Data Report: **Climate Disaster Trends and Risk Correlation Analysis**

Questions

1. How has the frequency of climate-related disasters evolved globally over the years?
2. Which nations or regions experience particular kinds of climate-related disasters most frequently?
3. Do more severe climate disasters occur in nations with higher inform risk indicators?

Data Sources

Data source 1: Climate Disasters Frequency

Data URL:
https://opendata.arcgis.com/datasets/b13b69ee0dde43a99c811f592af4e821_0.csv

Metadata URL:
https://climatedata.imf.org/datasets/b13b69ee0dde43a99c811f592af4e821_0/about

Data Type: .csv

This data source contains the data of frequency of disasters occurred from year 1980 to 2022. It contains data of several disaster categories occurred, including storms, floods, wildfires, droughts. It is sourced from EM-DAT by the Université catholique de Louvain.

Data source # 2: Climate-driven INFORM Risk

Data URL:
https://opendata.arcgis.com/datasets/7cae02f84ed547fbbd6210d90da19879_0.csv

Metadata URL:
https://climatedata.imf.org/datasets/7cae02f84ed547fbbd6210d90da19879_0/about

Data Type: .csv

This data source contains the data of risk associated with climate driven hazards divided in three types: hazard & exposure, vulnerability, and lack of coping capacity. The available data is from 2013 to 2021. This dataset aids in assessing the vulnerability of different countries.

What is the data structure and quality of your sources?

Climate-related Disasters Frequency

```
print(climate_disasters.head(10))
```
✓ 0.0s Python

Id	Country	ISO3	Indicator	Unit	2013	2014	2015	2016	2017	2018	2019	2020	2021
1	Afghanistan, Islamic Rep. of	AFG	Drought	Number of	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	1.0
2	Afghanistan, Islamic Rep. of	AFG	Extreme temperature	Number of	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Afghanistan, Islamic Rep. of	AFG	Flood	Number of	4.0	2.0	1.0	4.0	1.0	3.0	6.0	5.0	2.0
4	Afghanistan, Islamic Rep. of	AFG	Landslide	Number of	1.0	NaN	4.0	NaN	2.0	1.0	1.0	1.0	1.0
5	Afghanistan, Islamic Rep. of	AFG	Storm	Number of	NaN	1.0	NaN	NaN	2.0	NaN	NaN	1.0	NaN
6	Afghanistan, Islamic Rep. of	AFG	TOTAL	Number of	5.0	3.0	5.0	4.0	5.0	5.0	7.0	7.0	4.0
7	Afghanistan, Islamic Rep. of	AFG	Wildfire	Number of	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	Albania	ALB	Drought	Number of	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	Albania	ALB	Extreme temperature	Number of	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN
10	Albania	ALB	Flood	Number of	NaN	NaN	3.0	1.0	1.0	1.0	NaN	NaN	1.0

This dataset is of high quality and reliability sourced from International Monetary Fund. It was originally from 1980 to 2022 but due to correlation analysis with another dataset it was reduced into 2013 to 2021 to ensure result relevance and reliability. The dataset also contains nulls in the yearly columns, indicating that in those years no disaster occurred. The data is in a tabular format and was originally provided in CSV form. It has been transformed and loaded into an SQLite database for ease of analysis and integration.

Climate-driven INFORM Risk

```
print(inform_risk)
```
✓ 0.0s Python

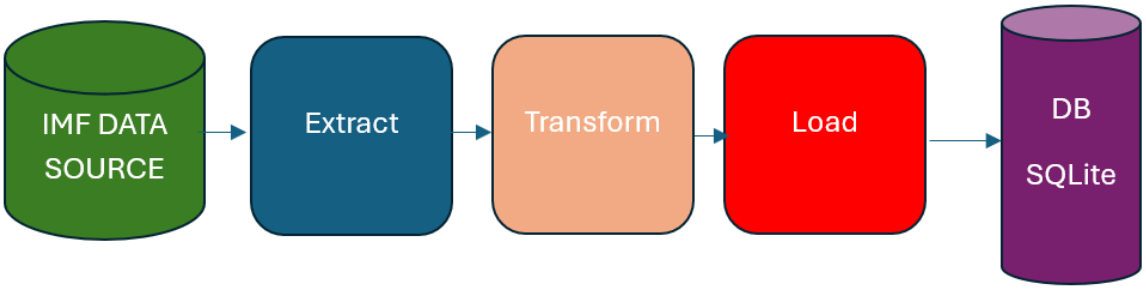
ObjectId	Country	ISO3	Indicator	Unit	2013	2014	2015	2016	2017	2018	2019	2020	2021
745	Zambia	ZMB	Climate-driven Hazard & Exposure	Index	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.8
746	Zambia	ZMB	Climate-driven INFORM Risk Indicator	Index	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.1	5.2
747	Zambia	ZMB	Lack of coping capacity	Index	6.1	6.2	6.2	6.1	5.9	5.9	5.9	5.9	5.9
748	Zambia	ZMB	Vulnerability	Index	5.7	5.6	5.6	5.6	5.7	5.7	5.7	5.9	6.2
749	Zimbabwe	ZWE	Climate-driven Hazard & Exposure	Index	6.6	6.6	6.6	6.6	6.6	6.6	6.6	6.6	6.6
750	Zimbabwe	ZWE	Climate-driven INFORM Risk Indicator	Index	6.3	5.8	6.1	5.8	5.9	5.8	6.2	6.2	6.2
751	Zimbabwe	ZWE	Lack of coping capacity	Index	7.2	5.9	5.9	5.9	5.8	5.8	5.9	5.8	5.8
752	Zimbabwe	ZWE	Vulnerability	Index	5.3	5.1	5.8	5.1	5.3	5.1	6.1	6.2	6.2

This dataset spans the years from 2013 to 2021 and originates from IMF Climate Data, guaranteeing its relevance and credibility. Initially provided in CSV format, the data has been transformed and loaded into an SQLite database to facilitate analysis and integration. The structure is tabular which makes data management efficient and enhances usability.

Describe the licenses of your data sources, why you are allowed to use the data and how you are planning to follow their obligations.

The data sources are open licensed, for more details you can refer to IMF's general terms and condition of use.

Data Pipeline



The data pipeline is developed with the help of three modules which are extract, transform, and load. In the first module, **extract.py**, data is fetched from the source. After that, in the second module **transform.py**, the data is typecast into pandas for necessary transformations, which includes handling missing values, standardizing country names, deleting irrelevant columns, and simplifying detailed indicator names into more understandable forms. Lastly, in the **load.py** module, the SQLite library is used to load data into an SQLite database for efficient querying and exploratory data analysis.

Technology: Based on the project and requirement, **python** is used as core programming language along with its libraries such as **pandas** for data manipulation and **sqlite3** for data storage. Moreover, for error handling, **try except** block and **pytest** is implemented in all modules, to ensure data is correctly fetched, transformed, and loaded into database.

Result and Limitations

Finally, after all the transformation, SQLite database **Climate.db** is created having two tables climate_disaster_freq and climate_inform_risk, the data pipeline processes efficiently and transforms data into a format that can be analysed. The dataset is correct and contain necessary information which will be used to answer questions. Despite data being reliable and robust, there is a limitation regarding available timeline, climate disasters frequency dataset contains data from 1980 to 2022, whereas the inform risk dataset contains data from 2013 to 2021. This restriction places a time limit on our correlation analysis. Therefore, the timeline of the climatic disasters' dataset will be shortened to match that of the inform risk dataset which is from 2013 to 2021 to guarantee a fair comparison and reliable correlation analysis.