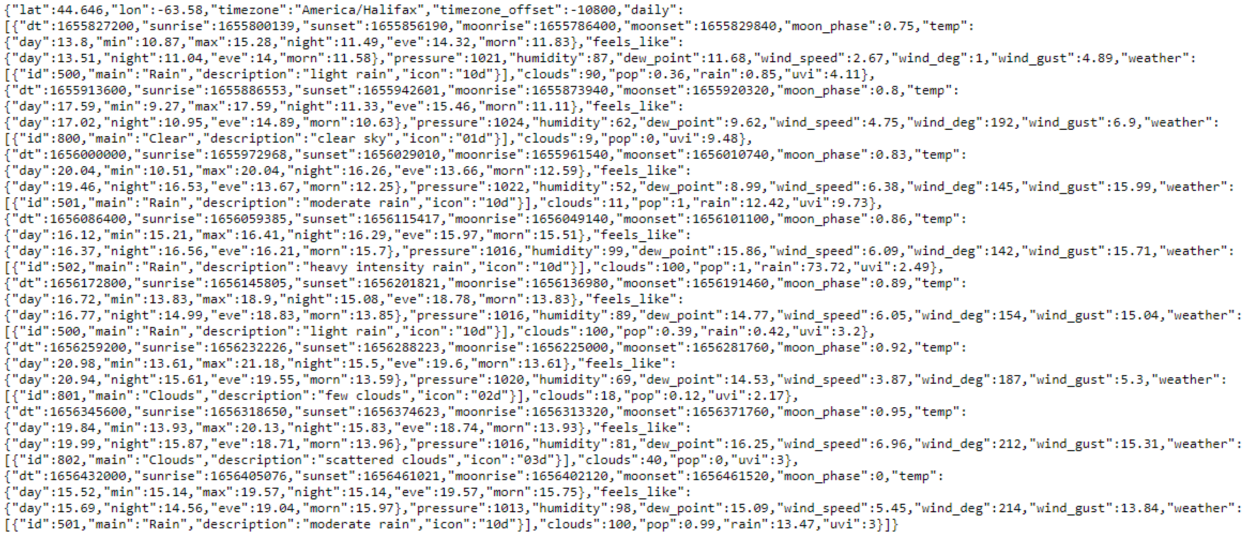
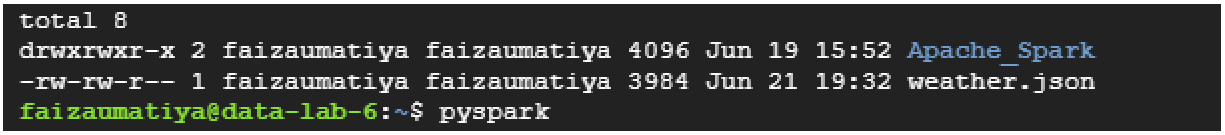
LAB TASK - 4

Initial Data:



```
{"lat":44.646,"lon":-63.58,"timezone":"America/Halifax","timezone_offset":-10800,"daily":
[{"dt":1655827200,"sunrise":1655800139,"sunset":1655856190,"moonrise":1655786400,"moonset":1655829840,"moon_phase":0.75,"temp":
{"day":13.8,"min":10.87,"max":15.28,"night":11.49,"eve":14.32,"morn":11.83},"feels_like":
{"day":13.51,"night":11.04,"eve":14,"morn":11.58},"pressure":1021,"humidity":87,"dew_point":11.68,"wind_speed":2.67,"wind_deg":1,"wind_gust":4.89,"weather":
[{"id":500,"main":"Rain","description":"light rain","icon":"10d"}],"clouds":90,"pop":0.36,"rain":0.85,"uvi":4.11},
{"dt":1655913600,"sunrise":1655886553,"sunset":1655942601,"moonrise":1655873940,"moonset":1655920320,"moon_phase":0.8,"temp":
{"day":17.59,"min":9.27,"max":17.59,"night":11.33,"eve":15.46,"morn":11.11},"feels_like":
{"day":17.02,"night":10.95,"eve":14.89,"morn":10.63},"pressure":1024,"humidity":62,"dew_point":9.62,"wind_speed":4.75,"wind_deg":192,"wind_gust":6.9,"weather":
[{"id":800,"main":"Clear","description":"clear sky","icon":"01d"}],"clouds":9,"pop":0,"uvi":9.48},
{"dt":1656000000,"sunrise":1655972968,"sunset":1656029010,"moonrise":1655961540,"moonset":1656010740,"moon_phase":0.83,"temp":
{"day":20.04,"min":10.51,"max":20.04,"night":16.26,"eve":13.66,"morn":12.59},"feels_like":
{"day":19.46,"night":16.53,"eve":13.67,"morn":12.25},"pressure":1022,"humidity":52,"dew_point":8.99,"wind_speed":6.38,"wind_deg":145,"wind_gust":15.99,"weather":
[{"id":501,"main":"Rain","description":"moderate rain","icon":"10d"}],"clouds":11,"pop":1,"rain":12.42,"uvi":9.73},
{"dt":1656086400,"sunrise":1656059385,"sunset":1656115417,"moonrise":1656049140,"moonset":1656101100,"moon_phase":0.86,"temp":
{"day":16.12,"min":15.21,"max":16.41,"night":16.29,"eve":15.97,"morn":15.51},"feels_like":
{"day":16.37,"night":16.56,"eve":16.21,"morn":15.7},"pressure":1016,"humidity":99,"dew_point":15.86,"wind_speed":6.09,"wind_deg":142,"wind_gust":15.71,"weather":
[{"id":502,"main":"Rain","description":"heavy intensity rain","icon":"10d"}],"clouds":100,"pop":1,"rain":73.72,"uvi":2.49},
{"dt":1656172800,"sunrise":1656145805,"sunset":1656201821,"moonrise":1656136980,"moonset":1656191460,"moon_phase":0.89,"temp":
{"day":16.72,"min":13.83,"max":18.9,"night":15.08,"eve":18.78,"morn":13.83},"feels_like":
{"day":16.77,"night":14.99,"eve":18.83,"morn":13.85},"pressure":1016,"humidity":89,"dew_point":14.77,"wind_speed":6.05,"wind_deg":154,"wind_gust":15.04,"weather":
[{"id":500,"main":"Rain","description":"light rain","icon":"10d"}],"clouds":100,"pop":0.39,"rain":0.42,"uvi":3.2},
{"dt":1656259200,"sunrise":1656232226,"sunset":1656288223,"moonrise":1656225000,"moonset":1656281760,"moon_phase":0.92,"temp":
{"day":20.98,"min":13.61,"max":21.18,"night":15.5,"eve":19.6,"morn":13.61},"feels_like":
{"day":20.94,"night":15.61,"eve":19.55,"morn":13.59},"pressure":1020,"humidity":69,"dew_point":14.53,"wind_speed":3.87,"wind_deg":187,"wind_gust":5.3,"weather":
[{"id":801,"main":"Clouds","description":"few clouds","icon":"02d"}],"clouds":18,"pop":0.12,"uvi":2.17},
{"dt":1656345600,"sunrise":1656318650,"sunset":1656374623,"moonrise":1656313320,"moonset":1656371760,"moon_phase":0.95,"temp":
{"day":19.84,"min":13.93,"max":20.13,"night":15.83,"eve":18.74,"morn":13.93},"feels_like":
{"day":19.99,"night":15.87,"eve":18.71,"morn":13.96},"pressure":1016,"humidity":81,"dew_point":16.25,"wind_speed":6.96,"wind_deg":212,"wind_gust":15.31,"weather":
[{"id":802,"main":"Clouds","description":"scattered clouds","icon":"03d"}],"clouds":40,"pop":0,"uvi":3},
{"dt":1656432000,"sunrise":1656405076,"sunset":1656461021,"moonrise":1656402120,"moonset":1656461520,"moon_phase":0,"temp":
{"day":15.52,"min":15.14,"max":19.57,"night":15.14,"eve":19.57,"morn":15.75},"feels_like":
{"day":15.69,"night":14.56,"eve":19.04,"morn":15.97},"pressure":1013,"humidity":98,"dew_point":15.09,"wind_speed":5.45,"wind_deg":214,"wind_gust":13.84,"weather":
[{"id":501,"main":"Rain","description":"moderate rain","icon":"10d"}],"clouds":100,"pop":0.99,"rain":13.47,"uvi":3}]}
```

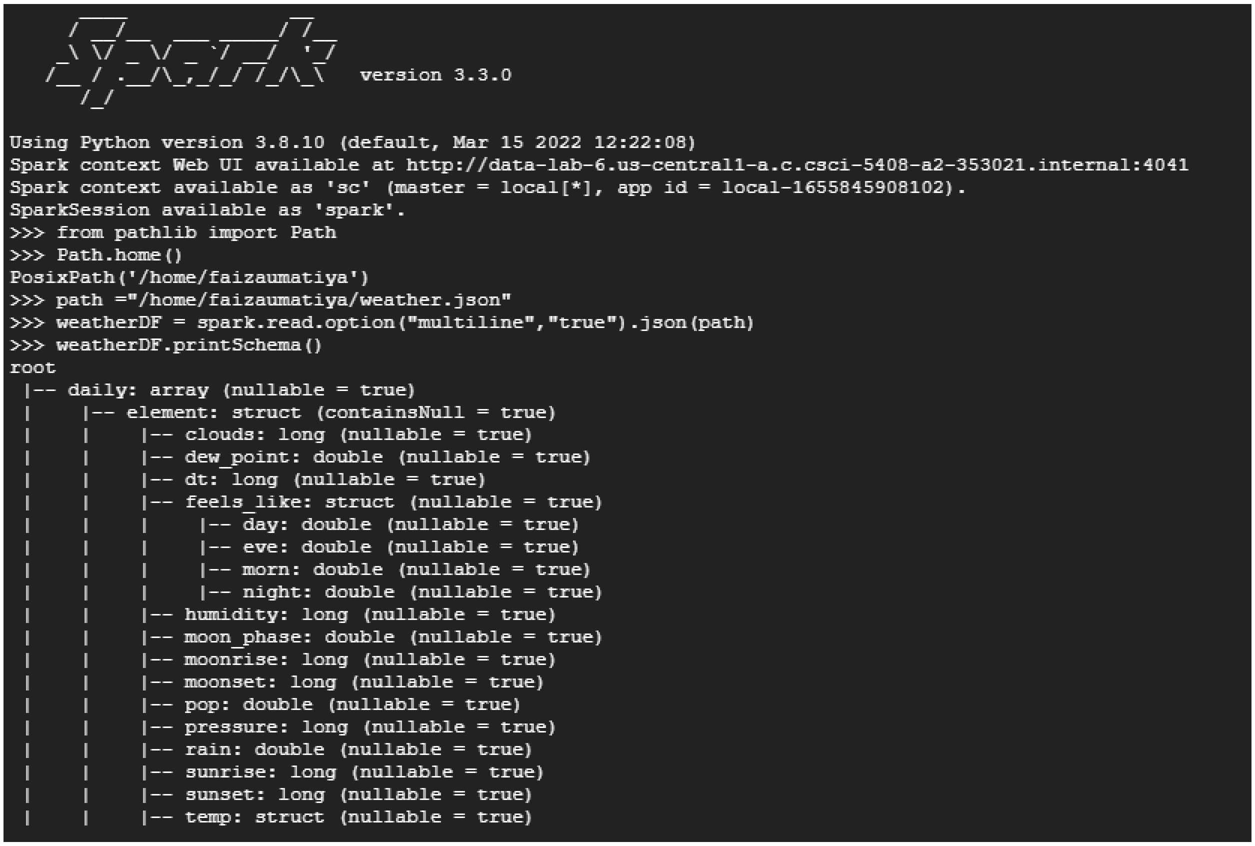


```
total 8
drwxrwxr-x 2 faizaumatiya faizaumatiya 4096 Jun 19 15:52 Apache_Spark
-rw-rw-r-- 1 faizaumatiya faizaumatiya 3984 Jun 21 19:32 weather.json
faizaumatiya@data-lab-6:~$ pyspark
```

Code:

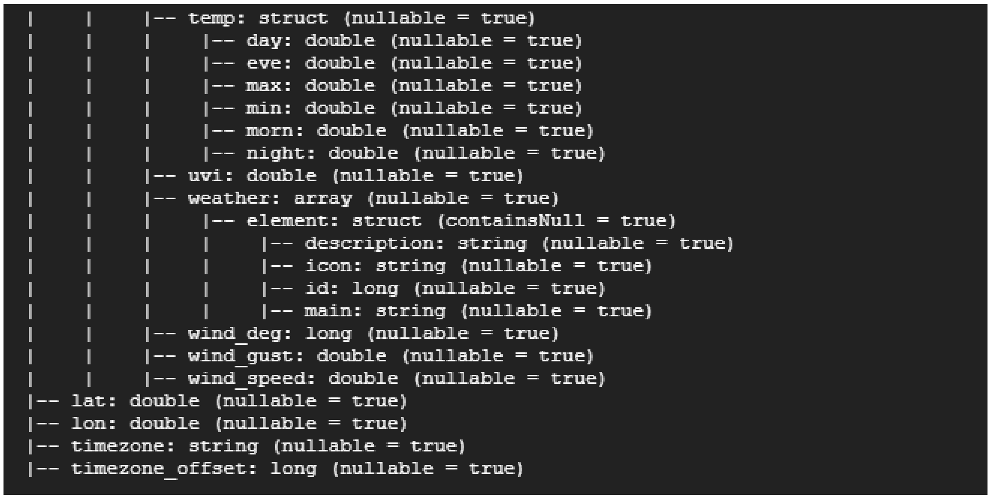
1. path ="/home/faizaumatiya/weather.json"
2. weatherDF = spark.read.option("multiline","true").json(path)
3. weatherDF.printSchema()
4. weatherDF.createOrReplaceTempView("weather")
5. halifaxWeatherDF=spark.sql("SELECT weather.daily FROM weather")
6. weatherDF.show()
7. from pyspark.sql.functions import expr
8. from pyspark.sql.functions import expr
9. halifaxTempDF = halifaxWeatherDF.withColumn("weather.daily.feels_like.day", expr("filter(weather.daily.feels_like.day, x -> (x < 16))"))
10. halifaxTempDF.show()
11. results = halifaxTempDF.toJSON().collect()
12. results

Final Output:

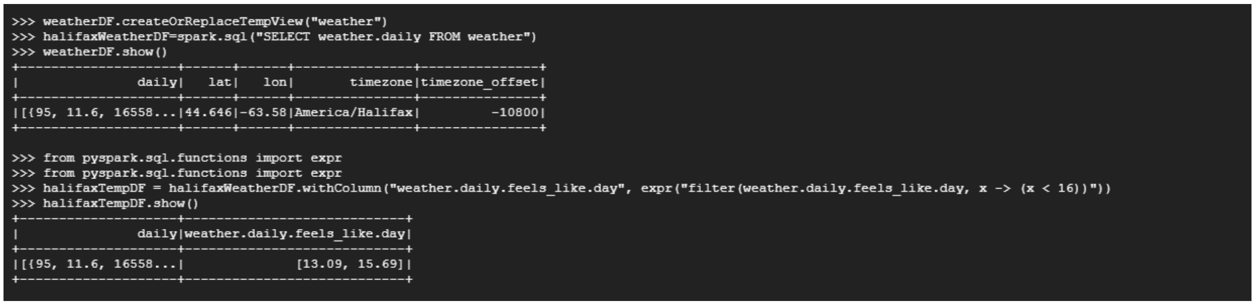


```
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 3.3.0
      /_/

Using Python version 3.8.10 (default, Mar 15 2022 12:22:08)
Spark context Web UI available at http://data-lab-6.us-central1-a.c.csci-5408-a2-353021.internal:4041
Spark context available as 'sc' (master = local[*], app id = local-1655845908102).
SparkSession available as 'spark'.
>>> from pathlib import Path
>>> Path.home()
PosixPath('/home/faizaumatiya')
>>> path ="/home/faizaumatiya/weather.json"
>>> weatherDF = spark.read.option("multiline","true").json(path)
>>> weatherDF.printSchema()
root
 |-- daily: array (nullable = true)
 |    |-- element: struct (containsNull = true)
 |    |    |-- clouds: long (nullable = true)
 |    |    |-- dew_point: double (nullable = true)
 |    |    |-- dt: long (nullable = true)
 |    |    |-- feels_like: struct (nullable = true)
 |    |    |    |-- day: double (nullable = true)
 |    |    |    |-- eve: double (nullable = true)
 |    |    |    |-- morn: double (nullable = true)
 |    |    |    |-- night: double (nullable = true)
 |    |    |-- humidity: long (nullable = true)
 |    |    |-- moon_phase: double (nullable = true)
 |    |    |-- moonrise: long (nullable = true)
 |    |    |-- moonset: long (nullable = true)
 |    |    |-- pop: double (nullable = true)
 |    |    |-- pressure: long (nullable = true)
 |    |    |-- rain: double (nullable = true)
 |    |    |-- sunrise: long (nullable = true)
 |    |    |-- sunset: long (nullable = true)
 |    |    |-- temp: struct (nullable = true)
```



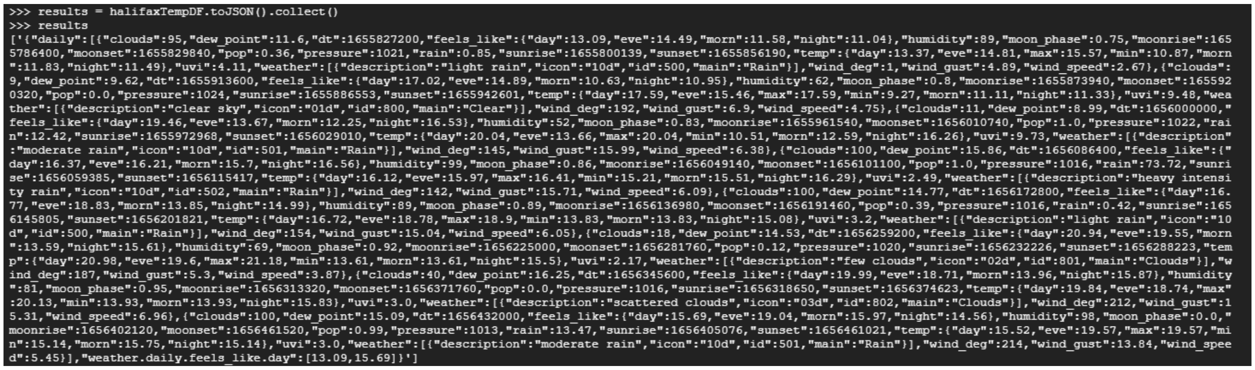
```
 |    |    |-- temp: struct (nullable = true)
 |    |    |    |-- day: double (nullable = true)
 |    |    |    |-- eve: double (nullable = true)
 |    |    |    |-- max: double (nullable = true)
 |    |    |    |-- min: double (nullable = true)
 |    |    |    |-- morn: double (nullable = true)
 |    |    |    |-- night: double (nullable = true)
 |    |    |-- uvi: double (nullable = true)
 |    |    |-- weather: array (nullable = true)
 |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |-- icon: string (nullable = true)
 |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |-- main: string (nullable = true)
 |    |    |-- wind_deg: long (nullable = true)
 |    |    |-- wind_gust: double (nullable = true)
 |    |    |-- wind_speed: double (nullable = true)
 |-- lat: double (nullable = true)
 |-- lon: double (nullable = true)
 |-- timezone: string (nullable = true)
 |-- timezone_offset: long (nullable = true)
```



```
>>> weatherDF.createOrReplaceTempView("weather")
>>> halifaxWeatherDF=spark.sql("SELECT weather.daily FROM weather")
>>> weatherDF.show()
+--------------------+------+------+--------------+---------------+
|               daily|   lat|   lon|      timezone|timezone_offset|
+--------------------+------+------+--------------+---------------+
|[{95, 11.6, 16558...|44.646|-63.58|America/Halifax|         -10800|
+--------------------+------+------+--------------+---------------+

>>> from pyspark.sql.functions import expr
>>> from pyspark.sql.functions import expr
>>> halifaxTempDF = halifaxWeatherDF.withColumn("weather.daily.feels_like.day", expr("filter(weather.daily.feels_like.day, x -> (x < 16))"))
>>> halifaxTempDF.show()
+--------------------+----------------------------+
|               daily|weather.daily.feels_like.day|
+--------------------+----------------------------+
|[{95, 11.6, 16558...|              [13.09, 15.69]|
+--------------------+----------------------------+
```

Final data (Filtered):



```
>>> results = halifaxTempDF.toJSON().collect()
>>> results
['{"daily":[{"clouds":95,"dew_point":11.6,"dt":1655827200,"feels_like":{"day":13.09,"eve":14.49,"morn":11.58,"night":11.04},"humidity":89,"moon_phase":0.75,"moonrise":165
5786400,"moonset":1655829840,"pop":0.36,"pressure":1021,"rain":0.85,"sunrise":1655800139,"sunset":1655856190,"temp":{"day":13.37,"eve":14.81,"max":15.57,"min":10.87,"morn
":11.83,"night":11.49},"uvi":4.11,"weather":[{"description":"light rain","icon":"10d","id":500,"main":"Rain"}],"wind_deg":1,"wind_gust":4.89,"wind_speed":2.67},{"clouds":
9,"dew_point":9.62,"dt":1655913600,"feels_like":{"day":17.02,"eve":14.89,"morn":10.63,"night":10.95},"humidity":62,"moon_phase":0.8,"moonrise":1655873940,"moonset":165592
0320,"pop":0.0,"pressure":1024,"sunrise":1655886553,"sunset":1655942601,"temp":{"day":17.59,"eve":15.46,"max":17.59,"min":9.27,"morn":11.11,"night":11.33},"uvi":9.48,"wea
ther":[{"description":"clear sky","icon":"01d","id":800,"main":"Clear"}],"wind_deg":192,"wind_gust":6.9,"wind_speed":4.75},{"clouds":11,"dew_point":8.99,"dt":1656000000,"
feels_like":{"day":19.46,"eve":13.67,"morn":12.25,"night":16.53},"humidity":52,"moon_phase":0.83,"moonrise":1655961540,"moonset":1656010740,"pop":1.0,"pressure":1022,"rai
n":12.42,"sunrise":1655972968,"sunset":1656029010,"temp":{"day":20.04,"eve":13.66,"max":20.04,"min":10.51,"morn":12.59,"night":16.26},"uvi":9.73,"weather":[{"description"
:"moderate rain","icon":"10d","id":501,"main":"Rain"}],"wind_deg":145,"wind_gust":15.99,"wind_speed":6.38},{"clouds":100,"dew_point":15.86,"dt":1656086400,"feels_like":{"
day":16.37,"eve":16.21,"morn":15.7,"night":16.56},"humidity":99,"moon_phase":0.86,"moonrise":1656049140,"moonset":1656101100,"pop":1.0,"pressure":1016,"rain":73.72,"sunri
se":1656059385,"sunset":1656115417,"temp":{"day":16.12,"eve":15.97,"max":16.41,"min":15.21,"morn":15.51,"night":16.29},"uvi":2.49,"weather":[{"description":"heavy intensi
ty rain","icon":"10d","id":502,"main":"Rain"}],"wind_deg":142,"wind_gust":15.71,"wind_speed":6.09},{"clouds":100,"dew_point":14.77,"dt":1656172800,"feels_like":{"day":16.
77,"eve":18.83,"morn":13.85,"night":14.99},"humidity":89,"moon_phase":0.89,"moonrise":1656136980,"moonset":1656191460,"pop":0.39,"pressure":1016,"rain":0.42,"sunrise":165
6145805,"sunset":1656201821,"temp":{"day":16.72,"eve":18.78,"max":18.9,"min":13.83,"morn":13.83,"night":15.08},"uvi":3.2,"weather":[{"description":"light rain","icon":"10
d","id":500,"main":"Rain"}],"wind_deg":154,"wind_gust":15.04,"wind_speed":6.05},{"clouds":18,"dew_point":14.53,"dt":1656259200,"feels_like":{"day":20.94,"eve":19.55,"morn
":13.59,"night":15.61},"humidity":69,"moon_phase":0.92,"moonrise":1656225000,"moonset":1656281760,"pop":0.12,"pressure":1020,"sunrise":1656232226,"sunset":1656288223,"tem
p":{"day":20.98,"eve":19.6,"max":21.18,"min":13.61,"morn":13.61,"night":15.5},"uvi":2.17,"weather":[{"description":"few clouds","icon":"02d","id":801,"main":"Clouds"}],"w
ind_deg":187,"wind_gust":5.3,"wind_speed":3.87},{"clouds":40,"dew_point":16.25,"dt":1656345600,"feels_like":{"day":19.99,"eve":18.71,"morn":13.96,"night":15.87},"humidity
":81,"moon_phase":0.95,"moonrise":1656313320,"moonset":1656371760,"pop":0.0,"pressure":1016,"sunrise":1656318650,"sunset":1656374623,"temp":{"day":19.84,"eve":18.74,"max"
:20.13,"min":13.93,"morn":13.93,"night":15.83},"uvi":3.0,"weather":[{"description":"scattered clouds","icon":"03d","id":802,"main":"Clouds"}],"wind_deg":212,"wind_gust":1
5.31,"wind_speed":6.96},{"clouds":100,"dew_point":15.09,"dt":1656432000,"feels_like":{"day":15.69,"eve":19.04,"morn":15.97,"night":14.56},"humidity":98,"moon_phase":0.0,"
moonrise":1656402120,"moonset":1656461520,"pop":0.99,"pressure":1013,"rain":13.47,"sunrise":1656405076,"sunset":1656461021,"temp":{"day":15.52,"eve":19.57,"max":19.57,"mi
n":15.14,"morn":15.75,"night":15.14},"uvi":3.0,"weather":[{"description":"moderate rain","icon":"10d","id":501,"main":"Rain"}],"wind_deg":214,"wind_gust":13.84,"wind_spee
d":5.45}],"weather.daily.feels_like.day":[13.09,15.69]}']
```

References:

1. https://sparkbyexamples.com/spark/spark-read-and-write-json-file/
2. https://stackoverflow.com/questions/43269244/pyspark-dataframe-write-to-single-json-file-with-specific-name
3. https://dal.brightspace.com/d2l/le/content/221749/viewContent/3049202/View