Customer Shopping Behavior Analysis

Web Template Link:

https://sudhamshasagar.github.io/customer_data_analysis_PYTHON-SQL-POWERBI/

Project Overview

End-to-End Data Analytics Case Study focused on Customer Behaviour and Segmentation within an e-commerce context.

It involved analysing a transactional dataset of 3,900 customer purchases to generate actionable business intelligence.

```
[3]: df.shape

[3]: (3900, 18)
```

3900 rows and 18 columns present in the dataset

The core methodology and deliverables include:

1. **Data Preparation (ETL):** Cleaning, transformation, and feature engineering (e.g., creating age and purchase frequency groups) to ready the data for advanced analysis.

```
[4]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             3900 non-null   int64
 1   Age                     3900 non-null   int64
 2   Gender                  3900 non-null   object
 3   Item Purchased          3900 non-null   object
 4   Category                3900 non-null   object
 5   Purchase Amount (USD)   3900 non-null   int64
 6   Location                3900 non-null   object
 7   Size                    3900 non-null   object
 8   Color                   3900 non-null   object
 9   Season                  3900 non-null   object
 10  Review Rating           3863 non-null   float64
 11  Subscription Status     3900 non-null   object
 12  Shipping Type           3900 non-null   object
 13  Discount Applied        3900 non-null   object
 14  Promo Code Used         3900 non-null   object
 15  Previous Purchases      3900 non-null   int64
 16  Payment Method          3900 non-null   object
 17  Frequency of Purchases  3900 non-null   object
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB
```

There are few missing values present in the review rating column

```
[5]: df.describe(percentiles=[0.80,0.90,0.95])
```

[5]:

	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
count	3900.000000	3900.000000	3900.000000	3863.000000	3900.000000
mean	1950.500000	44.068462	59.764359	3.750065	25.351538
std	1125.977353	15.207589	23.685392	0.716983	14.447125
min	1.000000	18.000000	20.000000	2.500000	1.000000
50%	1950.500000	44.000000	60.000000	3.800000	25.000000
80%	3120.200000	60.000000	84.000000	4.500000	40.000000
90%	3510.100000	65.000000	93.000000	4.700000	46.000000
95%	3705.050000	68.000000	96.050000	4.900000	48.000000
max	3900.000000	70.000000	100.000000	5.000000	50.000000

Custom percentiles have been used to check outliers.

Missing Data Handling: Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.

Column Standardization**:** Renamed columns to **snake case** for better readability and documentation.

Feature Engineering:

- Created age_group column by binning customer ages.
- Created purchase_frequency_days column from purchase data.

Data Consistency Check**:** Verified if discount_applied and promo_code_used were redundant; dropped promo_code_used.

Database Integration**:** Connected Python script to MySQL and loaded the cleaned Data Frame into the database for SQL analysis.

2. **Advanced Analytical Processing:** Implementing complex SQL queries utilizing Window Functions, CTEs, and conditional logic to establish key metrics, perform customer segmentation (New, Returning, Loyal), and evaluate product and subscription performance.
3. **Strategic Reporting:** Developing an interactive Power BI dashboard to visualize key findings (e.g., revenue contribution by age group, discount efficacy) and provide data-driven recommendations for optimizing marketing spend and guiding product inventory strategies.

Dashboard Preview

Strategic Business Recommendations

1. Maximize Subscription Value and Penetration

- Rationale: Data confirms that subscribed customers have a significantly higher average purchase amount compared to non-subscribers, validating the program's long-term value.
- Action: Transition from simply promoting the subscription to marketing the exclusive financial and service benefits (e.g., expedited returns, early access to sales, unique content) to incentivize conversion and reinforce the value proposition for existing members.

2. Implement Tiered Customer Loyalty Programs

- Rationale: The segmented customer analysis identifies distinct New, Returning, and Loyal customer groups based on purchase frequency. The "Returning" segment (2 to 10 previous purchases) represents a critical opportunity for retention.
- Action: Design targeted rewards for Repeat Buyers (the "Returning" segment) to successfully migrate them into the high-value "Loyal" segment, reducing churn risk and boosting lifetime value (LTV).

3. Optimize Promotional and Discount Policy

- Rationale: Queries analysing discount application by product showed that promotional codes are often used on items that are already highly popular or top-rated, potentially sacrificing margin unnecessarily.
- Action: Establish clear rules for discount application. Shift promotions away from top-rated, high-demand items and strategically target discounts toward clearing slow-moving inventory or driving first-time purchases for the "New" customer segment, balancing sales volume with profitability**.**

4. Strategic Product Positioning in Campaigns

- Rationale: Product ranking analysis provides clear identification of the top 5 highest-rated and best-selling items within each category. These items act as quality assurance and demand drivers.
- Action: Ensure marketing and on-site product displays prominently feature top-rated and best-selling products to leverage social proof, build trust with new customers, and maximize conversion rates.

5. Refine Demographic and Behavioural Targeting

- Rationale: Revenue contribution analysis revealed which Age Groups generate the highest total spend. Separately, the comparative shipping analysis highlights a specific, high-intent user group (e.g., Express Shipping users).
- Action: Allocate a larger portion of the marketing budget to channels and creatives that resonate with the highest-revenue Age Groups. Additionally, design specialized campaigns (e.g., fast-track delivery offers) to reward or re-engage customers who prioritize Expedited Shipping.