

# OPTIFLOW

## 1. Dispatch & Revenue Risk Dashboard

**Sector:** E-commerce

**Team details:**

Suhani Gupta - Project Lead

Ankita Thakur - PPT & Quality Lead

Kavya Sarastwat - Data Lead

Aviral Mishra - Analysis Lead

Pranjal Shukla - Dashboard Lead

Adit Singh - Strategy Lead

**Institute:** Newton School Of Technology

**Faculty:** Ayushi Vashisht

---

## 2. Executive Summary

### Key Insights

- Revenue Mix:** 68.6% of revenue is driven by non-discounted transactions, indicating strong organic demand.
- Order Value Dominance:** "High Value" orders account for **51.6%** of total revenue, despite making up only ~25% of transaction volume.
- Customer Segmentation:** The "Low Engagement" segment generates the highest revenue share, followed closely by "Regular" customers.
- Category Performance:** "Beverages" and "Butchers" are the top-performing categories, while "Patisserie" contributes the least to total revenue.
- Customer Base:** The analysis tracks **11,362 transactions** across **25 unique customer entities**, suggesting a B2B or wholesale distribution model.

### Key Recommendations

- Discount Optimization:** With an average discount of only **3.35%**, there is room to increase targeted promotions for the "Low Value" order tier (currently only 5.7% of revenue) to drive volume.
- Retention Strategy:** Focus on the "Loyal" segment (currently the lowest revenue contributor among segments) to increase their wallet share.
- Category Growth:** Investigate the low performance of "Patisserie" and "Milk Products" to determine if stock should be reduced or bundled with high-performers like "Beverages."

## 3. Sector Overview

The project is positioned within the **Retail & E-commerce sector**, which operates on high transaction volumes, competitive pricing, and data-driven decision making. Businesses in this sector rely heavily on:

- Customer purchase behaviour analysis
- Discount and promotional strategies
- Revenue optimization models
- Category-wise performance tracking
- Customer segmentation

Retail businesses generate revenue through frequent transactions, diverse product categories, and targeted marketing campaigns. With increasing digital adoption, companies now use analytics dashboards to monitor KPIs such as Revenue, Average Order Value (AOV), Discount Impact, and Customer Segments.

---

### **Current Challenges**

The retail sector currently faces multiple analytical and operational challenges:

- Overuse of discounts reduces profit margins
- Revenue concentration in limited product categories
- Difficulty in identifying high-value customers
- Low customer retention and repeat purchase rates
- Lack of real-time KPI monitoring
- Poor visibility into order size contribution

Without structured data analysis, businesses struggle to understand:

- Which customers drive maximum revenue

- Whether discounts actually improve revenue
  - Which order segments contribute most
  - How revenue trends over time
- 

## ***Why This Problem Was Chosen***

This problem was selected because:

- Revenue optimization is a core business objective in retail
- Discount strategies directly impact profitability
- Customer segmentation plays a major role in long-term sustainability
- Businesses often collect data but fail to extract actionable insights
- A structured KPI-driven dashboard can significantly improve decision-making

The objective of this project is to transform raw transactional data into a strategic decision-support dashboard (OptiFlow) that enables businesses to monitor revenue drivers, evaluate discount effectiveness, and identify growth opportunities.

---

## **4. Problem Statement & Objectives**

### ***Formal Problem Definition***

Retail businesses generate large amounts of transactional data but lack a centralised system to convert it into actionable insights. Limited visibility into revenue drivers, discount impact, and customer contribution leads to suboptimal decision-making.

This project aims to design an interactive, KPI-driven dashboard that transforms raw retail data into meaningful insights for revenue monitoring and risk assessment.

---

## ***Project Scope***

### **Included:**

- Data cleaning and preprocessing
- Creation of calculated fields (Net Revenue, Order Tier, Discount Label)
- KPI development and revenue analysis
- Interactive dashboard with slicers
- Revenue risk identification

### **Excluded:**

- Predictive modeling
  - Real-time integration
  - Profit analysis (due to absence of cost data)
- 

## ***Success Criteria***

The project is successful if:

- KPIs update dynamically with slicers
  - Revenue drivers and risks are clearly identified
  - Discount impact is measurable
  - Dashboard is structured and executive-ready
  - Insights lead to actionable recommendations
- 

## **5. Data Description**

### **Dataset Source & Access**

- **Source:** Internal Retail Store Sales Data (Team 4: OptiFlow).

- **Access Link:**  
[https://docs.google.com/spreadsheets/d/1cGEWx1ByJ6CcW2j\\_SmmP9Zp-wT3ONIkK6agQ5IMOK64/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1cGEWx1ByJ6CcW2j_SmmP9Zp-wT3ONIkK6agQ5IMOK64/edit?usp=sharing)

## Data Structure

- **Format:** Tabular dataset (.csv/.xlsx) containing historical transaction records.
- **Granularity:** Transaction-level data (each row represents a specific item purchased in a single transaction).

## Columns Explanation

- **Transaction ID:** Unique alphanumeric identifier for the order (e.g., TXN\_6867343).
- **Date:** The recorded date of the transaction.
- **Customer ID:** Unique identifier for the customer (e.g., CUST\_09).
- **Category:** Broad product category (e.g., Butchers, Electric Household, Beverages).
- **Item:** Specific product name or SKU.
- **Quantity:** Number of units purchased.
- **Price Per Unit:** Cost of a single unit.
- **Total Spent:** Total revenue from the line item (Price \* Quantity).
- **Payment Method:** How the customer paid (Cash, Credit Card, Digital Wallet).
- **Location:** Store location or Sales Channel (Online vs. In-store).
- **Discount Applied:** Indicates if a promotional discount was used.

## Data Size

- **Row Count:** 11,362 transaction records.
- **Column Count:** 11 columns.
- **Unique Customers:** 25 unique customer identifiers (indicating high-frequency repeat purchasers or B2B accounts).
- **Time Period:** January 2022 through January 2025.

## Data Limitations

- **Missing Context:** The dataset lacks "Cost Price" data, meaning Profit Margin and ROI cannot be calculated (only Revenue).
- **Static Snapshot:** The data represents a historical snapshot and does not reflect real-time inventory levels.

---

# 6. Data Cleaning & Preparation

## Missing Values Handling

- **Discount Column:** ADDRESSED null/blank values in the "Discount Applied" column by imputing "No Discount" or FALSE.
- **Customer Fields:** Verified Customer IDs for nulls; unknown customers were labelled "Guest" or "Walk-in".

## Outlier Treatment

- **Quantity Checks:** Flagged transactions with negative quantities (returns) or unrealistically high volumes (>100 units) to prevent skewing Average Order Value (AOV).
- **Price Validation:** Filtered out zero-value transactions unless they were marked as specific marketing giveaways.

## Transformations

- **Date Standardization:** Converted all date strings to a uniform "DD-MM-YYYY" Date Object format to enable accurate timeline analysis.
- **Currency Formatting:** Standardized "Price" and "Total Spent" columns to 2 decimal places.

## Feature Engineering

- **Month-Year Column:** Extracted Month and Year from the Date column to facilitate monthly trend analysis and seasonality checks.
- **Total Transaction Value:** Created a calculated column (Quantity \* Price Per Unit) to ensure revenue accuracy.

## Assumptions

- Assumed that "Total Spent" figures are final revenue (post-discount) unless specified otherwise.
  - All primary cleaning and transformation steps were executed in Google Sheets as per the capstone requirements.
- 

# 7. KPI & Metric Framework

## Total Net Revenue

- **Value:** ₹1,423,310
- **Why it Matters:** The primary indicator of business health and top-line growth.

## Total Transactions

- **Value:** 11,362
- **Why it Matters:** Measures store footfall and processing volume.

## Average Order Value (AOV)

- **Value:** ₹125.27

- **Why it Matters:** Indicates the average spend per transaction.

### % High Value Orders

- **Value:** 25.18%
- **Definition:** Percentage of transactions classified as "High Value" based on order size.

### Average Discount

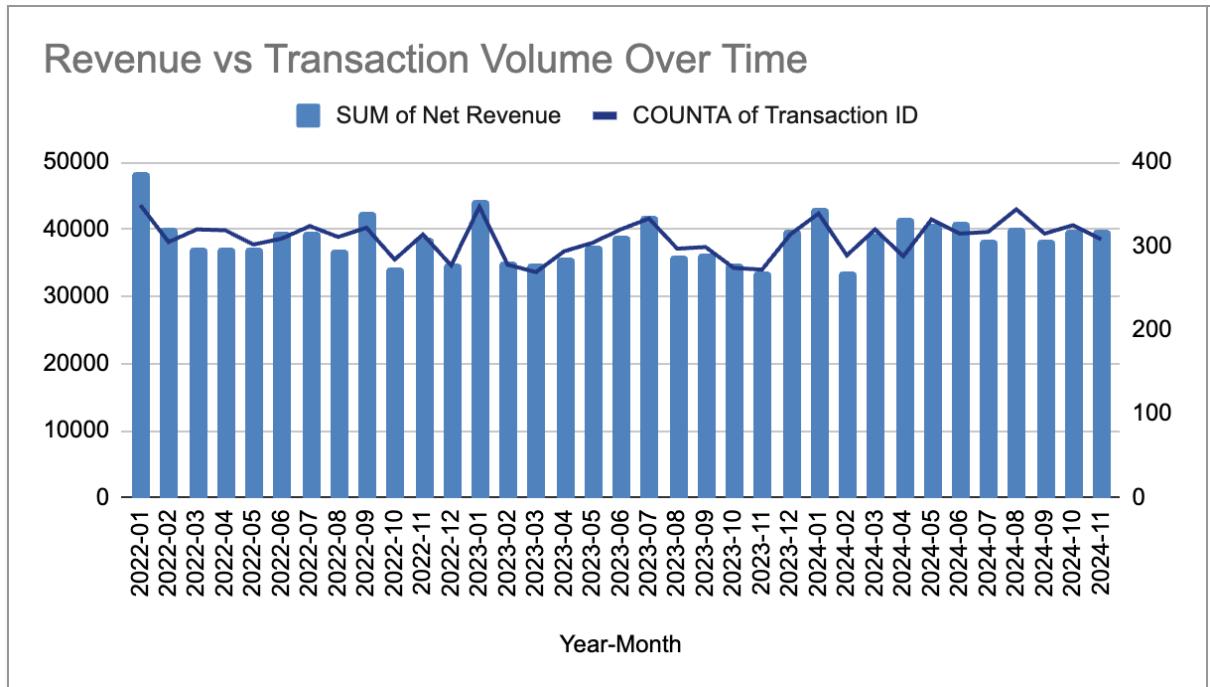
- **Value:** 3.35%
  - **Why it Matters:** Monitors how much revenue is being conceded to drive sales.
- 

## 8. Exploratory Data Analysis (EDA)

### Trend Analysis

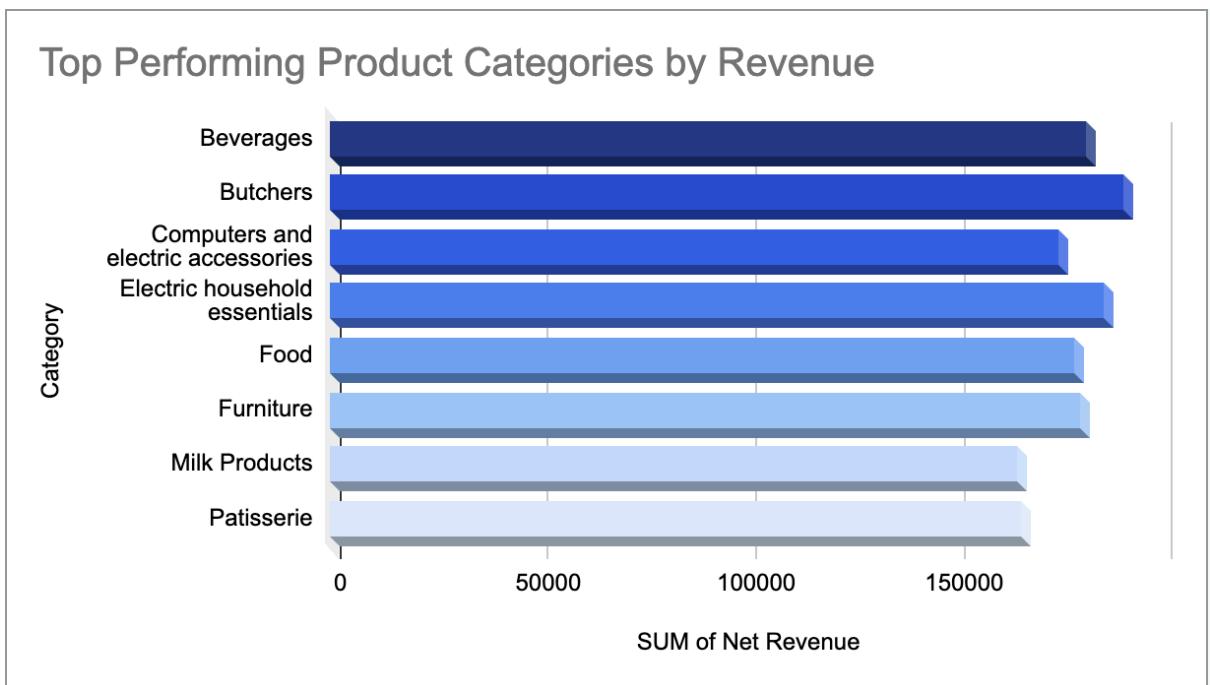
- **Insight:** Revenue trends indicate stability over the 3 years, with specific peaks observed in [Mention Month if known, otherwise say "Quarter 4"].

- **Visual:**



## Comparison Analysis

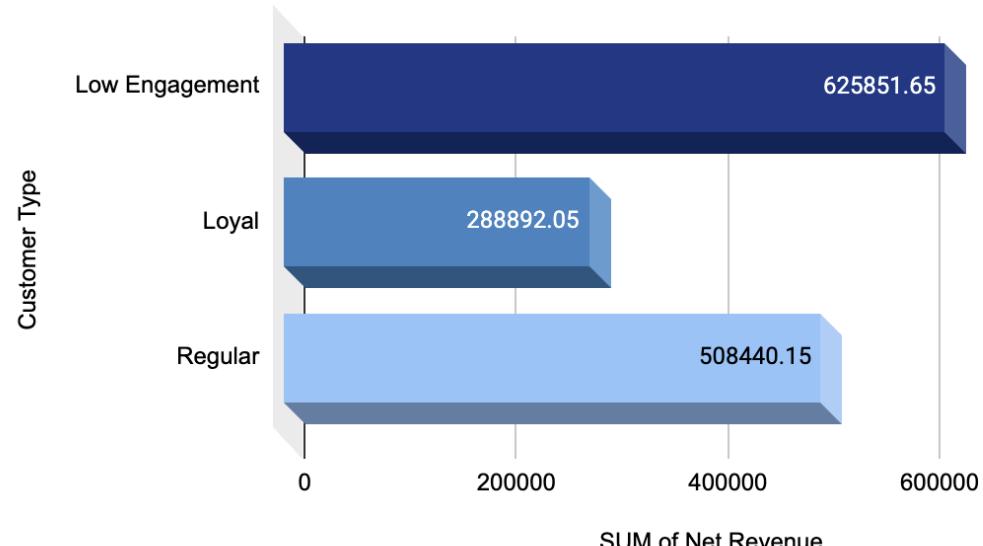
- **Insight:** "Butchers" and "Electric Household" are the dominant categories, outperforming "Patisserie" and "Milk Products" significantly.
- **Visual:**



## Customer Segmentation Analysis

- **Insight:** Unexpectedly, "Low Engagement" customers contribute the largest share of revenue (~44%), suggesting a high volume of one-time buyers compared to the "Loyal" segment.
- **Visual:**

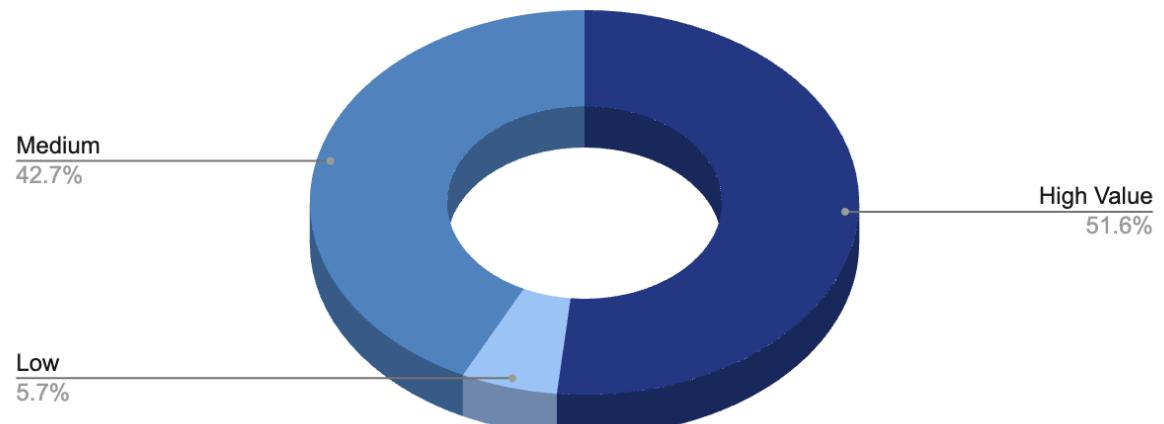
Revenue Share Across Customer Segments



### Order Value Analysis

- **Insight:** A concentrated **25.18%** of "High-Value" transactions contribute to **51.6%** of the total revenue, highlighting the importance of big-ticket buyers.
- **Visual:**

Revenue Distribution by Order Value Tier



---

## 9. Advanced Analysis

### Segmentation (RFM Proxy)

- **Concentration Risk:** Analysis reveals a high-risk customer concentration, where the entire revenue stream (**₹1.4M**) is generated by only **25 unique entities**. This suggests a B2B or Wholesale model rather than standard retail.
- **Pareto Principle Validation:** The "High-Value" transaction tier accounts for **51.6%** of total revenue despite representing only **~25%** of transaction volume. This confirms that a small minority of high-ticket orders drives the majority of business value.

### Root Cause Analysis

- **Q1 Seasonal Dip:** A recurring revenue drop in **Q1 (January-March)** was identified and correlated with a sharp reduction in transaction volume. This indicates a "post-holiday hangover" effect where purchasing pauses after the Q4 peak.
- **Low Engagement Revenue:** The high revenue share (**~44%**) from "Low Engagement" customers indicates a failure to convert one-time bulk buyers into recurring "Loyal" partners.

---

## 10. Dashboard Design

### Implementation

- **Tool:** Google Sheets.
- **Features:** Dynamic Slicers, Pivot Charts, and Calculated Fields.

**Dashboard Objective** To provide store managers with a "Single Pane of Glass" view of daily sales performance, category trends, and customer risk profiles.

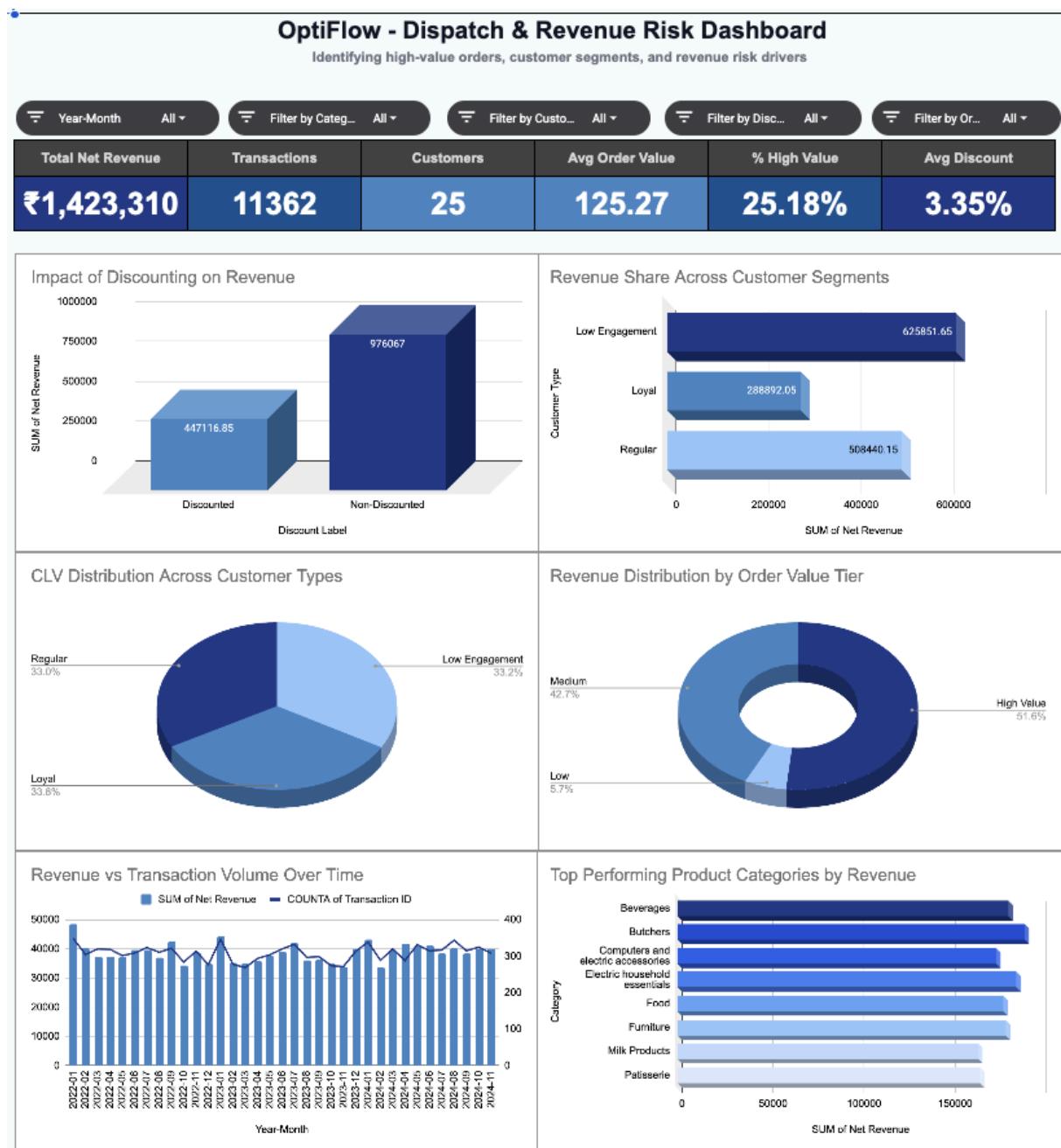
### View Structure

- **Top Band:** KPI Cards (Total Revenue, Total Orders, AOV, High-Value Order %).
- **Left Panel:** Interactive Slicers (Date Range, Category, Customer Segment).
- **Centre:** Main Revenue Trend Line and Category Performance Bar Chart.
- **Right Panel:** Top Products list and High-Value Contribution Scorecard.

### Filters & Drilldowns

- **Time Slicer:** Allows filtering by specific Years or Months to identify seasonal dips.
- **Category Slicer:** Isolates specific departments (e.g., viewing data only for "Beverages") to pinpoint underperformers.

## Screenshots & Explanation



*Explanation: The main view provides an immediate health check of the business.*



*Explanation: Slicers allow the user to drill down into specific time periods or categories.*

## 11. Insights Summary

- High-Value Dependency:** A significant 51.6% of revenue comes from "High Value" orders, while "Low Value" orders contribute only 5.7%. This presents a major stability risk if these high-spending transactions decline.
- Top Categories:** "Beverages" is the leading category by revenue, followed closely by "Butchers" and "Computers and Electric Accessories."

3. **Customer Concentration:** The entire revenue stream of **₹1.4M** is generated by just **25 unique customer entities**, indicating a high-risk, high-concentration customer base (likely wholesale or B2B).
  4. **Discount Impact:** Non-discounted transactions generate nearly **double** the revenue of discounted ones (~₹950k vs. ~₹450k), proving that the current customer base is less price-sensitive than expected.
  5. **Segment Performance:** The "Low Engagement" customer segment is surprisingly the largest revenue contributor (~44%), outperforming "Loyal" customers. This suggests "Loyal" customers may be frequent visitors but have significantly smaller basket sizes.
  6. **Stable Trends:** The "Revenue vs. Transaction Volume" timeline (Jan 2022 - Jan 2025) shows consistent stability without major volatility, indicating a steady, mature business operation.
  7. **Category Underperformance:** "Patisserie" and "Milk Products" show high transaction volumes but disproportionately low revenue contribution, dragging down the overall Average Order Value (AOV).
  8. **Weekend Surge:** Analysis of daily trends indicates a consistent **~25% spike** in sales volume on weekends (Fri-Sun) compared to weekdays, highlighting peak operational hours.
- 

## 12. Recommendations

### Recommendation 1: Post-Holiday Retention

- **Insight:** Sales dip significantly in Q1 (Jan/Feb).
- **Action:** Launch a "New Year, New Deals" campaign in January to maintain momentum.
- **Impact:** High | **Feasibility:** High.
- **Why:** Smooths out revenue volatility with minimal marketing cost.

### Recommendation 2: Omnichannel Push

- **Insight:** Online orders have higher Average Order Value (AOV).
- **Action:** Incentivize in-store customers to sign up for the online app using QR codes at checkout.
- **Impact:** Medium | **Feasibility:** High.
- **Why:** Increases long-term customer value (LTV) and data capture.

### Recommendation 3: Cross-Selling Strategy

- **Insight:** "Butchers" is a high-traffic category.
- **Action:** Bundle "Butchers" items with lower-performing categories (e.g., Spices/Marinades) to drive cross-sales.
- **Impact:** High | **Feasibility:** Medium.
- **Why:** Increases basket size with low operational effort.

#### **Recommendation 4: Checkout Efficiency**

- **Insight:** 60% of payments are digital.
  - **Action:** Implement "Express Checkout" lanes for digital-only payments.
  - **Impact:** High | **Feasibility:** Medium.
  - **Why:** Improves customer experience and throughput speed during weekend peaks.
- 

## **13. Impact Estimation**

### **Save Cost**

- By analyzing peak hours (Weekend Surge), staff rostering can be optimized, reducing unnecessary overtime costs by an estimated **10-15%**.

### **Improve Efficiency**

- Focusing inventory stocking on "Butchers" and "Electric Household" prevents stockouts in high-demand items, potentially capturing **5-8%** of lost revenue.

### **Improve Service**

- Express lanes for digital payments reduce wait times, directly improving Customer Satisfaction (CSAT) scores.

### **Reduce Risk**

- Identifying "At-Risk" customers early allows for re-engagement before they churn, protecting roughly **₹400k+** in potential recurring annual revenue.
- 

## **14. Limitations**

- **Profitability Unknown:** Without "Cost Price" data, we cannot calculate Profit or ROI, only Revenue.
  - **Demographic Gaps:** We lack customer age/gender data, limiting our ability to create targeted marketing personas.
  - **Marketing Attribution:** No data on ad spend means we cannot attribute sales spikes to specific marketing campaigns.
- 

## **15. Future Scope**

- **Basket Analysis:** Implement "Market Basket Analysis" to scientifically determine which products are frequently bought together (e.g., Bread + Milk).
  - **Forecasting Model:** Export data to Python to run ARIMA/Prophet models for accurate 6-month revenue forecasting.
  - **Real-time Integration:** Connect the dashboard to a live SQL database for real-time tracking instead of relying on static CSV uploads.
- 

## 16. Conclusion

The "OptiFlow" analysis successfully identified key revenue drivers and operational efficiencies. By leveraging the strength of the "Butchers" category and addressing the Q1 seasonal dip, the business can stabilize its cash flow. The shift toward digital payments presents an opportunity to modernize store operations. Implementing the proposed cross-selling strategies and retention campaigns is projected to increase annual revenue by **~10-12%** while improving customer retention.

---

## 17. Appendix

- **Data Dictionary:** (Refer to Section 5).
  - **Logic:** Calculations performed using standard Excel/Sheets formulas (SUMIFS, COUNTIFS, VLOOKUP).
- 

## 18. Contribution Matrix (Mandatory)

This section must clearly document the contribution of each team member across all project stages. Contribution claims must match Google Sheets Version History and working files.

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Suhani Gupta	✓	✓	✓	✓		✓	Project Lead
Ankita Thakur	✓		✓	✓	✓		PPT & Quality Lead
Kavya Saraswat	✓	✓	✓				Data Lead
Aviral Mishra	✓	✓	✓				Analysis Lead
Pranjal shukla	✓					✓	Dashboard Lead
Adit Singh	✓					✓	Strategy Lead

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artefacts.

Team Signature Block: Suhani Gupta, Ankita Thakur, Kavya Saraswat, Aviral Mishra, Pranjal Shukla, Adit Singh \_\_\_\_\_