Customer Segmentation Using K-Means Clustering

A Menternship Project on Online Retail Analytics prepared for Snapdeal

The Business Challenge

What Snapdeal Faced

- Wide variation in customer purchase behavior
- Low marketing efficiency due to lack of segmentation
- Difficulty identifying high-value or at-risk customers

The Core Problem

Snapdeal needed a way to **analyze real transaction data** to better **understand, group, and target** its customer base for improved retention and growth

Project Goal

- Segment customers using behavioral data
- Enable **personalized marketing, win-back strategies, and VIP targeting**
- Improve **customer retention, revenue, and marketing ROI**

Approach & Tools Used

Method

- Unsupervised learning using K-Means Clustering.
- Grouped customers based on transaction behavior

A

Key Features Used

- **Python** – Analysis scripting
- **Pandas** – Data handling
- **Scikit-learn** – Clustering model
- **Matplotlib & Seaborn** – Visualizations
- **PCA** – Visualizing clusters in 2D

B

Tools & Technologies

- **Python** – Analysis scripting
- **Pandas** – Data handling
- **Scikit-learn** – Clustering model
- **Matplotlib & Seaborn** – Visualizations
- **PCA** – Visualizing clusters in 2D

C

Steps Taken

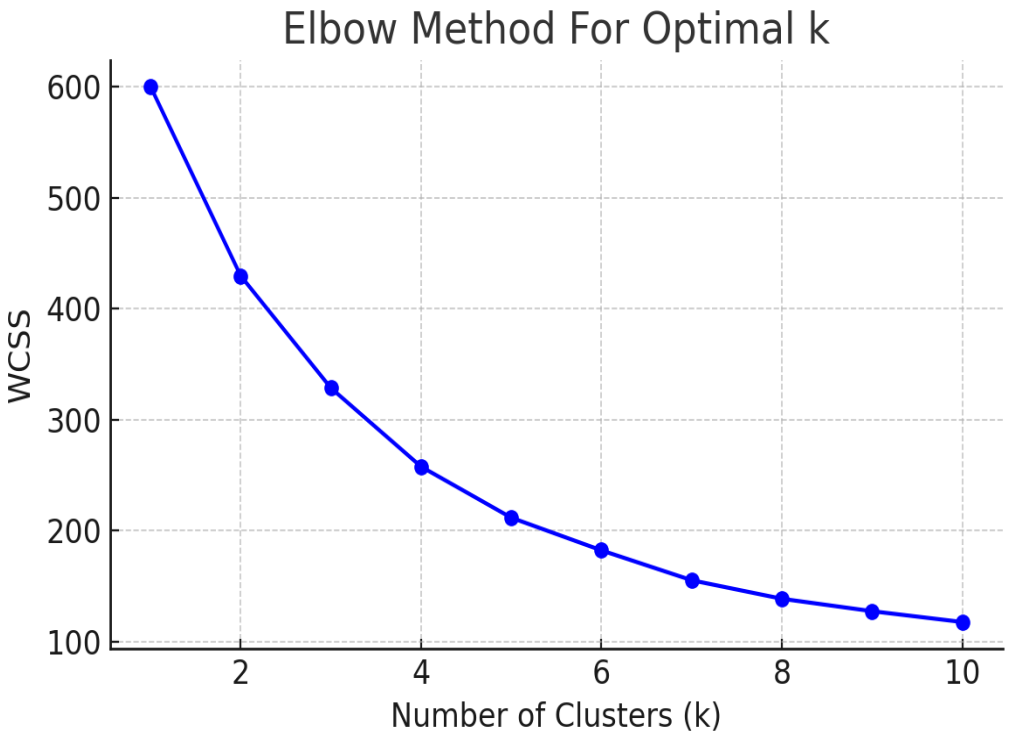
1. **Data preprocessing and cleaning:**
   Removed missing or duplicate records.
   Filtered relevant transaction fields

2. **Feature engineering(RFM):**
   Created three key variables:
   Recency (Days since last purchase**), Frequency(** Number of purchases**), Monetary(**Total spend**)**

3. **Standardized data and used Elbow method to determine optimal clusters**

4. **Applied K-Means algorithm:**
   Grouped customers into 4 segments based on RFM patterns

5. **Cluster Visualization:**
   Applied PCA (Principal Component Analysis) to reduce dimensions.Visualized customer segments in 2D for clear interpretation



Key Customer Segments Identified

Each segment displays distinct behaviors that can be leveraged to **personalize marketing and maximize customer lifetime value.**

0
Loyal Big Spenders

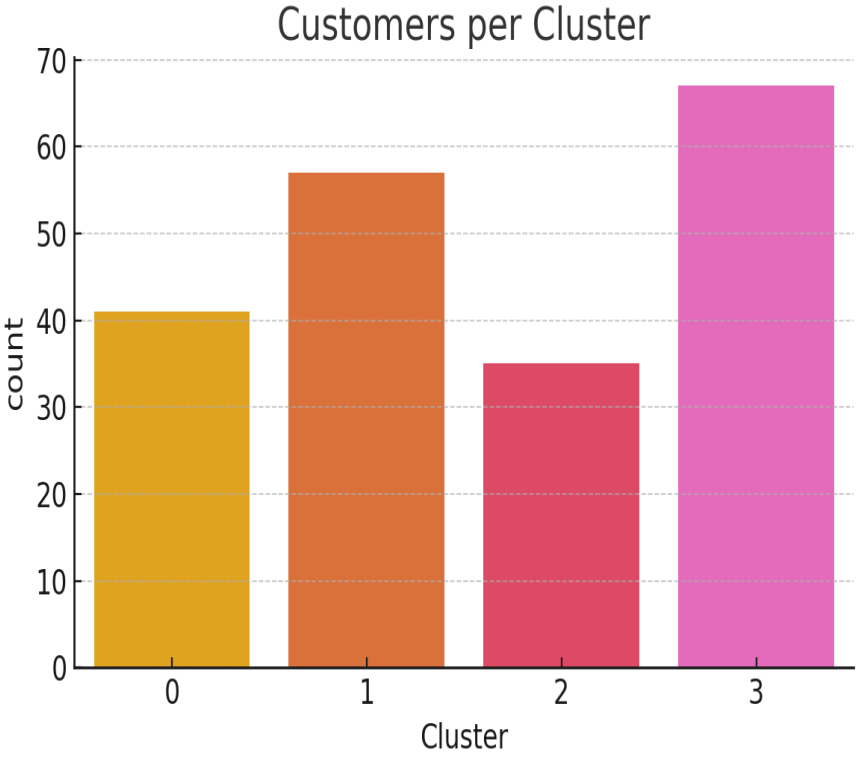
Frequent, recent, high-spending customers

Recommended Strategy:
Launch VIP programs and early-access deals

1
Occasional Buyers

Moderate frequency & spend, risk of churn

Recommended Strategy:
Use personalized emails and retention offers



3
Dormant Users

Inactive, low engagement and spend

Recommended Strategy:
Run win-back campaigns with time-limited deals

2
Bulk Buyers

Infrequent but high-volume purchases, price-sensitive

Recommended Strategy:
Offer bulk discounts and bundle promotions

PCA Cluster Visualization

Purpose of Visualization

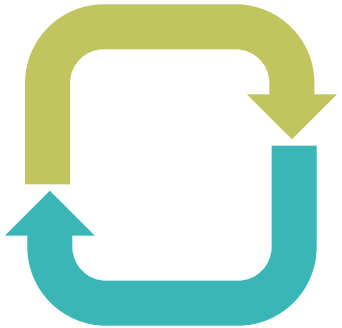
•Reduced high-dimensional data into 2D using Principal Component Analysis (PCA)
•Visualized how distinct the customer clusters are

Insights from PCA Plot

•**Four clear clusters** are visible, confirming meaningful segmentation
•Clusters show natural separation based on RFM patterns
•Visualization validates the effectiveness of the K-Means model

Business Impact

Data-driven segmentation enables Snapdeal to **personalize user experience, boost loyalty, and increase lifetime value.**



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

Project Outcomes

- Successfully segmented customers into **4 distinct groups** using K-Means clustering
- Enabled **data-driven strategies** for **retention**, **revenue growth**, and **personalized marketing**
- Delivered **actionable insights** backed by visualization and customer behavior analysis

Next Steps for Snapdeal

- Implement personalized campaigns based on customer segments
- Monitor performance and continuously refine the segmentation model
- Explore predictive modeling for customer churn or lifetime value

Skills & Experience Gained

- Applied **machine learning** to solve a real-world business challenge
- Enhanced skills in **data preprocessing, feature engineering (RFM), and model evaluation**
- Gained experience in **translating data insights into business strategy**

THANK YOU