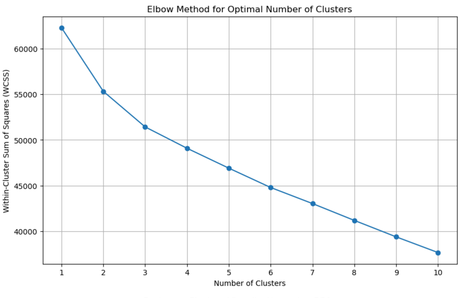
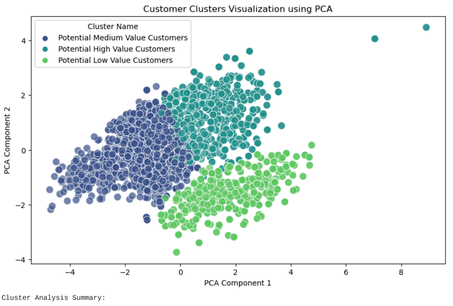
In [ ]:

In [ ]:

In [6]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA

# Load the dataset
file_path = r"C:\Users\manid\OneDrive\Desktop\churn_true.csv"
df = pd.read_csv(file_path)

# Ensure 'Churn' column is properly formatted
df['Churn'] = df['Churn'].astype(int)

# Select relevant numerical columns for clustering
numerical_cols = [
    'Account_Length', 'VMail_Message', 'Day_Mins', 'Day_Calls', 'Day_Charge',
    'Eve_Mins', 'Eve_Calls', 'Eve_Charge', 'Night_Mins', 'Night_Calls',
    'Night_Charge', 'Intl_Mins', 'Intl_Calls', 'Intl_Charge', 'CustServ_Calls'
]

X = df[numerical_cols]

# Standardize the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Determine the optimal number of clusters using the Elbow Method
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_scaled)
    wcss.append(kmeans.inertia_)

# Plot the Elbow Method
plt.figure(figsize=(10, 6))
plt.plot(range(1, 11), wcss, marker='o', linestyle='-')
plt.title('Elbow Method for Optimal Number of Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('Within-Cluster Sum of Squares (WCSS)')
plt.xticks(range(1, 11))
plt.grid(True)
plt.show()

# Choosing the optimal number of clusters, e.g., 3 for illustration purposes
optimal_clusters = 3
kmeans = KMeans(n_clusters=optimal_clusters, init='k-means++', max_iter=300, n_init=10, random_state=42)
df['Cluster'] = kmeans.fit_predict(X_scaled)

# Visualize the clusters using PCA
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)
df['PCA1'] = X_pca[:, 0]
df['PCA2'] = X_pca[:, 1]

# Calculate Customer Lifetime Value (CLV)
df['CLV'] = df['Day_Charge'] + df['Eve_Charge'] + df['Night_Charge'] + df['Intl_Charge']

# Define cluster names based on your analysis
cluster_names = {
    0: 'Potential Low Value Customers',
    1: 'Potential Medium Value Customers',
    2: 'Potential High Value Customers'
}

# Map clusters to descriptive names
df['Cluster_Name'] = df['Cluster'].map(cluster_names)

# Visualizing the clusters
plt.figure(figsize=(10, 6))
sns.scatterplot(x='PCA1', y='PCA2', hue='Cluster_Name', data=df, palette='viridis', s=100, alpha=0.7)
plt.title('Customer Clusters Visualization using PCA')
plt.xlabel('PCA Component 1')
plt.ylabel('PCA Component 2')
plt.legend(title='Cluster Name', loc='best')
plt.show()

# Cluster analysis: describe statistics and CLV
cluster_analysis = {}
for i in range(optimal_clusters):
    cluster_name = cluster_names.get(i, f'Cluster {i}')
    cluster_data = df[df['Cluster'] == i][numerical_cols].describe()
    cluster_clv_mean = df[df['Cluster'] == i]['CLV'].mean()
    churn_rate = df[df['Cluster'] == i]['Churn'].mean() * 100  # Convert to percentage
    cluster_analysis[cluster_name] = {
        'Data Description': cluster_data,
        'Average CLV': cluster_clv_mean,
        'Churn Rate': churn_rate
    }

# Count the number of records in each cluster
cluster_counts = df['Cluster_Name'].value_counts()

# Save the updated dataframe with clusters and CLV to a CSV file
output_path = r"C:\Users\manid\OneDrive\Desktop\churn_true_clus.csv"
df.to_csv(output_path, index=False)

# Results
print("Cluster Analysis Summary:")
for cluster_name, analysis in cluster_analysis.items():
    print(f"\n{cluster_name}:\n")
    print(analysis['Data Description'])
    print(f"Average CLV: {analysis['Average CLV']:.2f}")
    print(f"Churn Rate: {analysis['Churn Rate']:.2f}%")

print("\nNumber of Records in Each Cluster:")
print(cluster_counts)
```






```
Cluster Analysis Summary:

Potential Low Value Customers:

       Account_Length  VMail_Message    Day_Mins   Day_Calls  Day_Charge  \
count      815.000000     815.000000  815.000000  815.000000  815.000000
mean       100.150920       9.519018  267.742822  171.063804   40.132748
std        101.920288      21.453608   52.058728   93.199350   24.539723
min          0.000000       0.000000   82.200000   22.000000    9.890000
25%         86.000000       0.000000  228.800000   88.000000   22.960000
50%        208.000000       0.000000  282.800000  202.000000   28.890000
75%        282.000000       0.000000  296.100000  226.000000   32.890000
max        329.000000      88.000000  423.400000  329.000000   89.980000

        Eve_Mins   Eve_Calls  Eve_Charge  Night_Mins  Night_Calls  \
count  815.000000  815.000000  815.000000  815.000000  815.000000
mean   267.214969  157.452761   26.655755  268.176442  162.157055
std     41.364894   95.047718    4.814313   39.823880   93.681421
min    200.600000   20.000000    8.980000  202.000000   22.000000
25%    236.800000   88.000000   22.200000  228.200000   88.000000
50%    269.800000   98.000000   28.260000  282.000000  200.000000
75%    292.800000  222.000000   29.320000  292.100000  226.000000
max    329.900000  329.000000   32.990000  329.800000  329.000000

       Night_Charge   Intl_Mins  Intl_Calls  Intl_Charge  CustServ_Calls
count    815.000000  815.000000  815.000000   815.000000      815.000000
mean      13.069104   27.532147    5.419032     8.431018        2.787730
std        8.936546    4.237295    5.095463     0.390067        2.698229
min        2.020000   20.000000    2.000000     4.240000        0.000000
25%        8.020000   22.800000    2.000000     8.260000        2.000000
50%        9.220000   28.800000    2.000000     8.280000        2.000000
75%       20.820000   32.200000    8.000000     8.820000        2.000000
max       32.980000   32.900000   32.000000     8.920000        9.000000
Average CLV: 88.29
Churn Rate: 100.00%

Potential Medium Value Customers:

       Account_Length  VMail_Message     Day_Mins   Day_Calls  Day_Charge  \
count     2264.000000    2264.000000  2264.000000  2264.000000  2264.000000
mean       127.633392       5.367491   231.863825   134.278269    29.897182
std         83.871827      13.797318    46.589209    76.958370    12.227067
min          1.000000       0.000000    62.400000    22.000000    10.610000
25%         43.000000       0.000000   222.300000    75.000000    22.415000
50%         98.000000       0.000000   229.800000   106.000000    26.220000
75%        222.000000       0.000000   260.000000   220.000000    32.600000
max        298.000000      88.000000   429.400000   328.000000    89.080000

         Eve_Mins    Eve_Calls   Eve_Charge   Night_Mins  Night_Calls  \
count  2264.000000  2264.000000  2264.000000  2264.000000  2264.000000
mean    231.528313   131.469081    22.345760   230.372173   133.407244
std      32.142102    76.411958     4.268822    30.259936    77.409461
min     109.600000    20.000000     3.220000   123.500000    20.000000
25%     220.600000    77.750000    21.397500   222.000000    77.750000
50%     228.200000   100.000000    22.390000   226.200000   104.000000
75%     242.400000   220.000000    23.630000   242.200000   220.000000
max     404.200000   329.000000    32.890000   332.700000   326.000000

       Night_Charge    Intl_Mins   Intl_Calls  Intl_Charge  CustServ_Calls
count   2264.000000  2264.000000  2264.000000  2264.000000     2264.000000
mean      10.453467    14.372350     3.941696     2.814819        2.199205
std        6.936043     7.733242     3.710128     0.792529        1.770175
min        2.020000     2.000000     1.000000     0.320000        0.000000
25%        4.240000     8.700000     2.000000     2.260000        2.000000
50%        9.220000    13.200000     3.000000     2.620000        2.000000
75%       14.045000    22.200000     4.000000     3.230000        2.000000
max       24.220000    28.800000    28.000000     8.820000        9.000000
Average CLV: 65.51
Churn Rate: 100.00%

Potential High Value Customers:

       Account_Length  VMail_Message     Day_Mins    Day_Calls   Day_Charge  \
count     1074.000000    1074.000000  1074.000000  1074.000000  1074.000000
mean       101.934823       7.547486   282.258520   188.414339    42.256074
std        105.490314      21.238320    44.153311    93.528156    24.605756
min          2.000000       0.000000    82.900000    22.000000    20.030000
25%         82.000000       0.000000   262.200000    92.000000    28.020000
50%        209.000000       0.000000   282.900000   208.000000    29.620000
75%        282.000000       0.000000   322.000000   229.000000    44.020000
max        329.000000      88.000000   444.400000   329.000000   112.011000

         Eve_Mins    Eve_Calls   Eve_Charge   Night_Mins  Night_Calls  \
count  1074.000000  1074.000000  1074.000000  1074.000000  1074.000000
mean    282.326816   180.987896   28.043820   282.169209   171.635009
std      36.108108    94.873806   13.280099    35.602576    97.912437
min     202.200000    22.000000    8.320000   160.600000    20.000000
25%     262.900000    92.000000   23.300000   262.900000    88.000000
50%     286.200000   203.000000   28.280000   288.200000   202.000000
75%     299.200000   228.000000   29.660000   320.200000   226.000000
max     404.200000   329.000000   211.211000   354.900000   329.000000

       Night_Charge    Intl_Mins   Intl_Calls  Intl_Charge  CustServ_Calls
count   1074.000000  1074.000000  1074.000000  1074.000000     1074.000000
mean      14.968200    13.166918     5.539106     2.571038        2.761639
std        9.572395     8.369207     4.860482     0.422409        2.582647
min        2.080000     2.000000     2.000000     0.820000        0.000000
25%        8.020000     8.200000     2.000000     2.280000        2.000000
50%        9.820000     9.200000     3.000000     2.465000        2.000000
75%       22.255000    20.600000     8.000000     2.880000        2.000000
max       32.960000    32.900000    32.000000     4.440000        9.000000
Average CLV: 87.84
Churn Rate: 100.00%

Number of Records in Each Cluster:
Cluster_Name
Potential Medium Value Customers    2264
Potential High Value Customers      1074
Potential Low Value Customers        815
Name: count, dtype: int64
```

In [ ]: