Customer Churn Clustering Analysis Report

1- Executive Summary

This report presents the results of a clustering analysis conducted on our customer dataset to identify distinct segments based on their behavior and attributes. The primary aim was to understand customer churn within these segments, enabling targeted strategies for improving retention. The analysis identified three unique customer clusters, each exhibiting distinct characteristics and churn rates. Insights derived from these findings will guide our marketing and customer service strategies to reduce churn and enhance customer engagement.

2- Introduction

Background: Customer churn is a major concern as it affects revenue and long-term viability. This analysis segments customers into distinct groups based on their behaviors and attributes to identify which groups are more prone to churn.

Objective: The main goal is to use clustering techniques to uncover hidden patterns in customer behaviors, which can help develop more effective customer retention strategies

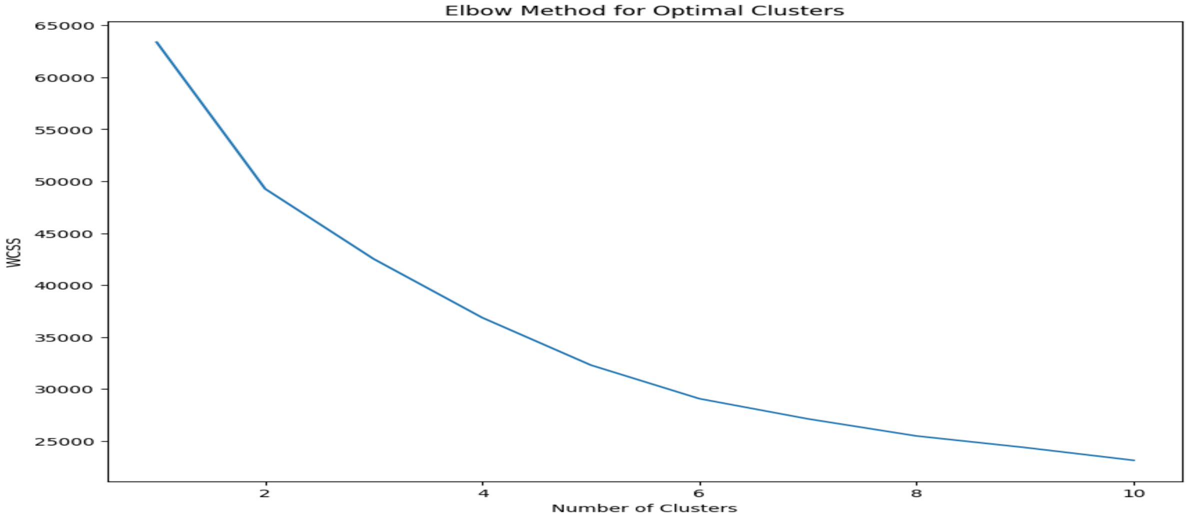
3-Determination of Optimal Clusters

Elbow Method: The elbow method was used to determine the optimal number of clusters by plotting the within-cluster sum of squares (WCSS) against the number of clusters. This method identifies the point at which the WCSS begins to decrease at a diminishing rate, indicating that adding more clusters does not significantly improve the model.

Elbow Plot Analysis

Visualization: The elbow plot clearly shows a bend at three clusters, suggesting that increasing the number of clusters beyond this point yields minimal gain in the variance explained by the clusters.

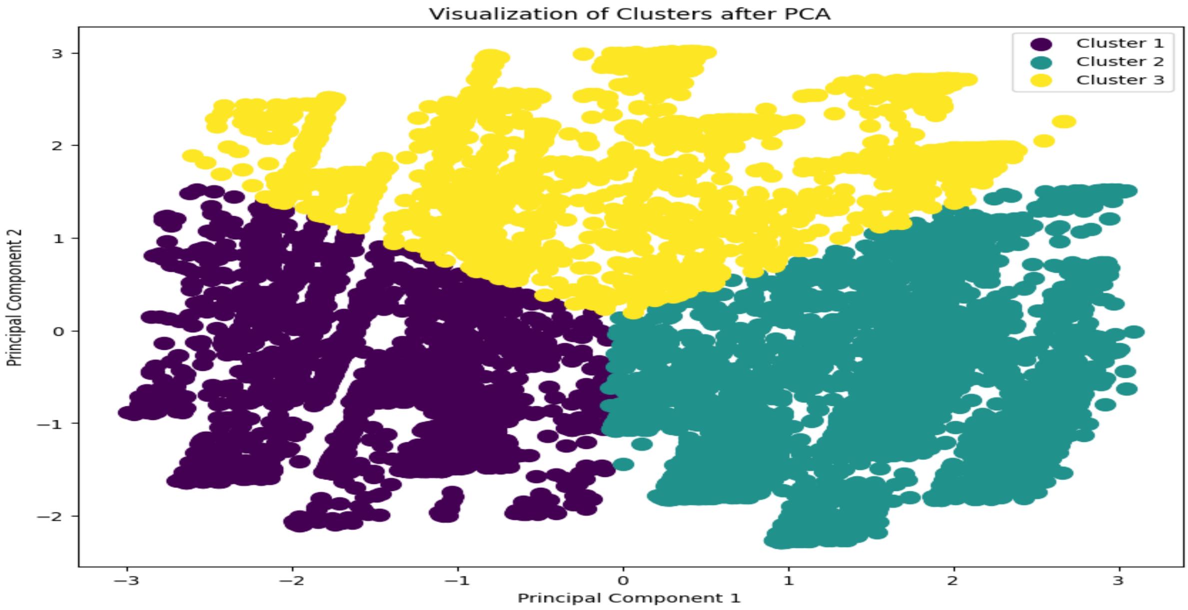
Interpretation: Based on the elbow plot, we selected three clusters as the optimal number. This choice balances complexity and clarity, providing a manageable number of segments that are distinct enough to be actionable.




```python
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import numpy as np
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
```

4- Clustering Algorithm

1- K-Means Clustering: The number of clusters was determined using the elbow method, which suggested three clusters as optimal to balance between within-cluster variance and the number of clusters.



```python
from sklearn.decomposition import PCA
from sklearn.metrics import silhouette_score

# PCA for dimensionality reduction
pca = PCA(n_components=2)
principal_components = pca.fit_transform(X)

optimal_clusters = 3

# K-Means clustering
kmeans = KMeans(n_clusters=optimal_clusters, init='k-means++', n_init=10, max_iter=300, random_state=42)
clusters = kmeans.fit_predict(principal_components)


silhouette_avg = silhouette_score(principal_components, clusters)

print(f"Silhouette Score: {silhouette_avg}")
```

4,1 Analysis Results

1- Cluster Characterization

Cluster 1 (High-Value Customers): Characterized by high monthly expenditures and long tenure. Predominantly subscribed to multiple high-value services.

Cluster 2 (Budget-Conscious New Customers): This cluster includes newer customers with lower monthly charges, primarily using basic service packages.

Cluster 3 (At-Risk Customers): Customers with medium tenure showing signs of churn, characterized by complaints and lower satisfaction scores.

2- Performance Evaluation

Silhouette Score: 0.44, indicating moderate separation between clusters.

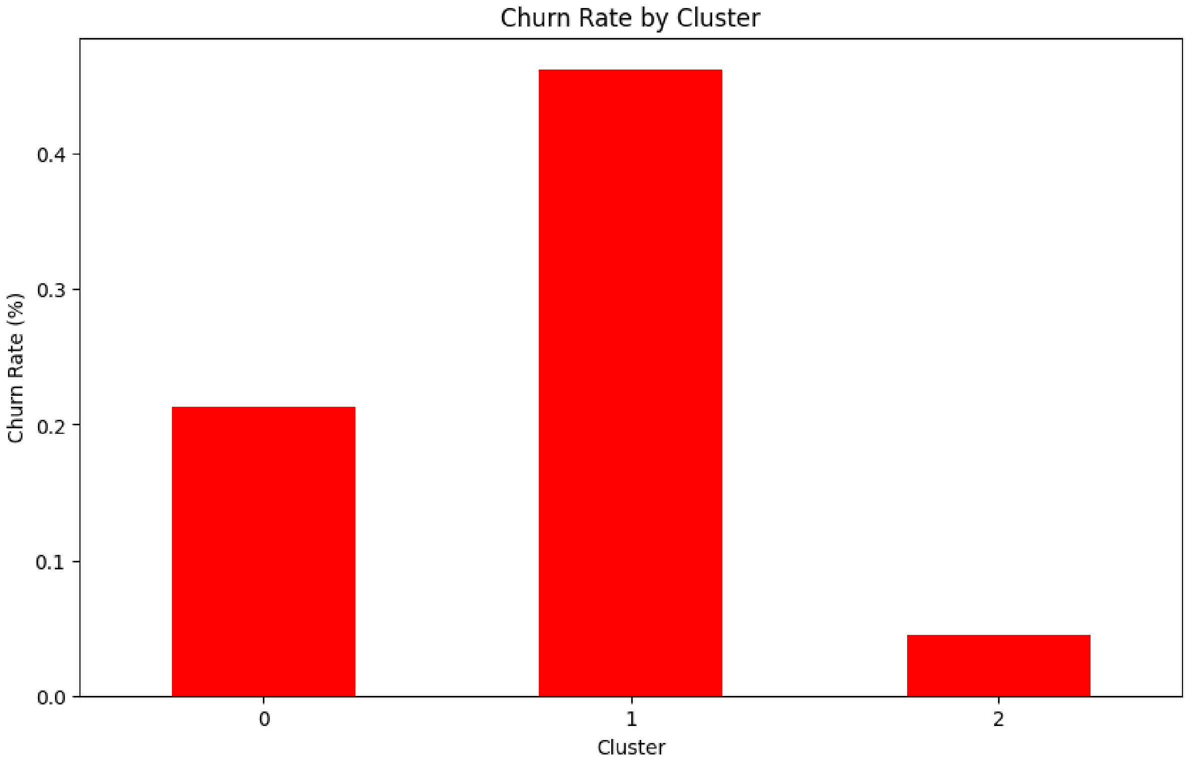
3- churn rate by cluster

Cluster

0 -> 0.212936

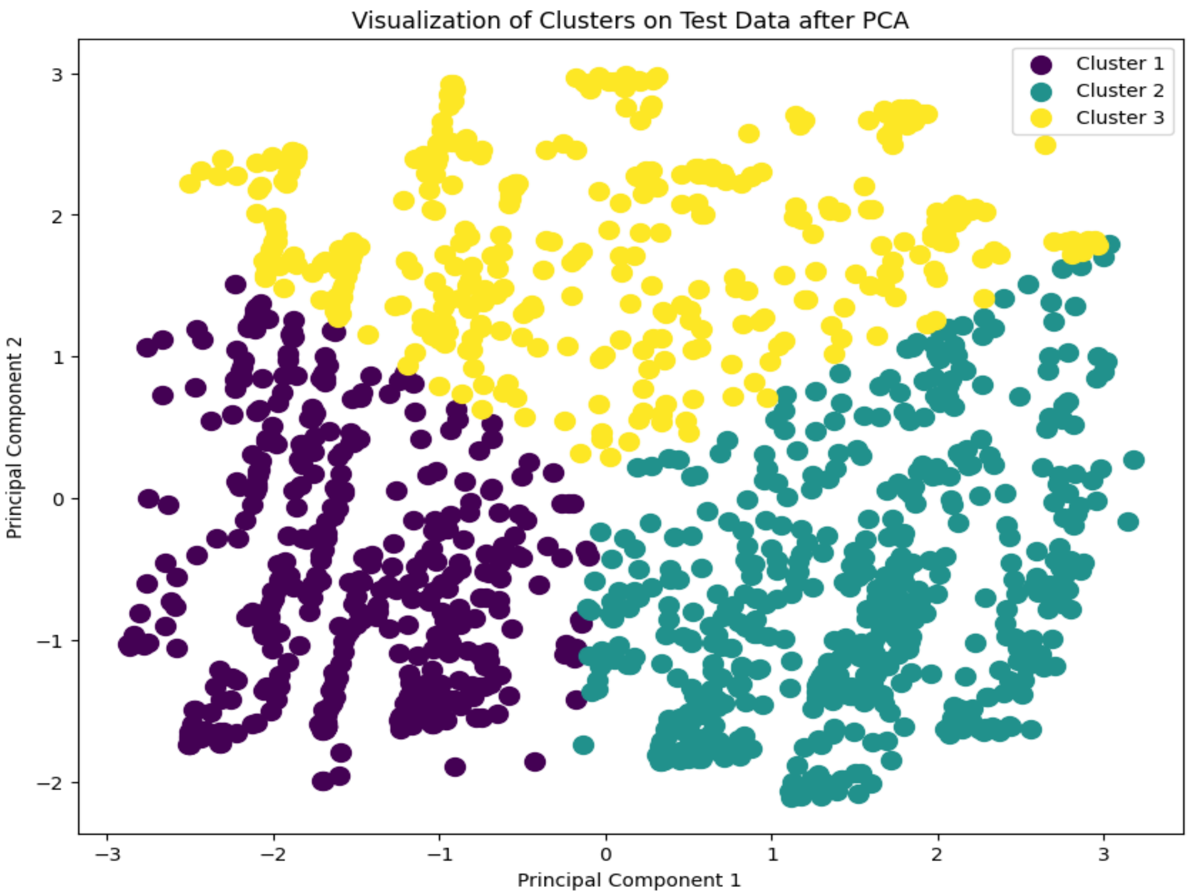
1 -> 0.462486

2 ->0.045192



5- Model on test data:

Silhouette Score: 0.43, indicating moderate separation between clusters.



6- Conclusion

The cluster analysis has successfully segmented the customer base into meaningful groups that can inform targeted marketing and service strategies. Continued refinement of clustering parameters and inclusion of additional behavioral data can further enhance the granularity and utility of customer segmentation.