Problem Set #6

ECON 833, Prof. Jason DeBacker
Due Thursday, October 16, 1:15 p.m.

This problem set is designed to provide you with hands on experience estimating a matching model. Here's the general framework:

- Kaggle is an online platform that hosts data science competitions. Teams can form and compete in these competitions. I want you to estimate a matching model that tell us how these teams form.
- Data on the competitions and teams are available at https://www.kaggle.com/datasets/kaggle/meta-kaggle
  - You will need to join some of these data tables together
  - Of interest are the Teams, Team Membership, Users, User Organizations, User Achievements tables (but you may want/need to use other tables also).
- I want you to include in your matching model specification, at least three covariates from the following list:
  - Organization
  - Gender (infer from first name (use a Python natural language processing package to help you with this)
  - Country
  - User experience (e.g., tier, ranking, number of followers, score in competitions)
- You can otherwise specify the model as you would like. Your specification may depend on the estimator you are using (more on that below). You might consider a specification like the matching models discussed in class, where the payoffs are some linear function of the covariates (e.g, Hitch, Hortascu, and Ariely (2010) or Akkus, Cookson, and Hortaschu (2016)). But all models need to consider (1) that people may choose to work alone (about 90% choose this option) and (2) that teams can have more than 2 members.

Two useful papers to reference are:

- Chan, Chen, and Wu (*Marketing Science*, 2023) – this paper estimates matching model using Kaggle data
- Lemus and Marshall (*Review of Economics and Statistics*, forthcoming) – this paper estimates the returns to teamwork in Kaggle competitions and provides a good description of the data

There will be two groups, each will employ a different estimator:

- Group 1: Maximum Score Estimator; members: Ayse, Prachi, David, Yunus
- Group 2: MLE/Mixed Logit Estimator; members: Emma, Gyumin, Narender

Each group will present their findings to class on Thursday, Oct 16. You will have 20 minutes to present findings. In addition, each gropu will turn in code that will enable me to replicate their results, as well as a paper describing the model and the main results.