Customer Behaviour Analysis

Executive Summary

This project executes a full-cycle data analysis pipeline to understand customer shopping habits, revenue drivers, and product performance. The solution involves data extraction and cleaning using **Python**, strategic analysis using **SQL**, and interactive reporting using **Power BI**.

The analysis focuses on a dataset of **3,900 customers**, uncovering critical insights regarding subscription inefficacy, gender-based revenue dominance, and opportunities for loyalty programs.

Project Structure

```
├── data/
│   └── shopping_behavior_updated.csv    # Raw dataset
├── notebooks/
│   └── Customer Behaviour Analysis.ipynb # Python data cleaning & feature engineering
├── sql/
│   └── customer_behaviour_analysis.sql   # SQL queries for business questions
├── dashboard/
│   └── Customer Behaviour Dashboard.pbix # Interactive Power BI dashboard
└── README.md
```

Methodology & Tech Stack

Phase I: Data Extraction & Transformation (Python)

File: `Customer Behaviour Analysis.ipynb`

- **Library:** `pandas`
- **Process:**
  - Loaded raw CSV data.
  - Performed statistical audits to ensure data integrity (checking for nulls, distributions).
  - **Feature Engineering:**
    - **Standardization:** Renamed columns (e.g., `Purchase Amount (USD)` → `purchase_amount`).
    - **Age Grouping:** Created buckets (Young Adult, Adult, Middle-aged, Senior) using `pd.qcut`.
    - **Frequency Mapping:** Converted text frequencies (e.g., "Fortnightly") to numeric values.

Phase II: Strategic Business Analysis (SQL)

File: `customer_behaviour_analysis.sql`

- **Objective:** Executed queries to answer specific business questions regarding demographics, revenue, and loyalty.
- **Key Logic:**
  - **Revenue Segmentation:** Grouped revenue by Gender and Age Group.
  - **Customer Segmentation:** Categorized users based on purchase history:
    - *New:* 1 purchase
    - *Returning:* 2-10 purchases
    - *Loyal:* >10 purchases
  - **Subscription Analysis:** Compared average spend between Subscribers and Non-Subscribers.

Phase III: Visualization (Power BI)

File: `Customer Behaviour Dashboard.pbix`

- **Executive Page:** High-level KPIs (Total Revenue, Avg Order Value).
- **Customer Loyalty Page:** Retention metrics and subscription breakdown.
- **Product Performance Page:** Inventory analysis and shipping impact.

Key Findings

Metric	Value	Insight
Total Customers	3,900	
Gender Split	68% Male	Male customers drive the majority of revenue ($157k vs $75k).
Avg Spend	~$60	Consistent across most segments.
Subscription	73% No / 27% Yes	**Critical Insight:** Subscribers do *not* spend more on average ($59.49) compared to non-subscribers ($59.87).
Top Category	Clothing	Dominant category across all seasons.

Strategic Recommendations

Based on the data analysis, the following actions are proposed:

1. **Marketing Pivot:** Shift marketing budget towards the **Male demographic** and **Adult/Middle-Aged groups (30-60)** as they are the highest revenue generators.
2. **Subscription Overhaul:** The current subscription model fails to drive higher spending. Implement **"Exclusive Bundles"** or **"Tiered Discounts"** to incentivize basket size growth for subscribers.
3. **Inventory Management:** Increase stock for high-rated Winter accessories (Gloves/Hats) and maintain high levels of Clothing inventory.

4. **Retention Program:** Launch a VIP program for the **2,400+ "Loyal" customers** identified in SQL analysis to prevent churn.

How to Run

1. **Python:** Run the Jupyter Notebook to process `shopping_behavior_updated.csv` and generate the clean dataset.
2. **SQL:** Import the cleaned data into your SQL database and run the queries in `customer_behaviour_analysis.sql` to extract insights.
3. **Power BI:** Open the `.pbix` file. Ensure the data source points to your local SQL instance or the cleaned CSV file to interact with the visualizations.

Author: [Karthik Mukkera]