**CS5542 Big Data Apps and Analytics**

**In Class Programming –7**

**Due 7th March 2022 (11:59 pm CST)**

**K-means clustering:**

**Use a different dataset (than the one we used in class) and use the model provided in ICP7 to perform clustering. You must try 5 different number of clusters (for example n\_clusters= 5 or n\_clusters=6,7,8,or 9 etc) based on elbow curve and for each cluster visualize the clustering results and report your findings in detail. Also provide the elbow curve screen shots in your report.**

ICP Requirements:

1. Successfully executing the code and making clusters with different value of K.
2. Using a new and good dataset
3. Providing the logical explanation of the observations and over all code quality
4. Wiki Report quality, video explanation

Submission Guidelines:

1. Sign in into your github account
2. Click on this link : <https://classroom.github.com/a/GdzGJY5X>
3. Accept the ICP-6 (Assignment 6)
4. Complete your ICP and create your wiki report (Pdf or Word doc).
5. Folders for ICP:
   1. Create two folders for source code and documentation
      1. Source code folder contains only the code and output file (if applicable)
      2. Documentation folder contains wiki report and the images of your results
6. The wiki report should have at least the following:
   1. What you learned in the ICP
   2. ICP description what was the task you were performing
   3. Challenges that you faced
   4. Screen shots that shows the successful execution of each required step of your code
   5. Out put file link if applicable
   6. Video link (YouTube or any other publicly available video platform)
   7. Any inside about the data or the ICP in general
7. Upload your ICP folders to your assignment GitHub repository
8. Click on the add a readme file on the next page and write done your and your partner name and email

You are all Done!!!!!!!