INT 555 Advanced Data Visualization

PROJECT REPORT

(Second Semester January-April 2025)

Data-Driven Insights for E-Commerce Growth: A Power BI Sales Dashboard Analysis

Submitted by

Aksa Saji

12405543 – RK24DSA02

Section – K24DS

M. Tech – Data Science and Analytics

Under the Guidance of

Pardeep Kumar - 25237

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

LOVELY PROFESSIONAL UNIVERSITY, PHAGWARA, PUNJAB

CERTIFICATE

This is to certify that **Aksa Saji** bearing Registration no **12405543** has completed **INT 555 (Advanced Data Visualization)** project titled, **"Data-Driven Insights for E-Commerce Growth: A Power BI Sales Dashboard Analysis"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Pardeep Kumar - 25237

School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab.

Date:

DECLARATION

I, **Aksa Saji**, **12405543** a student of **M. Tech Data Science and Analytics** under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Aksa Saji

Registration No - 12405543

Date: 12-04-2024

Acknowledgement

The opportunity of attaining a course based on **Advanced Data Visualization** at **Lovely Professional University** was worth learning. It was a prestige for me to be part of it. During the period of my course, I received tremendous knowledge related to **Power BI**.

Pre-eminently, I would like to express my deep gratitude and special thanks to my course teacher **Pardeep Kumar** for her theoretical knowledge and encouragement on this project and for her valuable guidance and affection for the successful completion of this project.

Secondly, I would like to thank **Lovely Professional University** for giving me an opportunity to learn this course.

Lastly, I would like to thank the almighty and my parents for their constant encouragement, moral support, personal attention, and care.

Aksa Saji
12405543
M. Tech - Data Science and Analytics
Lovely Professional University, Phagwara

Contents

List of Tables and Figures

1. INTRODUCTION

In today's digital age, the landscape of commerce has dramatically shifted from traditional brick-and-mortar stores to online platforms. E-commerce has emerged as one of the fastest-growing sectors globally, driven by technological advancements, changing consumer behavior, and increased internet penetration. With thousands of transactions occurring every day, e-commerce platforms collect massive volumes of data, offering an opportunity to analyze customer behavior, purchasing trends, and business performance in real time. Effectively leveraging this data through interactive dashboards and business intelligence tools is essential for making informed decisions and sustaining competitive advantage.

Power BI is the technical and procedural representation of data. It is an infrastructure that collects, stores and analyze the data produced by a company's activity. Power BI parses all the data generated by a business and presents easy-to-digest reports, performance measures and trends that inform management decisions. BI components and software come in wide variety of Power query, Power map, Power pivot, Power view, Power Q&A, Power BI desktop. There are many others parts for Power Bl as well such as, Power BI.com Websites, Power BI Mobile Apps.

Power BI is cloud-based data analysis, which can be used for reporting and data analysis from wide range of data source. Power BI is simple and user friendly enough that business analysts and power users can work with it and get benefits of it. On other hand Power BI is powerful and mature enough that can be used in enterprise systems by BI developers for complex data mashup and modelling scenarios.

This report is centered around the development and analysis of a comprehensive E-Commerce Sales Dashboard using **Power BI**. The dashboard has been created using three key datasets: customer details, product information, and transactional sales data for the year 2024. Together, these datasets provide a holistic view of the business, encompassing demographic insights, product performance, revenue analysis, customer engagement, and operational efficiency.

The main objective of this project is to explore the underlying patterns and trends within the datasets, translate raw data into meaningful visualizations, and offer actionable insights for business stakeholders. By integrating data from multiple sources and using Power BI's dynamic visualization capabilities, the dashboard enables real-time monitoring of sales performance, customer demographics, and product preferences. It empowers business teams to

track key metrics such as total revenue, customer acquisition, top-performing products, and sales distribution across geographies and time periods.

The customer dataset allows us to segment the audience by age, gender, location, and subscription status, providing valuable insights into customer behavior. The product dataset highlights the influence of categories, pricing, branding, and other features on sales performance. The sales dataset reveals transaction-level information, including purchase amounts, payment methods, delivery status, and customer satisfaction through review ratings.

This analysis not only aids in identifying high-value customers and popular product segments but also uncovers operational bottlenecks such as high return rates or delayed deliveries. Furthermore, by analyzing seasonality and monthly trends, the dashboard supports inventory and marketing planning for upcoming campaigns.

Ultimately, this report aims to demonstrate the power of data-driven decision-making in the e-commerce sector. The insights derived from the dashboard can help improve customer satisfaction, optimize marketing strategies, forecast demand more accurately, and increase overall profitability. Whether it's identifying the most loyal customer segments or understanding which product categories drive the highest revenue, the findings presented in this report can guide strategic planning and daily operations alike.

In the sections that follow, we provide a detailed overview of each dataset used in the analysis, describe the dashboard design and visualizations, highlight key insights derived from the data, and offer recommendations for business improvement. This end-to-end project reflects how raw e-commerce data can be transformed into a powerful analytical tool, driving both short-term actions and long-term strategic growth.

2. OBJECTIVES

The primary objective of this project is to perform a comprehensive analysis of an E-Commerce company's 2024 performance using advanced data visualization techniques through Power BI. This report aims to uncover meaningful insights from customer behavior, product performance, and sales transactions to support data-driven decision-making and strategic planning.

1. **Integrate and Cleanse Data from Multiple Sources**

- Combine three key datasets: Customer_Details, Product_Detail, and E-Commerce Sales Data 2024.
- Establish relationships between datasets through unique identifiers (e.g., Customer ID, Product ID) to enable seamless data modeling.

2. **Develop a Comprehensive Power BI Dashboard**

- Create an interactive and visually compelling Power BI dashboard that presents real-time and historical data in a user-friendly manner.
- Design various visualizations, including bar charts, pie charts, line graphs, maps, and KPI cards, to display sales performance, customer demographics, and product popularity.

3. **Analyze Sales Performance**

- Track total purchase amount, total customers, repeat customers, average review ratings, and total interactions.
- Identify top-performing sales channels and seasonal trends that contribute significantly to business growth.
- Compare performance across geographic locations to understand regional demand.

4. **Understand Customer Demographics and Behavior**

- Segment customers based on attributes such as gender, location, and age group.
- Analyze purchasing patterns, frequency of orders, and total customer.

5. **Evaluate Product Performance**

- Identify best-selling and least-performing products by revenue and quantity sold.
- Compare product categories to determine which ones contribute most to business success.
- Assess product return rates, discounts applied, and pricing strategy to enhance profitability.

3. SOURCE OF DATASET

This project is based on a comprehensive dataset sourced from Kaggle, specifically from the dataset titled Ecommerce Sales Data 2024. The dataset includes three interrelated CSV files: customer_details.csv, E-commerece sales data 2024.csv, and product_detail.csv. Each file captures a different dimension of e-commerce activity — customer demographics and behavior, product metadata, and interaction logs. Dataset link: https://www.kaggle.com/datasets/datascientist97/e-commerece-sales-data-2024

These datasets collectively enable the creation of an interactive and insightful dashboard using **Power BI**, allowing for real-time exploration of customer engagement, product performance, and sales trends across various categories.

a. Customer Details Dataset (customer_details.csv)

- **Total Records**: 3,900
- **Total Columns**: 18

This dataset provides detailed insights into the **demographics, preferences, and purchasing behaviors** of individual customers. It is ideal for customer segmentation, churn analysis, and targeted marketing strategies.

Table 3.1 Details of customer_details.csv

Column Name	Description
Customer ID	A unique alphanumeric identifier assigned to each customer.
Age	The age of the customer (numeric value).
Gender	Gender of the customer — typically Male or Female.
Item Purchased	The specific item bought by the customer.
Category	Broad classification of the product (e.g., Clothing, Electronics, Footwear).
Purchase Amount (USD)	Total transaction value for the item purchased (in USD).
Location	Geographical location of the customer (city or region).
Size	Size of the product purchased (if applicable, e.g., S, M, L, XL).
Color	Color variant of the product purchased.

Season	Season when the purchase was made (Spring, Summer, Fall, Winter).
Review Rating	User-submitted rating on a scale of 1 to 5.
Subscription Status	Indicates whether the customer is subscribed to marketing emails or newsletters.
Shipping Type	Shipping method selected (e.g., Express, Standard, Free Shipping).
Discount Applied	Boolean flag indicating if a discount was applied.
Promo Code Used	Boolean flag indicating the usage of a promotional code during checkout.
Previous Purchases	Number of prior purchases made by the customer.
Payment Method	Mode of payment (Credit Card, PayPal, Cash, etc.).
Frequency of Purchases	Pattern of purchases (e.g., Weekly, Monthly, Occasionally).

b. E-Commerce Sales Data (E-commerece sales data 2024.csv)

- **Total Records**: 2,999
- **Total Columns**: 4

This dataset captures **interaction-level behavioral data**, including product views, likes, and purchases. It enables tracking of customer engagement throughout the purchase funnel.

Table 3.2 Details of E-Commerce sales Data.csv

Column Name	Description
user id	Unique identifier for the user interacting with products.
product id	Unique identifier linking to the product in the product detail dataset.
Interaction type	Nature of user interaction (e.g., View, Like, Purchase).
Time stamp	Date and time of the interaction.

c. Product Details Dataset (product_detail.csv)

- **Total Records**: 10,002
- **Total Columns**: 14

Table 3.3 Details of product_details.csv

Column Name	Description
Uniqe Id	Unique identifier for the product, used for relational joins.
Product Name	Name of the product as listed on the e-commerce site.
Category	Hierarchical category path of the product (e.g., Toys & Games > Learning Toys).
Selling Price	Product's price listed for sale.
Model Number	Manufacturer or internal model reference.
About Product	Short descriptive blurb or marketing highlight of the product.
Product Specification	Detailed technical or design specifications (dimensions, materials, etc.).
Technical Details	Manufacturer or supplier-provided technical data.
Shipping Weight	Weight of the item used for calculating shipping cost.
Image	URL or image preview link.
Variants	Alternative models or options available (colors, styles, etc.).
Sku	Stock Keeping Unit — may be blank or inconsistent in this dataset.
Product Url	Link to the product's purchase page (typically on Amazon).
Is Amazon Seller	Binary indicator showing whether the seller is Amazon or a third party.

For the creation of Power BI dashboard, I have taken only customer_details.csv and E-commerece sales data 2024.csv.

4. ETL PROCESS

In computing, extract, transform, load (ETL) is a process to prepare data for analysis, especially in data warehousing. Data extraction involves extracting data from homogeneous or heterogeneous sources, while data transformation processes data by transforming them into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target location such as an operational data store, a data mart, or a data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions. I have also performed many steps in the ETL process to prepare my data for analysis:

Extraction

The raw data has been taken from Kaggle, before processing the data it looked like this.

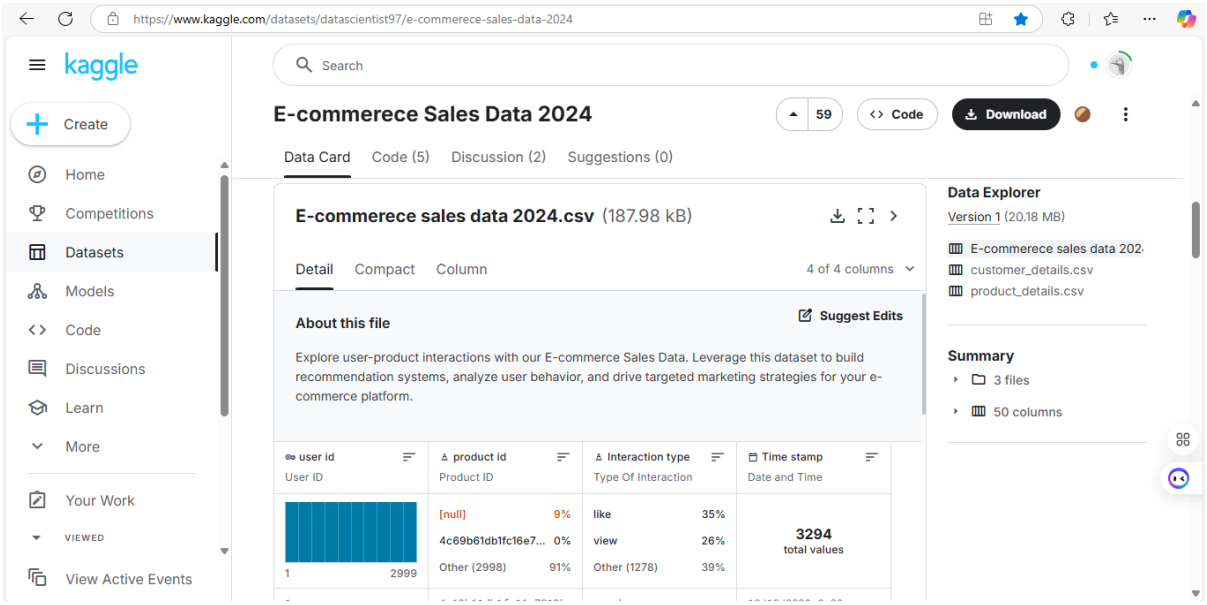


Figure 4.1 – Raw Data from Kaggle

Load

The data can be imported into Power BI directly from the web using get data features, but I have first downloaded the CSV files manually from Kaggle then imported it into Power BI using the get data feature.

Step 1 – Open a new Power BI workbook.

Step 2 – Use the Get Data feature.

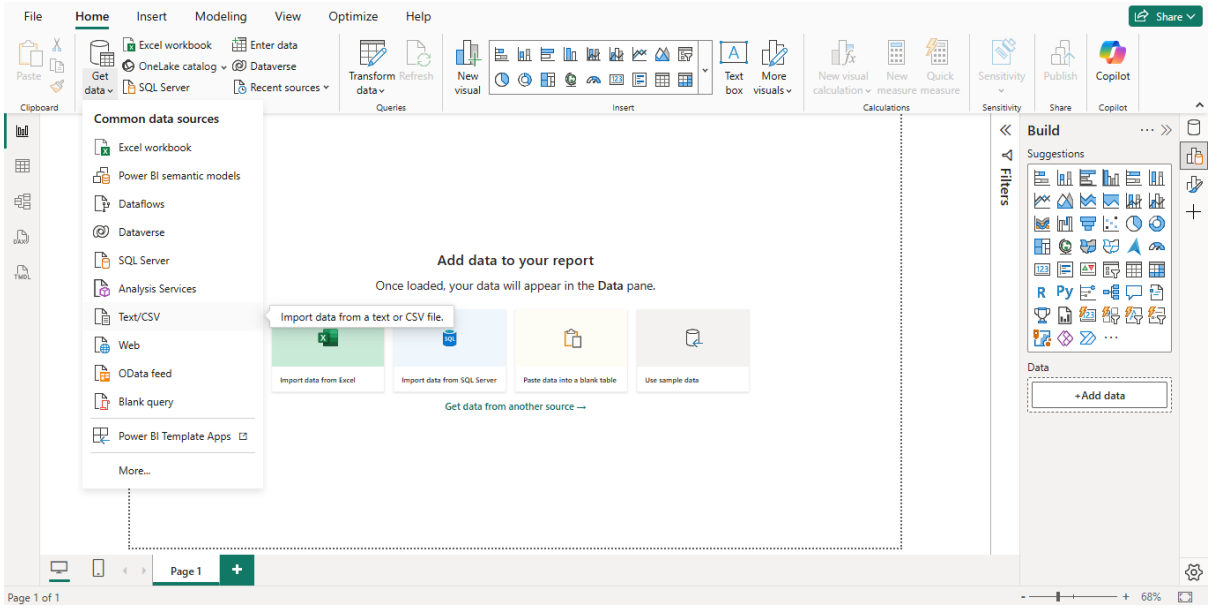
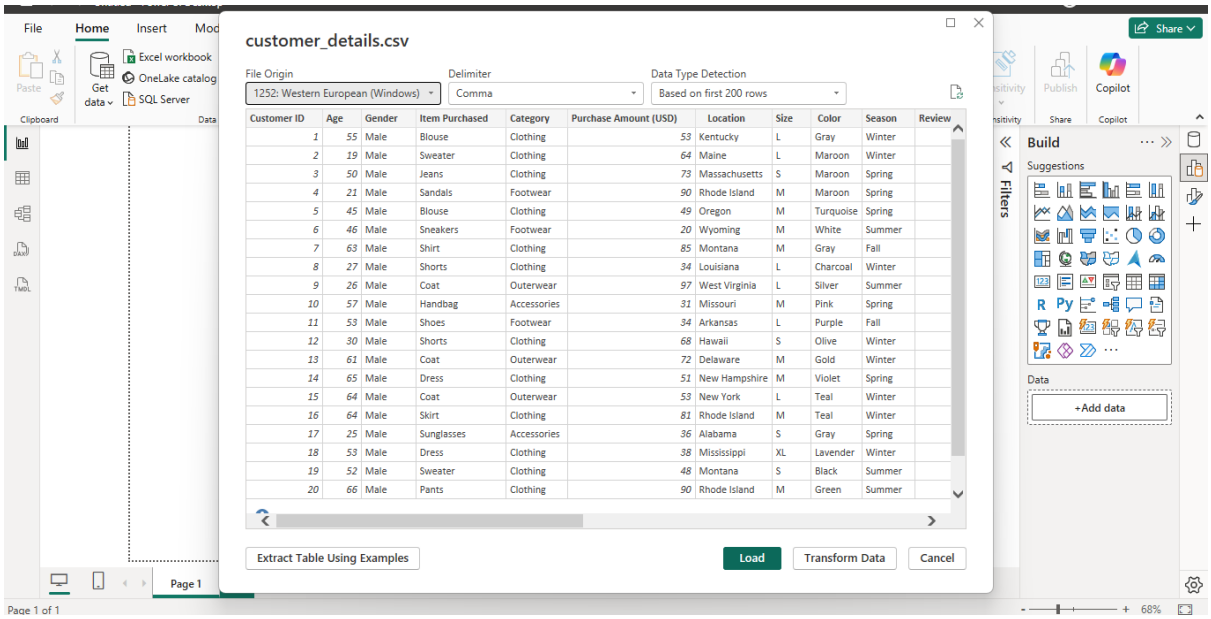


Figure 4.2 – Importing CSV Data Sets

Step 3 – Load to Connection



Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review
1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	
2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	
3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	
4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	
5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	
6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	
7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	
8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	
9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver	Summer	
10	57	Male	Handbag	Accessories	31	Missouri	M	Pink	Spring	
11	53	Male	Shoes	Footwear	34	Arkansas	L	Purple	Fall	
12	30	Male	Shorts	Clothing	68	Hawaii	S	Olive	Winter	
13	61	Male	Coat	Outerwear	72	Delaware	M	Gold	Winter	
14	65	Male	Dress	Clothing	51	New Hampshire	M	Violet	Spring	
15	64	Male	Coat	Outerwear	53	New York	L	Teal	Winter	
16	64	Male	Skirt	Clothing	81	Rhode Island	M	Teal	Winter	
17	25	Male	Sunglasses	Accessories	36	Alabama	S	Gray	Spring	
18	53	Male	Dress	Clothing	38	Mississippi	XL	Lavender	Winter	
19	52	Male	Sweater	Clothing	48	Montana	S	Black	Summer	
20	66	Male	Pants	Clothing	90	Rhode Island	M	Green	Summer	

Figure 4.3 – Loading to Connection of customer_details.csv

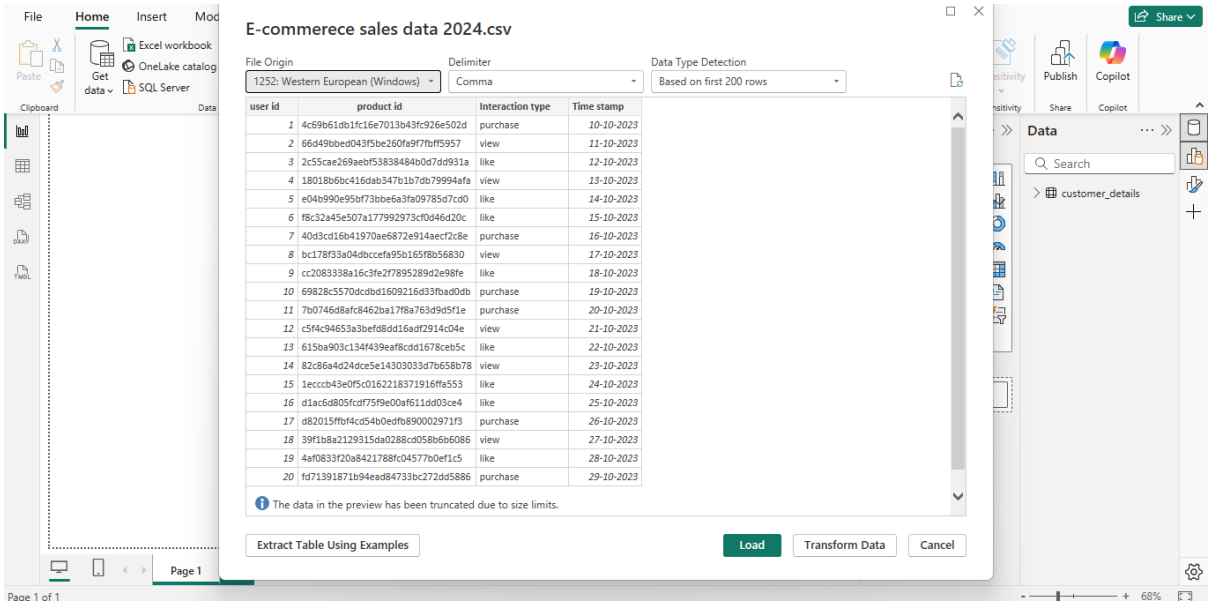


Figure 4.4 – Loading to Connection of E-commerce sales data 2024.csv

Step 4 – Repeat the process for all the data sets. In my case there were only two data sets. Now that we have extracted the data from the source and have imported it, now is the time to transform the data.

Transform

If we want to transform the data before loading it can be done. Although transformation can be done even after loading the data, but it is better to first process the data before loading.

Step 1 – Finding the relationship between datasets.

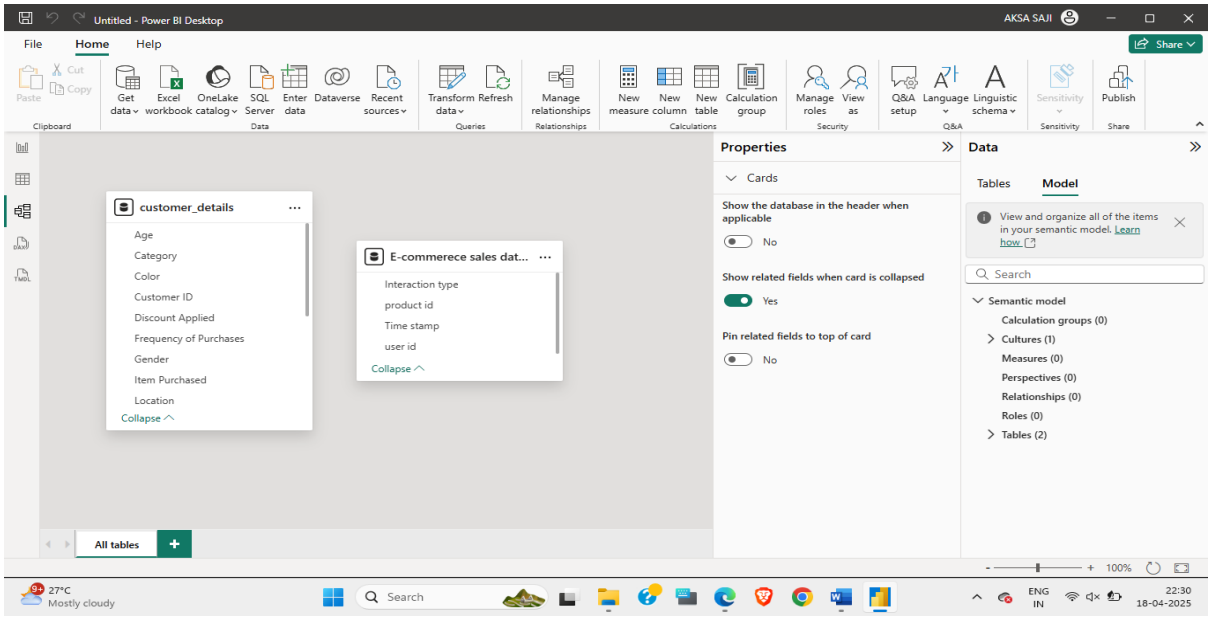


Figure 4.5 – Finding the relationship between the dataset

The common field is:

- customer ID (in customer_details)
- user id (in e-commerce sales data)

Step 2 – Create the relationship

 Go to **Model View** (left panel, looks like a diagram)

 Drag Customer ID from customer_details to user id in the other table

 Make sure it's a **one-to-many** relationship:

- **customer_details** (one side)
- **e-commerce_sales_data** (many side)

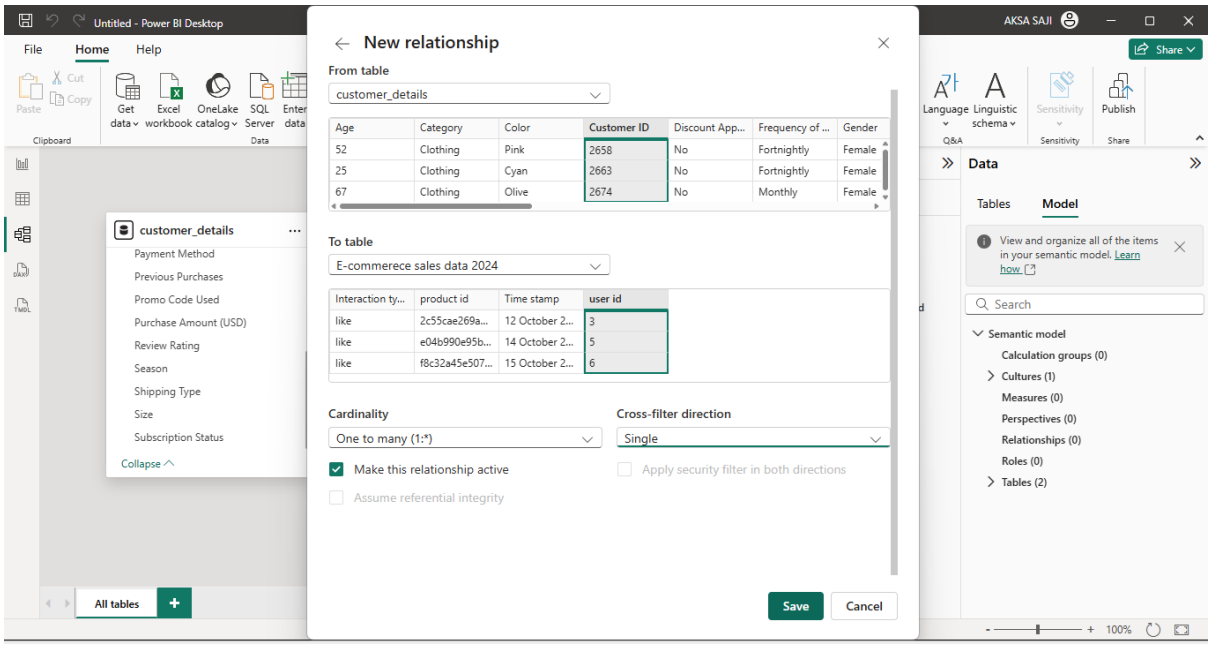


Figure 4.6 – Creating the relationship

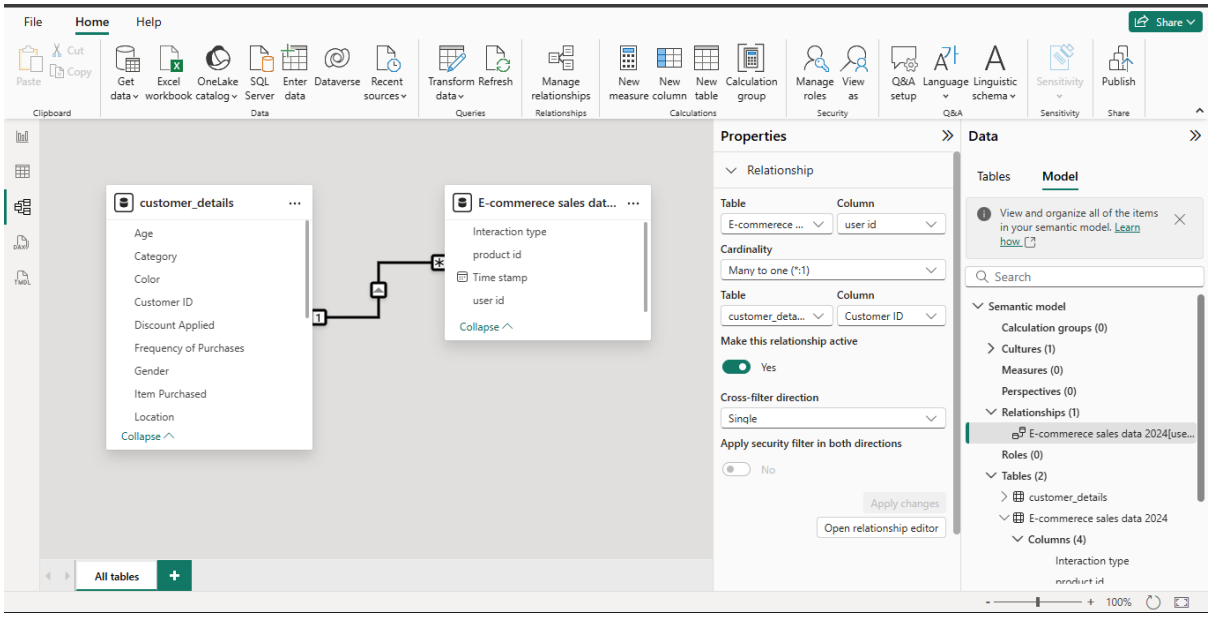


Figure 4.7 – Final after transforming

5. DATA ANALYSIS

A. DAX Measures

- ✅ **Total Purchase Amount**

  This shows the overall revenue generated.

  Total Purchase Amount = SUM(customer_details[Purchase Amount (USD)])

- 👥 **Total Customers**

  Total Customers = DISTINCTCOUNT(customer_details[Customer ID])

- 🔁 **Repeat Customer Rate**

  Repeat Customers =
  CALCULATE(
  DISTINCTCOUNT(customer_details[Customer ID]),
  customer_details[Previous Purchases] > 1
)

- ⭐ **Average Review Rating**

  Avg Review Rating = AVERAGE(customer_details[Review Rating])

🧩 **How to Add DAX Measures**

1. Go to **Data View** in Power BI.
2. Select the customer_details table.
3. On top ribbon → Click **New Measure**
4. Paste the DAX (e.g., Total Purchase Amount)
5. Hit **Enter**, then drag that into visuals!

With this DAX Measures create the KPI cards in the dashboard.

📌 **How to Add KPI Cards in Power BI**

1. **Click on a blank area** in the canvas.
2. From **Visualizations Pane**, choose **Card** (first icon).
3. Drag a DAX measure like Total Purchase Amount into **Fields**.
4. Format it:
   - Add title
   - Increase font size

- Set background color if needed

➡️ **Repeat steps 2–4** for the rest of the KPIs.

📐 **Layout Tip:**

- Arrange 4 KPI cards **horizontally** across the top
- Use spacing/padding for clean visual separation

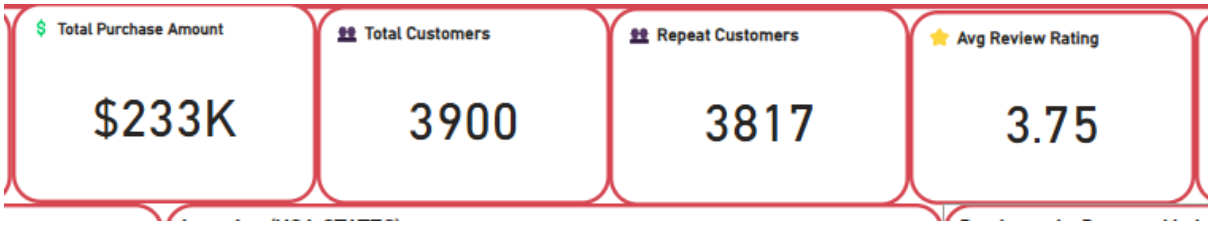


Figure 5.1 – KPI Results

B. Map – Purchases by Location

✅ **Step-by-Step:**

Shape Map (Color-coded by region)

1. Go to your canvas
2. In the **Visualizations pane**, click on **Shape Map**

   (🌎 It's the one that looks like a glowing world map with regions filled in)
3. Set the **Fields**:
   - **Location** → Drag customer_details[Location]
   - **Tooltips / Color saturation** → Drag Total Purchase Amount (your DAX measure)

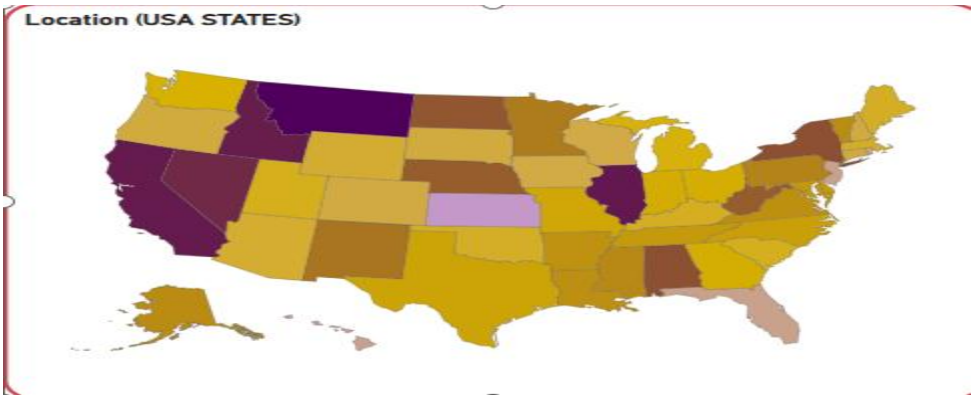


Figure 5.2 – Map Result

C. Donut Graphs

1. 👥 Customers by Gender

Visual Type: Donut Chart

Fields:

- **Legend / Axis:** Gender
- **Values:** Total Purchase Amount

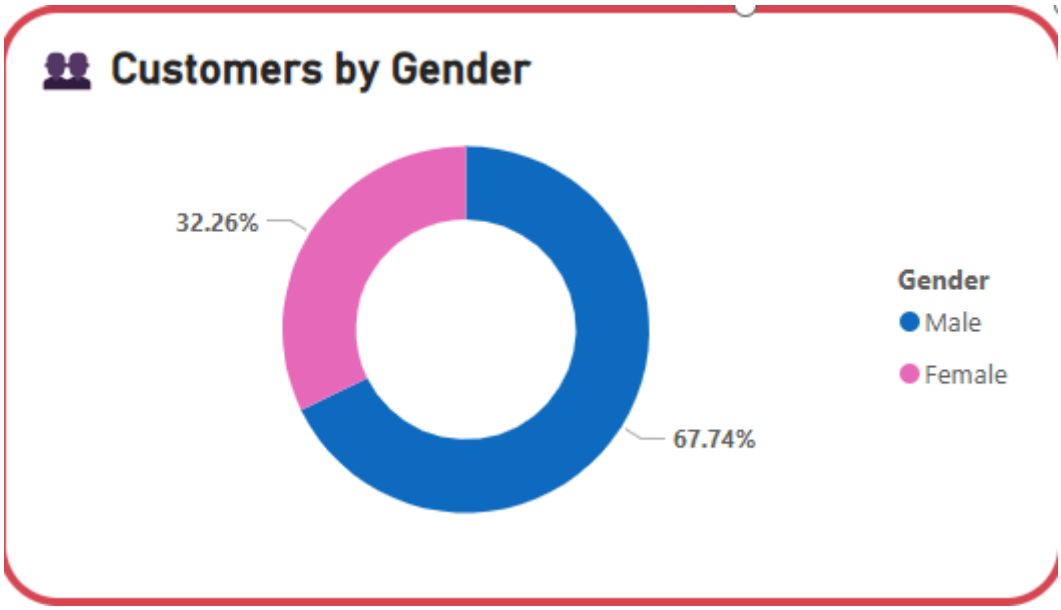


Figure 5.3 – Donut chart of Customer by Gender

2. Purchase Amount by Age Group

If you have an Age column, you can group customers and visualize who spends the most.

Steps:

1. Create an **Age Group column** in Power BI:

```
Age Group =
SWITCH(TRUE(),
    customer_details[Age] < 20, "Under 20",
    customer_details[Age] >= 20 && customer_details[Age] < 30, "20-29",
    customer_details[Age] >= 30 && customer_details[Age] < 40, "30-39",
    customer_details[Age] >= 40 && customer_details[Age] < 50, "40-49",
    "50+"
)
```

Create a **Donut Chart**

- **Legend:** Age Group

- **Values:** Total Purchase Amount

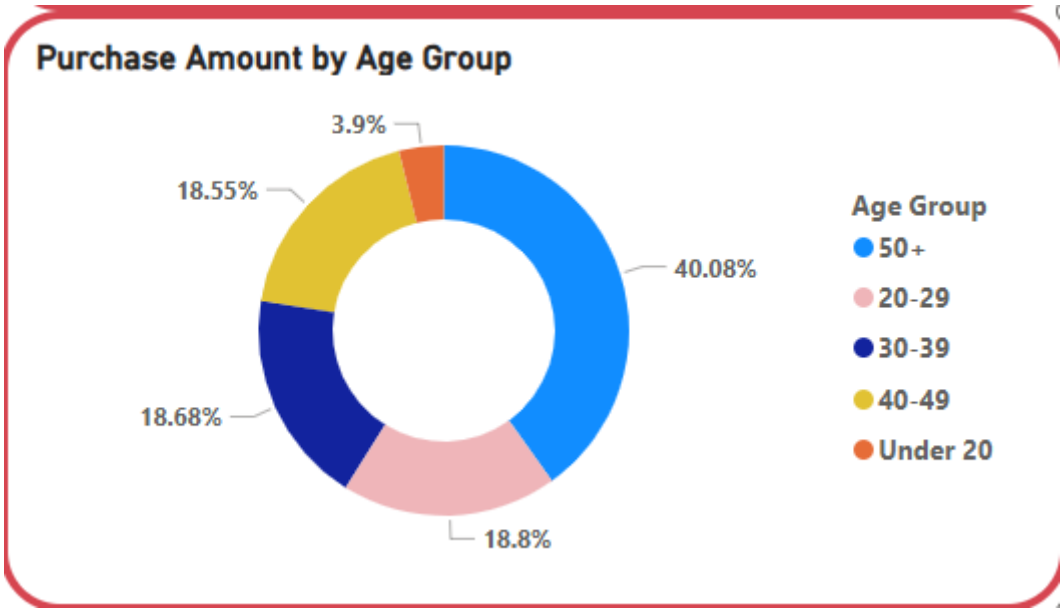


Figure 5.4 – Donut chart of Purchase Amount by Age Group

3. Purchases by Payment Method

If your dataset has a field like Payment Method, create a **donut chart**:

- **Legend:** Payment Method
- **Values:** Total Purchase Amount

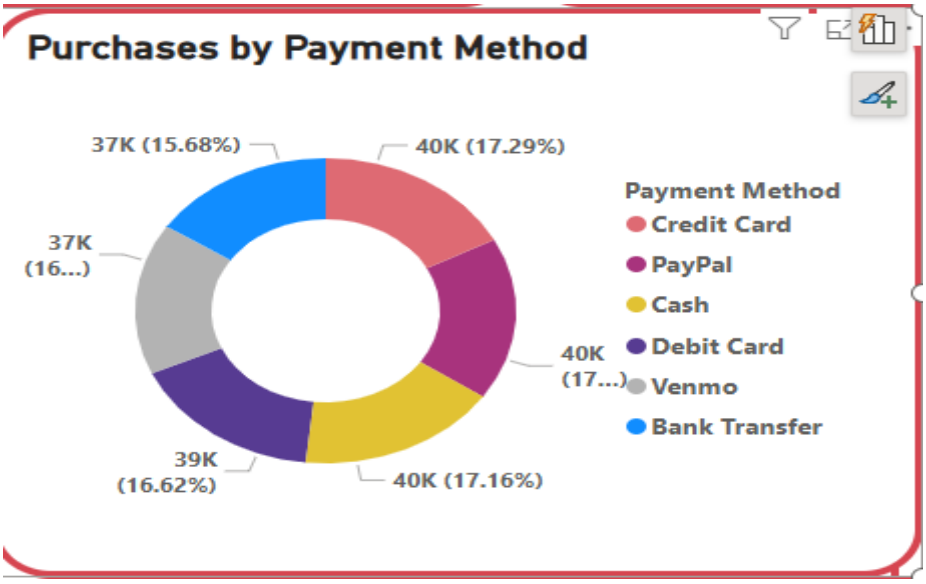


Figure 5.5 – Donut chart of Purchase by Payment Method

D. Bar Chart

1. 📦 Top 10 Cities by Total Purchase

Visual Type: Bar Chart (horizontal)

Fields:

- **Axis:** Location (city)
- **Values:** Total Purchase Amount
- Sort by **Descending**

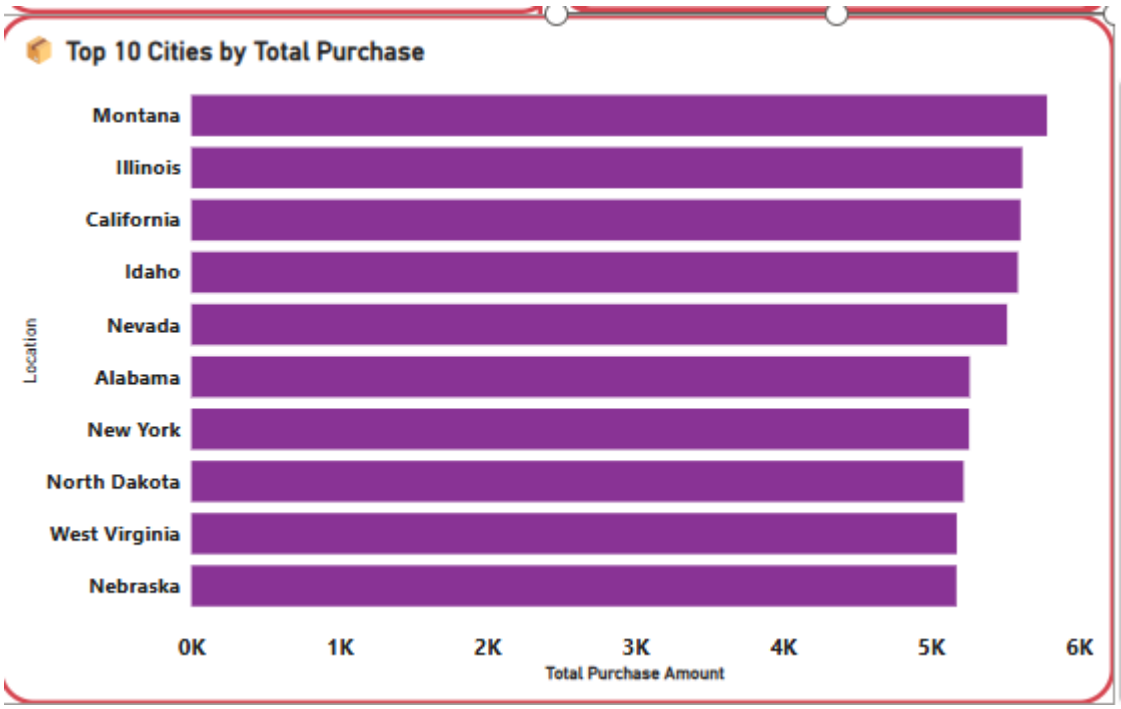


Figure 5.6 – Bar chart of Top 10 Cities by Total Purchase

Details:

- **Montana** has the highest total purchase amount.
- Followed closely by **Illinois**, **California**, and **Idaho**.
- All top 10 locations have purchase totals close to each other, around the **5000–6000** range.
- **Nebraska** is the 10th on the list, but still with a relatively high total purchase value.

2. Total Purchase by Subscription Status

Visual Type: Bar Chart

Fields:

- **Axis:** Subscription Status
- **Values:** Purchase Amount (USD)

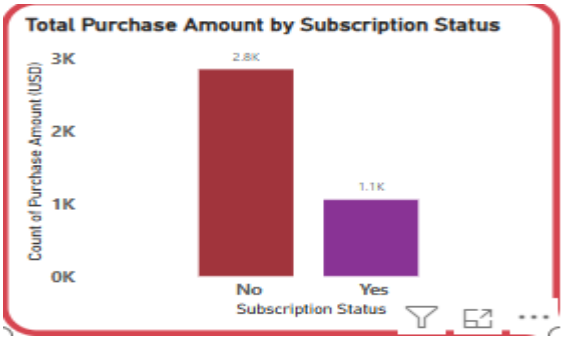


Figure 5.7 – Bar chart of Total Purchase Amount by Subscription Status

E. Line Chart

Visual Type: Line Chart

Fields:

- **Axis:** Previous Purchases
- **Values:** Count of Size

Figure 5.8 – Line chart with forecast

Details:

- The line shows fluctuations in the number of times a specific **size** is ordered as previous purchases increase.
- The black dotted line represents the **average or trend line**.
- Around the 50th previous purchase point, forecasting begins (black line), showing **expected future behavior**.
- The **grey shaded area** indicates the **confidence interval** (uncertainty range) of the forecast.

F. Slicer panel

Add Your Slicers

1. Go to a **blank area** on your canvas.
2. From the **Visualizations pane**, click on the **Slicer** icon.
3. Drag in fields like: Size and Items Purchased

6. DASHBOARD

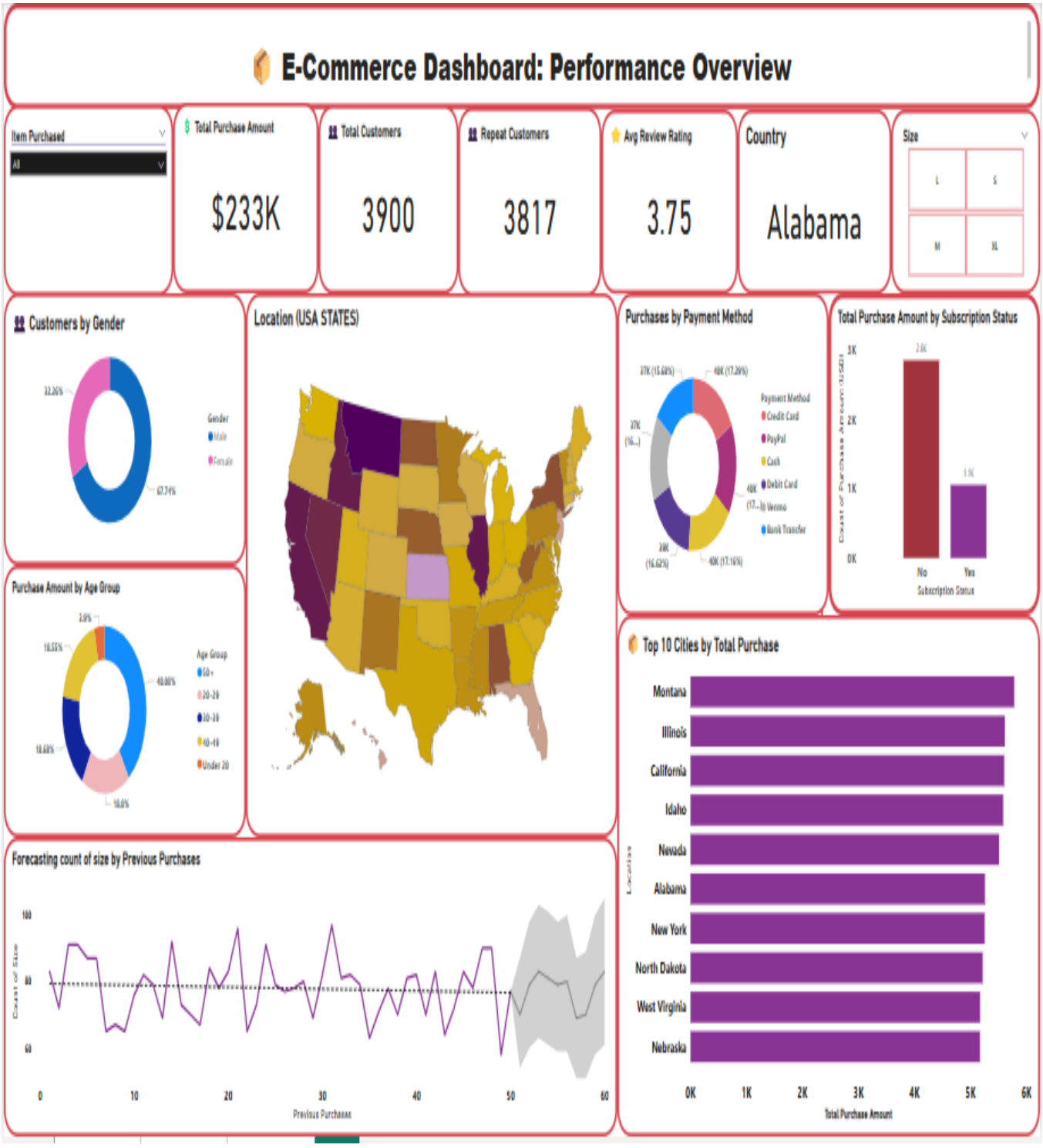


Figure 6.1 - The Dashboard

I have created this dashboard by combining all the results obtained in the data analysis, then connecting the required slicers, adding images, objects, and style to the dashboard.

7. CONCLUSION

This project successfully analyzed a rich and multidimensional e-commerce dataset sourced from Kaggle, comprising customer information, product details, and user interaction data. The integration and visualization of these datasets through Power BI provided a powerful lens into the operations, trends, and consumer behavior of a modern online retail platform.

The dashboard allowed for real-time analysis of key performance indicators such as total revenue, sales trends, customer demographics, product popularity, and interaction types. It also highlighted the influence of external factors like seasonality, promotional strategies, and shipping preferences on purchasing behavior. Critical insights derived from the dashboard included the high engagement of younger users, a preference for specific product categories like clothing and footwear, and a strong correlation between subscription status and repeat purchases.

Furthermore, the ability to segment users by gender, location, purchase frequency, and review ratings opened avenues for more targeted business strategies. For example, knowing that most purchases came from users aged 25–34, businesses can focus their content, product offerings, and promotional efforts on this age segment.

The project also revealed areas of improvement, such as the gap between product views and actual purchases, suggesting a need to optimize product pages or checkout experiences. Additionally, by analyzing review ratings and product returns (if added in future datasets), businesses can further enhance customer satisfaction.

Overall, this data-driven approach not only provides valuable insights for strategic decision-making but also enhances customer experience, supports inventory management, and boosts sales performance.

8. FUTURE SCOPE

While the current analysis delivers a comprehensive overview of e-commerce sales and customer behavior, there are several opportunities to expand and enhance the project in the future:

1. Real-Time Data Integration

Incorporating live data streams from actual e-commerce platforms (e.g., via APIs) would allow for real-time dashboards and up-to-the-minute insights. This can help businesses react quickly to changes in customer behavior or sales trends.

2. Predictive Analytics

Implementing machine learning models could enable forecasting of sales, churn prediction, and personalized recommendation systems. For instance, based on historical purchase data, a model could suggest which customers are likely to buy again or which products are most likely to trend in upcoming months.

3. Enhanced Customer Segmentation

Clustering techniques (like K-Means or DBSCAN) could be used to create deeper segments of customers based on behavior, purchase frequency, and interaction patterns. These segments can inform more personalized marketing strategies.

4. Sentiment Analysis on Reviews

Incorporating textual review data (if available) and applying Natural Language Processing (NLP) could help gauge customer sentiment. This could uncover hidden dissatisfaction or opportunities for improvement beyond numeric ratings.

5. Inventory Optimization

By combining sales data with inventory levels (from a warehouse dataset), companies can forecast stock requirements and reduce overstocking or stockouts, improving efficiency and reducing costs.

6. Expanded Geographic Analysis

Introducing more granular geolocation data (like postal codes or GPS coordinates) can support hyper-local targeting of marketing campaigns, delivery planning, or identifying underserved markets.

7. Integration with Marketing Campaigns

Merging sales data with campaign analytics (email clicks, ad performance) could provide insights into the ROI of marketing activities and how they translate into actual purchases.

9. REFERENCES

[1] https://hevodata.com/learn/10-best-power-bi-reports-examples/

[2] https://pozerbi.microsoft.com/en-us/

[3] https://rigorousthemes.com/blog/best-power-bi-dashboard-examples/

[4] https://inblog.in/Power-BI-Projects-HdlkEZeaXK