

Pandas COVID19 Practice - Hints - Unibs 2021

In this exercise, we will analyze some public data provided by the "Dipartimento della Protezione Civile". To inform the citizens and give the reached data, useful only to communicate and informate, Dipartimento della Protezione Civile has elaborated a dashboard reachable to the URL <http://arcg.is/C1unv> (desktop version) and <http://arcg.is/081a51> (mobile version) and give to everyone, under the licence CC-BY-4.0, the following infos updated daily at 18.30:

- National evolution data
- json data
- Regional data
- Provincial data
- Summaries
- Areas
- Notes
- Contracts data DPC furnitures
- Metrics

See <https://github.com/pcm-dpc/COVID-19>

Import pandas package under name `pd` and print version

```
In [ ]: import matplotlib.pyplot as plt
```

```
hint: import ... as, pd.__version__
```

```
In [ ]: import pandas as pd
```

```
In [ ]: pd.__version__
```

```
Out[ ]: '2.0.3'
```

Base

Download latest COVID19 csv of italian regions from official Italian Civil Protection github and show it

CSV data at: <https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni-latest.csv>

```
hint: pd.read_csv, sep=","
```

```
In [ ]: covid_data = pd.read_csv(  
        "https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-  
        sep=", "  
        )
```

```
In [ ]: print(covid_data)
```

	data	stato	codice_regione	denominazione_regione	\
0	2024-04-03T17:00:00	ITA	13	Abruzzo	
1	2024-04-03T17:00:00	ITA	17	Basilicata	
2	2024-04-03T17:00:00	ITA	18	Calabria	
3	2024-04-03T17:00:00	ITA	15	Campania	
4	2024-04-03T17:00:00	ITA	8	Emilia-Romagna	
5	2024-04-03T17:00:00	ITA	6	Friuli Venezia Giulia	
6	2024-04-03T17:00:00	ITA	12	Lazio	
7	2024-04-03T17:00:00	ITA	7	Liguria	
8	2024-04-03T17:00:00	ITA	3	Lombardia	
9	2024-04-03T17:00:00	ITA	11	Marche	
10	2024-04-03T17:00:00	ITA	14	Molise	
11	2024-04-03T17:00:00	ITA	21	P.A. Bolzano	
12	2024-04-03T17:00:00	ITA	22	P.A. Trento	
13	2024-04-03T17:00:00	ITA	1	Piemonte	
14	2024-04-03T17:00:00	ITA	16	Puglia	
15	2024-04-03T17:00:00	ITA	20	Sardegna	
16	2024-04-03T17:00:00	ITA	19	Sicilia	
17	2024-04-03T17:00:00	ITA	9	Toscana	
18	2024-04-03T17:00:00	ITA	10	Umbria	
19	2024-04-03T17:00:00	ITA	2	Valle d'Aosta	
20	2024-04-03T17:00:00	ITA	5	Veneto	

	lat	long	ricoverati_con_sintomi	terapia_intensiva	\
0	42.351222	13.398438	3	0	
1	40.639471	15.805148	0	0	
2	38.905976	16.594402	0	0	
3	40.839566	14.250850	70	3	
4	44.494367	11.341721	69	2	
5	45.649435	13.768136	3	0	
6	41.892770	12.483667	205	6	
7	44.411493	8.932699	2	0	
8	45.466794	9.190347	21	0	
9	43.616760	13.518875	4	0	
10	41.557748	14.659161	0	0	
11	46.499335	11.356624	1	0	
12	46.068935	11.121231	2	0	
13	45.073274	7.680687	48	1	
14	41.125596	16.867367	1	0	
15	39.215312	9.110616	23	1	
16	38.115697	13.362357	72	4	
17	43.769231	11.255889	12	0	
18	43.106758	12.388247	13	0	
19	45.737503	7.320149	2	0	
20	45.434905	12.338452	181	5	

	totale_ospedalizzati	isolamento_domiciliare	...	\
0	3	8440	...	
1	0	9874	...	
2	0	3431	...	
3	73	7276	...	
4	71	87	...	
5	3	275	...	
6	211	60441	...	
7	2	0	...	
8	21	223	...	
9	4	0	...	
10	0	2	...	
11	1	0	...	
12	2	4	...	

13	49	54006	...
14	1	91	...
15	24	9485	...
16	76	910	...
17	12	169	...
18	13	519	...
19	2	5	...
20	186	9552	...

	note	\
0	Il dato ''incremento casi confermati'' è compo...	
1	Il dato relativo al numero dei "Casi in isolam...	
2		NaN
3		NaN
4	Il dato odierno dei casi per la provincia, ris...	
5		NaN
6		NaN
7		NaN
8		NaN
9		NaN
10		NaN
11	0 nuovi positivi di cui 0 test antigenici conf...	
12		NaN
13		NaN
14		NaN
15		NaN
16		NaN
17		NaN
18	Si fa presente che dei 13 ricoveri non in T.I....	
19		NaN
20	Si segnala che è presente un delta negativo re...	

	ingressi_terapia_intensiva	note_test	\
0	0	NaN	
1	0	NaN	
2	0	NaN	
3	0	NaN	
4	0	NaN	
5	0	NaN	
6	0	NaN	
7	0	NaN	
8	0	NaN	
9	0	NaN	
10	0	NaN	
11	0	NaN	
12	0	NaN	
13	0	NaN	
14	0	NaN	
15	0	NaN	
16	0	NaN	
17	0	NaN	
18	0	NaN	
19	0	NaN	
20	0	NaN	

	note_casi	\
0	NaN	
1	NaN	
2	NaN	
3	NaN	

4	NaN
5	NaN
6	NaN
7	*di cui 61033 reinfezioni a partire dal 3/09/...
8	NaN
9	NaN
10	NaN
11	0 nuovi positivi di cui 0 test antigenici conf...
12	NaN
13	NaN
14	A seguito di verifiche, è stato eliminato un c...
15	NaN
16	NaN
17	NaN
18	NaN
19	NaN
20	NaN

	totale_positivi_test_molecolare	totale_positivi_test_antigenico_rapido	\
0	264190	424918	
1	71288	131249	
2	204361	450911	
3	959949	1588513	
4	1100133	1110398	
5	242484	356239	
6	1043026	1488387	
7	288791	401365	
8	1553114	2786503	
9	223417	513599	
10	31118	74454	
11	87646	213382	
12	45806	207664	
13	519880	1284275	
14	515236	1173083	
15	180340	346146	
16	543124	1291660	
17	732345	921320	
18	160678	298389	
19	16500	35971	
20	1150176	1691221	

	tamponi_test_molecolare	tamponi_test_antigenico_rapido	codice_nuts_1	\
0	2650397	5053546	ITF	
1	704040	671785	ITF	
2	1964365	2578566	ITF	
3	9756715	12163271	ITF	
4	10827943	9286122	ITH	
5	3895644	3993245	ITH	
6	9762086	17829492	ITI	
7	2714986	4434158	ITC	
8	17211174	30012512	ITC	
9	2020910	1784425	ITI	
10	555760	297885	ITF	
11	944722	4683817	ITH	
12	879836	2214657	ITH	
13	5152255	17406836	ITC	
14	4890999	9748274	ITF	
15	2247679	3337407	ITG	
16	5500120	11467697	ITG	
17	7147621	10190660	ITI	

18	1764365	3408820	ITI
19	146733	460264	ITC
20	12194201	27170459	ITH

	codice_nuts_2
0	ITF1
1	ITF5
2	ITF6
3	ITF3
4	ITH5
5	ITH4
6	ITI4
7	ITC3
8	ITC4
9	ITI3
10	ITF2
11	ITH1
12	ITH2
13	ITC1
14	ITF4
15	ITG2
16	ITG1
17	ITI1
18	ITI2
19	ITC2
20	ITH3

[21 rows x 30 columns]

Sort columns and show their types

```
hint: df.sort_index, axis=1, inplace=True, df.dtypes
```

```
In [ ]: covid_data.sort_index(axis=1, inplace=True)
```

```
In [ ]: print(covid_data.dtypes)
```

casi_da_screening	float64
casi_da_sospetto_diagnostico	float64
casi_testati	int64
codice_nuts_1	object
codice_nuts_2	object
codice_regione	int64
data	object
deceduti	int64
denominazione_regione	object
dimessi_guariti	int64
ingressi_terapia_intensiva	int64
isolamento_domiciliare	int64
lat	float64
long	float64
note	object
note_casi	object
note_test	float64
nuovi_positivi	int64
ricoverati_con_sintomi	int64
stato	object
tamponi	int64
tamponi_test_antigenico_rapido	int64
tamponi_test_molecolare	int64
terapia_intensiva	int64
totale_casi	int64
totale_ospedalizzati	int64
totale_positivi	int64
totale_positivi_test_antigenico_rapido	int64
totale_positivi_test_molecolare	int64
variazione_totale_positivi	int64
dtype:	object

Print first and last five rows of the data

hint: `df.head, df.tail`

```
In [ ]: print(covid_data.head())
```

	casi_da_screening	casi_da_sospetto_diagnostico	casi_testati	\
0	NaN	NaN	1393003	
1	NaN	NaN	410982	
2	NaN	NaN	3557762	
3	NaN	NaN	5532262	
4	NaN	NaN	3004494	

	codice_nuts_1	codice_nuts_2	codice_regione	data	deceduti	\
0	ITF	ITF1	13	2024-04-03T17:00:00	4092	
1	ITF	ITF5	17	2024-04-03T17:00:00	1055	
2	ITF	ITF6	18	2024-04-03T17:00:00	3676	
3	ITF	ITF3	15	2024-04-03T17:00:00	12142	
4	ITH	ITH5	8	2024-04-03T17:00:00	20035	

	denominazione_regione	dimessi_guariti	...	tamponi	\
0	Abruzzo	676573	...	7703943	
1	Basilicata	191608	...	1375825	
2	Calabria	648165	...	4542931	
3	Campania	2528971	...	21919986	
4	Emilia-Romagna	2190338	...	20114065	

	tamponi_test_antigenico_rapido	tamponi_test_molecolare	terapia_intensiva	\
0	5053546	2650397	0	
1	671785	704040	0	
2	2578566	1964365	0	
3	12163271	9756715	3	
4	9286122	10827943	2	

	totale_casi	totale_ospedalizzati	totale_positivi	\
0	689108	3	8443	
1	202537	0	9874	
2	655272	0	3431	
3	2548462	73	7349	
4	2210531	71	158	

	totale_positivi_test_antigenico_rapido	totale_positivi_test_molecolare	\
0	424918	264190	
1	131249	71288	
2	450911	204361	
3	1588513	959949	
4	1110398	1100133	

	variazione_totale_positivi
0	1
1	0
2	-2
3	-1409
4	-25

[5 rows x 30 columns]

In []: `print(covid_data.tail())`

	casi_da_screening	casi_da_sospetto_diagnostico	casi_testati	\
16	NaN	NaN	11303927	
17	NaN	NaN	5472005	
18	NaN	NaN	829222	
19	NaN	NaN	148734	
20	NaN	NaN	5449067	

	codice_nuts_1	codice_nuts_2	codice_regione	data	deceduti	\
16	ITG	ITG1	19	2024-04-03T17:00:00	13056	
17	ITI	ITI1	9	2024-04-03T17:00:00	12501	
18	ITI	ITI2	10	2024-04-03T17:00:00	2532	
19	ITC	ITC2	2	2024-04-03T17:00:00	591	
20	ITH	ITH3	5	2024-04-03T17:00:00	17486	

	denominazione_regione	dimessi_guariti	...	tamponi	\
16	Sicilia	1820742	...	16967817	
17	Toscana	1640983	...	17338281	
18	Umbria	456003	...	5173185	
19	Valle d'Aosta	51873	...	606997	
20	Veneto	2814173	...	39364660	

	tamponi_test_antigenico_rapido	tamponi_test_molecolare	\
16	11467697	5500120	
17	10190660	7147621	
18	3408820	1764365	
19	460264	146733	
20	27170459	12194201	

	terapia_intensiva	totale_casi	totale_ospedalizzati	totale_positivi	\
16	4	1834784	76	986	
17	0	1653665	12	181	
18	0	459067	13	532	
19	0	52471	2	7	
20	5	2841397	186	9738	

	totale_positivi_test_antigenico_rapido	totale_positivi_test_molecolare	\
16	1291660	543124	
17	921320	732345	
18	298389	160678	
19	35971	16500	
20	1691221	1150176	

	variazione_totale_positivi
16	-6
17	0
18	5
19	0
20	16

[5 rows x 30 columns]

**Print synthetic statistical description of the dataframe
(count, min, max, mean, etc.)**

hint: `df.describe`

```
In [ ]: print(covid_data.describe())
```

	casi_da_screening	casi_da_sospetto_diagnostico	casi_testati \
count	0.0	0.0	2.100000e+01
mean	NaN	NaN	4.538268e+06
std	NaN	NaN	6.172326e+06
min	NaN	NaN	1.487340e+05
25%	NaN	NaN	8.710820e+05
50%	NaN	NaN	3.004494e+06
75%	NaN	NaN	5.472005e+06
max	NaN	NaN	2.800304e+07

	codice_regione	deceduti	dimessi_guariti \
count	21.000000	21.000000	2.100000e+01
mean	11.857143	9356.523810	1.255296e+06
std	6.428730	10629.040279	1.123045e+06
min	1.000000	591.000000	5.187300e+04
25%	7.000000	2532.000000	4.560030e+05
50%	12.000000	6018.000000	6.841360e+05
75%	17.000000	13056.000000	1.820742e+06
max	22.000000	47886.000000	4.291487e+06

	ingressi_terapia_intensiva	isolamento_domiciliare	lat \
count	21.0	21.000000	21.000000
mean	0.0	7847.142857	43.046293
std	0.0	16870.901663	2.550241
min	0.0	0.000000	38.115697
25%	0.0	5.000000	41.125596
50%	0.0	275.000000	43.616760
75%	0.0	8440.000000	45.434905
max	0.0	60441.000000	46.499335

	long ...	tamponi	tamponi_test_antigenico_rapido \
count	21.000000 ...	2.100000e+01	2.100000e+01
mean	12.225955 ...	1.338697e+07	8.485424e+06
std	2.724611 ...	1.281471e+07	8.466755e+06
min	7.320149 ...	6.069970e+05	2.978850e+05
25%	11.121231 ...	4.542931e+06	2.578566e+06
50%	12.388247 ...	7.703943e+06	4.683817e+06
75%	13.768136 ...	2.011406e+07	1.146770e+07
max	16.867367 ...	4.722369e+07	3.001251e+07

	tamponi_test_molecolare	terapia_intensiva	totale_casi \
count	2.100000e+01	21.000000	2.100000e+01
mean	4.901550e+06	1.047619	1.272536e+06
std	4.626607e+06	1.856777	1.137433e+06
min	1.467330e+05	0.000000	5.247100e+04
25%	1.764365e+06	0.000000	4.590670e+05
50%	2.714986e+06	0.000000	6.901560e+05
75%	7.147621e+06	1.000000	1.834784e+06
max	1.721117e+07	6.000000	4.339617e+06

	totale_ospedalizzati	totale_positivi \
count	21.000000	21.000000
mean	35.904762	7883.047619
std	59.987419	16906.928528
min	0.000000	1.000000
25%	2.000000	7.000000
50%	4.000000	278.000000
75%	49.000000	8443.000000
max	211.000000	60652.000000

	totale_positivi_test_antigenico_rapido \	
count	2.100000e+01	
mean	7.995070e+05	
std	7.061507e+05	
min	3.597100e+04	
25%	2.983890e+05	
50%	4.509110e+05	
75%	1.284275e+06	
max	2.786503e+06	

	totale_positivi_test_molecolare	variazione_totale_positivi
count	2.100000e+01	21.000000
mean	4.730287e+05	-71.047619
std	4.472728e+05	306.921077
min	1.650000e+04	-1409.000000
25%	1.606780e+05	-2.000000
50%	2.641900e+05	0.000000
75%	7.323450e+05	1.000000
max	1.553114e+06	16.000000

[8 rows x 23 columns]

Count elements for each column

hint: `df.count`

```
In [ ]: print(covid_data.count())
```

```
casi_da_screening          0
casi_da_sospetto_diagnostico 0
casi_testati               21
codice_nuts_1              21
codice_nuts_2              21
codice_regione             21
data                      21
deceduti                   21
denominazione_regione      21
dimessi_guariti            21
ingressi_terapia_intensiva  21
isolamento_domiciliare    21
lat                        21
long                       21
note                       6
note_casi                  3
note_test                  0
nuovi_positivi             21
ricoverati_con_sintomi     21
stato                      21
tamponi                    21
tamponi_test_antigenico_rapido 21
tamponi_test_molecolare    21
terapia_intensiva          21
totale_casi                21
totale_ospedalizzati       21
totale_positivi            21
totale_positivi_test_antigenico_rapido 21
totale_positivi_test_molecolare 21
variazione_totale_positivi 21
dtype: int64
```

Select only "totale_positivi" and "nuovi_positivi" columns

hint: df[]

```
In [ ]: print(covid_data[['totale_positivi', 'nuovi_positivi']])
```

	totale_positivi	nuovi_positivi
0	8443	1
1	9874	0
2	3431	0
3	7349	12
4	158	4
5	278	4
6	60652	15
7	2	5
8	244	29
9	4	1
10	2	0
11	1	0
12	6	5
13	54055	6
14	92	7
15	9509	2
16	986	0
17	181	14
18	532	6
19	7	0
20	9738	29

Create the new column "precedenti_positivi" columns using the formula

$precedenti_positivi = totale_positivi - nuovi_positivi$
and show it

hint: df[] = df[] - df[]

```
In [ ]: covid_data['precedenti_positivi']=covid_data['totale_positivi']-covid_data['nuov  
print(covid_data)
```

```
File "<ipython-input-236-42ca4314ba33>", line 1
    covid_data['precedenti_positivi']=
    ^
SyntaxError: invalid syntax
```

Select only rows from 5 to 7

hint: df.loc[]

```
In [ ]: print(covid_data.loc[5:7])
```

	casi_da_screening	casi_da_sospetto_diagnostico	casi_testati	\
5	NaN	NaN	1280274	
6	NaN	NaN	6278624	
7	NaN	NaN	1550304	

	codice_nuts_1	codice_nuts_2	codice_region	data	deceduti	\
5	ITH	ITH4	6	2024-04-03T17:00:00	6452	
6	ITI	ITI4	12	2024-04-03T17:00:00	13267	
7	ITC	ITC3	7	2024-04-03T17:00:00	6018	

	denominazione_region	dimessi_guariti	...	\
5	Friuli Venezia Giulia	591993	...	
6	Lazio	2457494	...	
7	Liguria	684136	...	

	tamponi_test_antigenico_rapido	tamponi_test_molecolare	terapia_intensiva	\
5	3993245	3895644	0	
6	17829492	9762086	6	
7	4434158	2714986	0	

	totale_casi	totale_ospedalizzati	totale_positivi	\
5	598723	3	278	
6	2531413	211	60652	
7	690156	2	2	

	totale_positivi_test_antigenico_rapido	totale_positivi_test_molecolare	\
5	356239	242484	
6	1488387	1043026	
7	401365	288791	

	variazione_totale_positivi	precedenti_positivi
5	1	274
6	11	60637
7	-1	-3

[3 rows x 31 columns]

Select only "totale_positivi" and "nuovi_positivi" columns
and only rows from 5 to 7

hint: `df[], df.loc[]`

```
In [ ]: print(covid_data[['totale_positivi', 'nuovi_positivi']].loc[5:7])
```

	totale_positivi	nuovi_positivi
5	278	4
6	60652	15
7	2	5

Set "denominazione_region" as index and show it

hint: `df.set_index, inplace=True`

```
In [ ]: covid_data.set_index('denominazione_region', inplace=True)
print(covid_data)
```

	casi_da_screening	casi_da_sospetto_diagnostico	\
denominazione_regione			
Abruzzo	NaN	NaN	
Basilicata	NaN	NaN	
Calabria	NaN	NaN	
Campania	NaN	NaN	
Emilia-Romagna	NaN	NaN	
Friuli Venezia Giulia	NaN	NaN	
Lazio	NaN	NaN	
Liguria	NaN	NaN	
Lombardia	NaN	NaN	
Marche	NaN	NaN	
Molise	NaN	NaN	
P.A. Bolzano	NaN	NaN	
P.A. Trento	NaN	NaN	
Piemonte	NaN	NaN	
Puglia	NaN	NaN	
Sardegna	NaN	NaN	
Sicilia	NaN	NaN	
Toscana	NaN	NaN	
Umbria	NaN	NaN	
Valle d'Aosta	NaN	NaN	
Veneto	NaN	NaN	

	casi_testati	codice_nuts_1	codice_nuts_2	\
denominazione_regione				
Abruzzo	1393003	ITF	ITF1	
Basilicata	410982	ITF	ITF5	
Calabria	3557762	ITF	ITF6	
Campania	5532262	ITF	ITF3	
Emilia-Romagna	3004494	ITH	ITH5	
Friuli Venezia Giulia	1280274	ITH	ITH4	
Lazio	6278624	ITI	ITI4	
Liguria	1550304	ITC	ITC3	
Lombardia	9468091	ITC	ITC4	
Marche	28003043	ITI	ITI3	
Molise	804360	ITF	ITF2	
P.A. Bolzano	871082	ITH	ITH1	
P.A. Trento	610253	ITH	ITH2	
Piemonte	4532363	ITC	ITC1	
Puglia	3036106	ITF	ITF4	
Sardegna	1767670	ITG	ITG2	
Sicilia	11303927	ITG	ITG1	
Toscana	5472005	ITI	ITI1	
Umbria	829222	ITI	ITI2	
Valle d'Aosta	148734	ITC	ITC2	
Veneto	5449067	ITH	ITH3	

	codice_regione	data	deceduti	\
denominazione_regione				
Abruzzo	13	2024-04-03T17:00:00	4092	
Basilicata	17	2024-04-03T17:00:00	1055	
Calabria	18	2024-04-03T17:00:00	3676	
Campania	15	2024-04-03T17:00:00	12142	
Emilia-Romagna	8	2024-04-03T17:00:00	20035	
Friuli Venezia Giulia	6	2024-04-03T17:00:00	6452	
Lazio	12	2024-04-03T17:00:00	13267	
Liguria	7	2024-04-03T17:00:00	6018	
Lombardia	3	2024-04-03T17:00:00	47886	
Marche	11	2024-04-03T17:00:00	4560	

Molise	14	2024-04-03T17:00:00	798
P.A. Bolzano	21	2024-04-03T17:00:00	1670
P.A. Trento	22	2024-04-03T17:00:00	1681
Piemonte	1	2024-04-03T17:00:00	13928
Puglia	16	2024-04-03T17:00:00	10084
Sardegna	20	2024-04-03T17:00:00	2977
Sicilia	19	2024-04-03T17:00:00	13056
Toscana	9	2024-04-03T17:00:00	12501
Umbria	10	2024-04-03T17:00:00	2532
Valle d'Aosta	2	2024-04-03T17:00:00	591
Veneto	5	2024-04-03T17:00:00	17486

	dimessi_guariti	ingressi_terapia_intensiva	...	\
denominazione_regione			...	
Abruzzo	676573		0	...
Basilicata	191608		0	...
Calabria	648165		0	...
Campania	2528971		0	...
Emilia-Romagna	2190338		0	...
Friuli Venezia Giulia	591993		0	...
Lazio	2457494		0	...
Liguria	684136		0	...
Lombardia	4291487		0	...
Marche	732452		0	...
Molise	104772		0	...
P.A. Bolzano	299357		0	...
P.A. Trento	251783		0	...
Piemonte	1736172		0	...
Puglia	1678143		0	...
Sardegna	514000		0	...
Sicilia	1820742		0	...
Toscana	1640983		0	...
Umbria	456003		0	...
Valle d'Aosta	51873		0	...
Veneto	2814173		0	...

	tamponi_test_antigenico_rapido	\
denominazione_regione		
Abruzzo	5053546	
Basilicata	671785	
Calabria	2578566	
Campania	12163271	
Emilia-Romagna	9286122	
Friuli Venezia Giulia	3993245	
Lazio	17829492	
Liguria	4434158	
Lombardia	30012512	
Marche	1784425	
Molise	297885	
P.A. Bolzano	4683817	
P.A. Trento	2214657	
Piemonte	17406836	
Puglia	9748274	
Sardegna	3337407	
Sicilia	11467697	
Toscana	10190660	
Umbria	3408820	
Valle d'Aosta	460264	
Veneto	27170459	

denominazione_regione	tamponi_test_molecolare	terapia_intensiva	totale_casi \
Abruzzo	2650397	0	689108
Basilicata	704040	0	202537
Calabria	1964365	0	655272
Campania	9756715	3	2548462
Emilia-Romagna	10827943	2	2210531
Friuli Venezia Giulia	3895644	0	598723
Lazio	9762086	6	2531413
Liguria	2714986	0	690156
Lombardia	17211174	0	4339617
Marche	2020910	0	737016
Molise	555760	0	105572
P.A. Bolzano	944722	0	301028
P.A. Trento	879836	0	253470
Piemonte	5152255	1	1804155
Puglia	4890999	0	1688319
Sardegna	2247679	1	526486
Sicilia	5500120	4	1834784
Toscana	7147621	0	1653665
Umbria	1764365	0	459067
Valle d'Aosta	146733	0	52471
Veneto	12194201	5	2841397

denominazione_regione	totale_ospedalizzati	totale_positivi \
Abruzzo	3	8443
Basilicata	0	9874
Calabria	0	3431
Campania	73	7349
Emilia-Romagna	71	158
Friuli Venezia Giulia	3	278
Lazio	211	60652
Liguria	2	2
Lombardia	21	244
Marche	4	4
Molise	0	2
P.A. Bolzano	1	1
P.A. Trento	2	6
Piemonte	49	54055
Puglia	1	92
Sardegna	24	9509
Sicilia	76	986
Toscana	12	181
Umbria	13	532
Valle d'Aosta	2	7
Veneto	186	9738

denominazione_regione	totale_positivi_test_antigenico_rapido \
Abruzzo	424918
Basilicata	131249
Calabria	450911
Campania	1588513
Emilia-Romagna	1110398
Friuli Venezia Giulia	356239
Lazio	1488387
Liguria	401365
Lombardia	2786503
Marche	513599

Molise	74454
P.A. Bolzano	213382
P.A. Trento	207664
Piemonte	1284275
Puglia	1173083
Sardegna	346146
Sicilia	1291660
Toscana	921320
Umbria	298389
Valle d'Aosta	35971
Veneto	1691221

denominazione_regione	totale_positivi_test_molecolare \
Abruzzo	264190
Basilicata	71288
Calabria	204361
Campania	959949
Emilia-Romagna	1100133
Friuli Venezia Giulia	242484
Lazio	1043026
Liguria	288791
Lombardia	1553114
Marche	223417
Molise	31118
P.A. Bolzano	87646
P.A. Trento	45806
Piemonte	519880
Puglia	515236
Sardegna	180340
Sicilia	543124
Toscana	732345
Umbria	160678
Valle d'Aosta	16500
Veneto	1150176

denominazione_regione	variazione_totale_positivi	precedenti_positivi
Abruzzo	1	8442
Basilicata	0	9874
Calabria	-2	3431
Campania	-1409	7337
Emilia-Romagna	-25	154
Friuli Venezia Giulia	1	274
Lazio	11	60637
Liguria	-1	-3
Lombardia	-38	215
Marche	0	3
Molise	0	2
P.A. Bolzano	-2	1
P.A. Trento	1	1
Piemonte	-47	54049
Puglia	2	85
Sardegna	1	9507
Sicilia	-6	986
Toscana	0	167
Umbria	5	526
Valle d'Aosta	0	7
Veneto	16	9709

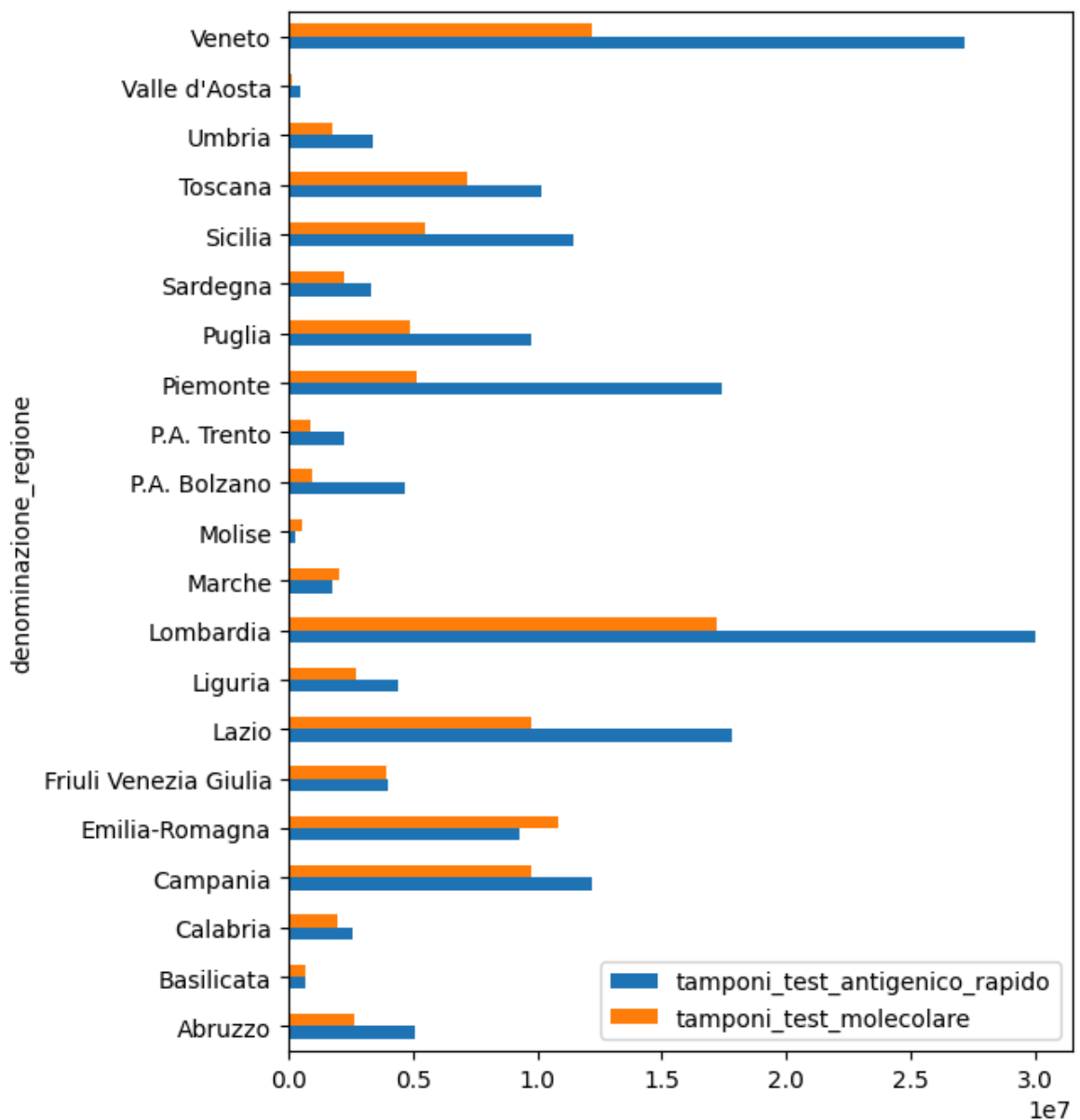
[21 rows x 30 columns]

Plot horizontal bars of "tamponi_test_antigenico_rapido" and "tamponi_test_molecolare" columns for each region

hint: `df[].plot.barh, figsize=()`

- Figure size of (6,8)

```
In [ ]: covid_data[['tamponi_test_antigenico_rapido',  
                  'tamponi_test_molecolare']].plot.barh(figsize=(6, 8))  
plt.show()
```

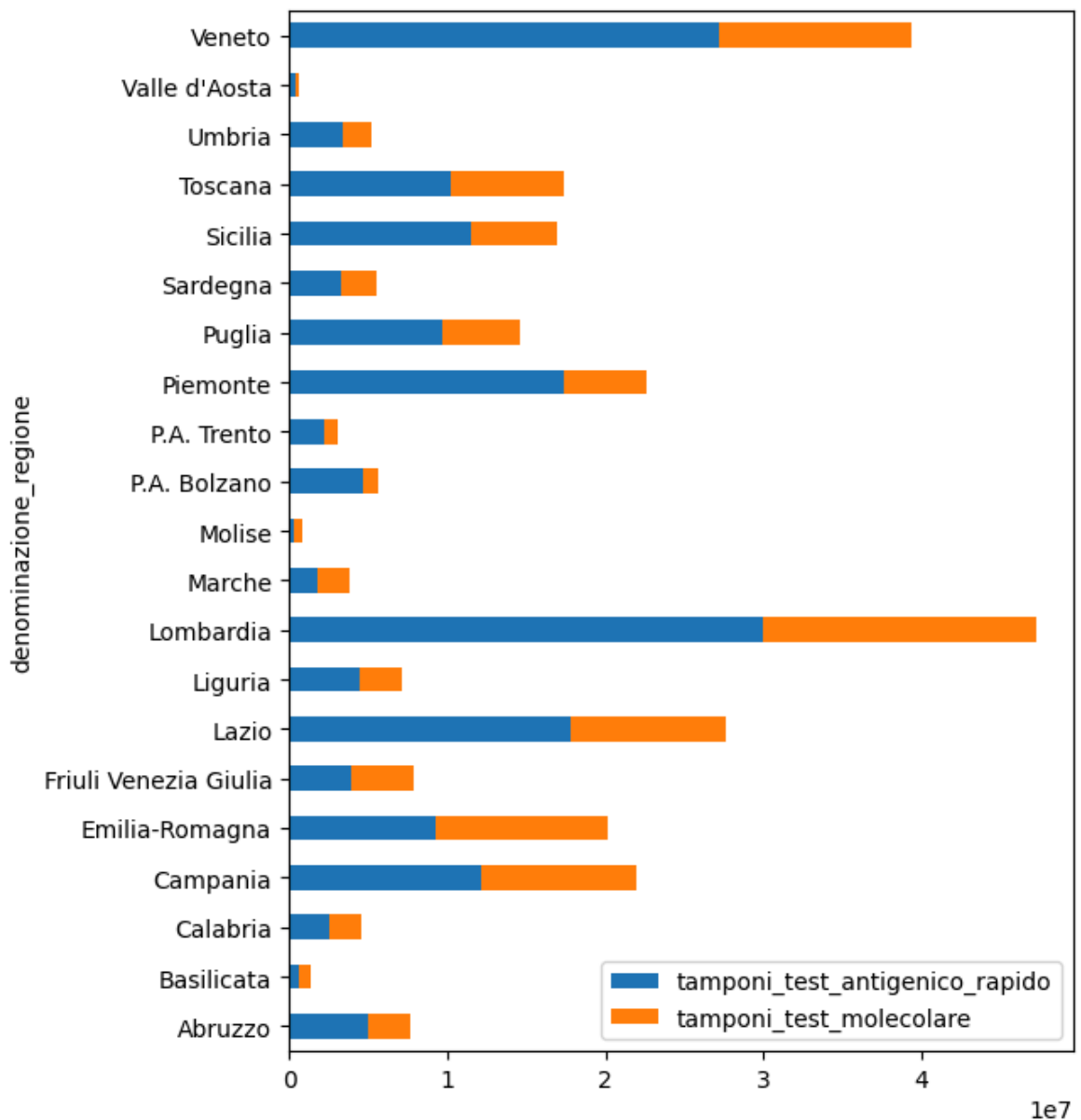


Plot horizontal bars of "tamponi_test_antigenico_rapido" and "tamponi_test_molecolare" stacked together

hint: `stacked=True`

- Figure size of (6,8)

```
In [ ]: covid_data[['tamponi_test_antigenico_rapido',
                    'tamponi_test_molecolare']].plot.barh(stacked=True, figsize=(6, 8))
plt.show()
```

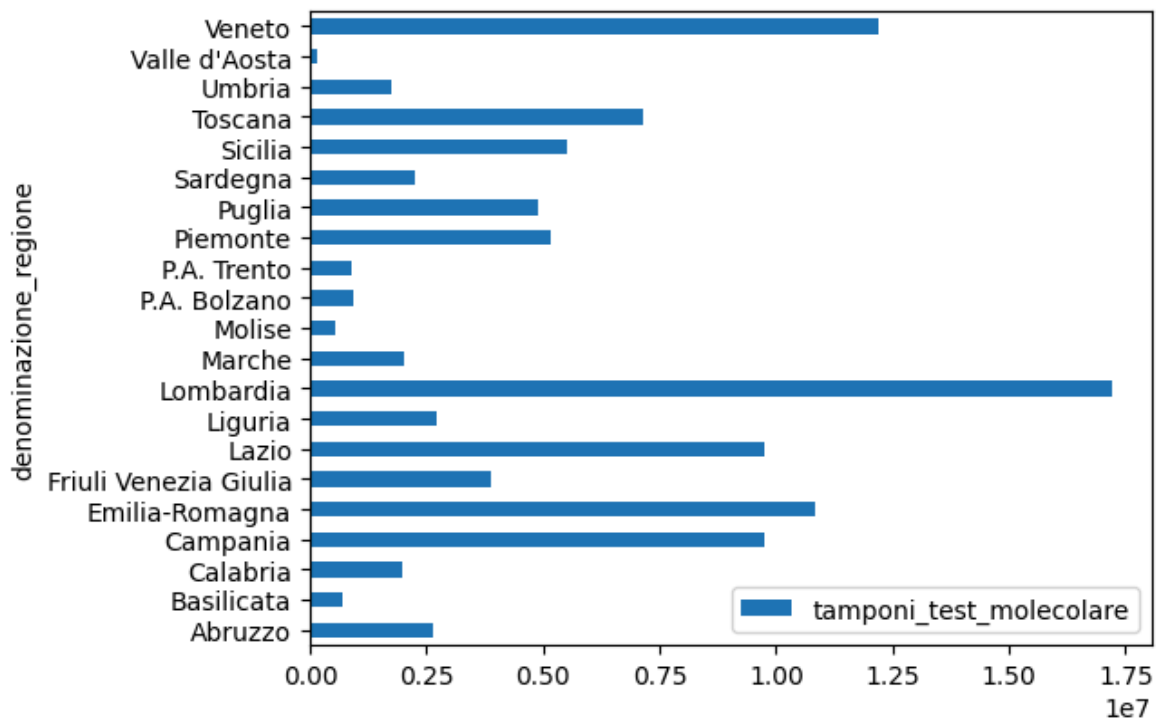
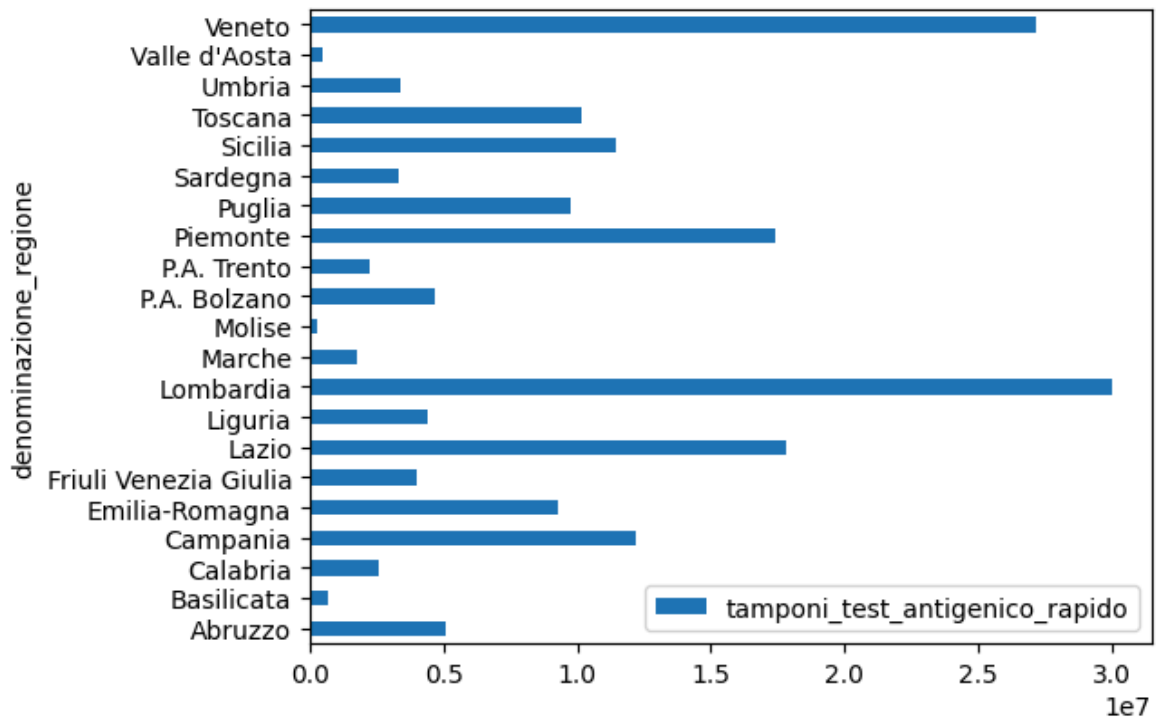


Plot horizontal bars of "tamponi_test_antigenico_rapido" and "tamponi_test_molecolare" columns for each region in different subplots

hint: `df[].plot.barh, figsize=()`

- Figure size of (6,10)

```
In [ ]: fig, axs = plt.subplots(2, 1, figsize=(6, 10))
covid_data[['tamponi_test_antigenico_rapido']].plot.barh(ax=axs[0], stacked=True)
covid_data[['tamponi_test_molecolare']].plot.barh(ax=axs[1], stacked=True)
plt.show()
```



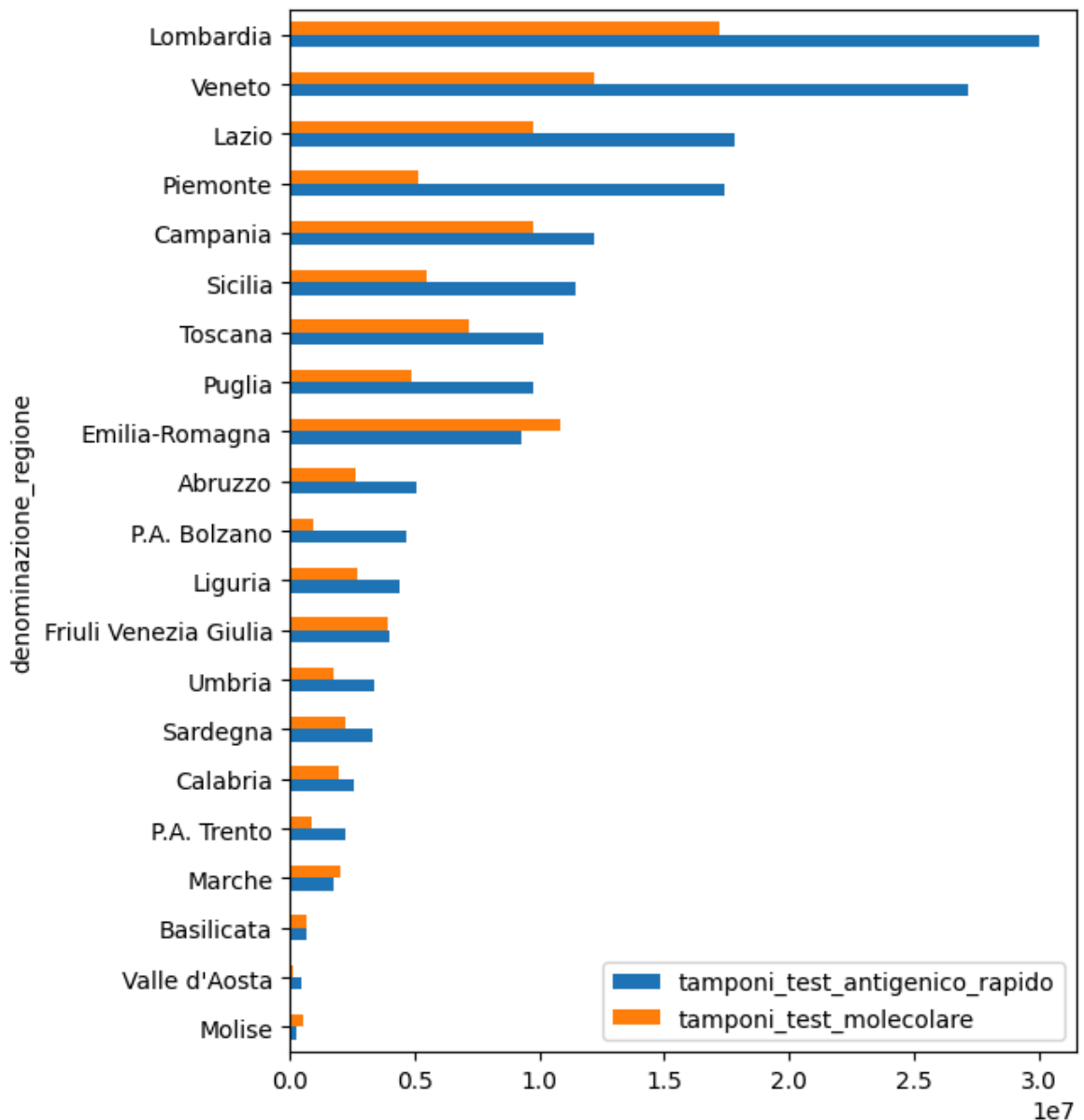
Plot horizontal bars of "tamponi_test_antigenico_rapido" and "tamponi_test_molecolare" columns for each region sorting by "tamponi_test_antigenico_rapido" column

hint: `df.sort_values`

- Figure size of (6,8)

```
In [ ]: covid_data[['tamponi_test_antigenico_rapido',
                  'tamponi_test_molecolare']].sort_values(
                  by='tamponi_test_antigenico_rapido').plot.barh(
```

```
figsize=(6, 8))
plt.show()
```



Intermediate

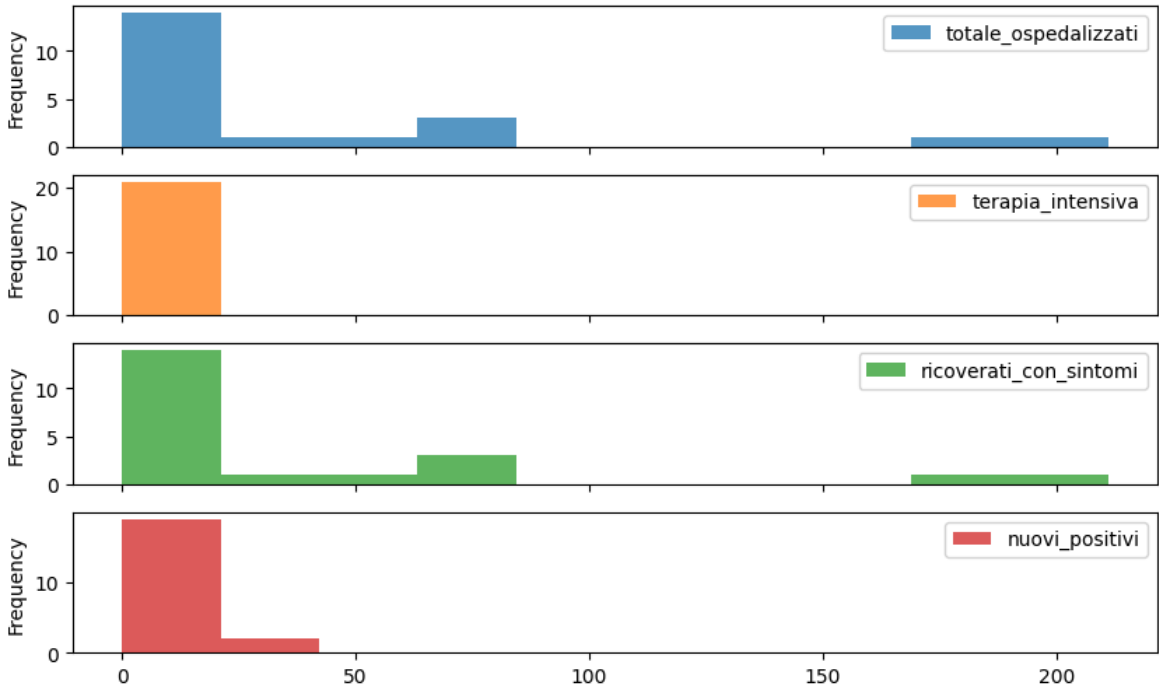
Plot histograms of "totale_ospedalizzati", "terapia_intensiva", "ricoverati_con_sintomi", "nuovi_positivi" in different subplots

```
hint: df[].plot.hist, bins=, alpha=
```

- Figure size of (10, 6)
- 10 Bins
- Alpha of 0.75

```
In [ ]: bins=10
alpha=0.75
covid_data[['totale_ospedalizzati',
```

```
'terapia_intensiva',  
'ricoverati_con_sintomi',  
'nuovi_positivi']] .plot .hist(  
    figsize=(10,6),  
    bins=bins,  
    alpha=alpha,  
    subplots=True)  
  
plt.show()
```

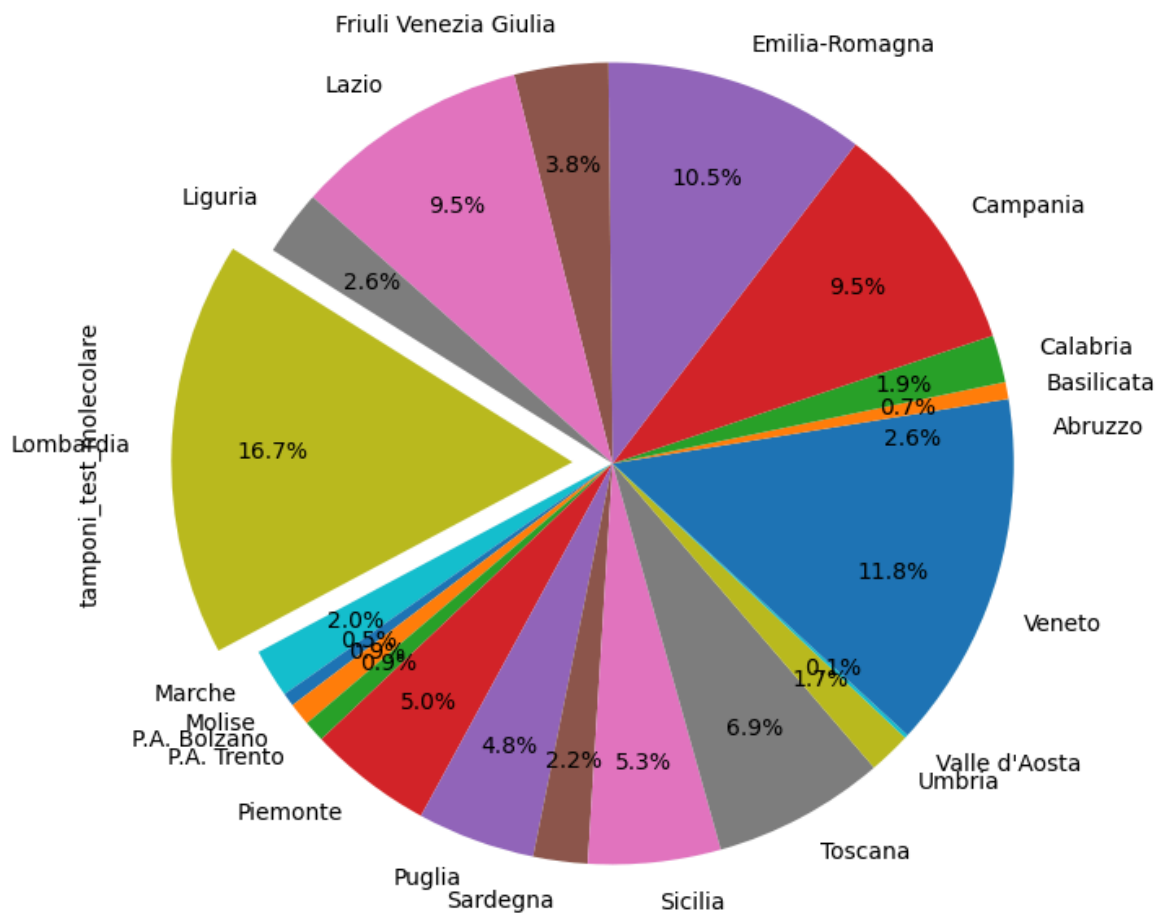


Plot in pie charts the number of "tamponi_test_molecolare" with percentage for each region with exploded slice for *Lombardia* region

```
hint: df.index, df[].plot.pie, figsize=, autopct="%1.1f%",
pctdistance=, explode=, ylabel=,
```

- Figure size of (8,8)
- Explode of 0.1 for Lombardia region
- Distance of the percentage of 0.75

[illegible]



Plot hexbin plot of "deceduti" by ("long", "lat) coordinates of regions

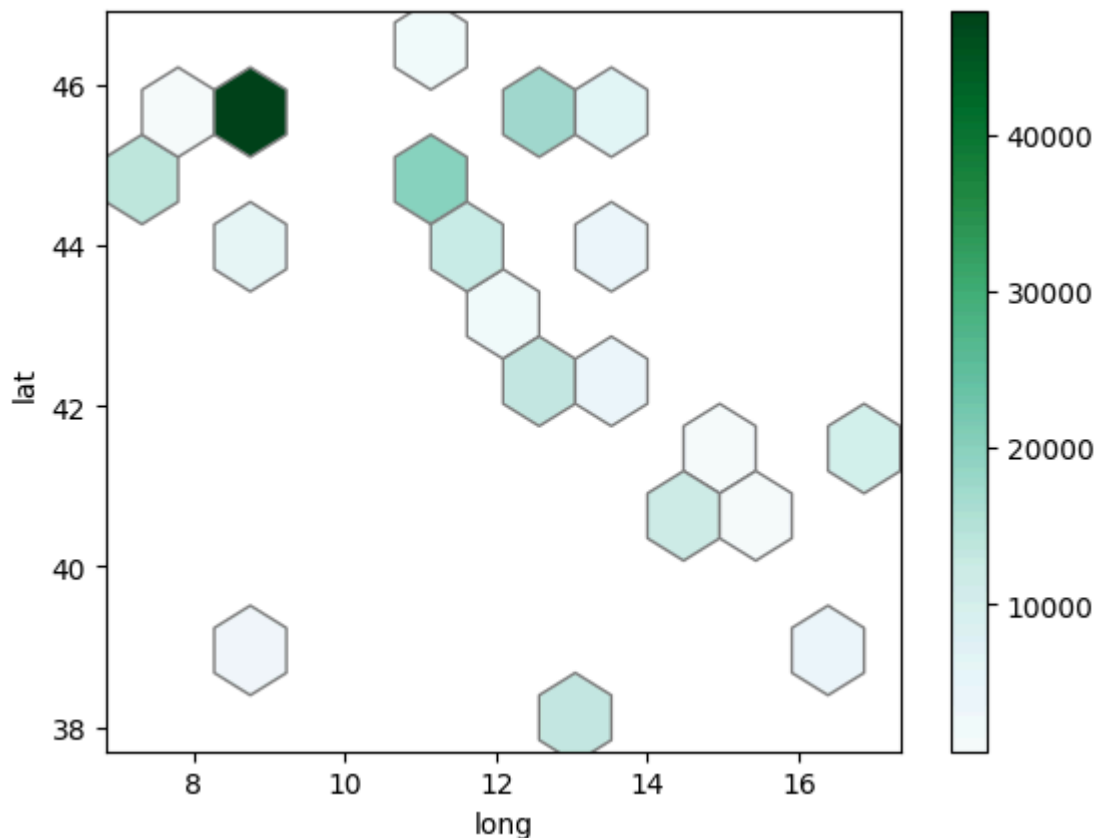
hint: df.plot.hexbin, x=, y=, C=, edgecolor=, gridsize=

- Figure size of (8, 6)
- Gridsize of 10

```
In [ ]: plt.figure(figsize=(8, 6))
covid_data.plot.hexbin(x='long',
                       y='lat',
                       C='deceduti',
                       edgecolor='gray',
                       gridsize=10)

plt.show()
```

<Figure size 800x600 with 0 Axes>



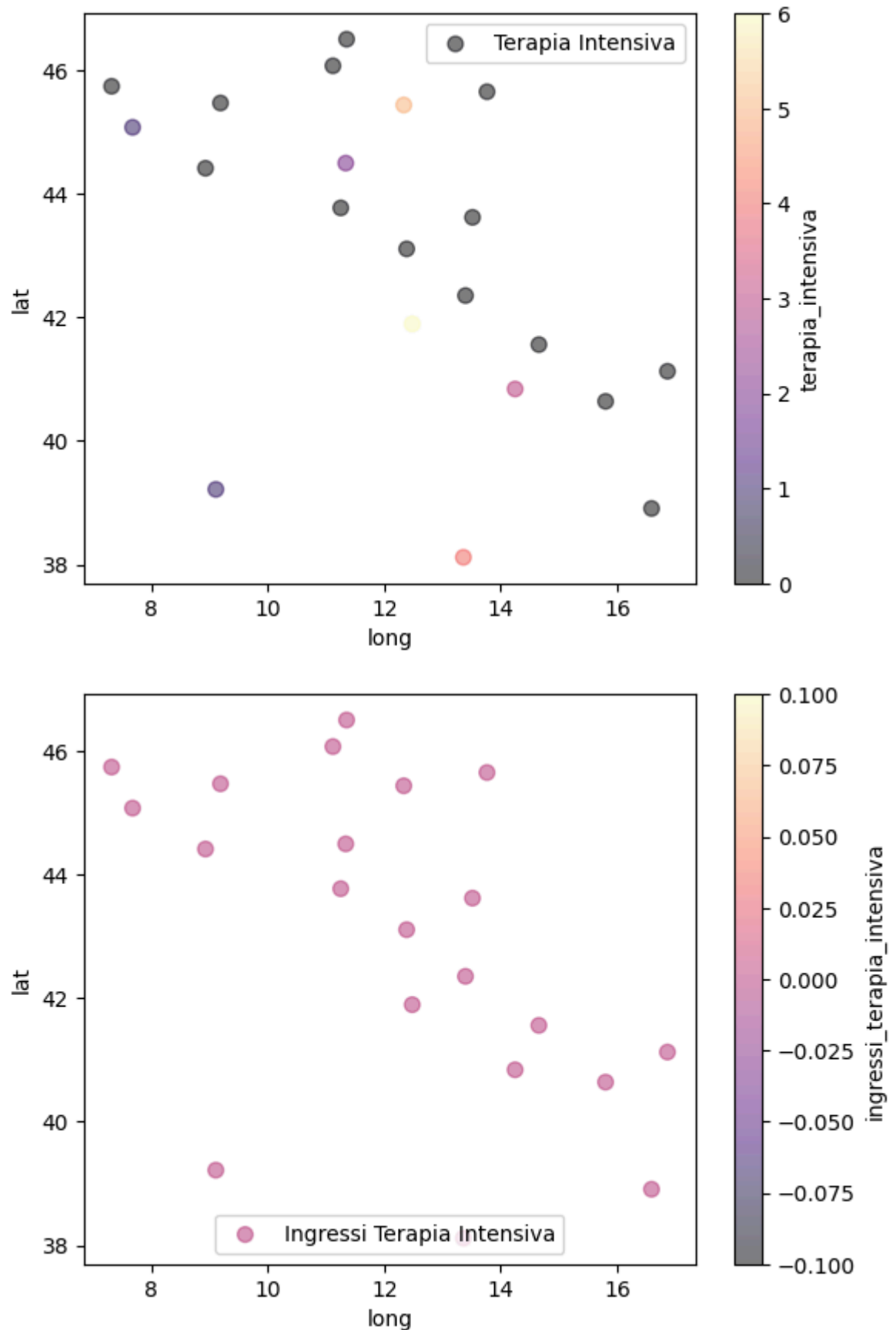
Plot scatterplots of "terapia_intensiva" and "ingressi_terapia_intensiva" by ("long", "lat") coordinates of regions both in the same plot

hint: `df.plot.scatter, x=, y=, color=, alpha=, s=, label=, ax=, ax.legend`

- Figure size of (8, 6)
- Alpha of 0.5
- Legend at the bottom

```
In [ ]: plt.figure(figsize=(8, 6))
alpha = 0.5
covid_data.plot.scatter(x='long',
                        y='lat',
                        c='terapia_intensiva',
                        colormap='magma',
                        alpha=alpha,
                        s=50,
                        label='Terapia Intensiva')
covid_data.plot.scatter(x='long',
                        y='lat',
                        c='ingressi_terapia_intensiva',
                        colormap='magma',
                        alpha=alpha,
                        s=50,
                        label='Ingressi Terapia Intensiva')
plt.legend(loc='lower center')
plt.show()
```


<Figure size 800x600 with 0 Axes>



Group the regions by color and plot bars of the mean value of "nuovi_positivi" column

hint: `pd.Series`, `df.groupby().mean().plot.bar()`, `title=`, `rot=`,

- Figure size (8, 6)

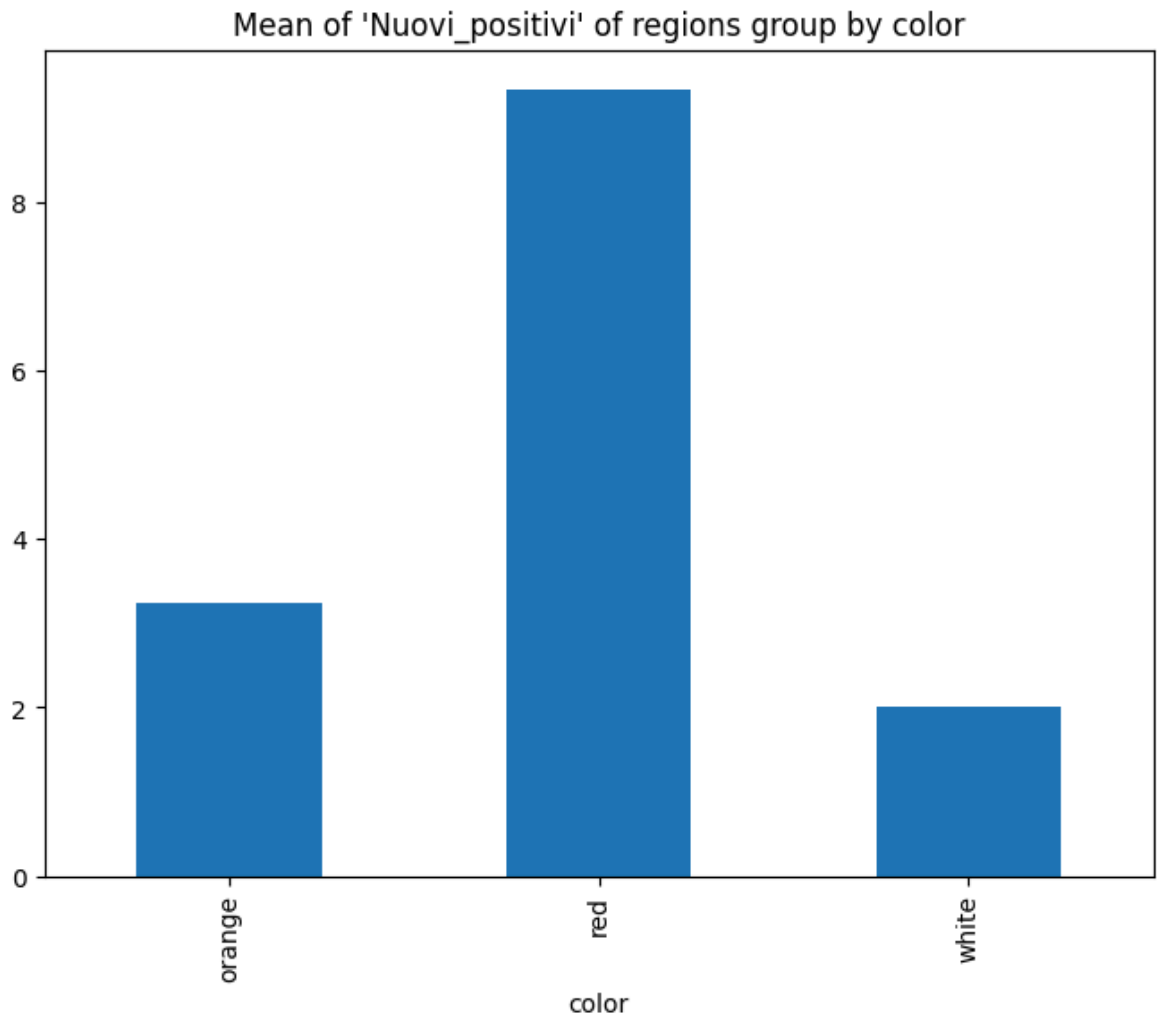
```
regions_colors = { "Abruzzo": "orange", "Basilicata": "orange", "Calabria": "orange",
"Campania": "red", "Emilia-Romagna": "red", "Friuli Venezia Giulia": "red", "Lazio": "red",
"Liguria": "orange", "Lombardia": "red", "Marche": "red", "Molise": "red", "P.A. Bolzano":
"orange", "P.A. Trento": "red", "Piemonte": "red", "Puglia": "red", "Sardegna": "white",
"Sicilia": "red", "Toscana": "orange", "Umbria": "orange", "Valle d'Aosta": "orange",
"Veneto": "red", }
```

```
In [ ]: regions_colors = { "Abruzzo": "orange", "Basilicata": "orange", "Calabria": "ora
colors_series = pd.Series(regions_colors,
                           name='color')
covid_data_colorized = covid_data.merge(colors_series,
                                         left_on='denominazione_regione',
                                         right_index=True)

print(covid_data_colorized['color'])
```

```
denominazione_regione
Abruzzo                orange
Basilicata              orange
Calabria               orange
Campania                red
Emilia-Romagna         red
Friuli Venezia Giulia  red
Lazio                  red
Liguria                orange
Lombardia               red
Marche                 red
Molise                  red
P.A. Bolzano           orange
P.A. Trento            red
Piemonte                red
Puglia                  red
Sardegna               white
Sicilia                red
Toscana                orange
Umbria                 orange
Valle d'Aosta          orange
Veneto                 red
Name: color, dtype: object
```

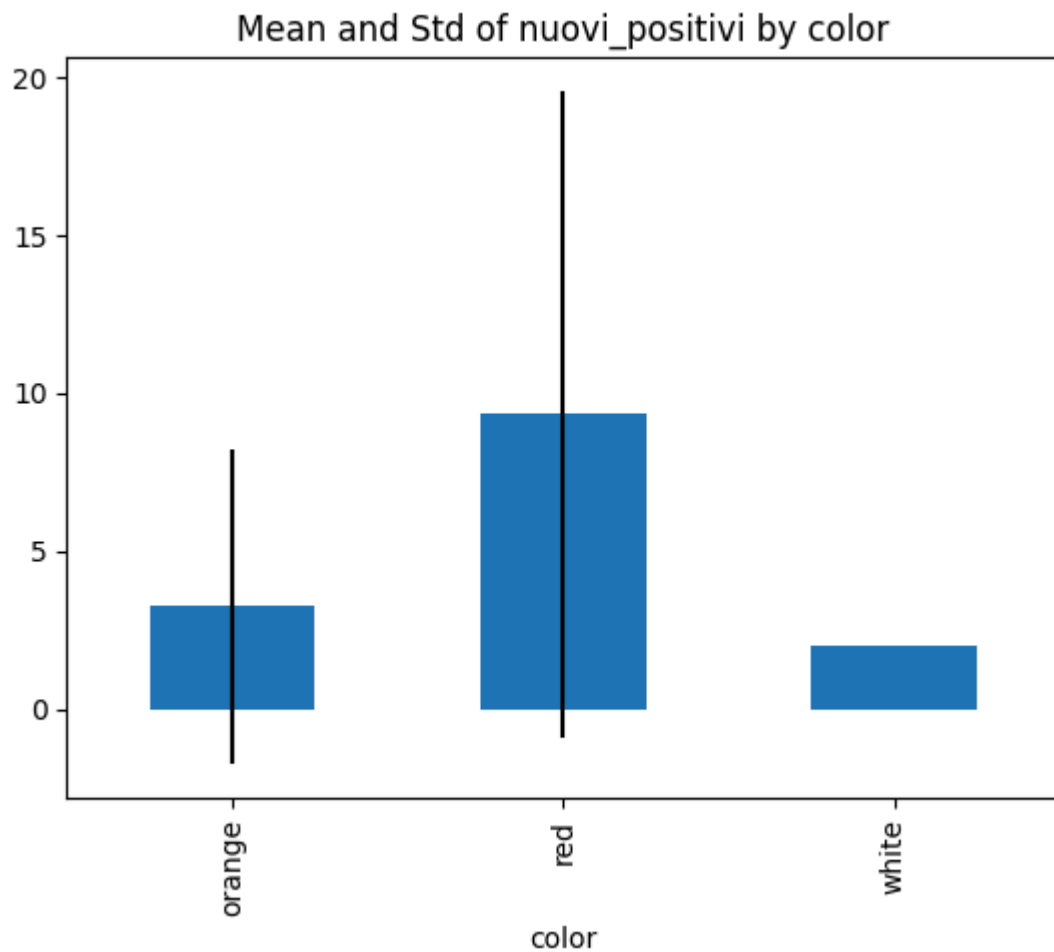
```
In [ ]: plt.figure(figsize=(8, 6))
covid_data_colorized.groupby('color')['nuovi_positivi'].mean().plot.bar(
    title="Mean of 'Nuovi_positivi' of regions group by color")
plt.show()
```



Group the regions by color and plot bars of mean *and error* of "nuovi_positivi" column

hint: `df.groupby().std()`, `yerr=`, `rot=`,

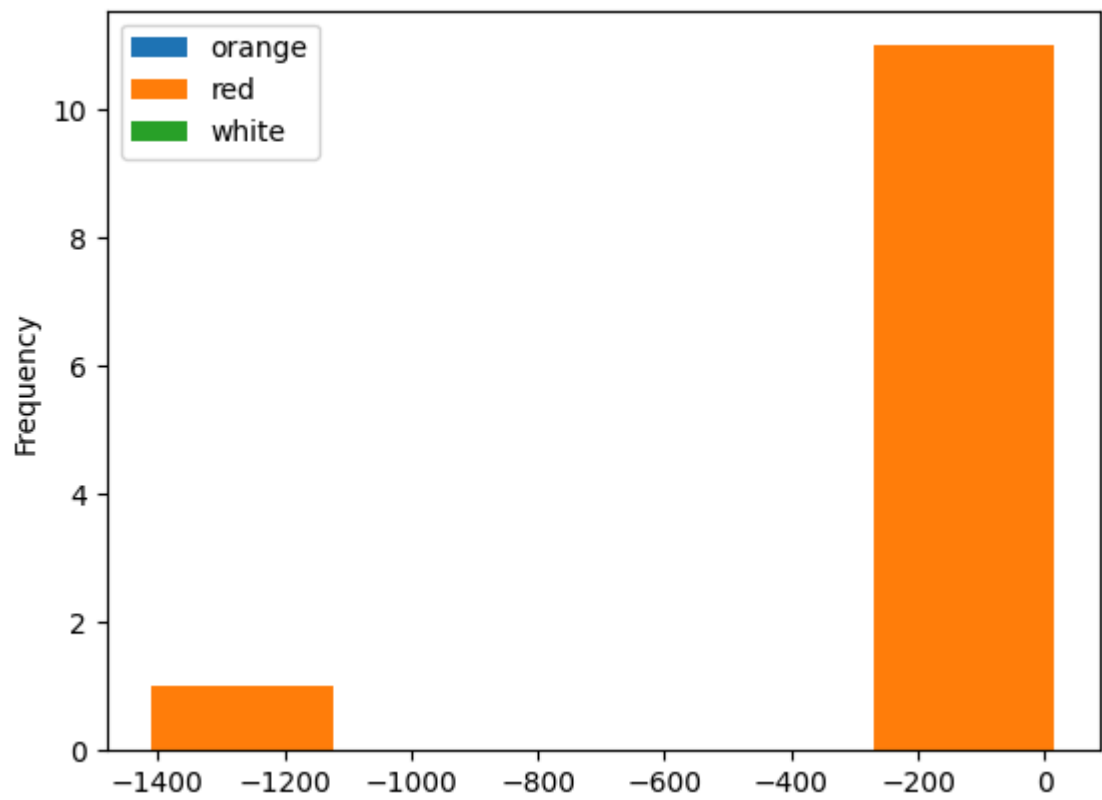
```
In [ ]: covid_data_colorized.groupby('color')['nuovi_positivi'].mean().plot.bar(
        yerr=covid_data_colorized.groupby('color')['nuovi_positivi'].std(),
        title='Mean and Std of nuovi_positivi by color')
plt.show()
```



Group the regions by color and plot a hist of "variazione_totale_positivi" using 5 bins

hint: `pd.Series, df.groupby().plot.hist(), legend=,`

```
In [ ]: covid_data_colorized.groupby('color')['variazione_totale_positivi'].plot.hist(  
        bins=5,  
        legend=True)  
plt.show()
```



Thanks to Matteo Olivato m.olivato@unibs.it