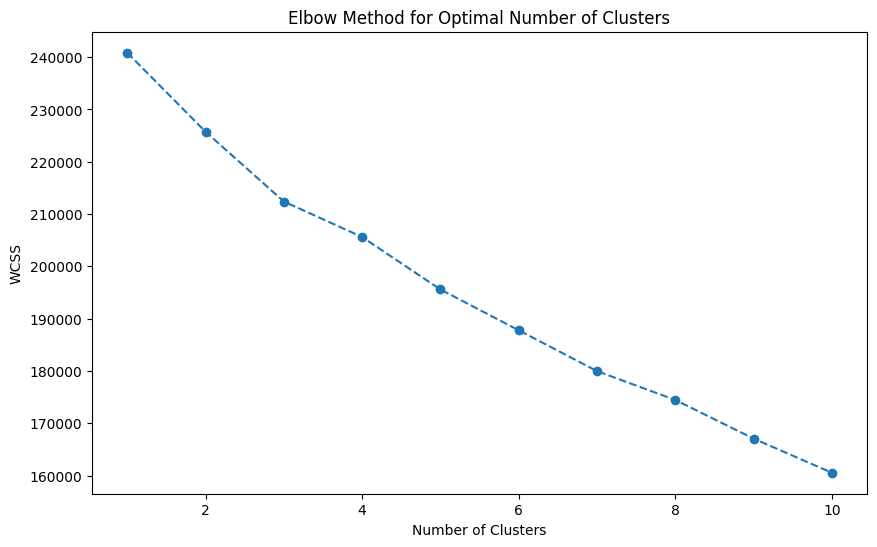
# Clustering Analysis

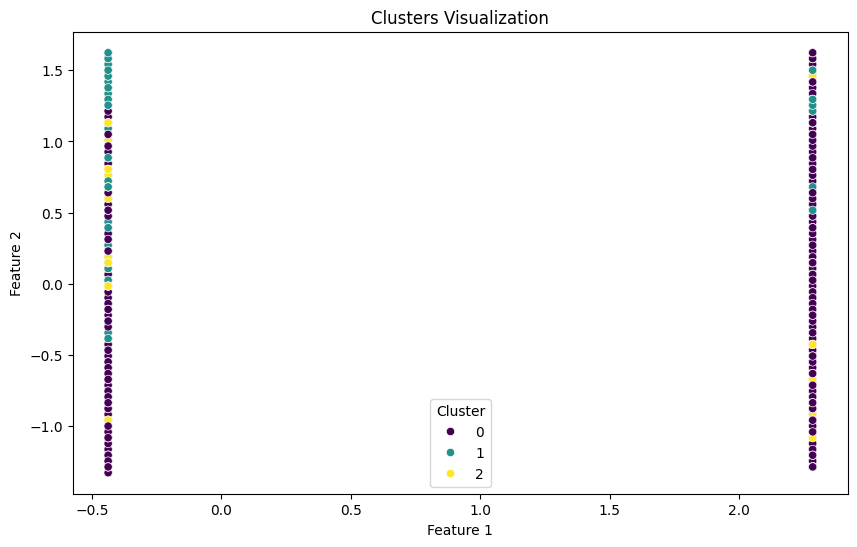
**Step 1: Identify the Optimal Number of Clusters using the Elbow Method**

The elbow method involves plotting the within-cluster sum of squares (WCSS) against the number of clusters and identifying the point where the rate of decrease sharply slows down (the "elbow").



The elbow plot provided indicates that the within-cluster sum of squares (WCSS) decreases as the number of clusters increases. The optimal number of clusters is typically at the "elbow point" where the rate of decrease sharply slows down. Based on the plot, it appears that the optimal number of clusters is around 3.

We'll use 3 clusters for the K-Means model and visualize the results.

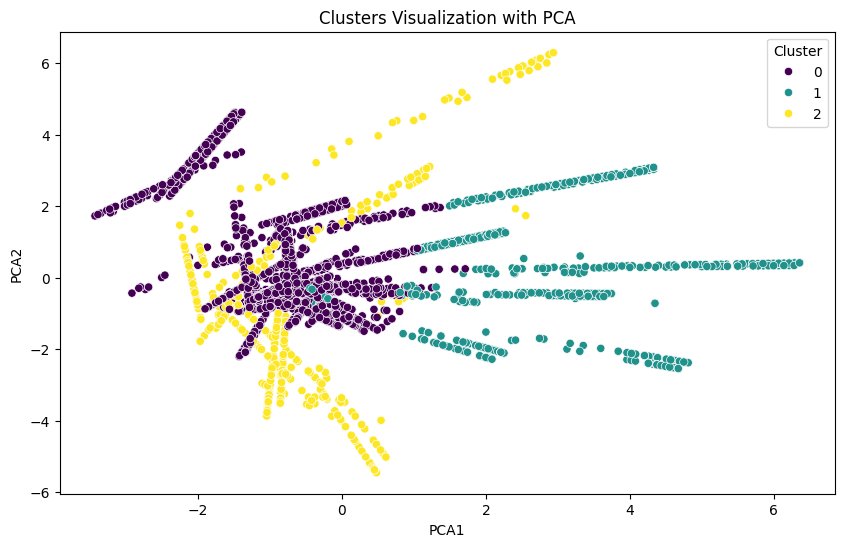


The clustering visualization attained indicates that the clusters are not well-separated. This might be due to the features used for the visualization. When visualizing clusters, it is important to choose features that clearly show the separation between clusters.

**Steps to Improve Visualization:**

1. **Dimensionality Reduction**: Use techniques like PCA (Principal Component Analysis) or t-SNE (t-distributed Stochastic Neighbor Embedding) to reduce the dimensionality of the data and visualize clusters in 2D or 3D space.
2. **Check Feature Importance**: Guarantee that the features used for visualization capture significant variations in the data.

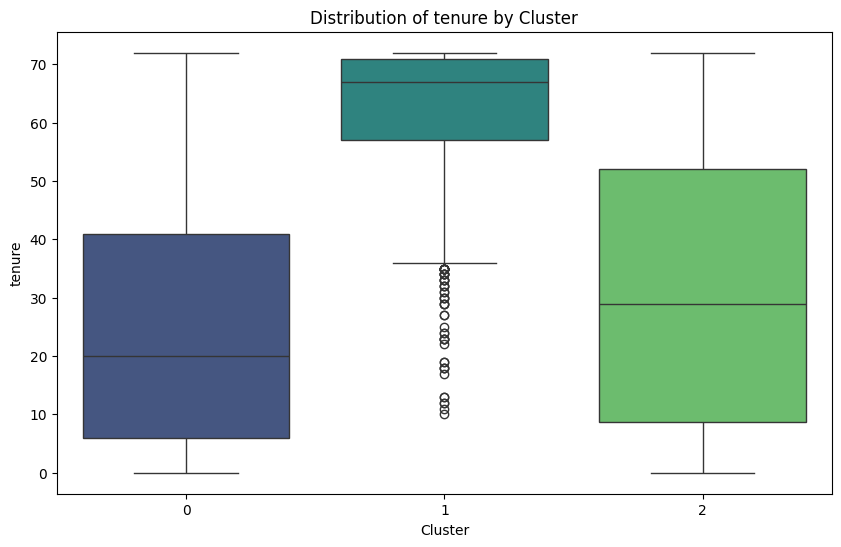
# Interpretation of PCA-based Clusters Visualization

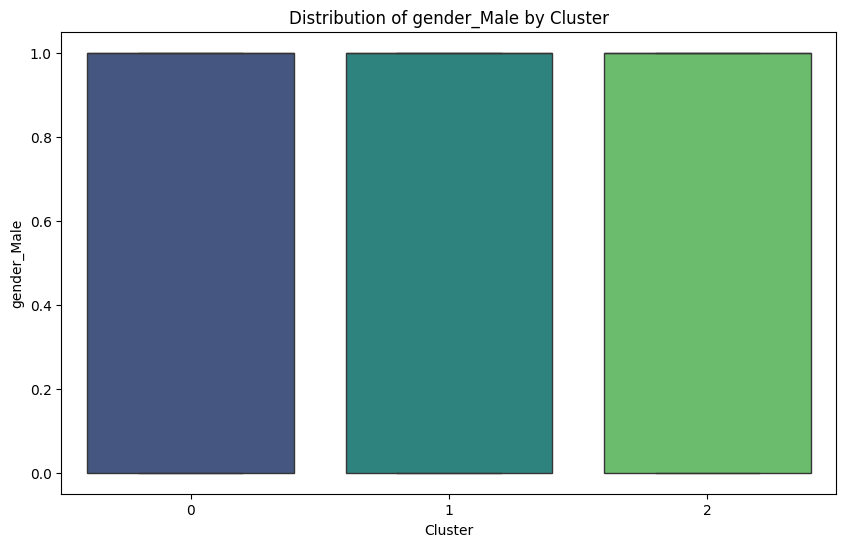


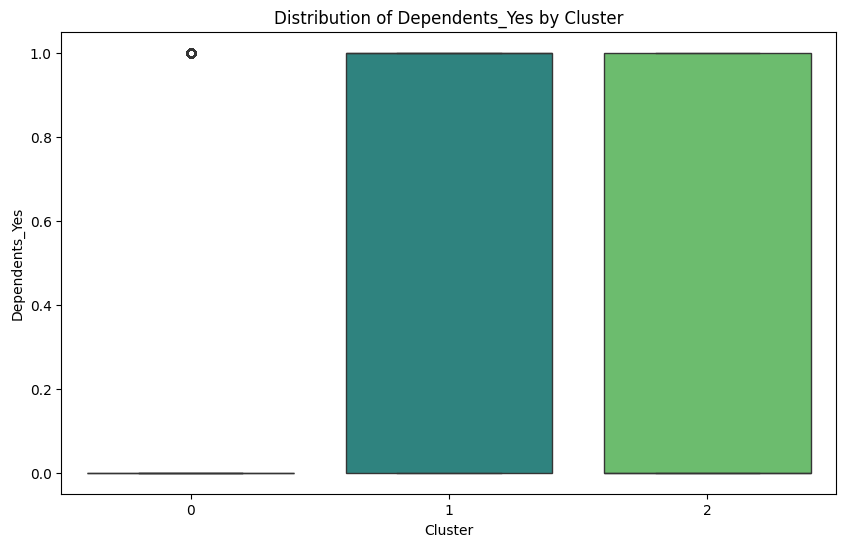
The PCA-based visualization shows a clearer separation of the clusters.

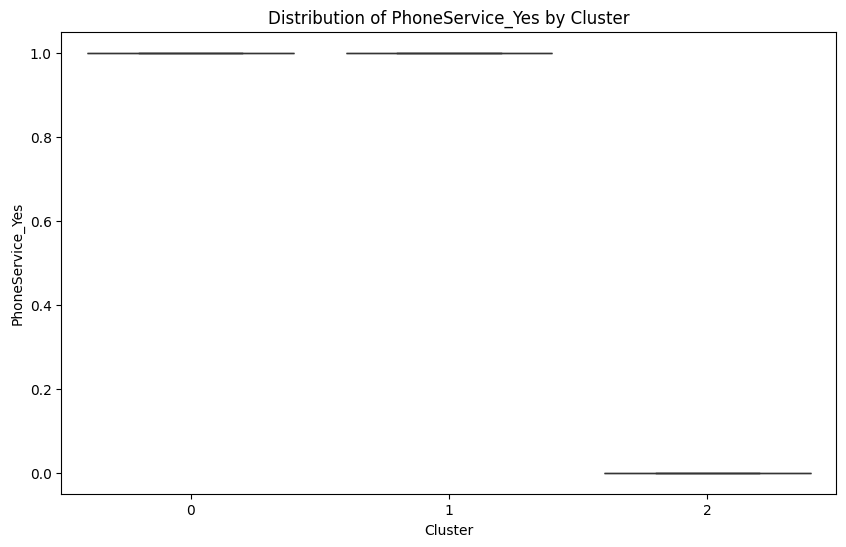
1. **Cluster 0 (Purple)**:
   * Appears to be a well-defined group with moderate variability along the PCA1 axis.
   * Members of this cluster are more densely packed towards the center, indicating similarity in certain features.
2. **Cluster 1 (Teal)**:
   * Spreads widely along the PCA1 axis.
   * This cluster shows more variability, indicating a diverse group of customers with varied characteristics.
3. **Cluster 2 (Yellow)**:
   * Has distinct separation from the other clusters, especially along the PCA1 axis.
   * Members of this cluster are more spread out, indicating a different set of characteristics compared to Clusters 0 and 1.

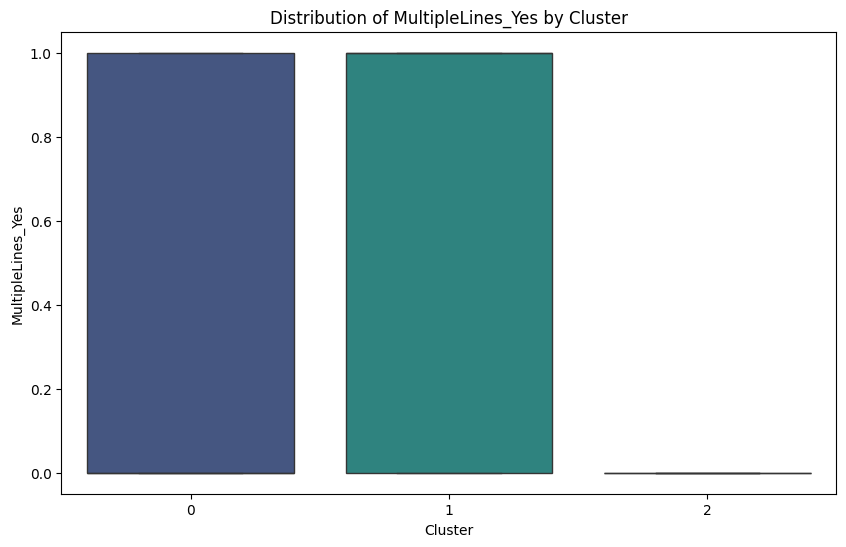
# Interpretation of Cluster Visualizations

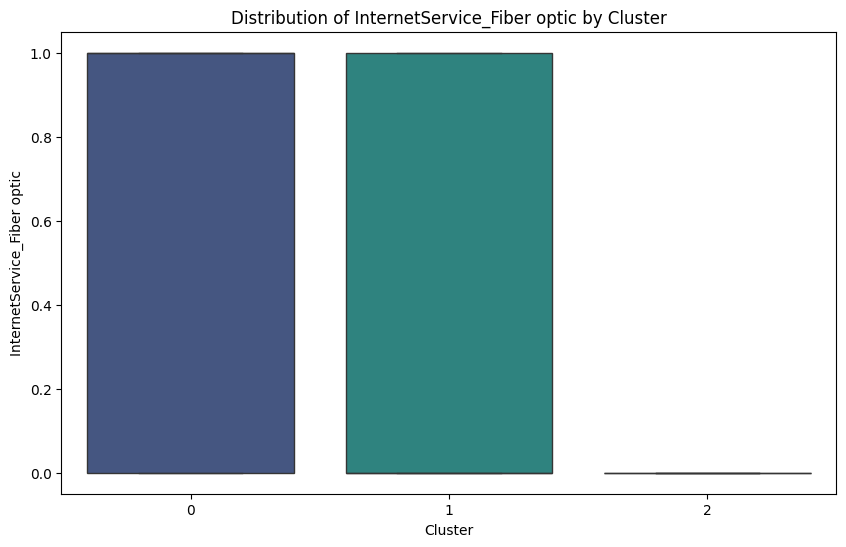


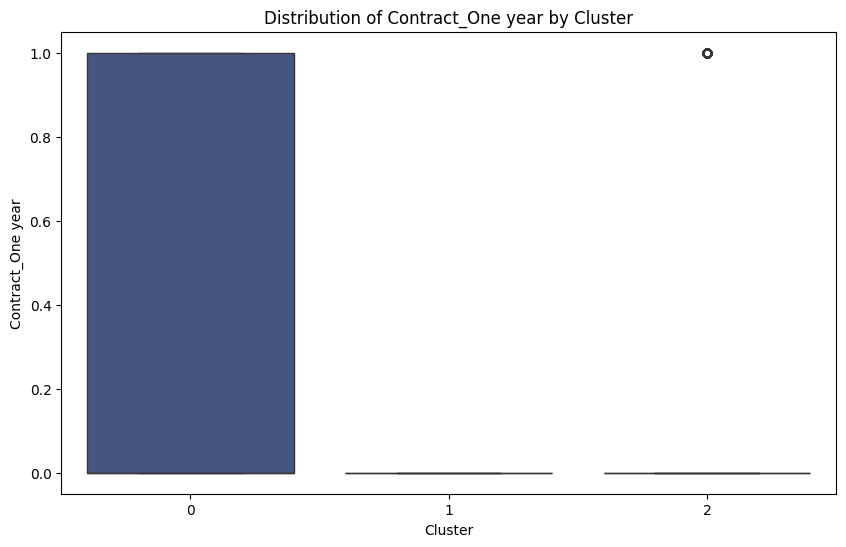


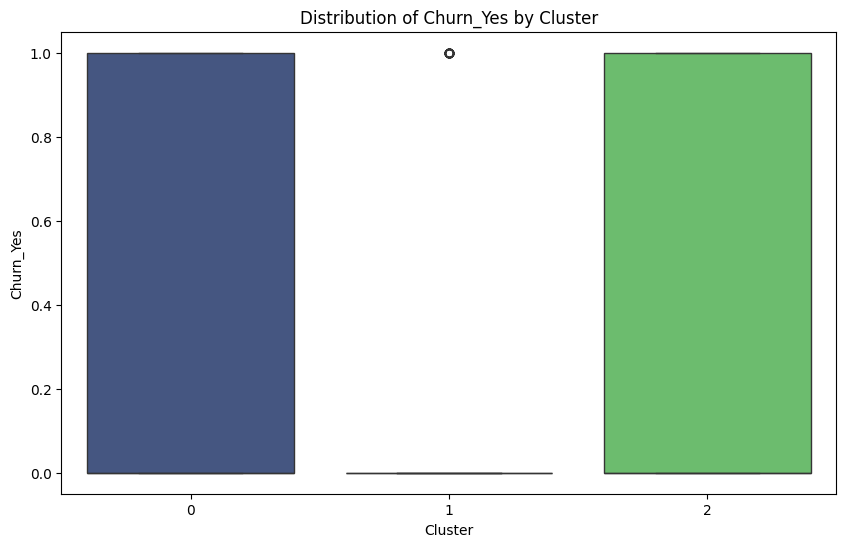












Based on the visualizations above, the interpretations and actionable insights for each cluster has been given below:

**Cluster 0**

* **SeniorCitizen**: Almost no senior citizens.
* **Tenure**: Moderate tenure, with a wide range from 0 to 70.
* **Gender**: Balanced gender distribution.
* **Dependents**: Almost no dependents.
* **PhoneService**: Almost all have phone service.
* **MultipleLines**: Majority have multiple lines.
* **InternetService**: Majority have fiber optic internet.
* **Contract**: All are on a one-year contract.
* **Churn**: All have churned.

**Label**: "Moderate Tenure, High Churn, Fiber Optic Users"

**Actionable Insights**:

* Focus on reducing churn by offering better customer service and retention plans for fiber optic users.
* Investigate why customers with moderate tenure and fiber optic internet are leaving.

**Cluster 1**

* **SeniorCitizen**: Similar distribution as Cluster 0.
* **Tenure**: High tenure, mostly above 60 months.
* **Gender**: Balanced gender distribution.
* **Dependents**: Majority have dependents.
* **PhoneService**: Almost all have phone service.
* **MultipleLines**: Majority have multiple lines.
* **InternetService**: Majority have fiber optic internet.
* **Contract**: Very few on a one-year contract.
* **Churn**: Very low churn.

**Label**: "High Tenure, Low Churn, Family Oriented"

**Actionable Insights**:

* These customers are highly loyal; continue offering family-friendly plans and promotions to retain them.
* Leverage this cluster to understand what drives customer satisfaction and apply these insights to other clusters.

**Cluster 2**

* **SeniorCitizen**: Similar distribution as Clusters 0 and 1.
* **Tenure**: Lower tenure, ranging from 0 to 70, but with a wider spread than Cluster 0.
* **Gender**: Balanced gender distribution.
* **Dependents**: Majority have dependents.
* **PhoneService**: Very few have phone service.
* **MultipleLines**: Very few have multiple lines.
* **InternetService**: Very few have fiber optic internet.
* **Contract**: Very few on a one-year contract.
* **Churn**: High churn.

**Label**: "Low Tenure, High Churn, Minimal Service Users"

**Actionable Insights**:

* Investigate why customers with low tenure and minimal services are leaving.
* Offer incentives for new customers to explore more services and become long-term users.

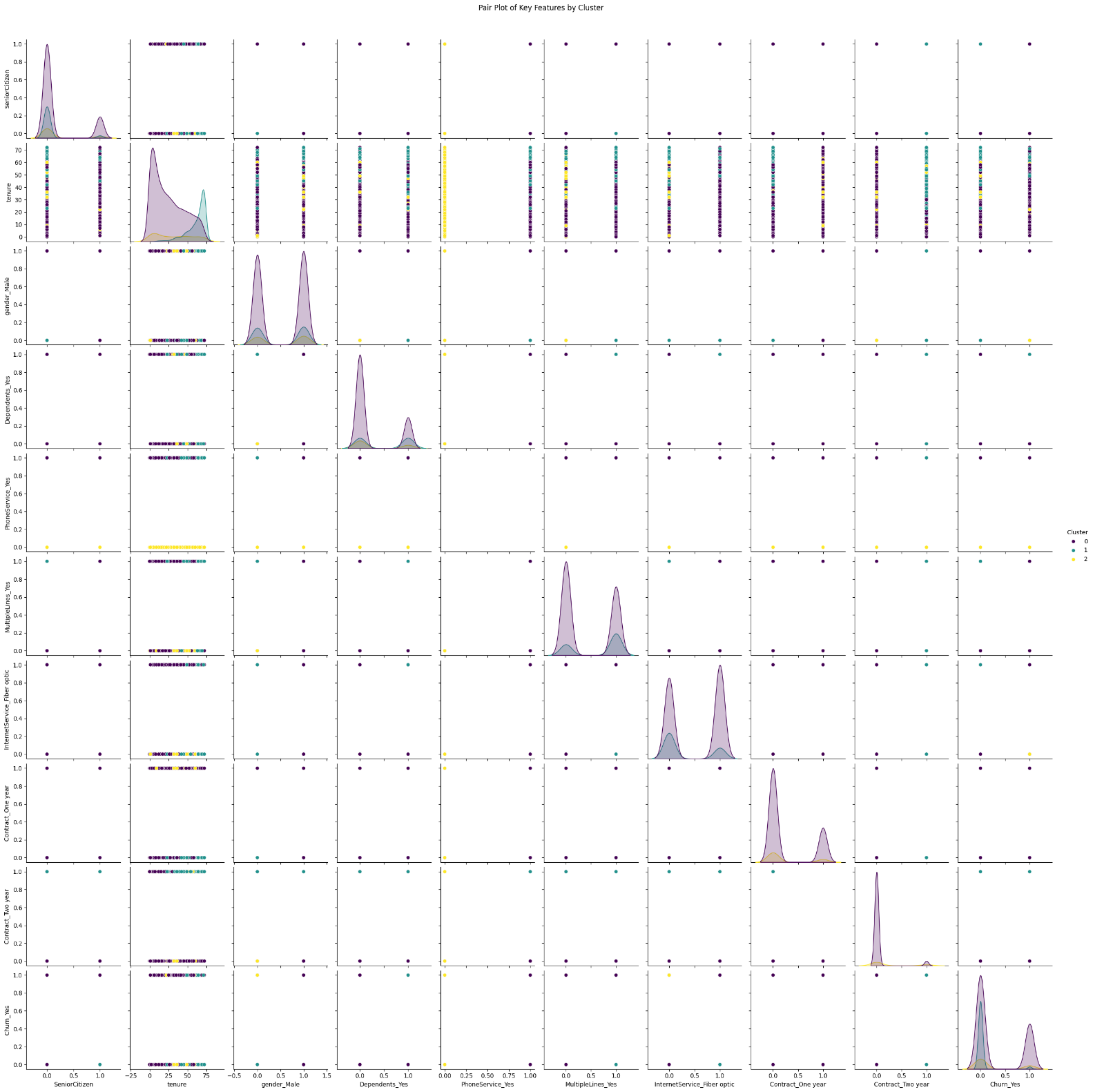
**Summary of Cluster Labels and Insights**

1. **Cluster 0: Moderate Tenure, High Churn, Fiber Optic Users**
   * Focus on reducing churn among fiber optic users with moderate tenure.
   * Investigate customer service issues and improve retention plans.
2. **Cluster 1: High Tenure, Low Churn, Family Oriented**
   * Continue offering family-friendly plans and promotions.
   * Use this cluster to derive insights for improving customer satisfaction.
3. **Cluster 2: Low Tenure, High Churn, Minimal Service Users**
   * Investigate reasons for high churn among minimal service users.
   * Provide incentives for new customers to explore and adopt more services.

# Some other Visualizations:

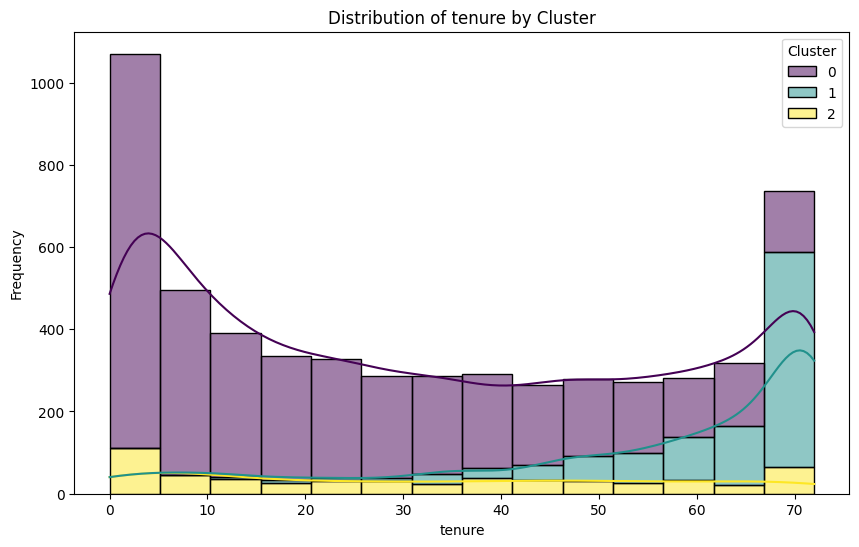
**1. Pair Plots**

Pair plots provide a way to visualize the relationships between pairs of features and the distribution of these features for different clusters.



**2. Histograms**

Histograms for each feature by cluster can help understand the distribution of numeric features.



**3. Heatmaps**

Heatmaps can visualize the correlation between features within each cluster.

