**INFO 7390 assignment I**

**Maximum marks = 50**

**Assignment date March 01, 2022**

**Assignment deadline is March 14 2022 (6:00 pm PST)**

(1) Implement a basic k means clustering algorithm, ideally as a set of functions\*. Run it on any dataset of your choice or on the attached online retail dataset. Provide a simple visualization of its results. Please write code in Python on a Jupyter notebook.

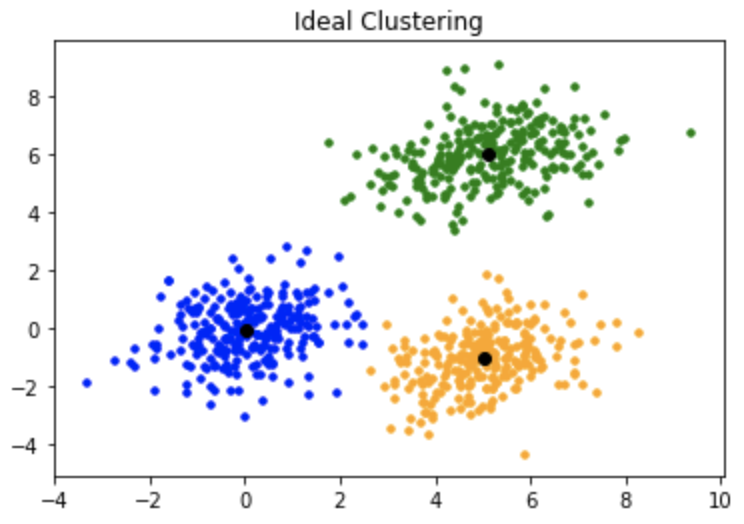
Note: Please submit the assignment before the due date. Each delayed day will accrue a 10% penalty

Rubric is given below:

Assignment1 Rubric

Code that runs all the way from reading data in to printing output, i.e, the visualization carries 25 marks. This will be tested during evaluation and an error will cost you all 25 marks :-(...

During code review, your code takes input (your choice of data) and produces the k-means cluster plot, each set or group should have a different color. See below for an example:



The visualization will carry a total of 15 marks split up this way: the plot will carry 5 marks, and the x and y axis labels (NOT SHOWN ABOVE) will carry another 5 marks. Show the above plot for a few values of k, k =2, k =3, k=4, This sub-part caries 5 marks.

Code review. Showing up and successfully completing code review carries 10 marks.

\****Please note that you are not allowed to use any machine learning packages or libraries such as scikit- learn to implement this algorithm. Use of each package will lead to deduction of 5 points. Please remind yourself of this very important point.***