Smart-Supply-Chain-Disruption-Predictor

Customer Trend Analysis – Technical & Business Report

1. Executive Summary

This project demonstrates a full-cycle data analytics workflow from raw customer shopping data to actionable business insights. Using Python (Pandas, SQLAlchemy), PostgreSQL, and SQL queries, we cleaned, transformed, and analyzed customer behavior data to uncover trends in spending, loyalty, product preferences, and promotional effectiveness.

Key findings include:

- **Male customers generate more revenue despite fewer in number.**
- **Subscribed customers spend less on average but contribute significantly to total revenue.**
- **High-rated products are consistent across categories, with Blouse, Jeans, and Sweater leading.**
- **Discount usage is correlated with higher spending, indicating effective promotion targeting.**
- **Loyalty and repeat purchases are strong predictors of subscription uptake.**

---

2. Data Preparation & Cleaning

Steps Taken:

1. **Loaded dataset (customer_shopping_behavior.csv) with 3,900 rows and 18 columns.**
2. **Handled missing values in Review Rating by filling with median per Category.**
3. **Standardized column names to lowercase with underscores.**
4. **Created new features:**
    - **age_group (Young Adult, Adult, Middle-aged, Senior)**
    - **purchase_frequency_days (mapped from textual frequency labels)**
5. **Removed redundant column promo_code_used (identical to discount_applied).**
6. **Exported cleaned data to PostgreSQL using SQLAlchemy.**

Data Quality Post-Cleaning:

- **Zero missing values.**
- **Consistent data types.**
- **Business-ready for analysis.**

---

3. Exploratory Data Analysis (EDA) – Key Insights

🔹 Customer Demographics:

- **Gender: 2,652 Male vs. 1,248 Female customers.**
- **Age: Mean = 44 years, range 18–70.**
- **Top Location: Montana (96 customers).**

🔹 Spending Behavior:

- **Average Purchase Amount: $59.76**
- **Most Common Category: Clothing (1,737 purchases)**
- **Popular Product: Blouse (171 purchases)**

🔹 Subscription & Discounts:

- **Non-subscribers: 2,847 vs. Subscribers: 1,053**
- **Discount Applied: 43% of transactions used discounts.**

---

4. SQL Analysis – Business Questions Answered

Q1: Revenue by Gender

```
SELECT gender, SUM(purchase_amount) AS revenue
FROM customer GROUP BY gender;
```

Insight: Male customers generate significantly higher revenue despite similar average spend.

Q2: High Spenders Using Discounts

```
SELECT customer_id, purchase_amount
FROM customer
```

WHERE discount_applied = 'Yes'

AND purchase_amount >= (SELECT AVG(purchase_amount) FROM customer);

Insight: Discounts attract above-average spenders, indicating effective promotion targeting.

Q3: Top 5 Rated Products

SELECT item_purchased, ROUND(AVG(review_rating),2) AS avg_rating

FROM customer

GROUP BY item_purchased

ORDER BY avg_rating DESC LIMIT 5;

Insight: High-rated products (Blouse, Jeans, Sweater) should be highlighted in marketing.

Q4: Shipping Type vs. Spend

SELECT shipping_type, ROUND(AVG(purchase_amount),2) AS avg_spend

FROM customer

WHERE shipping_type IN ('Standard', 'Express')

GROUP BY shipping_type;

Insight: Express shipping correlates with higher spend—consider bundling with premium products.

Q5: Subscription Impact on Revenue

SELECT subscription_status,

COUNT(customer_id) AS total_customers,

ROUND(AVG(purchase_amount),2) AS avg_spend,

ROUND(SUM(purchase_amount),2) AS total_revenue

FROM customer

GROUP BY subscription_status;

Insight: Subscribers are fewer but contribute substantially to revenue—focus on retention.

Q6: Products with Highest Discount Usage

```
SELECT item_purchased,
    ROUND(100.0 * SUM(CASE WHEN discount_applied = 'Yes' THEN 1 ELSE 0 END) /
COUNT(*),2) AS discount_rate
FROM customer
GROUP BY item_purchased
ORDER BY discount_rate DESC LIMIT 5;
```

Insight: High-discount products are popular—consider dynamic pricing or bundling.

Q7: Customer Segmentation by Purchase History

```
WITH customer_type AS (
  SELECT customer_id, previous_purchases,
      CASE
        WHEN previous_purchases = 1 THEN 'New'
        WHEN previous_purchases BETWEEN 2 AND 10 THEN 'Returning'
        ELSE 'Loyal'
      END AS customer_segment
  FROM customer
)
SELECT customer_segment, COUNT(*) AS customer_count
FROM customer_type
GROUP BY customer_segment;
```

Insight: Loyal customers are a significant segment—reward programs could enhance retention.

Q8: Top 3 Products per Category

```
WITH item_counts AS (
  SELECT category, item_purchased,
```

```
    COUNT(customer_id) AS total_orders,
    ROW_NUMBER() OVER (PARTITION BY category ORDER BY COUNT(customer_id)
DESC) AS item_rank
  FROM customer
  GROUP BY category, item_purchased
)
SELECT item_rank, category, item_purchased, total_orders
FROM item_counts
WHERE item_rank <= 3;
```

Insight: Category leaders are clear—optimize inventory and promotions for these items.

Q9: Repeat Buyers & Subscriptions

```
SELECT subscription_status, COUNT(customer_id) AS repeat_buyers
FROM customer
WHERE previous_purchases > 5
GROUP BY subscription_status;
```

Insight: Repeat buyers are more likely to subscribe—target them with subscription incentives.

Q10: Revenue by Age Group

```
SELECT age_group, SUM(purchase_amount) AS total_revenue
FROM customer
GROUP BY age_group
ORDER BY total_revenue DESC;
```

Insight: Middle-aged and Senior groups drive the most revenue—tailor marketing accordingly.

5. Recommendations for Stakeholders

Marketing:

- **Target male customers with premium product campaigns.**
- **Promote top-rated products (Blouse, Jeans, Sweater) in featured collections.**
- **Use discount strategies to attract high spenders without eroding margins.**

Sales & Inventory:

- **Stock more top-performing products per category.**
- **Bundle Express shipping with high-margin items to increase AOV.**

Customer Retention:

- **Launch loyalty programs for repeat buyers to boost subscription uptake.**
- **Segment communications by age group (focus on Middle-aged & Senior).**

Operations:

- **Analyze discount effectiveness—avoid over-discounting already popular items.**
- **Monitor review ratings to maintain product quality and customer satisfaction.**

---

6. Tools & Technologies Used

Tool	Purpose
Python (Pandas)	Data cleaning & transformation
PostgreSQL	Database storage & querying
SQLAlchemy	Python-to-Database connection
Jupyter Notebook	Interactive analysis & reporting
Git/GitHub	Version control & collaboration

Report Prepared By: Thirumagal Meena A
Date: January 12, 2025
Project: Customer Trend Analysis – Full-Cycle Data Analytics