Hierarchical Clustering (Agglomerative)

**Introduction:**

Hierarchical clustering is a technique in which clusters are created in a hierarchical fashion. This method can be visualized using a tree structure known as a dendrogram. The primary advantage of hierarchical clustering is that any number of clusters can be chosen by cutting the dendrogram at the right level.

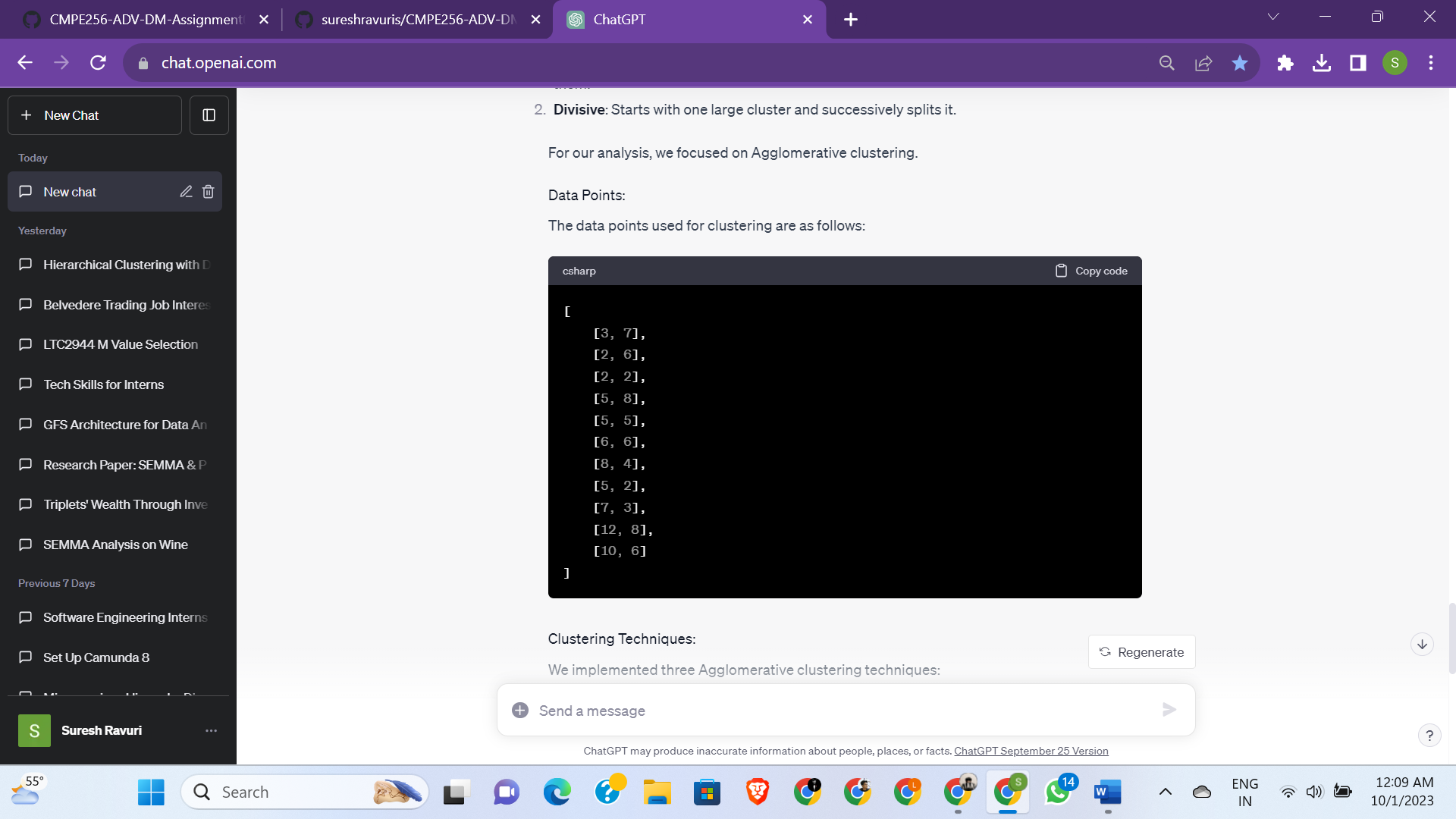
There are two primary strategies for hierarchical clustering:

1. **Agglomerative**: Starts with individual data points as clusters and successively merges them.
2. **Divisive**: Starts with one large cluster and successively splits it.

For our analysis, we focused on Agglomerative clustering.

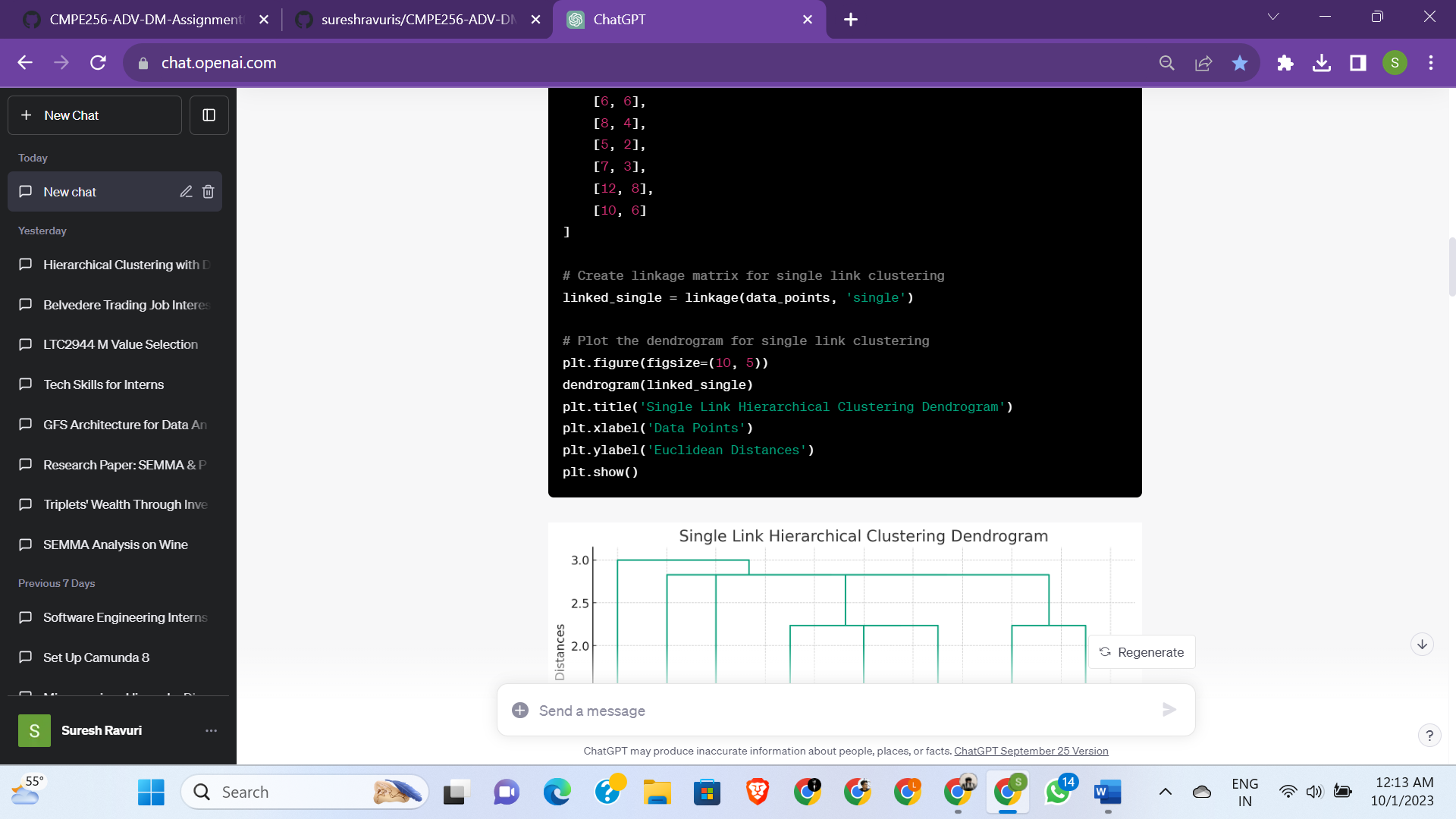
**Data Points:**

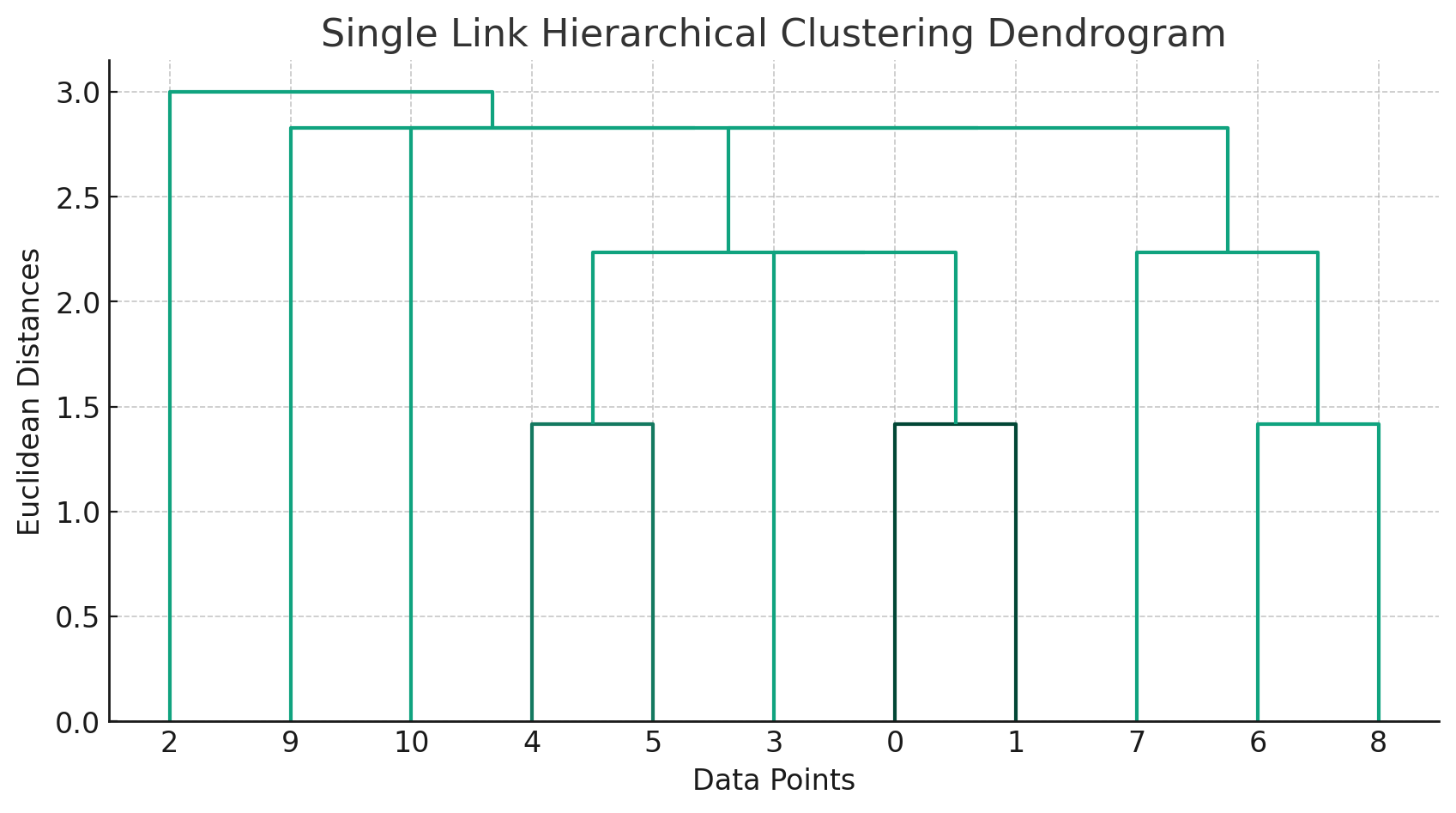
The data points used for clustering are as follows:



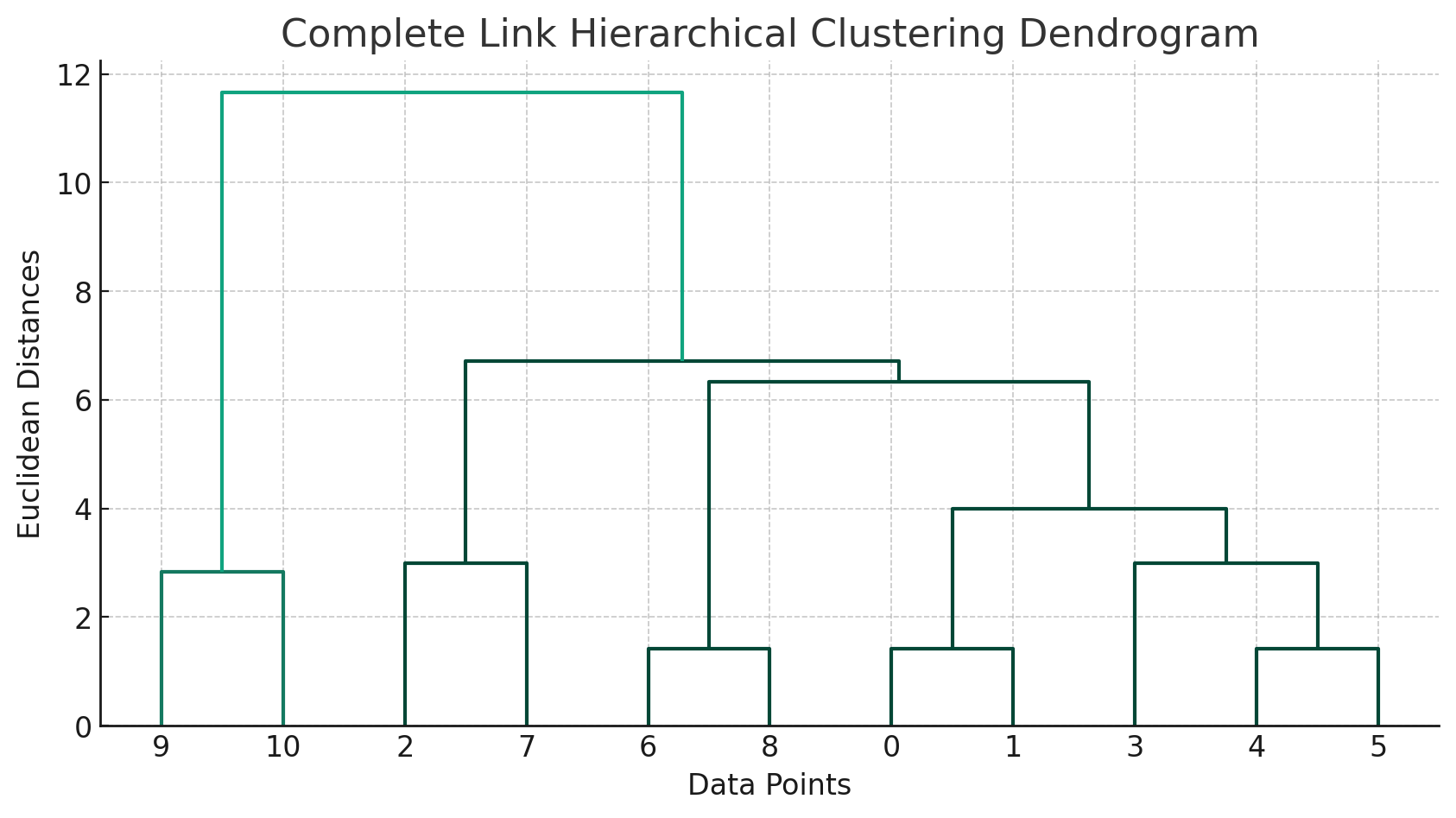
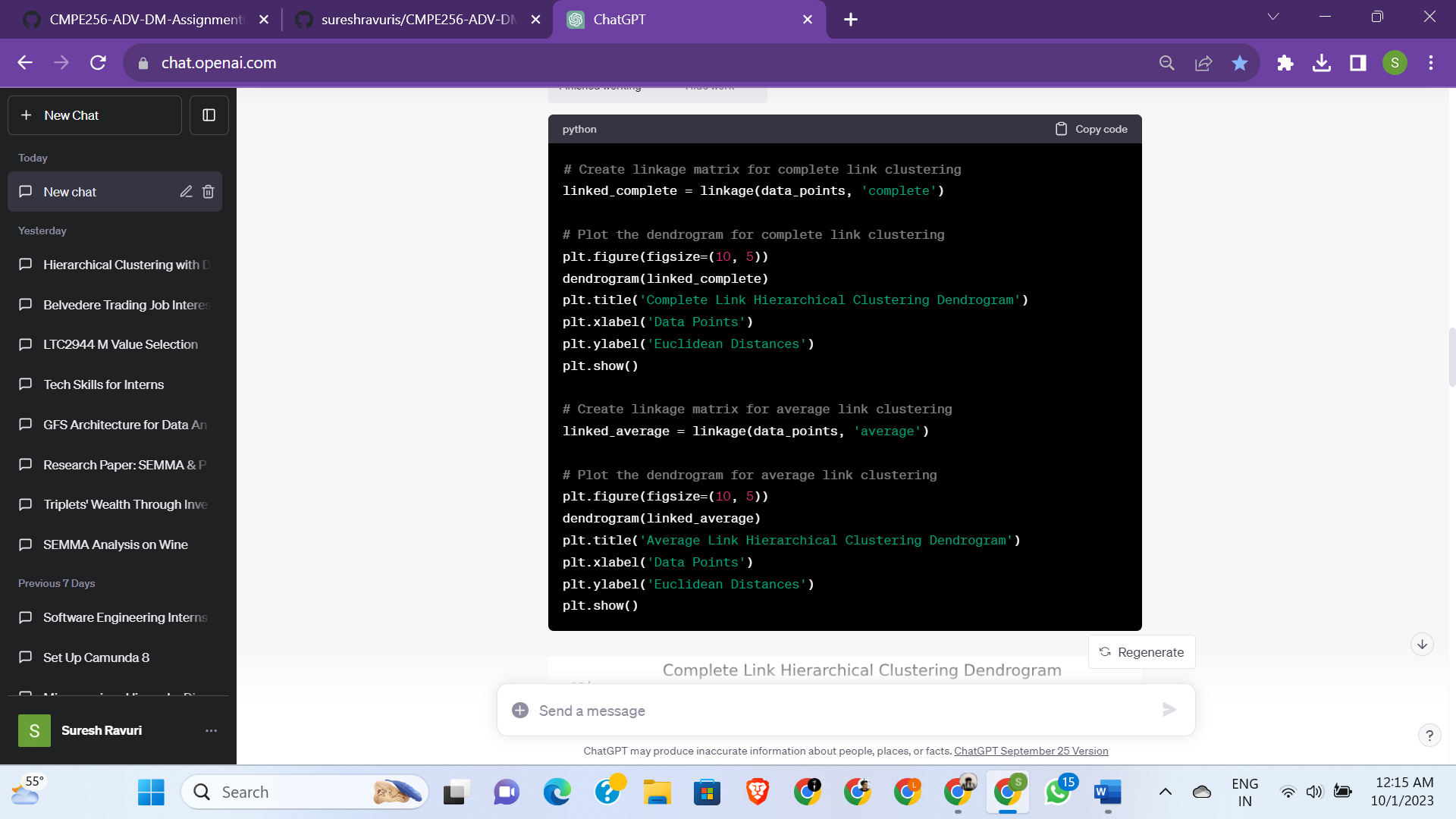
**Clustering Techniques:**

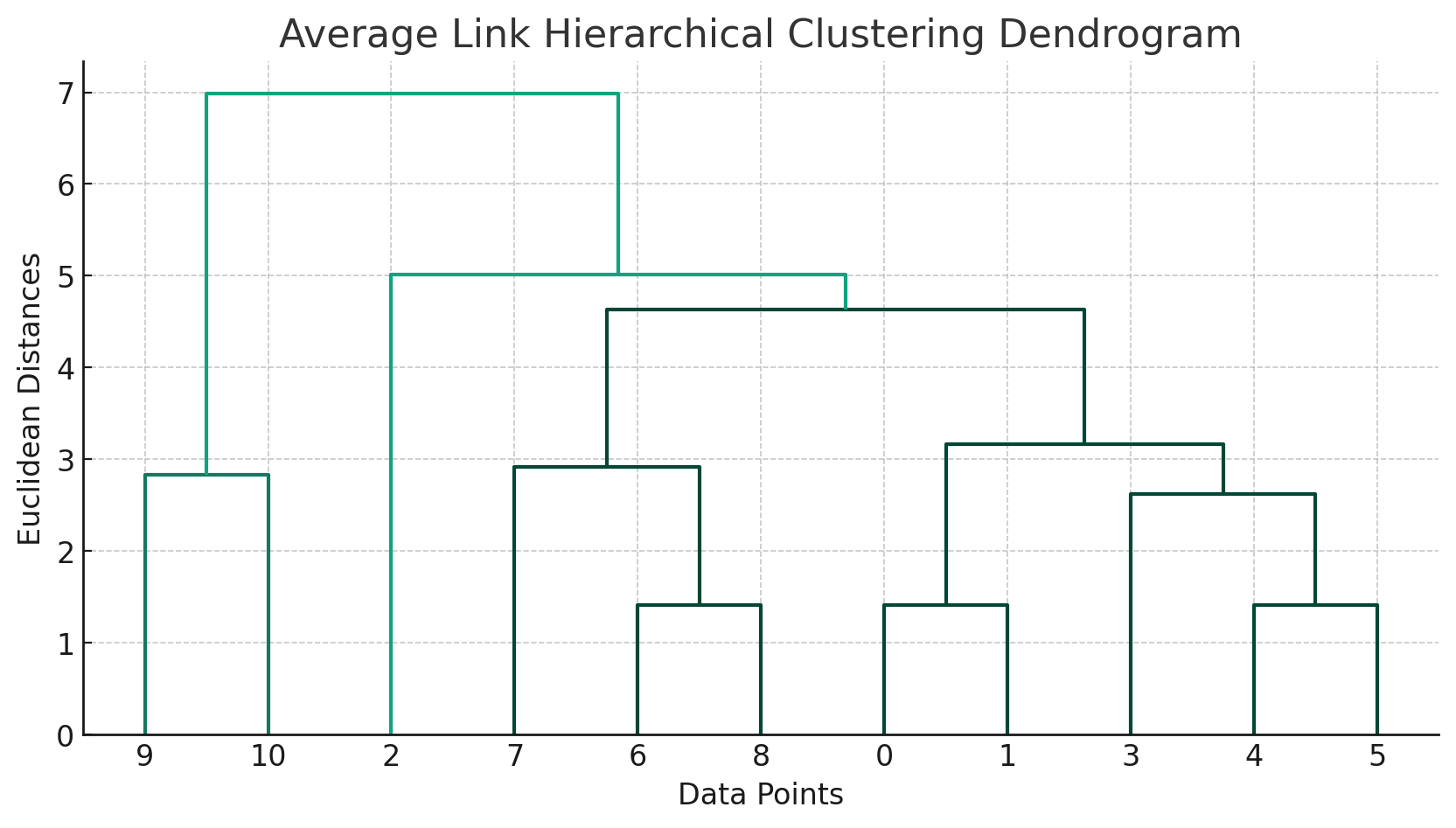
We implemented three Agglomerative clustering techniques:

1. **Single Link (Minimum Linkage)**: The distance between two clusters is the shortest distance between any two points, one from each cluster.



1. **Complete Link (Maximum Linkage)**: The distance between two clusters is the longest distance between any two points, one from each cluster.
2. **Average Link**: The distance between two clusters is the average distance between every pair of points, one from each cluster.



**Results:**

* **Single Link Hierarchical Clustering**: The dendrogram showcased how clusters are formed considering the shortest distance between them for merging. The y-axis represented the Euclidean distances at which data points or clusters are merged.
* **Complete Link Hierarchical Clustering**: The dendrogram depicted clusters formed based on the longest distance between data points in two different clusters. Clusters are merged based on the smallest of these maximum distances.
* **Average Link Hierarchical Clustering**: The dendrogram highlighted how clusters are formed based on the average distance between data points in two different clusters.

**Conclusion:**

Hierarchical clustering provides a comprehensive approach to understanding the structure of our data. By examining the dendrograms, we can determine how data points are related and decide on an appropriate number of clusters based on the application's requirements. The three techniques we implemented provided different perspectives, and the best method would depend on the specific context and requirements of the data analysis task.