Customer Behavior Analysis

Analyzing purchasing trends, spending habits, and the impact of marketing factors on sales through a data-driven approach.

Project Overview & Methodology

Project Goal

Uncover key customer purchasing trends and spending habits by integrating demographic and marketing data.

Core Technologies

Python

Data Cleaning & Preprocessing

SQL

Querying & Insight Extraction

Power BI

Interactive Visualization

Key Analytical Objectives

The analysis was structured around five core business questions to drive actionable insights.

Data Preprocessing

Clean and prepare the raw dataset for reliable analysis.

Identify Top Performers

Determine best-selling products and most valuable customer segments.

Subscriber Comparison

Compare spending behavior between subscribers and non-subscribers.

Marketing Impact

Examine the relationship between discounts, ratings, and customer loyalty.

Visualize Insights

Create an interactive BI dashboard for dynamic exploration.

Data Cleaning & Transformation (Python)

The raw dataset was rigorously cleaned and prepared using the Pandas library in Python to ensure data quality.

Initial Inspection

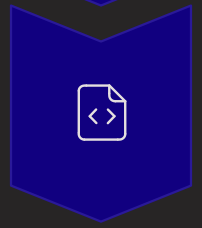
Reviewed data structure and types (e.g., column standardization).

Handle Missing Values

Filled missing review ratings with the median value specific to each product category.

Feature Engineering

Created new features, including standardized age groups and numeric mapping for purchase frequency.



Redundancy Check

Identified and removed redundant or unnecessary columns from the dataset.

SQL Analytical Queries: Insight Extraction

Key queries focused on identifying top revenue drivers, segment performance, and the effectiveness of marketing levers.

Revenue by Category

Determined which product categories contribute the most to overall sales volume and revenue.

Age Group Spending

Analyzed total revenue contribution across different customer age segments.

Discount Effectiveness

Measured the correlation between applied discounts and purchase volume/order value.

Shipping Preference

Compared transaction volume and average order value across different shipping types.

Power BI Dashboard: Interactive Visualization

A comprehensive Power BI dashboard was created to visualize key metrics and allow interactive data exploration.

Key Insights: Spending & Segmentation

25%

Subscriber Premium

Subscribers spend approximately 25% more than non-subscribers on average.

1

Top Revenue Category

Clothing and Accessories generate the highest revenue and sales volume.

High

Revenue Contribution

Young and middle-aged segments dominate the total revenue contribution.

Key Insights: Marketing & Logistics

Discount Impact

Discounts successfully increase purchase volume but do not necessarily lead to higher total revenue per order.

Shipping Analysis

- Standard shipping has the highest transaction volume.
- Express shipping correlates with a significantly higher average order value.

Future Work & Next Steps

To further enhance the analysis, the following steps are recommended:

Predictive Modeling

Implement models for churn prediction and Customer Lifetime Value (LTV) forecasting.

ETL Automation

Automate the data extraction, transformation, and loading pipelines for continuous data refreshes.

Advanced Segmentation

Apply clustering techniques (e.g., K-Means) to discover more nuanced customer segments.

Project details available on GitHub.