In [0]:
```
#LISTING OUT THE FILES

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

path	name	size
dbfs:/FileStore/tables/Covid_Filtered_Data.csv	Covid_Filtered_Data.csv	44987720
dbfs:/FileStore/tables/Crops_price-1.csv	Crops_price-1.csv	1813
dbfs:/FileStore/tables/Crops_price-2.csv	Crops_price-2.csv	1813
dbfs:/FileStore/tables/Crops_price.csv	Crops_price.csv	1835
dbfs:/FileStore/tables/bamboo.csv	bamboo.csv	664
dbfs:/FileStore/tables/bamboo_plantation.txt	bamboo_plantation.txt	666
dbfs:/FileStore/tables/bamboo_plantation_15_16.txt	bamboo_plantation_15_16.txt	666
dbfs:/FileStore/tables/clean_covid_filtered.csv	clean_covid_filtered.csv	57098243
dbfs:/FileStore/tables/cleaned123.csv/	cleaned123.csv/	0
dbfs:/FileStore/tables/new_arjun_credentials__1_.csv	new_arjun_credentials__1_.csv	209
dbfs:/FileStore/tables/new_user_credentials-1.csv	new_user_credentials-1.csv	202
dbfs:/FileStore/tables/new_user_credentials-10.csv	new_user_credentials-10.csv	199
dbfs:/FileStore/tables/new_user_credentials-2.csv	new_user_credentials-2.csv	207
dbfs:/FileStore/tables/new_user_credentials-3.csv	new_user_credentials-3.csv	209
dbfs:/FileStore/tables/new_user_credentials-4.csv	new_user_credentials-4.csv	207
dbfs:/FileStore/tables/new_user_credentials-5.csv	new_user_credentials-5.csv	202
dbfs:/FileStore/tables/new_user_credentials-6.csv	new_user_credentials-6.csv	203
dbfs:/FileStore/tables/new_user_credentials-7.csv	new_user_credentials-7.csv	207
dbfs:/FileStore/tables/new_user_credentials-8.csv	new_user_credentials-8.csv	203
dbfs:/FileStore/tables/new_user_credentials-9.csv	new_user_credentials-9.csv	202
dbfs:/FileStore/tables/new_user_credentials.csv	new_user_credentials.csv	201
dbfs:/FileStore/tables/new_user_credentials__1_.csv	new_user_credentials__1_.csv	219
dbfs:/FileStore/tables/new_user_credentials_new.csv	new_user_credentials_new.csv	219
dbfs:/FileStore/tables/occupation.csv	occupation.csv	23609

path	name	size
dbfs:/FileStore/tables/owid_covid_data.csv	owid_covid_data.csv	46387709
dbfs:/FileStore/tables/rubber.csv	rubber.csv	666
dbfs:/FileStore/tables/rubber_plantation.txt	rubber_plantation.txt	666
dbfs:/FileStore/tables/sales_data.csv	sales_data.csv	2129689
dbfs:/FileStore/tables/tea.csv	tea.csv	697
dbfs:/FileStore/tables/tea_plantation.txt	tea_plantation.txt	697
dbfs:/FileStore/tables/vaccinations.csv	vaccinations.csv	507055

In [0]:
```
# READING CSV FILES


import pandas as pd
df = pd.read_csv("/dbfs/FileStore/tables/owid_covid_data.csv")
```

In [0]:
```
#REPLACE ALL THE  NEGATIVE and NULL valued records with 0


df=df.fillna(0)


for i in range(len(df)):
    if (df.at[i,'new_cases'])<0:
        df.at[i,'new_cases']=0

    if (df.at[i,'new_cases_smoothed'])<0:
        df.at[i,'new_cases_smoothed']=0

    if (df.at[i,'new_deaths'])<0:
        df.at[i,'new_deaths']=0

    if (df.at[i,'new_deaths_smoothed'])<0:
        df.at[i,'new_deaths_smoothed']=0

    if (df.at[i,'new_cases_per_million'])<0:
        df.at[i,'new_cases_per_million']=0

    if (df.at[i,'new_cases_smoothed_per_million'])<0:
        df.at[i,'new_cases_smoothed_per_million']=0

    if (df.at[i,'new_deaths_per_million'])<0:
        df.at[i,'new_deaths_per_million']=0

    if (df.at[i,'new_deaths_smoothed_per_million'])<0:
        df.at[i,'new_deaths_smoothed_per_million']=0

    if (df.at[i,'reproduction_rate'])<0:
        df.at[i,'reproduction_rate']=0

    if (df.at[i,'excess_mortality_cumulative_per_million'])<0:
        df.at[i,'excess_mortality_cumulative_per_million']=0

    if (df.at[i,'excess_mortality'])<0:
        df.at[i,'excess_mortality']=0

    if (df.at[i,'excess_mortality_cumulative'])<0:
        df.at[i,'excess_mortality_cumulative']=0
```

```
    if (df.at[i,'excess_mortality_cumulative_absolute'])<0:
        df.at[i,'excess_mortality_cumulative_absolute']=0

    if (df.at[i,'excess_mortality_cumulative'])<0:
        df.at[i,'excess_mortality_cumulative']=0
```

In [0]:
```
# DROPING THE ROWN WHERE CONTINENT HAVING BLANKS OR ZERO'S


for i in range(len(df)):
    if (df.at[i,'continent'])==0 :     # (pd.isnull(df.at[i,'continent']) or
        df.drop([i], axis=0, inplace=True)
```

In [0]:
```
# import findspark

# findspark.init()

# import pyspark
# from pyspark.sql import SparkSession
# import pandas as pd

# # Create a spark session
# spark = SparkSession.builder.getOrCreate()

# spark_df = spark.createDataFrame(df)
```

In [0]:
```
# CHECKING


print(df[df["new_deaths"] < 0])
```

Empty DataFrame Columns: [iso_code, continent, location, date, total_cases, new_cases, new_cases_smoothed, total_deaths, new_deaths, new_deaths_smoothed, total_cases_per_million, new_cases_per_million, new_cases_smoothed_per_million, total_deaths_per_million, new_deaths_per_million, new_deaths_smoothed_per_million, reproduction_rate, icu_patients, icu_patients_per_million, hosp_patients, hosp_patients_per_million, weekly_icu_admissions, weekly_icu_admissions_per_million, weekly_hosp_admissions, weekly_hosp_admissions_per_million, new_tests, total_tests, total_tests_per_thousand, new_tests_per_thousand, new_tests_smoothed, new_tests_smoothed_per_thousand, positive_rate, tests_per_case, tests_units, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, new_vaccinations, new_vaccinations_smoothed, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, new_vaccinations_smoothed_per_million, new_people_vaccinated_smoothed, new_people_vaccinated_smoothed_per_hundred, stringency_index, population, population_density, median_age, aged_65_older, aged_70_older, gdp_per_capita, extreme_poverty, cardiovasc_death_rate, diabetes_prevalence, female_smokers, male_smokers, handwashing_facilities, hospital_beds_per_thousand, life_expectancy, human_development_index, excess_mortality_cumulative_absolute, excess_mortality_cumulative, excess_mortality, excess_mortality_cumulative_per_million] Index: []

In [0]:
```
#df.to_csv('dbfs/FileStore/tables/cleaned123.csv', index=False)

print(df[df["new_deaths"] > 0])
```

iso_code ... excess_mortality_cumulative_per_million 28 AFG ... 0.0 31 AFG ... 0.0 34 AFG ... 0.0 39 AFG ... 0.0 41 AFG ... 0.0 ... ... 163081 ZWE ... 0.0 163086 ZWE ... 0.0 163087 ZWE ... 0.0 163088 ZWE ... 0.0 163089 ZWE ... 0.0 [74907 rows x 67 columns]

In [0]:
```
display(dbutils.fs.ls("/FileStore/tables"))
```

path	name	size
dbfs:/FileStore/tables/Covid_Filtered_Data.csv	Covid_Filtered_Data.csv	44987720
dbfs:/FileStore/tables/Crops_price-1.csv	Crops_price-1.csv	1813
dbfs:/FileStore/tables/Crops_price-2.csv	Crops_price-2.csv	1813
dbfs:/FileStore/tables/Crops_price.csv	Crops_price.csv	1835
dbfs:/FileStore/tables/bamboo.csv	bamboo.csv	664
dbfs:/FileStore/tables/bamboo_plantation.txt	bamboo_plantation.txt	666
dbfs:/FileStore/tables/bamboo_plantation_15_16.txt	bamboo_plantation_15_16.txt	666
dbfs:/FileStore/tables/clean_covid_filtered.csv	clean_covid_filtered.csv	57098243
dbfs:/FileStore/tables/cleaned123.csv/	cleaned123.csv/	0
dbfs:/FileStore/tables/new_arjun_credentials__1_.csv	new_arjun_credentials__1_.csv	209
dbfs:/FileStore/tables/new_user_credentials-1.csv	new_user_credentials-1.csv	202
dbfs:/FileStore/tables/new_user_credentials-10.csv	new_user_credentials-10.csv	199
dbfs:/FileStore/tables/new_user_credentials-2.csv	new_user_credentials-2.csv	207
dbfs:/FileStore/tables/new_user_credentials-3.csv	new_user_credentials-3.csv	209
dbfs:/FileStore/tables/new_user_credentials-4.csv	new_user_credentials-4.csv	207
dbfs:/FileStore/tables/new_user_credentials-5.csv	new_user_credentials-5.csv	202
dbfs:/FileStore/tables/new_user_credentials-6.csv	new_user_credentials-6.csv	203
dbfs:/FileStore/tables/new_user_credentials-7.csv	new_user_credentials-7.csv	207
dbfs:/FileStore/tables/new_user_credentials-8.csv	new_user_credentials-8.csv	203
dbfs:/FileStore/tables/new_user_credentials-9.csv	new_user_credentials-9.csv	202
dbfs:/FileStore/tables/new_user_credentials.csv	new_user_credentials.csv	201
dbfs:/FileStore/tables/new_user_credentials__1_.csv	new_user_credentials__1_.csv	219
dbfs:/FileStore/tables/new_user_credentials_new.csv	new_user_credentials_new.csv	219
dbfs:/FileStore/tables/occupation.csv	occupation.csv	23609
dbfs:/FileStore/tables/owid_covid_data.csv	owid_covid_data.csv	46387709

path	name	size
dbfs:/FileStore/tables/rubber.csv	rubber.csv	666
dbfs:/FileStore/tables/rubber_plantation.txt	rubber_plantation.txt	666
dbfs:/FileStore/tables/sales_data.csv	sales_data.csv	2129689
dbfs:/FileStore/tables/tea.csv	tea.csv	697
dbfs:/FileStore/tables/tea_plantation.txt	tea_plantation.txt	697
dbfs:/FileStore/tables/vaccinations.csv	vaccinations.csv	507055

In [0]:
```
df1 = spark.read.csv("dbfs:/FileStore/tables/Covid_Filtered_Data.csv")
```

In [0]:
```
# CREATING TABLE FOR DATAFRAME


df1.createOrReplaceTempView("covid")
```

In [0]:
```
#3.display the continent,positivity rate where the strigency index is maximum

spark.sql( " SELECT _c1,_c31,max(_c47) from covid group by _c1,_c31,_c47  order by _c47 desc limit 2" ).show()
```

+-------------+-------------+---------------+ _c1| _c31| max(_c47)| +-------------+-------------+---------------+
continent|positive_rate|stringency_index| South America| 0.088| 98.15| +-------------+-------------+---------------+

In [0]:
```
# 4.which continent and location having max of deaths and male_smokers

spark.sql( " SELECT _c1,_c2,max(_c7) as max_deaths,max(_c58) as male_smokers from covid group by _c1,_c2 order b
```

+---------+--------+------------+------------+ _c1| _c2| max_deaths|male_smokers| +---------+--------+------------+------------+
continent|location|total_deaths|male_smokers| Europe| Austria| 9997| 30.9| +---------+--------+------------+------------+

In [0]:
```
#  5.which year having the highest number of deaths

spark.sql("select _c3,sum(_c7) from covid group by _c3,_c7 order by _c7 desc limit 2 ").show()
```

+---------+----------------------+ _c3|sum(CAST(_c7 AS DOUBLE))| +---------+----------------------+ date| null| 4/21/2021| 9997.0| +---------+----------------------+

In [0]:
```

```