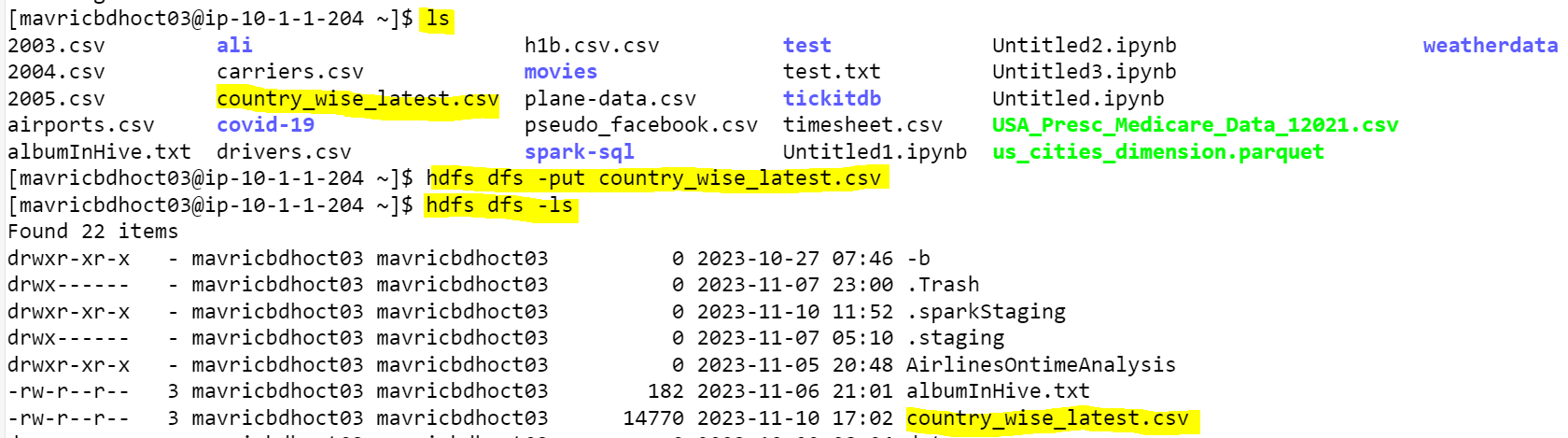
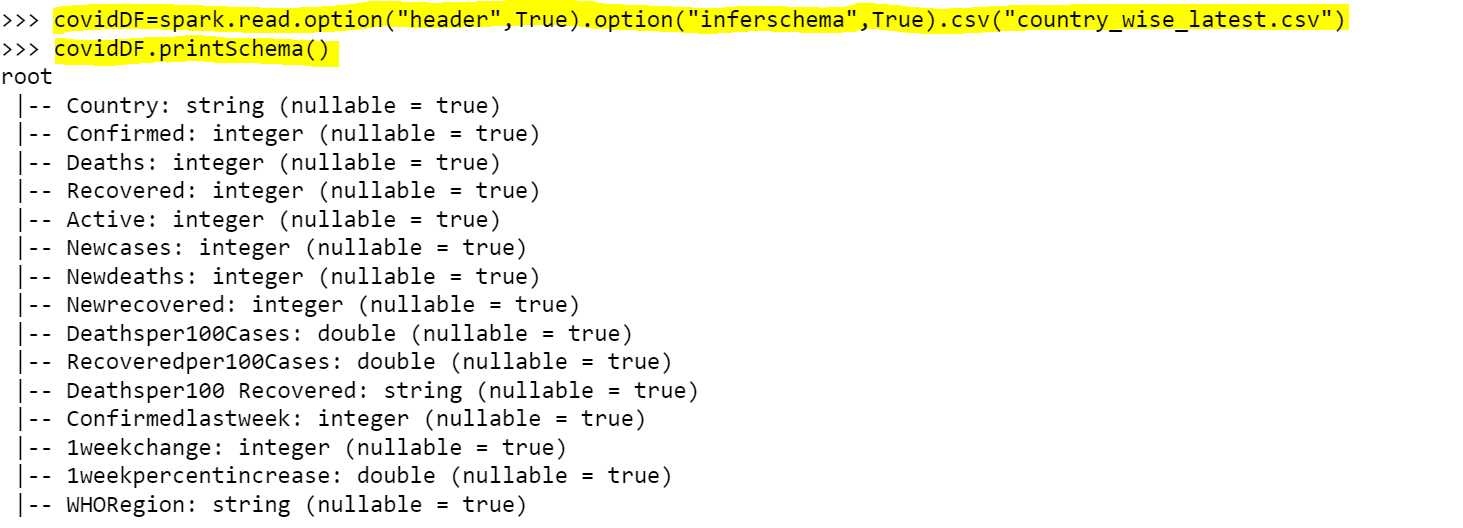
**Pyspark Assignment**

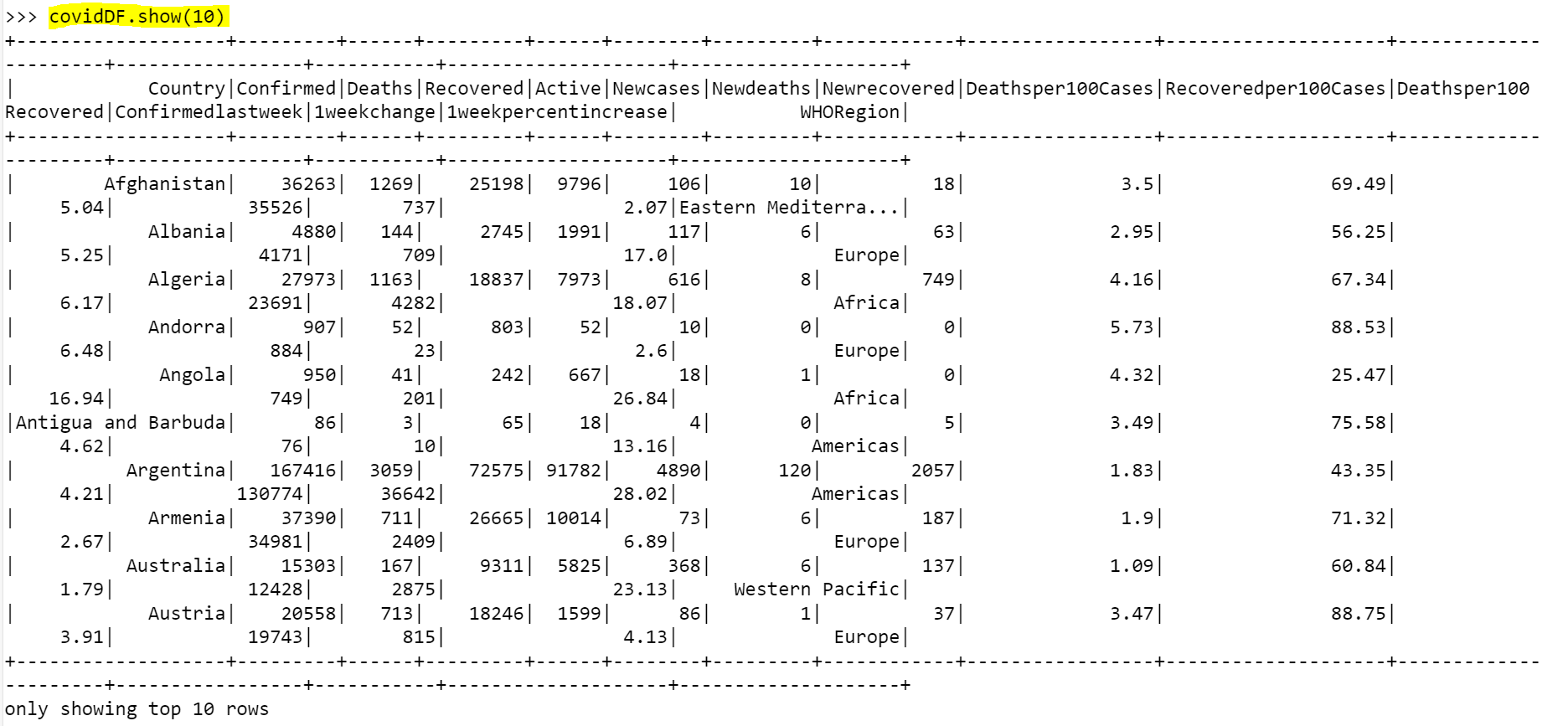
**Covid 19 Data analysis:**

1. **Load the covid 19 data into a spark dataframe (country\_wise\_latest.csv) with the correct schema definition.**

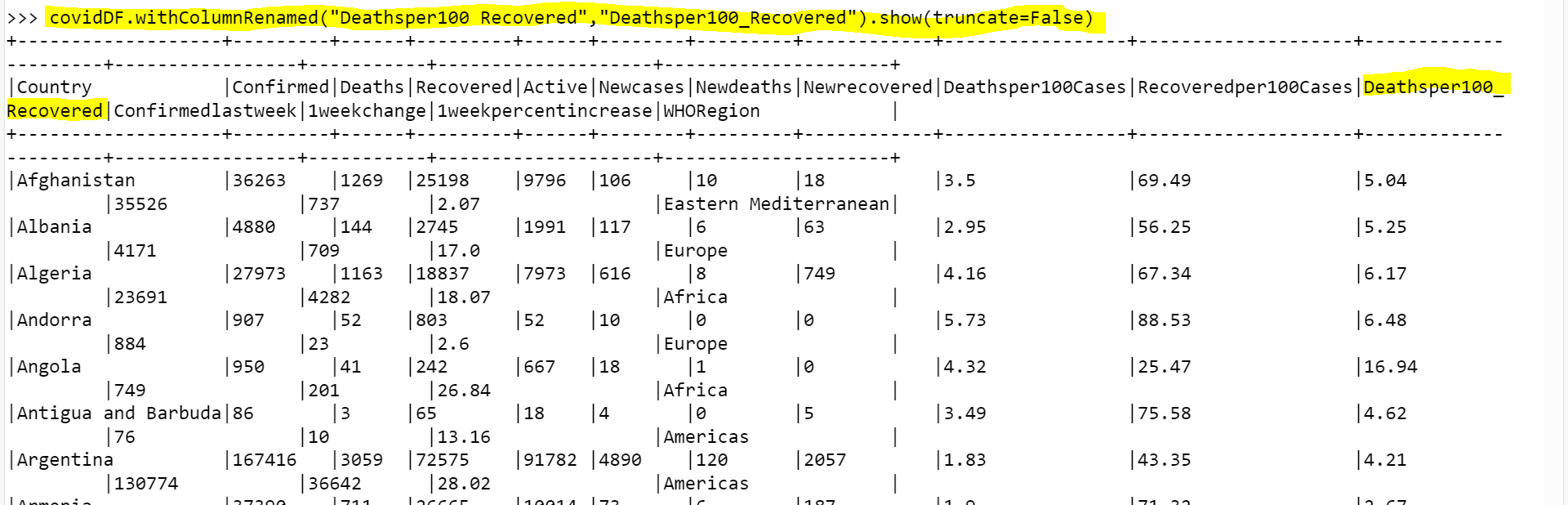


**Pyspark**





1. **The are some column names which are long, contains special characters, spaces etc. Rename all such column names accordingly. Example Country/Region → country New cases → New\_cases. Etc.**

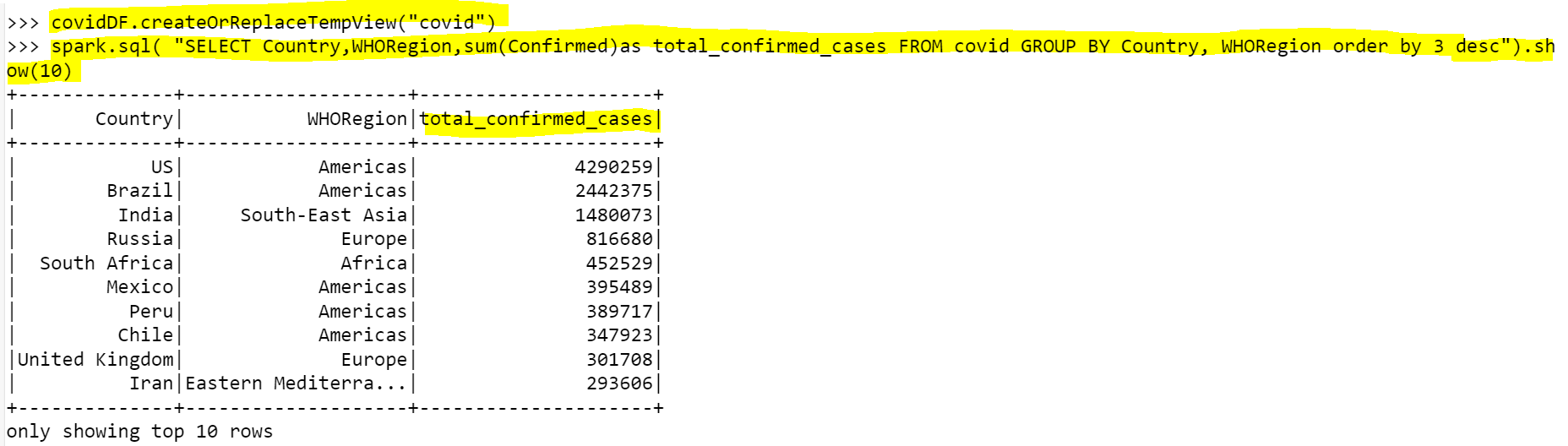


1. **Count and check if there any null values in any of the columns**

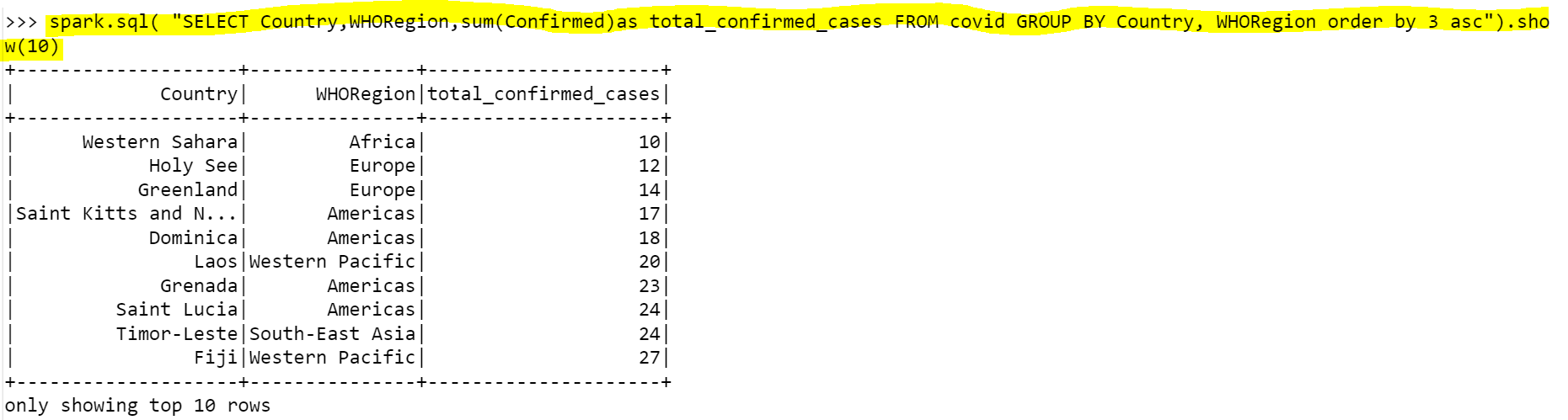
A screenshot of a computer

Description automatically generated

1. What are the top 10 countries under the WHO region with covid 19 confirmed cases.



1. What are the bottom 10 countries under the WHO region with covid 19 Confirmed Cases.



1. What are the total number of countries/ total no. of WHO regions and also list the various WHO regions.

