FROM DATA TO ACTION: AI-BASED CUSTOMER CLUSTERING FOR EFFECTIVE MARKETING CAMPAIGNS

Master Degree in Artifical Intelligence
Academic Year 2024-2025

Francesco Simbola

Table of Contents

1. Introduction and Context

In today's competitive landscape, companies continuously strive to refine their marketing strategies, seeking solutions that maximize the effectiveness of promotional campaigns and enhance long-term customer loyalty. With the widespread adoption of digital tools and the automation of numerous data collection processes, virtually every transaction, purchase, and customer interaction is recorded and stored. However, an abundance of information does not automatically translate into actionable knowledge, as it requires a structured approach to analysis, interpretation, and complexity reduction.

In this context, segmentation techniques serve as a strategic tool for understanding the diversity of purchasing behaviors and for designing targeted actions for groups of individuals with similar characteristics. The underlying motivation of this project stems from the need to "bring order" to a database containing numerous customer-related variables, including income, spending patterns, purchase frequency, time elapsed since the last order, and, importantly, responsiveness to promotional campaigns, such as the number of accepted promotions or reactions to specific marketing actions.

The objective of this study is to identify intelligent approaches to represent this variety of factors through a limited number of fundamental dimensions, reducing complexity via a dimensionality reduction technique, Principal Component Analysis (PCA), and subsequently clustering the data along these dimensions using the K-Means algorithm. The ultimate goal is to provide a concise and interpretable overview of different customer segments, highlighting the strengths of each group and potential areas for marketing interventions.

The reference context typically pertains to companies operating in the fast-moving consumer goods (FMCG) sector or e-commerce businesses that record orders and transactions. Without segmentation, marketing campaigns tend to adopt a "one-size-fits-all" approach, often failing to leverage the opportunities inherent in customer heterogeneity. By focusing instead on homogeneous groups, it becomes possible to design tailored marketing actions, thereby maximizing results and minimizing resource wastage.

2. State of the Art and Motivations

Since the inception of modern marketing, it has been widely recognized that not all customers are alike, nor can they be addressed with a single, uniform strategy. The formalization of concepts such as segmentation, targeting, and positioning dates back to the 1960s, and these principles continue to serve as the cornerstone of strategic marketing today. Over time, technological advancements and the availability of sophisticated information systems have enabled companies to collect and store an ever-growing volume of data. However, the true challenge does not lie in data collection but rather in the ability to process and synthesize information into a coherent, concise, and, most importantly, actionable form for decision-making.

PCA has established itself as a fundamental tool in cases where numerous potentially correlated variables are available. The dimensionality reduction helps eliminate redundancies and highlights the primary axes of variation within a dataset. In marketing contexts, this proves particularly useful when aiming to extract key insights from large datasets while ensuring greater manageability in data exploration.

For instance, when dealing with metrics such as total spending, income, purchase frequency, recency, and similar variables, PCA can identify two or three principal components that account for approximately 60-70% of the total variance. This not only simplifies subsequent analyses, such as clustering, but also enhances their robustness. Regarding segmentation, the K-Means clustering algorithm remains highly popular due to its ease of implementation and the interpretability of its results. Although alternative methods exist (such as DBSCAN, hierarchical clustering, and mixture models) K-Means is often preferred in marketing applications because it provides a clear partitioning of the data, offers a straightforward way to describe clusters through centroids, and, when applied to a well-structured or reduced set of variables, yields clear and relatively stable solutions. The primary challenges encountered in practice include the need to specify the number of clusters in advance and the algorithm's sensitivity to outliers. To address these issues, evaluation methods such as the Elbow Method (useful for identifying an "elbow" in the inertia reduction curve as k increases) and the Silhouette Score (measuring how well-defined and cohesive clusters are) are commonly employed.
The main motivation driving this study is the interest in providing a marketing department with an effective tool for both analysis and strategic action. For instance, if a group of high-income customers who respond enthusiastically to promotional campaigns is identified, it may be reasonable to offer them preferential treatment, exclusive promotions, or premium offers aimed at increasing their average spending. Conversely, if a cluster of low-income, infrequent buyers is detected, the company may decide whether to invest in targeted re-engagement strategies or to focus its efforts on more promising customer segments.

3. Dataset e Preprocessing

The reference dataset contains detailed information on a set of customers, including data on purchasing behavior, responses to marketing campaigns, and socio-demographic characteristics. The only variable with missing values is Annual Income (Income), for which null values have been removed from the dataset to prevent distortions in the analysis.

The variable Dt_Customer, which represents the customer registration date, was converted into a datetime format and used to determine Customer_Age, given that the dataset contains data collected up to 2014. Subsequently, we reviewed the values of Age, removing entries over 100 years old as they were not relevant for segmentation purposes.

```
[7] # Convert Dt_Customer column to datetime format
    df['Dt_Customer'] = pd.to_datetime(df['Dt_Customer'], format='%d-%m-%Y')

    # Find the most recent registration date in the dataset
    latest_date = df['Dt_Customer'].max()
    print("Most recent registration date in the dataset:", latest_date)

    Most recent registration date in the dataset: 2014-06-29 00:00:00

[8] # Calcolare gli anni da quando il cliente è stato registrato (rispetto al 2014)
    df['Customer_Years'] = 2014 - df['Dt_Customer'].dt.year

    # Eliminare la colonna originale
    df.drop(columns=['Dt_Customer'], inplace=True)

    # Verifica
    print("Distribuzione di Customer_Years:")
    print(df['Customer_Years'].describe())

    Mostra output nascosti

    # Create the Age column based on the year 2014
    df['Age'] = 2014 - df['Year_Birth']

    # Remove the original Year_Birth column
    df.drop(columns=['Year_Birth'], inplace=True)
    print(df[['Age']].describe())

    Mostra output nascosti

[10] # Remove customers with Age > 100 years
     df = df[df['Age'] <= 100]

     # Verify that there are no more outliers
     print("Customers with Age > 100 after cleaning:", df[df['Age'] > 100].shape[0])

     Customers with Age > 100 after cleaning: 0
```

The categorical variables Education and Marital_Status were transformed using one-hot encoding to ensure they could be utilized in the clustering models without introducing arbitrary ordering among categories

```
[ ] # Create a binary column for Education
    df['Education_Basic'] = df['Education'].apply(lambda x: 1 if x == 'Basic' else 0)

    # Remove the original Education column
    df.drop(columns=['Education'], inplace=True)
    print(df['Education_Basic'].value_counts())
```

Mostra output nascosti

```
# Create a binary column for Marital Status
df['Marital_Single'] = df['Marital_Status'].apply(lambda x: 1 if x in ['Single', 'Alone', 'Widow'] else 0)

# Remove the original Marital_Status column
df.drop(columns=['Marital_Status'], inplace=True)
print(df['Marital_Single'].value_counts())
```

Mostra output nascosti

Four aggregate variables were created to enhance the quality of the analysis, including Total_Spending (total expenditure), Total_Purchases (total number of purchases), Accepted_Campaigns (number of accepted campaigns), and Total_Children (number of children, obtained by summing Kidhome and Teenhome). These additions provide a more comprehensive view of customer behavior

```
[ ] # Create a new column with the total number of children
    df['Total_Children'] = df['Kidhome'] + df['Teenhome']

    # Remove the original columns
    df.drop(columns=['Kidhome', 'Teenhome'], inplace=True)
    print(df['Total_Children'].value_counts())
```

Mostra output nascosti

```
[ ] # Create a new column with the total spending
    df['Total_Spending'] = df[['MntWines', 'MntFruits', 'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts', 'MntGoldProds']].sum(axis=1)
    df.drop(columns=['MntWines', 'MntFruits', 'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts', 'MntGoldProds'], inplace=True)

[ ] # Create a new column with the total purchase
    df['Total_Purchases'] = df[['NumWebPurchases', 'NumCatalogPurchases', 'NumStorePurchases']].sum(axis=1)
    df.drop(columns=['NumWebPurchases', 'NumCatalogPurchases', 'NumStorePurchases'], inplace=True)

[ ] # Create a new column for accepted campaigns
    df['Accepted_Campaigns'] = df[['AcceptedCmp1', 'AcceptedCmp2', 'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5']].sum(axis=1)
    df.drop(columns=['AcceptedCmp1', 'AcceptedCmp2', 'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5'], inplace=True)
    print(df[['Accepted_Campaigns']].head(20))
```

Mostra output nascosti

To finalize the optimization process, certain columns that did not add value to the clustering analysis were removed. Specifically, the ID column was eliminated as it represents a unique identifier that does not contribute to the analysis. Additionally, the columns Z_CostContact and Z_Revenue were removed because they contained constant values, providing no useful information. Lastly, the Complain variable was excluded, as the number of complaints made by customers was not relevant to the clustering objective and did not provide meaningful insights into purchasing behavior or responses to marketing campaigns.

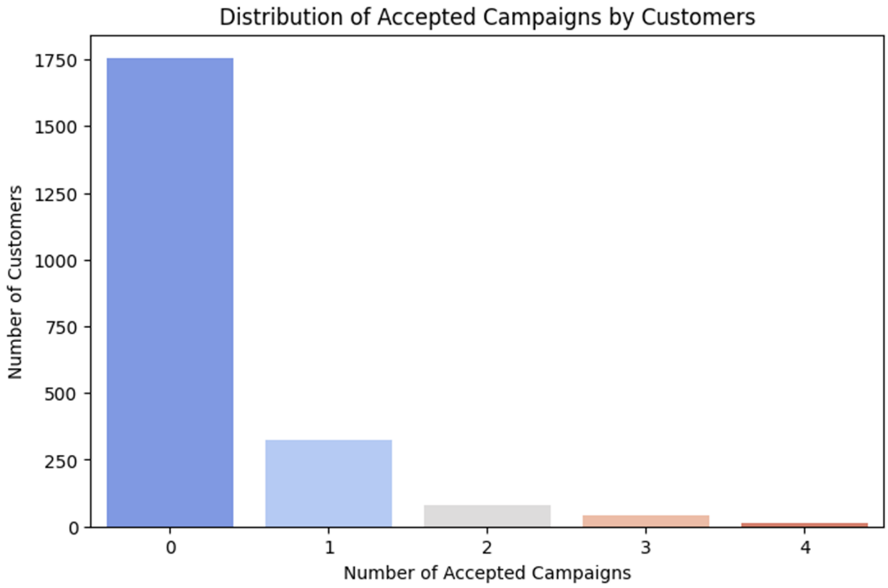
```
[18] # Remove ID, Z_CostContact, Z_Revenue columns
     df.drop(columns=['ID', 'Z_CostContact', 'Z_Revenue', 'Complain'], inplace=True)

[19] print(df.columns)

     Index(['Income', 'Recency', 'NumDealsPurchases', 'NumWebVisitsMonth',
            'Response', 'Customer_Years', 'Age', 'Education_Basic',
            'Marital_Single', 'Total_Children', 'Total_Spending', 'Total_Purchases',
            'Accepted_Campaigns'],
           dtype='object')
```

Following the preprocessing phase described above, the dataset was properly structured for the implementation of clustering techniques, ensuring greater reliability and interpretability of the results obtained.

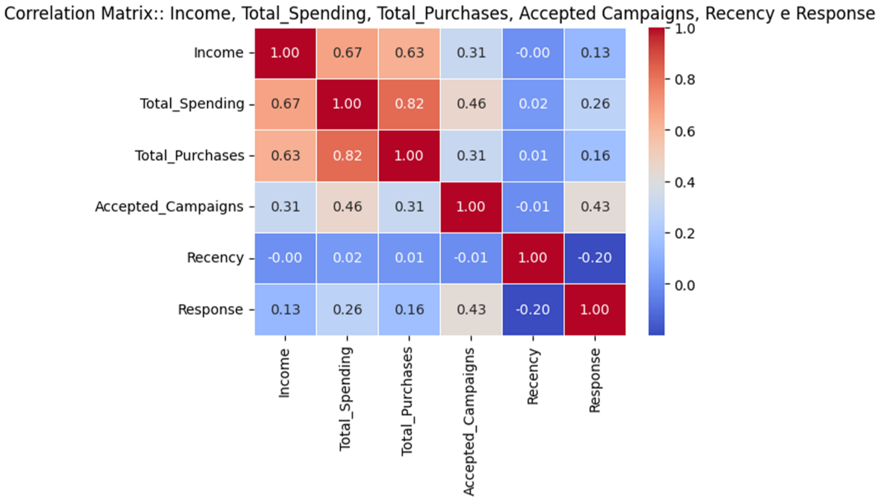
4. Preliminary Analysis

The graph clearly illustrates a highly imbalanced distribution in the number of marketing campaigns accepted by customers. The vast majority of users have never engaged with a promotional campaign, as evidenced by the predominant height of the first bar. This finding suggests a generally low inclination to interact with marketing offers, a factor that may be influenced by several elements, such as a lack of personalization in promotions, a disengaged target audience, or the limited effectiveness of the campaigns themselves.

As the number of accepted campaigns increases, the number of customers decreases drastically. Few customers have accepted even a single campaign, and even fewer have engaged with two, three, or four campaigns. This trend indicates that only a small niche of customers actively responds to marketing initiatives, while the majority disregards them entirely. The fact that so few customers have accepted multiple campaigns may suggest that the promotions do not foster long-term customer loyalty but rather sporadically engage a limited subset of individuals.



The correlation matrix was computed to clarify the relationships between variables related to marketing campaigns and customer purchasing behaviors, particularly in light of the initial analysis that revealed an imbalance in promotion acceptance. This observation alone was insufficient to fully understand the actual impact of the campaigns on purchasing patterns. For this reason, the correlation analysis provided additional insights into the extent to which accepted campaigns, spending, purchase frequency, and overall engagement were interconnected.



The resulting analysis highlights several key findings, one of the most relevant being the positive correlation (0.43) between Accepted_Campaigns and Response. This suggests that customers who have previously engaged with marketing campaigns are more likely to do so in the future, indicating a recurring behavioral pattern: those who have shown interest in a promotion may exhibit the same tendency when exposed to subsequent initiatives. However, the correlation is not strong enough to imply a deterministic relationship, leaving room for numerous exceptions in which customers who previously accepted campaigns later choose not to respond to new offers.

Another particularly noteworthy correlation (0.31) is observed between Accepted_Campaigns and Total_Purchases. The strength of this relationship is lower than expected if one were to assume a strong link between promotional engagement and increased purchase frequency. This finding becomes more comprehensible when considering that, on average, customers have been active for a limited period (Customer_Years around just one year). As a result, those who accept marketing campaigns do not necessarily display a higher purchasing frequency, possibly due to the short observation window or other factors unrelated to a single promotional initiative.

An additional significant aspect is the negative correlation (-0.20) between Recency and Response, indicating that customers who have made recent purchases are more likely to engage with new marketing campaigns, while those whose last purchase occurred a long time ago tend to be less responsive. This underscores the importance of maintaining frequent customer interactions and implementing promotional initiatives within a relatively short time frame after the most recent purchase to optimize response rates. However, the moderate correlation suggests that other factors may also influence a customer's decision to engage with a campaign. The relationship between Income and Accepted_Campaigns (0.31), while of some interest, is not the primary focus of this study, which instead centers on the connection between marketing efforts and purchasing behavior. Nonetheless, it is worth noting that income exhibits a moderate correlation with total spending and purchase frequency, indicating that higher economic availability does not automatically translate into a greater propensity to accept promotional campaigns.

Overall, the correlation matrix analysis supports some of the hypotheses proposed earlier and provides further insights into the relationship between promotional initiatives and spending behavior. In particular, the results suggest that campaign acceptance is not a direct predictor of increased purchases but rather a signal of higher engagement with the company's marketing efforts.

Customers who have accepted promotions in the past tend to respond positively in the future, although this does not necessarily translate into an increase in the number of purchases. Additionally, the negative impact of Recency on promotional response highlights the need to maintain consistent customer engagement to prevent a decline in their level of interaction. These findings suggest the need to reconsider the marketing strategy, favoring a segmented and personalized approach. Rather than relying solely on generalized campaigns, it would be more effective to target customers at moments when the likelihood of purchase is higher, for example, shortly after a recent transaction, thereby maximizing the impact of promotional initiatives.

5. Standard Scale & Dimensionality Reduction

Data standardization is a crucial step in preprocessing, particularly for models that are sensitive to the scale of variables, such as distance-based algorithms. The implemented code utilizes StandardScaler to transform several numerical variables (income, total spending, number of purchases, number of accepted campaigns, and recency) ensuring that they have a mean of zero and a standard deviation of one. This process guarantees that each feature contributes equally to the analysis, preventing variables with larger scales from dominating the results. To preserve the original data, the dataset is copied before the transformation, ensuring that any subsequent modifications do not affect the non-standardized version. An analysis of the standardization results confirms the effectiveness of the transformation: the standardized variables exhibit means close to zero and standard deviations around one, indicating that the rescaling has been correctly applied. While some values may exceed two or three standard deviations in the case of outliers, the process still maintains the relative structure of the data. The Recency variable, which represents the time elapsed since the last purchase, may display a less symmetric distribution compared to other features; however, standardization ensures that it is treated consistently alongside other metrics.

The PCA analysis identified two principal components, PC1 and PC2, which together explain 73.02% of the total variance in the dataset. Specifically, PC1 accounts for 52.99%, while PC2 contributes an additional 20.03%, allowing for a significant reduction in dimensionality while retaining a substantial portion of the original information.
The analysis of loadings (coefficients representing the contribution of the original variables to the formation of the principal components) provides insight into the meaning of the extracted components. PC1 exhibits high contributions from Total_Spending (0.572), Total_Purchases (0.542), and Income (0.503), indicating that this component synthesizes factors related to spending capacity and purchasing propensity. Accepted_Campaigns (0.353) also contributes, albeit to a lesser extent, suggesting that this component captures an axis that differentiates customers based on their level of economic engagement with the company. Conversely, Recency (0.0059) has a negligible contribution to this component, confirming that recent purchase frequency is not a key factor in determining the first direction of variability. In contrast, PC2 is almost entirely dominated by Recency (0.997), while all other variables exhibit coefficients close to zero. This finding indicates that the second principal component essentially represents a measure of purchase recency, distinguishing between customers who have made recent transactions and those who have not purchased for an extended period.

```python
from sklearn.preprocessing import StandardScaler

# Define the normalizer
scaler = StandardScaler()

# Copy the original dataset to avoid permanent modifications
df_scaled = df.copy()

# Select only continuous numerical features for standardization (excluding Response)
numerical_features = ['Income', 'Total_Spending', 'Total_Purchases', 'Accepted_Campaigns', 'Recency']

# Apply standardization
df_scaled[numerical_features] = scaler.fit_transform(df[numerical_features])

# Verify the newly standardized values
print(df_scaled[numerical_features].describe())
```

```
             Income  Total_Spending  Total_Purchases  Accepted_Campaigns  \
count  2.213000e+03    2.213000e+03     2.213000e+03        2.213000e+03
mean  -5.056958e-17   -4.013459e-17     9.391493e-17       -4.013459e-17
std    1.000226e+00    1.000226e+00     1.000226e+00        1.000226e+00
min   -2.006386e+00   -9.994508e-01    -1.744293e+00       -4.390416e-01
25%   -6.749549e-01   -8.932007e-01    -9.113234e-01       -4.390416e-01
50%   -3.430598e-02   -3.486691e-01    -7.835361e-02       -4.390416e-01
75%    6.455518e-01    7.320934e-01     7.546162e-01       -4.390416e-01
max    2.440836e+01    3.184146e+00     2.698212e+00        5.449437e+00

            Recency
count  2.213000e+03
mean  -4.254266e-17
std    1.000226e+00
min   -1.693697e+00
25%   -8.642613e-01
50%   -2.654845e-04
75%    8.637304e-01
max    1.727726e+00
```

```python
from sklearn.decomposition import PCA
import numpy as np
# Instantiate PCA with 2 principal components
pca = PCA(n_components=2, random_state=42)

# Apply PCA only on standardized numerical features
df_pca = pca.fit_transform(df_scaled[numerical_features])

# Create a DataFrame with principal components
df_pca_final = pd.DataFrame(df_pca, columns=['PC1', 'PC2'])

expl_var_ratio = pca.explained_variance_ratio_
print("Explained variance by PC1 and PC2:", expl_var_ratio)
print("Cumulative explained variance:", np.cumsum(expl_var_ratio))


# Loadings (PCA coefficients)
loadings = pd.DataFrame(pca.components_, columns=numerical_features, index=['PC1', 'PC2'])

print("PCA Loadings (Contribution of each feature to PC1 and PC2):")
print(loadings)
```

```
Explained variance by PC1 and PC2: [0.52993088 0.20030156]
Cumulative explained variance: [0.52993088 0.73023244]
PCA Loadings (Contribution of each feature to PC1 and PC2):
       Income  Total_Spending  Total_Purchases  Accepted_Campaigns   Recency
PC1  0.503427        0.572292         0.542465            0.353185  0.005991
PC2 -0.004980        0.020391         0.017512           -0.069755  0.997190
```

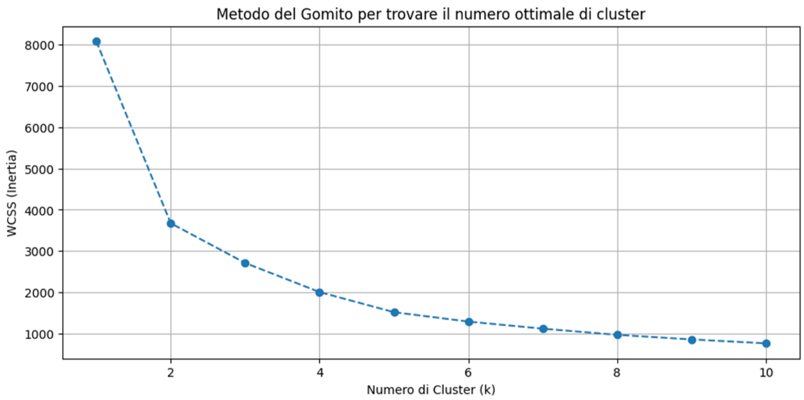
The fact that spending-related variables (Total_Spending and Total_Purchases) have near-zero weights in PC2 suggests that temporal purchasing behavior (Recency) operates independently of spending capacity and purchase volume. This confirms that these two factors function on distinct axes of variability, reinforcing the importance of considering both financial engagement and temporal activity when analyzing customer segmentation.

6. Clustering with the K-Means Algorithm

The K-Means algorithm was applied to the projected coordinates in (PC1, PC2) to identify groups of customers with similar characteristics. K-Means is an iterative method that, given a predefined number of clusters, progressively updates the centroids until a stable configuration is reached, where each data point is assigned to the nearest centroid based on Euclidean distance. A critical aspect of the clustering process is determining the optimal number of clusters, k, which must be chosen to balance internal cohesion within groups and separation between them. To achieve this, two evaluation criteria were employed:

- Elbow Method, which analyzes the variation in inertia (Within-Cluster Sum of Squares) as k increases, identifying the point beyond which additional clusters no longer yield significant improvements in intra-cluster variance reduction.
- Silhouette Score, a metric that assesses the quality of segmentation by simultaneously considering the internal density of clusters and their separability.

By leveraging these evaluation techniques, the optimal k value was determined, ensuring that the resulting segmentation effectively captures meaningful customer distinctions while avoiding excessive fragmentation or under-clustering.



The analysis revealed that the Silhouette Score for k = 3 is 0.4123, while for k = 4, it slightly decreases to 0.3996. Although both values indicate an acceptable segmentation, the higher score obtained with three clusters suggests that this configuration ensures better separation between groups while avoiding excessive data fragmentation. Furthermore, a limited number of clusters enhances the interpretability of the results and their applicability in decision-making contexts.

```
from sklearn.metrics import silhouette_score

# Test k=3 and k=4 after standardization and PCA
for k in [3, 4]:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=20)
    labels = kmeans.fit_predict(df_pca_final[['PC1', 'PC2']])
    silhouette_avg = silhouette_score(df_pca_final[['PC1', 'PC2']], labels)
    print(f"Silhouette Score per k={k}: {silhouette_avg:.4f}")

Silhouette Score per k=3: 0.4124
Silhouette Score per k=4: 0.3996
```

nce k = 3 was established, the algorithm was executed, assigning each customer to one of the three identified clusters. This classification enabled a clear segmentation of the customer base, distinguishing groups that are internally homogeneous but significantly different from each other. The outcome of this phase resulted in an enriched dataset with a new column, where each row is labeled with a numerical identifier representing the assigned cluster. This addition allowed for a direct connection between the clustering structure and the original dataset variables, facilitating a deeper interpretation of the identified groups.

7.Interpretation of Segments

Following the clustering process, a strategic interpretation of the segments was conducted to understand the operational implications for business decision-making. The primary goal of segmentation is not merely to identify homogeneous customer groups but also to translate these insights into actionable strategies aimed at improving customer portfolio management and resource allocation. To provide a more detailed description of each cluster's characteristics, the original dataset was referenced, analyzing the average values of key variables within each group. This analysis allowed for the identification of three distinct customer profiles, each characterized by specific needs and behaviors.

Cluster 0: High-Spending Customers with Strong Engagement

The analysis reveals that Cluster 0 is characterized by high values along the first principal component (PC1), which aggregates indicators such as Total_Spending, Total_Purchases, and Income. These results suggest that customers in this segment not only exhibit significant spending capacity but also demonstrate a high purchase frequency and, on average, have a higher income. From a strategic perspective, this segment represents the most economically valuable target for the company, offering strong potential for customer retention initiatives and long-term value generation. The distribution along the second principal component (PC2) indicates some temporal variability in purchasing behavior: while some customers have made recent purchases, the segment as a whole remains economically active. Although Recency is not the primary discriminating factor, maintaining a high level of engagement with the brand is crucial. In this regard, the implementation of VIP programs, exclusive offers, and personalized incentives can help strengthen customer loyalty and prevent potential declines in purchase frequency.

Cluster 1: At-Risk Customers with Low Spending

Cluster 1 is characterized by negative values on both PC1 and PC2, indicating a segment of customers who, despite having a low overall spending level, have made recent purchases, as evidenced by a low Recency score. This configuration suggests that, although the economic profile of this segment is less robust than that of Cluster 0, these customers remain active and may still be responsive to targeted marketing efforts. From a growth perspective, strategies such as personalized offers, promotions aimed at encouraging repeat purchases, and cross-selling initiatives could help increase purchase frequency. If effectively implemented, these strategies could lead a portion of this segment to evolve over time, adopting characteristics more similar to those of Cluster 0. This transition would have positive implications for the company's profitability in the medium to long term..

Cluster 2: Customers with Growth Potential and High Recency

Cluster 2 is characterized by negative values on PC1 combined with positive values on PC2, defining a segment with limited spending capacity where Recency plays a predominant role. This indicates that customers in this group have not made purchases for a considerable period, placing them at high risk of churn. The strong influence of Recency in the construction of PC2 highlights their low level of activity and reduced engagement with the brand.

From a marketing perspective, the response rate to promotional campaigns in this segment is generally low. While in some cases, re-engagement strategies and targeted discounts may incentivize a return to activity, the cost-benefit ratio must be carefully evaluated. If the cost of reactivating these customers exceeds their potential economic value, a more strategic investment approach may involve reducing resources allocated to this segment. An alternative strategy would be to proactively identify customers in the early stages of disengagement, before their level of inactivity reaches a critical threshold. This would allow for timely intervention with retention initiatives, maximizing the chances of maintaining long-term customer relationships.

8. Conclusions and Future Developments

The study conducted highlights the effectiveness of an integrated approach that combines dimensionality reduction with K-Means clustering for customer segmentation. This strategy enables efficient management of a set of relevant variables while providing a more interpretable representation of the data. The identification of three clusters proves to be sufficiently flexible for distinguishing different customer behaviors in terms of spending capacity and responsiveness to offers, without leading to excessive fragmentation that could compromise interpretability. For future developments, additional variables could be incorporated, such as demographic attributes or temporal trends (e.g., monthly spending evolution), to further explore the dynamic nature of the customer-business relationship. Additionally, adopting alternative clustering algorithms, such as DBSCAN or hierarchical models, could facilitate the identification of latent subgroups or anomalous patterns within the analyzed population. From an applied perspective, implementing predictive classification techniques that leverage cluster membership as an explanatory variable could offer further opportunities to estimate the likelihood of customer responses to specific promotional offers.

In conclusion, the presented case study confirms that the combination of these two techniques, along with a thorough interpretation of key variables, represents a solid and flexible methodology for capturing the main dimensions of variability among customers and translating them into more effective and personalized marketing initiatives.