CLOUD DATA MANAGEMENT PROJECT: COVID19 analysis

Gianluca Molteni mat 730113 e Vladi Valsecchi 730030

Università degli Studi dell'Insubria - a.a. 2020/2021

SOMMARIO

INTRODUCTION 4

DESCRIPTION OF THE DATASETS 4

NOTEBOOK 6

QUERIES 8

Query 0: show last day recorded for each region 9

Query 1: show top 5 regions for most cases 9

Query 2: show top 5 regions for less cases 10

Query 3: show average number of cases daily for each region 11

Query 4: show top 5 regions for most diagnostic tests done 12

Query 5: show top 5 regions for less diagnostic tests done **Errore. Il segnalibro non è definito.**

Query 6: show average number of diagnostic tests done daily for each region **Errore. Il segnalibro non è definito.**

Query 7: show top 5 regions for most deaths 15

Query 8: show top 5 regions for less deaths 15

Query 9: show average number of deaths daily for each region 17

Query 10: show top 5 regions for most healed 18

Query 11: show top 5 regions for less healed 18

Query 12: show average number of healed daily for each region 20

Query 13: show top 5 regions for most people in intensive care in a single day 21

Query 14: show top 5 regions for less people in intensive care in the last day recorded 22

Query 15: show average number of people in intensive care daily for each region 23

Query 16: show region with most cases in a single day and date 24

Query 17: show date with biggest increase of cases in Italy and number of cases 24

Query 18: show biggest increase of cases in a single region and date 25

Query 19: show biggest decrease of cases in a single region and date 25

Query 20: show top 5 regions for most people in home isolation of all time 26

Query 21: show top 5 regions for less people in home isolation in the last day recorded 27

Query 22: show average number of people in home isolation daily for each region 28

Query 23: show top 5 regions for total cases and number of first doses, second doses and total doses of vaccine done 29

Query 24: show top 5 regions for total healed and number of first doses, second doses and total doses of vaccine done 31

Query 25: show top 5 regions for total deaths and number of first doses, second doses and total doses of vaccine done 33

Query 26: show for each vaccine the number of first doses, second doses and total doses done 35

Query 27: show for each region, for each vaccine the number of first doses, second doses and total doses done 36

Query 28: show the number of doses administered to males, females and the total number 39
Query 29: show for each region the number of doses administered to males, females and the total number 40
Query 30: show for each age group the number of first doses, second doses and total doses done 41
Query 31: show for each region, for each age group the number of first doses, second doses and total doses done 42

BIBLIOGRAPHY 47

INTRODUCTION

The project aims to analyze the SARS-CoV-19 pandemic, commonly known as Coronavirus or Covid19 for short.
Through queries on various datasets, relating to the infection and the vaccination campaign, we will investigate different consequences of this pandemic.

DESCRIPTION OF THE DATASETS

The data comes from two different datasets.
The first dataset, *dpc_covid19_ita_regioni.csv*, comes from the Github repository of the Dipartimento della Protezione Civile and refers to the progress of the Covid19 pandemic in Italy.
The data are collected daily starting from the first positive cases discovered on Italian soil.
Each individual entity has the following attributes:

- data: Date of notification
- stato: Country of reference
- codice_regione: Code of the Region (ISTAT 2019)
- denominazione_regione: Name of the Region
- lat: Latitude coordinates
- long: Longitude coordinates
- ricoverati_con_sintomi: Hospitalised patients with symptoms
- terapia_intensiva: Intensive Care
- totale_ospedalizzati: Total hospitalised patients
- isolamento_domiciliare: Home confinement
- totale_positivi: Total amount of current positive cases (Hospitalised patients + Home confinement)
- variazione_totale_positivi: News amount of current positive cases (totale_positivi current day - totale_positivi previous day)
- nuovi_positivi: News amount of current positive cases (totale_casi current day - totale_casi previous day)
- dimessi_guariti: Recovered
- deceduti: Death
- casi_da_sospetto_diagnostico: Positive cases emerged from clinical activity (No longer populated)
- casi_da_screening: Positive cases emerging from surveys and tests, planned at national or regional level (No longer populated)
- totale_casi: Total amount of positive cases
- tamponi: Tests performed (processed with molecular tests)
- casi_testati: Total number of people tested
- note: Notes
- ingressi_terapia_intensiva: Daily admissions to intensive care
- note_test,note_casi: Notes on the tests carried out
- totale_positivi_test_molecolare: Total number of positive molecular test
- totale_positivi_test_antigenico_rapido: Total number of positive rapid antigen testing
- tamponi_test_molecolare: Total number of molecular test

- tamponi_test_antigenico_rapido: Total number of rapid antigen testing
- codice_nuts_1: European classification of NUTS territorial units: NUTS level 1
- codice_nuts_2

Not all fields are used for the execution of the project.

The second dataset, *somministrazione_vaccini_latest*, comes from the official website of the Italian government and shows the daily progress of the vaccination campaign.
Each update consists of:

- data_somministrazione: Date of administration
- fornitore: Supplier
- area: Abbreviation region name
- fascia_anagrafica: Personal range
- sesso_maschile: Male
- sesso_femminile: Female
- prima_dose: Number of first administrations
- seconda_dose: Number of second administrations.
- pregressa_infezione: Number of administrations carried out to subjects with previous covid-19 infection in the period 3-6 months and who, therefore, end the vaccination cycle with a single dose
- codice_NUTS1: European classification of NUTS territorial units: NUTS level 1
- codice_NUTS2: European classification of NUTS territorial units: NUTS level 2
- codice_regione_ISTAT: Code of the Region (ISTAT 2019)
- nome_area: Name of the Region

Some considerations. Currently the suppliers for the Italian vaccination campaign are Pfizer/Biontech, Janssen (Johnson & Johnson group), Moderna and Vaxzevria (ex AstraZeneca). There are no other vaccines known as Sputnik V, Novavax e Sinovac.
Based on the personal data range and on the basis of personal medical history, one brand is administered rather than another. Example: Moderna is recommended for seniors, so most doses will be given in the 70-79 and 80+ range.
Janssen is single-dose, so 0 appears in the second_dose column.
Due to the expiry of the contract between Vaxzevria and the EU, the doses administered will be lower and lower in the coming months. The current availability is reserved for those who have yet to do the recall or will be used for a hypothetical third dose.
It can also be seen that the administration of the vaccines went in order of age following the range shown in the table. Hence the older groups have a greater number of vaccinated than the younger groups.

During the design phase, to query both tables, a JOIN operation was performed with the following fields keys:

- data/data_somministrazione;
- codice_regione/codice_regione_ISTAT;
- denominazione_regione.

NOTEBOOK

The code was written on a notebook in .ipynb format in Python language. For the execution of the queries we used different libraries. Specifically, these are Numpy, Pandas and PySpark.

```
#Import
from pyspark.sql import SparkSession
import pandas as pd
import numpy as np
```

```
#Datasets urls
url_region="https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni.csv"
url_vaccine="https://raw.githubusercontent.com/italia/covid19-opendata-vaccini/master/dati/somministrazioni-vaccini-latest.csv"
```

After carrying out the phase of importing the necessary libraries and setting the reference links, we proceed to create the Pandas DataFrame.

```
#Create pandas data frame from url
pdf_r=pd.read_csv(url_region)
pdf_v=pd.read_csv(url_vaccine)
```

Pandas and Numpy are used for management and synchronization between the two DataFrames. In *dpc_covid19_ita_regioni* we have changed the date from a format “yyyy-mm-ddThh: mm: ss” to “yyyy-mm-dd”, so that it is in the same format as *somministrazioni_vaccini_latest*. The codice_region of the Autonomous Province of Trento and the Autonomous Province of Bolzano is subsequently standardized from 4 to 21 and 22 in the *somministrazioni_vaccini_latest* dataset so that they correspond to *dpc_covid19_ita_regioni*.

```
#fix regions dataset (keep only date yyyy-mm-dd)
pdf_r['data']=pdf_r['data'].str[:10] #keep only the first 10 characters
#fix vaccine dataset (Provincia Autonoma di Trento and Bolzano from region code 4 to 21 and 22)
unique_areas=pdf_v.nome_area.unique()#create unique area names list
pdf_v.loc[pdf_v.nome_area == unique_areas[11], "codice_regione_ISTAT"] = np.int64(21)#change unique
area name region code in position 11 from 4 to 21
pdf_v.loc[pdf_v.nome_area == unique_areas[12], "codice_regione_ISTAT"] = np.int64(22)#change unique
area name region code in position 12 from 4 to 22
```

PySpark is used through SparkSession which provides the entry point to programming Spark with the Dataset and DataFrame API. A SparkSession can be used to create DataFrame, register DataFrame as tables (through createOrReplaceTempView), execute SQL over tables, cache tables, and read parquet files.

```
#Spark setup
spark = SparkSession.builder.appName("Query").getOrCreate()
spark.conf.set("spark.sql.execution.arrow.enabled", "true")

#Create spark dataframe from pandas data frame
dpc_covid19_ita_regioni = spark.createDataFrame(pdf_r)#create spark dataframe for
dpc_covid19_ita_regioni
somministrazioni_vaccini_latest=spark.createDataFrame(pdf_v)#create spark dataframe for
somministrazioni_vaccini_latest

#Create temporary views that will be used in the queries
dpc_covid19_ita_regioni.createOrReplaceTempView("base")
somministrazioni_vaccini_latest.createOrReplaceTempView("base2")
```

QUERIES

Below are the 32 queries that we have established and then the individual queries commented and their results.

- **Query 0**: show last day recorded for each region
- **Query 1**: show top 5 regions for most cases
- **Query 2**: show top 5 regions for less cases
- **Query 3**: show average number of cases daily for each region
- **Query 4**: show top 5 regions for most diagnostic tests done
- **Query 5**: show top 5 regions for less diagnostic tests done
- **Query 6**: show average number of diagnostic tests done daily for each region
- **Query 7**: show top 5 regions for most deaths
- **Query 8**: show top 5 regions for less deaths
- **Query 9**: show average number of deaths daily for each region
- **Query 10**: show top 5 regions for most healed
- **Query 11**: show top 5 regions for less healed
- **Query 12**: show average number of healed daily for each region
- **Query 13**: show top 5 regions for most people in intensive care in a single day
- **Query 14**: show top 5 regions for less people in intensive care in the last day recorded
- **Query 15**: show average number of people in intensive care daily for each region
- **Query 16**: show region with most cases in a single day and date
- **Query 17**: show date with biggest increase of cases in Italy and number of cases
- **Query 18**: show biggest increase of cases in a single region and date
- **Query 19**: show biggest decrease of cases in a single region and date
- **Query 20**: show top 5 regions for most people in home isolation of all time
- **Query 21**: show top 5 regions for less people in home isolation in the last day recorded
- **Query 22**: show average number of people in home isolation daily for each region
- **Query 23**: show top 5 regions for total cases and number of first doses, second doses and total doses of vaccine done
- **Query 24**: show top 5 regions for total healed and number of first doses, second doses and total doses of vaccine done
- **Query 25**: show top 5 regions for total deaths and number of first doses, second doses and total doses of vaccine done
- **Query 26**: show for each vaccine the number of first doses, second doses and total doses done
- **Query 27**: show for each region, for each vaccine the number of first doses, second doses and total doses done
- **Query 28**: show the number of doses administered to males, females and the total number
- **Query 29**: show for each region the number of doses administered to males, females and the total number
- **Query 30**: show for each age group the number of first doses, second doses and total doses done
- **Query 31**: show for each region, for each age group the number of first doses, second doses and total doses done

Query 0: show last day recorded for each region

Order all rows by data and show the first 21 rows.

```
SELECT * --select all
FROM base --from dpc_covid19_ita_regioni
ORDER BY data DESC --order from the biggest to the smallest value
LIMIT 21 --limit to 21 to get only 1 result for each region
```

Query 1: show top 5 regions for most cases

Group rows by denominazione_regione, extract max value of totale_casi from each group, order all results and show the first 5 rows.

```
SELECT
MAX(totale_casi) AS top_5_regions_for_most_cases,
--get max value of totale_casi alias top_5_regions_for_most_cases
denominazione_regione AS region --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MAX(totale_casi) DESC  --order from biggest to smallest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_most_cases	region
840689	Lombardia
425009	Veneto
423381	Campania
386265	Emilia-Romagna
362594	Piemonte

Query 2: show top 5 regions for less cases

Group rows by denominazione_regione, extract min value of totale_casi from each group, order all results and show the first 5 rows.

```
SELECT
MIN(totale_casi) AS top_5_regions_for_less_cases,
--get min value of totale_casi alias top_5_regions_for_less_cases
denominazione_regione AS region  --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MIN(totale_casi) ASC --order from smallest to biggest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_less_cases	region
11674	Valle d'Aosta
13682	Molise
26827	Basilicata
45731	P.A. Trento
56771	Umbria

Query 3: show average number of cases daily for each region

Group rows by denominazione_regione, extract avg value of nuovi_positivi from each group, order all results.

```
SELECT
AVG(nuovi_positivi) AS avg_new_cases_daily,
--get average value of nuovi_positivi alias avg_new_cases_daily
denominazione_regione AS region --denominazione_regione alias region
FROM base --from dpc_covid19_ita_regioni
GROUP BY denominazione_regione --group each region name
ORDER BY AVG(nuovi_positivi) DESC --order from biggest to smallest value
```

avg_new_cases_daily	region
1740.606625258799	Lombardia
879.9337474120083	Veneto
876.4782608695652	Campania
802.51966873706	Emilia-Romagna
756.0745341614906	Piemonte
714.7267080745341	Lazio
523.5320910973085	Puglia
504.8944099378882	Toscana
478.2277432712215	Sicilia
222.9648033126294	Friuli Venezia Gi...
214.37060041407867	Marche
213.5631469979296	Liguria
154.84886128364388	Abruzzo
142.1262939958592	Calabria
130.04140786749483	P.A. Bolzano
118.29606625258799	Sardegna
117.5383022774327	Umbria
93.87991718426501	P.A. Trento
55.57556935817805	Basilicata
28.335403726708076	Molise
24.113871635610767	Valle d'Aosta

Query 4: show top 5 regions for most diagnostic tests done

Group rows by denominazione_regione, extract max value of tamponi from each group, order all results and show the first 5 rows.

```
SELECT
MAX(tamponi) AS top_5_regions_for_most_DiagnosticTests_done, --
get max value of tamponi alias top_5_regions_for_most_DiagnosticTests_done
denominazione_regione AS region --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MAX(tamponi) DESC --order from biggest to smallest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_most_DiagnosticTests_done	region
11317204	Lombardia
8472946	Veneto
7423680	Lazio
6569590	Emilia-Romagna
5171077	Campania

Query 5: show top 5 regions for less diagnostic tests done

Group rows by denominazione_regione, extract min value of tamponi from each group, order all results and show the first 5 rows.

```
SELECT
MIN(tamponi) AS top_5_regions_for_less_DiagnosticTests_done, --
get min value of tamponi alias top_5_regions_for_less_DiagnosticTests_done
denominazione_regione AS region --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MIN(tamponi) ASC --order from smallest to biggest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_less_DiagnosticTests_done	region
136736	Valle d'Aosta
237156	Molise
386912	Basilicata
855808	P.A. Trento
911859	Calabria

Query 6: show average number of diagnostic tests done daily for each region

Group rows by denominazione_regione, extract avg value of tamponi from each group, order all results.

```
SELECT
AVG(tamponi) AS avg_new_DiagnosticTests_daily,
 --get average value of tamponi alias avg_new_DiagnosticTests_daily
denominazione_regione AS region --denominazione_regione alias region
FROM base --from dpc_covid19_ita_regioni
GROUP BY denominazione_regione --group each region name
ORDER BY AVG(tamponi) DESC --order from the biggest to the smallest value
```

avg_new_DiagnosticTests_daily	region
3883739.50310559	Lombardia
2964353.0973084886	Veneto
2263808.6314699794	Lazio
2191687.5817805384	Emilia-Romagna
1610114.122153209	Campania
1581074.0559006212	Toscana
1527242.6480331263	Piemonte
1322873.685300207	Sicilia
856617.9130434783	Puglia
750665.2298136646	Friuli Venezia Gi...
583793.0579710145	Liguria
512040.87163561076	Abruzzo
459970.5093167702	Marche
447631.21532091097	P.A. Bolzano
445729.98136645963	Umbria
427223.3084886128	Sardegna
346796.9399585921	P.A. Trento
335481.0910973085	Calabria
140469.3188405797	Basilicata
87183.7950310559	Molise
49621.03519668737	Valle d'Aosta

Query 7: show top 5 regions for most deaths

Group rows by denominazione_regione, extract max value of deceduti from each group, order all results and show the first 5 rows.

```
SELECT
MAX(deceduti) AS top_5_regions_for_most_deaths,
 --get max value of deceduti alias top_5_regions_for_most_deaths
denominazione_regione AS region --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MAX(deceduti) DESC --order from the biggest to the smallest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_most_deaths	region
33753	Lombardia
13246	Emilia-Romagna
11690	Piemonte
11603	Veneto
8304	Lazio

Query 8: show top 5 regions for less deaths

Group rows by denominazione_regione, extract min value of deceduti from each group, order all results and show the first 5 rows.

```
SELECT
MIN(deceduti) AS top_5_regions_for_less_deaths,
--get min value of deceduti alias top_5_regions_for_less_deaths
denominazione_regione AS region
--denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
 --from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MIN(deceduti) ASC --order from smallest to the biggest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_less_deaths	region
472	Valle d'Aosta
491	Molise
589	Basilicata
1180	P.A. Bolzano
1219	Calabria

Query 9: show average number of deaths daily for each region

Group rows by denominazione_regione, extract avg value of deceduti from each group, order all results.

```
SELECT
AVG(deceduti) AS avg_deaths_daily, --get average value of deceduti alias avg_deaths_daily
denominazione_regione AS region --denominazione_regione alias region
FROM base --from dpc_covid19_ita_regioni
GROUP BY denominazione_regione --group each region name
ORDER BY AVG(deceduti) DESC --order from the biggest to the smallest value
```

avg_deaths_daily	region
20602.48447204969	Lombardia
6618.739130434783	Emilia-Romagna
6046.590062111802	Piemonte
4971.836438923396	Veneto
2950.6293995859214	Lazio
2680.6749482401656	Toscana
2297.3002070393377	Liguria
2271.2981366459626	Campania
2099.6356107660454	Puglia
1929.4182194616976	Sicilia
1476.0621118012423	Marche
1366.2111801242236	Friuli Venezia Gi...
1005.5590062111801	Abruzzo
707.9751552795032	P.A. Trento
564.9565217391304	P.A. Bolzano
548.1138716356107	Sardegna
486.1966873706004	Umbria
376.7639751552795	Calabria
254.0952380952381	Valle d'Aosta
187.01656314699792	Basilicata
163.57556935817806	Molise

Query 10: show top 5 regions for most healed

Group rows by denominazione_regione, extract max value of dimessi_guariti from each group, order all results and show the first 5 rows.

```
SELECT
MAX(dimessi_guariti) AS top_5_regions_for_most_healed,
--get max value of dimessi_guariti
denominazione_regione AS region --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MAX(dimessi_guariti) DESC
--order from the biggest to the smallest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_most_healed	region
793279	Lombardia
408194	Veneto
405623	Campania
367429	Emilia-Romagna
349206	Piemonte

Query 11: show top 5 regions for less healed

Group rows by denominazione_regione, extract min value of dimessi_guariti from each group, order all results and show the first 5 rows.

```
SELECT
MIN(dimessi_guariti) AS top_5_regions_for_less_healed,
--get max value of dimessi_guariti alias top_5_regions_for_less_healed
denominazione_regione AS region --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MIN(dimessi_guariti) ASC
--order from the smallest to the biggest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_less_healed	region
11145	Valle d'Aosta
13101	Molise
24938	Basilicata
44055	Sardegna
44221	P.A. Trento

Query 12: show average number of healed daily for each region

Group rows by denominazione_regione, extract avg value of dimessi_guariti from each group, order all results.

```
SELECT
AVG(dimessi_guariti) AS avg_healed_daily,
--get average value of dimessi_guariti alias avg_healed_daily
denominazione_regione AS region --denominazione_regione alias region
FROM base --from dpc_covid19_ita_regioni
GROUP BY denominazione_regione --group each region name
ORDER BY AVG(dimessi_guariti) DESC
--order from the biggest to the smallest value
```

avg_healed_daily	region
273208.2960662526	Lombardia
131294.69565217392	Veneto
115180.2132505176	Piemonte
104589.8612836439	Emilia-Romagna
93134.89855072464	Campania
85536.96066252587	Lazio
69984.06418219462	Toscana
56114.35817805383	Sicilia
52417.97101449275	Puglia
35560.977225672876	Liguria
30153.604554865426	Friuli Venezia Gi...
28713.71635610766	Marche
21316.75569358178	P.A. Bolzano
19665.96894409938	Abruzzo
17265.111801242238	Umbria
15441.16977225673	P.A. Trento
14392.260869565218	Calabria
12310.244306418219	Sardegna
5407.5859213250515	Basilicata
4010.927536231884	Valle d'Aosta
3956.215320910973	Molise

Query 13: show top 5 regions for most people in intensive care in a single day

Group rows by denominazione_regione, extract max value of terapia_intensiva from each group, order all results and show the first 5 rows.

```
SELECT
MAX(terapia_intensiva) AS top_5_regions_for_most_IC,
--get max value of terapia_intensiva alias top_5_regions_for_most_IC
denominazione_regione AS region --denominazione_regione alias region
FROM base --from dpc_covid19_ita_regioni
GROUP BY denominazione_regione --group each region name
ORDER BY MAX(terapia_intensiva) DESC
--order from the biggest to the smallest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_most_IC	region
1381	Lombardia
453	Piemonte
402	Emilia-Romagna
398	Lazio
372	Veneto

Query 14: show top 5 regions for less people in intensive care in the last day recorded

Group rows by denominazione_regione, extract min value of terapia_intensiva from each group, order all results and show the first 5 rows.

```
SELECT
MIN(terapia_intensiva) AS
top_5_regions_for_less_IC_in_the_last_day,
--get min value of terapia_intensiva alias top_5_regions_for_less_IC_in_the_last_day
denominazione_regione AS region --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MIN(terapia_intensiva) ASC
--order from the smallest to the biggest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_less_IC_in_the_last_day	region
0	Basilicata
0	Molise
0	Valle d'Aosta
1	Abruzzo
1	Friuli Venezia Gi...

Query 15: show average number of people in intensive care daily for each region

Group rows by denominazione_regione, extract avg value of terapia_intensiva from each group, order all results.

```
SELECT
AVG(terapia_intensiva) AS avg_IC_daily,
--get average value of terapia_intensiva alias avg_IC_daily
denominazione_regione AS region --denominazione_regione alias region
FROM base --from dpc_covid19_ita_regioni
GROUP BY denominazione_regione --group each region name
ORDER BY AVG(terapia_intensiva) DESC
--order from the biggest to the smallest value
```

avg_IC_daily	region
414.5879917184265	Lombardia
163.39544513457557	Lazio
153.98343685300208	Piemonte
153.66045548654245	Emilia-Romagna
130.935817805383	Veneto
123.00414078674949	Toscana
91.59627329192547	Puglia
86.93167701863354	Sicilia
76.07867494824016	Campania
52.17184265010352	Marche
49.515527950310556	Liguria
30.546583850931675	Abruzzo
28.815734989648032	Friuli Venezia Gi...
28.163561076604555	Umbria
24.4824016563147	Sardegna
21.70600414078675	P.A. Trento
15.461697722567287	P.A. Bolzano
15.066252587991718	Calabria
5.9937888198757765	Basilicata
5.3768115942028984	Molise
4.46583850931677	Valle d'Aosta

Query 16: show region with most cases in a single day and date

Order rows by nuovi_positivi and show row 1 result.

```
SELECT
denominazione_regione AS region_with_most_cases_in_a_single_day,
--denominazione_regione alias region_with_most_cases_in_a_single_day
data AS date, --data alias date
nuovi_positivi AS new_cases --nuovi_positivi alias new_cases
FROM base --from dpc_covid19_ita_regioni
ORDER BY nuovi_positivi DESC --order from the biggest to the smallest value
LIMIT 1 --limit to 1 to get only the first result
```

```
+-------------------------------------+----------+---------+
|region_with_most_cases_in_a_single_day|      date|new_cases|
+-------------------------------------+----------+---------+
|                            Lombardia|2020-11-07|    11489|
+-------------------------------------+----------+---------+
```

Query 17: show date with biggest increase of cases in Italy and number of cases

Group rows by date, sum nuovi_positivi for each group, order results and show the first row.

```
SELECT
data AS date_with_most_cases_in_Italy,
--data alias date_with_most_cases_in_Italy
SUM(nuovi_positivi) AS cases --sum nuovi_positivi alias cases
FROM base --from dpc_covid19_ita_regioni
GROUP BY data --group each date
ORDER BY SUM(nuovi_positivi) DESC
--order from the biggest to the smallest value
LIMIT 1 --limit to 1 to get only the first result
```

```
+----------------------------------------+-----+
|date_with_biggest_increase_cases_in_Italy|cases|
+----------------------------------------+-----+
|                              2020-11-13|40902|
+----------------------------------------+-----+
```

Query 18: show biggest increase of cases in a single region and date

Order rows by variazione_totale_positivi from the biggest value to the smallest and show the first row.

```
SELECT
variazione_totale_positivi AS biggest_increase_of_cases, --
variazione_totale_positivi alias biggest_increase_of_cases
denominazione_regione AS region, --denominazione_regione alias region
data AS date --data alias date
FROM base --from dpc_covid19_ita_regioni
ORDER BY variazione_totale_positivi DESC
--order from the biggest to the smallest value
LIMIT 1 --limit to 1 to get only the first result
```

```
+-------------------------+---------+----------+
|biggest_increase_of_cases|   region|      date|
+-------------------------+---------+----------+
|                    10263|Lombardia|2020-11-07|
+-------------------------+---------+----------+
```

Query 19: show biggest decrease of cases in a single region and date

Order rows by variazione_totale_positivi from the smallest value to the biggest and show the first row.

```
SELECT
variazione_totale_positivi AS biggest_decrease_of_cases,
--variazione_totale_positivi alias biggest_decrease_of_cases
denominazione_regione AS region, --denominazione_regione alias region
data AS date --data alias date
FROM base --from dpc_covid19_ita_regioni
ORDER BY variazione_totale_positivi ASC
 --order from the smallest to the biggest value
LIMIT 1 --limit to 1 to get only the first result
```

```
+-------------------------+--------+----------+
|biggest_decrease_of_cases|  region|      date|
+-------------------------+--------+----------+
|                   -48092|Campania|2021-06-15|
+-------------------------+--------+----------+
```

Query 20: show top 5 regions for most people in home isolation of all time

Group rows by denominazione_regione, extract max value of isolamento_domiciliare from each group, order all results and show the first 5 rows.

```
SELECT
MAX(isolamento_domiciliare) AS
top_5_regions_for_most_home_isolation_of_all_time,
--get max value of isolamento_domiciliare alias
top_5_regions_for_most_home_isolation_of_all_time
denominazione_regione AS region --denominazione_regione alias region
FROM base --from dpc_covid19_ita_regioni
GROUP BY denominazione_regione --group each region name
ORDER BY MAX(isolamento_domiciliare) DESC
--order from the biggest to the smallest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_most_home_isolation_of_all_time	region
155066	Lombardia
102438	Campania
101041	Veneto
90735	Lazio
73011	Piemonte

Query 21: show top 5 regions for less people in home isolation in the last day recorded

Group rows by denominazione_regione, extract min value of isolamento_domiciliare from each group, order all results and show the first 5 rows.

```
SELECT
MIN(isolamento_domiciliare) AS
top_5_regions_for_less_home_isolation_in_the_last_day,
--get min value of isolamento_domiciliare alias
top_5_regions_for_less_home_isolation_in_the_last_day
denominazione_regione AS region --denominazione_regione alias region
FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21)
--from rows with the last date recorded
GROUP BY denominazione_regione --group each region name
ORDER BY MIN(isolamento_domiciliare) ASC
--order from the smallest to the biggest value
LIMIT 5 --limit to 5 to get only the first 5 results
```

top_5_regions_for_less_home_isolation_in_the_last_day	region
53	Valle d'Aosta
84	Molise
138	P.A. Trento
236	Liguria
267	P.A. Bolzano

Query 22: show average number of people in home isolation daily for each region

Group rows by denominazione_regione, extract avg value of isolamento_domicilare from each group, order all results.

```
SELECT
AVG(isolamento_domiciliare) AS avg_home_isolation_daily,
--get average value of isolamento_domiciliare alias avg_home_isolation_daily
denominazione_regione AS region --denominazione_regione alias region
FROM base --from dpc_covid19_ita_regioni
GROUP BY denominazione_regione --group each region name
ORDER BY AVG(isolamento_domiciliare) DESC
--order from the biggest to the smallest value
```

avg_home_isolation_daily	region
39121.923395445134	Campania
38351.57556935818	Lombardia
25591.08695652174	Lazio
24195.836438923394	Emilia-Romagna
21674.376811594204	Veneto
19377.31469979296	Puglia
14000.490683229813	Piemonte
13142.991718426501	Sicilia
10539.409937888198	Toscana
7058.271221532091	Sardegna
5195.416149068323	Abruzzo
5078.803312629399	Friuli Venezia Gi...
4703.5859213250515	Marche
4560.966873706004	Calabria
3303.0662525879916	Liguria
2960.7204968944097	P.A. Bolzano
2543.6521739130435	Umbria
2362.8385093167703	Basilicata
1145.0910973084885	P.A. Trento
639.8530020703934	Molise
369.14906832298135	Valle d'Aosta

Query 23: show top 5 regions for total cases and number of first doses, second doses and total doses of vaccine done

Show region (td), total_cases (td), first_dose (fsd), second_dose (fsd) and total_doses (td) by joining fsd and td and then order the results by total_cases (td).

- fsd groups rows by denominazione_regione (base) then extracts denominazione_regione (base), sums prima_dose (base2), sums seconda_dose (base2) and sums the two sums, all made by joining dpc_covid19_ita_regioni and somministrazioni_vaccini_latest;
- td extracts denominazione_regione (lr), totale_casi (lr), dosi (sd), all made by joining query_0 (lr) and sd;
- sd groups each row by denominazione_regione (base) then extracts denominazione_regione (base), sum of sums of prima_dose (base2) and seconda_dose (base2) all made by joining dpc_covid19_ita_regioni and somministrazioni_vaccini_latest;

```
--td total doses
--lr last record
--sd sum doses
--fsd first second doses
SELECT
td.region, --region name from td
td.total_cases, --total cases from td
fsd.first_dose, --first dose from fsd
fsd.second_dose, --second dose from fsd
td.total_doses --total doses from td
FROM
      (SELECT
      base.denominazione_regione,
       --denominazione_regione from dpc_covid19_ita_regioni
      SUM(base2.prima_dose) AS first_dose,
       --sum prima_dose from somministrazioni_vaccini_latest alias first_dose
      SUM(base2.seconda_dose) AS second_dose
       --sum seconda_dose from somministrazioni_vaccini_latest alias second_dose
      FROM base --from dpc_covid19_ita_regioni
      JOIN base2 --inner join somministrazioni_vaccini_latest
      ON base.data=base2.data_somministrazione --matching based on dates
      AND base.codice_regione=base2.codice_regione_ISTAT
      --and matching based on region codes
      GROUP BY base.denominazione_regione) AS fsd
      --group each region name, alias fsd
```

following in the next page

```
JOIN --inner join
	(SELECT
	lr.denominazione_regione AS region,
	 --denominazione_regione from lr alias region
	lr.totale_casi AS total_cases, --totale_casi from lr alias total_cases
	sd.dosi AS total_doses --dosi from sd alias total_doses
	FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21) AS lr
--from rows with the last date recorded, alias lr
	JOIN --inner join
		(SELECT
		base.denominazione_regione,
		  --denominazione_regione from dpc_covid19_ita_regioni
		SUM(base2.prima_dose)+SUM(base2.seconda_dose) AS dosi
--sum sums of prima_dose and seconda_dose alias dosi
		FROM base --from dpc_covid19_ita_regioni
		JOIN base2 --inner join somministrazioni_vaccini_latest
		ON base.data=base2.data_somministrazione --matching based on dates
		AND base.codice_regione=base2.codice_regione_ISTAT
		--and matching based on region codes
		GROUP BY base.denominazione_regione) AS sd
		 --group each region name, alias sd
	ON lr.denominazione_regione=sd.denominazione_regione --
matching based on region names
	ORDER BY lr.totale_casi DESC --order from biggest to smallest value
	LIMIT 5) AS td --limit to 5 to get only the first 5 results, alias td
ON td.region=fsd.denominazione_regione --matching based on region names
ORDER BY td.total_cases DESC --order from biggest to smallest value
```

region	total_cases	first_dose	second_dose	total_doses
Lombardia	840689	5472122	2493543	7965665
Veneto	425009	2529540	1173458	3702998
Campania	423381	3143876	1292624	4436500
Emilia-Romagna	386265	2286165	1147212	3433377
Piemonte	362594	2241788	1032040	3273828

Query 24: show top 5 regions for total healed and number of first doses, second doses and total doses of vaccine done

Show region (base), total_healed (td), first_dose (fsd), second_dose (fsd) and total_doses (td) by joining fsd and td and then order the results by total_healed (td).

- fsd groups rows by denominazione_regione (base) then extracts denominazione_regione (base), sums prima_dose (base2), sums seconda_dose (base2) and sums the two sums, all made by joining dpc_covid19_ita_regioni and somministrazioni_vaccini_latest;
- td extracts denominazione_regione (lr), dimessi_guariti (lr), dosi (sd), all made by joining query_0 (lr) and sd;
- sd groups each row by denominazione_regione (base) then extracts denominazione_regione (base), sum of sums of prima_dose (base2) and seconda_dose (base2)all made by joining dpc_covid19_ita_regioni and somministrazioni_vaccini_latest;

```
--td total doses
--lr last record
--sd sum doses
--fsd first second doses
SELECT
td.region, --region name from td
td.total_healed,  --total healed from td
fsd.first_dose, --first dose from fsd
fsd.second_dose, --second dose from fsd
td.total_doses --total doses from td
FROM
      (SELECT
      base.denominazione_regione,
--denominazione_regione from dpc_covid19_ita_regioni
      SUM(base2.prima_dose) AS first_dose,
--sum prima_dose from somministrazioni_vaccini_latest alias first_dose
      SUM(base2.seconda_dose) AS second_dose
--sum seconda_dose from somministrazioni_vaccini_latest alias second_dose
      FROM base --from dpc_covid19_ita_regioni
      JOIN base2  --inner join somministrazioni_vaccini_latest
      ON base.data=base2.data_somministrazione
--matching based on dates
      AND base.codice_regione=base2.codice_regione_ISTAT
--and matching based on region codes
      GROUP BY base.denominazione_regione) AS fsd
--group each region name, alias fsd
```

following in the next page...

```
JOIN --inner join
      (SELECT
      lr.denominazione_regione AS region,
--denominazione_regione from lr alias region
      lr.dimessi_guariti AS total_healed,
 --dimessi_guariti from lr alias total_healed
      sd.dosi AS total_doses --dosi from sd alias total_doses
      FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21) AS lr
--from rows with the last date recorded, alias lr
      JOIN --inner join
            (SELECT
            base.denominazione_regione,
--denominazione_regione from dpc_covid19_ita_regioni
            SUM(base2.prima_dose)+SUM(base2.seconda_dose) AS dosi
--sum sums of prima_dose and seconda_dose alias dosi
            FROM base --from dpc_covid19_ita_regioni
            JOIN base2 --inner join somministrazioni_vaccini_latest
            ON base.data=base2.data_somministrazione
--matching based on dates
            AND base.codice_regione=base2.codice_regione_ISTAT --
and matching based on region codes
            GROUP BY base.denominazione_regione) AS sd --group each
region name, alias sd
      ON lr.denominazione_regione=sd.denominazione_regione --
matching based on region names
      ORDER BY lr.totale_casi DESC --order from biggest to smallest value
      LIMIT 5) AS td --limit to 5 to get only the first 5 results, alias td
ON td.region=fsd.denominazione_regione --matching based on region names
ORDER BY td.total_healed DESC --order from biggest to smallest value
```

region	total_healed	first_dose	second_dose	total_doses
Lombardia	793279	5472122	2493543	7965665
Veneto	408194	2529540	1173458	3702998
Campania	405623	3143876	1292624	4436500
Emilia-Romagna	367429	2286165	1147212	3433377
Piemonte	349206	2241788	1032040	3273828

Query 25: show top 5 regions for total deaths and number of first doses, second doses and total doses of vaccine done

Show region (base), total_deaths (td), first_dose (fsd), second_dose (fsd) and total_doses (td) by joining fsd and td and then order the results by total_deaths (td).

- fsd groups rows by denominazione_regione (base) then extracts denominazione_regione (base), sums prima_dose (base2), sums seconda_dose (base2) and sums the two sums, all made by joining dpc_covid19_ita_regioni and somministrazioni_vaccini_latest;
- td extracts denominazione_regione (lr), deceduti (lr), dosi (sd), all made by joining query_0 (lr) and sd;
- sd groups each row by denominazione_regione (base) then extracts denominazione_regione (base), sum of sums of prima_dose (base2) and seconda_dose (base2)all made by joining dpc_covid19_ita_regioni and somministrazioni_vaccini_latest;

```
--td total doses
--lr last record
--sd sum doses
--fsd first second doses
SELECT
td.region, --region name from td
td.total_deaths, --total deaths from td
fsd.first_dose, --first dose from fsd
fsd.second_dose, --second dose from fsd
td.total_doses --total doses from td
FROM
      (SELECT
      base.denominazione_regione,
--denominazione_regione from dpc_covid19_ita_regioni
      SUM(base2.prima_dose) AS first_dose,
--sum prima_dose from somministrazioni_vaccini_latest alias first_dose
      SUM(base2.seconda_dose) AS second_dose
--sum seconda_dose from somministrazioni_vaccini_latest alias second_dose
      FROM base --from dpc_covid19_ita_regioni
      JOIN base2 --inner join somministrazioni_vaccini_latest
      ON base.data=base2.data_somministrazione
--matching based on dates
      AND base.codice_regione=base2.codice_regione_ISTAT
--and matching based on region codes
      GROUP BY base.denominazione_regione) AS fsd
--group each region name, alias fsd
```

following in the next page...

```
JOIN --inner join
      (SELECT
      lr.denominazione_regione AS region,
--denominazione_regione from lr alias region
      lr.deceduti AS total_deaths, --deceduti from lr alias total_deaths
      sd.dosi AS total_doses --dosi from sd alias total_doses
      FROM (SELECT * FROM base ORDER BY data DESC LIMIT 21) AS lr
--from rows with the last date recorded, alias lr
      JOIN --inner join
            (SELECT
            base.denominazione_regione,
--denominazione_regione from dpc_covid19_ita_regioni
            SUM(base2.prima_dose)+SUM(base2.seconda_dose) AS dosi --
-sum sums of prima_dose and seconda_dose alias dosi
            FROM base --from dpc_covid19_ita_regioni
            JOIN base2 --inner join somministrazioni_vaccini_latest
            ON base.data=base2.data_somministrazione
--matching based on dates
            AND base.codice_regione=base2.codice_regione_ISTAT
-and matching based on region codes
            GROUP BY base.denominazione_regione) AS sd
--group each region name, alias sd
      ON lr.denominazione_regione=sd.denominazione_regione --
matching based on region names
      ORDER BY lr.totale_casi DESC --order from biggest to smallest value
      LIMIT 5) AS td --limit to 5 to get only the first 5 results, alias td
ON td.region=fsd.denominazione_regione --matching based on region names
ORDER BY td.total_deaths DESC --order from biggest to smallest value
```

region	total_deaths	first_dose	second_dose	total_doses
Lombardia	33753	5472122	2493543	7965665
Emilia-Romagna	13246	2286165	1147212	3433377
Piemonte	11690	2241788	1032040	3273828
Veneto	11603	2529540	1173458	3702998
Campania	7386	3143876	1292624	4436500

Query 26: show for each vaccine the number of first doses, second doses and total doses done

Group each row by fornitore, sum prima_dose, seconda_dose for each group, sum the two sums.

```
SELECT
fornitore AS vaccine, --fornitore as vaccine
SUM(prima_dose) AS first_dose, --sum prima_dose alias first_dose
SUM(seconda_dose) AS second_dose, --sum seconda_dose alias second_dose
SUM(prima_dose)+SUM(seconda_dose) AS total_doses
--sum sums of prima_dose and seconda_dose alias total_doses
FROM base2 --from somministrazioni_vaccini_latest
GROUP BY fornitore --group each vaccine name
```

vaccine	first_dose	second_dose	total_doses
Vaxzevria (AstraZ...	6441695	1824002	8265697
Janssen	1184490	0	1184490
Moderna	2861917	1466710	4328627
Pfizer/BioNTech	20880824	11250315	32131139

Query 27: show for each region, for each vaccine the number of first doses, second doses and total doses done

Group each row by nome_area, codice_regione_ISTAT and fornitore, sum prima_dose, sum seconda_dose, sum of sums the two sums, order by codice_regione_ISTAT and fornitore.

```
SELECT
nome_area AS region, --nome_area alias region
codice_regione_ISTAT AS reg_cod, --codice_regione_ISTAT alias reg_cod
fornitore AS vaccine, --fornitore as vaccine
SUM(prima_dose) AS first_dose, --sum prima_dose alias first_dose
SUM(seconda_dose) AS second_dose, --sum seconda_dose alias second_dose
SUM(prima_dose)+SUM(seconda_dose) AS total_doses
--sum sums of prima_dose and seconda_dose alias total_doses
FROM base2 --from somministrazioni_vaccini_latest
GROUP BY nome_area, codice_regione_ISTAT, fornitore
--group each region name, region code and vaccine name
ORDER BY codice_regione_ISTAT,fornitore
--order from the smallest to the biggest value by default codice_regione_ISTAT then fornitore
```

region	reg_cod	vaccine	first_dose	second_dose	total_doses
PIE	1	Janssen	61154	0	61154
PIE	1	Moderna	205615	107356	312971
PIE	1	Pfizer/BioNTech	1556982	825061	2382043
PIE	1	Vaxzevria (AstraZ...	418037	99623	517660
VDA	2	Janssen	895	0	895
VDA	2	Moderna	4631	2753	7384
VDA	2	Pfizer/BioNTech	42962	25129	68091
VDA	2	Vaxzevria (AstraZ...	13513	3105	16618
LOM	3	Janssen	259282	0	259282
LOM	3	Moderna	486711	258043	744754
LOM	3	Pfizer/BioNTech	3492640	1989337	5481977
LOM	3	Vaxzevria (AstraZ...	1233489	246163	1479652
VEN	5	Janssen	57713	0	57713
VEN	5	Moderna	229544	125815	355359
VEN	5	Pfizer/BioNTech	1744092	888741	2632833
VEN	5	Vaxzevria (AstraZ...	498191	158902	657093
FVG	6	Janssen	21934	0	21934
FVG	6	Moderna	60573	24530	85103
FVG	6	Pfizer/BioNTech	426328	234287	660615
FVG	6	Vaxzevria (AstraZ...	141246	36299	177545
LIG	7	Janssen	31602	0	31602
LIG	7	Moderna	72977	46197	119174
LIG	7	Pfizer/BioNTech	517325	341943	859268

LIG	7	Vaxzevria (AstraZ...	178664	30117	208781
EMR	8	Janssen	82961	0	82961
EMR	8	Moderna	212518	123508	336026
EMR	8	Pfizer/BioNTech	1536364	892506	2428870
EMR	8	Vaxzevria (AstraZ...	454322	131198	585520
TOS	9	Janssen	59949	0	59949
TOS	9	Moderna	184381	87997	272378
TOS	9	Pfizer/BioNTech	1310023	592226	1902249
TOS	9	Vaxzevria (AstraZ...	368765	133713	502478
UMB	10	Janssen	15877	0	15877
UMB	10	Moderna	42670	20559	63229
UMB	10	Pfizer/BioNTech	309638	144726	454364
UMB	10	Vaxzevria (AstraZ...	99366	28767	128133
MAR	11	Janssen	22084	0	22084
MAR	11	Moderna	74723	42420	117143
MAR	11	Pfizer/BioNTech	543537	278398	821935
MAR	11	Vaxzevria (AstraZ...	167022	42326	209348
LAZ	12	Janssen	157459	0	157459
LAZ	12	Moderna	245901	127604	373505
LAZ	12	Pfizer/BioNTech	2004734	1158003	3162737
LAZ	12	Vaxzevria (AstraZ...	683553	248545	932098
ABR	13	Janssen	21663	0	21663
ABR	13	Moderna	59982	33078	93060
ABR	13	Pfizer/BioNTech	447728	249054	696782
ABR	13	Vaxzevria (AstraZ...	166997	39892	206889
MOL	14	Janssen	5058	0	5058
MOL	14	Moderna	13858	6867	20725
MOL	14	Pfizer/BioNTech	113462	55709	169171
MOL	14	Vaxzevria (AstraZ...	34142	9052	43194
CAM	15	Janssen	105239	0	105239
CAM	15	Moderna	321896	122330	444226
CAM	15	Pfizer/BioNTech	2051047	986288	3037335
CAM	15	Vaxzevria (AstraZ...	665694	184006	849700
PUG	16	Janssen	101654	0	101654
PUG	16	Moderna	188587	94227	282814
PUG	16	Pfizer/BioNTech	1415286	748905	2164191
PUG	16	Vaxzevria (AstraZ...	454774	160734	615508
BAS	17	Janssen	9418	0	9418
BAS	17	Moderna	19366	16525	35891
BAS	17	Pfizer/BioNTech	183159	118765	301924
BAS	17	Vaxzevria (AstraZ...	58638	10371	69009
CAL	18	Janssen	48716	0	48716
CAL	18	Moderna	79743	44674	124417
CAL	18	Pfizer/BioNTech	604316	348331	952647
CAL	18	Vaxzevria (AstraZ...	196890	28884	225774
SIC	19	Janssen	100322	0	100322
SIC	19	Moderna	227419	112564	339983
SIC	19	Pfizer/BioNTech	1671210	875574	2546784
SIC	19	Vaxzevria (AstraZ...	319305	148063	467368
SAR	20	Janssen	14632	0	14632
SAR	20	Moderna	80873	47012	127885

SAR	20	Pfizer/BioNTech	560884	266007	826891
SAR	20	Vaxzevria (AstraZ...	191605	42607	234212
PAB	21	Janssen	3640	0	3640
PAB	21	Moderna	25655	14001	39656
PAB	21	Pfizer/BioNTech	166569	107670	274239
PAB	21	Vaxzevria (AstraZ...	60285	21166	81451
PAT	22	Janssen	3238	0	3238
PAT	22	Moderna	24294	8650	32944
PAT	22	Pfizer/BioNTech	182538	123655	306193
PAT	22	Vaxzevria (AstraZ...	37197	20469	57666

Query 28: show the number of doses administered to males, females and the total number

Sum sesso_maschile, sum sesso_femminile and sum the two sums.

```
SELECT
SUM(sesso_maschile) AS males, --sum sesso_maschile alias males
SUM(sesso_femminile) AS females, --sum sesso_femminile alias females
SUM(sesso_maschile)+SUM(sesso_femminile) AS combined
--sum sums of sesso_maschile and sesso_femminile alias combined
FROM base2 -- from somministrazioni_vaccini_latest
```

males	females	combined
21069306	24840647	45909953

Query 29: show for each region the number of doses administered to males, females and the total number

Group rows by nome_area, sum sesso_maschile, sum sesso_femminile and sum the two sums, ordered by total_doses_received.

```
SELECT
nome_area AS region, --nome_area alias region
SUM(sesso_maschile) AS total_doses_received_males,
--sum sesso_maschile alias total_doses_received_males
SUM(sesso_femminile) AS total_doses_received_females,
 --sum sesso_femminile alias total_doses_received_females
SUM(sesso_maschile)+SUM(sesso_femminile) AS total_doses_received
--sum of sums of sesso_maschile and sesso_femminile alias total_doses_received
FROM base2 --from somministrazioni_vaccini_latest
GROUP BY nome_area --group each region name
ORDER BY total_doses_received DESC --order from the biggest to the smallest value
```

region	total_doses_received_males	total_doses_received_females	total_doses_received
LOM	3618669	4346996	7965665
LAZ	2112939	2512860	4625799
CAM	2099258	2337242	4436500
VEN	1693047	2009951	3702998
SIC	1629288	1825169	3454457
EMR	1547329	1886048	3433377
PIE	1478136	1795692	3273828
PUG	1455086	1709081	3164167
TOS	1231611	1505443	2737054
CAL	647063	704491	1351554
LIG	554711	664114	1218825
SAR	553187	650433	1203620
MAR	535519	634991	1170510
ABR	465755	552639	1018394
FVG	432990	512207	945197
UMB	294927	366676	661603
BAS	193307	222935	416242
PAT	182313	217728	400041
PAB	190533	208453	398986
MOL	110052	128096	238148
VDA	43586	49402	92988

Query 30: show for each age group the number of first doses, second doses and total doses done

Group rows by fascia_anagrafica, Sum prima_dose, sum seconda_dose and sum the two sums, order by fascia_anagrafica.

```
SELECT
fascia_anagrafica AS age_group, --fascia_anagrafica alias age_group
SUM(prima_dose) AS first_dose, --sum prima_dose alias first_dose
SUM(seconda_dose) AS second_dose, --sum seconda_dose alias second_dose
SUM(prima_dose)+SUM(seconda_dose) AS total_doses
--sum sums of prima_dose and seconda_dose alias total_doses
FROM base2 --from somministrazioni_vaccini_latest
GROUP BY fascia_anagrafica --group each age group
ORDER BY fascia_anagrafica
--order from the smallest to the biggest value by default
```

age_group	first_dose	second_dose	total_doses
12-19	487806	50371	538177
20-29	1859481	644337	2503818
30-39	2466575	901535	3368110
40-49	4589595	1380886	5970481
50-59	6624991	2375986	9000977
60-69	5973750	2659379	8633129
70-79	5180607	2674298	7854905
80-89	3400378	3147437	6547815
90+	785743	706798	1492541

Query 31: show for each region, for each age group the number of first doses, second doses and total doses done

Group rows by nome_area, fascia_anagrafica, sum prima_dose, sum seconda_dose and sum the two sums, order by nome_area, fascia_anagrafica.

```
SELECT
nome_area AS region, --nome_area alias region
fascia_anagrafica AS age_group, --fascia_anagrafica alias age_group
SUM(prima_dose) AS first_dose, --sum prima_dose alias first_dose
SUM(seconda_dose) AS second_dose, --sum seconda_dose alias second_dose
SUM(prima_dose)+SUM(seconda_dose) AS total_doses
--sum sums of prima_dose and seconda_dose alias total_doses
FROM base2 --from somministrazioni_vaccini_latest
GROUP BY nome_area, fascia_anagrafica
--group each region name then each age group
ORDER BY nome_area, fascia_anagrafica
--order alphabetically region names then from the smallest to the biggest value the age groups
```

region	age_group	first_dose	second_dose	total_doses
ABR	12-19	13252	899	14151
ABR	20-29	45153	12680	57833
ABR	30-39	53025	20197	73222
ABR	40-49	99648	32342	131990
ABR	50-59	137496	50976	188472
ABR	60-69	135818	58065	193883
ABR	70-79	116852	59250	176102
ABR	80-89	76496	70980	147476
ABR	90+	18630	16635	35265
BAS	12-19	4296	1130	5426
BAS	20-29	14477	7824	22301
BAS	30-39	18607	11197	29804
BAS	40-49	36955	17478	54433
BAS	50-59	52799	24202	77001
BAS	60-69	57049	22085	79134
BAS	70-79	46479	24289	70768
BAS	80-89	32046	30175	62221
BAS	90+	7873	7281	15154
CAL	12-19	24210	1464	25674
CAL	20-29	76142	16232	92374
CAL	30-39	87856	27365	115221
CAL	40-49	134876	41601	176477
CAL	50-59	176103	66685	242788
CAL	60-69	178658	88794	267452
CAL	70-79	144082	82605	226687

CAL	80-89	88434	80228	168662
CAL	90+	19304	16915	36219
CAM	12-19	96723	6024	102747
CAM	20-29	323731	63398	387129
CAM	30-39	339612	90337	429949
CAM	40-49	506473	135617	642090
CAM	50-59	642783	323590	966373
CAM	60-69	546565	224747	771312
CAM	70-79	420818	210898	631716
CAM	80-89	224453	201924	426377
CAM	90+	42718	36089	78807
EMR	12-19	26333	3503	29836
EMR	20-29	106389	49965	156354
EMR	30-39	150950	68713	219663
EMR	40-49	290113	103077	393190
EMR	50-59	499963	168424	668387
EMR	60-69	450015	226033	676048
EMR	70-79	402259	185634	587893
EMR	80-89	285577	272481	558058
EMR	90+	74566	69382	143948
FVG	12-19	8702	889	9591
FVG	20-29	44806	16474	61280
FVG	30-39	48386	20034	68420
FVG	40-49	95245	31961	127206
FVG	50-59	129216	60172	189388
FVG	60-69	114035	34006	148041
FVG	70-79	112691	39857	152548
FVG	80-89	78170	74327	152497
FVG	90+	18830	17396	36226
LAZ	12-19	49936	2646	52582
LAZ	20-29	130575	47596	178171
LAZ	30-39	199618	80343	279961
LAZ	40-49	512294	128346	640640
LAZ	50-59	706278	254590	960868
LAZ	60-69	602115	294116	896231
LAZ	70-79	497425	360071	857496
LAZ	80-89	321733	301150	622883
LAZ	90+	71673	65294	136967
LIG	12-19	12452	984	13436
LIG	20-29	49292	15556	64848
LIG	30-39	57003	19239	76242
LIG	40-49	82433	29617	112050
LIG	50-59	152607	59423	212030
LIG	60-69	151255	57023	208278
LIG	70-79	151254	101440	252694
LIG	80-89	115335	109414	224749
LIG	90+	28937	25561	54498
LOM	12-19	46283	12705	58988
LOM	20-29	266891	132577	399468
LOM	30-39	494649	168951	663600
LOM	40-49	841061	248440	1089501

LOM	50-59	1222464	382050	1604514
LOM	60-69	1005885	491756	1497641
LOM	70-79	877853	380199	1258052
LOM	80-89	586749	555753	1142502
LOM	90+	130287	121112	251399
MAR	12-19	16351	972	17323
MAR	20-29	56446	14249	70695
MAR	30-39	61420	21767	83187
MAR	40-49	104373	34346	138719
MAR	50-59	153928	51337	205265
MAR	60-69	152576	60005	212581
MAR	70-79	137423	68884	206307
MAR	80-89	99587	89102	188689
MAR	90+	25262	22482	47744
MOL	12-19	3160	169	3329
MOL	20-29	7699	2695	10394
MOL	30-39	12061	4287	16348
MOL	40-49	23556	7301	30857
MOL	50-59	34566	11503	46069
MOL	60-69	34091	12904	46995
MOL	70-79	27273	10404	37677
MOL	80-89	18923	17794	36717
MOL	90+	5191	4571	9762
PAB	12-19	4591	507	5098
PAB	20-29	22135	6774	28909
PAB	30-39	25286	9154	34440
PAB	40-49	40037	16855	56892
PAB	50-59	54011	36930	90941
PAB	60-69	43635	25236	68871
PAB	70-79	36734	20702	57436
PAB	80-89	24193	21824	46017
PAB	90+	5527	4855	10382
PAT	12-19	745	297	1042
PAT	20-29	9198	5534	14732
PAT	30-39	11066	6667	17733
PAT	40-49	29620	9708	39328
PAT	50-59	59562	25107	84669
PAT	60-69	53970	35392	89362
PAT	70-79	45986	36388	82374
PAT	80-89	29140	26474	55614
PAT	90+	7980	7207	15187
PIE	12-19	17462	2878	20340
PIE	20-29	112696	46589	159285
PIE	30-39	158751	60878	219629
PIE	40-49	296596	88314	384910
PIE	50-59	469833	133177	603010
PIE	60-69	431209	194979	626188
PIE	70-79	402595	178983	581578
PIE	80-89	287280	268364	555644
PIE	90+	65366	57878	123244
PUG	12-19	39877	3494	43371

PUG	20-29	107097	42288	149385
PUG	30-39	161332	63633	224965
PUG	40-49	334280	104979	439259
PUG	50-59	458824	161935	620759
PUG	60-69	430939	191421	622360
PUG	70-79	364758	194941	559699
PUG	80-89	215256	197998	413254
PUG	90+	47938	43177	91115
SAR	12-19	9110	1770	10880
SAR	20-29	41954	14115	56069
SAR	30-39	56612	24490	81102
SAR	40-49	138653	39628	178281
SAR	50-59	176764	55471	232235
SAR	60-69	172811	62015	234826
SAR	70-79	145340	62736	208076
SAR	80-89	87470	79121	166591
SAR	90+	19280	16280	35560
SIC	12-19	65440	4498	69938
SIC	20-29	210895	52100	262995
SIC	30-39	225854	72527	298381
SIC	40-49	345171	117364	462535
SIC	50-59	441767	219315	661082
SIC	60-69	418743	210597	629340
SIC	70-79	351077	230200	581277
SIC	80-89	214504	191844	406348
SIC	90+	44805	37756	82561
TOS	12-19	21822	2118	23940
TOS	20-29	98319	40159	138478
TOS	30-39	144698	54617	199315
TOS	40-49	214992	79325	294317
TOS	50-59	373494	102928	476422
TOS	60-69	387242	90555	477797
TOS	70-79	361855	147279	509134
TOS	80-89	255683	236457	492140
TOS	90+	65013	60498	125511
UMB	12-19	5304	573	5877
UMB	20-29	19728	9327	29055
UMB	30-39	26456	14367	40823
UMB	40-49	57126	21968	79094
UMB	50-59	100944	30406	131350
UMB	60-69	93615	22840	116455
UMB	70-79	86640	21576	108216
UMB	80-89	61393	57815	119208
UMB	90+	16345	15180	31525
VDA	12-19	572	95	667
VDA	20-29	3498	1023	4521
VDA	30-39	3828	1517	5345
VDA	40-49	7358	2454	9812
VDA	50-59	14372	4791	19163
VDA	60-69	12380	5172	17552
VDA	70-79	11137	7764	18901

VDA	80-89	7155	6589	13744
VDA	90+	1701	1582	3283
VEN	12-19	21185	2756	23941
VEN	20-29	112360	47182	159542
VEN	30-39	129505	61255	190760
VEN	40-49	398735	90165	488900
VEN	50-59	567217	152974	720191
VEN	60-69	501144	251638	752782
VEN	70-79	440076	250198	690274
VEN	80-89	290801	257623	548424
VEN	90+	68517	59667	128184

BIBLIOGRAPHY

Notebook:
https://colab.research.google.com/drive/1Xeetx08UP_g_w6Sbobc3bFU7lFj4gU0N?usp=sharing
Protezione Civile Github Repository:
https://github.com/pcm-dpc/COVID-19
Raw dpc_covid19_ita_regioni:
https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni.csv
Government Github Repository:
https://github.com/italia/covid19-opendata-vaccini
Raw somministrazioni_vaccini_latest dataset:
https://raw.githubusercontent.com/italia/covid19-opendata-vaccini/master/dati/somministrazioni-vaccini-latest.csv