Final Project Rubric

PPOL670 – Introduction to Data Science

Fall 2019

Student:_____

Project Name: _____

Total Score: _____ / 52

Project Materials

- Report was posted on Canvas on time? (1 point)
  - .Rmd file and all figures were included with the submission?
- Report was generated using R? (1 point)
- Was the student's entire project was version controlled on Github?
  - Does the git record extend back to when (at least) the proposal was due? (1 point)
  - Is all code for the project contained within? (1 point)
  - Were the materials organized as an `R` Project? (1 point)
- Did the student provide a link to the data? (1 point)
  - Via Dropbox (if greater than a Gigabyte)
  - Via Canvas (if less than a Gigabyte)

Presentation

- **Student used professional looking visualizations in the report:**
  - Figures were easy to read and understand? (1 point)
  - Figures made sense within the context of the report? (1 point)
  - Student described the purpose and the insight drawn from the figure in the text? (1 point)
  - Color scheme made sense; easy to differentiate between colored items (1 point)
  - Figures were appropriately proportioned (1 point)
- **Student used professional looking tables:**
  - When presenting data and/or figures, student formatted the data as a clean table (i.e. latex)? (1 point)
    - see `stargazer` package or `gt`
  - Table made sense within the context of the report? (1 point)
  - Table was clear and easy to read. (1 point)
  - Table was concise and did not contain unnessary information. (1 point)
- **Student used `R` Markdown for a professional looking report:**
  - Report was rendered without errors or warnings. (1 point)
  - No `R` was visible in the report. (1 point)
  - Report contained a table of contents. (1 point)
  - Report contained no (or few) gramatical/spelling errors. (1 point)
  - Report reads as a single cohesive document. (1 point)

  - Report is 12 pages in length (double-spaced; 12 pt font). (1 point)
  - Student cited academic, data, and package sources. (1 point)
    * To cite a package, use `citation("package_name")`, e.g. `citation("ggplot2")`

Content

The student's project sufficiently addressed these general areas.

- **Introduction** (5 point)
  - What is the aim of the project?
    * Summarize the problem
    * State your goals
  - What do you do in this report?
    * offer a roadmap of the project
- **Problem Statement and Background** (5 point)
  - Give a clear and complete statement of the problem and/or aim of your analysis.
  - Include a brief summary of any related work that has tried to tackle a project similar to yours (i.e. a *light* literature review)
- **Data** (5 point)
  - Where does the data come from?
  - What is the unit of observation?
  - What are the variables of interest?
  - What steps did you take to wrangle the data?
- **Analysis** (5 point)
  - Describe the methods/tools you explored in your project.
  - Outline in detail our entire analysis.
    * Justify the tools/methods that you used.
    * Assume the reader is smart but doesn't know `R`/Machine Learning well. That is, be crystal clear about what you're doing and why.
- **Results** (5 point)
  - Give a detailed summary of your results.
  - Present your results clearly and concisely.
  - Please use visualizations and tables whenever possible.
- **Discussion** (5 point)
  - Speak on the "success" of your project (as defined in your proposal).
    * Did you achieve what you set out to do? If not why?
  - What tools/methods did you consider but *not* use in the final analysis?

– How would you expand the analysis if given more time?