

Stat480_Homework 6

Bin Feng

The Pig scripts for exercise 1 to 4 are all in the one *.pig* script submitted with this report. Please refer to that file when checking code details.

Exercise 1

The joined relation results (first 5 entries) are shown below:

usaf	wban	Year	Temperature (x10 Celsius)	Location
11060	99999	1929	0	ROEST III
11060	99999	1929	11	ROEST III
11060	99999	1929	-50	ROEST III
11060	99999	1929	-39	ROEST III
11060	99999	1929	-28	ROEST III

Note that the temperature results in the table are in tenths of degrees Celsius.

Exercise 2

The grouped by station results (first 10 entries) are shown below:

Location	Count	Average (x10 Celsius)	Max (x10 Celsius)	Min (x10 Celsius)
LVIV	332	60.653614457831324	289	-400
RIGA	166	-63.433734939759034	50	-278
ROME	61	116.55737704918033	189	22
WICK	368	64.70108695652173	156	-17
BREST	268	129.3320895522388	239	39
LARNE	275	77.47636363636363	167	0
PINSK	159	68.00628930817611	278	-100
TIREE	276	80.3731884057971	139	22
AACHEN	225	10.848888888888888	122	-200
BREMEN	303	106.13531353135313	278	-72

Note that the average, minimum, and maximum temperature results in the table are in tenths of degrees Celsius.

Exercise 3

The station name with the lowest minimum temperature is: *SODANKYLA*. The grouped by year results for this station are shown below:

Year	Min (x10 Celsius)	Average (x10 Celsius)	Max (x10 Celsius)
1920	-344	14.904371584699453	272
1921	-417	-5.7967032967032965	233
1922	-400	-9.691666666666666	267
1923	-394	-5.237877401646844	233
1924	-456	0.8650865998176845	289
1925	-378	-2.5712971481140756	311
1926	-411	-12.684692942254813	261

Note that the average, minimum, and maximum temperature results in the table are in tenths of degrees Celsius. Also find that temperature results for year 1927, 1928, and 1929 are not available for this station.

Exercise 4

Programmatically (using ORDER...DESC..., and LIMIT...1) find the station name with the largest maximum deviation above the average:

Name: ABERDEEN/DYCE AIRPO;

Maximum deviation above average: 423.15497382198953 (tenths of degrees Celsius)

The grouped by year results for this station are shown below:

Year	Maximum deviation below average
1926	97.5126050420168
1927	382.4566473988439
1928	241.248730964467
1929	127.47424892703863

Note that the temperature results in the table are in tenths of degrees Celsius. Also find that temperature results for year 1920 to 1925 are not available for this station.