Pipeline:

- [ ] Understand dataset and problem
- [ ] Write a quick proposal
    - [ ] 3 - 4 questions or goals we want answered
        - [ ] How do we plan on addressing these questions?
        - [ ] Provide a generic workflow
    - [ ] What is the motive?
    - [ ] What significance does the problem have?
    - [ ] How will this add value
        - [ ] Think ML
    - [ ] Discuss in detail what the data is
        - [ ] Is the data in one table or spread across many? How many columns? How many rows?
        - [ ] Is the data clean and ready to be analyzed or do you need to do it?
- [ ] Presentation
    - [ ] Create a slide deck
        - [ ] Templates: https://www.slidescarnival.com/
        - [ ] Follow the BLUF method
    - [ ] If time permits, create a simple website with all of the information
    - [ ] Record a video of the presentation
        - [ ] We only have **5 minutes** to present

BLUF METHOD:

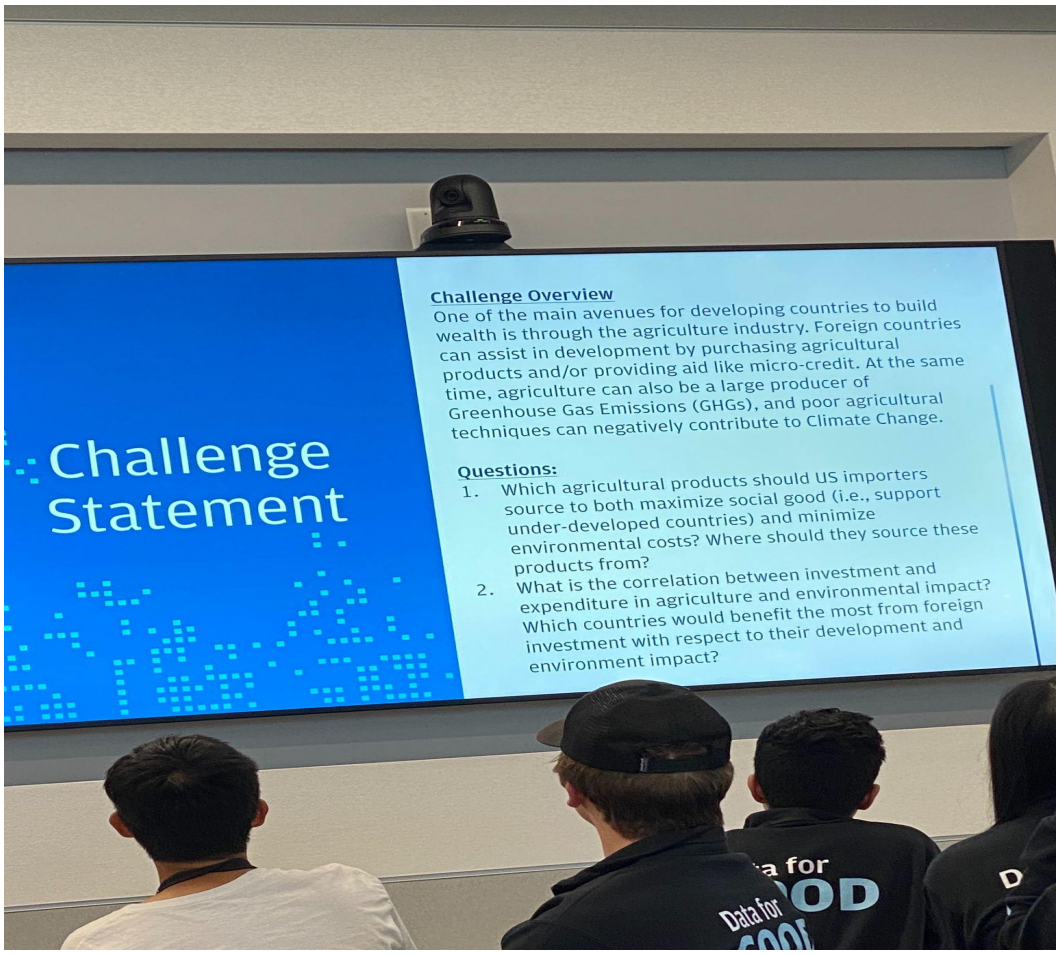
- [ ] Key findings
- [ ] Project Context
- [ ] Datasets
- [ ] Exploratory Data Analysis
- [ ] Key Project Object #1
- [ ] Key Project Object #2
- [ ] Key Project Object #3
- [ ] Future Work
    - [ ] Objective → Finding → Future Action
- [ ] Project Continuation

CODING PLAN / IDEAS!!!!!!

- [ ] Join agriculture and employment datasets on one of the same variables (natural join)

Notes about the case:

- Environmental concerns
- Helping non-profits
- We can find additional datasets (just keep country in mind)
- Run time optimization does not matter
    - Do not OVEROPTIMIZE
- Investment = foreign dollars spent



Libraries we're using:

- Pandas

https://git-scm.com/book/en/v2/Git-Branching-Basic-Branching-and-Merging

Virtual Server: https://bit.ly/3k0OmEB
(Scott and Kiana will help if password problems)

Usernames:
user-1: izzy
user-2: patricia
user-3: kevin
user-5: mentor
user-7: meghana
User-8: nikan

Sleeping Schedule:

- ☐ Nikan → 10 hrs
- ☐ Izzy → 2.5 hrs
- ☐ Meghana → 3 hrs/ 1.5 hr splits → 2 am, 7am
- ☐ Kevin → No sleep he came late
- ☐ Patricia → 1 hr repeated 3 times

Schedule:

- **MAKE SURE PROJECTS ARE SUBMITTED BY 12PM SATURDAY**
- First round of judging from **12 - 1:15 pm**
- Final top 4 presentation from **1:30 - 2 pm (we got this!!!)**
- **5 minute** presentations

Timeline:

Time	Key Goals
5 PM	Plan Check in with mentor if our plan is good and what we should add or remove
6 PM	EDA - for data visualizations Code
7 PM	Dinner
8 PM	
9 PM	9:00-9:30: Reconvene and discuss problems + new ideas Check in with mentor about right path
10 PM	
11 PM	Create team project page (so we don't have to worry about it tmr)
12 AM	Midnight pizza
1 AM	
2 AM	If mentor is awake: check in with mentor
3 AM	
4 AM	
5 AM	
6 AM	Check in with mentor on our progress
7 AM	CHANGE GRAPH NAMES TO RATES OF POVERTY RATHER THAN POOR PEOPLE
8 AM	Breakfast! Create a slidedeck for presentation Make templates so all we have to do is fill stuff in
9 AM	Start adding data to presentation
10 AM	
11 AM	Split up presentation slide content

	Write out speaking parts for presentation 11:30 - practice
12 PM	CODE SUBMISSION ON https://c4g.bemyapp.com JUDGING STARTS
1 PM	
2 PM	2:15- Closing ceremony
3 PM	

Understanding the Questions:

1. Using both economic and environmental data, focus on the following questions:
   1) Which agricultural products should US importers source to both maximize social good (i.e., support under-developed countries) and minimize environmental costs?
   2) Where should they source these products from?

- What are we defining as under-developed countries?
  - Use poverty rate csv: find average of poverty rates, compare that to each country

- What is our definition of maximization?

- Minimize environmental costs:
  - Find average of co2eq, compare to other data points, get countries with lowest CO2eq
  - Lowest temperature change

2. What is the correlation between investment and expenditure in agriculture and environmental impact? Which countries would benefit the most from foreign investment with respect to their development and environment impact?

Visualizations:

1. Regression Line graph
   a. X = Poor people
   b. Y = Aggie rate

2. Bar Graph

    a. X = country name
    b. Y = Poor peeps

3. Line graph
    a. X = Population
    b. Aggies

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline


mew = new.copy()
mew = mew.sort_values(by='PPeeps', ascending=True).head(5)

plt.figure(figsize=(10,5))
sns.barplot(x="Area",y = 'PPeeps', data=pew).set_title(label='Top 5 countries with Highest # of Poor
People');

agCopy = agData.copy()

def graphbyYear (Year, CO2Int):
        largest = agData[agData['Year']==Year].nlargest(3,CO2[CO2Int])
        smallest = agData[agData['Year']==Year].nsmallest(3,CO2[CO2Int])
        ax = largest[[CO2[CO2Int],'Area']].plot(kind='bar', figsize=(15, 10), legend=True,
fontsize=12)

sns.barplot(x="Area",y = 'CO2[CO2Int]', data=largest ).set_title(label='Top 5 countries with Highest # of
Poor People');
```

Where we are:

- We have CO2 emissions general trends
    - products with smallest emissions
- We have poverty rates correlations w/ amount of agriculture jobs

Social Good Metric:

Environmental Impact Metric:

- sum last 5 years CO2 emissions of country x/ sum of ALL countries CO2 emissions

Sorting Algorithm:

- Takes in Social Good Metric and Environmental Metric (combine the 2 dataframes on country name?)
- Process:
    - has list of rankings of different agr. products in terms of CO2 emissions
    - Picks top 5 countries to invest in based on which country has the (highest?lowest?) overall metric #
    - on EACH country, runs an algorithm to determine the lowest CO2 product that country can reasonably produce
        - Reasonably produce = country produces (in tons) is in the at least the 25% quartile
- Output:
    - Dictionary or Pandas
    - {Country Name: Crop to Invest in}
    - or?

CountryName	CropName	SocialMetric	Environmental Metric	Overall Metric

    -