K means Cluster analysis

```
K-means clustering with 3 clusters of sizes 131, 201, 268

Cluster means:
         Trust Social_Influence Innovativeness    Mobility Perceived_Enjoyment
1  0.244737028        0.1277306      0.3583080  0.84039995           1.0733931
2  0.006495194        0.2060478      0.1177320  0.03095592          -0.2251212
3 -0.124500316       -0.2169713     -0.2634421 -0.43400945          -0.3558401
  Involvement Perceived_Usefulness Perceived_Easeofuse Customer_Satisfaction
1   0.6564766           -0.2908949          -0.4075394            -0.2904174
2   0.0465200            0.9675349           0.9776167             0.9560028
3  -0.3557797           -0.5834600          -0.5340048            -0.5750443
       Loyalty
1 -0.112083417
2  0.008759614
3  0.048217333
```

Cluster Means Interpretation:

Each cluster's means for various variables are presented in the table. Here's a breakdown:

Cluster 1:

- Trust: 0.245 (Moderately high)
- Social Influence: 0.128 (Low)
- Innovativeness: 0.358 (Moderate)
- Mobility: 0.840 (High)
- Perceived Enjoyment: 1.073 (Very high)
- Involvement: 0.656 (Moderate)
- Perceived Usefulness: -0.291 (Low)
- Perceived Ease of Use: -0.408 (Low)
- Customer Satisfaction: -0.290 (Low)
- Loyalty: -0.112 (Low)

Interpretation: This cluster exhibits high perceived enjoyment and mobility but lower perceived usefulness, ease of use, customer satisfaction, and loyalty.

Cluster 2:

- Trust: 0.006 (Very low)

- Social Influence: 0.206 (Moderate)
- Innovativeness: 0.118 (Low)
- Mobility: 0.031 (Very low)
- Perceived Enjoyment: -0.225 (Low)
- Involvement: 0.046 (Very low)
- Perceived Usefulness: 0.968 (High)
- Perceived Ease of Use: 0.978 (High)
- Customer Satisfaction: 0.956 (High)
- Loyalty: 0.009 (Very low)

Interpretation: This cluster shows very high perceived usefulness, ease of use, and customer satisfaction, but very low levels in trust, enjoyment, and loyalty. This group might prioritize functionality over enjoyment.

Cluster 3:

- Trust: -0.125 (Low)
- Social Influence: -0.217 (Low)
- Innovativeness: -0.263 (Low)
- Mobility: -0.434 (Low)
- Perceived Enjoyment: -0.356 (Low)
- Involvement: -0.356 (Low)
- Perceived Usefulness: -0.583 (Low)
- Perceived Ease of Use: -0.534 (Low)
- Customer Satisfaction: -0.575 (Low)
- Loyalty: 0.048 (Moderate)

Interpretation: This cluster demonstrates overall low values across nearly all dimensions, suggesting a group of customers who may not feel satisfied or loyal to the product or service.

Summary:

- **Cluster 1** may represent users who find enjoyment in the product or service but struggle with its practical aspects.
- **Cluster 2** consists of users who are highly satisfied and find the service useful and easy to use but lack trust and engagement.
- **Cluster 3** indicates users who are generally dissatisfied and disengaged.

These insights can inform targeted strategies for marketing, product development, and customer support, focusing on enhancing user experience, addressing pain points, and increasing engagement based on cluster characteristics.

```
Clustering vector:
  [1] 1 2 3 3 3 3 1 3 1 1 3 3 2 3 3 3 2 3 1 2 3 3 1 3 2 1 3 2 3 1 2 2 3 1 2 3 2
 [38] 3 2 3 1 3 1 2 3 2 2 3 3 3 1 3 3 2 1 2 2 2 2 1 3 2 3 3 3 3 1 3 3 1 3 3 3 1
 [75] 2 3 2 3 2 3 3 3 3 3 3 2 3 2 3 3 1 1 2 3 3 1 3 1 2 1 1 3 2 1 2 1 3 2 2 3 3
[112] 3 2 1 3 1 1 3 3 1 3 3 3 1 3 2 1 2 3 2 3 3 1 3 1 3 3 3 3 3 3 2 3 2 1 3 3 3
[149] 3 3 2 1 1 3 1 3 3 2 2 3 3 3 1 2 3 2 3 2 3 3 3 3 2 1 2 2 2 2 3 3 2 3 3 2 2
[186] 1 2 2 3 2 2 3 1 2 2 2 3 1 3 1 1 3 3 3 3 2 2 3 3 3 3 2 2 3 3 2 1 2 1 3 3 3
[223] 2 3 3 3 2 3 1 1 3 3 2 3 2 3 3 1 2 3 3 1 1 2 2 1 2 1 3 3 1 3 2 2 3 1 3 3 1
[260] 3 1 2 3 1 3 3 3 2 2 1 2 2 3 2 3 1 3 3 2 1 3 2 3 3 2 3 2 2 2 2 1 1 2 3 3 3
[297] 1 2 3 2 2 2 3 1 3 3 1 2 2 1 3 1 2 2 2 2 3 2 3 1 1 2 2 2 2 2 2 2 1 2 1 3 2
[334] 1 3 2 2 1 3 1 3 1 1 1 1 2 3 3 3 2 2 1 2 3 1 1 3 1 2 2 2 3 3 2 3 2 3 2 3 2
[371] 3 2 2 2 1 2 3 2 1 2 2 2 3 1 3 1 2 1 2 2 1 3 3 3 1 3 3 2 1 3 3 2 3 3 2 3 2
[408] 1 3 3 3 2 1 3 1 2 3 2 3 2 2 2 1 2 2 2 2 3 2 3 1 2 3 2 3 2 1 2 1 1 3 1 1 2
[445] 3 3 1 3 3 3 1 3 1 3 3 1 2 3 2 1 3 1 3 1 3 3 1 3 1 2 3 1 3 3 3 3 3 3 3 2 3
[482] 2 3 3 3 1 2 2 2 3 3 1 2 3 3 2 2 3 2 1 3 2 2 2 3 3 1 1 1 3 2 3 1 2 3 3 3 2
[519] 2 2 3 3 3 3 1 1 3 2 3 2 2 1 3 2 1 2 2 2 3 3 1 3 3 2 2 2 2 1 2 2 3 1 3 2 3
[556] 3 2 1 3 2 2 3 3 2 2 2 3 3 3 3 1 3 2 1 2 1 3 3 3 2 3 2 2 3 3 3 1 1 3 3 3 2
[593] 2 3 3 3 2 2 3 3

Within cluster sum of squares by cluster:
[1] 1159.482 1746.145 1715.386
 (between_SS / total_SS =  22.9 %)
```

Within Cluster Sum of Squares by Cluster

The "Within cluster sum of squares by cluster" section indicates how much variation exists within each cluster:

- **Cluster 1:** 1159.482
- **Cluster 2:** 1746.145
- **Cluster 3:** 1715.386

Interpretation of Within Cluster Sum of Squares

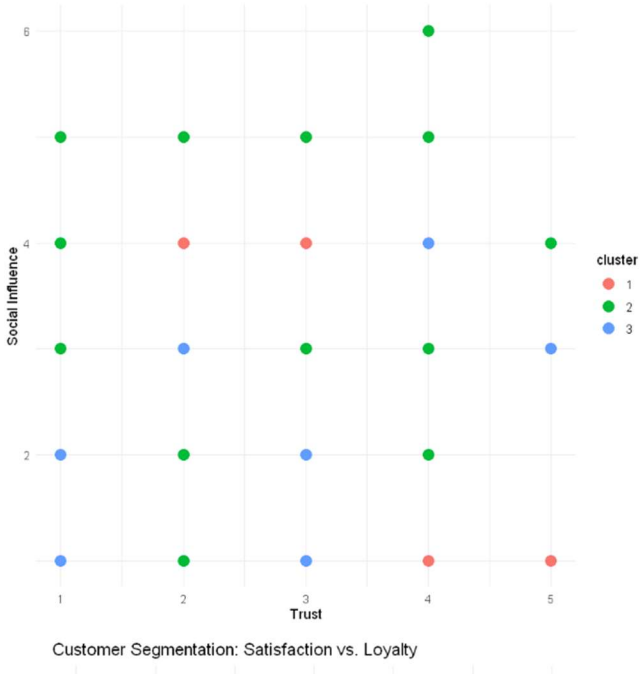
- The **within-cluster sum of squares** measures the compactness of each cluster. A lower value indicates that the observations within the cluster are closer to each other (more homogenous).
- In this case, **Cluster 1** has the lowest within-cluster sum of squares (1159.482), suggesting it is the most compact cluster.
- **Clusters 2** and **3** have higher sums of squares (1746.145 and 1715.386, respectively), indicating that the observations in these clusters are more spread out and potentially less homogeneous than those in Cluster 1.

```
Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

K-means Clustering

The scatter plot titled "Customer Segmentation: Satisfaction vs. Loyalty" uses three clusters (represented by different colors: red, green, and blue) to analyze the relationships between variables labeled as **Trust** (x-axis) and **Social Influence** (y-axis). Here's a breakdown of what this visualization indicates:



Customer Segmentation: Satisfaction vs. Loyalty

Key Observations:

1. Cluster Analysis:

- The plot is segmented into three distinct clusters:

- Cluster 1 (Red): This group is relatively small, and the data points are scattered at lower values of Trust and Social Influence.

- Cluster 2 (Green): This cluster has the highest concentration of points across different values of Trust and Social Influence, suggesting a broad distribution across all levels of these two variables.

- Cluster 3 (Blue): Points in this cluster are more evenly distributed across mid to high values of Trust and Social Influence.

2. Patterns and Grouping:

- Cluster 1 might represent customers with low satisfaction and loyalty, as seen from the lower range of values.

- Cluster 2 appears to cover a broad spectrum, possibly indicating a mix of satisfaction and loyalty levels.

- Cluster 3 has a spread across high Trust and Social Influence, indicating potentially highly satisfied and loyal customers.

3. Axes Interpretation:

- The x-axis labeled Trust ranges from 1 to 5.

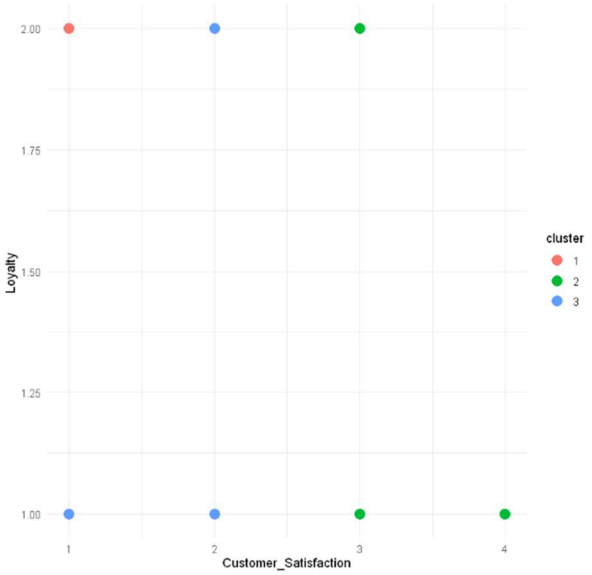
- The y-axis labeled Social Influence ranges from 1 to 6.

- The graph does not directly show "Satisfaction" or "Loyalty" on the axes, which might be inferred through combinations of Trust and Social Influence.

Summary:

The plot provides a clear segmentation of customer groups based on their trust and social influence levels, aiding in identifying segments such as low-engagement customers (Cluster 1), broadly distributed customers (Cluster 2), and high-trust, high-influence customers (Cluster 3). Further context would be needed to definitively label these clusters with respect to satisfaction and loyalty.

The scatter plot titled "Customer Satisfaction vs. Loyalty" presents three clusters (Cluster 1 in red, Cluster 2 in green, and Cluster 3 in blue) to analyze the relationship between **Customer Satisfaction** (x-axis) and **Loyalty** (y-axis). Here's a summary of the visualization:



Key Observations:

1. Cluster Analysis:

- Cluster 1 (Red): This cluster has a single point at a higher Loyalty value of 2, but a lower Customer Satisfaction value of 1.

- Cluster 2 (Green): Data points in this cluster span from mid to high levels of Customer Satisfaction (2 to 4) but remain at the lowest Loyalty value of 1.

- Cluster 3 (Blue): Points in this cluster are mostly concentrated at low Satisfaction values (1 to 2) but include varying Loyalty levels (1 and 2).

2. Patterns and Grouping:

- Cluster 1 seems to represent customers who have higher loyalty despite lower satisfaction, which might indicate external factors driving their loyalty (e.g., lack of alternatives, brand commitment).

- Cluster 2 indicates customers who are relatively satisfied (Customer Satisfaction = 3 or 4) but do not show high loyalty, which might suggest a gap in brand engagement or perceived value.

- Cluster 3 includes customers with low satisfaction and low-to-mid loyalty, likely representing disengaged or at-risk customers.

3. Distribution:

- There is no clustering observed at high values for both satisfaction and loyalty. The clusters appear sparse, which might suggest that most customers fall into segments with some disparity between satisfaction and loyalty.

This plot provides insights into how satisfaction and loyalty vary across different customer clusters. It shows that high loyalty does not always correlate with high satisfaction (as seen with Cluster 1) and highlights groups that may need strategic interventions to align satisfaction and loyalty levels. Further investigation could focus on the reasons behind these disparities to improve customer engagement strategies.

```
  1   2   3
131 201 268
```

The table provided shows the number of customers in each of the three clusters based on Customer Satisfaction. The clusters are represented as:

- Cluster 1: 131 customers

- Cluster 2: 201 customers

- Cluster 3: 268 customers

Interpretation:

1. Cluster 1 (131 Customers): This cluster has the smallest number of customers compared to the other two. Customers in this group might represent those with unique needs or specific dissatisfaction points. Understanding what differentiates them from other clusters would be critical in forming targeted engagement strategies.

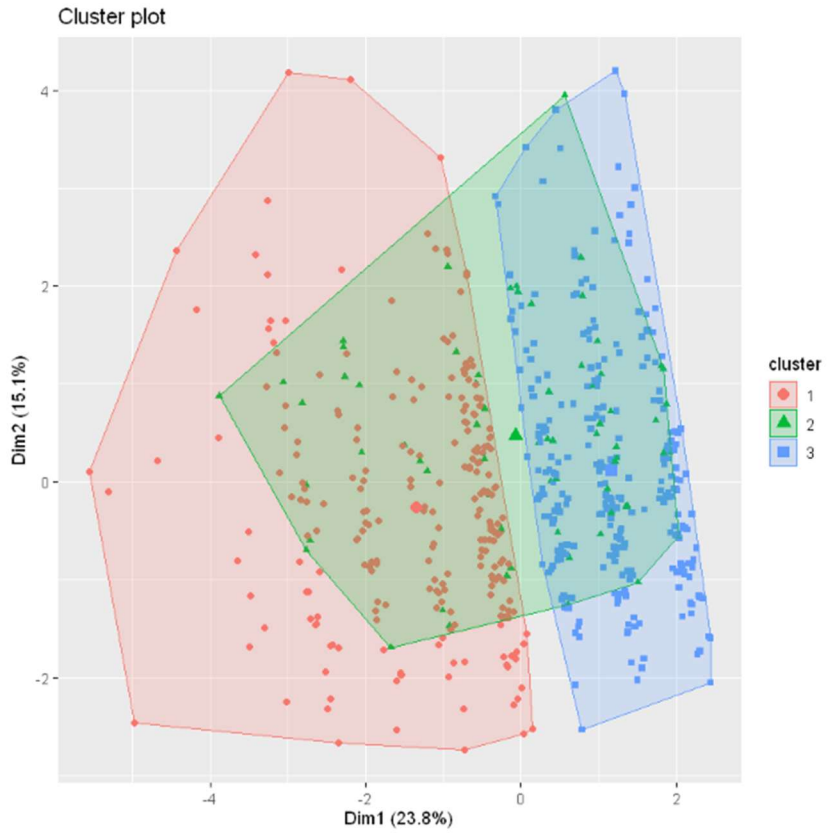
2. Cluster 2 (201 Customers): This cluster has a moderate number of customers. They could be moderately satisfied but not loyal or have varying experiences with the product/service. Analyzing their satisfaction levels in detail could reveal opportunities for converting them into loyal customers.

3. Cluster 3 (268 Customers): This is the largest cluster, indicating that a significant portion of the customer base shares similar satisfaction levels. This group could include customers who are satisfied but not entirely loyal or exhibit mixed behavior patterns. Further segmentation might be necessary to optimize engagement with this cluster.

Strategic Insight:

The larger size of Cluster 3 suggests that understanding the motivations and pain points of these customers should be a priority, as they constitute the majority. Special attention should also be given to Cluster 1, as they might be outliers or at-risk customers needing immediate intervention to prevent mix.

The image shows a cluster plot with three distinct clusters identified using some clustering algorithm (likely K-means or a similar method). Here's an interpretation of the plot:



1. Axes Interpretation:

- The `Dim1` (x-axis) and `Dim2` (y-axis) represent the first two principal components (PCs) or dimensions from a dimensionality reduction technique, capturing a significant amount of variance in the data.

- `Dim1` explains 23.8% of the variance, and `Dim2` explains 15.1%, as indicated by the axis labels.

2. Clusters:

- There are three clusters, each represented by a unique color and shape:

  - Cluster 1 (red diamonds): Positioned mostly to the left of the plot.

  - Cluster 2 (green triangles): Occupies the middle region, overlapping slightly with the other two clusters.

  - Cluster 3 (blue squares): Positioned to the right, with some overlap with Cluster 2.

- Each cluster has its own boundary, represented by a shaded convex hull, indicating the extent of data points in each cluster.

3. Cluster Separation:

- The boundaries of the clusters show how the data points are grouped. There is some overlap between clusters, particularly between **Clusters 2 and 3**. This overlap suggests that the distinction between these two groups may not be very strong in these dimensions.

- Cluster 1 has a broader spread and appears to be more separated from the other two clusters, indicating that it may have more distinct characteristics.

4. Variance Explained:

- The plot does not capture the entire variance of the original data since `Dim1` and `Dim2` only explain a combined 38.9% of the variance. However, it still provides a useful representation of the clustering structure.

Insights:

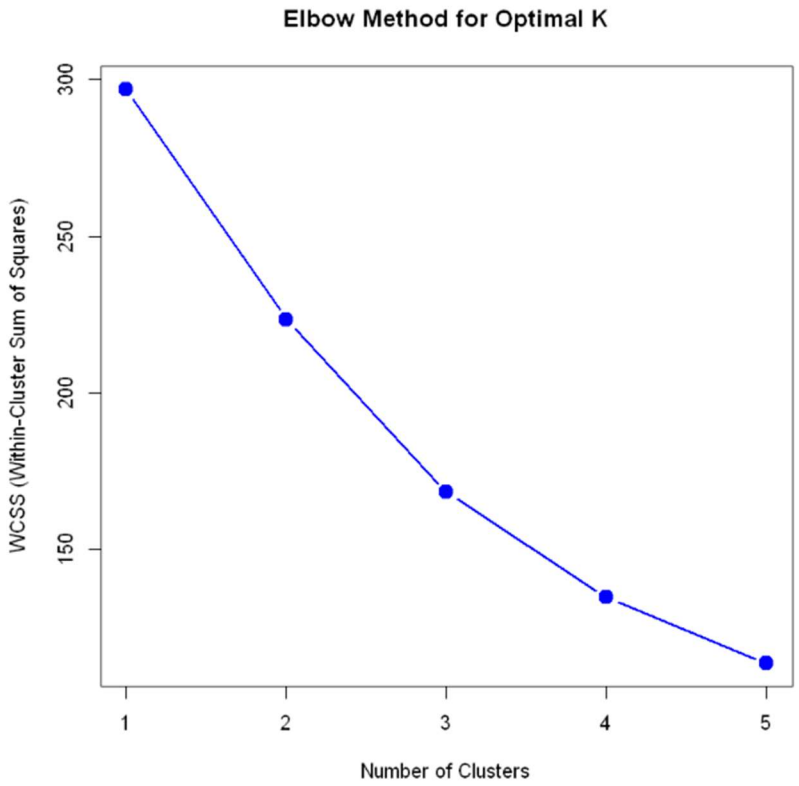
- Cluster Overlap: There is noticeable overlap between Cluster 2 and Cluster 3. This could suggest that these two clusters share common features or that the separation in these dimensions is not as distinct.

- Cluster Shape and Size: The shape and spread of each cluster indicate their internal variance. Cluster 1 is more dispersed, while Cluster 3 has a more condensed shape, suggesting lower within-cluster variance.

- Potential Action: If more separation is needed, consider using additional features or dimensions to improve cluster distinction or use different clustering algorithms.

Overall, this plot provides a clear visual representation of the customer segments based on the identified clusters and their distribution in the reduced-dimensional space.

The image shows an Elbow Method plot used to determine the optimal number of clusters (K) for a K-means clustering algorithm.



The image shows an Elbow Method plot used to determine the optimal number of clusters (K) for a K-means clustering algorithm. Here's a detailed interpretation:

1. **X-axis (Number of Clusters):**

- Represents the number of clusters (K) tested, ranging from 1 to 5.

2. **Y-axis (WCSS: Within-Cluster Sum of Squares):**

- The Within-Cluster Sum of Squares (WCSS) measures the total variance within each cluster. A lower WCSS value indicates that the points within each cluster are closer to each other.

3. **Elbow Plot:**

- The plot shows the WCSS value for each number of clusters (1 through 5). As the number of clusters increases, the WCSS decreases because adding more clusters reduces the variance within each cluster.

Interpretation:

1. **Pattern Analysis:**

- The WCSS decreases rapidly from K = 1 to K = 3, indicating that adding clusters significantly reduces within-cluster variance.

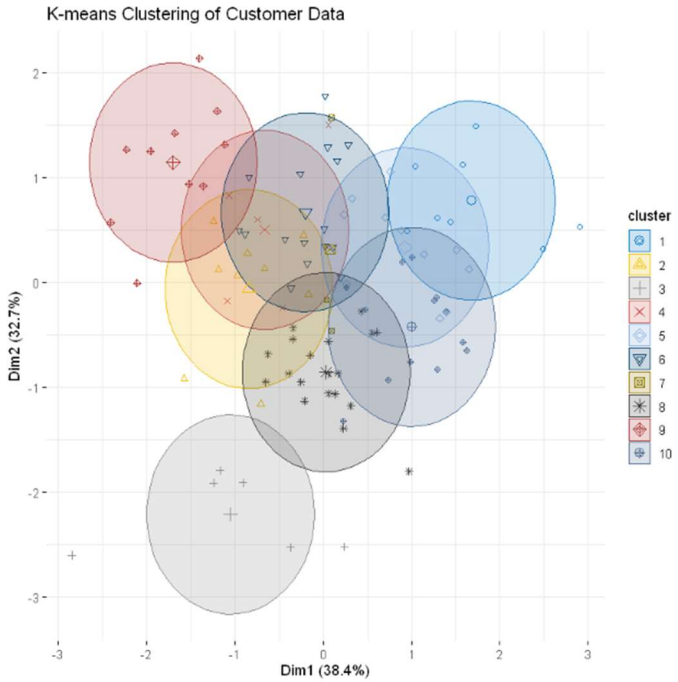
- After K = 3, the reduction in WCSS slows down, showing a diminishing return in variance reduction as more clusters are added.

2. **Optimal Number of Clusters (Elbow Point):**

- The "elbow point" of the plot is typically considered the optimal number of clusters. This point is where the curve starts to flatten, indicating that adding more clusters beyond this point results in only a slight reduction in WCSS.

- In this plot, the elbow appears around K = 3, suggesting that using 3 clusters is a good choice for capturing the underlying structure of the data without overfitting.

The optimal number of clusters is likely 3, as it provides a good balance between minimizing within-cluster variance and not over-complicating the model with too many clusters.



The image shows a K-means clustering plot of customer data. Here's an interpretation of the visual:

1. **Axes Interpretation:**

- The `Dim1` (x-axis) and `Dim2` (y-axis) represent the first two principal components (PCs) from a dimensionality reduction technique (likely PCA) applied to the customer data. These dimensions are used to project high-dimensional data into a two-dimensional space, capturing the most variance.

2. **Clusters:**

- There are 10 distinct clusters, each represented by different colors and shapes (e.g., circles, triangles, crosses, etc.). The plot shows how data points are grouped into clusters based on similarity.

- The clusters are labeled from 1 to 10, as indicated in the legend on the right.

3. **Cluster Separation:**

- Each cluster is represented by a colored circle that outlines its boundary. The circles show the approximate area occupied by each cluster in this two-dimensional space.

- Some clusters overlap, indicating that the customer groups share similarities in these dimensions. This overlap might suggest that the distinction between some clusters is not very strong in the projected space.

4. **Variance Explained:**

- The plot indicates the amount of variance explained by each dimension:

- `Dim1` (38.4%): Explains 38.4% of the variance in the dataset.

- `Dim2` (32.7%): Explains 32.7% of the variance.

- Combined, these two dimensions explain around 71.1% of the variance in the data, which is a good representation in a reduced space.

Insights:

- **Cluster 2** (yellow triangle) is overlapping significantly with multiple clusters, indicating that the customers in this group share characteristics with those in other groups.

- **Clusters 1, 4, and 5** (different shades of blue) are relatively separated, indicating more distinct characteristics within these customer groups.

- **Cluster 10 (dark blue)** appears to have a broader spread, suggesting that it may contain more diverse data points within it.

Overall, this visualization helps in identifying distinct customer segments based on their attributes and understanding the relationships between different clusters in a simplified two-dimensional space.