Assignment 6
Due Date: Sunday, 7 Feb 2021

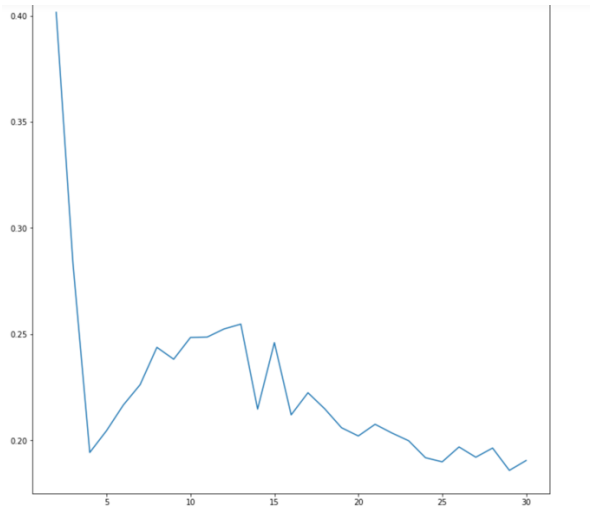
Important: Only PDF outputs and IPYNB files will be considered for evaluation.

.py files will NOT be graded: you will be awarded a zero score if you end up submitting a .py file!

Clustering task: Use the Customer Dataset provided to perform exploratory data analysis and find some interesting clusters; You may take code snippets from the labs uploaded.

Q1. Check if you observe any hints about the number of clusters you should create from either elbow method or silhouette method with a visualization. **(20 points)**

Expected outcome: an explanation with a sample graph like below. It is okay if you don't get a perfectly smooth shape or an elbow.



Q2. Create 2 clustering algorithms and observe if your dataset can be clustered into balanced clusters. Explain the differences between clusters. **(50 points)**

Expected outcome: you should be able to run your clustering algorithm using scikit learn and explain 5 differentiating points between your clusters. You can use the number of clusters inferred from the first question.

Sample code to build 1 model (you will have to change/ customize it as per the choice of your algorithms):
kmeans = KMeans(n_clusters=n)
kmeans.fit(X)

Q3. Explore the possibility if PCA could help you differentiate clusters which you observed and plot a relevant visualization. **(30 points)**
Expected Outcome (your visualization may vary depending on your feature selection and number of principal components you decide to visualize):