

● Définir l'objectif de votre projet :

Ce projet vise à analyser et visualiser les données COVID-19 depuis Google BigQuery pour identifier les tendances temporelles, géographiques et les corrélations entre indicateurs clés (hospitalisations, vaccinations). **L'objectif** est d'améliorer la compréhension et d'aider à la prise de décisions éclairées pour une gestion optimale des ressources et des stratégies sanitaires.

● Décrivez le dataset et le type d'analyse que vous effectuerez pour obtenir des informations à partir de ce dataset :

Ce dataset est une source ouverte contenant des informations détaillées et agrégées sur la pandémie de COVID-19. Les colonnes principales incluent des données sur :

- Les cas confirmés, décès, et hospitalisations.
- Les vaccinations
- Les informations géographiques et démographiques
- Les dates et indicateurs temporels
- Autres indicateurs contextuels

Types d'Analyse à Effectuer :

1. Analyse temporelle des cas et vaccinations

- ****Objectif :**** Identifier les tendances globales et locales dans la progression des cas confirmés, des décès, et des vaccinations afin d'explorer la progression des cas confirmés depuis le début de la pandémie.

3. Analyse des taux de croissance

Objectif : Mesurer la vitesse de propagation du virus ou l'augmentation des vaccinations pour vérifier à quels moments le virus s'est propagé le plus rapidement. Ainsi que l'impact des campagnes de vaccination sur le ralentissement de la propagation.

**4. Comparaisons entre régions ou périodes

- ****L'objectif :** Sera de mettre en évidence des différences entre plusieurs régions ou périodes de temps. Nous allons chercher à savoir quels pays ont des taux de mortalité plus faibles, et le Pourquoi. Vérifier également comment la vaccination a-t-elle influencé les taux d'infection et de mortalité.

5. Analyse des corrélations

Objectif : Identifier des relations entre différentes variables :
Vérifier la corrélation entre taux de vaccination et diminution des cas confirmés. Analyse des impacts des indicateurs géographiques (densité de population) sur la propagation du virus.

- **Quelles transformations, si nécessaire, sont effectuées**

1. Nettoyage des Données

- Gérer les valeurs manquantes :

Supprimer les colonnes inutiles (comme `url`, qui est vide).

Remplir les valeurs manquantes dans `value`

2. Conversion de Données

• Dates et temporalité :

- Convertir la colonne date en format de type datetime pour permettre des analyses temporelles.
- Extraire des informations telles que l'année, le mois, ou le jour de la semaine.

3. Filtrage et Sélection

• Filtrer les indicateurs spécifiques :

- Se concentrer sur les vaccinations ou les cas confirmés.

• Sélectionner les périodes ou régions pertinentes :

- Comparer des périodes spécifiques avant/après vaccination) ou des pays d'intérêt.

4. Agrégation

• Par semaine ou mois :

- Agréger les données pour obtenir des moyennes ou des totaux hebdomadaires/mensuels.

• Par pays ou région :

- Calculer des statistiques globales ou par région géographique.

5. Création de nouvelles colonnes

• Calcul de ratios ou de proportions :

- Exemples :
 - Taux de mortalité = décès / cas confirmés.

- $\text{Taux de vaccination} = \frac{\text{vaccinations effectuées}}{\text{population totale}}$.

Indicateurs dérivés :

- Exemples :
 - $\text{Taux de croissance journalier} = \frac{(\text{valeur actuelle} - \text{valeur précédente})}{\text{valeur précédente}}$.
 - Jours depuis le premier cas dans un pays donné

6. Filtrage et segmentation des données

- **Sélection des périodes ou indicateurs pertinents :**
 - Filtrer les données pour se concentrer sur des périodes, pays ou indicateurs spécifiques.
- **Segmentation par région ou groupe :**
 - Ajouter une colonne pour classer les pays en fonction de leur continent ou d'autres caractéristiques (densité de population, revenus).

• Comment les données sont intégrées dans BigQuery :

Préparation des Données :

- Les données sont collectées depuis des sources externes fichiers CSV.
- Elles sont nettoyées et transformées pour s'assurer qu'elles respectent le schéma attendu dans BigQuery (types de données corrects, colonnes requises).

Création de la Table dans BigQuery :

- Une table est créée dans un dataset BigQuery avec un schéma défini (nom des colonnes, types de données).
- Cela a été fait via l'interface web de BigQuery.

3- Chargement des Données :

Chargement des Données :

- La clé JSON est utilisée pour authentifier l'action.

- Les données sont chargées via :
 - **Interface Web** : Téléchargement direct dans une table.
 - **Script Python** : Automatisation avec le SDK `google-cloud-bigquery`

Validation et Vérification :

- Une fois les données intégrées, des requêtes SQL sont exécutées pour vérifier la qualité et la cohérence des données dans la table.

- **Quels sont les résultats que vous cherchez à trouver**