Turning Transaction Logs into **Actionable Insight**

Customer trends across products and demographics
Analysis used Python, PostgreSQL, and Power BI to inform marketing, inventory, and retention

Lakshya Shree H

Project Overview — Scope and Approach

Scope, key questions, and technical approach

Scope: analyze shopping data for trends by demographics, product categories, discounts, subscriptions, and feedback

Identify patterns that drive sales and satisfaction

Key questions: who buys what, which discounts work, impact of subscriptions, and sentiment signals

Focus on customer behavior and business impact

Data pipeline: cleaning and preprocessing in Python, storage and SQL analysis in PostgreSQL

Ensure reliable, queryable datasets

Delivery: interactive storytelling with Power BI for business user exploration

Enable self-serve insights and decision making

Outcome: actionable insights to improve sales performance and customer satisfaction

Translate analysis into business actions

Turning Data into Decisions

Convert raw customer transactions into actionable retail insights

Understand Customer Spending Behaviour

Identify Purchase Frequency, Basket Size, And Lifetime Value

Assess Product Performance

Measure Sell-Through, Returns, And Product-Level Margins

Evaluate Impact Of Discounts And Subscriptions

Quantify Uplift, Cannibalization, And Retention Effects

Attribute Revenue By Customer Segments

Segment By Value, Cohort, Channel, And Subscription Status

Implication: Operationalize Insights

Build Pipelines And Align Stakeholders To Turn Analysis Into Action

Project Objectives: Clean data, repeatable analysis, and actionable insights

Prepare customer shopping data for SQL reporting and Power BI self-serve dashboards

Clean and preprocess customer shopping data using Python

- Standardize fields and fix missing values for accuracy
- Normalize product and customer identifiers for joins
- Create reproducible ETL scripts for ongoing use
- Improve data quality to enable reliable segmentation

Store and analyze data with SQL in PostgreSQL

- Design normalized schema for transactional data
- Implement indexed tables for query performance
- Build repeatable SQL queries for business KPIs
- Enable scheduled extracts for reporting workflows

Extract meaningful business insights from transactions

- Compute customer lifetime value and purchase frequency
- Identify high-value segments for targeted actions
- Surface trends that inform pricing and promotions
- Translate data into prioritized business recommendations

Visualize trends and patterns using Power BI dashboards

- Create interactive dashboards for stakeholder exploration
- Expose KPIs and drilldowns for self-serve analysis
- Use clear visuals to highlight seasonality and cohorts
- Enable exportable reports for regular decision meetings

Dataset Scope and Key Fields

3,900 records, 18 columns — primary fields and validation priorities

Metric	Value / List
Total rows	3,900
Total columns	18
Important columns	Customer ID, Age, Gender, Item Purchased, Product Category, Purchase Amount, Location, Shipping Type, Discount Applied, Subscription Status, Review Rating, Previous Purchases

01 

Validate distributions for numeric fields

Check skewness, outliers, and missing values

02 

Check class balance for categorical fields

Ensure meaningful group sizes for Product Category and Gender

03 

Assess sample representativeness

Compare demographics and locations to target population

04 

Verify data quality and consistency

Detect duplicates, inconsistent IDs, and incorrect formats

05 

Confirm readiness for modeling

Impute or exclude missing values and encode categorical features

Data Preparation and Cleaning: Practical Process

Reproducible steps from load to database insertion

- Load dataset
  - Import CSV into a Pandas DataFrame in Python environment
- Standardize columns
  - Convert names to snake case and remove irrelevant columns
- Check categorical uniformity
  - Normalize categorical values to consistent labels
- Handle missing review_rating
  - Identify and impute or flag missing review_rating entries
- Verify data types
  - Confirm and correct types for numeric, datetime, and categorical fields
- Generate statistics
  - Create descriptive stats to validate distributions and outliers
- Prepare for insertion
  - Export cleaned dataset ready for database insertion

Data Quality: Statistical Summaries and Missing Values

Quick view of distributions and handling of missing review ratings

34

Median age

Central tendency for user ages

52.4

Avg purchase amount

Mean transaction value

4.2

Avg review rating

Mean rating across reviewed orders

18

Missing review_rating

Proportion of records without rating

Database Integration and SQL Analysis Implementation

PostgreSQL load, repeatable views, and key query set

ETL: Data Cleaning and Load

Cleaned data uploaded with Python and SQLAlchemy; table contains 3,900 records.

Database: PostgreSQL

Persisted cleaned table in PostgreSQL for analytic queries and views.

Query Layer: Repeatable SQL Views

Encapsulate business questions as views and parameterized queries for re-use and validation.

Dashboard: Business Consumption

Expose view outputs to dashboards for analyst and stakeholder consumption.

Key Insights from SQL Analysis

Data-driven actions to boost revenue and retention

Certain product categories drive higher revenue

Focus inventory and promotions on these categories

Customers using discounts still add strong revenue

Calibrate discounting to protect margins

Subscription customers show consistent engagement

Invest in subscription and loyalty programs

Middle-aged and young adults contribute the highest revenue

Target marketing to these age segments

Repeat customers form a valuable segment

Increase repeat purchase frequency

Business implications

Prioritize high-revenue categories; protect margins; grow subscriptions

Customer Behavior Dashboard

Subscription Status

No | Yes

Gender

Female | Male

Category

- Accessories
- Clothing
- Footwear
- Outerwear

Shipping Type

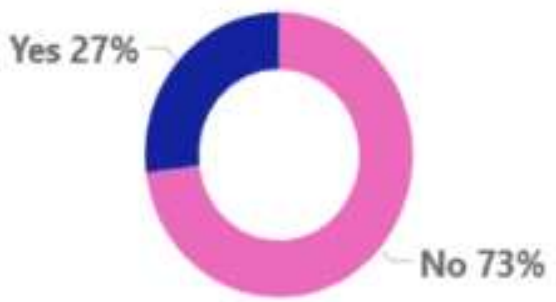
- 2-Day Shipping
- Express
- Free Shipping
- Next Day Air
- Standard
- Store Pickup

3.9K
Number of Customers

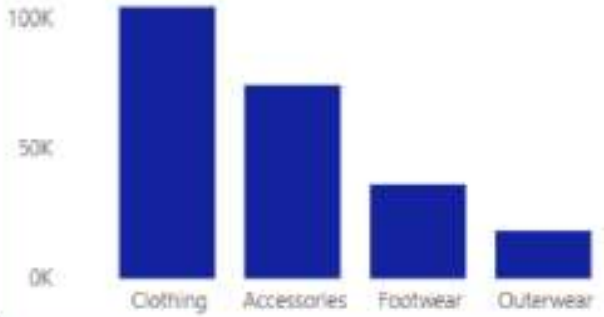
$59.76
Average Purchase Amount

3.75
Average Review Rating

% of Customers by Subscription Status



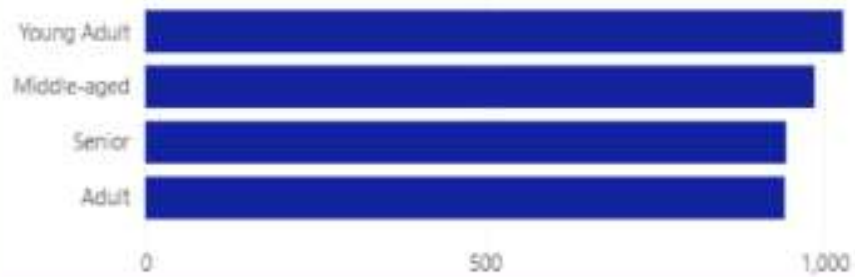
Revenue by Category



Sales by Category

Revenue by Age Group

Sales by Age Group



Power BI Dashboard: Components and User Interactivity

Top KPIs and recommended charts with interactive filters for business users

01

Top-row KPIs: Total number of customers, Average purchase amount, Average customer review rating

High-visibility metrics for quick decisions

02

Category revenue: Revenue by product category chart

Stacked bar or treemap to compare categories

03

Sales by category: Volume and trend view

Line or combo chart to show sales trends

04

Age-group analysis: Revenue and sales by age group

Segmented bars or age cohort trend lines

05

Subscription distribution: Subscription-based customer breakdown

Donut or stacked bar to show subscription share

06

Interactive filters: Gender, Category, Shipping type

Slicers and cross-filtering for exploration

07

Design tips: Consistent color coding and explanatory tooltips

Add a how to use panel for business users

Business Recommendations and Conclusion

Actionable steps to convert insights into measurable retail gains

Recommendations

1. Strengthen loyalty and subscription programs to boost retention
2. Target high-value age groups through focused marketing
3. Optimize discount strategies to balance revenue and profitability
4. Promote high-rated products to increase customer trust
5. Use dashboards for continuous performance monitoring

Actions and Next Steps

1. Operationalize dashboards with regular refresh schedules
2. A/B test targeted campaigns for top age groups
3. Set up KPI monitoring and alerts to track performance over time
4. Define discount thresholds and run margin impact analyses
5. Integrate product ratings into promotional workflows

Thank You