

# REPORT ON CUSTOMER SEGEMENTATION IN AN ONLINE RETAIL MARKET

## 1. Introduction

The primary objective of this project was to design and implement a sophisticated customer segmentation system tailored for online retailers. This system leverages a hybrid approach, combining the strengths of both k-means and hierarchical clustering algorithms. The ultimate aim was to create an effective solution for building targeted promotional strategies, thereby enhancing customer engagement and increasing sales revenue.

**2. Rationale for Hybrid Approach:** The hybrid approach was chosen due to a strategic combination of advantages offered by both k-means and hierarchical clustering algorithms. K-means was selected for its superior performance and efficiency, especially when dealing with larger datasets. Its popularity and widespread usage in segmentation tasks further validated its selection. On the other hand, hierarchical clustering was preferred for its ability to produce more accurate results with higher values of k (number of clusters). Moreover, the results obtained from hierarchical clustering are inherently more interpretable and easier to analyze.

**3. Data Pre-processing and Cluster Selection:** It's worth noting that the success of the segmentation system heavily relies on proper data pre-processing techniques and intelligent cluster selection. While the choice of algorithm is crucial, the quality of data and the number of clusters selected play equally pivotal roles. The hybrid approach capitalizes on this synergy, utilizing both algorithms to extract the most meaningful insights from the data

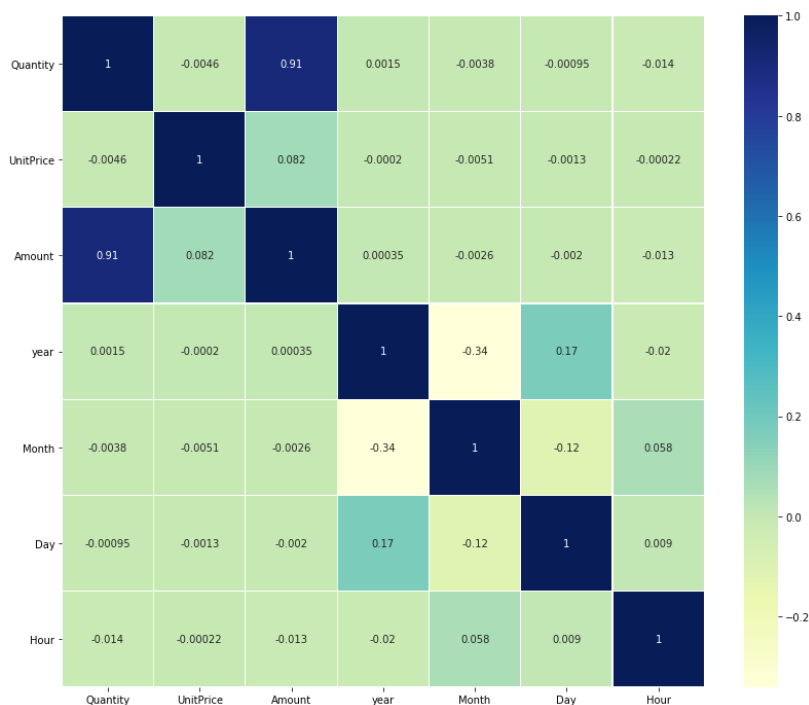
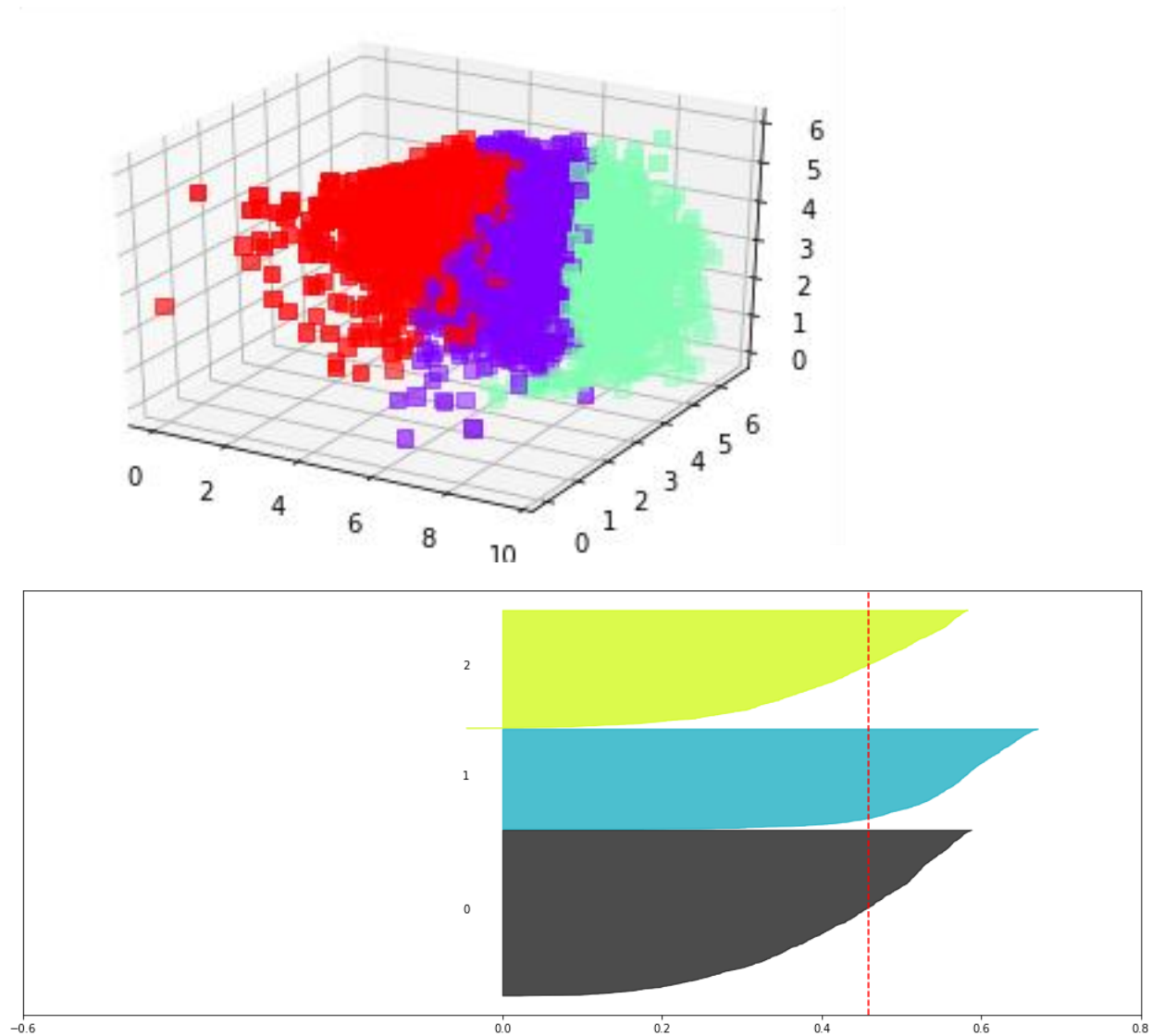


FIGURE: Correlation matrix for different features

## 4. Results and Performance

The hybrid customer segmentation model yielded impressive results, achieving an accuracy rate of 86.8%. This accuracy is a testament to the effectiveness of the hybrid approach in correctly classifying customers into distinct segments. It is important to highlight that hierarchical clustering demonstrated exceptional performance with smaller datasets, but its computational time increased significantly when dealing with larger datasets. In contrast, k-means exhibited greater efficiency with larger datasets. The adjusted random score, a metric that quantifies the similarity between different clustering solutions, indicated a strong level of agreement between the results of k-means and hierarchical clustering, with a score of approximately 96%. This reinforces the idea that the hybrid approach has effectively leveraged the strengths of both algorithms to create a more robust and reliable segmentation system.



**Figure: Silhouette for different clusters**

**5. Conclusion:** In conclusion, the hybrid k-means and hierarchical clustering-based customer segmentation system developed for online retailers has proven to be a valuable asset for targeted

promotions. By combining the speed and efficiency of k-means with the accuracy and interpretability of hierarchical clustering, this system has demonstrated its ability to effectively categorize customers into meaningful segments. The choice of approach has been well-justified based on the respective strengths of the algorithms, and the achieved results highlight the synergy achieved through this hybrid strategy. As the landscape of online retail continues to evolve, this hybrid approach can serve as a foundation for refining and enhancing customer engagement strategies. It presents an innovative solution for online retailers seeking to optimize their marketing efforts and maximize returns on promotional campaigns.