

# 1 Questions

- How do changes in atmospheric CO<sub>2</sub> concentrations correlate with changes in sea level rise over time?
- Is there a correlation between rising mean surface temperatures and land cover changes?

## 2 Data Sources

### 2.1 Description of Data Sources

- **Dataset 1: Annual Surface Temperature Change**

This data source provides the mean surface temperature change for the period 1961–2021 for each country. It uses temperatures from 1951 and 1980 as a baseline. [2]

- **Dataset 2: World Monthly Atmospheric CO<sub>2</sub> Concentrations**

In this data source, world-wide average concentrations of CO<sub>2</sub> are present, which have been observed on a monthly basis since 1958. [1]

- **Dataset 3: Change in Mean Sea Levels**

The indicator provided in this data source is global sea level rise for different seas and oceans, observed monthly since 1993. [3]

- **Dataset 4: Land Cover Altering Indicator**

This data source looks at the changes in land cover over time from 1992 to 2020 for each country. The indicator is grouped into categories on the basis of climate influence: altering, regulating, and neutral. [4]

### 2.2 Data Structure and Quality

- **Annual Surface Temperature Change** The data is structured in a time series for each country in a horizontal manner. Temperature change is in degree-Celsius units. The main columns to use are country, iso3 (i.e., country code mapping required for generating maps), and indicator values for all the years. At most, 8% of the data in one of the years is unobserved, which is fine as that can be imputed with zeros.

Table: Annual\_Surface\_Temperature\_Change

	Country	ISO3	1992	1993	1994	1995	1996	1997	1998	1999	...	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
0	Afghanistan, Islamic Rep. of	AFG	-0.294	0.220	0.430	0.359	-0.116	0.471	0.675	1.198	...	1.397	0.223	1.281	0.456	1.093	1.555	1.540	1.544	0.910	0.498
1	Albania	ALB	0.106	0.076	1.330	-0.172	-0.038	0.075	0.795	0.670	...	1.055	1.487	1.333	1.198	1.569	1.464	1.121	2.028	1.675	1.498
2	Algeria	DZA	-0.312	0.552	0.732	0.595	0.846	1.059	1.109	1.476	...	1.398	1.147	1.192	1.690	1.121	1.757	1.512	1.210	1.115	1.926
3	American Samoa	ASM	0.344	-0.069	0.189	0.755	0.784	NaN	NaN	0.242	...	0.854	0.924	1.257	1.170	1.009	1.539	1.435	1.189	1.539	1.430
4	Andorra, Principality of	AND	0.386	0.174	1.508	1.279	0.570	1.788	1.018	1.055	...	1.677	1.265	0.831	1.946	1.690	1.990	1.925	1.919	1.964	2.562

5 rows × 31 columns

Figure 1: First 5 rows of annual surface temperature change dataset.

- **World Monthly Atmospheric CO<sub>2</sub> Concentrations** This indicator of CO<sub>2</sub> concentration is structured in a global time series with a monthly frequency. The main columns are date and value. The indicator value is expressed in parts per million (ppm). Nulls are 0% in this dataset.

	Country	Unit	Date	Value
0	World	Parts Per Million	1958-03-01 00:00:00	315.70
1	World	Parts Per Million	1958-04-01 00:00:00	317.45
2	World	Parts Per Million	1958-05-01 00:00:00	317.51
3	World	Parts Per Million	1958-06-01 00:00:00	317.24
4	World	Parts Per Million	1958-07-01 00:00:00	315.86

Figure 2: First 5 rows of world monthly atmospheric CO<sub>2</sub> concentrations dataset.

- **Change in Mean Sea Levels** In this dataset main columns are measure, date and value in millimeters. It is again a time series observed with a monthly frequency but it is categorized into major seas and oceans in measure column. There are 0% nulls in this dataset.

	Country	Measure	Date	Value
0	World	Andaman Sea	1992-12-17 00:00:00	-10.34
1	World	Arabian Sea	1992-12-17 00:00:00	-18.46
2	World	Atlantic Ocean	1992-12-17 00:00:00	-15.41
3	World	Baltic Sea	1992-12-17 00:00:00	196.85
4	World	Bay Bengal	1992-12-17 00:00:00	3.27

Figure 3: First 5 rows of change in mean sea levels dataset.

- **Land Cover Altering Indicator** This dataset is the land cover estimation of each country into three main categories of climate influence: altering, regulating, and neutral. Annual values are in a time series for each country in horizontal manner. At most 0.3% of the data in one of the years is unobserved, which is fine as that can be imputed with zeros.

	Country	ISO3	Indicator	Unit	Climate_Influence	1992	1993	1994	1995	1996	...	2011	2012	2013
0	Advanced Economies	ACTMP	Climate Altering Land Cover Index	Index	Climate altering	95.942016	97.027517	97.030493	97.339731	97.462165	...	99.903233	99.963105	99.978710
1	Advanced Economies excluding US	EAEUSTMP	Climate Altering Land Cover Index	Index	Climate altering	94.851414	96.800658	96.767106	97.090049	97.172806	...	100.051718	100.104101	100.069837
2	Africa	AFRTMP	Climate Altering Land Cover Index	Index	Climate altering	91.590697	93.557008	93.500134	94.266184	94.532599	...	98.793984	100.076777	100.086539
3	Americas	AMETMP	Climate Altering Land Cover Index	Index	Climate altering	92.924554	93.023248	92.940142	94.263051	94.632936	...	99.025837	99.287572	99.647269
4	Asia	ASIATMP	Climate Altering Land Cover Index	Index	Climate altering	100.417597	100.022963	99.578377	99.613406	99.621526	...	100.133256	99.990573	99.915897

Figure 4: First 5 rows of land cover altering indicator dataset.

## 2.3 Licenses and Permissions

The data sources are publicly available on IMF under open-data licenses. Detailed license information can be found at: [License](#)

### 3 Data Pipeline

The data pipeline has three main modules: extractor, transform, and loader. Each of the modules has their respective functions. First `extract_csv` from extractor module is used to extract the data source from URL, then `delete_columns` from transform module deletes the list of useless columns specified for every dataset, then a flag of "date\_column" is present in configs which only applies `standardize_date_column` to standardize the date format across necessary datasets, then renaming of the date columns is done using `rename_year_columns` function, it is only triggered for those datasets where flag of "rename\_year\_columns" is equals to true, once all the transformations have been applied, dataset is then loaded to sqlite database using `load_df_to_sqlite` from loader module.



Figure 5: ETL Pipeline Diagram

### 4 Result and Limitations

Output datasets of the pipeline for all data sources are stored in sqlite database as tables as it was faster and easier to handle as a collective database, The pipeline is coded in a way that data quality dimensions were of the upmost priority and that the output datasets of the pipeline

- reflect the real word and are correct indicators
- contain all necessary information which is required to answer selected questions
- are consistent in their formats
- time period of datasets are appropriate and intersecting
- presentation of the datasets aligns with the requirements of the questions need to be answered

Annual surface temperature change and land cover altering indicator can be compared and checked for correlation and similarly the other two datasets can be compared. The only limitation is in the sea level dataset where the measure column is not friendly to be mapped on the world map for oceans and seas and no code mappings were found for them.

### References

- [1] NOAA/GML Dr. Pieter Tans and Scripps Institution of Oceanography Dr. Ralph Keeling. World monthly atmospheric carbon dioxide concentrations. [https://climatedata.imf.org/datasets/9c3764c0efcc4c71934ab3988f219e0e\\_0/about](https://climatedata.imf.org/datasets/9c3764c0efcc4c71934ab3988f219e0e_0/about), 2024.
- [2] FAO. Temperature change. [https://climatedata.imf.org/datasets/4063314923d74187be9596f10d034914\\_0/about](https://climatedata.imf.org/datasets/4063314923d74187be9596f10d034914_0/about), 2024.
- [3] NOAA Laboratory for Satellite Altimetry. Sea level rise. <https://climatedata.imf.org/datasets/b84a7e25159b4c65ba62d3f82c605855/about>, 2024.
- [4] FAO; IMF staff calculations. Land cover. <https://climatedata.imf.org/datasets/b1e6c0ea281f47b285addae0cbb28f4b/about>, 2024.