Customer Shopping Behavior Analysis

Uncovering patterns for better business decisions.

Project Overview

01

Data Cleaning

Python (Jupyter Notebook)

02

Data Storage & Querying

MySQL

03

Visualization & Insights

Power BI

Actionable insights for improved customer engagement, marketing, and revenue.

Business Problem Statement

Retail company seeks deeper insights into customer interactions.

Key Challenges

- Identify high-value customers
- Impact of discounts/subscriptions
- Best performing products/categories
- Demographic effects on spending

Core Business Question

How can shopping data optimize marketing, retention, and product strategies?

Project Objectives

1 **Clean & Prepare Data**

Using Python for analysis.

2 **Store Structured Data**

In MySQL for efficient querying.

3 **Perform SQL Analysis**

Answer real business questions.

4 **Build Interactive Dashboard**

Visualize insights with Power BI.

5 **Provide Recommendations**

Data-driven business strategies.

Dataset Description

Overview

- 3,900 Total Records
- 18 Total Columns

Key Data Fields

- Customer demographics (Age, Gender, Location)
- Purchase details (Item, Category, Amount, Season)
- Shopping behavior (Discount, Subscription, Review)
- Transaction behavior (Previous Purchases, Frequency, Shipping)

Data Quality Notes

- Missing values in Review Rating
- Data types and column naming standardized

Data Preparation & Cleaning (Python)

Data Loading

Imported dataset using pandas.

Initial Exploration

.info() and .describe() for structure.

Handle Missing Values

Median rating per category for reviews.

Column Standardization

Converted to snake_case.

Feature Engineering

Age groups, derived fields.

Data Export

Cleaned data saved for database.

Database Integration (MySQL)

Steps Performed

- Created MySQL database
- Defined tables for shopping data
- Loaded cleaned dataset
- Verified data insertion with SQL

Foundation for SQL analysis and Power BI connectivity.

Data Analysis Using SQL

1. Revenue by Gender
2. High Spending Discount Users
3. Top Rated Products
4. Shipping Type Comparison
5. Subscribers vs. Non-Subscribers
6. Discount Dependent Products
7. Customer Segmentation
8. Top Products per Category
9. Repeat Buyers & Subscriptions
10. Revenue by Age Group

Power BI Dashboard

Interactive visualization for stakeholders.

Key Performance Indicators

- 3.9K Customers
- 3.75 Avg. Review Rating
- $59.76 Avg. Purchase Amount

Subscription Status

27% Subscribers, 73% Non-Subscribers. Opportunity for conversion!

Key Insights & Recommendations

Loyalty & Subscriptions

Drive significant revenue.

Product Categories

Certain categories drive most sales.

Discounts

Influence behavior, use carefully.

Age Groups

Younger groups are key revenue drivers.

Business Recommendations

- Improve Subscription Offers
- Strengthen Loyalty Programs
- Optimize Discount Strategies
- Focus on High-Performing Categories
- Target High-Value Segments