# databricks CASE STUDY

(https://databricks.com)

## FETCHING DATA FROM COVID-19 API

```
import requests
import pandas as pd

# Fetching global COVID-19 data
url = 'https://disease.sh/v3/covid-19/countries'
response = requests.get(url)
data = response.json()

# Convert data to a Pandas DataFrame
covid_df = pd.DataFrame(data)

# Convert to Spark DataFrame for visualizations
spark_df = spark.createDataFrame(covid_df)

#temporary view to perform SQL queries
spark_df.createOrReplaceTempView('covid_data')

#to create visualizations
display(spark_df)

#CONCLUSIONS

#CHART 1: POPULATION AND CASES PER CONTINENT
#It gives a brief overview of the affected out of the total population

#CHART 2: CASES AND RECOVERED (PER MILLION) PER CONTINENT
#It shows that recovery rate of Europe and North America is lower than others, this is due to lack of hospital beds and more smokers in said count
PMC8157209/ (reference article)

#CHART 3: CASES, TESTS, POPULATION PER CONTINENT
#It shows that Europe and North America did more tests and probably that's why reported higher cases, but Asia and Africa despite having greater p
reported lower cases. this shows lack of testing capacity due to them having more developing countries and lesser facilities. https://www.ncbi.nlm
```

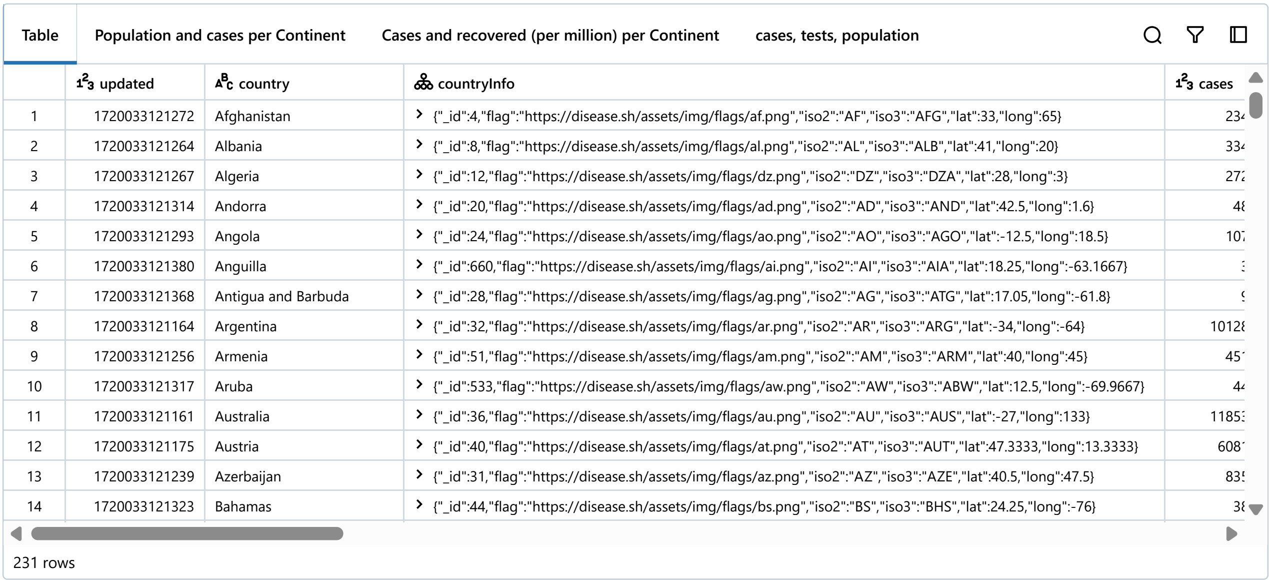

Table
Population and cases per Continent
Cases and recovered (per million) per Continent
cases, tests, population

| | updated | country | countryInfo | cases |
|---|---|---|---|---|
| 1 | 1720033121272 | Afghanistan | {"_id":4,"flag":"https://disease.sh/assets/img/flags/af.png","iso2":"AF","iso3":"AFG","lat":33,"long":65} | 234 |
| 2 | 1720033121264 | Albania | {"_id":8,"flag":"https://disease.sh/assets/img/flags/al.png","iso2":"AL","iso3":"ALB","lat":41,"long":20} | 334 |
| 3 | 1720033121267 | Algeria | {"_id":12,"flag":"https://disease.sh/assets/img/flags/dz.png","iso2":"DZ","iso3":"DZA","lat":28,"long":3} | 272 |
| 4 | 1720033121314 | Andorra | {"_id":20,"flag":"https://disease.sh/assets/img/flags/ad.png","iso2":"AD","iso3":"AND","lat":42.5,"long":1.6} | 48 |
| 5 | 1720033121293 | Angola | {"_id":24,"flag":"https://disease.sh/assets/img/flags/ao.png","iso2":"AO","iso3":"AGO","lat":-12.5,"long":18.5} | 107 |
| 6 | 1720033121380 | Anguilla | {"_id":660,"flag":"https://disease.sh/assets/img/flags/ai.png","iso2":"AI","iso3":"AIA","lat":18.25,"long":-63.1667} | 3 |
| 7 | 1720033121368 | Antigua and Barbuda | {"_id":28,"flag":"https://disease.sh/assets/img/flags/ag.png","iso2":"AG","iso3":"ATG","lat":17.05,"long":-61.8} | 9 |
| 8 | 1720033121164 | Argentina | {"_id":32,"flag":"https://disease.sh/assets/img/flags/ar.png","iso2":"AR","iso3":"ARG","lat":-34,"long":-64} | 10128 |
| 9 | 1720033121256 | Armenia | {"_id":51,"flag":"https://disease.sh/assets/img/flags/am.png","iso2":"AM","iso3":"ARM","lat":40,"long":45} | 451 |
| 10 | 1720033121317 | Aruba | {"_id":533,"flag":"https://disease.sh/assets/img/flags/aw.png","iso2":"AW","iso3":"ABW","lat":12.5,"long":-69.9667} | 44 |
| 11 | 1720033121161 | Australia | {"_id":36,"flag":"https://disease.sh/assets/img/flags/au.png","iso2":"AU","iso3":"AUS","lat":-27,"long":133} | 11853 |
| 12 | 1720033121175 | Austria | {"_id":40,"flag":"https://disease.sh/assets/img/flags/at.png","iso2":"AT","iso3":"AUT","lat":47.3333,"long":13.3333} | 6081 |
| 13 | 1720033121239 | Azerbaijan | {"_id":31,"flag":"https://disease.sh/assets/img/flags/az.png","iso2":"AZ","iso3":"AZE","lat":40.5,"long":47.5} | 835 |
| 14 | 1720033121323 | Bahamas | {"_id":44,"flag":"https://disease.sh/assets/img/flags/bs.png","iso2":"BS","iso3":"BHS","lat":24.25,"long":-76} | 38 |

231 rows

## RETRIEVING TOP 10 COUNTIRES WITH HIGHEST COVID CASES

```
# SQL query to get top 10 countries with highest COVID cases
query = """
SELECT country, cases
FROM covid_data
ORDER BY cases DESC
LIMIT 10
"""

# Executing the query and display the result for visualization
top_10_countries = spark.sql(query)
display(top_10_countries)
```

Table | Top 10 countires with highest cases

| | country | cases |
|---|---|---|
| 1 | USA | 111820082 |
| 2 | India | 45035393 |
| 3 | France | 40138560 |
| 4 | Germany | 38828995 |
| 5 | Brazil | 38743918 |
| 6 | S. Korea | 34571873 |
| 7 | Japan | 33803572 |
| 8 | Italy | 26723249 |
| 9 | UK | 24910387 |
| 10 | Russia | 24124215 |

10 rows

ADDING DROPDWON WIDGET FOR SELECTING COUNTIRES AND CREATING DATETIME DATA

Table | TOTAL CASES TIMELINE

| | date | cases |
|---|---|---|
| 1 | 2020-01-22T00:00:00.000+00:... | 0 |
| 2 | 2020-01-23T00:00:00.000+00:... | 0 |
| 3 | 2020-01-24T00:00:00.000+00:... | 0 |
| 4 | 2020-01-25T00:00:00.000+00:... | 0 |
| 5 | 2020-01-26T00:00:00.000+00:... | 0 |
| 6 | 2020-01-27T00:00:00.000+00:... | 0 |
| 7 | 2020-01-28T00:00:00.000+00:... | 0 |
| 8 | 2020-01-29T00:00:00.000+00:... | 0 |
| 9 | 2020-01-30T00:00:00.000+00:... | 1 |
| 10 | 2020-01-31T00:00:00.000+00:... | 1 |
| 11 | 2020-02-01T00:00:00.000+00:... | 1 |
| 12 | 2020-02-02T00:00:00.000+00:... | 2 |
| 13 | 2020-02-03T00:00:00.000+00:... | 3 |
| 14 | 2020-02-04T00:00:00.000+00:... | 3 |
| 15 | 2020-02-05T00:00:00.000+00:... | 3 |

1,143 rows