Task:

We have a microbiome dataset that has 16S rRNA gene count for 178 samples and 7 features. The samples have been grouped into distinct clusters [CST1, CST2, CST3, CST4, CST5, CST6, CST7].

Please create a report to answer the following questions

1. Does the data fit any distribution?
2. How will you normalize the data?
3. Perform EDA for entire dataset. Share graphs and explanations
4. Perform EDA for data split by clusters. Share graphs and explanations
5. We would like to write a classifier to classify samples into the clusters.
   1. What corrective methods do you need to perform to the data?
   2. What are the challenges?
   3. What will be you choice for classification method and why?