

sparksessionpractice

```
%pyspark
# Numero de recuperados a nivel mundial
```

```
%pyspark
```

```
from pyspark.sql import SparkSession
```

```
%pyspark
```

```
spark = SparkSession.builder.appName('Basics').getOrCreate()
```

```
%pyspark
```

```
SensorDataRDD = sc.textFile('s3://aws-logs-061895363576-us-east-1/mundial/time_series_covid19_recove
```

```
SensorDataCount = SensorDataRDD.count()
```

```
SensorDataCount
```

254

```
%pyspark
```

```
df = spark.read.csv('s3://aws-logs-061895363576-us-east-1/mundial/time_series_covid19_recovered_glob
```

```
%pyspark
```

```
df.columns
```

```
['Province/State', 'Country/Region', 'Lat', 'Long', '1/22/20', '1/23/20', '1/24/20', '1/25/20', '1/26/20', '1/27/20', '1/28/20', '1/29/20', '1/30/20', '1/31/20', '2/1/20', '2/2/20', '2/3/20', '2/4/20', '2/5/20', '2/6/20', '2/7/20', '2/8/20', '2/9/20', '2/10/20', '2/11/20', '2/12/20', '2/13/20', '2/14/20', '2/15/20', '2/16/20', '2/17/20', '2/18/20', '2/19/20', '2/20/20', '2/21/20', '2/22/20', '2/23/20', '2/24/20', '2/25/20', '2/26/20', '2/27/20', '2/28/20', '2/29/20', '3/1/20', '3/2/20', '3/3/20', '3/4/20', '3/5/20', '3/6/20', '3/7/20', '3/8/20', '3/9/20', '3/10/20', '3/11/20', '3/12/20', '3/13/20', '3/14/20', '3/15/20', '3/16/20', '3/17/20', '3/18/20', '3/19/20', '3/20/20', '3/21/20', '3/22/20', '3/23/20', '3/24/20', '3/25/20', '3/26/20', '3/27/20', '3/28/20', '3/29/20', '3/30/20', '3/31/20', '4/1/20', '4/2/20', '4/3/20', '4/4/20', '4/5/20', '4/6/20', '4/7/20', '4/8/20', '4/9/20', '4/10/20', '4/11/20', '4/12/20', '4/13/20', '4/14/20', '4/15/20', '4/16/20', '4/17/20', '4/18/20', '4/19/20', '4/20/20', '4/21/20', '4/22/20', '4/23/20', '4/24/20', '4/25/20', '4/26/20', '4/27/20', '4/28/20', '4/29/20', '4/30/20', '5/1/20', '5/2/20', '5/3/20', '5/4/20', '5/5/20', '5/6/20', '5/7/20', '5/8/20', '5/9/20', '5/10/20', '5/11/20', '5/12/20', '5/13/20', '5/14/20', '5/15/20']
```

```
%pyspark
```

```
df.printSchema()
```

```
-- 2/27/20: string (nullable = true)
```

```
-- 2/28/20: string (nullable = true)
```

```
-- 2/29/20: string (nullable = true)
```

```
-- 3/1/20: string (nullable = true)
```

```
-- 3/2/20: string (nullable = true)
```

sparksessionpractice

```
df.head(5)
```

```
[Row(Province/State=None, Country/Region='Afghanistan', Lat='33.0', Long='65.0', 1/22/20='0', 1/23/20='0', 1/24/20='0', 1/25/20='0', 1/26/20='0', 1/27/20='0', 1/28/20='0', 1/29/20='0', 1/30/20='0', 1/31/20='0', 2/1/20='0', 2/2/20='0', 2/3/20='0', 2/4/20='0', 2/5/20='0', 2/6/20='0', 2/7/20='0', 2/8/20='0', 2/9/20='0', 2/10/20='0', 2/11/20='0', 2/12/20='0', 2/13/20='0', 2/14/20='0', 2/15/20='0', 2/16/20='0', 2/17/20='0', 2/18/20='0', 2/19/20='0', 2/20/20='0', 2/21/20='0', 2/22/20='0', 2/23/20='0', 2/24/20='0', 2/25/20='0', 2/26/20='0', 2/27/20='0', 2/28/20='0', 2/29/20='0', 3/1/20='0', 3/2/20='0', 3/3/20='0', 3/4/20='0', 3/5/20='0', 3/6/20='0', 3/7/20='0', 3/8/20='0', 3/9/20='0', 3/10/20='0', 3/11/20='0', 3/12/20='0', 3/13/20='0', 3/14/20='0', 3/15/20='0', 3/16/20='1', 3/17/20='1', 3/18/20='1', 3/19/20='1', 3/20/20='1', 3/21/20='1', 3/22/20='1', 3/23/20='1', 3/24/20='1', 3/25/20='2', 3/26/20='2', 3/27/20='2', 3/28/20='2', 3/29/20='2', 3/30/20='2', 3/31/20='5', 4/1/20='5', 4/2/20='10', 4/3/20='10', 4/4/20='10', 4/5/20='15', 4/6/20='18', 4/7/20='18', 4/8/20='29', 4/9/20='32', 4/10/20='32', 4/11/20='32', 4/12/20='32', 4/13/20='32', 4/14/20='40', 4/15/20='43', 4/16/20='54', 4/17/20='99', 4/18/20='112', 4/19/20='131', 4/20/20='135', 4/21/20='150', 4/22/20='166', 4/23/20='179', 4/24/20='188', 4/25/20='188', 4/26/20='207', 4/27/20='220', 4/28/20='228', 4/29/20='252', 4/30/20='260', 5/1/20='310', 5/2/20='331', 5/3/20='345', 5/4/20='397', 5/5/20='421', 5/6/20='458', 5/7/20='468', 5/8/20='472', 5/9/20='502', 5/10/20='558', 5/11/20='558', 5/12/20='610', 5/13/20='648', 5/14/20='691', 5/15/20='745'), Row(Province/State=None, Country/Region='Albania', Lat='41.1533', Long='20.1683', 1/22/20='0', 1/23/20='0', 1/24/20='0', 1/25/20='0', 1/26/20='0', 1/27/20='0', 1/28/20='0', 1/29/20='0', 1/30/20='0', 1/31/20='0', 2/1/20='0', 2/2/20='0', 2/3/20='0', 2/4/20='0', 2/5/20='0', 2/6/20='0', 2/7/20='0', 2/8/20='0', 2/9/20='0', 2/10/20='0', 2/11/20='0', 2/12/20='0', 2/13/20='0', 2/14/20='0', 2/15/20='0', 2/16/20='0', 2/17/20='0', 2/18/20='0', 2/19/20='0', 2/20/20='0', 2/21/20='0', 2/22/20='0', 2/23/20='0', 2/24/20='0', 2/25/20='0', 2/26/20='0', 2/27/20='0', 2/28/20='0', 2/29/20='0', 3/1/20='0', 3/2/20='0', 3/3/20='0', 3/4/20='0', 3/5/20='0', 3/6/20='0', 3/7/20='0', 3/8/20='0', 3/9/20='0', 3/10/20='0', 3/11/20='0', 3/12/20='0', 3/13/20='0', 3/14/20='0', 3/15/20='0', 3/16/20='0', 3/17/20='0', 3/18/20='0', 3/19/20='0', 3/20/20='0', 3/21/20='0', 3/22/20='0', 3/23/20='0', 3/24/20='0', 3/25/20='0', 3/26/20='0', 3/27/20='0', 3/28/20='0', 3/29/20='0', 3/30/20='0', 3/31/20='0', 4/1/20='0', 4/2/20='0', 4/3/20='0', 4/4/20='0', 4/5/20='0', 4/6/20='0', 4/7/20='0', 4/8/20='0', 4/9/20='0', 4/10/20='0', 4/11/20='0', 4/12/20='0', 4/13/20='0', 4/14/20='0', 4/15/20='0', 4/16/20='0', 4/17/20='0', 4/18/20='0', 4/19/20='0', 4/20/20='0', 4/21/20='0', 4/22/20='0', 4/23/20='0', 4/24/20='0', 4/25/20='0', 4/26/20='0', 4/27/20='0', 4/28/20='0', 4/29/20='0', 4/30/20='0', 5/1/20='0', 5/2/20='0', 5/3/20='0', 5/4/20='0', 5/5/20='0', 5/6/20='0', 5/7/20='0', 5/8/20='0', 5/9/20='0', 5/10/20='0', 5/11/20='0', 5/12/20='0', 5/13/20='0', 5/14/20='0', 5/15/20='0', 5/16/20='0', 5/17/20='0', 5/18/20='0', 5/19/20='0', 5/20/20='0', 5/21/20='0', 5/22/20='0', 5/23/20='0', 5/24/20='0', 5/25/20='0', 5/26/20='0', 5/27/20='0', 5/28/20='0', 5/29/20='0', 5/30/20='0', 5/31/20='0', 6/1/20='0', 6/2/20='0', 6/3/20='0', 6/4/20='0', 6/5/20='0', 6/6/20='0', 6/7/20='0', 6/8/20='0', 6/9/20='0', 6/10/20='0', 6/11/20='0', 6/12/20='0', 6/13/20='0', 6/14/20='0', 6/15/20='0', 6/16/20='0', 6/17/20='0', 6/18/20='0', 6/19/20='0', 6/20/20='0', 6/21/20='0', 6/22/20='0', 6/23/20='0', 6/24/20='0', 6/25/20='0', 6/26/20='0', 6/27/20='0', 6/28/20='0', 6/29/20='0', 6/30/20='0', 7/1/20='0', 7/2/20='0', 7/3/20='0', 7/4/20='0', 7/5/20='0', 7/6/20='0', 7/7/20='0', 7/8/20='0', 7/9/20='0', 7/10/20='0', 7/11/20='0', 7/12/20='0', 7/13/20='0', 7/14/20='0', 7/15/20='0', 7/16/20='0', 7/17/20='0', 7/18/20='0', 7/19/20='0', 7/20/20='0', 7/21/20='0', 7/22/20='0', 7/23/20='0', 7/24/20='0', 7/25/20='0', 7/26/20='0', 7/27/20='0', 7/28/20='0', 7/29/20='0', 7/30/20='0', 7/31/20='0', 8/1/20='0', 8/2/20='0', 8/3/20='0', 8/4/20='0', 8/5/20='0', 8/6/20='0', 8/7/20='0', 8/8/20='0', 8/9/20='0', 8/10/20='0', 8/11/20='0', 8/12/20='0', 8/13/20='0', 8/14/20='0', 8/15/20='0', 8/16/20='0', 8/17/20='0', 8/18/20='0', 8/19/20='0', 8/20/20='0', 8/21/20='0', 8/22/20='0', 8/23/20='0', 8/24/20='0', 8/25/20='0', 8/26/20='0', 8/27/20='0', 8/28/20='0', 8/29/20='0', 8/30/20='0', 8/31/20='0', 9/1/20='0', 9/2/20='0', 9/3/20='0', 9/4/20='0', 9/5/20='0', 9/6/20='0', 9/7/20='0', 9/8/20='0', 9/9/20='0', 9/10/20='0', 9/11/20='0', 9/12/20='0', 9/13/20='0', 9/14/20='0', 9/15/20='0', 9/16/20='0', 9/17/20='0', 9/18/20='0', 9/19/20='0', 9/20/20='0', 9/21/20='0', 9/22/20='0', 9/23/20='0', 9/24/20='0', 9/25/20='0', 9/26/20='0', 9/27/20='0', 9/28/20='0', 9/29/20='0', 9/30/20='0', 10/1/20='0', 10/2/20
```

```
df.describe().show()
```

A handwriting practice sheet featuring 15 rows of dashed lines. Each row contains a series of plus signs (+) positioned at regular intervals along the dashed line, intended for tracing practice. The plus signs are small and black, and the dashed lines are thin and grey. The rows are evenly spaced and cover the majority of the page area.

```
%pyspark
sparksessionpractice
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
--+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      null| Korea, South|      36.0|      128.0|      0|      0|      0|      0|
0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      0|      1|      1|
3|      3|      3|      7|      7|      7|      9|      9|     10|     12|     12|     16|     16|
16|     18|     18|     22|     22|     22|     22|     27|     30|     30|     30|     41|     41|     1
35|    135|    118|    118|    247|    288|    333|    510|    510|    510|    1137|    1407|    1540|
1540|    1540|    1540|    2909|    2909|    3507|    3730|    4144|    4528|    4811|    5033|    5228|    54
08|    5567|    5828|    6021|    6325|    6463|    6598|    6694|    6776|    6973|    7117|    7243|    7368|    7447|
7534|    7616|    7757|    7829|    7937|    8042|    8114|    8213|    8277|    8501|    8635|    8717|    87
64|    8854|    8922|    9059|    9072|    9123|    9183|    9217|    9283|    9333|    9419|    9484|    9568|    9610|
9632|    9670|    9695|    9762|    9821|    9851|
|      Anhui|      China|    31.8257|    117.2264|      0|      0|      0|      0|
```

```
%pyspark
from pyspark.sql.functions import mean
```

```
%pyspark
df.select(mean("5/15/20")).show()
# Promedio de recuperados el 15 de mayo

-----+-----+
|      avg(5/15/20)|
+-----+-----+
|6470.620553359684|
+-----+-----+
```