ECON860 Data Analysis For Economics(Fall 2021)
Final Exam 1

Due: 12th December 2021 (11:59 pm)

Name: _____

UID: _____

---

1. You are given a dataset with 21639 individuals. The dataset contains the answers to a questionnaire with 40 questions. Your task is to cluster these individuals into groups.
   (a) The questions are similar to the "Big Five Inventory" in the lecture (But not the same). However, you do not have the cookbook, so you do not know which questions correspond to which personality traits.
   (b) The answers are on a scale of 1 to 5. If the individual refuses to answer that particular question, the value would be zero.
   (c) Use factor analysis to get a measure of several personality traits from the questionnaire. Notice that this is not the same as the "Big Five Inventory", so you may not have exactly five traits. You need to follow the procedure introduced in the lecture to find out what is the suitable number of traits (factors).
   (d) Use the personality traits to cluster the individuals. You may use KMean clustering, Gaussian mixture model, Agglomerative Hierarchical clustering or any other unsupervised learning techniques.
   (e) Find the optimal number of clusters. Your answer may be different for each of the clustering algorithms you have chosen.
   (f) Which algorithm gives you a better result? Explain your answer or explain why it is not possible to evaluate which algorithm is better.
   (g) Are your answers different for different countries?
   (h) You must hand in your homework via Github. Create a repository named "ECON860_final". In your repository, you should have the code and a .gitignore file. You should also include a file named README, which includes step-by-step instructions on how to run your Python code to collect the data you collected. This is especially important if you have multiple Python files. You should have a written answer committed in your repository. It can be included in the README file or it can be a separate file.
   (i) You can commit and push to Github as many times as you like. Only your last commit before the deadline is graded. I can read your previous commits, but they will not be graded.