# *End-to-end data analysis project – PROPOSAL*

*Data Analytics – Andreea Ioana Mihai*

- ***What dataset or datasets do you plan to use? What are the features, rows, and data types of each? (Note: The data that you use shouldn't be from the curriculum.)***

I am using two **datasets** found on Github.com:
https://github.com/owid/co2-data/blob/master/owid-co2-data.csv
https://github.com/hongtaoh/olymvis-data/blob/master/data_sources/continent.csv

All the data is scraped from the database of Our World In Data:
https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions
Our World in Data (OWID) is a scientific online publication that focuses on large global problems such as poverty, disease, hunger, climate change, war, existential risks, and inequality.

**The main dataset contains 23708 rows,** each row representing a country by year and **55 columns** with a collection of key metrics about **$CO_2$**(coal, oil, cement, gas, flaring, other industries) and **Greenhouse Gas**(methane, nitrous oxide, primary energy consumption) **emissions** for all the countries worldwide since year of 1750 until 2019. The second dataset contains **6 columns** containing data about continents and country codes.
New data is published on an annual basis.

- ***What research or business questions do you want to answer?***

Human emissions of carbon dioxide and other greenhouse gases – are a primary driver of climate change – and present one of the world's most pressing challenges.
**My main research questions** focus around looking at what is driving $CO_2$ and Greenhouse Gas(GHG) emissions the most: population, fossil fuels, or energy-intensive industries? And how these emissions are changing in each country, which are the countries with most emissions and how much they affect the climate change worldwide, so as a result the countries will implement commitments to combat this major problem.

- ***What are your hypotheses going in? How will you use your data to test your hypotheses?***

**My goal** is to review total $CO_2$ and Greenhouse Gas emissions for USA and China over the 5 year period (from 2012 to 2016) in order to determine if there is an increase in $CO_2$ emissions as well if there is an increase in Greenhouse Gas emissions between United States and China, the most powerful countries in the world, as as this would show growing need to adopt legislation and programs by the countries and could have greater impact on slowing the rate and limit the amount of global warming.

**Hyphothesis 1: Carbon dioxide ($CO_{2)}$ emissions.**

$H_o$: There is no statistically significant difference in the mean $CO_2$ emissions for United States vs. China.

$H_a$: There is statistically significant difference in the mean $CO_2$ emissions for the year of United States vs. China.

**Hyphothesis 2: Greenhouse Gas (GHG) emissions.**

$H_o$: There is no statistically significant difference in the mean CHG emissions for United States vs. China.

$H_a$: There is statistically significant difference in the mean GHG emissions for United States vs. China.

***Who will find your findings valuable, and how will they use them?***

United States and China's governments and people will benefit from the findings, so they can make progress in changing programs and policies to reduce emissions and suppress the climate change. Not only them, all the countries can benefit from the analysis.