Amazon sales analysis

1.Introduction

In today's data-driven e-commerce landscape, understanding sales performance is critical for making informed business decisions. This project focuses on the creation of an Amazon Sales Analysis Dashboard using Microsoft Power BI, designed to transform raw sales data into meaningful insights.

The dashboard provides a comprehensive view of key performance indicators such as total sales, revenue trends, order volume, product performance, and regional sales distribution. By leveraging Power BI's interactive visualizations and data modelling capabilities, the dashboard enables users to monitor performance over time, identify high-performing products, analyse customer purchasing patterns, and uncover areas for improvement.

Overall, this Amazon Sales Analysis Dashboard serves as a powerful decision-support tool that helps stakeholders track business performance efficiently, optimize sales strategies, and drive data-backed growth.

2. Problem Statement

Amazon sales data is often large and complex, making manual analysis difficult and time-consuming. Traditional reporting methods fail to provide quick insights into trends, product performance, and regional sales distribution. There is a need for an interactive and dynamic reporting system that can simplify data analysis and present insights clearly. This project aims to address this challenge by developing an Amazon Sales Analysis Dashboard using Power BI.

Objectives

1. How is our overall sales performance trending over time, and are there identifiable peak periods or seasonal patterns?
2. Which product categories are driving the highest revenue, and which category is top performing.
3. Which individual products are our top performers.
4. How does sales performance vary across different cities, and which regions present opportunities for growth?
5. What is the distribution of orders by city, and how can we optimize logistics and delivery operations accordingly?
6. How are month-to-month sales changing, and what factors are contributing to growth or decline?
7. What payment methods do our customers prefer, and how can this inform payment strategy and customer experience improvements?

3.Existing System

- In the existing system, Amazon sales data is primarily analysed using traditional methods such as spreadsheets and static reports. Data is manually collected and processed using tools like Microsoft Excel, where calculations and summaries are performed using basic formulas and pivot tables. These reports are mostly static and do not provide real-time or interactive analysis.
- Due to the large volume of sales data, the existing system faces several limitations. Manual analysis is time-consuming, prone to errors, and makes it difficult to track trends, compare performance across regions, or identify top-selling products efficiently. Additionally, the lack of dynamic visualizations and drill-down capabilities restricts deeper analysis and timely decision-making.

4.Technology Used

This project utilizes modern business intelligence and data visualization tools to analyse Amazon sales data effectively. The primary technologies used in this project are explained below:

A. Microsoft Power BI

- Microsoft Power BI is the core technology used for developing the Amazon Sales Analysis Dashboard. It is a powerful business intelligence tool that enables data transformation, modelling, visualization, and reporting.
  Power BI was used to:

- Import and integrate sales data from Excel/CSV files
- Clean and transform raw data using Power Query
- Create data models and relationships between tables
- Develop interactive dashboards with charts, cards, slicers, and filters
- Perform sales analysis using dynamic visuals
- Power BI's interactive features allow users to explore data by applying filters such as month and category, enabling quick and efficient analysis.

B. Power Query

- Power Query is a data transformation engine within Power BI used for data cleaning and preparation. It helps convert raw data into a structured and usable format before analysis.
- Power Query was used to:
- Remove duplicate and null values
- Change data types (date, numeric, text)
- Rename columns for better readability
- Filter unnecessary rows and columns
- Prepare clean datasets for analysis
- This step ensures data accuracy and consistency throughout the dashboard.

C. DAX (Data Analysis Expressions)

- DAX is a formula language used in Power BI to create calculated measures and columns. It allows advanced data analysis and helps derive meaningful insights from the dataset.
- DAX was used to:
- Calculate Total Revenue
- Compute Total Orders and Product Count
- Perform Month-over-Month Sales calculations
- Calculate Sales Variance Percentage
- Create dynamic KPIs
- DAX plays a crucial role in enabling real-time calculations based on user interactions.

D. Microsoft Excel

- Microsoft Excel files were used as the primary data source for this project. These files contain structured Amazon sales data such as order details, product information, sales amount, quantity, category, city, and dates.
- Excel was used for:
- Initial data storage
- Basic data validation
- Easy integration with Power BI

E. Data Visualization Components

- Various Power BI visual elements were used to represent data clearly and effectively, including:
- KPI Cards for revenue and orders
- Line charts for monthly sales trends
- Bar and column charts for city-wise and product-wise sales
- Donut charts for category contribution
- Slicers for interactive filtering
- These visuals enhance understanding and support quick decision-making.

F. Operating System

The project was developed on a Windows operating system, which supports Power BI Desktop and related tools efficiently.

G. Hardware Requirements

- Minimum 4 GB RAM (8 GB recommended)
- Intel i3 processor or higher
- Minimum 5 GB free disk space

5. Data Transformation Steps (Power BI – Power Query)

Data transformation was carried out using Power Query Editor to prepare clean, structured, and analysis-ready data. Each transformation step is documented below:

1. Removing Missing Values

Action Performed:

- Identified columns with null or blank values such as Sales, Quantity, Order Date, Product, and City.
- Removed rows containing missing or null values using Remove Rows → Remove Blank Rows and column-level filtering.

Reason:

Missing values can:

- Distort total sales and order counts
- Cause errors in DAX calculations
- Lead to incorrect visual insights

Outcome

- Dataset contains only complete and valid records.
- Improved reliability of aggregations and KPIs.

2. Changing Data Types

Action Performed

- Converted columns to appropriate data types:
    - Order Date → Date
    - Sales → Decimal Number
    - Quantity → Whole Number
    - Order ID → Text
    - City / Product / Category → Text

Reason:

Correct data types are required for:

- Accurate calculations
- Time intelligence functions
- Proper sorting and filtering

Outcome

- No calculation errors
- Better performance and correct visual behaviour

3. Extracting Year, Month, and Day from Order Date

Action Performed

- Created new columns from Order Date:
    - Year (using Date. Year)
    - Month Name (using Date. Month Name)
    - Month Number (using Date. Month)
    - Day (using Date. Day)

Reason

These columns are required for:

- Monthly sales trend analysis
- Year-wise comparison
- Month-to-Month (MoM) analysis

Outcome

- Enabled time-based slicers and visuals
- Improved flexibility in trend analysis

4. Appending Two Years of Data

Action Performed

- Loaded two separate yearly datasets (2024 and 2025).
- Used Append Queries → Append as New to combine both tables into one consolidated table.

Reason

- Allows unified reporting across multiple years
- Enables year-over-year and month-over-month analysis

Outcome

- Single fact table containing both years of sales data
- Simplified data model and DAX calculations

5. Removing Duplicate Records

Action Performed

- Identified duplicate rows using key columns (Order ID, Product, City).
- Removed duplicates using Remove Rows → Remove Duplicates.

Reason

Duplicate data can:

- Inflate sales values
- Increase order counts incorrectly
- Mislead business decisions

Outcome

- Accurate order counts and sales totals
- Clean and consistent dataset

6. Creating Product Dimension Table

Action Performed

- Duplicated the main query.
- Kept only the Product column.
- Removed duplicate product names.
- Renamed the query as Dim_ Product.

Reason

- Avoids repeating product names in the fact table
- Improves data model structure using star schema

Outcome

- Separate Product dimension table
- Reduced data redundancy
- Better filtering and slicing

7. Creating City Dimension Table

Action Performed

- Duplicated the fact table.
- Retained only the City column.
- Removed duplicate city values.
- Renamed the table as Dim_ City.

Reason

- Enables city-wise analysis
- Supports clean one-to-many relationships

Outcome

- Unique city list
- Improved performance and relationship management

8. Segregating Category Data

Action Performed

- Ensured Category column was cleaned and standardized.
- Removed unnecessary spaces and formatting issues.

Reason

- Prevents incorrect grouping in visuals
- Ensures accurate category-wise sales distribution

Outcome

- Clean and consistent category reporting

9. Final Data Validation

Action Performed

- Verified row counts before and after transformations
- Cross-checked total sales and order counts
- Ensured all relationships were correctly mapped

Reason

- To ensure transformations did not impact data accuracy

Outcome

- Final dataset validated and ready for reporting

6. Data Modelling

1. Overview of Data Model

The data model for the Amazon Sales Analysis project is designed using a Star Schema architecture to ensure optimal performance, simplified relationships, and accurate analytical reporting.

The model consists of:

- 1 Fact Table
- 3 Dimension Tables

This structure enables efficient slicing, filtering, and aggregation of sales data across multiple business dimensions such as product, city, and date.

2. Tables Used in the Data Model

2.1 Fact Table – Amazon Sales

Table Name: Amazon Sales
Table Type: Fact Table

This table stores transactional-level sales data and acts as the central table in the data model.

Key Columns:

- Order_ID
- Order_Date
- Product_ID
- City_ID
- Product_Category
- Payment_Method
- Delivery_Status
- Sales Amount
- Quantity

Purpose:

- Stores all measurable business metrics such as sales and orders
- Acts as the base table for all calculations and KPIs

2.2 Product Dimension Table

Table Name: Product
Table Type: Dimension Table

Key Columns:

- Product_ID (Primary Key)
- Product_Name

Purpose:

- Contains unique product details
- Enables product-wise analysis such as top products and category performance
- Avoids duplication of product names in the fact table

2.3 City Dimension Table

Table Name: City
Table Type: Dimension Table

Key Columns:

- City_ID (Primary Key)
- City

Purpose:

- Stores unique city information
- Enables city-wise sales and order analysis
- Improves filtering and drill-down capabilities

2.4 Calendar Dimension Table

Table Name: Calendar
Table Type: Dimension Table

Key Columns:

- Order_Date (Primary Key)
- Day
- Month
- Month Name
- Year
- Order_ID

Purpose:

- Supports time intelligence analysis
- Enables month-wise, year-wise, and trend-based reporting

- Used for Month-to-Month (MoM) and Year-to-Year comparisons

3. Relationships Established

The relationships are created following best practices using one-to-many (1:*) cardinality.

3.1 Product → Amazon Sales

- Relationship: One-to-Many (1:*)
- Join Column: Product_ID
- Direction: Single Direction
- Purpose: Enables product-level sales analysis

3.2 City → Amazon Sales

- Relationship: One-to-Many (1:*)
- Join Column: City_ID
- Direction: Single Direction
- Purpose: Enables city-wise sales and order insights

3.3 Calendar → Amazon Sales

- Relationship: Many-to-Many (*:*)
- Join Column: Order_Date
- Direction: Both Direction
- Purpose: Enables time-based analysis such as trends and MoM comparison

4. Schema Type Used

Star Schema

Why Star Schema?

- Improves report performance
- Simplifies DAX calculations
- Reduces model complexity
- Enables faster filtering and aggregation

Structure:

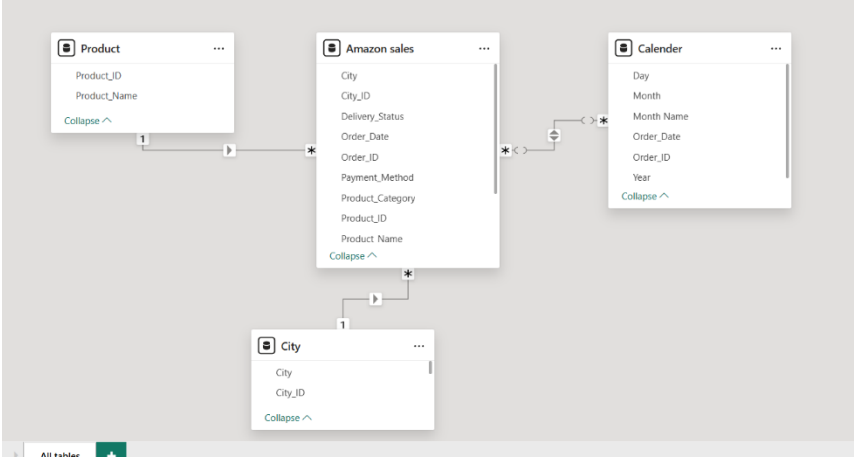
- Central Fact Table: Amazon Sales
- Surrounding Dimension Tables: Product, City, Calendar

5. Data Model Benefits

✓ Reduced data redundancy
✓ Faster report performance
✓ Simplified relationships

✓ Accurate aggregations
✓ Scalable for future data

6.Overview Of Data Modelling



7.DAX Measures

1. Total Sales

Definition:
Total Sales calculates the overall revenue generated by summing the sales amount from all transactions in the dataset.

Purpose:
Used to understand total business revenue and track overall sales performance.

2. Total Orders

Definition:
Total Orders counts the number of unique orders placed by customers using the Order ID.

Purpose:
Helps measure customer activity and demand without counting duplicate orders.

3. Total Quantity Sold

Definition:
Total Quantity Sold calculates the total number of products sold across all orders.

Purpose:
Used to analyse product demand and sales volume.

Measures

- Total Sales = SUM('Amazon sales'[Total_Sales])
- Total Orders = DISTINCTCOUNT('Amazon sales'[Order_ID])
- Total Quantity Sold = sum('Amazon sales'[Quantity])

8.KPI SUMMARY

Key Performance Indicators (KPIs)

The following KPIs were designed to provide a high-level overview of business performance and support data-driven decision-making.

1. Total Sales

Definition:
Represents the total revenue generated from all orders during the selected period.

Purpose:

- Measures overall business performance
- Helps track revenue growth and trends over time

Insight from Dashboard:

- Total sales amount reached $216M, indicating strong overall revenue performance.

2. Total Orders

Definition:
Total number of unique orders placed by customers.

Purpose:

- Indicates customer purchasing activity
- Helps analyse demand and order volume trends

Insight from Dashboard:

- A total of 2K orders were placed during the analysis period.

3. Sold Quantity

Definition:
Total number of products sold across all orders.

Purpose:

- Measures product movement and demand
- Helps identify high-volume sales periods

Insight from Dashboard:

- 7K units were sold, showing consistent product demand.

4. Month-over-Month (MoM) Sales Growth

Definition:
Compares current month sales with previous month sales to measure growth or decline.

Purpose:

- Tracks sales performance changes over time
- Identifies seasonal patterns and sales fluctuations

Insight from Dashboard:

- Sales show noticeable month-to-month variation, highlighting peak and low-performing months.

5. Top Selling Category

Definition:
Category contributing the highest share of total sales.

Purpose:

- Identifies most profitable product categories
- Supports inventory and marketing strategy decisions

Insight from Dashboard:

- Certain categories dominate sales contribution, indicating strong customer preference.

6. Top 10 Products by Sales

Definition:
Top 10 products ranked based on total sales value.

Purpose:

- Identifies best-performing products
- Helps optimize pricing, promotions, and stock levels

Insight from Dashboard:

- Products such as Power Bank, USB Cable, and LED Bulb are among the top revenue generators.

7. City-wise Sales Performance

Definition:
Total sales generated by each city.

Purpose:

- Identifies high-performing and low-performing regions
- Supports regional sales and expansion strategies

Insight from Dashboard:

- Cities like Delhi, Kolkata, and Bangalore contribute significantly to total sales.

9. Orders by City

Definition:
Total number of orders placed in each city.

Purpose:

- Measures customer engagement at a regional level
- Helps compare demand across locations

Insight from Dashboard:

- Order distribution varies by city, with metro cities leading in order count.

10. Orders by Payment Method

Definition:
Distribution of orders across different payment methods.

Purpose:

- Analyses customer payment preferences
- Helps optimize payment options and user experience

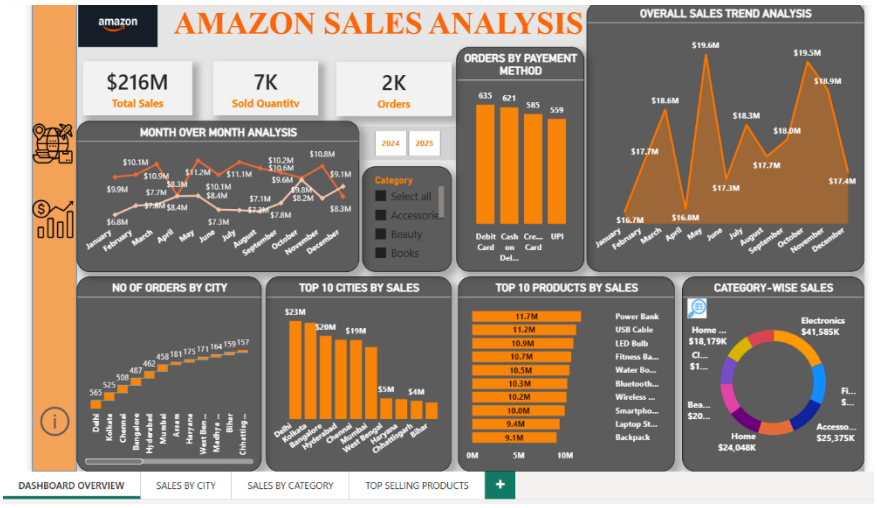
Insight from Dashboard:

- Debit Card and Cash on Delivery are the most commonly used payment methods.

9.Charts Used

Visualizations are an essential part of any Power BI dashboard, as they make complex data easier to understand and analyse. The Amazon Sales Analysis Dashboard uses a variety of charts to provide insights into sales performance, product performance, and regional trends. The charts used in this project are described below:

The overview of dashboard

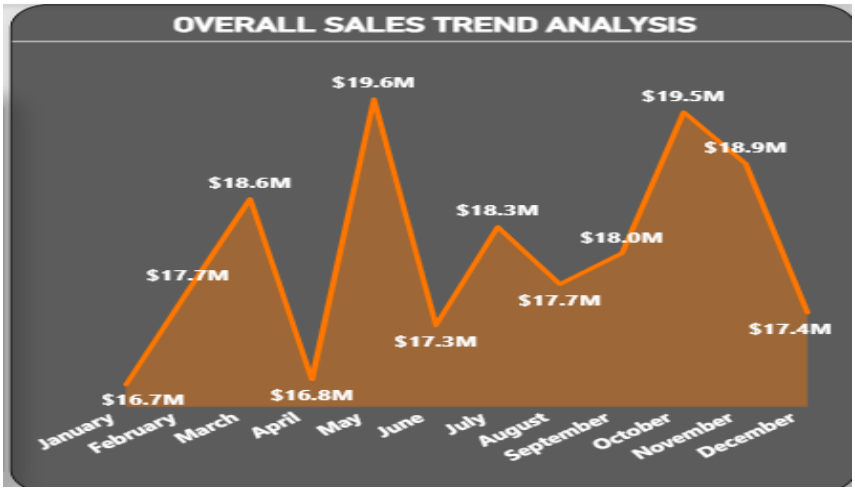


1. KPI Cards

- Purpose: To display key metrics such as Total Revenue, Total Orders, and Number of Products.
- Insight: Provides a quick overview of the business performance at a glance.
- Type: Single-number cards with dynamic values.

2. Area Chart(Overall Sales Trend)

- Purpose: To show sales trends over the months.
- Insight: Helps identify seasonal fluctuations, peak sales periods, and growth trends.
- Type: Area chart with months on the x-axis and revenue on the y-axis.
- Legend: Year



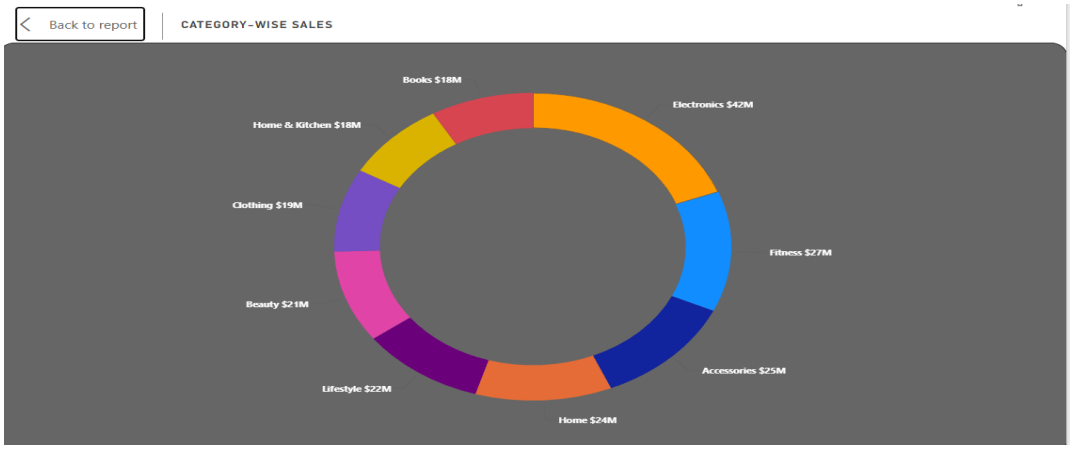
2. Top Selling Category Based on Sales
Chart Type: Donut Chart
Description:
This chart represents the contribution of each product category to total sales, helping identify the top-performing categories.
X-Axis:

- Product Category (Accessories, Beauty, Books, Clothing, Electronics, Fitness)

Y-Axis:

- Total Sales (Percentage / Value)

Insight Provided:
Helps understand which product categories generate the highest revenue.



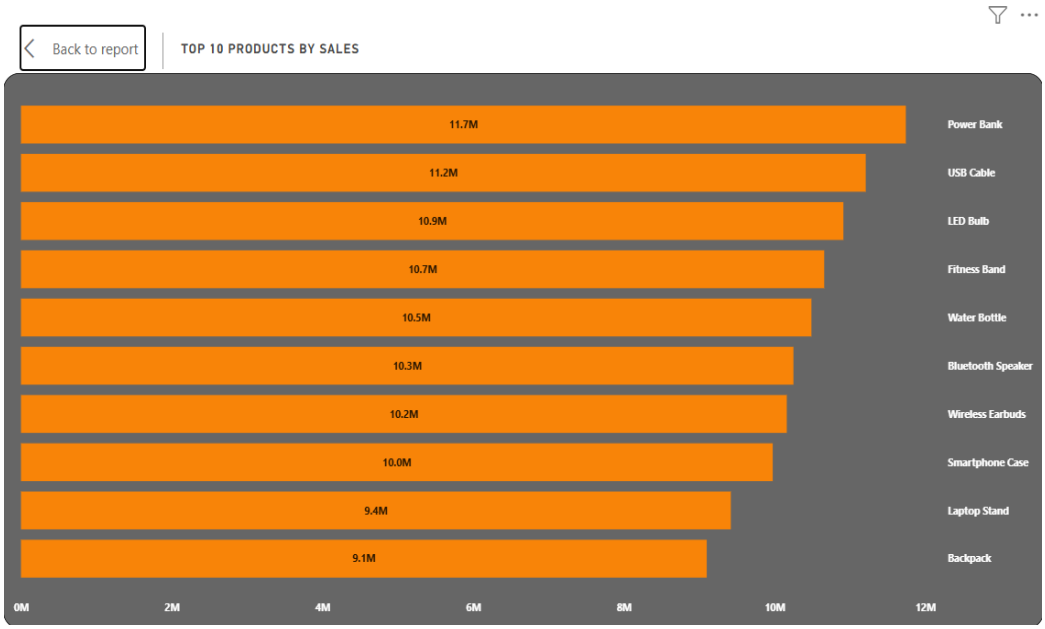
3. Top 10 Products by Sales
Chart Type: Horizontal Bar Chart
Description:
Displays the top 10 products ranked by total sales value.
X-Axis:

- Total Sales Amount

Y-Axis:

- Product Name

Insight Provided:
Identifies best-selling products that drive maximum revenue.



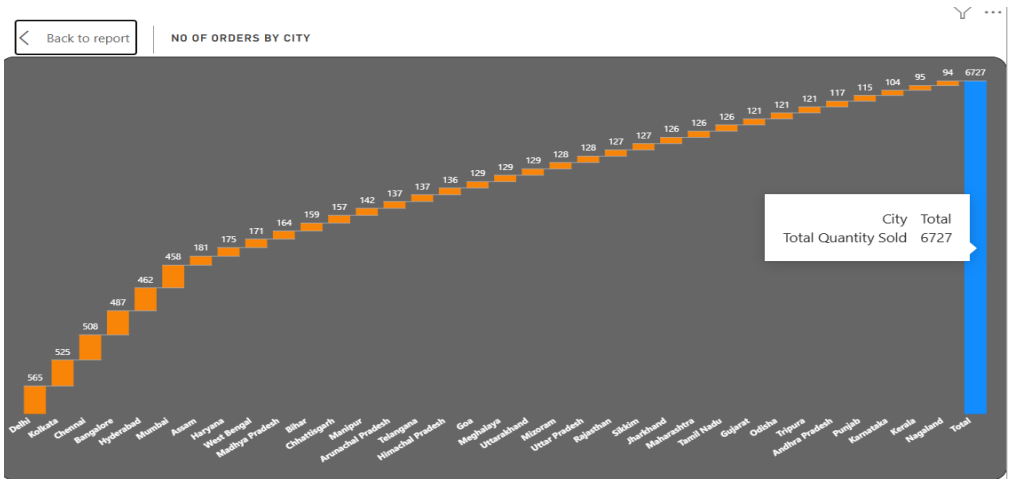
4. City-Wise Sales Trend
Chart Type: Column Chart
Description:
This chart shows total sales generated by each city, enabling geographical sales comparison.
X-Axis:

- City Name

Y-Axis:

- Total Sales Amount

Insight Provided:
Helps identify high-performing and low-performing cities.



5. Orders by Each City
Chart Type: Column Chart

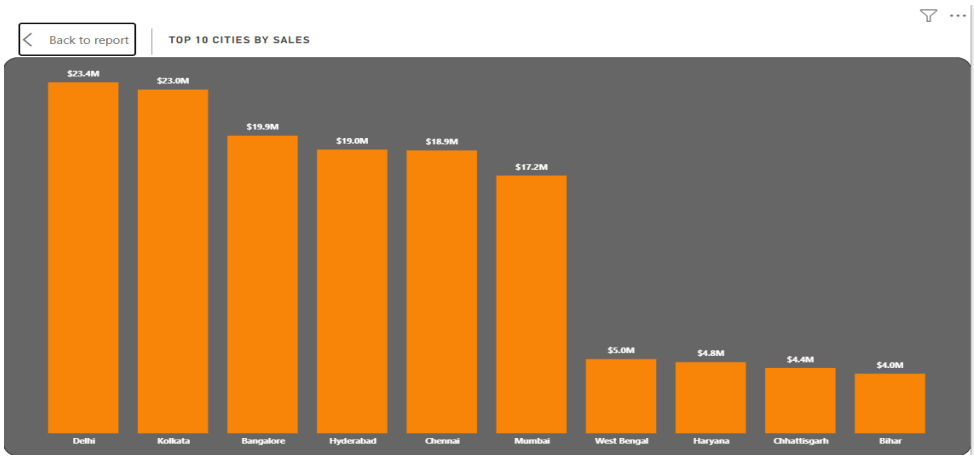
Description:
Displays the total number of orders placed from each city.
X-Axis:

- City Name

Y-Axis:

- Number of Orders

Insight Provided:
Shows customer demand and order distribution across cities.



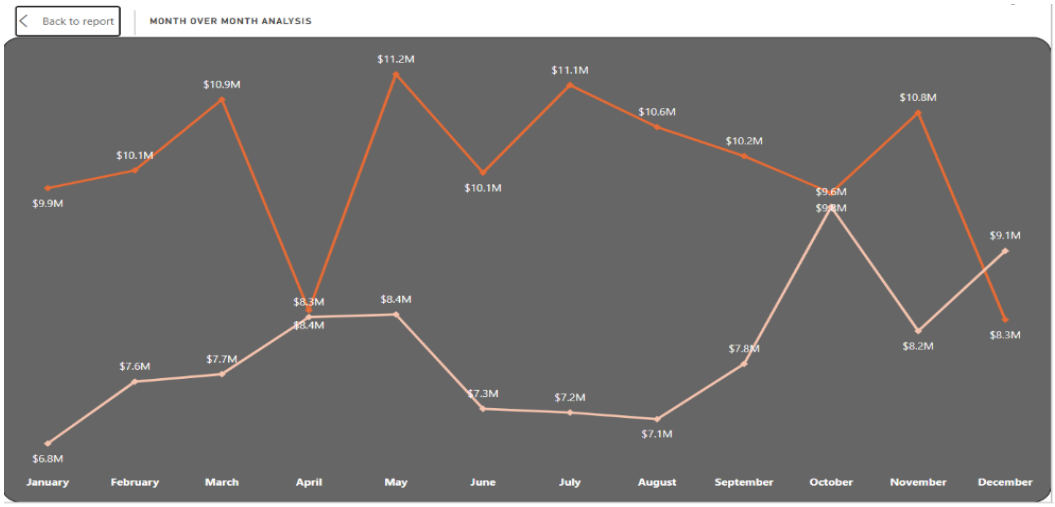
6. Month-to-Month Analysis
Chart Type: Line Chart
Description:
This chart shows the month wise sales for both the years.
X-Axis:

- Month

Primary Y-Axis (Column):

- Total Sales

Legend:Year
Insight Provided:
Helps track month-over-month performance and identify growth or decline trends.



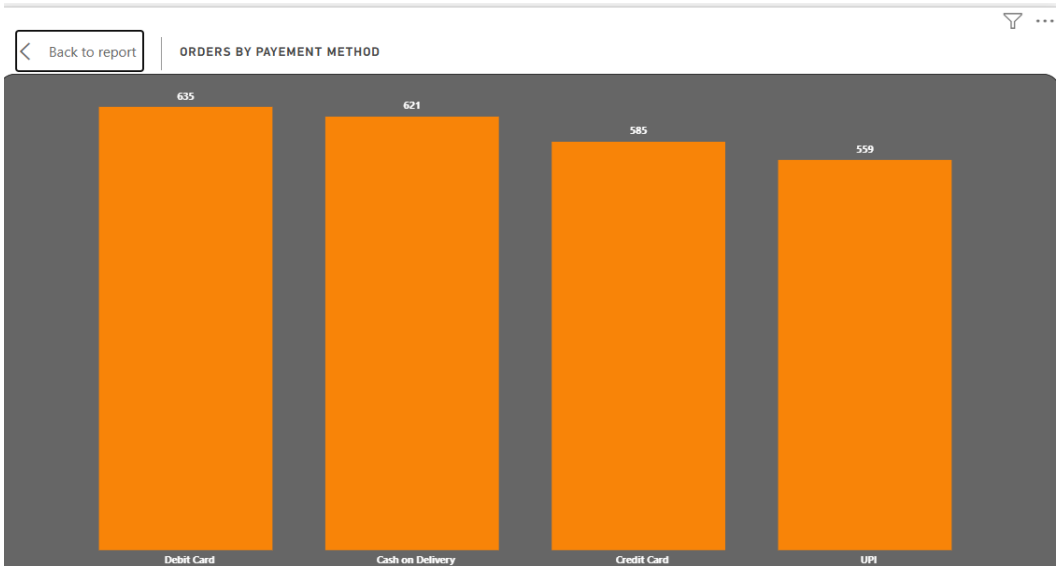
7. Orders Based on Payment Method
Chart Type: Column Chart
Description:
Shows the number of orders placed using different payment methods.
X-Axis:

- Payment Method (Debit Card, Credit Card, Cash on Delivery, UPI)

Y-Axis:

- Number of Orders

Insight Provided:
Helps understand customer payment preferences.



9.Conclusion

- The Amazon Sales Analysis Dashboard developed using Microsoft Power BI provides a comprehensive and interactive platform to monitor and analyse sales performance. By transforming raw sales data into meaningful insights, the dashboard enables stakeholders to track total revenue, order volume, product performance, regional sales, and category-wise contributions efficiently.
- Through dynamic visualizations such as KPI cards, line charts, bar charts, and donut charts, the dashboard offers both high-level summaries and detailed views, allowing users to identify trends, seasonal patterns, and top-performing products quickly. The inclusion of interactive slicers and filters further enhances user experience by enabling customized analysis for specific months, cities, or product categories.
- Overall, this project demonstrates the power of Power BI as a business intelligence tool. It not only simplifies complex data analysis but also supports data-driven decision-making, helping businesses optimize sales strategies, improve operational efficiency, and plan future growth effectively.