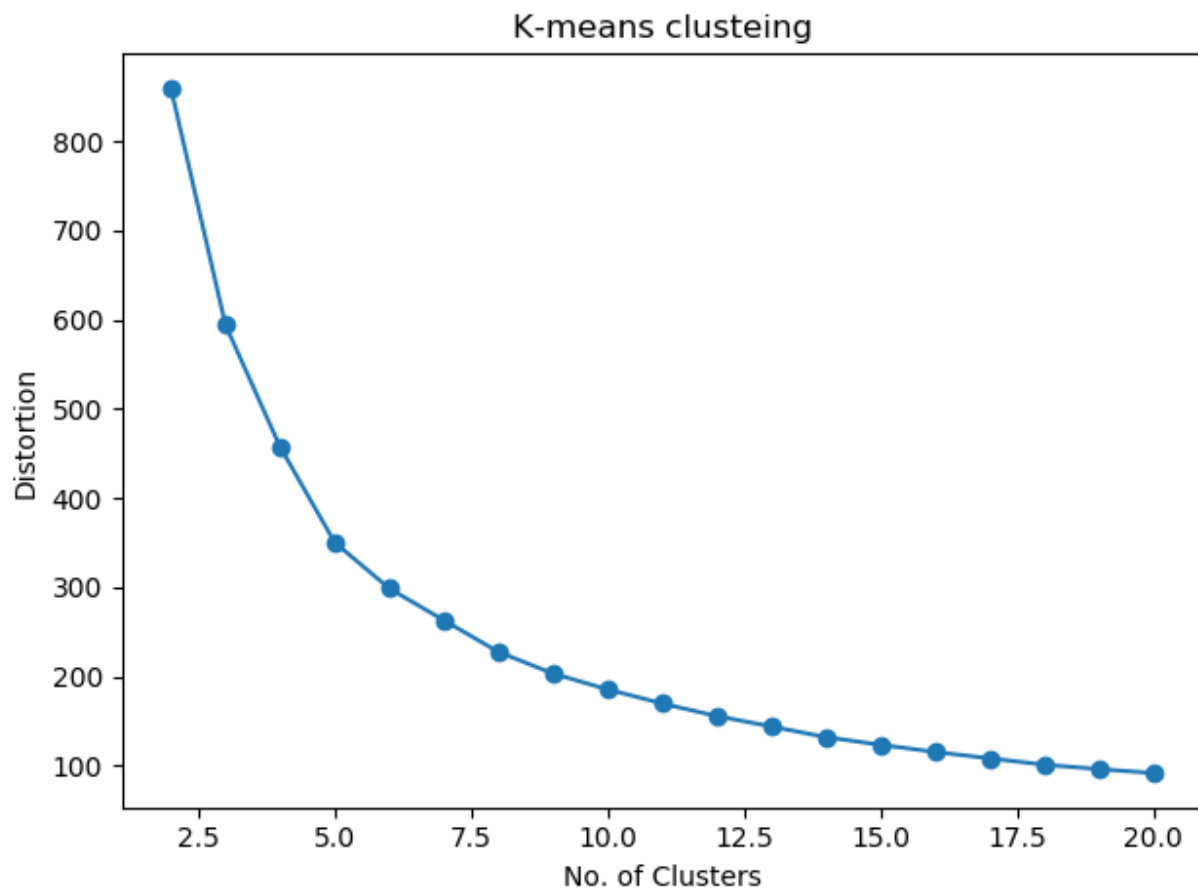


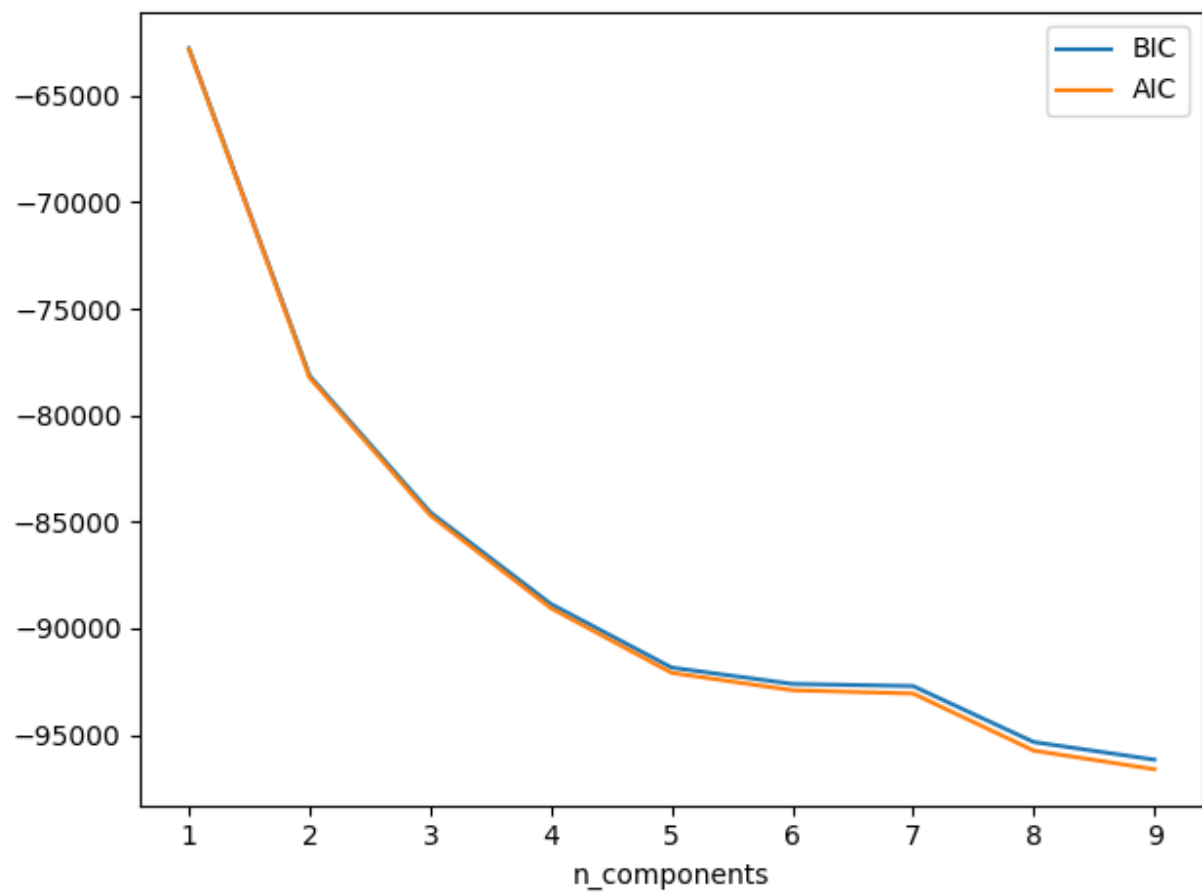
No of clusters given the data:

2 approaches were used to find out the optimal no of clusters given the data - K Means clustering and Gaussian Mixture Model.

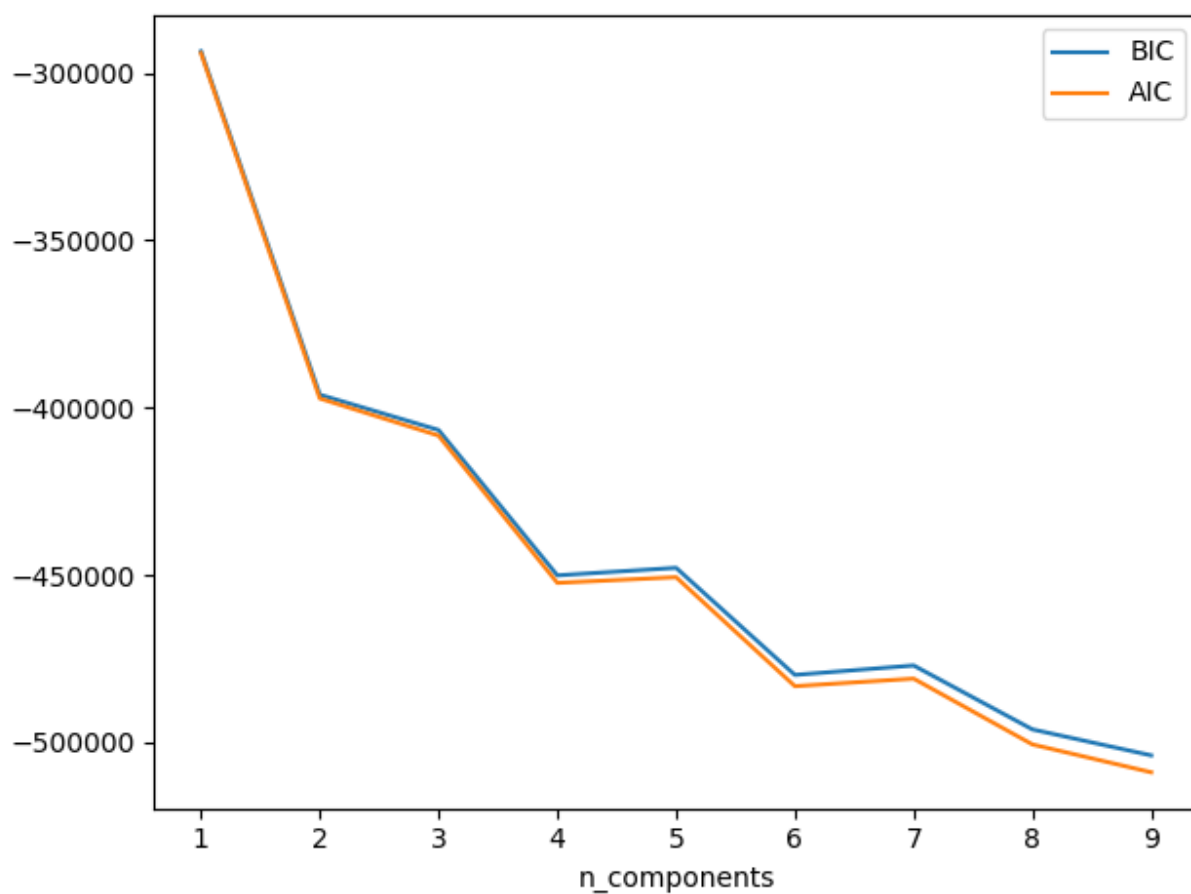
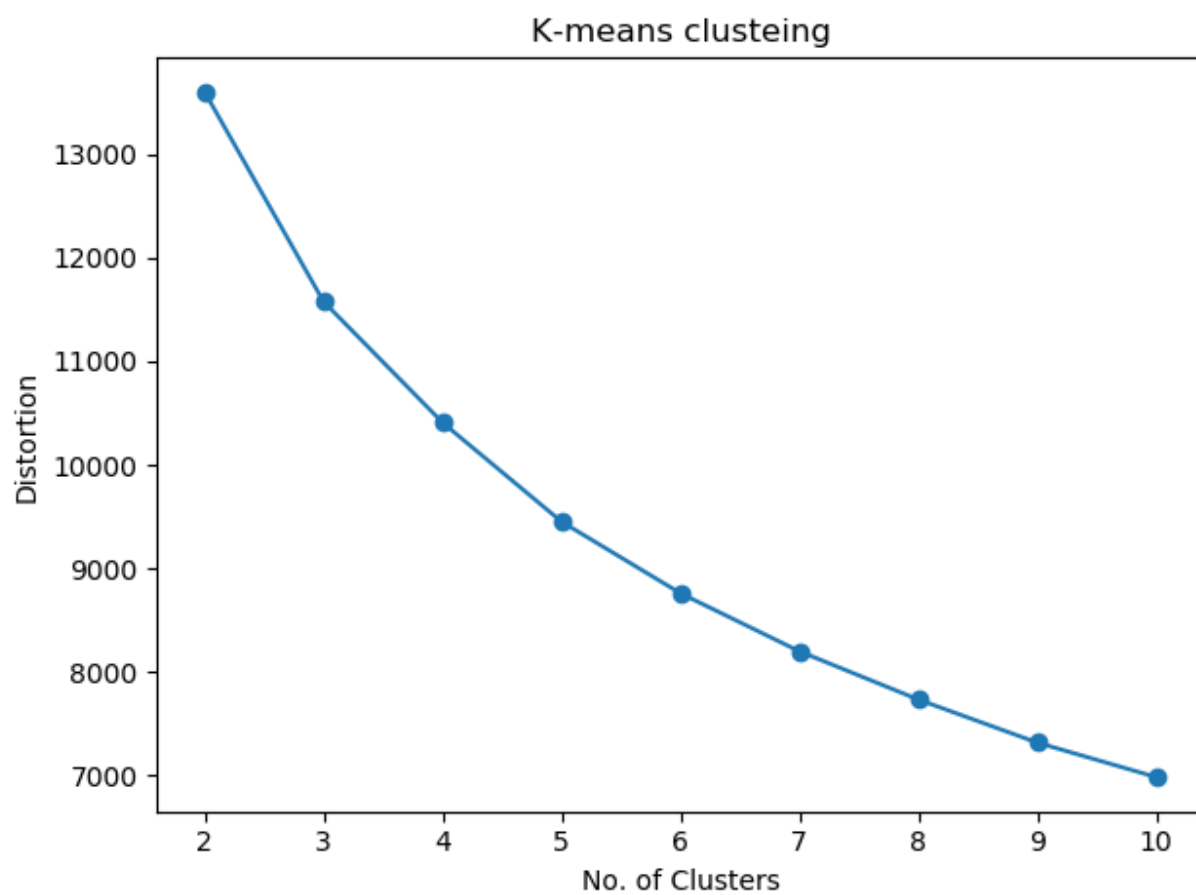
Since attributes d1-d13 pointed to the similar category of sensors, a maximum of the recorded values was taken for these sensor for every observation and an additional attributed was 'maxValue' created to reflect this additional information. 2 approaches were used: take the maxValue and the radiation attributes to find out the clusters, and to take all attributes.

When only maxValue attribute and radiation attribute was taken, the k-means clustering showed a distinctive elbow at k=5, indicating that there might be 5 clusters. This hypothesis was tested with Gaussian Mixture Model. However, the objective to find the cluster that minimizes the AIC and BIC criterion failed as both the AIC and BIC were decreasing as the no of clusters increased. Following are the two graphs: the first one indicates the distortion as the no of clusters increase. The second indicates the area under the AIC and the BIC criterion.





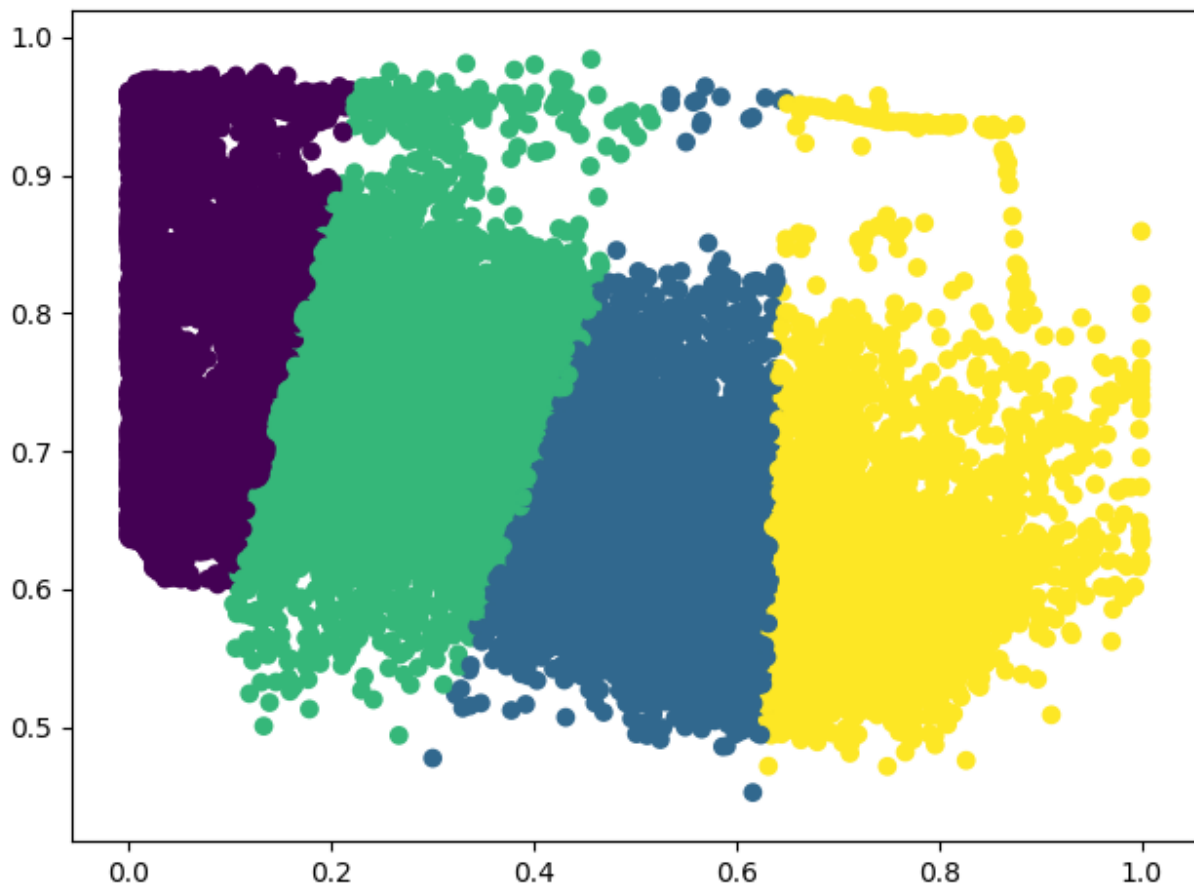
The models were also tested with all the attributes taken into consideration (with the exception of the 1st 13 sensor attributes being replaced with the max of the 13 sensor data). Following are the results:



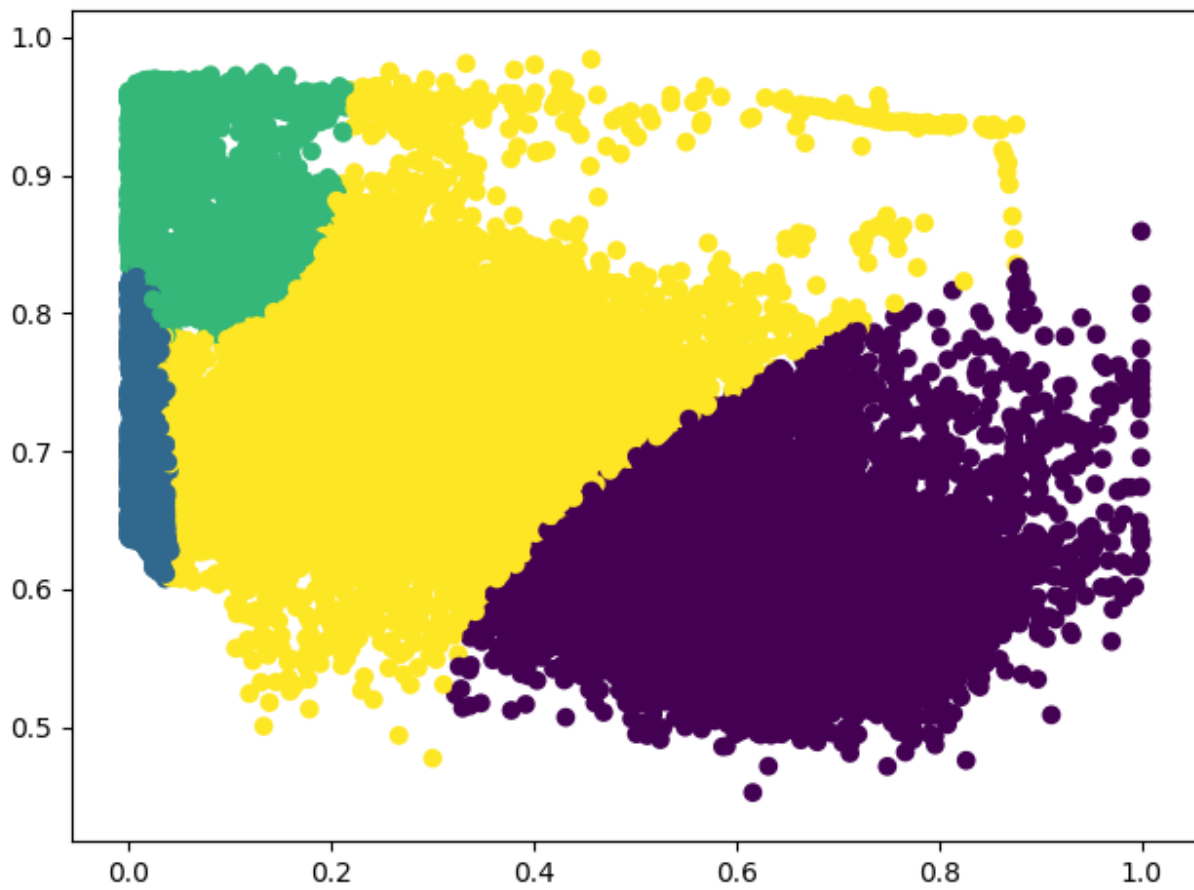
Not much can be inferred from the above two graphs. Hence, cluster = 5 was chosen as the optimum cluster.

Following are the results for the complete data when no. of clusters = 5

#### 1. K-Means



## 2. GMM



For the above two graphs, the x-axis indicates the maxValue of the sensors and the y-axis indicates the radiation.