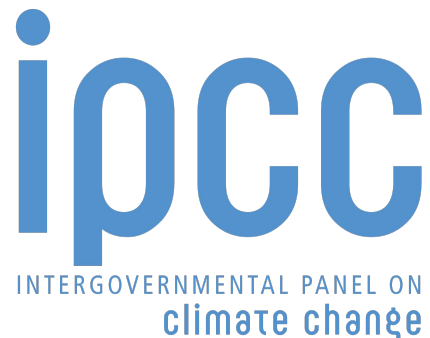


Climate Change

Data Lab

Thème

Le GIEC vous demande de les aider à comprendre le réchauffement climatique ainsi que ces potentielles causes.



Contexte

Depuis la révolution industrielle, une hausse des températures est mesurée partout sur la planète. Cette hausse semble être corrélée avec l'augmentation de la population industrielle et du CO2 présent dans l'atmosphère.

Source: <https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>



Objectif

Mise en place d'une solution d'analyse de données

1. Construction d'un Data Warehouse suivant un modèle multi-dimensionnel.
2. Mise en place d'un ETL qui permettra d'ajouter à l'avenir les nouvelles données au fur et mesure.
3. Exploiter le nouveau modèle et ses données afin d'élaborer des rapports permettant de répondre aux questions qui sont posées dans le slide suivant (voire d'autres). Ces rapports peuvent être réalisés à l'aide d'outils tels que Power BI, Qlik Sense, Tableau, etc.

Requirements

1. Aider à comprendre s'il y a bien une hausse des températures depuis la révolution industrielle.
2. Aider à comprendre s'il semble bien y avoir une corrélation avec la révolution industrielle.
3. Aider à comprendre s'il y a bien une hausse du CO2 dans l'atmosphère.
4. Aider à comprendre s'il semble bien y avoir une corrélation avec la possible hausse du CO2 dans l'atmosphère.

Sources de données

Une base de donnée de staging comprenant :

- **GlobalLandTemperature**

- La liste de toutes les températures par mois dans toutes les villes du monde depuis la révolution industrielle.
- Nombre de lignes : 8 516 158.
- Source :

<https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>

Sources de données

Mise à disposition d'une base de staging contenant déjà d'autres données

- **CountryGDP**

- Des informations sur le PIB de 193 pays par personne et par an en USD allant du 1er janvier 1901 au 1er janvier 2011.
- <https://www.kaggle.com/divyansh22/gdp-per-person-19012011>

- **CO2Emission**

- Des informations sur les émissions de CO2 depuis la Révolution Industrielle.
- <https://www.kaggle.com/srikantsahu/co2-and-ghg-emission-data>

Ces sources des données peuvent être consolidées afin, par exemple, de comparer la vitesse de propagation des différentes épidémies.

Sources de données supplémentaires

Il s'agit d'un projet pour lequel une partie des données est fournie, toutefois seule une des sources de données doit être obligatoirement utilisée et ce, afin que tous les stagiaires partent d'un socle commun.

Les stagiaires sont absolument libres de rajouter d'autres sources de données et de les explorer au gré de leurs envies.

Délivrables

1. Backup de votre Data Warehouse.
2. Fichier de reporting Power BI, Qlik Sense, Tableau, etc.
3. Slides de présentation expliquant votre démarche et vos résultats.
 - Comment vous avez choisi de traiter le problème.
 - Quels sont les données que vous avez choisi.
 - Quels sont les constats que vous avez pu mettre en évidence.
 - Quels sont les conclusions que vous pouvez apporter au sujet.