

Temperature Dataset length

Data is consistent during Write parquet files. OK!
Start process the World Temperature data ...

to see how many rows in dfS.count: 8599212

Schema of staging Temp dfS-----

```
root
|-- dt: timestamp (nullable = true)
|-- AverageTemperature: double (nullable = true)
|-- AverageTemperatureUncertainty: double (nullable = true)
|-- City: string (nullable = true)
|-- Country: string (nullable = true)
|-- Latitude: string (nullable = true)
|-- Longitude: string (nullable = true)
```

```
+-----+-----+-----+-----+-----+-----+
|          dt|AverageTemperature|AverageTemperatureUncertainty| City|Country|Latitude|Longitude|
+-----+-----+-----+-----+-----+-----+
|1743-11-01 00:00:00|        6.068|        1.7369999999999999|Århus|Denmark| 57.05N| 10.33E|
|1743-12-01 00:00:00|         null|                        null|Århus|Denmark| 57.05N| 10.33E|
+-----+-----+-----+-----+-----+-----+
```

only showing top 2 rows

Immigration Datasets length (length for only first data file)

Start process the Immigration Data sas7dat ...

Accumulating Read into sas_data Schema for 1 sas data files

3096313

Schema of staging Immigra_data -----

```
root
|-- cicaid: double (nullable = true)
|-- i94yr: double (nullable = true)
|-- i94mon: double (nullable = true)
|-- i94cit: double (nullable = true)
|-- i94res: double (nullable = true)
|-- i94port: string (nullable = true)
|-- arrdate: double (nullable = true)
|-- i94mode: double (nullable = true)
|-- i94addr: string (nullable = true)
|-- depdate: double (nullable = true)
|-- i94bir: double (nullable = true)
|-- i94visa: double (nullable = true)
|-- count: double (nullable = true)
|-- dtadfile: string (nullable = true)
|-- visapost: string (nullable = true)
|-- occup: string (nullable = true)
|-- entdepa: string (nullable = true)
|-- entdepd: string (nullable = true)
|-- entdepu: string (nullable = true)
|-- matflag: string (nullable = true)
|-- biryear: double (nullable = true)
|-- dtaddto: string (nullable = true)
|-- gender: string (nullable = true)
|-- insnum: string (nullable = true)
|-- airline: string (nullable = true)
|-- admnum: double (nullable = true)
|-- fltno: string (nullable = true)
|-- visatype: string (nullable = true)
```

```
+-----+-----+-----+-----+-----+-----+
+-----+-----+
|cicaid| i94yr|i94mon|i94cit|i94res|i94port|arrdate|i94mode|i94addr|depdate
um|fltno|visatype|
+-----+-----+-----+-----+-----+-----+
```

Sql TEST run for count Immration case, group by Foregin Born in uscites

+-----+	
count_F_Imm	
+-----+	
	485762
	343193
	310001
	152462
	149164
	130464
	101433
	95898
	92528
	89148
	71781
	57275
	47687
	38847
	37669
	25607
	24957
	18230
	17502
	16220
	12666
	9804
	9063
	8928
	7066
	5114
	5100
	4490
	4271
	3501
	3469
	2970
	2175
	1719
	1605
	1000
	943
	777
	621
	586
	562
	542
	512
	419
	392
	354
	337
	329
	244
	238

+-----+

only showing top 50 rows