Customer Segmentation Using K-Means Clustering

Intern Name: Krishna Kumar
Intern ID: 217256336

Company: Indolike

Project Duration: July 2025

1. Objective

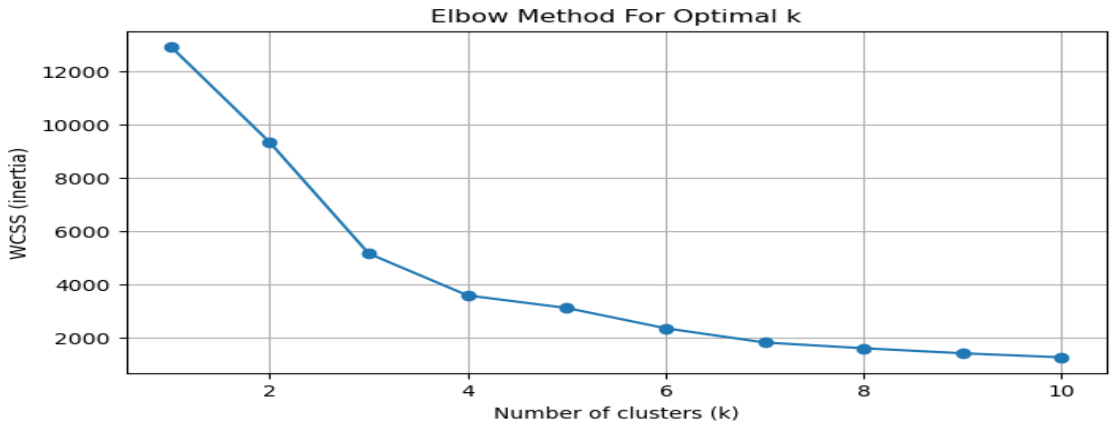
To segment customers based on purchasing behavior using RFM (Recency, Frequency, Monetary) analysis and K-Means clustering. This helps Indolike understand customer segments better and make data-driven marketing decisions.

2. Dataset

- Source: Online Retail Dataset from Kaggle
- Size: ~400,000 records
- Fields: Invoice No, Product Description, Quantity, Invoice Date, Unit Price, Customer ID, Country

3. Methodology

- Data Cleaning: Removed missing values, returns, and duplicates
- RFM Feature Engineering: Calculated Recency, Frequency, Monetary values
- Scaling: Standardized data using StandardScaler
- Clustering: Applied K-Means with k=4 (using Elbow Method)
- Visualization: PCA used for 2D cluster plotting



4. Cluster Insights

Cluster	Recency	Frequency	Monetary	Num_Customers	Segment Type
0	43.03	4.46	₹1710.65	3204	Loyal Customers
1	242.98	1.66	₹593.54	1047	At-Risk Customers
2	5.60	113.60	₹2,15,535	5	Super VIPs
3	14.91	47.02	₹28,896.42	56	High-Value Active

5. Recommendations

- Loyal Customers: Offer loyalty programs or early access
- At-Risk Customers: Re-engagement campaigns or discounts
- Super VIPs: Assign personal managers and exclusive deals
- High-Value Active: Upsell premium items, engage regularly

6. Conclusion

K-Means helped identify 4 distinct customer segments. These insights will support targeted marketing and customer retention efforts at Indolike.