CANalitica

Data Format Document
For
Parquet Weather Data

January 14, 2025

CANalitica Professional Services Inc.

CANalitica

Introduction

The following is a description of summarized weather data that is based on the Global Summary of the Day weather data provided by National Centers for Environmental Information (NCEI) in Asheville, NC.(https://www.ncei.noaa.gov/). Additional documentation and source data and can be accessed via their links:

Web Accessible Folder -- https://www.ncei.noaa.gov/data/global-summary-of-the-day/

Common Access -- https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-day

The data described here are intended for free and unrestricted use in research, education, and other non-commercial activities.

The data is stored in parquet format which is well-suited for big data processing frameworks, making it a popular choice for scalable data storage and analysis in cloud environments. Data covers the years from 1950 to present day and is generally updated once per month.

Our data cleansing process includes excluding data with no position data (latitude or longitude), normalizing figures and using appropriate data types and was then organized by continent and year. Therefore, filenames indicate the continent which the weather station belongs to as well as the year on when the observation date was made. For example, data for a weather station in Canada for the year 1998 would be in the file: `North-America_1998.parquet`.

*** Note that weather stations that are near continental borders were intentionally added to each of the respective continents (for example, Europe and Asia). Therefore**, weather data from stations near continental borders will appear in multiple files and care should be taken when querying the data.**

CANalitica

The below field descriptions are in alphabetical order. The original text that the below is based on can be found here:

https://www.ncei.noaa.gov/data/global-summary-of-the-day/doc/readme.pdf

Conversion Formulas are as follows:

Fahrenheit to Celsius:
$C=(F-32)\times 5/9$

Inches to mm:
$mm=inches\times 25.4mm$

Feet to Meters:
$meters=feet\times 0.3048$

Psi to millibars:
$millibars=psi\times 68.9476$

Knots to meters per second (m/s):
$m/s=knots\times 0.514444$

Field	Description	Notes
DEWP	Mean dew point for the day in degrees Fahrenheit to tenths. Missing = 9999.9	
DEWP_ATTRIBUTES	Number of observations used in calculating mean dew point.	
ELEVATION	Given in meters.	
FRSHTT	Indicators (1 = yes, 0 = no/not reported) for the occurrence during the day of:	

	Fog ('F' - 1st digit). Rain or Drizzle ('R' - 2nd digit). Snow or Ice Pellets ('S' - 3rd digit). Hail ('H' - 4th digit). Thunder ('T' - 5th digit). Tornado or Funnel Cloud ('T' - 6th digit).	
GUST	Maximum wind gust reported for the day in knots to tenths. Missing = 999.9	
LATITUDE	Given in decimated degrees (Southern Hemisphere values are negative)	
LONGITUDE	Given in decimated degrees (Western Hemisphere values are negative)	
MAX	Maximum temperature reported during the day in Fahrenheit to tenths. Missing = 9999.9	Time of minimum temperature report varies by country and region, so this will sometimes not be the maximum for the calendar day.
MAX_ATTRIBUTES	Blank indicates maximum temperature was taken from the explicit maximum temperature report and not from the 'hourly' data.	* indicates maximum temperature was derived from the hourly data (i.e. highest hourly or synoptic-reported temperature).
MIN	Minimum temperature reported during the day in Fahrenheit to tenths. Missing = 9999.9	Time of minimum temperature report varies by country and region, so this will sometimes not be the maximum for the calendar day.

MIN_ATTRIBUTES	Blank indicates minimum temperature was taken from the explicit minimum temperature report and not from the 'hourly' data.	* indicates minimum temperature was derived from the hourly data (i.e. highest hourly or synoptic-reported temperature).
MXSPD	Maximum sustained wind speed reported for the day in knots to tenths. Missing = 999.	
NAME	Name of station/airport/military base	
OBSERVATION_DATE	Given in mm/dd/yyyy format	
PRCP	Total precipitation (rain and/or melted snow) reported during the day in inches and hundredths; will usually not end with the midnight observation (i.e. may include latter part of previous day). “0” indicates no measurable precipitation (includes a trace). Missing = 99.99	Many stations do not report “0” on days with no precipitation, therefore “99.99” will often appear on these days. Also, for example, a station may only report a 6-hour amount for the period during which rain fell. See attribute field for source of data.
PRCP_ATTRIBUTES	A = 1 report of 6-hour precipitation amount. B = Summation of 2 reports of 6-hour precipitation amount. C = Summation of 3 reports of 6-hour precipitation amount. D = Summation of 4 reports of 6-hour precipitation amount. E = 1 report of 12-hour precipitation amount. F = Summation of 2 reports of 12-hour precipitation amount. G = 1 report of 24-hour precipitation amount. H = Station reported '0' as the amount for the day (eg, from 6-hour reports), but also reported at least one occurrence of precipitation in hourly observations. This could indicate a trace occurred, but should be considered as incomplete data for the day.	

	I = Station did not report any precipitation data for the day and did not report any occurrences of precipitation in its hourly observations. It's still possible that precipitation occurred but was not reported.	
SLP	Mean sea level pressure for the day in millibars to tenths. Missing = 9999.9	
SLP_ATTRIBUTES	Number of observations used in calculating mean sea level pressure.	
SNDP	Snow depth in inches to tenths. It is the last report for the day if reported more than once. Missing = 999.9	Most stations do not report “0” on days with no snow on the ground, therefore, “999.9” will often appear on these days.
STATION	Station number.	
STP	Mean station pressure for the day in millibars to tenths. Missing = 9999.9	
STP_ATTRIBUTES	Number of observations used in calculating mean station pressure.	
TEMP	Mean temperature for the day in degrees Fahrenheit to tenths. Missing = 9999.9	
TEMP_ATTRIBUTES	Number of observations used in calculating mean temperature.	
VISIB	Mean visibility for the day in miles to tenths. Missing = 999.9	
VISIB_ATTRIBUTES	Number of observations used in calculating mean visibility.	
WDSP	Mean wind speed for the day in knots to tenths. Missing = 999.9	
WDSP_ATTRIBUTES	Number of observations used in calculating mean wind speed.	