

# Shopping Analysis

## Title: Customer Shopping Analysis

### Project Summary:

This document describes the Shopping Report dashboard created to provide a consolidated view of retail sales performance across channels, payment methods, categories and shopping malls. The dashboard is designed for business analysts and retail managers to monitor revenue, customer behavior, and key performance indicators (KPIs).

### Problem Statement:

Retail stakeholders need a single interactive dashboard to answer the following questions:

- What is the total revenue and how is it distributed across online and offline channels?
- Which payment methods and categories contribute most to revenue?
- How do different shopping malls perform in terms of revenue?
- What are the customer age group distributions and average basket sizes?

Without a unified view, decision making for promotions, inventory and channel strategies is slow.

### Objective:

Primary objectives:

- Provide real-time KPIs for total revenue, average spending, average basket size, and channel split.
- Allow filtering by Category, Gender, Shopping Mall, and Payment Type Category.
- Visualize distribution by payment method, payment category, customer age group, and monthly trends.
- Offer actionable insights for marketing and operations teams.

## Table of Contents

Sr. No.	Section Name
1	Scope
2	Data Sources
3	Data Preparation & Cleaning
4	Dashboard Design & Layout
5	Metrics, Measures & Calculations
6	Filters and Interactions
7	Visualizations (page-wise)
8	Key Insights & Findings
9	Recommendations
10	Limitations
11	Future Work
12	Appendix: Data Dictionary & Files
13	Dashboard Screenshot
14	Appendix: Data Transformation Summary (Before & After Cleaning):
15	Data Transformation & Wrangling
16	Data Cleaning Process
17	Storytelling, KPIs & Charts
18	Storytelling, KPIs & Charts:
19	Visuals / Chars
20	Filter / Slicer
21	Conclusion

## Scope:

In scope:

- Interactive Power BI dashboard with 1 main page (Shopping Report) showing KPIs and visuals.
- Data ingestion from CSV/transactional source and processed in Power Query / Power BI data model.

Out of scope:

- Real-time streaming connectors and transactional writes back to source systems.

## Data Sources:

List of data sources used:

- customer\_shopping\_data.csv (transactions): fields include Date, Total Amount, Payment Type, Payment Category, Shopping Mall, Category, Gender, Age, Quantity, Item, etc.
- (Optional) Master data tables: Product Catalog, Mall Metadata, Payment Method Lookup.

## Data Preparation & Cleaning:

Steps performed during data preparation:

1. Imported transactional CSV into Power BI (Power Query).
2. Parsed date fields and created Year, Month, MonthNumber columns for time analysis.
3. Standardized text fields (Payment Type, Category, Mall names) using Trim and Clean.
4. Handled missing values: removed or imputed rows with null amounts; filled missing gender with 'Unknown'.
5. Created calculated columns: Age Group (Youth/Adult/Senior) based on Age, Basket Size = Quantity per transaction.
6. Created relationships: transaction table linked to lookup tables (if any) using key fields.

## Dashboard Design & Layout:

Design principles followed:

- Dark theme with high-contrast KPI cards for quick skim.
- Left-side filter panel for persistent filtering across visuals.
- Top row contains KPIs (Total Revenue, Average Spending, Average Basket Size, Online Sales %, Offline Sales %).
- Middle and bottom rows contain detailed visuals: donut and pie charts, bar charts, line chart for monthly trend, and mall-wise revenue.
- Used consistent colors and card borders to group related visuals.

## Metrics, Measures & Calculations:

Key measures (DAX examples) used in Power BI:

1) Total Revenue

DAX:

Total Revenue = SUM('Shopping\_Data'[Total Amount])

2) Average Spending

DAX:

Average Spending = DIVIDE([Total Revenue],  
DISTINCTCOUNT('Shopping\_Data'[CustomerID]))

3) Average Basket Size

DAX:

Average Basket Size = AVERAGE('Shopping\_Data'[Quantity])

4) Online Sales %

DAX:

Online Sales % = DIVIDE(SUMX(FILTER('Shopping\_Data',  
'Shopping\_Data'[Payment Type Category] = "Online"), 'Shopping\_Data'[Total  
Amount]), [Total Revenue]) \* 100

5) Offline Sales % = 100 - [Online Sales %]

## Filters and Interactions:

Available filters on the left panel:

- Category (All / specific category)
- Gender (All / Male / Female / Unknown)
- Shopping Mall (All / specific mall)
- Payment Type Category (Online / Offline)

### Interactions:

- Clicking a segment in pie/donut charts filters other visuals (cross-filter).
- Selecting a month on the line chart filters the KPI cards and other visuals to that month.

## Visualizations (page-wise):

Breakdown of visuals (as seen in the dashboard):

- KPI Cards (top row): Total Revenue, Average Spending, Average Basket Size, Online Sales %, Offline Sales %.
- Donut Chart: Total Amount by Payment Method (Online vs Offline)
- Horizontal Bar: Customer by Age Group (Youth, Adult, Senior)
- Column Chart: Total Amount by Payment Category (Cash, Credit Card, Debit Card)
- Column Chart: Total Amount by Shopping Mall
- Pie Chart: Total Revenue by Category (Clothing, Shoes, Technology, etc.)

- Line Chart: Total Amount by Month (monthly trend)

### Key Insights & Findings (example):

Example insights that business users can derive:

- Online channel contributes ~55% of revenue — consider investing in online promotions.
- Top categories by revenue: Clothing and Shoes — prioritize inventory and promotions for these.
- Certain malls (e.g., Mall of Ista..., Kanyon) have much higher revenue — explore store-level merchandising strategies.
- Monthly trend shows peak in January with drop in subsequent months — investigate seasonality or marketing campaigns.

### Recommendations:

- Allocate marketing budget toward high-performing channels and categories.
- Run targeted campaigns for underperforming months and malls.
- Improve payment options and user experience on the online channel to increase conversion.
- Enrich data with customer demographics and repeat purchase tracking for cohort analysis.

### Limitations:

- Data quality issues if source transaction data is incomplete or inconsistent.
- Age group inferred from available age column; inaccuracies if age is missing.
- Dashboard shows aggregated data; drill-through and store-level details require more granular data.

### Future Work:

- Add drill-through pages (e.g., Customer Details, Product Performance).
- Integrate real-time sales data or scheduled refresh.
- Add predictive analytics (forecast monthly revenue, churn risk) using Python/R visuals.
- Implement role-based views for managers vs analysts.

### Appendix: Data Dictionary & Files:

Example Data Dictionary (sample columns):

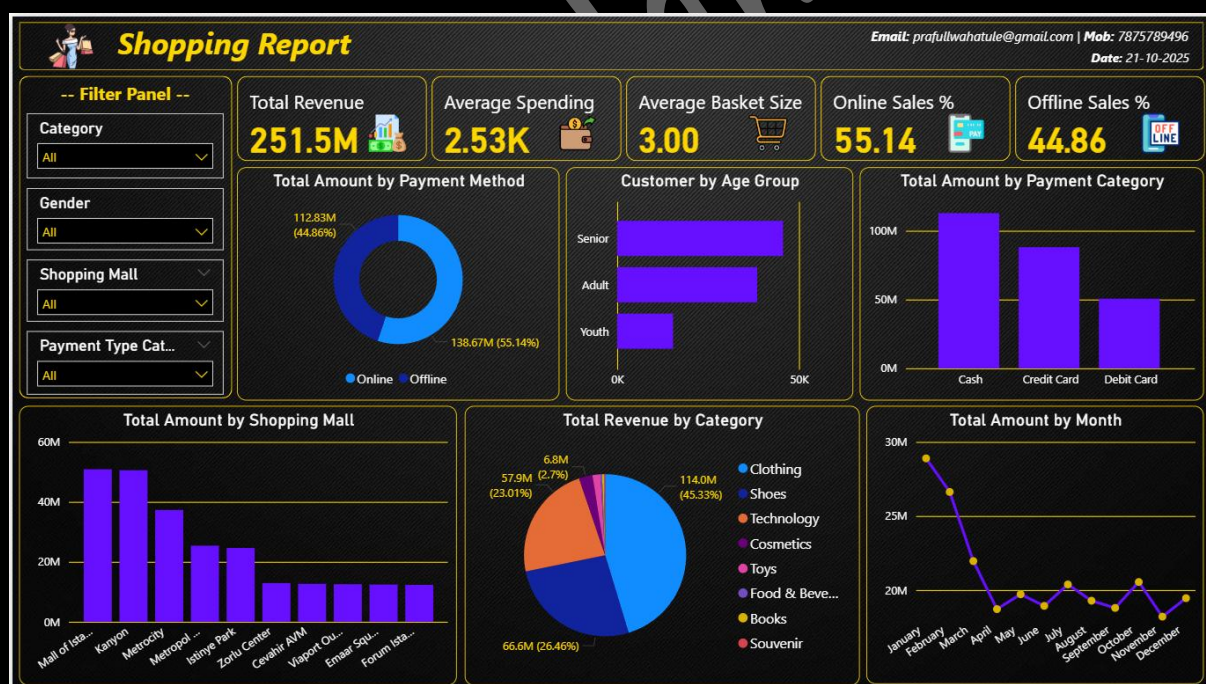
- TransactionID: Unique transaction identifier
- Date: Transaction date
- CustomerID: Unique customer identifier
- Total Amount: Numeric, transaction amount

- Quantity: Number of items in the transaction
- Payment Type: e.g., Cash, Credit Card, Debit Card
- Payment Type Category: Online / Offline
- Category: Product category (Clothing, Shoes, etc.)
- Shopping Mall: Mall name
- Gender: Male/Female/Unknown
- Age: Numeric

### Files included:

- customer\_shopping\_data.csv
- Shopping Analysis Dashboard.pbix
- Shopping Analysis Report.xlsx
- Shopping\_Analysis.docx

### Dashboard Screenshot:



### Appendix: Data Transformation Summary (Before & After Cleaning):

Before performing data cleaning and transformation, the dataset contained raw and unstructured columns imported from the CSV file.

Each column was analyzed and modified in Power BI using Power Query to make the data more consistent, meaningful, and ready for analysis.



The following summarizes the key changes made during transformation:

- The "invoice\_no" column was renamed to "Invoice No" to make it more readable. It represents a unique number for every shopping transaction.
- The "customer\_id" column became "Customer ID", used to track each customer's purchase history.
- The "gender" and "age" columns were cleaned for missing or inconsistent values, and used to analyze shopping behavior by gender and age group.
- The "category" column was standardized to show product types such as *Clothing, Shoes, Technology, etc.*, for better category-wise insights.
- The "quantity" and "price" columns were retained but verified for numeric consistency. These were later used to calculate the "Total Amount" (Quantity × Price).
- The "payment\_method\_type" column was split and classified into a new column called "Payment Type Category", which distinguishes between Online (Card/UPI) and Offline (Cash) payments.
- The "invoice\_date" column was renamed to "Invoice Date" and used for time-based analysis like month and year trends.
- The "shopping\_mall" column was renamed to "Shopping Mall" to ensure proper formatting for mall-level performance analysis.
- Several new calculated columns were added for deeper insights, such as:
  - Total Amount: Quantity × Price
  - Average Price per Item: Price ÷ Quantity
  - Payment Method: Specifies Cash, Debit Card, or Credit Card
  - Age Group: Categorizes customers as Teen, Youth, Adult, or Senior

Overall, this transformation process helped convert raw data into a clean, structured format, making it easier to build accurate visuals and KPIs in Power BI.

Col No. (Before)	Column Name (Before)	Description / Purpose (Before)	Col No. (After)	Column Name (After)	Description / Purpose (After Transformation)
1	invoice_no	Each shopping bill has a unique invoice number. It helps identify each	1	Invoice No	Unique number for each transaction — helps identify every purchase.

		transaction separately.			
2	customer_id	Unique ID assigned to each customer to track their purchases.	2	Customer ID	Unique ID assigned to each customer — used to track their purchase history.
3	gender	Gender of the customer — used for analyzing shopping patterns between males and females.	3	Gender	Gender of the customer — used for gender-based shopping analysis.
4	age	Age of the customer — helps in understanding which age group shops more.	4	Age	Age of the customer — helps in finding which age group shops the most.
5	category	Type of product purchased (e.g., Clothing, Shoes, etc.). Useful for product-wise analysis.	5	Category	Product category purchased (e.g., Clothing, Shoes, etc.). Useful for product-wise insights.
6	quantity	Number of items purchased	6	Quantity	Number of items purchased in a



		in that transaction.			single transaction.
7	price	Total amount spent by the customer for that invoice.	7	Price	Price of the items purchased in that transaction.
8	payment_method_type	Mode of payment used — e.g., Cash, Debit Card, Credit Card, etc. Helps study payment preferences.	8	Payment Type Category	Classifies payments as Online (Card/UPI) or Offline (Cash) — helps compare payment preferences.
9	invoice_date	Date when the purchase was made — useful for time-based analysis (month/year trends).	9	Invoice Date	Date of purchase — used for trend analysis by day, month, or year.
10	shopping_mall	Name of the mall where the purchase happened — helps analyze mall-wise performance.	10	Shopping Mall	Name of the mall where the purchase happened — useful for mall performance comparison.
—	—	—	11	Total Amount	Calculated as Quantity ×

					Price — total amount spent in each transaction.
—	—	—	12	Average Price per Item	Shows average cost per product: Price ÷ Quantity.
—	—	—	13	Payment Method	Specific payment mode like Cash, Credit Card, Debit Card, etc.
—	—	—	14	Age Group	Groups customers based on age (e.g., Teen, Youth, Adult, Senior) for better demographic analysis.

### Data Cleaning Process:

The raw dataset was imported from a CSV file containing 99,457 rows and 10 columns.

Data cleaning was performed in Power BI Power Query Editor to ensure accuracy, consistency, and readiness for analysis.

The step-by-step process is summarized below:

1. Extract Data from CSV:

The data was loaded directly from the CSV file into Power BI without any structural change.

(Rows: 99,457 → 99,457 | Columns: 10 → 10)

2. Remove Errors:

Checked for invalid or error values in all columns. No errors were found.

(No change in row or column count.)

3. Remove Blank Rows:

Verified for missing or blank rows in the dataset. None were found.

*(No effect on row count.)*

4. Remove Duplicates:

Checked for duplicate transactions using the "Invoice No" column. No duplicates existed.

*(Rows remained constant at 99,457.)*

5. Add New Calculated Columns:

- Total Amount (Quantity × Price)
- Average Price per Item (Price ÷ Quantity)
- Payment Method (Derived from Payment Type Category)
- Age Group (Categorized as Teen, Youth, Adult, Senior)

These steps gradually increased the column count from 10 to 14 while keeping rows unchanged.

Overall, the data structure improved from 10 columns to 14 columns, maintaining 99,457 records.

This ensured the dataset was clean, complete, and suitable for visual analysis and DAX calculations.

### **Data Transformation & Wrangling:**

After cleaning and validating the raw data, additional transformation steps were applied to enrich the dataset and make it more insightful for analysis.

These transformations were mainly performed in Power Query and Power BI DAX. The following key activities were done during data wrangling:

1. Total Amount:

A new calculated column was created using the formula

Total Amount = Quantity × Price

This helps calculate the total amount spent per transaction and is used for KPI metrics like Total Revenue.

2. Average Price per Item:

Another calculated column was added using the formula

Average Price per Item = Price ÷ Quantity

This helps understand the average cost of items purchased within a single transaction.

3. Payment Method:

Text values under payment types were standardized (e.g., "Credit Card", "Debit Card", "Cash") by removing extra spaces and correcting spelling variations.

This ensures data consistency and accurate grouping in visualizations.

4. Age Group:

A custom column was created using conditional logic to segment customers based on age:

```
if [Age] < 18 then "Teen"
else if [Age] >= 18 and [Age] <= 25 then "Youth"
else if [Age] > 25 and [Age] <= 45 then "Adult"
else "Senior"
```

This categorization made it easier to analyze shopping behavior across different age groups.

Overall, these transformation steps improved the dataset's analytical value and ensured the creation of accurate visuals and KPIs in the dashboard.

### Storytelling, KPIs & Charts:

In this Power BI Shopping Analysis Dashboard, several KPIs and visuals were created to understand business performance and customer behavior.

The key metrics include Total Revenue, Average Spending per Customer, Average Basket Size, and Online/Offline Sales Percentage. These KPIs help to measure sales growth, customer value, and purchasing patterns.

Different visuals were used to make the insights more interactive and easy to understand.

A Column Chart was used to show *Total Revenue by Mall*, helping identify which shopping mall performs the best.

A Donut Chart was created for *Revenue by Category* to highlight which product category contributes most to total sales.

The *Age Group Distribution Chart* shows the number of customers in each age group for demographic understanding.

A Pie Chart compares *Online vs Offline Sales Percentage* to track customer payment trends, and a Line Chart displays the *Monthly Revenue Trend* for time-based performance tracking.

Lastly, a Column Chart was used for *Payment Method Analysis* to visualize the popularity of different payment types like Cash, Credit, or Debit Cards.

Multiple slicers were added — such as Mall, Category, Gender, Age Group, and Payment Method Filters — to allow users to interact with the dashboard and explore the data dynamically.

### Key Performance Indicators (KPIs):

KPIs highlight the most important metrics of the report. They help track business performance, trends, and targets at a glance.

Sr . No .	KPI Name	Description	Formula
1	Total Revenue	Shows the total sales amount generated from all transactions — helps measure overall business performance.	Total Revenue = SUM('Shopping_Data'[Total Amount])
2	Average Spending per Customer	Tells how much on average each customer spends — helps understand customer value.	Average Spending = DIVIDE(SUM('customer_shopping_data'[Total Amount]), DISTINCTCOUNT('customer_shopping_data'[Customer ID]))
3	Average Basket Size	Measures the average number of items (quantity) purchased per customer — helps understand purchase behavior.	Average Basket Size = DIVIDE(SUM('customer_shopping_data'[Quantity]), DISTINCTCOUNT('customer_shopping_data'[Customer ID]))

4	Online Sales %	Compares total revenue from online payments (Credit/Debit Card, UPI) — shows payment trends.	Online Sales % = $\text{DIVIDE}(\text{SUMX}(\text{FILTER}('Shopping\_Data', 'Shopping\_Data'[Payment Method] = "Online" ), 'Shopping\_Data'[Total Amount] ), \text{SUM}('Shopping\_Data'[Total Amount])) * 100$
5	Offline Sales %	Compares total revenue from offline payments (Cash) — helps analyze customer payment preferences.	Offline Sales % = $100 - [\text{Online Sales \%}]$

### Visuals & Charts:

Visuals and charts represent data graphically. They make trends, comparisons, and patterns easy to understand, providing clear insights quickly.

Sr. No.	Chart Name	Columns Used	Visual Type	Insight / Purpose
1	Total Revenue by Mall	Shopping Mall, Total Amount	Column Chart	Shows which mall generates the highest revenue.
2	Revenue by Category	Category, Total Amount	Pie / Donut Chart	Understand revenue contribution by each product category.
3	Age Group Distribution	Age Group, Customer ID	Bar / Column	Shows number of customers in each age

			Chart	group for demographic analysis.
4	Online vs Offline Sales %	Payment Type Category, Total Amount	Doughnut / Pie Chart	Compare revenue from online vs offline payments.
5	Monthly Revenue Trend	Invoice Date, Total Amount	Line Chart	Shows revenue trends month by month for time-based analysis.
6	Payment Method Analysis	Payment Method, Total Amount	Column Chart	Displays popularity of each payment mode (Cash, Credit, Debit).

### Filters / Slicers:

Filters and slicers make the report interactive. Users can view data based on specific criteria or segments for customized analysis.

Sr. No.	Filter / Slicer Name	Used Column
1	Mall Filter	Shopping Mall
2	Category Filter	Category
3	Gender Filter	Gender
4	Age Group Filter	Age Group
5	Payment Method Filter	Payment Method



## Conclusion:

Creating professional Power BI documentation involves organizing your report in a clear and structured way so that readers can quickly understand the insights, metrics, and interactivity. From our discussion, the key takeaways are:

1. **Separate Sections for Clarity:**
  - KPIs: Highlight key business metrics to track performance at a glance.
  - Visuals & Charts: Represent data graphically to show trends, comparisons, and patterns.
  - Filters / Slicers: Allow users to interact with the report and perform customized analysis.
2. **Descriptions Matter:**
  - Each section should have a concise description explaining its purpose.
  - Descriptions help readers understand why a chart or filter exists and how to use it.
3. **Structured Tables for Reference:**
  - Use tables to list KPIs, visuals, and filters with columns like Name, Columns Used, Type, Insight/Purpose.
  - This makes the documentation organized and easy to follow.
4. **Interactive and Insightful Reporting:**
  - Proper documentation ensures that users not only see the data but also understand the story behind it.
  - It makes your Power BI report professional, user-friendly, and ready for presentation or analysis.

In short: Good documentation is a combination of clear descriptions, organized tables, and interactive explanations that help anyone quickly grasp the key insights from your Power BI report.