Install Required Library

```
pip install snowflake-snowpark-python
```

```
Collecting snowflake-snowpark-python
  Downloading snowflake_snowpark_python-1.14.0-py3-none-any.whl (419 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 419.7/419.7 kB 4.1 MB/s eta 0:00:00
Requirement already satisfied: setuptools>=40.6.0 in /usr/local/lib/python3.10/dist-packages (from snowflake-snowpark-python) (67.7
Requirement already satisfied: wheel in /usr/local/lib/python3.10/dist-packages (from snowflake-snowpark-python) (0.43.0)
Collecting snowflake-connector-python<4.0.0,>=3.6.0 (from snowflake-snowpark-python)
  Downloading snowflake_connector_python-3.9.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.6 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.6/2.6 MB 11.0 MB/s eta 0:00:00
Requirement already satisfied: typing-extensions<5.0.0,>=4.1.0 in /usr/local/lib/python3.10/dist-packages (from snowflake-snowpark-p
Requirement already satisfied: pyyaml in /usr/local/lib/python3.10/dist-packages (from snowflake-snowpark-python) (6.0.1)
Requirement already satisfied: cloudpickle!=2.1.0,!=2.2.0,<=2.2.1,>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from snowflake
Collecting asn1crypto<2.0.0,>0.24.0 (from snowflake-connector-python<4.0.0,>=3.6.0->snowflake-snowpark-python)
  Downloading asn1crypto-1.5.1-py2.py3-none-any.whl (105 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 105.0/105.0 kB 9.6 MB/s eta 0:00:00
Requirement already satisfied: cffi<2.0.0,>=1.9 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<4.0.0,>=
Requirement already satisfied: cryptography<43.0.0,>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-pyth
Requirement already satisfied: pyOpenSSL<25.0.0,>=16.2.0 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python
Requirement already satisfied: pyjwt<3.0.0 in /usr/lib/python3/dist-packages (from snowflake-connector-python<4.0.0,>=3.6.0->snowfla
Requirement already satisfied: pytz in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<4.0.0,>=3.6.0->snowfl
Requirement already satisfied: requests<3.0.0 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<4.0.0,>=3
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<4.0.0,>=3.6.0->
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<4.0.0,>=3.6
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<4.0.0,
Requirement already satisfied: filelock<4,>=3.5 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<4.0.0,>=
Requirement already satisfied: sortedcontainers>=2.4.0 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-python<4
Requirement already satisfied: platformdirs<5.0.0,>=2.6.0 in /usr/local/lib/python3.10/dist-packages (from snowflake-connector-pytho
Collecting tomlkit (from snowflake-connector-python<4.0.0,>=3.6.0->snowflake-snowpark-python)
  Downloading tomlkit-0.12.4-py3-none-any.whl (37 kB)
Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-packages (from cffi<2.0.0,>=1.9->snowflake-connector-pyth
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3.0.0->snowflake-connect
Installing collected packages: asn1crypto, tomlkit, snowflake-connector-python, snowflake-snowpark-python
Successfully installed asn1crypto-1.5.1 snowflake-connector-python-3.9.0 snowflake-snowpark-python-1.14.0 tomlkit-0.12.4
```

Create Snowpark Session

```
from snowflake.snowpark.session import Session

username = 'MAGICDASH91'
password = '*****************'
account = 'tk11073.europe-west4.gcp'
warehouse = 'COMPUTE_WH'
database = 'COVID19_EPIDEMIOLOGICAL_DATA'
schema = 'PUBLIC'

def snowpark_session_create():
  connection_params = {
      "user": username,
      "password": password,
      "account": account,
      "warehouse": warehouse,
      "database": database,
      "schema": schema
  }

  # Create the session
  session = Session.builder.configs(connection_params).create()
  return session

demo_session = snowpark_session_create()
```

Start Build Data Pipelines and Data Transformation

```
df = demo_session.sql('SELECT * FROM OWID_VACCINATIONS')
df.show()
```

```
-----------------------------------------------------------------------------------------------------------------------------------
CINATIONS_PER_MILLION"  |"VACCINES"                                          |"LAST_OBSERVATION_DATE"  |"SOURCE_NAME"              |"
-----------------------------------------------------------------------------------------------------------------------------------
                        |CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26               |World Health Organization  |h
                        |CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26               |World Health Organization  |h
```

```
|CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26             |World Health Organization  |
|CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26             |World Health Organization  |
|CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26             |World Health Organization  |
|CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26             |World Health Organization  |
|CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26             |World Health Organization  |
|CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26             |World Health Organization  |
|CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26             |World Health Organization  |
|CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |2023-11-26             |World Health Organization  |
--------------------------------------------------------------------------------------------------------------------
```

```
# Check the datatypes
df.dtypes
```

```
[('DATE', 'date'),
 ('COUNTRY_REGION', 'string(100)'),
 ('ISO3166_1', 'string(2)'),
 ('TOTAL_VACCINATIONS', 'bigint'),
 ('PEOPLE_VACCINATED', 'bigint'),
 ('PEOPLE_FULLY_VACCINATED', 'bigint'),
 ('DAILY_VACCINATIONS_RAW', 'bigint'),
 ('DAILY_VACCINATIONS', 'bigint'),
 ('TOTAL_VACCINATIONS_PER_HUNDRED', 'double'),
 ('PEOPLE_VACCINATED_PER_HUNDRED', 'double'),
 ('PEOPLE_FULLY_VACCINATED_PER_HUNDRED', 'double'),
 ('DAILY_VACCINATIONS_PER_MILLION', 'double'),
 ('VACCINES', 'string(1000)'),
 ('LAST_OBSERVATION_DATE', 'date'),
 ('SOURCE_NAME', 'string(500)'),
 ('SOURCE_WEBSITE', 'string(2000)'),
 ('LAST_UPDATE_DATE', 'timestamp'),
 ('LAST_REPORTED_FLAG', 'boolean')]
```

```
# Extract year from the DATE column
df = df.withColumn("DATE", df["DATE"].substr(1, 4))
df.show()
```

```
---------------------------------------------------------------------------------------------------------------------------------
|"COUNTRY_REGION"  |"ISO3166_1"  |"TOTAL_VACCINATIONS"  |"PEOPLE_VACCINATED"  |"PEOPLE_FULLY_VACCINATED"  |"DAILY_VACCINATIONS_RAW"
---------------------------------------------------------------------------------------------------------------------------------
|Afghanistan       |AF           |0                     |0                    |NULL                       |NULL
|Afghanistan       |AF           |NULL                  |NULL                 |NULL                       |NULL
|Afghanistan       |AF           |NULL                  |NULL                 |NULL                       |NULL
|Afghanistan       |AF           |NULL                  |NULL                 |NULL                       |NULL
|Afghanistan       |AF           |NULL                  |NULL                 |NULL                       |NULL
|Afghanistan       |AF           |NULL                  |NULL                 |NULL                       |NULL
|Afghanistan       |AF           |8200                  |8200                 |NULL                       |NULL
|Afghanistan       |AF           |NULL                  |NULL                 |NULL                       |NULL
|Afghanistan       |AF           |NULL                  |NULL                 |NULL                       |NULL
|Afghanistan       |AF           |NULL                  |NULL                 |NULL                       |NULL
---------------------------------------------------------------------------------------------------------------------------------
```

```
# Sum of Vaccination by each country
from snowflake.snowpark.functions import col, lit, sum as sum_, max as max_
df.group_by("COUNTRY_REGION").agg(sum_("TOTAL_VACCINATIONS")).show()
```

```
---------------------------------------------------
|"COUNTRY_REGION"     |"SUM(TOTAL_VACCINATIONS)"  |
---------------------------------------------------
|Afghanistan          |1659147079                 |
|Albania              |395364296                  |
|Andorra              |5846260                    |
|Angola               |1398489266                 |
|Anguilla             |1290079                    |
|Antigua and Barbuda  |7482788                    |
|Argentina            |86986241297                |
|Armenia              |58724412                   |
|Aruba                |98025212                   |
|Austria              |1897828020                 |
---------------------------------------------------
```

```
# Sum of Vaccination by each year
from snowflake.snowpark.functions import col, lit, sum as sum_, max as max_
df.group_by("DATE").agg(sum_("TOTAL_VACCINATIONS")).show()
```

```
-------------------------------------
|"DATE"  |"SUM(TOTAL_VACCINATIONS)"  |
-------------------------------------
|2021    |1200757165302              |
|2022    |3307282050895              |
|2023    |1335368675780              |
```

```
|2020    |54476506                  |
--------------------------------------
```

Use Group By to get the Vaccines that used in that country

```
# Define the SQL query with window function and filtering
sql = """
SELECT
  COUNTRY_REGION,
  VACCINES
FROM (
  SELECT
    COUNTRY_REGION,
    VACCINES,
    ROW_NUMBER() OVER (PARTITION BY COUNTRY_REGION ORDER BY LENGTH(VACCINES) DESC) AS rn
  FROM OWID_VACCINATIONS
) AS ranked_vaccines
WHERE rn = 1
"""

# Execute the SQL query using the session
df_with_longest_vaccine = demo_session.sql(sql)

# Display the results
df_with_longest_vaccine.show()
```

```
-------------------------------------------------------------------------------
|"COUNTRY_REGION"           |"VACCINES"                                       |
-------------------------------------------------------------------------------
|Trinidad and Tobago        |Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bio...  |
|Zimbabwe                   |Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac,...  |
|Iran                       |COVIran Barekat, CanSino, Covaxin, Johnson&John...  |
|Fiji                       |Moderna, Oxford/AstraZeneca, Pfizer/BioNTech        |
|Dominican Republic         |Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/...  |
|El Salvador                |Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/...  |
|Maldives                   |Johnson&Johnson, Moderna, Oxford/AstraZeneca, P...  |
|Sint Maarten (Dutch part)  |Moderna, Oxford/AstraZeneca, Pfizer/BioNTech        |
|Belize                     |Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bio...  |
|Mozambique                 |Johnson&Johnson, Oxford/AstraZeneca, Sinopharm/...  |
-------------------------------------------------------------------------------
```

```
# Sum of Vaccination by each country
from snowflake.snowpark.functions import col, lit, sum as sum_, max as max_
df_vaccines_sum = df.group_by("COUNTRY_REGION").agg(sum_("TOTAL_VACCINATIONS").alias("TOTAL_VACCINATIONS"))
df_vaccines_sum.show()
```

```
-----------------------------------------------
|"COUNTRY_REGION"     |"TOTAL_VACCINATIONS"  |
-----------------------------------------------
|Afghanistan          |1659147079            |
|Albania              |395364296             |
|Andorra              |5846260               |
|Angola               |1398489266            |
|Anguilla             |1290079               |
|Antigua and Barbuda  |7482788               |
|Argentina            |86986241297           |
|Armenia              |58724412              |
|Aruba                |98025212              |
|Austria              |1897828020            |
-----------------------------------------------
```

Join Both Column

```
from snowflake.snowpark.functions import col

df_join = df_vaccines_sum.join(df_with_longest_vaccine, "COUNTRY_REGION").select(df_vaccines_sum.COUNTRY_REGION.alias("COUNTRY_REGIOI
                                                    df_vaccines_sum.TOTAL_VACCINATIONS, df_with_longest_vaccine.VACCINES)
df_join.show()
```

```
-------------------------------------------------------------------------------------------
|"COUNTRY_REGION"  |"TOTAL_VACCINATIONS"  |"VACCINES"                                     |
-------------------------------------------------------------------------------------------
|Mozambique        |561270607             |Johnson&Johnson, Oxford/AstraZeneca, Sinopharm/...  |
|Albania           |395364296             |Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, S...  |
|Tajikistan        |324439943             |Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, S...  |
|Zimbabwe          |3211422214            |Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac,...  |
```

```
|Tokelau           |18977                |Pfizer/BioNTech                                  |
|Oman              |202915483            |CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |
|Cayman Islands    |12327845             |Oxford/AstraZeneca, Pfizer/BioNTech              |
|Kuwait            |1044825806           |CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...  |
|Luxembourg        |684641413            |Johnson&Johnson, Moderna, Novavax, Oxford/Astra...  |
|Senegal           |107224427            |Johnson&Johnson, Oxford/AstraZeneca, Sinopharm/...  |
-----------------------------------------------------------------------------------------------
```

Do Data Transformation for 2nd Data

```
df2 = demo_session.sql('SELECT * FROM WHO_SITUATION_REPORTS')
df2.show()
```

```
-----------------------------------------------------------------------------------------------------------------------------------------------
|"COUNTRY"      |"TOTAL_CASES"  |"CASES_NEW"  |"DEATHS"  |"DEATHS_NEW"  |"TRANSMISSION_CLASSIFICATION"  |"DAYS_SINCE_LAST_REPORTED_(
-----------------------------------------------------------------------------------------------------------------------------------------------
|South Africa   |553188         |7712         |10210     |301           |Community transmission         |0
|Nigeria        |46140          |453          |942       |6             |Community transmission         |0
|Ghana          |40533          |436          |206       |0             |Community transmission         |0
|Algeria        |34693          |538          |1293      |11            |Community transmission         |0
|Kenya          |25837          |699          |418       |5             |Community transmission         |0
|Ethiopia       |22253          |801          |390       |10            |Community transmission         |0
|Cameroon       |18042          |0            |395       |0             |Community transmission         |1
|Côte d'Ivoire  |16620          |96           |104       |0             |Community transmission         |0
|Madagascar     |12922          |214          |141       |6             |Community transmission         |0
|Senegal        |11003          |116          |229       |4             |Community transmission         |0
-----------------------------------------------------------------------------------------------------------------------------------------------
```

```
# Check the datatypes
df2.dtypes
```

```
[('COUNTRY', 'string(16777216)'),
 ('TOTAL_CASES', 'bigint'),
 ('CASES_NEW', 'bigint'),
 ('DEATHS', 'bigint'),
 ('DEATHS_NEW', 'bigint'),
 ('TRANSMISSION_CLASSIFICATION', 'string(16777216)'),
 ('DAYS_SINCE_LAST_REPORTED_CASE', 'bigint'),
 ('ISO3166_1', 'string(16777216)'),
 ('COUNTRY_REGION', 'string(16777216)'),
 ('DATE', 'date'),
 ('SITUATION_REPORT_NAME', 'string(16777216)'),
 ('SITUATION_REPORT_URL', 'string(16777216)'),
 ('LAST_UPDATE_DATE', 'timestamp'),
 ('LAST_REPORTED_FLAG', 'boolean')]
```

```
# Group by COUNTRY_REGION and calculate sums
df_cases_and_death_sum = df2.groupBy("COUNTRY_REGION") \
                          .agg(sum_(col("TOTAL_CASES")).alias("TOTAL_CASES"),  # Calculate sum of TOTAL_CASES
                               sum_(col("DEATHS")).alias("DEATHS"))  # Calculate sum of DEATHS

# Display the results
df_cases_and_death_sum.show()
```

```
-------------------------------------------------------------------
|"COUNTRY_REGION"                     |"TOTAL_CASES"  |"DEATHS"  |
-------------------------------------------------------------------
|South Africa                         |11879542       |196895    |
|Nigeria                              |1504842        |37748     |
|Ghana                                |1112680        |6107      |
|Algeria                              |1032520        |61547     |
|Ethiopia                             |413832         |6964      |
|Côte d'Ivoire                        |532443         |3852      |
|Madagascar                           |245945         |2239      |
|Senegal                              |379824         |6423      |
|Congo, The Democratic Republic of the  |335702         |8281      |
|Zambia                               |138227         |3379      |
-------------------------------------------------------------------
```

Do Join Table Between df_cases_and_death_sum and df_join

```
from snowflake.snowpark.functions import col

df_join2 = df_cases_and_death_sum.join(df_join, "COUNTRY_REGION").select(df_cases_and_death_sum.COUNTRY_REGION.alias("COUNTRY_REGION")
                                                                         df_cases_and_death_sum.TOTAL_CASES, df_cases_and_death_sum.DEATHS,
                                                                         df_join.TOTAL_VACCINATIONS, df_join.VACCINES)
```

```
6 df_join2.show()
```

```
-----------------------------------------------------------------------------------------------------------------------------------
|"COUNTRY_REGION"        |"TOTAL_CASES"  |"DEATHS"  |"TOTAL_VACCINATIONS"  |"VACCINES"                                                |
-----------------------------------------------------------------------------------------------------------------------------------
|Lithuania               |197936         |7742      |3318934224            |Johnson&Johnson, Moderna, Novavax, Oxford/Astra...        |
|Thailand                |397600         |6848      |45547040066           |Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, S...        |
|Norway                  |1025544        |27431     |5163271667            |Moderna, Pfizer/BioNTech                                  |
|Maldives                |216528         |877       |214353973             |Johnson&Johnson, Moderna, Oxford/AstraZeneca, P...        |
|Barbados                |11667          |790       |128989451             |Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bio...        |
|Cyprus                  |111396         |2323      |188496747             |Johnson&Johnson, Moderna, Novavax, Oxford/Astra...        |
|Togo                    |54860          |1437      |114080933             |Oxford/AstraZeneca                                        |
|Bosnia and Herzegovina  |519635         |19876     |13550064              |Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, S...        |
|Italy                   |29000446       |4026385   |112298543444          |Johnson&Johnson, Moderna, Novavax, Oxford/Astra...        |
|Iraq                    |4514214        |174372    |1082797012            |CanSino, Covaxin, Johnson&Johnson, Moderna, Oxf...        |
-----------------------------------------------------------------------------------------------------------------------------------
```

Next thing is do other join and lastly make a Dashboard