Final Project, Spring 2018

Python for Data Management And Analytics

# Guidelines

* The final project is team-based
* Each team can consist of 2-3 students
* All written reports should be submitted in PDF format
* All code must be submitted as Python Jupyter Notebook scripts

# Milestones

* **(Due 3/19)** The abstract is just a short paragraph describing:
  + The project name and member(s)
  + The research question you have chosen
  + Why is the research question interesting to you
  + What is your overall plan for data analysis (this may change as you work on your project, but we want you to at least start planning ahead)
* **(Due 3/26)** Data story  
  *Note: by this point you must merge, clean, and transform your data*
  + The title and member(s)
  + The research question
  + Why is the research question interesting to you
  + Data story - tell a story of your data using descriptive statistics and visualizations. This data story must also include the list of variables that you are planning on using in your analysis and the justification explaining why you are using these particular variables.
* **(Due 4/16)** Final presentation
* **(Due 4/23)** Final report
  + The final paper should be 4 - 6 pages in length following and contain the following sections:
    1. Abstract
    2. Introduction
    3. Methodology
    4. Results
    5. Discussion
  + You should describe your problem, approach, dataset, data analysis, evaluation, discussion, references, and so on, in sufficient details, and you need to show supporting evidence in tables and/or figures.
  + You need to provide captions for all tables and figures.
  + You should also briefly describe how each member contributes to the total work in the end of the report.

**The final project will be graded largely based on the final presentation and final report.**

**Late submissions will not be accepted for the final report.**

# The Data

You must use the data sources listed below, but you can also find any relevant datasets on your own if you think that they are useful to your analysis. For example, depending on your research question, you may want to find a dataset of wealth (income) by zip code and include it in your analysis.

* **Accidental Drug Related Deaths 2012-2017**. A listing of each accidental death associated with drug overdose in Connecticut from 2012 to June 2017. <https://catalog.data.gov/dataset/accidental-drug-related-deaths-january-2012-sept-2015>
* **Opioid Related Treatment Admissions by Town in Department of Mental Health and Addiction Services Programs**. Town level data on depicting the number of admissions, and individuals served per year in treatment programs funded or operated by the Department of Mental Health and Addiction Services, where the primary drug at admission was an opioid. Clients entering these programs were treated for their opioid related disorder as a result of it being the primary drug at the time of admission. This dataset does not reflect ALL individuals in Connecticut who have participated in an opioid related treatment program, but only service provided by the Department of Mental Health and Addiction Services, nor are these programs specifically designated as "opioid treatment." <https://catalog.data.gov/dataset/opioid-related-treatment-admsissions-by-town-in-department-of-mental-health-and-addiction->
* **Pharmacies offering Narcan, Evzio and other brands of Naloxone**. Pharmacies that are able to dispense Naloxone. Naloxone is a medication approved by the Food and Drug Administration (FDA) to prevent overdose by opioids such as heroin, morphine, and oxycodone. <https://catalog.data.gov/dataset/pharmacies-offering-narcan-evzio-and-other-brands-of-naloxone>
* **U.S. Opiate Prescriptions**. This dataset contains summaries of prescription records for 250 common opioid and non-opioid drugs written by 25,000 unique licensed medical professionals in 2014 in the United States for citizens covered under Class D Medicare as well as some metadata about the doctors themselves. <https://www.kaggle.com/greenmaverick/exploratory-analysis-on-opioid-prescriptions/data>
* Other Resources:
  + Articles and references to help orient you to the topic and inspire research questions
  + **Short Answers to Hard Questions About the Opioid Crisis.** <https://www.nytimes.com/interactive/2017/08/03/upshot/opioid-drug-overdose-epidemic.html>
  + **Jump In Overdoses Shows Opioid Epidemic Has Worsened** <https://www.npr.org/sections/health-shots/2018/03/06/590923149/jump-in-overdoses-shows-opioid-epidemic-has-worsened>
  + **CDC releases grim new opioid overdose figures: 'We're talking about more than an exponential increase'** <http://www.chicagotribune.com/lifestyles/health/ct-opioid-overdose-increase-20171221-story.html>
  + Opioid Abuse And Poisoning: Trends In Inpatient And Emergency Department Discharges <https://www.healthaffairs.org/doi/10.1377/hlthaff.2017.0260>

# Evaluation Rubric

|  | **Technical strength and experimental evaluation** | **Presentation** | **Implementation** | **Overall project management** | **Total** |
| --- | --- | --- | --- | --- | --- |
| **Peer review** | 15% | 15% |  |  | 30% |
| **Instructor review** | 20% | 20% | 20% | 10% | 70% |

## Technical strength and experimental evaluation

***Technical strength***: The pedagogical goal of this project is for you to practice how to design, implement and evaluate data mining techniques and how to apply them in practical scenarios. In order to show the technical strength in your project, you need to provide sufficient details in your oral presentation and written report. Specifically, what do you try to accomplish? What method (data mining techniques) did you adopt? What are the strengths of your method? Describe the systematic approach how you achieve your specific goal, and justify your method with supporting argument through evaluation.

***Experimental evaluation***: Describe your experimental results with systematic evaluation in your oral presentation and written report. Specifically, describe the evaluation criteria, including performance measures and baseline methods. Explicitly state the performance of your method, and provide whatever results and insight you have gained from the evaluation.

## Presentation

You need to clearly present all the required components as described above in both your oral presentation and written report. Present your project to the entire class and you will receive peer review from the class. In your written report, provide sufficient details within proper page length. In the written report, cite and provide reference for any source of information you have used in your project, including publications, online information, code or software packages.

***Honor code***: You may consult any papers, books, online references, or publicly available implementations for ideas and code that you may want to incorporate into your project, as long as you clearly cite the sources in your code and your written report.

## Implementation

You need to submit Python code for your project that is working and is able to produce the same experimental results as described in your final report. Hence, the experimental procedure and the testing data must be submitted together with necessary documentation describing how to run the experimentation.

## Overall project management

You need to submit required materials to meet each milestone as described in the course webpage “Final project guideline.” These milestones are to help you seek resources you may need, make sure you develop a concrete direction, and better manage your project. Before each milestone, if you find it difficult to meet the requirement, you should make an appointment and discuss with me in advance.