Capstone Proposal

The Ascending Cost Of Disaster Mitigation

Business Understanding

- *What problem are you trying to solve, or what question are you trying to answer?*
    - UN Crisis Relief needs to anticipate higher costs and scarcity of relief personnel and resources to provide relief for future crises.
- *Why this topic? AKA why do you care, and why should anyone else?*
    - Climatological disasters are unpredictable in location, scale, and date. However, the trends of disaster frequency and rising scale are predictable. Since we know disaster scale and frequency will rise as the earth gets hotter, we need to be prepared.
- *What industry/realm/domain does this apply to?*
    - Disaster readiness, humanitarian aid, medical relief, and politics.
- *Who is your target audience?*
    - UN Crisis Relief 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 crisis relief 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. Also, I hope the UN crisis relief organization will increase both stockpiles of supplies and disaster relief personnel.
- *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: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation
        - The Uninhabitable Earth by Davis Wallace-Wells
        - The Climate Book edited by Greta Thunberg and Others
    - As a Houstonian, I volunteered at the Astrodome when it housed Katrina refugees, and later had to flee the city myself due to the threat of Hurricane Rita.
- *The motivation for the project is described. (Saying you needed to do a capstone project for Flatiron is not an appropriate motivation)*
    - The anti-refugee sentiment that has grown in the Western world and America disturbs me right now in the present. The knowledge that the crisis will escalate as time moves on terrifies me. These uneasy feelings encourage me to use whatever skills I have to draw attention to the fact that we should be preparing to help more people.

Data Understanding

- *What data will you collect?*
    - **Disaster Stats:**
        - International Disaster Database: EM-Dat, Column Definitions
    - **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
- *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. Though there are estimates expressed as text of how much worse disasters will get, to my knowledge I have not seen this same data expressed in escalating line graphs.

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 multiple climate-related disasters, there's a column that defines whether it was a tornado, hurricane, etc…

- *What kind of preprocessing steps do you foresee?*
    - Culling which disasters are not relevant to the project. 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 10K rows of disaster data. I was able to isolate this to be exclusively climate-related disasters already, so I don't anticipate much of it going away.
- *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 costs.

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 cost of disaster relief 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 global temperature rise plus disaster frequency and cost.
- *What are your level-up stretch goals? How will you improve your project between MVP and presentation?*
    - Combine this information with estimates on population growth and 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 of disasters 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 of the need to prepare for disaster mitigation.

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.