Main Question

How has climate change affected the Indian Import and Export?

Data Sources

Data Source 1: Temperature Change in India

- ***Metadata URL:*** World Bank Climate Change Knowledge Portal Metadata
- ***Data URL:*** Kaggle Dataset
- ***Description:*** Monthly and annual temperature data of India and its states and UTs from 1901 to 2020.
- ***Reason for Choosing:*** Provides comprehensive historical climate data essential for analyzing long-term temperature trends and their impacts on the Indian subcontinent.

Data Source 2: Floods in India

- ***Metadata URL:*** Springer Article Metadata
- ***Data URL:*** Kaggle Dataset
- ***Description:*** Detailed records of floods in India, designed for comprehensive flood research.
- ***Reason for Choosing:*** Critical for understanding the impact of climate change on extreme weather events and their frequency in India.

Data Source 3: Import and Export Data of India

- ***Metadata URL:*** Kaggle Dataset Metadata
- ***Data URL:*** Kaggle Dataset
- ***Description:*** Trade data (in Million USD) from January 1997 to July 2022.
- ***Reason for Choosing:*** Enables analysis of the economic impact of climate change on India's trade, especially in port cities affected by climate events.

Data Quality and Structure

- ***Temperature Change Data*:** The dataset is well-structured with monthly and annual temperature readings across various states and UTs. The quality is high, with detailed metadata provided, ensuring accuracy and reliability.

```
data1.head()
```
✓ 0.0s — Python

	States	Period	1901	1902	1903	1904	1905	1906	1907	1908	...	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
0	Andaman and Nicobar	Jan	27.68	26.75	26.70	26.02	26.12	27.62	26.78	26.40	...	26.53	26.78	26.83	25.88	26.29	26.80	26.95	26.79	27.30	27.56
1	Andaman and Nicobar	Feb	28.50	26.67	27.54	26.49	27.29	27.40	26.92	26.29	...	26.87	27.00	27.73	26.46	26.75	26.86	26.83	26.61	27.48	27.14
2	Andaman and Nicobar	Mar	27.03	28.09	27.66	27.55	27.89	26.88	27.04	27.59	...	26.39	27.53	27.84	27.54	27.86	28.29	27.78	27.63	27.88	27.90
3	Andaman and Nicobar	Apr	28.52	29.35	29.53	27.72	28.79	29.62	28.75	29.23	...	27.94	28.48	28.73	28.98	28.39	29.74	28.31	28.18	29.39	28.96
4	Andaman and Nicobar	May	28.28	28.25	29.10	28.02	27.80	29.67	27.42	27.70	...	27.81	27.94	28.40	28.74	28.78	28.60	28.39	27.74	29.06	29.28

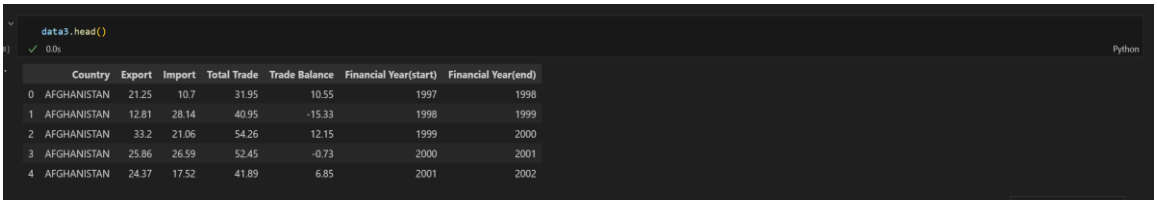
5 rows × 122 columns

- ***Flood Data:*** This dataset includes geospatial data and records of flood events, structured to support detailed flood analysis. The quality is excellent, with thorough documentation.

```
data2.head()
```
[7] ✓ 0.0s — Python

	Start Date	End Date	Duration(Days)	Main Cause	Location	Districts	State	Latitude	Longitude	Severity	Area Affected	Human fatality	Human injured	Human Displaced	Animal Fatality	Description of Casualties/injured	Extent of damage	Event Source	Event Source ID
0	2015-06-20	2015-06-21	1	Heavy rains	NaN	East Godavari, Srikakulam, Visakhapatnam and W...	ANDHRA PRADESH	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	i) Damage to 2000 hectares of crops reported. ...	IMD	NaN
1	2015-11-15	2015-11-23	8	Heavy rains	NaN	Anantapur, Chittoor, East Godavari, Krishna, N...	ANDHRA PRADESH	NaN	NaN	NaN	NaN	88.0	NaN	NaN	16710.0	88 persons died. 16710 animals perished. (844 ...	i) Extensive damage to Agricultural crops (mor...	IMD	NaN
2	2015-12-22	2015-12-22	0	Heavy rains	NaN	Vishakhapatnam	ANDHRA PRADESH	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	4 persons died due to landslips	Landslips	IMD	NaN
3	2015-10-06	2015-10-06	0	Heavy rains	NaN	Parts of Arunachal Pradesh	ARUNACHAL PRADESH	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	2 persons died.	NaN	IMD	NaN
4	2015-02-19	2015-02-19	0	Heavy rains	NaN	Parts of Assam	ASSAM	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	2 persons died due to landslide	NaN	IMD	NaN

- ***Trade Data:*** Structured as time-series data, this dataset includes trade values over several years. The quality is robust, with clear and consistent data entries.



```
data3.head()
```

✓ 0.0s Python

	Country	Export	Import	Total Trade	Trade Balance	Financial Year(start)	Financial Year(end)
0	AFGHANISTAN	21.25	10.7	31.95	10.55	1997	1998
1	AFGHANISTAN	12.81	28.14	40.95	-15.33	1998	1999
2	AFGHANISTAN	33.2	21.06	54.26	12.15	1999	2000
3	AFGHANISTAN	25.86	26.59	52.45	-0.73	2000	2001
4	AFGHANISTAN	24.37	17.52	41.89	6.85	2001	2002

Licenses and Permissions

All datasets are available under open data licenses. The Climate Change Knowledge Portal and the flood dataset from Springer provide comprehensive usage rights for research purposes. The trade data from Kaggle is also freely available for analysis. I will ensure to attribute the sources properly in all outputs and follow any specific usage guidelines mentioned in the metadata.

Data Pipeline

Technologies Used:

- ***Python:*** For data cleaning, transformation, and analysis.
- ***Pandas:*** For data manipulation.
- ***Jupyter Notebook:*** For data exploration and visualization.
- ***Matplotlib and Seaborn:*** For visualizations.

Pipeline Description:

1. ***Data Ingestion:*** Datasets downloaded from Kaggle and unzipped.
2. ***Data Cleaning:***
    - Temperature Data: Dropped incomplete rows, standardized date formats.
    - Flood Data: Renamed columns, removed unnecessary columns.
    - Trade Data: Dropped incomplete rows, standardized country names and dates.
3. ***Data Transformation:***
    - Normalized data for comparisons.
    - Aggregated data to annual summaries.
    - Merged datasets by year and region.
4. ***Error Handling:***
    - Implemented consistency checks.
    - Logged and handled processing errors.
5. ***Output Generation:*** Cleaned data exported to CSV files.

Steps Taken:

1. ***Dataset Download and Unzipping:***
    - Downloaded datasets using Kaggle API, unzipped, and stored in the data directory.
    - Renamed files to remove double extensions.
2. ***Reading and Cleaning Data:***
    - Used Pandas to read CSV files, handled encoding issues.
    - Specific cleaning operations for each dataset:
        - Temperature data: Dropped rows with missing values.
        - Flood data: Renamed columns, removed irrelevant columns.
        - Trade data: Dropped rows with missing values, standardized formats.

3. ***Renaming Files:***
    - Renamed cleaned files for clarity (e.g., 'mean_temperature_data' to 'INDTemperature.csv').
4. ***Post-Cleaning Adjustments:***
    - Replaced "till now" with "2023" in the trade data to standardize temporal references.

Problems Encountered and Solutions

- ***Missing Data:*** Dropped incomplete rows or used interpolation.
- ***Inconsistent Formats:*** Standardized date and text formats.
- ***Data Merging Issues:*** Used outer joins to ensure no loss of critical information.

Results and Limitations

Output Data:

- ***Structure and Quality:*** Clean, well-structured data ready for analysis.
- ***Format:*** Exported as CSV files for further analysis.

Limitations:

- ***Temporal Resolution:*** Different datasets have varying resolutions.
- ***Data Completeness:*** Some regions and periods may have less reliable data.
- ***Future Changes:*** The pipeline may need adjustments for new data types.

Critical Reflection(s)

The data pipeline effectively handles current datasets, but future work might involve integrating more data sources and refining the pipeline to handle evolving research questions. Issues such as data bias and regional disparities will be considered in the final analysis.