Data Science for Economics Research Project

Project overview

- In groups of 3-4, prepare a short original research paper using tools from class to answer an economic question
- Key requirements:
    - Use of tools presented in class to create and/or prepare analysis dataset
        - Focus on second half of course: spatial data, remote sensing, supervised and unsupervised machine learning, text analysis/NLP, web APIs/scraping
        - Encouraged to go beyond concepts/techniques that were covered in the course material
    - Apply dataset to answer an economically interesting question
        - Could be causal or descriptive, but threshold for meaningful descriptive work is higher
    - Meaningful visualization of analysis
        - Both descriptive and analysis figures/tables
    - Connect research question and analysis to an economic framework
        - Objective is to present and justify the theoretical relationships between variables in the analysis, and use this to both motivate the analysis and interpret results
        - The framework can be presented in words or using equations
        - This is NOT the same thing as a econometric specification
    - Novel intellectual contribution
    - Well-documented code allowing reproduction of results
        - Python is not required; any combination of software packages is allowed
        - Documentation should include list of required libraries/modules as well as information on how raw data can be accessed
- Possible approaches:
    - Develop/create a new data source measuring a feature of economic interest, and apply that feature (on its own or combined with other data) to answering some question
    - Combine existing data sources in a novel way to answer an economic question
- Focus is on quality and contribution of data work, connection to an economic framework, and clarity of interpretation and visualization of results
    - Econometric approach is not a focus of this assignment but any limitations and concerns should be clearly presented and discussed in interpretation of results

Deliverables

- Proposal: List of group members and proposed idea in 2-3 sentences
    - Email due January 30 by 17h
- Presentation: 5-minute presentation of proposed research on February 6
    - Slides due February 6 by 12h
- Meeting: Discussion of project progress and feedback
    - Sign-ups for week of February 10
- Paper: 12-15 pages including references and exhibits, 15-20 pages for groups of 4
    - Paper due February 28 by 17h

- Code: Submit the set of code files used to produce the paper exhibits from original data, along with readme file describing required packages/libraries and outlining how raw data can be obtained or generated
    - Code files due February 28 by 17h

Evaluation rubric

- 5%: research question is well-defined and motivated
- 5%: intellectual contribution is clearly stated and non-trivial
- 10%: research question is clearly linked to an economic framework, both to motivate the analysis with theory and to guide interpretation of the results
- 45%: clear and appropriate use of tools and methods presented in class
    - Use of at least one tool or method from the following: spatial data analysis, remote sensing, big data analysis, supervised and unsupervised machine learning, text analysis/NLP, web APIs/scraping
    - Demonstrate a clear understanding of how to apply particular tools/methods to a given problem; not just using a tool/method but explaining why it is appropriate and how it is implemented
    - Preparation of dataset aligns with key principles for data cleaning/preprocessing and merging
- 5%: empirical approach/analytical method is appropriate for the research question, clearly described, and faithfully implemented
- 20%: figures and tables are clean and clear and summarize the key points of the paper
- 5%: paper is well-organized and written
- 5%: code is well-documented and results are reproducible