MGT 6203 Group Project Proposal

Team #: 4

Team Members:

1. **Name**: Liam O'Donnell
    - **EdX username:** lodonnell930
    - **Professional Background:** Project Management on large construction sites. Worked within mining, heavy industrial, and automotive industries
    - **Educational Background:** Civil Engineering from Iowa State University (Graduated S2021)
    - **Previous Analytics Projects:** NLP Invoice Classification and XGBoost construction production forecasting.
2. **Name:** Will Coughlin
    - **EdX username:** wcoughlin
    - **Professional Background:** 3.5 years in software engineering. Worked in fintech and now in energy.
    - **Educational Background:** Computer Science at Auburn University (2019)
    - **Previous Analytics Projects:** Factors in printed circuit board (PCB) trace lifespan (school), Cryptocurrency sentiment-price visualization and regression analysis (personal).
3. **Name:** Joshua (Josh) Geiger
    - **EdX username:** gte975n
    - **Professional Background:** I have 20 years of product design, test, and validation experience in the automotive industry.
    - **Educational Background:** I have a bachelor's degree in mechanical engineering from Georgia Tech (SP2004), a master's degree in engineering with a focus in integrated vehicle systems from Purdue University (SP2008). I have been accepted to the OMSA program at Georgia Tech with a Fall 2024 starting semester.
    - **Previous Analytics Projects:** My previous analytics projects have focused on descriptive analytics in passenger car tire design working on many extract, transform, and load pipelines for disparate datasets to create dashboards aimed at creating job efficiency by connecting datasets.
4. **Name:** Stephanie Poole
    - **EdX username:** Poole1
    - **Professional Background:** 10 years of experience within the Oil & Gas industry spanning High Performance Computing applications and infrastructure support, reservoir engineering supporting subsurface workflows and hydraulic engineer modeling fluid flow behavior under steady state and transient operation.
    - **Educational Background:** Bachelor's degree in Chemical Engineering, with minor in Mathematics (2014). I have taken several EdX & coursera courses in Machine Learning and languages. Currently enrolled in the MicroMasters program from EdX and Georgia Tech and plan on pursuing the full masters in Data analytics from Georgia Tech.

- **Previous Analytics Projects:** Most analytics projects have been within industry performing regression analysis / machine learning models based on features when physics-based models no longer work.

5. **Name:** Jonathan Ho
    - **EdX username:** jonathanho42
    - **Professional Background:** I have been teaching Mathematics/Statistics at community colleges since 2012.
    - **Educational Background:** I have a bachelor's degree in Theoretical Physics from the University of Puerto Rico and a Master's degree in Pure Mathematics from The University of Iowa**.** I also completed three semesters of a master's program in Statistics but did not complete it. I completed ISYE 6501 in SU 23.
    - **Previous Analytics Projects:** Several short projects in Stats graduate courses. These projects were on multiple linear regression, variable selection, coding Buffon's Needle Problem in R to estimate the value of pi, etc.

Project Title:

Analyzing the Possible Influence of a Developed Country's Energy Sources and Demand on Neighboring Developing Countries

Background Information on chosen project topic:

Energy production and consumption have been a politically and emotionally driven issue, especially within the last decade. Energy has broad impacts on the human experience due to its impact on both the local economy metrics via job and infrastructure while offering the potential for improved quality of life through market product creation and availability. As energy demand increases, production must increase, but energy can come in many forms with varying degrees of environmental impact and efficiency. Many countries have made commitments to evolve energy production towards more environmentally favorable sources. Will an evolution in energy production towards clean energy in highly developed countries impact developing countries?

Problem Statement:

Through analysis of global energy production and Gross domestic product (GDP) data, our group intends to investigate and model if energy:

1. Sources and consumption have evolved globally
2. 'Habits' of developed countries affect neighboring less developed countries

State your Primary Research Question (RQ):

How do energy source changes for developed countries affect less developed countries?

Supporting Research Questions:

1. What effect does changing the primary energy source have on the country's GDP?
2. What is the projection for energy use by source for the year 2050?

Business Justification:

Countries in the same region may exhibit similar trends in energy consumption by type over time. Government and private energy sectors would be interested to understand the relationship between energy and GDP to drive investment.

If a developed country's energy consumption and production type can impact a less developed neighboring country's GDP, partnerships and investments can be strategically optimized. By analyzing these trends, leaders of less developed countries could also focus on which alternative sources of energy to pursue and predict potential demand and GDP increases.

Nonprofits in the energy sector could use the data to determine opportunities in energy source type by country to drive investment and development strategy.

Data Sources (links, attachments, etc.):

1. Hosted copy of all datasets: https://1drv.ms/f/s!AidNWjxCQLvGlGqTcG3Mksmakbde
2. Links to original sources:
    a. https://ember-climate.org/data-catalogue/yearly-electricity-data/
    b. https://ourworldindata.org/grapher/gdp-per-capita-worldbank?tab=table
    c. https://hdr.undp.org/data-center/documentation-and-downloads

Data Description:

1. This dataset includes information regarding the capacity, demand, and generation of energy (by type) for each country in the world since the year 2000. This dataset will be the primary data backbone of our analysis.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Area	Country cc	Year	Area type	Continent	Ember regi	EU	OECD	G20	G7	ASEAN	Category	Subcatego	Variable	Unit	Value	Y
2	Afghanista	AFG	2000	Country	Asia	Asia	0	0	0	0	0	Capacity	Aggregate	Clean	GW	0.19	
3	Afghanista	AFG	2000	Country	Asia	Asia	0	0	0	0	0	Capacity	Aggregate	Fossil	GW	0.03	
4	Afghanista	AFG	2000	Country	Asia	Asia	0	0	0	0	0	Capacity	Aggregate	Gas and O	GW	0.03	
5	Afghanista	AFG	2000	Country	Asia	Asia	0	0	0	0	0	Capacity	Aggregate	Hydro, Bio	GW	0.19	

2. This dataset shows GDP over time for each country, dating back to 1990. We plan to use this dataset to associate GDP with energy consumption at the country level. We then plan to use this relationship for causal studies and inference.

	A	B	C	D	E	F	G	H
1	Entity	Code	Year	GDP per capita, PPP (constant 2017 international $)				
2	Afghanista	AFG	2002	1280.463				
3	Afghanista	AFG	2003	1292.334				
4	Afghanista	AFG	2004	1260.061				
5	Afghanista	AFG	2005	1352.321				

3. This dataset contains time series for Human Development Index (HDI) since 1990 for every country in the world. HDI may serve as an indicator for the "development" or wealth of a country, which may be used to correlate energy production and consumption over time.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	iso3	country	hdicode	region	hdi_rank_2	hdi_1990	hdi_1991	hdi_1992	hdi_1993	hdi_1994	hdi_1995	hdi_1996	hdi_1997	hdi_1998	hdi_1999	hdi_2000	hdi_2001	hdi_2002	hdi_2003	hdi_2004	hdi_2005	hdi_2006	hdi_2007	hd
2	AFG	Afghanista	Low	SA	180	0.273	0.279	0.287	0.297	0.292	0.31	0.319	0.323	0.324	0.332	0.335	0.337	0.362	0.376	0.392	0.4	0.409	0.424	
3	AGO	Angola	Medium	SSA	148										0.364	0.375	0.386	0.403	0.42	0.433	0.447	0.459	0.475	
4	ALB	Albania	High	ECA	67	0.647	0.629	0.614	0.617	0.624	0.634	0.645	0.642	0.657	0.669	0.677	0.684	0.689	0.696	0.7	0.711	0.718	0.73	
5	AND	Andorra	Very High		40											0.818	0.825	0.832	0.841	0.833	0.833	0.848	0.847	
6	ARE	United Ara	Very High	AS	26	0.728	0.739	0.742	0.748	0.755	0.762	0.767	0.773	0.779	0.787	0.796	0.8	0.804	0.814	0.818	0.822	0.827	0.831	
7	ARG	Argentina	Very High	LAC	47	0.723	0.73	0.735	0.739	0.744	0.745	0.751	0.756	0.762	0.773	0.779	0.784	0.785	0.793	0.798	0.802	0.814	0.817	

Key Variables:

- Dependent variable(s): energy production or demand by energy type for less-developed nations.
- Independent variable(s): energy production or demand by energy type for higher-developed nations, GDP by country by year, indicator variables for country clustering, and HDI by country by year.

Planned Approach

The first steps in our approach will need to be in an effort to clean and join/merge the data. Where there is missing data we will need to determine if we want to remove the data points or impute the missing data. The next step will be to connect the different datasets through identifying the correct key (like country). The goal is to get a flat set of data that can be analyzed across different domains (for example energy consumption and gdp).

Once the data is cleaned, transformations may be necessary to guarantee that the required conditions of the chosen analytical models are satisfied. Upon further inspection of the data we may also decide to normalize or standardize certain predictors. This could be done for model performance or for model interpretability.

A clustering methodology will need to be created to determine what is considered a neighboring country. Similarly, we'll need to determine how to classify a country as higher developed and lower developed.

In any model we create, we plan to split the data into training and test datasets. Cross validation will be used where applicable to estimate the true performance of the models. The hyperparameters of the models will be determined by using the predicted performance using the training and test datasets targeting model performance without overfitting.

Anticipated Conclusions/Hypothesis:

It is anticipated that we will see a positive correlation between energy production/consumption and GDP. This correlation will be stronger when the source of the GDP is primarily from energy heavy sectors like manufacturing and weaker when the GDP is primarily from less energy heavy sectors like services.

It is also anticipated that as a country in a region increases energy availability that neighboring countries also see an increase in energy production/consumption.

For countries where GDP is not high, we expect to see a quality of life increase due to the increased availability of energy.

It is anticipated that the GDP and HDI increase at a higher rate for countries that install and utilize cleaner energy production sources.

Potential Business Decisions and Impacts:

If the data and models align with our expected results, it could create a predictive roadmap for countries and companies seeking to invest in the energy sector. It could provide suggestions on which countries to invest in and which type of clean energy to invest in.

If the data suggests a strong relationship between an increase in GDP with an increase in clean energy production, the environmental impacts could be positive locally and globally on different time scales.

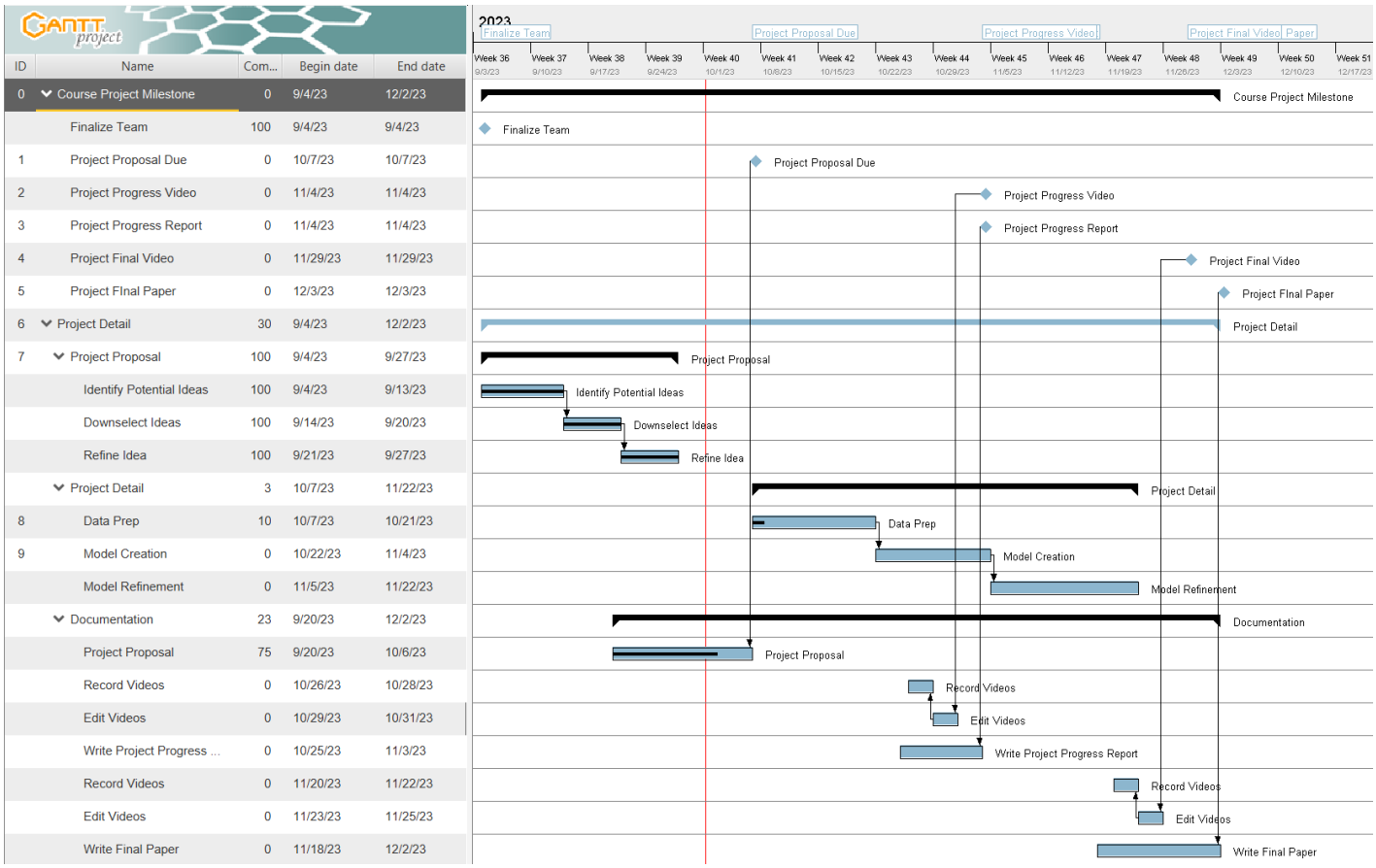
Project Planning:

For efficiency of task division and specialization to support project deliverables, we have organized into the following roles:

- Project Leader: Will Coughlin
- Data Engineer: Will Coughlin
- Business Data Analyst: Stephanie Poole, Liam O'Donnell
- Data Analyst: Jonathan Ho, Josh Geiger

Team has a 1 hour standing weekly meeting on Wednesday nights. Other meetings to be called between affected team members as needed.

Project Timeline:



ID	Name	Com...	Begin date	End date
0	Course Project Milestone	0	9/4/23	12/2/23
	Finalize Team	100	9/4/23	9/4/23
1	Project Proposal Due	0	10/7/23	10/7/23
2	Project Progress Video	0	11/4/23	11/4/23
3	Project Progress Report	0	11/4/23	11/4/23
4	Project Final Video	0	11/29/23	11/29/23
5	Project FInal Paper	0	12/3/23	12/3/23
6	Project Detail	30	9/4/23	12/2/23
7	Project Proposal	100	9/4/23	9/27/23
	Identify Potential Ideas	100	9/4/23	9/13/23
	Downselect Ideas	100	9/14/23	9/20/23
	Refine Idea	100	9/21/23	9/27/23
	Project Detail	3	10/7/23	11/22/23
8	Data Prep	10	10/7/23	10/21/23
9	Model Creation	0	10/22/23	11/4/23
	Model Refinement	0	11/5/23	11/22/23
	Documentation	23	9/20/23	12/2/23
	Project Proposal	75	9/20/23	10/6/23
	Record Videos	0	10/26/23	10/28/23
	Edit Videos	0	10/29/23	10/31/23
	Write Project Progress ...	0	10/25/23	11/3/23
	Record Videos	0	11/20/23	11/22/23
	Edit Videos	0	11/23/23	11/25/23
	Write Final Paper	0	11/18/23	12/2/23