

# Weather and Climate Analysis

AWS has provided public Data for ML for NOAA Historical Weather Conditions.

AWS Open Data Repository: <https://registry.opendata.aws/>

AWS Blog for this Demo:

<https://aws.amazon.com/blogs/big-data/visualize-over-200-years-of-global-climate-data-using-amazon-athena-and-amazon-quicksight/>

NOAA Source Data: <https://docs.opendata.aws/noaa-ghcn-pds/readme.html>

NOAA Global Forecast Dataset:

<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>

## Import Weather Data from NOAA

As part of the import of the public data, we had to create our own version of it stored in the local dlzp account.

I created a Glue database called: ghcnblog

The initial import of the data was into a table called: tblallyears

S3 Bucket Name: dlzp-weather-data

The data was copied into the DLZP Database using the following script.

```
create table ghcnblog.tblallyears_qa
with (
  format='PARQUET',
  external_location='s3://dlzp-weather-data/ghcnblog/allyearsqa/'
) as
select * from ghcnblog.tblallyears
where q_flag = '';
```

In addition, we had to download and import the weather station data

```
create table ghcnblog.tblghcnd_stations_qa
with (
  format='PARQUET',
  external_location='s3://dlzp-weather-data/ghcnblog/stations/'
) as
select * from ghcnblog.tblghcnd_stations;
```

# Exploratory data analysis and visualization

With our two tables created, we are now ready to query and visualize to gain insight.

Exploratory data analysis In the Athena query window, run the following queries to get an idea of the size of the dataset.

Query #1: the total number of observations since 1763:

```
select count(*) as Total_Number_of_Observations
from ghcncblog.tblallyears_qa;
```

Query #2: the number of stations since 1763:

```
select count(*) as Total_Number_of_Stations
from ghcncblog.tblghcnd_stations_qa;
```

## Average weather parameters for the Earth

The following shows a query that calculates the average maximum temperature (Celsius), average minimum temperature (Celsius), and average rainfall (mm) for the Earth since 1763.

In the query, we must convert the data\_value from a String variable to a Real variable. We also must divide by 10, because the temperature and precipitation measurements are in tenths of their respective units. For more information about these details and the element codes (TMIB, TMAX and PRCP), see the readme file.

```
select element, round(avg(CAST(data_value AS real)/10),2) as value
from ghcncblog.tblallyears_qa
where element in ('TMIN','TMAX','PRCP')
group by element;
```

element	value
TMAX	16.35
PRCP	2.34
TMIN	4.42

From: <https://wiki.cloud.dlzpgroup.com/> -

Permanent link: <https://wiki.cloud.dlzpgroup.com/doku.php?id=ml:noa>

Last update: **2019/02/17 13:37**

