# Weather Prediction Model

## Objective:

Create a toy model of the environment that evolves over time across various locations and predict weather

## Design Approach:

1, Obtain historic weather data

2, Clean the Data

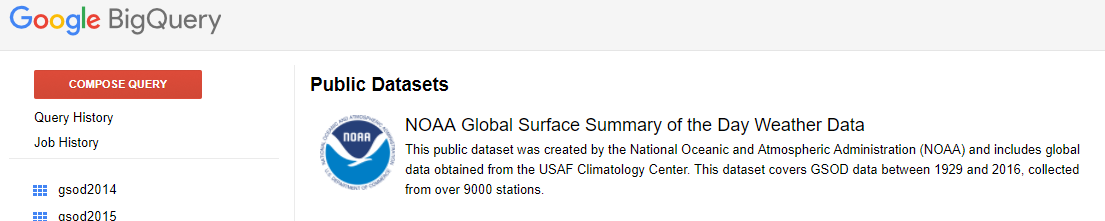
3, Create Model for Temperature, Humidity, Pressure and Condition with respect to location details

4, Predict the weather for feature dates based on the model generated

## Implementation Details:

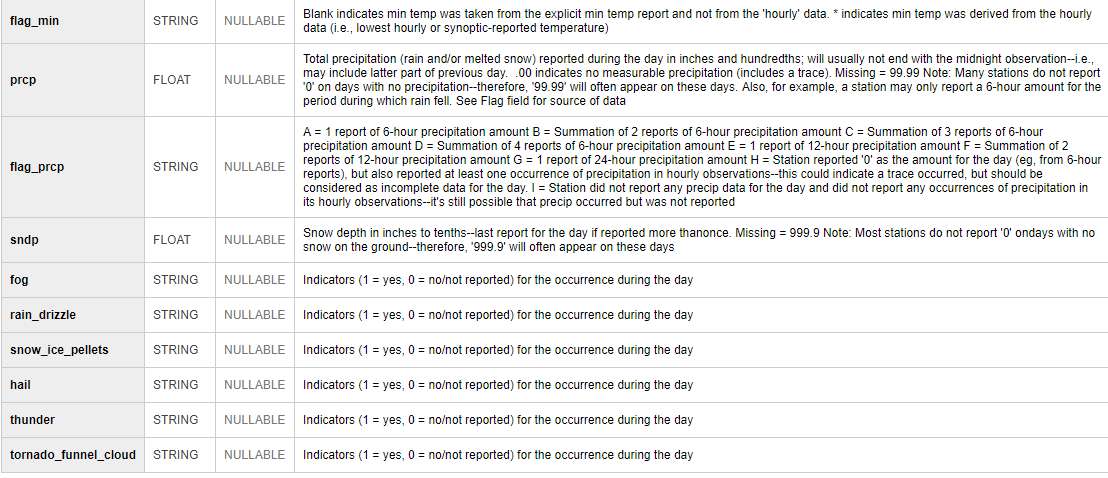
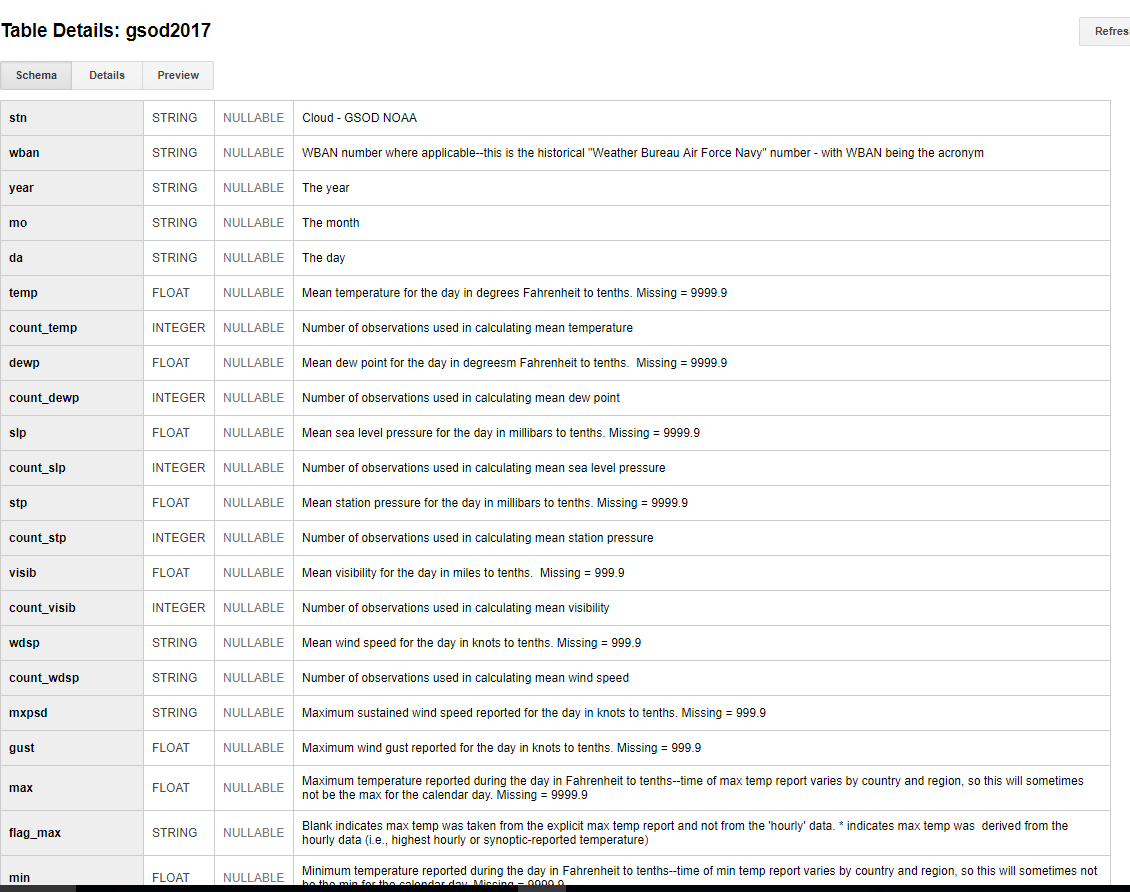
### Obtaining Historic Weather Data

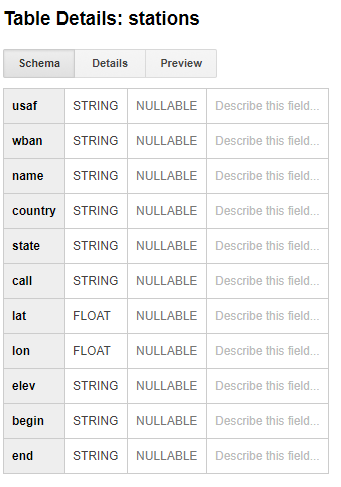
Using Google Big Query to get the historic weather data NOOA



Query the historic datas from gsod and stations tables from Jan 2016 till date.

gsod table will have the weather datas and station tables will have the station location details, schemas are as below

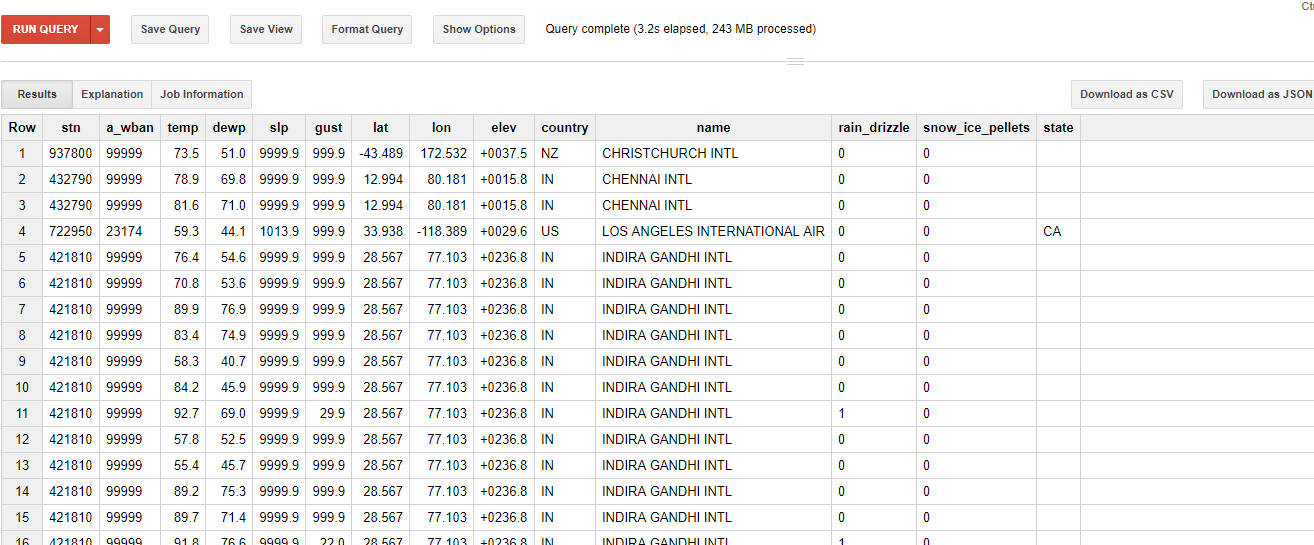




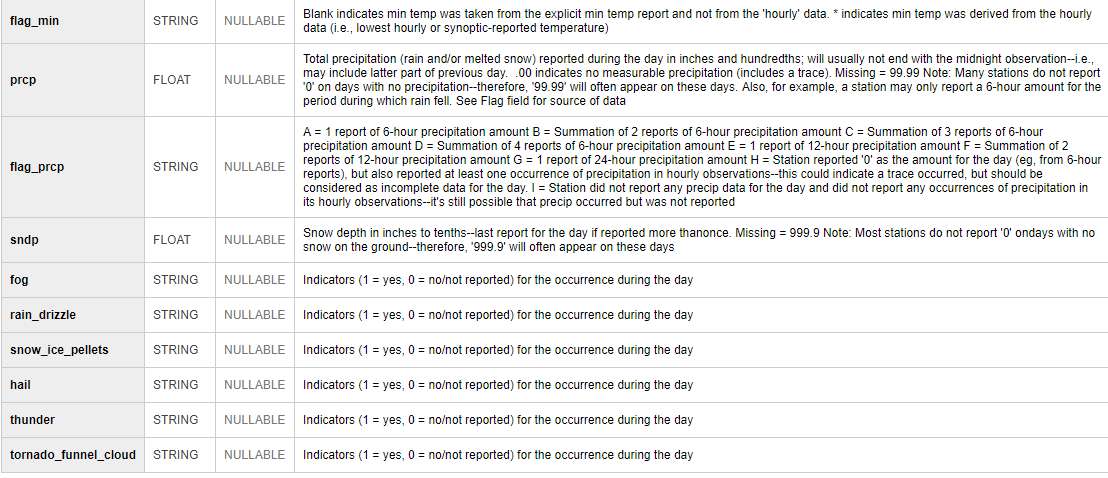
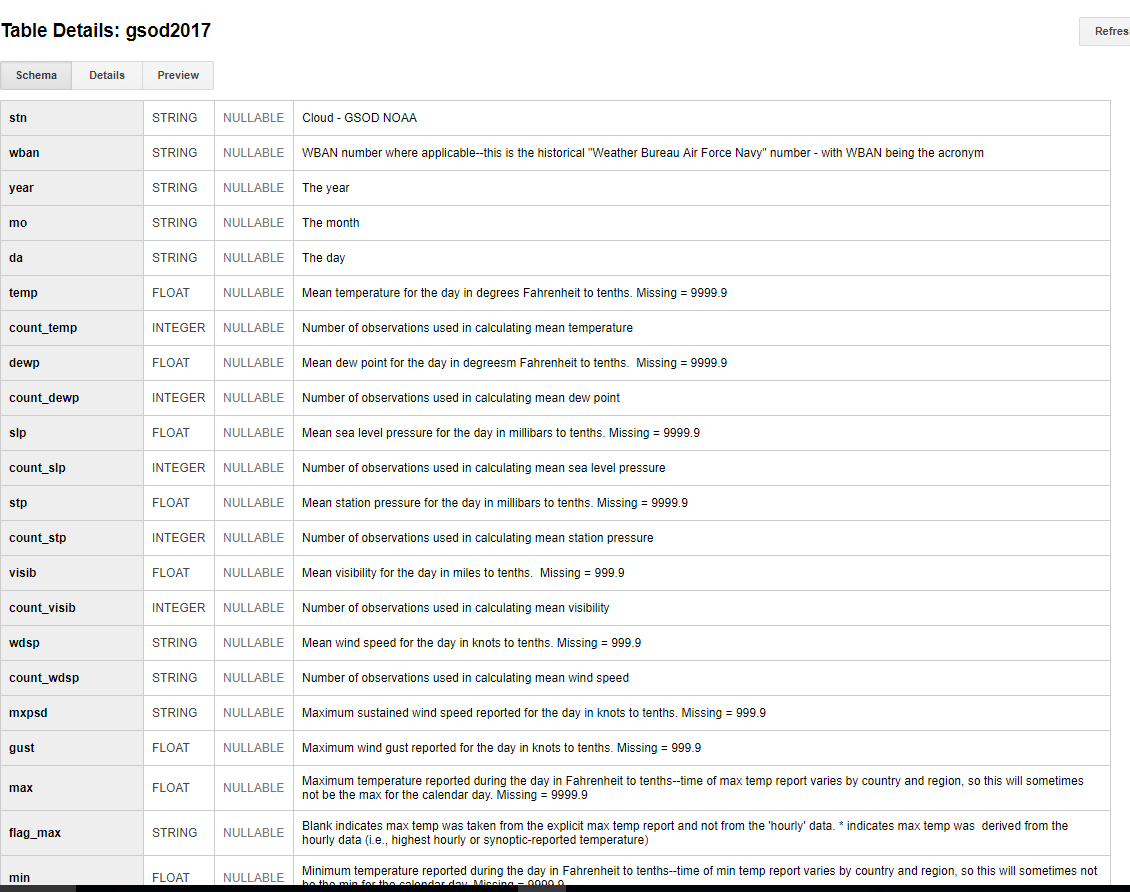
Sample query is as follows



Obtain the records for the various locations, the output will be as below



Save the result as CSV file for Data Analysis



### Clean the Data

If the datas are not available then dummy value of 9999.9 will be updated for the record.

Clean the data by removing those records in the obtained results



Add column for every record as condition to identify whether day was Rain, Snow or Sunny

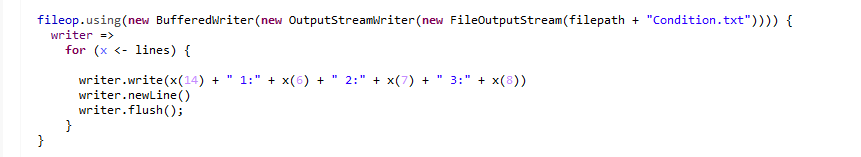
1, If results has snow\_ice\_pellets 🡪 Mark day as Snow

2, If result has rain\_drizzle 🡪 Mark day as Rain

3, Else mark as Sunny

Convert the temperatures in Fahrenheits to Celsius

Obtain only necessary parameters for analysis in the record as below



### Analyse the Data

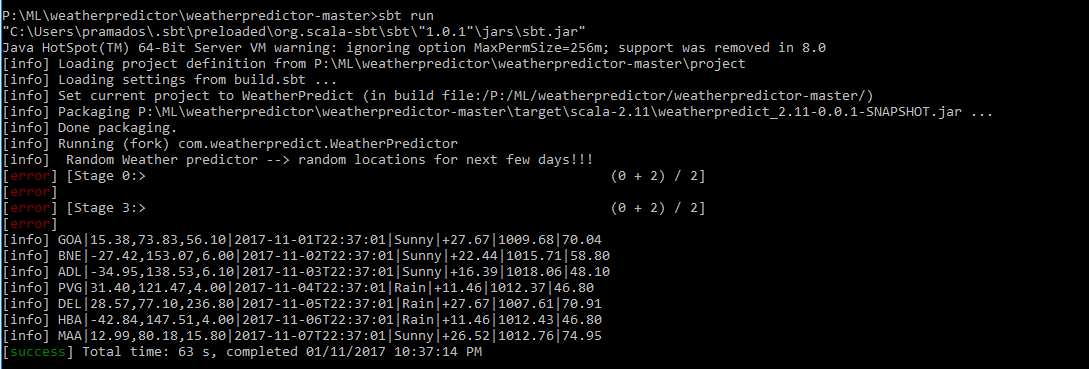
Random Forest Model is selected for analysis of the data since we have Regression & Classification type of analysis.

For Condition use classification model else use Regression Model as below

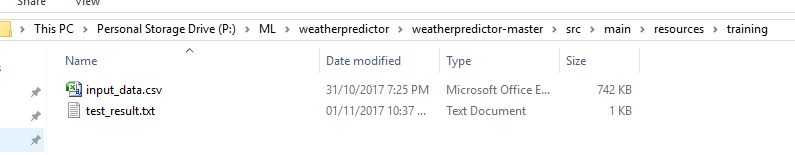


## Results:

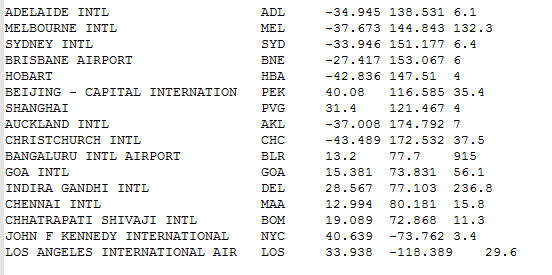
The output of the result will be as below



The results will saved in test file as well



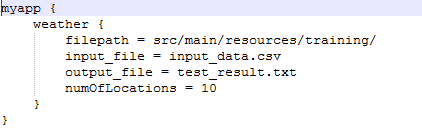
Station IATA mapping is as below



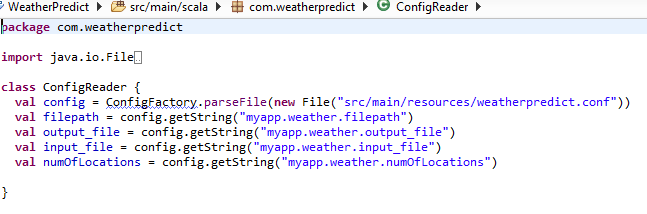
## How to Run:

Download the code from git and execute using **sbt run** !!!

If any config want to change, change the cofing file



Location of config file will be in ConfigReader.scala as below



Ensure Hadoop\_Home property is set correctly to avoid any errors

## Improvements:

The model application can be improved further by implementing

1, Rather removing the unavailable datas, we can predict the data and use the same

2, Daily summary data is used, can use hourly summary data with larger historic data to predict the weather on hourly basis

3, The model prediction can consider seasons as well to predict the data.