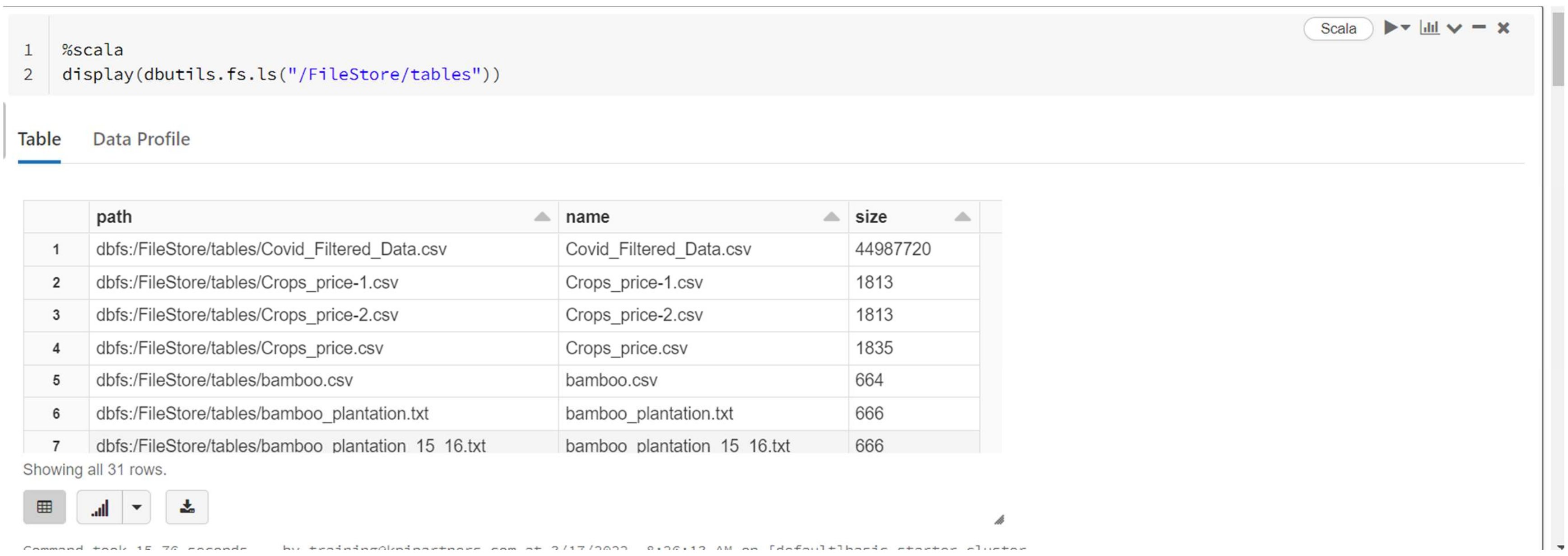
COVID DATA ON DATABRICKS – SCALA SPARK

LISTING OUT THE FILES

%scala

display(dbutils.fs.ls("/FileStore/tables"))



```
%scala
display(dbutils.fs.ls("/FileStore/tables"))
```

	path	name	size
1	dbfs:/FileStore/tables/Covid_Filtered_Data.csv	Covid_Filtered_Data.csv	44987720
2	dbfs:/FileStore/tables/Crops_price-1.csv	Crops_price-1.csv	1813
3	dbfs:/FileStore/tables/Crops_price-2.csv	Crops_price-2.csv	1813
4	dbfs:/FileStore/tables/Crops_price.csv	Crops_price.csv	1835
5	dbfs:/FileStore/tables/bamboo.csv	bamboo.csv	664
6	dbfs:/FileStore/tables/bamboo_plantation.txt	bamboo_plantation.txt	666
7	dbfs:/FileStore/tables/bamboo_plantation_15_16.txt	bamboo_plantation_15_16.txt	666

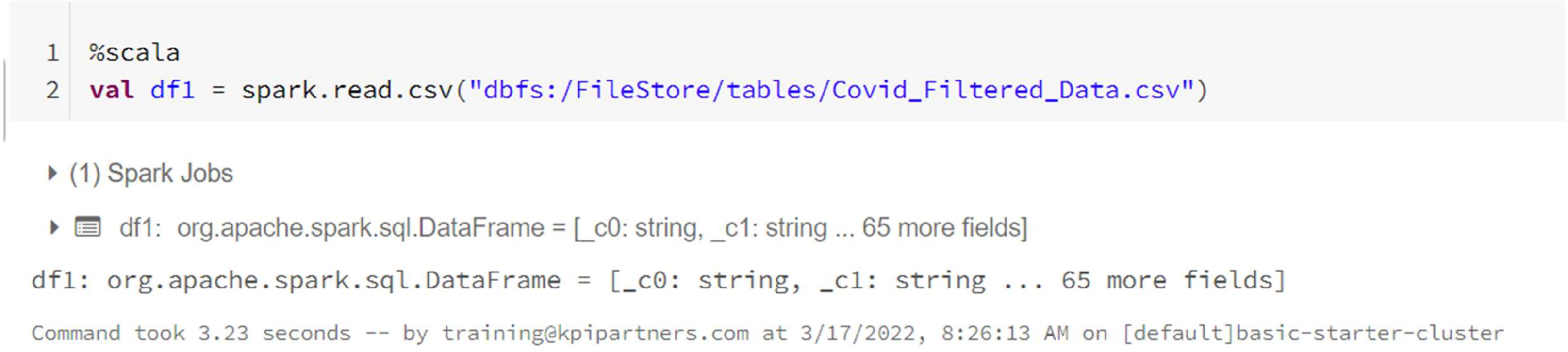
Showing all 31 rows.

```
Command took 15.76 seconds -- by training@kpipartners.com at 3/17/2022, 8:26:13 AM on [default]basic-starter-cluster
```

READING THE CSV FILE

%scala

val df1 = spark.read.csv("dbfs:/FileStore/tables/Covid_Filtered_Data.csv")



```
%scala
val df1 = spark.read.csv("dbfs:/FileStore/tables/Covid_Filtered_Data.csv")
```

▸ (1) Spark Jobs

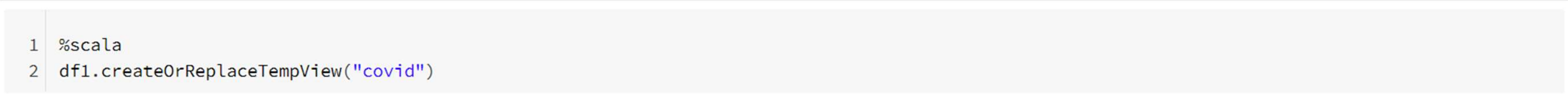
▸ df1: org.apache.spark.sql.DataFrame = [_c0: string, _c1: string ... 65 more fields]

```
df1: org.apache.spark.sql.DataFrame = [_c0: string, _c1: string ... 65 more fields]
Command took 3.23 seconds -- by training@kpipartners.com at 3/17/2022, 8:26:13 AM on [default]basic-starter-cluster
```

CREATING TABLE FOR DATAFRAME

%scala

df1.createOrReplaceTempView("covid")

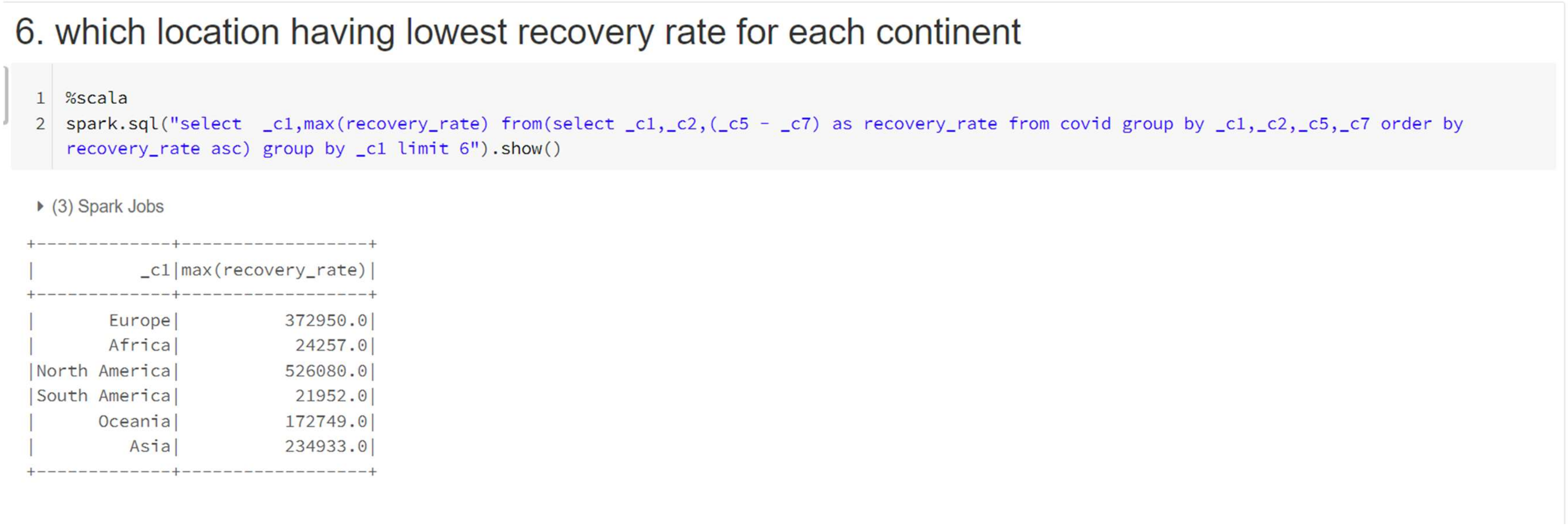


```
%scala
df1.createOrReplaceTempView("covid")
```

SPARK –SQL QUERIES

which location having lowest recovery rate for each continent

```
%scala
spark.sql("select  _c1,max(recovery_rate) from(select _c1,_c2,(_c5 - _c7) as recovery_rate from covid group by _c1,_c2,_c5,_c7 order by recovery_rate asc) group by _c1 limit 6").show()
```



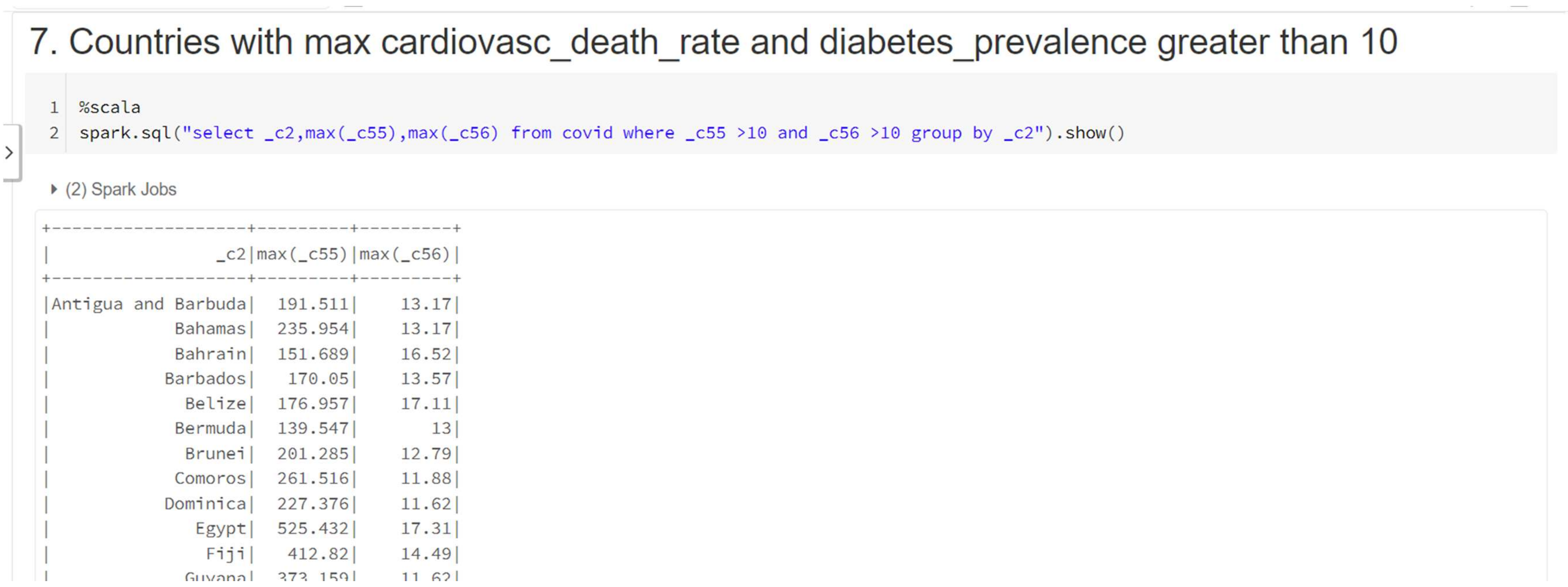
6. which location having lowest recovery rate for each continent

```
%scala
spark.sql("select  _c1,max(recovery_rate) from(select _c1,_c2,(_c5 - _c7) as recovery_rate from covid group by _c1,_c2,_c5,_c7 order by
recovery_rate asc) group by _c1 limit 6").show()
```

```
+-------------+------------------+
|          _c1|max(recovery_rate)|
+-------------+------------------+
|       Europe|          372950.0|
|       Africa|           24257.0|
|North America|          526080.0|
|South America|           21952.0|
|      Oceania|          172749.0|
|         Asia|          234933.0|
+-------------+------------------+
```

Countries with max cardiovasc_death_rate and diabetes_prevalence greater than 10

```
%scala
spark.sql("select _c2,max(_c55),max(_c56) from covid where _c55 >10 and _c56 >10 group by _c2").show()
```



7. Countries with max cardiovasc_death_rate and diabetes_prevalence greater than 10

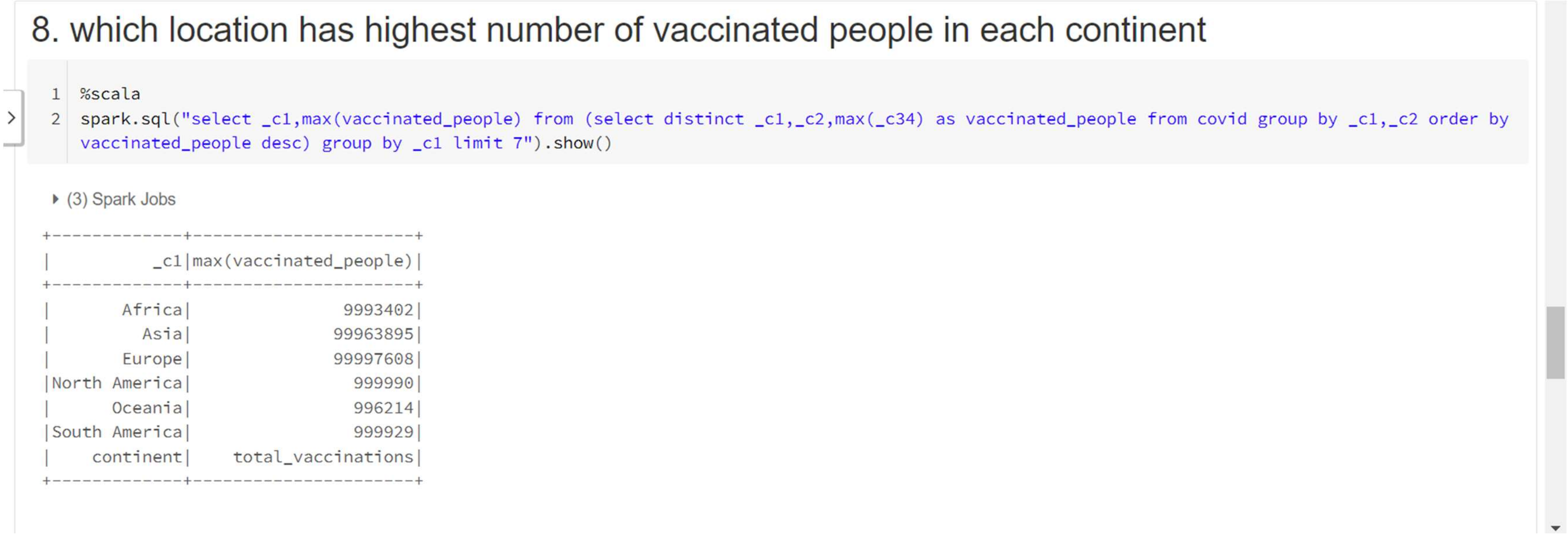
```
%scala
spark.sql("select _c2,max(_c55),max(_c56) from covid where _c55 >10 and _c56 >10 group by _c2").show()
```

```
+-------------------+---------+---------+
|                _c2|max(_c55)|max(_c56)|
+-------------------+---------+---------+
|Antigua and Barbuda|  191.511|    13.17|
|            Bahamas|  235.954|    13.17|
|            Bahrain|  151.689|    16.52|
|           Barbados|   170.05|    13.57|
|             Belize|  176.957|    17.11|
|            Bermuda|  139.547|       13|
|             Brunei|  201.285|    12.79|
|            Comoros|  261.516|    11.88|
|           Dominica|  227.376|    11.62|
|              Egypt|  525.432|    17.31|
|               Fiji|   412.82|    14.49|
|             Guyana|  373.159|    11.62|
```

Which location has highest number of vaccinated people in each continent

```
%scala

spark.sql("select _c1,max(vaccinated_people) from (select distinct _c1,_c2,max(_c34) as vaccinated_people from covid group by _c1,_c2 order by vaccinated_people desc) group by _c1 limit 7").show()
```



8. which location has highest number of vaccinated people in each continent

```
%scala
spark.sql("select _c1,max(vaccinated_people) from (select distinct _c1,_c2,max(_c34) as vaccinated_people from covid group by _c1,_c2 order by vaccinated_people desc) group by _c1 limit 7").show()
```

```
+-------------+---------------------+
|          _c1|max(vaccinated_people)|
+-------------+---------------------+
|       Africa|              9993402|
|         Asia|             99963895|
|       Europe|             99997608|
|North America|               999990|
|      Oceania|               996214|
|South America|               999929|
|    continent|     total_vaccinations|
+-------------+---------------------+
```