Customer Segmentation Report

1. Introduction

This project aims to segment customers based on their purchasing behavior using K-Means clustering. Segmentation allows businesses to understand their customers better, personalize marketing strategies, and improve customer satisfaction.

2. Methodology

The following steps were performed:
- Loaded and cleaned transaction data from an online retail dataset.
- Engineered a 'TotalSpend' feature (UnitPrice x Quantity).
- Scaled the data for consistency using StandardScaler.
- Applied K-Means clustering to group customers into meaningful segments.
- Used Principal Component Analysis (PCA) to visualize the clusters in 2D.
- Summarized cluster behavior based on UnitPrice, Quantity, and TotalSpend.

3. Cluster Summary

The table below shows the average behavior of each cluster:

Cluster	Avg. UnitPrice	Avg. Quantity	Avg. TotalSpend
0	3.36	12.06	20.50
1	1.56	-77605.00	-122826.60
2	1.56	77605.00	122826.60
3	38970.00	-1.00	-38970.00

4. Interpretation

Each cluster represents a group of customers with similar purchasing behavior. For example, Cluster 0 represents average customers with typical spending habits, while Clusters 1 and 2 contain large negative and positive values respectively, which may indicate returns or wholesale transactions. Cluster 3 seems to contain outliers with unusual behavior.

5. Business Recommendations

- Cluster 0: Consider loyalty programs or targeted promotions to increase spending.
- Cluster 1: Investigate large returns or cancellations for quality or policy issues.
- Cluster 2: These may be bulk buyers; offer wholesale discounts or personalized service.
- Cluster 3: Review for data errors or outliers and exclude from general campaigns.

6. Conclusion

K-Means clustering provided meaningful segmentation of customers. These insights can guide targeted business strategies, leading to more efficient marketing and improved customer relationships.