INTERNSHIP REPORT

An Internship Report
On
Unveiling Trends: Holistic Retail Management and Customer Data Analysis

by

E.Pranith Kumar– 21951A67AA6
Haasini Ragam – 21951A0556
V.Gowtham– 21951A12D0

INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad – 500 043, Telangana

May, 2024

DECLARATION

I certify that

a. The work contained in this report is original and has been done by me under the guidance of my supervisor.

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute for preparing the report.

d. I have conformed to the norms and guidelines given in the Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Place: Hyderabad	**Signature of the Student**
Date:	**Roll No. 21951A67A6**

Acknowledgement

I extend my heartfelt gratitude to **Makri Solutions (OPC) Private Limited, Hyderabad -** 500038, for providing me with the opportunity to undertake this internship project. I am deeply thankful to **Dr. D V Ramana**, Project Manager, for his invaluable guidance, mentorship, and unwavering support throughout the internship. His expertise and insights have been instrumental in shaping my understanding and skills in [mention specific area, e.g., "Human Resource Analytics"]. I am also grateful to the entire team at Makri Solutions for their encouragement and assistance, which significantly contributed to the success of this project. This internship experience has been invaluable in my professional development, and I am privileged to have been part of Makri Solutions (OPC) Private Limited during this transformative journey.

With sincere regards,

E.Pranith Kumar	21951A67A6
Haasini Ragam	21951A0556
V.Gowtham	21951A12D0

Abstract

Holistic retail management and customer data analysis are crucial for modern retailers aiming to enhance efficiency, satisfaction, and profitability. By integrating supply chain coordination, inventory control, and customer service, businesses ensure seamless operations and a consistent customer experience. This study, analysing 25,000 transactions, reveals that personalized marketing, driven by customer data insights, boosts engagement and conversion rates. Inventory optimization benefits from demand forecasting, while customer service and product development improve through behavioral insights. Python aids in data processing, and Power BI enhances data visualization. This integrated approach helps retailers optimize operations, personalize experiences, and drive sustainable growth.

Keywords: Holistic Retail Management, Customer Data Analysis, Supply Chain Coordination

Inventory Control.

TABLE OF CONTENTS

Chapter 1

Introduction

Retailers today face the challenge of optimizing operations, ensuring consistent customer experiences, and driving profitability. Leveraging holistic retail management and customer data analysis, supported by technologies like Python and Power BI, enables businesses to make informed decisions, personalize interactions, and stay competitive in a dynamic market.

1.1 Background

Overview of Data Analytics and Visualization:

Data Analytics involves examining datasets to draw conclusions about the information they contain, using specialized systems and software. It encompasses various techniques from simple statistical analysis to complex data mining and predictive analytics.

What is Data Visualization?

Data Visualization refers to the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Importance of Data Analytics and Visualization

1.Enhanced Decision Making: Helps in making informed decisions by providing insights derived from data.
2.Identifying Trends and Patterns: Facilitates the recognition of trends, patterns, and anomalies that might not be apparent in raw data.
3.Increased Efficiency: Streamlines the decision-making process and helps in solving problems quickly.
4.Improved Communication: Makes it easier to communicate complex data insights to stakeholders through visual representation.
5.Data-Driven Strategies: Enables the development of strategies based on data-driven insights, leading to better outcomes.
6.Performance Tracking: Assists in monitoring and evaluating performance across different areas of business or operations.

Advantages of Data Visualization

- Simplifies Complex Data: Makes large and complex data sets comprehensible by presenting them visually.
- Quick Insights: Allows for quick and accurate interpretation of data.
- Interactive Exploration: Tools often allow interactive exploration of data, helping users to uncover hidden insights.
- Enhanced Data Storytelling: Enables effective storytelling through visual means, making data more engaging.

Disadvantages of Data Visualization

- Misleading Information: Poor design choices can lead to misinterpretation or distortion of data.
- Oversimplification: Important details might be lost in the process of simplifying data for visualization.
- Data Quality Dependency: The accuracy of visualizations depends heavily on the quality of the underlying data.

Importance of Power BI and Python in Data Analysis

Power BI is a powerful business analytics tool by Microsoft that offers a comprehensive solution for data visualization and business intelligence. It enables users to transform raw data into actionable insights through its easy-to-use interface and robust capabilities.

Key Benefits of Power BI in Data Analysis:

1. **User-Friendly Interface**:
   - Power BI’s drag-and-drop functionality makes it accessible to users with varying levels of technical expertise.
   - The intuitive design allows users to quickly create and customize visual reports and dashboards.
2. **Integration with Multiple Data Sources:**
   - Power BI supports a wide range of data sources including Excel, SQL Server, Azure, and cloud services like Salesforce and Google Analytics.
   - The ability to connect to various data sources enables comprehensive data analysis from disparate systems.
3. **Real-Time Data Processing:**
   - Power BI can handle real-time data, providing up-to-date insights and allowing for timely decision-making.
   - Integration with streaming data sources facilitates real-time monitoring and alerting.

4. **Interactive Visualizations:**
    - Power BI offers a variety of interactive visualizations, such as bar charts, line charts, scatter plots, and maps.
    - Users can drill down into data, apply filters, and interact with visual elements to explore data in-depth.
5. **Collaboration and Sharing:**
    - Reports and dashboards can be easily shared within an organization through the Power BI Service.
    - Features like workspaces and shared datasets promote collaboration among teams.
6. **Customization and Extensibility:**
    - Power BI supports custom visuals and third-party add-ons from the Power BI marketplace.
    - Users can extend functionality by embedding Python and R scripts for advanced data manipulation and visualization.
7. **Security and Compliance:**
    - Power BI provides robust security features, including row-level security to control access to data.
    - Compliance with industry standards and regulations ensures data protection and governance.

Python is a versatile programming language widely used for data analysis and visualization due to its powerful libraries and ease of use. It enables advanced data manipulation and the creation of custom, complex visualizations.

Key Benefits of Python in Data Analysis:

1. **Extensive Libraries for Data Analysis:**
    - Python offers a rich ecosystem of libraries such as Pandas for data manipulation, NumPy for numerical operations, and SciPy for scientific computing.
    - These libraries simplify the process of data cleaning, transformation, and analysis.
2. **Advanced Data Visualization:**
    - Libraries like Matplotlib, Seaborn, and Plotly allow for the creation of detailed and interactive visualizations.
    - Customization options in these libraries enable users to tailor visualizations to specific requirements.

3. **Automation and Efficiency:**
    - Python scripts can automate repetitive data processing tasks, improving efficiency and accuracy.
    - This automation capability is crucial for handling large datasets and complex workflows.
4. **Machine Learning and AI Integration:**
    - Python's machine learning libraries, such as Scikit-learn, TensorFlow, and PyTorch, enable the development of predictive models and AI applications.
    - Integration of these models into data analysis workflows provides deeper insights and forecasts.
5. **Flexibility and Scalability:**
    - Python's flexibility allows for the analysis of data from diverse sources and formats.
    - The language can scale to handle large volumes of data, making it suitable for both small and large datasets.
6. **Community Support and Resources:**
    - Python has a large and active community, providing extensive documentation, tutorials, and support.
    - Continuous development and updates ensure that Python remains a cutting-edge tool for data analysis.

Combining Power BI and Python

Enhanced Data Visualization:

- While Power BI provides robust data visualization capabilities, Python can be used to create custom visuals and perform advanced data manipulation before importing the results into Power BI.

Advanced Analytics and Machine Learning:

- Python's powerful libraries for statistical analysis and machine learning can be leveraged to perform complex analyses, with the results integrated into Power BI dashboards for visualization and reporting.

Automation of Data Workflows:

- Python scripts can automate data extraction, transformation, and loading (ETL) processes, which can then feed into Power BI for visualization and analysis.

Extending Power BI Functionality:

- Python can be used within Power BI to extend its capabilities, such as by using Python scripts in Power BI Desktop for advanced data manipulations that are not possible with DAX or Power Query alone.

Compare Power BI Visualization and Python Visualization Power BI

Visualization

Strengths:

1. **Ease of Use:**
    - User-friendly interface with drag-and-drop functionality.
    - No coding required, making it accessible to non-technical users.
2. **Integration with Data Sources:**
    - Seamless integration with numerous data sources like Excel, SQL Server, Azure, and cloud services.
    - Real-time data connectivity.
3. **Interactive Dashboards:**
    - Highly interactive dashboards with built-in features like filters, slicers, and drill-through capabilities.
4. **Collaboration and Sharing:**
    - Easy sharing and collaboration through the Power BI Service.
    - Role-based access control for secure sharing.
5. **Pre-built Visualizations:**
    - Extensive library of pre-built visualizations such as bar charts, pie charts, line charts, maps, etc.
    - Custom visuals available from the Power BI marketplace.
6. **Continuous Updates:**
    - Regular updates and new features from Microsoft.
7. **Integration with Other Microsoft Tools:**
    - Deep integration with other Microsoft products like Excel, Teams, and SharePoint.

Weaknesses:

1. **Customization Limitations:**
    - Limited customization compared to programming languages.
    - Some complex visualizations may not be achievable with default options.
2. **Cost:**
    - While there is a free tier, advanced features and sharing capabilities require a paid subscription.
3. **Performance with Large Datasets:**
    - Can experience performance issues with very large datasets or complex models.

Use Cases:

- Business reporting and dashboarding.
- Quick insights and ad-hoc analysis.
- Collaboration and sharing insights across teams.

Python Visualization Strengths:

1. **Flexibility and Customization:**
    - Highly customizable visualizations tailored to specific needs.
    - Ability to create complex and unique visualizations using libraries like Matplotlib, Seaborn, Plotly, and Bokeh.
2. **Advanced Analytics:**
    - Integration with powerful data analysis and machine learning libraries like Pandas, NumPy, Scikit-learn, TensorFlow.
3. **Scalability:**
    - Capable of handling large datasets and complex data transformations.
    - Automation of data processing workflows**.**
4. **Open Source:**
    - Free and open-source libraries with extensive community support.
5. **Reproducibility:**
    - Scripts can be easily shared and reproduced by other users.

Integration with Other Tools:

- ○Can be used in conjunction with other data processing and storage tools, such as SQL databases and cloud storage.

Weaknesses:

1. **Learning Curve:**
    - Requires knowledge of programming and familiarity with Python.
    - Steeper learning curve for users without a coding background.
2. **Less Interactive:**
    - By default, visualizations are static, though interactivity can be added with libraries like Plotly and Bokeh.
3. **Manual Process:**
    - More manual effort required to set up and create visualizations compared to Power BI's drag-and-drop interface.

Use Cases:

- Detailed and customized data analysis and visualization.
- Research and exploratory data analysis.
- Complex data transformations and automation.
- Integration with machine learning workflows.

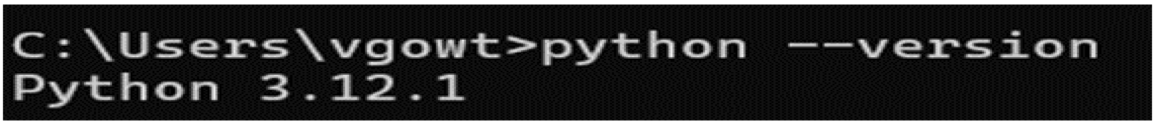
1.2 Getting Started:

❖ **Installing Jupyter Notebook and Necessary Libraries**

1. Install Python:

   Before installing Jupyter, you need to have Python installed on your system. You can download the latest version of Python from the official Python website. Follow the installation instructions for your operating system. Make sure to check the box that says "Add Python to PATH" during installation.

   To verify that Python is installed, open your terminal (or command prompt) and run:

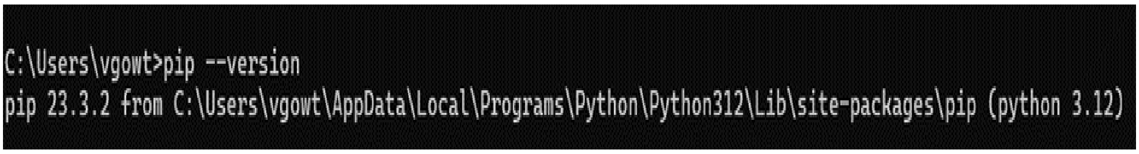


```
C:\Users\vgowt>python --version
Python 3.12.1
```

   This should display the installed Python version. If it doesn't, you might need to add Python to your system PATH.

2. Install pip:

   pip is the package installer for Python. It should be installed by default with Python. You can check if pip is installed by running the following command in your terminal or command prompt:

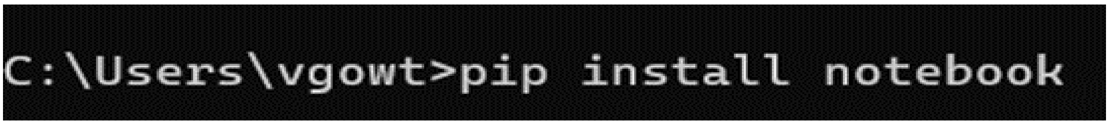


```
C:\Users\vgowt>pip --version
pip 23.3.2 from C:\Users\vgowt\AppData\Local\Programs\Python\Python312\Lib\site-packages\pip (python 3.12)
```

   If pip is not installed, you can install it by following the instructions here.

3. Install Jupyter Notebook:

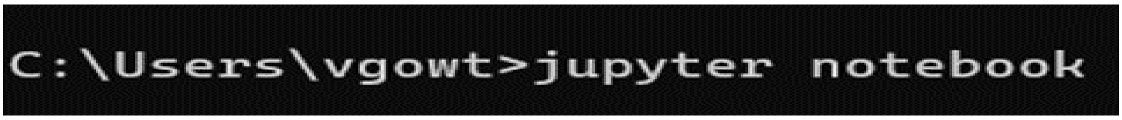
   Open a terminal or command prompt. Install Jupyter Notebook using pip.



```
C:\Users\vgowt>pip install notebook
```

4. Create a New Jupyter Notebook:

Launch Jupyter Notebook from the terminal or command prompt by typing Jupyter Notebook.



```
C:\Users\vgowt>jupyter notebook
```

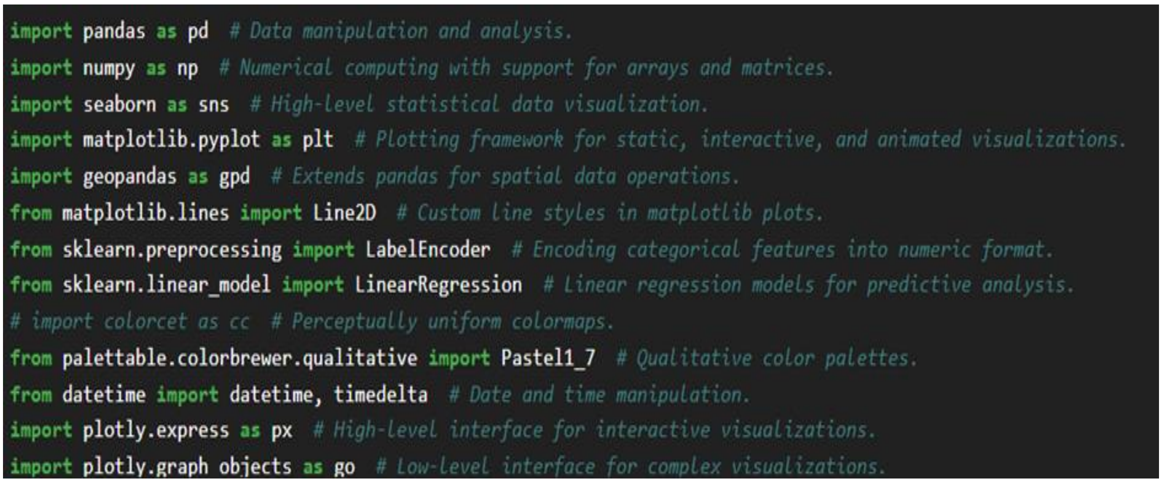
Create a new Python notebook from the Jupyter interface.

5. Install Essential Libraries:

   Open a new code cell in your Jupyter Notebook. Use pip to install the following libraries:

   - pandas-2.2.2   # Data manipulation.
   - numpy-1.26.4   # Numerical computing.
   - seaborn-0.13.2   # Statistical visualization.
   - matplotlib-3.7.2   # Plotting framework.
   - geopandas-0.14.4 # Spatial data operations.
   - scikit-learn-1.4.2   # Machine learning.
   - palettable-3.3.3   # Colour palettes.
   - plotly-5.22.0   # Interactive visualizations.
   - Powerbi-2.130.754.0 #Intractive visualizations

6. Import Libraries



```python
import pandas as pd  # Data manipulation and analysis.
import numpy as np  # Numerical computing with support for arrays and matrices.
import seaborn as sns  # High-level statistical data visualization.
import matplotlib.pyplot as plt  # Plotting framework for static, interactive, and animated visualizations.
import geopandas as gpd  # Extends pandas for spatial data operations.
from matplotlib.lines import Line2D  # Custom line styles in matplotlib plots.
from sklearn.preprocessing import LabelEncoder  # Encoding categorical features into numeric format.
from sklearn.linear_model import LinearRegression  # Linear regression models for predictive analysis.
# import colorcet as cc  # Perceptually uniform colormaps.
from palettable.colorbrewer.qualitative import Pastel1_7  # Qualitative color palettes.
from datetime import datetime, timedelta  # Date and time manipulation.
import plotly.express as px  # High-level interface for interactive visualizations.
import plotly.graph objects as go  # Low-level interface for complex visualizations.
```

7. Load and Explore Data

❖ Setting Up Power BI Desktop

Download Power BI Desktop

Visit the Power BI Website:

- Go to the official Power BI website: Power BI Desktop Download.

Download:

- Click on the “Download Free” button.
- You will be redirected to the Microsoft Store or a direct download link will be provided.
- Download the installer file.

Install Power BI Desktop

Run the Installer:

- Locate the downloaded installer file (usually in your Downloads folder).
- Double-click the installer file to run it.

Installation Process:

- Follow the on-screen instructions.
- Accept the license agreement and select the installation location if prompted.
- Click on “Install” and wait for the installation to complete.

Complete Installation:

- Once the installation is complete, click on “Finish” or “Launch” to open Power BI Desktop.

Initial Setup and Configuration

Open Power BI Desktop:

- If Power BI Desktop doesn’t open automatically, find it in your Start Menu or Desktop and open it.

Sign In:

- If prompted, sign in with your Microsoft account. If you don’t have one, you may need to create an account.

Optional: Update Settings:

- You can customize your settings by clicking on the gear icon (⚙) in the top right corner.
- Adjust regional settings, data load options, and privacy settings as needed.

Connecting to Data

Get Data:

- Click on the “Get Data” button on the Home ribbon.
- Select the data source you want to connect to (e.g., Excel, SQL Server, Web, etc.).

Load Data:

- Follow the prompts to connect to your data source.
- Preview the data and click “Load” to import it into Power BI Desktop.

Building Queries and Data Models:

- Use the Power Query Editor to clean and transform your data. You can remove duplicates, handle missing values, and perform other data transformation tasks.
- Create relationships between different data tables by dragging and dropping fields in the Relationships view.
- Use DAX to create calculated columns and measures for advanced calculations.

Start Creating Reports

Begin with Visualizations:

- Use the visualization pane on the right to drag and drop fields into the report canvas.
- Customize your visualizations with various options available.

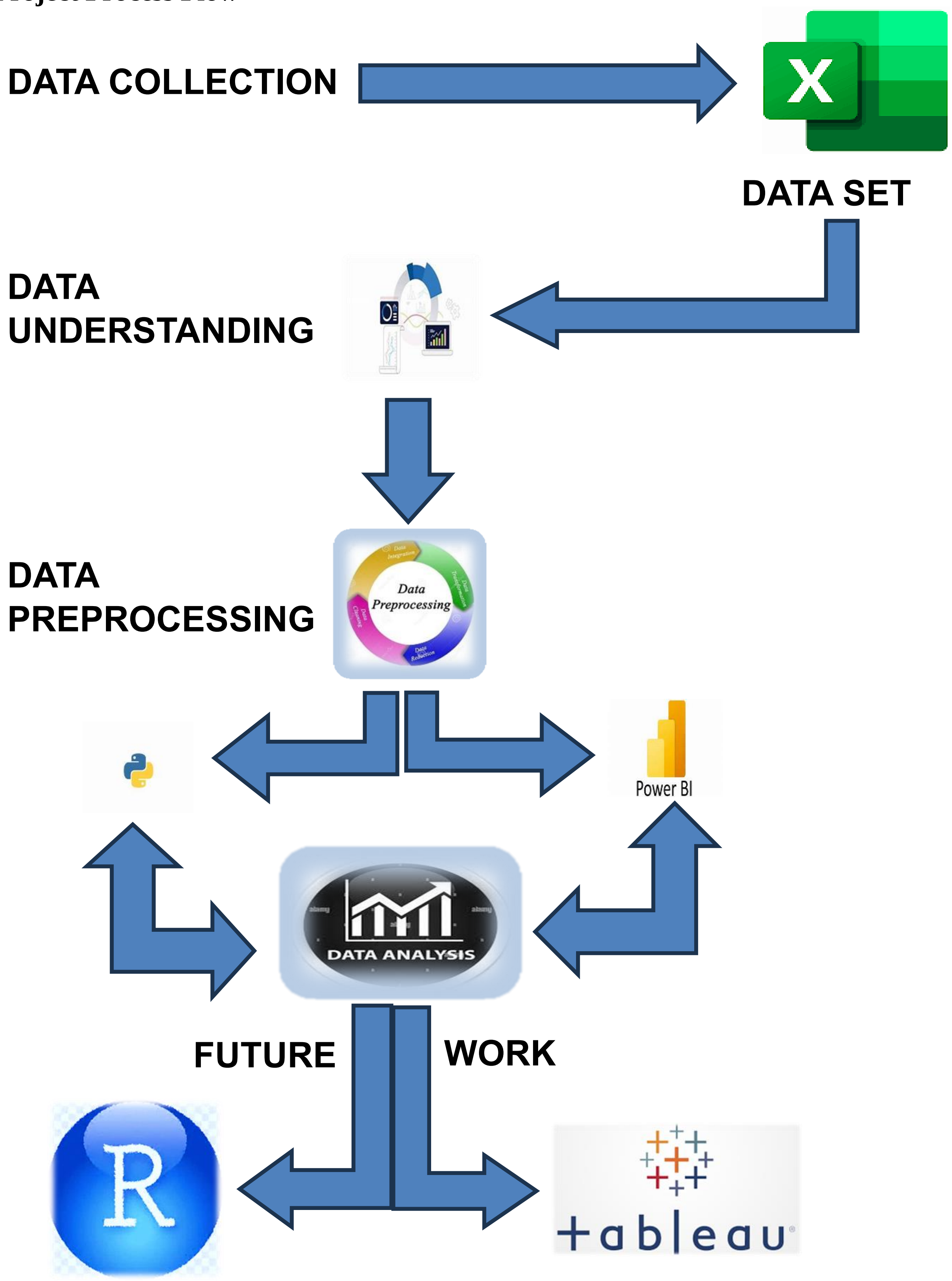
Save Your Work

1.3 Libraries Used

- python 3.12.1
- sacremoses 0.1.1
- scikit-learn 1.4.2
- scipy 1.12.0
- seaborn 0.13.2
- numpy 1.26.4
- geopandas 0.14.4
- plotly 5.22.0
- pandas 2.2.2
- pip 23.3.2
- palettable 3.3.3
- power Bi 2.129.1229.

Chapter 2

Project Process Flow



2.1 Data Preparation and Data Preprocessing:

2.1.1 Data Acquisition:

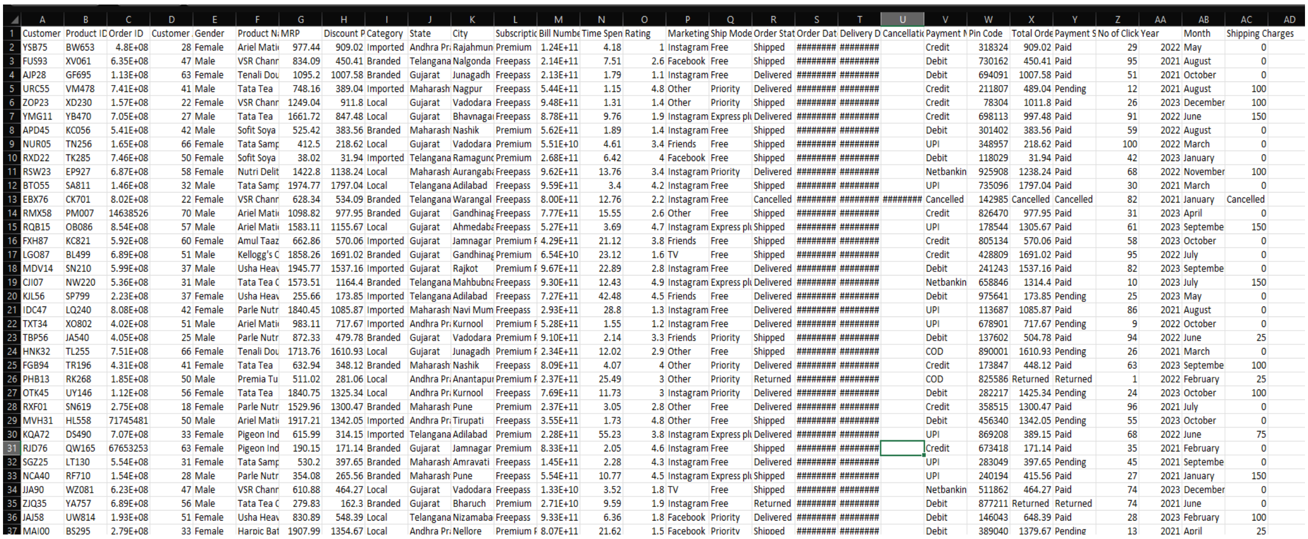
Source of the Data

The dataset used in this project was obtained from the GitHub repository. The specific file used is an Excel (.xl) file that contains relevant data for the analysis.

Accessing the Data

To access the dataset, follow these steps:

1. Navigate to the GitHub repository URL: [https://www.kaggle.com/datasets/praneethkumar007/dmart-ready-online-store].
2. Locate the Excel file within the repository. The file is named [final test1.xlsx].
3. Download the file to your local machine by clicking the "Download" button or using GitHub's raw file view to save the file.

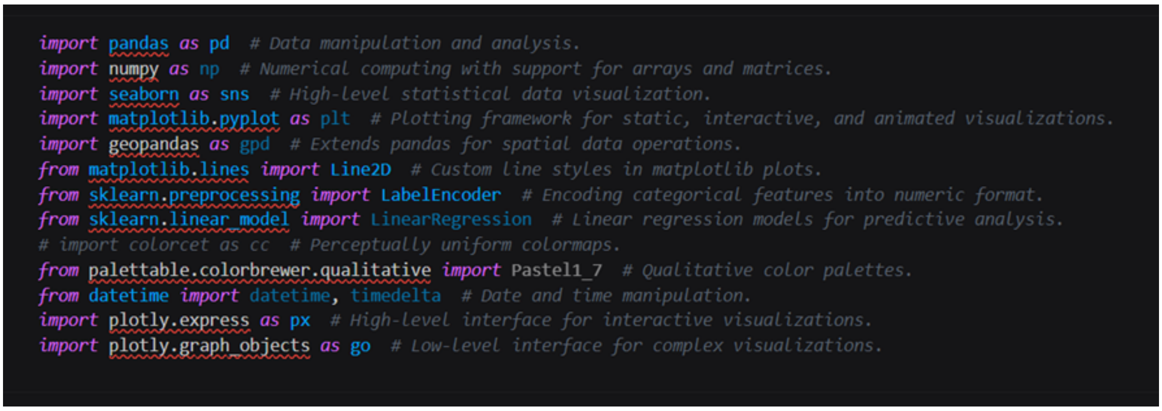


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	Customer	Product ID	Order ID	Customer .	Gender	Product Na	MRP	Discount P	Category	State	City	Subscriptio	Bill Numbe	Time Spen	Rating	Marketing	Ship Mode	Order Stat	Order Dat	Delivery D	Cancellatio	Payment M	Pin Code	Total Orde	Payment S	No of Click	Year	Month	Shipping Charges	
2	YSB75	BW653	4.8E+08	28	Female	Ariel Mati	977.44	909.02	Imported	Andhra Pra	Rajahmun	Premium	1.24E+11	4.18	1	Instagram	Free	Shipped	########	########		Credit	318324	909.02	Paid	29	2022	May	0	
3	FUS93	XV061	6.35E+08	47	Male	VSR Chann	834.09	450.41	Branded	Telangana	Nalgonda	Freepass	2.14E+11	7.51	2.6	Facebook	Free	Shipped	########	########		Debit	730162	450.41	Paid	95	2021	August	0	
4	AJP28	GF695	1.13E+08	63	Female	Tenali Dou	1095.2	1007.58	Branded	Gujarat	Junagadh	Freepass	2.13E+11	1.79	1.1	Instagram	Free	Delivered	########	########		Debit	694091	1007.58	Paid	51	2021	October	0	
5	URC55	VM478	7.41E+08	41	Male	Tata Tea	748.16	389.04	Imported	Maharash	Nagpur	Freepass	5.44E+11	1.15	4.8	Other	Priority	Delivered	########	########		Credit	211807	489.04	Pending	12	2021	August	100	
6	ZOP23	XD230	1.57E+08	22	Female	VSR Chann	1249.04	911.8	Local	Gujarat	Vadodara	Freepass	9.48E+11	1.31	1.4	Other	Priority	Shipped	########	########		Credit	78304	1011.8	Paid	26	2023	December	100	
7	YMG11	YB470	7.05E+08	27	Male	Tata Tea	1661.72	847.48	Local	Gujarat	Bhavnagar	Freepass	8.78E+11	9.76	1.9	Instagram	Express plu	Delivered	########	########		Credit	698113	997.48	Paid	91	2022	June	150	
8	APD45	KC056	5.41E+08	42	Male	Sofit Soya	525.42	383.56	Branded	Maharash	Nashik	Premium	5.62E+11	1.89	1.4	Instagram	Free	Shipped	########	########		Debit	301402	383.56	Paid	59	2022	August	0	
9	NUR05	TN256	1.65E+08	66	Female	Tata Samp	412.5	218.62	Local	Gujarat	Vadodara	Premium	5.51E+10	4.61	3.4	Friends	Free	Shipped	########	########		UPI	348957	218.62	Paid	100	2022	March	0	
10	RXD22	TK285	7.46E+08	50	Female	Sofit Soya	38.02	31.94	Imported	Telangana	Ramagunc	Premium	2.68E+11	6.42	4	Facebook	Free	Shipped	########	########		Debit	118029	31.94	Paid	42	2023	January	0	
11	RSW23	EP927	6.87E+08	58	Female	Nutri Delit	1422.8	1138.24	Local	Maharash	Aurangaba	Freepass	9.62E+11	13.76	3.4	Instagram	Priority	Delivered	########	########		Netbankin	925908	1238.24	Paid	68	2022	November	100	
12	BTO55	SA811	1.46E+08	32	Male	Tata Samp	1974.77	1797.04	Local	Telangana	Adilabad	Freepass	9.59E+11	3.4	4.2	Instagram	Free	Shipped	########	########		UPI	735096	1797.04	Paid	30	2021	March	0	
13	EBX76	CK701	8.02E+08	22	Female	VSR Chann	628.34	534.09	Branded	Telangana	Warangal	Freepass	8.00E+11	12.76	2.2	Instagram	Free	Cancelled	########	########	########	Cancelled	142985	Cancelled	Cancelled	82	2021	January	Cancelled	
14	RMX58	PM007	14638526	70	Male	Ariel Mati	1098.82	977.95	Branded	Gujarat	Gandhinag	Freepass	7.77E+11	15.55	2.6	Other	Free	Shipped	########	########		Credit	826470	977.95	Paid	31	2023	April	0	
15	RQB15	OB086	8.54E+08	57	Male	Ariel Mati	1583.11	1155.67	Local	Gujarat	Ahmedaba	Freepass	5.27E+11	3.69	4.7	Instagram	Express plu	Shipped	########	########		UPI	178544	1305.67	Paid	61	2023	Septembe	150	
16	FXH87	KC821	5.92E+08	60	Female	Amul Taaz	662.86	570.06	Imported	Gujarat	Jamnagar	Premium P	4.29E+11	21.12	3.8	Friends	Free	Shipped	########	########		Credit	805134	570.06	Paid	58	2023	October	0	
17	LGO87	BL499	6.89E+08	51	Male	Kellogg's C	1858.26	1691.02	Branded	Gujarat	Gandhinag	Premium	6.54E+10	23.12	1.6	TV	Free	Shipped	########	########		Credit	428809	1691.02	Paid	95	2022	July	0	
18	MDV14	SN210	5.99E+08	37	Male	Usha Heav	1945.77	1537.16	Imported	Gujarat	Rajkot	Premium P	9.67E+11	22.89	2.8	Instagram	Free	Delivered	########	########		Debit	241243	1537.16	Paid	82	2023	Septembe	0	
19	CJI07	NW220	5.36E+08	31	Male	Tata Tea C	1573.51	1164.4	Branded	Telangana	Mahbubna	Freepass	9.30E+11	12.43	4.9	Instagram	Express plu	Delivered	########	########		Netbankin	658846	1314.4	Paid	10	2023	July	150	
20	KJL56	SP799	2.23E+08	37	Female	Usha Heav	255.66	173.85	Imported	Telangana	Adilabad	Freepass	7.27E+11	42.48	4.5	Friends	Free	Delivered	########	########		Debit	975641	173.85	Pending	25	2023	May	0	
21	IDC47	LQ240	8.08E+08	42	Female	Parle Nutr	1840.45	1085.87	Imported	Maharash	Navi Mum	Freepass	2.93E+11	28.8	1.3	Instagram	Free	Delivered	########	########		UPI	113687	1085.87	Paid	86	2021	August	0	
22	TXT34	XO802	4.02E+08	51	Male	Ariel Mati	983.11	717.67	Imported	Andhra Pra	Kurnool	Premium P	5.28E+11	1.55	1.2	Instagram	Free	Delivered	########	########		UPI	678901	717.67	Pending	9	2022	October	0	
23	TBP56	JA540	4.05E+08	25	Male	Parle Nutr	872.33	479.78	Branded	Gujarat	Vadodara	Premium P	9.10E+11	2.14	3.3	Friends	Priority	Shipped	########	########		Debit	137602	504.78	Paid	94	2022	June	25	
24	HNK32	TL255	7.51E+08	66	Female	Tenali Dou	1713.76	1610.93	Local	Gujarat	Junagadh	Premium P	2.34E+11	12.02	2.9	Other	Free	Shipped	########	########		COD	890001	1610.93	Pending	26	2021	March	0	
25	FGB94	TR196	4.31E+08	41	Female	Tata Tea	632.94	348.12	Branded	Maharash	Nashik	Freepass	8.09E+11	4.07	4	Other	Priority	Delivered	########	########		Credit	173847	448.12	Paid	63	2023	Septembe	100	
26	PHB13	RK268	1.85E+08	50	Male	Premia Tu	511.02	281.06	Local	Andhra Pra	Anantapur	Premium P	2.37E+11	25.49	3	Other	Priority	Returned	########	########		COD	825586	Returned	Returned	1	2022	February	25	
27	OTK45	UY146	1.12E+08	56	Female	Tata Tea	1840.75	1325.34	Local	Andhra Pra	Kurnool	Freepass	7.69E+11	11.73	3	Instagram	Priority	Delivered	########	########		Debit	282217	1425.34	Pending	24	2023	October	100	
28	RXF01	SN619	2.75E+08	18	Female	Parle Nutr	1529.96	1300.47	Branded	Maharash	Pune	Premium	2.37E+11	3.05	2.8	Other	Free	Delivered	########	########		Credit	358515	1300.47	Paid	96	2021	July	0	
29	MVH31	HL558	71745481	50	Male	Ariel Mati	1917.21	1342.05	Imported	Andhra Pra	Tirupati	Freepass	3.55E+11	1.73	4.8	Other	Free	Shipped	########	########		Debit	456340	1342.05	Pending	55	2023	October	0	
30	KQA72	DS490	7.07E+08	33	Female	Pigeon Ind	615.99	314.15	Imported	Telangana	Adilabad	Premium	2.28E+11	55.23	3.8	Instagram	Express plu	Delivered	########	########		UPI	869208	389.15	Paid	68	2022	June	75	
31	RJD76	QW165	67653253	63	Female	Pigeon Ind	190.15	171.14	Branded	Gujarat	Jamnagar	Premium	8.33E+11	2.05	4.6	Instagram	Free	Shipped	########	########		Credit	673418	171.14	Paid	35	2021	February	0	
32	SGZ25	LT130	5.54E+08	31	Female	Tata Samp	530.2	397.65	Branded	Maharash	Amravati	Freepass	1.45E+11	2.28	4.3	Instagram	Free	Delivered	########	########		UPI	283049	397.65	Pending	45	2021	Septembe	0	
33	NCA40	RF710	1.54E+08	28	Male	Parle Nutr	354.08	265.56	Branded	Maharash	Pune	Freepass	5.54E+11	10.77	4.5	Instagram	Express plu	Shipped	########	########		UPI	240194	415.56	Paid	27	2021	January	150	
34	JJA90	WZ081	6.23E+08	47	Male	VSR Chann	610.88	464.27	Local	Gujarat	Vadodara	Freepass	1.33E+10	3.52	1.8	TV	Free	Shipped	########	########		Netbankin	511862	464.27	Paid	74	2023	December	0	
35	ZJQ35	YA757	6.89E+08	56	Male	Tata Tea C	279.83	162.3	Branded	Gujarat	Bharuch	Premium	2.71E+10	9.59	1.9	Instagram	Free	Returned	########	########		Debit	877211	Returned	Returned	74	2021	June	0	
36	JAJ58	UW814	1.93E+08	51	Female	Usha Heav	830.89	548.39	Local	Telangana	Nizamaba	Freepass	9.33E+11	6.36	1.8	Facebook	Priority	Delivered	########	########		Debit	146043	648.39	Paid	28	2023	February	100	
37	MAI00	BS295	2.79E+08	33	Female	Harpic Bat	1907.99	1354.67	Local	Andhra Pra	Nellore	Premium P	8.07E+11	21.62	1.5	Facebook	Priority	Shipped	########	########		Debit	389040	1379.67	Pending	13	2021	April	25	

2.2 Data Preprocessing & Cleaning:

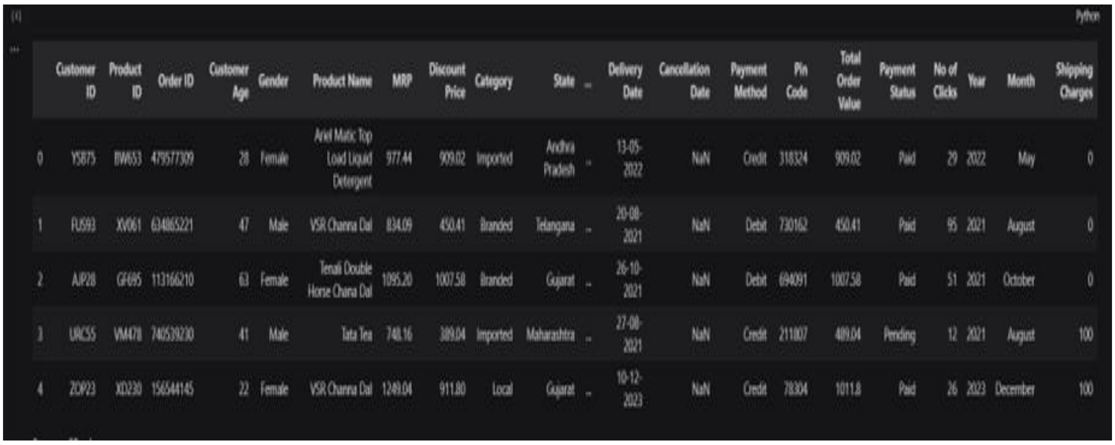
Data Preprocessing:

- pandas (pd): Handles data frames, like spreadsheets, for analysis and manipulation.
- numpy (np): Provides powerful tools for working with numbers in arrays and matrices.
- seaborn (sns): Creates informative and visually appealing statistical charts.
- matplotlib.pyplot (plt): Makes various plots, from basic to complex and interactive.
- geopandas (gpd): Adds geographic data handling capabilities to pandas.
- LabelEncoder: Converts text labels (like categories) into numerical values.
- LinearRegression: Creates models to predict values based on relationships between features.
- datetime: Works with dates and times for data processing.
- plotly.express (px): Builds interactive charts for exploration.
- plotly.graph_objects (go): Offers more control for creating complex visualizations.



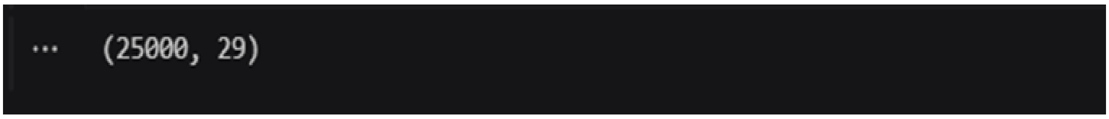
```python
import pandas as pd  # Data manipulation and analysis.
import numpy as np  # Numerical computing with support for arrays and matrices.
import seaborn as sns  # High-level statistical data visualization.
import matplotlib.pyplot as plt  # Plotting framework for static, interactive, and animated visualizations.
import geopandas as gpd  # Extends pandas for spatial data operations.
from matplotlib.lines import Line2D  # Custom line styles in matplotlib plots.
from sklearn.preprocessing import LabelEncoder  # Encoding categorical features into numeric format.
from sklearn.linear_model import LinearRegression  # Linear regression models for predictive analysis.
# import colorcet as cc  # Perceptually uniform colormaps.
from palettable.colorbrewer.qualitative import Pastel1_7  # Qualitative color palettes.
from datetime import datetime, timedelta  # Date and time manipulation.
import plotly.express as px  # High-level interface for interactive visualizations.
import plotly.graph_objects as go  # Low-level interface for complex visualizations.
```

- Import: pd.read_csv function from the pandas library is used to read data from a CSV file.
- File path: The file path is specified as r'D:\project\New folder\final test1.csv'.
- Data frame creation: The loaded data is converted into a pandas dataframe, a tabular data structure, and assigned to the variable data.
- Data Exploration: data.head() method displays the first few rows of the dataframe, providing a glimpse into the data's contents.



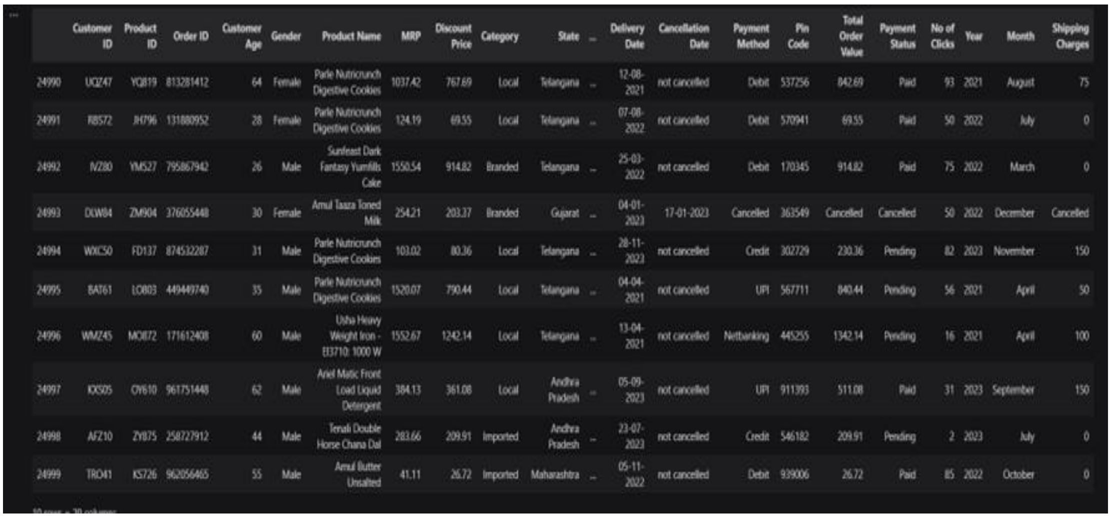
	Customer ID	Product ID	Order ID	Customer Age	Gender	Product Name	MRP	Discount Price	Category	State	...	Delivery Date	Cancellation Date	Payment Method	Pin Code	Total Order Value	Payment Status	No of Clicks	Year	Month	Shipping Charges
0	YS875	BW653	479577309	28	Female	Ariel Matic Top Load Liquid Detergent	977.44	909.02	Imported	Andhra Pradesh	...	13-05-2022	NaN	Credit	318324	909.02	Paid	29	2022	May	0
1	FU593	XV061	634865221	47	Male	VSR Channa Dal	834.09	450.41	Branded	Telangana	...	20-08-2021	NaN	Debit	730162	450.41	Paid	95	2021	August	0
2	AJP28	GF695	113166210	63	Female	Tenali Double Horse Chana Dal	1095.20	1007.58	Branded	Gujarat	...	26-10-2021	NaN	Debit	694091	1007.58	Paid	51	2021	October	0
3	URC55	VM478	740539230	41	Male	Tata Tea	748.16	389.04	Imported	Maharashtra	...	27-08-2021	NaN	Credit	211807	409.04	Pending	12	2021	August	100
4	ZOP23	XD230	156544145	22	Female	VSR Channa Dal	1249.04	911.80	Local	Gujarat	...	10-12-2023	NaN	Credit	78304	1011.8	Paid	26	2023	December	100

- print(data.shape): This line prints the shape of the data, which is the number of rows and columns in the dataframe. It helps understand the size of the data you're working with.



```
···   (25000, 29)
```

- data.tail(): This line displays the last few rows of the dataframe by default (typically 5 rows). This gives you a glimpse of what the data looks like towards the end.



	Customer ID	Product ID	Order ID	Customer Age	Gender	Product Name	MRP	Discount Price	Category	State	...	Delivery Date	Cancellation Date	Payment Method	Pin Code	Total Order Value	Payment Status	No of Clicks	Year	Month	Shipping Charges
24990	UQZ47	YQ819	813281412	64	Female	Parle Nutricrunch Digestive Cookies	1037.42	767.69	Local	Telangana	...	12-08-2021	not cancelled	Debit	537256	842.69	Paid	93	2021	August	75
24991	RBS72	JH796	131880952	28	Female	Parle Nutricrunch Digestive Cookies	124.19	69.55	Local	Telangana	...	07-08-2022	not cancelled	Debit	570941	69.55	Paid	50	2022	July	0
24992	IVZ80	YMS27	795867942	26	Male	Sunfeast Dark Fantasy Yumfills Cake	1550.54	914.82	Branded	Telangana	...	25-03-2022	not cancelled	Debit	170345	914.82	Paid	75	2022	March	0
24993	DLW84	ZM904	376055448	30	Female	Amul Taaza Toned Milk	254.21	203.37	Branded	Gujarat	...	04-01-2023	17-01-2023	Cancelled	363549	Cancelled	Cancelled	50	2022	December	Cancelled
24994	WXC50	FD137	874532287	31	Male	Parle Nutricrunch Digestive Cookies	103.02	80.36	Local	Telangana	...	28-11-2023	not cancelled	Credit	302729	230.36	Pending	82	2023	November	150
24995	BAT61	LO803	449449740	35	Male	Parle Nutricrunch Digestive Cookies	1520.07	790.44	Local	Telangana	...	04-04-2021	not cancelled	UPI	567711	840.44	Pending	56	2021	April	50
24996	WMZ45	MO872	171612408	60	Male	Usha Heavy Weight Iron - EI3710: 1000 W	1552.67	1242.14	Local	Telangana	...	13-04-2021	not cancelled	Netbanking	445255	1342.14	Pending	16	2021	April	100
24997	KXS05	OY610	961751448	62	Male	Ariel Matic Front Load Liquid Detergent	384.13	361.08	Local	Andhra Pradesh	...	05-09-2023	not cancelled	UPI	911393	511.08	Paid	31	2023	September	150
24998	AFZ10	ZY875	258727912	44	Male	Tenali Double Horse Chana Dal	283.66	209.91	Imported	Andhra Pradesh	...	23-07-2023	not cancelled	Credit	546182	209.91	Pending	2	2023	July	0
24999	TRO41	KS726	962056465	55	Male	Amul Butter Unsalted	41.11	26.72	Imported	Maharashtra	...	05-11-2022	not cancelled	Debit	939006	26.72	Paid	85	2022	October	0

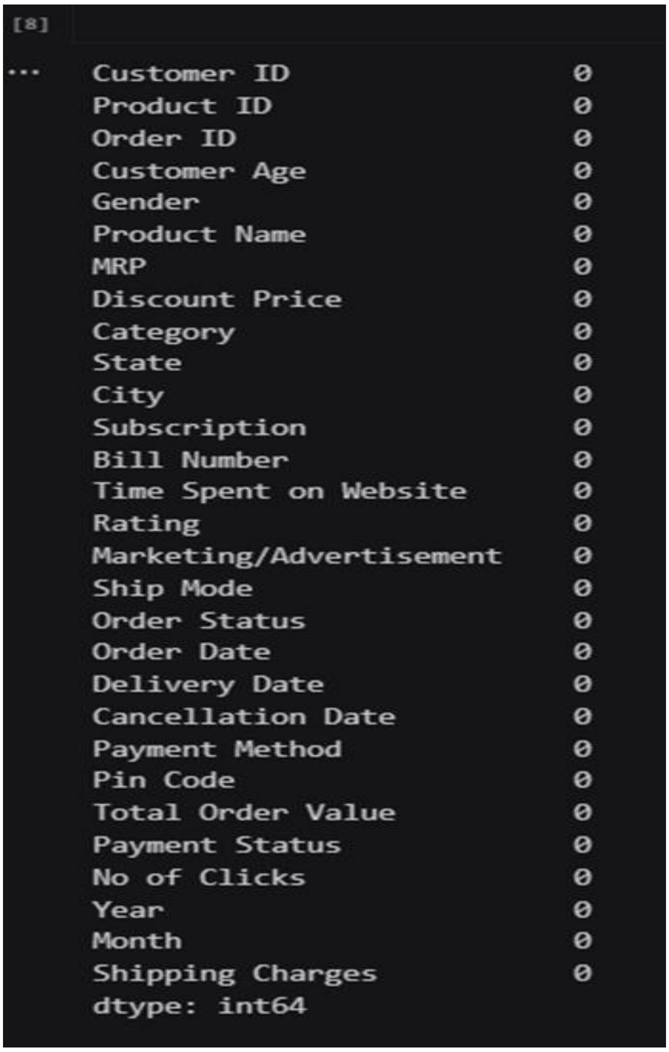
10 rows × 30 columns

- data.sample(5): This line displays a random sample of 5 rows from the dataframe. This helps get a general idea of the data's content without looking at all rows.

Data Cleaning:

- Checks for missing values: data.isnull().sum() counts missing values in each column of the pandas dataframe data.
- Fills missing cancellation dates: It replaces missing values in the 'Cancellation Date' column with the text 'not cancelled' using data['Cancellation Date'] = data['Cancellation Date'].fillna('not cancelled').
- Shows last 10 rows: data.tail(10) displays the last 10 rows of the dataframe after the cleaning operations.
- # Check for missing values
- data.isnull().sum()   # Basic statistics
- data.describe()   # Display cleaned data
- data.info()

- Check for missing values: data.isnull().sum() counts missing values in each column of the data frame data.
- Basic statistics: data.describe() generates summary statistics of the data, like mean, standard deviation, quartiles, etc., for numerical columns.
- Display data info: data.info() provides information about the data, including data types and non-null value counts for each column.



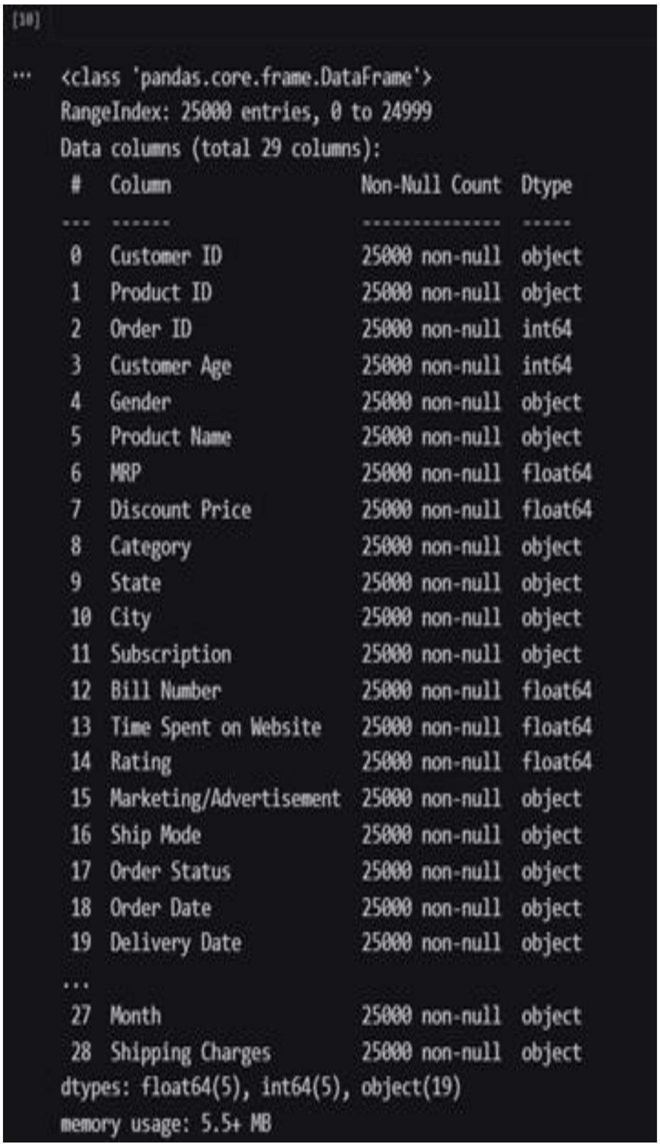
```
[8]

...   Customer ID                  0
      Product ID                   0
      Order ID                     0
      Customer Age                 0
      Gender                       0
      Product Name                 0
      MRP                          0
      Discount Price               0
      Category                     0
      State                        0
      City                         0
      Subscription                 0
      Bill Number                  0
      Time Spent on Website        0
      Rating                       0
      Marketing/Advertisement      0
      Ship Mode                    0
      Order Status                 0
      Order Date                   0
      Delivery Date                0
      Cancellation Date            0
      Payment Method               0
      Pin Code                     0
      Total Order Value            0
      Payment Status               0
      No of Clicks                 0
      Year                         0
      Month                        0
      Shipping Charges             0
      dtype: int64
```

Handling Missing Numerical Values:

- The code attempts to convert the 'Total Order Value' column into numerical format using pd.to_numeric.
- The errors='coerce' argument instructs the function to replace any conversion errors (likely due to non-numeric values) with NaN (Not a Number).

Removing Rows with Missing Values:

- The dataframe is then filtered using df.dropna(subset=['Total Order Value']).
- This keeps only rows where the 'Total Order Value' column has valid numerical data (not NaN).



```
[10]

...  <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 25000 entries, 0 to 24999
     Data columns (total 29 columns):
      #   Column                    Non-Null Count  Dtype
     ---  ------                    --------------  -----
      0   Customer ID               25000 non-null  object
      1   Product ID                25000 non-null  object
      2   Order ID                  25000 non-null  int64
      3   Customer Age              25000 non-null  int64
      4   Gender                    25000 non-null  object
      5   Product Name              25000 non-null  object
      6   MRP                       25000 non-null  float64
      7   Discount Price            25000 non-null  float64
      8   Category                  25000 non-null  object
      9   State                     25000 non-null  object
      10  City                      25000 non-null  object
      11  Subscription              25000 non-null  object
      12  Bill Number               25000 non-null  float64
      13  Time Spent on Website     25000 non-null  float64
      14  Rating                    25000 non-null  float64
      15  Marketing/Advertisement   25000 non-null  object
      16  Ship Mode                 25000 non-null  object
      17  Order Status              25000 non-null  object
      18  Order Date                25000 non-null  object
      19  Delivery Date             25000 non-null  object
     ...
      27  Month                     25000 non-null  object
      28  Shipping Charges          25000 non-null  object
     dtypes: float64(5), int64(5), object(19)
     memory usage: 5.5+ MB
```

In simpler terms, this code cleans a data column by converting text values (if any) to numbers and then removes rows with missing numbers. This prepares the data for numerical analysis.

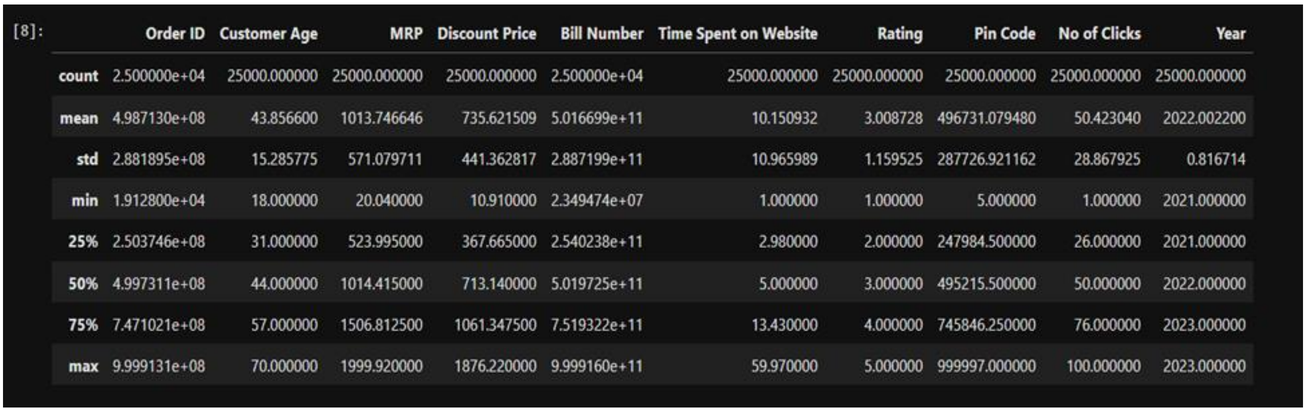
The dataset includes 28 columns, providing a comprehensive range of data points for analysis. Each column represents a different variable or attribute, contributing to a detailed and multifaceted dataset. This diversity of data enhances the potential for in-depth insights and thorough examination across multiple dimensions of the dataset.

1. **Customer ID:** Unique identifier for each customer.
2. **Product ID:** Unique identifier for each product.
3. **Order ID:** Unique identifier for each order.
4. **Customer Age:** Age of the customer.
5. **Gender:** Gender of the customer.
6. **Product Name:** Name of the product.
7. **MRP:** Maximum Retail Price of the product.
8. **Discount Price:** Discounted price of the product.
9. **Category: C**ategory of the product (e.g., Local, Branded).
10. **State:** State from which the order was placed.
11. **City:** City from which the order was placed.
12. **Subscription:** Subscription status of the customer.
13. **Bill Number:** Bill number for the order.
14. **Time Spent on Website:** Time spent by the customer on the website.
15. **Rating:** Rating given by the customer.
16. **Marketing/Advertisement:** Marketing or advertisement details.
17. **Ship Mode:** Shipping mode used for the order.
18. **Order Status:** Current status of the order.
19. **Order Date:** Date on which the order was placed.
20. **Delivery Date:** Date on which the order was delivered.
21. **Cancellation Date:** Date on which the order was cancelled (if applicable).
22. **Payment Method:** Method used for payment.
23. **Pin Code:** Pin code of the delivery address.
24. **Total Order Value:** Total value of the order.
25. **Payment Status:** Status of the payment (e.g., Paid, Pending).
26. **No of Clicks:** Number of clicks made by the customer.
27. **Year:** Year in which the order was placed.
28. **Month:** Month in which the order was placed.

Chapter 3

Data Analysis Techniques

In this report, we present a comprehensive analysis of the D-Mart ready sales database, focusing on uncovering trends, patterns, and insights that can drive informed decision-making. Our analysis begins with an essential step: data cleaning. The raw sales data is meticulously prepared by addressing missing values, removing duplicates, and correcting inconsistencies. This process ensures the integrity and accuracy of the data, laying a solid foundation for subsequent analysis.



	Order ID	Customer Age	MRP	Discount Price	Bill Number	Time Spent on Website	Rating	Pin Code	No of Clicks	Year
count	2.500000e+04	25000.000000	25000.000000	25000.000000	2.500000e+04	25000.000000	25000.000000	25000.000000	25000.000000	25000.000000
mean	4.987130e+08	43.856600	1013.746646	735.621509	5.016699e+11	10.150932	3.008728	496731.079480	50.423040	2022.002200
std	2.881895e+08	15.285775	571.079711	441.362817	2.887199e+11	10.965989	1.159525	287726.921162	28.867925	0.816714
min	1.912800e+04	18.000000	20.040000	10.910000	2.349474e+07	1.000000	1.000000	5.000000	1.000000	2021.000000
25%	2.503746e+08	31.000000	523.995000	367.665000	2.540238e+11	2.980000	2.000000	247984.500000	26.000000	2021.000000
50%	4.997311e+08	44.000000	1014.415000	713.140000	5.019725e+11	5.000000	3.000000	495215.500000	50.000000	2022.000000
75%	7.471021e+08	57.000000	1506.812500	1061.347500	7.519322e+11	13.430000	4.000000	745846.250000	76.000000	2023.000000
max	9.999131e+08	70.000000	1999.920000	1876.220000	9.999160e+11	59.970000	5.000000	999997.000000	100.000000	2023.000000

Figure: Statistical Analysis of the dataset

Customer Lifetime Value (CLV) Analysis

Customer lifetime value (CLV) analysis is another critical component of our report. This analysis focuses on understanding the long-term value of different customer segments by examining historical purchase data and customer interactions. Estimating the future value each customer brings to the company helps in identifying high-value customers, optimizing retention strategies, and improving targeted marketing efforts.

Sentiment Analysis

Additionally, we conduct sentiment analysis on customer feedback, reviews, and social media interactions to gain insights into customer satisfaction and brand perception. By analyzing sentiments expressed in customer reviews and social media posts, we identify areas for improvement, address customer concerns, and enhance the overall customer experience.

Descriptive Statistics

The dataset consists of 25,000 entries with various attributes related to retail transactions. Here are some key descriptive statistics:

1. Customer Age ranges from 18 to 70 years, with a mean age of 44.08.

2. Total Order Value ranges from 10.73 to 1888.26, with a mean value of 730.38.

The dataset includes both numerical and categorical variables, providing a comprehensive view of retail transactions, customer demographics, and product details.

Numerical Attributes

For numerical attributes, we calculated the mean, median, standard deviation, minimum, maximum, and quartiles.

1.Order ID

- Mean: 498,713,020.8
- Median: 499,731,100
- Standard Deviation: 288,189,479.7
- Min: 19,128
- Max: 999,913,100
- Quartiles:

  25%: 250,374,600

  75%: 747,102,100

2.Customer Age

- Mean: 43.9 years
- Median: 44 years
- Standard Deviation: 15.3 years
- Min: 18 years
- Max: 70 years
- Quartiles:

  25%: 31 years

  75%: 57 years

3.MRP (Maximum Retail Price)

- Mean: ₹1,013.75
- Median: ₹1,014.42
- Standard Deviation: ₹571.08
- Min: ₹20.04
- Max: ₹1,999.92

- Quartiles:

  25%: ₹523.99

  75%: ₹1,506.81

4.Discount Price

- Mean: ₹735.62
- Median: ₹713.14
- Standard Deviation: ₹441.36
- Min: ₹10.91
- Max: ₹1,876.22
- Quartiles:

  25%: ₹367.67

  75%: ₹1,061.35

5.Bill Number

- Mean: 501,669,925,947
- Median: 501,972,475,013
- Standard Deviation: 288,719,913,506
- Min: 23,494,736
- Max: 999,916,008,244
- Quartiles:

  25%: 254,023,782,185

  75%: 751,932,231,770

6.Time Spent on Website

- Mean: 10.15 minutes
- Median: 5 minutes
- Standard Deviation: 10.97 minutes
- Min: 1 minute
- Max: 59.97 minutes
- Quartiles:

  25%: 2.98 minutes

  75%: 13.43 minutes

7.Rating

- Mean: 3.01
- Median: 3
- Standard Deviation: 1.16
- Min: 1
- Max: 5

- Quartiles:

  25%:2

  75%: 4

8.Pin Code

- Mean: 496,731.08
- Median: 495,215.5
- Standard Deviation: 287,726.92
- Min: 5
- Max: 999,997
- Quartiles:

  25%: 247,984.5

  75%: 745,846.25

9.No of Clicks

- Mean: 50.42
- Median: 50
- Standard Deviation: 28.87
- Min: 1
- Max: 100
- Quartiles:

  25%: 26

  75%: 76

10.Year

- Mean: 2022
- Median: 2022
- Standard Deviation: 0.82
- Min: 2021
- Max: 2023
- Quartiles:

  25%: 2021

  75%: 2023

Analysis

- Order ID: The mean Order ID is approximately 498.7 million, with a wide range from 19,128 to 999.9 million, indicating a large number of transactions.
- Customer Age: The average customer age is about 44 years, with ages ranging from 18 to 70 years, showing a broad age distribution among customers.
- MRP: The mean MRP is around ₹1,013.75, with a maximum of ₹1,999.92, reflecting a wide price range for products.
- Discount Price: On average, the discount price is ₹735.62, with discounts ranging from ₹10.91 to ₹1,876.22.
- Bill Number: The average bill number is 501.7 billion, with a wide distribution, indicating a large volume of billing entries.
- Time Spent on Website: Customers spend an average of 10.15 minutes on the website, with some spending up to nearly 60 minutes, suggesting varying engagement levels.
- Rating: The average rating is 3, indicating moderate customer satisfaction.
- Pin Code: The mean pin code is 496,731, showing transactions across a wide range of geographical locations.
- No of Clicks: The average number of clicks is 50.42, with a maximum of 100 clicks, indicating varied interaction levels with the website.
- Year: The data spans from 2021 to 2023, with a mean year of 2022, indicating a three-year analysis period.

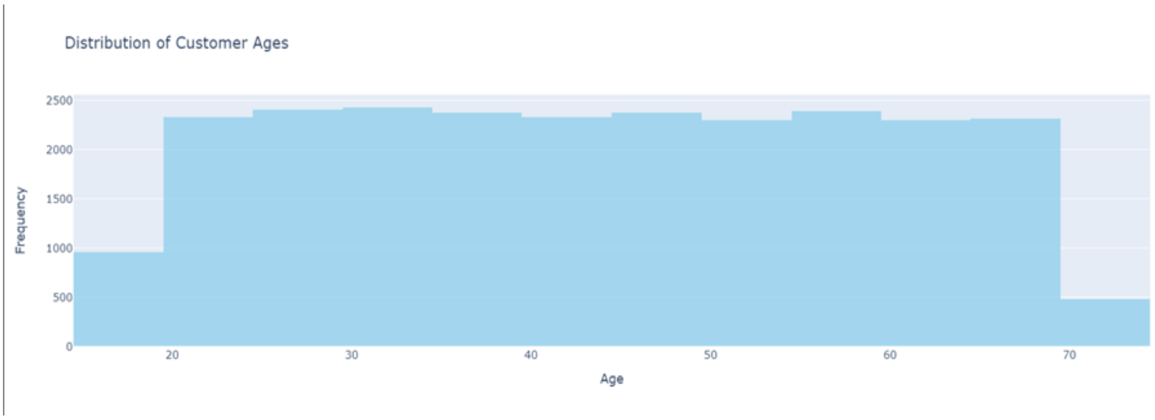
Chapter 4

Data Visualisation

4.1 Using Python:

1. **Histogram**

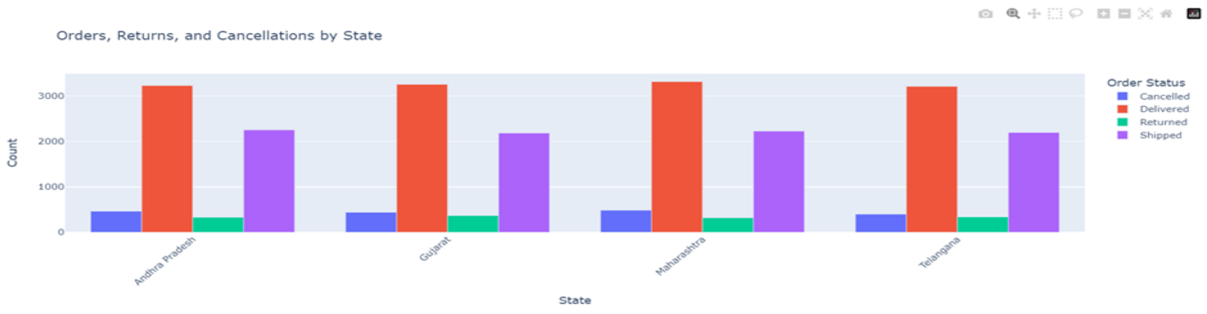
   The histogram visualizes the age distribution of D-Mart customers, displaying a relatively uniform spread across ages 20 to 70. Each age group, particularly between 20 and 60, shows a high frequency of around 2,000 to 2,500 customers, indicating a diverse and evenly distributed customer age range. This suggests that D-Mart's customer base is broad and appeals consistently across various age demographics. The consistency in customer age frequency highlights the brand's wide-reaching appeal and the potential for age-targeted marketing strategies to further capitalize on this evenly distributed customer demographic.



2. **Stacked-Bar Graph**

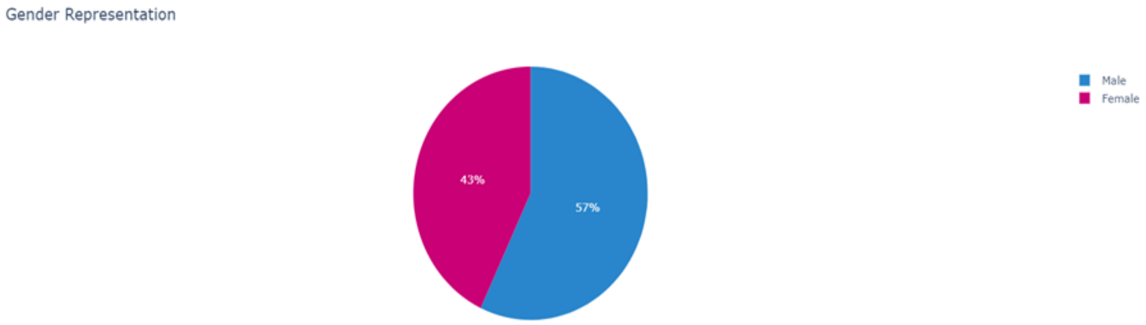
   This Stacked-bar graph represents the distribution of orders, returns, and cancellations by state. The data is grouped by State and Order Status, with the count of each status displayed for each state.

   This Stacked-bar graph shows the counts of orders, returns, and cancellations across four states: Andhra Pradesh, Gujarat, Maharashtra, and Telangana. Each state has a high number of delivered and shipped orders, with Telangana having the highest shipped orders. The cancellation rate is significant in all states, suggesting an area for potential improvement.



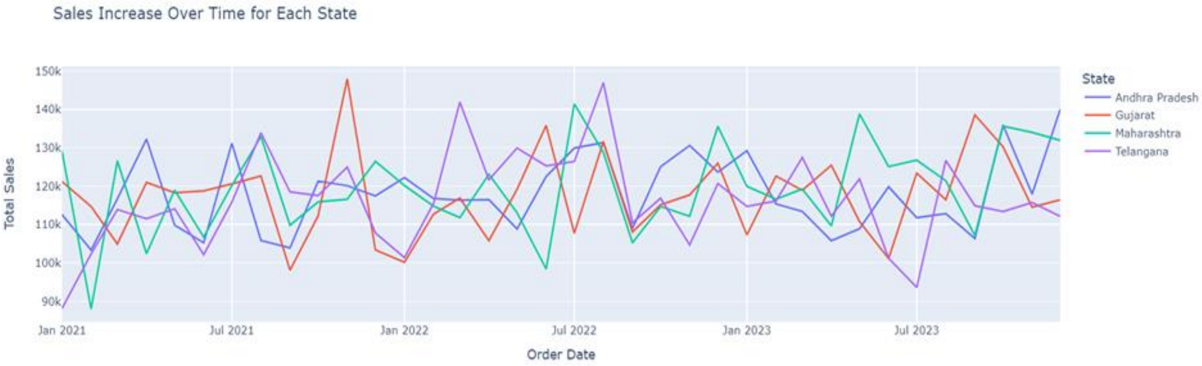
3. **Pie Chart**

This pie chart illustrates the gender representation of D-Mart customers, with 57% male and 43% female. The chart uses distinct colours to differentiate between genders: light blue for males and light pink for females. The visual highlights a higher proportion of male customers, indicating a potential focus area for marketing strategies aimed at increasing female customer engagement. This representation provides a clear and concise view of the gender distribution within the customer base.



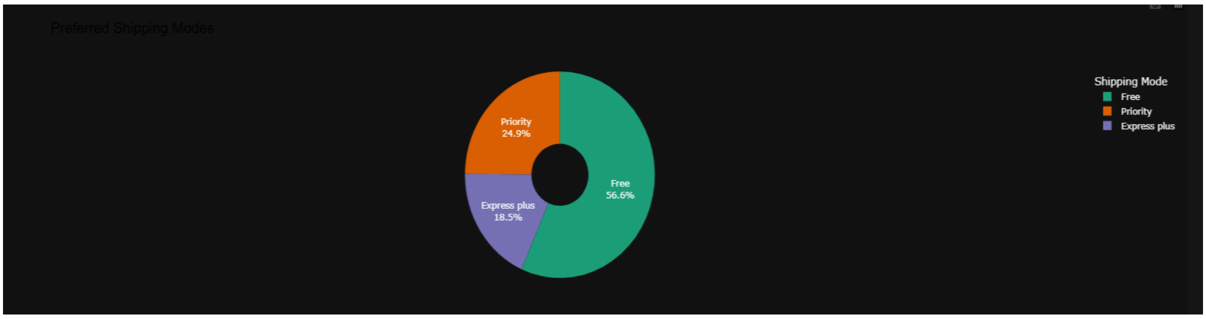
4. **Grouped-Line Chart**

This Grouped line chart illustrates the total sales over time for D-Mart across four Indian states: Andhra Pradesh, Gujarat, Maharashtra, and Telangana. The data reveals significant fluctuations in sales across all states, with no consistent upward or downward trend. Notably, Gujarat stands out with the highest peaks, particularly around January 2021 and late 2022, indicating periods of exceptional sales. In contrast, Andhra Pradesh and Maharashtra display more stable sales patterns with moderate variations, suggesting a more consistent sales performance. Telangana shows the highest volatility, with sharp increases and decreases, indicating a more unpredictable sales trend. Overall, the sales trends across all states appear to follow similar seasonal patterns, suggesting the influence of common factors such as holidays, festivals, or economic conditions during specific periods.



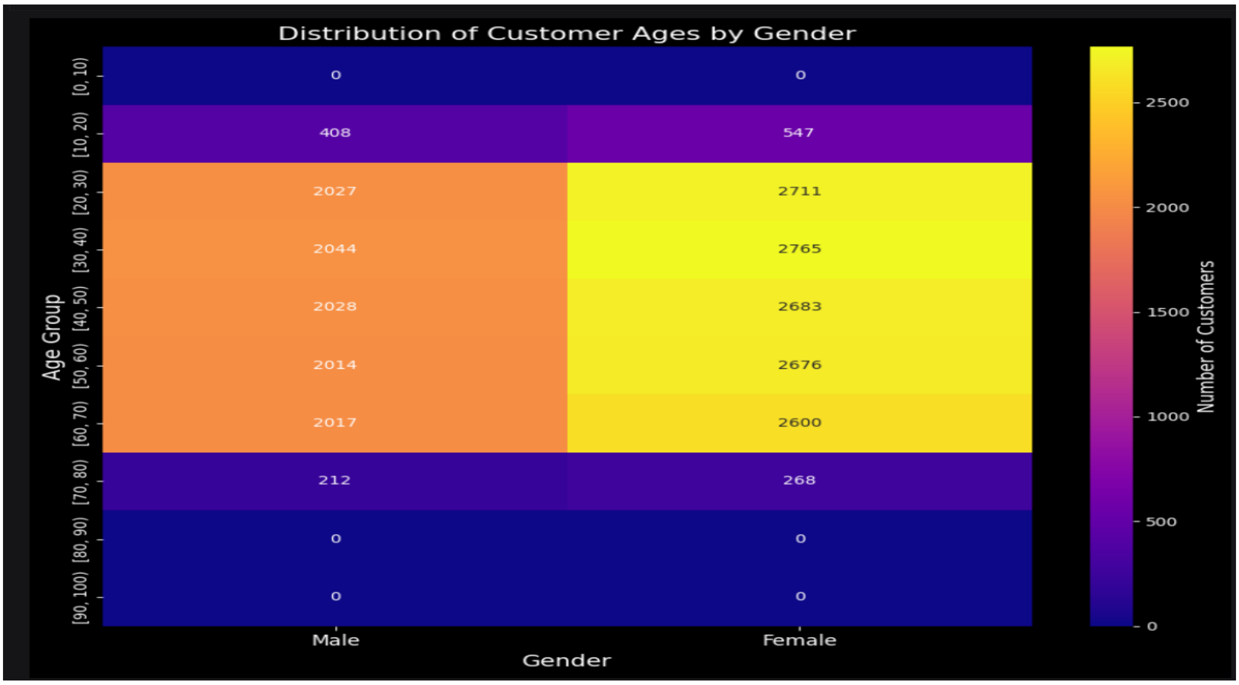
5. **Donut Chart**

The graph is a donut chart, a variation of a pie chart with a hollow center, representing the distribution of preferred shipping modes among customers. It reveals that a majority of customers (56.6%) prefer free shipping, indicating a strong preference for cost-saving options. Priority shipping is the second most popular choice, accounting for 24.9% of the total, suggesting that a significant portion of customers value quicker delivery times. Express plus shipping is chosen by 18.5% of customers, further highlighting the demand for expedited shipping services. Overall, while free shipping dominates, a considerable number of customers are willing to pay extra for faster delivery options.



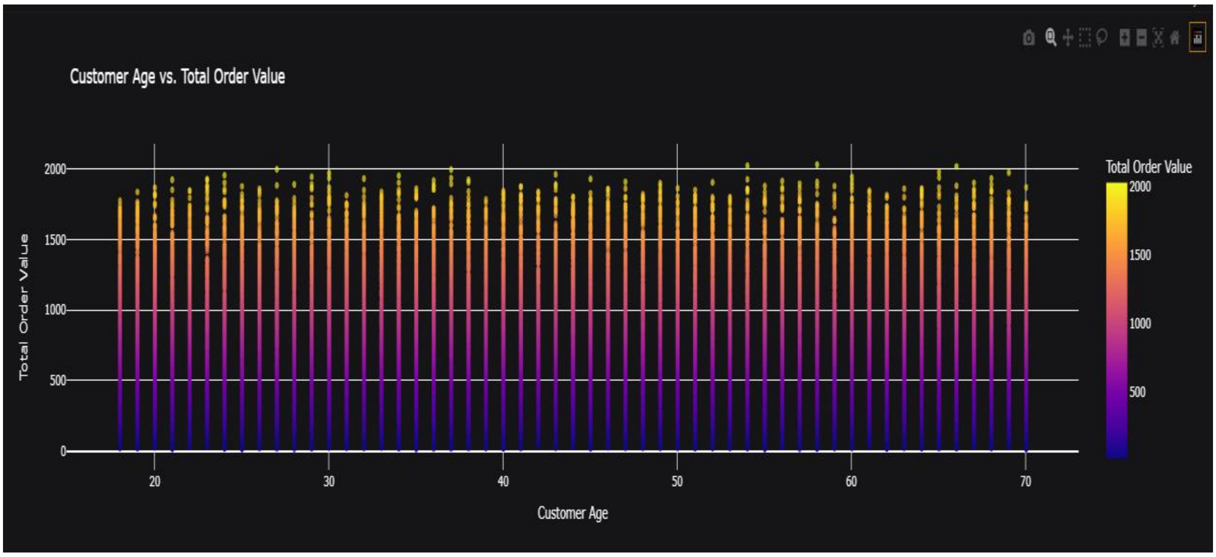
6. **Heat Map**

The heatmap visualizes the distribution of customer ages by gender, using color intensity to represent the number of customers in each age group. It shows that the 30-40 age group has the highest customer count, with females slightly outnumbering males. Most customers are between 20 and 80 years old, with very few in the 0-10 and 90-100 ranges. Females generally have higher counts in most age groups, except for the 10-20 range where both genders have low numbers. The 50-60 and 60-70 age groups have nearly equal distribution between genders.



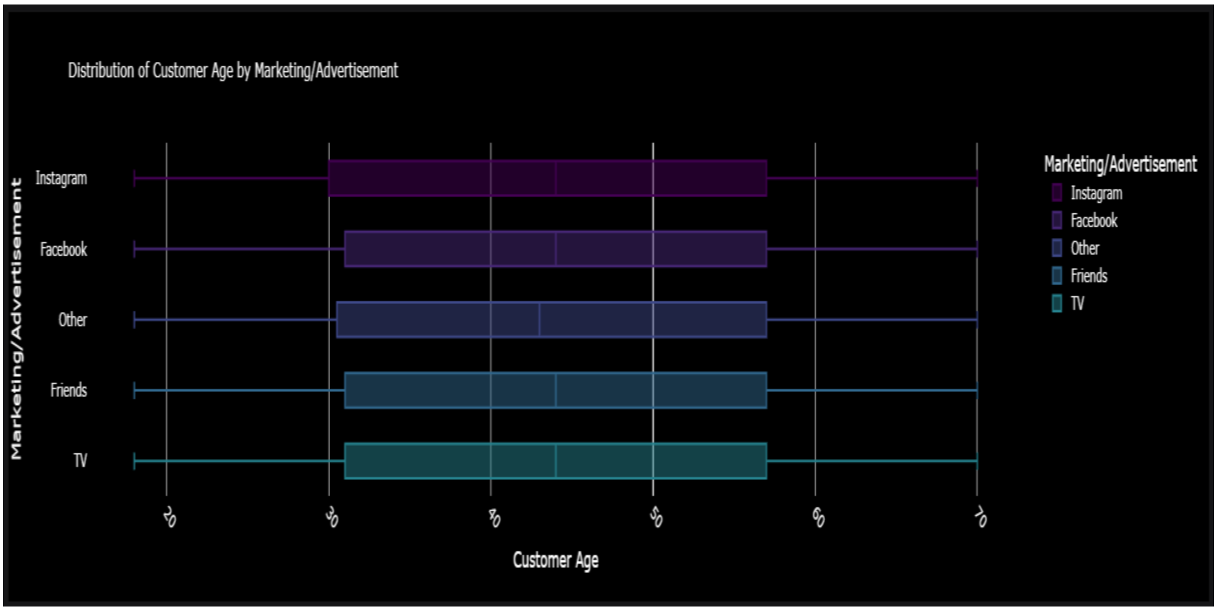
7. **Scatter plot**

The 3D scatter plot visualizes data points in a three-dimensional space, providing a comprehensive view of customer interaction metrics. In this plot, the x-axis represents customer age, offering insight into the age distribution of the customer base. The y-axis indicates the time spent on the website, showing how long customers engage with the site. The z-axis represents the number of clicks, reflecting customer activity and interaction levels. Each data point is colored based on the time spent on the website, creating a color gradient that highlights variations in engagement. This visualization effectively combines age, engagement time, and interaction frequency, allowing for a detailed analysis of customer behavior patterns.



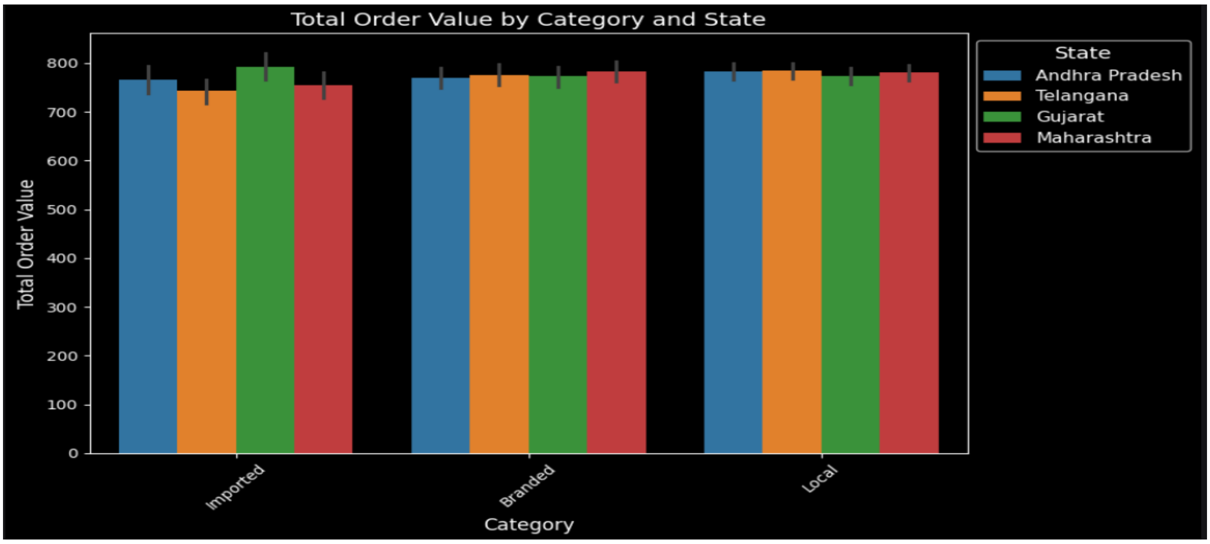
8. **Box Plot**

The box plot displays the distribution of customer age for each marketing/advertisement category. The x-axis represents the customer age, and the y-axis represents the marketing/advertisement categories



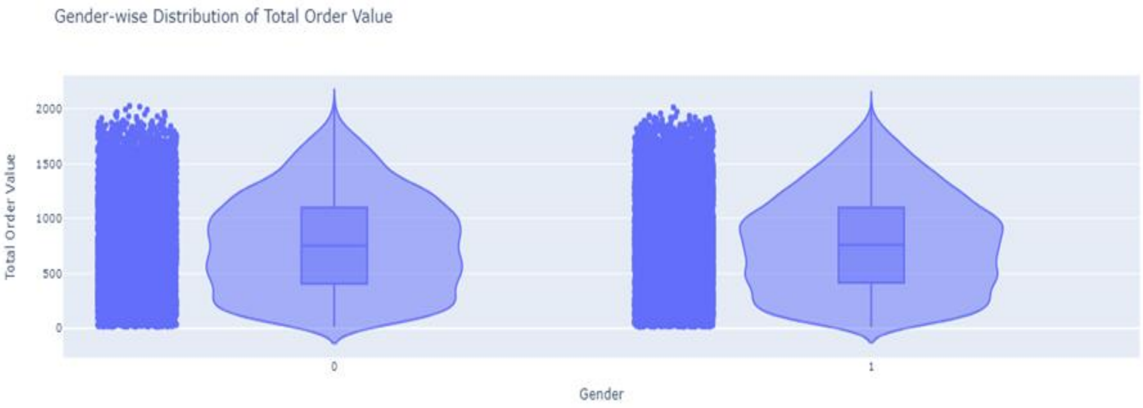
9. **Grouped Bar Plot**

A grouped bar plot illustrates the total order values across different categories, categorized by states. Each category is represented by a set of bars, where each bar shows the total value of orders. States within each category are distinguished by different colours, providing a clear visual comparison of order values across categories and states simultaneously. This visualization helps in identifying patterns of order distribution across various states within each category effectively.



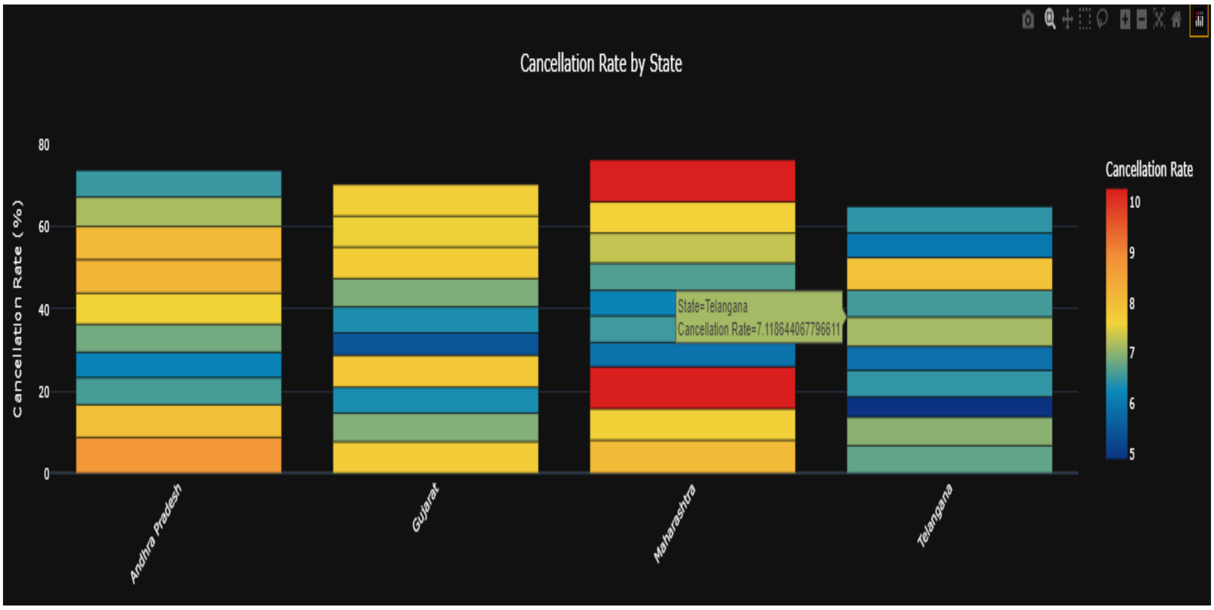
10. **Violin Plot**

A violin plot offers a comprehensive view of the distribution of total order values between genders. For each gender category (Male and Female) on the x-axis, the plot shows the distribution of order values on the y-axis. It combines the statistical summary of a box plot (median, quartiles, outliers) with a density plot, illustrating the probability density of the data at different values. This detailed visualization aids in comparing the spread, skewness, and central tendency of order values across genders, offering insights into the variability and shape of the distribution within each group.

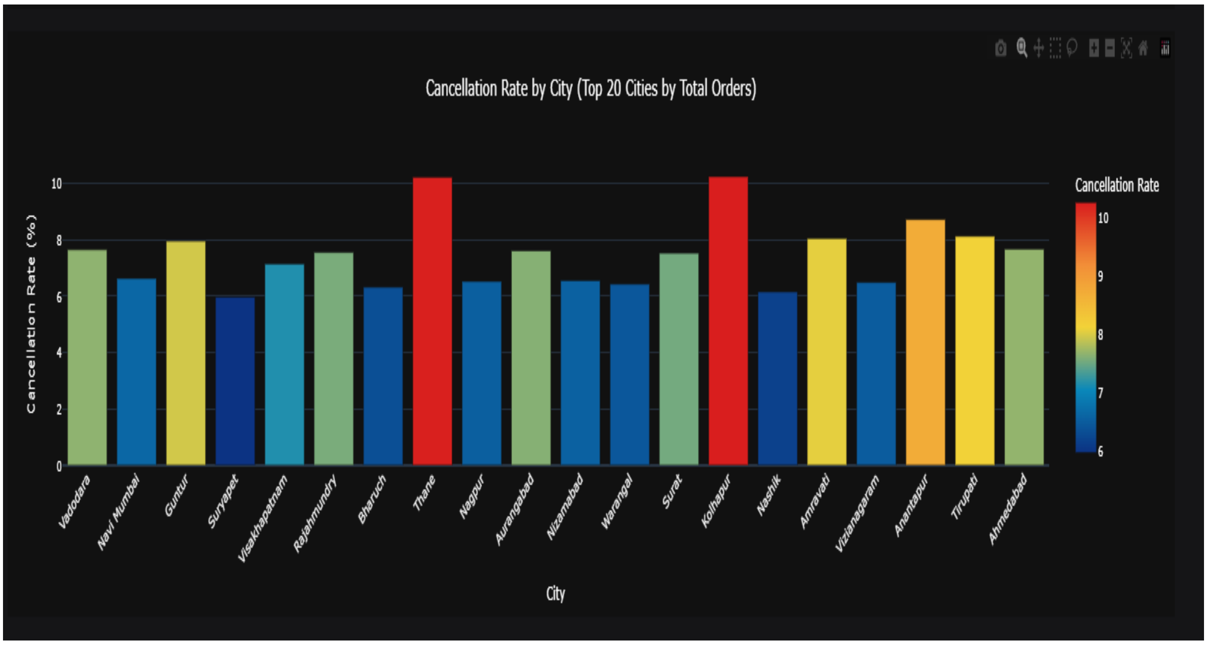


11. Bar Plot

The bar plot visually represents the count of cancelled orders categorized by states, offering a straightforward comparison across different regions. Each bar corresponds to a state, with its height indicating the number of cancelled orders. This allows for immediate identification of states with higher cancellation rates, highlighting potential areas for enhancing customer service or optimizing order management processes. Such insights can drive targeted interventions to improve operational efficiency and customer satisfaction, based on specific regional patterns revealed by the visualization.

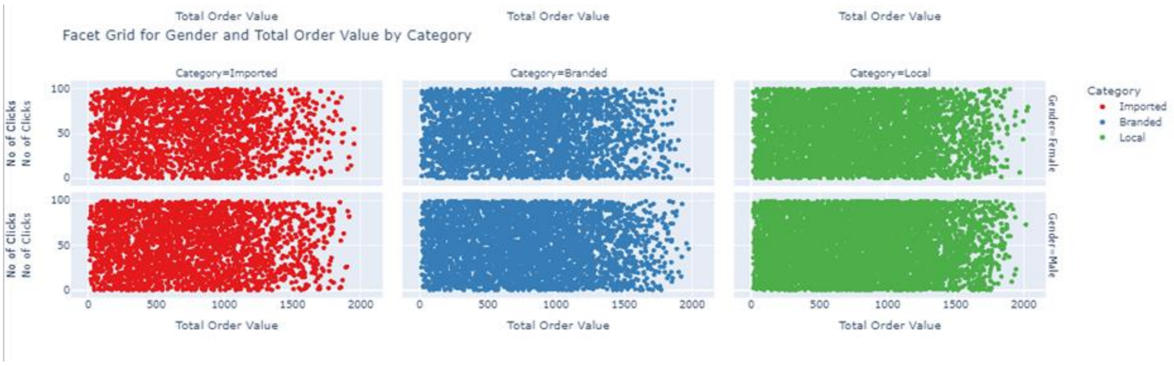


Cancellation by city



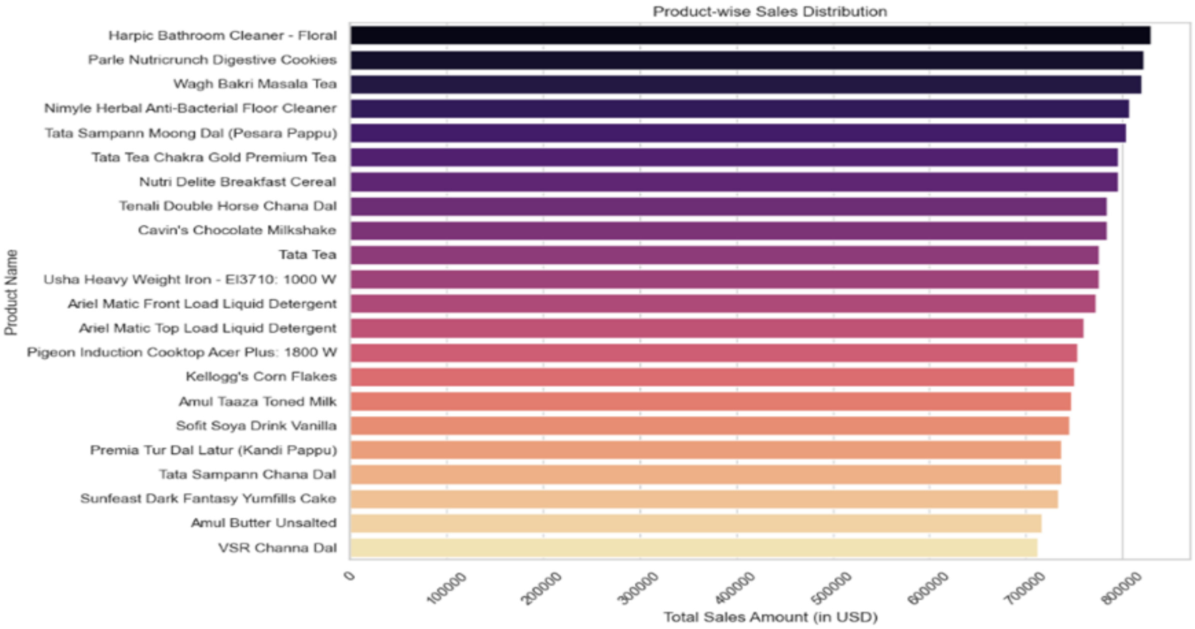
12. **Facet Grid**

The facet grid presents a grid of scatter plots, where each plot examines the relationship between total order value and number of clicks for a distinct combination of gender and category. This structured visualization allows for detailed exploration of how these variables interact within specific demographic and product categories. By segregating the data into individual facets, insights can be gleaned regarding any potential correlations, disparities, or trends unique to each gender-category pair, facilitating targeted analysis and informed decision-making in areas such as marketing strategies or customer segmentation.



13. **Horizontal Bar Plot**

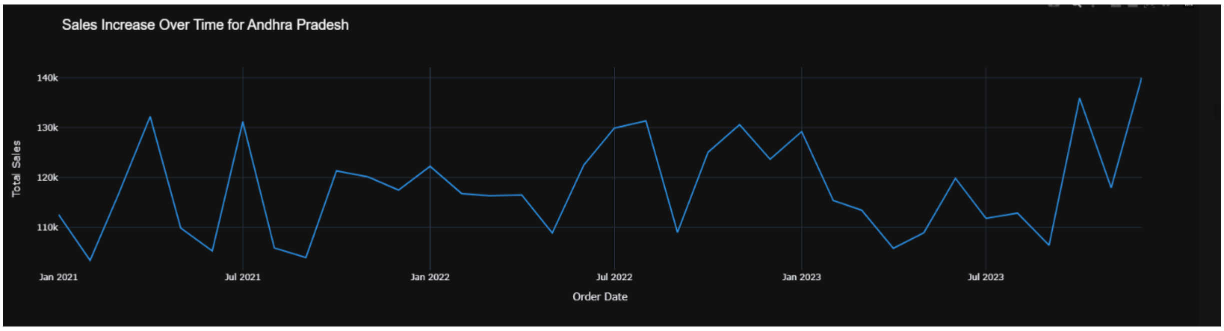
The horizontal bar plot visualizes the total sales amount attributed to each product, where the x-axis denotes the monetary values and the y-axis lists individual product names. Each bar represents the cumulative sales for a specific product, offering a clear comparison of revenue generation across different items. This visualization is effective for identifying top-selling products as well as outliers in sales performance, enabling strategic decisions related to inventory management, marketing strategies, and product development based on the revenue contributions of each product within the dataset

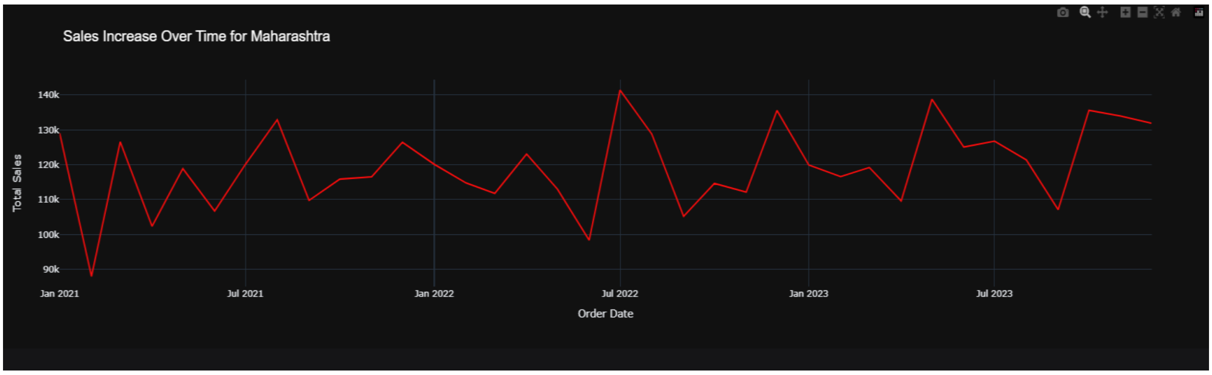


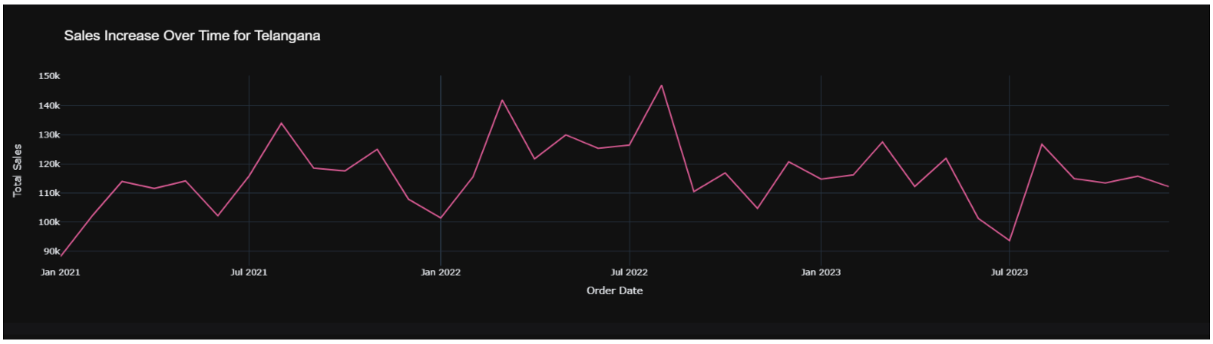
.

14. **Line Graph**

The line graph tracks monthly sales trends over time specifically for Andhra Pradesh, providing a detailed depiction of fluctuations and patterns in total sales. Each point on the graph represents sales figures for a specific month, offering a clear visual representation of how sales have evolved over the period analysed. This visualization enables the identification of seasonal variations, growth trends, or periods of decline in sales within Andhra Pradesh, aiding in strategic planning, resource allocation, and decision-making related to sales and market performance in that region.

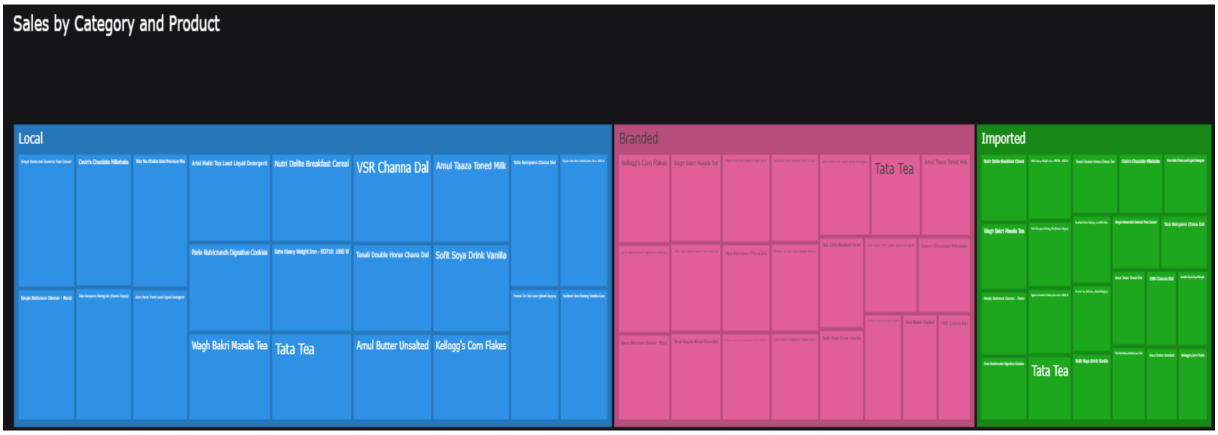






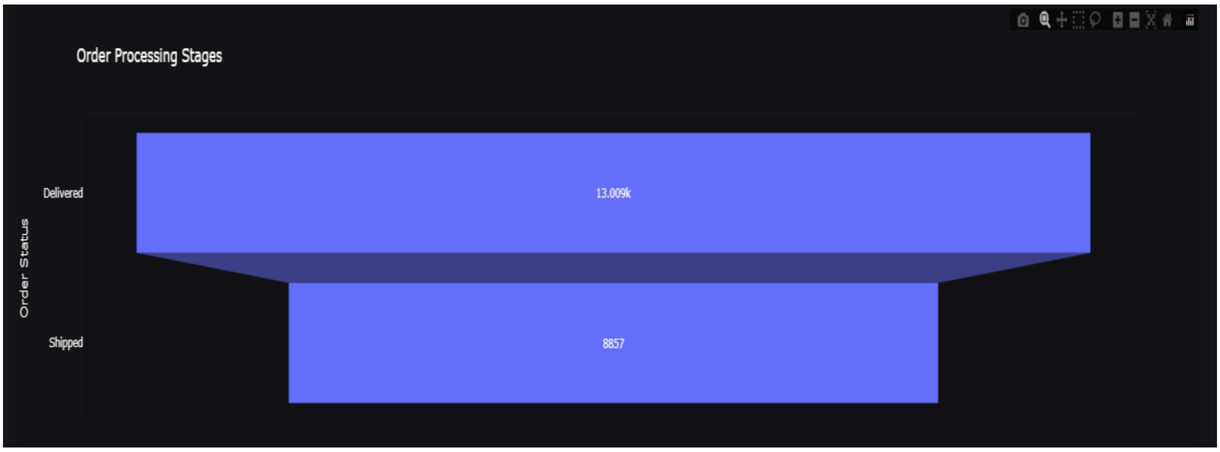
15. Tree map:

A tree map visually represents sales data by category and product, illustrating hierarchical relationships. Each rectangle's size corresponds to sales volume, with larger rectangles indicating higher sales. This visualization facilitates the quick identification of top-performing categories and products, supporting strategic decision-making and efficient inventory management in data analysis.

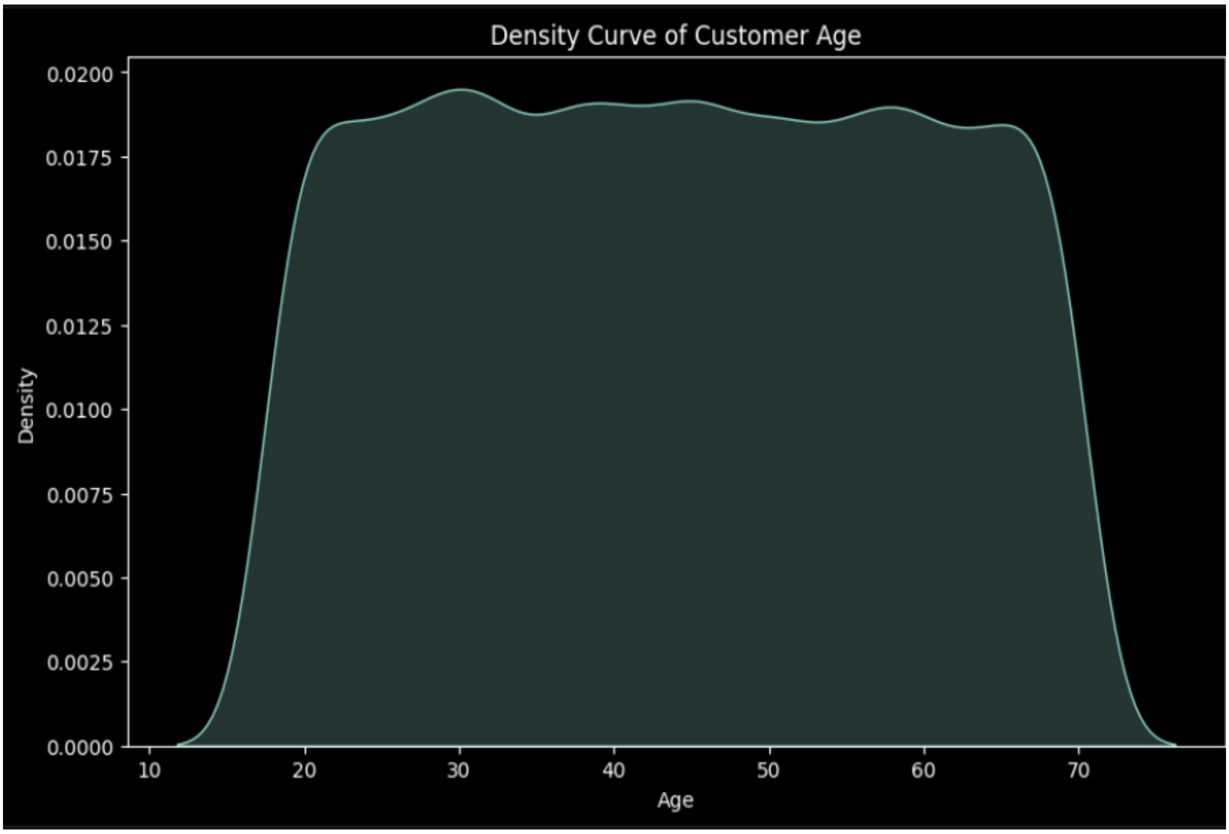


16. Funnel chart:

A funnel chart represents the customer journey from awareness to purchase, narrowing as potential customers progress through stages such as discovery, consideration, and decision. A tree map visually illustrates the relation of order processing stages, highlighting the proportion and hierarchy of each stage order placement, processing, shipping, and delivery within the overall process.



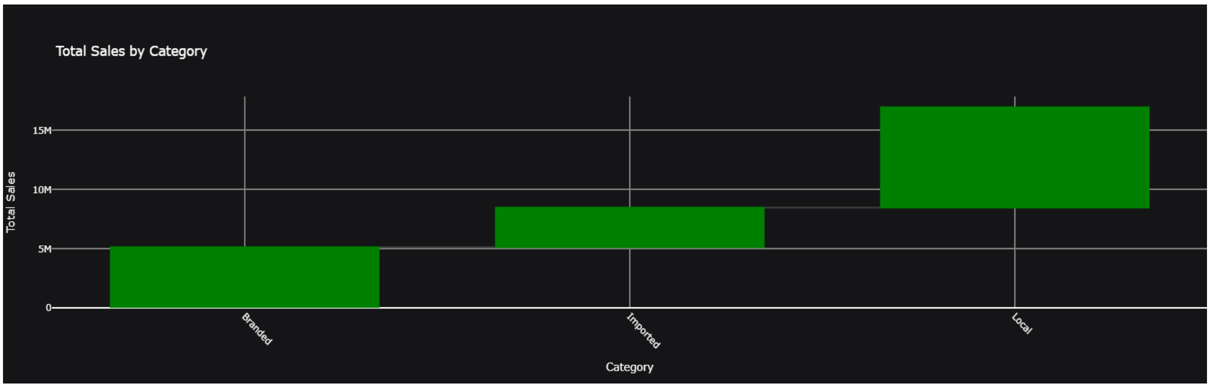
17. **Density Curve:**
A density curve represents the distribution of customer data, such as age, across a continuum. It shows the frequency of customers within different age groups, highlighting patterns and trends. A tree map visualizes the relationship between customer age and other variables, offering insights into age-based purchasing behaviours and demographics.



On the other hand, a tree map is a visualization tool that displays the relationship between customer age and other variables in a hierarchical manner. Each rectangle in the tree map represents a category, such as age group, with the size and colour of the rectangles providing additional insights. For instance, a tree map can reveal age-based purchasing behaviours, showing which age groups are more likely to buy certain products. By combining density curves and tree maps, businesses can gain comprehensive insights into customer demographics and purchasing patterns, enabling more effective marketing strategies and product offerings.

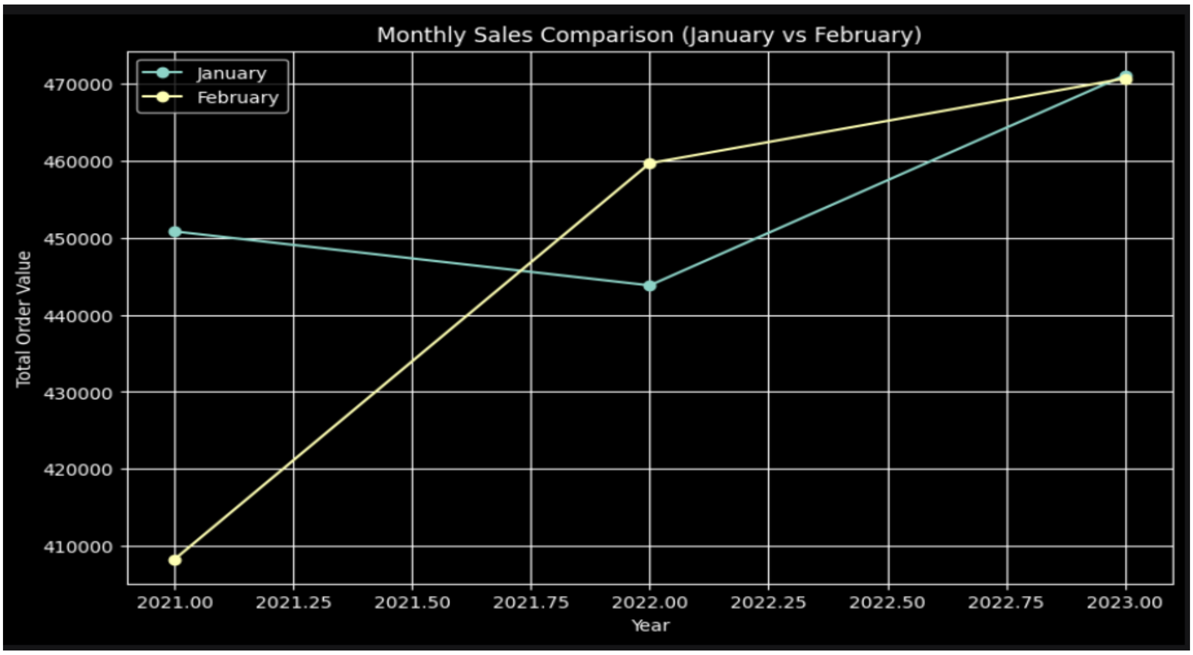
18. Waterfall chart:

A waterfall chart visually tracks total sales changes across categories, demonstrating cumulative impacts of positive and negative contributions. Python libraries such as Matplotlib and Plotly are utilized to generate these charts, portraying how product categories influence overall sales. This visualization aids in understanding revenue distribution, identifying key contributors, and evaluating trends in sales performance.



19. Slope graph:

A slope graph visually illustrates changes in monthly sales comparisons across categories or products. It effectively showcases trends, highlighting sales increases or decreases over time. When combined with a treemap a visualization depicting sales proportions by area size—it offers a clear, comparative analysis of sales performance among different categories or products across multiple months.

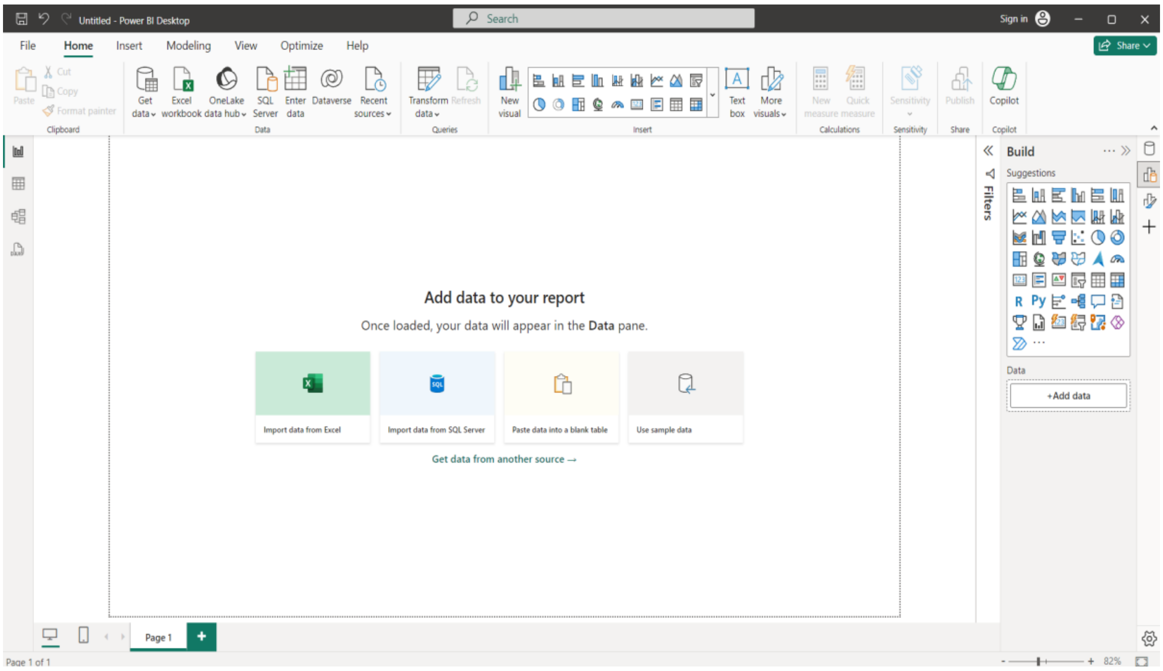


4.2 Data visualization using Power BI

Introduction to Power BI

Interface Overview:

- Ribbon: Contains various tabs and commands for data connection, modelling, and visualization.
- Report View: The main canvas where you create and arrange your visualizations.
- Data View: Allows you to view and manage the data tables loaded into Power BI.
- Model View: Helps you define relationships between different data tables.
- Visualization Pane: Contains various visual elements that you can drag onto the report canvas.
- Fields Pane: Lists all the data fields available for creating visualizations.
- Filters Pane: Enables you to apply filters at the report, page, or visual level.
- Connecting to Data Sources: Step-by-step instructions on how to connect Power BI to different data sources such as Excel, SQL Server, Azure, and online services.
- Building Queries and Data Models: Demonstration of using Power Query Editor to transform and shape data, and creating relationships and calculations in the data model.



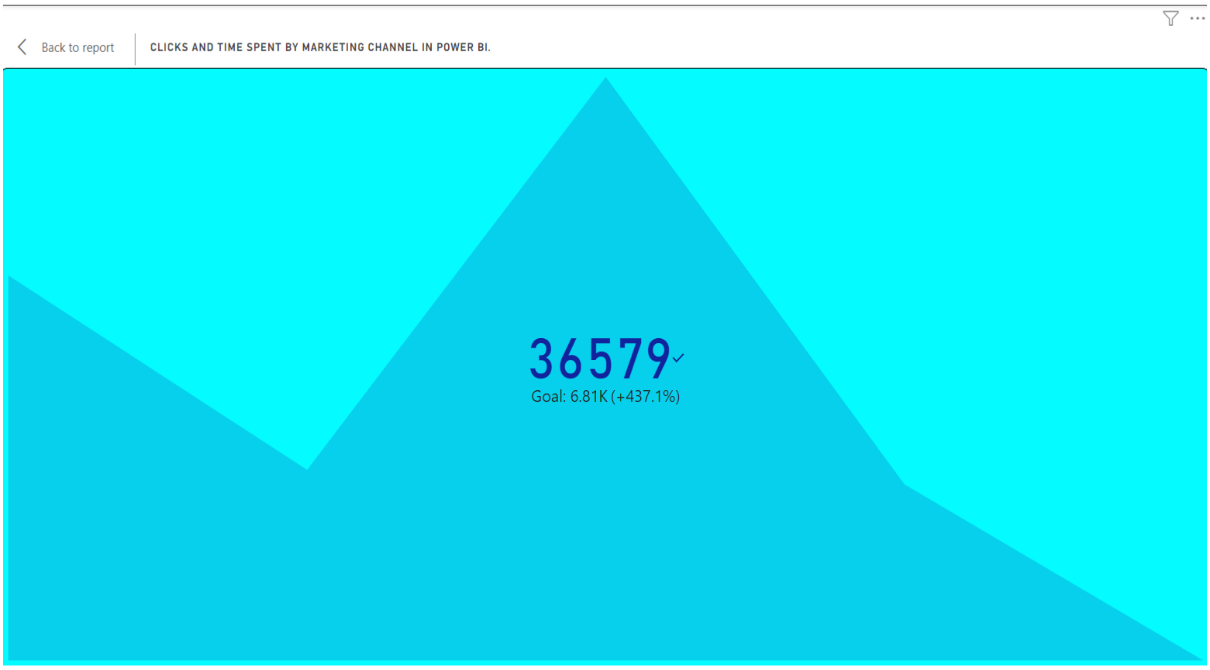
Add data to your report

Once loaded, your data will appear in the **Data** pane.

Power BI Visualizations:

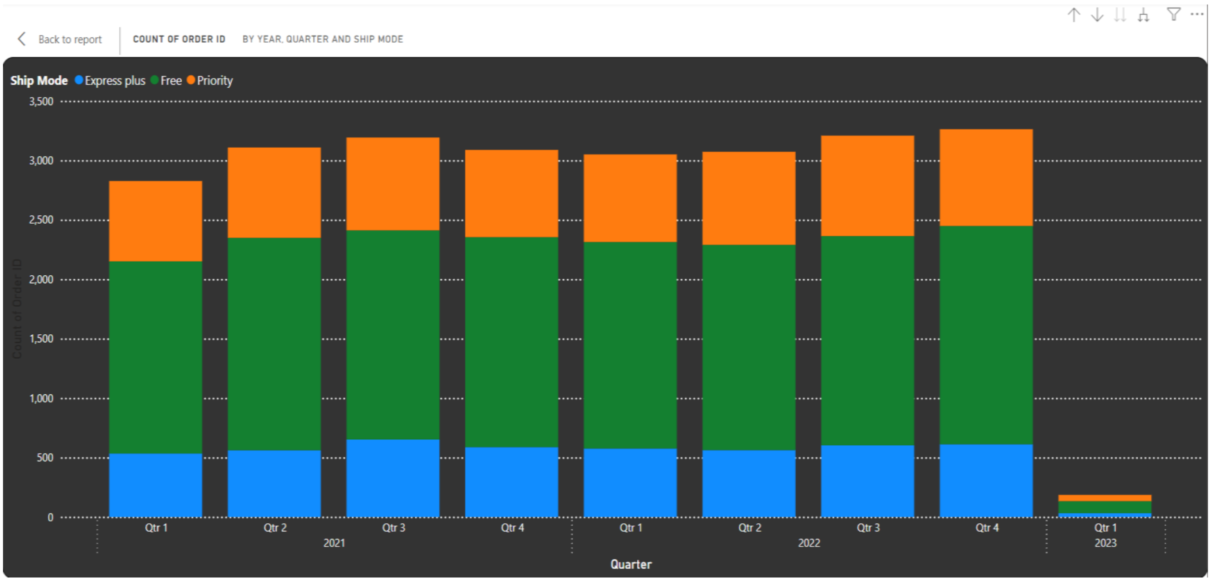
1 Card

A Card visual displays a single value, such as a total, average, or other summary measure. It is ideal for highlighting key performance indicators (KPIs) or important metrics at a glance. Use cards to emphasize critical data points like total revenue, number of orders, or average sales price.



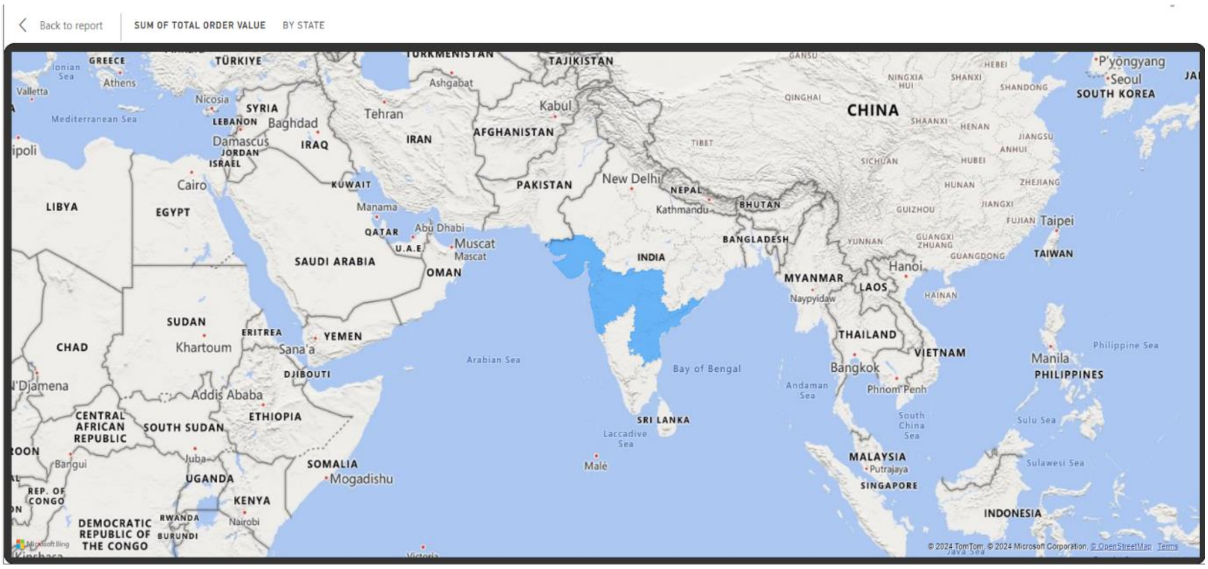
2 Stacked Column Chart

A Stacked Column Chart visualizes data with rectangular bars stacked on top of each other, representing different categories. It is useful for showing the cumulative contribution of each category over a period. For example, you can use it to display quarterly sales by product category, highlighting both individual and total sales.



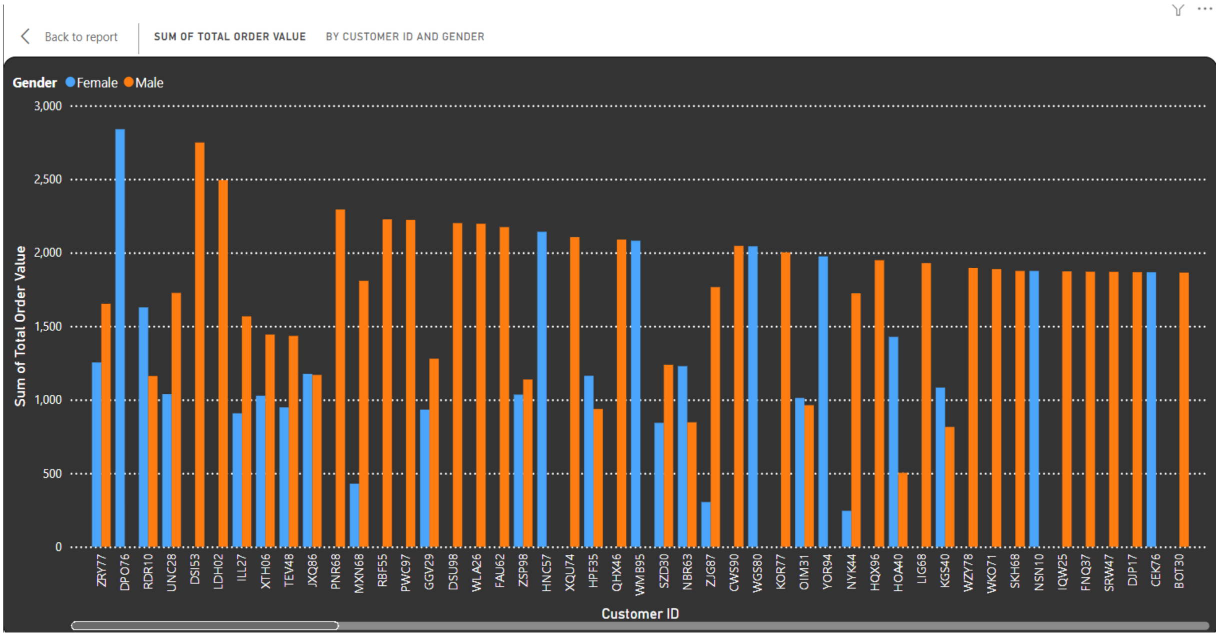
3 Filled Map

A Filled Map uses colour shading to represent data distribution across geographical regions. It is useful for showing regional performance metrics, such as sales volume or customer count by state or country. Filled maps provide a clear, visual way to identify trends and patterns based on location.



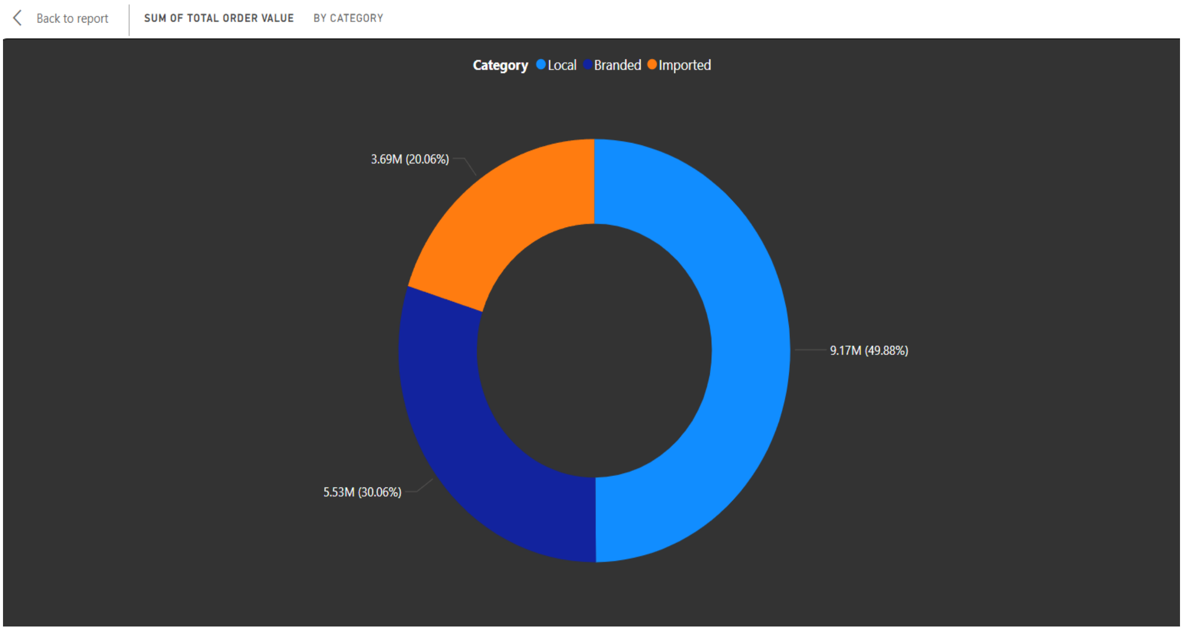
4 Line and Clustered Column Chart

Line and Clustered Column Chart combines a line graph and a clustered column chart in one visual. It is ideal for comparing two related metrics with different scales, such as sales volume (columns) and profit margin (line) over time, enabling a comprehensive analysis of trends and relationships.



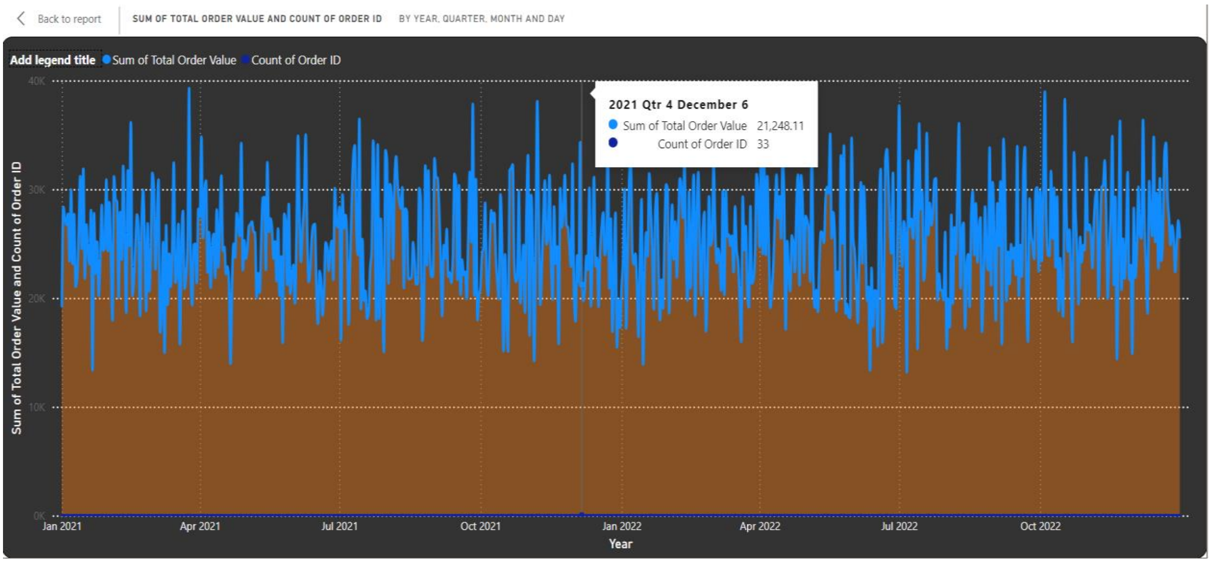
5 Donut Chart

A Donut Chart is similar to a pie chart but with a central hole. It displays proportions of a whole, making it easier to read and interpret. Use donut charts to show percentage breakdowns of categories, such as market share by product or customer distribution by segment.



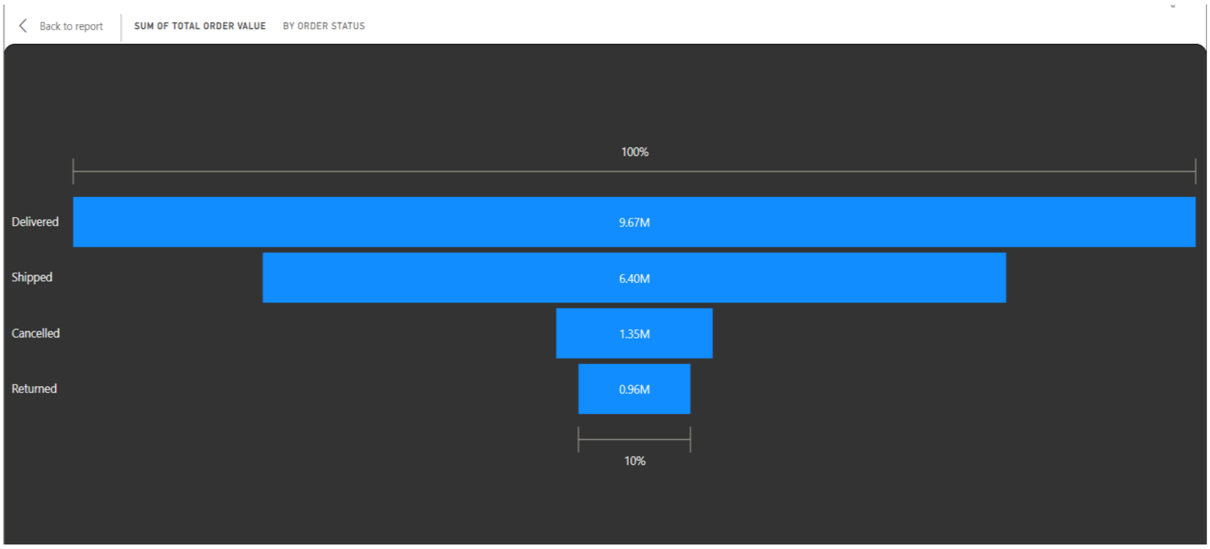
6 Line Chart

A Line Chart displays data points connected by straight lines, effectively showing trends over time. It is ideal for visualizing time-series data, such as monthly revenue or daily website traffic, helping to identify patterns, trends, and fluctuations over a specific period.



7 Funnel Chart:

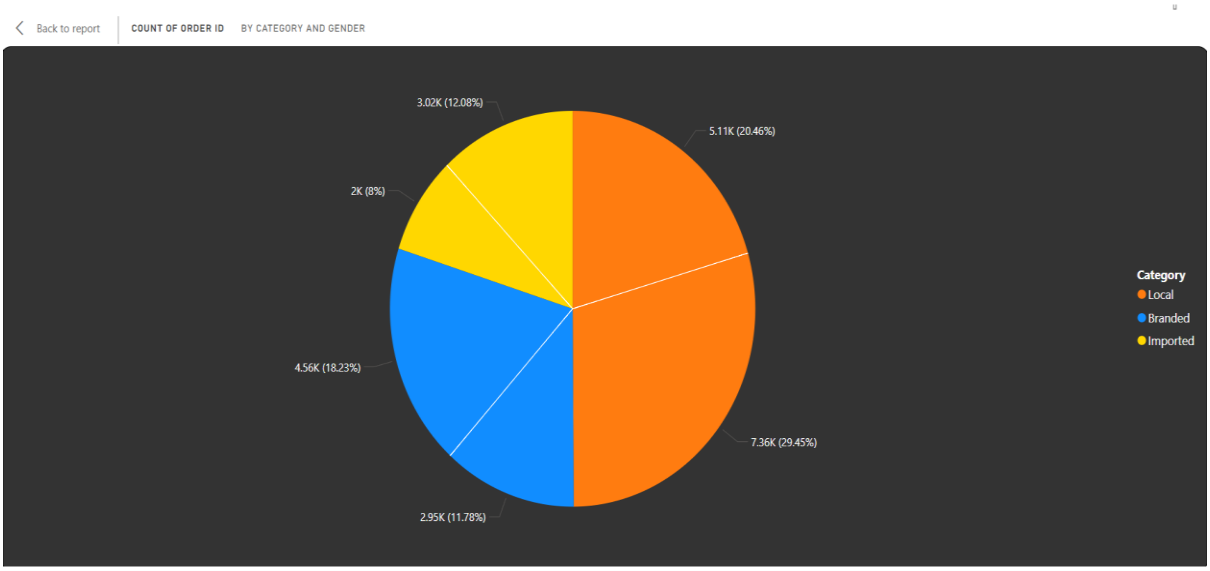
A Funnel Chart visualizes stages in a linear process, highlighting drop-off points and conversion rates between stages. It is commonly used in sales and marketing to track the customer journey from lead generation to sales conversion, identifying areas for improvement.



8 Pie Chart

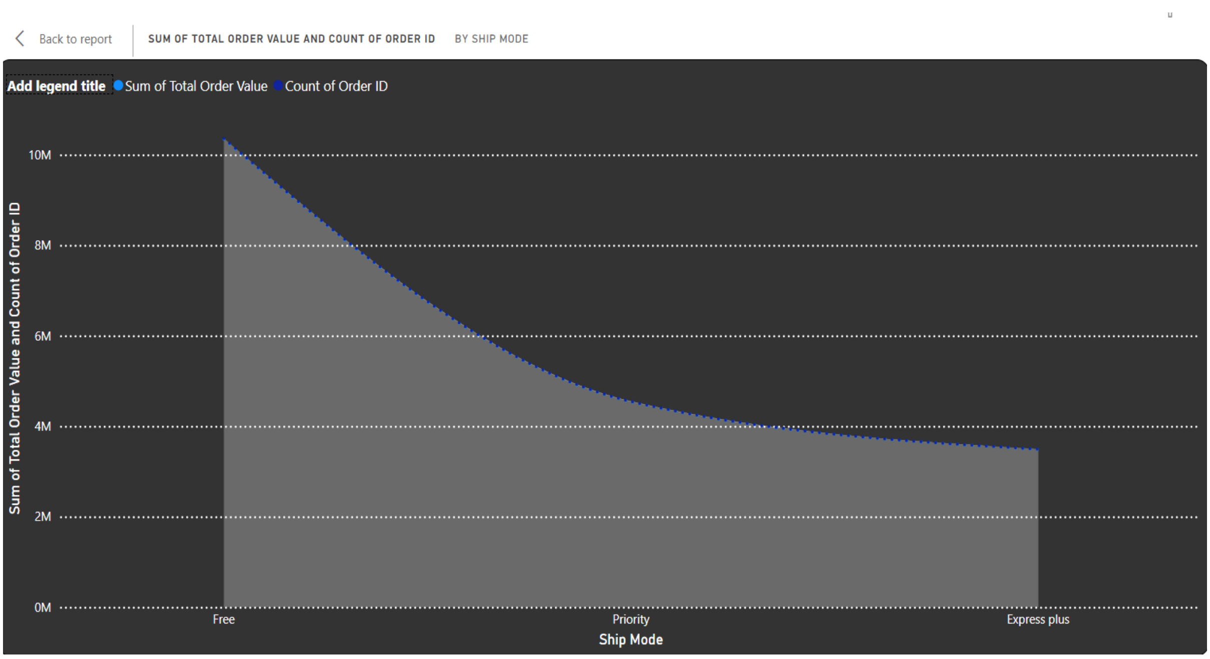
A Pie Chart represents data as slices of a circle, showing proportions of a whole. It is useful for visualizing percentage distribution among categories, such as market share by product or budget allocation by department. Each slice's size reflects its relative contribution to the total.

A pie chart showing the distribution of products by category. Notable products include Parle Nutr... (4.85%), Harpic Bat... (4.37%), and Wagh Bak... (4.33%).



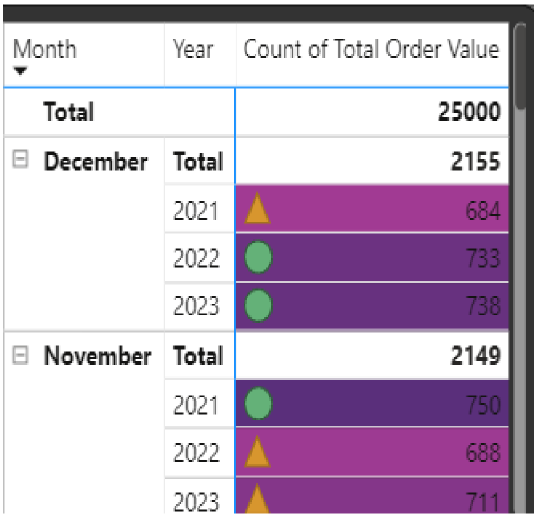
9 Stacked Area Chart

A **Stacked Area Chart** displays cumulative totals over time with shaded areas, representing different categories stacked on top of each other. It is useful for showing the composition of data over a period, such as sales by product category, and how each category contributes to the total.



10 Matrix

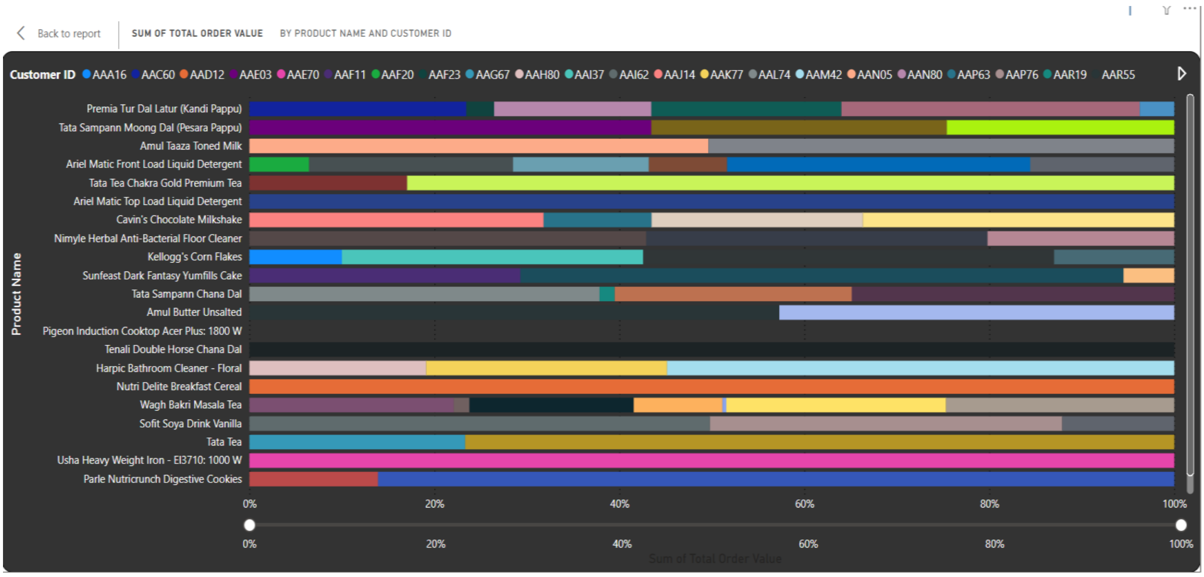
A Matrix visual is a table with rows and columns that allows for hierarchical data grouping and drill-down capabilities. It is ideal for presenting detailed, multi-dimensional data, such as sales by region, product, and time period, enabling a comprehensive analysis.



Month	Year	Count of Total Order Value
Total		25000
⊟ December	Total	2155
	2021	▲ 684
	2022	● 733
	2023	● 738
⊟ November	Total	2149
	2021	● 750
	2022	▲ 688
	2023	▲ 711

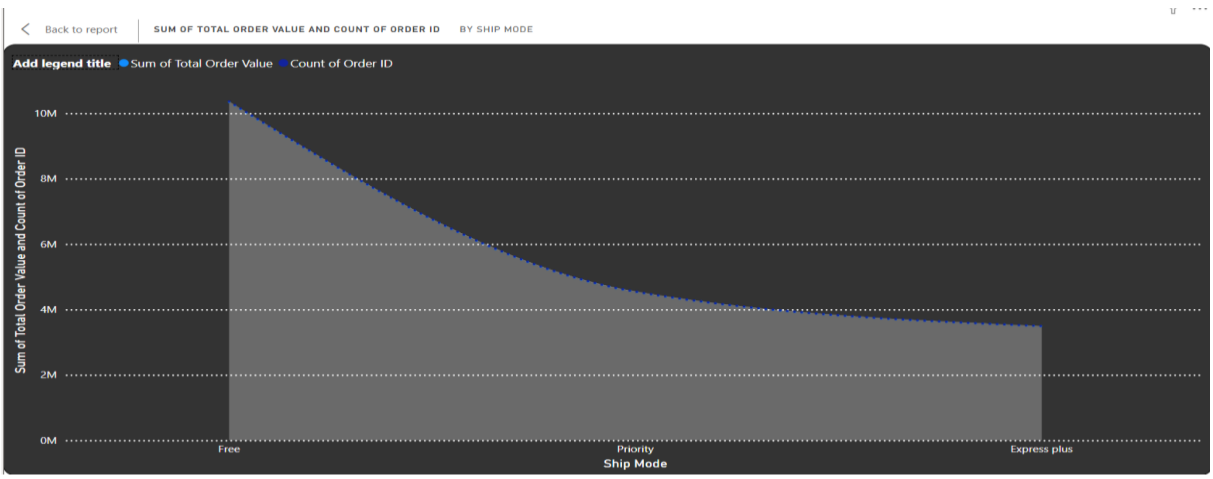
11 100% Stacked Bar Chart

A 100% Stacked Bar Chart displays the relative percentage contribution of categories to the whole, with each bar representing 100%. It is useful for comparing the proportional makeup of categories across different groups, such as market share by product in different regions.



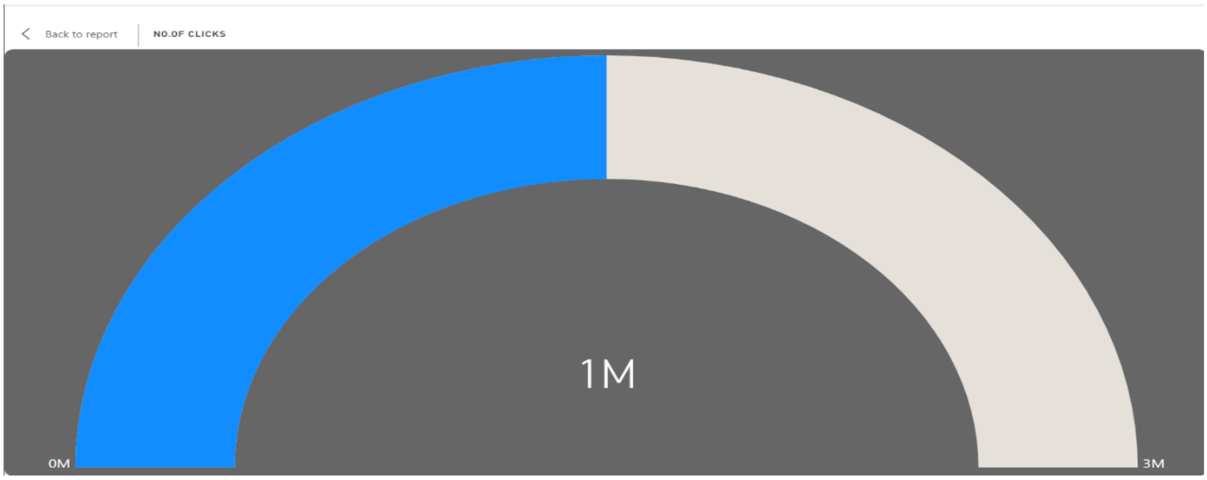
12 Area Chart

An Area Chart is similar to a line chart but with the area below the line filled in. It emphasizes the magnitude of change over time and is useful for showing cumulative data, such as revenue growth or stock prices, highlighting the overall trend and volume.



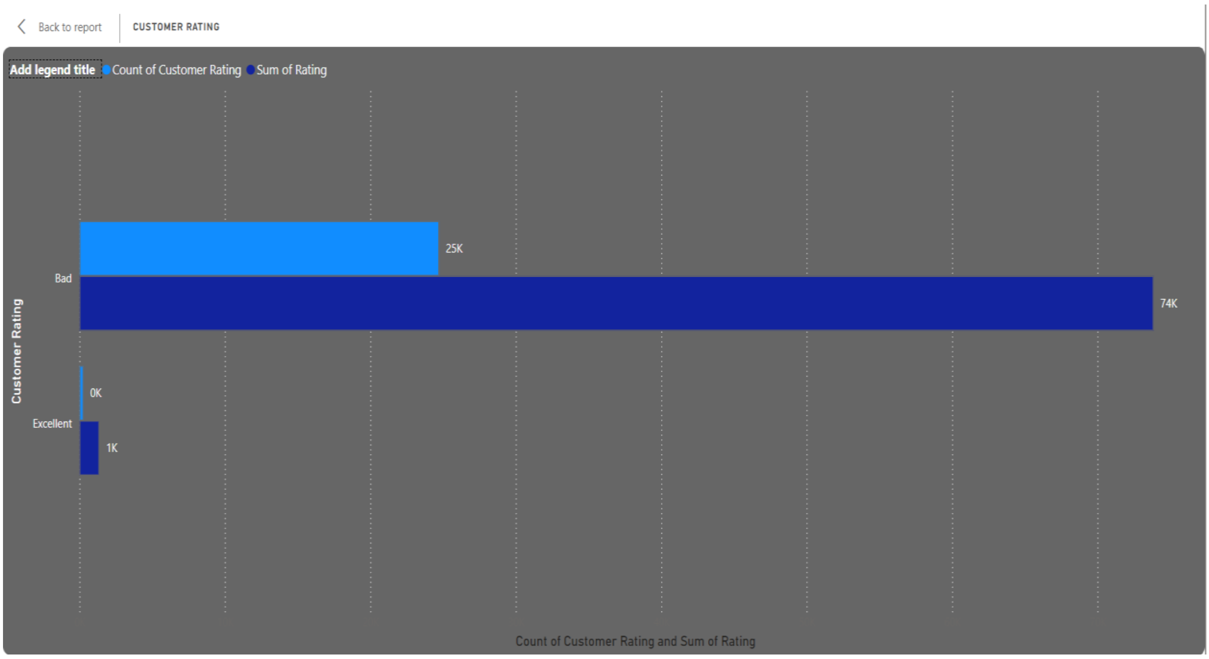
13 Gauge Chart

A Gauge Chart displays a single measure against a range of values, typically using a speedometer-like needle. It is useful for showing progress towards a goal, such as sales target achievement or customer satisfaction score, providing a quick visual status check.



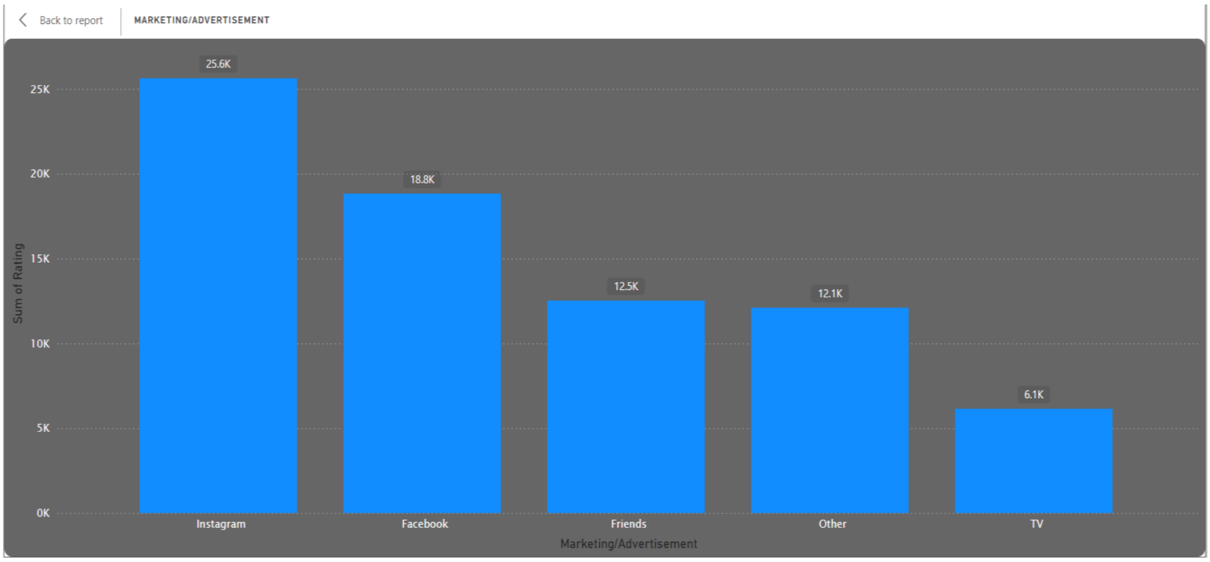
14 Clustered Column Chart

A Clustered Column Chart displays multiple series of data with vertical bars grouped side-by-side. It is useful for comparing multiple categories or groups, such as sales by product and region, showing how different series relate to each other within the same axis.



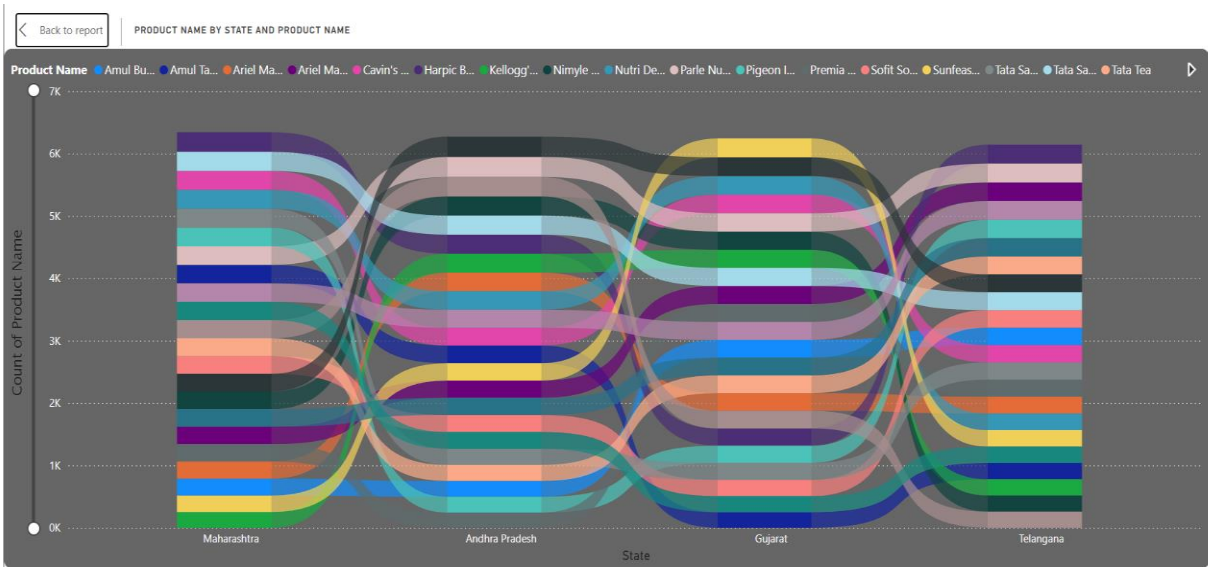
15 Clustered Bar Chart

A Clustered Bar Chart is similar to a clustered column chart but with horizontal bars. It is useful for comparing multiple categories side-by-side, such as revenue by department and region, highlighting differences and similarities across groups.



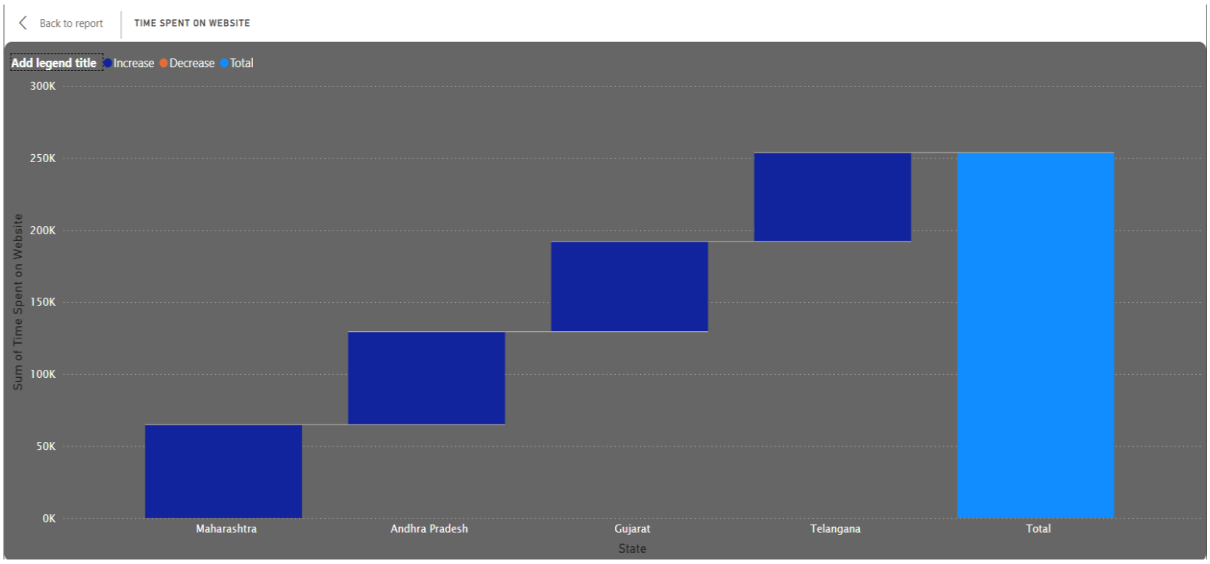
16 Ribbon Chart

A Ribbon Chart shows data over time with ribbons that flow and shift positions, representing changes in rank or value. It is useful for visualizing how categories compare to each other over a period, such as sales rankings of products or market position changes.



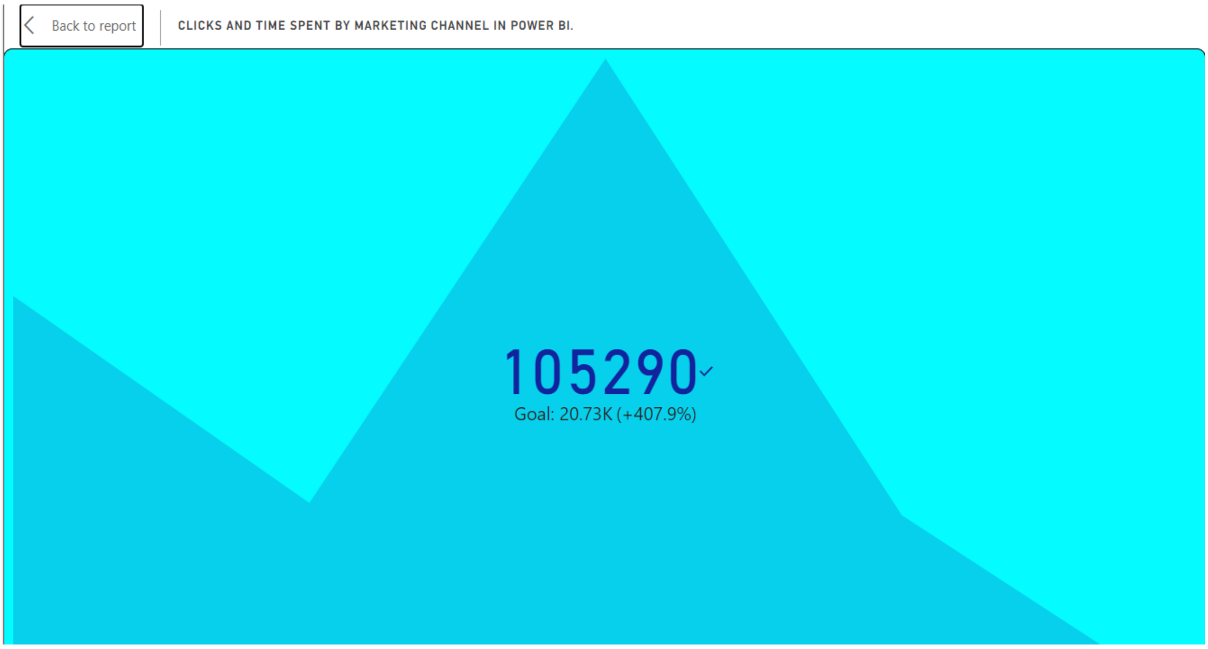
17 Waterfall Chart

A Waterfall Chart visualizes incremental changes to a cumulative total, showing how individual contributions add up to a final result. It is useful for financial analysis, such as tracking the impact of revenue and expenses on net profit, highlighting increases and decreases.



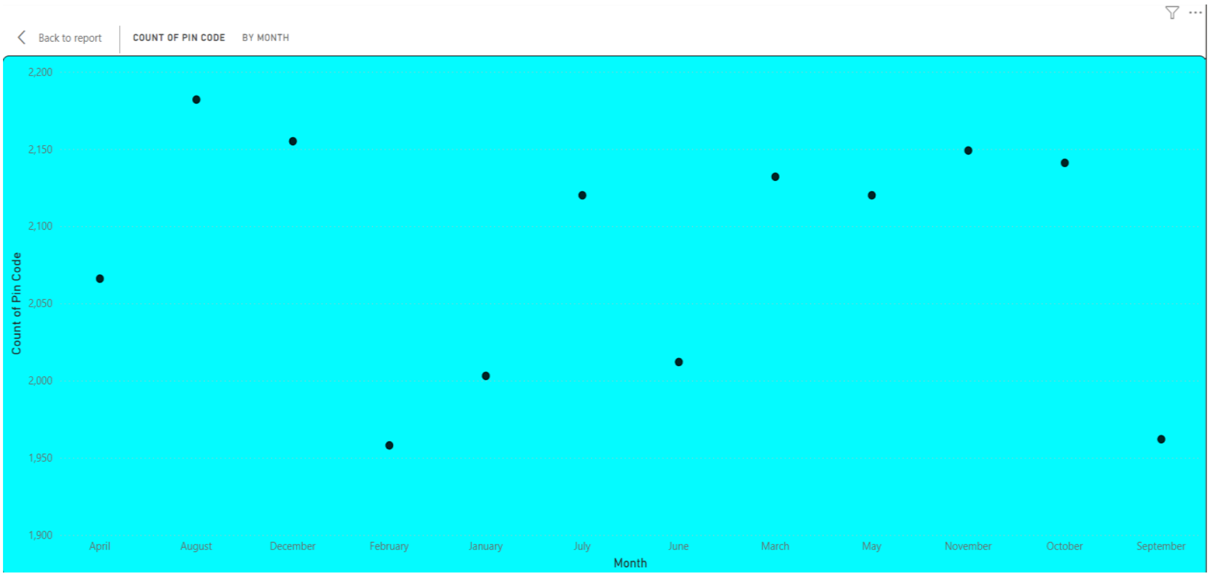
18 KPI

A KPI (Key Performance Indicator) visual displays progress towards a measurable goal. It typically shows the current value, target, and status, such as sales target achievement or customer satisfaction levels, providing a quick overview of performance metrics.



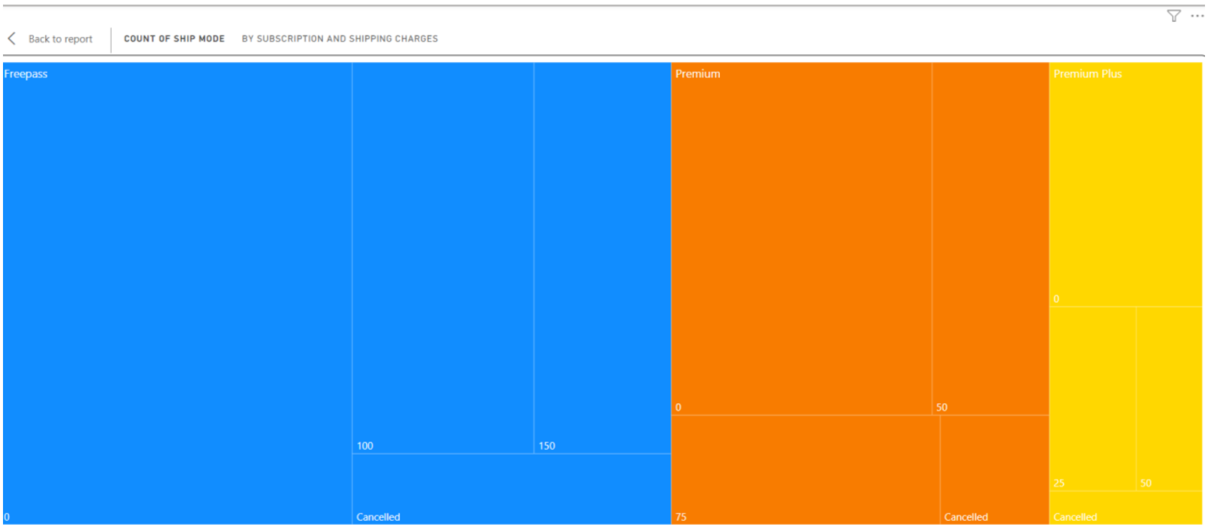
19 Scatter Plot

A Scatter Plot displays individual data points on a Cartesian plane, showing the relationship between two variables. It is useful for identifying correlations, trends, and outliers, such as the relationship between advertising spend and sales revenue.



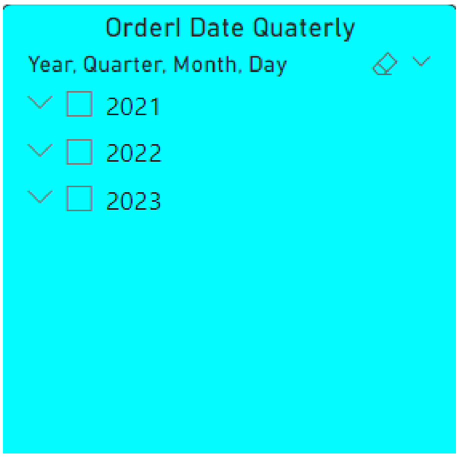
20 Tree map

A Tree map visualizes hierarchical data with nested rectangles, where the size and colour of each rectangle represent different dimensions. It is useful for displaying the proportion of categories within a whole, such as sales by product and sub-product, highlighting relative sizes



21 Slicer

A Slicer is an interactive filtering tool that allows users to segment and view specific portions of data. It is useful for dynamically filtering reports and visuals by criteria such as date, category, or region, providing a more focused and customized analysis.



Orderl Date Quaterly

Year, Quarter, Month, Day

□ 2021

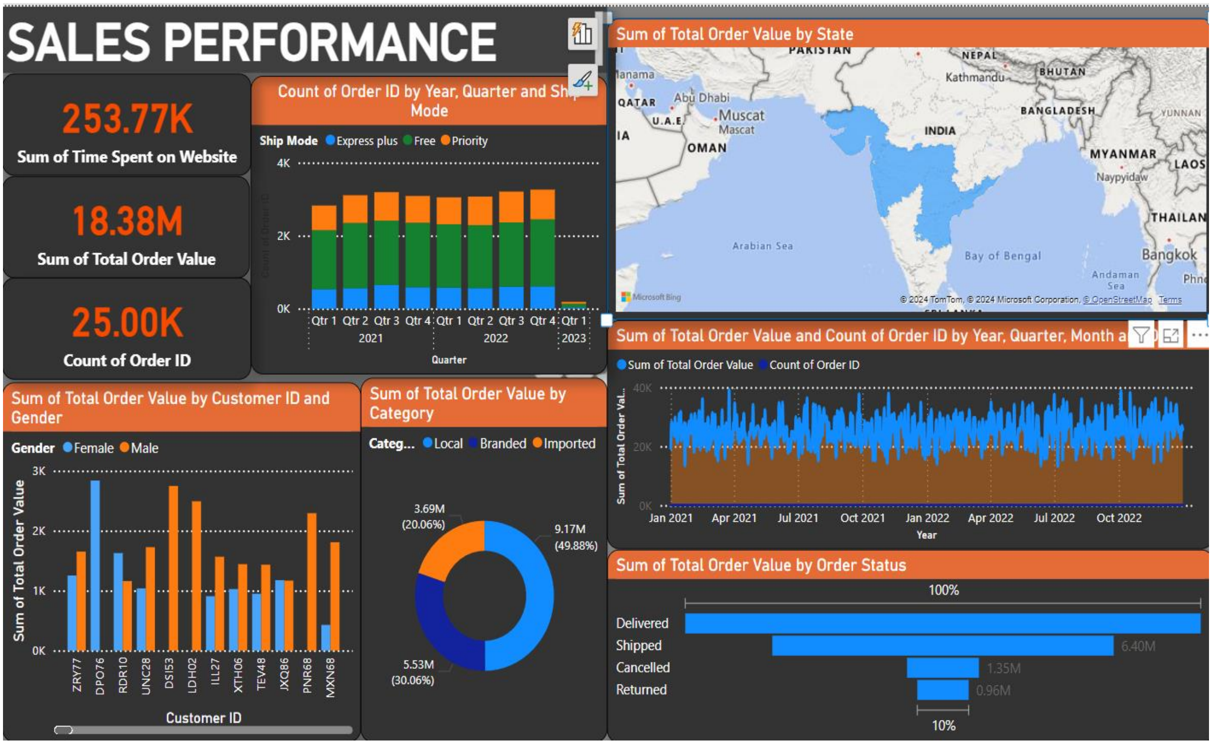
□ 2022

□ 2023

22 Multi-row Card

A Multi-row Card displays multiple data points or metrics in a compact, card-like format. It is useful for presenting key statistics or summaries, such as sales figures, customer counts, and profit margins, enabling quick and easy comparison of multiple values.

Sales Performance Dashboard:


SALES PERFORMANCE
253.77K
Sum of Time Spent on Website
18.38M
Sum of Total Order Value
25.00K
Count of Order ID
Count of Order ID by Year, Quarter and Ship Mode
Ship Mode ● Express plus ● Free ● Priority
4K
2K
0K
Qtr 1 Qtr 2 Qtr 3 Qtr 4 Qtr 1 Qtr 2 Qtr 3 Qtr 4 Qtr 1
2021
2022
2023
Quarter
Sum of Total Order Value by State
PAKISTAN
NEPAL
Kathmandu
BHUTAN
QATAR
Abu Dhabi
U.A.E.
Muscat
Mascat
OMAN
BANGLADESH
INDIA
YUNNAN
MYANMAR
Naypyidaw
LAOS
THAILAN
Arabian Sea
Bay of Bengal
Andaman Sea
Bangkok
Microsoft Bing
© 2024 TomTom, © 2024 Microsoft Corporation, © OpenStreetMap Terms
Sum of Total Order Value and Count of Order ID by Year, Quarter, Month a
● Sum of Total Order Value ● Count of Order ID
Sum of Total Order Val...
40K
20K
0K
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Jan 2022
Apr 2022
Jul 2022
Oct 2022
Year
Sum of Total Order Value by Customer ID and Gender
Gender ● Female ● Male
Sum of Total Order Value
3K
2K
1K
0K
ZRY77
DPO76
RDR10
UNC28
DSI53
LDH02
ILL27
XTH06
TEV48
JXQ86
PNR68
MXN68
Customer ID
Sum of Total Order Value by Category
Categ... ● Local ● Branded ● Imported
3.69M (20.06%)
9.17M (49.88%)
5.53M (30.06%)
Sum of Total Order Value by Order Status
100%
Delivered
Shipped
Cancelled
Returned
6.40M
1.35M
0.96M
10%

Product Performance

The "Product Performance" is a comprehensive and interactive visualization tool designed to provide a detailed overview of various performance metrics for a range of products. The dashboard is divided into several sections, each offering unique insights into key aspects of product performance and customer engagement.

Key Metrics:

- Total sum of sales: 16.95 million (17M).
- Category by Product Name: Donut chart.
- Order Count by Time and Ship Mode: Stacked column chart.
- Sum of Total Order Value based on States: Filled map chart.

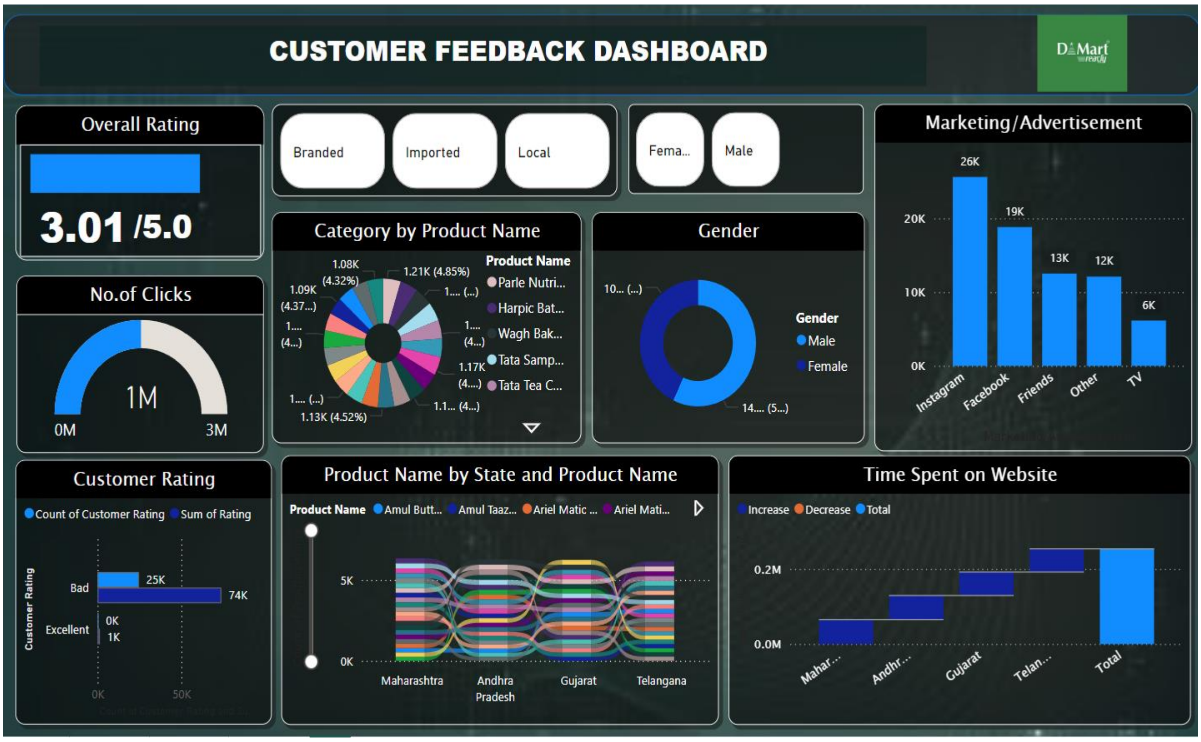
Detailed Visualizations

The main section of the dashboard features various detailed visualizations that offer deeper insights into specific performance aspects:

- **Gender Distribution**
  - A donut chart displaying the gender distribution of customers, with a majority being Male.
- **Marketing/Advertisement Channels**
  - A bar chart illustrating the effectiveness of different marketing channels. Instagram leads with 26K interactions, followed by Facebook (19K) and Friends (13K).
- **Customer Rating**
  - A bar chart showing the count and sum of customer ratings. There are 25K ratings classified as 'Bad' and 1K ratings classified as 'Excellent'.
- **Product Name by State and Product Name**
  - A parallel coordinates chart detailing product performance across different states. The chart highlights popular products like Amul Butt... and Amul Taaz... across Maharashtra, Andhra Pradesh, Gujarat, and Telangana.
- **Time Spent on Website**
  - A stacked bar chart representing the time spent on the website by state, with metrics for increase, decrease, and total time. The total time spent is highest, with notable time spent in Maharashtra and Telangana

Interactive Elements

Interactive elements such as filters for Branded, Imported, Local products, and Gender (Male, Female) allow users to customize their view and focus on specific data points of interest. This enhances the usability of the dashboard by enabling targeted analysis.



Conclusion

In conclusion, the "Product Performance Dashboard" offers a comprehensive and detailed view of product performance and customer engagement. It empowers stakeholders to make informed decisions by analysing critical metrics such as overall rating, number of clicks, customer demographics, and marketing effectiveness. Through its interactive and visually engaging design, the dashboard serves as a powerful tool for optimizing marketing strategies and improving product performance.

You can customize this summary with specific product names, metrics, and descriptions as needed. This structure ensures that you cover all key aspects and provide a detailed overview of your Power BI dashboard.

Advanced Power BI Features

DAX (Data Analysis Expressions) for Calculations

To create a comprehensive summary that includes an overview of DAX (Data Analysis Expressions) used in Power BI dashboard, we considered different DAX functions and calculations that are involved.

The dashboard utilized several DAX expressions to calculate and present key metrics.

1) **Overall Rating:**

   Overall Rating = AVERAGE (Reviews [Rating])

   This formula calculates the average rating from the Reviews table.

2) **Number of Clicks:**

   Total Clicks = SUM (Clicks [Count])

   This formula sums the click counts from the Clicks table.

3) **Customer Rating Count:**

   Count of Ratings = COUNT (Ratings [CustomerID])

   This formula counts the number of customer ratings.

4) **Marketing/Advertisement Interactions:**

   Total Interactions = SUM (Marketing [Interactions])

   This formula calculates the total interactions from different marketing channels.

5) **Time Spent on Website:**

   Time Spent = SUM(WebsiteData[TimeSpent])

   This formula sums the total time spent on the website.

Challenges:

Performance Issues:

- **Execution Speed:** Python scripts may run slower compared to native Power BI functions, especially with large datasets.
- **Resource Intensive:** Python can be resource-intensive, potentially affecting the performance of Power BI dashboards.

Integration Complexity:

- **Setup and Configuration:** Integrating Python with Power BI requires proper installation and configuration of Python on the local machine.
- **Library Management:** Ensuring that all necessary Python libraries are installed and compatible with Power BI can be cumbersome.

Security Concerns:

- **Execution of Scripts:** Running Python scripts can pose security risks, especially when dealing with sensitive or confidential data.
- **Access Permissions:** Proper permissions need to be managed to allow the execution of Python scripts within Power BI.

Maintenance and Debugging:

- **Script Maintenance:** Maintaining and updating Python scripts within Power BI can be more complex compared to using Power BI’s native functions.
- **Debugging:** Debugging Python scripts in Power BI is less straightforward, often requiring external tools or environments.

Version Compatibility:

- **Power BI Updates:** Updates to Power BI or Python can cause compatibility issues, requiring adjustments to existing scripts.
- **Library Updates:** Python libraries frequently update, which can lead to compatibility issues or require code changes.

User Expertise:

- **Learning Curve:** Users need to have knowledge of both Python programming and Power BI, which may require additional training.
- **Best Practices:** Ensuring that scripts follow best practices for performance and efficiency can be challenging, particularly for users who are new to Python.

Environment Consistency:

- **Local vs. Service:** Python scripts run in different environments when used in Power BI Desktop versus Power BI Service, potentially leading to inconsistencies.
- **Dependency Management:** Ensuring that all dependencies are consistently managed across different environments can be difficult.

Key Future Trends in Data Analytics and Visualization

1. AI and Machine Learning Integration
    - Predictive Analytics: Leveraging AI for forecasting future trends.
    - Automated Insights: AI-driven insights and narratives from data.
2. Real-Time Data Analytics
    - Streaming Data: Instant insights from real-time data streams.
    - Edge Analytics: Processing data at the edge for faster decision-making.
3. Advanced Visualization Techniques
    - Immersive Analytics: Utilizing VR and AR for enhanced data visualization.
    - 3D Visualizations: Improved spatial representation of complex data sets.
4. Self-Service and Augmented Analytics
    - Self-Service Tools: Enabling non-technical users to perform data analysis.
    - Augmented Discovery: AI recommendations to uncover hidden patterns.
5. Ethics and Data Governance
    - Data Privacy: Protecting personal data and ensuring compliance.
    - Ethical AI: Developing transparent and fair AI models.

Chapter 5

Challenges:

1. Integrating Python with PowerBI will be applied only with POWERBI service.
2. Due to large number of job titles, the entire data is not fitting into an entire graph. Because of the large numbers of job title, we are unable to display the some of the insights accurately.
3. Extracting meaningful features (variables) from raw data, particularly job titles containing unstructured text, is crucial for analysis but can be complex.

Chapter 6

Glossaries:

1. Heat Map:A graphical representation of data where individual values are represented by colors. It is used to visualize the distribution of customer ages by gender, with color intensity indicating the number of customers in each age group.

2. 3D Scatter Plot: A plot that visualizes data points in a three-dimensional space, representing customer age, time spent on the website, and number of clicks. It provides a comprehensive view of customer interaction metrics.

3. Violin Plot: A method of plotting numeric data that shows the density of the data at different values. It combines aspects of box plots and kernel density plots and is used to compare the distribution of total order values between genders.

4. Facet Grid: A grid of scatter plots that allows for detailed exploration of the relationship between different variables across multiple subsets of data. It helps in examining how variables interact within specific demographic and product categories.

5. Ribbon (Power BI): A component of the Power BI interface that contains various tabs and commands for data connection, modelling, and visualization.

6. Data View (Power BI): Part of the Power BI interface where users can view and manage the data tables loaded into the application.

7. Model View (Power BI): A feature in Power BI that helps define relationships between different data tables.

8. Parallel Coordinates Chart: A type of chart used to visualize multivariate data by plotting multiple variables as lines on parallel axes. It is used to detail product performance across different states in the Product Performance Dashboard.

9.Holistic Retail Management: This involves the integration of various aspects of retail operations such as supply chain coordination, inventory control, and customer service to ensure seamless operations and a consistent customer experience.

10. Customer Data Analysis: Analysing customer data to gain insights into customer behaviour, preferences, and trends. This analysis helps in personalized marketing, improving customer service, and developing products that meet customer needs.

11.Supply Chain Coordination: The process of managing and synchronizing the flow of goods, information, and finances as they move from supplier to manufacturer to wholesaler to retailer to consumer. Effective supply chain coordination enhances efficiency and reduces costs.

12.Inventory Control: Techniques and strategies used to manage inventory levels, track inventory movements, and ensure the right amount of stock is available at the right time. This helps in minimizing costs and meeting customer demand efficiently.

13.Data Analytics: The process of examining datasets to draw conclusions using specialized systems and software. It includes techniques ranging from simple statistical analysis to complex data mining and predictive analytics.

14. Data Visualization: The graphical representation of information and data using visual elements like charts, graphs, and maps. It makes trends, outliers, and patterns in data more accessible and understandable.

15.Customer Lifetime Value (CLV) Analysis: Analyzing historical purchase data and customer interactions to estimate the future value each customer brings to the company. It helps in identifying high-value customers and optimizing retention strategies.

16.Sentiment Analysis: Analyzing customer feedback, reviews, and social media interactions to gain insights into customer satisfaction and brand perception. It helps in identifying areas for improvement and enhancing the customer experience.

17.Descriptive Statistics: The analysis of data that helps describe, show, or summarize data in a meaningful way such as calculating the mean, median, standard deviation, minimum, maximum, and quartiles of numerical attributes.

18. DAX (Data Analysis Expressions): A collection of functions, operators, and constants used in Power BI to perform advanced calculations on data in the model. It is used for creating custom metrics and aggregations in Power BI dashboards.

Chapter 7

CONCLUSION:

The "Product Performance Dashboard" provides a robust and detailed analysis of product performance and customer engagement. By leveraging advanced data visualization techniques such as heat maps, 3D scatters plots, and violin plots, it offers stakeholders critical insights into customer behaviour and product metrics. Key performance 0078indicators like overall rating, number of clicks, customer demographics, and marketing effectiveness are clearly illustrated, enabling informed decision-making. Interactive elements, including filters for branded, imported, and local products, as well as gender-specific analysis, enhance the usability of the dashboard, making it a powerful tool for targeted analysis and strategy optimization. The integration of Power BI and Python significantly contributes to data processing and visualization, supporting the depth and accuracy of the analysis. The dashboard reveals a comprehensive view of customer interactions with products, highlighting significant engagement through various marketing channels, particularly Instagram and Facebook. Additionally, a detailed regional analysis showcases the performance of key products across different areas, identifying trends and opportunities for targeted marketing efforts. Overall, the dashboard empowers stakeholders to optimize marketing strategies, improve product performance, and drive sustainable growth by providing a visually engaging and interactive platform for analyzing complex data sets.

Chapter 9

References:

https://www.python.org/
https://www.microsoft.com/en-us/power-platform/products/power-bi#tabs-pill-bar-ocb9d418_tab1
https://www.kaggle.com/datasets/praneethkumar007/dmart-ready-online-store
pandas - Python Data Analysis Library

GeoPandas 0.14.4 — GeoPandas 0.14.4+0.g60c9773.dirty documentation

Top 50 matplotlib Visualizations - The Master Plots (w/ Full Python ...

Guy in a Cube - YouTube

How to Choose the Right Data Visualization | Atlassian

APPENDIX 1

For this project we have used the following system configuration:

- Windows 11
- POWERBI DESKTOP
- Matplotlib: 3.6.3
- Plotly (io):5.3.0
- python 3.12.1
- sacremoses 0.1.1
- scikit-learn 1.4.2
- scipy 1.12.0
- seaborn 0.13.2
- numpy 1.26.4
- geopandas 0.14.4
- plotly 5.22.0
- pandas 2.2.2
- pip 23.3.2
- palettable 3.3.3
- power Bi 2.129.1229.