

Weather Data Analysis:-

- The Source code is just the python file. In order to execute the program you will need to run it on Pyspark notebook or on NYU Hpc Dumbo cluster.
- The data set for this project is too big to attach with the file so the data can be found at : - <https://www.kaggle.com/noaa/gsod>
- We have only taken data of the last decade (2008-2018). So if you want to get that data, you will have to run a query on Kaggle kernel and download the dataset.
- The query and the program for extraction is there in the report.

Thank You