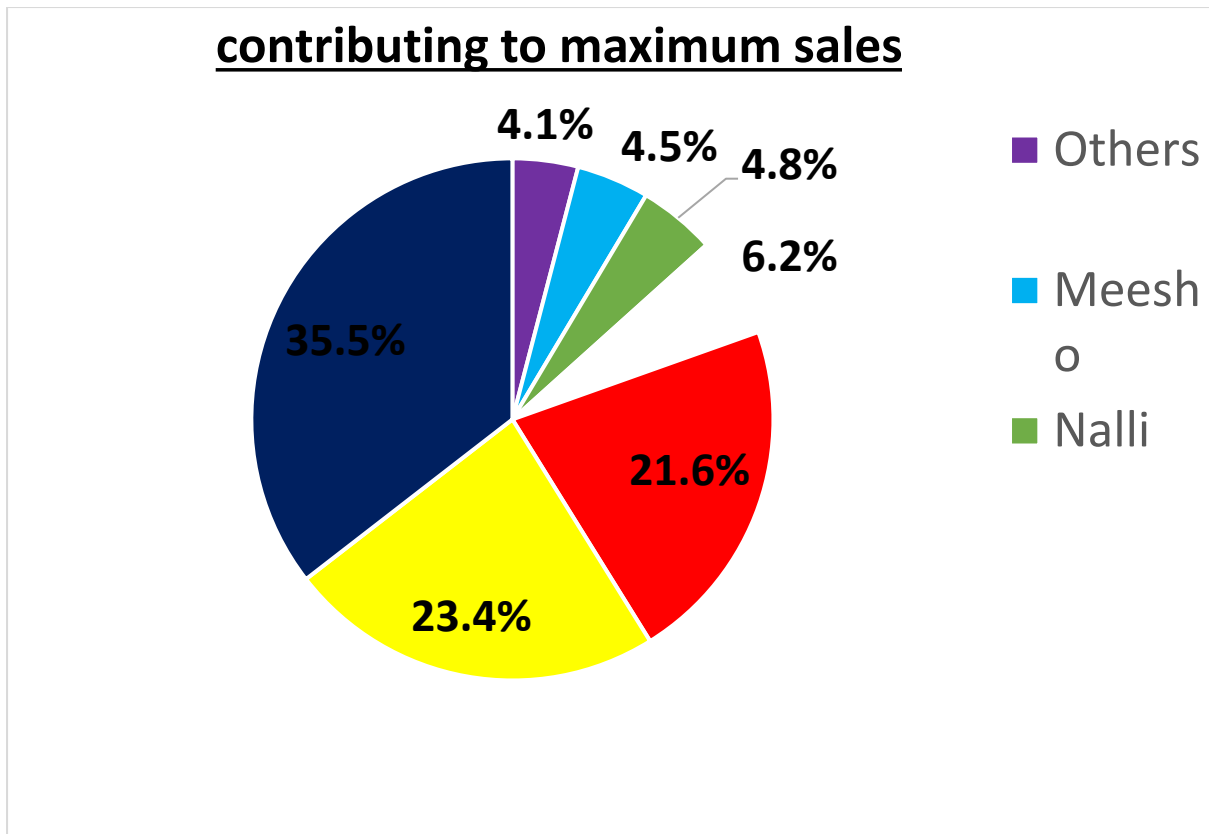| Row Labels	Count of Order ID
Others	4.1%
Meesho	4.5%
Nalli	4.8%
Ajio	6.2%
Flipkart	21.6%
Myntra	23.4%
Amazon	35.5%

Fig: Pie Chart of sales channel contributed the most to overall sales revenue

8. Which product category had the highest sales revenue in 2023?

Answer: "Set" was the leading product category in terms of sales revenue, followed by Western Dresses and Kurtas. Below pivot table and the chart show the clear picture of which category had the highest sales revenue in 2023

Pivot Table:

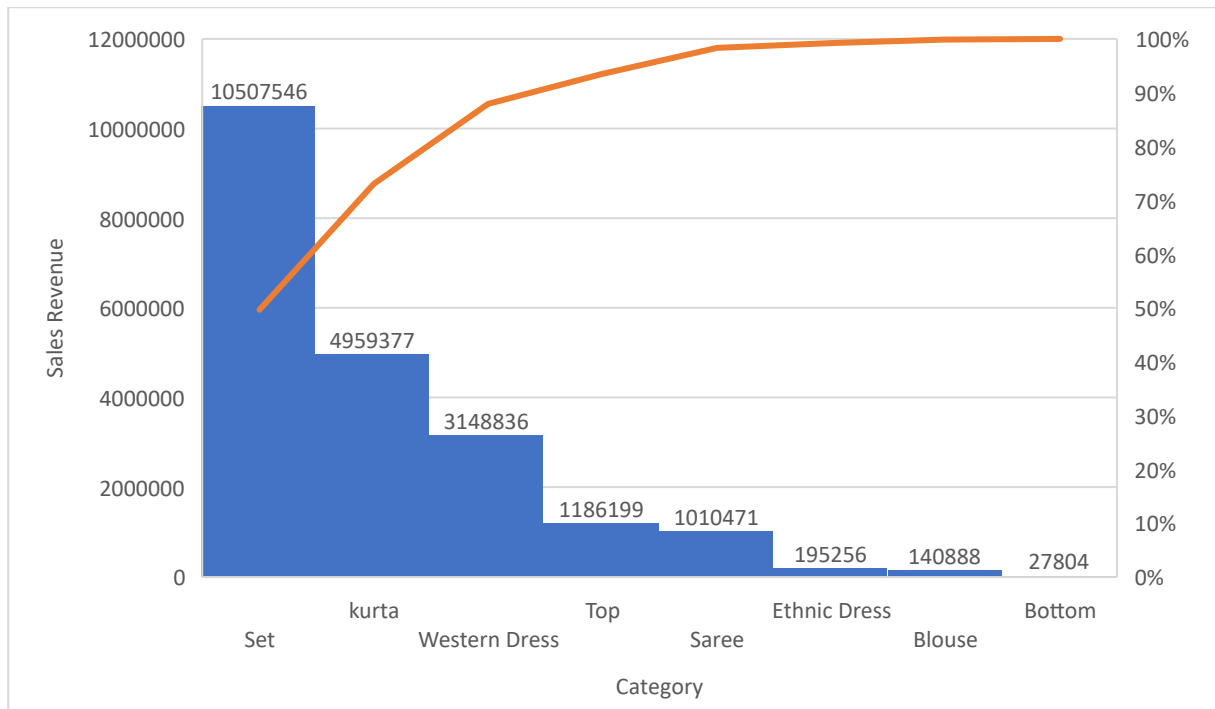| Row Labels	Sum of Amount
Western Dress	3148836
Top	1186199
Set	10507546
Saree	1010471
kurta	4959377
Ethnic Dress	195256
Bottom	27804
Blouse	140888
Grand Total	**21176377**

Fig: Bar Chart of which the product category had the highest sales revenue in

Creating Interactive dashboard:

Creating an Interactive Dashboard involves a step-by-step process that transforms raw data into actionable insights through visualization. The key stages are Data Collection, Data Cleaning, Data Processing, Data Analysis, and Building the Interactive Dashboard itself. Here's how each step works

1. **Data Collection:** Gather relevant data from various sources (e.g., sales records, customer data).
2. **Data Cleaning:** Prepare the data by removing errors, handling missing values, and standardizing formats.
3. **Data Processing:** Organize and transform the data for analysis, such as aggregating or calculating new metrics.
4. **Data Analysis:** Extract insights from the processed data, looking for trends, correlations, and key patterns.
5. **Interactive Dashboard:** Build a dashboard where users can interact with the data, applying filters and exploring visualizations like charts and graphs.

Below is an example of the interactive dashboard that allows users to explore sales trends, customer demographics, order statuses etc.
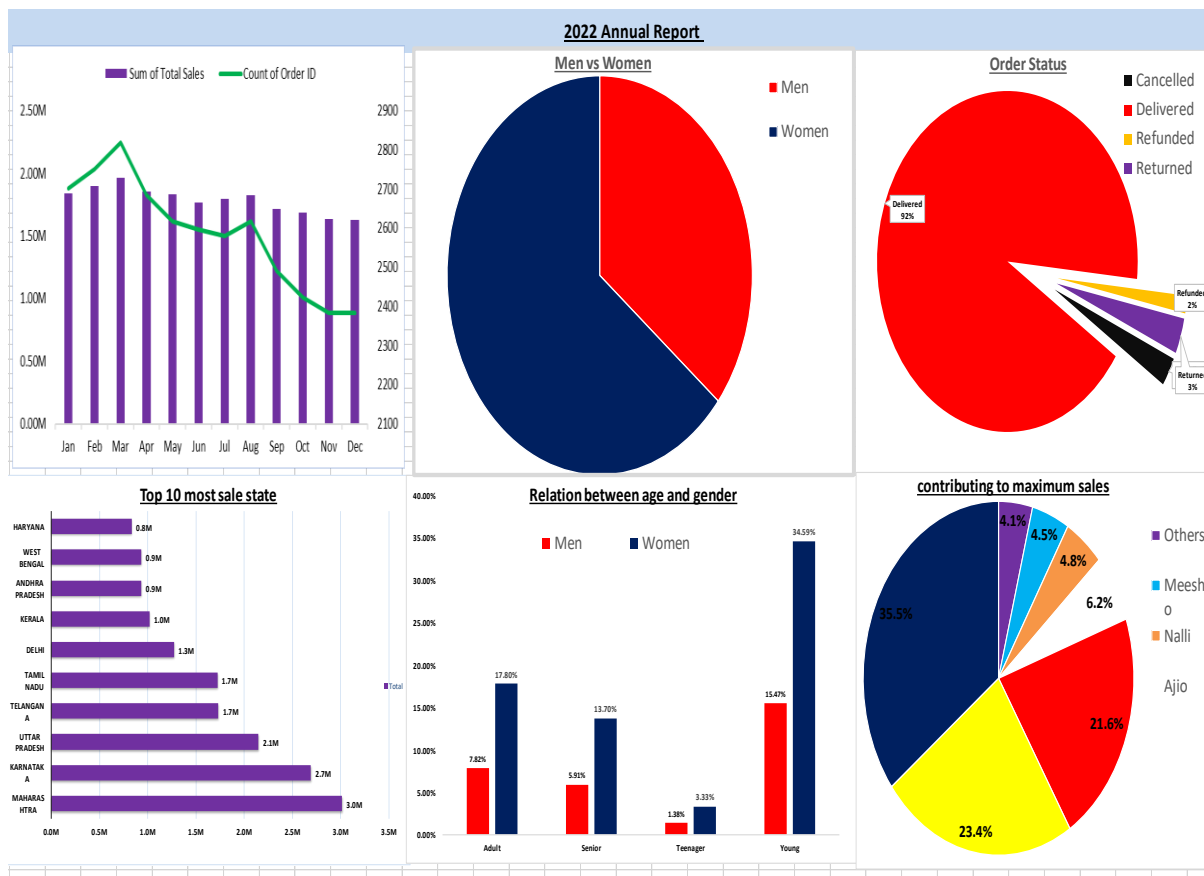

Fig: Interactive Dashboard

Conclusion:

The dataset provides key insights into sales trends, customer behavior, and regional performance. March emerged as the top month for both sales revenue and order volume. Female customers placed more frequent orders, while male customers spent more overall. Maharashtra was the leading state in sales, and Amazon was the top sales channel. "Set" was the highest-grossing product category.

The interactive dashboard enables users to explore trends and visualize data, helping decision-makers optimize strategies for 2022.

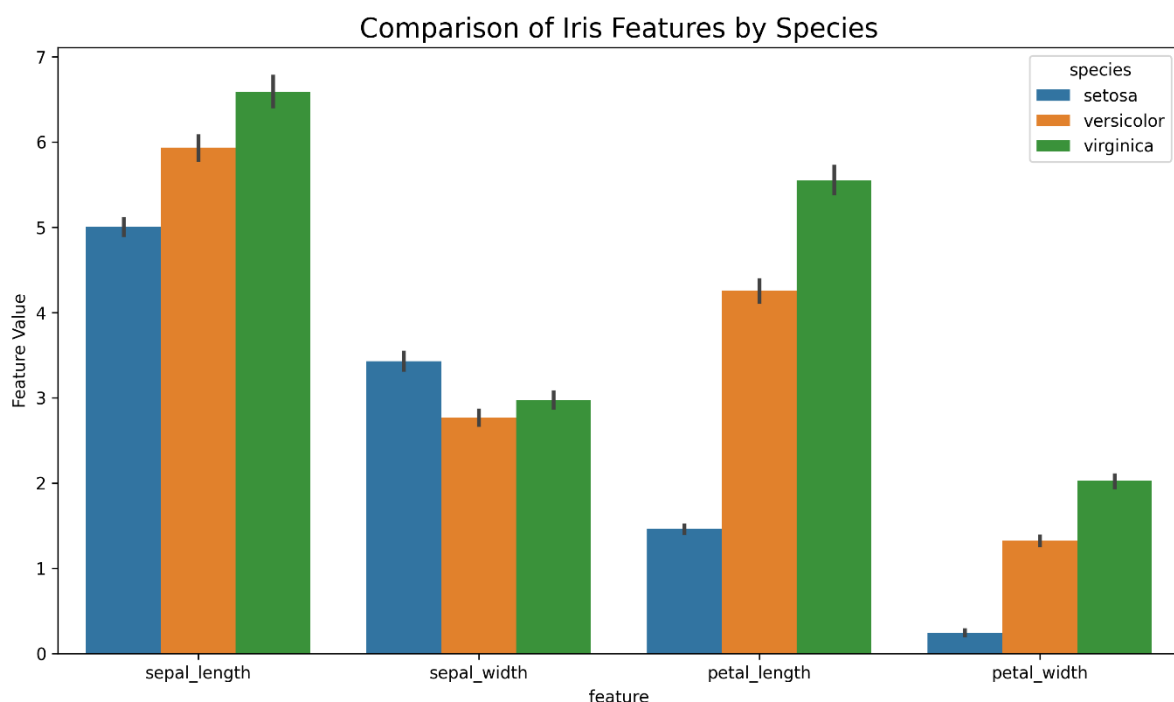**Lab 2: To implement an overview of Iris Dataset using Tableau**

Theoretical Background:

The **Iris dataset** is one of the most widely recognized datasets in the field of data science. It consists of 150 samples of iris flowers, split into three species: **Setosa**, **Versicolor**, and **Virginica**. Each flower is measured across four attributes:

- **Sepal Length** (in cm)

- **Sepal Width** (in cm)

- **Petal Length** (in cm)

- **Petal Width** (in cm)

These measurements provide a basis for classifying flowers into species, making it a classic example in classification tasks. The dataset is often used for practice and demonstrations in machine learning and data visualization.

Visualizations

Conclusion: In this way, Iris data overview and visualization were implemented. Also, this lab provides an overview of the Iris dataset and demonstrates how to visualize data.

**Lab 3: To implement an overview of the Detroit Dataset using Tableau**

Theoretical Background:

The Detroit dataset is a synthetic data collection representing various aspects of the zoning districts in Detroit. It includes different attributes related to zoning areas, including their descriptions, primary zoning types, and geographical measurements. Key attributes include:

- **Global ID**: A unique identifier for each district.

- **Zdescr**: A textual description of the zoning district.

- **Zoning**: The zoning type (e.g., residential, commercial).

- **Shape_Area**: The area of the zoning district in square units.

- **Shape_Length**: The perimeter of the zoning district.

This dataset is used for analysis and educational purposes, typically to visualize spatial information and zoning data.

Visualization
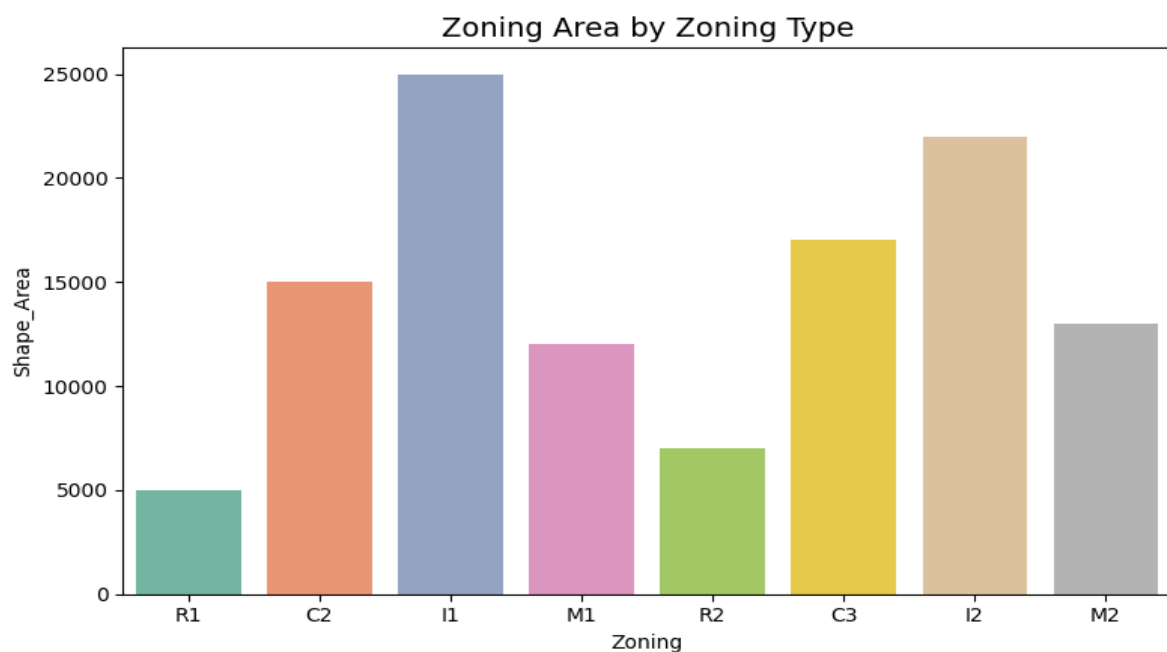
Conclusion: The Detroit dataset was explored through various visualizations, including bar charts for zoning area comparisons. This exercise highlights the ability to analyze geographical and zoning-related data. in this way, Iris data overview and visualization were implemented.

Lab 4: To implement an overview of Breakfast Cereal Dataset using Tableau

Theoretical Background:

The **Breakfast Cereal dataset** provides details about different cereals, focusing on their nutritional values and consumer ratings. The dataset includes attributes like:

- **Calories**: Number of calories per serving.

- **Protein**: Protein content per serving.

- **Fat**: Fat content per serving.

- **Fiber**: Fiber content per serving.

- **Sugars**: Sugar content per serving.

- **Rating**: Consumer rating of the cereal.

This dataset is useful for nutritional analysis and comparing the healthiness of various cereal brands.

Visualization

Calories per Serving of Different Cereals

Fiber Content vs Rating
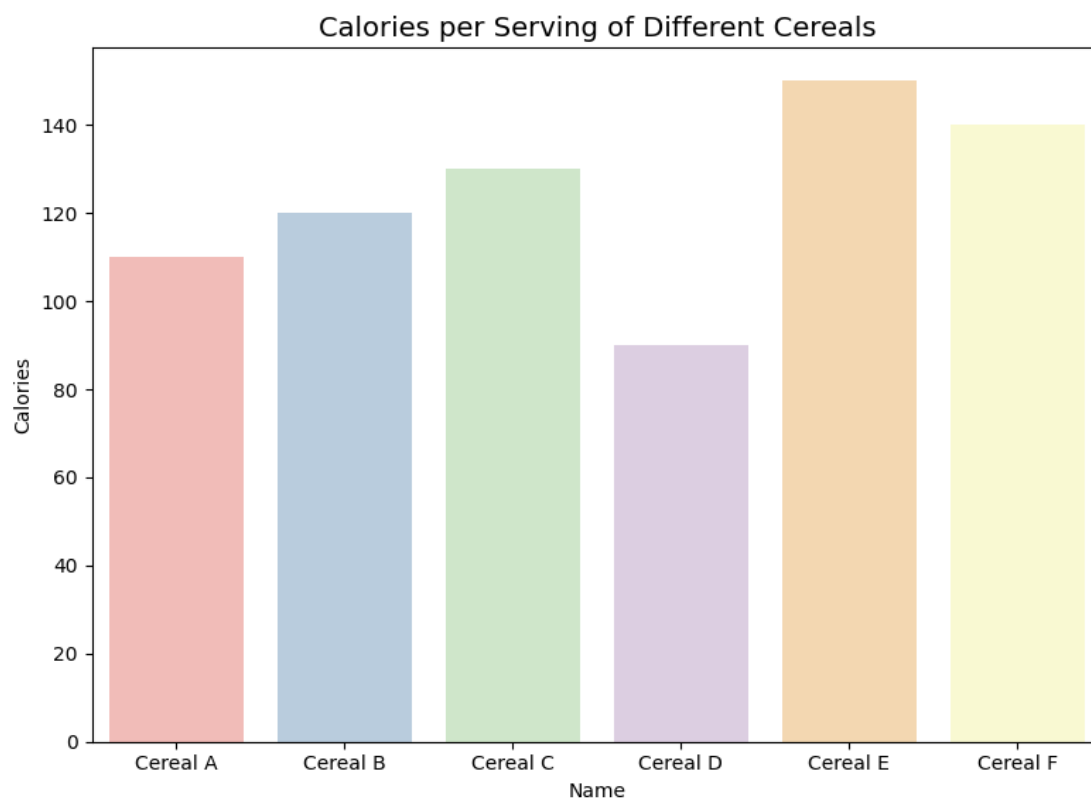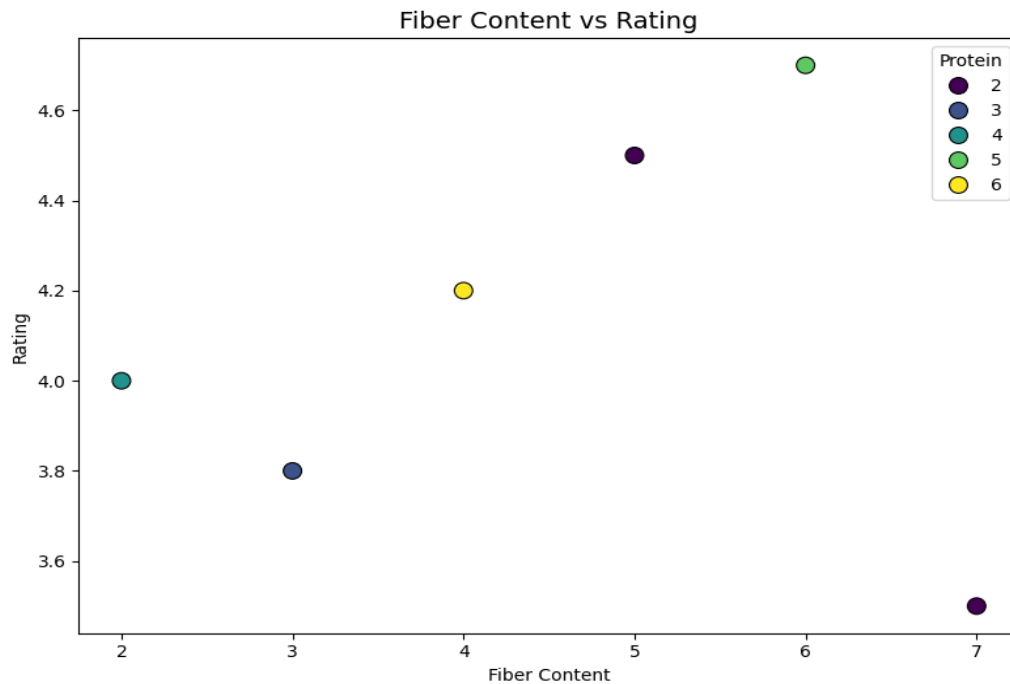
Conclusion: This lab demonstrates how to visualize the Breakfast Cereal dataset and how to use bar and scatter plots to analyze nutritional content and consumer ratings.

Lab 5: To implement an overview of the Dow Jones Industrial Dataset using Tableau
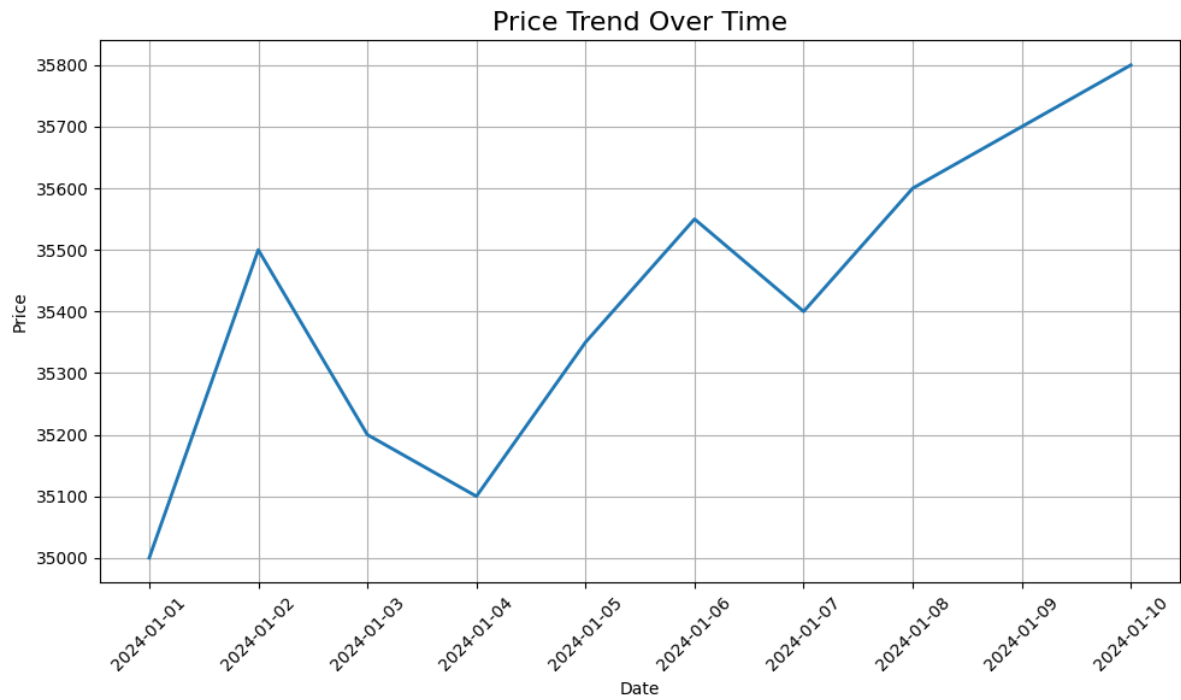
Theoretical Background:

The **Dow Jones Industrial Average (DJIA)** dataset represents the stock market performance over time, specifically for the DJIA index. The dataset includes:

- **Date**: Date of the trading day.

- **Index Value**: Value of the DJIA index.

- **Open Price**: The opening price of the DJIA index.

- **Close Price**: The closing price of the DJIA index.

- **Volume**: The number of shares traded.

This dataset is typically used to analyze stock market trends and index performance.

Visualizations

Price Trend Over Time

Conclusion: This lab explored the DJIA dataset, visualizing stock market price trends through line chart and generating summary statistics.

Lab 6: To implement a basic spatial visualization.

Theoretical Background:

Spatial data visualization represents geographic information in a visual format. This could involve points (locations), lines (routes), polygons (regions), or raster data (e.g., satellite images). It enables us to understand, analyze, and communicate patterns, relationships, and trends present in data that has a spatial or geographic component. This type of visualization is particularly useful for datasets that involve geographic coordinates, maps, and geospatial attributes.

Here are some key aspects and concepts related to spatial data visualization:

1. **Geographic Context:** Spatial data visualization places data within its geographic context. It allows us to see how data points relate to each other in the physical world.
2. **Types of Spatial Data**: Spatial data can include various types of information, such as points (individual locations), lines (routes, rivers), polygons (areas, regions), and raster data (gridded data like satellite images or elevation models).
3. **Mapping Tools and Libraries**: Various tools and libraries facilitate spatial data visualization, including geographic information systems (GIS) software, web-based mapping platforms (like Google Maps), and programming libraries such as Matplotlib, Plotly, GeoPandas, and Folium.
4. **Spatial Attributes**: Data associated with each location can be visualized through attributes like color, size, shape, and opacity, allowing for the representation of multiple dimensions simultaneously.
5. **Spatial Aggregation**: Spatial data can be aggregated to show larger patterns. For example, points can be aggregated into density maps, or polygons can represent statistical averages.
6. **Choropleth Maps**: Choropleth maps use color-coding to represent data values within predefined areas or regions. They are commonly used to show population density, income levels, election results, and other regional attributes.

Visualizations:

Global Sales of products using order data
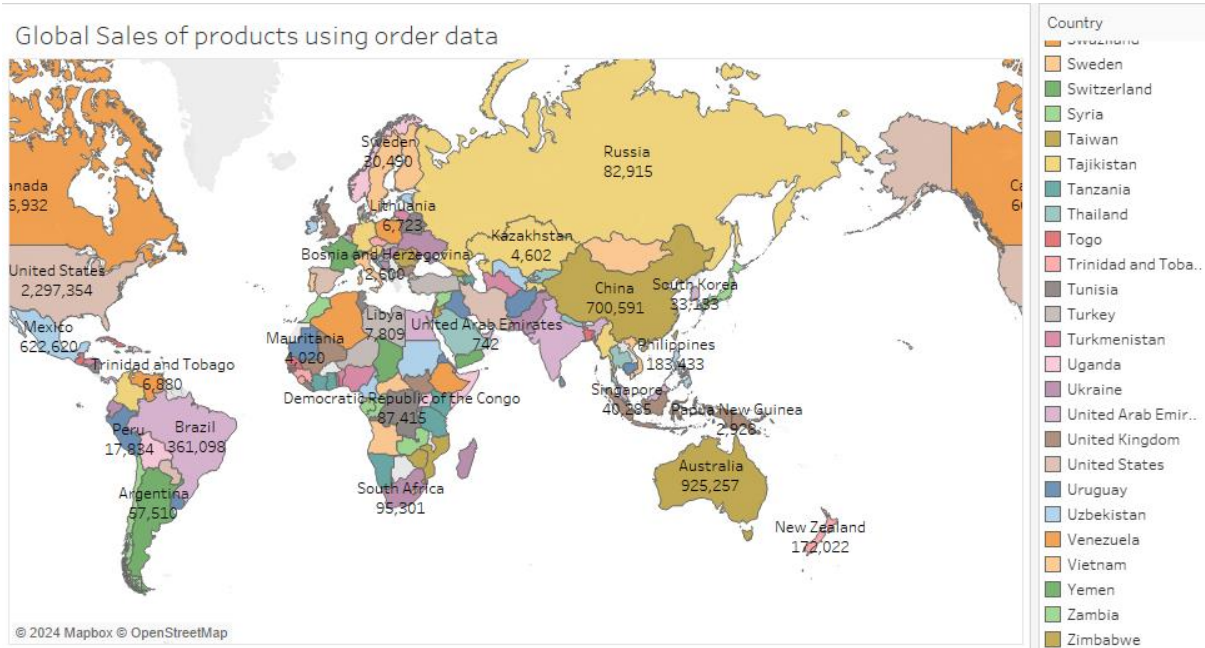
Conclusion:

In this way, spatial visualization was implemented.