SummaryStatistics

November 30, 2017

```python
In [72]: from pyspark.sql.types import StructType, StructField, FloatType, LongType, StringType
         from pyspark.shell import spark

         feats = []
         f = open('features.txt')
         for line_num, line in enumerate(f):
             if line_num == 0:
                 # Timestamp
                 feats.append(StructField(line.strip(), LongType(), True))
             elif line_num == 1:
                 # Geohash
                 feats.append(StructField(line.strip(), StringType(), True))
             else:
                 # Other features
                 feats.append(StructField(line.strip(), FloatType(), True))

         schema = StructType(feats)

In [73]: df = spark.read.format('csv').option('sep', '\t').schema(schema).load('inputs/nam_min

In [74]: len(df.columns)

Out[74]: 58

In [75]: import pyspark.sql.functions as sf
         summary_values = []
         feature_names = []
         for i in range(2,len(df.columns)):
             feature_names.append(df.columns[i])
             summary_values.append(df.select(sf.max(df.columns[i]).alias("Max"),
                     sf.min(df.columns[i]).alias("Min"),
                     sf.avg(df.columns[i]).alias("Avg"),
                     sf.stddev(df.columns[i]).alias("Std_Dev")))

In [76]: for i in range(0,len(summary_values)):
             print("Feature: " + feature_names[i] + "\n")
             df_summ = summary_values[i]
             df_summ.select(df_summ.Max,df_summ.Min,df_summ.Avg,df_summ.Std_Dev).show()
```

Feature: geopotential_height_lltw

Max	Min	Avg	Std_Dev
4902.578	-5817.172	1571.8196034524215	1850.1112628212659

Feature: water_equiv_of_accum_snow_depth_surface

Max	Min	Avg	Std_Dev
5501.0	0.0	20.543379364668635	68.09250334632227

Feature: drag_coefficient_surface

Max	Min	Avg	Std_Dev
11800.0	0.0	0.08109455386769765	18.554648814462873

Feature: sensible_heat_net_flux_surface

Max	Min	Avg	Std_Dev
918.42236	-354.10474	19.861206073268022	62.41369246411668

Feature: categorical_ice_pellets_yes1_no0_surface

Max	Min	Avg	Std_Dev
1.0	0.0	3.529407287893087E-4	0.01878341016092709

Feature: visibility_surface

Max	Min	Avg	Std_Dev
24226.518	21.58885	21453.45861394169	6530.964319791838

Feature: number_of_soil_layers_in_root_zone_surface

Max	Min	Avg	Std_Dev
4.0	0.0	1.5993218971658194	1.7434861124729084

Feature: categorical_freezing_rain_yes1_no0_surface

Max	Min	Avg	Std_Dev
1.0	0.0	0.001784681421047...	0.04220779582082851

Feature: pressure_reduced_to_msl_msl

Max	Min	Avg	Std_Dev
103267.0	97514.0	101446.47825847058	852.5249028233586

Feature: upward_short_wave_rad_flux_surface

Max	Min	Avg	Std_Dev
544.875	0.0	31.775694632066426	63.021314040222634

Feature: relative_humidity_zerodegc_isotherm

Max	Min	Avg	Std_Dev
100.0	1.0	54.886496735060426	33.020286215580654

Feature: categorical_snow_yes1_no0_surface

Max	Min	Avg	Std_Dev
1.0	0.0	0.037243333130384484	0.18935760185067188

Feature: u-component_of_wind_tropopause

Max	Min	Avg	Std_Dev
84.40288	-26.128845	23.976830985869892	18.93582483968142

Feature: surface_wind_gust_surface

Max	Min	Avg	Std_Dev
33.857338	0.017012283	7.903872391106444	5.443834759774358

Feature: total_cloud_cover_entire_atmosphere

Max	Min	Avg	Std_Dev
100.0	0.0	59.6390491339158	45.6339232119349

Feature: upward_long_wave_rad_flux_surface

Max	Min	Avg	Std_Dev
602.3677	168.97278	364.58511873634154	64.33154516625235

Feature: land_cover_land1_sea0_surface

Max	Min	Avg	Std_Dev
1.0	0.0	0.468058388383208	0.49897892771902136

Feature: vegitation_type_as_in_sib_surface

Max	Min	Avg	Std_Dev
19.0	0.0	3.4780548875156017	5.110653406732627

Feature: v-component_of_wind_pblri

Max	Min	Avg	Std_Dev
37.2527	-24.354431	1.0066287453840368	6.8018571086992585

Feature: albedo_surface

Max	Min	Avg	Std_Dev
78.5	6.0	18.960229002404944	18.229718719127998

Feature: lightning_surface

Max	Min	Avg	Std_Dev
1.0	0.0	0.013511636579500138	0.11545165627527863

Feature: ice_cover_ice1_no_ice0_surface

Max	Min	Avg	Std_Dev
1.0	0.0	0.05187753051843283	0.22177984384922358

Feature: convective_inhibition_surface

Max	Min	Avg	Std_Dev
0.2319336	-463.76807	-7.679669304094945	32.00044118898448

Feature: pressure_surface

Max	Min	Avg	Std_Dev
102952.0	66432.0	97775.1470040945	6017.298326363637

Feature: transpiration_stress-onset_soil_moisture_surface

```
+-----+---+-----------------+------------------+
|  Max|Min|             Avg|          Std_Dev|
+-----+---+-----------------+------------------+
|0.405|0.0|0.15496598628838015|0.16673748577601488|
+-----+---+-----------------+------------------+
```

Feature: soil_porosity_surface

```
+---+---+----------------+-----------------+
|Max|Min|             Avg|          Std_Dev|
+---+---+----------------+-----------------+
|0.5|0.0|0.234029194191604|0.24948946385951068|
+---+---+----------------+-----------------+
```

Feature: vegetation_surface

```
+----+---+----------------+----------------+
| Max|Min|             Avg|         Std_Dev|
+----+---+----------------+----------------+
|89.0|0.0|8.029887468035557|14.376710704380757|
+----+---+----------------+----------------+
```

Feature: categorical_rain_yes1_no0_surface

```
+---+---+------------------+----------------+
|Max|Min|             Avg|         Std_Dev|
+---+---+------------------+----------------+
|1.0|0.0|0.09887382416511918|0.2984926729769393|
+---+---+------------------+----------------+
```

Feature: downward_long_wave_rad_flux_surface

```
+--------+---------+----------------+----------------+
|     Max|      Min|             Avg|         Std_Dev|
+--------+---------+----------------+----------------+
|477.64667|128.84128|317.2842871058952|66.07543316435003|
+--------+---------+----------------+----------------+
```

Feature: planetary_boundary_layer_height_surface

```
+-------+-----+----------------+----------------+
|    Max|  Min|             Avg|         Std_Dev|
+-------+-----+----------------+----------------+
|16691.5|49.75|1474.2216218054582|1204.0192087738408|
+-------+-----+----------------+----------------+
```

Feature: soil_type_as_in_zobler_surface

Max	Min	Avg	Std_Dev
16.0	0.0	2.438837559743067	3.1838219305490516

Feature: geopotential_height_cloud_base

Max	Min	Avg	Std_Dev
13283.5	-5000.0	35.28252762641176	3872.367803112988

Feature: friction_velocity_surface

Max	Min	Avg	Std_Dev
2.4021146	2.7896842E-4	0.364367133141722	0.27074228535090705

Feature: maximumcomposite_radar_reflectivity_entire_atmosphere

Max	Min	Avg	Std_Dev
51.3125	-20.0	-5.305052153889768	17.49404787327044

Feature: plant_canopy_surface_water_surface

Max	Min	Avg	Std_Dev
0.5	0.0	0.08494268427975549	0.17656010997633576

Feature: v-component_of_wind_maximum_wind

Max	Min	Avg	Std_Dev
73.77826	-70.763824	-1.4368067607294932	24.31172414090126

Feature: geopotential_height_zerodegc_isotherm

Max	Min	Avg	Std_Dev
5480.0	0.0	2660.4940979938506	1820.3260335654372

Feature: mean_sea_level_pressure_nam_model_reduction_msl

Max	Min	Avg	Std_Dev
102986.0	97514.0	101425.10075649183	838.6100678960836

Feature: temperature_surface

Max	Min	Avg	Std_Dev
320.9536	233.63518	283.4239539954792	13.741775470569852

Feature: snow_cover_surface

Max	Min	Avg	Std_Dev
100.0	0.0	26.49275187220311	44.11821455509774

Feature: geopotential_height_surface

Max	Min	Avg	Std_Dev
3576.32	-73.92998	321.01270061114224	549.5768832623813

Feature: convective_available_potential_energy_surface

Max	Min	Avg	Std_Dev
5640.0	0.0	159.70642180888308	413.3867443276708

Feature: latent_heat_net_flux_surface

Max	Min	Avg	Std_Dev
1400.7657	-192.73425	55.34242500623483	67.73013290284702

Feature: surface_roughness_surface

Max	Min	Avg	Std_Dev
2.750016	1.5900003E-5	0.48349228123311166	0.8125509365105341

Feature: pressure_maximum_wind

Max	Min	Avg	Std_Dev
49926.734	11126.578	21138.468413811468	6261.820944189558

Feature: temperature_tropopause

Max	Min	Avg	Std_Dev
236.33627	191.76529	210.6579736611292	8.532582478390049

Feature: geopotential_height_pblri

Max	Min	Avg	Std_Dev
4103.4277	15.164593	520.3162076844575	470.2662848360115

Feature: pressure_tropopause

Max	Min	Avg	Std_Dev
47669.184	8124.4873	19267.652348282718	7090.561550204382

Feature: snow_depth_surface

Max	Min	Avg	Std_Dev
14.492	0.0	0.08273362823361356	0.24630959619901469

Feature: v-component_of_wind_tropopause

Max	Min	Avg	Std_Dev
60.24814	-52.25186	-0.7253900709288382	19.231001260173503

Feature: downward_short_wave_rad_flux_surface

Max	Min	Avg	Std_Dev
1009.05884	0.0	180.21073217396813	273.24860529533936

Feature: u-component_of_wind_maximum_wind

Max	Min	Avg	Std_Dev
92.16766	-49.616287	28.469879072700817	21.527677396011985

Feature: wilting_point_surface

Max	Min	Avg	Std_Dev
0.13499999	0.0	0.031106272933598897	0.03738218278900426

Feature: precipitable_water_entire_atmosphere

Max	Min	Avg	Std_Dev
59.385254	0.7078148	16.54948786552142	10.593789793864588

Feature: u-component_of_wind_pblri

```
+---------+----------+----------------+----------------+
|      Max|       Min|             Avg|         Std_Dev|
+---------+----------+----------------+----------------+
|27.820816|-30.641708|0.514368695782223|6.255974012101941|
+---------+----------+----------------+----------------+
```

Feature: direct_evaporation_cease_soil_moisture_surface

```
+----------+---+------------------+------------------+
|       Max|Min|               Avg|           Std_Dev|
+----------+---+------------------+------------------+
|0.13499999|0.0|0.031106272933598897|0.03738218278900426|
+----------+---+------------------+------------------+
```