# Retail Store Sales Analysis

## 1. Project Overview

This project analyzes sales performance and customers behavior using transactional data from 5000 purchases across various product categories, The goal is to uncover insights into sales performance in each category, customers spending patterns to guide strategic business decisions.

## 2. Dataset Summary

- Rows: 5,000
- Columns: 11
- Ket Features:
    - Customers Demographics (CustomerID, Age, Gender)
    - Purchase Details (Category, ItemPurchased, Amount, Season, PaymentMethod, ItemRating)
    - Shopping Behavior (DiscountAppiled)
    - Frequency (PreviousPurchases)
- Missing Data: 0 Values

## 3. Tools & skills used

- Python (data cleaning & EDA)
- SQL (aggregation, segmentation, KPI queries)
  Power BI (dashboard & interactive visualizations)
- Analytical methods: segmentation (age/gender), seasonality, discount-effect analysis, payment-method satisfaction, AOV (average order value), repeat-customer analysis

# 4. Methodology

## 4.1. Exploratory data using Python

I began with Data preparation and cleaning in Python

- **Data loading:** Imported the dataset using *pandas*
- **Setting Dataset:**
  (1) Set the column to the standard (Lowercase, No spaces)
  (2) Check Data types used *df.info()*
- **Initial Exploration:** Used *df.info()* to check structure and describe() to check summary statistic

| | customerid | age | gender | category | itempurchased | amount | season | paymentmethod | itemrating | discountapplied(%) | previousp |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 5000.000000 | 5000.000000 | 5000 | 5000 | 5000 | 5000.000000 | 5000 | 5000 | 5000.000000 | 5000.000000 | 500 |
| unique | NaN | NaN | 2 | 9 | 30 | NaN | 4 | 2 | NaN | NaN | |
| top | NaN | NaN | Female | Footwear | Sandals | NaN | Spring | Card | NaN | NaN | |
| freq | NaN | NaN | 2504 | 983 | 339 | NaN | 1300 | 4009 | NaN | NaN | |
| mean | 2500.500000 | 45.224800 | NaN | NaN | NaN | 285.090522 | NaN | NaN | 3.784160 | 14.983600 | |
| std | 1443.520003 | 14.564995 | NaN | NaN | NaN | 551.454382 | NaN | NaN | 0.681796 | 5.988063 | |
| min | 1.000000 | 20.000000 | NaN | NaN | NaN | 5.080000 | NaN | NaN | 1.100000 | 0.000000 | |
| 25% | 1250.750000 | 33.000000 | NaN | NaN | NaN | 70.547500 | NaN | NaN | 3.300000 | 11.000000 | |
| 50% | 2500.500000 | 45.000000 | NaN | NaN | NaN | 122.485000 | NaN | NaN | 3.800000 | 15.000000 | |
| 75% | 3750.250000 | 58.000000 | NaN | NaN | NaN | 184.535000 | NaN | NaN | 4.300000 | 19.000000 | |
| max | 5000.000000 | 70.000000 | NaN | NaN | NaN | 2997.940000 | NaN | NaN | 5.000000 | 36.000000 | 1 |

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 11 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   customerid          5000 non-null   int64
 1   age                 5000 non-null   int64
 2   gender              5000 non-null   object
 3   category            5000 non-null   object
 4   itempurchased       5000 non-null   object
 5   amount              5000 non-null   float64
 6   season              5000 non-null   object
 7   paymentmethod       5000 non-null   object
 8   itemrating          5000 non-null   float64
 9   discountapplied(%)  5000 non-null   int64
 10  previouspurchases   5000 non-null   int64
dtypes: float64(2), int64(4), object(5)
memory usage: 429.8+ KB
```

- **Verify the completeness of the Data:** used *.isnull().sum()* to check NULL value in any columns and rows
- **Export Data:** Used *.to_csv()* for export data to csv

## 4.2. Asking questions to support business decision

- Which customer segments (by age group and gender) generate the highest value for the business?
- Which product categories and items perform best in each season?
- How do different age segments differ in purchase behavior and satisfaction, and which segment should we prioritize?
- Which gender has the highest purchase frequency and customer value?
- Which categories or items have high purchase volume but low customer ratings and need service improvement?
- What level of discount is most effective in increasing customer purchases?
- Which payment method is associated with the highest customer satisfaction?
- Which season generates the highest business performance and deserves more investment?

## 4.3 Data Analysis using SQL

1. Which customer segments (by age group and gender) generate the highest value for the business?

| age_ranges | gender | total_revenue | total_order | AOV |
|---|---|---|---|---|
| 20-29 | Male | 198736.11 | 480 | 414.03 |
| 30-39 | Female | 53520.15 | 481 | 111.27 |
| 30-39 | Male | 238729.84 | 465 | 513.4 |
| 40-49 | Female | 56530.33 | 525 | 107.68 |
| 40-49 | Male | 245992.7 | 514 | 478.59 |
| 50-59 | Female | 55656.01 | 520 | 107.03 |
| 50-59 | Male | 226350.08 | 494 | 458.2 |
| 60-70 | Female | 59678.87 | 520 | 114.77 |
| 60-70 | Male | 239905.23 | 543 | 441.81 |

| age_ranges | gender | total_revenue | total_order | AOV |
|---|---|---|---|---|
| 40-49 | Male | 245992.7 | 514 | 478.59 |

Answer:

*Men aged 40-49 are the customer group that generates the most value for businesses, Although it had fewer orders than some other groups, it also had the highest AOV (Average Order Value)*

2. Which product categories and items perform best in each season?

| season | category | item | total_revenue | total_order | AOV |
|---|---|---|---|---|---|
| Autumn | Electronics | Mobile Phone | 214345.77 | 126 | 1701.16 |
| Spring | Electronics | Smart Watch | 236964.91 | 131 | 1808.89 |
| Summer | Electronics | Smart Watch | 222637.35 | 126 | 1766.96 |
| Winter | Electronics | Laptop | 221454.32 | 124 | 1785.92 |

Answer:

*Electronics can make the most total revenue in every each season but differences by some items are*

*Autumn = Mobile Phone,*

*Spring = Smart Watch*

*Summer = Smart Watch,*

*Winter = Laptop*

3. How do different age segments differ in purchase behavior and satisfaction, and which segment should we prioritize?

| age_range | total_revenue | total_order | AOV | average_rating |
|---|---|---|---|---|
| 20-29 | 249089.4 | 938 | 265.55 | 3.78 |
| 30-39 | 292249.99 | 946 | 308.93 | 3.79 |
| 40-49 | 302523.03 | 1039 | 291.17 | 3.78 |
| 50-59 | 282006.09 | 1014 | 278.11 | 3.78 |
| 60-70 | 299584.1 | 1063 | 281.83 | 3.79 |

Answer:

*The Customers group Age 40-49 years old, They make the most total revenue and had huge order numbers.*

***Age group 20-29:** We were able to achieve the lowest revenue, number of orders and AOV (Average Order Value) in this group, I don't think this customer group is our primary target market.*

*However, we can use various marketing strategies to increase the number of orders from this customer group.*

***Age group 30-39:*** *This customer segment has the highest AOV (Average Order Value) among all our clients, but the number of orders is low. We need to implement marketing strategies to increase the order volume for this customer segment.*

***Age group 40-49:*** *This customer segment generates the highest total revenue. We can develop a marketing plan to increase their Average Order Value (AOV).*

***Age group 50-59:*** *This customer segment already has a relatively high number of orders. We should develop a marketing plan to increase the Average Order Value (AOV) for this customer group.*

***Age group 60-70:*** *This customer segment already has a relatively high number of orders. We should develop a marketing plan to increase the Average Order Value (AOV) for this customer group.*

*However, The customers had low satisfaction in every age group (lower than 4.0). We should review the operations team, stores, and the logistics team to ensure they are providing appropriate services, and urgently plan for service improvements, as this directly impacts customers.*

4. Which gender has the highest purchase frequency and customer value?

| gender | total_revenue | total_order | AOV | frequency |
|---|---|---|---|---|
| Male | 1149713.96 | 2496 | 460.62 | 12471 |
| Female | 275738.65 | 2504 | 110.12 | 12575 |

Answer:

*The customers Male group had frequency more than the Female group, But the differences weren't significant enough to be substantial.*

5. Which categories or items have high purchase volume but low customer ratings and need service improvement?

| category | item | total_revenue | total_order | average_rating_item |
|---|---|---|---|---|
| Footwear | Sandals | 46042.49 | 339 | 3.83 |
| Footwear | Formal Shoes | 47301.48 | 324 | 3.76 |
| Footwear | Sneakers | 47798.27 | 320 | 3.78 |
| Sports | Yoga Mat | 38717.61 | 283 | 3.74 |
| Sports | Dumbbells | 39047.34 | 277 | 3.7 |
| Sports | Football | 37833.16 | 275 | 3.78 |
| Beauty | Skincare Kit | 16275.77 | 193 | 3.88 |
| Beauty | Makeup Set | 15003.84 | 176 | 3.69 |
| Accessories | Wallet | 17881.13 | 170 | 3.74 |
| Beauty | Perfume | 12607.45 | 160 | 3.88 |
| Womens Clothing | Top | 18783.2 | 160 | 3.75 |
| Accessories | Handbag | 15491.03 | 152 | 3.87 |
| Accessories | Sunglasses | 14261.67 | 148 | 3.73 |
| category | item | total_revenue | total_order | average_rating_item |
| Electronics | Laptop | 265862.26 | 142 | 3.79 |
| Electronics | Mobile Phone | 233026.68 | 140 | 3.88 |
| Mens Clothing | Jacket | 16304.97 | 138 | 3.72 |
| Womens Clothing | Skirt | 15649.82 | 136 | 3.76 |
| Mens Clothing | T-Shirt | 15444.71 | 133 | 3.91 |
| Mens Clothing | Shirt | 14695.18 | 130 | 3.75 |
| Womens Clothing | Jeans | 13675.23 | 128 | 3.73 |
| Womens Clothing | Dress | 13610.16 | 123 | 3.79 |
| Electronics | Smart Watch | 191853.07 | 113 | 3.7 |
| Electronics | Headphones | 204660.34 | 112 | 3.74 |
| Mens Clothing | Jeans | 13235.57 | 110 | 3.84 |

Answer:

*There are Categories and Items that had more than 100 pieces but had review ratings lower than 4.0 points. The majority of the products are clothing and apparel.*

6. What level of discount is most effective in increasing customer purchases?

| discount_range_percent | total_revenue | total_order | AOV |
|---|---|---|---|
| 0-12 % | 516845.91 | 1807 | 286.02 |
| 13-24 % | 827127.85 | 2911 | 284.14 |
| 25-36 % | 81478.85 | 282 | 288.93 |

| discount_range_percent | total_revenue | total_order | AOV |
|---|---|---|---|
| 13-24 % | 827127.85 | 2911 | 284.14 |

Answer:

*13-24% of discount could make total revenue more than others discount percent groups*

7. Which payment method is associated with the highest customer satisfaction?

| payment_method | average_rating_method |
|---|---|
| Card | 3.79 |
| Cash on Delivery | 3.75 |

Answer:

*There was no significant difference in customer satisfaction between the two methods. However, it was evident that customer satisfaction was low with both methods. We should check our payment system for bugs or delays.*

8. Which season generates the highest business performance and deserves more investment?

| season | total_revenue | total_order | AOV |
|---|---|---|---|
| Spring | 376029.4 | 1300 | 289.25 |
| Winter | 354884.96 | 1244 | 285.28 |
| Summer | 350860.36 | 1225 | 286.42 |
| Autumn | 343677.89 | 1231 | 279.19 |

| season | total_revenue | total_order | AOV |
|---|---|---|---|
| Spring | 376029.4 | 1300 | 289.25 |

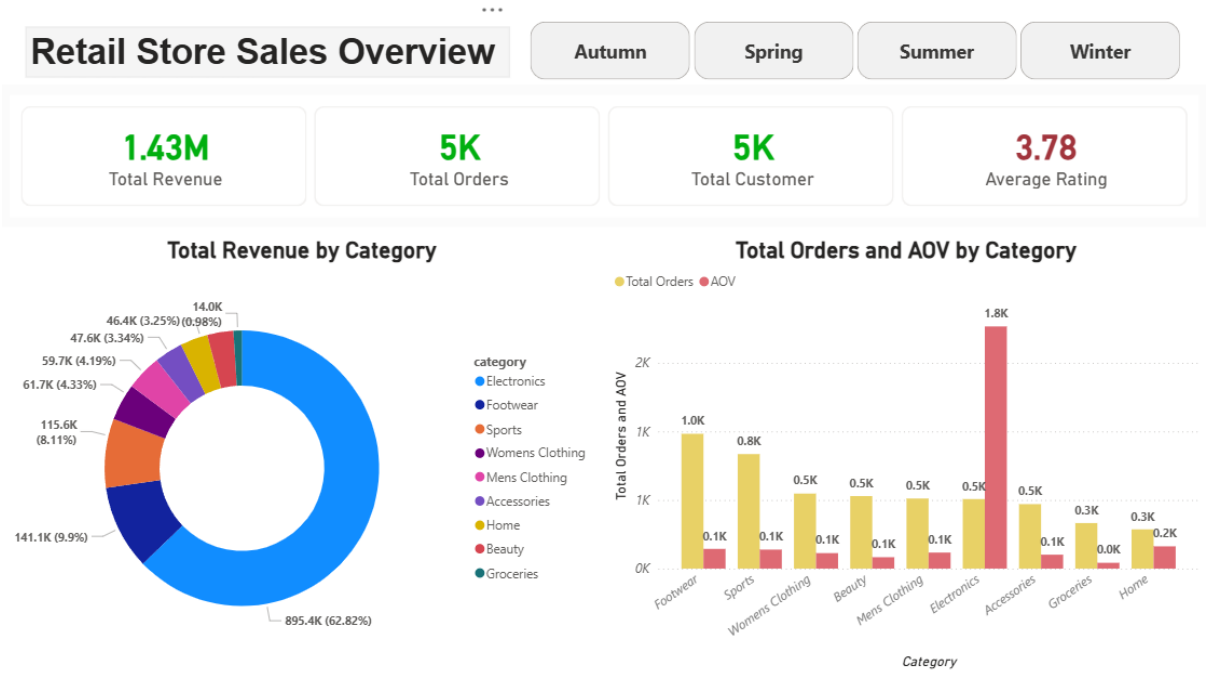
Answer:

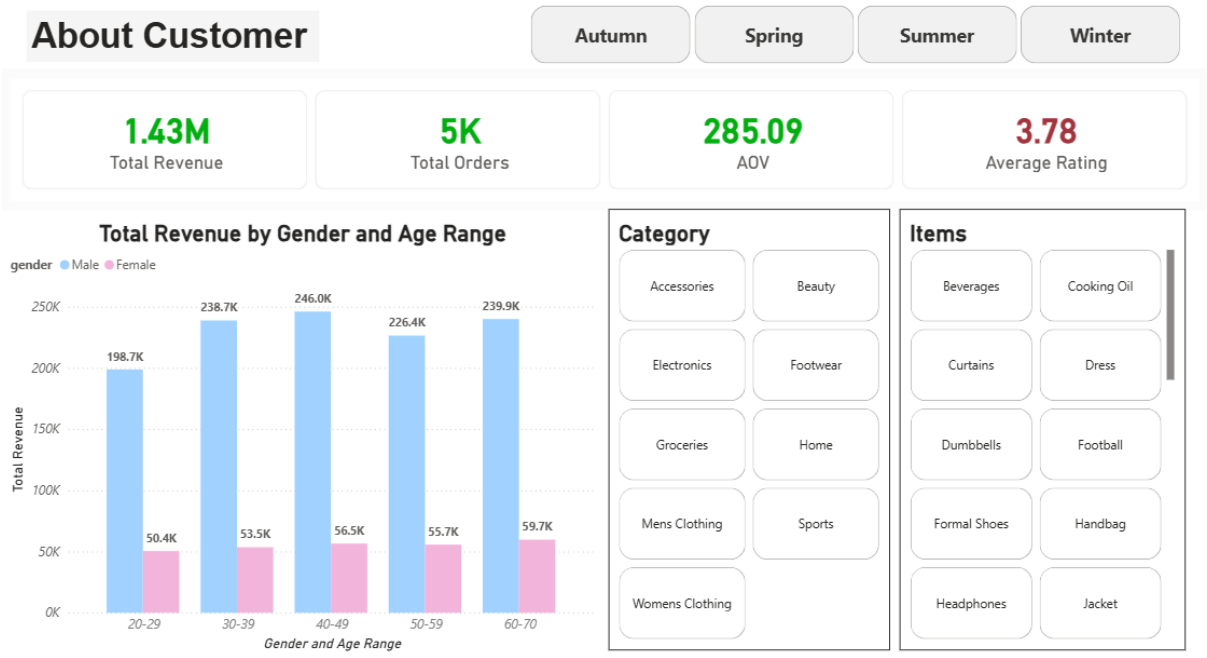
*We could make the most total revenue in the Spring season, However, there isn't much difference between seasons.*

## Suggestions:

*"We can observe that our business has relatively low customer satisfaction. This may be the reason why our business is stuck in some rut. We should examine every process that directly affects customer satisfaction, such as product quality, store operations, system issues or delays, logistics performance, and post-service customer response. Only then should we develop various marketing plans to increase sales."*

## 4.4. Make Dashboard using Microsoft Power BI



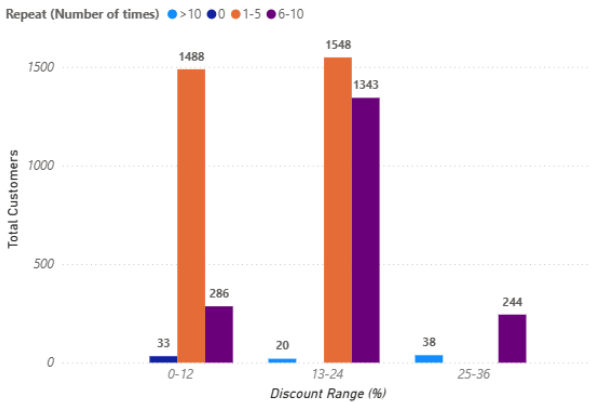



## About Marketing

Autumn | Spring | Summer | Winter

| 1.43M | 5K | 3.78 | 285.09 |
|---|---|---|---|
| Total Revenue | Total Orders | Average Rating | AOV |

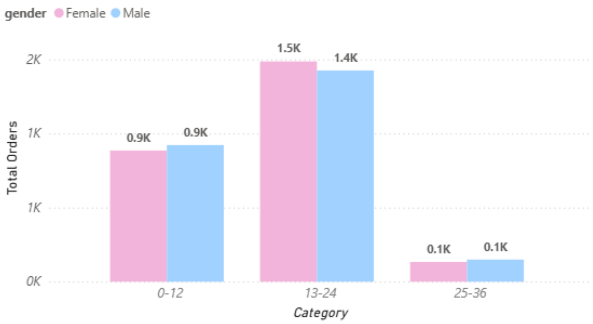
### Total Customers Repeat Order by Discount Range (%)




### Age Ranges

20-29 | 30-39 | 40-49 | 50-59 | 60-70

### Total Orders by Discount Rate (%)

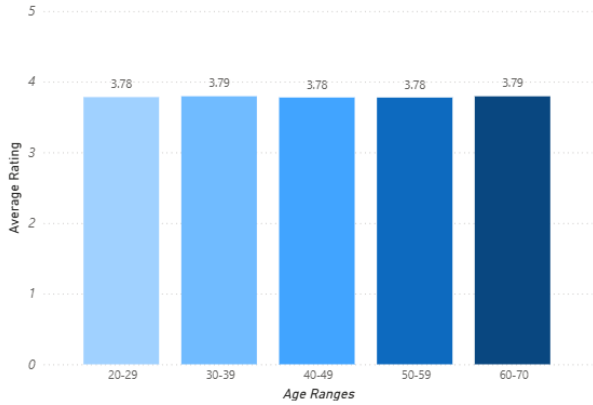



## Others

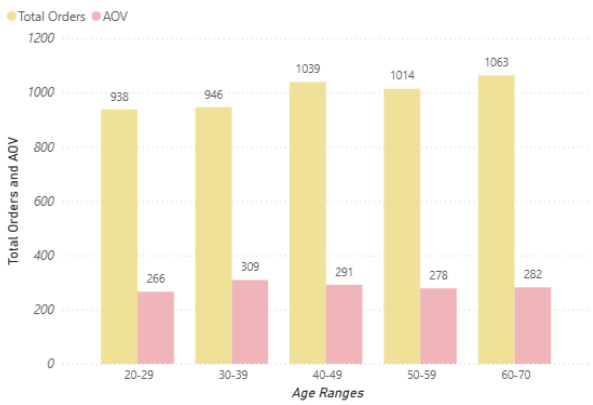
Autumn | Spring | Summer | Winter

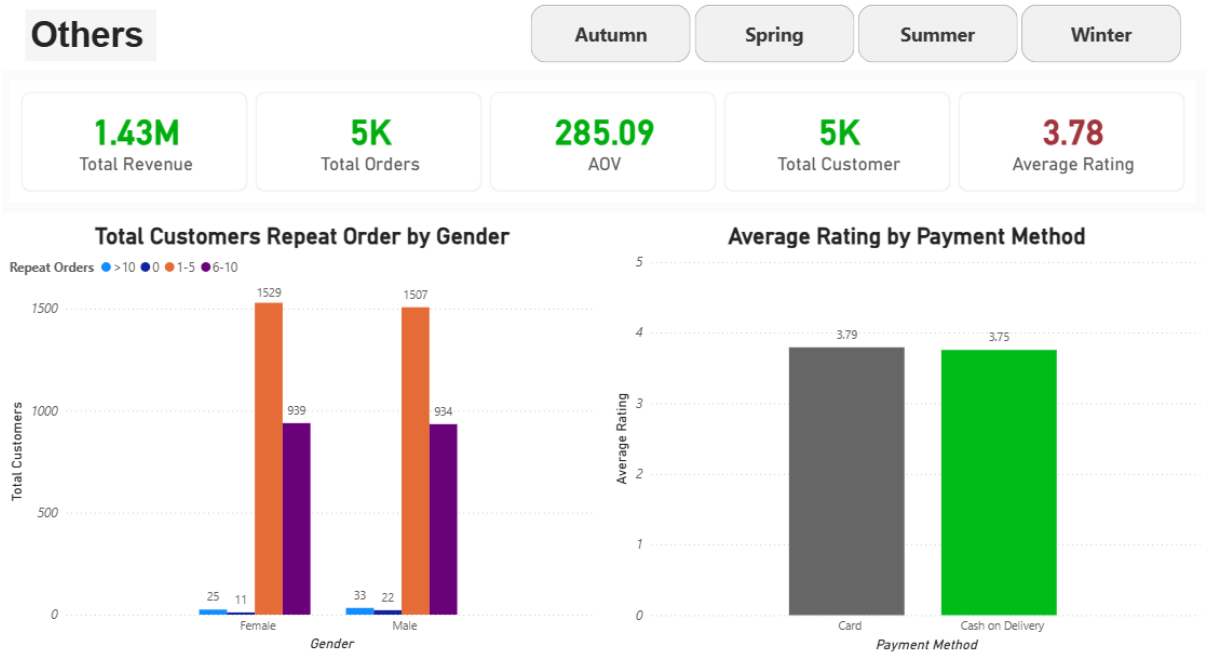
| 1.43M | 5K | 285.09 | 5K | 3.78 |
|---|---|---|---|---|
| Total Revenue | Total Orders | AOV | Total Customer | Average Rating |

### Average Rating by Age Ranges




### Total Orders and AOV by Age Ranges






# 5. Business Recomments

## 5.1 Prioritize high-value segment: Men aged 40–49

- Action: Launch targeted CRM (personalized offers, bundles, premium service) to increase purchase frequency and AOV.
- KPI: Revenue & AOV from segment, repeat-purchase rate.

## 5.2 Scale electronics assortment by season

- Action: Allocate inventory & marketing to seasonal top SKUs (mobile, smartwatch, laptop); offer accessory bundles and pre-season promotions.
- KPI: Category revenue by season, stockout rate, GM%.

## 5.3 Immediate CX & quality improvement (rating < 4.0)

- Action: Root-cause analysis (post-purchase survey, returns review), tighten QC for low-rating categories, improve returns/complaint SLA.
- KPI: Avg. item rating, CSAT/NPS, complaint resolution time, retention.

## 5.4 Optimize discounting (focus 13–24% cohort)

- Action: Run A/B tests for cart recovery and cohort discounts; measure revenue vs. margin impact before scale.

- KPI: Revenue lift per discount cohort, margin impact, conversion uplift.

## 5.5 Remediate fashion category (high volume, low rating)

- Action: Audit suppliers/QC, improve product pages (images/size guidance), simplify returns.

- KPI: Rating & return rate in apparel, repeat purchase in category.

## 5.6 Monitor with a weekly KPI dashboard

- Action: Build Power BI dashboard showing revenue by segment, AOV, repeat rate, ratings by category, discount cohort performance, cart abandonment & payment failures. Set alerts for KPI drops.

- KPI: Dashboard availability + leading metrics above.