1. Executive Summary (The Elevator Pitch) :

This project focused on analyzing the purchasing habits of customers to identify key drivers of revenue and loyalty. The main goal was to segment customers and uncover actionable insights for targeted marketing. We found that subscribed customers and certain age groups generate significantly higher revenue, and we established clear segments for prioritizing marketing efforts (New, Returning, Loyal). These findings allow the business to design promotions that maximize sales while protecting profitability.

2. Introduction and Problem Statement :

2.1 Background: The e-commerce company needed a deeper understanding of its customer base to move beyond simple sales tracking. They wanted to know who spends the most, what they buy, and how different factors (like shipping or discounts) affect their spending.

2.2 Analytical Role: The primary objective was to answer key business questions related to customer value, product demand, and promotional effectiveness. This included:

- Measuring revenue contribution by demographics (Gender, Age).
- Evaluating the impact of subscription status and discounts.
- Identifying top-performing products and creating customer loyalty segments.

2.3 Success Metrices: Success was measured by the clarity and actionability of the insights generated from the 10 key questions.

3. Data Description and Preparation :

3.1 Data Sources: The analysis utilized the Customer Shopping Behavior Dataset from Kaggle, an open-source dataset containing records of individual customer purchases.

3.2 Data Structure: The dataset is a flat file containing information on customer demographics (Age, Gender), purchase details (Item Purchased, Category, Purchase Amount), behavioral data (Review Rating, Subscription Status, Previous Purchases), and transaction details (Discount, Shipping Type, Payment Method).

3.3 Data Cleaning & Preprocessing:

- **Missing Values:** Missing values were checked across all key financial columns. For the 'Purchase Amount (USD)' column, any records with missing values were removed to ensure that all revenue calculations were accurate and unbiased.
- **Feature Engineering:** A key part of the project was segmenting customers into "New" (e.g., 0-1 previous purchase), "Returning" (e.g., 2-5 previous purchases), and "Loyal" (e.g., 6+ purchases), based on the Previous Purchases column.

4. Methodology and Analysis :

4.1 Technology Stack: The analysis utilized Python (for initial data manipulation and feature engineering), SQL (for cleaning, complex aggregation, and answering the core questions), and Power BI (for visualization and interactive dashboard creation).

4.2 Exploratory Data Analysis (EDA): Initial EDA confirmed the distribution of key variables. For example, the data showed "The 'Clothing' category accounted for the largest volume of purchases," indicating it is the most popular product area. High-impact visualizations were used to compare total revenue between male and female customers (Q1) and visualize the difference in purchase amounts by shipping type (Q4).

4.3 Analytical Approach: The analysis used aggregation and grouping techniques (as implemented in SQL and Python) to answer the 10 targeted questions.

- **Segmentation:** Used conditional logic to group customers by loyalty (Q7) and then cross-referenced this segment with subscription status (Q9).
- **Revenue Metrics:** Calculated sums and averages (Q1, Q4, Q5, Q10) to compare customer groups and segments directly.
- **Product Performance:** Used grouping and filtering to find top products by different criteria (Rating - Q3, Category - Q8, Discount Use - Q6).

5. Results and Key Findings :

Based on the 10 questions, the most impactful results were :

5.1 Subscriber Value (Q5): Subscribed customers spend significantly more (in both average spend and total revenue) than non-subscribers. This confirms that the subscription program is highly effective in increasing Customer Lifetime Value.

5.2 Gender and Revenue (Q1, Q10): Analysis showed that Female customers generated a significantly higher total revenue contribution than male customers, emphasizing the importance of targeted strategies for this demographic. Furthermore, the 25-35 and 45-55 age groups were the top two contributors to overall revenue.

5.3 Customer Loyalty Segments (Q7, Q9):

- The segmentation resulted in a clear count of New, Returning, and Loyal customers.
- There is a strong correlation between being a repeat buyer and subscribing (Q9). This suggests that customers who value the products eventually opt into the subscription for recurring benefits.

5.4 Discount Effectiveness (Q2, Q6): We successfully identified customers who used a discount but still had an above-average purchase amount (Q2). This group is key because they are high-value buyers who respond well to promotions. The top products with applied discounts (Q6) showed where promotions are currently concentrated.

6. Conclusion and Recommendation :

6.1 Project Summary: The project successfully analyzed customer data, providing clear quantitative answers to the most important business questions. We now have a solid understanding of which customer segments are the most valuable and which products are performing best.

6.2 Recommendations:

- **Prioritize Subscribers and Loyal Customers:** Focus marketing and loyalty efforts primarily on the Loyal and Subscribed segments, as they provide the highest returns (Q5, Q7).
- **Targeted Promotions:** Use the list from Q2 (high spenders who use discounts) to create highly targeted, personalized discount campaigns, rather than offering site-wide discounts that might hurt profit margins.
- **Optimize Shipping:** Use the comparison of average purchase amounts between Standard and Express Shipping (Q4) to determine if offering different shipping thresholds or prices would encourage higher spend.

7. Future Work :

- **Predictive Modeling:** The next logical step would be to build a predictive model (like a classification model) to forecast which non-subscribed customers are most likely to convert to a subscriber, allowing for highly efficient marketing spend
- **Profitability Analysis:** Analyze the actual cost of goods sold (COGS) to move beyond just *revenue* and identify the most profitable products and customer segments.

8. Appendix :

→ **Link to Code Repository (GitHub):** https://github.com/GouravBerwal

(This repository contains the Python notebooks for EDA and the SQL scripts used for aggregation.)