**Task 2: Customer Segmentation Using K-Means Clustering**

**1. Objective**

The objective of this task is to segment customers into meaningful groups using unsupervised machine learning techniques. The segmentation aims to help businesses understand customer characteristics and design targeted marketing strategies, particularly for high-value customer segments.

Due to the absence of transactional variables required for RFM analysis (such as transaction date and monetary value), **K-Means clustering** was selected as the appropriate segmentation method.

**2. Dataset Description**

The dataset consists of **6,665 customer records** after preprocessing. Each record contains demographic and behavioural attributes such as:

- Age
- Work Experience
- Family Size
- Spending Score (categorical)

These variables collectively describe customer profiles and spending behaviour.

**3. Data Preprocessing**

The following preprocessing steps were applied before clustering:

1. **Handling Missing Values**
   Rows containing missing values were removed to ensure consistency and algorithm compatibility.
2. **Data Type Corrections**
   - Family_Size was converted from float to integer.
   - Spending_Score (categorical) was converted into numeric form using **Label Encoding**.
3. **Feature Selection**
   The following features were selected for clustering:

- Age
- Work Experience
- Family Size
- Spending Score (Encoded)

4. **Feature Scaling**
   All selected variables were standardized using **StandardScaler** to ensure equal contribution to distance-based clustering.

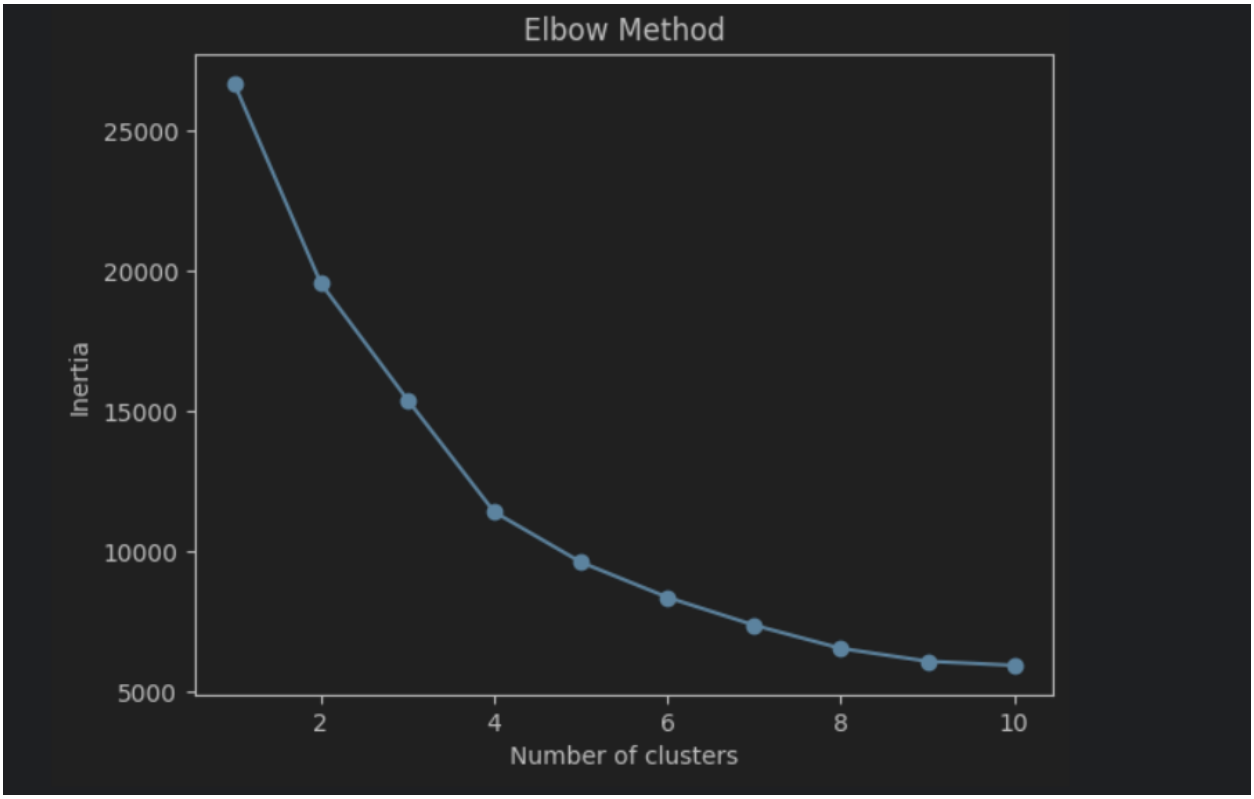
## 4. Methodology

### 4.1 Clustering Technique

K-Means clustering was used to group customers based on similarity across selected features.

### 4.2 Optimal Cluster Selection (Elbow Method)

The Elbow Method was applied by computing inertia values for cluster sizes ranging from 1 to 10.
A noticeable reduction in inertia was observed up to **3 clusters**, after which the improvement diminished. **Figure 1: Elbow Method Plot**




Based on this analysis, **3 clusters** were selected as the optimal number.

**5. Cluster Profiling Results**

After applying K-Means clustering, customer segments were profiled using average values of key variables.

| Cluster | Avg Age | Avg Work Experience | Avg Family Size | Avg Spending Score | Customer Count |
|---|---|---|---|---|---|
| 0 | 59.22 | 1.22 | 1.68 | 1.73 | 1,850 |
| 1 | 47.43 | 2.36 | 3.25 | 0.18 | 2,021 |
| 2 | 30.34 | 3.76 | 3.32 | 1.95 | 2,794 |

**6. Segment Interpretation**

**Cluster 0 – Affluent Low-Family Segment**

- Older customers
- Smaller family size
- Moderate-to-high spending behaviour
  This segment likely values stability, premium offerings, and personalised services.

**Cluster 1 – Conservative Mid-Age Segment**

- Middle-aged customers
- Moderate work experience
- Lowest spending score among all clusters
  This group appears price-sensitive and cautious in spending.

**Cluster 2 –Young High-Spending Segment (High-Value)**

- Youngest customer group
- Highest work experience
- Largest family size
- Highest spending score
  This cluster represents the **most valuable and growth-oriented customer segment**.

**7. Targeted Marketing Strategies (High-Value Segment – Cluster 2)**

**1. Personalised Digital Campaigns**

Design customised offers, product bundles, and promotions tailored to younger customers with high spending capacity.

**2. Loyalty & Rewards Programs**

Introduce reward-based programs offering discounts, cashback, or exclusive benefits to increase retention and lifetime value.

**3. Cross-Selling and Upselling**

Promote complementary products and premium upgrades to maximise revenue from this segment.

**8. Business Impact**

Customer segmentation enables organizations to:

- Optimize marketing spend
- Improve customer engagement
- Increase revenue through targeted strategies
- Enhance long-term customer relationships

By focusing on high-value segments, businesses can significantly improve overall profitability.

**9. Conclusion**

This analysis successfully segmented customers into three meaningful groups using K-Means clustering. The approach was data-driven, justified using the Elbow Method, and resulted in actionable insights. The findings provide a strong foundation for targeted marketing strategies and customer relationship management.