🧾 Customer Behaviour Analysis – End-to-End Data Analytics Project

📌 Project Overview

This project focuses on understanding **customer shopping behaviour** using a real-world dataset. The main goal was to perform **end-to-end analysis** — from data cleaning and exploratory data analysis in Python, solving business-related queries in MySQL, and finally visualizing all the insights through Power BI.

Through this project, I wanted to understand how data analysis helps businesses make data-driven decisions — right from raw data to meaningful insights.

🧩 Step 1: Data Analysis and Feature Engineering in Python

Tools Used: Python, Pandas, NumPy, Matplotlib, Seaborn

I started the project by performing **Exploratory Data Analysis (EDA)** on the customer dataset.
The initial steps included:

- Loading the dataset into Python and checking data types, missing values, and duplicates.
- Cleaning the data by handling null values and standardizing columns (like gender, shipping type, discount applied).
- Performing **feature engineering** such as:
    - Creating a new column for **Age Group** (e.g., 18–25, 26–35, etc.).
    - Categorizing customers as **New**, **Returning**, and **Loyal** based on their previous purchases.
    - Deriving **Total Revenue per Customer** and other summary metrics.

After that, I used various **visualization libraries** to identify patterns and trends. Some important plots and analyses included:

- **Spending trend by age group and gender**
- **Effect of discount usage on total spending**
- **Comparison between subscribed and non-subscribed customers**
- **Distribution of review ratings across products**

🧠 Key Observations from EDA

- Female customers had slightly higher average purchase values than male customers.
- Loyal customers with more past purchases contributed the most to overall revenue.

- Discounts helped increase sales volume, but not always total revenue.
- Subscribed customers showed a higher average spend compared to non-subscribers.

This stage helped in understanding the data patterns and what kind of questions could be answered through business queries.

💾 Step 2: Solving Business Problems using MySQL

Tool Used: MySQL Workbench

After data cleaning and exploration, I imported the dataset into a **MySQL database** named customer_behavior.
I created and executed **SQL queries** to answer real business questions. Some of the key queries included:

1. **Revenue by Gender:**
   Found that both male and female customers contributed significantly, with females slightly ahead in total revenue.
2. **High-Value Discount Users:**
   Identified customers who used discounts but still spent more than the average amount — showing strong purchasing interest even during promotions.
3. **Top Rated Products:**
   Retrieved the top 5 products with the highest average review ratings.
4. **Shipping Type Comparison:**
   Compared average spending between Standard and Express shipping — Express shipping customers showed higher purchase amounts on average.
5. **Subscribers vs Non-Subscribers:**
   Checked whether subscribers spend more. The analysis confirmed subscribers not only spent more but also contributed higher overall revenue.
6. **Top Discounted Products:**
   Found products that had the highest percentage of discounted purchases.
7. **Customer Segmentation:**
   Grouped customers as New, Returning, and Loyal — with most falling under the Returning segment.
8. **Top Products per Category:**
   Used window functions to find the top 3 most purchased products in each category.
9. **Repeat Buyers and Subscription Correlation:**
   Observed that most repeat buyers were already subscribed.
10. **Revenue by Age Group:**
    Found that the 26–35 age group generated the highest total revenue.

💡 **Insights from SQL Analysis**

- **Subscribers and loyal customers** drive most of the business revenue.
- **Product ratings** strongly influence purchase frequency.
- **Discounts** attract a large volume of customers but may reduce per-order value.
- **Age and shipping preferences** can be used to create better-targeted marketing campaigns.

📊 Step 3: Data Visualization in Power BI

Tool Used: Microsoft Power BI

The final step was connecting the MySQL database to **Power BI** for creating an **interactive dashboard**.
I designed multiple visual reports to represent the insights clearly:

- Revenue by **Gender**, **Age Group**, and **Subscription Status**
- Top-selling **Products and Categories**
- Spending trend by **Shipping Type** and **Discount Usage**
- Customer **Segmentation Breakdown (New / Returning / Loyal)**

I also added **KPIs** such as:

- Total Revenue
- Average Purchase Amount
- Total Customers

This dashboard gives a clear, business-friendly view of how different customer groups behave and how their actions impact sales and revenue.

🧠 Learnings & Takeaways

This project helped me gain a clear understanding of the **entire data analytics process**, from raw data to business insight generation.
I learned how to:

- Perform **EDA and feature engineering** effectively using Python.
- Solve **real-world business queries** through SQL.
- Build **interactive dashboards** that communicate insights visually.

It also strengthened my skills in understanding data from a **business perspective** — identifying what metrics matter and how analytics supports better decision-making.

▓ Final Summary

The **Customer Behaviour Analysis** project was an excellent exercise in combining **Python, SQL, and Power BI** to perform end-to-end data analysis.
It provided me with practical exposure to transforming raw data into actionable insights that can help businesses improve their marketing, sales, and customer retention strategies.

Special thanks to the Amlan Mohanty's Youtube channel that guided me through this project — it really helped me understand the complete workflow of solving a business problem through data analysis.