

Customer Shopping Behavior Analysis

📌 Project Overview

This project analyzes **customer shopping behavior** using transactional data from **3,900 purchases** across multiple product categories. The objective is to extract actionable insights related to **spending patterns, customer segments, product preferences, discounts, and subscription behavior** to support data-driven business decisions.

The project demonstrates an **end-to-end data analytics workflow** using **Python, SQL (PostgreSQL), and Power BI**.



Dataset Summary

- **Total Rows:** 3,900
 - **Total Columns:** 18
 - **Key Features:**
 - Customer demographics: Age, Gender, Location, Subscription Status
 - Purchase details: Item Purchased, Category, Purchase Amount, Season, Size, Color
 - Shopping behavior: Discount Applied, Previous Purchases, Purchase Frequency, Review Rating, Shipping Type
 - **Missing Data:** 37 missing values in the *Review Rating* column
-



Data Cleaning & Preparation (Python)

Performed using **Pandas** in Python:

- Loaded dataset and explored structure using `df.info()` and `df.describe()`
 - Handled missing values by imputing *Review Rating* using the **median rating per product category**
 - Standardized column names into **snake_case**
 - Feature Engineering:
 - Created `age_group` by binning customer ages
 - Derived `purchase_frequency_days`
 - Identified redundancy between `discount_applied` and `promo_code_used` and removed unnecessary column
 - Loaded cleaned data into **PostgreSQL** for SQL-based analysis
-

Data Analysis Using SQL (PostgreSQL)

Key business questions answered using SQL queries:

1. Revenue comparison by gender
 2. High-spending customers who used discounts
 3. Top 5 products by average review rating
 4. Average purchase amount by shipping type
 5. Subscriber vs non-subscriber spending analysis
 6. Products most dependent on discounts
 7. Customer segmentation: New, Returning, Loyal
 8. Top 3 products within each category
 9. Subscription likelihood of repeat buyers
 10. Revenue contribution by age group
-



Dashboard (Power BI)

An interactive **Power BI dashboard** was created to visualize:

- Revenue trends
 - Customer segments
 - Product performance
 - Subscription impact
 - Discount and shipping insights
-



Business Recommendations

- Promote **subscription plans** with exclusive benefits
 - Implement **loyalty programs** for repeat customers
 - Optimize **discount strategies** to protect profit margins
 - Highlight **top-rated and best-selling products**
 - Run **targeted marketing campaigns** for high-revenue age groups and express-shipping users
-



- ## Tools & Technologies
- **Python:** Pandas, NumPy
 - **Database:** PostgreSQL
 - **Visualization:** Power BI
 - **IDE:** Jupyter Notebook
-

Project Structure (Suggested)

```
├── data/
│   └── raw_dataset.csv
├── notebooks/
│   └── data_cleaning_eda.ipynb
├── sql/
│   └── analysis_queries.sql
├── powerbi/
│   └── dashboard.pbix
└── README.md
```

Conclusion

This project showcases a complete analytics pipeline—from raw data to insights and business recommendations—making it suitable for **data analyst portfolios, academic projects, and entry-level data roles**.