

Report

NAKKA KIRAN KUMAR

ME20B117

Steps followed in Airflow:

Pipeline-1:

- Setting the variables and lists required like url, years list, number of stations data to fetch, required fields list (weather fields).
- Created a for loop to iterate over years.
- Then for each year url contents are fetched and extracted the csv files links to download them and zip them.

Pipeline-2:

- Setting the variables and lists required like url, years list, required fields list (weather fields).
- Created a for loop to iterate over years.
- Unzip the file after checking its validity and processed the csv files to fetch the required data and plotting plots for every month.
- Then using the monthly plots, created a gif file.
- Then deleted all the files except the gif file.

Additional things:

- To add additional library dependancies, like matplotlib library and etc., other libraries, I extended the docker image.
- I had run the pipelines over fewer stations and for one year only. My system is getting too hot and hanging for processing more years and stations.
- If code works for few stations, it should also work for many too!

I have attached a GIF file with period showing MM-YYYY.