COVID DATA ON DATABRICKS –SQL

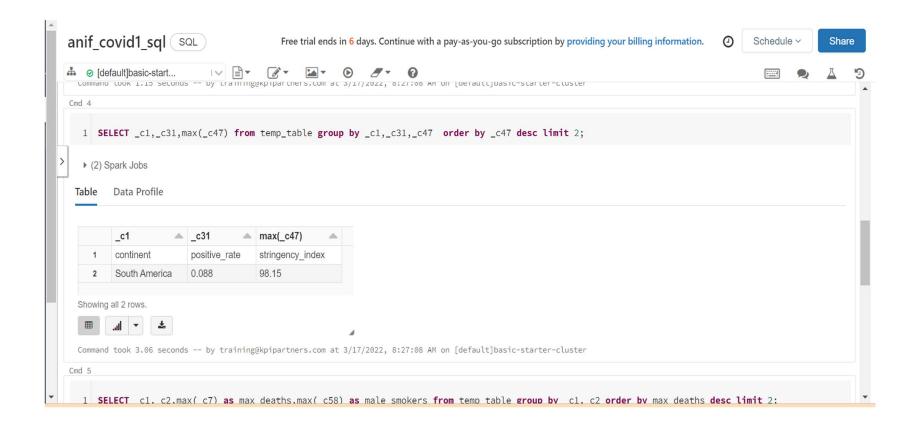## LISTING OUT THE FILES AND READING THE CSV FILE

```
Cmd 1
1  %python
2  df = spark.read.csv("dbfs:/FileStore/tables/Covid_Filtered_Data.csv")
```

▸ (1) Spark Jobs

▸ ▤ df: pyspark.sql.dataframe.DataFrame = [_c0: string, _c1: string ... 65 more fields]

Command took 0.82 seconds -- by training@kpipartners.com at 3/17/2022, 8:27:08 AM on [default]basic-starter-cluster

```
Cmd 2                                                                    Python  ▶▾ ∨ ─ ✕
1  %python
2
3  df.createOrReplaceTempView("temp_table")
4
```

Command took 0.08 seconds -- by training@kpipartners.com at 3/17/2022, 8:27:08 AM on [default]basic-starter-cluster

select * from temp_table;

```
1  select * from temp_table;
2
```

▸ (1) Spark Jobs

Table   Data Profile

	_c0	_c1	_c2	_c3	_c4	_c5	_c6	_c7	_c8	_c9
1	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
2	AFG	Asia	Afghanistan	2/24/2020	5	5	0	0	0	0
3	AFG	Asia	Afghanistan	2/25/2020	5	0	0	0	0	0
4	AFG	Asia	Afghanistan	2/26/2020	5	0	0	0	0	0
5	AFG	Asia	Afghanistan	2/27/2020	5	0	0	0	0	0
6	AFG	Asia	Afghanistan	2/28/2020	5	0	0	0	0	0
7	AFG	Asia	Afghanistan	2/29/2020	5	0	0.714	0	0	0

Truncated results, showing first 1000 rows.
Click to re-execute with maximum result limits.

SELECT _c1,_c31,max(_c47) from temp_table group by _c1,_c31,_c47 order by _c47 desc limit 2;

anif_covid1_sql `SQL`

[default]basic-start...

Cmd 4

```
1  SELECT _c1,_c31,max(_c47) from temp_table group by _c1,_c31,_c47  order by _c47 desc limit 2;
```

▸ (2) Spark Jobs

Table   Data Profile

	_c1	_c31	max(_c47)
1	continent	positive_rate	stringency_index
2	South America	0.088	98.15

Showing all 2 rows.

Command took 3.06 seconds -- by training@kpipartners.com at 3/17/2022, 8:27:08 AM on [default]basic-starter-cluster

Cmd 5

```
1  SELECT _c1, _c2,max(_c7) as max_deaths,max(_c58) as male_smokers from temp_table group by _c1, _c2 order by max_deaths desc limit 2;
```

SELECT _c1,_c2,max(_c7) as max_deaths,max(_c58) as male_smokers from temp_table group by _c1,_c2 order by max_deaths desc limit 2;

anif_covid1_sql `SQL`

[default]basic-start...

Cmd 5

```
1  SELECT _c1,_c2,max(_c7) as max_deaths,max(_c58) as male_smokers from temp_table group by _c1,_c2 order by max_deaths desc limit 2;
```

▸ (2) Spark Jobs

Table   Data Profile

	_c1	_c2	max_deaths	male_smokers
1	continent	location	total_deaths	male_smokers
2	Europe	Austria	9997	30.9

Showing all 2 rows.

Command took 1.86 seconds -- by training@kpipartners.com at 3/17/2022, 8:27:08 AM on [default]basic-starter-cluster

Cmd 6