

# Sprints 1-3 Feedback

## Prof GS

---

### SPRINT 1-2

- Good job following the instructions from the video!
  - You all seem to have good communication and a solid plan in place. You have really talked as a team and thought through your answers (commitment, conflict resolution, preferences, etc).
  - For your branching strategy, you mention that you will use git branching, but that is not a particular strategy and for Sprint 3 you did not do any branches for any new features or any particular strategy.
- 

### SPRINT 3

#### Git/GitHub

- Make sure to commit and push more frequently. The idea behind going through sprint 2 is for you to keep using these principles when working as a team via GitHub. I do not see much of that in your git log.
- Make explicit commit messages
- Make sure everyone is working as a team and contributing equally, I mostly see commits from one team member.
- The second half of your README has a very different style than the first part. Go over it to make it more legible.
- The links you provided for your dataset has multiple files, and it's unclear which one you want us to download. Please upload your dataset to a gdrive and share the link in the README for the next assignment.
- You are also missing links! You mention "access link", but there is no link.

#### Feedback on questions:

- How do environmental factors correlate with mortality rates from heart disease and diabetes?
  - The way this question is phrased sounds like you are only going to do a correlation analysis instead of a model.
  - Thank you for including the target variable
- Can we predict future mortality rates from heart disease and diabetes based on socio-economic and environmental factors?
  - This is a yes/no answer. Similar to EDA you do not want yes/no questions, you want question that will render you an analysis or in this case a model.
  - The dataset has future mortality rates? How were these calculated?
  - Remember I said to not have any time series in your projects? Unless you are planning to consider it like our HW1 data. I don't want to see any forecasting, that's another class.
- What role do socio-economic factors play in the vulnerability of populations to heart disease and diabetes?
  - Again, we are not going to do correlation and causation analysis; that's another class. Think about the prediction and classification models we've done so far.
- Rephrase your questions to align with this course.
- Consider that each question will have a different target variable.
- Please keep in mind that we are using all the features for our models and then we decide if some of them need to be removed, according to the contribution to the model.
- Consider if you want to have an unsupervised learning question.

### **Feedback on Cleaning:**

- Include more explanations when cleaning your data.
- You checked for value counts, but nothing else. Are there any missing values? What did you learn about your dataset.
- Include a description for your last visualization.
- From your README it looked like you had multiple datasets but you only have one in your notebook.
- Please cleanup your README and notebook to reflect the correct data or datasets.

---

## **Steps moving forward**

For the next sprint:

- Use more GitHub as a team and push smaller changes.
- Make sure you are branching and following a strategy.
- Make sure everyone in the team is equally contributing to the repo.
- Revisit your team contract and check if your strategies are working. Feel free to change them if they are not.
- Add more explanations to your notebook.
- Revise your questions
- Meet up with your assigned TA.
- Assigned TA: Meghan