Key KPIs:

1. Total Revenue: Sum of the Amount or Total_Amount column.
2. Total Transactions: Count of unique Transaction_ID.
3. Average Transaction Value: Total_Amount / Total_Purchases or Total_Amount / count of Transaction_ID.
4. Average Customer Age: Average of the Age column.
5. Customer Retention Rate: Percentage of returning customers (those with multiple Transaction_ID entries) relative to all customers.
6. Product Popularity: Count of purchases by Product_Category or Product_Brand to find the most popular items.
7. Order Fulfillment Rate: Percentage of Order_Status marked as "Completed" relative to total orders.
8. Customer Satisfaction: Average of Ratings column or analysis of the Feedback column if it's qualitative.
9. Revenue by Customer Segment: Sum of Amount for each Customer_Segment.

GitHub Link:

git@github.com:kumkumbaswal003/Cloudthat_Project.git

AZURE DATA PIPELINE

For TechRetail

Group 4:
- Arnab Saha
- Debadrita Acharjee
- Kumkum Baswal

Background:

- **TechRetail**, a mid-sized retail company, wants to create a data pipeline retail data from various sources, process it using advanced analytics, and visualize the results in a dashboard. The goal is to gain insights into sales trends and improve decision-making. The company wants to leverage Azure Databricks for data processing and Microsoft Fabric for data integration and visualization.

Objectives:
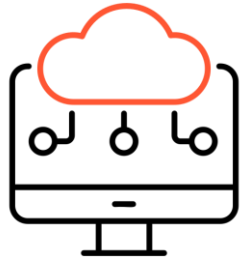
Data Ingestion

Azure Data Factory (ADF)

ADLS Gen2

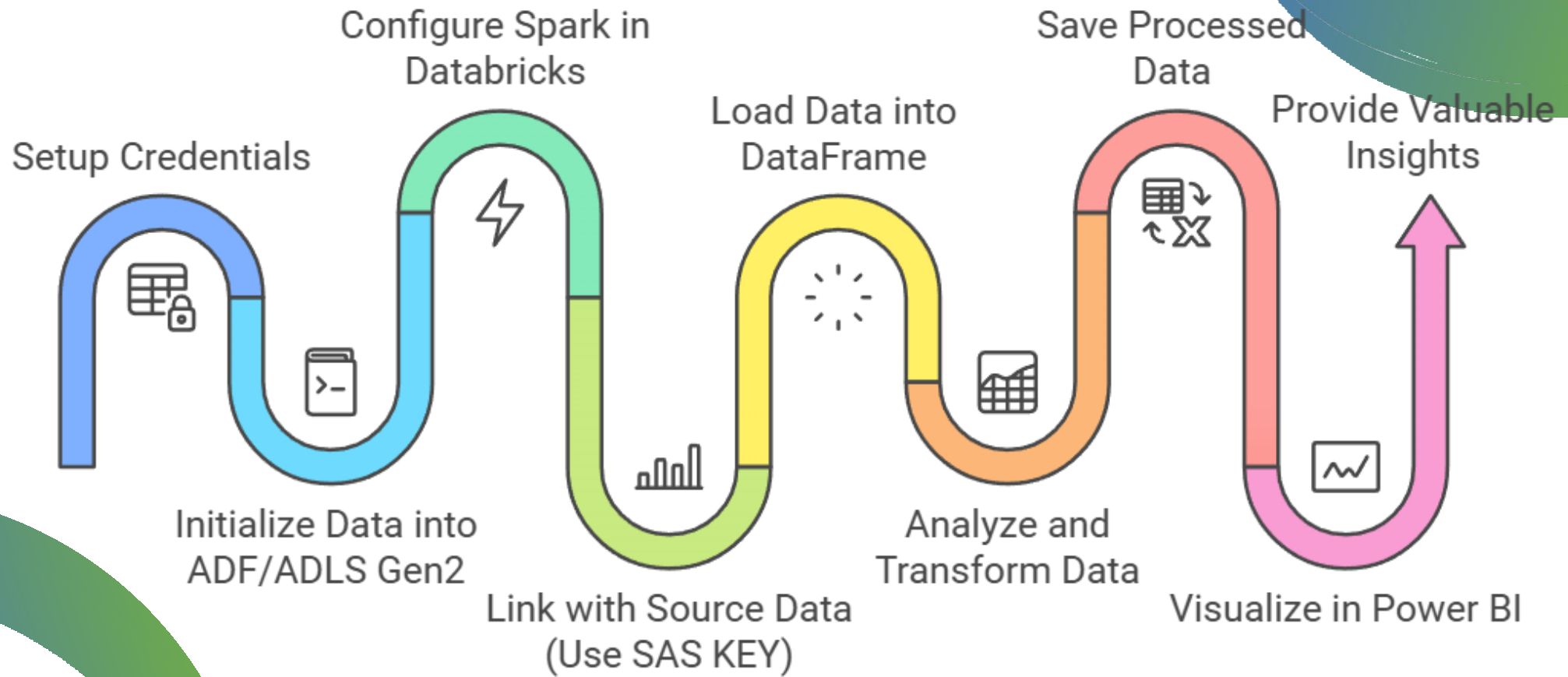**Data Processing**

Synapse Data Engineering

Databricks

Data Storage

ADLS Delta Table

Data Visualization

Microsoft PowerBI
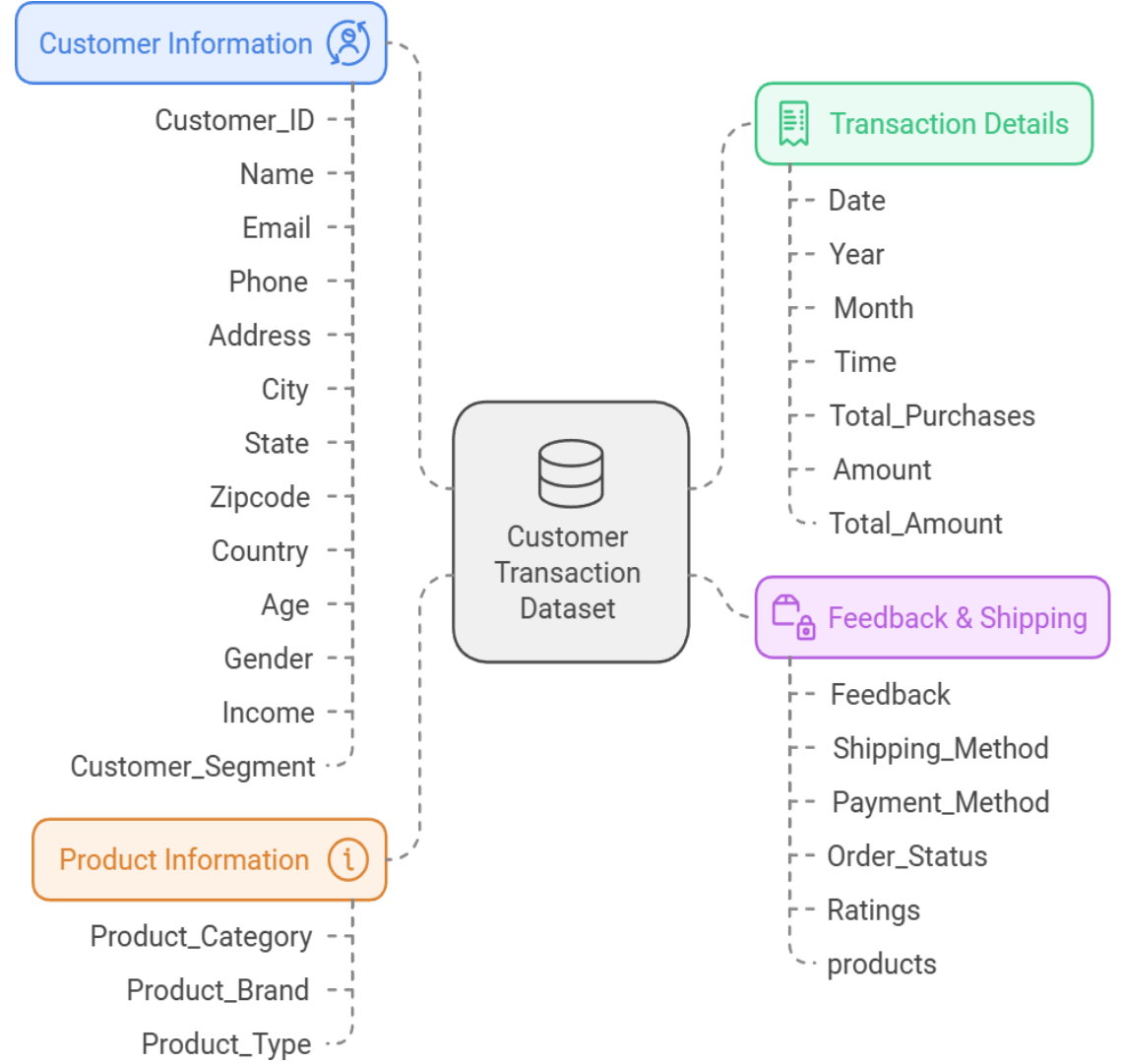
Architectural Process Flow:

Setup Credentials

Configure Spark in Databricks

Load Data into DataFrame

Save Processed Data

Provide Valuable Insights

Initialize Data into ADF/ADLS Gen2

Link with Source Data (Use SAS KEY)

Analyze and Transform Data

Visualize in Power BI

KNOW YOUR DATA

Dataset Insights

DATASET OVERVIEW

Customer Information
- Customer_ID
- Name
- Email
- Phone
- Address
- City
- State
- Zipcode
- Country
- Age
- Gender
- Income
- Customer_Segment

Product Information
- Product_Category
- Product_Brand
- Product_Type

Customer Transaction Dataset

Transaction Details
- Date
- Year
- Month
- Time
- Total_Purchases
- Amount
- Total_Amount

Feedback & Shipping
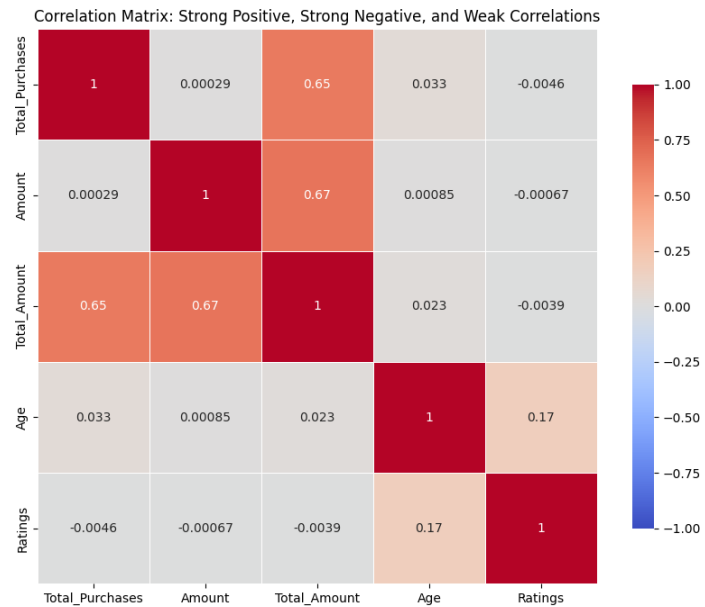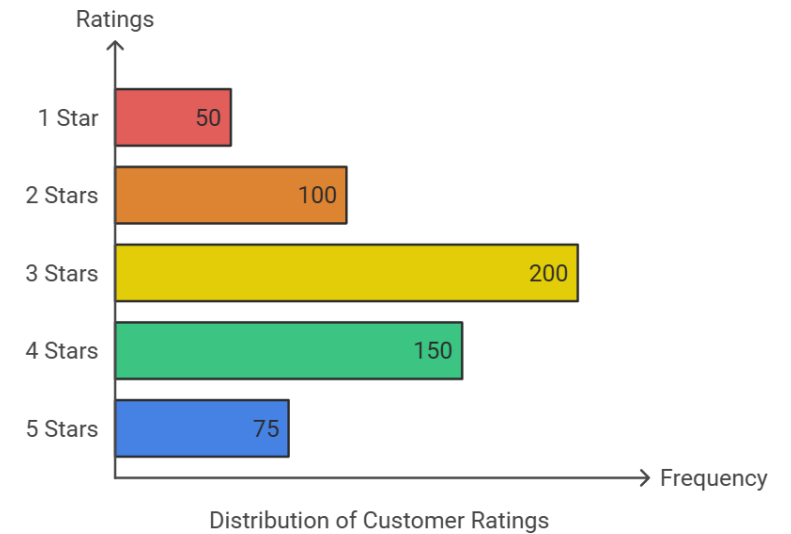- Feedback
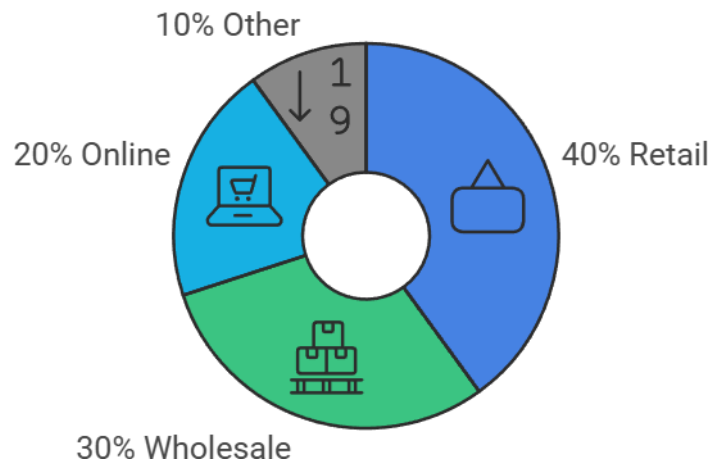- Shipping_Method
- Payment_Method
- Order_Status
- Ratings
- products

Data Cleaning and Transformation Process

Raw Dataset

Handle Missing Values

Convert Data Types

Remove Unnecessary Columns

Encode and Scale Data (Optional)

Cleaned and Transformed Dataset
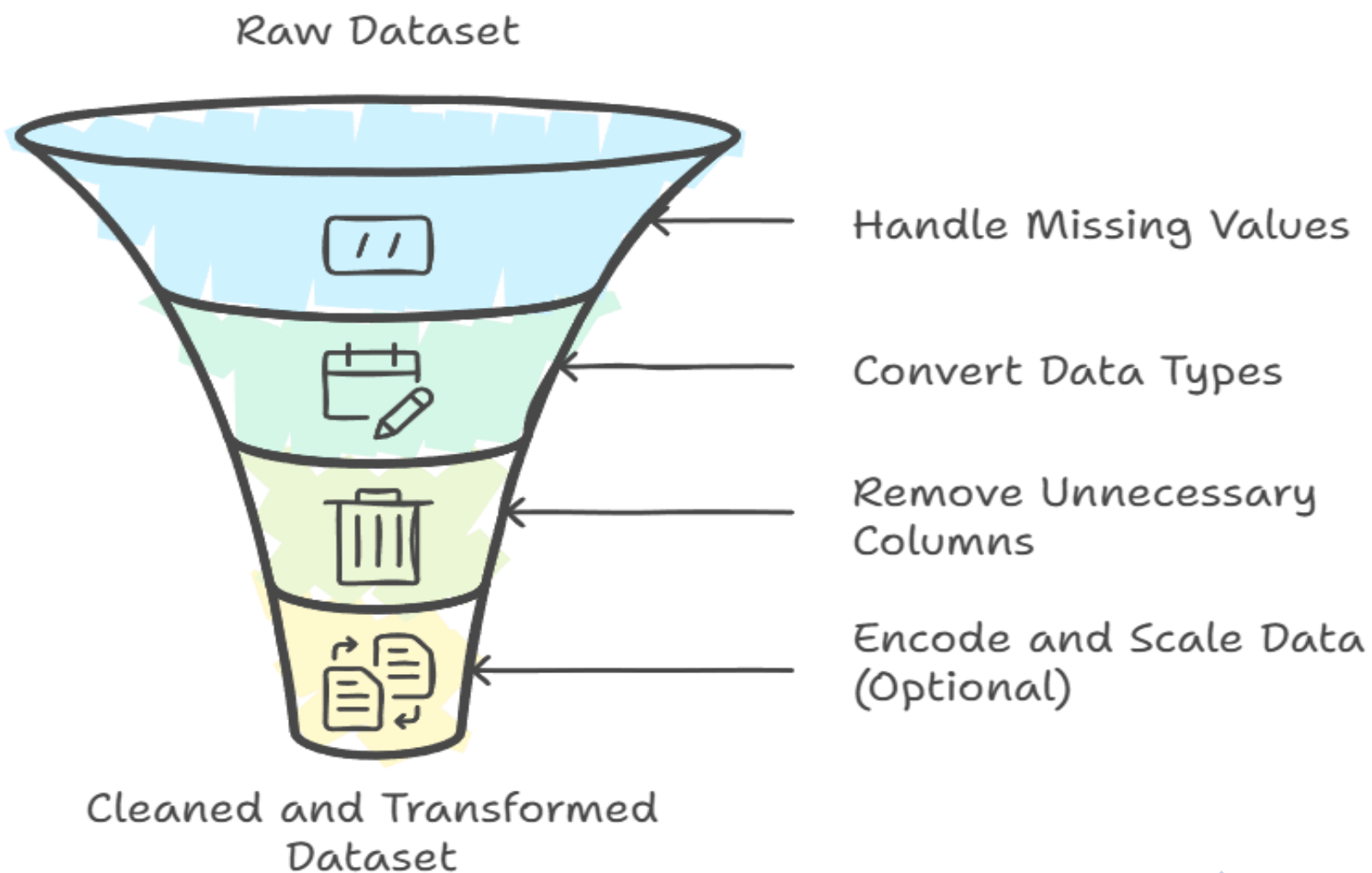
Dataset Insights Visualization

Correlation Matrix: Strong Positive, Strong Negative, and Weak Correlations

	Total_Purchases	Amount	Total_Amount	Age	Ratings
Total_Purchases	1	0.00029	0.65	0.033	-0.0046
Amount	0.00029	1	0.67	0.00085	-0.00067
Total_Amount	0.65	0.67	1	0.023	-0.0039
Age	0.033	0.00085	0.023	1	0.17
Ratings	-0.0046	-0.00067	-0.0039	0.17	1

Distribution of Customer Segments

- 40% Retail
- 30% Wholesale
- 20% Online
- 10% Other

Distribution of Customer Ratings

- 1 Star: 50
- 2 Stars: 100
- 3 Stars: 200
- 4 Stars: 150
- 5 Stars: 75

Key Distributions In The Data

Total Purchases Distribution

Total Amount Distribution
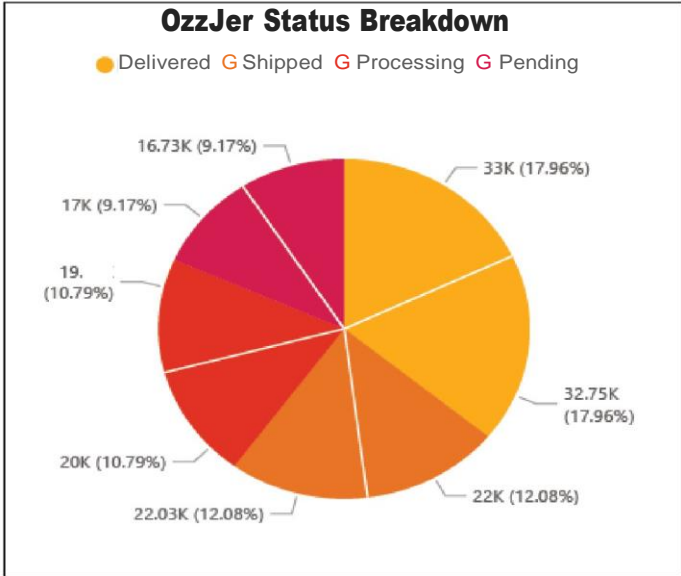
Product Category Distribution

Customer Segment Distribution
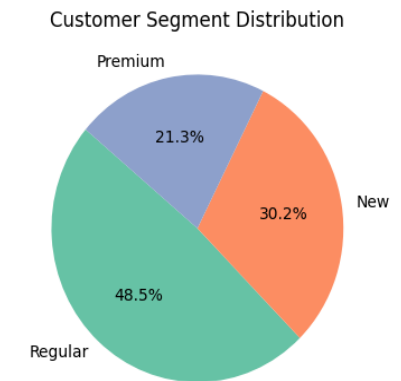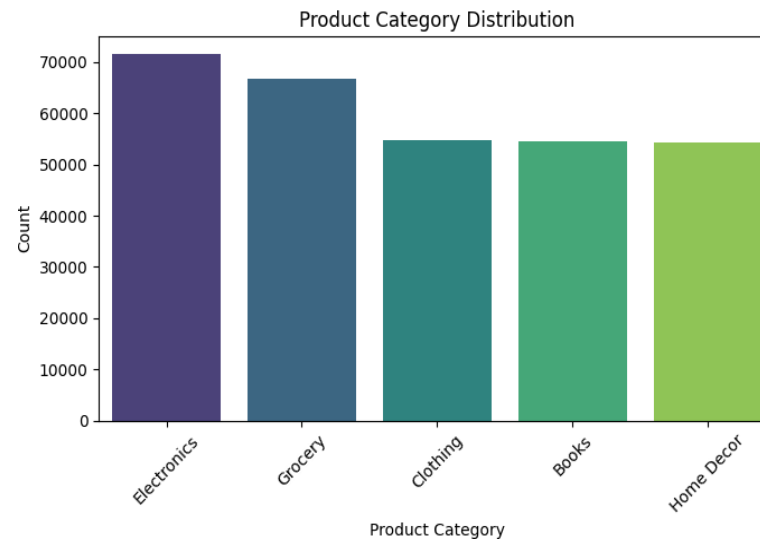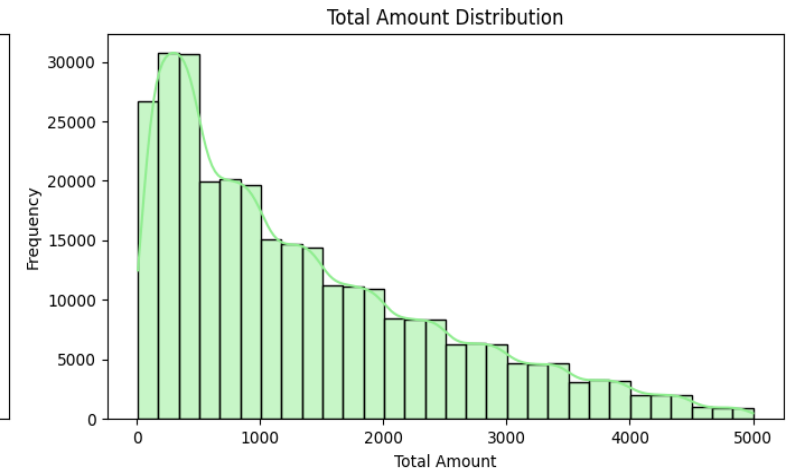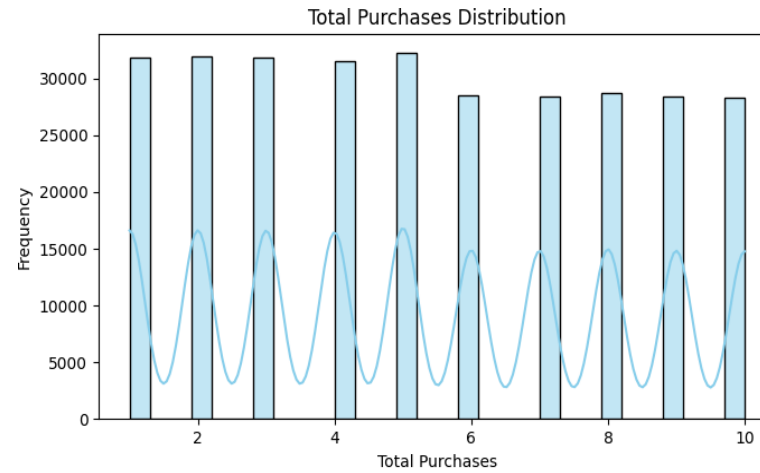
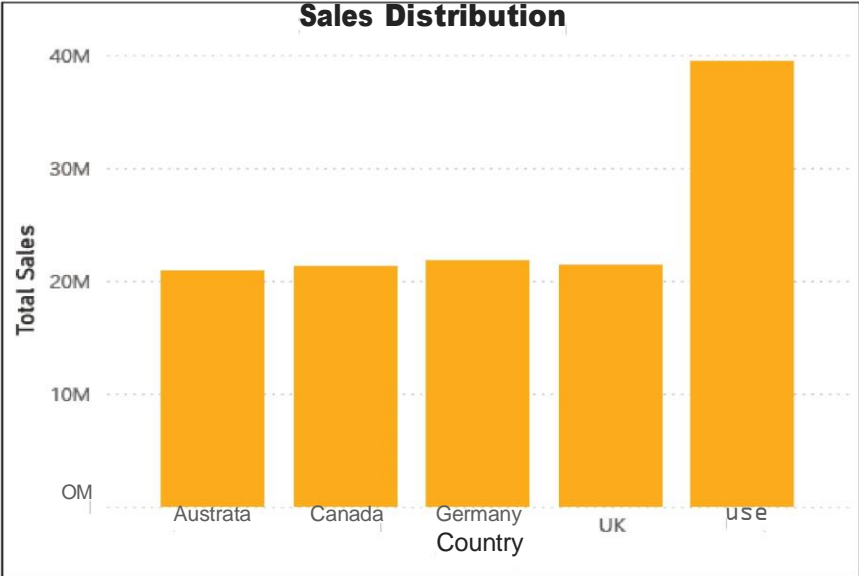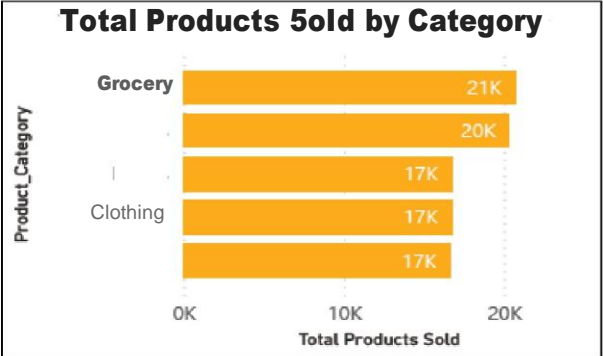# TechRetail Sales Analytics Dashboard

TechRetaiI Sales Analysis

Key Metrics for Business Insights

Total Revenue	Total Transactions
124.70M	90K

ATV	Average Age
J.38K	35.83

Monthly 5ale Tzends

Revenue Per Quarter

Customer Retention Rat:e

0.42

Total Products 5old by Category

- Grocery — 21K
- 20K
- 17K
- Clothing — 17K
- 17K

Total Products Sold

Date	Customer_Seg.
All	New

Product_Catego...	Country, City, St...
All	All

Sales Distribution

Country: Austrata, Canada, Germany, UK, use

OzzJer Status Breakdown

● Delivered G Shipped G Processing G Pending

- 16.73K (9.17%)
- 33K (17.96%)
- 17K (9.17%)
- 19. (10.79%)
- 32.75K (17.96%)
- 20K (10.79%)
- 22K (12.08%)
- 22.03K (12.08%)

THANK YOU