Capstone Proposal

The Projected Rise of Food Insecurity

- *Why this topic? AKA why do you care, and why should anyone else?*
  - In a book referenced below, I read that for each degree of Celsius of warming over the pre-industrial era that the world experiences, crop yields drop by 10%. Combined with our current trajectory of 3.2 °C over the industrial era and the continued rise in population, this alarms me.
- *What industry/realm/domain does this apply to?*
  - Agriculture, humanitarian aid, and politics.
- *Who is your target audience?*
  - UN World Food Program and their prospective donors.
- *What impact would your answer/solution have on the real world, if your analysis were to be used/put into production?*
  - UN World Food Program is keenly aware that they need more money and resources already. The hope is that the analysis will help the UN better communicate the needs to donors and the general population.
- *What pre-existing projects/research/papers in this field have you explored, or what domain knowledge are you relying upon?*
  - Major resources I've explored:
    - IPCC Report Chapter 5: Food Insecurity
    - The Uninhabitable Earth by Davis Wallace-Wells
    - The Climate Book edited by Greta Thunberg and Others
  - My wife and I lived in Uganda for a year. As part of my time with the organization I worked for, Engineering Ministries International, I traveled to rural parts of Uganda and saw multiple consequences of merely financial insecurity, which contributes to food insecurity. Imagining food simply not growing at high enough rates would exacerbate the social ills I witnessed.
- *The motivation for the project is described. (Saying you needed to do a capstone project for Flatiron is not an appropriate motivation)*
  - World hunger is already a big enough problem. It creates migration and war, which leads to more migration, which leads to anti-immigrant populism, which leads to more war. Raising awareness of the oncoming exacerbation of these problems is important.

Data Understanding

- *What data will you collect?*
  - **Refugee Stats:**
    - All the refugee info you could ever want, split any way you want: UNHCR
    - More displacement data on refugees Internal-Displacement

        - A mega dive into the reasoning behind the shocking estimates of future climate refugees: Migration Data Portal
    - **Population Stats**
        - Population from 1500–2000: Our World in Data
        - Population from 1950–2021 and future projections: UN's Website, Dataset
    - **Average Global Temperature Stats**
        - Annual average global land-ocean surface temperature from 1880–2022 compared to the long-term average from 1951–1980. NASA, Dataset
        - A brilliant graphic from Bloomberg from back in 1995. Possibly update it? Project it towards the future? Bloomberg (Multiple datasets and links in this article)
    - **Greenhouse Gas Emissions**
        - Various emissions data by country and by source. Integrated Carbon Observation System
    - **Food Production**
        - Crop yields under three different climate scenarios: NASA via Data World
        - Another dataset regarding crop yields: Kaggle
- *Where did your raw data come from?*
        - Sources linked above.
    - *Is there a plan for how to get the data?*
        - The data is all available in downloadable format. A mixture of CSV and XLSX files.
    - *Are the features that will be used described clearly?*
        - There is documentation available that explains all the features in the data.
    - *Has anyone else worked on this specific problem/dataset? If so, how will your work build on theirs?*
        - Like many topics related to the climate crisis, this has been explored. However, since there is still not enough investment and policies to address this crisis, we cannot stop talking about it.

          I will add to the existing conversation by creating compelling graphics and a presentation that will get the attention of the public.

Data Preparation

- *In what form is the data stored?*
    - CSV and XLSX files.
- *What data types are the variables? Do you have any frequency counts or descriptive statistics yet?*
    - I don't have frequency counts or descriptive stats yet. I know that the majority of the data is in floats and integers. Most of the text data is for labeling purposes. For instance, in the data that lists population numbers, there's a column that defines which country the numbers apply to.
- *What kind of preprocessing steps do you foresee?*
    - Getting every dollar amount on the same inflation index and year. Combining the different types of data on each country into a single DataFrame.

- *What are some of the cleaning/pre-processing challenges for this data?*
    - Since I will be combining stats from different sources, different organizations will likely list different countries in their analysis. Do we really need population growth for Vatican City?
- *What is the minimum number of rows you will have? (this can be a ballpark estimate)*
    - There will be over 20K rows of data. The data separates different crops in different countries.
- *How are you planning to visualize the important aspects of this data to bring it to life?*
    - Line graphs. World maps. Bar charts to show the deficit of nutrition.

Modeling

- *What modeling techniques are most appropriate for your problem?*
    - Time-Series modeling.
- *What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)*
    - The target variable is the crop yields separated by country projected for 2050.
- *What model are you planning to use as a baseline?*
    - Baseline will be White Noise. The final model will likely be ARMA.
- *Is this a regression or classification problem?*
    - Regression.

Evaluation

- *What metrics will you use to determine success?*
    - AIC and BIC values and coefficients of ARMA.
- *What does the minimum viable product (MVP) involve? What is the smaller project that you can accomplish in a week that your overall project is based on?*
    - Minimum Viable Product will exclusively look at crop yields as the temperature has risen combined with the rise of the demand for food as the population rises.
- *What are your level-up stretch goals? How will you improve your project between MVP and presentation?*
    - Combine this information with estimates of the continued rise of greenhouse gas.

Deployment

- *Is the method for reporting final results described?*
    - The presentation will include a mixture of real-life photos illustrating hunger and charts produced through the analysis.
- *Is there a plan for deployment? (web app)*
    - I don’t have plans for a web application. It’s my understanding that providing documentation for reproducing my results will meet the reproducibility requirement.
- *What is the functionality?*
    - Raise awareness to the need for continued development funds to go to nations where high percentages of the population suffer from food insecurity. Raise awareness to eliminate the use of GHG emissions.

Tools/Methodologies

- *What are some of the Python libraries you are planning to use to gather, clean, explore, and model your data?*
    - Pandas, Numpy, ARMA, Matplotlib, Statsmodels, Arima
- *What modeling algorithms are you planning to use?*
    - White Noise, Random Walk, Arima
- *Where will you be performing your analysis - on your machine or in the cloud?*
    - My machine.
- *Will your data be stored on your machine or in the cloud?*
    - My machine, and uploaded to GitHub.