**NAMES: Faris Alquaddoomi, Ethan Frisch, Megan Morrissey, Stella Zhou, Christian North**

**Problem Statement**

**Propose in the clearest way possible, a problem you would like to solve.**

We are currently designing a waste heat recovery (WHR) system as a solution to utilizing wasted heat at float glass factories. Our system will passively produce electricity from heat; we plan to charge for the electricity produced.

**Problem:** (1)We would like to determine how much to charge the float glass plant in comparison to the electricity price they currently pay from their electricity provider. (2) We would also like to quantify the expected revenue from each plant, assuming we can harvest 20% of the wasted heat and convert it into electricity.

Our WHR system will mitigate carbon dioxide emissions from float glass manufacturing. Assuming we can decrease natural gas fuel consumption by 25%, we can decrease carbon emissions by 20%.

**Problem:** (1) We would like to quantify how much carbon dioxide we can mitigate at each float glass plant and determine the equivalent amount of carbon credits associated with each mitigation scenario. (2) We then want to quantify the size of the carbon credit revenue stream and the opportunity at each float glass plant.

**Describe the project problem statement and argue how it can be tackled within the scope of the class. A few things to consider:**

Glass factory float lines generate significant amounts of wasted heat, which represents an opportunity for heat reclamation. The problem lies in converting the heat to useful electricity with an economically viable waste heat recovery system. Glass factories currently source electricity from the grid at competitive rates for industrial consumption. For our solution to be cost-competitive with electricity generated from fossil fuel plants, it must supply electricity at a lower comparative price.

Glass factories also generate significant carbon dioxide emissions. Our waste heat recovery solution will mitigate carbon dioxide emissions by reducing the amount of natural gas used by float glass factories. Float glass factories are granted a certain number of carbon credits based on their historic carbon dioxide emissions. Because we are reducing emissions from the float glass factories, we can sell off unused carbon credits as an additional source of revenue for our company. Consequently, we would like to predict carbon credit prices in the United States to quantify the size of our carbon credit revenue stream going forward.

**What questions would you like to explore?**

* How much carbon emissions does the glass industry contribute?
* How much electricity do glass plants use?
* What are the electrical costs for glass plants?
* How much can TPV technology save on electrical costs and CO2 emissions?
* How does this translate to carbon credits?
* What are the future trends for carbon credit prices?
* What are the future trends for fossil fuel plant electrical rates?
* What are the future trends for government emission regulations?
* What is the competitive landscape for waste heat recovery now and how can we position ourselves in low, medium, and high-end profit scenarios to be competitively priced for our electrical rates and our carbon mitigation?

**Why are these questions the right questions to consider?**

These questions take into account the amount of carbon dioxide emissions the glass industry produces during their manufacturing processes as well as the electricity used in the manufacturing processes. Knowing how much the waste heat recovery system can save on electricity will translate to the amount of carbon emissions that can be reduced. Knowing how much electricity we can save will directly factor into the profitability of a waste heat recovery system. Furthermore, knowing the amount of carbon credits we can sell from reducing carbon dioxide emissions will factor into the size of the expected carbon credit revenue stream.

**State previous attempts at this question if any and how your approach will be different. Note, you are expected to build on existing Kaggle notebooks and add novel approaches.**

Because our waste heat recovery system generates electricity for the glass plants, we want to charge them for electricity at a rate lower than what they currently pay. In order to know how much we can charge to remain competitive with on-grid electrical prices, we need to find out how much is currently being paid at each float glass plant; this depends on the geographic location of each plant, since electricity price varies by state.

We will approach predicting carbon credit prices through linear or polynomial regression (whichever is more accurate), sentiment analysis, and historical price trends in Europe. Previous attempts at answering these questions rely on interpreting planned government emission regulations. These attempts also rely on analyses of future climate change mitigation goals - in other words, they seek to answer how much regulation should be implemented to reach specific climate change mitigation goals. Our approach is different in that it is predictive, and it uses machine learning approaches to generating the predictions. Our approach is also prescriptive in that it influences our design of a novel waste heat recovery system based on predictions of a potential carbon credit revenue stream.

**Dataset**

**What datasets will you use to help address and derive insights upon answering questions in the problem statement? Do you need any additional datasets other than those in the example?**

Float Glass Factory Dataset:

*Source*

<https://members.glass.org/cvweb/cgi-bin/msascartdll.dll/ProductInfo?productcd=WOGFLOAT>

*Location*

<https://docs.google.com/spreadsheets/d/184kJd3PJ4hn2CUh795bUs3U1rP0pWbG-/edit?rtpof=true#gid=1366517478>

Electrical Costs Dataset:

*Source*

<https://www.eia.gov/electricity/monthly/xls/table_5_06_a.xlsx>

*Location*

<https://drive.google.com/file/d/1LlzoEiJHyTIZF-enDpgKA0eiupHm2TRK/view>

Carbon Pricing Dataset:

*Source*

<https://carbonpricingdashboard.worldbank.org/>

*Location*

<https://drive.google.com/file/d/1LlzoEiJHyTIZF-enDpgKA0eiupHm2TRK/view>

**Data Cleaning**

With the proposed datasets, list some steps you might need to carry out to get it to a form that will be more amenable to analysis. Please be as specific as possible to your dataset and less generic.

Float Glass Factory Dataset:

* US located plants
* Establish WCSS (Global coordinates) for each plant location
  + For mapping the location of the plants onto a US layer
* Largest plants (tons glass/size)
* Final product for analysis will be a pandas dataframe containing information about the plant location, number of float lines at the plant, geographic coordinates, and the cluster the plant belongs to.

Electrical Cost Dataset

* Pure electrical costs $/kWh
* Geographic coordinates for mapping cost of electricity to each state
* Dataframe containing electricity costs for industrial manufacturing delineated by state
* Final product will be a dataframe mapping states to their industrial electricity price ($/kWh)

Carbon Pricing Dataset:

* Filter data in terms of US and EU
* Organize data in terms of where carbon prices are the highest and the lowest
* One dataframe containing carbon credit prices mapped to their respective country/region for the US
* One dataframe containing carbon credit prices mapped to their respective country/region for the EU
* Final product will be the separate dataframes containing geographical information (US or EU; country/region in US or EU) and carbon credit prices as a time series (price each year since their introduction in a country/region)

**Visualizations**

**What plots would you graph as you explore your dataset? Here I want you to think about the plots that would be most useful for yourself to make sense of your dataset as it relates to your problem state.**

* Maps of US electricity prices for industrial manufacturers by state
* Maps of locations of US-based float plants and number of float lines per plant
* Overplotted maps of electricity prices and US-based float plants
* Interactive map of electricity prices and float lines that displays the expected revenue at each line given the electricity price in the state, the number of float lines, and the carbon credit revenue stream
* Histogram of electricity prices distribution in the US
* Linear or polynomial regression plots of US versus EU carbon credit prices
* K-means, DBSCAN, and hierarchical clustering visualizations of float plants in the US
* Elbow method and other validation methods plotted
* Sentiment analysis graphics showing positive or negative sentiment surrounding carbon credit price increase or decrease

**Ethical Considerations**

**What ethical concerns, if any, may arise as you consider your problem statement?**

* Incorrectly judging our electrical rate advantage over industrial rates could lead to our customers paying more than their best bet.
* TPVs Rare earth materials needed by mining: Gallium, Germanium
  + These materials are obtained through mining, but in small amounts. As TPVs are scaled this will become a greater ethical concern
* Ethical concerns that arise are CO2 emissions in regard to Global Warming.

**Who might be affected? Why?**

* Our customers
  + Overestimating will result in falsely optimistic forecasts on our revenue and could lead us to integrate more costly systems with our customers paying for the hardware upfront. This would predict an incorrect payback period on the customer’s investment and consequently return to them less over the described period.
  + Underestimating will result in falsely pessimistic forecasts which would lead us to adopt a more risk-averse approach to implementing our system. We might implement a more cost-effective system that trades performance for a lower payback period as described to your customers. Our customer would then break even within a shorter period as described and generate less profit past break even.
  + Carbon emissions affect customers through different factors including weather, water, air quality, and more.
* Our employees
  + Our employees stand to lose from incorrect predictions from our algorithms. Incorrect electrical rate advantage predictions would lower our profit, given our revenue is dependent on our customers’ proven cost savings.
  + Lower revenue to our company from inaccurate customer cost savings would adversely affect our reputation and our ability to gain more clients. This would result in lower future profits and potentially the inability to pay salaries. In the worst-case scenario, we would have to terminate certain positions and layoff employees to meet our margin requirements.
* Electricity companies
  + By predicting what the electric input would be needed for the power plants where we implement our technology. If predicted wrong we might need to pay for more power than we are taking from the grid.
* The World
  + Carbon emissions

**How can that be mitigated?**

* Utilizing advanced technology like TPVs, we can mitigate the effects of global warming.
* Valid data sets
  + Training set, testing set for data
* Research
* Code of ethics