Course Project

Introduction

The goal of the course project is to design and share a data-focused Streamlit app.

Logistics

The following are requirements to receive a passing score on the course project.

- Due date: Monday, December 6th 2021, 3:30pm California time.
- For the submission, give a link to the app on Streamlit.
- Share the app using GitHub and Streamlit cloud (Community tier). The code should be viewable on GitHub, and the app itself should be shared by Streamlit cloud. Include a link to the GitHub repository somewhere in the Streamlit app.
- The primary focus of the app must be on something involving data。
- The project should clearly build on the Course material.
- Anything that is taken from another source (either the idea for the project or a piece of code that is longer than a few lines, even if you edit that code) should be referenced with a link, both in the source code and in the final Streamlit app. (One way to do this is to have a **References** section at the bottom of your app. Then the references should be of the form "This portion of the app was taken from [link]").

Rubric

1. (5 points) Is the project clear and well-organized? Does it explore one or more clearly described datasets, and include links specifying where those datasets came from? Use text throughout the project to help the reader understand what is going on. Is the reasoning clear? (It's fine if you have negative results like, "so there is no clear connection between these variables". In fact, those sorts of results are better than over-stated claims.)
2. (4 points) Does the project make essential use of pandas to explore the data? Using pandas to read and store the data is not sufficient. It should also be used either to clean the data, or to analyze the data.
3. (4 points) Does the project explore the data using a variety of tools from scikit-learn and/or Keras?
4. (4 points) Does the project include a variety of interesting charts, made using Altair? Do we learn something about the data from these charts?
5. (4 points) Does the project include *interactivity,* made possible through the use of multiple different Streamlit widgets? Are additional Streamlit tools used to improve the layout or flow or appearance of the app?
6. (4 points) Does the project include material that was not mention above? (This could include different libraries, different machine learning algorithms.)

Advice

Here is some general advice about the project.

- Don't repeat the same technique multiple times unless it's really essential. For example, it would be better to have one interesting chart and one less interesting chart, than to have the same interesting chart made for two different datasets.
- Don't spend too much time searching for the perfect dataset. Having a great dataset is less important for the project than exploring the dataset in an interesting way.
- Keep your statements reasonable. It's perfectly fine to conclude something like, "Therefore, we did not find a connection between A and B."
- I already wrote it above but it's worth repeating. Please err on the side of referencing everything. If your project is based on an idea you saw somewhere else, that's totally fine, but include a clear link in your app (not just in the source code) to the original source.

What should I focus on?

- The primary content should be from one or more of the Course tools (NumPy, pandas, Altair, scikit-learn, Keras, Streamlit). If you like Streamlit much more than pandas, for example, then it's okay to go more in-depth in the use of Streamlit, and less in-depth in the use of pandas.

Do I need to get original research results?

- No! If you explore the data in an interesting way, but can't find any interesting conclusions, that's fine. In fact, I prefer that to making claims that the data does not support.