Project #1 – Group 8

Comparing available SIC categories for analysis project

Below are snips from the three spreadsheets that were being reviewed.

* \_COE28282\_Emissions2C\_Emissions\_Intensities2C\_and\_Emissions\_Multipliers.csv
* final\_raw\_sample\_0\_percent.csv
* final\_raw\_sample\_3\_percent.csv

Will need to choose which ‘final\_raw\_sample’ file will be used. The 3% file has a number of duplicated SIC categories, which I chose to avoid and hopefully avoid complications.

I detailed on the Information sheet how I narrowed down the selection, if interested. The final 8 categories were determined by:

1. Category had to be present in each csv file.
2. Category had to have more than 8 countries in the dataset (for matching other file data such as GDP later).
3. Category had to have data from 2010 to 2018.
4. The categories in green have a greater magnitude, hopefully to show greater effects. There are 5 green categories.
5. The categories in yellow are ‘backups’ with less data magnitude. Chosen if 8 clusters to be analyzed (total).

Many columns of data were removed to narrow down the categories that may have matches to fields in other datasets. Indexing of SIC and country codes to show environmental impact in dollars, carbon dioxide emissions, gross domestic product, population data.

MOD-final\_raw\_sample\_0\_percent.xlsx below:

Graphical user interface, application

Description automatically generated

MOD-final\_raw\_sample\_3\_percent.xlsx below:

Graphical user interface, table

Description automatically generated with medium confidence

MOD\_COE28282\_Emissions2C\_Emissions\_Intensities2C\_and\_Emissions\_Multipliers.xlsx:

Graphical user interface, text, table

Description automatically generated with medium confidence

I removed the emission intensity and emission multiplier values as they were grouped with the emissions and were using different units. That would mess up any dimensional analysis.

The spreadsheets have pivot tables that can filter for various SIC categories. The above was my attempt to find something highly compatible with other data sets.

**Need to choose what SIC categories will be used, how many, and keeping the 2010 to 2018 time frame?**