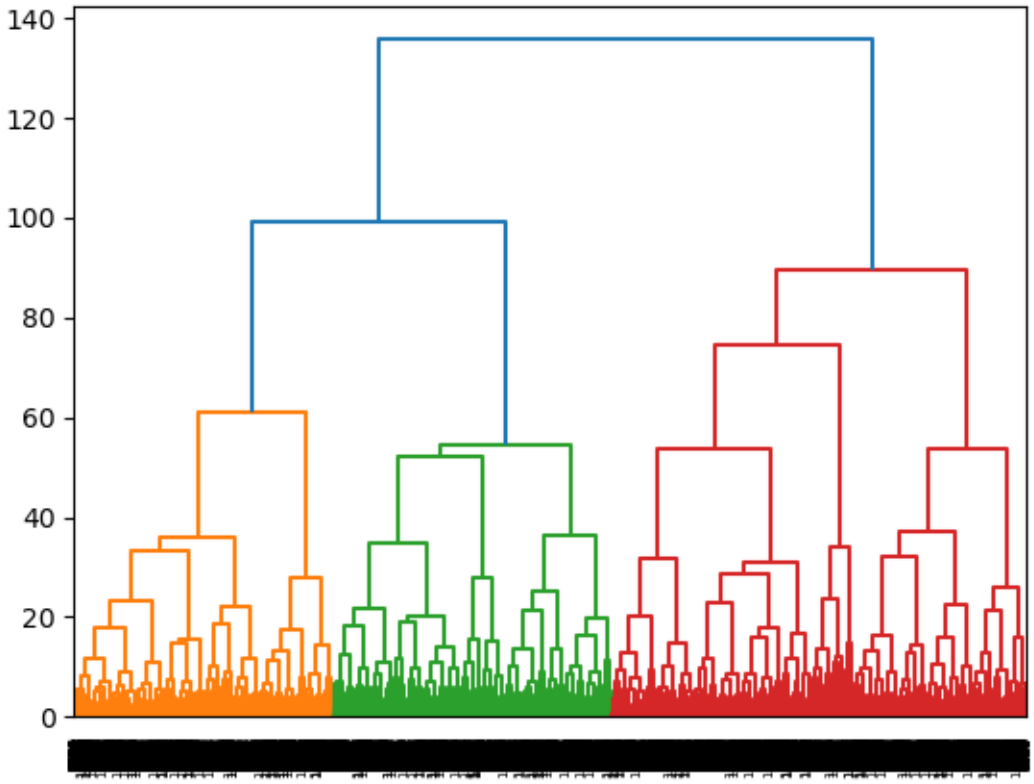
4

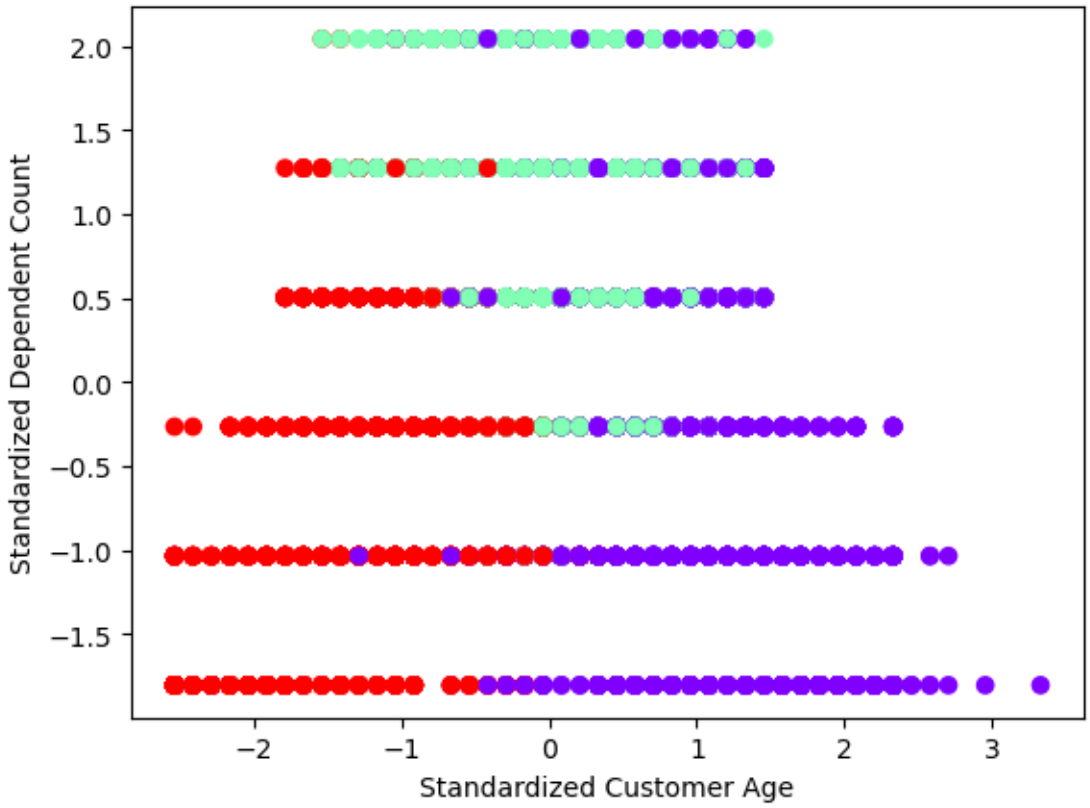
November 3, 2023

Implement Agglomerative hierarchical clustering algorithm using appropriate dataset.

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import AgglomerativeClustering
from scipy.cluster.hierarchy import dendrogram, linkage
```

[2]:
```
# Load the credit card dataset
data = pd.read_csv('BankChurners.csv')

# Select relevant features
X = data[['Customer_Age', 'Dependent_count', 'Months_on_book',␣
 ↪'Total_Relationship_Count', 'Months_Inactive_12_mon']]

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

# Apply Agglomerative Clustering
agg_clustering = AgglomerativeClustering(n_clusters=3)
agg_labels = agg_clustering.fit_predict(X_scaled)

# Plot the dendrogram
linked = linkage(X_scaled, 'ward')
dendrogram(linked, orientation='top', distance_sort='descending',␣
 ↪show_leaf_counts=True)
plt.show()

# Plot the clusters (using 2D projection for visualization)
plt.scatter(X_scaled[:, 0], X_scaled[:, 1], c=agg_labels, cmap='rainbow')
plt.xlabel('Standardized Customer Age')
plt.ylabel('Standardized Dependent Count')
plt.show()
```


140
120
100
80
60
40
20
0


Standardized Dependent Count
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2
−1
0
1
2
3
Standardized Customer Age