



INTRODUCTION

- Le présent projet vise à analyser les données relatives à la pandémie de COVID-19, une maladie infectieuse qui a profondément marqué le monde depuis son émergence en 2019. Face à cette crise sanitaire d'envergure mondiale, il est essentiel de comprendre les multiples facettes de cette pandémie et d'explorer les questions clés qui se posent :
- Quels sont les facteurs influençant la propagation et la gravité du COVID-19?
- Quel est l'effet de la vaccination sur la propagation et la mortalité liées au virus ?

MÉTHODOLOGIE

Partie 0: Importation des données actualisées datant du 15 Mai 2023 depuis le site web OurWorldInData

Partie 1: Nettoyage des données :

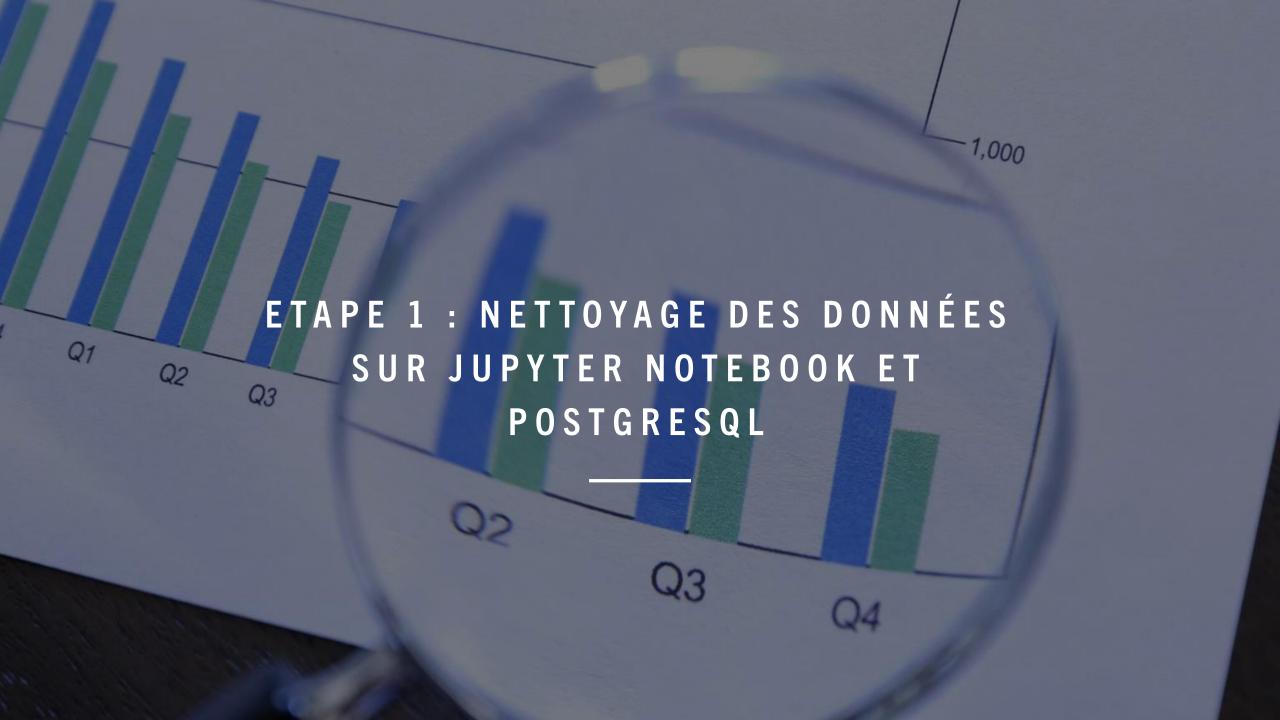
- Identification et suppression des valeurs manquantes.
- Suppression des doublons.
- Prétraitement des données pour assurer leur cohérence et leur qualité.

Partie2: Analyse des données sur PostgreSQL:

- Importation des données nettoyées dans une base de données PostgreSQL.
- Utilisation de requêtes SQL pour extraire des informations spécifiques, telles que le nombre de cas confirmés et de décès par année, le pourcentage de personnes infectées par région, etc.

Partie 3: Visualisations sur Tableau:

- Transfert des données analysées depuis PostgreSQL vers Tableau.
- Création de différentes visualisations, telles que des cartes interactives, des graphiques, et des comparaisons temporelles.
- -Utilisation des fonctionnalités de Tableau pour explorer et présenter les tendances et les corrélations identifiées.



SUPPRESSION DES VALEURS NULLES DANS LA BASE DE DONNÉES DES DÉCÈS

- Nous avons supprimé toutes les valeurs nulles de la colonne "continent" car leur localisation ne semblait pas pertinente.
- Nous avons choisi de remplacer les valeurs nulles dans les colonnes commençant par "new" par 0, ce qui indique l'absence de nouveaux cas enregistrés ce jour.
- Pour les autres colonnes, nous avons opté pour le remplacement par la valeur précédente du même pays, et si cette valeur n'est pas disponible, elle est remplacée par la valeur suivante.
- Les valeurs nulles pour toutes les lignes d'un pays ont été remplacées par -1, indiquant ainsi leur non-disponibilité.
- Le lien vers le notebook est le suivant : [insérer le lien vers le notebook].

```
# Tri des données par pays et date
df.sort values(['location', 'date'], inplace=True)
# Liste des colonnes numériques
numeric columns = df.select dtypes(include=['float64']).columns
# Groupe par pays
grouped = df.groupby('location')
# Boucle for pour remplacer les valeurs null
for col in numeric_columns:
   if col.startswith('new'):
        # Remplacement des valeurs null par 0 pour les colonnes commençant par 'new'
        df[col].fillna(0, inplace=True)
   else:
        #remplacement des valeurs null par la valeur precedente ou suivante du meme pays
        group_mask = df[col].isnull().groupby(df['location']).transform('any')
       df[col] = df.groupby('location')[col].ffill().bfill().where(group_mask, df[col])
# Affichage du DataFrame mis à jour
print(df.isnull().sum())
iso_code
continent
location
population
total cases
new cases smoothed
total deaths
new deaths
new deaths smoothed
total_cases_per_million
new cases per million
```

SUPPRESSION DES VALEURS NULLES DANS LA BASE DE DONNÉES DES VACCINATIONS

iso code

- Le même principe de nettoyage expliqué précédemment a été appliqué dans ce cas.
- Le lien vers le Notebook :

```
•[7]: # Tri des données par pays et date
      df.sort values(['location', 'date'], inplace=True)
      # Liste des colonnes numériques
      numeric columns = df.select dtypes(include=['float64']).columns
      # Groupe par pays
      grouped = df.groupby('location')
      # Boucle for pour remplacer les valeurs null
      for col in df:
          if col.startswith('new'):
              # Remplacement des valeurs null par 0 pour les colonnes commençant par 'new'
              df[col].fillna(0, inplace=True)
              #remplacement des valeurs null par la valeur precedente ou suivante du meme pays
              group mask = df[col].isnull().groupby(df['location']).transform('any')
              df[col] = df.groupby('location')[col].ffill().bfill().where(group_mask, df[col])
      # Affichaae du DataFrame mis à iour
      print(df.isnull().sum())
```

continent location date total tests new_tests total_tests_per_thousand new tests per thousand new tests smoothed new_tests_smoothed_per_thousand positive rate tests per case tests units total_vaccinations people vaccinated people fully vaccinated new vaccinations new_vaccinations_smoothed total vaccinations per hundred people vaccinated per hundred people_fully_vaccinated_per_hundred new_vaccinations_smoothed_per_million new people vaccinated smoothed new people vaccinated smoothed per hundred

ÉLIMINATION DES DOUBLONS DANS LES TABLES DES VACCINATIONS ET DES DÉCÈS

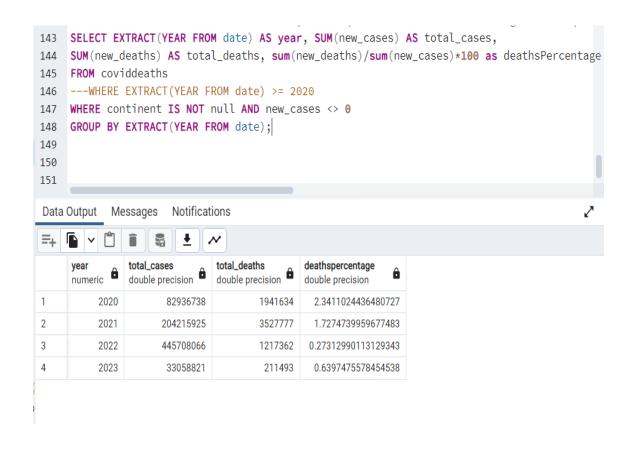
- Nous avons entrepris d'éliminer les valeurs en double de notre table en suivant les étapes suivantes :
- Nous avons créé une copie temporaire de notre table qui contient uniquement les valeurs uniques des champs de la table.
- Ensuite, nous avons supprimé notre table d'origine.
- Nous avons donné à la copie temporaire le même nom que notre table d'origine pour la remplacer.
- Le lien vers le notebook est le suivant : [insérer le lien vers le notebook].

```
--Etapes pour supprimer les doublons de date dans la table des vaccinations
-- Créer une copie temporaire de la table
CREATE TABLE temp table AS
SELECT DISTINCT *
FROM covidvaccinations;
-- Supprimer l'ancienne table
DROP TABLE covidvaccinations;
-- Renommer la copie temporaire avec le nom original
ALTER TABLE temp_table RENAME TO covidvaccinations;
--Etapes pour supprimer les doublons de date dans la table des deces
-- Créer une copie temporaire de la table
CREATE TABLE temp_table AS
SELECT DISTINCT *
FROM coviddeaths;
-- Supprimer l'ancienne table
DROP TABLE coviddeaths:
-- Renommer la copie temporaire avec le nom original
ALTER TABLE temp_table RENAME TO coviddeaths;
```



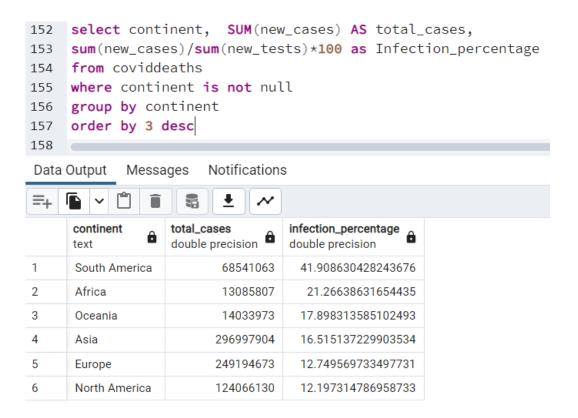
ÉVOLUTION ANNUELLE DES CAS CONFIRMÉS ET DES DÉCÈS LIÉS À LA COVID-19

• Globalement, nous observons une diminution du taux de mortalité au fil des années pour la pandémie de COVID-19. Cependant, une observation intéressante est que le nombre de cas enregistrés en 2020 est nettement inférieur à ceux de 2021 et 2022, mais le pourcentage de décès est significativement plus élevé. Cette constatation soulève une fois de plus la question des facteurs qui ont influencé cette régression du taux de mortalité.



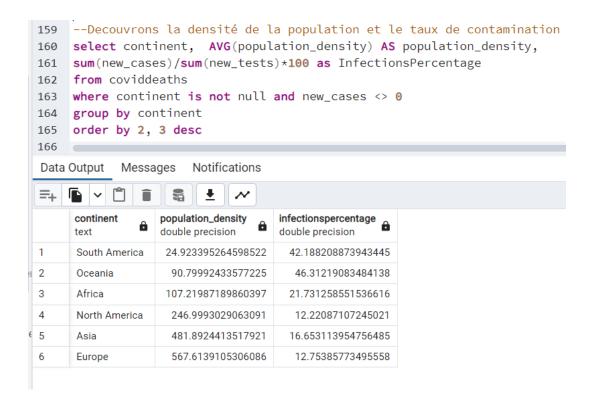
ANALYSE COMPARATIVE DES CAS CONFIRMÉS ET DU POURCENTAGE DE PERSONNES INFECTÉES PAR RÉGION DANS LE MONDE

 Lorsqu'on analyse la corrélation entre le nombre de cas confirmés et le nombre de personnes testées, une observation frappante se dégage : le pourcentage d'infection par la Covid-19 est de 40% pour les Sud-Américains et de 21% pour les Africains. Ces chiffres laissent supposer que la génétique pourrait jouer un rôle dans la susceptibilité à la contamination par la Covid-19.



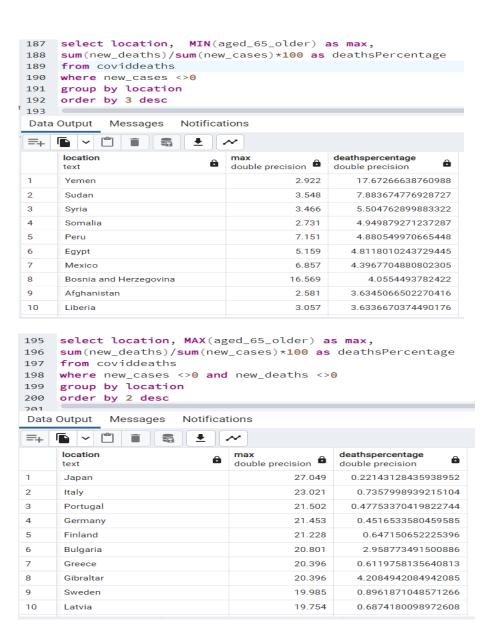
IMPACT DE LA DENSITÉ DE POPULATION SUR LE TAUX DE CONTAGION DE LA COVID-19

 Une observation intéressante émerge de nos résultats: les continents ayant une densité de population élevée présentent en réalité un taux de contagion plus faible par rapport à ceux ayant une densité plus faible. Cela suggère que la densité de la population ne semble pas être un facteur déterminant influençant le taux de contagion de la Covid-19.



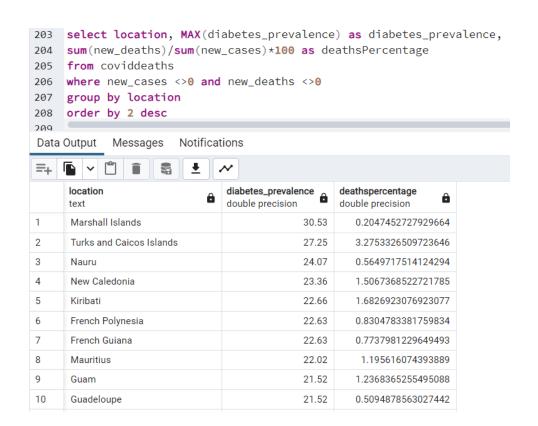
ANALYSE DE LA CORRÉLATION ENTRE LE POURCENTAGE DE PERSONNES ÂGÉES ET LE TAUX DE MORTALITÉ

Lorsque l'on compare le pourcentage de la population âgée de 65 ans ou plus dans les 10 pays présentant les taux de décès les plus élevés liés à la COVID-19 (au-dessus) avec les 10 pays ayant le pourcentage le plus élevé de personnes âgées de 65 ans ou plus (en dessous), on constate que les pays avec un taux de décès très élevé ont un pourcentage de personnes âgées assez faible. De même, les pays présentant un pourcentage élevé de personnes âgées ont un taux de décès beaucoup plus faible. Par conséquent, on peut conclure que l'âge avancé ne semble pas être un facteur prédominant influençant la mortalité liée à la COVID-19



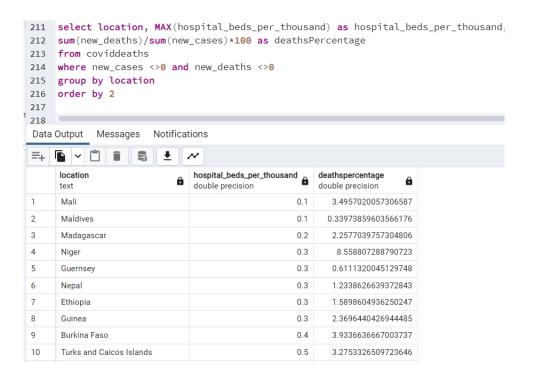
ANALYSE DES TAUX DE MORTALITÉ DANS LES PAYS À FORTE PRÉVALENCE DU DIABÈTE

Lors de l'analyse des taux de mortalité dans les 10 pays présentant les taux de prévalence les plus élevés du diabète, nous avons constaté que ces taux étaient généralement similaires ou inférieurs à la moyenne mondiale (0,90 %). Cette observation suggère que le diabète ne semble pas être un facteur prépondérant de la mortalité liée à la COVID-19.



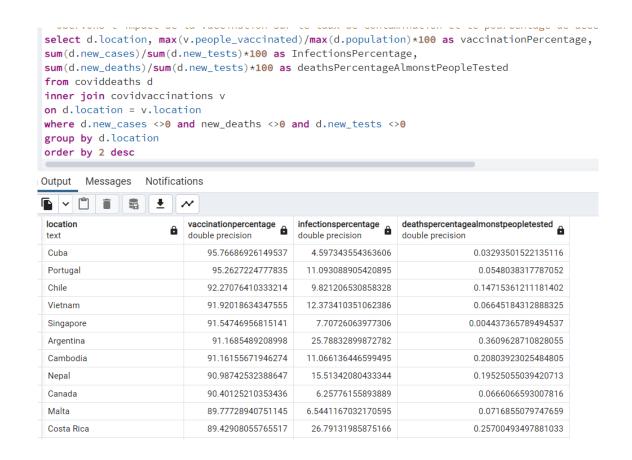
ANALYSE DE LA CORRÉLATION ENTRE LA CAPACITÉ DE PRISE EN CHARGE ET LE TAUX DE MORTALITÉ

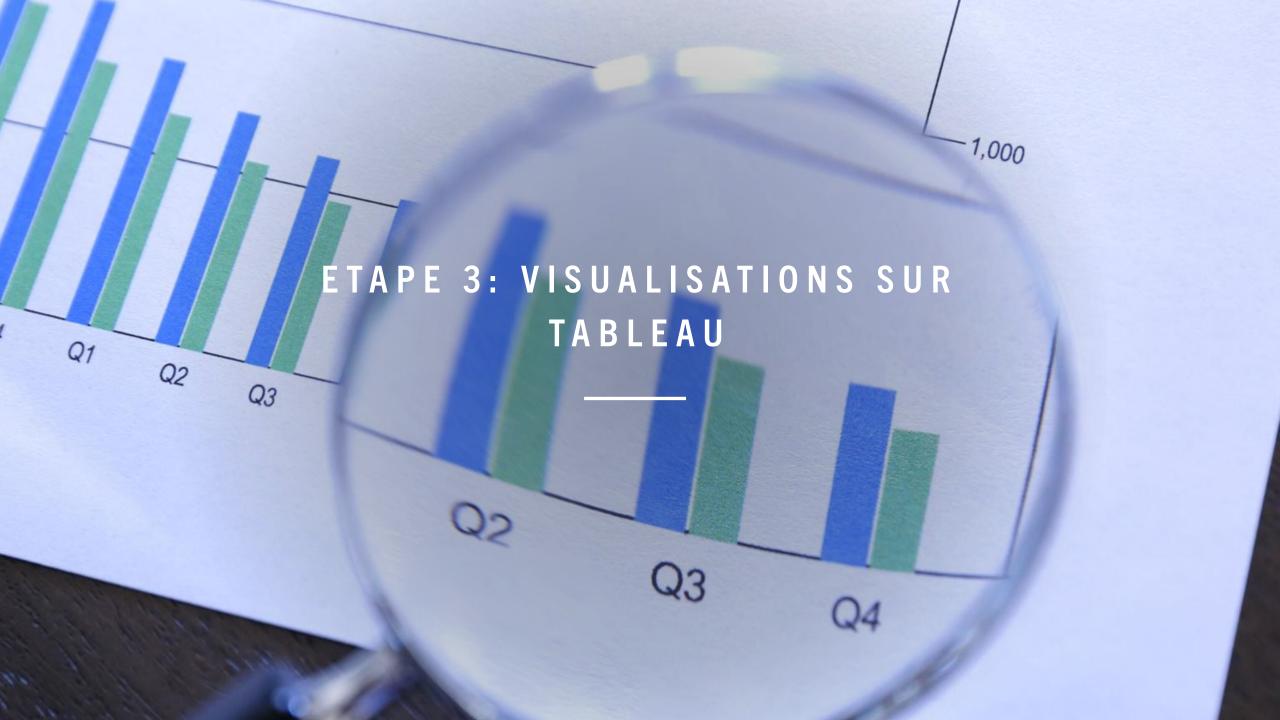
L'observation des taux de décès liés à la COVID-19 dans les 10 pays présentant les pourcentages les plus bas de lits d'hôpital par tranche de 1000 personnes révèle des taux de mortalité très élevés, atteignant jusqu'à 8,5 %. Ce chiffre représente plus de neuf fois la moyenne mondiale. Ces résultats indiquent clairement que la capacité de prise en charge des personnes infectées dans un pays a une influence considérable sur le taux de mortalité à la COVID-19.



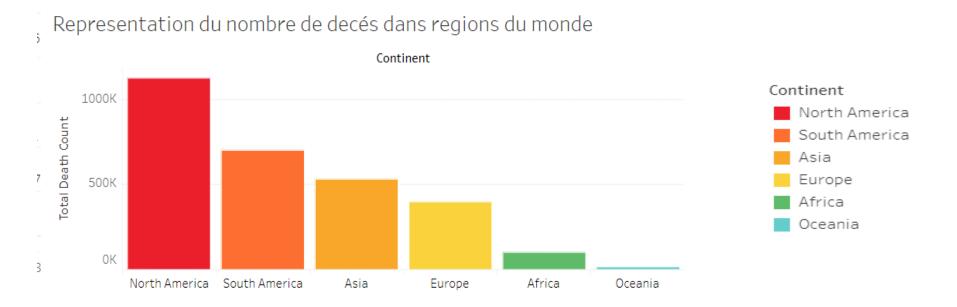
ANALYSE DE L'IMPACT DE LA VACCINATION SUR LE TAUX DE CONTAMINATION ET LE TAUX DE DÉCÈS

- L'analyse des 10 pays affichant les taux de vaccination les plus élevés révèle des tendances encourageantes. Malgré le fait que seulement environ 10% des personnes testées se révèlent positives à la COVID-19, le taux de décès parmi les personnes testées est d'environ 0,2%. Ces résultats indiquent clairement que la vaccination joue un rôle majeur dans la prévention des contaminations et des décès liés à la COVID-19.
- En d'autres termes, la vaccination a un impact significatif sur la réduction de la propagation du virus et la diminution des conséquences graves de la maladie.





RÉPARTITION MONDIALE DES CAS CONFIRMÉS DE COVID-19

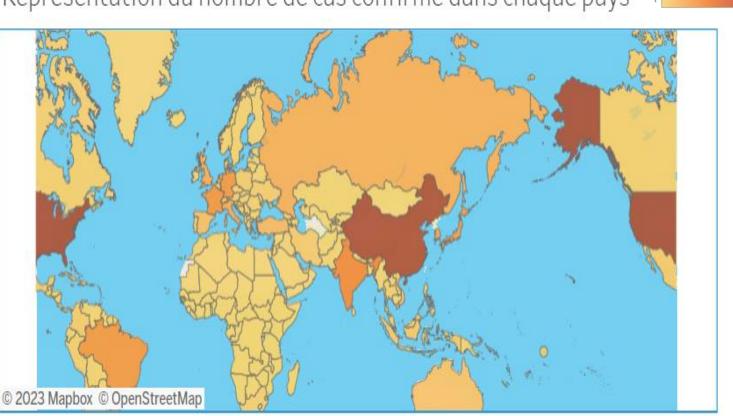


• Ce graphique illustre la répartition inégale des décès par régions du monde, mettant en évidence les disparités dans les taux de mortalité liés à la COVID-19. Ces disparités peuvent être influencées par divers facteurs précédemment identifiés. Cela souligne l'importance de prendre en compte ces différences régionales dans la gestion de la pandémie et d'adapter les stratégies en conséquence pour réduire les effets négatifs de la maladie.

CARTE INTERACTIVE DE LA COVID-19 : IMPACT MONDIAL ET STATISTIQUES PAR PAYS

- la compréhension de l'impact de la Covid-19 dans différents pays en utilisant des couleurs pour représenter les niveaux faibles, moyens et élevés de contamination. En pointant le curseur sur un pays, on peut également connaître le nombre de personnes infectées.
- Lien Tableau Public: https://public.tableau.com/app/profile/chr is.essomba/viz/CovidDeathsVisualization_ 16845886904710/Tableaudebord1

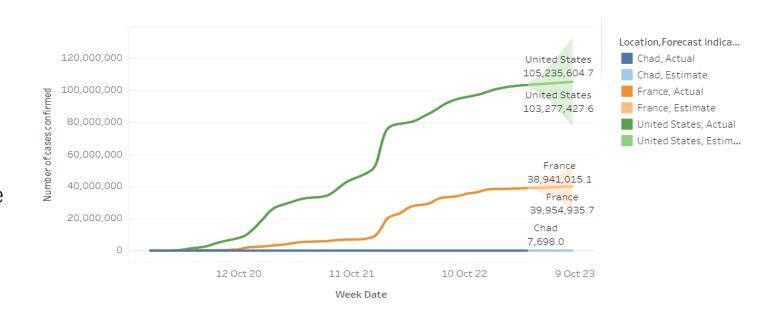
Representation du nombre de cas confirmé dans chaque pays



Number of people infected

COMPARAISON DE L'ÉVOLUTION DES CAS DE COVID-19 ENTRE LA FRANCE, LES ÉTATS-UNIS ET LE TCHAD

Nous avons analysé l'évolution du nombre total de personnes infectées en France par rapport aux États-Unis (le pays ayant enregistré le plus de cas confirmés) et au Tchad (le pays ayant enregistré le moins de cas). Nous observons une augmentation constante de la courbe entre octobre 2020 et octobre 2022, qui semble maintenant se stabiliser à une certaine valeur, indiquant une très faible incidence de nouveaux cas de contamination. Cependant, les prévisions indiquent qu'il pourrait y avoir environ 1000 nouveaux cas d'ici la fin de cette année.



CONCLUSION

- Cette analyse approfondie des données liées à la pandémie de COVID-19 a permis de répondre à des questions clés. Nous avons identifié les facteurs qui influencent la propagation et la gravité de la maladie, tels que la densité de population, l'âge moyen et la capacité de prise en charge des patients infectés. De plus, nous avons examiné l'impact de la vaccination sur la propagation et la mortalité, en évaluant l'efficacité des campagnes de vaccination.
- En somme, cette étude approfondie a contribué à une meilleure compréhension des dynamiques de la pandémie de COVID-19 et des facteurs qui influencent significativement sa propagation et son taux de mortalité. Ces informations fournissent des bases solides pour orienter les mesures de prévention et les décisions stratégiques visant à atténuer l'impact de la maladie. Il est essentiel de continuer à suivre et à analyser les données afin de rester informé et de s'adapter aux défis évolutifs posés par la COVID-19.

RÉFÉRENCE

- Nom du site web : OurWorldInData
- URL du site web : https://ourworldindata.org/covid-deaths
- Date de récupération des données : 15 mai 2023

