Dan, Eric, Irina

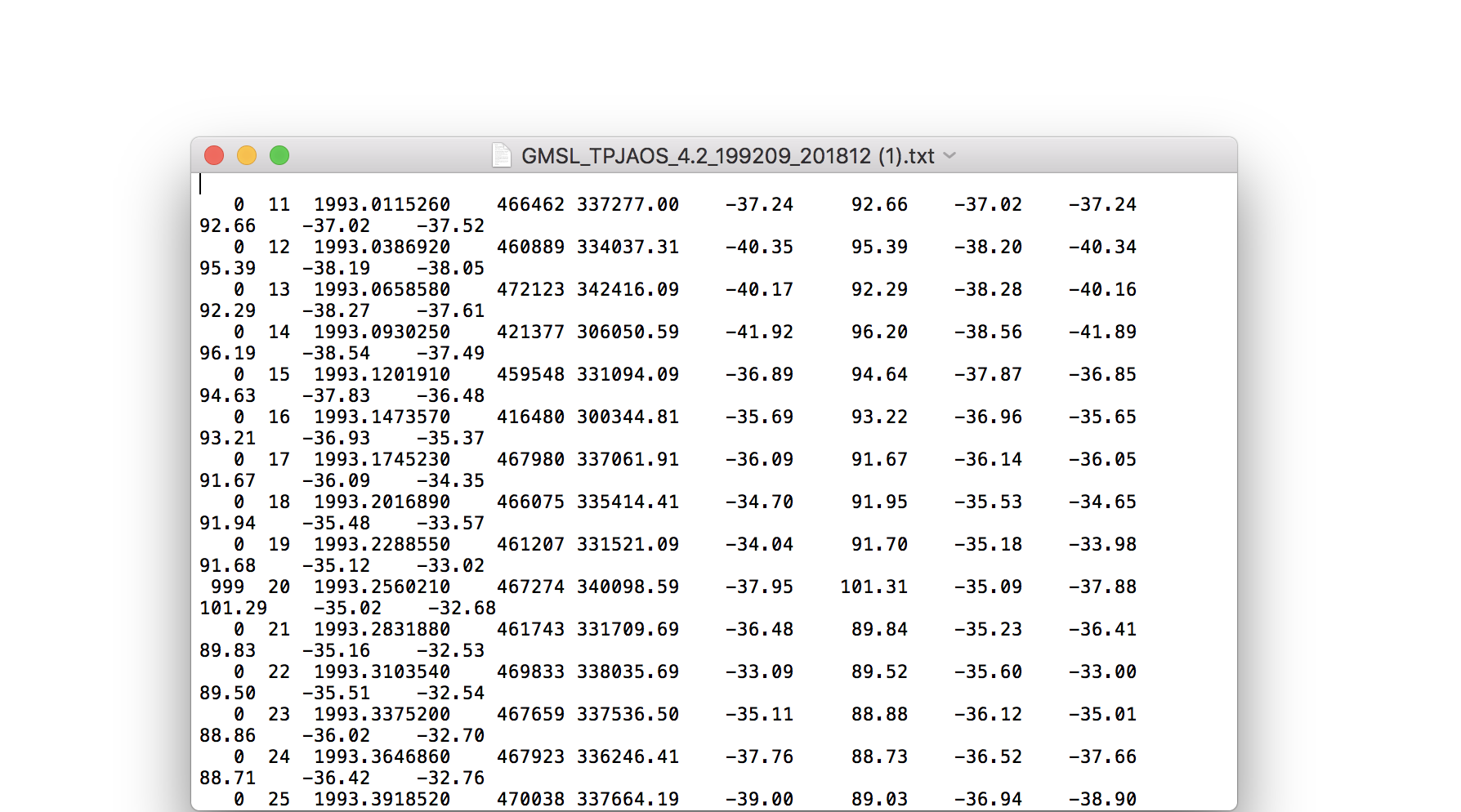
GLOBAL WARMING ETL PROJECT

Data Sources & Cleaning Process

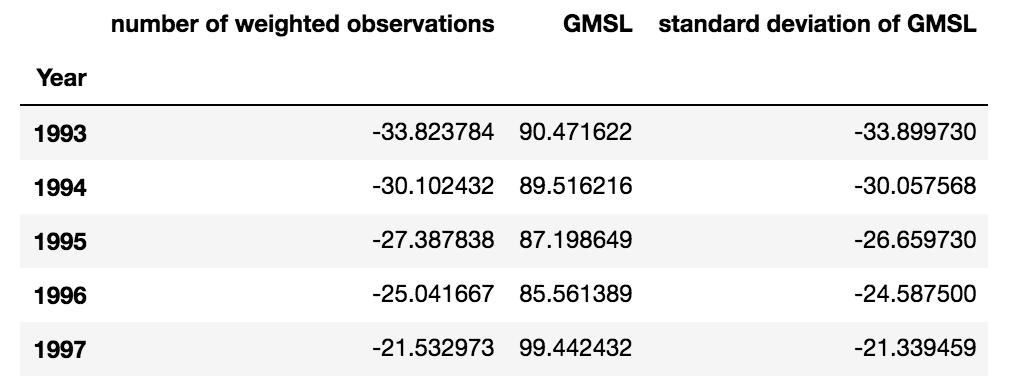
For our project, we were looking for data sources that could be used to analyse the effects of Global Warming. We searched various websites and we were able to find the following data sets.

1. **Sea Level Rise**

This data set was found on the [NASA website](https://climate.nasa.gov/vital-signs/sea-level/) and we retrieved it as a text file. The method we use to clean this file was to load it into jupyter notebook and shape the file into a tabular database. The columns in the text file are defined by their positions where position 0 starts where the first red arrow has been marked and the last position where the second red arrow is marked. The values that we used are the found under indexes 2,5,6,7 which represent Year, Global Mean Sea Level (GMSL), Standard Deviation of GMSL, Smoothed (60-day Gaussian type filter) respectively.

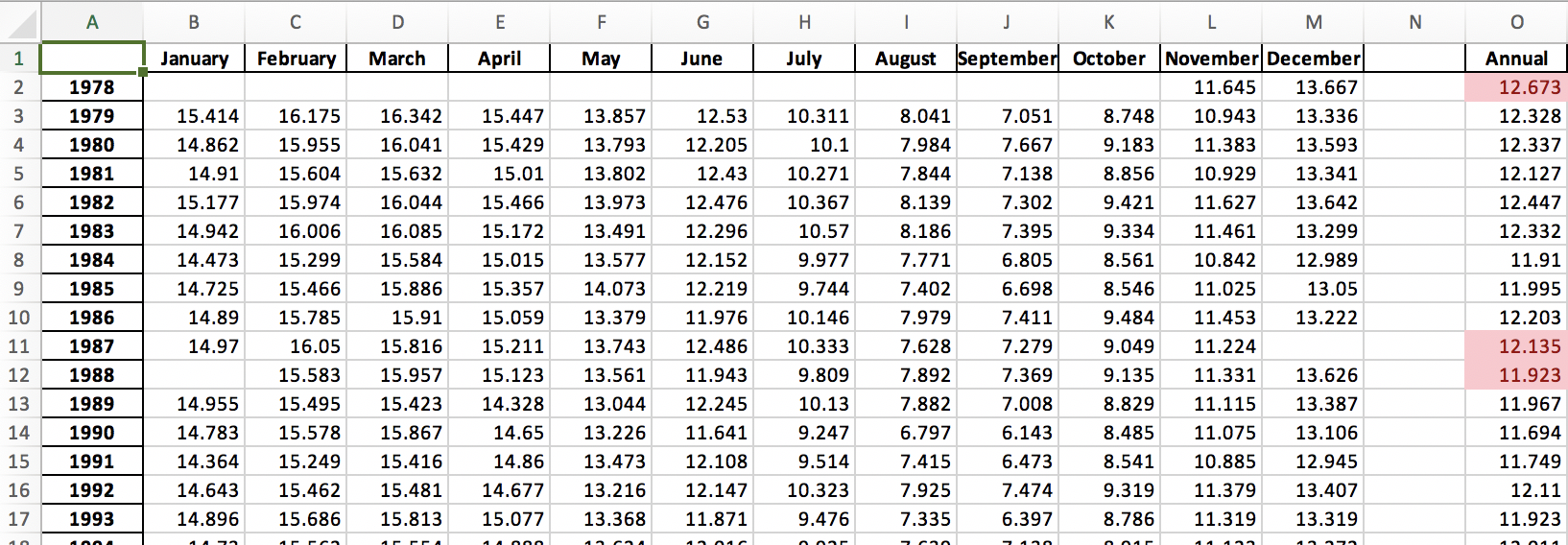


In order to format the data the way in which would serve us best for analysis, we had to perform a number of functions. Firstly, the years are given in decimals so we needed to use “pandas.Series.str.split” function within the pandas library to split the years from their decimals. Then we needed to set the column names to label what the values were representing as opposed to index values. The final product appeared as:



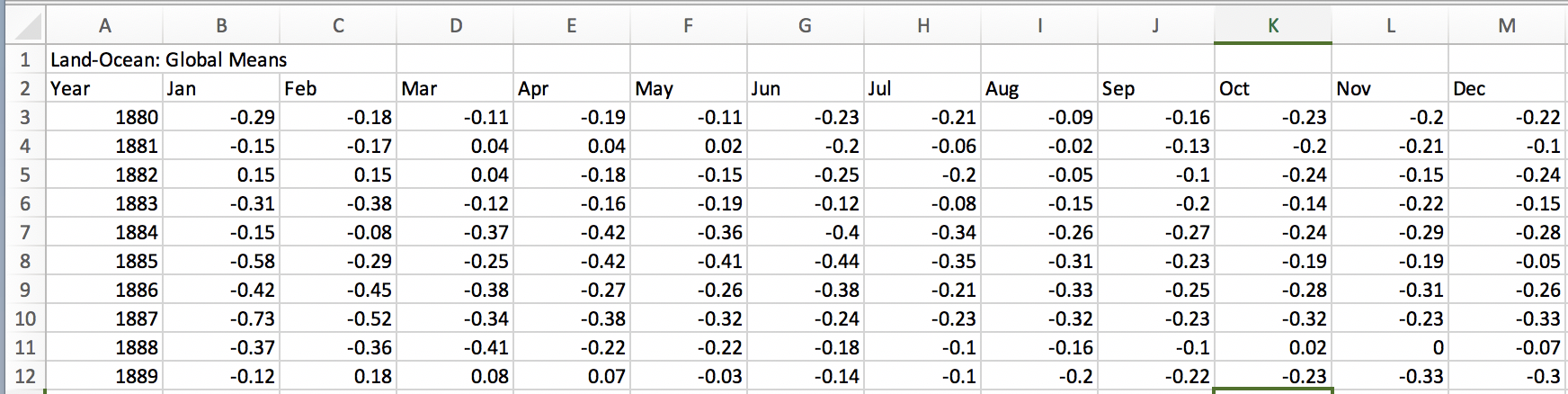
1. **Sea Ice Index**

The sea ice index dataset was a CSV file that was found on the [National Snow & Ice Data Center](https://nsidc.org/data/seaice_index). We loaded this file into pandas with “pd.read” function and from there we were able to drop the columns we do not need such as the sea level index and just take the Year and Annual Columns.



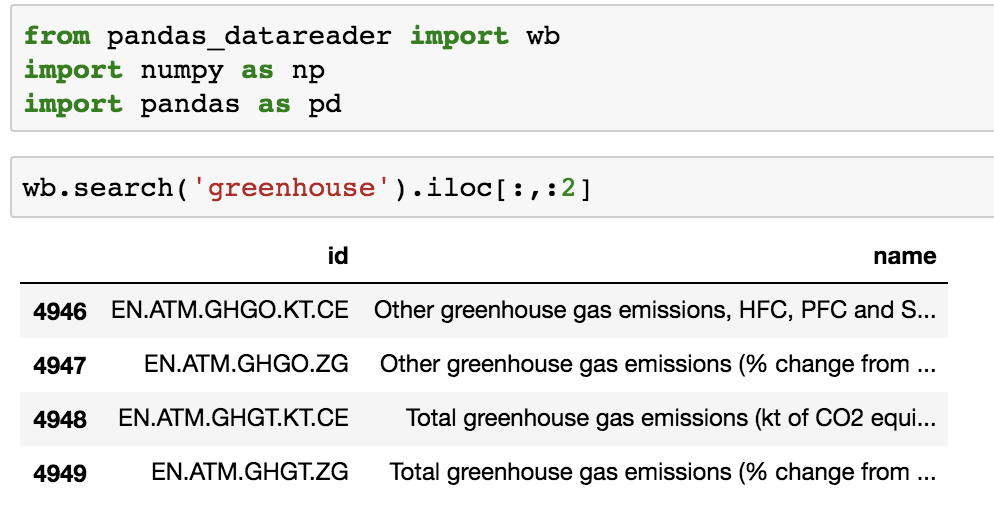
1. **Land-Ocean: Global Mean**

This dataset was another dataset found on the [NASA website](https://data.giss.nasa.gov/gistemp/) and we retrieved it in CSV format and loaded the file into the pandas library. For this one, we needed to find the annual ratios of Land to Ocean area and to do this we needed to find the mean of all the months, for each year.



In order to take the mean for each year we needed to set the year as an Index and then take the mean of each row, this was to ensure that the year was not being considered in the calculation of the mean.

1. **World Bank Web Scrape**

****The final dataset was retrieved by performing a web scrape off the World Bank website. With the scrape we were able to do a web search on any results containing “greenhouse” and found the level of greenhouse emissions by country. We needed to group this data so that we could see the total global greenhouse emission levels in order to merge with our other data sets that is also on a global level.

****

Conclusion

With the data sets that we were able to find and then clean, we are able to combine one large data sets that demonstrate the levels of Sea Level Rise, Sea Ice Index, Land-Ocean mean and the level of Greenhouse Emissions on a global level. This data can be used to draw conclusions on the severity the effects of Global Warming has on the planet. Topics that could be explored using this data could be the study of certain arctic species that rely on arctic temperatures. Alternatively the cities that would prosper and the cities that would perish economically due to the rising sea levels and global temperatures.