CS5380

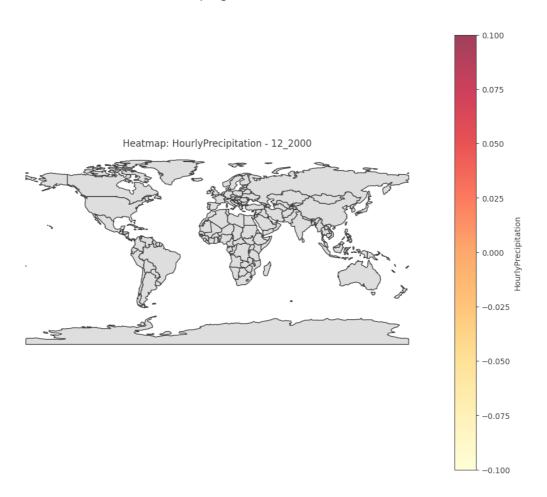# BIG DATA LABORATORY

ASSIGNMENT - 2

SARATH S

BS20B033

TASK-1

Refer to the attached code(pipeline_1.py) for the full pipeline. The climate data was successfully scraped using BeautifulSoup library(scrape function). The data was archived into a zip file (data.zip) using an archive task and placed in climate data directory.

TASK-2

The extracted CSV files were processed to filter the dataframe based on required fields and extract Lat/Long values to create tuples of hourly data. Monthly averages of the required fields are computed and finally, using 'geopandas' and 'geodatasets' packages, heatmaps for the fields - [['HourlyDewPointTemperature','HourlyPrecipitation'] at different lat/lon positions were generated.
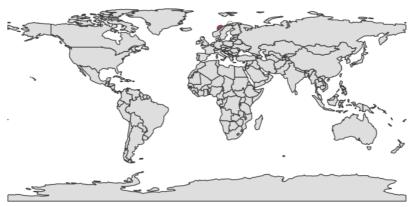Refer attachements for all the png files.

Heatmap: HourlyPrecipitation - 12_2000

Heatmap: HourlyPrecipitation - 12_2018

Heatmap: HourlyDewPointTemperature - 12_2000

Heatmap: HourlyDewPointTemperature - 12_2018