11/16/16: Applied Data Analytics

Presenter:

- Julia Lane, Robert F. Wagner Graduate School of Public Service, NYU; julia.lane@nyu.edu
- Frauke Kreuter, Professor, Joint Program in Survey Methodology, University of Maryland; fkreuter@umd.edu

Relevant Attachments:
- Applied Data Analytics Presentation

Key Takeaway: *We want to empower and train people on the ground to use our approaches to the maximum extent possible to increase community saving and decrease recidivism. We are not just teaching people how coding works but training them on how to deal with social problems by using real data in the local community.*

With 11 million people moving through the CJ system, we need to build tools that trace people at the different stages, from arrest to community reintegration. It costs over 22 billion dollars to provide services to those incarcerated that are dealing with mental illness and other health issues. We are developing a training program to advance data and social science based solutions in criminal justice – how to work with data and how to develop policy analysis from raw data and how machine learning can help reduce administrative costs. We want to empower and train people in on the ground to use our approaches to the maximum extent possible to increase community savings and decrease recidivism. We are not just teaching people how coding works but training them on how to deal with social problems by using real data in the local community. One of the distinct facets of our program is that our students can use information in the real world, not abstract theory.

Taking micro data from the Department of Corrections and pairing it with data collected from the United States Department of Housing and Urban Development helped to identify trends that could be used to improve the community with changes to the infrastructure at the local level. For example, if you change a city bus line, how many jobs would be created or eliminated? One of the unique things about our program is that we take local data and generate scenarios for our students to work with – generating "real world" solutions to community problems.

The goal is not just to take this course but to work with the data to develop new data products. Our focus is on learning about the challenges in the integration of data. After 3 rounds of teaching this course, we developed a textbook to blend code, theory, and examples. Linking data sources to answer broader questions, from where the target population you are trying to help lives to how they get to work can be answered by analyzing city infrastructure and the likelihood of accepting jobs. Our data sets fit into three silos, ex-offenders, services, and veterans and we cross reference this information with housing data.

Q. Do you have to be in government, or can could those in the private sector apply?

A. You don't need to be affiliated with a government agency to participate in the course but would need to be a grantee from the Arnold Foundation.

Q. Is the course offered online?
A. No, we prefer in person interaction to build cross-agency networks and hands-on experience.

Q. Can counties use their own data sets in the course?
A. Our course will use set data sets taught around data on ex-offenders, welfare recipients and veterans, but counties can bring in teams and work on their own data. However, the curriculum is built around a particular module and sets of data.

Q. Does the course require background knowledge into data and analytics or Python?
A. Participants should have a basic knowledge of data, an understanding of statistics, and how to draw inferences. Additionally, the class is taught with the Python platform but will teach a functional use of Python to the class on day one.