

# Capstone Proposal

## What would it take to Meet the Paris Agreement Energy Production Targets?

### Business Understanding

- *What problem are you trying to solve, or what question are you trying to answer?*
  - The UN's [International Panel on Climate Change \(IPCC\)](#) needs to know if their warnings are being heeded, or if governments around the world are failing to implement their promised changes. If countries are not meeting their targets, then they need to know the best ways to get reduce greenhouse gas (GHG) emissions as soon as possible.
- *Why this topic? AKA why do you care, and why should anyone else?*
  - The ongoing threat of climate change scares me for the future more than anything else. I don't want to inherit a worse world than the one I was born into. I want to do everything I can to help humanity turn back from the perilous path we're on.
- *What industry/realm/domain does this apply to?*
  - Climatology, electricity production, humanitarian aid, and politics.
- *Who is your target audience?*
  - The [UN International Panel on Climate Change \(IPCC\)](#) and the heads of state they work with.
- *What impact would your answer/solution have on the real world, if your analysis were to be used/put into production?*
  - The IPCC needs to know that, despite their excellent work, the public still doesn't care enough about the climate crisis to demand change, and the politicians they elect act accordingly. 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 AR6 Report Climate Change 2023](#)
    - [The Uninhabitable Earth](#) by Davis Wallace-Wells
    - [The Climate Book](#) edited by Greta Thunberg and Others
- *The motivation for the project is described. (Saying you needed to do a capstone project for Flatiron is not an appropriate motivation)*
  - The enormity of the climate crisis is staggering. There is no other ill in our world of greater concern at this time. It's bigger than war, natural disasters, animal extinctions, pandemics, and poverty, combined because it will exacerbate all of these tragedies and more.

## Data Understanding

- What data will you collect?
  - **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](#)
  - **Energy Production / Consumption**
    - The chart and data motherload: [Our World Data](#)
    - Energy Production Wattage by Country by Source: [Our World In Data](#)
    - Energy Consumption Percentage Increases (Will probably go with projections below instead): [Our World In Data](#)
    - Electricity Demand by Country: [Our World In Data](#)
    - Death Rates by electricity source [Our World In Data](#)
    - Usable energy demand projection up until 2050 from IEA: [IEA](#)
  - **The 2030 target, keeping with the Paris Agreement\***
    - The Paris Agreement is a legally binding document signed by 186 countries intended to limit global temperature rise to 1.5 °C: [The UN website](#)
      - "The Paris Agreement is a legally binding international treaty on climate change. It was adopted by 196 Parties at the UN Climate Change Conference (COP21) in Paris, France, on 12 December 2015. It entered into force on 4 November 2016.  
"Its overarching goal is to hold "the increase in the global average temperature to well below 2°C above pre-industrial levels" and pursue efforts "to limit the temperature increase to 1.5°C above pre-industrial levels.""
    - In 2018, the gauntlet was thrown, to cut emissions by 45% from 2010 levels by 2030: [IPCC 2018 Special Report – Global Warming of 1.5 °C](#)
      - "C.1. In model pathways with no or limited overshoot of 1.5°C, global net anthropogenic CO2 emissions decline by about 45% from 2010 levels by 2030 (40–60% interquartile range), reaching net zero around 2050 (2045–2055 interquartile range)."
    - *\*The links in this bullet are not datasets. Instead, they are references for the targets that countries need to meet.*
- 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 a downloadable format. It comes in 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 labeling GHG emissions by country, there's a column of strings that identify which country the data is about.
- *What kind of preprocessing steps do you foresee?*
  - 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 emissions information 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 data.
- *How are you planning to visualize the important aspects of this data to bring it to life?*
  - Line graphs. World maps.

## 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 continued rise in temperature as Greenhouse Gases rise. There will be multiple projections showing how the model will change based on different scenarios of eliminated GHG emissions.
- *What model are you planning to use as a baseline?*
  - The 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?*
  - RMSE, AIC and BIC values and coefficients of ARMA. I'll use some test data to see if the model properly predicts historical data based on GHG emissions.
- *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 against greenhouse gases and compared with what countries agreed to in the Paris Agreement. It will then breakdown the different types of power generation rather than grouping all dirty and clean energy sources together and express where the most effective cuts can take place.
- *What are your level-up stretch goals? How will you improve your project between MVP and presentation?*
  - Expressing what types of power generation can feasibly replace the lost power generation due to eliminating certain types of power generation.

## Deployment

- *Is the method for reporting final results described?*
  - The presentation will include a mixture of real-life photos that show the current negative impacts of the climate crisis and charts produced during the analysis.
- *Is there a plan for deployment? (web app)*
  - I don't have plans for a web application. It's my understanding that exporting my Conda environment and providing documentation for reproducing my results will meet the reproducibility requirement.
- *What is the functionality?*
  - Hopefully, give the IPCC a new tool to raise awareness of the need to cut greenhouse gases as fast as possible.

## 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.

## Timeline:

- Mon, April 10th: All Data Cleaning completed
- Mon, April 17th: All modeling and visualizations completed
- Mon, April 24th: All items are 100% Complete and ready for presentation
- Mon, May 1st: Week of the presentation