

bp



Projet Chef d'œuvre

ETUDE SUR LA PRODUCTION ET LA CONSOMMATION
DES ENERGIES FOSSILES ET ENERGIES RENOUVELABLES

CERTIFICATIONS “DEVELOPPER UNE BASE DE DONNEES” (CNCP 3497)
ET “EXPLOITER UNE BASE DE DONNEES” (CNCP 3508)

KALONJI Jimmy | Développeur Data Nanterre | 2020/2021

Chapitre 1 Introduction

1.1 Exploitation des ressources naturelles	2
1.2 L'énergie fossile	3
1.3 L'énergie renouvelable	3
1.4 La transition énergétique	4

Chapitre 2 Cahier des charges

2.1 Contexte	4
2.2 L'objectif	4
2.3 Contraintes	4
2.4 Tâches des techniques	5
2.5 Langage et outils utilisés	6

Chapitre 3 Architecture en général

6

Chapitre 4 Les différentes étapes de la préparation de la donnée

4.1 Collecte de Données	7
4.2 Description des données	7
4.3 Nettoyer et valider les données	7

Chapitre 5 Modèle physique de données (MCD)

10

Chapitre 6 Modèle organisationnel des données

12

Chapitre 7 Requêtes

15

Chapitre 8 Sécurisation des données

16

Chapitre 9 Visualisation

17

Conclusion

20

Gestion du projet

20

REMERCIEMENT

22

BIBLIOTHEQUE

23

Chapitre 1 Introduction

Le chef d'œuvre a pour but de montrer l'ensemble des compétences acquises durant 7 mois de formation à Simplon Nanterre au métier de développeur data.

Commencé en décembre 2020, cette formation intensive délivre principalement des certifications reconnues par l'état et des organismes professionnels tels que :

1. Développer une base de données (CNCP 3497)
2. Exploiter d'une base de données (CNCP 3508)

À cette date, ce rapport s'inscrit dans le processus d'obtention de ces certifications. Il décrit de manière conceptuelle et technique mon projet de fin de formation.

Présentation du projet

Pour ce projet d'étude, J'ai choisi comme sujet le monde des ressources naturelles plus particulièrement les énergies fossiles et les énergies renouvelables dans le monde.

Le but de ce chef d'œuvre est de comprendre les rapports énergétiques de chaque pays c'est-à-dire connaître la consommation ou la production des énergies et de savoir s'il y a un lien avec l'augmentation de la population mondial. Le jeu de données que j'ai trouvé tire d'une entreprise pétrolière dans une période entre 1965 à 2019.

Ce projet se compose en 2 parties.

- ❖ Structure et stockage des données.
- ❖ Datavisualisation

Contexte

Y a-t-il une solution à l'exploitation des ressources naturelles ?

En 2021, l'économie mondiale extrait plus de 92 milliards de tonnes de ressources naturelles par an pour fonctionner, alors que le chiffre n'était que de 27 milliards de tonnes à la fin des années 60. Selon le rapport du Groupe international d'experts sur les ressources « International Resource Panel » nous utilisons 3 fois plus de matières à l'échelle du globe. Il y a diverse raison pourquoi cette exploitation accrue des ressources (nouvelle technologie, augmentation de la population...)

Pour mieux connaître le contexte, je vais vous donner quatre définitions clés de compréhension que j'utiliserai souvent dans mon rapport.

1.1 Exploitation des ressources naturelles

L'exploitation des ressources naturelles et une activité constante est vital pour tout être vivant, que ce soit animal, végétal ou humaine. Quand une ressource limitée naturellement dans le temps, il peut y avoir une surexploitation, voir une disparition de cette ressource.

1.2 L'énergie fossile

L'énergie fossile est une décomposition d'éléments organique provenant principalement d'animaux ou de plantes. Les énergies fossiles sont tirées principalement du charbon, du pétrole et du gaz naturel.

1.3 L'énergie renouvelable

L'énergie renouvelable est une énergie dont la nature se renouvelle en permanence. Et les tirés principalement de 2 grandes sources d'énergie naturelles que soit le soleil (via les cycles de l'eau, les marées, les vents ou les croissances végétaux), ou bien de la terre (géothermie, biomasse...).

On le surnomme l'énergie propre ou l'énergie verte, car l'exploitation engendre très peu de déchets ou d'émissions polluantes. Cependant, leur pouvoir énergétique est beaucoup plus faible que celui de l'énergie dites non renouvelable.

1.4 La transition énergétique

La transition énergétique a pour une transformation profonde le système de production et de consommation d'énergie afin de réduire l'impact environnemental et de favoriser un modèle énergétique plus durable.

Chapitre 2 Cahier des charges

Situation – Connaître l'états des lieux dans le monde la transition énergétique

2.1 Contexte

Enjeux

Comprendre le comportement de chaque pays de la consommation des énergies fossiles et de connaître les pays qui font la transition énergétique.

Stratégie

Observer l'extraction énergétique pour chaque période.

- ❖ Le nombre de populations si c'est en rapport avec leurs consommations
- ❖ Connaître les ressources
- ❖ Le volume par unité d'énergie

Domaine d'application : les top 10 pays qui auront une forte consommation d'énergie.

2.2 L'objectifs

Avoir le bilan pour chaque pays :

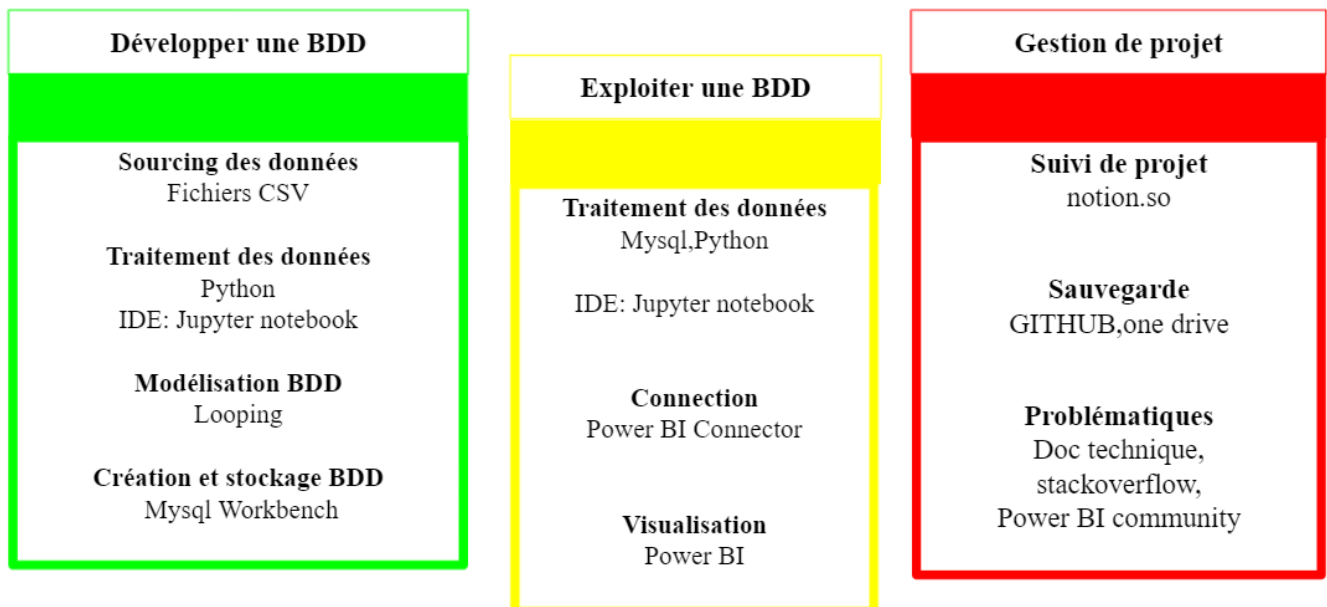
- ❖ Une visualisation de la consommation d'énergie fossile (le pétrole)
- ❖ Une visualisation de la consommation d'énergies propres
- ❖ Le nombre de population par pays

2.3 Contraintes

- **Contraintes sur le choix du sujet** : Avoir une bonne compréhension et lecture du sujet choisi
- **Contrainte de délais** : Correction des outils utilisés
- **Contrainte sur la base de données** : Trop grande quantité de donnée à traiter en la visualisation

2.4 Tâches des techniques

C'est un écosystème de données utilisés pour développer un programme.



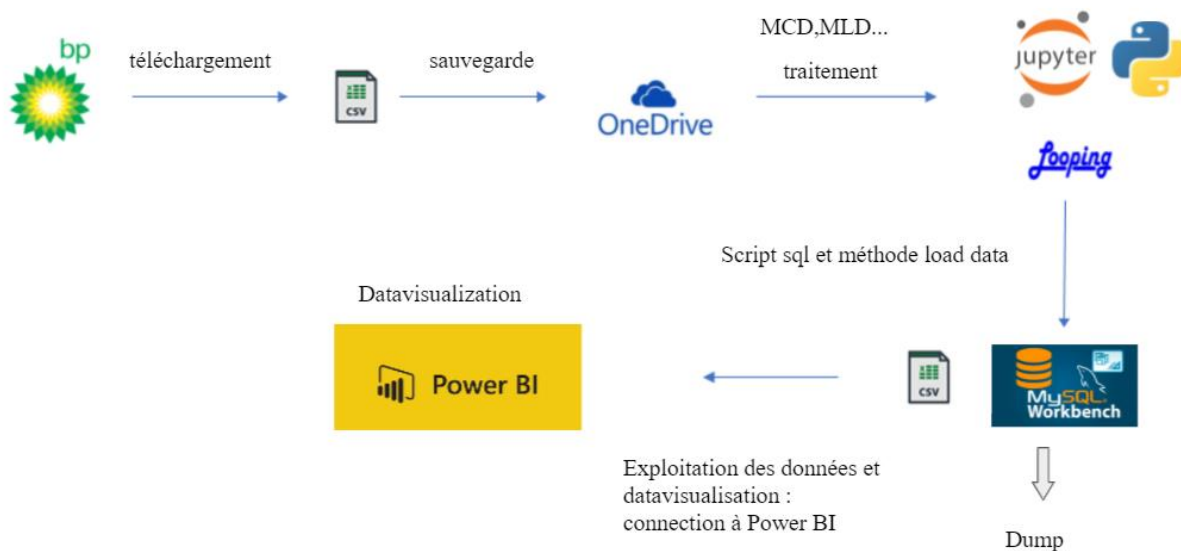
2.5 Langage et outils utilisés

<u>Outils</u>	<u>Utilisation</u>
Python	Langage orienté objet pour traiter la Data et développer des applications.
Jupyter Notebook	Application web pour exécuter du code : sert aux chargement, exploration, traitement et nettoyage des données.

MySQL Workbench 8.0	IDE SQL pour la conception de la base de données, la visualisation et administration de la base de données.
Power BI	Visualisation des données (Data Visualisation).
Notion.so	Gestion de projet, permet de visualiser les tâches selon la méthode Agile.
Github	GitHub est un site web conçu pour partager le code source d'un projet de développement d'application mis à œuvre par plusieurs programmeurs.
One Drive	OneDrive permet de stocker des fichiers personnels au sein d'un emplacement unique et aussi de partager avec d'autres personnes à n'importe quel appareil connecté à Internet.

Chapitre 3 Architecture en général

L'architecture générale, décrit de manière schématique le fonctionnement de mon programme.



Chapitre 4 Les différentes étapes de la préparation de la donnée

4.1 Collecte des données

La source des données pour constituer ma database proviennent de la source suivante : <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html>



- ❖ **Format** : CSV
- ❖ **Contenu** : Le nom des énergies par unités et les noms des pays figurant entre 1965 et 2019.
- ❖ **Détails** : 87 colonnes x 6508 lignes
- ❖ **Chemin du document** : Overview document/select a consolidated dataset/ Consolidated dataset – panel format csv

4.2 Description des données

Voici les exemples d'unités d'énergies suivante :

Nom de la colonne	Signification/Traduction	Description
oilcons_ej	oil consumption exajoule/Consommation pétrole en exajoule	Représente une unité d'énergie de consommation de pétrole d'un pays qui la possède.
oilprod_ej	oil production exajoule/ Production de pétrole en exajoule	Représente une unité d'énergie de production de pétrole d'un pays qui la possède.

Unité de valeur

Units

1 metric tonne = 2204.62 lb = 1.1023 short tons

1 kilolitre = 6.2898 barrels

1 kilolitre = 1 cubic metre

1 kilocalorie (kcal) = 4.1868 kJ = 3.968 Btu

1 kilojoule (kJ) = 0.239 kcal = 0.948

1 petajoule (PJ) = 1 quadrillion joules (1 x 10.15)

1 exajoule (EJ) = 1 quintillion joules (1 x 10.18)

Dans ce document, qu'il y a des pays qui n'ont pas les mêmes dispositions énergétiques. Nous trouvons des valeurs nulles à une certaine périodes. Que toutes les valeurs énergétiques numériques sont en float.

Afin de comprendre ma problématique, je me suis concentré les choses suivantes.

- ❖ La notion énergétique : Il y a une reconnaissance du préfixe comme gas/gaz, oil/pétrole ou suffixe comme prod / production, cons / consommation
- ❖ La notion géographique : le pays est identifiable par son nom mais surtout par le code ISO 3 qui est un code unique et international définissant le pays

4.3 Nettoyer et valider les données

Pour le nettoyage, j'ai importé les fichiers CSV sur un Jupyter Notebook et puis, j'ai sélectionné les colonnes pour créer les tables de la base de données.

Pour que mon csv puisse être valide, j'ai dû effectuer les choses suivantes :

- ❖ Supprimés les lignes en trop
- ❖ Accepté les valeurs zéro pour l'importation des données

Lorsque les données ont été nettoyées, j'ai pu répéter sur ces tables suivantes : Key_Minerals, Fossil_Energy et Clean_Energy.

❖ Accepté les valeurs zéros

```
Entrée [23]: df['lithium_kt']=df['lithium_kt'].str.replace(',','').astype(float).fillna(0)
df['cobalt_kt']=df['cobalt_kt'].str.replace(',','').astype(float).fillna(0)
df['graphite_kt']=df['graphite_kt'].str.replace(',','').astype(float).fillna(0)
df['rareearths_kt']=df['rareearths_kt'].str.replace(',','').astype(float).fillna(0)
```

❖ Supprimés les redondances, les choix des colonnes et définir le csv

```
Entrée [24]: df5 = df[['Country','ISO3166_alpha3','lithium_kt','cobalt_kt','graphite_kt','rareearths_kt','Years']].drop_duplicates()
df5
```

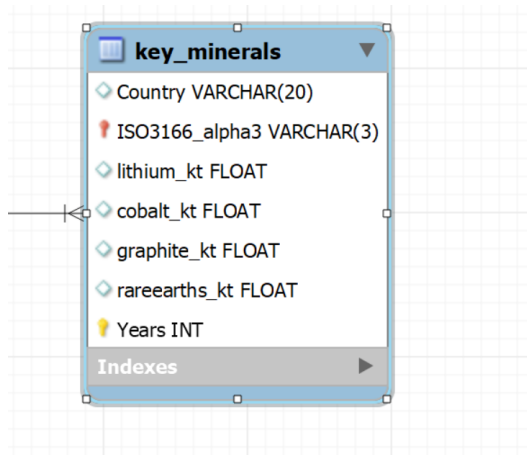
```
Out[24]:
```

	Country	ISO3166_alpha3	lithium_kt	cobalt_kt	graphite_kt	rareearths_kt	Years
0	Algeria	DZA	0.0	0.0	0.00	0.0	1965
1	Algeria	DZA	0.0	0.0	0.00	0.0	1966
2	Algeria	DZA	0.0	0.0	0.00	0.0	1967
3	Algeria	DZA	0.0	0.0	0.00	0.0	1968
4	Algeria	DZA	0.0	0.0	0.00	0.0	1969
...
6501	Zimbabwe	ZWE	0.9	0.0	7.00	0.0	2015
6502	Zimbabwe	ZWE	1.0	0.0	6.00	0.0	2016
6503	Zimbabwe	ZWE	0.8	0.0	1.58	0.0	2017
6504	Zimbabwe	ZWE	1.6	0.0	2.00	0.0	2018
6505	Zimbabwe	ZWE	1.6	0.0	2.00	0.0	2019

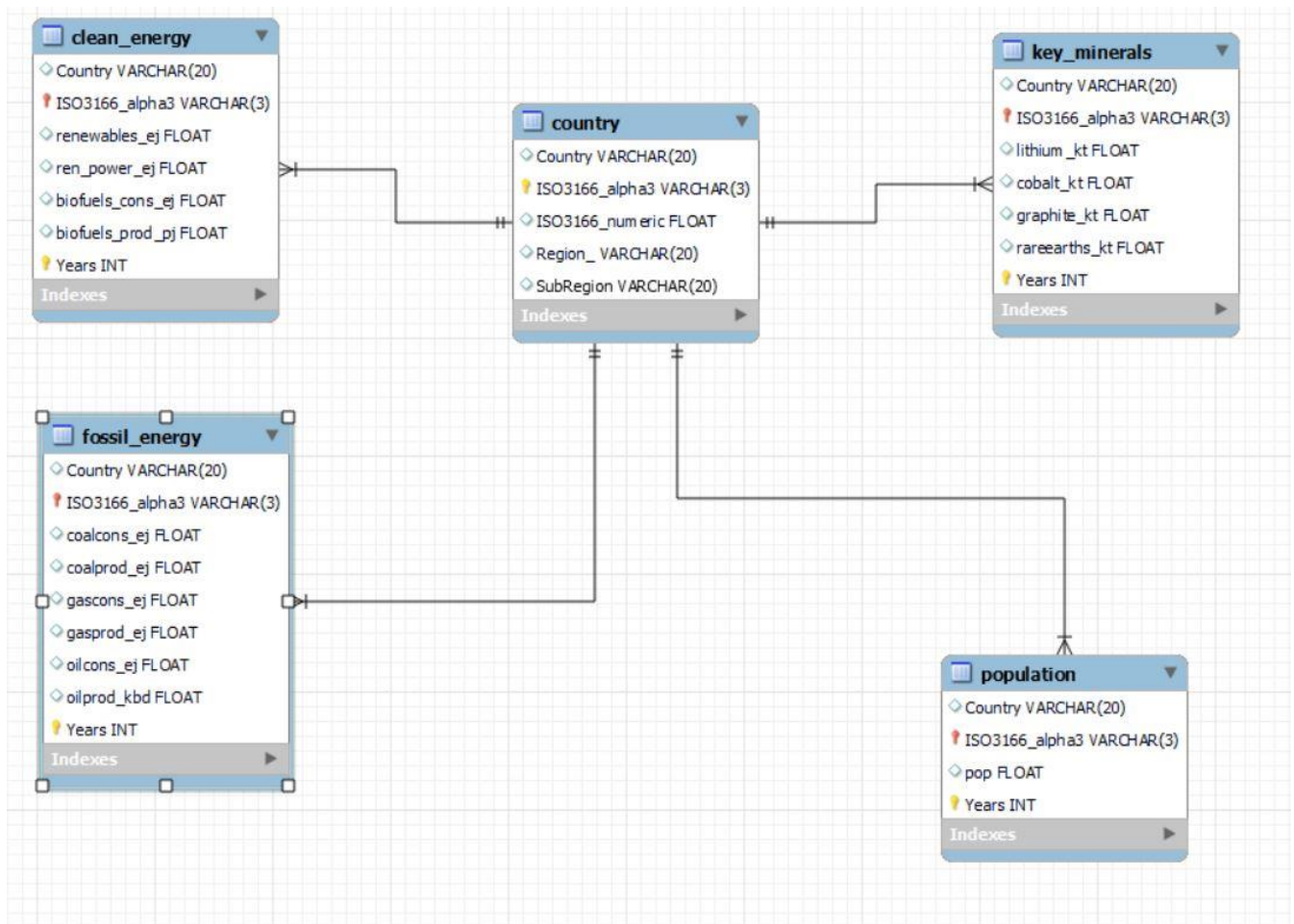
6506 rows × 7 columns

```
Entrée [25]: df5.to_csv('Key Minerals.csv',index=False)
```

❖ Insertion sur MySQL Workbench



Chapitre 5 Modèle physique de données (MCD)



CHOIX DU MODELE RELATIONNELS

- L'intégrité des données est primordial.
- Les CSV stocker sont intégrés avec les tables.

CLÉS, CONTRAINTES ET CARDINALITÉ

La clé primaire permet d'identifier de manière unique chaque enregistrement dans une table. Elle doit être UNIQUE et ne doit pas contenir de valeur NULL. Elle sert d'index.

Pour identifier de manière unique chaque enregistrement, les clés primaires sont les suivantes :

Nom des tables	Nom des colonnes	Fonctions
Country	❖ ISO3166	Clé primaire
Population	❖ ISO3166 ❖ Years	Clé composite (composée de plusieurs colonnes)
Fossil_energy	❖ ISO3166 ❖ Years	Clé composite (composée de plusieurs colonnes)
Clean_energy	❖ ISO3166 ❖ Years	Clé composite (composée de plusieurs colonnes)
Key_minerals	❖ ISO3166 ❖ Years	Clé composite (composée de plusieurs colonnes)

Les clés étrangères ont pour fonction principale la vérification de l'intégrité de votre base. Elles permettent de gérer des relations entre plusieurs tables, et garantissent la cohérence des données.

Les clés étrangères (fk pour foreign key) sont les suivantes :

Colonne dans les tables	Colonne de référence	Cardinalité
Population	Country (ISO3166)	1.n
Fossil_energy	Country (ISO3166)	1.n
Clean_energy	Country (ISO3166)	1.n
Key_minerals	Country (ISO3166)	1.n

Chapitre 6 Modèle organisationnel des données

Voici le descriptif du modèle organisationnel des données :

Table country

Nom des Colonnes	Numéro de colonne	Nature de la colonne	Descriptif
Country	1	Varchar	Nom des pays
ISO3166_alpha3	2	Varchar	Code ISO 3166-1 alpha-3 en trois lettres concernant le pays
ISO3166_numeric	3	Float	Code ISO 3166-1 alpha-3 à deux chiffres concernant le pays
Region_	4	Varchar	Région du pays
SubRegion	5	Varchar	Sous-région du pays

Table Fossil_Energy

Nom des Colonnes	Numéro de colonne	Nature de la colonne	Descriptif
Country	1	Varchar	Nom des pays
ISO3166_alpha3	2	Varchar	Code ISO 3166-1 alpha-3 en trois lettres concernant le pays
coalcons_ej	3	Float	Consommation de charbon par pays concernés
coalprod_ej	4	Float	Production de charbon par pays concernés
gascons_ej	5	Float	Consommation de gaz par pays concernés
gasprod_ej	6	Float	Production de gaz par pays concernés
oilcons_ej	7	Float	Consommation de pétrole par pays concernés
oilprod_kbd	8	Float	Production de pétrole par pays concernés
Years	9	Int	Données de 1965 à 2019

Table Clean_Energy

Nom des Colonnes	Numéro de colonne	Nature de la colonne	Descriptif
Country	1	Varchar	Nom des pays
ISO3166_alpha3	2	Varchar	Code ISO 3166-1 alpha-3 en trois lettres concernant le pays
renewables_ej	3	Float	Consommation d'énergie renouvelable par pays concernés
ren_power_ej	4	Float	Production d'énergie renouvelable par pays concernés
biofuels_cons_ej	5	Float	Consommation de biocarburant par pays concernés
biofuels_prod_pj	6	Float	Production de biocarburant par pays concernés
Years	7	Int	Données de 1965 à 2019

Table Key Minerals

Nom des Colonnes	Numéro de colonne	Nature de la colonne	Descriptif
Country	1	Varchar	Nom des pays
ISO3166_alpha3	2	Varchar	Code ISO 3166-1 alpha-3 en trois lettres concernant le pays
lithium_kt	3	Float	Matériaux nécessaires pour les énergies renouvelables (éolienne, solaire...)
cobalt_kt	4	Float	Matériaux nécessaires pour les énergies renouvelables (éolienne, solaire...)
graphite_kt	5	Float	Matériaux nécessaires pour les énergies renouvelables (éolienne, solaire...)
rareearths_kt	6	Float	Matériaux nécessaires pour les énergies renouvelables (éolienne, solaire...)
Years	7	Int	Données de 1965 à 2019

Table population

Nom des Colonnes	Numéro de colonne	Nature de la colonne	Descriptif
Country	1	Varchar	Nom des pays
ISO3166_alpha3	2	Varchar	Code ISO 3166-1 alpha-3 en trois lettres concernant le pays
Pop	3	Float	Nombre d'habitant par an
Years	4	Int	Données de 1965 à 2019

Chapitre 7 Requêtes

Comme objectif fixé au départ, voici un exemple une liste d'interrogation à requêter dans la base de données :

1. Vision globale :

- ❖ Évolution de la consommation de pétrole dans le monde.

2. Classement :

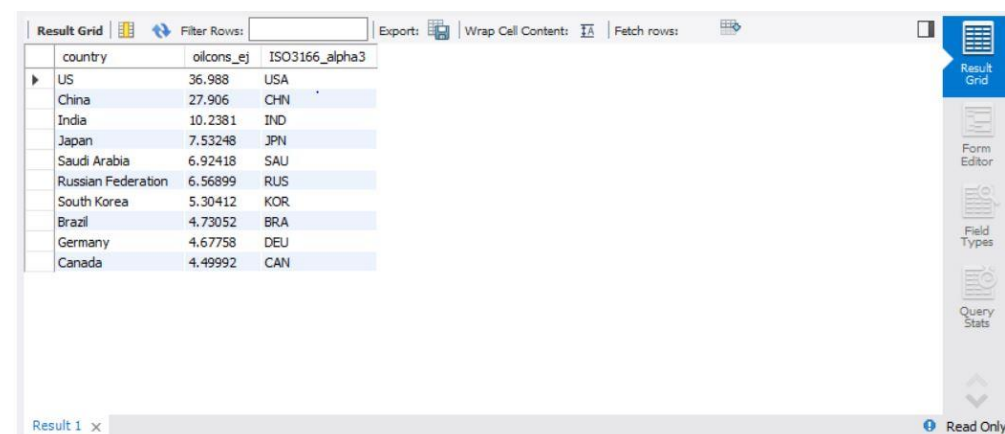
- ❖ Mesurer les top 10 des pays en 2019.



```
1 • select N.country,N.oilcons_ej,N.ISO3166_alpha3
2 FROM country AS C
3 JOIN fossil_energy AS N
4 on N.ISO3166_alpha3 = C.ISO3166_alpha3
5 WHERE N.country NOT LIKE '%Total%'
6 AND N.Years = 2019
7 ORDER BY N.oilcons_ej DESC
8 LIMIT 10;
9
```

Analyse de la requête :

- ❖ Select : la liste des colonnes utilisés
- ❖ From : la table choisie définit comme C
- ❖ Join : La connexion avec une table définit comme N
- ❖ On : relation entre les deux clés
- ❖ Where : résultat attendu
- ❖ And : l'année ciblée
- ❖ Order by : L'ordre par décroissant
- ❖ Limit : limité à 10 pays



country	oilcons_ej	ISO3166_alpha3
US	36.988	USA
China	27.906	CHN
India	10.2381	IND
Japan	7.53248	JPN
Saudi Arabia	6.92418	SAU
Russian Federation	6.56899	RUS
South Korea	5.30412	KOR
Brazil	4.73052	BRA
Germany	4.67758	DEU
Canada	4.49992	CAN

Chapitre 8 Sécurisation des données

Exportation des données

The screenshot shows the 'Data Export' wizard for a 'Local instance MySQL80'. The 'Object Selection' tab is active. Under 'Tables to Export', a list of schemas is shown with 'bibliothèque', 'bp_stats', 'fil_rouge1', 'foody', 'sakila', 'schema_name2', 'sys', and 'universite' selected. To the right, 'Schema Objects' for 'clean_energy', 'country', 'fossil_energy', 'key_minerals', and 'population' are also selected. The 'Export Options' section shows 'Export to Self-Contained File' selected, with the file path 'C:\Users\maxim\OneDrive\Documents\dumps\Dump20210819.sql'. Other options include 'Export to Dump Project Folder', 'Dump Stored Procedures and Functions', 'Dump Events', 'Dump Triggers', 'Create Dump in a Single Transaction', and 'Include Create Schema'. A 'Start Export' button is at the bottom right.

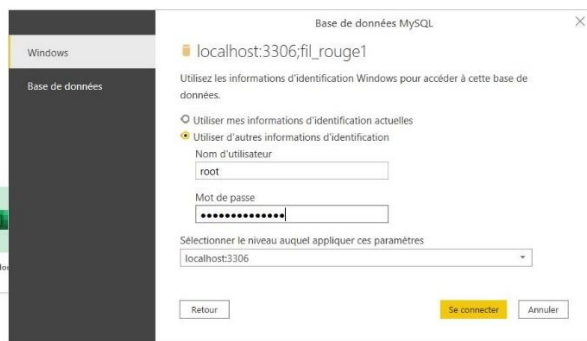
Importation des données

The screenshot shows the 'Data Import' wizard for a 'Local instance MySQL80'. The 'Import from Disk' tab is active. Under 'Import Options', 'Import from Self-Contained File' is selected, with the file path 'C:\Users\maxim\OneDrive\Documents\dumps\export.sql'. The 'Default Schema to be Imported To' section shows a dropdown menu and a 'New...' button. The 'Select Database Objects to Import' section has two empty lists for 'Schema' and 'Schema Objects'. At the bottom, there is a 'Start Import' button.

Chapitre 9 Visualisation

Power BI permet de créer du contenu attrayant et interactif ainsi qu'une visualisation pertinente améliorée. J'ai créé d'abord une connexion MySQL Workbench 8.0 qui est ma base de données et par la suite connecter à Power BI.

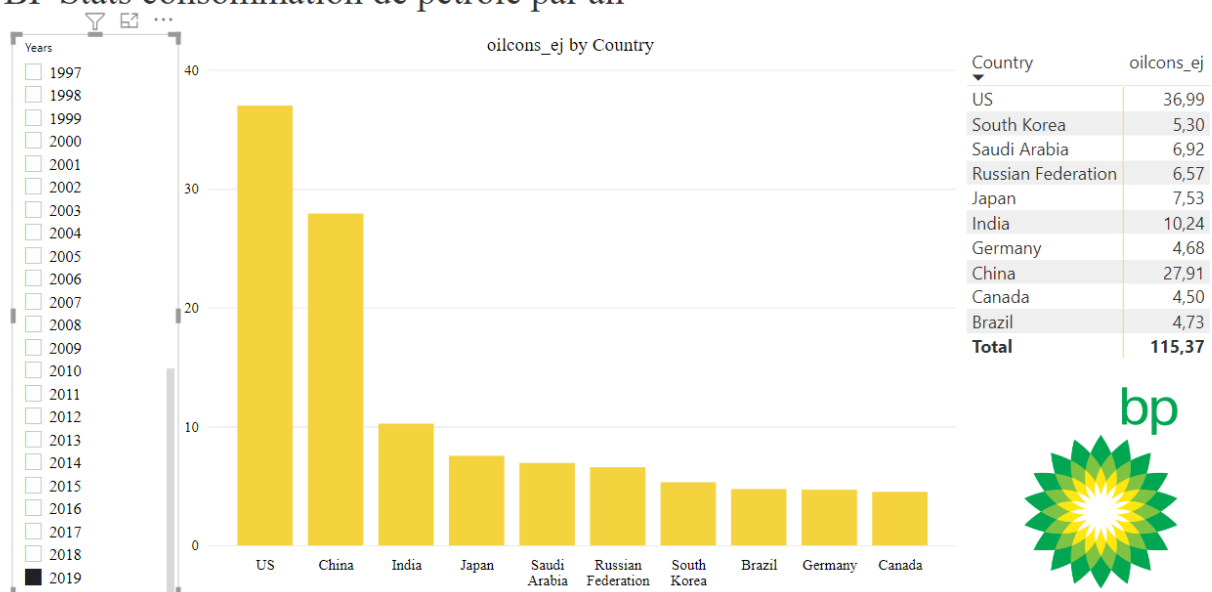
❖ Exemple d'une connexion d'une base de données



Dans la première image contient les résultats par an concernant le pétrole. Il permet de connaître l'évolution des top 10 pays en 2019.

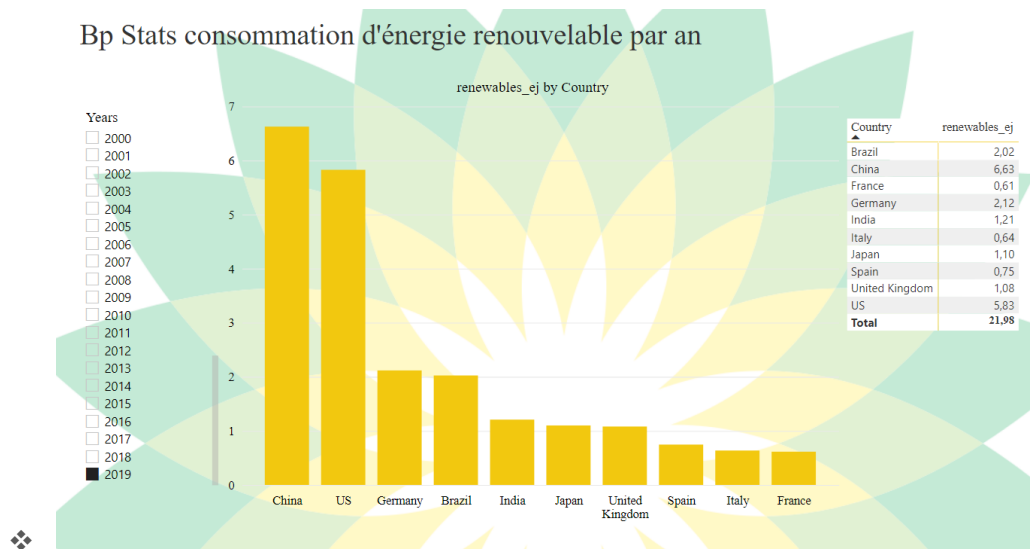
❖ Image 1

BP Stats consommation de pétrole par an



Cette deuxième image contient les résultats par an concernant la consommation d'énergie propre (géothermie, biomasse, solaire...). A la même fonction précédente, il permet de connaître l'évolution pour chaque pays.

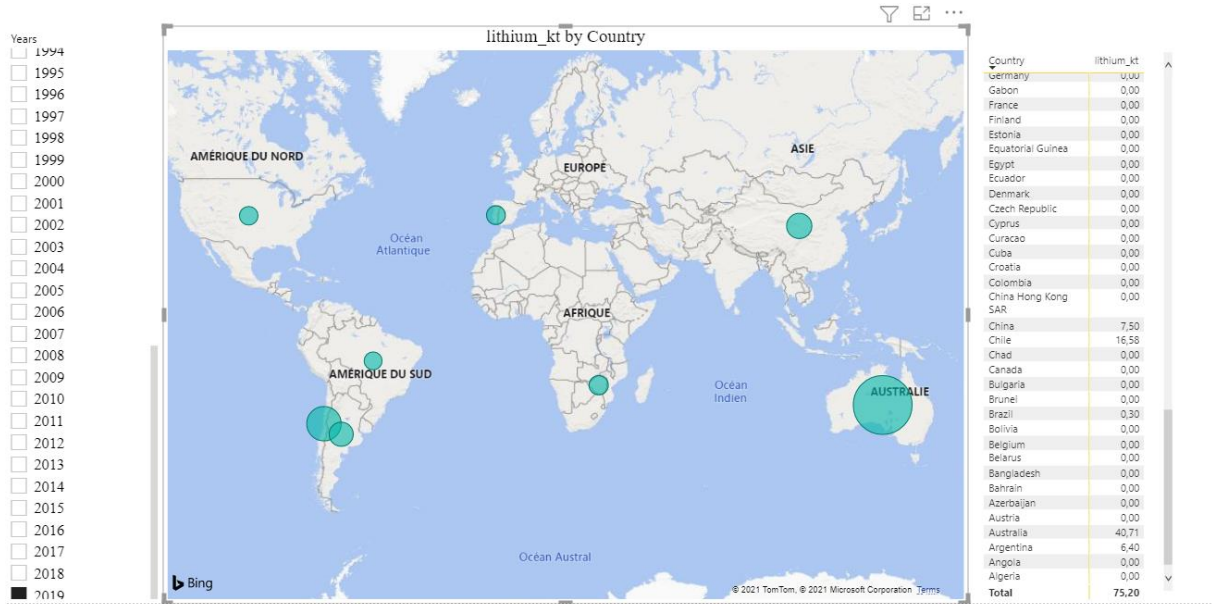
❖ Image 2



Dans la troisième image, une carte montre un aperçu d'où l'extraction est la plus forte. Ce type de ressource fait office de transition écologique (le lithium par exemple) qui est la clé de la fabrication des matériaux durables à l'énergie tel que les panneaux solaires ou l'éolienne.

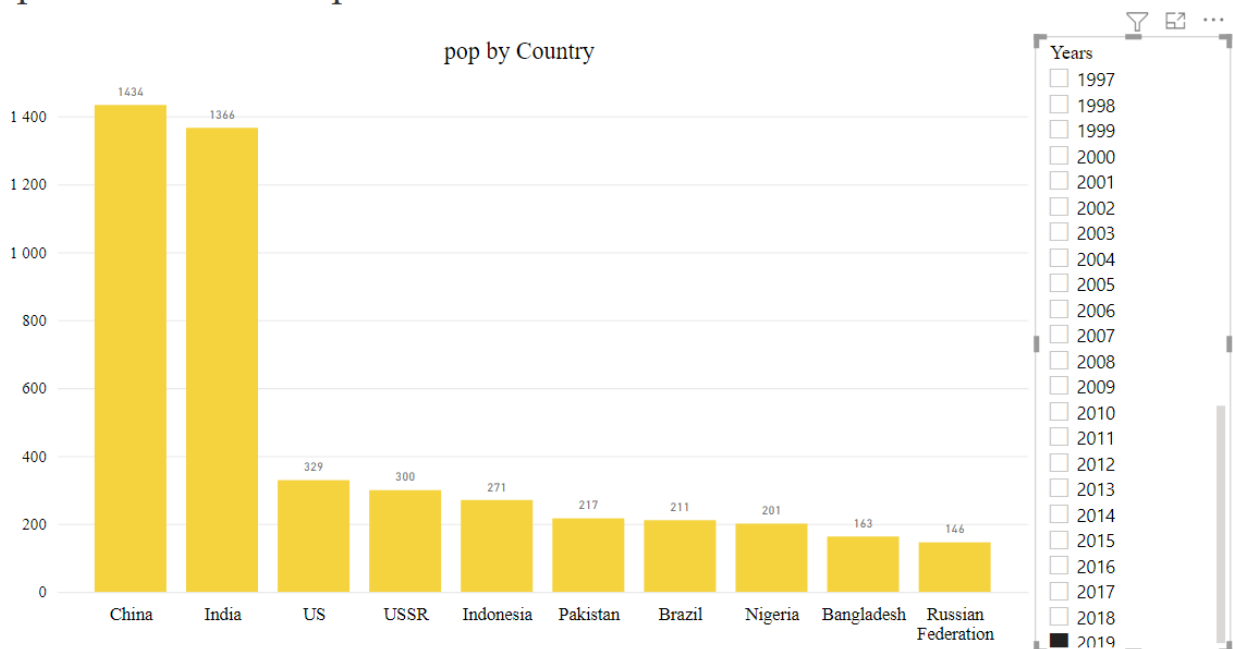
❖ Image 3

Transition d'énergie par an



Et enfin la dernière image, comprend le nombre de population par an

Population mondiale par an



Conclusion

Pour rappel, le but était d'apporter une vision plus claire d'un état des lieux sur la consommation à travers les pays.

Pour cela, les points abordés étaient :

- ❖ La consommation d'énergie fossile (le pétrole)
- ❖ La consommation d'énergies propres
- ❖ Le nombre de population par pays

Il y a une énorme variation entre le pays. Par exemple, les 3 pays en haut du podium des consommateurs de pétrole en sont également les pays les plus peuplés (voir image 1 et la dernière image). Le dernier constat est que ce sont ces mêmes pays sont à la tête du basculement vers l'énergie durable (voir image 2).

Des axes d'amélioration sont à apporter pour parfaire le projet :

- ❖ Connaître des pays producteurs d'énergie fossile ou d'énergie propres
- ❖ Comparer les pays consommateurs de bio carburant et de pétrole

Gestion du projet

Lien : [Planification Projet chef d'œuvre \(notion.so\)](https://www.notion.so/Planification-Projet-chef-d'œuvre)

Notion

La gestion du projet a été utilisée comme méthode agile (Kanban). Le site de planification est en anglais.

- ❖ To do/ A faire : regroupe de manière hiérarchique l'ensemble des tâches à faire. Celle qui est prioritaire est au-dessus.
- ❖ Doing / En cours : tâches en cours. Tant qu'elles ne sont pas terminées ou déplacées (abandon, blocage), aucune autre tâche n'est tirée de la rubrique à faire.
- ❖ Done/ Terminé : regroupe l'ensemble des tâches terminées.
- ❖ Archived/ Ressources : cette rubrique permet de retrouver pour chaque sujet ayant nécessité une recherche d'informations, de retrouver les liens ou références.
- ❖ Deserted/ Abandon : regroupe l'ensemble des tâches abandonnées.
- ❖ Competences / Compétences : regroupe l'ensemble des compétences nécessaires pour valider les certifications en cours.

Remerciement

Je souhaite remercier les 4 formateurs qui ont accompagné la promotion 2020/2021 Développeur.ses Data de Nanterre, sans qui ce projet n'aurait pu aboutir :

- MME Manel BOUMAIZA,
- M. Josselin TOBELEM,
- M. David AZRIA,
- M. Nicolas ZANFORLINI

Enfin, un grand merci aux 20 apprenants de la promotion 2020/2021 de Développeur.ses Data de Nanterre, avec qui j'ai passé 7 mois intenses mais passionnants.

Bibliothèque

Resources: [Downloads](#) | [Energy economics](#) | [Home \(bp.com\)](#)

Approximate conversion factors: [Approximate conversion factors \(bp.com\)](#)

Natural gas : [Natural gas – Statistical Review of World Energy 2021 \(bp.com\)](#)

Power BI: [Data Stories Gallery - Microsoft Power BI Community](#)