CUSTOMER SHOPPING DATA ANALYSIS REPORT

Project

Customer Shopping Data Analysis

Power BI Dashboard & MySQL Analysis

Prepared By

Akash Tiwari

Data Analyst (Aspirant)

Tools Used

Power BI | MySQL | Kaggle Dataset | MS Excel

Date of Submission

November – 2025

This report is part of a personal data analytics portfolio project demonstrating analytical, visualization, and insight-generation skills using real-world shopping data.

CUSTOMER SHOPPING DATA ANALYSIS REPORT

TABLE OF CONTENT

CUSTOMER SHOPPING DATA ANALYSIS REPORT

1. EXECUTIVE SUMMARY

1. PROJECT SUMMARY

- **Dataset Source:** Kaggle (Customer Shopping Data, ~100k records)
- **Tools Used:** MySQL, Power BI, MS Excel
- **Analysis Type:** Descriptive & Correlation Analytics
- **Key KPIs:** Total Revenue, Average Basket Value, Customer Retention Rate

2. PROJECT OVERVIEW

The Customer Shopping Data Analysis project aims to explore purchasing patterns and customer behavior across multiple shopping malls using transactional data from a large retail dataset. The analysis was conducted using **MySQL** for data transformation and **Power BI** for visualization, uncovering key trends in sales, gender preferences, and category performance.

The dataset consists of **99,457 transactions**, encompassing variables such as customer demographics (age, gender), product categories, payment methods, and mall locations. The study focuses on identifying dominant categories, customer spending behavior, and factors influencing revenue performance.

The analysis revealed that the **dataset exhibits notable bias** — with female customers (≈60%) dominating the dataset and the **clothing category** accounting for a disproportionate share of sales. Despite this skew, meaningful insights were extracted, such as the **51–69 age group** being the most active shoppers, and **cash transactions** contributing the largest share of total payments (~44.7%).

Top-performing malls included **Mall of Istanbul**, **Kanyon**, and **Metrocity**, contributing the highest revenue figures. Across all demographics, the **clothing**, **cosmetics**, and **food & beverage** categories emerged as the most preferred product types.

Overall, this project demonstrates a comprehensive approach to retail analytics — from database querying to visualization — providing actionable insights for business decisions while emphasizing the importance of **data quality and balance** in analytical interpretation.

3. PROJECT OBJECTIVES

1. Analyze customer shopping patterns using SQL queries.
2. Develop interactive dashboards in Power BI.
3. Identify trends and correlations between demographics and spending behavior.
4. Generate actionable business insights for retail optimization.

"Without data, you're just another person with an opinion." *– W. Edward Deming*

CUSTOMER SHOPPING DATA ANALYSIS REPORT

2. METHODOLOGY

This project follows a structured data analytics workflow to ensure accuracy, consistency, and meaningful insights from customer shopping data. The process involved multiple stages – from raw data understanding to business visualization – as summarized below.

1. PROJECT WORKFLOW

1. Data Collection

The dataset titled **"Customer Shopping Data**" was sourced from Kaggle, containing ~100k transaction records from various shopping malls across multiple cities.

2. Data Cleaning and Preparation

The dataset was cleaned and standardized in Microsoft Excel and MySQL, where duplicate entries, null values, and inconsistent data formats were handled. A cleaned version named **"Cleaned_R2(no_invoice_date)"** was used for all analyses.

3. Exploratory Data Analysis (EDA)

Key metrics such as revenue distribution, transaction frequency, and customer demographics were analyzed using SQL queries. The focus was on understanding customer purchasing behavior and identifying major sales drivers.

4. Visualization and Dashboarding

Using **Power BI,** multiple visuals were created – including bar charts, pie charts, KPIs, and Pareto analysis – to identify key patterns in customer shopping behavior and mall performance.

5. Insights and Reporting

Findings were summarized in this report to support data-driven decision, followed by a Power BI dashboard for visual storytelling.

2. TOOLS USED

- Microsoft Excel: Initial data cleaning
- MySQL: Data analysis and aggregation queries
- Power BI: Visualization and interactive dashboard creation
- Microsoft Word: Report documentation

CUSTOMER SHOPPING DATA ANALYSIS REPORT

3. DATA DESCRIPTION AND PREPARATION

1. DATA OVERVIEW

The dataset used for this project is titled "Customer Shopping Data", originally sourced from Kaggle. It contains over ~100k transactional record representing customer purchases across multiple shopping malls. Each record represents a unique invoice generated for a single purchase.

2. KEY ATTRIBUTES

1. **invoice_no** – Unique identifier for each transaction
2. **customer_id** – Unique ID representing each customer
3. **gender** – Gender of the customer (male/female)
4. **age** – Age of the customer at the time of purchase
5. **category** – Type of product purchased (e.g., Clothing, Cosmetics, Technology, etc.)
6. **quantity** – Number of items purchased in that transaction
7. **price** – Unit price of the product
8. **payment_method** – Mode of payment used (Cash, Credit Card, Debit Card)
9. **shopping_mall** – Name of the mall where the purchase occurred
10. **total_price** – Total revenue generated from that transaction
11. **age_group** – Categorized age group of customer's ages from the transaction

3. DATA CLEANING AND TRANSFORMATION

To ensure data reliability and accuracy, several cleaning and transformation steps were performed using MySQL and Power BI:

- Removed missing and duplicate records
- Standardized column names and data formats
- Verified data consistency (e.g., correct gender labels, valid payment types)
- Calculated additional metrics such as:
  - Total Revenue (SUM(total_price))
  - Average Basket Value
  - Cumulative Revenue Percentage (for Pareto Analysis)
- Exported the cleaned dataset as Cleaned_R2(no_invoice_date) for visualization in Power BI

4. TOOLS AND TECHNOLOGIES USED

- **MySQL** – For data cleaning, structuring, and analytical querying
- **Power BI** – For visualization and dashboard creation
- **Microsoft Excel** – For intermediate data handling and export
- **Kaggle** – Source of the raw dataset

CUSTOMER SHOPPING DATA ANALYSIS REPORT

4. ANALYSIS AND KEY FINDINGS

1. KEY PERFORMANCE INDICATORS (KPIs)

- **Total Revenue** – ₹251.51 Million (₹251505794.25)
- **Total Quantity Sold** – 298.7 thousand (298712)
- **Total Customers** – 99457
- **Average Basket Value** – ₹2528

Insight:

The mall ecosystem generated over 251.51 million in total sales from 99457 customers, with each customer spending an average of ₹2528 per visit. The large total quantity sold indicates healthy product turnover and consistent buying patterns across stores. *See figure 1.1, 1.2, 1.3 & 1.4.*

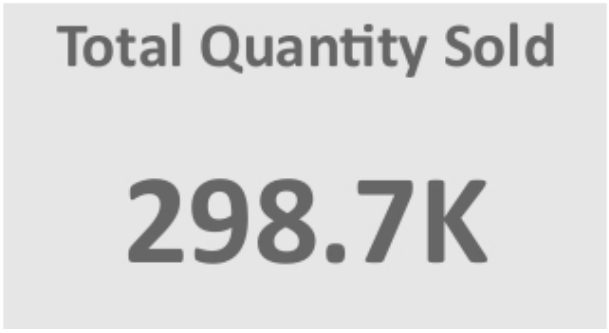

fig – 1.1

fig – 1.2

fig – 1.3

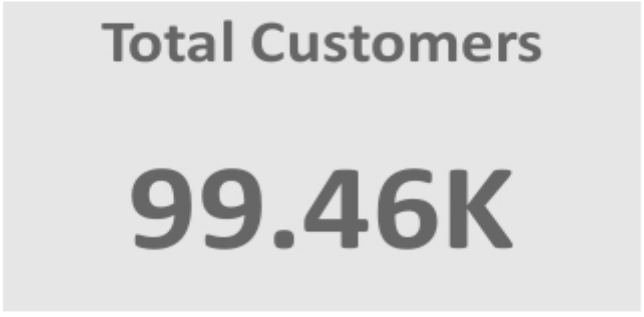

fig – 1.4

CUSTOMER SHOPPING DATA ANALYSIS REPORT

2. CATEGORY-WISE REVENUE CONTRIBUTION

Category	Total Revenue (₹)	Quantity Sold	Revenue Share (%)
Clothing	~114M	103,558	45.33
Shoes	~66.55M	30,217	26.46
Technology	~57.86M	15,021	23.01
Cosmetics	~6.79M	45,465	2.70
Toys	~3.98M	30,321	1.58
Foods & Beverages	~0.85M	14,982	0.34
Books	~0.83M	44,277	0.33
Souvenir	~0.64M	14,871	0.25

Table 2.1

Insight:

Although Technology ranks third in total revenue, it achieves this with a much lower sales volume (15k units) compared to categories like Clothing (103k) and Cosmetics (~45k units). This indicates a high unit price and premium positioning of Technology items, whereas Clothing dominates due to high purchase frequency and accessibility. *For more reference see figure 2.1 & 2.2.*

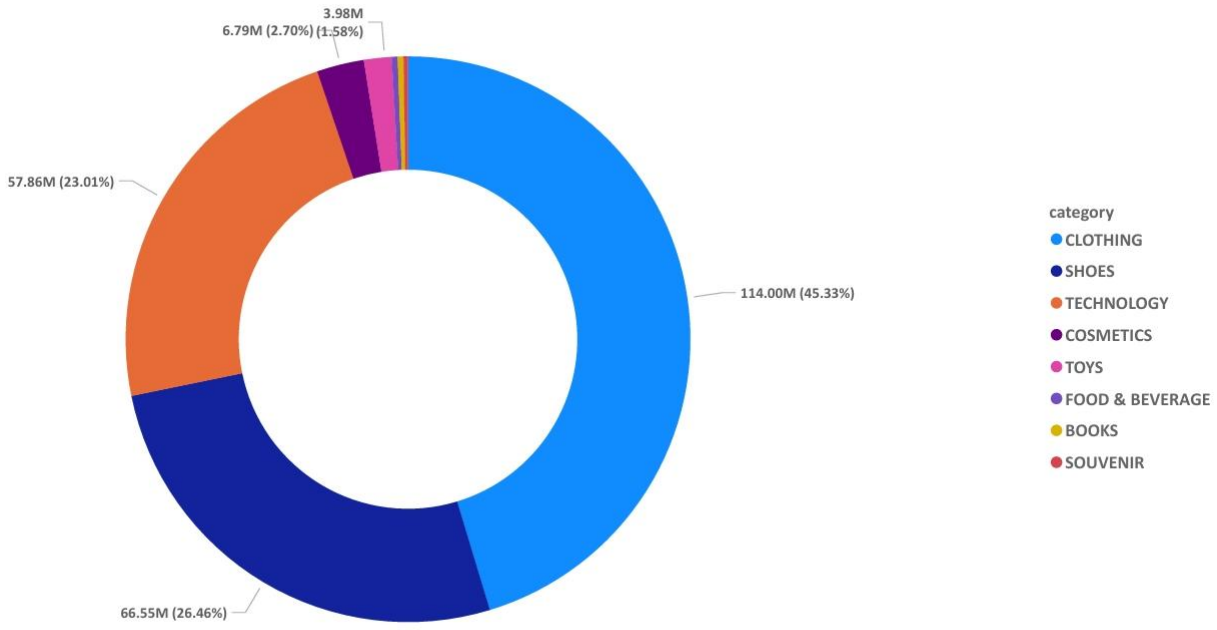

Fig – 2.1: Distribution of total revenue across Product Categories

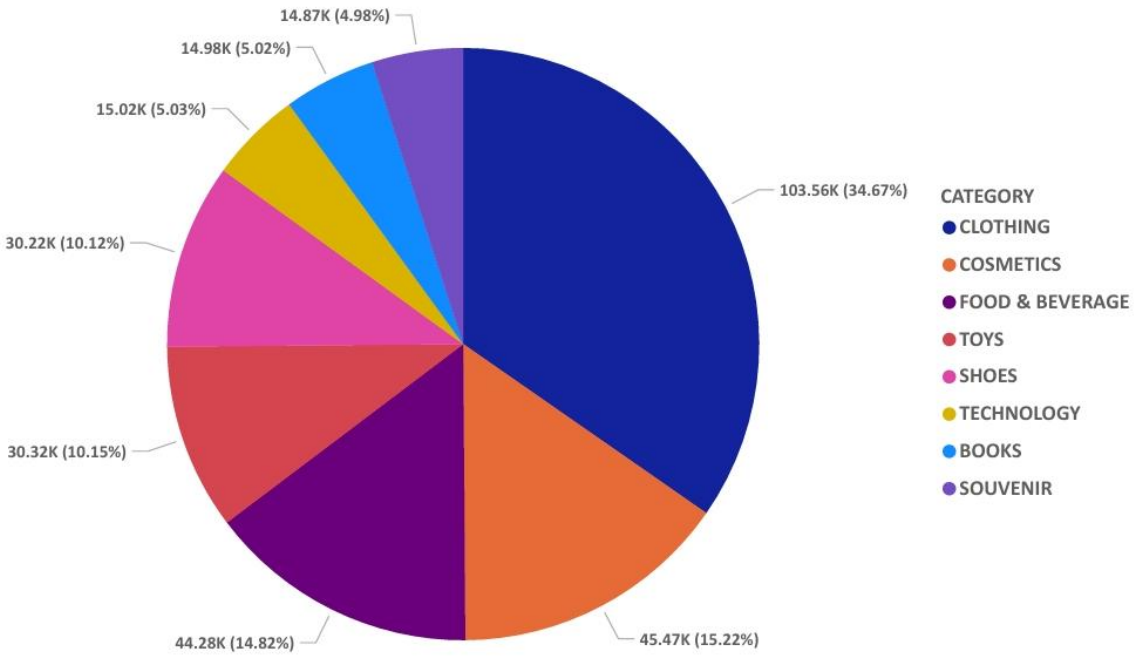

Fig – 2.2: No. of overall units purchased per product category

3. PAYMENT METHOD ANALYSIS

Payment Method	No of Payments	Payment Share (%)	Revenue (₹)
Cash	44447	44.69	112832243
Credit Card	34931	35.12	88077124
Debit Card	20079	20.19	50596427

Table 3.1

Observations:

- Cash generated the highest revenue across payment methods and is the most preferred payment method across all customers.
- Credit Card is second most preferred and Debit Card is the least preferred payment method across all.
- The gap between payment share and revenue share is negligible which shows that the preference is not influenced by any specific standard of customer. *(see figure 3.1 for reference).*

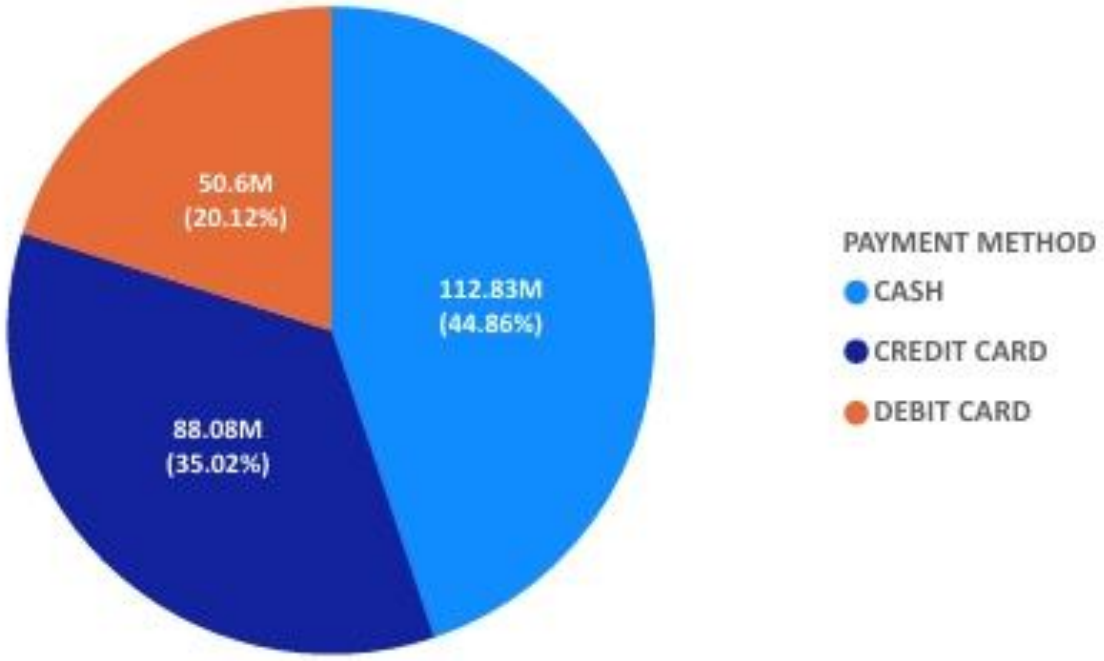

Fig – 3.1: Revenue contribution across payments methods

Insights:

- The given data conveys that the payment method preference among customers is not influenced by any specific standard or class of people as the revenue share and no. of payments share is almost same.
- Credit card and debit card is less preferred payment method than cash. In order to encourage customers to prefer credit card and debit card for payment, they can be offered by special discounts or seasonal discounts and gifts or tokens on payment.
- Other than credit card and debit card, the digital payment system can also be introduced for customers to make payment more convenient, secure and less need to carry cash along. (*See figure 3.1 and table 3.1 for more references).*

CUSTOMER SHOPPING DATA ANALYSIS REPORT

4. CUSTOMER DEMOGRAPHICS

Highlight:

Customers are divided in two genders specifically Male and Female. The dataset contains around ~60% of female customers and ~40% male customers. The sales revenue too is dominated by female customers over male customers where females contribute ~₹150 million in sales revenue whereas males contribute ~₹101 million in sales revenue. The customers are divided into 4 general age groups namely 18-25, 26-35, 36-50, 51-69 for each gender. *For more references see figure 4.1 and 4.2.*

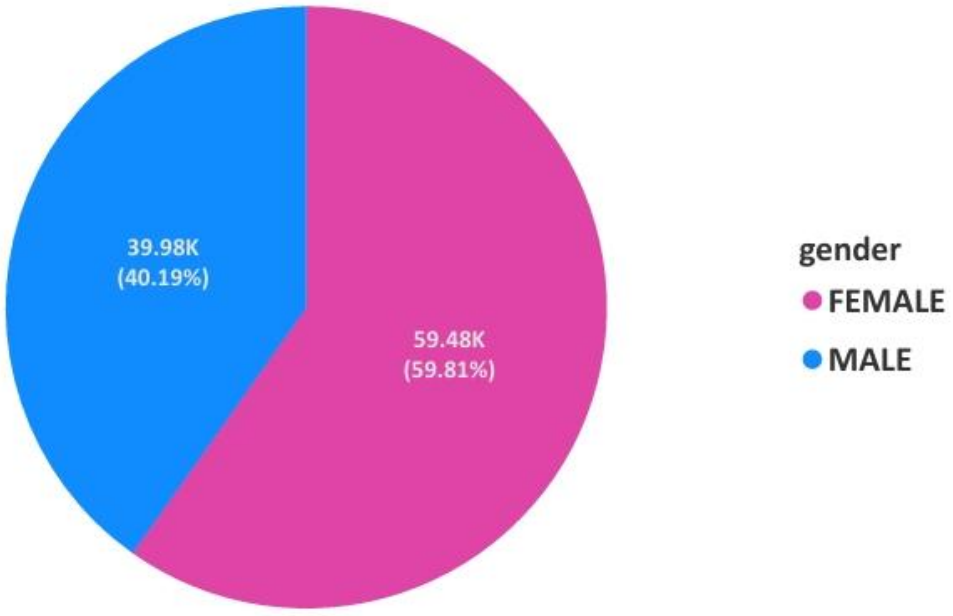

Fig – 4.1: Gender preferences

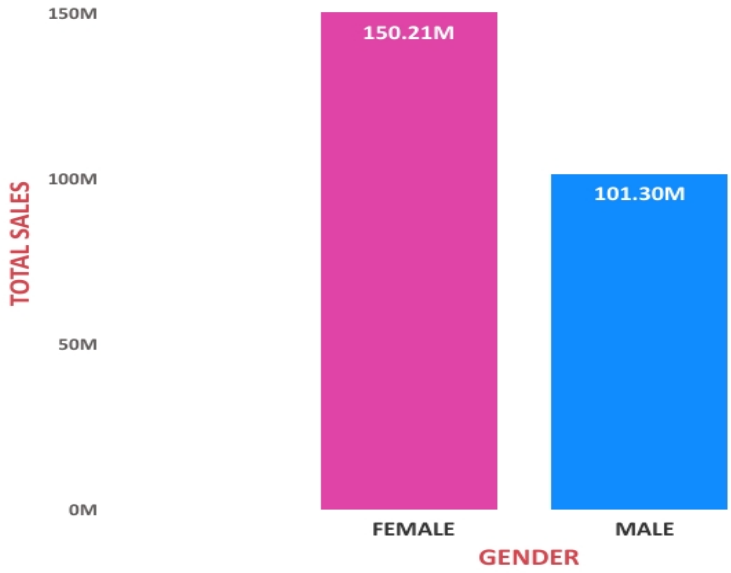

Fig – 4.2: Gender-wise revenue contribution

CUSTOMER SHOPPING DATA ANALYSIS REPORT

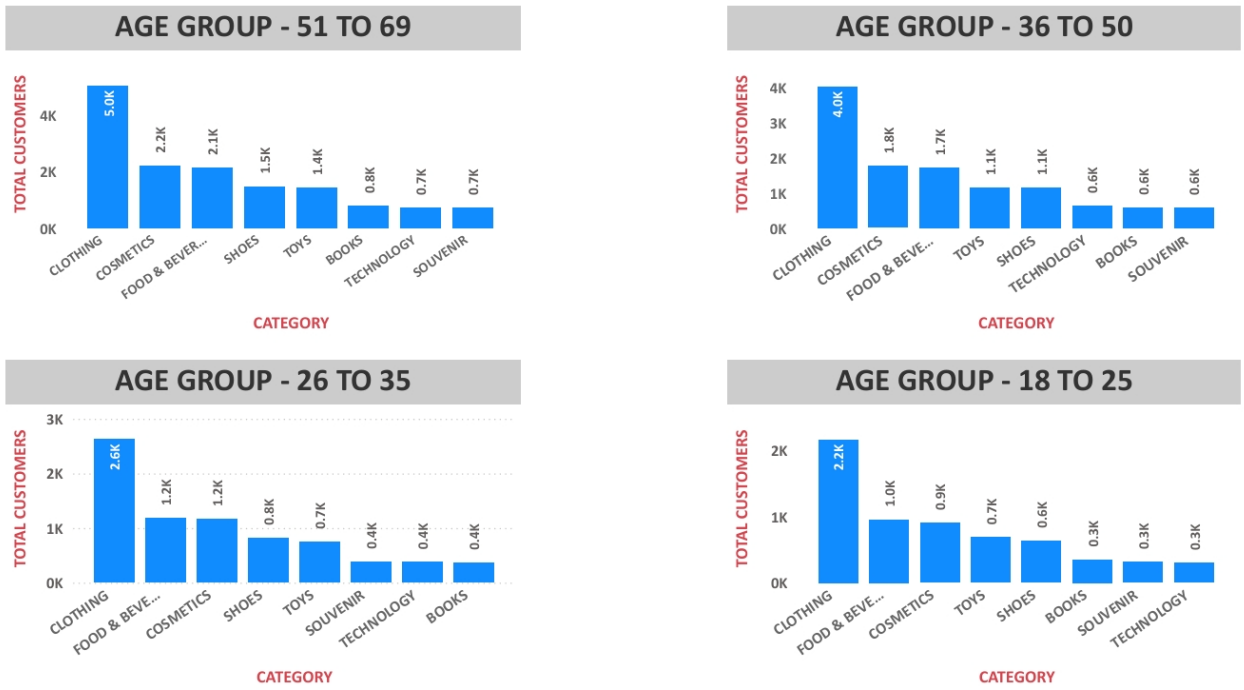

Figure – 4.3: Product category preferences by male across age groups

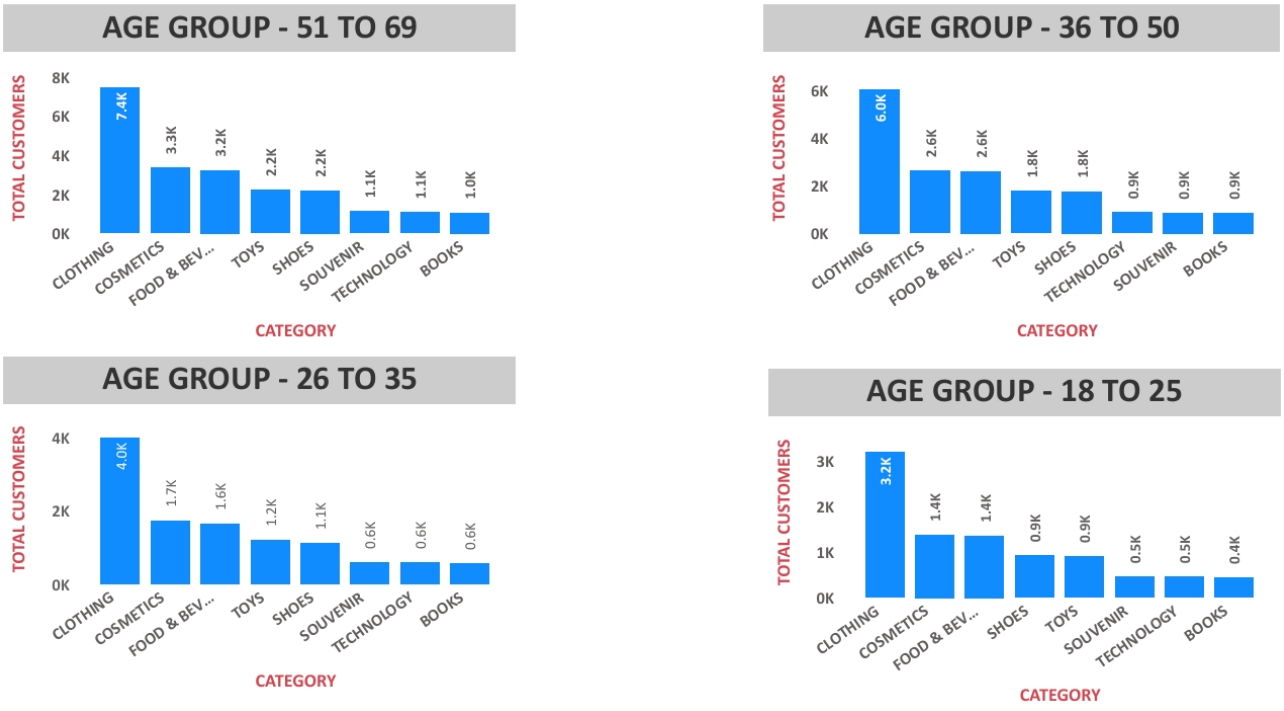

Fig – 4.4: Product category preferences by females across age groups

CUSTOMER SHOPPING DATA ANALYSIS REPORT

Observations:

- Category preference remains almost same for each gender and every age groups. There's not seen much deviations in category preferences and numbers.
- Clothing product category remains the first preference of all age group across genders in big numbers as it is the most sold product category.
- Cosmetics is the second most dominated product category across female age groups. However, it deviates with foods & beverages for male preferences but still remains competent in male preference too highlighting a type of biasness which is not seen mostly specially for male category of gender.

Insights

- **Dominant Age Group:**
  The **51–69 age group** clearly dominates across all product categories, especially **Clothing** — more than double that of any other segment. This age group forms the **core customer base** and the **highest revenue driver** overall.
- **Strong Mid-Age Engagement (36–50 years):**
  The **36–50 age group** follows closely, with **Clothing** leading again. Their shopping pattern mirrors the 51–69 group, indicating consistent preferences toward **fashion and lifestyle** spending.
- **Younger Age Groups (18–35 years):**

- **26–35 years:** Show moderate participation, with **Clothing** being the main purchase driver but noticeable drops in other categories.
- **18–25 years:** Represent the **least engaged segment**, averaging around **1–3K customers per category**. Their interest remains focused mainly on **Clothing, Cosmetics, and Food & Beverages**.

- **Category Consistency Across Ages:**
  Across all demographics, **Clothing remains the top-performing category**, followed by **Cosmetics** and **Food & Beverages**, highlighting a **universal preference pattern** irrespective of age.
- **Opportunity Insight:**
  Lower involvement of younger shoppers suggests a **marketing gap or affordability factor**. Retailers and mall management could explore **digital engagement, loyalty programs, and trendy product lines** to attract this demographic. *For more references, refer to figure 4.3 and 4.4.*

CUSTOMER SHOPPING DATA ANALYSIS REPORT

5. MALL PERFORMANCE

The analysis of shopping mall performance reveals a clear hierarchy in sales and category influence across the retail landscape. **Mall of Istanbul** and **Kanyon** emerge as the top performers, driving a significant and share of total revenue and customer traffic. Mid-tier malls such as **Metrocity, Metropol AVM,** and **Istinye Park** maintain strong yet stable performance levels, while others like **Viaport Outlet**, **Emaar Square Mall**, and **Forum Istanbul** contribute modestly. Overall, the findings suggest that large, centrally located malls with a diverse retail mix and high-footfall brands outperform smaller centers, emphasizing the importance of strategic location and category diversity in sustaining competitive advantage.

a. Product category dominance across shopping malls

The analysis of top-performing categories across various shopping malls reveals clear consumer preferences that align closely with the overall spending patterns. **Clothing** consistently emerges as the most dominant category in almost every mall, reflecting its universal appeal among shoppers. Secondary categories such as **Cosmetics**, **Shoes**, and **Technology** show variable dominance depending on the mall, hinting at each mall's customer demographics and brand portfolio. This insight suggests that malls with a balanced mix of fashion and lifestyle outlets tend to attract higher and more diverse customer engagement. *See Table 5.1 for more references*.

Shopping mall	Category	Total Quantity Sold
Mall Of Istanbul	Clothing	20813
Kanyon	Clothing	20513
Metrocity	Clothing	15729
Metropol AVM	Clothing	10552
Istinye Park	Clothing	10165
Forum Istanbul	Clothing	5239
Zorlu Center	Clothing	5228
Cevahir AVM	Clothing	5180
Viaport Outlet	Clothing	5101
Emaar Square Mall	Clothing	5038

Table – 5.1: Mall-wise top product category sold

b. Shopping Mall Revenue Contribution

A comparative assessment of mall-wise revenue indicates that a few malls account for a disproportionately large share of total sales. **Mall of Istanbul**, **Kanon, and Metrocity** together contributes nearly half of the total revenue, showcasing their market leadership and consumer trust. Mall such as **Istinye Park** and **Metropol AVM** maintain moderate but steady performance, while smaller centers like **Viaport Outlet**, **Emaar Square**, and **Forum Istanbul** contribute lower shares. This pattern emphasizes the concentration of consumer activity with high-traffic urban malls and reflects their strong retail ecosystems and accessibility advantages. *See figure 5.1 for more references.*

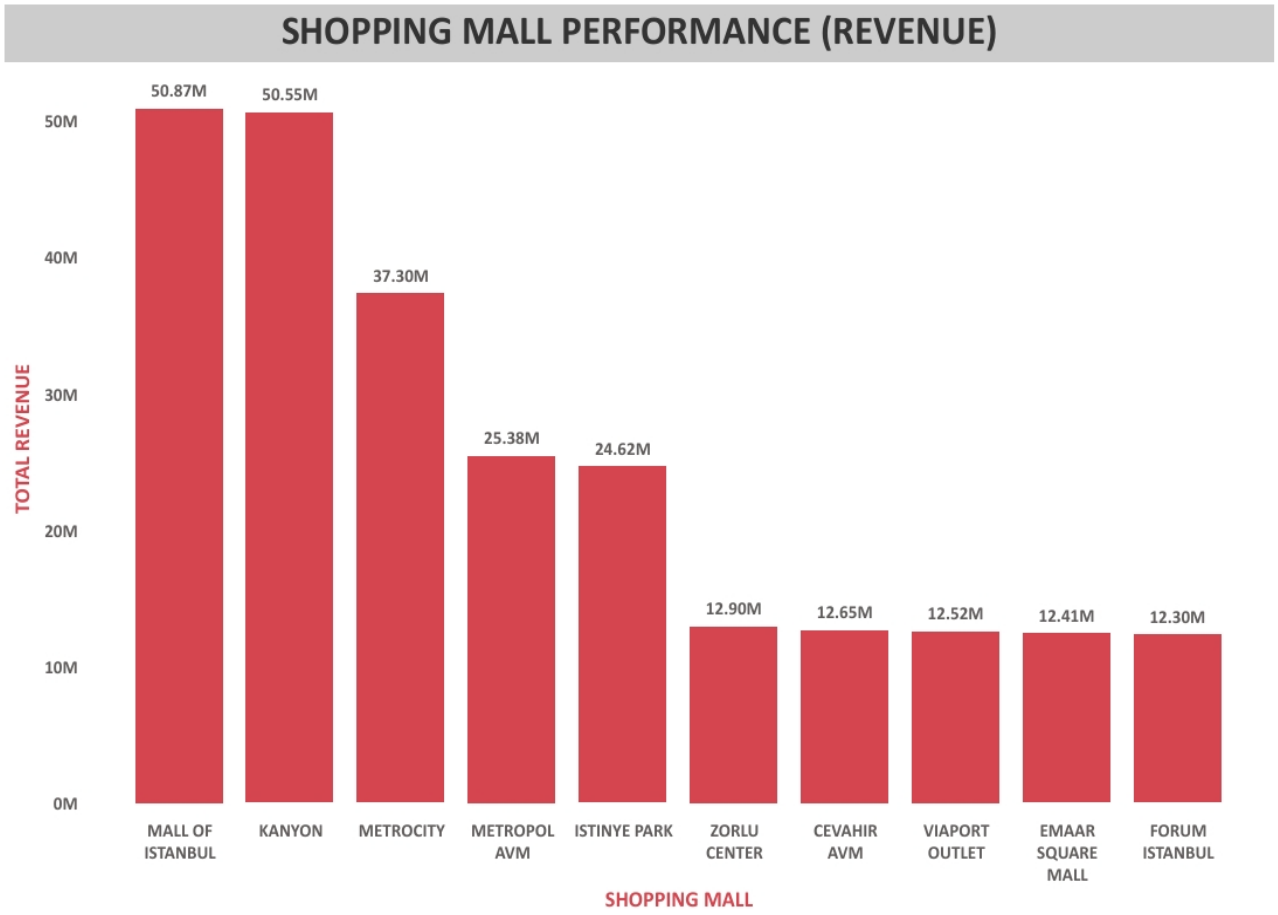

Figure – 5.1: Revenue contribution by shopping malls

c. Category-wise top shopping malls

The dominance analysis reveals that **Mall of Istanbul** leads in nearly all major product categories, highlighting its diverse retail offerings and strong customer retention. **Kanyon** stands out in **Souvenir,** suggesting a niche yet high-value segment focus. Other malls show limited category leadership, often reflecting localized shopping preferences or smaller retail footprints. The overall trend points to market concentration, where a handful of large malls capture both customer volume and revenue through broader category dominance. *See table 5.2 for more references.*

Category	Shopping mall	Total Quantity Sold
Clothing	Mall Of Istanbul	20813
Cosmetics	Mall Of Istanbul	9193
Food & Beverage	Mall Of Istanbul	8878
Shoes	Mall Of Istanbul	6112
Toys	Mall Of Istanbul	6031
Books	Mall Of Istanbul	3099
Technology	Mall Of Istanbul	3067
Souvenir	Kanyon	3025

Table – 5.2: Category-wise top shopping malls

KEY OBSERVATIONS AND INSIGHTS

a. Product category dominance across shopping malls

- Most malls exhibit dominance in **Clothing, Shoes, and Technology** categories, indicating these are the top revenue drivers across locations.
- **Mall of Istanbul** leads in almost every major category, highlighting its strong brand appeal and diverse retail mix.
- **Kanyon** uniquely outperforms others in **Souvenir**, suggesting niche category strength or targeted customer attraction.
- The category distribution pattern across malls remains relatively consistent, showing that consumer demand is largely uniform citywide. *(see table 5.1)*

b. Revenue contribution by shopping malls

- **Mall of Istanbul**, **Kanyon**, and **Metrocity** collectively contribute a major portion of total mall revenue — together accounting for over half of total sales.
- A steep revenue drop is visible beyond the top three performers, signaling **high sales concentration** among a few malls.
- Mid-tier malls such as **Metropol AVM** and **Istinye Park** maintain moderate sales, indicating balanced customer reach.
- The remaining malls (like **Forum Istanbul**, **Emaar Square Mall**, and **Viaport Outlet**) show comparatively lower revenue footprints — likely due to smaller customer bases or limited category spread. *(see figure 5.1)*

c. Category-wise top shopping malls

- Certain malls demonstrate **category specialization**, with stronger performance in specific product lines — for instance, **Technology** at Metrocity and **Fashion & Apparel** at Mall of Istanbul.
- This specialization could reflect targeted tenant mix strategies or demographic alignment with customer preferences.
- The dominance of a few malls across multiple categories suggests a **centralized consumer flow**, whereas smaller malls may benefit from niche positioning or localized marketing efforts. *(see table 5.2)*

CUSTOMER SHOPPING DATA ANALYSIS REPORT

6. CATEGORY REVENUE CONCENTRATION ANALYSIS – PARETO CHART

The Pareto analysis clearly illustrates a strong **revenue concentration pattern**, where **three key categories — Clothing, Shoes, and Technology — collectively contribute nearly 98% of total sales revenue**. This confirms the classic **80/20 principle**, indicating that a small number of product categories drive the majority of overall business performance. *See figure 6.1 for more references.*

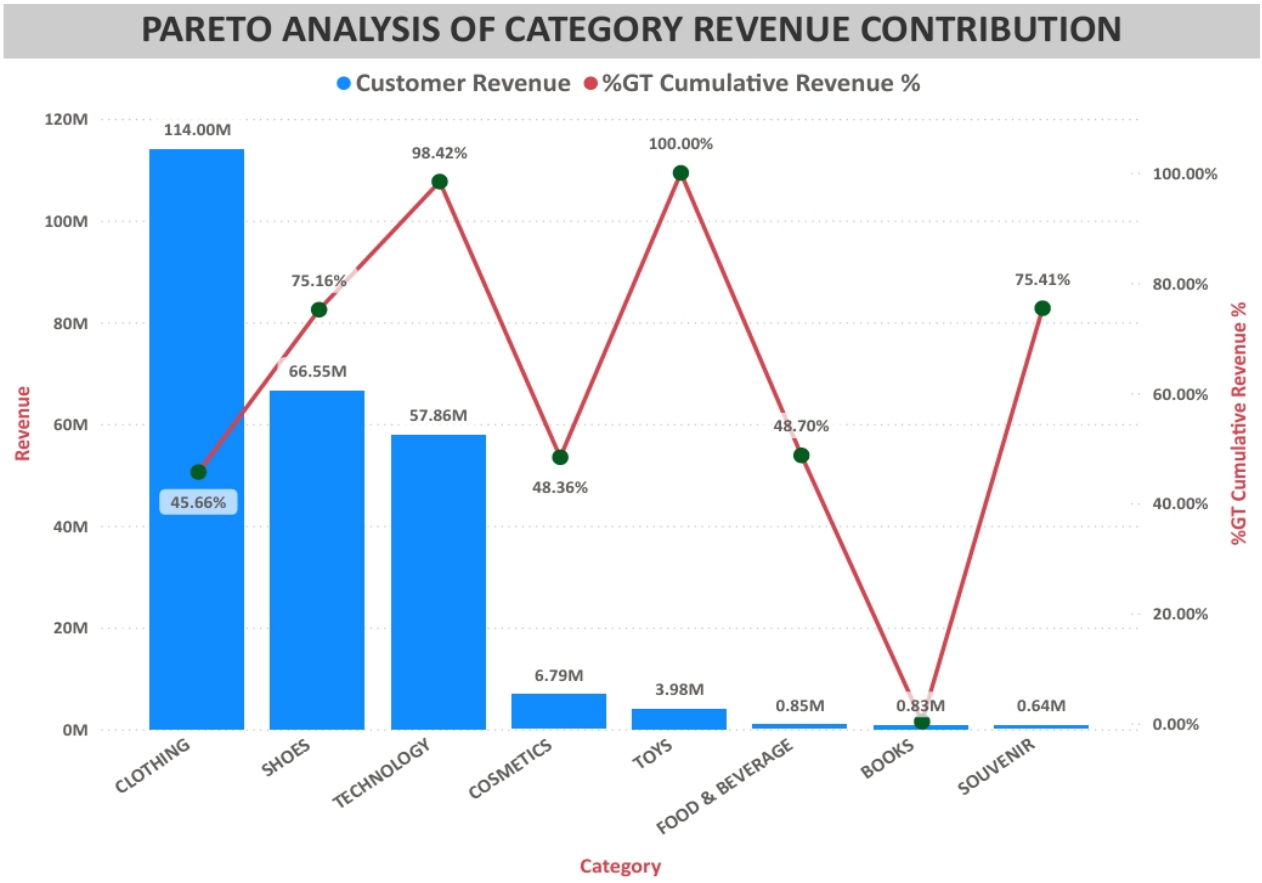

Figure – 6.1: Pareto Analysis of Category Revenue Contribution

Key Observations

1. Top three categories **— Clothing, Shoes, and Technology** — together contribute nearly 98% of total revenue.
2. **Clothing alone accounts for around 46%** of total sales, establishing it as the **primary revenue driver**.
3. **Shoes add another 29–30%**, bringing the cumulative share to about 75%.
4. **Technology completes the 80/20 Pareto threshold** by taking the cumulative to 98%.
5. Remaining categories like **Cosmetics, Toys, Food & Beverage, Books, and Souvenir** collectively account for **just 2% of total revenue**.
6. **The revenue gap between top and bottom categories is enormous**, showing high product concentration.

CUSTOMER SHOPPING DATA ANALYSIS REPORT

Insights & Interpretation

- The business is **heavily reliant on three major categories**, which means maintaining performance in these segments is **crucial for sustaining total revenue**.
- **Diversification** across weaker categories could reduce risk and increase market share in underperforming segments.
- **Marketing and promotional efforts** should focus primarily on Clothing, Shoes, and Technology — but a **growth strategy** can target Cosmetics and Toys for potential expansion.
- The Pareto distribution suggests **a typical 80/20 trend**, confirming that a small set of categories drive most of the business.
- The company can optimize **inventory, pricing, and shelf space** around top categories to maximize profitability.

CUSTOMER SHOPPING DATA ANALYSIS REPORT

5. CONCLUSION

- This analysis of customer shopping data offers a comprehensive understanding of purchasing behavior, sales distribution, and revenue concentration across multiple dimensions such as malls, categories, payment methods, and customer demographics.
- The study highlights that **a few top-performing categories and malls dominate overall revenue generation**, aligning with the **Pareto principle**. Specifically, *Clothing, Shoes, and Technology* contribute almost the entire revenue share, demonstrating strong consumer inclination toward fashion and tech-driven products.
- From the payment method perspective, **cash is still outperforming other modern payment methods like credit card and debit cards overwhelmingly (~45%)**, signaling that there needs to encourage people still for modern payment methods in order to enhance the system and let people make the payment system easier and quicker leading to efficiency.
- The **Mall Performance** and **Category-wise Analysis** further reveal that sales concentration is linked to both product relevance and customer footfall diversity. High-performing malls show better alignment between their category offerings and customer demographics, underscoring the importance of data-driven merchandising and strategic space allocation.
- Overall, the findings affirm that **targeted marketing, optimized inventory planning, and customer-centric strategies** can significantly improve retail performance. Leveraging such data insights enables management to focus resources on high-value segments while exploring underperforming areas for potential growth.
- In essence, this project reinforces the power of analytics in **transforming raw sales data into actionable business intelligence**, driving efficiency, profitability, and long-term customer satisfaction.