Customer Personality Analysis through Clustering and Fitting: A Data-driven Approach

Introduction:
Customer Personality Analysis is pivotal for businesses aiming to understand and segment their customer base effectively. By deciphering customer needs and behaviors, companies can tailor their marketing strategies to enhance customer satisfaction and operational efficiency (Wedel & Kamakura, 2000). Methods such as K-Means and Agglomerative Clustering automate the process of identifying distinct customer segments, enabling targeted marketing efforts and personalized product recommendations (Kotler & Keller, 2006).

Dataset Overview:
This dataset provides a comprehensive view of customer demographics, purchasing behavior, and responses to marketing campaigns. It serves as a foundation for customer segmentation, empowering businesses to optimize marketing expenditures and improve product offerings based on specific customer preferences (Customer Personality Analysis, 2021).

Data Visualization:
Pair Plot: Examining relationships across income, recency, and spending reveals nuanced customer behaviors. Notably, both parents and non-parents exhibit substantial spending patterns across different income levels, challenging assumptions about spending behavior solely based on income.

Barplot: Displays average spending by education level, indicating higher expenditures among postgraduates compared to other educational groups. This insight underscores the influence of education on consumer behavior, guiding targeted marketing strategies.
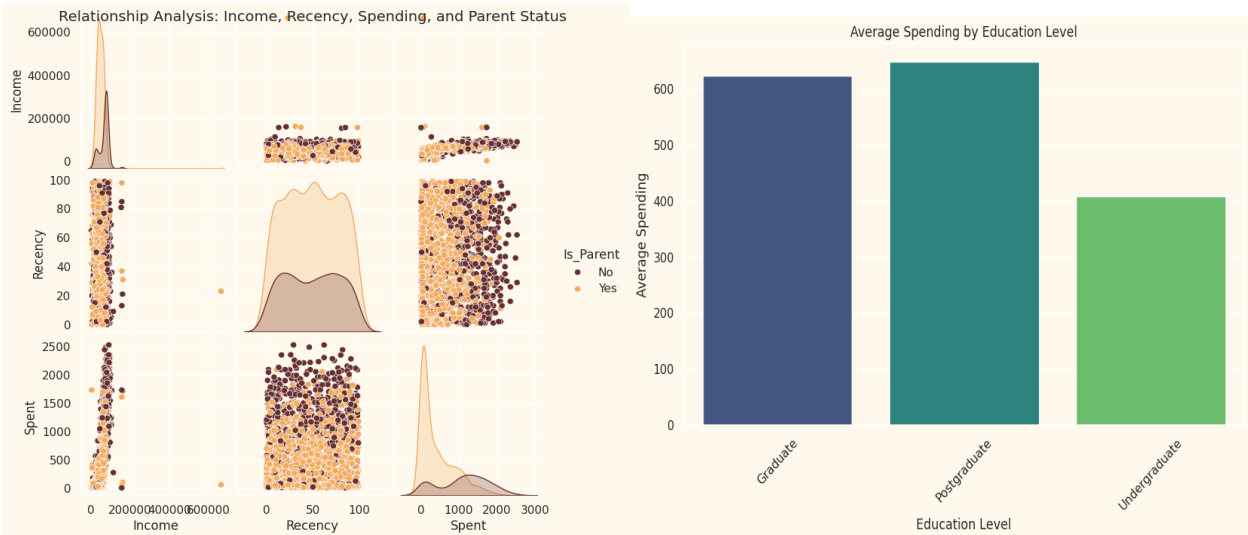

Fig 1: Pair Plot for Relationship Analysis   Fig 2: Bar plot of Average Spending by Education Level

Preprocessing:
Rigorous data preprocessing ensured the dataset's reliability and accuracy for subsequent analyses. Techniques included handling missing data, standardizing numerical features, encoding categorical variables, and employing dimensionality reduction methods like Principal Component Analysis (PCA). These steps optimized data quality and prepared it for effective clustering and fitting analyses.

Summary Statistics:

Key statistical insights, such as average income and recency of purchases, provided foundational understanding of customer demographics and behavior. These insights are essential for segmenting customers effectively and tailoring marketing strategies to specific customer segments.

Clustering:

Elbow Method: Based on inertia, the optimal number of clusters (2) was determined, balancing the need for explanatory power and interpretability (Umargono et al., 2020). This step facilitated the segmentation of customers into meaningful groups, uncovering distinct patterns in customer behavior.

KMeans: Achieved a silhouette score of 0.315, indicating reasonable cluster separation and consistency in segmenting customers based on their attributes (Li et al., 2012). This method identified clear customer segments while highlighting some data points that almost formed a separate cluster.

Agglomerative: Yielded a silhouette score of 0.319, demonstrating effective cluster separation and confirming results comparable to KMeans. This hierarchical clustering method provided additional insights into customer segmentation, supporting targeted marketing strategies (Davidson et al., 2005).
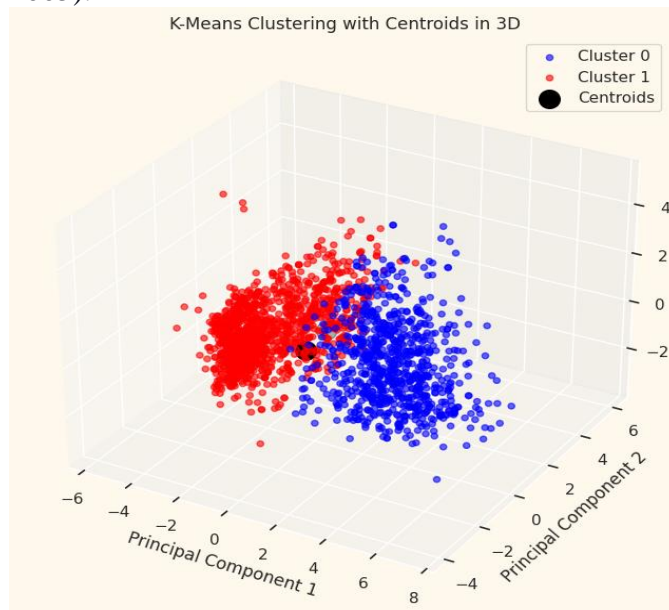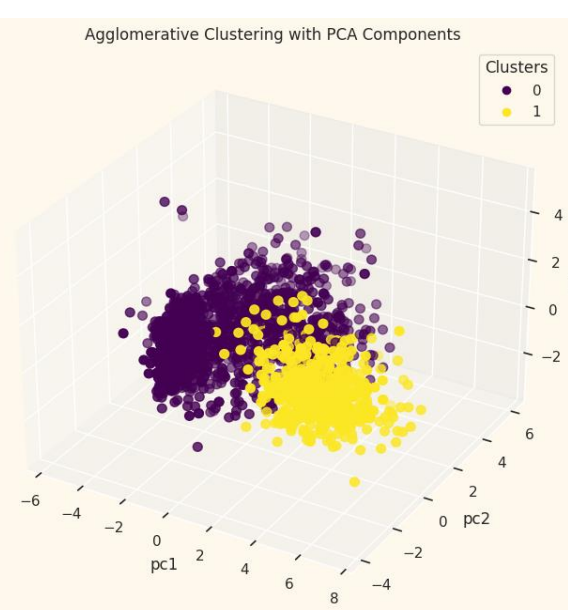

Fig 3: KMeans clustering   Fig 4: Agglomerative Clustering

Fitting: Leveraged Exponential Smoothing for time series forecasting of customer spending patterns. The model effectively captured trends and seasonal variations, enabling businesses to forecast future spending with confidence. This forecasting capability supports proactive marketing strategies and resource allocation, ensuring alignment with customer demand and preferences (Patangia & Soham, 2020).
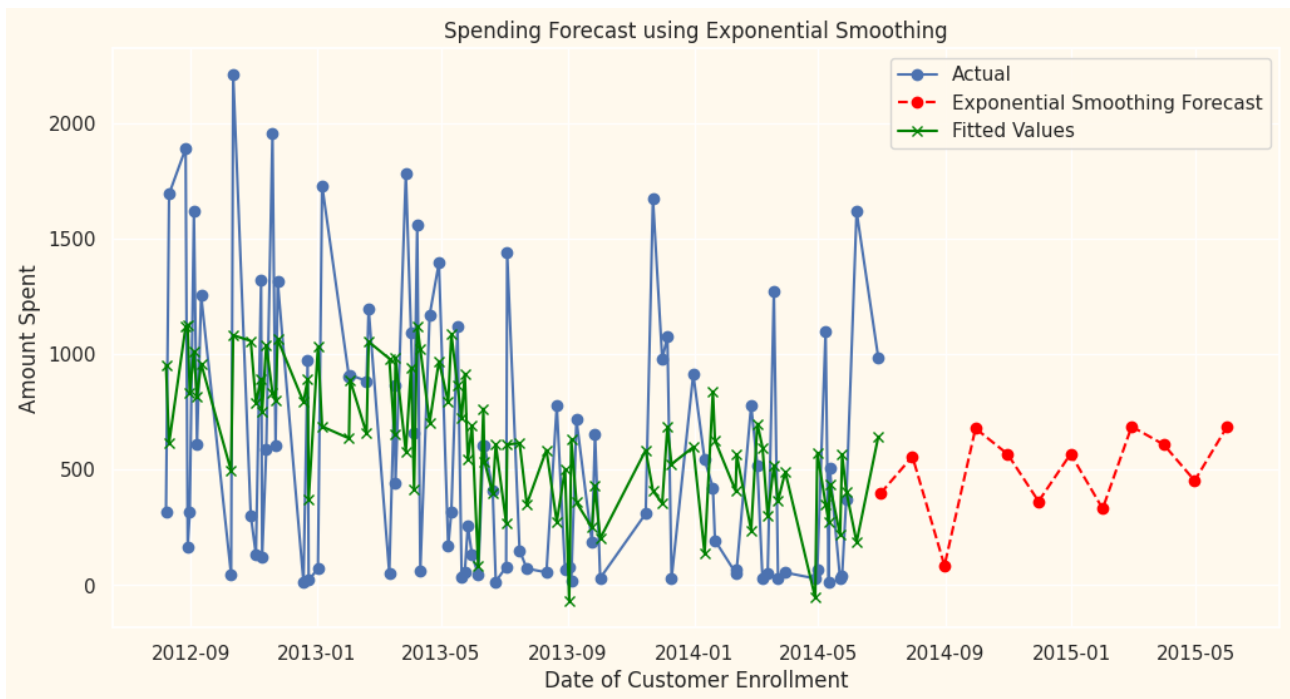
Fig 5: Spending Forecasting

Conclusion: Integrating clustering and fitting techniques in Customer Personality Analysis enhances customer segmentation and predictive analytics capabilities. By delving deeper into customer preferences and behaviors, we can gain a better understanding. businesses can refine their marketing strategies, improve customer engagement, and drive operational efficiency. This data-driven technique is important for staying competitive in dynamic markets and sustaining long-term growth.

GitHub Link :

References

Customer Personality Analysis. (2021, August 22). Kaggle.https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis/data

Kotler, Philip & Keller, K. (2006). Marketing Management. Upper Saddle River, New Yersey.

Li, Youguo & Wu, Haiyan. (2012). A Clustering Method Based on K-Means Algorithm. Physics Procedia. 25, 1104-1109.

Patangia, Soham. (2020). Sales Prediction of Market using Machine Learning. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS090345.

Umargono, Edy & Suseno, Jatmiko & Gunawan, S.K. (2020). K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. https://doi.org/10.2991/assehr.k.201010.019.

Wedel, Michel & Kamakura, Wagner. (2000). Market Segmentation: Conceptual and Methodological Foundations.