Overview

This analysis examines customer's behaviour for an e-commerce company using transactional data from thousands of across multiple product categories. The objective is to uncover insights into *spending patterns, customer segmentation, product preferences, and subscription trends*. The findings support *data-driven decision-making, optimized marketing strategies, and improved customer engagement and retention.*

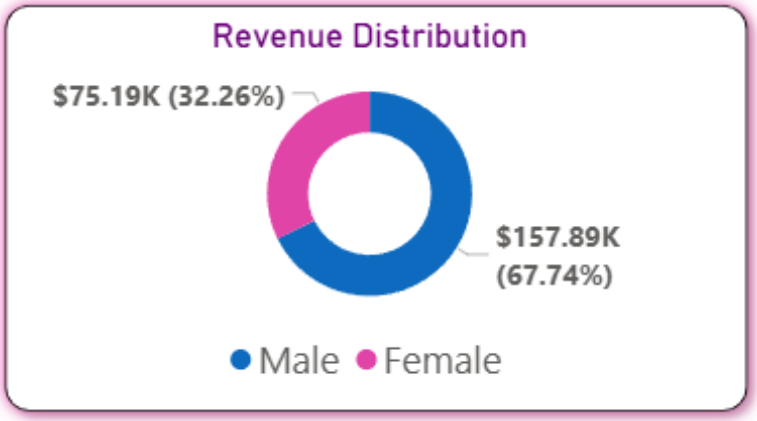
Problem Statement

A leading retail company wants to better understand its customers' shopping behaviour in order to improve sales, customer satisfaction, and long-term loyalty. The management team has noticed changes in purchasing patterns across demographics, product categories, and sales channels (online vs. offline). They are particularly interested in uncovering which factors, such as discounts, reviews, seasons, or payment preferences, drive consumer decisions and repeat purchases.

Insights From Analysis

1. **Revenue by different Genders**
   - ❖ Compared total revenue generated and Top Categories of products bought by male and female customers.

- Female customers generated $ 75,191 in revenue , which is over 32.27 % of the Total Revenue generated.
- The average amount spent on the purchase by the female customers is $ 60.25.
- The top category of items purchased by female customers is 'Clothing', followed by 'Accessories' and 'Footwear'. Almost 44.5% Female customers purchased items of 'Clothing' category And 31.4% Purchased items of 'Footwear' category.



- Male customers generated over $1,57,000 in revenue , which is over 67.72 % of the Total Revenue generated.
- The average amount spent on the purchase by the Male customers is $ 59.25 which is almost same as female customers.
- The top category of items purchased by female customers is ‘Clothing’, followed by ‘Accessories’ and ‘Footwear’. Almost 45% Male customers purchased items of ‘Clothing’ category And 31% Purchased items of ‘Footwear’ category.

2. Top 5 Products by Rating

❖ Found products with the highest average review ratings.

item_purchased	Average Product Rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.8
Skirt	3.78

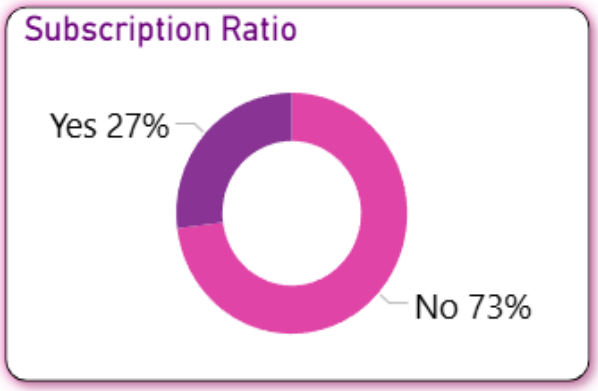
- The product with the Highest Average review Ratings is ‘Gloves’ with average rating of 3.86 .

- The product with 2nd Highest Average review Ratings is ‘Sandals’ with average rating of 3.84 .
- The list continues to ‘Boots’ with the 3rd highest average rating Of 3.82 , And ‘Hat’ with the 4th highest average rating of 3.8, Followed by ‘Skirt’ with the 5th highest average rating of 3.78.

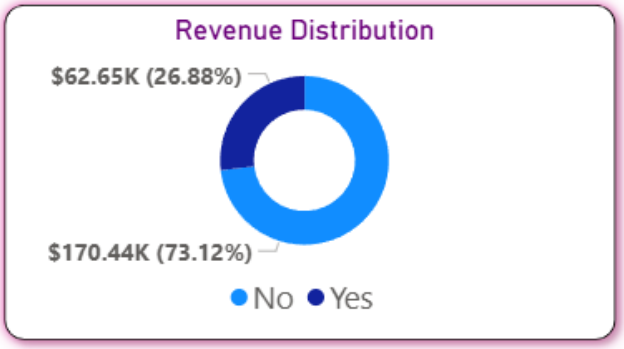
3. Subscribers vs. Non-Subscribers

❖ Compared average spend and total revenue across subscription status.

- 27% of the total customers have bought the subscription.



- Revenue Distribution between Subscribers and Non-Subscribers:



- Important insights of customers with subscription :

- Important insights of customers with No subscription :

4. **Customer Segmentation**
   - ❖ Classified customers into New, Returning, and Loyal segments based on purchase history.

	customer_segment text	Number of Customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

- The segmentation is based on number of orders placed by the customer.
- If no. of purchases = 1 , it's a 'New' Customer.
- If no. of purchases is between 2 and 10 , it's a 'Returning' Customer.
- If no. of purchases >= 10 , it's a 'Loyal' Customer.

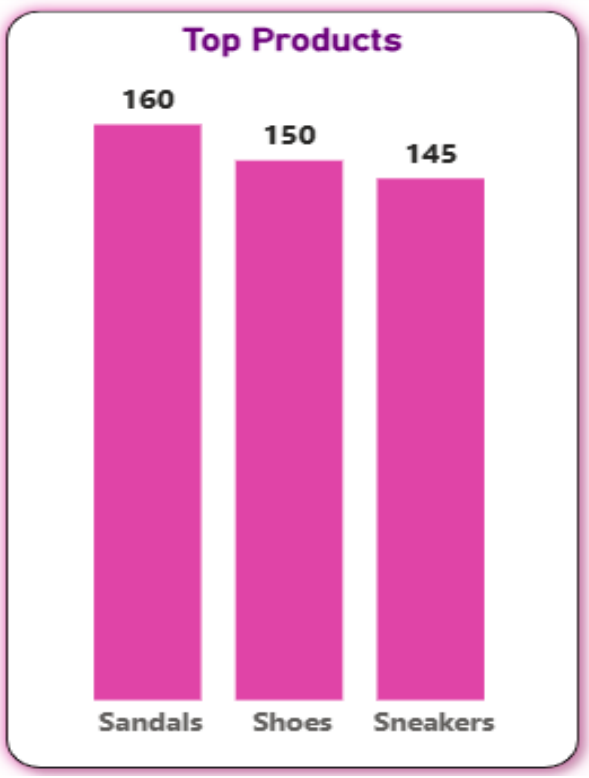
5. Top 3 Products per Category

❖ Listed the most purchased products within each category:

- Accessories

- Clothing

- Footwear



	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

6. Discount-Dependent Products

- ❖ Identified 5 products with the highest percentage of discounted purchases.

	item_purchased text	discount_rate numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

7. Shipping Type Comparison

- ❖ Compared average purchase amounts between Standard and Express shipping.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

8. Discount-Dependent Products

❖ Identified 5 products with the highest percentage of discounted purchases.

	item_purchased text	discount_rate numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

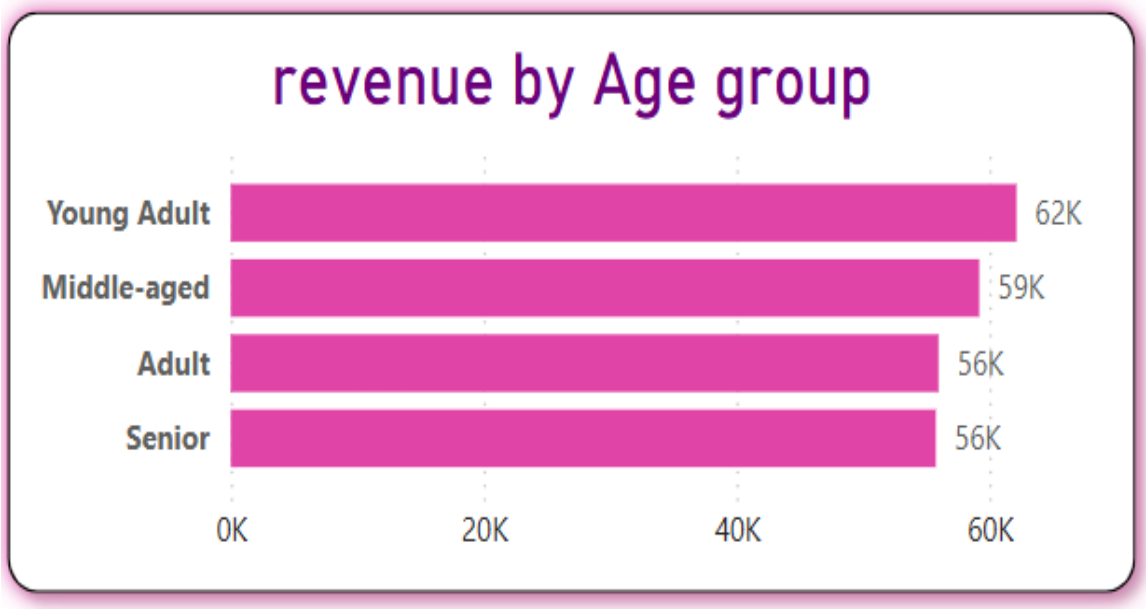
9. Repeat Buyers & Subscriptions

❖ Checked whether customers with >5 purchases are more likely to subscribe.

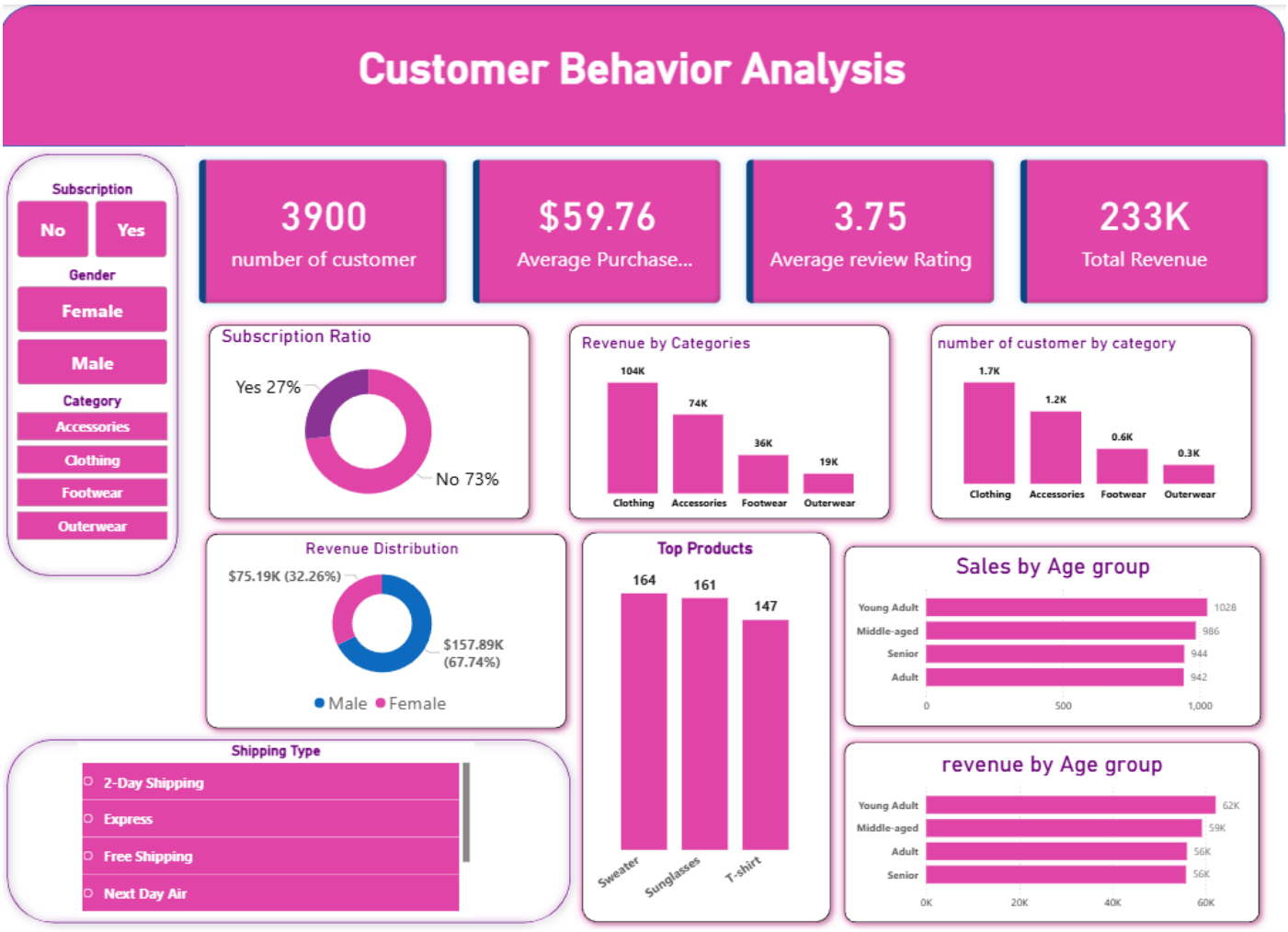
	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

10. Revenue by Age Group

❖ Calculated total revenue contribution of each age group.



Dashboard For Visualization



Key Performance Indicators (KPIs)

These are the main metrics used to measure customer behavior and business performance in the analysis:

1. **Total Revenue by Gender** – Male: $157,000 (67.7%), Female: $75,191 (32.3%)
2. **Average Order Value (AOV)** – Male: $59.25, Female: $60.25
3. **Top Performing Product Categories** – Clothing, Accessories, and Footwear
4. **Subscription Rate** – 27% of customers are subscribers
5. **Customer Segmentation:**
    - New Customers: 1 purchase
    - Returning Customers: 2–10 purchases
    - Loyal Customers: 10+ purchases
6. **Top Rated Products** – Gloves (3.86), Sandals (3.84), Boots (3.82), Hat (3.80), Skirt (3.78)
7. **Revenue by Age Group** – Identified key age segments contributing the most to total sales
8. **Shipping Type Comparison** – Comparison of average spend for Standard vs. Express users
9. **Discount Dependency** – Products most influenced by discounts identified

Business Recommendations

- **Boost Subscriptions** – Promote exclusive benefits for subscribers.
- **Customer Loyalty Programs** – Reward repeat buyers to move them into the “Loyal” segment.

- **Review Discount Policy** – Balance sales boosts with margin control.
- **Product Positioning** – Highlight top-rated and best-selling products in campaigns.
- **Targeted Marketing** – Focus efforts on high-revenue age groups and express-shipping users.

Dataset Summary

- Rows: 3,900
- Columns: 18
- Key Features:
  - *Customer demographics* (Age, Gender, Location, Subscription Status)
  - *Purchase details* (Item Purchased, Category, Purchase Amount, Season, Size, Colour)
  - *Shopping behaviour* (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
- Missing Data: 37 values in Review Rating column

Exploratory Data Analysis (EDA)

❖ We began with data preparation and cleaning in Python:

- **Data Loading:** Imported the dataset using pandas.
- **Initial Exploration:** Used df.info() to check structure and .describe() for summary statistics.
- **Missing Data Handling:** Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.

- **Column Standardization:** Renamed columns to snake case for better readability and documentation.
- **Feature Engineering:**
    - Created ‘age_group’ column by binning customer ages.
    - Created ‘purchase_frequency_days’ column from purchase data.
- **Data Consistency Check:** Verified if ‘discount_applied’ and ‘promo_code_used’ were redundant; dropped ‘promo_code_used’.
- **Database Integration:** Connected Python script to PostgreSQL and loaded the cleaned *DataFrame* into the database for SQL analysis.