

**Predictive Analytics**

**Project**

**BATCH - 7 B. Tech CSE (AI &ML)**

**Submitted to-**

**Achala Shakya**

**Submitted by-**

1. **NAME -Anshuman Agarwal**

**SAP ID- 500125828**

1. **NAME-Adit Jain**

**SAP ID: 500101867**

**Project Report: Customer Segmentation and Consumer Behaviour Analysis**

**1. Introduction**

This project focuses on **segmenting customers based on their purchasing behaviour** to uncover insights and enable data-driven decisions for targeted marketing strategies. Using real-world sales data from a UK-based retailer, the project aims to group customers with similar purchasing behaviours to help the business identify high-value segments and locations.

**Objectives**

1. Segment customers based on purchasing behaviour by applying the **k-means clustering algorithm**.
2. Identify the top 20% of:
   * Customers,
   * Products, and
   * Geographic locations that contribute to 80% of sales revenue, following the **Pareto Principle (80/20 Rule)**.

**2. Data Description**

The dataset contains sales data from a UK retailer, with various features related to customer purchases, including transaction amounts, customer location, product categories, and other relevant details. The key steps include loading, cleaning, and preprocessing the data for analysis.

**3. Methodology**

The project uses **k-means clustering** for segmentation analysis. Here’s a step-by-step summary of the approach:

**Step 1: Data Loading and Preprocessing**

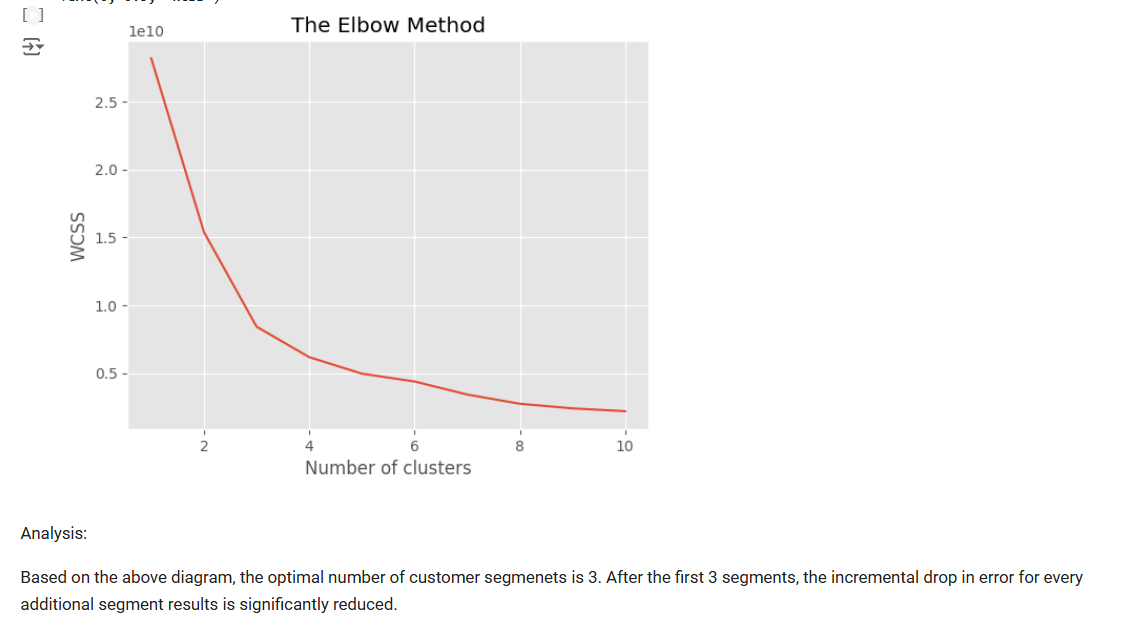
* Imported essential libraries (NumPy, Matplotlib, Pandas) for data handling, visualization, and analysis.
* Loaded and mounted the dataset from Google Drive.
* Cleaned the data to remove any missing values or inconsistencies to ensure accuracy.

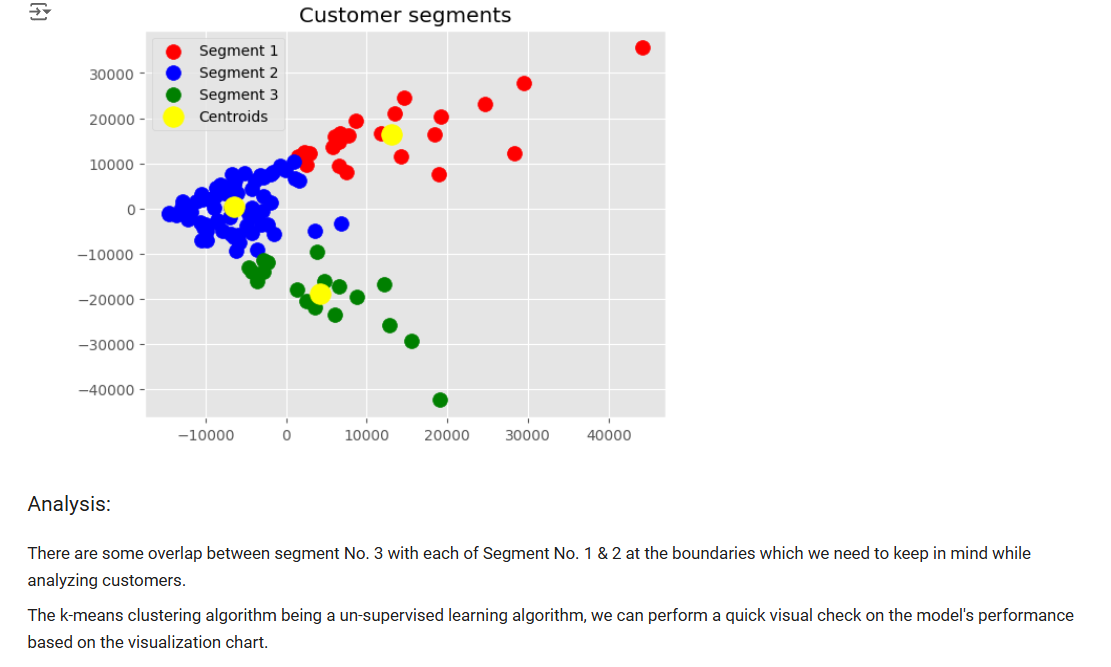
**Step 2: Feature Engineering**

* Extracted relevant features and constructed new variables based on customer spending patterns, purchase frequency, and geographic data.
* Normalized data to ensure all features are on a similar scale, enhancing the performance of the k-means algorithm.

**Step 3: Applying K-Means Clustering**

* Implemented the k-means clustering algorithm to identify customer segments.
* Used the **Elbow Method** to determine the optimal number of clusters, which helps in balancing cluster cohesion and separation.





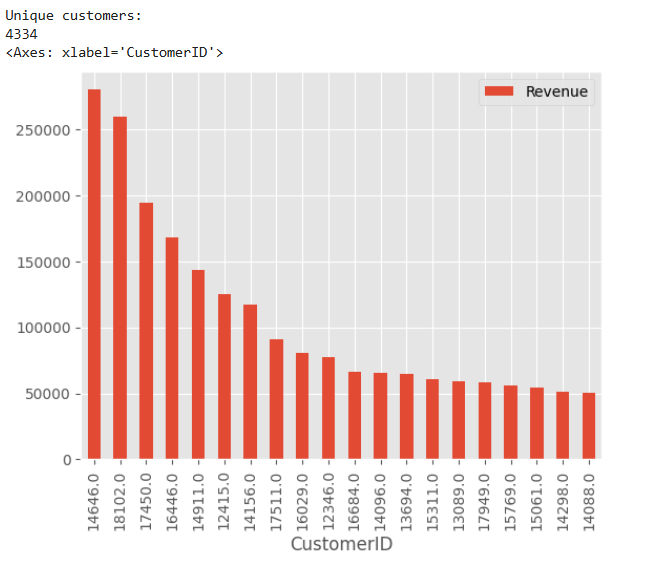
**Step 4: Analysis of Results**

* Analysed the resulting clusters to understand the distinct customer groups based on purchase behaviours.
* Applied the **80/20 rule** to identify the top customers, products, and geographic regions contributing to most of the revenue.

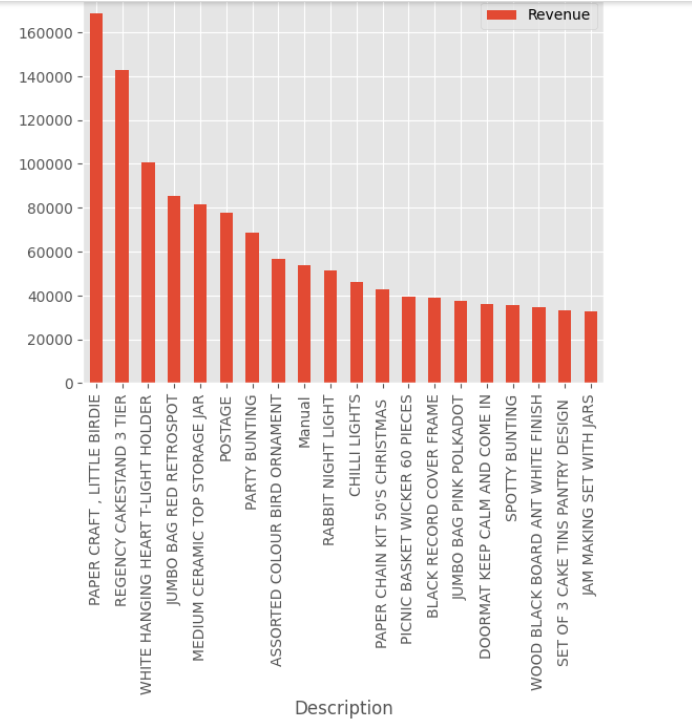
**4. Results and Findings**

* Identified key customer segments, each representing unique buying behaviours and characteristics.
* Identified the top 20% of customers, products, and locations that account for the majority of the revenue, aligning with the 80/20 rule. These high-value segments can help the retailer focus marketing efforts and resources on the most profitable areas.

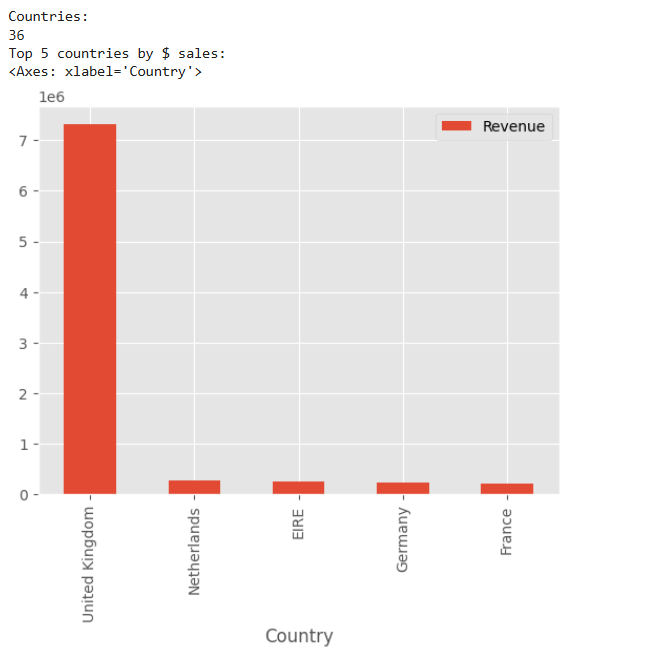
**CUSTOMERS:**



**PRODUCTS:**



**GEOGRAPHICAL LOCATIONS:**



**5. Conclusion**

This analysis provides valuable insights for the retailer, enabling them to:

* Tailor marketing campaigns to specific customer segments,
* Optimize product offerings, and
* Focus resources on high-revenue locations.

The clustering approach is effective in revealing distinct customer behaviours, assisting the retailer in creating targeted strategies to improve sales and customer satisfaction.