

HottestTemperature

November 29, 2017

```
In [1]: 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)
```

Welcome to

```
  _--_
 /  _/  _--_  _--_  _--_  _--_  _--_
 \  \ /  _ \  _ \  _ \  _ \  _ \
/_  /  . _/ \_/_/_/_/_/_/_/_/_/_  version 2.2.0
/_/_/
```

Using Python version 3.6.3 (default, Oct 6 2017 12:04:38)
SparkSession available as 'spark'.

```
In [2]: df = spark.read.format('csv').option('sep', '\t').schema(schema).load('inputs/nam_2015')

In [3]: sorted_temp_df = df.sort(df.temperature_surface.desc())

In [4]: sorted_temp_df.select(sorted_temp_df.Geohash,sorted_temp_df.temperature_surface).show(1)
```

```
+-----+-----+
|      Geohash|temperature_surface|
+-----+-----+
```

d5f0jqerq27b	330.67432
d5f0vd8eb80p	330.64062
9g77js659k20	330.6045

+-----+-----+

only showing top 3 rows