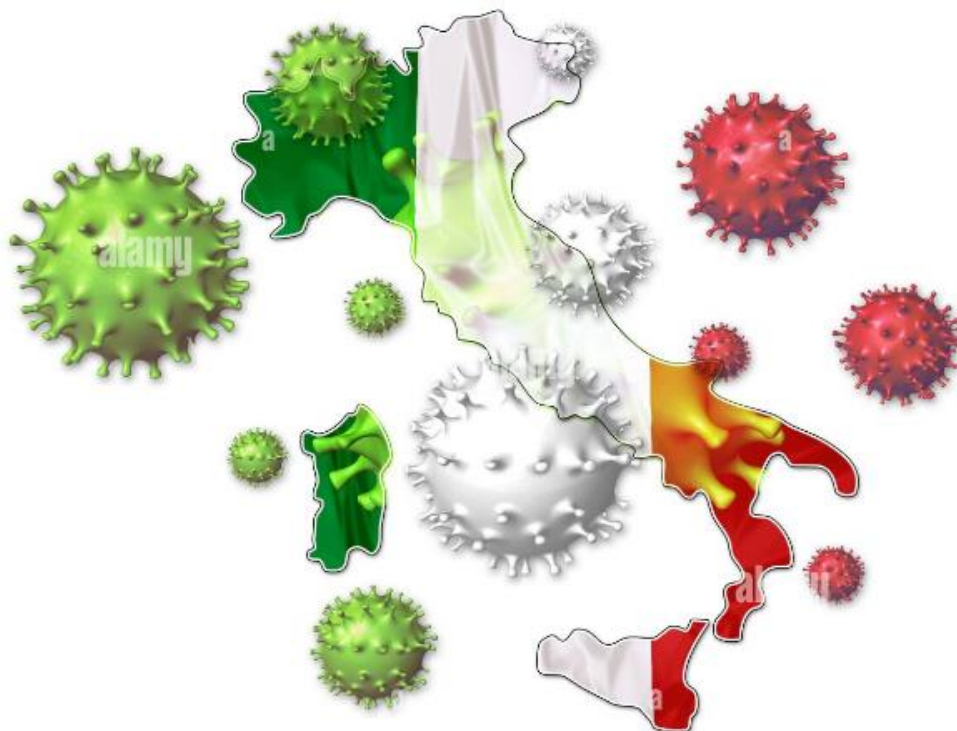


## **Projet de Modélisation et Visualisation des données :**

Evolution du COVID 19 en Italie :

L'impact de la pandémie sur les régions en 2020



Introduction :

Mon projet concerne l'évolution du COVID 19 en Italie pendant la période du 24 février 2020 jusqu'au 26 décembre 2020. Mon objectif avec cette base de données était d'observer la propagation du virus dans les différentes régions d'Italie et d'en déduire, quelle était la région la plus/moins impactée par le virus.

Pour déduire la région la plus/moins impactée par le covid, je vais me servir des indicateurs de suivi de propagation de l'OMS qui sont :

- Activité épidémique (nombre de nouveau cas positifs)
- Tension hospitalière (nombre de patients hospitalisés)
- Nombre de décès

### 1.1 Les données :

Les données proviennent du département de la protection civile d'Italie.

Les régions sont au nombre de 21, le territoire italien comprend 19 régions et 2 régions autonomes (Trento et Bolzano) :

Abruzzo : Abruzzes	Marche : Marches
Basilicata : Basilicate	Molise : Molise
Calabria : Calabre	Piemonte : Piémont
Campania : Campanie	Puglia : Les Pouilles
Emilia-Romagna : Emilie-Romagne	Sardegna : Sardaigne
Friuli-Venezia Giulia: Frioul-Vénétie julienne	Sicilia : Sicile
Veneto : Venetie	Toscana : Toscane
Lazio : Latium	Trentino-Alto-Adige : Trentin-haut-Adige
Liguria : Ligurie	Umbria : Ombrie
Lombardia : Lombardie	Valle d'Asta : Val d'Aoste

Les variables de la base de données :

'SNo', 'Date', 'Latitude', 'Longitude' ,	
Country : pays	CurrentPositiveCases : Cas actuels positifs
RegionCode : Code region	NewPositiveCases : Nouveaux cas positifs
RegionName : nom region	Recovered : rétabli
HospitalizedPatients : Patients hospitalisés	Deaths : décès
IntensiveCarePatients : Patients en soins intensifs	TotalPositiveCases : Total des cas positifs
TotalHospitalizedPatients : Total des patients hospitalisés	
HomeConfinement : confinement à domicile	

### 1.2 Traitements des données sur Python

J'ai commencé par l'importation de mon fichier covid19\_italy\_region.csv afin de pouvoir récupérer mes données et de pouvoir commencer le traitement. Puis, je les stocke dans un DataFrame que j'ai nommé « df ». Je vérifie que l'importation de mes variables a bien fonctionné en affichant le tableau de données (annexe tableau 1), j'ai actuellement 6027 lignes et 17 colonnes.

Puis, je m'intéresse à savoir s'il y a des données manquantes. Je constate qu'il y a des données manquantes. Dans la variable « TestsPerformed », il manque 1155 données. On n'a pas de données car au début de la pandémie, il n'y avait pas de test à disposition.

```

TestsPerformed      1155
HospitalizedPatients    0
Date                 0
Country              0
RegionCode           0
RegionName           0
Latitude             0
Longitude            0
IntensiveCarePatients  0
TotalPositiveCases     0
TotalHospitalizedPatients 0
HomeConfinement        0
CurrentPositiveCases   0
NewPositiveCases       0
Recovered             0
Deaths                0
SNo                   0

```

J'ai choisi de remplacer la valeur « nan » par la valeur 0. J'affiche mon nouveau DataFrame pour vérifier que l'opération a bien fonctionné. Je constate que les données ont été mise à jour. La colonne « TestsPerformed » affiche 0, donc il n'y a plus de données manquantes.

```

: SNo      0
Date      0
Country    0
RegionCode 0
RegionName 0
Latitude   0
Longitude  0
HospitalizedPatients 0
IntensiveCarePatients 0
TotalHospitalizedPatients 0
HomeConfinement 0
CurrentPositiveCases 0
NewPositiveCases 0
Recovered 0
Deaths 0
TotalPositiveCases 0
TestsPerformed 0
dtype: int64

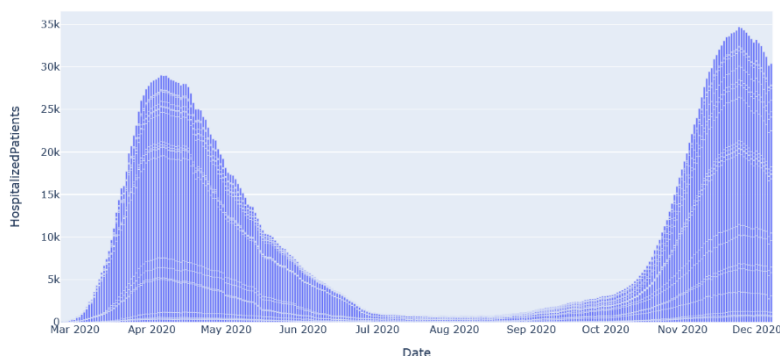
```

Je cherche à savoir quelles sont les régions les plus impactées. Pour ce fait, je regroupe toutes les données qui ont les mêmes modalités dans un nouveau DataFrame. Ce DataFrame se nomme region, ce qui me permet d'avoir un tableau des 21 régions (annexe : tableau2). Pour rendre plus lisible le tableau des statistiques descriptives, je décide d'enlever certaines variables ('RegionCode', 'SNo', 'Latitude', 'Longitude', 'HomeConfinement', 'IntensiveCarePatients').

## 1.2 Evolution du covid 19 en Italie :

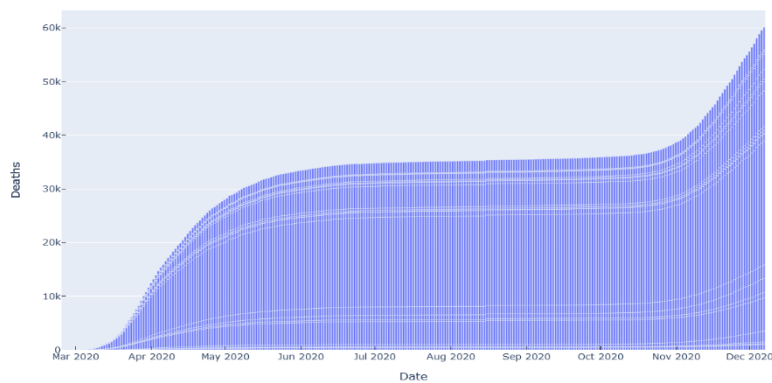
Avant d'observer l'impact sur les régions, on peut analyser l'évolution globale sur cette période à l'aide d'histogrammes.

### A) Histogramme des hospitalisations de l'année 2020 en l'Italie, toute région confondue



Elle montre bien deux vagues. La première commence en mars jusqu'en juillet. Son taux d'hospitalisation avoisine presque les 30000 patients hospitalisés lors de son pic. Et la deuxième vague commence vers septembre, elle met plus de temps à démarrer que la première. Elle décolle mi-octobre, son niveau le plus élevé a été d'environ 35 000 patients hospitalisés puis elle continue jusqu'en décembre.

## B) Histogramme des morts de l'année 2020 en l'Italie, toute région confondue



On remarque que les décès ont commencé légèrement après l'entrée des patients hospitalisés, mi-mars 2020. Le nombre de décès a augmenté durant la première vague pour se stabiliser vers 32000-33000. La deuxième vague relance le nombre de décès, elle atteint 60 000 en décembre 2020. On observe une stabilité entre les deux vagues.

### 1.3 statistiques descriptives pour les Régions :

	HospitalizedPatients	TotalHospitalizedPatients	CurrentPositiveCases	NewPositiveCases	Recovered	Deaths	TotalPositiveCases
<b>count</b>	21.000000	21.000000	21.000000	21.000000	21.000000	21.000000	21.000000
<b>mean</b>	1966.190476	2173.476190	41440.952381	2214.095238	43499.714286	2860.857143	82327.523810
<b>std</b>	2697.338292	2964.437866	43251.242876	2592.544578	63667.419390	4972.920071	98148.383247
<b>min</b>	69.000000	78.000000	2363.000000	176.000000	2397.000000	135.000000	5286.000000
<b>25%</b>	436.000000	482.000000	11572.000000	783.000000	11008.000000	531.000000	24186.000000
<b>50%</b>	1001.000000	1168.000000	18192.000000	1209.000000	18619.000000	1327.000000	35467.000000
<b>75%</b>	2331.000000	2532.000000	72526.000000	2822.000000	59432.000000	2622.000000	130255.000000
<b>max</b>	12077.000000	13328.000000	164406.000000	11489.000000	289706.000000	23024.000000	429109.000000

On voit qu'en moyenne pour les 21 régions, les patients hospitalisés sont aux nombres de 1966. Pour le premier quartile, 25% des régions ont 436 hospitalisations ou moins. On voit que 50% des régions ont environ 1001 patients hospitalisés. Pour le 3ème quartile, on voit que 25% des régions ont un taux de d'hospitalisation supérieure ou égale à 2331. Le maximum pour une région s'élevé à 12077 patients.

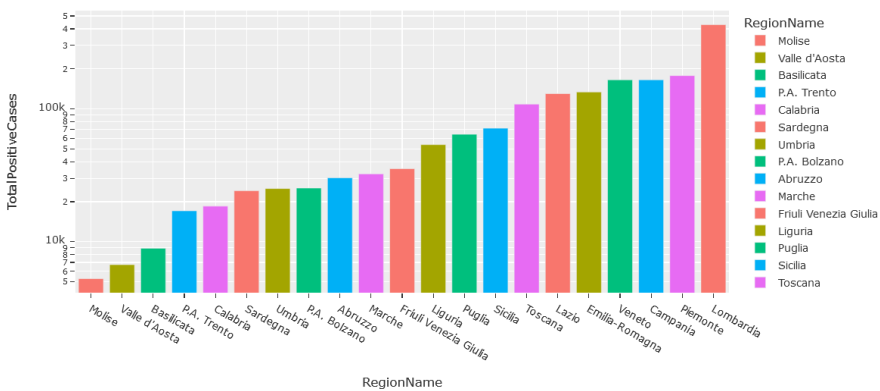
Le nombre de décès est en moyenne de 2860. Pour le premier quartile, 25% des régions ont 531 décès ou moins. On voit que 50% des régions ont au moins 1371 décès. Pour le 3ème quartile (75%), on voit que 25% des régions ont un taux de décès supérieure ou égale à 2622. Le maximum pour une région s'élevé à 23024 décès alors que pour autre région, le minimum est de 135 décès.

Les nouveaux cas positifs dans les régions sont de 2214 en moyenne. Pour le premier quartile, 25% des régions ont 783 nouveaux cas positifs ou moins. On voit que 50% des régions ont au moins 1209 cas. Pour le 3ème quartile (75%), on voit que 25% des régions ont un taux de décès supérieure ou égale à 2822. Le maximum pour une région s'élevé à 11409 cas alors que pour autre région, le minimum est de 176 nouveaux cas positifs.

### 1.4 les histogrammes :

Les histogrammes ci-dessous montre l'impact de la propagation du virus sur l'ensemble des régions en Italie. Les régions sont :

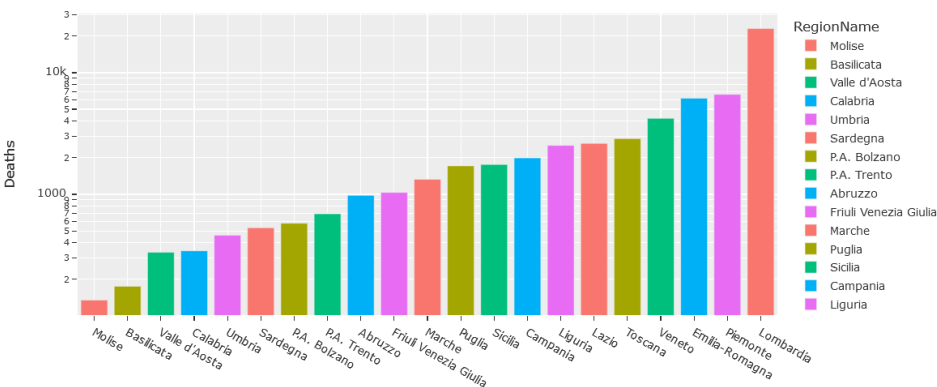
## Nombre des cas positifs par Région :



On observe que les 3 régions les plus touchées par le nombre de cas positifs sont :

- Lombardia (429 109 cas)
- Piemonte (177 788 cas)
- Campania (165 293 cas)

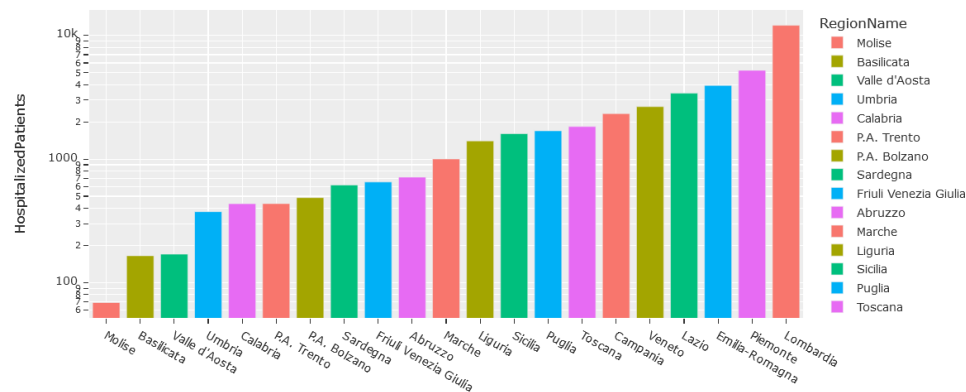
## Le nombre de décès par région :



On observe que les régions les plus touchées par le nombre de décès sont :

- Lombardia (23024 décès)
- Piemonte (6623 décès)
- Emilia- Romagna (6162 décès)

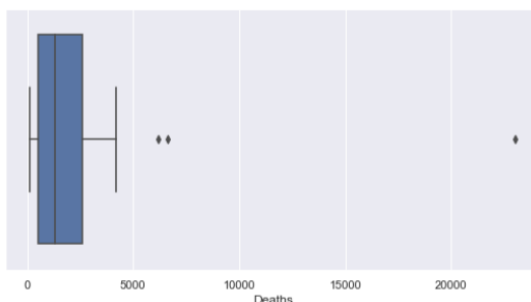
## Le nombre de patients hospitalisés par région :



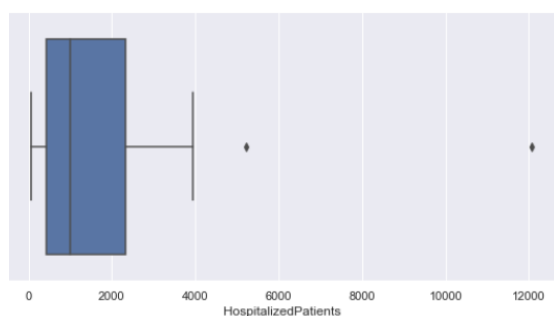
On observe que les régions les plus touchées par le nombre de patients hospitalisés sont :

- Lombardia (12077 patients)
- Piemonte (5225 patients)
- Emilia-Romagna (3944 patients)

## 1.5 La boite à moustache :

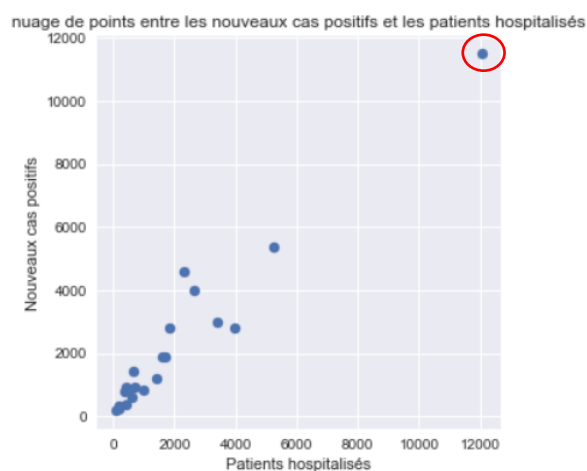


La première boite à moustache représente la distribution des décès parmi les 21 régions, on peut constater qu'il y a 3 valeurs aberrantes. La plus forte, celle qui avoisine les 23000 concerne la région de Lombodia et les deux autres, proche des 6000 sont les régions de Piemonte et Emilia-Romagna. On remarque que 75% des régions de dépassent pas les 3000.



La deuxième boîte à moustache représente la distribution des patients hospitalisés. Cette fois-ci, on a deux valeurs aberrantes. La première valeur est à 12000, elle concerne la région de Lombardia et la deuxième valeur est à 5000 avec la région de Piemonte. 50% des régions ne dépassent pas les 1000 hospitalisations.

### 1.6 Nuage de points :



Grace aux deux nuages de points, on observe bien la différence entre une région et les autres. La région de la Lombardia est celle qui a le plus grand écart envers les autres régions. On constate que les autres régions sont très proches de 0.

### Conclusion

En début de ce dossier, je me suis posé comme question quelle était la région la plus/moins impactée par la pandémie du covid en 2020. Après l'analyse des graphiques et histogrammes, je peux conclure ce dossier en disant que la région qui a été le plus impactée par la covid était la Lombardia/Lombardie. Cette région recense plus de 23000 décès, 12 077 patients hospitalisés et de 429 109 cas positifs.

On peut aller plus loin sur l'analyse des régions. La deuxième région l'a plus impacté était le Piemonte/Piémont avec plus de 6600 décès, 5225 patients hospitalisés et de 177 788 cas positifs et à contrario, la région qui a été le moins impactée, c'était la région du Molise avec 135 décès, 69 patients hospitalisés et 5286 cas positifs.

## ANNEXES

Tableau 1 :

	SNo	Date	Country	RegionCode	RegionName	Latitude	Longitude	HospitalizedPatients	IntensiveCarePatients	TotalHospitalizedPatients
0	0	2020-02-24T18:00:00	ITA	13	Abruzzo	42.351222	13.398438	0	0	
1	1	2020-02-24T18:00:00	ITA	17	Basilicata	40.639471	15.805148	0	0	
2	2	2020-02-24T18:00:00	ITA	18	Calabria	38.905976	16.594402	0	0	
3	3	2020-02-24T18:00:00	ITA	15	Campania	40.839566	14.250850	0	0	
4	4	2020-02-24T18:00:00	ITA	8	Emilia-Romagna	44.494367	11.341721	10	2	1
...	...	...	...	...	...	...	...	...	...	...
6022	6022	2020-12-06T17:00:00	ITA	19	Sicilia	38.115697	13.362357	1367	213	1580
6023	6023	2020-12-06T17:00:00	ITA	9	Toscana	43.769231	11.255889	1360	252	1612
6024	6024	2020-12-06T17:00:00	ITA	10	Umbria	43.106758	12.388247	332	60	392
6025	6025	2020-12-06T17:00:00	ITA	2	Valle d'Aosta	45.737503	7.320149	102	8	110
6026	6026	2020-12-06T17:00:00	ITA	5	Veneto	45.434905	12.338452	2508	308	2816

6027 rows × 17 columns

Tableau 2 :

	SNo	Date	Country	RegionCode	Latitude	Longitude	HospitalizedPatients	IntensiveCarePatients	TotalHospitalizedPatients	HospitalizedPatientsPer1000
RegionName										
Abruzzo	6006	2020-12-06T17:00:00	ITA	13	42.351222	13.398438	714	77	790	15.8
Basilicata	6007	2020-12-06T17:00:00	ITA	17	40.639471	15.805148	165	30	185	3.7
Calabria	6008	2020-12-06T17:00:00	ITA	18	38.905976	16.594402	435	53	482	9.8
Campania	6009	2020-12-06T17:00:00	ITA	15	40.839566	14.250850	2331	227	2532	51.6
Emilia-Romagna	6010	2020-12-06T17:00:00	ITA	8	44.494367	11.341721	3944	375	4310	89.4
Friuli Venezia Giulia	6011	2020-12-06T17:00:00	ITA	6	45.649435	13.768136	652	62	711	14.4
Lazio	6012	2020-12-06T17:00:00	ITA	12	41.892770	12.483667	3408	364	3762	77.2
Liguria	6013	2020-12-06T17:00:00	ITA	7	44.411493	8.932699	1402	179	1510	30.8
Lombardia	6014	2020-12-06T17:00:00	ITA	3	45.466794	9.190347	12077	1381	13328	275.8
Marche	6015	2020-12-06T17:00:00	ITA	11	43.616760	13.518875	1001	169	1168	23.8
Molise	6016	2020-12-06T17:00:00	ITA	14	41.557748	14.659161	69	14	78	1.6
P.A. Bolzano	6017	2020-12-06T17:00:00	ITA	21	46.499335	11.356624	487	65	530	10.8
P.A. Trento	6018	2020-12-06T17:00:00	ITA	22	46.068935	11.121231	436	81	478	9.7