Task 1 Covid 19 Analysis:

1.Load the covid 19 data in to a spark dataframe

(country\_wise\_latest.csv) with the correct schema definition

A black and white text

Description automatically generated

A white screen with black text

Description automatically generated

A white sheet of paper with black text

Description automatically generated

A number of numbers on a white background

Description automatically generated

2.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

A screenshot of a computer

Description automatically generated

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

A black and white text

Description automatically generated

4. What are the top 10 countries under the WHO region with covid 19 Confirmed cases

spark.sql( "SELECT Country,WHORegion,sum(Confirmed) FROM covid GROUP BY Country, WHORegion order by 3 desc").show(10)

A screenshot of a computer

Description automatically generated

5. What are the bottom 10 countries under the WHO region with covid 19 Confirmed cases

spark.sql( "SELECT Country,WHORegion,sum(Confirmed) FROM covid GROUP BY Country, WHORegion order by 3 asc").show(10)

A screenshot of a computer

Description automatically generated

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

spark.sql("SELECT count(distinct Country), count(distinct WHORegion),WHORegion from covid group by WHORegion").show()

