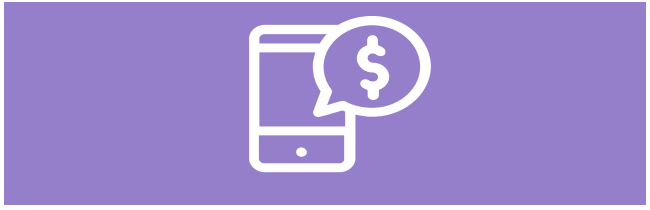
Customer Segmentation Using K-Means Clustering

A Menternship Project on
Online Retail Analytics

The Challenge

Retailers struggle to personalize marketing for diverse customer bases.



Needed a way to understand and segment customers based on real transaction data.

Goal: Improve **targeting**, **retention**, and **revenue** using customer insights.

Approach & Tools Used

Method

A

Unsupervised learning using K-Means Clustering.

Tools

B

Python, Pandas, Matplotlib, Seaborn, Scikit-learn.

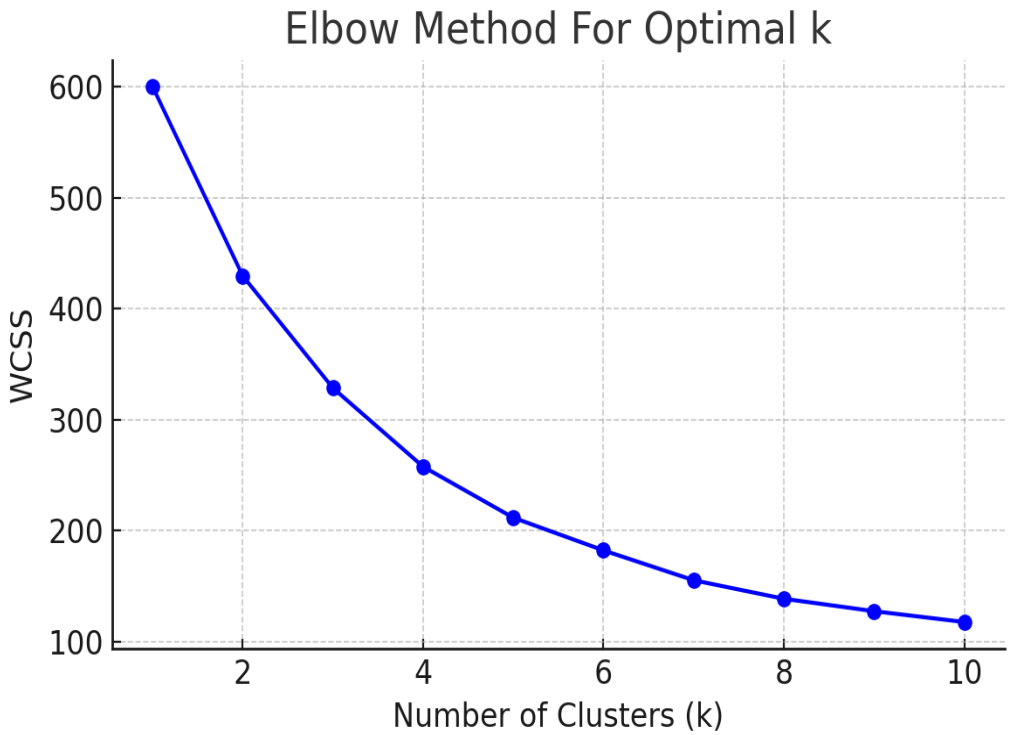
Key Features Used

C

- Recency (last purchase)
- Frequency (number of purchases)
- Monetary Value (total spent)

Steps Taken

- Data preprocessing and cleaning
- Feature engineering: created RFM variables`
- Standardized data and used Elbow method to determine optimal clusters
- Applied K-Means algorithm
- Visualized results using PCA plots and cluster summaries



Key Findings – 4 Customer Segments

0
Loyal Big Spenders

Frequent, recent, high-spending customers

1
Occasional Buyers

Moderate behavior, churn risk

3
Dormant Users

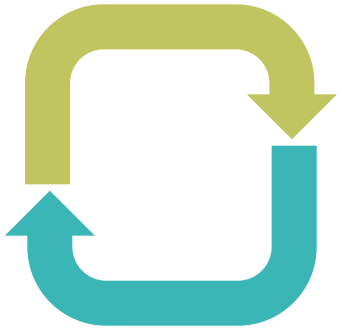
Inactive, need re-engagement

2
Bulk Buyers

Price-sensitive, large quantity buyers

PCA Cluster Visualization

Business Impact



Enabled targeted strategies:

- **VIP programs** for top spenders
- **Win-back campaigns** for dormant users
- **Discounts** for bulk buyers

Improved potential for

customer retention, revenue growth, and marketing ROI

Conclusion

Learned hands-on application of clustering to solve real-world business problems

Gained skills in data analysis, modeling, and insight generation

Excited to apply these insights to future roles in data science and analytics!