**Group Project 1 Proposal**

Nupur Murthy, Randy Payano, Maria Jose Siles Navarro, and Victor Dituro

Our proposal is to analyze data collected by the World Bank to produce visualizations highlighting differences between global economies and a scoring system on a variety of Environmental, Social, and Governance factors.

We will utilize the World Bank API to create a database of factors we have selected.

Methodology:

Create a standard score to measure relative ESG performance across the globe

* Environmental, Social, and Governance factors will be weighted equally
* Environmental measures focus on Pollution, Clean Energy, and Wildlife Protection
* Social measures focus on Education, Equality, Health, and Poverty
* Governance measures focus on Corruption, Rule of Law, Human Rights, and Ease of Doing Business
* Country wealth creates a higher expectation of performance

Questions:

* Where are the greatest improvements and deterioration of Poverty happening?
  + Social: Scatter Plot of Infant Mortality over time
* What countries are serial polluters?
  + Environmental: Top CO2 emission producers
* What countries have the worst record on law and human rights?
  + Governance: Correlation of Low Rule and Law to High Human Rights violations
* What countries are making the greatest move to clean energy and which are not?
  + Environmental: Top CO2 emission producers
* World Map of our ESG Score and the most responsible countries
* Regression of Global Scores to View Trend Line – is the world getting better?

Datasets:

<https://www.quandl.com/data/WWDI-World-Bank-World-Development-Indicators/documentation>

<https://www.quandl.com/data/WWGI-World-Bank-Worldwide-Governance-Indicators/documentation>

<https://datacatalog.worldbank.org/dataset/doing-business>

What Measures:

Pollution

Organic water pollutant (BOD) emissions (kg per day) EE\_BOD\_TOTL\_KG

CO2 emissions (kt) EN\_ATM\_CO2E\_KT

We recommend focusing your analysis on techniques such as aggregation, correlation, comparison, summary statistics, sentiment analysis, and time series analysis.

\* Use Pandas to clean and format your data set(s)

\* Create a Jupyter Notebook describing the \*\*data exploration and cleanup\*\* process

\* Create a Jupyter Notebook illustrating the \*\*final data analysis\*\*

\* Use Matplotlib to create a total of 6-8 visualizations of your data (ideally, at least 2 per "question" you ask of your data)

\* Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation

\* Optionally, use at least one API, if you can find an API with data pertinent to your primary research questions

\* Create a write-up summarizing your major findings. This should include a heading for each "question" you asked of your data, and under each heading, a short description of what you found and any relevant plots.

Goals:

* Learn how to extract data
* Select indicators
* Create database
* Create visualizations
  + How to plot a World Map Heatmap?
  + How to Plot regression?
  + How to create an interactive scatter plot?
* Presentation

**Presentation**

**Title**

Include the name of the Project and Group Members

**Motivation & Summary**

\* Define the core message or hypothesis of your project.

\* Describe the questions you asked, and why you asked them

\* Describe whether you were able to answer these questions to your satisfaction

\* Briefly summarize your findings

**Questions & Data**

\* Elaborate on the questions you asked, describing what kinds of data you needed to answer them, and where you found it

**Data Cleanup & Exploration**

\* Describe the exploration and cleanup process

\* Discuss insights you had while exploring the data that you didn't anticipate

\* Discuss any problems that arose after exploring the data, and how you resolved them

\* Present and discuss interesting figures developed during exploration with Jupyter Notebook

**Data Analysis**

\* Discuss the steps you took to analyze the data and answer each question in your proposal

\* Present and discuss interesting figures developed during analysis with Jupyter Notebook

**Discussion**

\* Discuss your findings. Did you find what you expected to find? If not, why not? What inferences or general conclusions can you draw from your analysis?

**Post Mortem**

\* Discuss any difficulties that arose, and how you dealt with them

\* Discuss any additional questions that came up, but which you didn't have time to answer

**Questions**

\* Open-floor Q&A with the audience