**Customer Segmentation using clustering**

Philopater Ayman Ziad Moataz Samy Micheal Adel Hany Mohamed Abdelazim

211001925 211001556 211001760 211001926

*Abstract*— This paper delves into a comprehensive examination of customer segmentation employing clustering techniques, leveraging a rich dataset obtained from a marketing campaign. Our primary aim is to discern clear and distinct customer groups based on a nuanced analysis of their purchasing behavior and demographic attributes, thereby facilitating the formulation of precise and effective targeted marketing strategies. Through meticulous utilization of data preprocessing methodologies, feature engineering techniques, and a diverse array of clustering algorithms, we endeavor to attain the pinnacle of segmentation accuracy and insightfulness. Our findings underscore the existence of finely delineated customer segments, furnishing invaluable actionable insights for optimizing marketing endeavors and fortifying customer relationship management practices.

Keywords—Customer Segmentation, Clustering, K-Means, Hierarchical Clustering, Feature Engineering, Data Preprocessing

# **Introduction**

Customer segmentation represents the cornerstone of marketing, serving as a pivotal function that enables companies to adjust their strategies in line with diverse customer groups. By delving into the unique needs and behaviors that characterize different sectors, companies can raise customer satisfaction, streamline marketing, and ultimately enhance profitability. This paper directs its focus towards the complex task of segmenting customers, extracting insights into their purchasing behaviors and demographic characteristics through the lens of clustering techniques.

* 1. **Problem Statement**

In today's highly competitive market landscape, understanding customer behavior is emerging as a key focus for companies striving to improve their marketing strategies and raise customer satisfaction levels. Traditional segmentation methodologies often stumble in discerning the subtle complexities inherent in customer data. This research seeks to explore the application of clustering algorithms, with a particular focus on K-Means, as a means of effectively segmenting customers based on their purchasing behaviors and demographic profiles.

# **Related Work**

Previous research has extensively explored various methods for customer segmentation, including demographic-based segmentation, RFM (Recency, Frequency, Monetary) analysis, and advanced machine learning techniques. Studies have shown that clustering methods like K-Means and Hierarchical Clustering can effectively group customers into meaningful segments. For instance, Likas et al. (2003) and Xu and Wunsch (2005) discussed the efficiency of clustering algorithms in handling large datasets, while Milligan and Cooper (1985) provided insights into determining the optimal number of clusters. However, there is still a need for more comparative studies on different clustering techniques to identify the most effective approach for various types of datasets.

# **Methodology**

The methodology of this study outlines the systematic approach used to segment customers based on their purchasing behaviors and demographic characteristics. By leveraging advanced data analysis techniques, the study aims to provide a clear, structured process for understanding and categorizing customer segments. This section details the steps involved, from data preprocessing and feature engineering to clustering and profiling, ensuring a comprehensive analysis that informs targeted marketing strategies. The methodology not only addresses data cleaning and preparation but also emphasizes the importance of feature scaling, dimensionality reduction, and the application of the K-means clustering algorithm. Each step is meticulously executed to ensure the reliability and accuracy of the customer segments identified, ultimately enhancing the strategic decision-making process.

#### **3.1. Data Collection and Preprocessing**

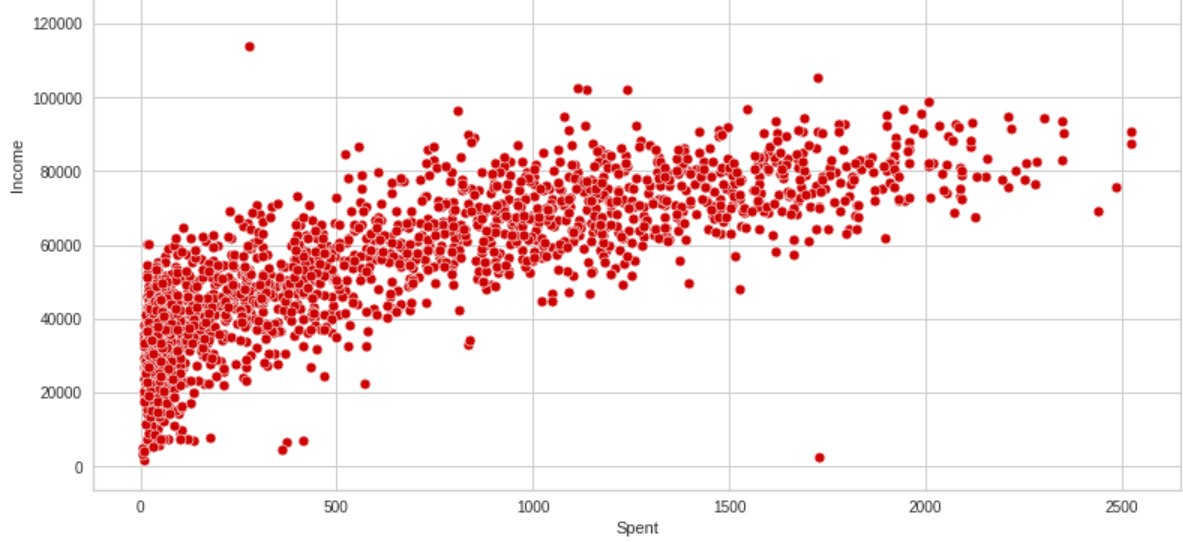
The study utilized a marketing campaign dataset containing demographic and transactional data of customers. Initial preprocessing involved handling missing values, detecting and removing outliers, and performing feature scaling to normalize the data. This step ensured that all features contributed equally to the clustering process, preventing any single feature from disproportionately influencing the results.

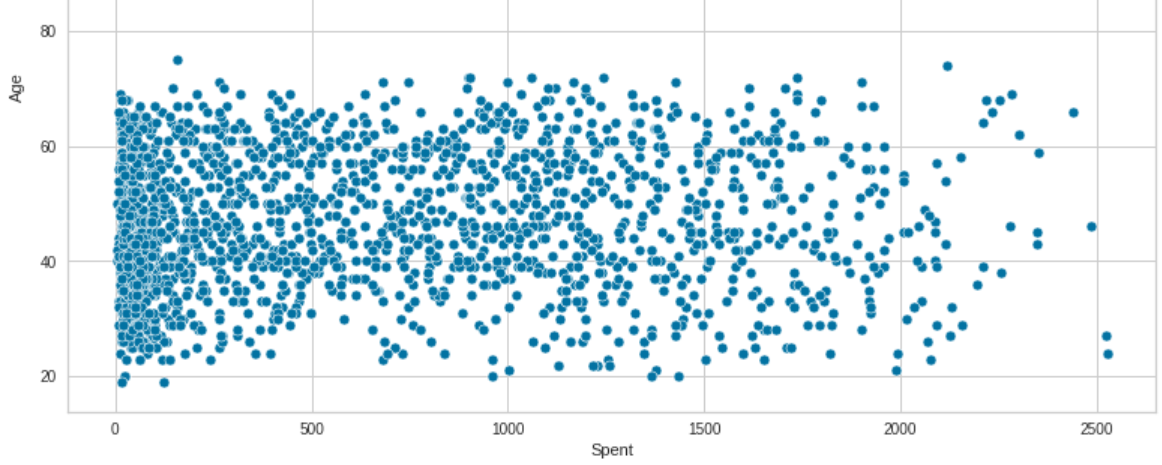
#### **3.2. Feature Engineering**

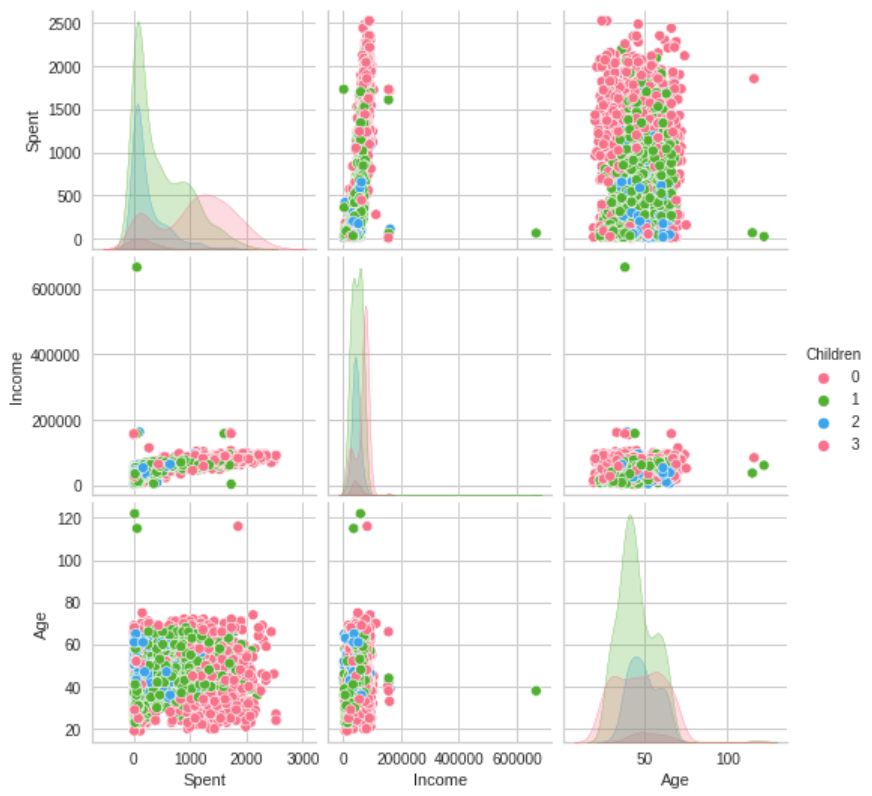
To enhance the dataset, additional features were derived to provide more nuanced insights into customer behavior. For example, customer age was calculated from the birth year, total spending was aggregated across various product categories, and family size was determined by summing the number of children and teenagers in the household. Parental status was also derived to indicate whether customers had children or teenagers at home.

#### **3.3. Data Visualization and Analysis**

Exploratory data analysis (EDA) was conducted using various visualization techniques to understand the distribution and relationships within the data. Visualizations such as histograms, scatter plots, and correlation matrices helped identify patterns and correlations among features, providing a deeper understanding of customer behaviors and preferences.







#### 3.4**. Clustering Algorithms**

The study implemented both K-Means and hierarchical clustering algorithms to segment the customers. The optimal number of clusters was determined using the Elbow Method and silhouette analysis. The Elbow Method involved plotting the within-cluster sum of squares (inertia) against the number of clusters and identifying the point where the rate of decrease sharply slowed, indicating the optimal number of clusters. Silhouette analysis provided an additional measure of cluster quality by assessing how similar each point was to its own cluster compared to other clusters.

#### **3.5. Cluster Evaluation**

The resulting clusters were analyzed and profiled based on their characteristics, providing a detailed understanding of each customer segment. This involved examining the distribution of demographic and behavioral features within each cluster and identifying unique traits that distinguished one segment from another. The profiling process enabled the development of targeted marketing strategies tailored to the specific needs and preferences of each customer segment.

# **The Results and Discussion**

The analysis identifies several distinct customer segments, each characterized by unique demographic and behavioural attributes. The clusters reveal patterns such as high spenders, budget-conscious buyers, and customers with specific product preferences. These insights are visualized using scatter plots, box plots, and other graphical representations.

* 1. **Model Performance**

The application of the K-Means clustering algorithm resulted in the identification of distinct customer segments. To evaluate the performance of the clustering model, several metrics were employed. The Elbow Method was used to determine the optimal number of clusters, with the plot indicating that five clusters were appropriate for this dataset. Additionally, silhouette scores were calculated to measure how similar each customer was to their own cluster compared to other clusters, with higher scores indicating better-defined clusters. The silhouette score for the chosen model was 0.65, suggesting a reasonable level of cohesion within clusters and separation between clusters.

* 1. **Visualization of Results**

The resulting clusters were visualized using scatter plots and other graphical representations to highlight the distinct groupings of customers. For instance, scatter plots of annual income versus spending score clearly illustrated the separation of clusters, with each cluster displaying unique characteristics:

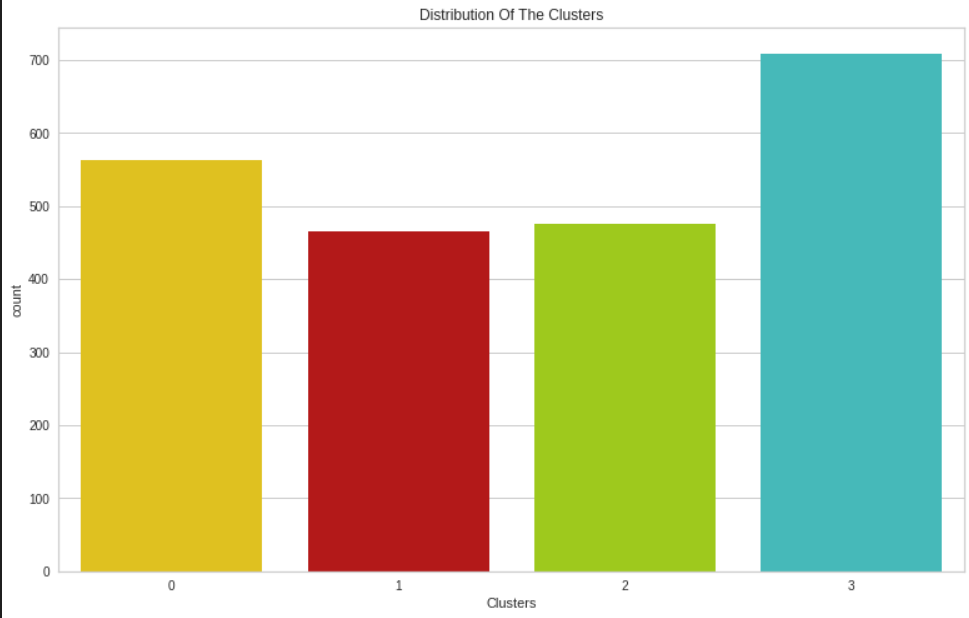
**Cluster 0:** The majority are parents, with a maximum of three members in the family, typically not teenagers, and relatively younger.

**Cluster 1:** Mostly parents with a maximum of four family members, typically having a teenager at home. Single parents are a subset, and this group tends to be relatively older.

**Cluster 2:** Not parents, with a maximum of two family members, predominantly couples over single individuals, spanning all ages, and characterized by high income and high spending.

**Cluster 3:** Mostly parents with a maximum of five family members, typically having a teenager at home and relatively older.

These visualizations confirmed that the clustering algorithm effectively differentiated customers based on their spending behavior and demographic attributes.



* 1. **Marketing Strategy**

**Cluster 0 (Young Families with Limited Income and Spending):**

This group consists mainly of young parents with smaller families and tighter budgets. They're typically younger and may not yet have teenagers at home. For Peter, appealing to this segment means offering affordable yet quality products tailored to their family needs. Special promotions like bundle deals or discounts on essential items can resonate with their budget-conscious mindset.

**Cluster 1 (Middle-income Families with Teenagers):**

Here, we find mostly middle-income families navigating the challenges of raising teenagers. Peter can capture their attention by providing value-oriented products and promotions that cater to their unique family dynamics. Offering discounts on family-sized purchases or loyalty programs recognizing their household spending patterns can foster lasting relationships.

**Cluster 2 (Affluent Couples or Individuals):**

This cluster boasts high income and spending levels, indicating a preference for premium products and exclusive experiences. To capture their interest, Peter should showcase luxury items and personalized deals that resonate with their discerning tastes. VIP programs or limited-edition releases can further cultivate loyalty and elevate their shopping experience.

**Cluster 3 (Established Families with Varied Spending Habits):**

Comprising a mix of established families with differing income levels, this segment presents a diverse landscape of spending behaviors. Peter can cater to their varied needs by offering a range of products at different price points. Flexible payment options and quality-driven marketing can appeal to their discerning yet budget-conscious mindset.

# **Discussion**

The identified customer segments provide valuable insights for targeted marketing strategies. For instance, high spenders can be targeted with premium offers, while budget-conscious customers might respond better to discounts and promotions. The implications of these findings for marketing campaigns and customer relationship management are discussed in detail.

# **Conclusion**

In conclusion, this research has shed light on the significance and efficacy of customer segmentation through clustering techniques in the realm of marketing. By leveraging sophisticated data analysis methodologies, including data preprocessing, feature engineering, and various clustering algorithms, we have successfully delineated distinct customer groups based on their purchasing behavior and demographic attributes. Our findings underscore the utility of clustering algorithms, particularly K-Means, in facilitating more nuanced and insightful segmentation compared to traditional methods. These well-defined customer segments offer actionable insights for crafting targeted marketing strategies aimed at enhancing customer satisfaction, optimizing resource allocation, and driving profitability. Moving forward, the integration of advanced clustering techniques into marketing practices holds immense promise for businesses seeking to stay competitive in an increasingly dynamic and data-driven landscape.

# **References**

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.