RATIONAL STATEMENT

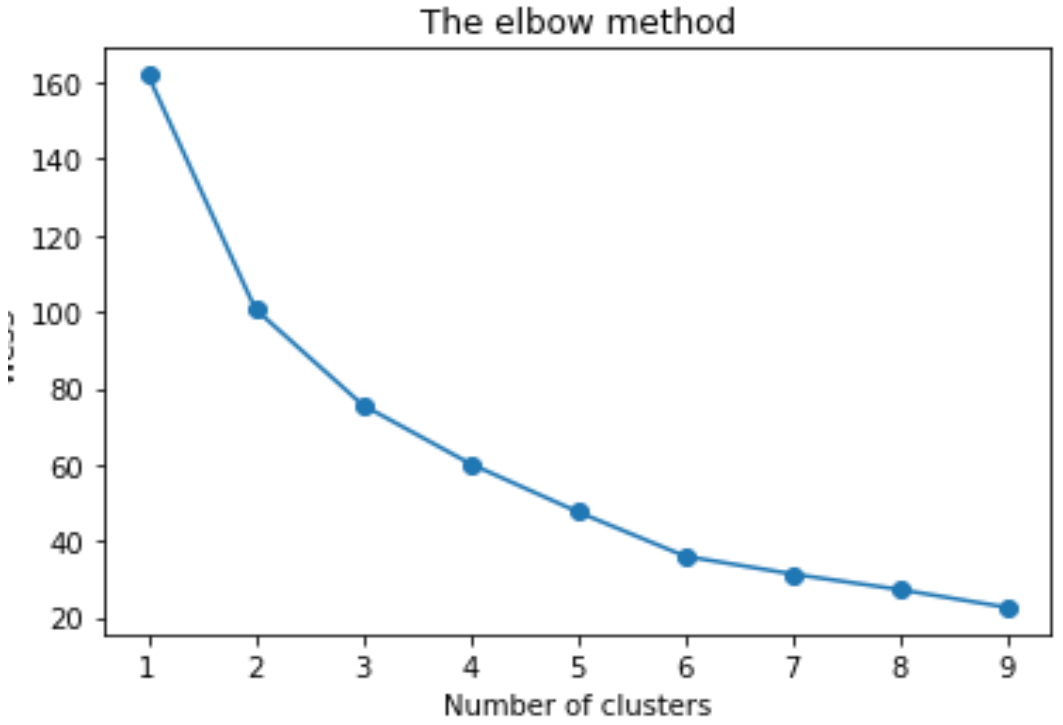
- To understand how clustering techniques are used for K-Means model. Using the ‘leuanalysis’ dataset to carryout this experiment.

Identify and Justify the number of clusters required to optimize a K-means Model by using both the Elbow and Silhouette Methods.

- ELBOW METHOD

The number of clusters needed to optimize a K-means Model is (6), as can be seen from the plotting, where the slope takes a bend knee to indicate.

Last point, possible overlap or non-spherical clusters

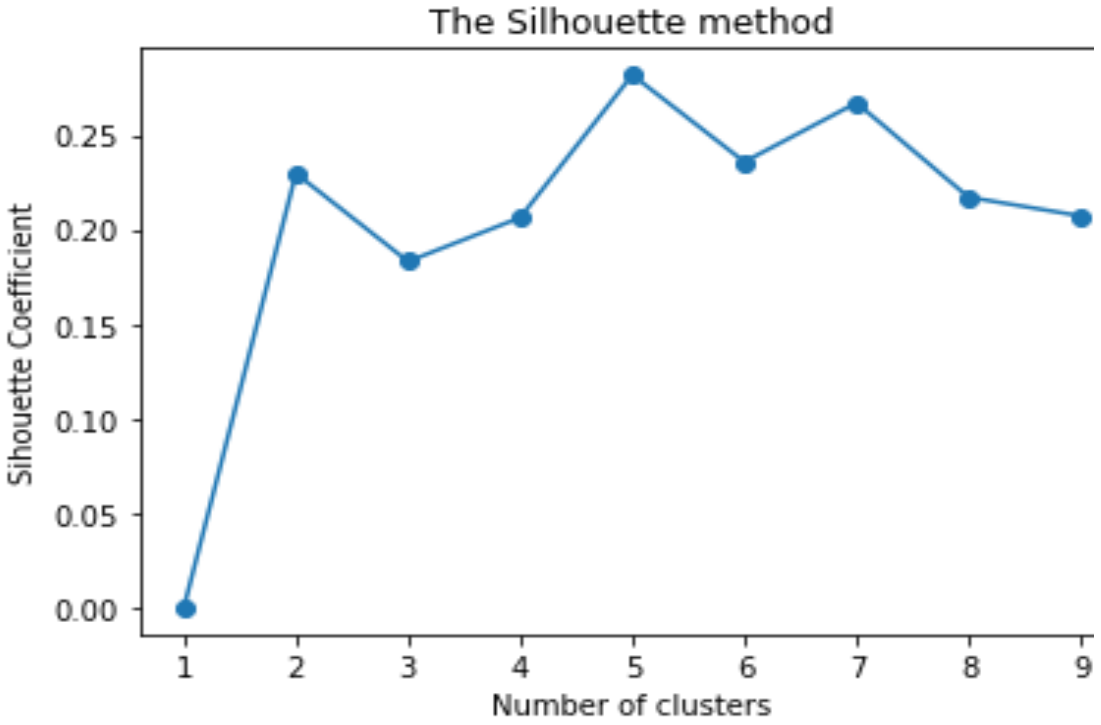


Identify and Justify the number of clusters required to optimize a K-means Model by using both the Elbow and Silhouette Methods.

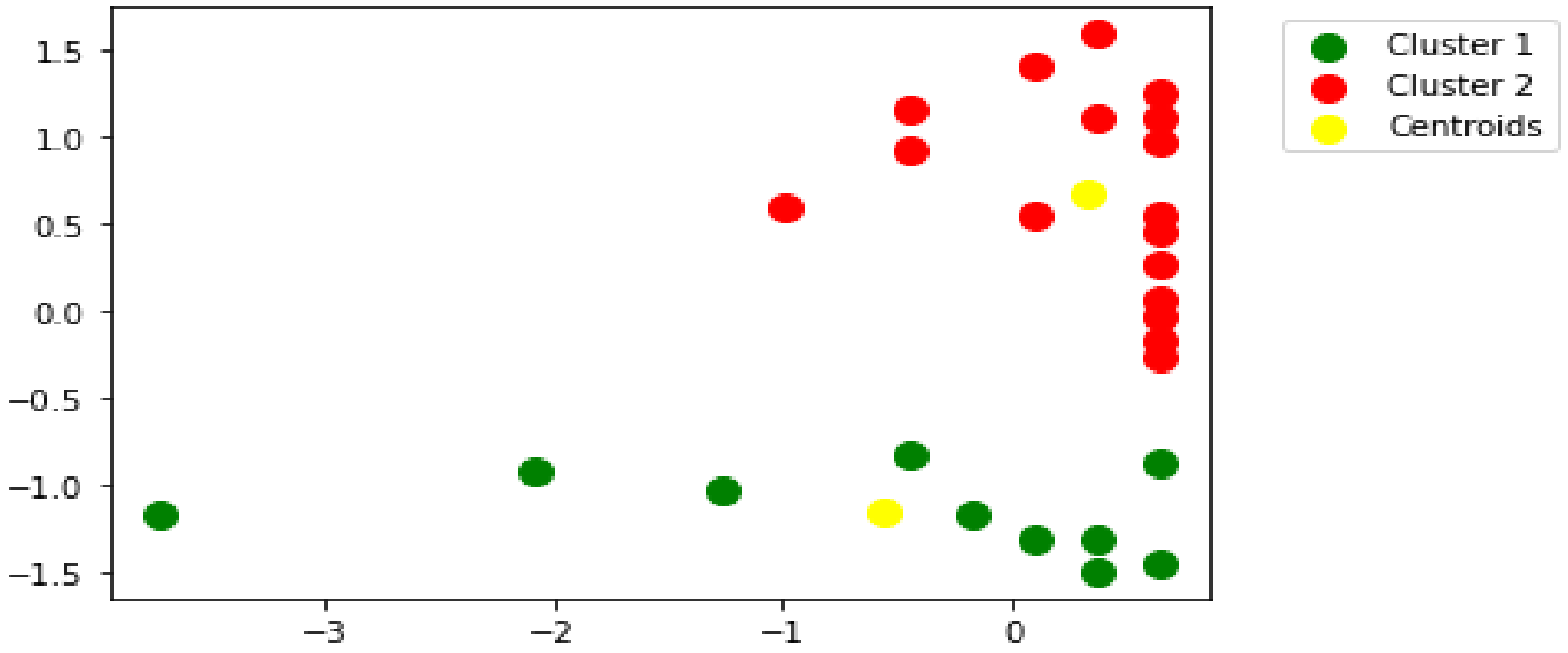
- SILHOUETTE METHOD

To determine the number of clusters needed, I will look at the maximum point on the plot, which stands at (5) for the right number of clusters required. At (5) the K-means model will be optimized

Output does not show negative but comes at zero(0.28) showing an overlap in the dataset



Create and Present a two (2) cluster visualization of the dataset



Identify and Explain **two (2)** insights from the newly created clusters in 2 d)

- The Centroids are found somewhat in the centre of cluster 1 and cluster 2. the groupings and separations are distanced with identification
- Optimized based on 2 clusters
- Observing the newly created clusters, **n_cluster should be 7**

```
#PLOT SILHOUETTE
plt.plot(range(1,10), scores, marker='o')
plt.title('The Silhouette method')
plt.xlabel('Number of clusters')
plt.ylabel('Sihouette Coefficient')
plt.show()
```

```
For n_clusters= 2, The Silhouette Coefficient is 0.22972085065748718
For n_clusters= 3, The Silhouette Coefficient is 0.18310534303313306
For n_clusters= 4, The Silhouette Coefficient is 0.20631955740350025
For n_clusters= 5, The Silhouette Coefficient is 0.28185069808024893
For n_clusters= 6, The Silhouette Coefficient is 0.23580631229453766
For n_clusters= 7, The Silhouette Coefficient is 0.26707532254416966
For n_clusters= 8, The Silhouette Coefficient is 0.21708486525696882
For n_clusters= 9, The Silhouette Coefficient is 0.20720625848456434
```