Problem Set #4

MACS 30000, Dr. Evans

Due Friday, Dec. 2 at 11:59pm

1. **Kaggle open call projects (3 points).** Kaggle.com is a data science website that hosts open call projects, datasets, data science competitions, code, and more.

   (a) Create an account with Kaggle.com and sign in.

   (b) Explore the Kaggle Competitions page, which maintains a list of open call projects. Describe one that is of interest to you. What is the goal of the competition? What would you have to do make a submission?

   (c) Explore the available datasets on the Kaggle Datasets page. Download one of the datasets and write a Python script that creates a descriptive plot that highlights some instructive characteristic of the data. Make sure your plot has a title, labeled axes and axis titles, and a legend if necessary. Submit both your Python script and your plot. Make sure your Python script is properly commented.

2. **Improving a journal article (3 points).** Look through a recent issue of a journal in your field. Find a paper that you think could have been reformulated as human computation project. How would you formulate the data collection as a human computation project? How might this improve the study?

3. **Amazon Mechanical Turk (4 points).** Many human computation projects rely on participants from Amazon Mechanical Turk. Sign-up to become a worker on Amazon Mechanical Turk by clicking on the "sign in as a worker" link in the upper-right corner of the home page. Spend one hour working there in some HITs (human intelligence tasks). Show evidence of your hour of work by attaching a screen print of your "Your Account" page and your "Qualifications" page.