# WEEK 5-8 ASSIGNMENT

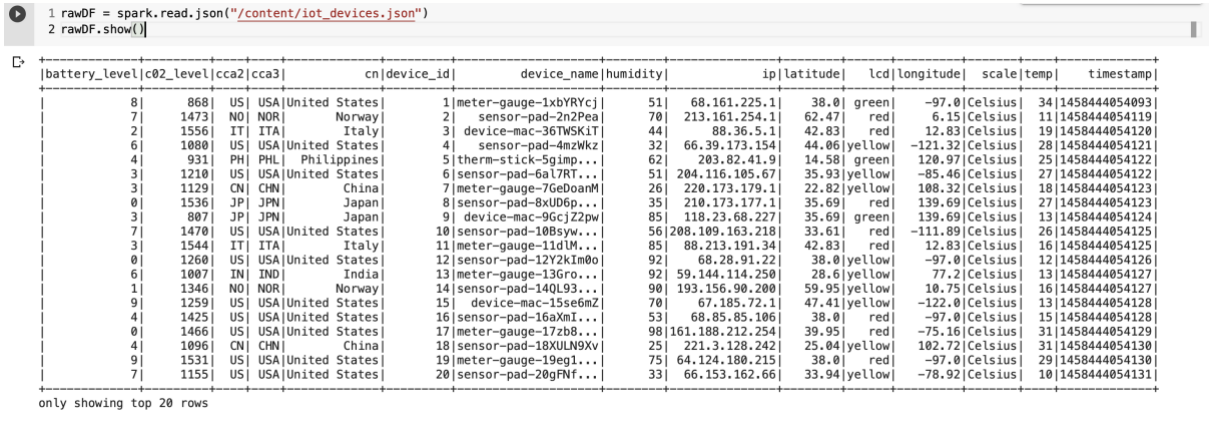
## PART 1:

1. Read the data into a Data frame.

Command:

```
rawDF = spark.read.json("/content/iot_devices.json")
rawDF.show()
```

Screenshot:



```
1 rawDF = spark.read.json("/content/iot_devices.json")
2 rawDF.show()
```

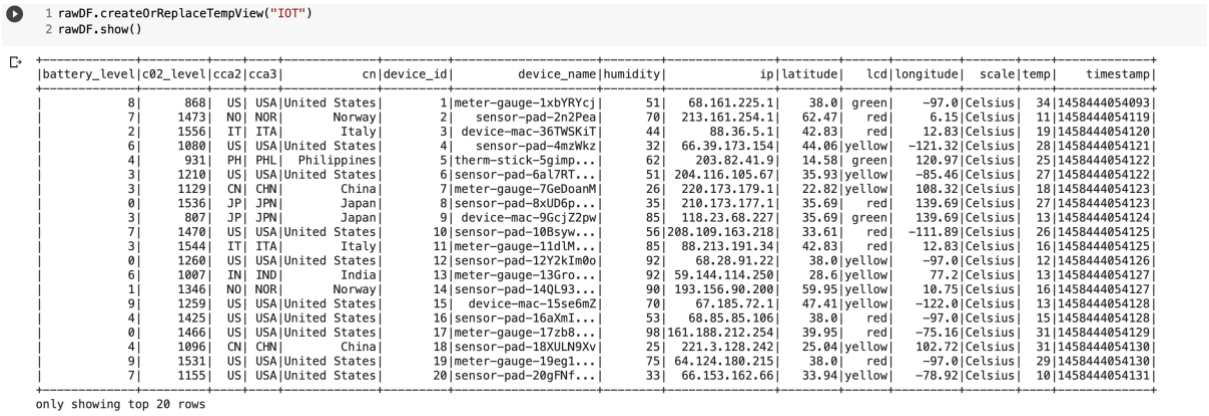
```
+-------------+----------+-----+-----+--------------+---------+--------------------+--------+---------------+--------+------+----------+-----+----+-------------+
|battery_level|c02_level|cca2|cca3|            cn|device_id|         device_name|humidity|             ip|latitude|   lcd|longitude|  scale|temp|    timestamp|
+-------------+----------+-----+-----+--------------+---------+--------------------+--------+---------------+--------+------+----------+-----+----+-------------+
|            8|       868|   US|  USA|United States|        1|meter-gauge-1xbYRYcj|      51|   68.161.225.1|    38.0| green|     -97.0|Celsius|  34|1458444054093|
|            7|      1473|   NO|  NOR|        Norway|        2|   sensor-pad-2n2Pea|      70|  213.161.254.1|   62.47|   red|      6.15|Celsius|  11|1458444054119|
|            2|      1556|   IT|  ITA|         Italy|        3| device-mac-36TWSKiT|      44|       88.36.5.1|   42.83|   red|     12.83|Celsius|  19|1458444054120|
|            6|      1080|   US|  USA|United States|        4|   sensor-pad-4mzWkz|      32|  66.39.173.154|   44.06|yellow|   -121.32|Celsius|  28|1458444054121|
|            4|       931|   PH|  PHL|   Philippines|        5|therm-stick-5gimp...|      62|     203.82.41.9|   14.58| green|    120.97|Celsius|  25|1458444054122|
|            3|      1210|   US|  USA|United States|        6|sensor-pad-6al7RT...|      51| 204.116.105.67|   35.93|yellow|    -85.46|Celsius|  27|1458444054122|
|            3|      1129|   CN|  CHN|         China|        7|meter-gauge-7GeDoanM|      26|  220.173.179.1|   22.82|yellow|    108.32|Celsius|  18|1458444054123|
|            0|      1536|   JP|  JPN|         Japan|        8|sensor-pad-8xUD6p...|      35|  210.173.177.1|   35.69|   red|    139.69|Celsius|  27|1458444054123|
|            3|       807|   JP|  JPN|         Japan|        9| device-mac-9GcjZ2pw|      85|  118.23.68.227|   35.69| green|    139.69|Celsius|  13|1458444054124|
|            7|      1470|   US|  USA|United States|       10|sensor-pad-10Bsyw...|      56|208.109.163.218|   33.61|   red|   -111.89|Celsius|  26|1458444054125|
|            3|      1544|   IT|  ITA|         Italy|       11|meter-gauge-11dlM...|      85|  88.213.191.34|   42.83|   red|     12.83|Celsius|  16|1458444054125|
|            0|      1260|   US|  USA|United States|       12|sensor-pad-12Y2kIm0o|      92|    68.28.91.22|    38.0|yellow|     -97.0|Celsius|  12|1458444054126|
|            6|      1007|   IN|  IND|         India|       13|meter-gauge-13Gro...|      92| 59.144.114.250|   28.6|yellow|      77.2|Celsius|  13|1458444054127|
|            1|      1346|   NO|  NOR|        Norway|       14|sensor-pad-14QL93...|      90| 193.156.90.200|   59.95|yellow|     10.75|Celsius|  16|1458444054127|
|            9|      1259|   US|  USA|United States|       15|  device-mac-15se6mZ|      70|    67.185.72.1|   47.41|yellow|    -122.0|Celsius|  13|1458444054128|
|            4|      1425|   US|  USA|United States|       16|sensor-pad-16aXmI...|      53|   68.85.85.106|    38.0|   red|     -97.0|Celsius|  15|1458444054128|
|            0|      1466|   US|  USA|United States|       17|meter-gauge-17zb8...|      98|161.188.212.254|   39.95|   red|    -75.16|Celsius|  31|1458444054129|
|            4|      1096|   CN|  CHN|         China|       18|sensor-pad-18XULN9Xv|      25|  221.3.128.242|   25.04|yellow|    102.72|Celsius|  31|1458444054130|
|            9|      1531|   US|  USA|United States|       19|meter-gauge-19eg1...|      75| 64.124.180.215|    38.0|   red|     -97.0|Celsius|  29|1458444054130|
|            7|      1155|   US|  USA|United States|       20|sensor-pad-20gFNf...|      33|  66.153.162.66|   33.94|yellow|    -78.92|Celsius|  10|1458444054131|
+-------------+----------+-----+-----+--------------+---------+--------------------+--------+---------------+--------+------+----------+-----+----+-------------+
only showing top 20 rows
```

2. Convert the Dataframe into a temporary view called iot.

Command:

```
rawDF.createOrReplaceTempView("IOT")
```

Screenshot:



```
1 rawDF.createOrReplaceTempView("IOT")
2 rawDF.show()
```

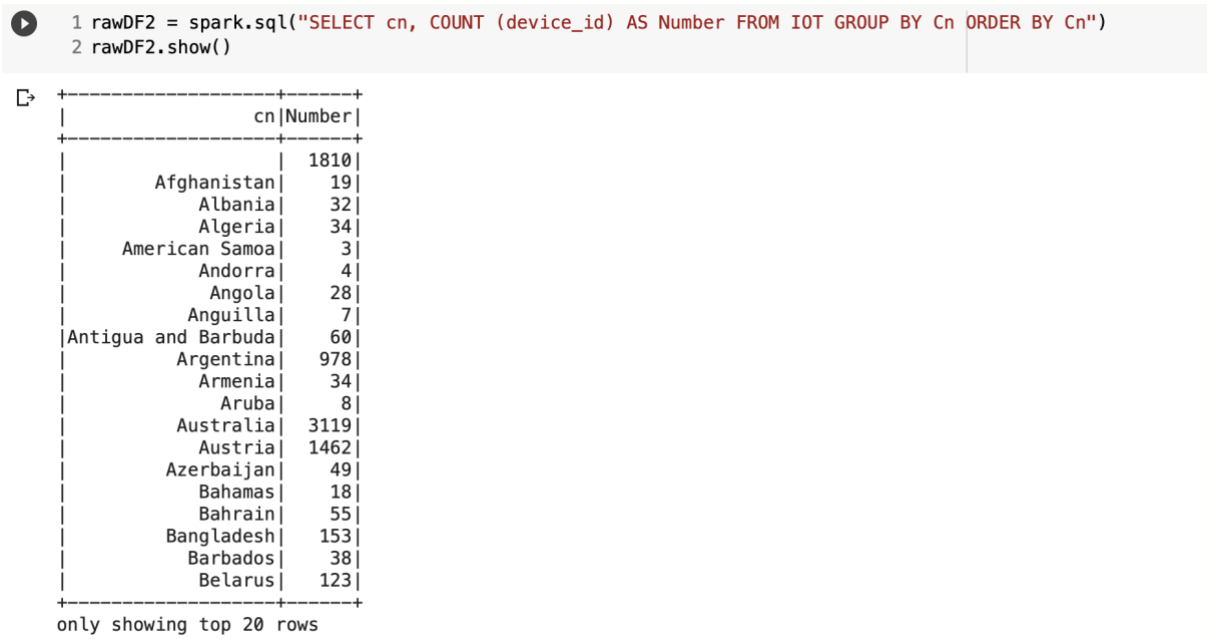
```
+-------------+----------+-----+-----+--------------+---------+--------------------+--------+---------------+--------+------+----------+-----+----+-------------+
|battery_level|c02_level|cca2|cca3|            cn|device_id|         device_name|humidity|             ip|latitude|   lcd|longitude|  scale|temp|    timestamp|
+-------------+----------+-----+-----+--------------+---------+--------------------+--------+---------------+--------+------+----------+-----+----+-------------+
|            8|       868|   US|  USA|United States|        1|meter-gauge-1xbYRYcj|      51|   68.161.225.1|    38.0| green|     -97.0|Celsius|  34|1458444054093|
|            7|      1473|   NO|  NOR|        Norway|        2|   sensor-pad-2n2Pea|      70|  213.161.254.1|   62.47|   red|      6.15|Celsius|  11|1458444054119|
|            2|      1556|   IT|  ITA|         Italy|        3| device-mac-36TWSKiT|      44|       88.36.5.1|   42.83|   red|     12.83|Celsius|  19|1458444054120|
|            6|      1080|   US|  USA|United States|        4|   sensor-pad-4mzWkz|      32|  66.39.173.154|   44.06|yellow|   -121.32|Celsius|  28|1458444054121|
|            4|       931|   PH|  PHL|   Philippines|        5|therm-stick-5gimp...|      62|     203.82.41.9|   14.58| green|    120.97|Celsius|  25|1458444054122|
|            3|      1210|   US|  USA|United States|        6|sensor-pad-6al7RT...|      51| 204.116.105.67|   35.93|yellow|    -85.46|Celsius|  27|1458444054122|
|            3|      1129|   CN|  CHN|         China|        7|meter-gauge-7GeDoanM|      26|  220.173.179.1|   22.82|yellow|    108.32|Celsius|  18|1458444054123|
|            0|      1536|   JP|  JPN|         Japan|        8|sensor-pad-8xUD6p...|      35|  210.173.177.1|   35.69|   red|    139.69|Celsius|  27|1458444054123|
|            3|       807|   JP|  JPN|         Japan|        9| device-mac-9GcjZ2pw|      85|  118.23.68.227|   35.69| green|    139.69|Celsius|  13|1458444054124|
|            7|      1470|   US|  USA|United States|       10|sensor-pad-10Bsyw...|      56|208.109.163.218|   33.61|   red|   -111.89|Celsius|  26|1458444054125|
|            3|      1544|   IT|  ITA|         Italy|       11|meter-gauge-11dlM...|      85|  88.213.191.34|   42.83|   red|     12.83|Celsius|  16|1458444054125|
|            0|      1260|   US|  USA|United States|       12|sensor-pad-12Y2kIm0o|      92|    68.28.91.22|    38.0|yellow|     -97.0|Celsius|  12|1458444054126|
|            6|      1007|   IN|  IND|         India|       13|meter-gauge-13Gro...|      92| 59.144.114.250|   28.6|yellow|      77.2|Celsius|  13|1458444054127|
|            1|      1346|   NO|  NOR|        Norway|       14|sensor-pad-14QL93...|      90| 193.156.90.200|   59.95|yellow|     10.75|Celsius|  16|1458444054127|
|            9|      1259|   US|  USA|United States|       15|  device-mac-15se6mZ|      70|    67.185.72.1|   47.41|yellow|    -122.0|Celsius|  13|1458444054128|
|            4|      1425|   US|  USA|United States|       16|sensor-pad-16aXmI...|      53|   68.85.85.106|    38.0|   red|     -97.0|Celsius|  15|1458444054128|
|            0|      1466|   US|  USA|United States|       17|meter-gauge-17zb8...|      98|161.188.212.254|   39.95|   red|    -75.16|Celsius|  31|1458444054129|
|            4|      1096|   CN|  CHN|         China|       18|sensor-pad-18XULN9Xv|      25|  221.3.128.242|   25.04|yellow|    102.72|Celsius|  31|1458444054130|
|            9|      1531|   US|  USA|United States|       19|meter-gauge-19eg1...|      75| 64.124.180.215|    38.0|   red|     -97.0|Celsius|  29|1458444054130|
|            7|      1155|   US|  USA|United States|       20|sensor-pad-20gFNf...|      33|  66.153.162.66|   33.94|yellow|    -78.92|Celsius|  10|1458444054131|
+-------------+----------+-----+-----+--------------+---------+--------------------+--------+---------------+--------+------+----------+-----+----+-------------+
only showing top 20 rows
```

3. Count how many devices are there from each country and display the output.

Command:

```
rawDF2 = spark.sql("SELECT cn, COUNT (device_id) AS Number FROM IOT
GROUP BY Cn ORDER BY Cn")
rawDF2.show()
```

Screenshot:



```
1 rawDF2 = spark.sql("SELECT cn, COUNT (device_id) AS Number FROM IOT GROUP BY Cn ORDER BY Cn")
2 rawDF2.show()

+--------------------+------+
|                  cn|Number|
+--------------------+------+
|                    |  1810|
|         Afghanistan|    19|
|             Albania|    32|
|             Algeria|    34|
|      American Samoa|     3|
|             Andorra|     4|
|              Angola|    28|
|            Anguilla|     7|
|Antigua and Barbuda|    60|
|           Argentina|   978|
|             Armenia|    34|
|               Aruba|     8|
|           Australia|  3119|
|             Austria|  1462|
|          Azerbaijan|    49|
|             Bahamas|    18|
|             Bahrain|    55|
|          Bangladesh|   153|
|            Barbados|    38|
|             Belarus|   123|
+--------------------+------+
only showing top 20 rows
```

4. Display all the countries whose carbon dioxide level is more than 1400. Sort the output in descending order.

Command:

```
rawDF2 = spark.sql("SELECT cn, c02_level FROM IOT WHERE c02_level >
1400 ORDER BY c02_level DESC")
rawDF2.show()
```

Screenshot:

```
1 rawDF2 = spark.sql("SELECT cn, c02_level FROM IOT WHERE c02_level > 1400 ORDER BY c02_level DESC")
2 rawDF2.show()
```

```
+--------------+---------+
|            cn|c02_level|
+--------------+---------+
|        Poland|     1599|
|         Spain|     1599|
| United States|     1599|
|         Japan|     1599|
|       Germany|     1599|
|   Philippines|     1599|
|        Canada|     1599|
|        Canada|     1599|
| United States|     1599|
| United States|     1599|
| United States|     1599|
|        France|     1599|
| United States|     1599|
|Czech Republic|     1599|
|         China|     1599|
|         Japan|     1599|
|        Turkey|     1599|
| United States|     1599|
|Czech Republic|     1599|
| United States|     1599|
+--------------+---------+
only showing top 20 rows
```

5. Select all countries' devices with high-levels of C02 and group by cca3 and order by device_ids (Hint: For high CO2 level, the LCD status will be RED).

Command:

```
rawDF3 = spark.sql("select cca3, count(distinct device_id) as device_id from IOT where lcd == 'red' group by cca3 order by device_id desc limit 100")
rawDF3.show()
```

Screenshot:

```
1 rawDF3 = spark.sql("select cca3, count(distinct device_id) as device_id from IOT where lcd == 'red' group by cca3 order by device_id desc limit 100")
2 rawDF3.show()
```

```
+----+---------+
|cca3|device_id|
+----+---------+
| USA|    17489|
| CHN|     3616|
| KOR|     2942|
| JPN|     2935|
| DEU|     1966|
| GBR|     1660|
| CAN|     1564|
| RUS|     1508|
| FRA|     1353|
| BRA|      856|
| AUS|      769|
| SWE|      724|
| ITA|      713|
| POL|      664|
| NLD|      646|
| ESP|      586|
| TWN|      542|
| IND|      446|
| NOR|      399|
| UKR|      373|
+----+---------+
only showing top 20 rows
```

6. find out all devices in countries whose batteries need replacements.

Command:

```
rawDF3 = spark.sql("select cca3, count(distinct device_id) as device_id from IOT where battery_level == 0 group by cca3 order by device_id desc limit 100")
rawDF3.show()
```

Screenshot:

```
[14]  1 rawDF3 = spark.sql("select cca3, count(distinct device_id) as device_id from IOT where battery_level == 0 group by cca3 order by device_id desc limit 100")
      2 rawDF3.show()

+----+---------+
|cca3|device_id|
+----+---------+
| USA|     7043|
| CHN|     1415|
| KOR|     1217|
| JPN|     1210|
| DEU|      760|
| GBR|      650|
| CAN|      612|
| RUS|      600|
| FRA|      582|
| BRA|      374|
| AUS|      322|
| SWE|      293|
| ITA|      287|
| POL|      278|
| NLD|      251|
| ESP|      223|
| TWN|      207|
| IND|      189|
| UKR|      149|
| HKG|      149|
+----+---------+
only showing top 20 rows
```