

```
In [ ]: # List files in DBFS
```

```
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
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 [ ]: # Reading the required file
```

```
In [0]: import pandas as pd  
df = pd.read_csv("/dbfs/FileStore/tables/owid_covid_data.csv")
```

```
In [ ]: # 1. Remove the records where the continent has blanks
```

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

```
In [ ]: # 2. Replace all the NEGATIVE and NULL valued records with 0
```

```
In [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]: 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
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]: df1.createOrReplaceTempView("covid")
```

```
In [ ]: # 3.Display the continent,positivity rate where the stringency index is maximum
```

```
In [0]: spark.sql( " SELECT _c1,_c31,max(_c47) from covid group by _c1,_c31,_c47 order by max(_c47) desc limit 1")
```

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

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

```
In [0]: spark.sql( " SELECT _c1,_c2,max(_c7) as max_deaths,max(_c58) as male_smokers from covid group by _c1,_c2 order by max(_c7) desc limit 1")
```

_c1	_c2	max_deaths	male_smokers
continent	location	total_deaths	male_smokers
Europe	Austria	9997	30.9

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

```
In [0]: spark.sql("select _c3,sum(_c7) from covid group by _c3,_c7 order by _c7 desc limit 1")
```

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