May 7, 2024

Point by Point Breakdown of Project 4

1. hosted in github
   1. decide which repo on whose account is “of record”
   2. branches for everyone
   3. push pull merge etc.
2. define your target audience
   1. good way to get to the heart of this is to ask “What is the point of this presentation?”
3. make a dummy readme with the 5(ish) points
   1. A problem statement.
   2. A succinct formulation of the question your analysis seeks to answer.
   3. A table of contents, which should indicate which notebook or scripts a stakeholder should start with, and a link to an executive summary.
   4. A paragraph description of the data you used, plus your data acquisition, ingestion, and cleaning steps.
   5. A short description of software requirements (e.g., Pandas, Scikit-learn) required by your analysis.
   6. try to directly address every point of the requirements in the readme
   7. state or at least imply the target audience
4. notebooks (set these up a generic fill-in-the-blank thing)
   1. set random seed at the beginning of each notebook
   2. provide relative paths so that everything runs without error
   3. check them
      1. “Restart & Run All”
      2. fork to a new location and try running from there
5. build some kind of model
   1. how can you make the situation less of a poopshow?
   2. state some reason for making the model that you do
   3. explain any limitations of the model
   4. explain your reason for any researcher DOF decisions
   5. interpret appropriately – don’t over- or under-generalize
6. presentations
   1. must include (create dummy slides)
      1. A summary of the problem you tackled.
      2. A walkthrough of how you set out to solve the problem.
         1. summarize how you found the data
         2. summarize how you ingested the data
      3. A demonstration of your solution. (i.e. You may demonstrate an app you developed, an example of how a model may be used, etc.)
      4. A summary of any models you fit and, if applicable, their performance.
      5. A brief discussion of limitations to your process. (i.e. data collection issues, missing values)
      6. A brief discussion of next steps.
   2. requirements
      1. audience
         1. frame it to their interests
         2. use an appropriate level of technicality
         3. within a reasonable amount of time for the context in which we’re giving this presentation
         4. loud and clear
      2. likely aim for somewhere between 15 and 20 minutes
      3. must include a slideshow
      4. visuals
         1. appropriately scaled
         2. appropriately formatted
         3. readable to the audience
      5. clear recommendations
         1. explicitly state the flow from data to model to rec
         2. repeat the original problem statement question and explicitly state its answer
         3. make clear, specific, actionable recommendations (SMART heuristic)
7. summary:
   1. to get organized, set up blank
      1. readme with the sections
      2. basic notebooks with the sections
      3. slideshows with blank slides
      4. blank .py files too
   2. avoid merge conflicts
      1. tag everything with -emily
   3. set up blank folder structure in the repo
8. grading summary
   1. project requirements – 0 or 1 (THIS IS WHERE YOU COULD FAIL)
      1. an appropriate readme
      2. reproducible, error-free notebooks
      3. build some kind of model that adds value
   2. presentation is appropriate for the audience – 0-3 (even a 0 will pass)
   3. used the right methods and explained well why you used them – 0-3 (even a 0 will pass)
   4. how valuable is the thing you did – 0-3 (even a 0 will pass)