



AIML CA2

CUSTOMER

CLUSTERING

Goh Kun Ming
DAAA/FT/1B/02



Data Cleaning and Analysis

Checking For Missing Data

Missing Values	
Gender	0
Age	0
Income (k\$)	0
How Much They Spend	0

Checking For Duplicated Data

Count	
Duplicated Rows	0

Dropping Exclusive Data

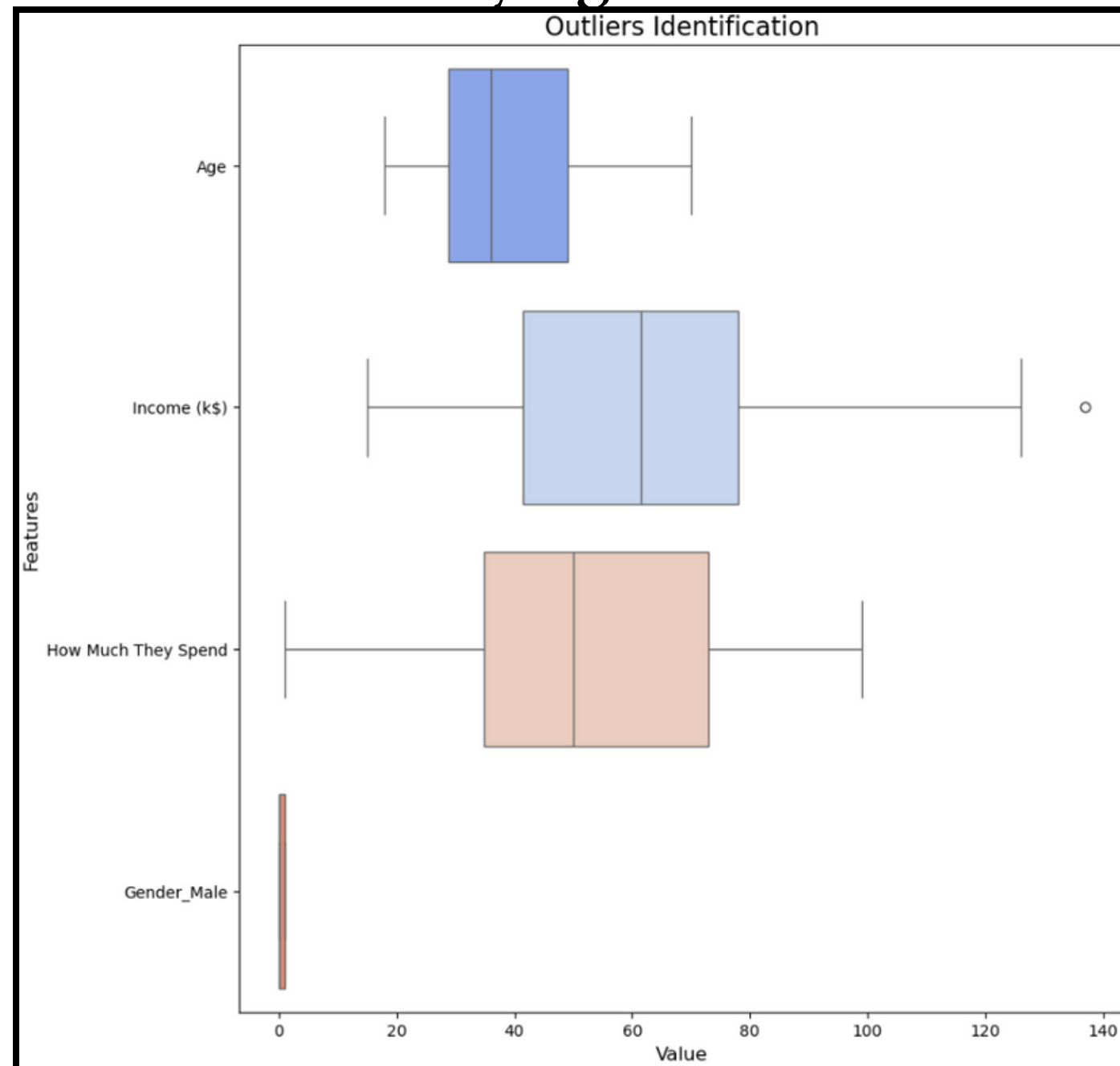
	Gender	Age	Income (k\$)	How Much They Spend
0	Male	19	15	39
1	Male	21	15	81
2	Female	20	16	6
3	Female	23	16	77
4	Female	31	17	40
5	Female	22	17	76
6	Female	35	18	6
7	Female	23	18	94
8	Male	64	19	3
9	Female	30	19	72

Creating Dummy Variable

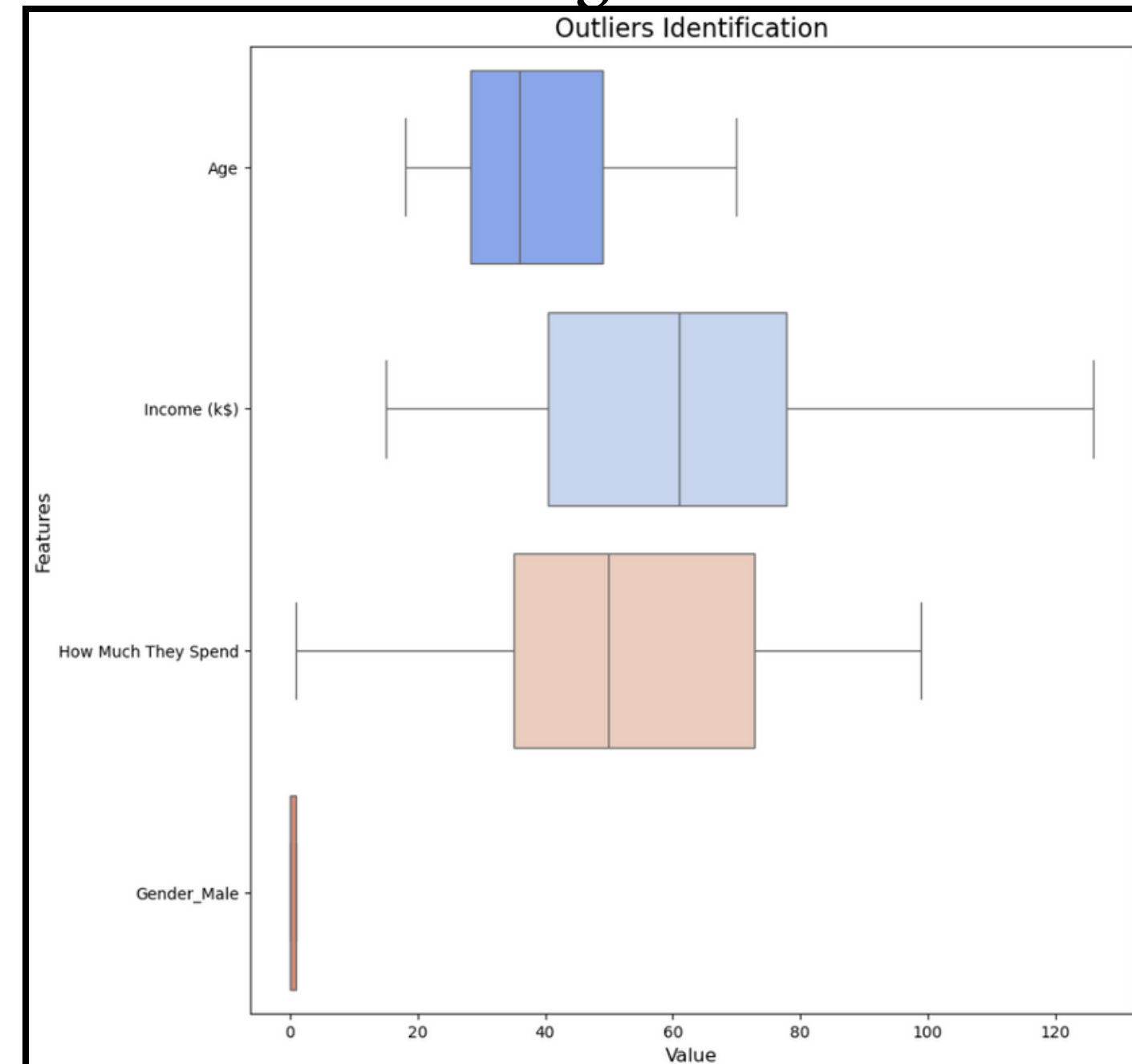
	Age	Income (k\$)	How Much They Spend	Gender_Male
0	19	15	39	1
1	21	15	81	1
2	20	16	6	0
3	23	16	77	0
4	31	17	40	0

Data Cleaning and Analysis

Identifying Outliers

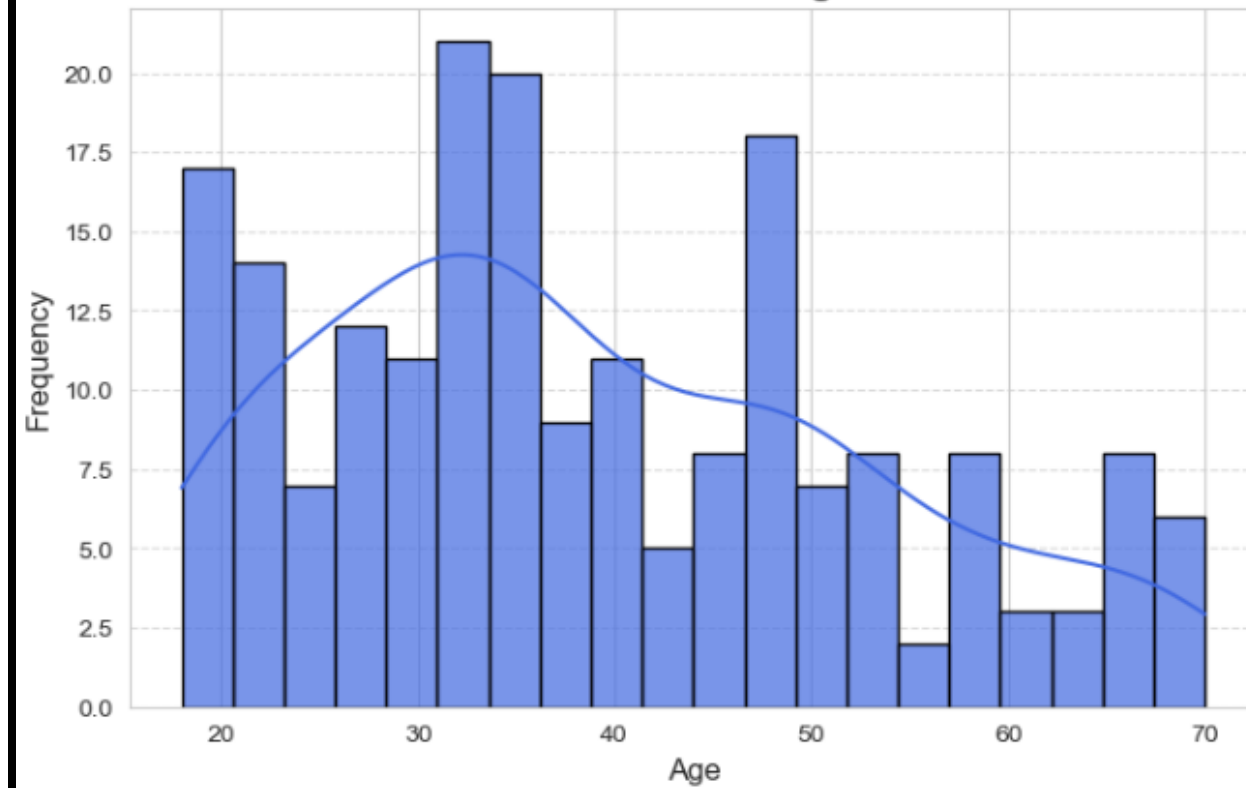


Removing Outliers

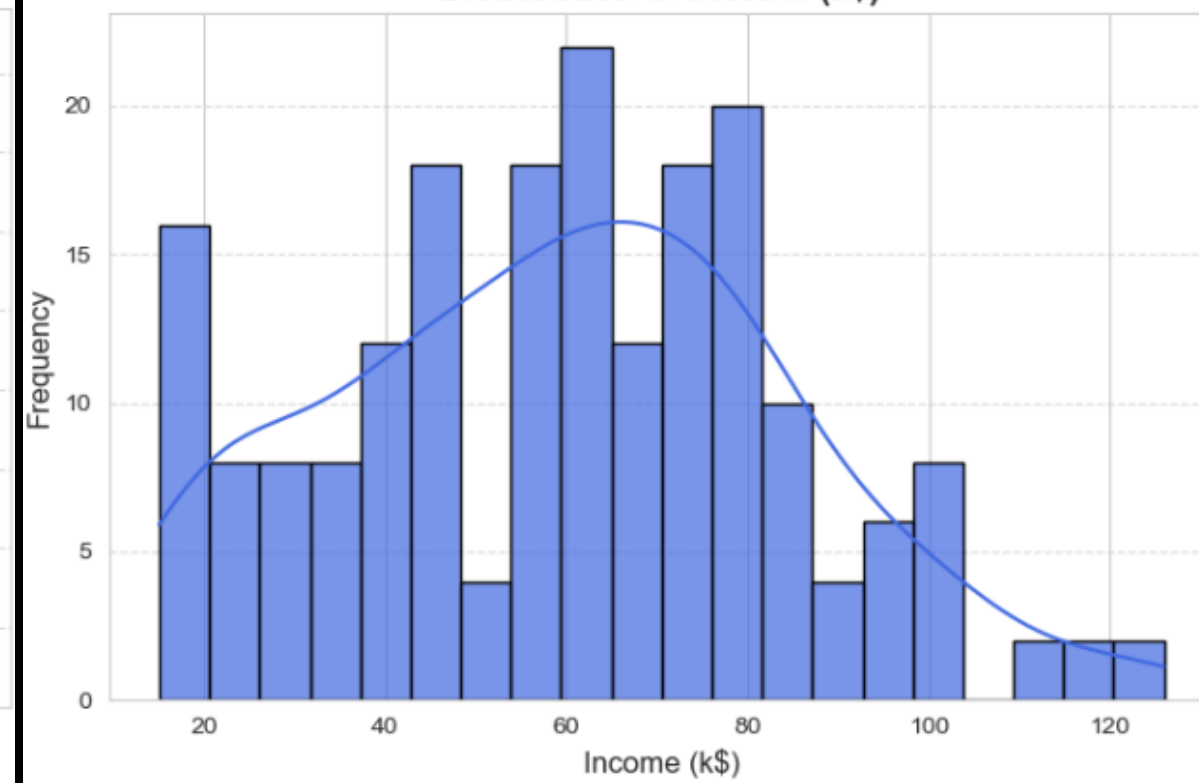


Univariate Analysis

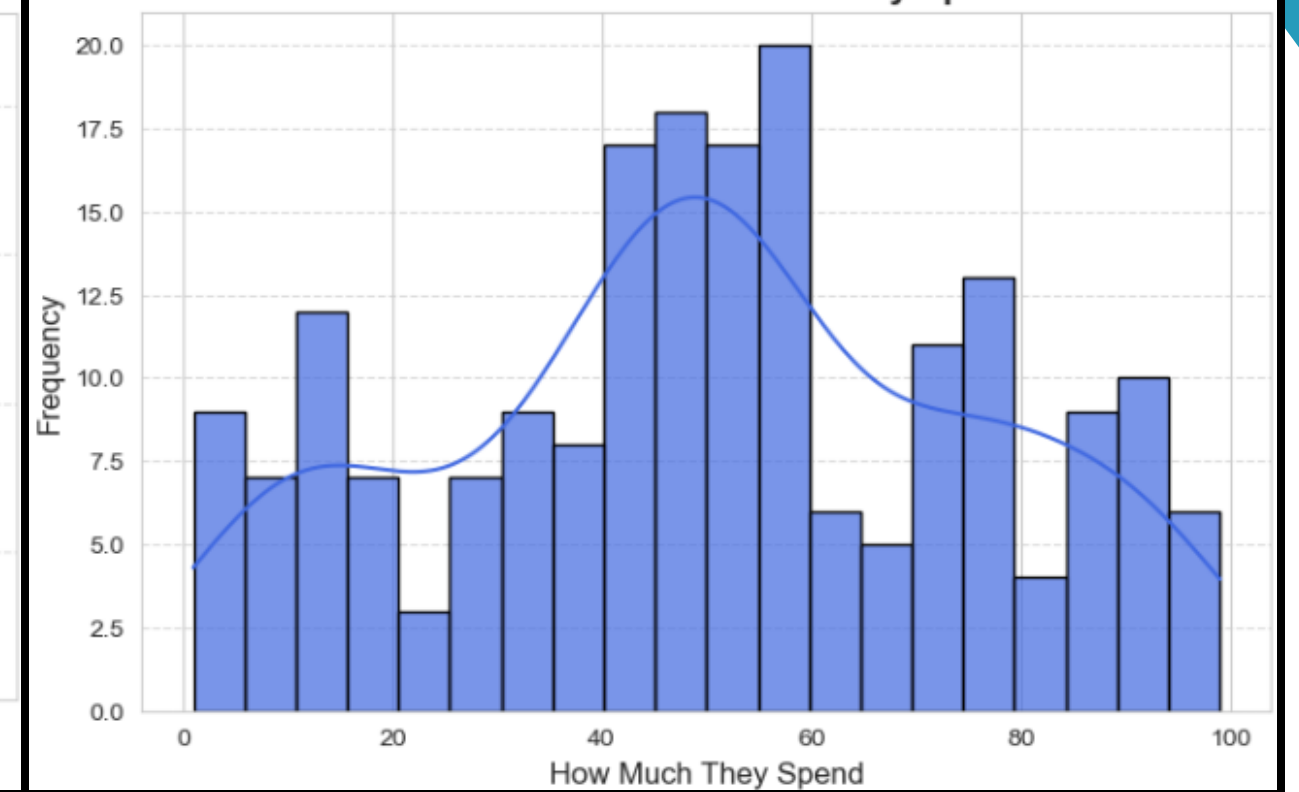
Distribution of Age



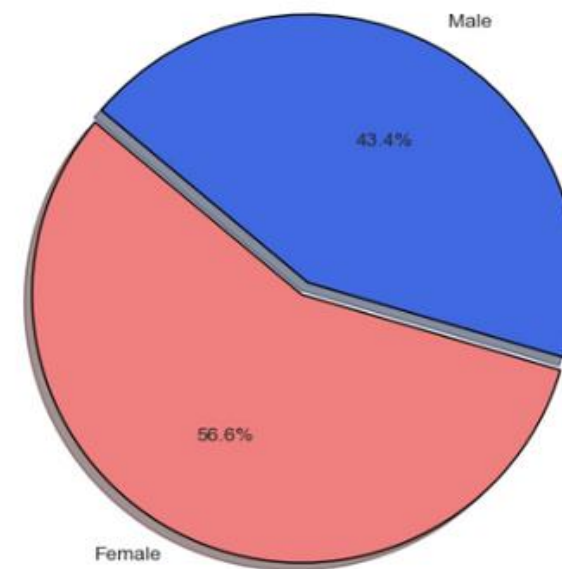
Distribution of Income (k\$)



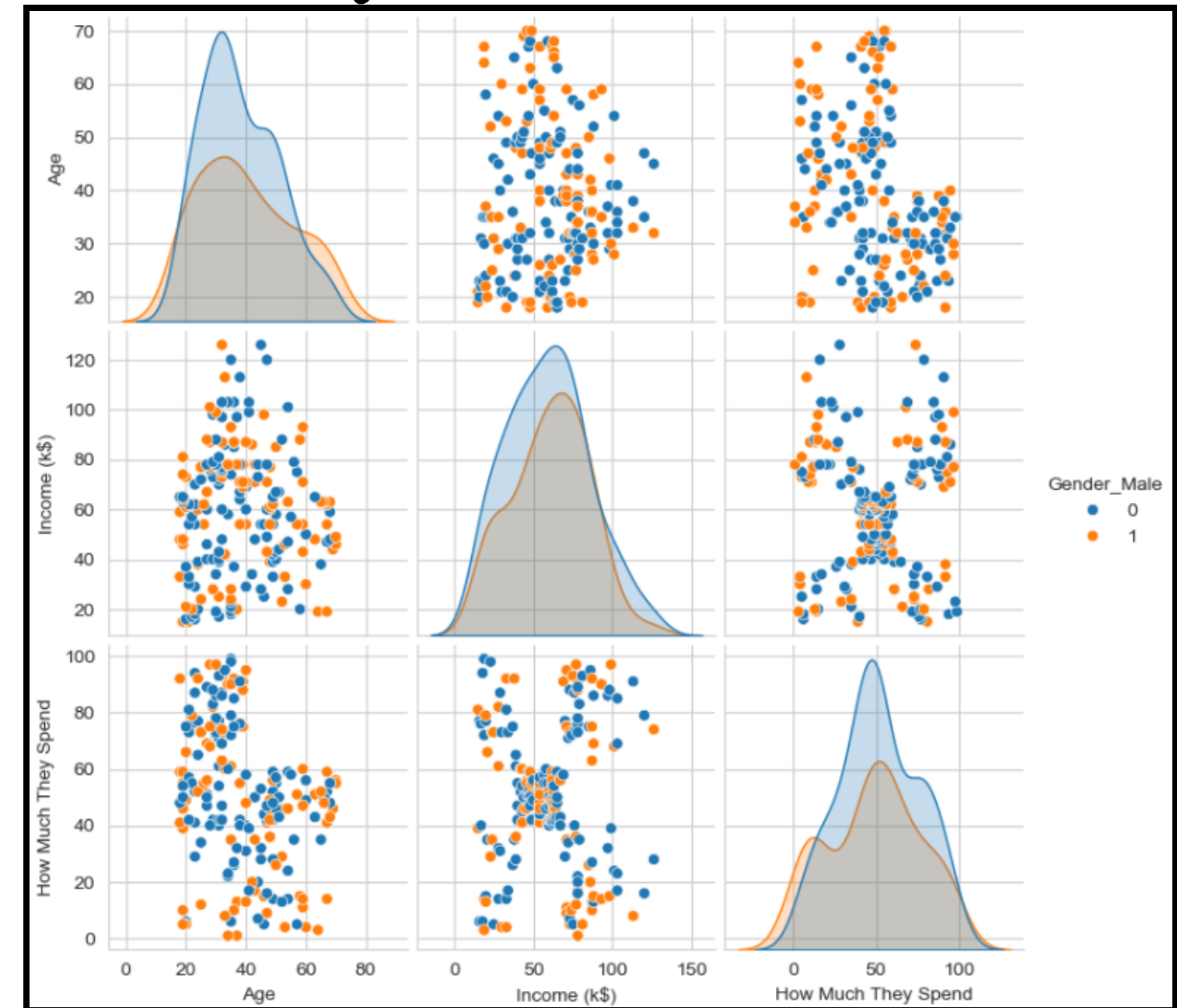
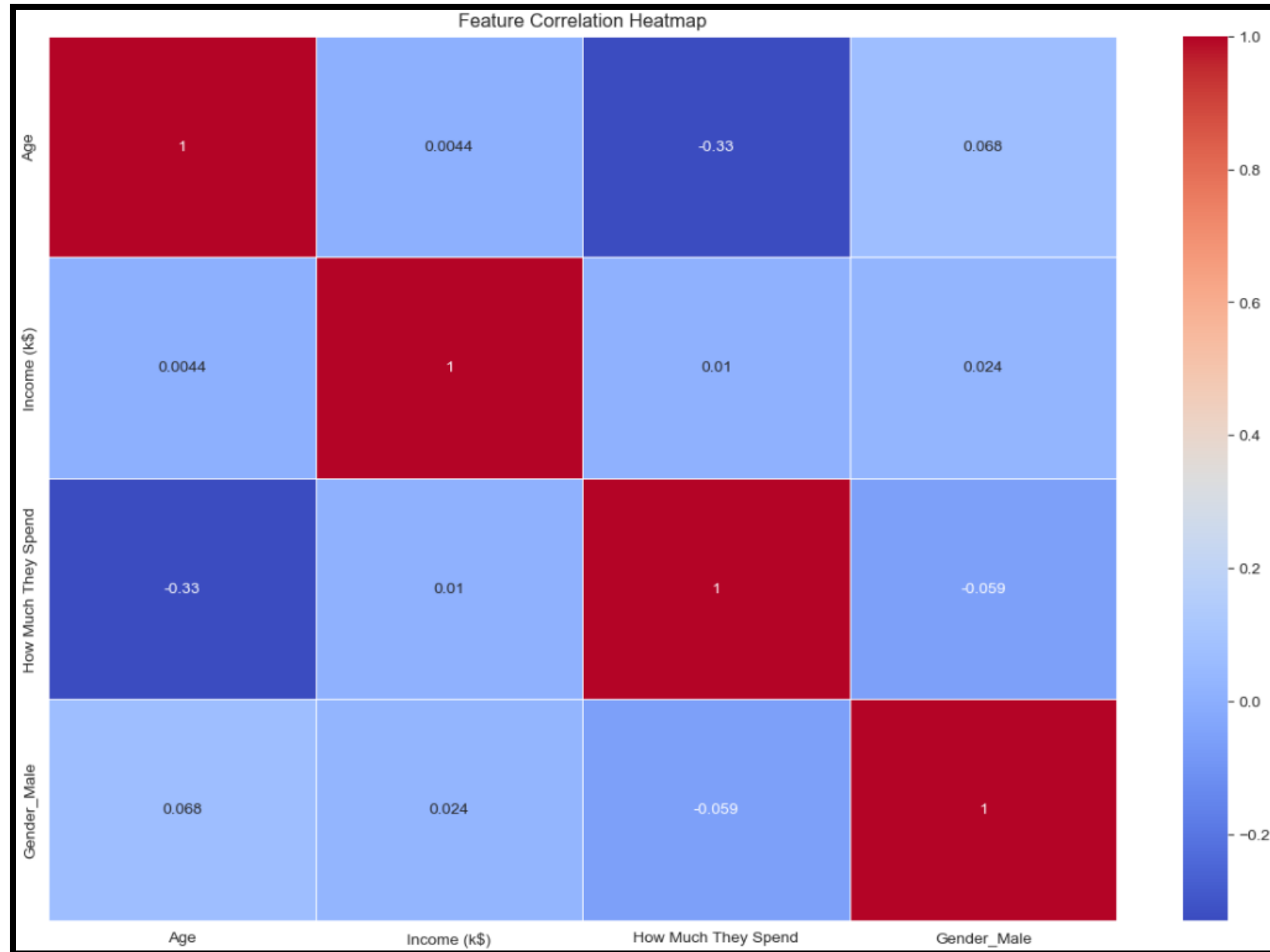
Distribution of How Much They Spend



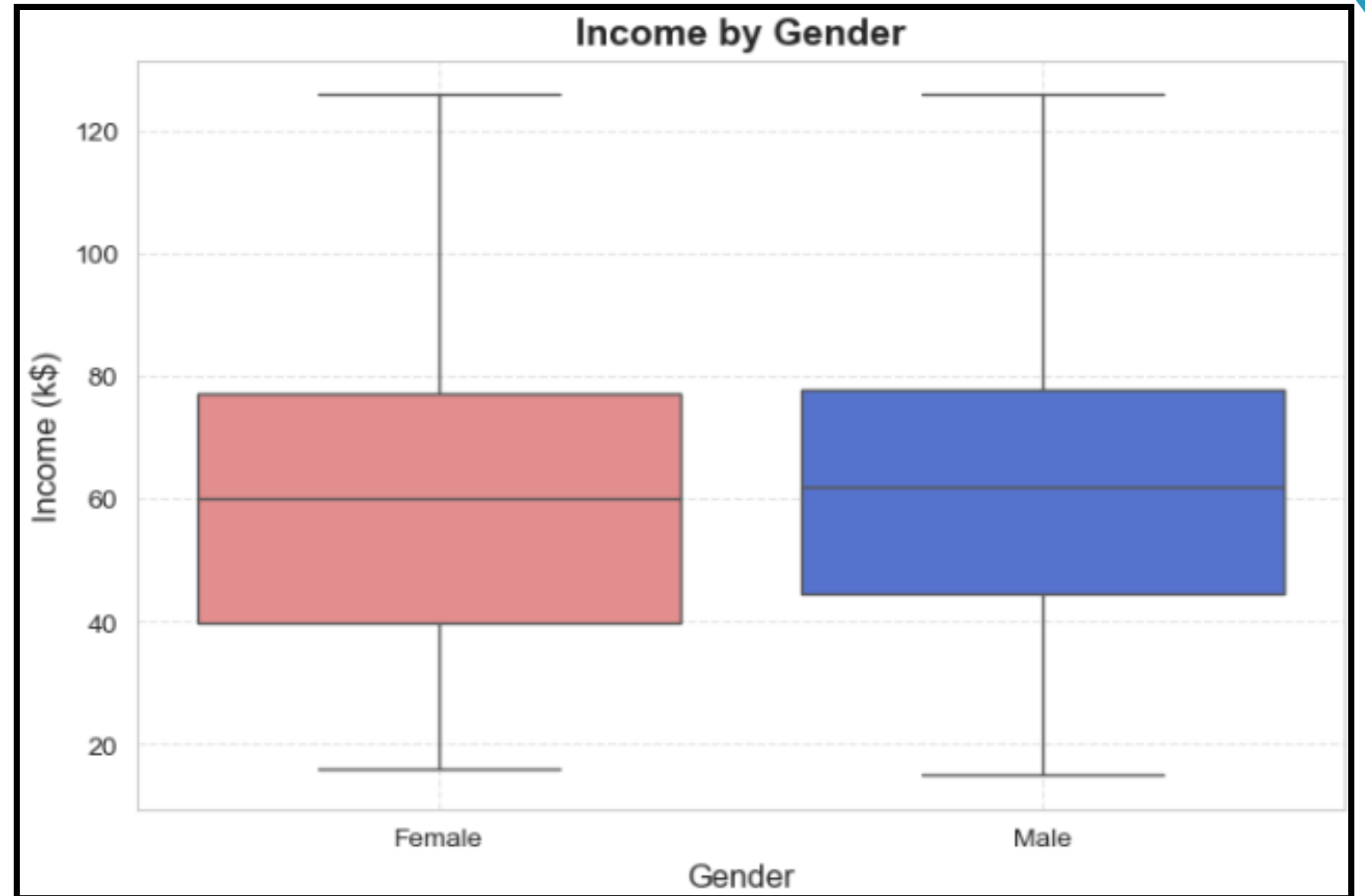
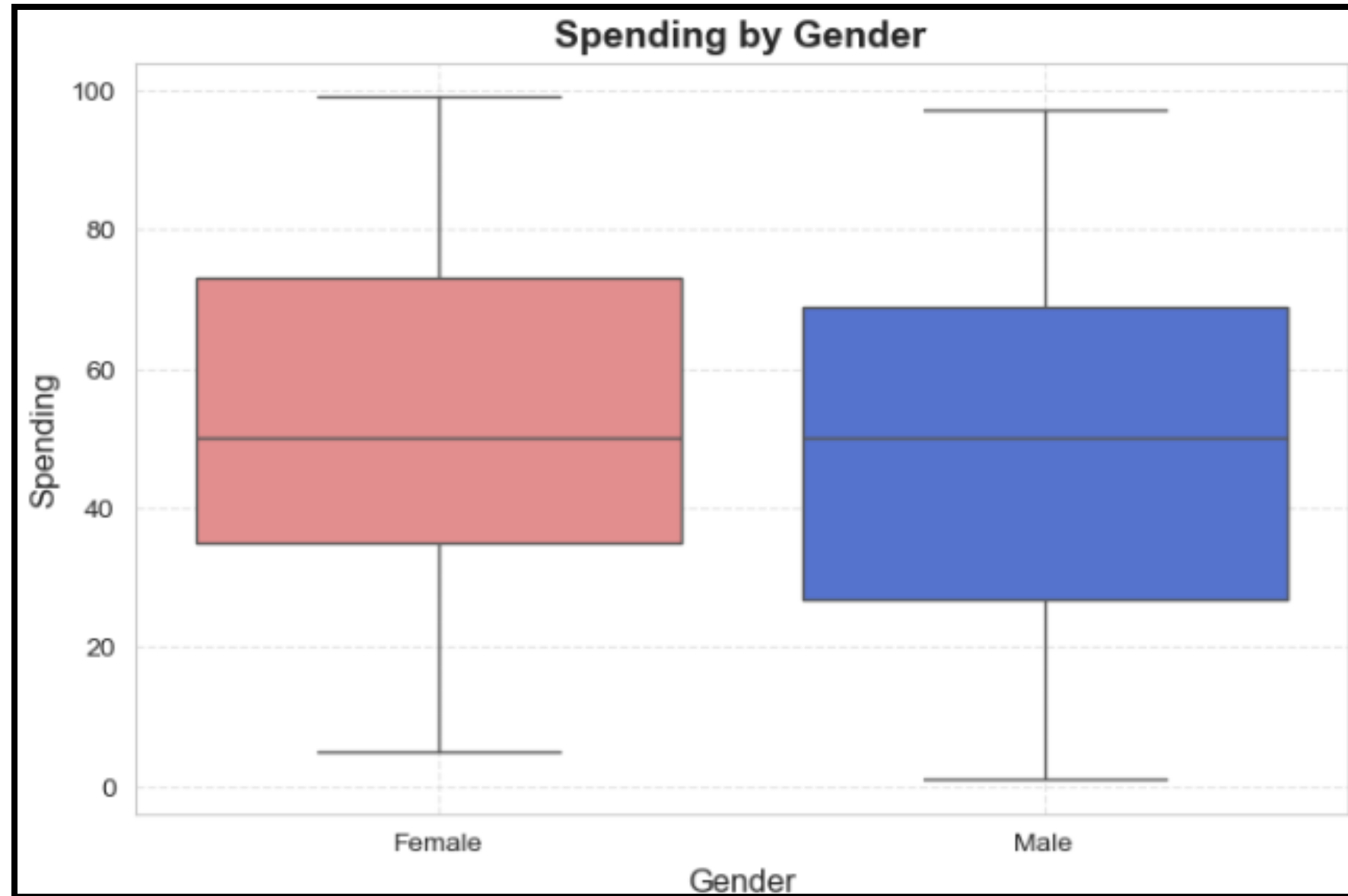
Gender Distribution



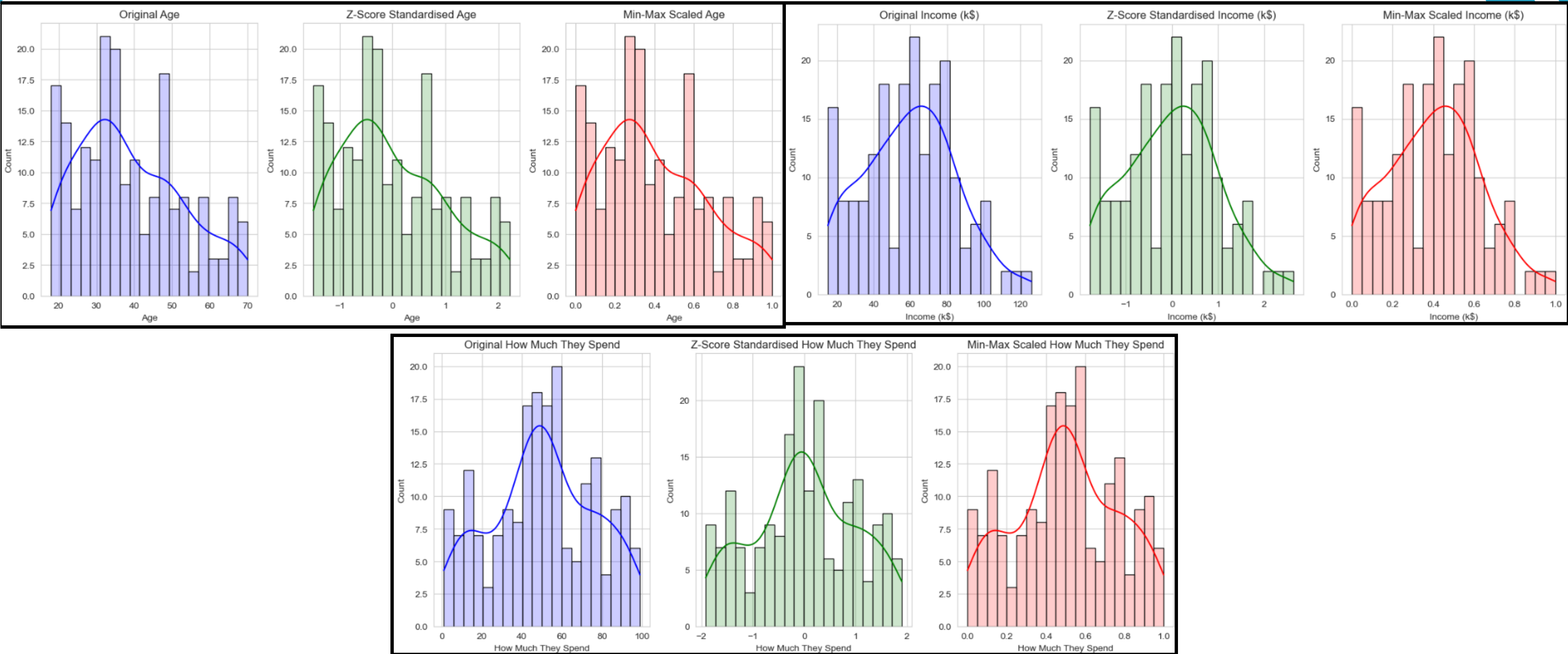
Bivariate Analysis



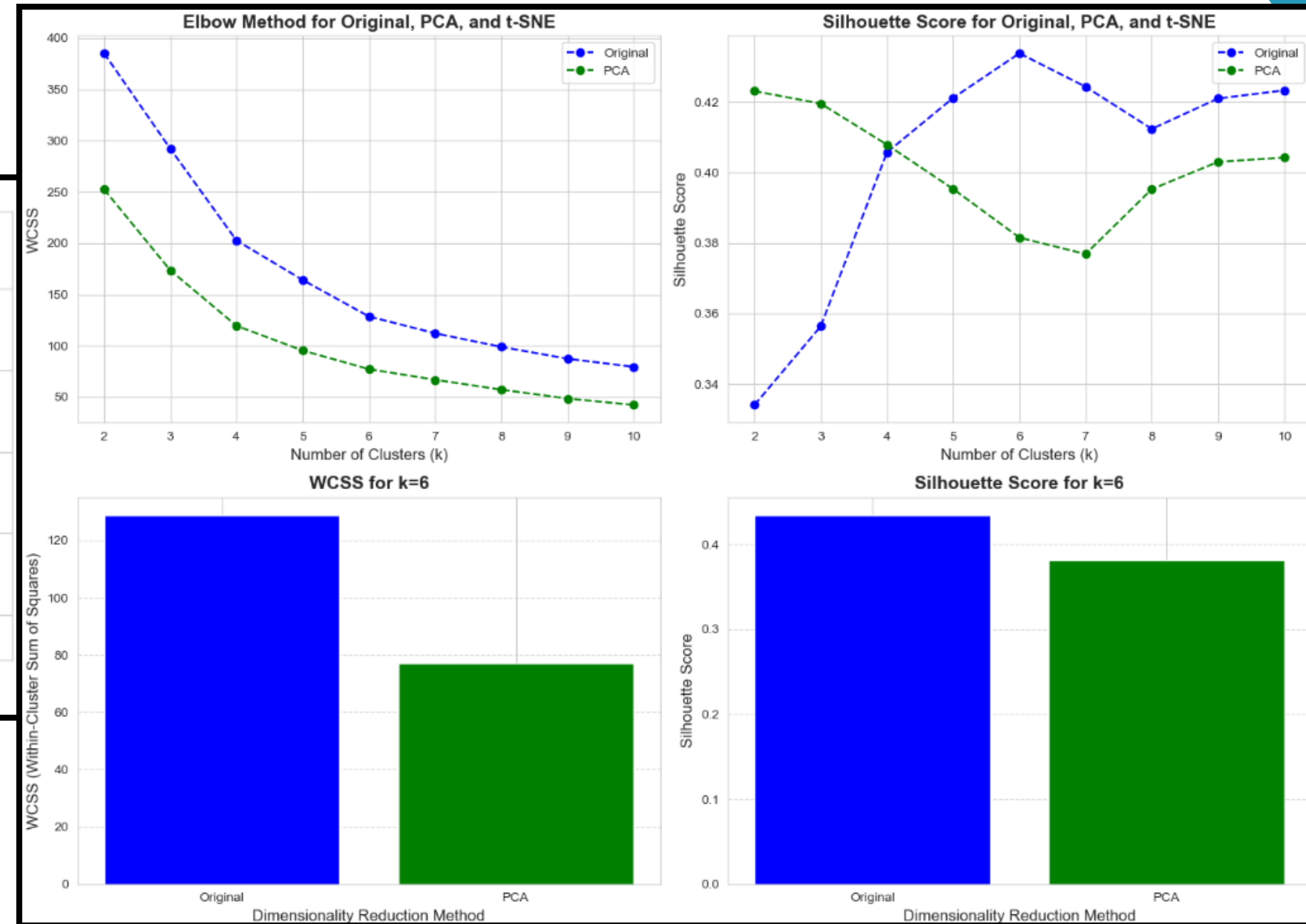
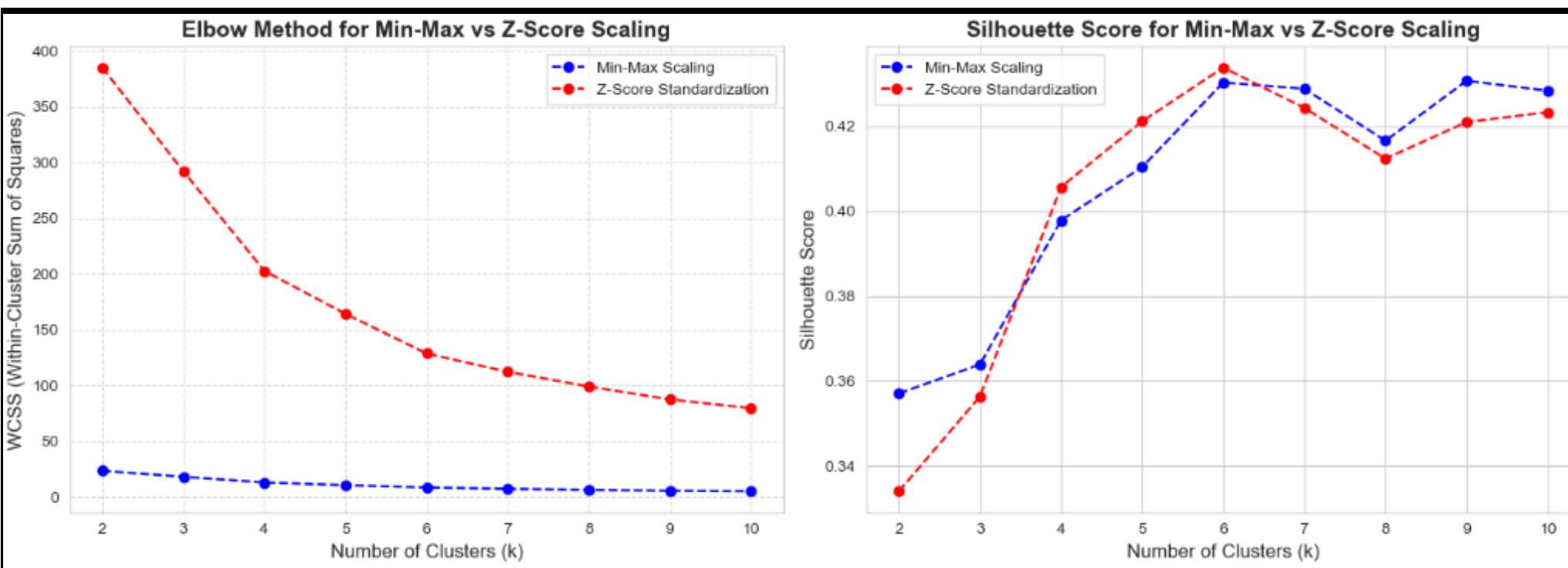
Bivariate Analysis



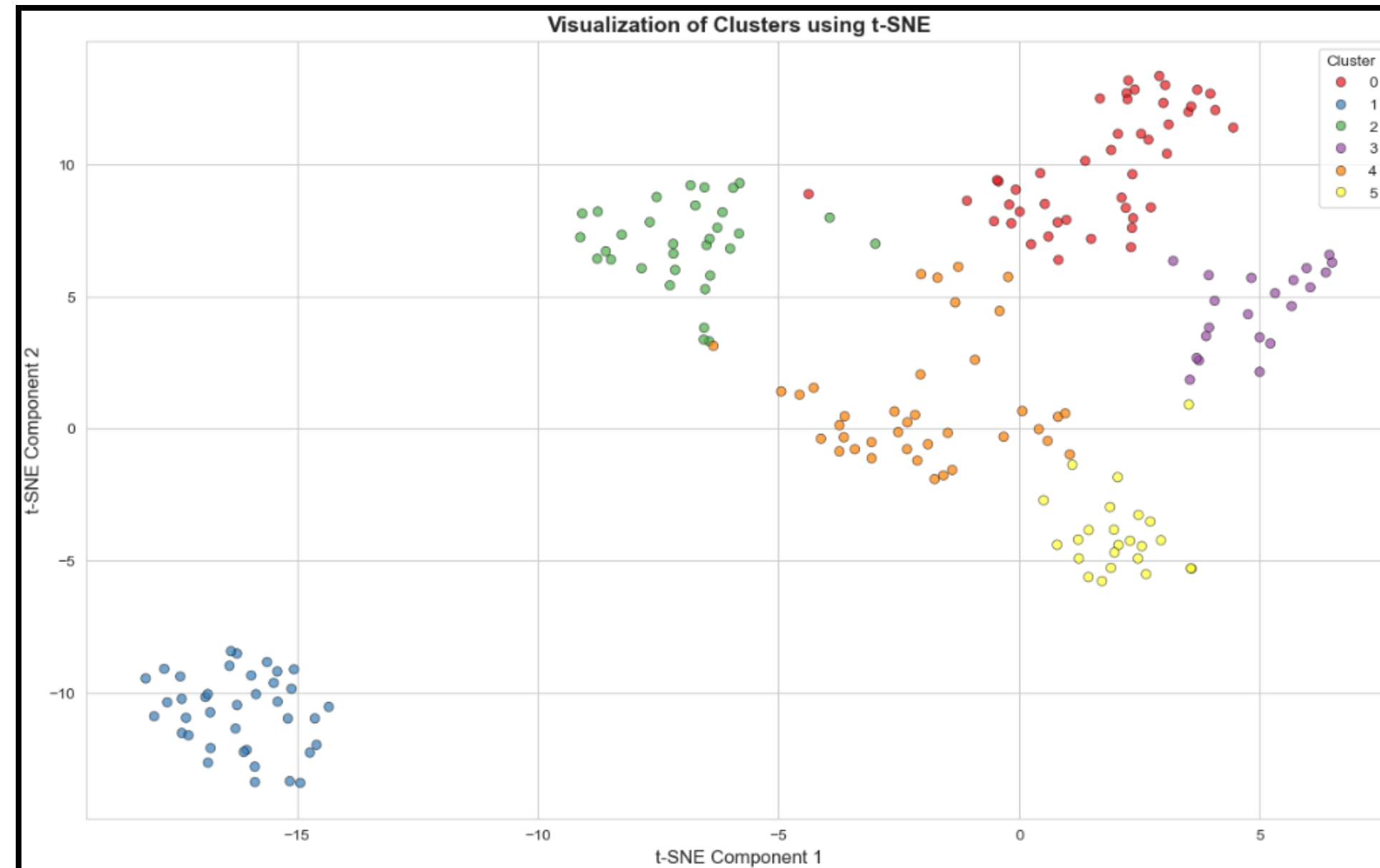
Feature Scaling



K-Means

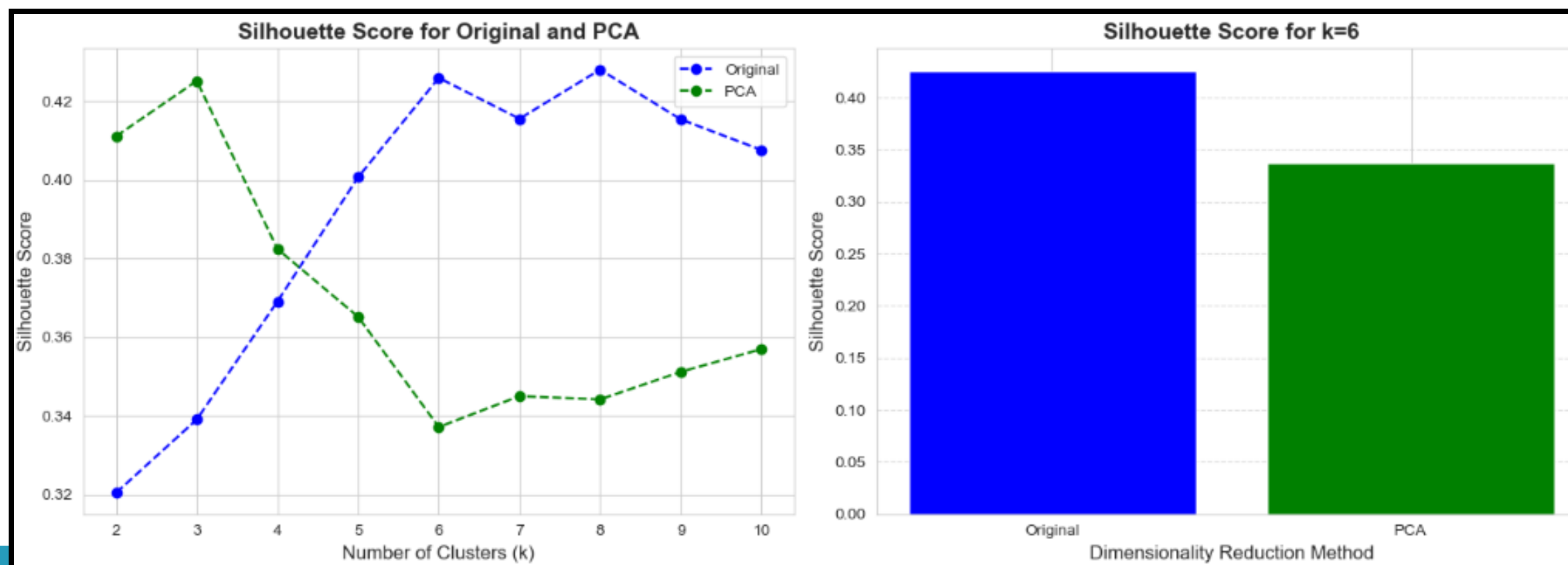
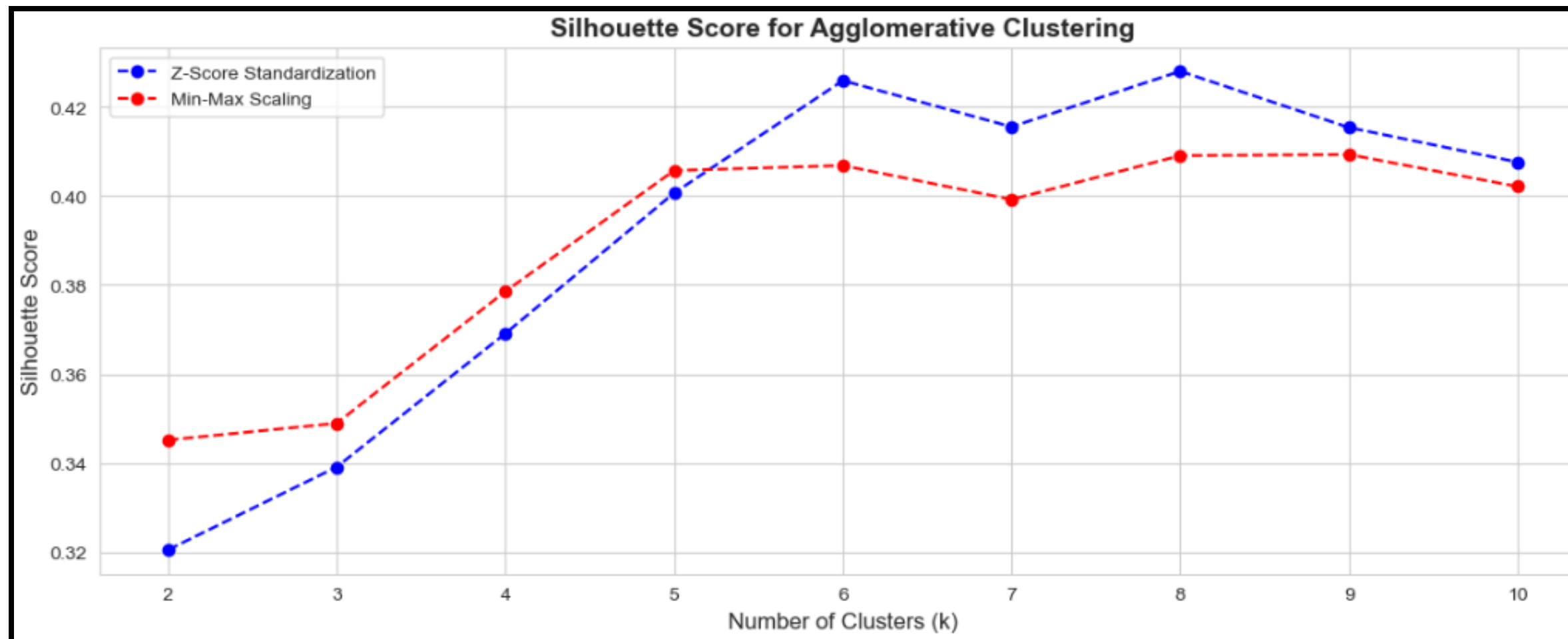


K-Means

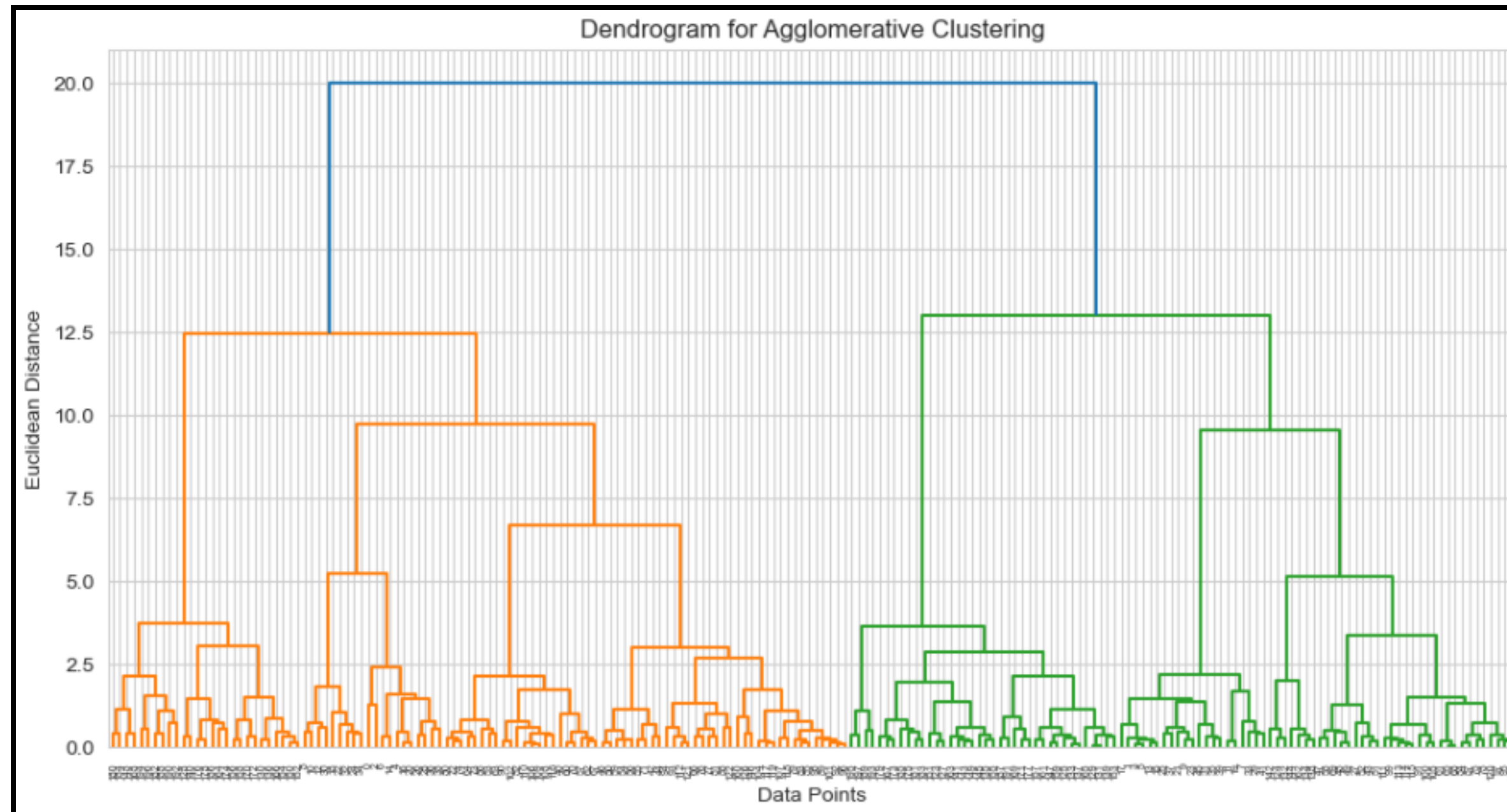


	init	max_iter	n_init	algorithm	silhouette_score	davies_bouldin_score	calinski_harabasz_score
36	random	100	10	lloyd	0.436898	0.821283	138.307770

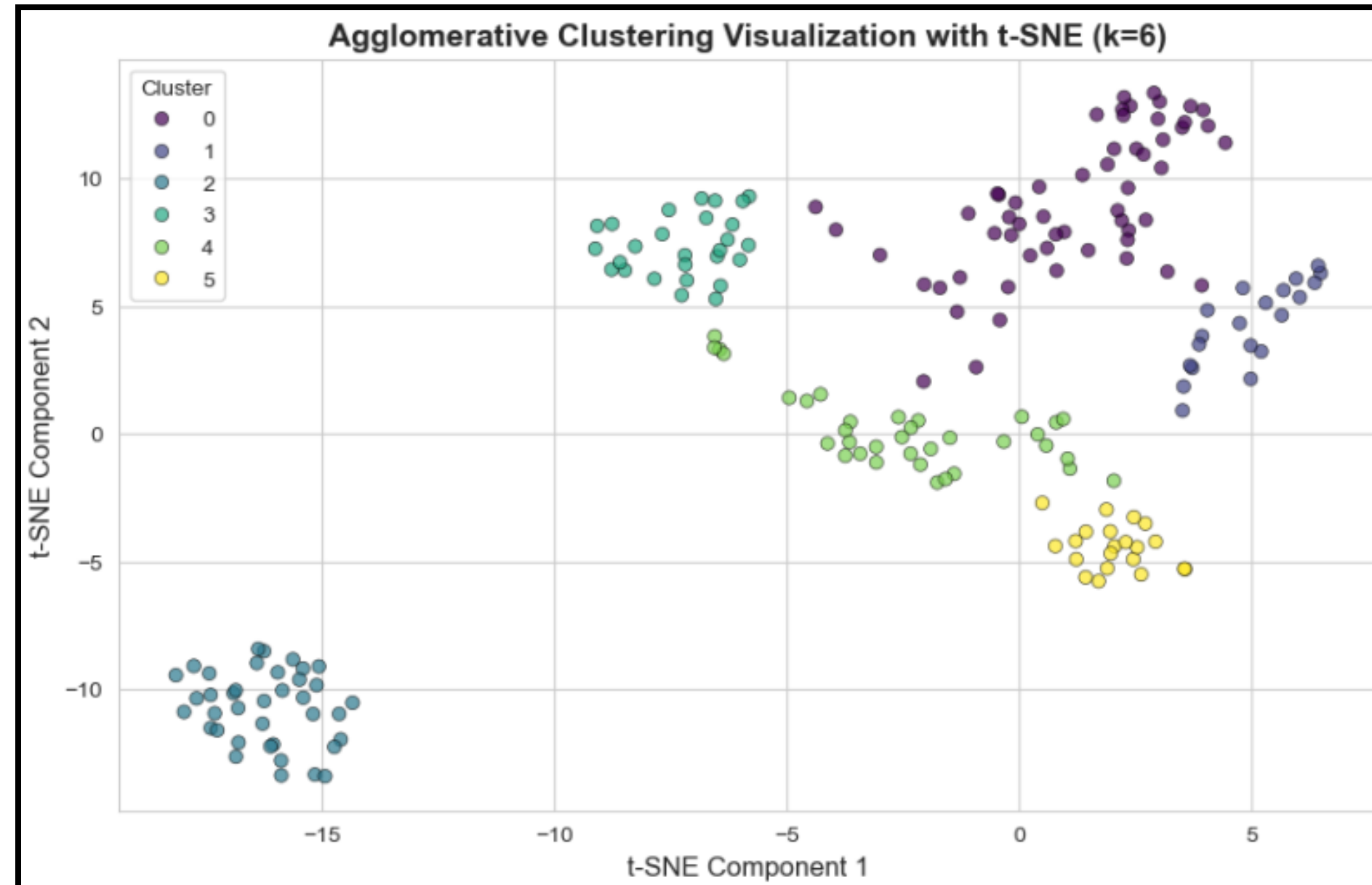
Agglomerative Clustering



Agglomerative Clustering

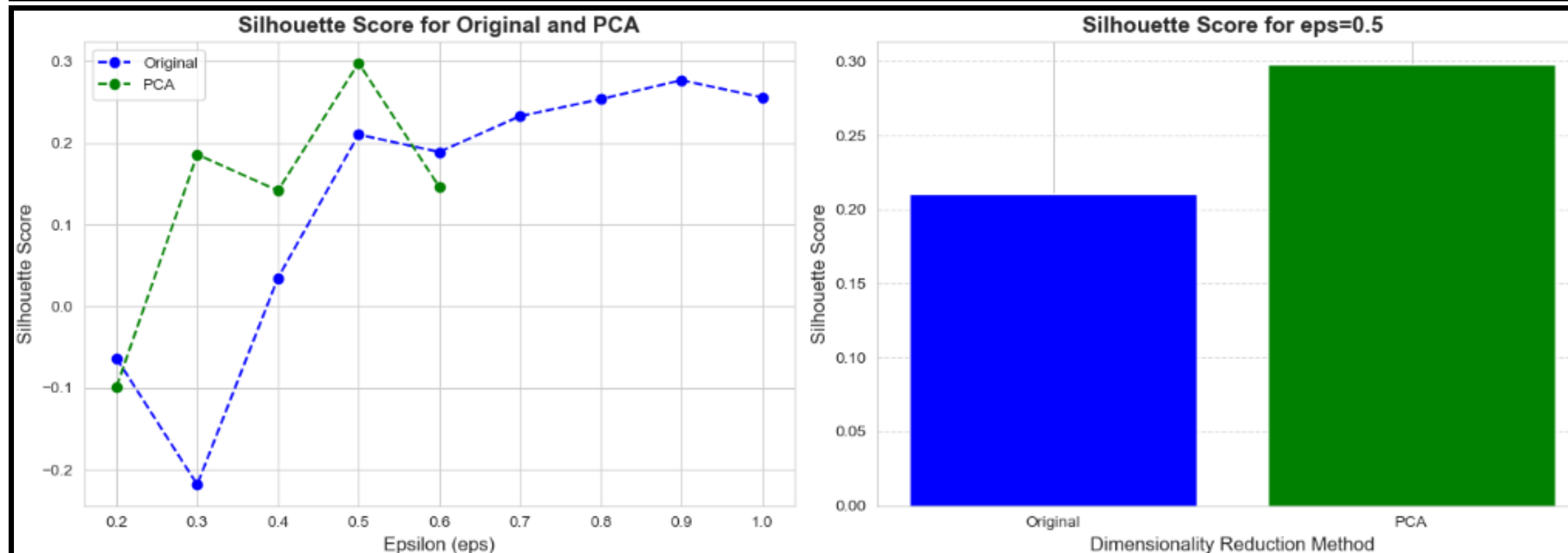
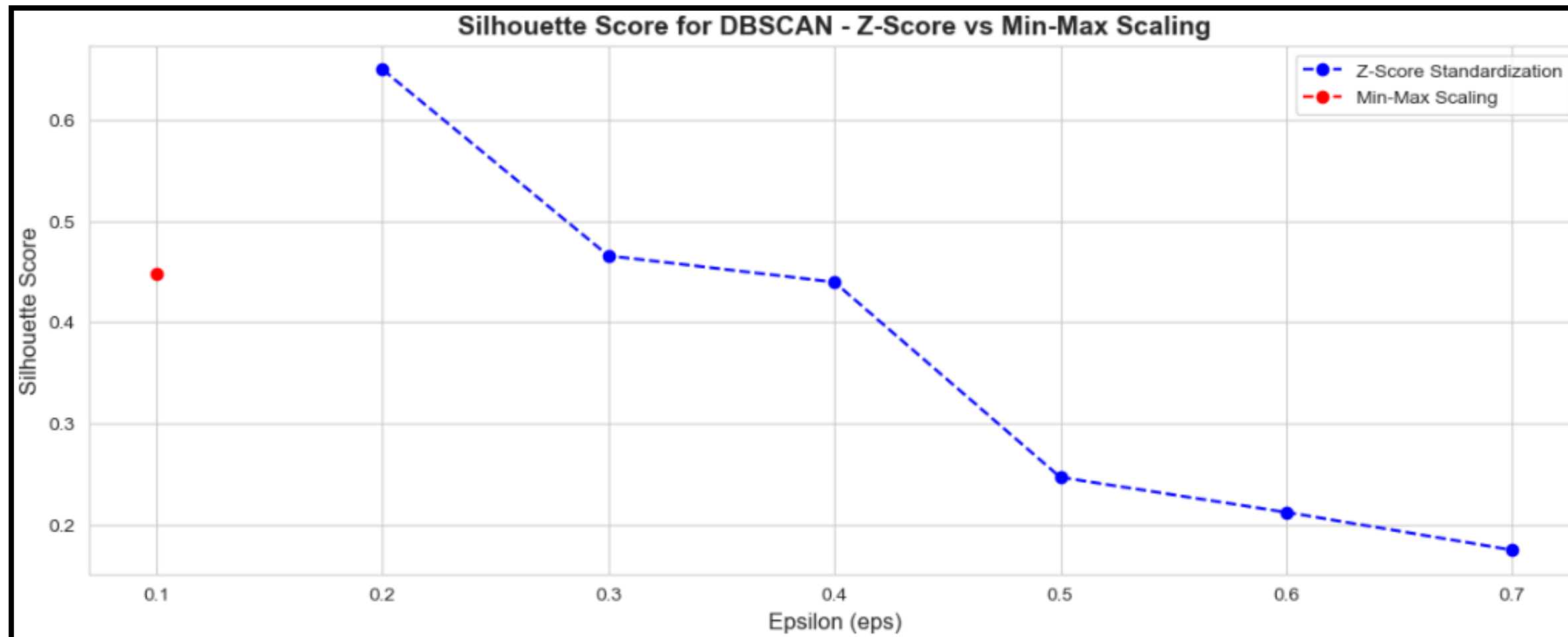


Agglomerative Clustering

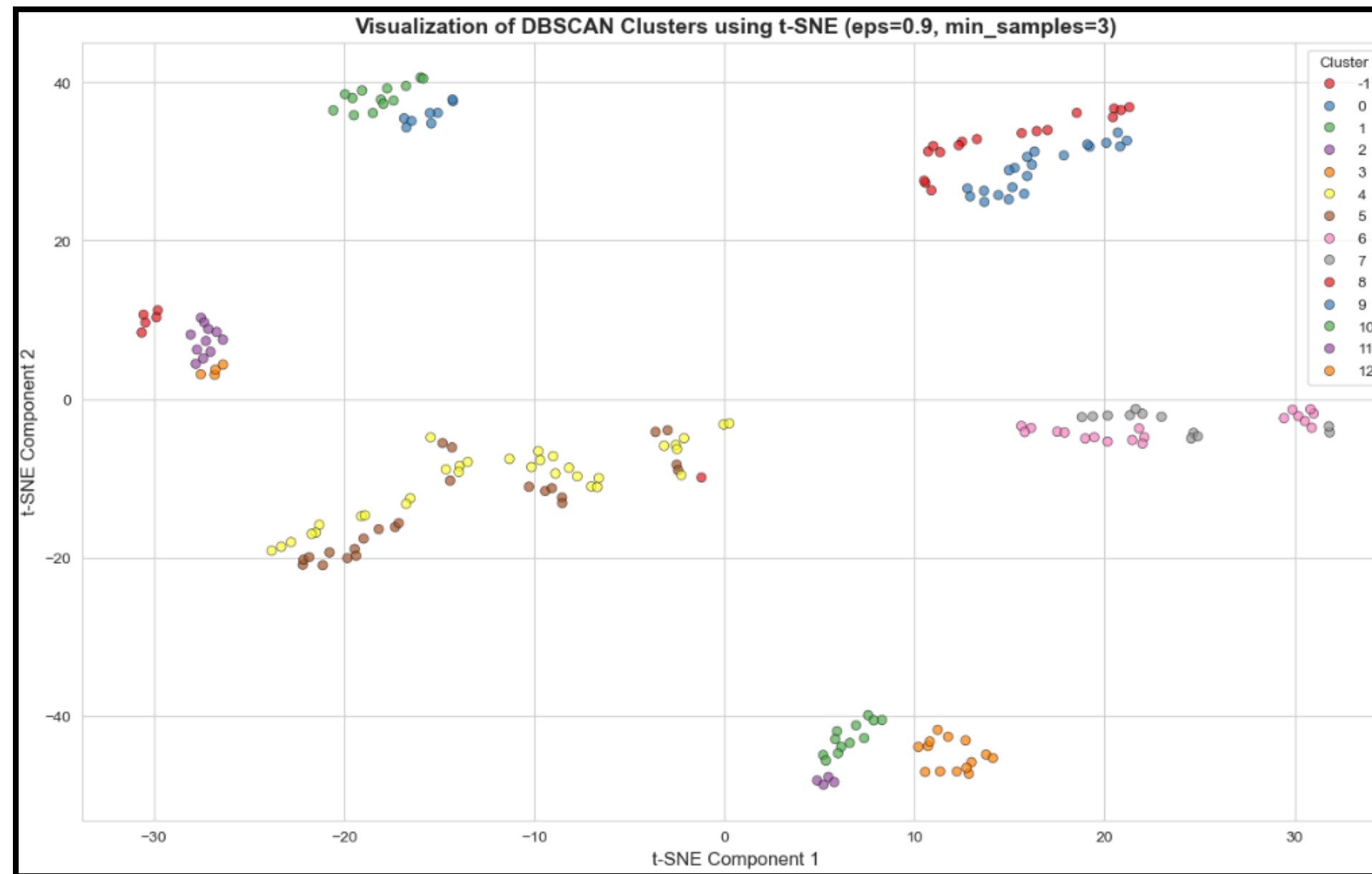


	n_clusters	linkage	compute_full_tree	silhouette_score	davies_bouldin_score	calinski_harabasz_score
0	6	ward	True	0.425948	0.844619	124.891588

DBSCAN



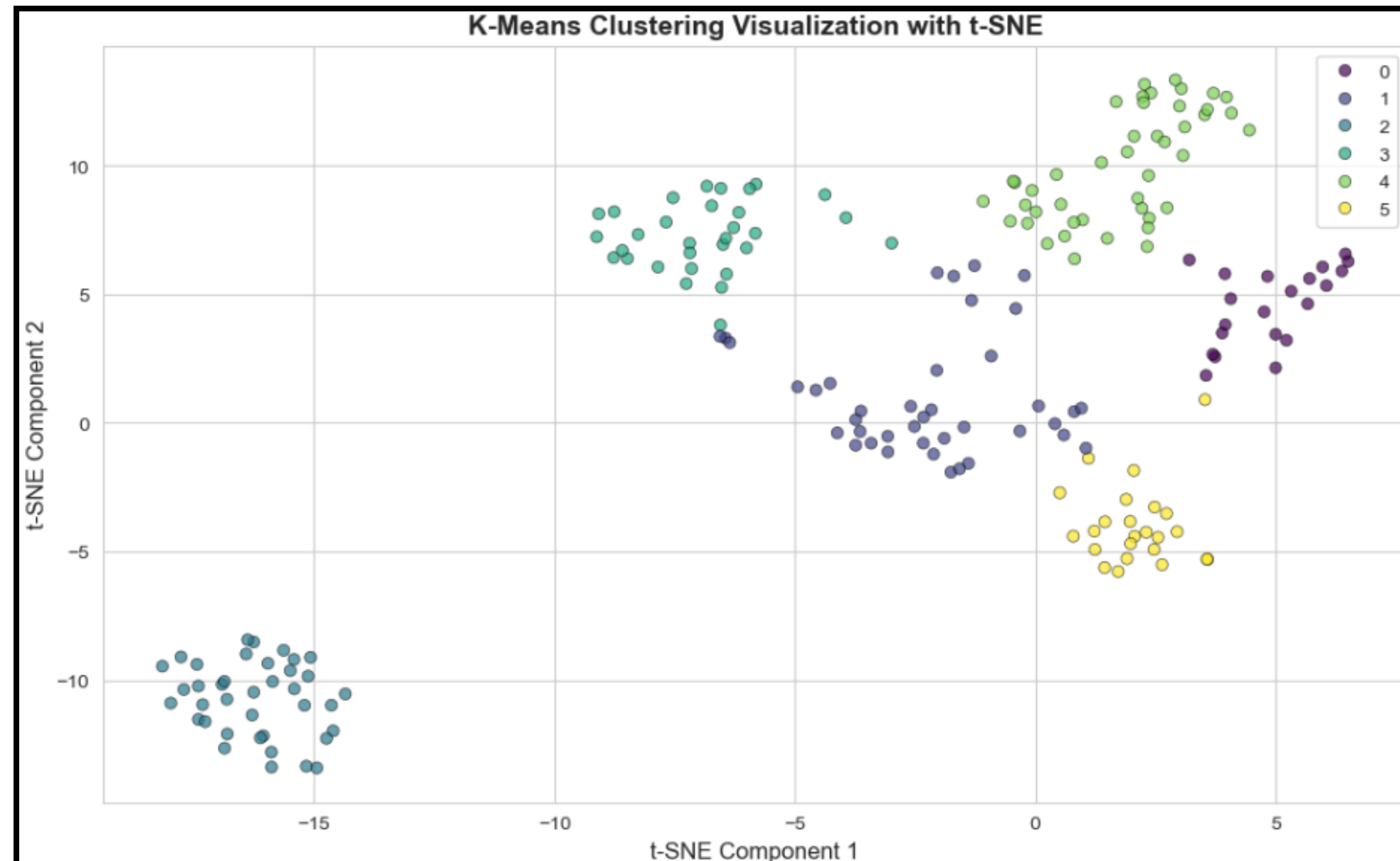
DBSCAN



	eps	min_samples	metric	algorithm	silhouette_score	davies_bouldin_score	calinski_harabasz_score
2	0.900000	7	euclidean	auto	0.012085	4.275562	56.064419

Model Evaluation

	silhouette_score	davies_bouldin_score	calinski_harabasz_score
K-Means	0.436898	0.821283	138.307770
Agglomerative Clustering	0.425948	0.844619	124.891588
DBSCAN	0.012085	4.275562	56.064419



Saving of Model

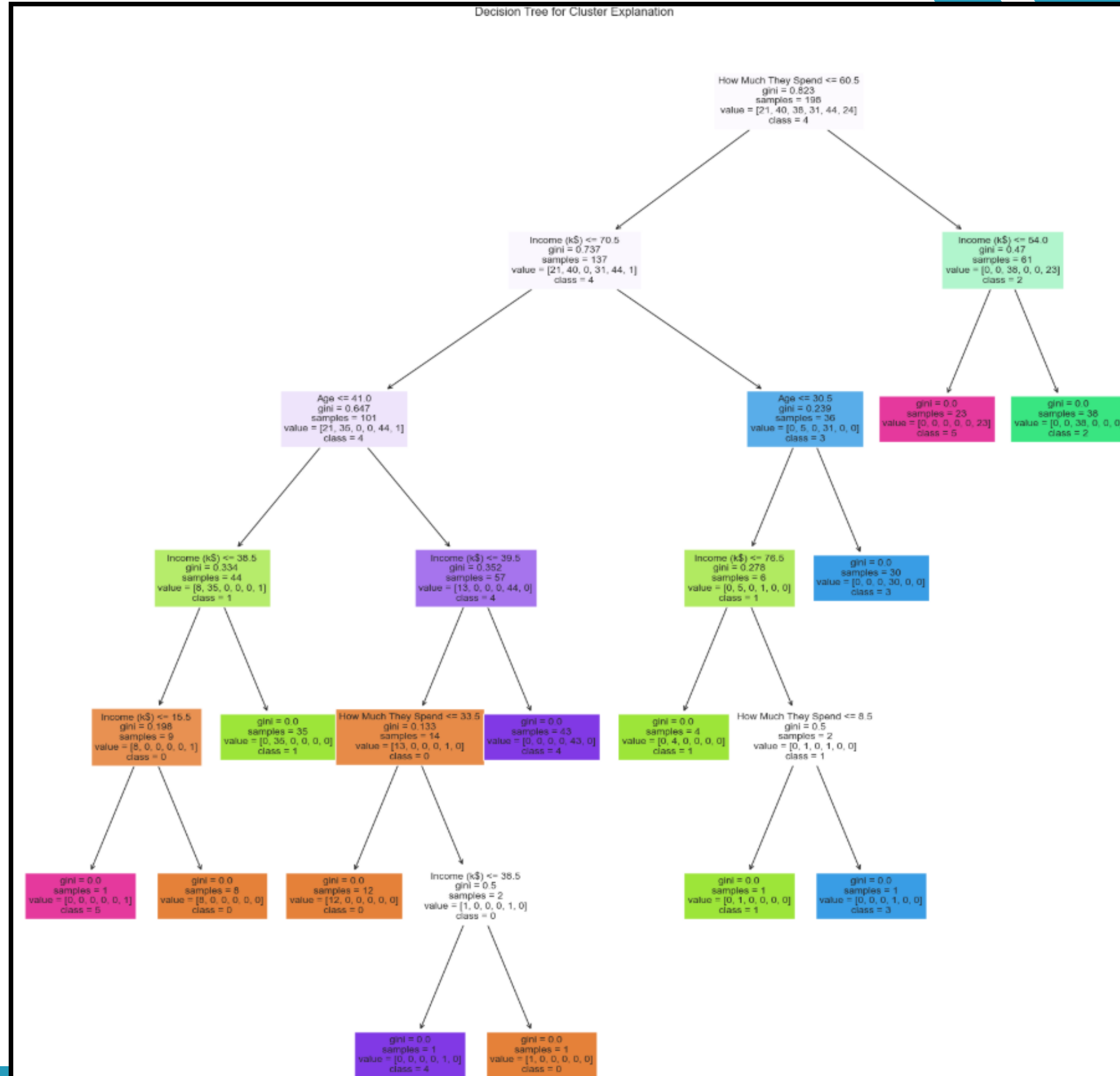
```
# ----- Saving Of Model ----- #  
joblib.dump(kmeans, 'CA2-Saved-Models/kmeans_model.pkl')  
  
# ----- Confirmation Message ----- #  
print("Model Saved Successfully")  
Model Saved Successfully
```

Interpretation

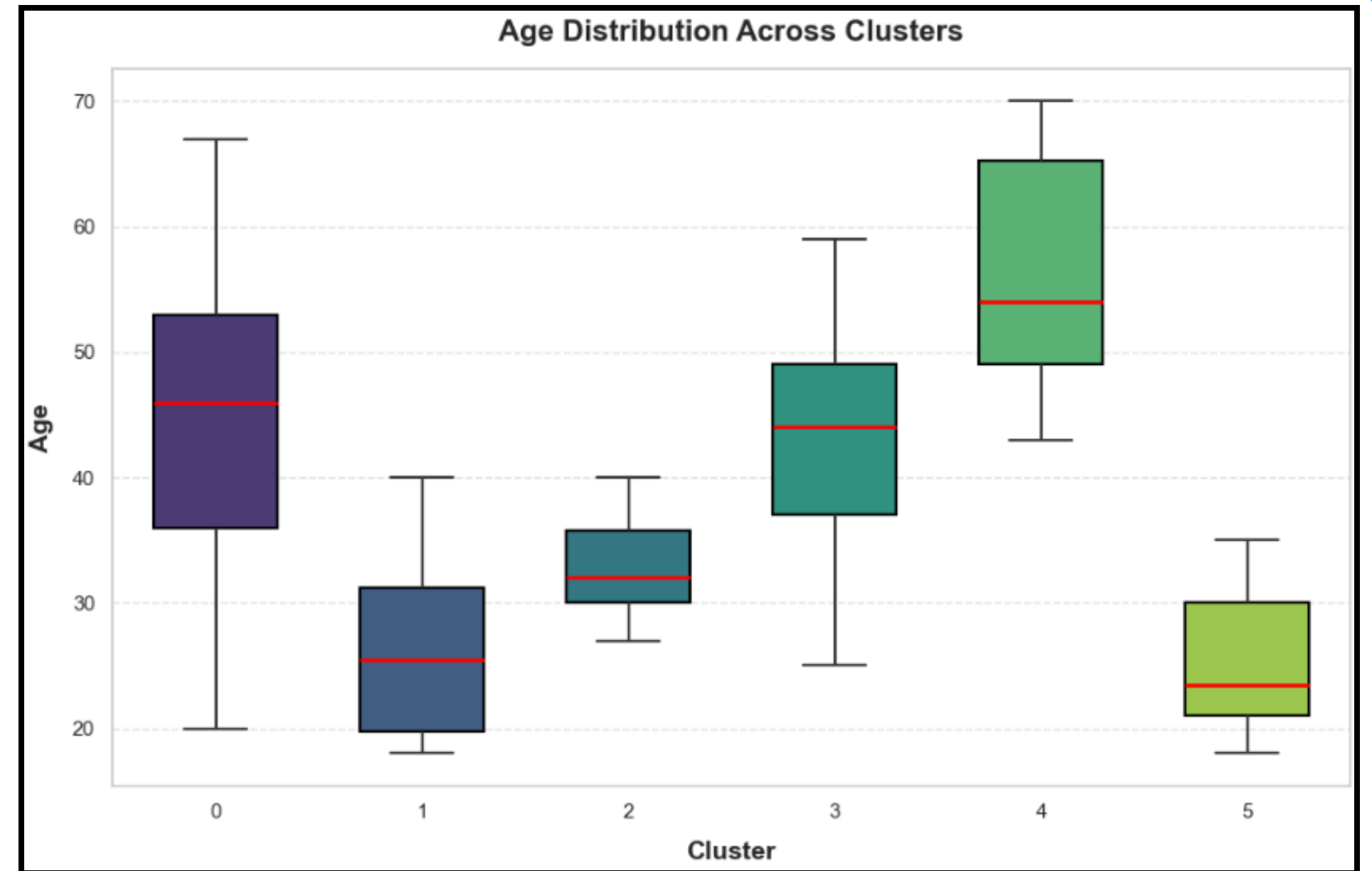
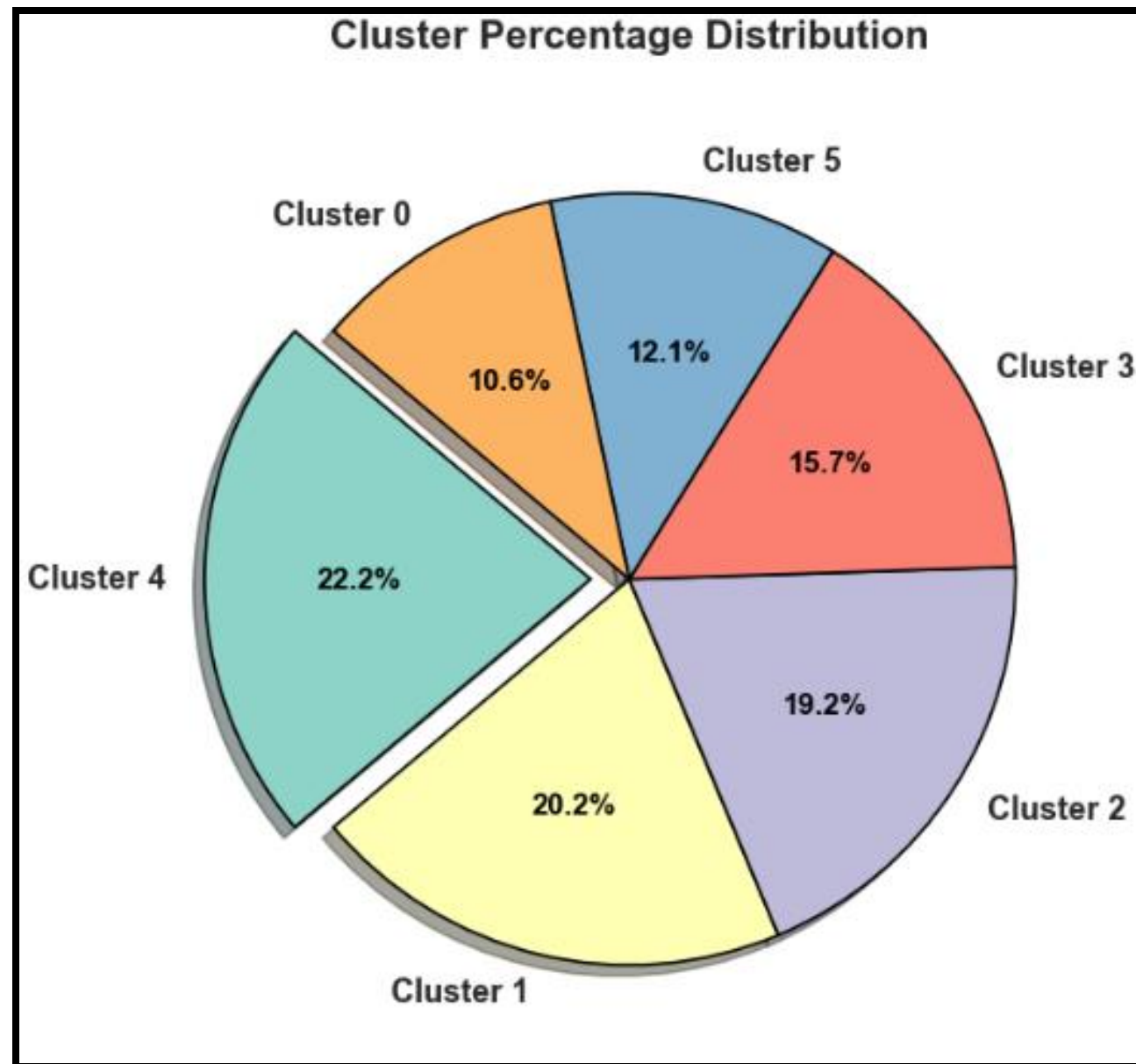
Reverse Scaling

	Age	Income (k\$)	How Much They Spend	Gender_Male	Cluster
0	19.000000	15.000000	39.000000	1	5
1	21.000000	15.000000	81.000000	1	5
2	20.000000	16.000000	6.000000	0	0
3	23.000000	16.000000	77.000000	0	5
4	31.000000	17.000000	40.000000	0	0

- Cluster 0: Lowest Income-Group With Lowest Spending
- Cluster 1: Younger Demographic With Moderate Income With Moderate Spending
- Cluster 2: High Income And High Spending
- Cluster 3: High Income Group With Low Spending
- Cluster 4: Older Demographic With Moderate Income And Moderate Spending
- Cluster 5: Low Income With High Spending

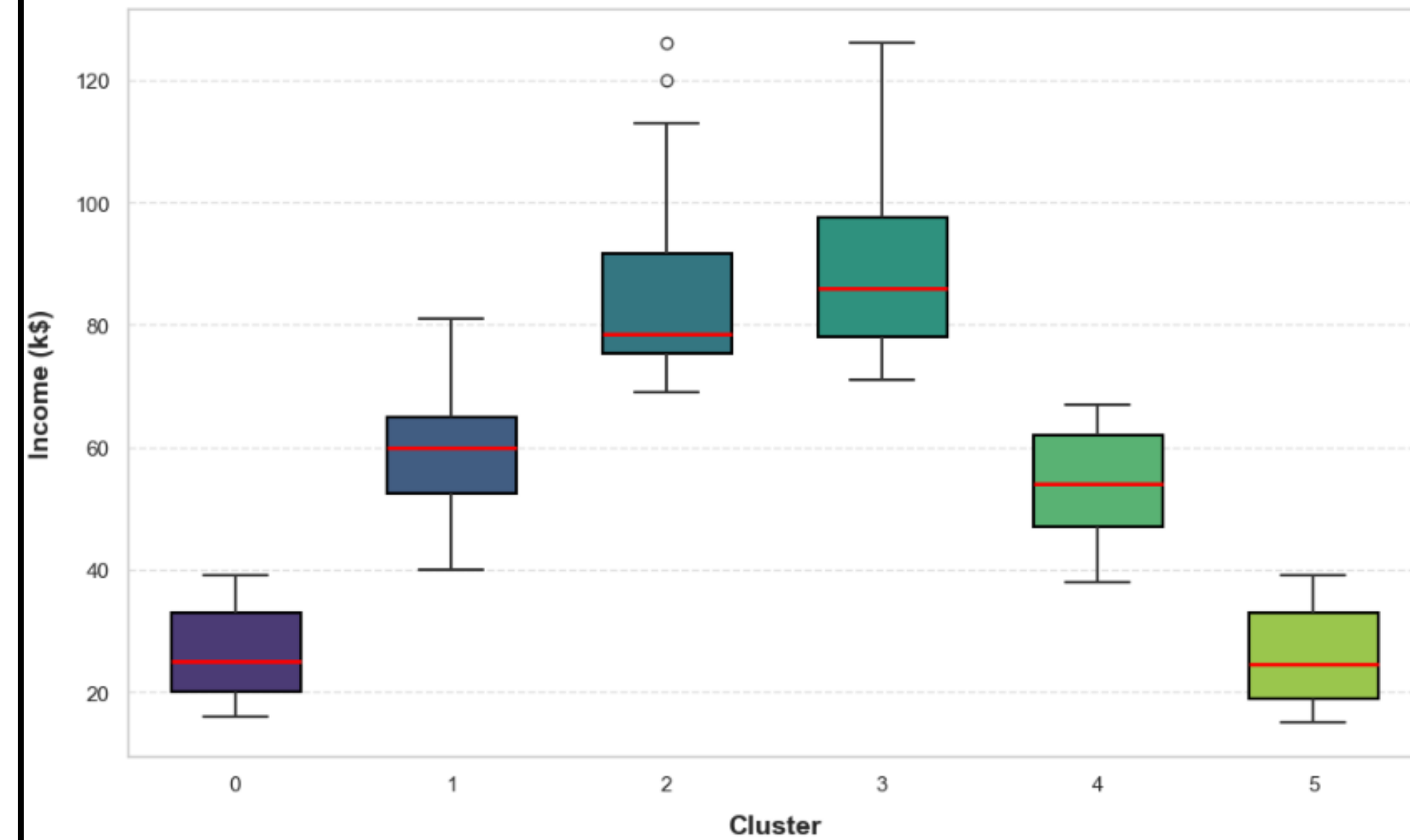


Implications

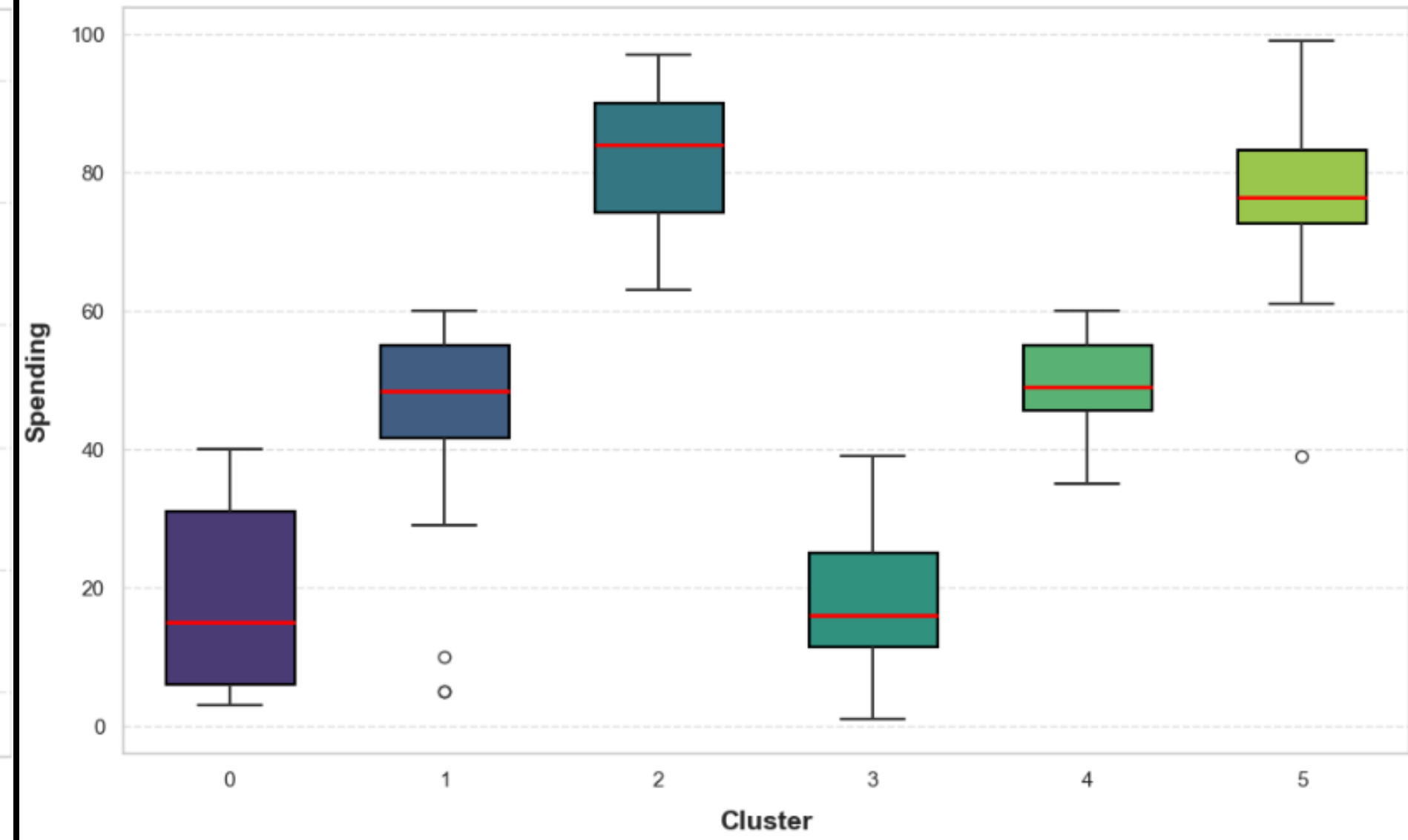


Implications

Income Distribution Across Clusters



Spending Distribution Across Clusters





AIML CA 2

THANK YOU
FOR YOUR ATTENTION

Goh Kun Ming