# INT374:DATA ANALYTICS WITH POWER BI

**PROJECT REPORT**

(Project Semester October-December 2025)

# *"Amazon Sales Analysis"*

Submitted by


Mahak

Registration No: 12319253

Programme and Section: B.tech. - CSE and K23KY

Course Code INT374


Under the Guidance of

**Dr. Baljinder Kaur**


**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

# CERTIFICATE

This is to certify that Mahak, bearing Registration No. 12319253, has completed the INT374 project titled **"**Amazon Sales 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.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of …….**

Lovely Professional University

Phagwara, Punjab.

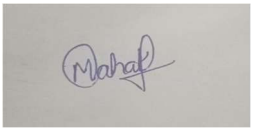
Date:

# DECLARATION

I, Mahak, student of Computer Science Engineering 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.

Date: 14-12-2025

Signature



Registration No: 12319253

Name of the student
Mahak

# Table of Content:

# 1. INTRODUCTION:

With the rapid growth of e-commerce, platforms like Amazon generate massive volumes of transactional and customer-related data on a daily basis. This data includes information about sales, customer behaviour, product categories, regional demand, and revenue trends. Analysing such large-scale data is essential for businesses to understand market dynamics, identify profitable products, optimize operations, and enhance customer satisfaction. Data-driven decision-making has therefore become a crucial aspect of modern business strategies.

Business Intelligence (BI) tools play a significant role in transforming raw and unstructured data into meaningful and actionable insights. Among these tools, Microsoft Power BI is widely used due to its powerful data visualization capabilities, ease of use, and ability to handle large datasets efficiently. Power BI allows users to perform data cleaning, modelling, and analysis, and present insights through interactive dashboards and reports. Features such as Power Query and DAX enable users to create advanced calculations and perform in-depth analysis.

This project focuses on analysing Amazon sales data using Power BI to gain a comprehensive understanding of sales performance across different products, categories, regions, and time periods. The dataset is processed and cleaned using Power Query Editor to ensure data accuracy and consistency. Various DAX measures are created to calculate key performance indicators such as total sales, profit, and quantity sold.

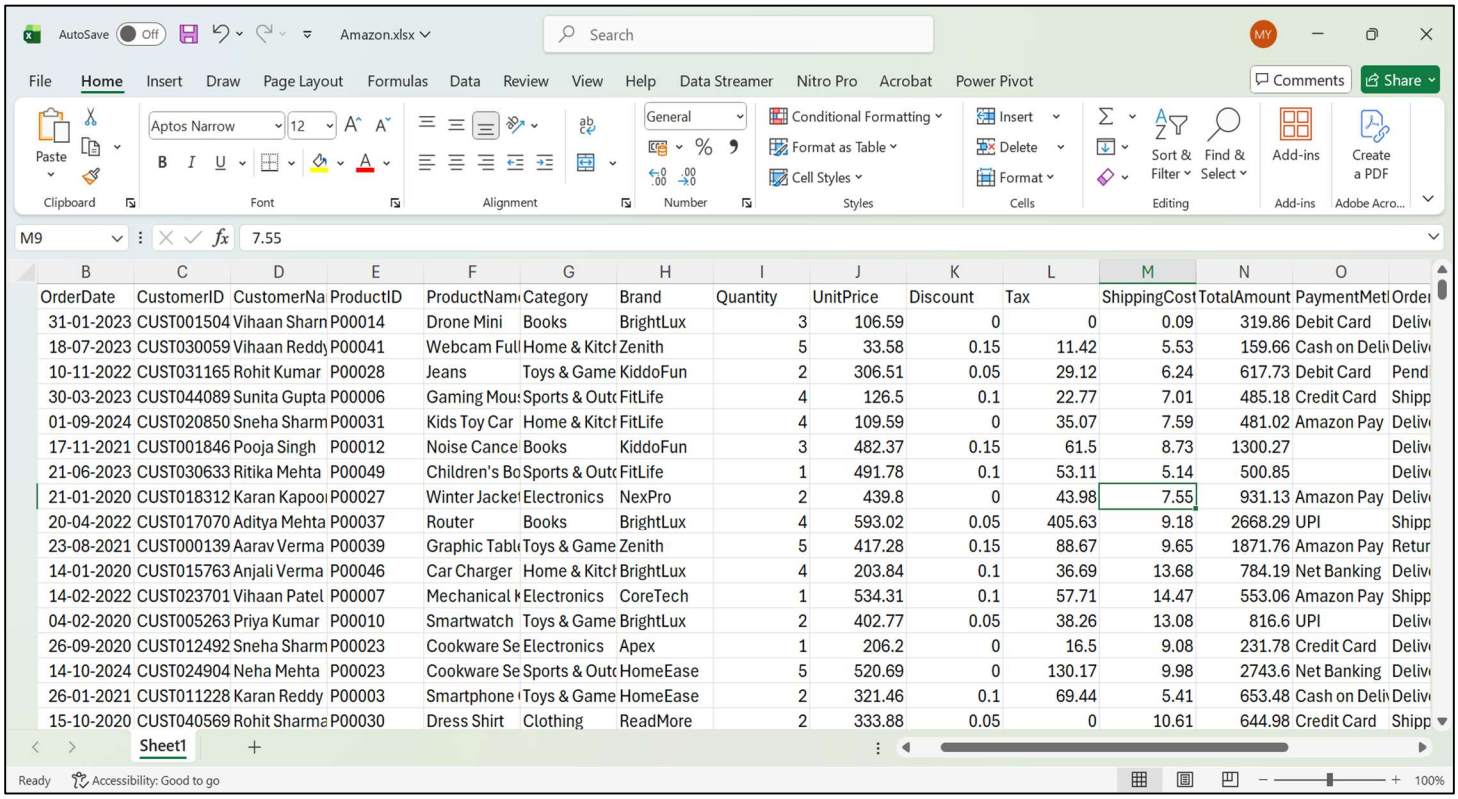
Interactive dashboards are designed using charts, KPIs, maps, and slicers to visually represent sales trends and patterns. These visualizations help in identifying high-performing product categories, seasonal sales trends, and regional performance variations. Overall, this project demonstrates how Power BI can be effectively used.

# 2. SOURCE OF DATASET:

Link: https://www.kaggle.com/datasets/rohiteng/amazon-sales-dataset

The dataset used for this project consists of Amazon sales data provided in Excel format. It contains information related to order details, product categories, sales amount, quantity sold, profit, customer segments, and geographical regions. The dataset serves as a realistic representation of e-commerce sales data suitable for analytical and visualization purposes.

Sample Data Set:



AutoSave Off | Amazon.xlsx | Search

File Home Insert Draw Page Layout Formulas Data Review View Help Data Streamer Nitro Pro Acrobat Power Pivot | Comments | Share

Paste | Aptos Narrow | 12 | General | Conditional Formatting | Format as Table | Cell Styles | Insert | Delete | Format | Sort & Filter | Find & Select | Add-ins | Create a PDF

Clipboard | Font | Alignment | Number | Styles | Cells | Editing | Add-ins | Adobe Acro...

M9 | 7.55

| | B | C | D | E | F | G | H | I | J | K | L | M | N | O | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OrderDate | CustomerID | CustomerNa | ProductID | ProductNam | Category | Brand | Quantity | UnitPrice | Discount | Tax | ShippingCost | TotalAmount | PaymentMetl | Order |
| | 31-01-2023 | CUST001504 | Vihaan Sharn | P00014 | Drone Mini | Books | BrightLux | 3 | 106.59 | 0 | 0 | 0.09 | 319.86 | Debit Card | Deliv |
| | 18-07-2023 | CUST030059 | Vihaan Reddy | P00041 | Webcam Full | Home & Kitcl | Zenith | 5 | 33.58 | 0.15 | 11.42 | 5.53 | 159.66 | Cash on Deli | Deliv |
| | 10-11-2022 | CUST031165 | Rohit Kumar | P00028 | Jeans | Toys & Game | KiddoFun | 2 | 306.51 | 0.05 | 29.12 | 6.24 | 617.73 | Debit Card | Pend |
| | 30-03-2023 | CUST044089 | Sunita Gupta | P00006 | Gaming Mou | Sports & Outc | FitLife | 4 | 126.5 | 0.1 | 22.77 | 7.01 | 485.18 | Credit Card | Shipp |
| | 01-09-2024 | CUST020850 | Sneha Sharm | P00031 | Kids Toy Car | Home & Kitcl | FitLife | 4 | 109.59 | 0 | 35.07 | 7.59 | 481.02 | Amazon Pay | Deliv |
| | 17-11-2021 | CUST001846 | Pooja Singh | P00012 | Noise Cance | Books | KiddoFun | 3 | 482.37 | 0.15 | 61.5 | 8.73 | 1300.27 | | Deliv |
| | 21-06-2023 | CUST030633 | Ritika Mehta | P00049 | Children's Bo | Sports & Outc | FitLife | 1 | 491.78 | 0.1 | 53.11 | 5.14 | 500.85 | | Deliv |
| | 21-01-2020 | CUST018312 | Karan Kapoo | P00027 | Winter Jacke | Electronics | NexPro | 2 | 439.8 | 0 | 43.98 | 7.55 | 931.13 | Amazon Pay | Deliv |
| | 20-04-2022 | CUST017070 | Aditya Mehta | P00037 | Router | Books | BrightLux | 4 | 593.02 | 0.05 | 405.63 | 9.18 | 2668.29 | UPI | Shipp |
| | 23-08-2021 | CUST000139 | Aarav Verma | P00039 | Graphic Tabl | Toys & Game | Zenith | 5 | 417.28 | 0.15 | 88.67 | 9.65 | 1871.76 | Amazon Pay | Retur |
| | 14-01-2020 | CUST015763 | Anjali Verma | P00046 | Car Charger | Home & Kitcl | BrightLux | 4 | 203.84 | 0.1 | 36.69 | 13.68 | 784.19 | Net Banking | Deliv |
| | 14-02-2022 | CUST023701 | Vihaan Patel | P00007 | Mechanical k | Electronics | CoreTech | 1 | 534.31 | 0.1 | 57.71 | 14.47 | 553.06 | Amazon Pay | Shipp |
| | 04-02-2020 | CUST005263 | Priya Kumar | P00010 | Smartwatch | Toys & Game | BrightLux | 2 | 402.77 | 0.05 | 38.26 | 13.08 | 816.6 | UPI | Deliv |
| | 26-09-2020 | CUST012492 | Sneha Sharm | P00023 | Cookware Se | Electronics | Apex | 1 | 206.2 | 0 | 16.5 | 9.08 | 231.78 | Credit Card | Deliv |
| | 14-10-2024 | CUST024904 | Neha Mehta | P00023 | Cookware Se | Sports & Outc | HomeEase | 5 | 520.69 | 0 | 130.17 | 9.98 | 2743.6 | Net Banking | Deliv |
| | 26-01-2021 | CUST011228 | Karan Reddy | P00003 | Smartphone | Toys & Game | HomeEase | 2 | 321.46 | 0.1 | 69.44 | 5.41 | 653.48 | Cash on Deli | Deliv |
| | 15-10-2020 | CUST040569 | Rohit Sharma | P00030 | Dress Shirt | Clothing | ReadMore | 2 | 333.88 | 0.05 | 0 | 10.61 | 644.98 | Credit Card | Shipp |

Sheet1

Ready | Accessibility: Good to go | 100%

# 3. DATA PREPROCESSING

Data preprocessing was performed using the Power Query Editor in Power BI, which serves as the data transformation and cleaning layer of the tool. The primary objective of this step was to prepare the raw Amazon sales dataset for accurate analysis and visualization by ensuring data quality, consistency, and reliability.

Initially, the Amazon sales dataset was imported into Power BI from an Excel file. After loading the data, each column was carefully examined to verify its data type. Columns such as Order Date were converted to the appropriate date format, while numerical fields including Sales, Profit, and Quantity were converted to numeric data types. Correct data typing was essential to avoid calculation errors and to ensure proper functioning of time-based and numerical visualizations.

Next, the dataset was reviewed for duplicate and irrelevant records. Columns that did not contribute to the analytical objectives were removed to optimize model performance and reduce data redundancy. Duplicate records, if present, were eliminated to maintain data accuracy and prevent misleading results in the analysis.

Missing and null values were handled using appropriate techniques such as replacing values or removing incomplete records, depending on the importance of the column. This step helped maintain consistency across the dataset and ensured that aggregations and calculations performed using DAX produced reliable results.

To enhance analytical depth, derived columns were created within Power Query. Columns such as Year and Month were extracted from the Order Date to support time-based trend analysis. Additionally, a Profit Category column was created to classify products into profit-based groups, enabling better comparison and segmentation during visualization.

Overall, these preprocessing steps transformed the raw dataset into a clean, well-structured, and analysis-ready format. Proper data preprocessing played a crucial role in improving the accuracy of insights generated and ensured smooth integration with data modelling, DAX calculations, and dashboard development in Power BI.

# 4.DATA MODELING AND DAX CALCULATIONS

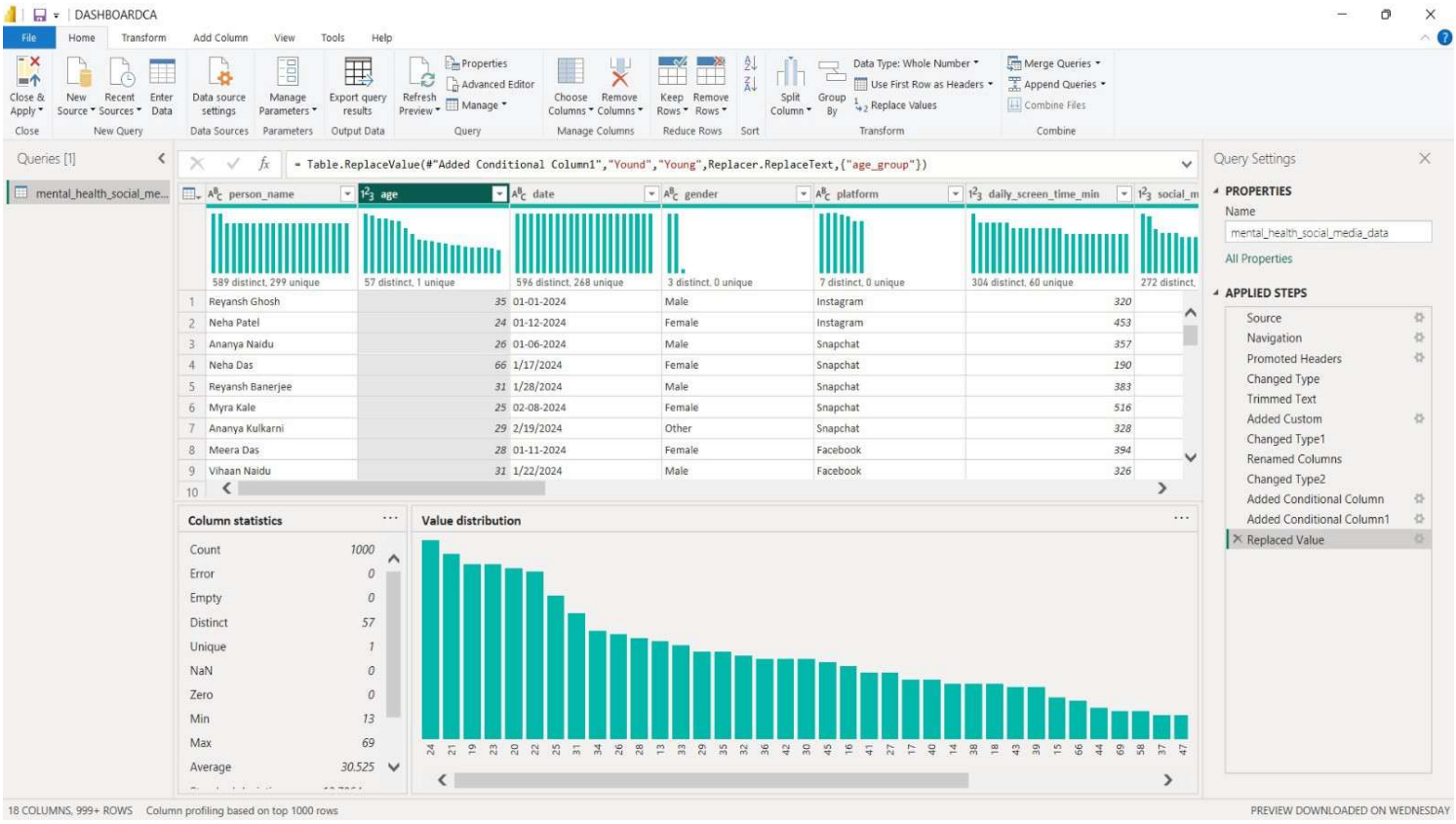
After completing the data preprocessing stage, data modelling was performed to support efficient analysis and visualization. Since the Amazon sales dataset consisted of a single main fact table containing all relevant sales and transactional information, complex relationships between multiple tables were not required. A simple and optimized data model was maintained to ensure better performance and ease of analysis.

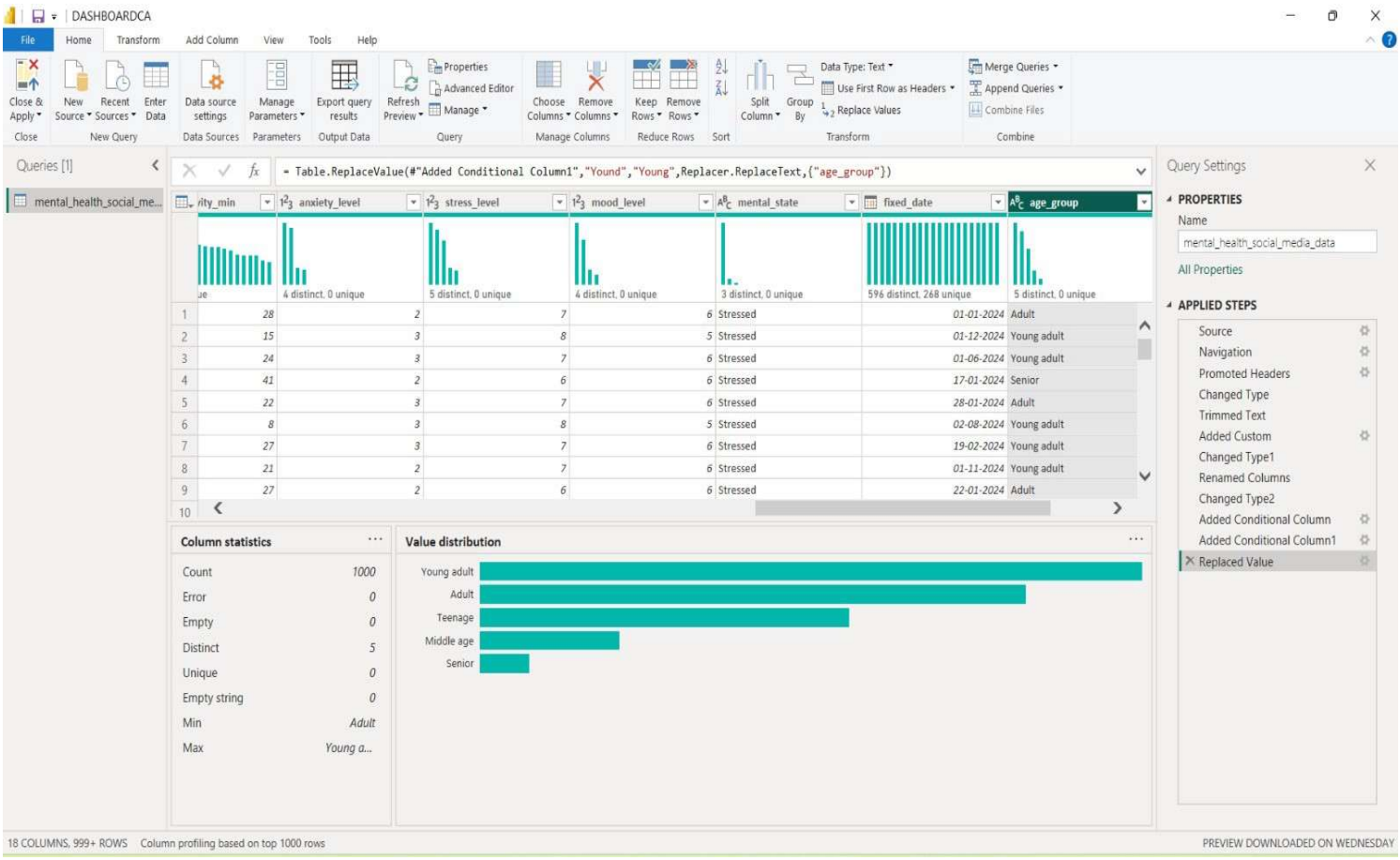
During the data modelling phase, attention was given to ensuring that all fields were correctly categorized and formatted to support visualizations such as charts, KPIs, and slicers. Date-related columns were utilized for time-based analysis, while numerical columns were optimized for aggregation and calculations.

To enable advanced analysis and dynamic reporting, DAX (Data Analysis Expressions) was used to create several calculated measures. These measures allow real-time calculations that automatically update based on user-selected filters and slicers. The key DAX measures created include:

- **Total Sales:** Calculates the overall revenue generated.
- **Total Profit:** Measures the net profit earned.
- **Total Quantity Sold:** Represents the total number of products sold.
- **Average Sales per Order:** Helps analyse average order value.
- **Profit Margin:** Evaluates profitability efficiency.

These DAX measures form the analytical foundation of the dashboard. They allow the dashboard to respond dynamically to filters such as year, region, and product category, making the analysis interactive, flexible, and insightful.


DASHBOARDCA
File
Home
Transform
Add Column
View
Tools
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Export query results
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Whole Number
Use First Row as Headers
Replace Values
Merge Queries
Append Queries
Combine Files
Close
New Query
Data Sources
Parameters
Output Data
Query
Manage Columns
Reduce Rows
Sort
Transform
Combine
Queries [1]
mental_health_social_me...
= Table.ReplaceValue(#"Added Conditional Column1","Yound","Young",Replacer.ReplaceText,{"age_group"})
person_name
age
date
gender
platform
daily_screen_time_min
social_m
589 distinct, 299 unique
57 distinct, 1 unique
596 distinct, 268 unique
3 distinct, 0 unique
7 distinct, 0 unique
304 distinct, 60 unique
272 distinct,
1 Reyansh Ghosh 35 01-01-2024 Male Instagram 320
2 Neha Patel 24 01-12-2024 Female Instagram 453
3 Ananya Naidu 26 01-06-2024 Male Snapchat 357
4 Neha Das 66 1/17/2024 Female Snapchat 190
5 Reyansh Banerjee 31 1/28/2024 Male Snapchat 383
6 Myra Kale 25 02-08-2024 Female Snapchat 516
7 Ananya Kulkarni 29 2/19/2024 Other Snapchat 328
8 Meera Das 28 01-11-2024 Female Facebook 394
9 Vihaan Naidu 31 1/22/2024 Male Facebook 326
10
Column statistics
Count 1000
Error 0
Empty 0
Distinct 57
Unique 1
NaN 0
Zero 0
Min 13
Max 69
Average 30.525
Value distribution
24 21 19 23 20 22 25 31 34 26 28 13 33 29 35 32 36 42 30 45 16 41 27 17 40 14 38 18 43 39 15 66 44 69 58 37 47
Query Settings
PROPERTIES
Name
mental_health_social_media_data
All Properties
APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
Trimmed Text
Added Custom
Changed Type1
Renamed Columns
Changed Type2
Added Conditional Column
Added Conditional Column1
Replaced Value
18 COLUMNS, 999+ ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED ON WEDNESDAY


DASHBOARDCA
File
Home
Transform
Add Column
View
Tools
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Export query results
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Text
Use First Row as Headers
Replace Values
Merge Queries
Append Queries
Combine Files
Close
New Query
Data Sources
Parameters
Output Data
Query
Manage Columns
Reduce Rows
Sort
Transform
Combine
Queries [1]
mental_health_social_me...
= Table.ReplaceValue(#"Added Conditional Column1","Yound","Young",Replacer.ReplaceText,{"age_group"})
ity_min
anxiety_level
stress_level
mood_level
mental_state
fixed_date
age_group
4 distinct, 0 unique
5 distinct, 0 unique
4 distinct, 0 unique
3 distinct, 0 unique
596 distinct, 268 unique
5 distinct, 0 unique
1 28 2 7 6 Stressed 01-01-2024 Adult
2 15 3 8 5 Stressed 01-12-2024 Young adult
3 24 3 7 6 Stressed 01-06-2024 Young adult
4 41 2 6 6 Stressed 17-01-2024 Senior
5 22 3 7 6 Stressed 28-01-2024 Adult
6 8 3 8 5 Stressed 02-08-2024 Young adult
7 27 3 7 6 Stressed 19-02-2024 Young adult
8 21 2 7 6 Stressed 01-11-2024 Young adult
9 27 2 6 6 Stressed 22-01-2024 Adult
10
Column statistics
Count 1000
Error 0
Empty 0
Distinct 5
Unique 0
Empty string 0
Min Adult
Max Young a...
Value distribution
Young adult
Adult
Teenage
Middle age
Senior
Query Settings
PROPERTIES
Name
mental_health_social_media_data
All Properties
APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
Trimmed Text
Added Custom
Changed Type1
Renamed Columns
Changed Type2
Added Conditional Column
Added Conditional Column1
Replaced Value
18 COLUMNS, 999+ ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED ON WEDNESDAY

# 5.OBJECTIVES AND ANALYSIS

## OBJECTIVE 1:

### Analyse overall sales and order performance

**Visualizations Used:**

- KPI Cards (Total Orders, Total Quantity, Total Shipping Cost, Average Tax, Average Shipping Cost)

**Type of Analysis:**
Descriptive and performance analysis

**Analysis:**
This objective focuses on evaluating the overall sales and order performance of the Amazon dataset. KPI cards provide a high-level summary of key metrics such as total number of orders, total quantity sold, total shipping cost, and average tax and shipping cost. These indicators help in understanding the scale of business operations and overall transaction volume. The KPIs offer an immediate snapshot of business performance before moving to detailed analysis are used to display key metrics such as Total Sales, Total Profit, and Total Quantity Sold. These KPIs provide a quick, high-level overview of business performance and help users understand the overall sales health of the organization. This visualization acts as a snapshot and sets the context for deeper analysis.

# OBJECTIVE 2:

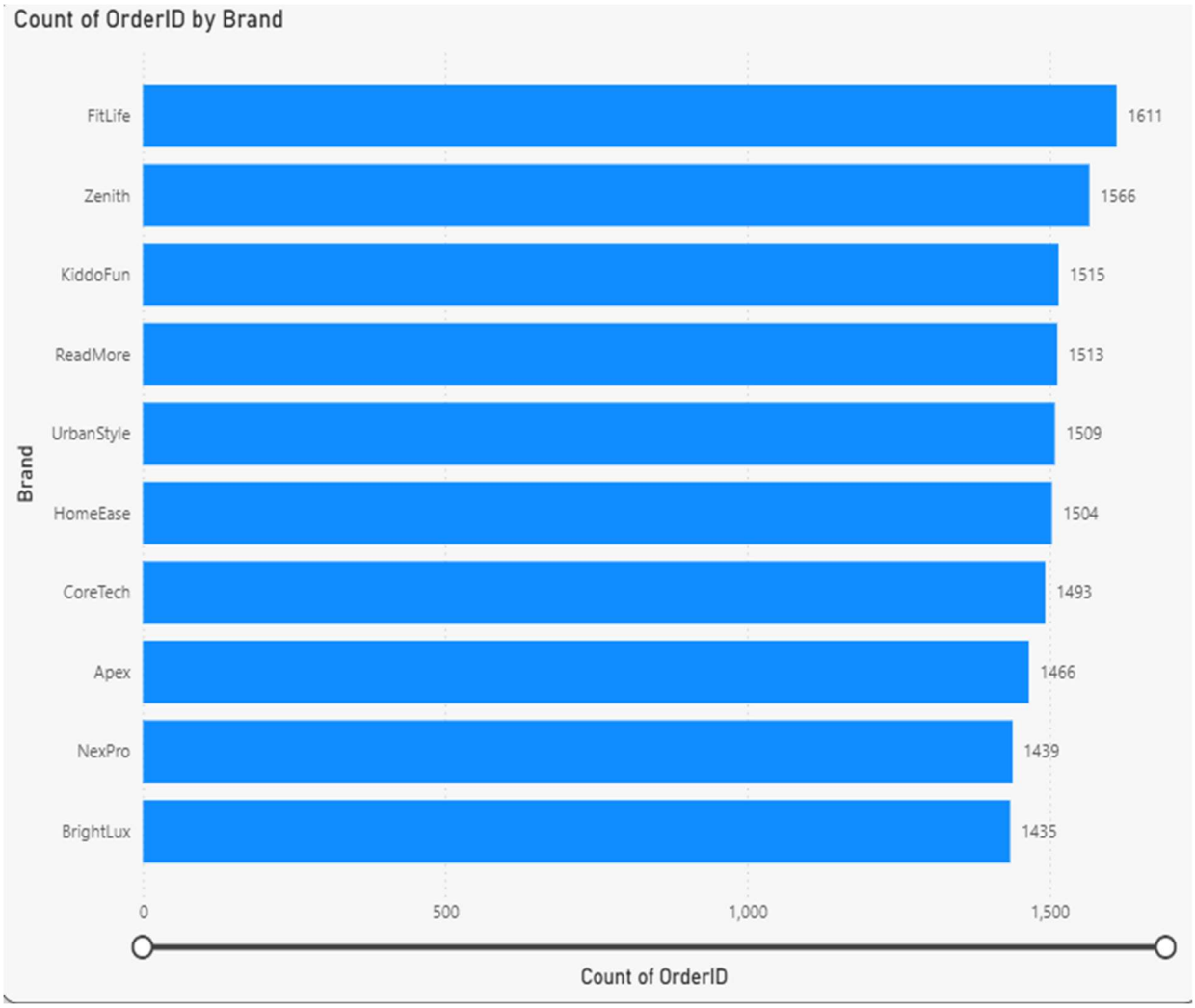
## Analyse brand-wise order distribution

### Visualization Used:

- Bar Chart (Count of Order ID by Brand)

**Type of Analysis:**
Comparative analysis

**Analysis:**
This objective aims to examine how orders are distributed across different brands. The bar chart highlights brands with the highest number of orders, indicating strong customer preference and demand. Brands with lower order counts may require improved visibility or promotional strategies. This analysis helps in identifying top-performing and underperforming brands.



# OBJECTIVE 3:

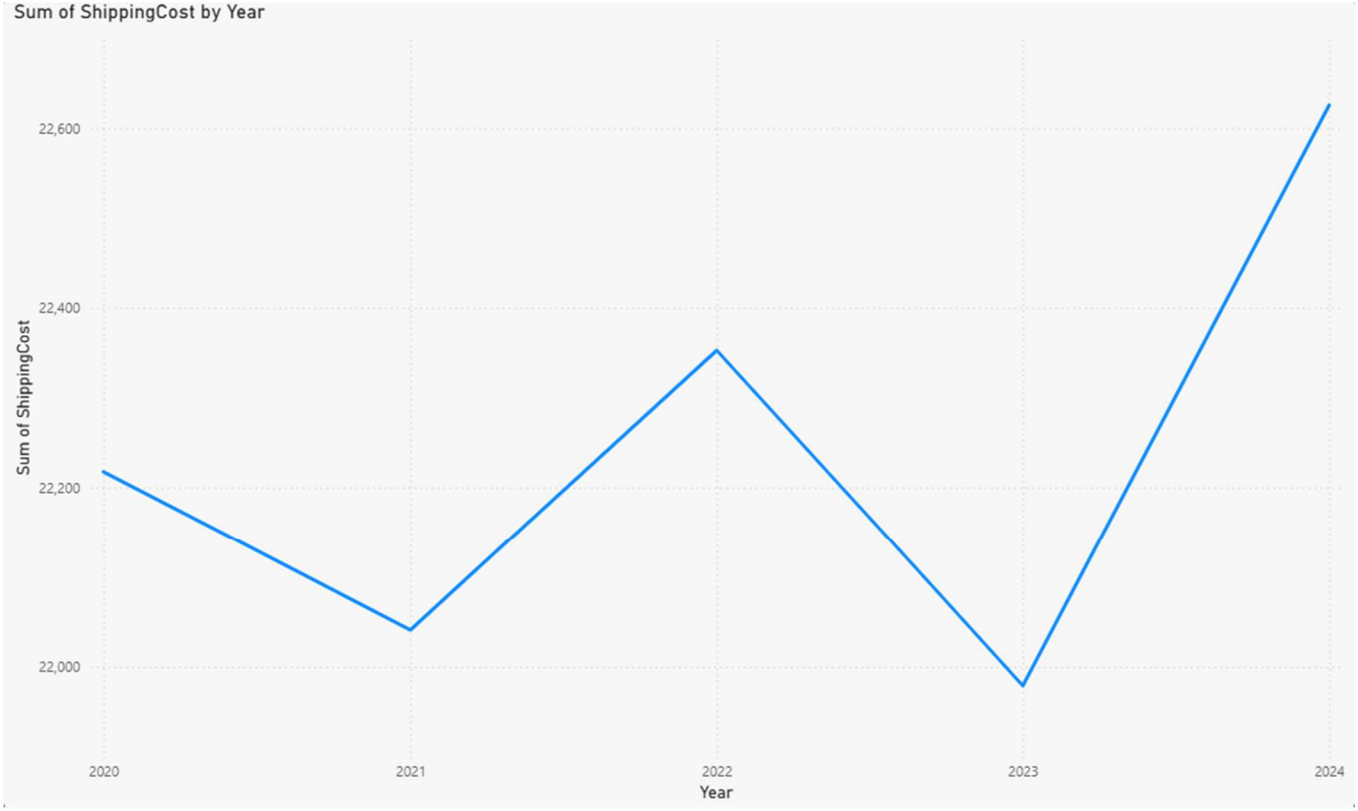
## Study year-wise shipping cost trends

### Visualization Used:

- Line Chart (Sum of Shipping Cost by Year)

**Type of Analysis:**
Trend and time-series analysis

**Analysis:**
This objective focuses on understanding how shipping costs have changed over different years. The line chart clearly shows fluctuations and trends in shipping expenses over time. An increase or decrease in shipping cost may indicate changes in logistics strategy, fuel costs, or order volume. This analysis helps in monitoring cost efficiency over the years.



# OBJECTIVE 4:

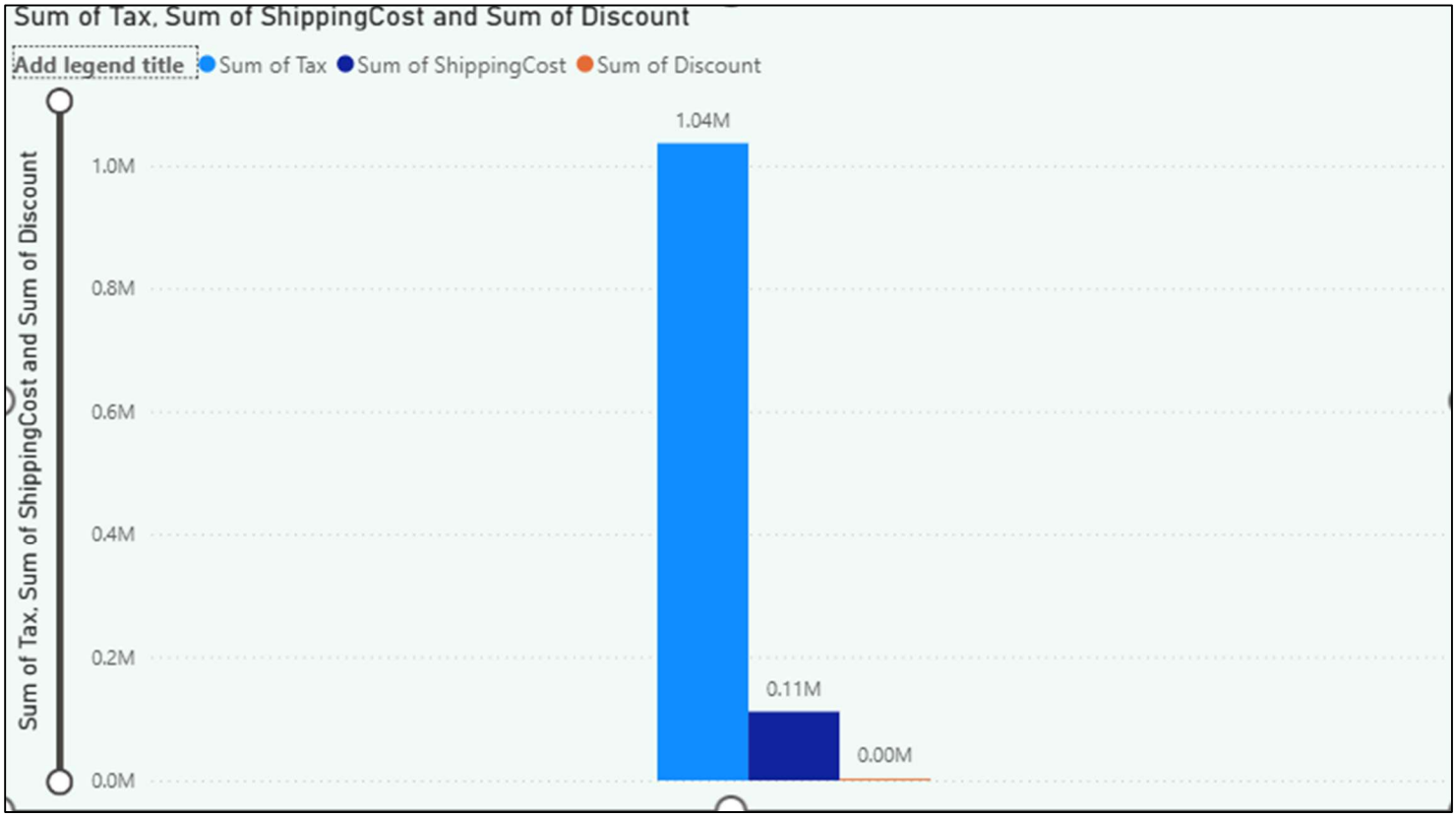
## Compare tax, shipping cost, and discount amounts

**Visualization Used:**

- Clustered Column Chart (Sum of Tax, Sum of Shipping Cost, Sum of Discount)

**Type of Analysis:**
Comparative financial analysis

**Analysis:**
This objective aims to compare major cost-related components such as tax, shipping cost, and discounts. The column chart helps in understanding which component contributes the most to the overall transaction value. This comparison supports better financial planning and pricing strategy decision



# OBJECTIVE 5:

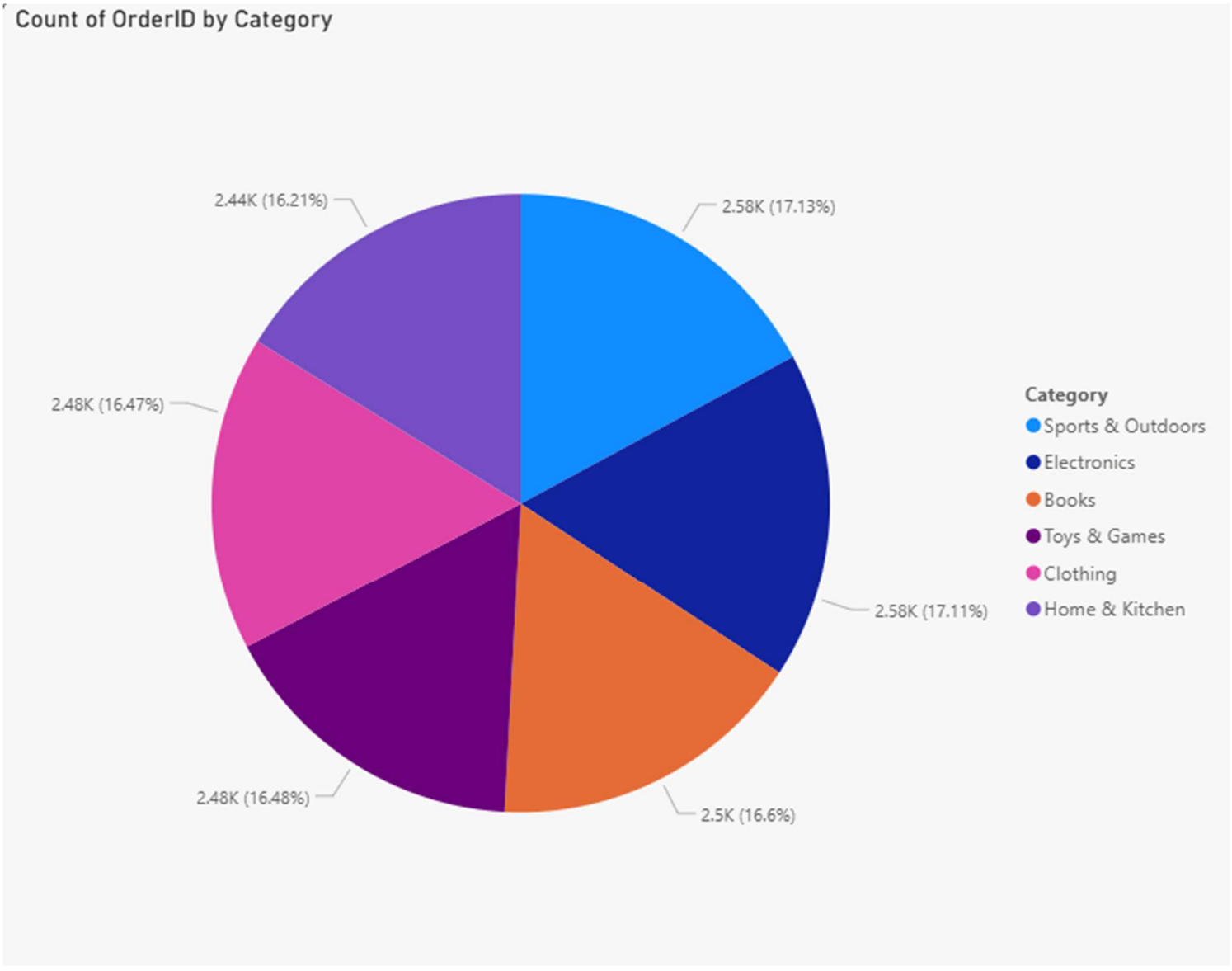
## Analyse category-wise quantity distribution

**Visualization Used:**

- Pie Chart (Sum of Quantity by Category)

**Type of Analysis:**
Proportional distribution analysis

**Analysis:**
This objective analyses how product quantities are distributed across different categories. The pie chart visually represents the contribution of each category to total quantity sold. Categories with higher shares indicate strong demand, while lower shares may point toward niche or low-performing categories.



# OBJECTIVE 6:

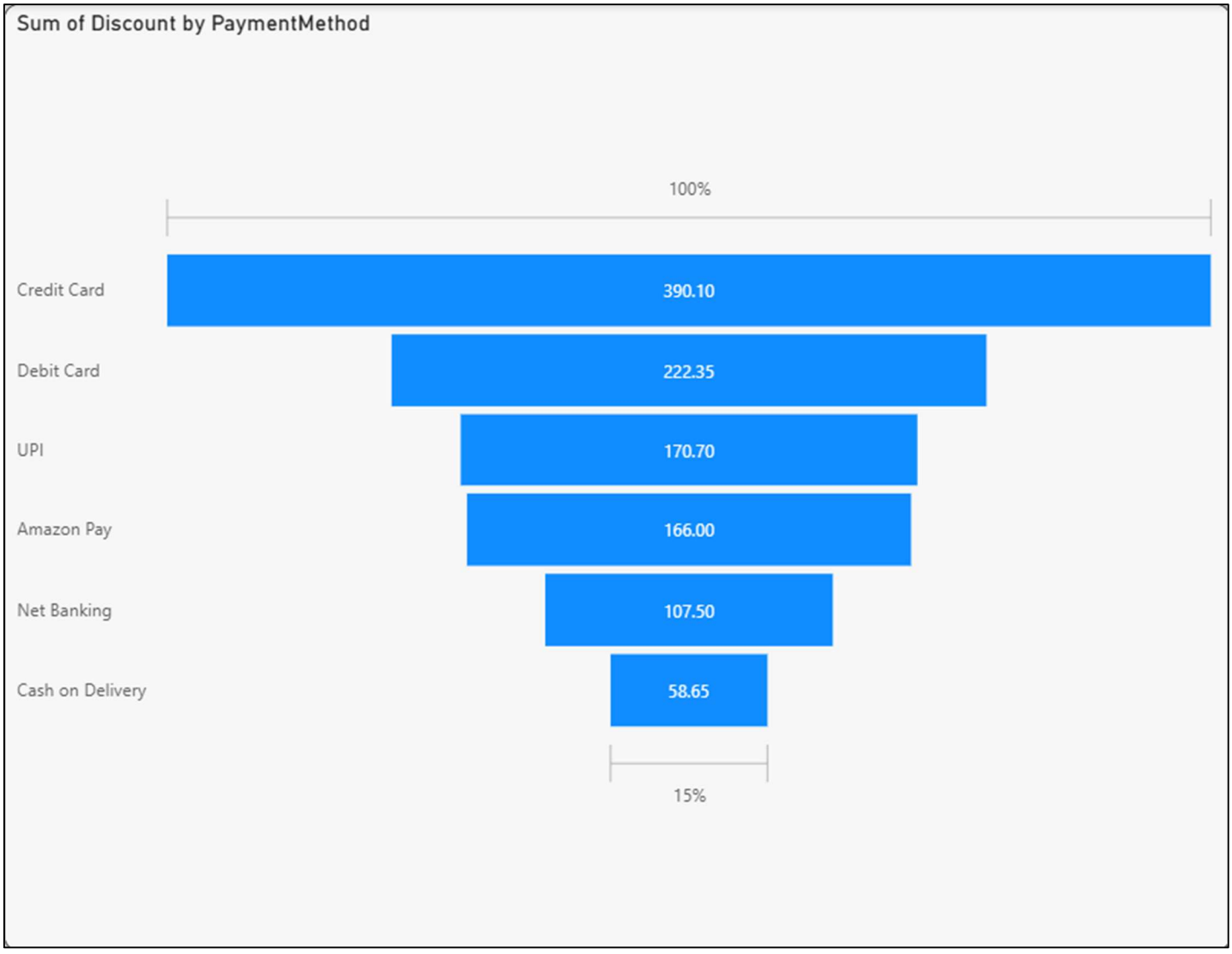
## Analyse discount distribution by payment method

**Visualization Used:**

- Funnel Chart / Bar Chart (Sum of Discount by Payment Method)

**Type of Analysis:**
Payment behaviour and discount analysis

**Analysis:**
This objective focuses on understanding how discounts vary across different payment methods such as credit card, debit card, UPI, and cash on delivery. The visualization highlights payment methods that receive higher discounts, helping businesses evaluate promotional strategies and customer payment preferences.



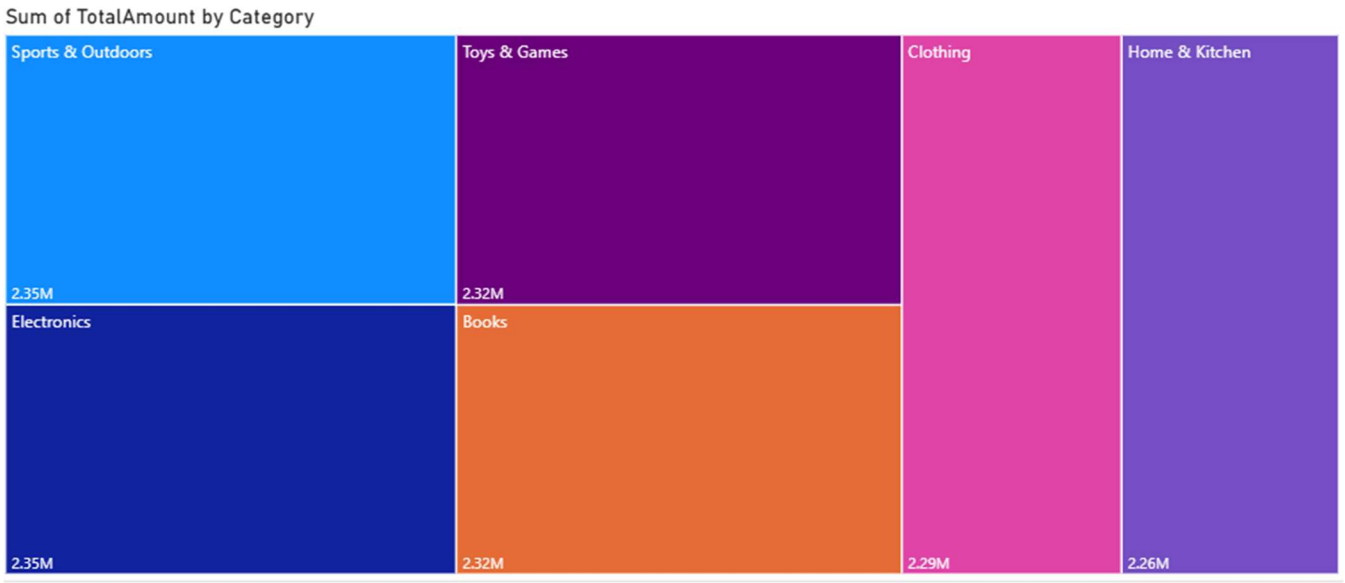
# OBJECTIVE 7:
# Analyse category-wise revenue contribution

**Visualization Used:**

- Tree Map (Sum of Total Amount by Category)

**Type of Analysis:**
Revenue contribution analysis

**Analysis:**
This objective examines the contribution of each product category to the total revenue. The tree map highlights high-revenue-generating categories such as electronics, sports, or clothing. Larger blocks represent categories with higher revenue contribution, helping in identifying profitable segments.



# OBJECTIVE 8:

## Enable interactive and dynamic analysis using slicers

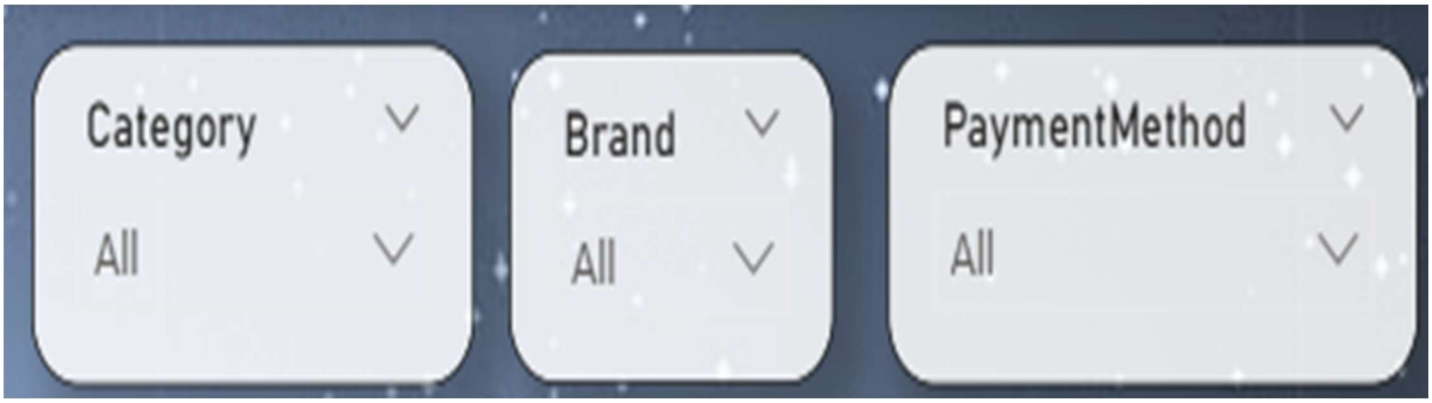
**Visualization Used:**

- Slicers (Category, Brand, Payment Method)

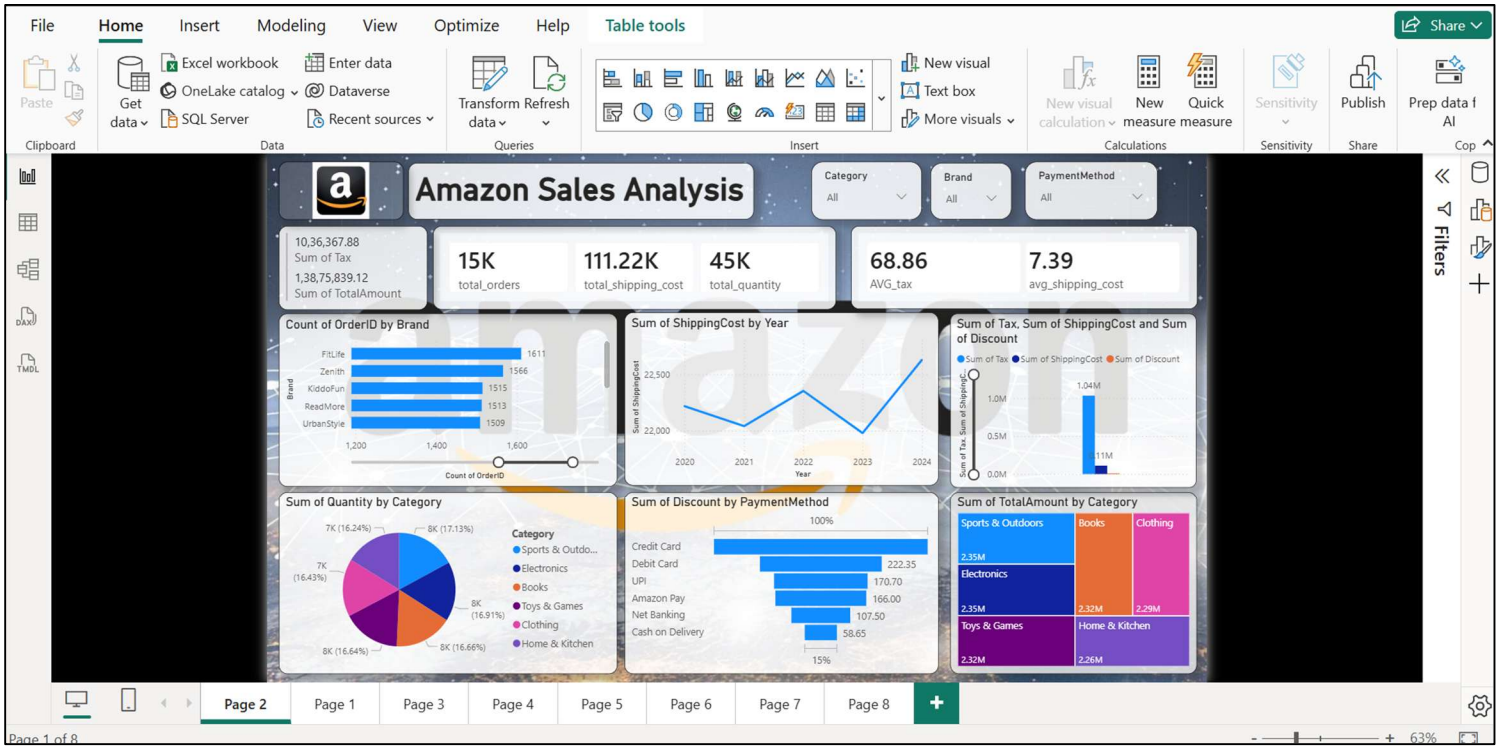
**Type of Analysis:**
Interactive and exploratory analysis

**Analysis:**

This objective enhances user interaction by enabling dynamic filtering of the dashboard. Slicers allow users to analyse sales performance based on specific categories, brands, or payment methods. This interactivity improves exploratory analysis and helps users derive customized insights according to their requirements.



# 6.Dashboard:




The Amazon Sales Power BI dashboard is designed to provide a comprehensive, interactive, and user-friendly overview of sales performance. The dashboard integrates multiple visual elements on a single page to ensure clarity and ease of interpretation.

At the top of the dashboard, KPI cards present key performance indicators such as Total Sales, Total Profit, and Quantity Sold, offering an immediate summary of overall performance. These KPIs help users quickly assess business health before exploring detailed insights.

Interactive slicers are included for attributes such as Year, Region, and Product Category. These slicers allow users to dynamically filter the entire dashboard and analyze specific segments of data. This feature enhances exploratory analysis and makes the dashboard flexible and user-driven.

Visualizations such as line charts, bar charts, tree maps, and maps are strategically placed to represent trends, comparisons, and distributions. The dashboard follows a clean layout with consistent colour themes and logical data flow, ensuring a smooth analytical experience.

Overall, the dashboard enables users to explore sales patterns, identify growth opportunities, and support data-driven decision-making effectively.

# 7.CONCLUSION

This Power BI project titled **“**Amazon Sales Analysis**”** successfully demonstrates how e-commerce sales data can be transformed into meaningful and actionable insights using business intelligence and data visualization techniques. By analysing transactional data related to orders, quantities, shipping costs, taxes, discounts, brands, categories, and payment methods, the project provides a comprehensive understanding of sales performance and customer purchasing behaviour.

A key strength of this project lies in the effective use of Power Query Editor for data preprocessing. Raw data was cleaned, structured, and transformed to ensure accuracy, consistency, and reliability. Proper handling of data types, missing values, and derived columns enabled smooth data modelling and laid a strong foundation for analysis. This step ensured that all visualizations and calculations produced accurate and reliable results.

The implementation of DAX measures played a crucial role in enabling dynamic and interactive analysis. Measures such as total orders, total quantity, shipping cost, average tax, and discount values allowed real-time calculations that respond to slicers and filters. This dynamic behaviour enhanced dashboard interactivity and provided flexibility for exploratory analysis.

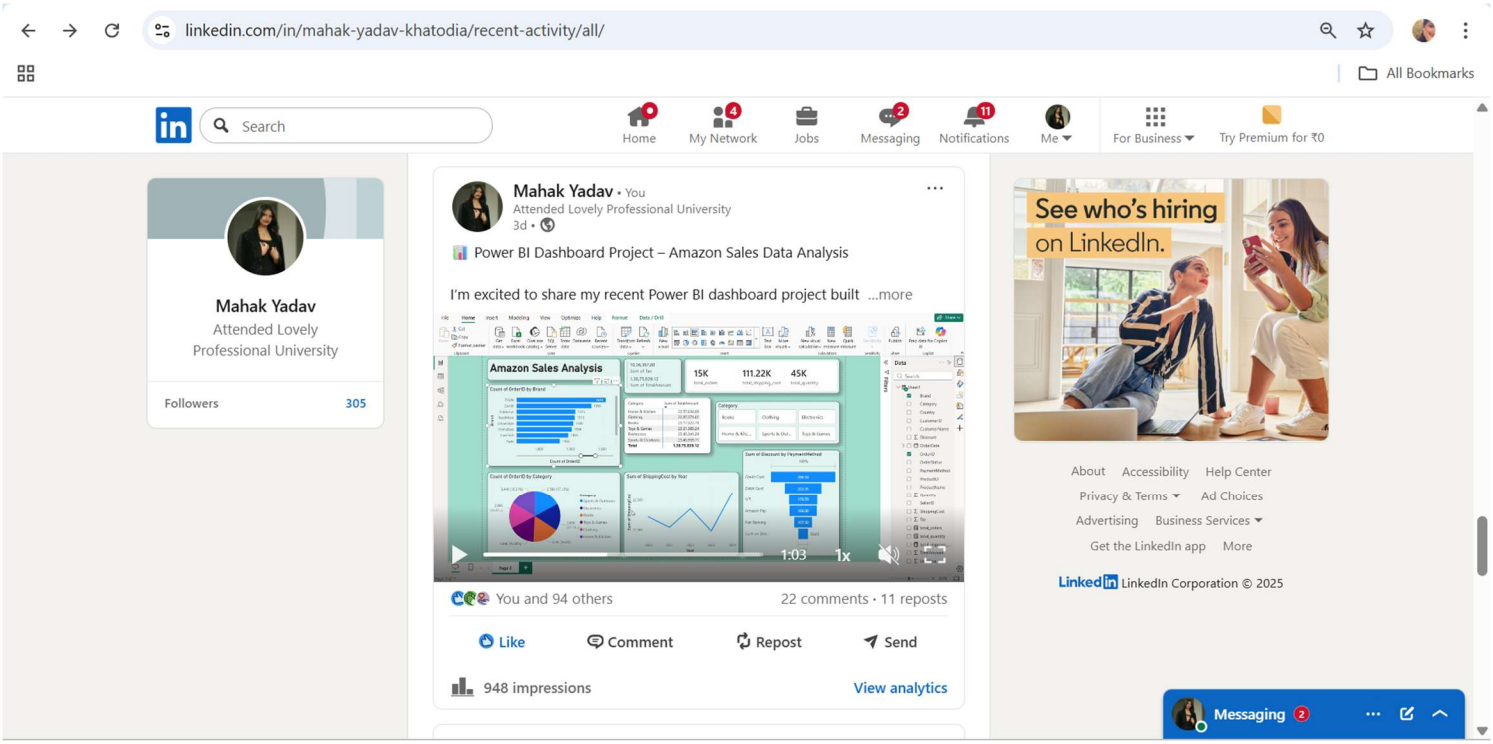
The interactive dashboard was designed with a clear layout and consistent visual theme, integrating KPI cards, bar charts, line charts, pie charts, funnel charts, and tree maps. Each visualization was carefully chosen to align with specific analytical objectives, such as brand-wise order comparison, year-wise shipping cost trends, category-wise quantity and revenue distribution, and discount analysis by payment method. The use of slicers further enhanced user engagement by enabling customized filtering based on category, brand, and payment method.

Overall, the dashboard provides both a high-level summary and detailed insights into Amazon sales performance. It helps identify top-performing brands and categories, understand cost and discount patterns, and evaluate trends over time. This project highlights the importance of data-driven decision-making in the e-commerce industry and demonstrates how Power BI can be effectively used as a business intelligence tool for analysing large datasets.

In conclusion, this project not only fulfills academic requirements but also strengthens practical skills in data preprocessing, DAX calculations, dashboard design, and analytical thinking. The insights derived from this analysis can support strategic business decisions such as optimizing pricing strategies, improving logistics efficiency, and enhancing customer experience. This project serves as a strong example of applying Power BI to real-world.

# 8. LinkedIn Link:

https://www.linkedin.com/in/mahak-yadav-khatodia/




**GitHub Repository:**