COVID ASSIGNEMENT BY USING DATABRICKS

Using SQL in Databricks

List out the files

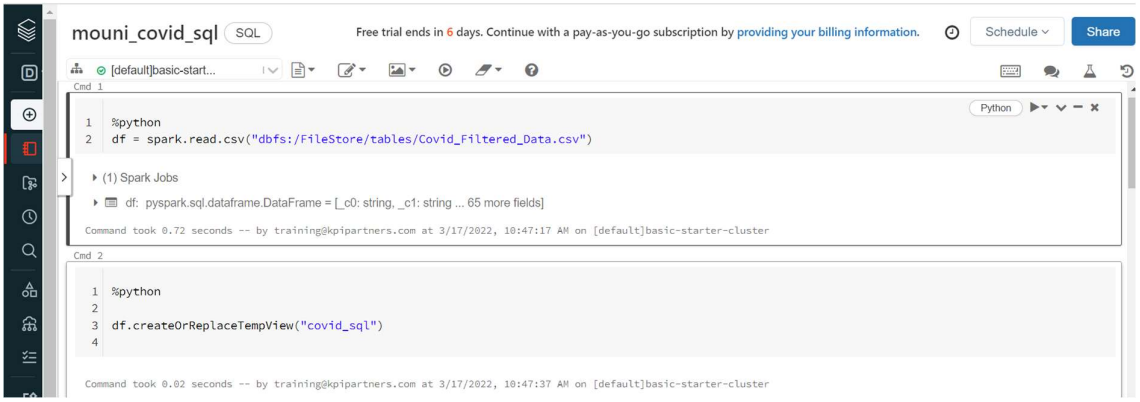
%python

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

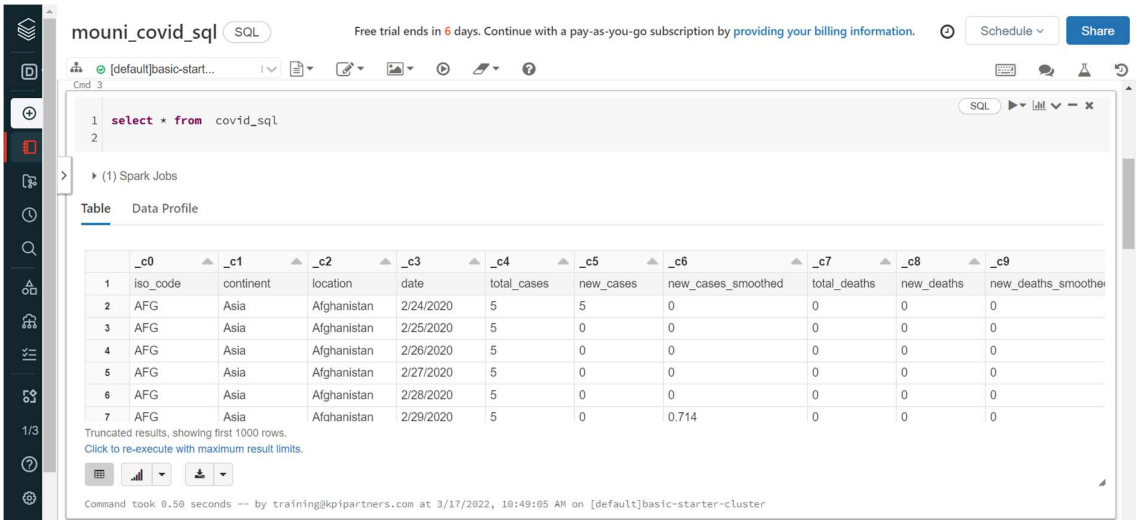
create a tempview

%python

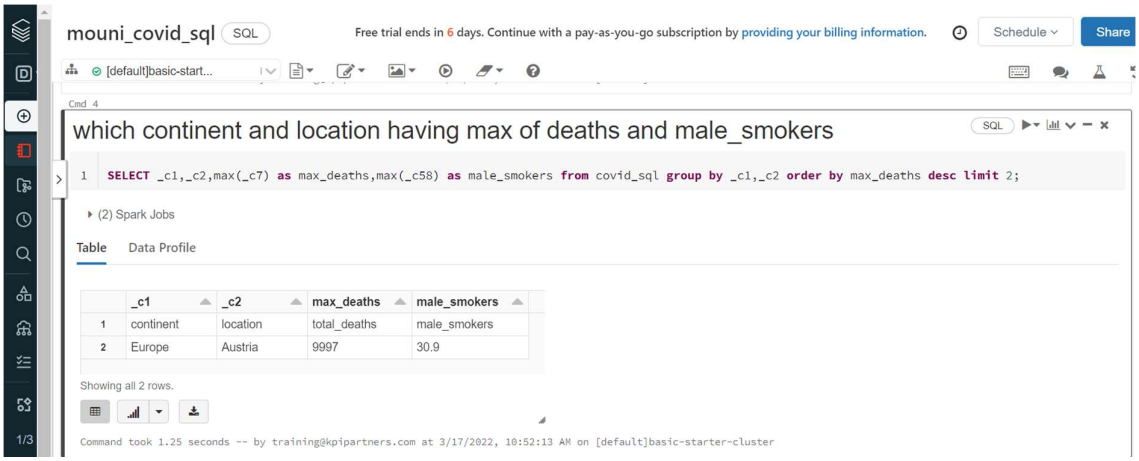
df.createOrReplaceTempView("covid_sql")



Select * from covid_sql



which continent and location having max of deaths and male_smokers



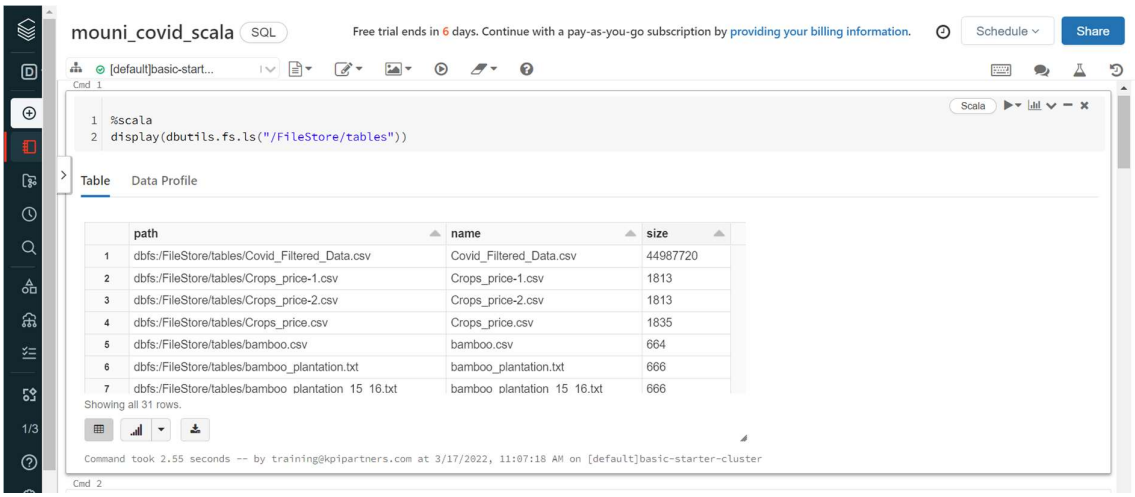
display the continent,positivity rate where the strigency index is maximum

which year having the highest number of deaths

Using scala:

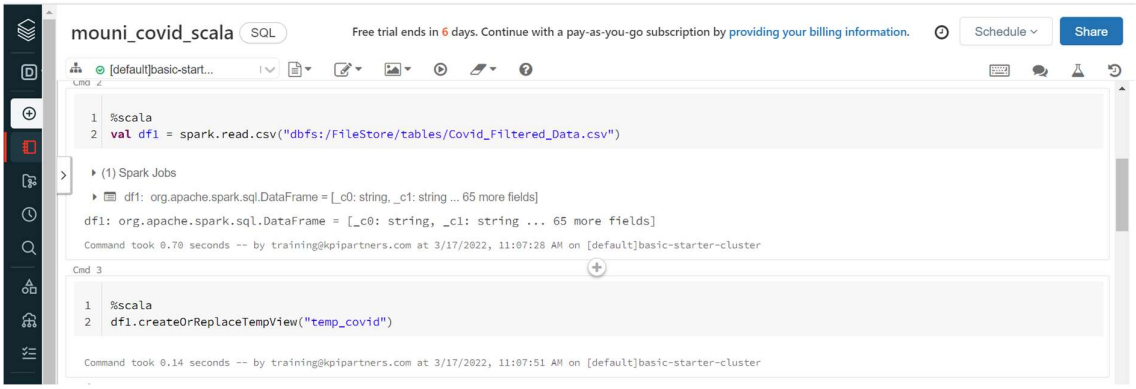
%scala

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



%scala

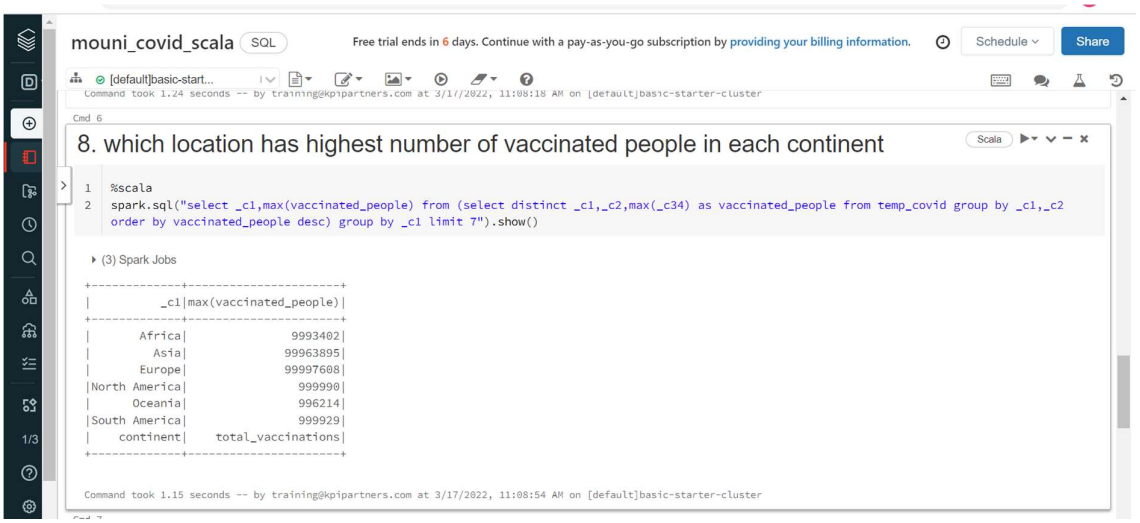
df1.createOrReplaceTempView("temp_covid")



which location having lowest recovery rate for each continent

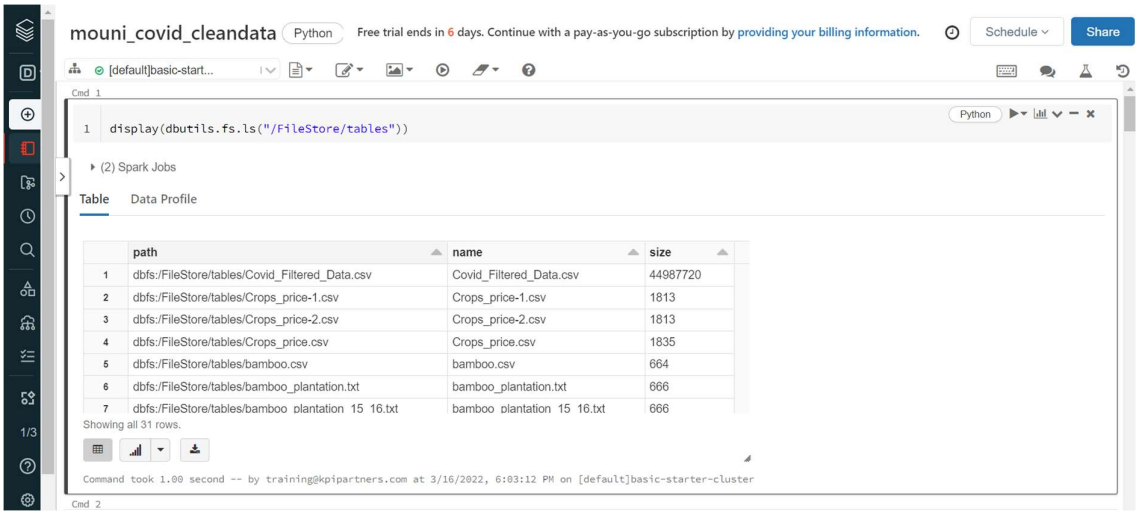
Countries with max cardiovasc_death_rate and diabetes_prevalence greater than 10

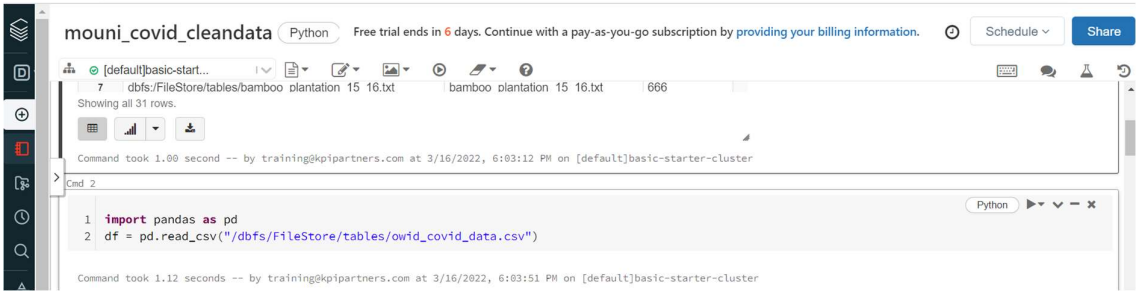
which location has highest number of vaccinated people in each continent



Uising pyspark

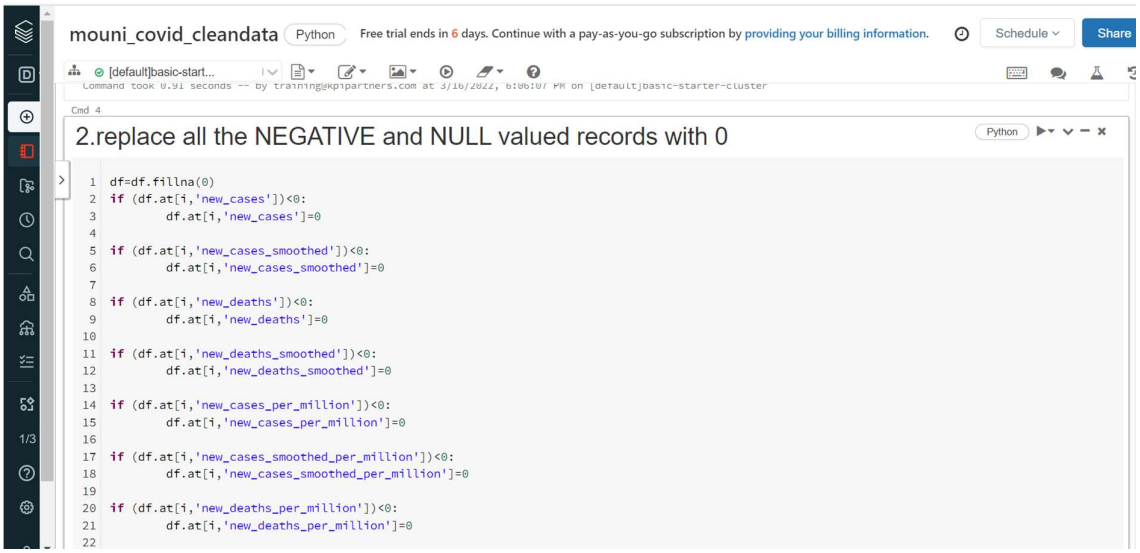
List out the file

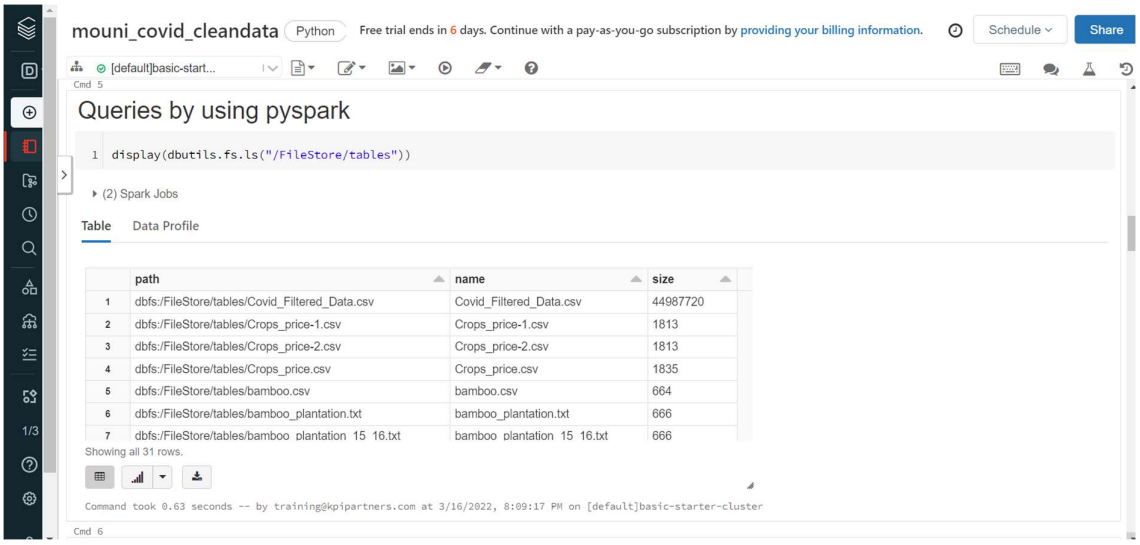


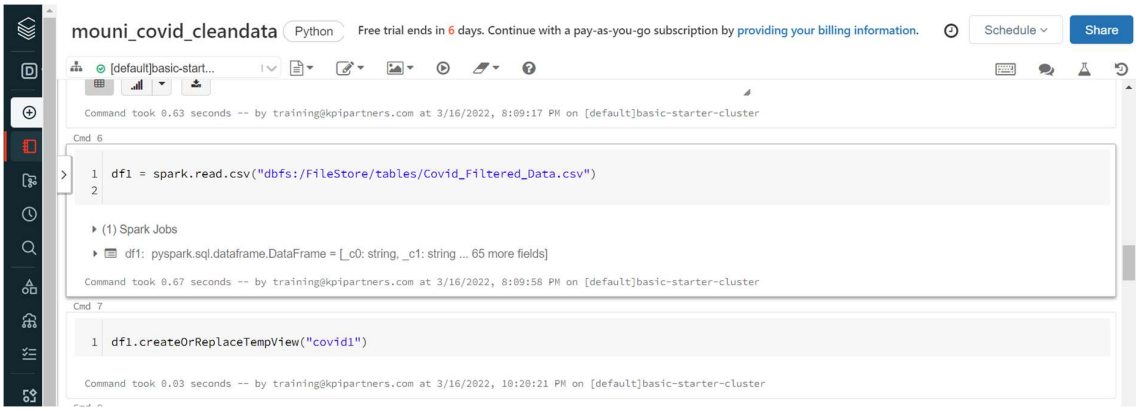


1. remove the records where the continent column having blanks

2.replace all the NEGATIVE and NULL valued records with 0

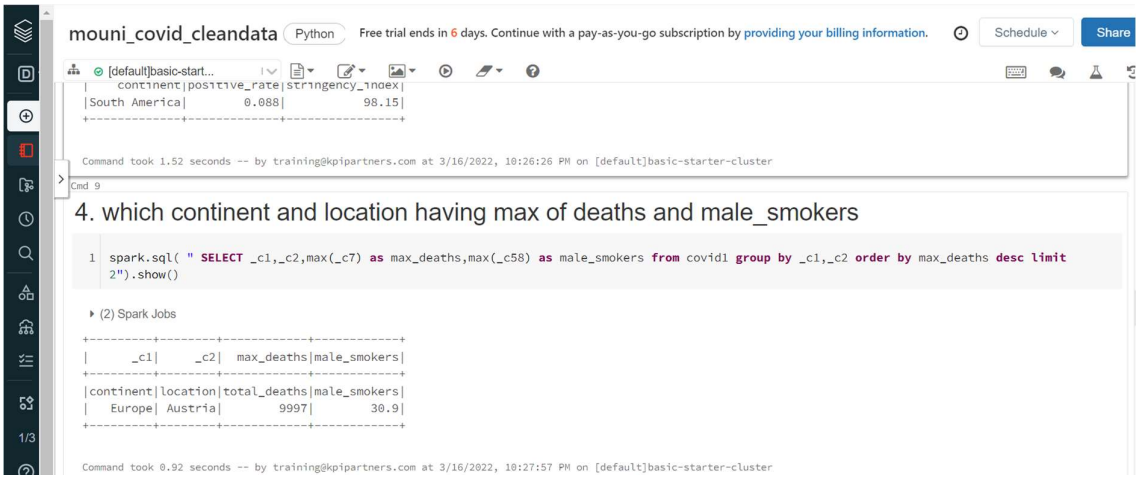






3. display the continent,positivity rate where the strigency index is maximum

4.which continent and location having max of deaths and male_smokers



5. which year having the highest number of deaths