Contents

[Objective 1](#_Toc194190390)

[Clustering Analysis 1](#_Toc194190391)

[Elbow Method 1](#_Toc194190392)

[Training K-Means Clustering Model 2](#_Toc194190393)

[Principal Component Analysis (PCA) for Dimensionality Reduction 2](#_Toc194190394)

[Visualizing Clusters 2](#_Toc194190395)

[Labelling Clusters 3](#_Toc194190396)

[Cluster 0 - Flexible Contract, Low-Dependency Users 3](#_Toc194190397)

[Cluster 1 - Mid-Tier Users with 1-Year Plans 3](#_Toc194190398)

[Cluster 2 - Family-Oriented, High-Internet Usage 4](#_Toc194190399)

[Cluster 3 - Loyal Long-Term Contract Customers 4](#_Toc194190400)

[Observations 5](#_Toc194190401)

[Conclusion 5](#_Toc194190402)

**ACS WIL Data Analytics Project**

**Customer Churn Analysis Report**

--

# Objective

The main objective of this project is to apply K-Means clustering to a scaled customer churn dataset to identify distinct customer segments. The analysis aims to determine the optimal number of clusters using the Elbow Method, train the clustering model, visualize high-dimensional data using Principal Component Analysis (PCA), label the clusters, and derive actionable business recommendations.

# Clustering Analysis

Clustering is a powerful unsupervised machine learning technique that groups customers based on feature similarity. Our goal is to use clustering to identify key customer personas that can be used for churn reduction strategies and targeted campaigns.

## Elbow Method

We applied the Elbow Method to determine the optimal number of clusters. Inertia (within-cluster sum of squares) was plotted for k = 1 to 10. The elbow in the curve appears at k = 4, indicating this as the optimal cluster count.

The following is the elbow plot visualizing the inertia for different cluster numbers:

A graph with a line

AI-generated content may be incorrect.

# Training K-Means Clustering Model

K-Means was trained on the scaled dataset (X\_train) with 4 clusters. Each customer was assigned to one cluster based on feature similarity. This allowed us to segment customers without relying on predefined labels, using patterns inherent in the data.

## Principal Component Analysis (PCA) for Dimensionality Reduction

To overcome visual clutter from high-dimensional data, PCA was used to reduce the dataset to 2 dimensions while retaining most of the variance. This made it possible to clearly plot and visually distinguish the four clusters in a 2D scatter plot.

The following is our scatter plot diagram of PCA-reduced data:

A chart of colored dots

AI-generated content may be incorrect.

# Visualizing Clusters

After PCA, a scatter plot showed clearly separated clusters, confirming that the dataset was successfully segmented into meaningful groups using the K-Means model. Each point was colored based on cluster assignment.

# Labelling Clusters

Average values of features within each cluster were computed to interpret customer characteristics. Below is a summary of each cluster:

A chart of customer persona distribution

AI-generated content may be incorrect.

## Cluster 0 - Flexible Contract, Low-Dependency Users

* Senior Citizen (0.199): 19.9% senior citizens
* Tenure (0.420): Medium tenure (42% of maximum)
* Monthly Charges (0.477): Mid-range charges
* Gender\_Male (0.704): 70.4% male customers
* Dependents\_Yes (0.101): 10.1% have dependents
* Phone Service (0.905): 90.5% have phone service
* Multiple Lines (0.285): 28.5% have multiple lines
* Internet Service - Fiber Optic (0.338): 33.8% use fiber optic
* Contract Type: 100% on month-to-month contracts
* **Recommendations:**
  + Offer long-term contract discounts
  + Provide loyalty-based upgrades

## Cluster 1 - Mid-Tier Users with 1-Year Plans

* Senior Citizen (0.179): 17.9% senior citizens
* Tenure (0.452): Moderate tenure
* Monthly Charges (0.447): Average charges
* Gender\_Male (0.498): Balanced gender
* Dependents\_Yes (0.305): 30.5% have dependents
* Phone Service (0.894): 89.4% have phone service
* Multiple Lines (0.416): 41.6% have multiple lines
* Internet Service - Fiber Optic (0.365): 36.5% use fiber optic
* Contract Type: 100% on one-year contracts
* **Recommendations:**
  + Offer 2-year plan upgrades
  + Run satisfaction surveys
  + Encourage bundle adoption

## Cluster 2 - Family-Oriented, High-Internet Usage

* Senior Citizen (0.105): 10.5% senior citizens
* Tenure (0.490): Medium-high tenure
* Monthly Charges (0.444): Average charges
* Gender\_Male (0.226): Mostly female
* Dependents\_Yes (0.571): 57.1% have dependents
* Phone Service (0.907): 90.7% have phone service
* Multiple Lines (0.504): 50.4% have multiple lines
* Internet Service - Fiber Optic (0.832): 83.2% use fiber optic
* Contract Type: 100% on month-to-month contracts
* **Recommendations:**
  + Upsell premium data/internet plans
  + Offer family-oriented bundles
  + Encourage contract commitment

## Cluster 3 - Loyal Long-Term Contract Customers

* Senior Citizen (0.160): 16% senior citizens
* Tenure (0.448): Medium tenure
* Monthly Charges (0.467): Medium-high charges
* Gender\_Male (0.526): Slightly more male
* Dependents\_Yes (0.297): 29.7% have dependents
* Phone Service (0.898): 89.8% have phone service
* Multiple Lines (0.516): 51.6% have multiple lines
* Internet Service - Fiber Optic (0.254): 25.4% use fiber optic
* Contract Type: 100% on two-year contracts
* **Recommendations:**
  + Maintain loyalty through VIP perks
  + Introduce referral programs
  + Provide personalized exclusive offers

# Observations

* **Cluster 0:** This group includes younger, flexible customers on month-to-month contracts, with low tenure and minimal service usage such as no internet and limited dependents. Their commitment to the service is low, making them a high-risk segment for churn. Strategic interventions like loyalty programs or contract upgrade incentives may help retain them.
* **Cluster 1:** These are mid-tier customers with moderate tenure and charges, mostly on one-year contracts. They show a balanced gender mix and a moderate level of dependents and services used. While relatively stable, they still have room for improvement in service engagement and contract length. Targeted engagement campaigns can solidify their loyalty.
* **Cluster 2:** Representing family-oriented users, this cluster includes customers with high dependents, long tenure, and a strong preference for fiber optic internet and multiple lines. Their monthly charges are high, indicating high service usage. This segment is ideal for premium service upsells and bundling family plans. Retention potential is high if value is consistently provided.
* **Cluster 3:** These are long-term, loyal customers primarily on two-year contracts. While their internet usage is lower compared to Cluster 2, they maintain consistent service use and have moderate tenure and charges. Their churn risk is minimal, but maintaining loyalty is crucial. Offering exclusive perks, personalized experiences, and referral incentives will keep this group engaged and advocate for the brand.

# Conclusion

This project successfully applied K-Means clustering to segment telecom customers into four actionable personas. Through PCA visualization and feature analysis, we labeled each group and identified unique strategies for marketing, retention, and engagement. These insights serve as a foundation for customer-centric business planning and churn reduction.

"Turning data into insight, and insight into action."