

Number of Cluster Comparison

Clustering: Number of Clusters	Age:Chol	Age:Rest_BP	Age:Max_HB	Age:ST_deprsn	Inference
KMeans 1. Elbow Method 2. Silhouette Score	4 0.357	4 0.361	4 0.368	4 0.390	Silhouette score is better in K-means than in GMM, thus the graph is denser in K-Means where as in GMM it uses probabilistic points.
GMM 1. Silhouette Score	0.321	0.328	0.348	0.350	Silhouette Score is less in GMM, thus the graph is different, points are scattered, and as an inference people with high BP/Cholesterol/HB/ST and medium or small BP/Cholesterol/HB/ST are also coming under same cluster.
Agglomerative 1. Dendogram	4	5	5	5	Number of clusters are more while using Dendogram, and thus the graphs are clearer, denser and outliers are more defined.

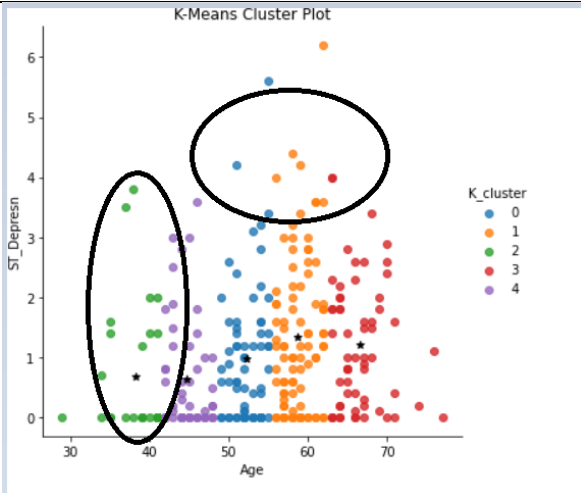
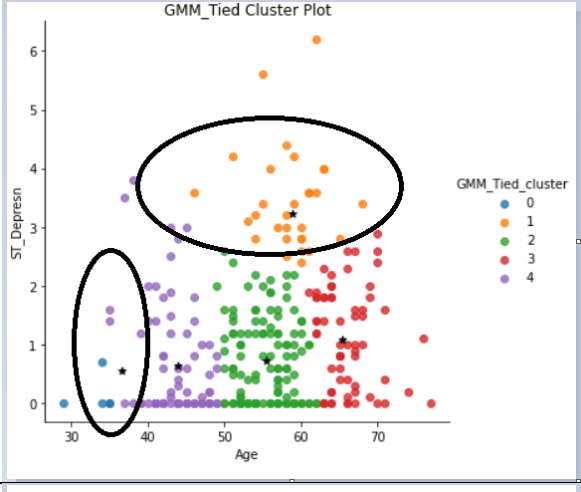
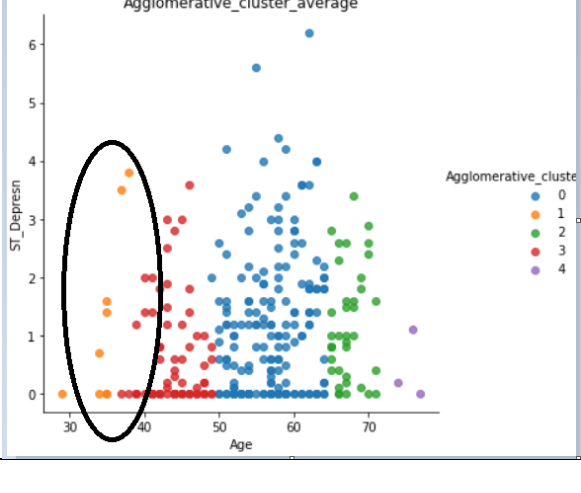
Cluster Visualization (Age vs ST_Depression)

Considering the data attributes of Age to ST-Depression and Age to Resting-BP, I am trying to find if the heart disease is possible for the patient. To predict the heart disease I referred to some of the documents to see how the behavior of these attributes contributes to an occurrence of heart disease.

With reference to:


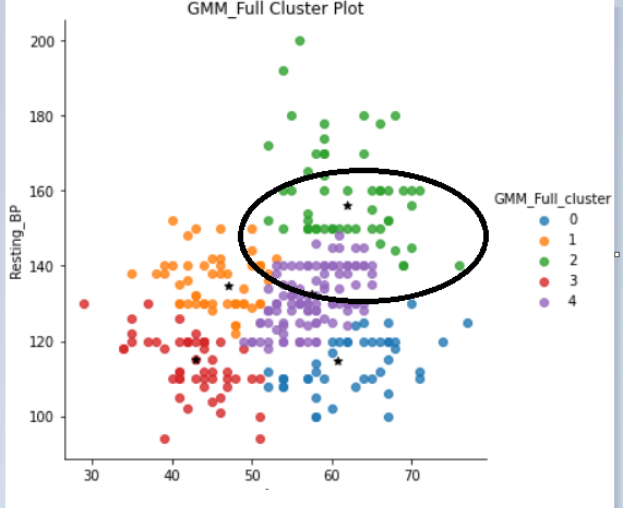
<https://pubmed.ncbi.nlm.nih.gov/11741361/#:~:text=Conclusions%3A%20In%20unstable%20coronary%20artery,substantially%20decreases%20death%2Fmyocardial%20infarction.>

It says that if the ST_Depression value is more than the patient is more prone for a heart disease.

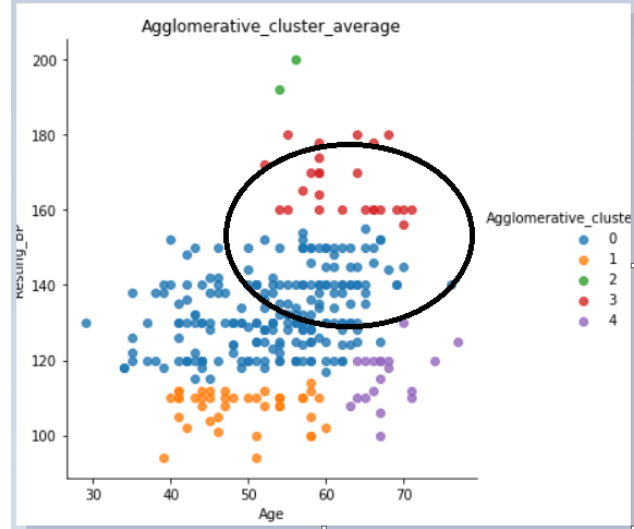
Clustering	Age:St_Depresn	Inference
KMeans	 <p>K-Means Cluster Plot</p>	The data points in the cluster for patients with high St Depression is a distributed in different clusters
GMM	 <p>GMM_Tied Cluster Plot</p>	The data points in the cluster for patients with high st_depression is in one cluster as compared to K-Means
Agglomerative	 <p>Agglomerative_cluster_average</p>	<p>Here, for age 30-40 there are two different clusters, which means in this clustering the behavior is different with respect to age as well and not only ST_Depression.</p> <p>Comparing all these, different clustering methods gives different insights of the data. It could be on the basis of X,Y or both.</p>

Cluster Visualization (Age vs Resting-BP)

According to https://www.medicinenet.com/what_is_stroke-level_high_blood_pressure/article.htm it says that people with 180/120 mm Hg are prone for heart stroke.

Clustering	Age:Resting_BP	Inference
KMeans	 <p>K-Means Cluster Plot</p> <p>Resting_BP</p> <p>Age</p> <p>K_cluster</p> <ul style="list-style-type: none">01234	If we see here, the resting bp value above 160 is a completely different cluster
GMM	 <p>GMM_Full Cluster Plot</p> <p>Resting_BP</p> <p>Age</p> <p>GMM_Full_cluster</p> <ul style="list-style-type: none">01234	If we see here, the resting bp value above 160 cluster is mixed with below values as well

Agglomerative



The density of bp value between 150-120 is more and is coming under one cluster.

Thus, these differences show how the values of BP are clustered differently in all the above methods.

Data Narrative:

Both the attributes show that between age 50-60/65 people have more ST_Depression and BP. So as a result even though the cluster itself has different insights but the overall business logic is still the same that is in between the ages 50-65 people are more prone for heart disease.