# MALL SALES DATA ANALYSIS DASHBOARD

**A PROJECT REPORT**

Submitted by

## DIVAGARAN S (22ADR025)
## SYED MUHAMMAD ABDUL BAKIR (22ADL134)

*For*
*22ADT51 DATA ANALYSIS*

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE**



**KONGU ENGINEERING COLLEGE (Autonomous)**

**PERUNDURAI, ERODE – 638 060**
**NOVEMBER 2024**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE**


**KONGU ENGINEERING COLLEGE (Autonomous)**
**PERUNDURAI, ERODE – 638 060**
**NOVEMBER 2024**


*Department of Artificial Intelligence*


**22ADT51 – Data Analysis Project Report**


Signature of course in-charge                Signature of the HOD


Submitted for the Continuous Assessment viva voce Examination held on _____


**EXAMINER I**                                          **EXAMINER II**

# ABSTRACT

This report provides an in-depth analysis of mall sales data, with the objective of uncovering insights into customer behaviors, product performance, and overall sales patterns within a retail environment. The dataset includes a range of attributes for each transaction, such as the transaction date, customer demographics (including gender and age), product categories, price per unit, and total purchase amount. By analyzing these attributes, this study seeks to identify key factors that influence sales trends and customer purchasing decisions, offering data-driven insights that can support strategic decision-making for retail management. The analysis begins with a data preparation phase, where missing values and anomalies are addressed to ensure the accuracy and reliability of results. Following this, an exploratory data analysis (EDA) is conducted, focusing on several dimensions of the data: temporal sales trends, demographic influences on purchasing behavior, product category performance, and price sensitivity. Temporal analysis highlights peak sales periods, revealing seasonal trends and periods of high demand. Demographic analysis explores the spending habits of various customer segments, identifying age and gender groups with higher average purchase amounts and frequent buying patterns. Furthermore, the product category analysis ranks categories by popularity and revenue contribution, providing insights into customer preferences and potential areas for inventory optimization. Advanced data visualization techniques are used to present findings clearly and effectively, enabling easy interpretation of trends and correlations. Key insights reveal patterns such as the impact of pricing on total sales, the preferences of different demographic groups, and the influence of time (e.g., weekdays versus weekends) on purchasing behavior. These insights lead to actionable recommendations for retail managers, such as implementing targeted marketing campaigns to attract high-value customer segments, adjusting product assortments to meet demand more effectively, and planning promotional activities around peak sales periods to maximize revenue.

In conclusion, this report serves as a practical resource for understanding the dynamics of mall sales, offering specific, data-backed strategies to enhance customer engagement, optimize inventory, and drive revenue growth in a competitive retail environment. The findings support a more targeted and customer-centric approach to retail management, with a focus on improving both sales performance and customer satisfaction.

# TABLE OF CONTENTS

# CHAPTER 1
## INTRODUCTION

## 1.1   INTRODUCTION

❖ In the competitive retail landscape, understanding customer behaviors, purchasing patterns, and product performance is essential for strategic decision-making. Retail malls, in particular, benefit from insights into sales trends and customer demographics, enabling them to make data-driven decisions that enhance customer satisfaction and optimize revenue. This report focuses on analyzing mall sales data to uncover patterns that can inform targeted marketing strategies, product assortment decisions, and inventory management.

❖ The dataset used for this analysis includes key attributes from mall transaction records: the transaction date, customer demographics (such as gender and age), product categories, unit prices, and total sales amounts. By examining these attributes, this report aims to identify important trends and insights that can help the mall better serve its customers and improve overall business performance. For instance, understanding demographic influences on purchasing behavior allows the mall to tailor its marketing strategies to specific age and gender groups, while analyzing sales by product category helps in optimizing inventory based on demand. Additionally, temporal trends (such as sales peaks and lows) can guide promotional timing and staffing requirements.

❖ This report is structured as follows: an introduction to the problem and objectives, a discussion of the data collection and preprocessing steps, an exploration of the main challenges faced by the mall, and specific business objectives. Each section aims to provide a clear understanding of the data analysis process, leading to actionable recommendations that align with the mall's goal to maximize revenue and improve customer satisfaction. Through this analysis, we aim to empower mall management with the insights needed to make informed decisions in the retail environment.

## 1.2 DATA COLLECTION

The data used in this analysis consists of transaction records from a mall, providing a rich source of information to understand customer preferences, purchasing patterns, and product performance. The dataset includes several key attributes that support in-depth exploration of customer and sales metrics. These attributes are:

- **Date**: The date of each transaction, which allows us to examine trends over time. Daily, weekly, and monthly sales analyses can reveal important insights, such as peak shopping days or seasonal trends that impact revenue.
- **Customer Information**: Demographics, including gender and age, provide valuable information for customer segmentation. Understanding how different demographic groups behave allows the mall to create tailored marketing campaigns and product offerings.
- **Product Category**: This attribute specifies the type of product purchased, enabling us to determine which categories drive the most revenue and are most popular among customers.
- **Price per Unit and Total Amount**: These financial attributes provide insights into the average spend per customer and allow for analyses on how pricing impacts total revenue.

Before analysis, the data was prepared to ensure its accuracy and consistency. Data pre-processing involved handling missing values, identifying and treating outliers, and ensuring that all data fields were in compatible formats. Missing data was managed by imputing values based on reasonable estimates or by removing incomplete records, depending on the context. Outliers, such as unusually high or low purchase amounts, were reviewed to determine if they reflected valid transactions or data entry errors. Once these steps were completed, the dataset was ready for analysis.

Data preparation was crucial to improve the reliability and validity of the analysis. A clean dataset allowed us to conduct meaningful analyses, uncover genuine insights, and provide accurate recommendations for mall management. This careful approach ensures that findings are based on a solid, trustworthy foundation of data.

## 1.3 PROBLEM STATEMENT

This report addresses several key questions that are essential to overcoming these challenges. These include:

- **Who are the mall's most valuable customers?** Identifying customer groups that contribute most to revenue can help direct targeted marketing efforts.
- **Which products and categories are performing best?** Understanding product demand at the category level allows for more strategic inventory management and assortment planning.
- **How do sales vary by time and demographics?** Examining how factors such as age, gender, and transaction timing influence purchasing behavior provides insights that can improve the timing of promotions and customer engagement.

## 1.4 BUSINESS OBJECTIVE

The primary business objective of this report is to help mall management make data-driven decisions that boost revenue and enhance customer engagement. By analyzing sales data, we aim to:

1. **Identify High-Value Customer Segments**: Use demographic insights (age, gender) to target customer groups with higher spending potential.
2. **Optimize Product Offerings**: Determine which product categories drive the most revenue to improve inventory management and ensure high-demand items are well-stocked.
3. **Improve Promotional Timing**: Leverage peak sales periods and seasonal patterns to plan more effective promotions and attract more customers.
4. **Enhance Customer Satisfaction**: Align the shopping experience and product selection with customer preferences, fostering loyalty and repeat visits.

## ABOUT POWER BI:

1. Power BI, a business analytics tool from Microsoft, offers interactive data visualization BI features that let users see and share information throughout their organization. By using data interactively and visualizing it, Power BI offers insight data. Utilize the data models to produce reports and visuals.

2. A business user can use it to centralize measurements and significant company goals so they can track their progress. In addition, it promotes cooperation and interaction on the site while being simple to use and aesthetically pleasing.

3. In the modern world people are very busy with their duties so they don't have enough time to look into the entertainment especially looking into cricket score and analysis takes more time so it is necessary to summarize all the events that happened in a cricket needed to be visualized attractively and understand to everyone easily.

4. So, for this purpose it is necessary to prepare dashboards. Power BI is a tool that helps users easily visualize dynamic and interactive Reports/Dashboards by utilizing its Business Intelligence Capabilities.

5. Power BI is a tool that makes decision-making easier as it offers a wide range of interactive visualizations along with Business Intelligence Capabilities.

# CHAPTER 2

## DATA PREPARATION AND MODELING

### 2.1 DATA CLEANING

Data cleaning is the process of removing errors from the data by filling in missing values, smearing noisy data, analyzing and removing outliers, and smoothing noisy data. Data at various degrees of detail may occasionally diverge from what is needed Missing Values – Appropriate values are substituted for missing values. The strategies listed below

✔ When a tuple contains many attributes with empty values, it is

✔ Disregarded.

✔ For the missing value, the values are manually filled in.

✔ The values may be filled with the same global constant.

✔ The attribute mean can replace the values that are absent.

✔ The most likely value can be used to fill in the blanks.

### 2.2 DATA TRANSFORMATION

1. The process of changing data from one format or structure to another is known as data transformation. It is a crucial component of the majority of data management and integration jobs, including

2. application **integration**, data **wrangling**, data **warehousing**, and data **integration**. Depending on the required modifications to the data between the source (initial data) and the destination (final data), data transformation can

be straightforward or difficult. The process of data transformation often involves both manual and automated procedures.

3. Depending on the format, structure, complexity, and amount of the data being changed, a broad range of tools and technologies may be employed. For decades, corporations have benefited greatly from using conventional data transformation techniques.

4. Since the development of the various tools and technologies (data profiling, data visualization, data purification, data integration, etc.), most (if not all) businesses now transform massive volumes of data that feed internal and external applications, data warehouses, and other data repositories.

5. So, Data Transformation is a required process in order to preprocess the loaded data set as per our requirement and apply those changes for future use. It is while Data Analysis and creating DAX functions of those relations respectively.

6. Since the dataset was already preprocessed and small, we didn't have to do much preprocessing.

## 2.3. DATA MODELING

Data modeling is one of the aspects used in BI tools to establish relationships between various data sources. When using several data sources, you can construct engaging data visualizations by defining the relationships between them.

It can create unique calculations on the already-existing tables using the modeling capability, and these columns can then be easily displayed in Power BI visualizations. This enables companies to create new measures and perform unique calculations for them.

Data Modeling is used to create relationships among the different tables in order to access the data of different tables to visualize them. There are four types of relations that we can create as,

❖ One to One relationship

❖ One to Many relationship

❖ Many to One relationship

❖ Many to Many relationship

# CHAPTER 3

## DATA ANALYSIS AND INTERPRETATION

### 3.1 DATA ANALYSIS

To turn raw data into insightful information, data analysis is the process of analyzing, manipulating, and monitoring. Making the necessary decisions for a business or company's growth is made easier with the use of data insights. Deep data analysis is crucial if we want to manage a firm that is data-driven. Then it is needed to find learning different Power BI data analysis approaches fascinating and useful.

Data analysis includes the following results

- Used to create various charts from Power Bi visuals

- Select data from various tables, analyze it and convert it into visuals.

- From the analyzed result infer the result or final solution.

# CHARTS

## 1. What are the overall sales trends over time?
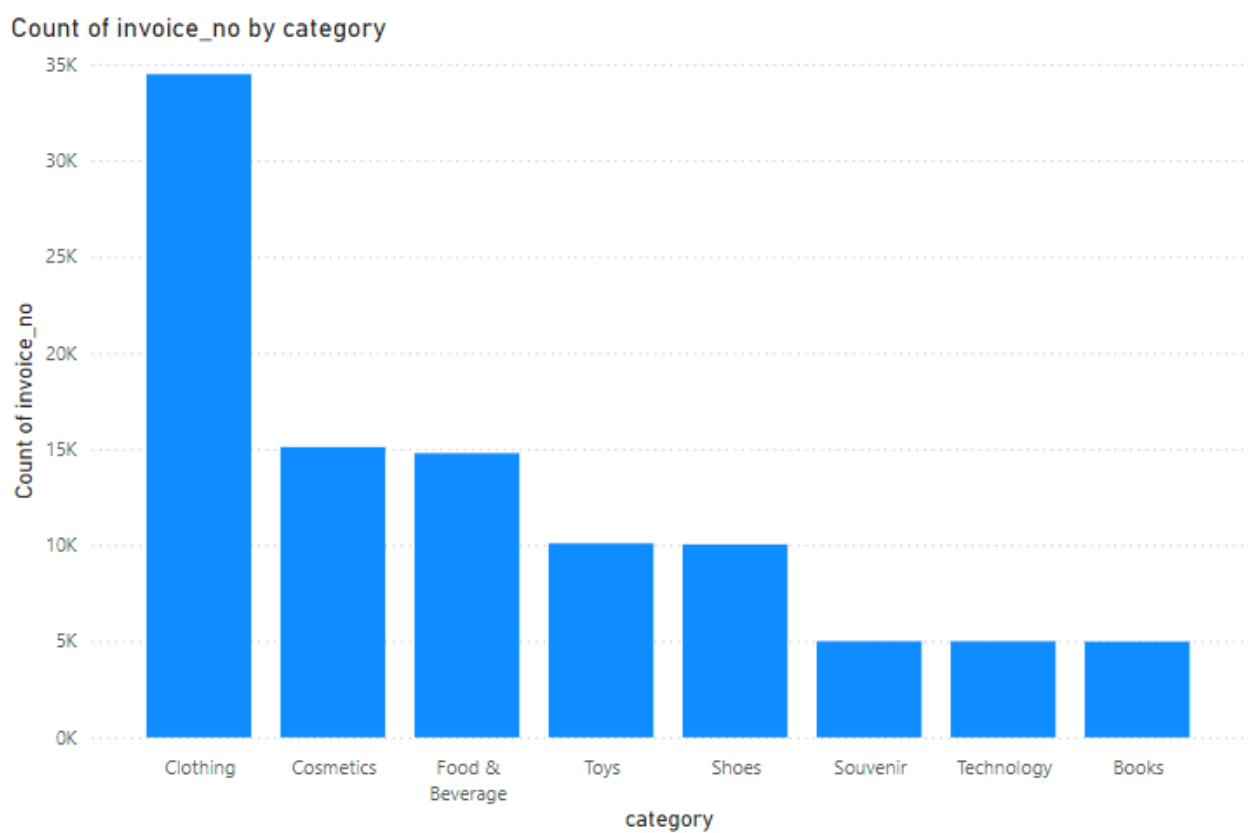
- **Procedure**:
  1. Create a line chart.
  2. Drag the **Date** field to the X-axis and **Total Amount** to the Y-axis.
  3. Set the date hierarchy (day, week, month, year) as needed to identify peak periods.



Count of invoice_no by month
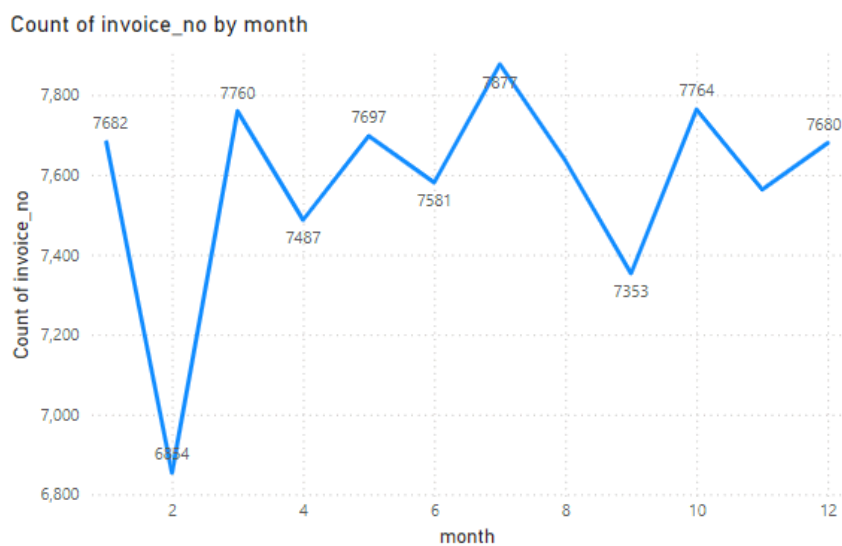
**2. Which product categories generate the highest revenue?**

- **Procedure**:

    1. Use a bar or column chart.

    2. Place **Product Category** on the X-axis and **Total Amount** on the Y-axis.

    3. Sort by **Total Amount** in descending order to see top-performing categories.
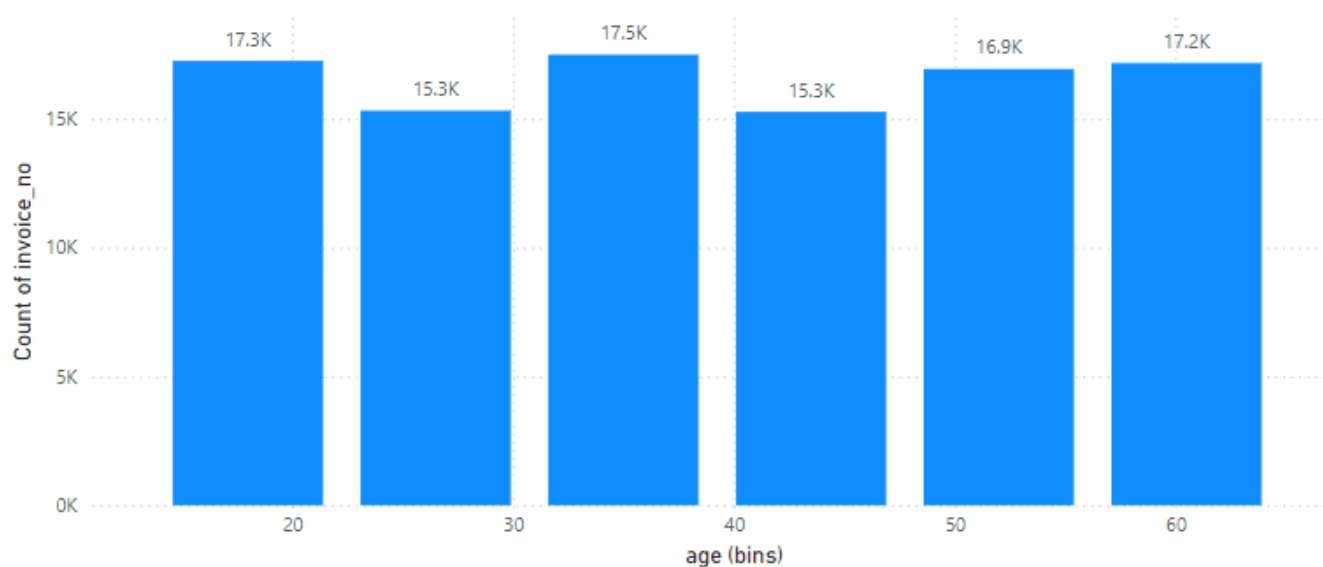
**Count of invoice_no by category**

**3. How does customer spending vary by age group?**

● **Procedure**:

1. Create a clustered column chart.

2. Group the **Age** field into age ranges (e.g., 20, 30) using Power BI's binning feature.

3. Add **Total Amount** as the Y-axis to analyze spending by age group.



Count of invoice_no by age (bins)

## 4. What are the gender-based purchasing patterns?

- **Procedure**:
    1. Use a bar chart with **Gender** on the X-axis and **Total Amount** on the Y-axis.
    2. Add **Product Category** as a legend to see category preferences by gender.
    3. Analyze total sales or category preference differences.



Count of invoice_no by gender and category
category ● Books ● Clothing ● Cosmetics ● Food & Beverage ● Shoes ● Souvenir ● Technology ● Toys
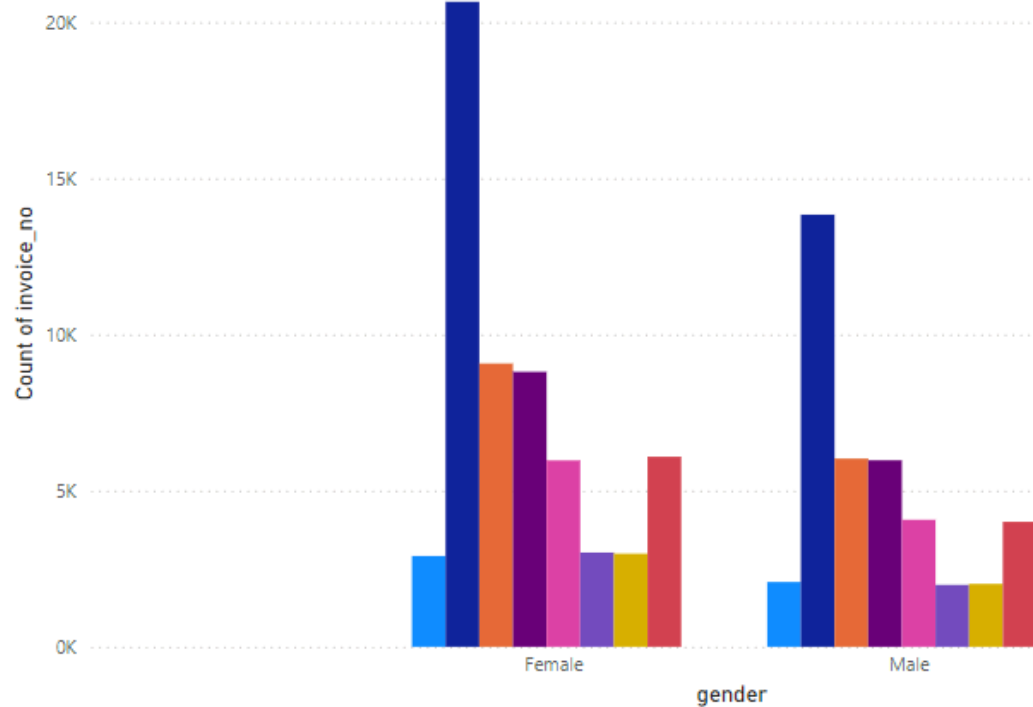
**5. What is the average spend per customer?**

- **Procedure**:

  1. Create a measure for average spend: <span style="color:green">Average Spend =
     AVERAGE('Sales'[Total Amount])</span>.

  2. Display this as a card visualization to see the overall average spend per
     transaction.
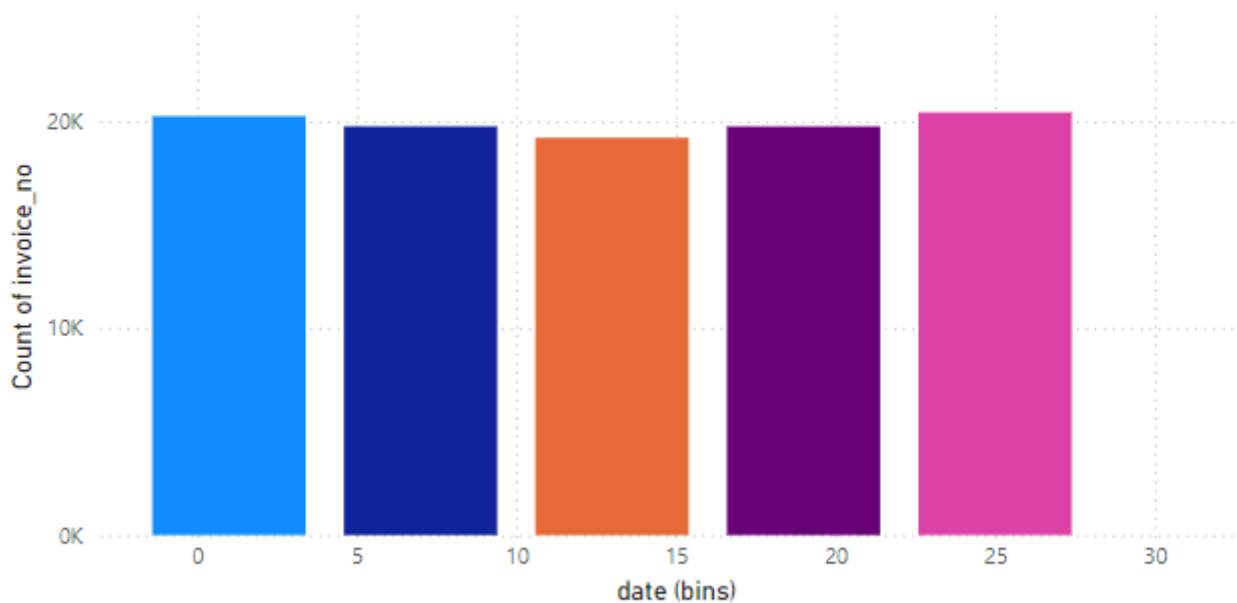
<div align="center">

## 2.53K
AverageSpend

</div>

**6. Which days of the week or times of day see the highest sales?**

- **Procedure**:

  1. Create a bar chart with **Day of the Week** (extracted from **Date**) on the X-axis and **Total Amount** on the Y-axis.

  2. For time-based analysis, extract hour information from **Date** and use it on the X-axis in a similar way.

### Count of invoice_no by date (bins) and date (bins)
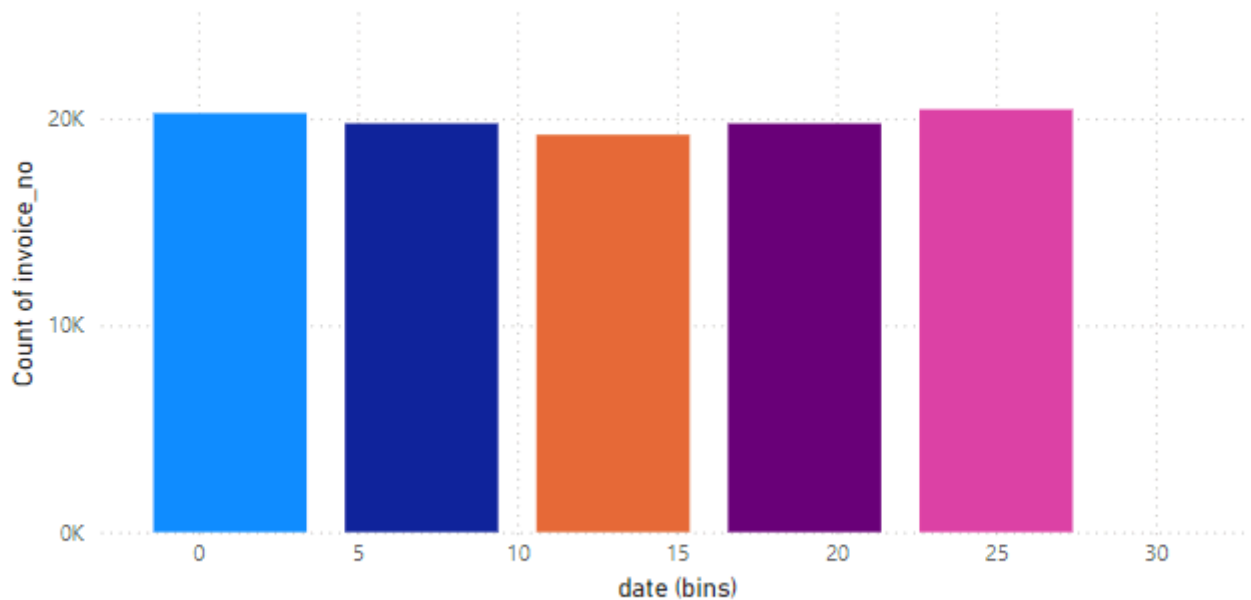
date (bins) ● 1 ● 7 ● 13 ● 19 ● 25

## 7. Are there any seasonal trends in sales?

- **Procedure**:
    1. Use a line chart.
    2. Set **Date** on the X-axis with **Total Amount** on the Y-axis.
    3. Group the date field by month or quarter to identify any seasonal peaks.

**Count of invoice_no by date (bins) and date (bins)**

date (bins) ●1 ●7 ●13 ●19 ●25

**8. What is the correlation between age and product category preferences?**

- **Procedure**:

  1. Create a stacked bar chart with **Product Category** on the X-axis and **Total Amount** on the Y-axis.

  2. Use **Age Group** as a legend to show category preference by age.

**Count of invoice_no by age (bins) and category**

category ● Books ● Clothing ● Cosmetics ● Food & Beverage ● Shoes ● Souvenir ● Technology ● Toys
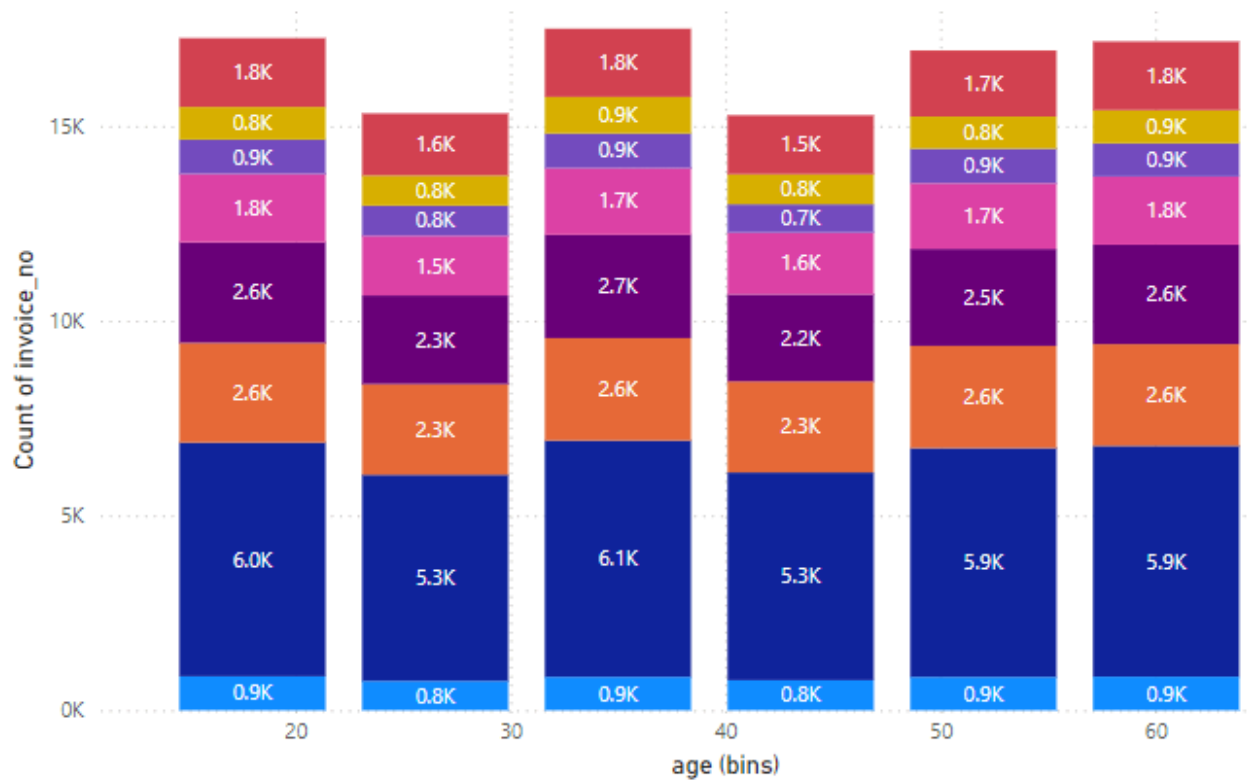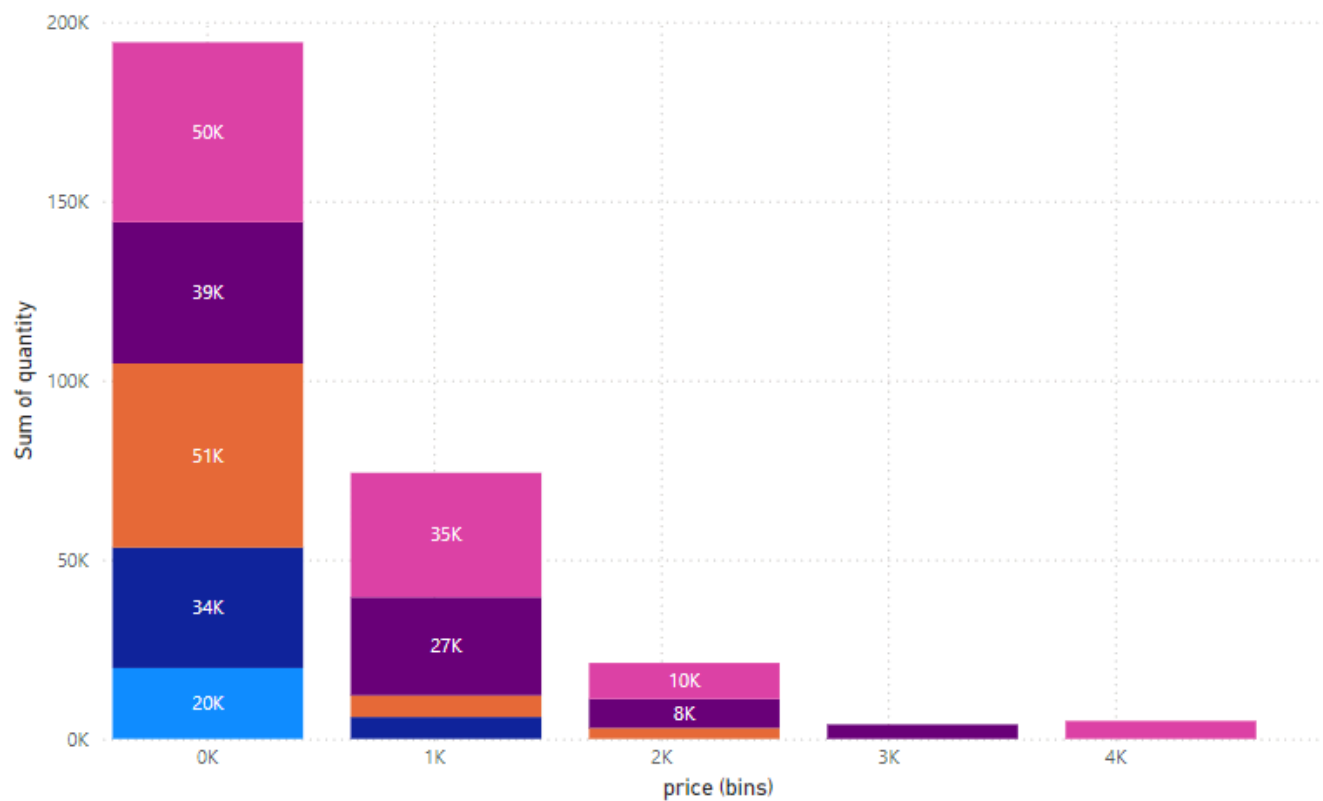
**9. How does price per unit affect the quantity purchased or total sales?**

- **Procedure**:

    1. Create a scatter plot.

    2. Set **Price per Unit** on the X-axis and **Total Amount** or **Quantity** on the Y-axis.

    3. Analyze the trend to see if higher or lower prices affect sales volume.

Sum of quantity by price (bins) and quantity
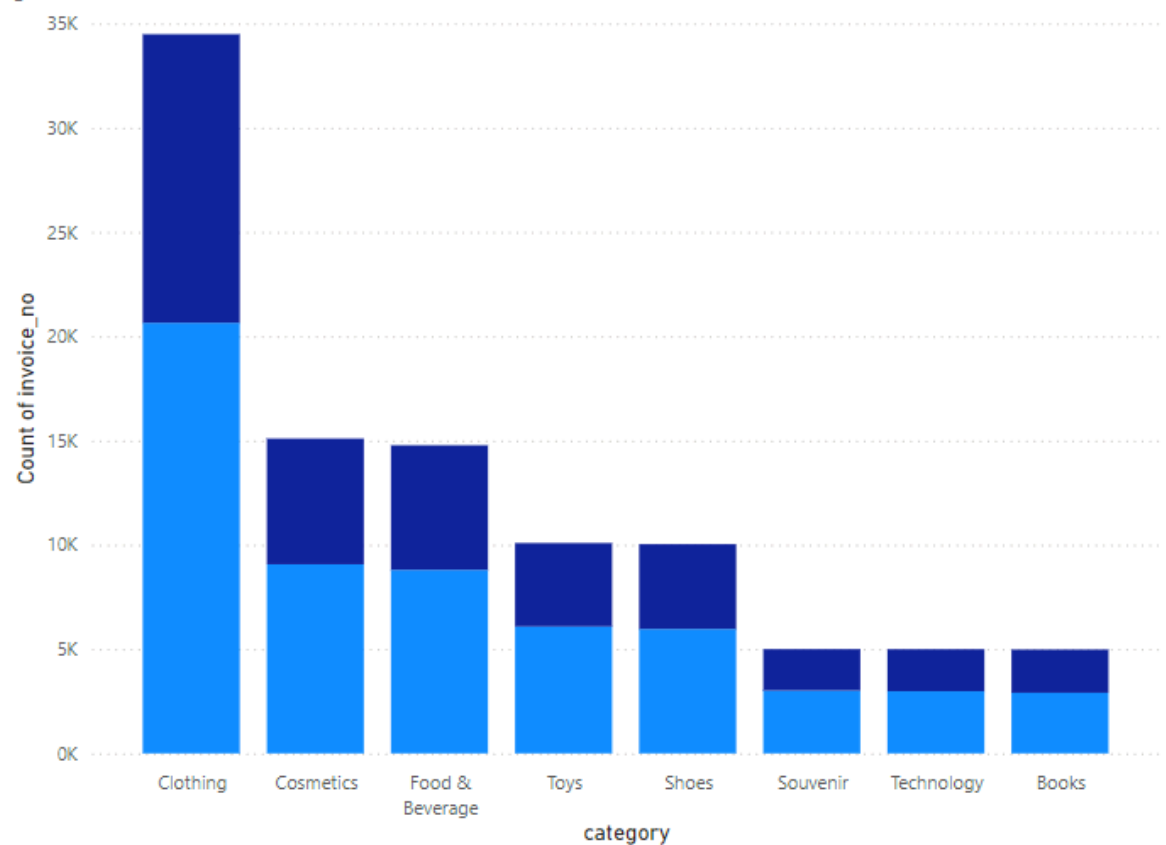
quantity ●1 ●2 ●3 ●4 ●5

## 10. Which product categories are most popular among different gender groups?

- **Procedure**:
    1. Use a stacked bar chart with **Product Category** on the X-axis and **Total Amount** on the Y-axis.
    2. Set **Gender** as the legend to show gender-based preferences within each product category.

**Count of invoice_no by category and gender**
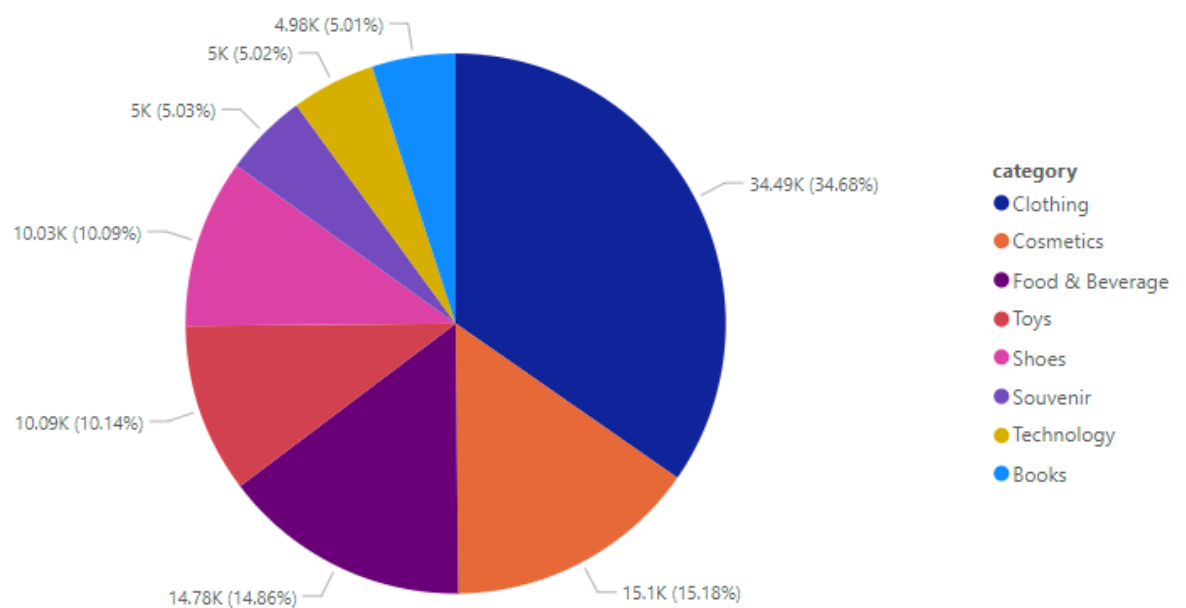
gender ● Female ● Male

**11. What is the percentage contribution of each product category to overall revenue?**

- **Procedure**:
    1. Create a pie chart or a stacked bar chart.
    2. Set **Product Category** as the category and **Total Amount** as the value.
    3. Use **Show Data Labels** to display percentage values for each category.

Count of invoice_no by category

**12. How many unique customers purchase from each product category?**

- **Procedure**:

    1. Create a measure: `Unique Customers =`
       `DISTINCTCOUNT('Sales'[Customer ID])`.

    2. Use a bar chart with **Product Category** on the X-axis and **Unique Customers** as the value.

<div align="center">
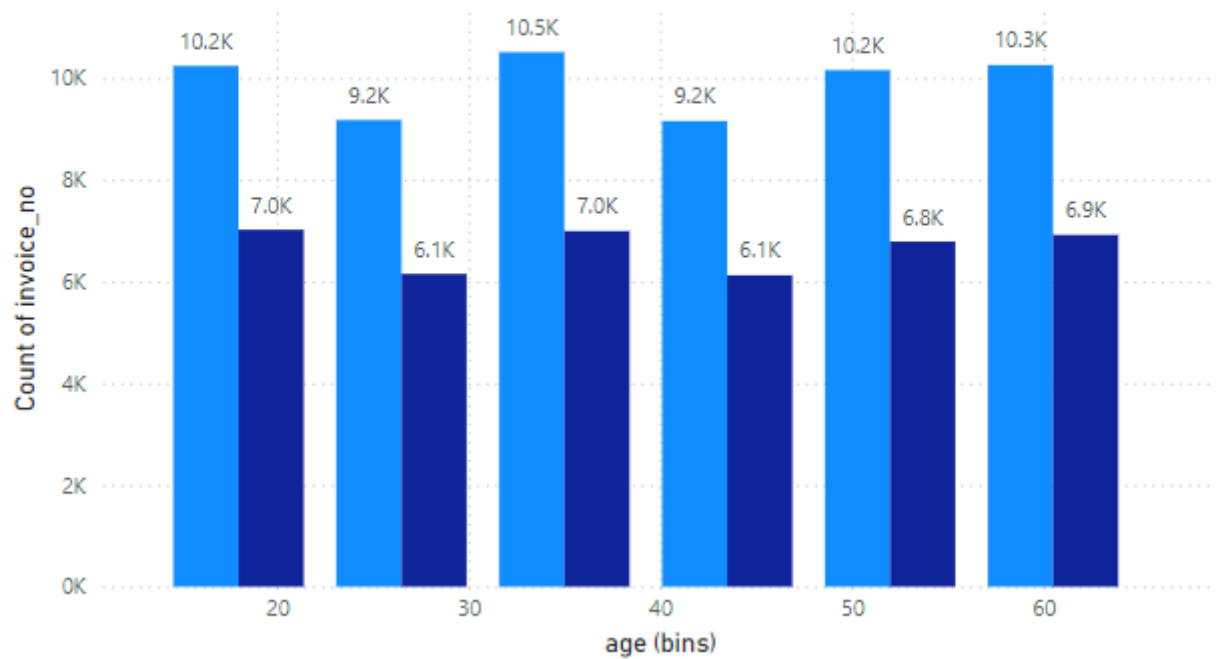
## 99K

Unique Customers

</div>

**13. What is the distribution of sales by customer demographics (age and gender)?**

- **Procedure**:
  1. Use a clustered bar chart with **Gender** and **Age Group** on the X-axis, and **Total Amount** as the Y-axis.
  2. This chart will show the distribution of sales across different age and gender groups.



Count of invoice_no by age (bins) and gender

**14. How does payment method influence sales?**

- **Procedure**:
    1. Use a line or bar chart, with **Payment Method** on the X-axis and **Total Amount** or **Quantity** as the Y-axis.
    2. Analyze if different payment methods correlate with increased sales.

Count of invoice_no by payment_method



20.08K (20.19%)

44.45K (44.69%)

34.93K (35.12%)

payment_method
- Cash
- Credit Card
- Debit Card
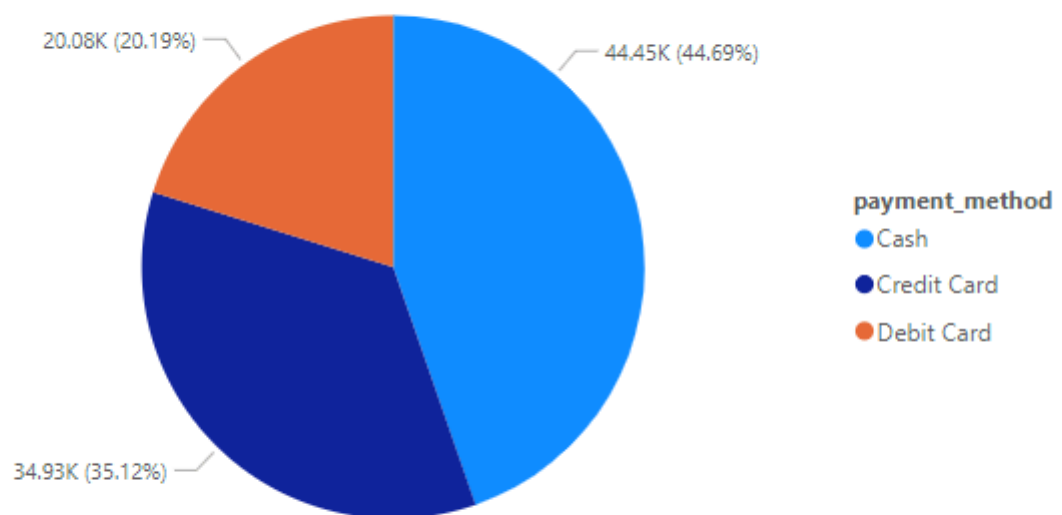
## 3.2 PUBLISHING DASHBOARD

❖ Often referred to as a canvas, a Power BI dashboard is a single page that employs visuals to convey a story. A well-designed dashboard only includes the key components of the tale because it is only one page long. The dashboard's tiles—the visuals you see there—are placed there by report creators.

❖ The report page where the visualization was made is often the page you land on after picking a tile. A dashboard's visuals are derived from reports, and each report is built using a single dataset. A dashboard may really be thought of as a portal to the underlying reports and statistics.

❖ Then it may get the report that was used to produce a visualization by selecting Dashboards as an excellent method to keep an eye on your company, search for solutions, and quickly view all of your most crucial indicators.

❖ A dashboard's visualizations might be drawn from a single underlying dataset or several, as well as a single underlying report or many.

❖ Regardless of where the data is stored, a dashboard may mix on-premises and cloud data to provide a consolidated picture. A dashboard is interactive, and the tiles refresh as the underlying data changes. It is more than simply a lovely picture.

**Link for dashboard**

https://app.powerbi.com/groups/me/reports/4f086f24-b75b-4efb-ad21-d143027d
bebb/dde567f36c2dbe5a43ad?experience=power-bi

## Creating a Dashboard

**STEP 1**

1. Open Power BI serviced in a web browser.

2. From that interface click on get data at the left bottom.

3. Select import data from device or local disk.

4. Then import the created Power BI file to the website.
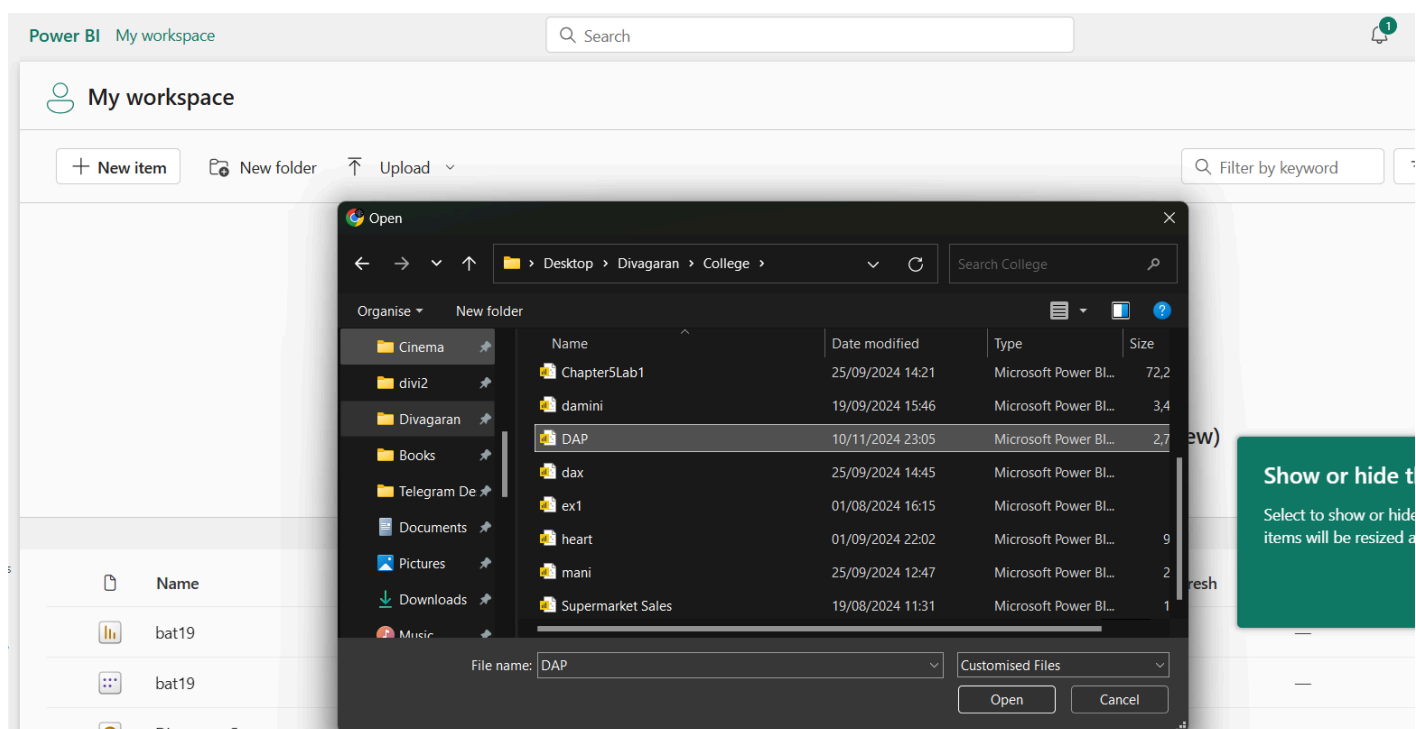
## 3.3 INFERENCES

### 1. What are the overall sales trends over time?

- The Sales are highest during the 7th month.
- The Sales are lowest during the 2nd month.

### 2. Which product categories generate the highest revenue?

- Clothing generates the highest revenue.

### 3. How does customer spending vary by age group?

- Ages 35-40 are the highest spenders.
- Ages 43-52 are the lowest spenders.

### 4. What are the gender-based purchasing patterns?

- Women predominantly buy Clothing.
- Men predominantly buy Clothing.

### 5. What is the average spend per customer?

- The average spend per customer is 2,530.

### 6. Which days of the week or times of day see the highest sales?

- The first week of the month sees the highest sales.

### 7. Are there any seasonal trends in sales?

- The first and last weeks of the month sees the highest sales.
- The middle of the month has lower sales comparatively.

### 8. What is the correlation between age and product category preferences?

- All ages prefer clothing the most.
- Book sales increase with age.

**9. How does price per unit affect the quantity purchased or total sales?**

- Lower price per unit has higher sales rate.

**10. Which product categories are most popular among different gender groups?**

- Clothing is the most popular between both Male and Female customers.

**11. What is the percentage contribution of each product category to overall revenue?**

- Clothing contributes the most.
- Books contribute the least.

**12. How many unique customers purchase from each product category?**

- There are around 1 lakh (or hundred thousand) unique customers.

**13. What is the distribution of sales by customer demographics (age and gender)?**

- Females spend the most.
- Males spend less than females.

**14. How does payment method influence sales?**

- Most customers pay through cash.
- But cashless payment is preferred by most.

# CHAPTER 4

## CONCLUSION AND FUTURE WORK

The Mall Sales Data Analysis has delivered in-depth insights into the various factors influencing customer behaviors, purchasing patterns, and overall sales performance. This analysis highlighted the importance of understanding customer demographics, product preferences, and time-based sales patterns for making informed business decisions in a competitive retail environment. By leveraging this data, mall management can adopt a more strategic approach to enhancing customer experience, optimizing inventory, and boosting revenue.

One of the most significant findings from this analysis is the impact of customer demographics on spending behaviors. Analysis of age and gender data revealed distinct patterns; certain age groups and genders were more inclined toward specific product categories, indicating opportunities for targeted marketing. By focusing on high-value demographics, the mall can design more personalized marketing campaigns and promotions, creating a shopping experience that resonates more effectively with different customer groups.

In terms of product performance, the analysis identified top-performing categories that drive the majority of sales revenue. These insights allow the mall to optimize inventory management, ensuring that popular products are always available and reducing investment in underperforming categories. This focus on product demand aligns with customer preferences and prevents stockouts of popular items, which can improve overall customer satisfaction and loyalty. Additionally, insights into price sensitivity and purchasing behaviors can support better pricing strategies, such as identifying the most effective discount levels that attract customers without eroding profit margins.

Temporal analysis revealed seasonal and time-based trends that provide a strategic edge in planning sales events and promotions. Peak sales periods, identified through day, week, and month-based analyses, allow management to time promotions to coincide with high-traffic times, such as weekends, holidays, or specific months. This knowledge can guide decisions

around staffing, promotional timing, and stock levels, ensuring that the mall is well-prepared to meet demand during peak periods while optimizing operational efficiency during off-peak times.

Furthermore, understanding the average spend per customer and identifying repeat versus one-time customers provide insights into customer loyalty. Repeat customers often contribute more to the mall's long-term revenue, and thus recognizing their value can help tailor loyalty programs or special offers to increase repeat visits. Insights from this analysis can be used to create targeted incentives or loyalty rewards, encouraging one-time shoppers to become regular customers and fostering a sense of brand loyalty.

In conclusion, this analysis empowers mall management with a data-driven approach to addressing key business objectives: enhancing customer satisfaction, optimizing inventory, and maximizing revenue. By tailoring strategies based on clear, data-backed insights, the mall can respond to customer needs more effectively, align product offerings with demand, and improve the shopping experience. This analysis underscores the importance of data analytics in modern retail management, providing a roadmap for future growth and enabling the mall to remain competitive and customer-focused in an evolving market.

Through this analytical approach, mall management is better equipped to make decisions that lead to improved profitability and customer loyalty, positioning the mall as a preferred shopping destination.