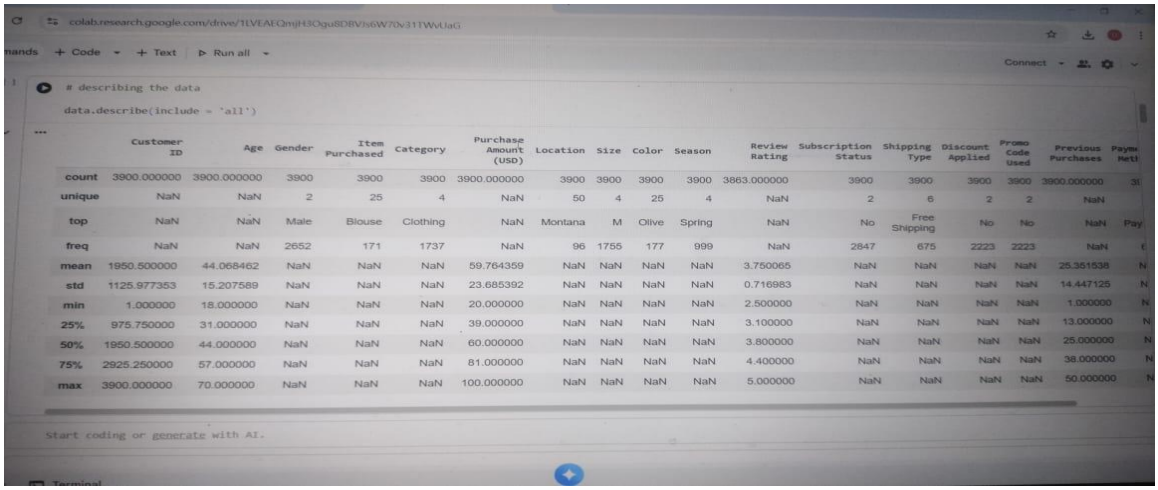
## Customer Behavior Analysis – Project Documentation

This document presents an end-to-end customer behavior analysis using Python, SQL, and Power BI. The project focuses on understanding customer purchasing patterns, preferences, and revenue trends to support data-driven business decisions.

## Data Cleaning & Transformation

Python libraries Pandas and NumPy were used to clean and transform the dataset. Duplicate records were removed, missing values were handled, data types were standardized, and derived metrics such as total purchase amount and purchase frequency were created.



```
# describing the data
data.describe(include = 'all')
```

| | Customer ID | Age | Gender | Item Purchased | Category | Purchase Amount (USD) | Location | Size | Color | Season | Review Rating | Subscription Status | Shipping Type | Discount Applied | Promo Code Used | Previous Purchases | Paym Meth |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 3900.000000 | 3900.000000 | 3900 | 3900 | 3900 | 3900.000000 | 3900 | 3900 | 3900 | 3900 | 3863.000000 | 3900 | 3900 | 3900 | 3900 | 3900.000000 | 3 |
| unique | NaN | NaN | 2 | 25 | 4 | NaN | 50 | 4 | 25 | 4 | NaN | 2 | 6 | 2 | 2 | NaN | |
| top | NaN | NaN | Male | Blouse | Clothing | NaN | Montana | M | Olive | Spring | NaN | No | Free Shipping | No | No | NaN | Pay |
| freq | NaN | NaN | 2652 | 171 | 1737 | NaN | 96 | 1755 | 177 | 999 | NaN | 2847 | 675 | 2223 | 2223 | NaN | |
| mean | 1950.500000 | 44.068462 | NaN | NaN | NaN | 59.764359 | NaN | NaN | NaN | NaN | 3.750065 | NaN | NaN | NaN | NaN | 25.351538 | N |
| std | 1125.977353 | 15.207589 | NaN | NaN | NaN | 23.685392 | NaN | NaN | NaN | NaN | 0.716983 | NaN | NaN | NaN | NaN | 14.447125 | N |
| min | 1.000000 | 18.000000 | NaN | NaN | NaN | 20.000000 | NaN | NaN | NaN | NaN | 2.500000 | NaN | NaN | NaN | NaN | 1.000000 | N |
| 25% | 975.750000 | 31.000000 | NaN | NaN | NaN | 39.000000 | NaN | NaN | NaN | NaN | 3.100000 | NaN | NaN | NaN | NaN | 13.000000 | N |
| 50% | 1950.500000 | 44.000000 | NaN | NaN | NaN | 60.000000 | NaN | NaN | NaN | NaN | 3.800000 | NaN | NaN | NaN | NaN | 25.000000 | N |
| 75% | 2925.250000 | 57.000000 | NaN | NaN | NaN | 81.000000 | NaN | NaN | NaN | NaN | 4.400000 | NaN | NaN | NaN | NaN | 38.000000 | N |
| max | 3900.000000 | 70.000000 | NaN | NaN | NaN | 100.000000 | NaN | NaN | NaN | NaN | 5.000000 | NaN | NaN | NaN | NaN | 50.000000 | N |

## Data Analysis Using SQL

SQL was used to fetch, filter, aggregate, and segment customer data. Queries were written to analyze customer behavior, revenue contribution, and category-wise performance. The cleaned and structured data was then connected to Power BI for visualization.

## Power BI Dashboard & Visualization

Power BI was used to create interactive dashboards with slicers, KPI cards, bar charts, and donut charts. These visuals allow dynamic exploration of customer behavior across subscription status, gender, product category, and shipping type.


## Key Insights

- Clothing category generates the highest revenue and customer count
- Non-subscribed customers form the majority of users
- Certain products such as boots and t-shirts contribute higher revenue
- Overall customer ratings indicate moderate to positive satisfaction

## Conclusion

The Customer Behavior Analysis provides valuable insights that support informed business decision-making. By analyzing customer purchase patterns, subscription status, product categories, and revenue distribution, the business can clearly identify which products and customer segments contribute the most to overall revenue. The analysis shows that certain categories, such as clothing, consistently generate higher sales and customer engagement, indicating opportunities for targeted promotions and inventory optimization.

The dashboard also highlights differences between subscribed and non-subscribed customers, helping the business evaluate the effectiveness of subscription-based strategies. Understanding customer preferences and purchasing frequency enables the organization to design personalized marketing campaigns, improve customer retention, and enhance overall customer experience. Additionally, insights into shipping preferences and product performance allow businesses to streamline logistics and improve operational efficiency.