Global E-Commerce Analysis Report
by Ayeni Joshua Adaviriku

1. EXECUTIVE SUMMARY

Project Overview

This project explores a global e-commerce dataset to generate actionable business insights using three analytical tools: MySQL, Python, and Power BI. The data represents transactional records, including sales, product Id, categories, customer IDs, payment methods, and regional information.

The analysis follows a structured 10-day workflow, focusing on three phases:
i.   Data exploration and querying using MySQL
ii. Data enrichment, statistical analysis, and visualizations in Python
iii. Dashboard development and storytelling using Power BI

The ultimate goal is to uncover patterns in sales, customer behavior, product performance, discount effectiveness, and regional opportunities to support data-driven business decisions.

Key Business Questions

The analysis was driven by several business-critical questions:
i.   What are the monthly revenue and sales trends?
ii.   Who are the top-performing customers and products?
iii. Which regions generate the most revenue?
iv.   How do discounts influence purchasing behavior?
v.   What are the most common and profitable payment methods?
vi.   Are customers repeat buyers or mostly one-time?
vii. How can customers be segmented for better targeting?
viii. Which categories drive the most volume vs. profit?

Summary of Insights

i.   Revenue peaked in specific months, aligning with potential seasonal or promotional spikes.
ii. Top 10 customers contributed a disproportionately high percentage of total revenue, highlighting the importance of loyalty programs.
iii. Most revenue was generated from a handful of regions and product categories, indicating key market drivers.
iv. Higher discounts increased quantity sold but negatively impacted overall revenue and profitability.
v.   Bank Payment and PayPal dominated as payment preferences, with variations in revenue-per-transaction.
vi. The customer base was a mix of one-time and repeat buyers, opening opportunities for segmentation and re-engagement.
vii. Four customer segments were identified: Low, Medium, High, and VIP, based on total spending.

Tools Used
i.   MySQL: Data import, cleaning, and exploratory SQL queries.
ii. Python (Pandas, Seaborn, Matplotlib): In-depth analysis, feature engineering, and trend visualizations.
iii. Power BI: Development of executive dashboards and interactive reporting for stakeholders.

2. DATA OVERVIEW

Source & Description
The dataset was sourced from Kaggle:
Title: Global E-Commerce Sales
Link: Global E-Commerce Sales Dataset on Kaggle
It represents sales transactions from an e-commerce platform, encompassing various aspects such as customer behavior, product performance, pricing strategy, and payment preferences. The dataset provides a holistic view ideal for operational, strategic, and marketing analysis.

Data Fields (Column Details)

S/No	Column Name	Description
i.	Transaction Date	Date of transaction in yyyy-mm-dd format. Used to extract time-based trends
ii.	Customer ID	Unique identifier for each customer. Useful for segmentation and retention.
iii.	Region	Geographical region of the purchase. Used to assess location-based performance.
iv.	Product ID	Unique identifier for each product sold.
v.	Category	Product Category. Allows aggregation of performance across product types.
vi.	Price	Price per unit of the product.
vii.	Quantity	Number of units purchased in the transaction
viii.	Discount	Discount applied (percentage). Useful for price sensitivity and impact studies.
ix.	Total Revenue	Total value of the transaction after discount.
x.	Payment Method	Method used to complete the purchase. Useful for user behavior analysis.

Data Cleaning & Preparation
a) Data Consistency Checks
Verified column types:
- Dates converted to datetime format
- Prices and revenue converted to numeric types
- Missing values assessed for all fields, and none was found

b) Duplicate Removal
Checked duplicates and none was found

c) Derived Columns (In Python)
- Month-Year: Extracted from transaction_date to analyze monthly trends.
- Customer Segment: Created using total spend thresholds (Low, Medium, High, VIP).
- Revenue per Product: Calculated where necessary for analysis and dashboard visuals.

d) Export for Multi-Tool Analysis
After initial preparation in MySQL and Python, the cleaned dataset was exported to CSV and loaded into Power BI for final visualization and reporting.

3. MySQL ANALYSIS

Database Setup & Import Process
The initial phase of the project involved structuring the dataset for SQL querying and relational data operations using MySQL.
Steps Taken:
a) Table Creation:
A table named customers was created to match the dataset's structure.

```
CREATE TABLE customers (
  transaction_date DATE,
  customer_id VARCHAR(50),
  region VARCHAR(100),
  product_id VARCHAR(50),
  category VARCHAR(100),
  price DECIMAL(10,2),
  quantity INT,
  discount DECIMAL(5,2),
  total_revenue DECIMAL(10,2),
  payment_method VARCHAR(50)
);
```

b) Data Upload:
The dataset was saved as a .csv file and imported using table import wizard

Key SQL Queries & Insights
Below are selected SQL queries and the insights they generated:
i. Total Revenue Over Time

```
SELECT DATE_FORMAT(transaction_date, '%Y-%m') AS month,
  SUM(total_revenue) AS monthly_revenue FROM customers
GROUP BY month
ORDER BY month;
```

Insight: Revenue fluctuated across months. Showing peak sales, seasons, and potential marketing campaign effects.
ii. Top 10 Customers by Revenue

```
SELECT customer_id, SUM(total_revenue) AS total_spent
FROM customers
GROUP BY customer_id
ORDER BY total_spent DESC
LIMIT 10;
```

Insight: A small subset of customers generated a large portion of revenue valuable for loyalty and VIP targeting.
iii. Revenue by Region

```
SELECT region, SUM(total_revenue) AS regional_revenue
FROM customers
GROUP BY region
ORDER BY regional_revenue DESC;
```

Insight: Certain regions far outperformed others, suggesting where marketing or logistics efforts could be focused.

4. Category Performance
SELECT category, SUM(total_revenue) AS category_revenue, SUM(quantity) AS total_units_sold
FROM customers
GROUP BY category
ORDER BY category_revenue DESC;
Insight: Identified high-performing product categories both in terms of revenue and volume.

5. Payment Method Trends
SELECT payment_method, COUNT(*) AS transaction_count, SUM(total_revenue) AS method_revenue
FROM customers
GROUP BY payment_method
ORDER BY method_revenue DESC;
Insight: Bank Transfer and PayPal dominated, but revenue-per-transaction varied significantly.

SQL Insights Summary
i.   Time-based revenue trends were evident, enabling the identification of high and low-performing months.
ii. Top customers contributed disproportionately to overall sales — strong candidates for targeted engagement.
iii. Geographic disparities in revenue provided direction for market expansion or localized strategies.
iv. Discounting behavior could be isolated for future pricing experiments.
v.   Payment method preferences gave insights into transaction friction and user experience optimization.

4. PYTHON ANALYSIS

After initial querying and data structure validation in MySQL, the dataset was exported to CSV and loaded into Python for deeper exploration, transformation, and visualization using Pandas, NumPy, Matplotlib, and Seaborn.

a) Data Transformation & Feature Engineering
Import & Setup

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("ecommerce_cleaned.csv")
df['transaction_date'] = pd.to_datetime(df['transaction_date'])
df['month'] = df['transaction_date'].dt.to_period('M')
```

New Features Created
i.   Month-Year (month): For time series analysis.
ii. Revenue per Unit: df['unit_revenue'] = df['total_revenue'] / df['quantity']
iii. Discount Range Category: Grouped into bins (e.g., 0-10%, 10-20%).
iv. Customer Segment:

```
def segment(revenue):  if revenue < 100: return "Low"
    elif revenue < 500: return "Medium"
    elif revenue < 1000: return "High"
    return "VIP"
customer_revenue = df.groupby("customer_id")["total_revenue"].sum().reset_index()
customer_revenue["segment"] = customer_revenue["total_revenue"].apply(segment)
```
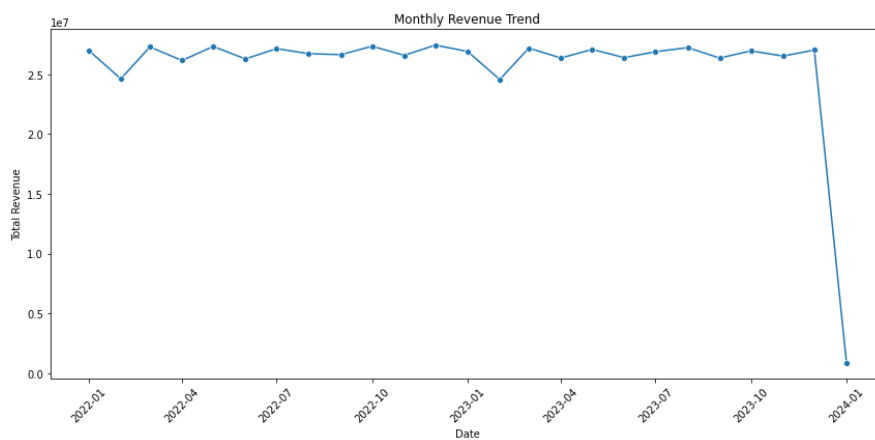
b) Exploratory Data Analysis (EDA)

i. Revenue & Sales Trends
Line plot of revenue by month:

```
monthly = df.groupby("month")["total_revenue"].sum()
monthly.plot(title="Monthly Revenue Trend")
```
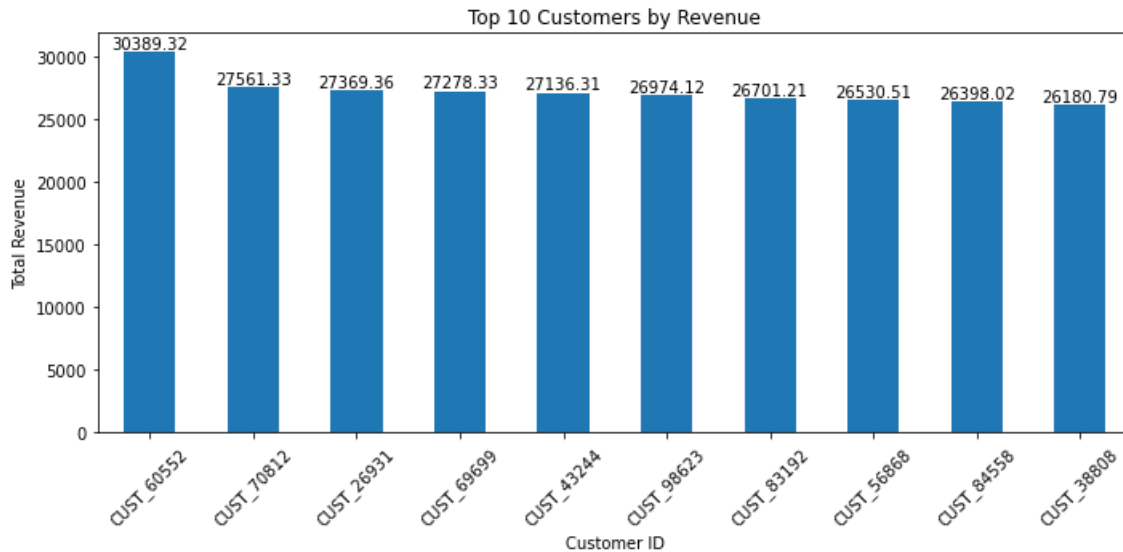

Insight: Seasonality observed, suggesting influence of promotions or holidays.

ii. Customer Behavior
Top 10 customers by revenue.

Average order value (AOV) per customer.

df.groupby('customer_id')['total_revenue'].sum().nlargest(10).plot(kind='bar')
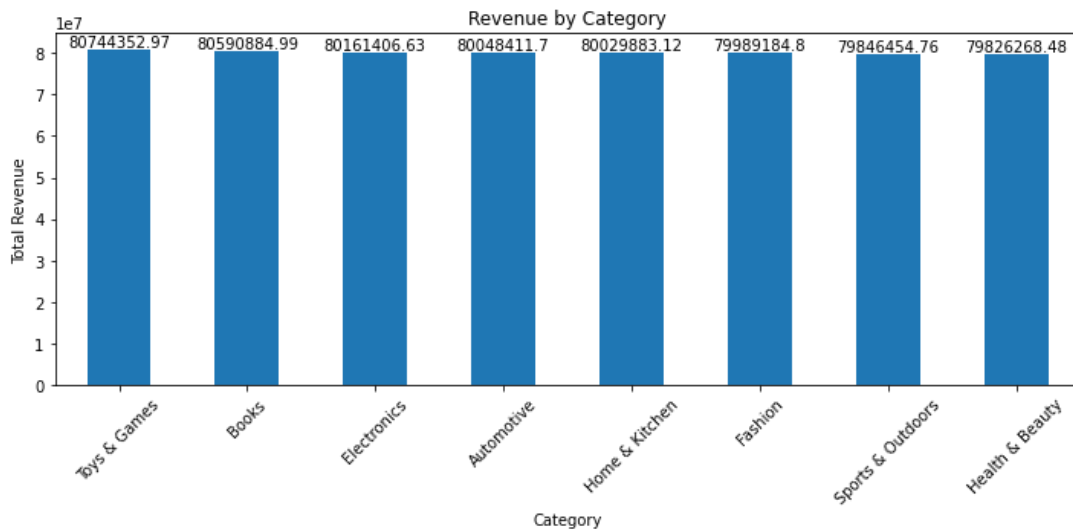

Insight: Revenue is skewed toward a few high-value customers.

iii. Product Performance

Quantity sold and revenue by category.

category_perf = df.groupby("category")[["quantity", "total_revenue"]].sum().sort_values("total_revenue", ascending=False)
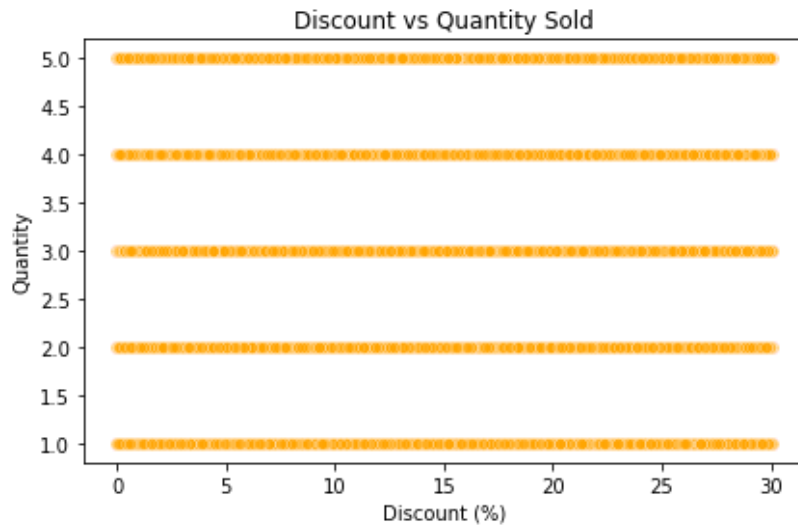

Insight: Some categories drive high volume but lower unit revenue, and vice versa.

iv. Discount Analysis

Impact of discount percentage on quantity sold.

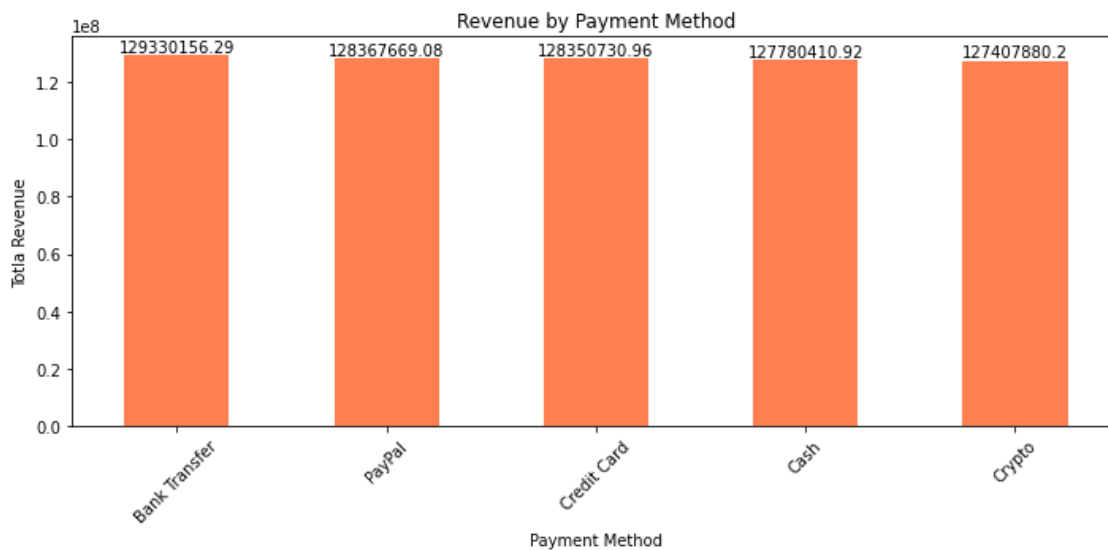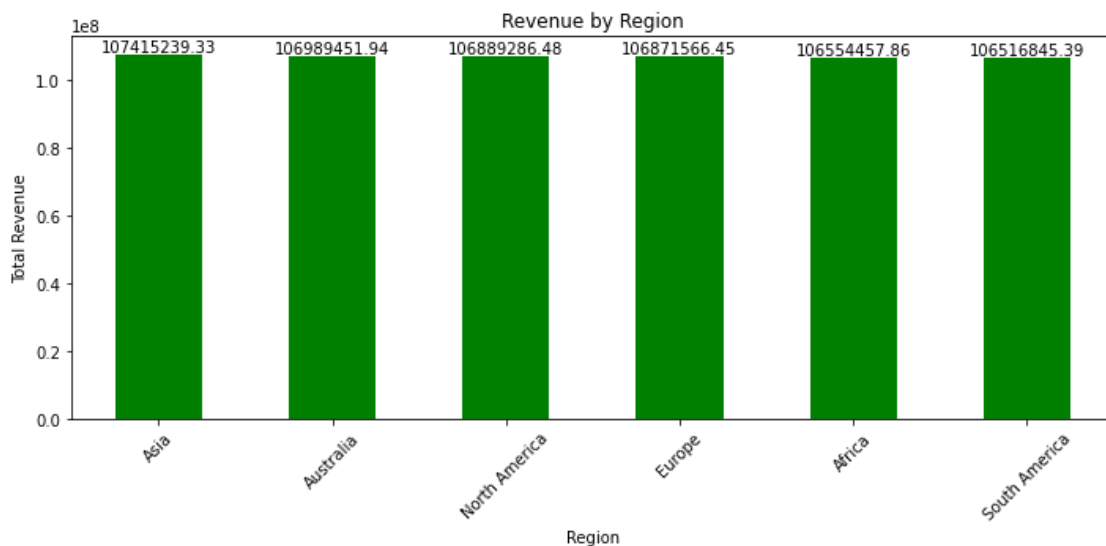sns.scatterplot(x='discount', y='quantity', data=df)

Insight: Higher discounts correlated with increased quantity but didn't always raise total revenue.

v. Regional & Payment Analysis

Revenue per region and payment method.

region_rev = df.groupby("region")["total_revenue"].sum().sort_values(ascending=False)

payment_pref = df.groupby("payment_method")["total_revenue"].sum()

Insight: Certain regions and payment methods dominate revenue contribution.

c) Visualizations & Key Findings

Used Seaborn for heatmaps, distribution plots, and boxplots.

Visuals supported strategic questions:

- Which segments are most profitable?
- Where should the business focus growth efforts?

Conclusion of Python Analysis

Python enabled:

- Granular segmentation of customers and discount effectiveness
- Clear visual storytelling with custom plots
- Easy integration with advanced logic for calculated fields

These insights formed the foundation for dashboard building in Power BI and supported data-driven recommendations for strategic planning.

5. POWER BI DASHBOARDS

Power BI was used to transform the Python-enriched dataset into an interactive, visually appealing dashboard for stakeholders. It allowed non-technical users to explore insights, filter data dynamically, and make informed decisions from well-structured visual summaries.

a) Dashboard Design Process
Data Import
The final cleaned dataset was exported from Python as a .csv file and imported into Power BI Desktop.
Data types were verified (especially dates and numerics), and relationships were checked.
New DAX measures were added for KPIs and segmentation (e.g., Total Revenue, AOV, Segments).
Pages Created:
Page 1: Executive Summary
Page 2: Customer & Product Analysis
Page 3: Regional & Discount Insights
Page 4: Customer Segmentation Overview
Each page served a different audience: executives, marketing teams, sales, and product managers.

b) KPI Overview & Global Metrics
Visuals:
Cards: Total Revenue, Total Orders, Avg. Order Value (AOV), Unique Customers
Trend Line: Monthly Revenue
Bar Chart: Top 5 Customers by Revenue, Top 5 Category by Revenue
Column Chart: Top 3 Product ID by Revenue
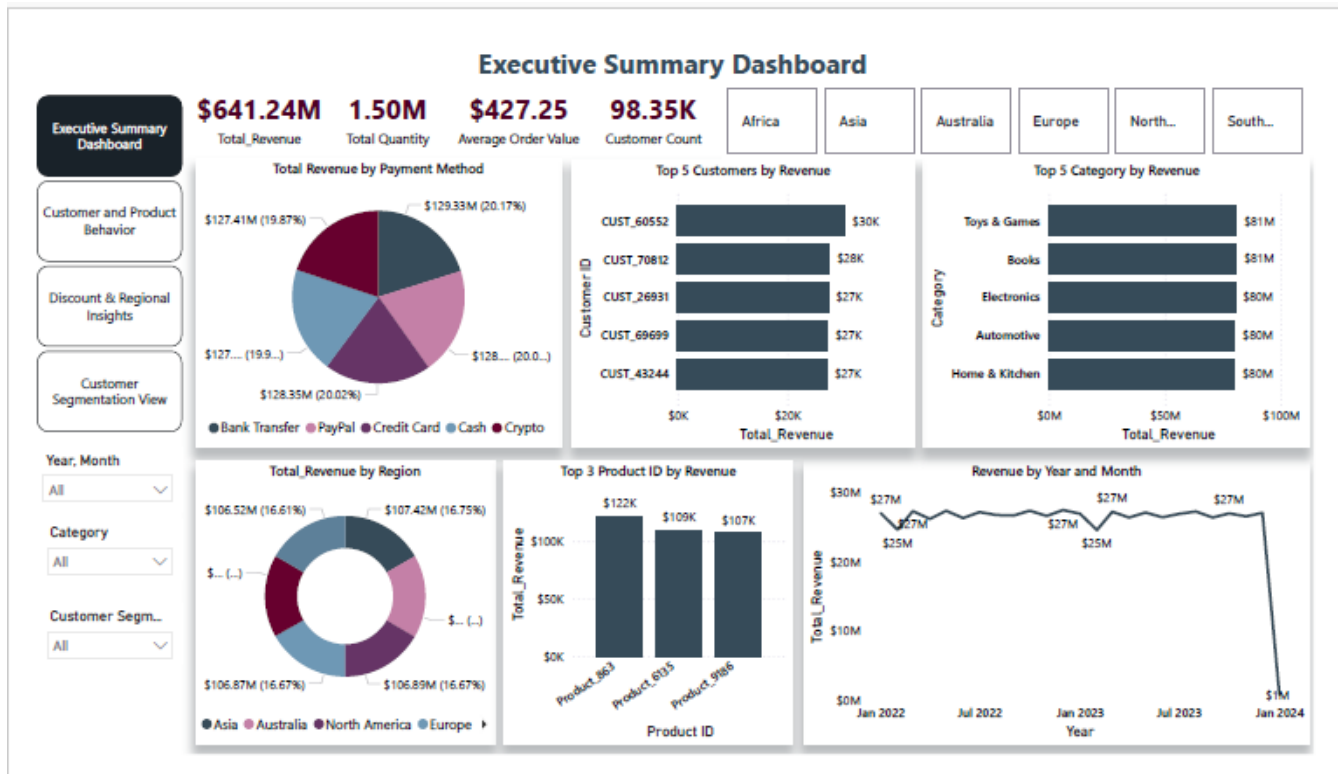**Pie Chart:** Total Revenue by Payment Method
Slicer: Year & Month, Category, Customer Segment, and Region
Key Metrics Defined (DAX):
Total Revenue = SUM(customers[total_revenue])
Average Order Value = [Total Revenue] / COUNT(customers[transaction_date])

Executive Summary Dashboard

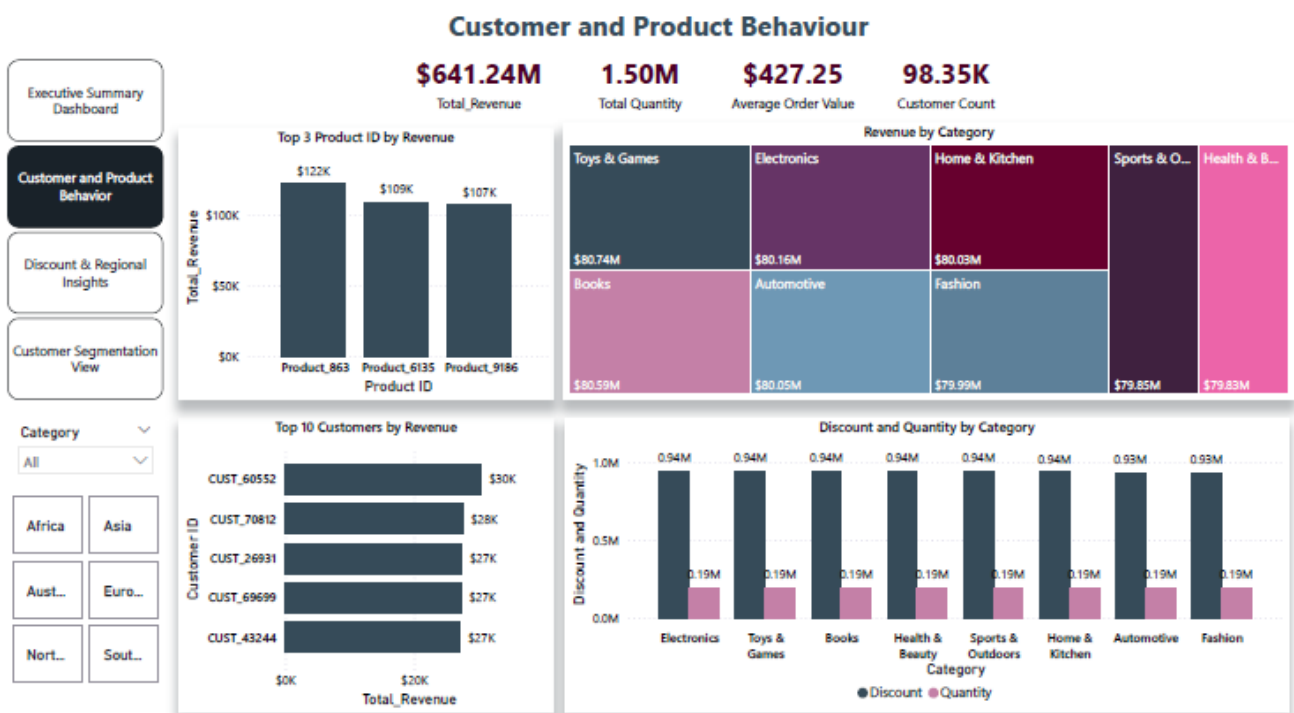c) **Customer & Product Insights**

Visuals:

Bar Chart: Top 5 Customers by Revenue

Treemap: Revenue by Product Category

Stacked Column: Discount and Quantity per Category, Top 3 Product ID by Revenue

Slicer: Category and Region

Insight: Certain customers and categories consistently outperform others; this page helps profile them.

d) Regional & Discount Insights

Visuals:

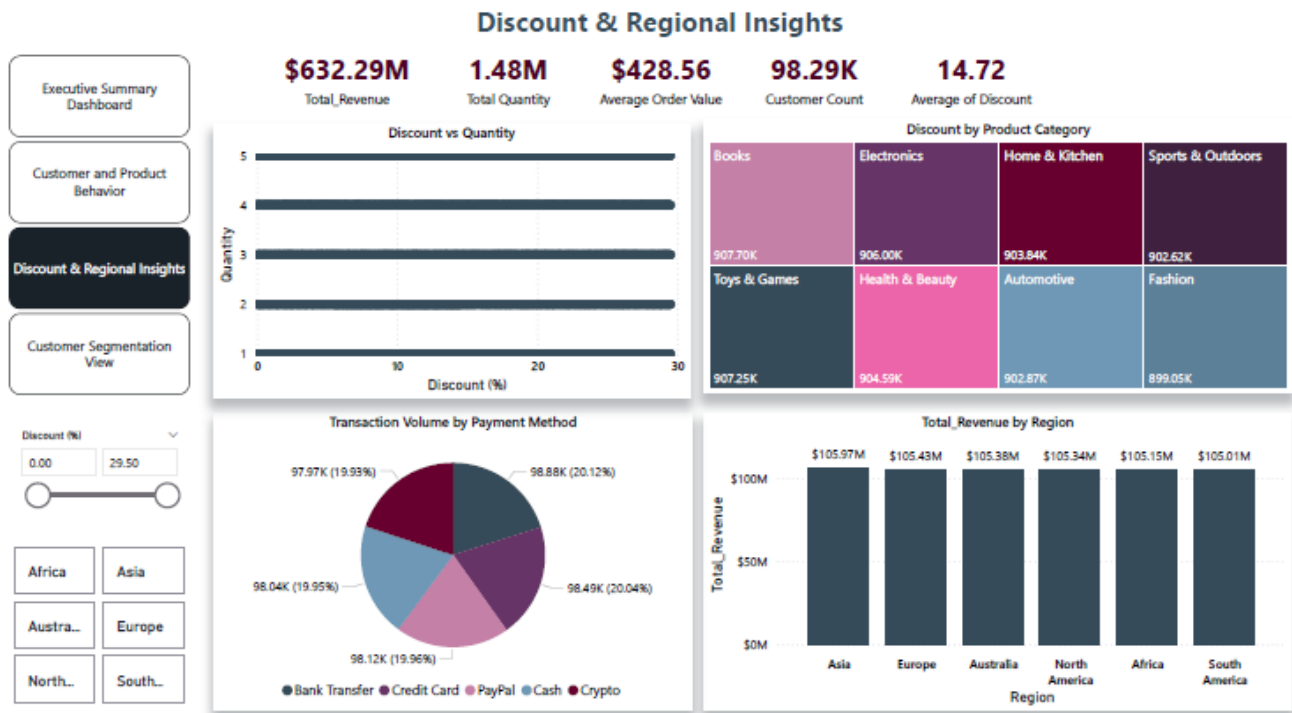**Tree Map:** Discount by Product Category

Pie Chart: Payment Method Distribution

Scatter Plot: Discount vs Quantity Sold

Column Chart: Total Revenue by Region

Cards: Average Discount and Revenue per Region

Slicers: Discount & Region

Interactivity: Selecting a region or discount range updates all other visuals dynamically.

e) Customer Segmentation View

Segmentation Logic (Created in DAX or Imported from Python):

Customer Segment =

SWITCH(TRUE(),

  [Total Revenue] < 100, "Low",

  [Total Revenue] < 500, "Medium",

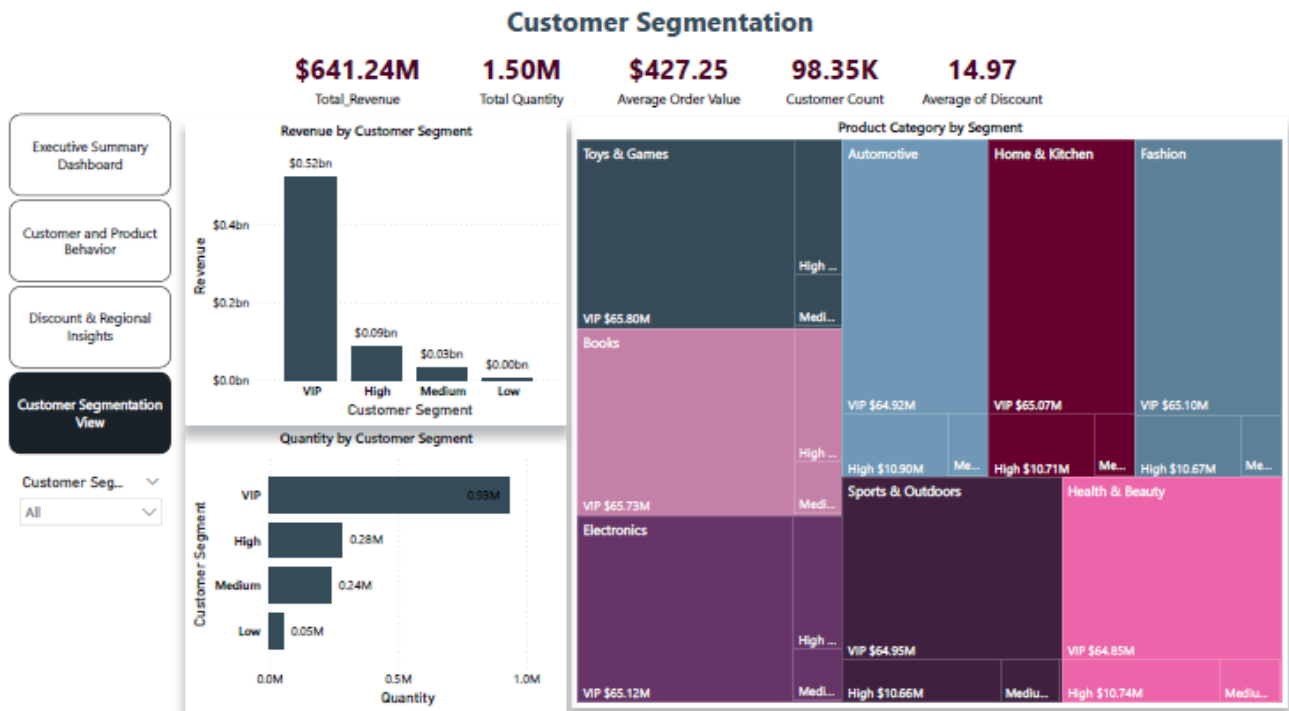 [Total Revenue] < 1000, "High",

  "VIP"

)

Visuals:

Bar Chart: Quantity Ordered by Customer Segment

Column Chart: Revenue per SCustomer Segment

Treemap: Segment vs Category

Slicer: Customer Segment

Insight: Enables targeted campaign planning and retention strategy for different tiers.

f) Publishing & Sharing

Exported as PDF for static reports.

Published to Power BI Service for dynamic sharing (if licensed).

Used bookmarks and tooltips to guide viewers through insights.

Design Principles Applied:

- Minimalism and readability
- Color coding by category or region
- Consistent chart labeling and formatting
- Power BI Summary

Power BI was essential in:

- Making the analysis accessible to non-technical stakeholders
- Empowering users with filters and slicers for custom exploration
- Summarizing weeks of work into four user-friendly, insight-rich pages

6. BUSINESS RECOMMENDATIONS

Based on the insights derived from MySQL queries, Python analysis, and Power BI dashboards, several strategic and operational recommendations are proposed to optimize the e-commerce business model. These recommendations span marketing, customer engagement, pricing strategy, and product positioning.

i. Focus on High-Value Customers

Insight: A small segment of customers (VIPs) generates a large portion of revenue.

Action:

- Launch a loyalty or rewards program for top-tier customers.
- Create exclusive offers and early access sales for high-spending segments.
- Use CRM tools to nurture and retain high-LTV (lifetime value) users.

ii. Regional Strategy Optimization

Insight: Certain regions consistently outperformed others in total revenue.

Action:

- Increase marketing investments in high-revenue regions.
- Improve shipping/logistics infrastructure in underperforming but high-potential regions.
- Localize promotions based on regional trends and preferences (e.g., holidays, payment methods).

iii. Pricing and Discount Strategy

Insight: Discounts increase quantity sold but can reduce profit margins.

Action:

- Implement targeted discounting (e.g., only on slow-moving inventory or during sales events).
- Analyze price elasticity for different categories and customer segments.
- A/B test different discount ranges to find optimal pricing for conversions and profit.

iv. Product Portfolio Management

Insight: Not all high-volume categories yield the highest revenue.

Action:

- Promote high-margin products more aggressively.
- Bundle underperforming products with top-sellers to increase exposure.
- Use demand data to adjust inventory and forecast seasonal sales trends.

v. Customer Segmentation for Campaigns

Insight: Four key customer segments were identified: Low, Medium, High, and VIP.

Action:

- Design personalized marketing campaigns per segment (email, SMS, retargeting).
- Offer upsell and cross-sell opportunities to Medium and High-value customers.
- Build segment-based lifetime value models for more accurate budgeting and ROI estimation.

vi. Payment Method Optimization

Insight: Digital wallets and credit cards dominate transactions.

Action:

- Streamline the checkout process for these methods.
- Introduce buy-now-pay-later (BNPL) options to capture price-sensitive customers.
- Offer incentives (e.g., cashback) for preferred payment methods to reduce cart abandonment.

vii. Visual Analytics Culture

Insight: Power BI dashboards improved visibility and decision-making.

Action:

- Regularly update and share dashboards across departments.
- Train staff in reading and interpreting dashboards.
- Use alerts and KPIs to monitor key metrics in real-time.

Strategic Takeaway

The e-commerce platform has rich, actionable data that when analyzed effectively, can power revenue growth, customer loyalty, and operational efficiency.

A data-driven culture must be established, combining business intelligence tools, customer insight, and agile marketing strategies.