

MON PROJET BIGQUERY

BIGQUERY → Google Cloud

Projet:

**Quelle est la station essence la moins chère autour de chez moi pour le SP95?
Soit une société de transport de camion, il s'agit de voir tout au long du
chemin la station la plus proche pour permettre de prendre du carburant
moins cher/**

- **Celle qui ont des prix de sp95 au dessus de 1.95 sont interdites**

Outils:

Google Sheet /Bigquery Datawarehouse de Google / SQL ou Pandas (PYTHON) /
Dashboard de data analyse.

Etapes:

1. Récupérer les données carburant (csv)
2. Nettoyer les données dans Google Sheet
3. Chargée les données dans Bigquery Datawarehouse
4. Analyse dans SQL

1-Récupérer les données carburant (csv)

Sur le site de Data [gouv.fr](https://data.economie.gouv.fr); on click sur

<https://data.economie.gouv.fr/explore/dataset/prix-des-carburants-en-france-flux-instantane-v2/table/?flg=fr-fr> → on va sélectionner tableau

- Dans l'onglet **information**, je relie le modèle de données (la description des colonnes)
- je relis le **tableau** notamment les colonnes nécessaires
- Dans l'onglet **export** j'ouvre pour sélectionner le fichier à télécharger : je le télécharge en csv

2-Nettoyer les données dans Google Sheet

- Importer les données dans Google Sheet → Google Sheet: Fichier →importer
→glisser le fichier → Importer les données
- Nettoyage des données
 - ❖ Faire du cleaning (enlever les colonnes dont j'ai pas besoin ceux en JSON)
 - ❖ Mettre les noms de colonnes en minuscules, séparé par un trait d'union lorsque qu'il y a 2 mots (snake case);
 - ❖ google: latitude longitude à partir d'une adresse→ click Coordonnées GPS → sélectionne la 1ère ligne géom que l'on place sur la barre long,lat de la feuille GPS , puis on valide

- ❖ Donc on va supprimer les colonnes longitude et latitude en gardant juste geom que l'on va scinder en 2: Google Sheet: donnée → scinder la colonne en 2; → écrire (latitude (F)longitude(G));

3-Charger les données dans BigQuery

- Aller sur mon compte BigQuery: on tape sur Google: Cloud Bigquery → console → = > click droit bigquery → Créer un nouveau projet nommé Mon projet bigquery valider
 - ❖ click sur les 3 points → create dataset nommé fr_carburants
 - ❖ sur la table fr_carburants click sur les 3 petits points → create table → importer(upload), format: CSV, table fr_carburants, auto detect (on coche),
 - ❖ Sur CSV, nommer la feuille: fr_carburants
- Table créée → click sur preview

4 A-Analyse SQL dans BigQuery

- Afficher toute la table
`SELECT * FROM `fr_carburants.fr_carburants` → (la table .dans le dossier)`
- Afficher seulement quelques colonnes de la table
- Trouver le prix min/max du SP95 (documentation fonction d'agrégation bigquery, que je peux traduire en français)
- Renommer les colonnes
- Afficher les stations essence du département: 83
- Compter les stations essences du département : 83
- Afficher les stations essences du département 83 qui ont du SP95
- Trouver la station la moins chère pour le SP95 du département 83

B-Analyse SQL dans BigQuery

Calculer la distance entre les stations essence et chez moi ?

- Latitude et longitude
- Récupérer la latitude/ longitude de mon adresse avec Google Maps
 Mon adresse 6 Avenue de Verdun, 92260 Fontenay aux roses, France
 - ❖ Sur Google Coordonnées GPS: je saisis mon adresse (6 avenue de verdun 92260 Fontenay aux roses)
 - ❖ Sur Chatgpt on tape : Donne latitude longitude

4 CRÉER LE DASHBOARD (ÉTAPES)

Macro (problématique)

- ❖ Répartition des stations essence dans la région (83) et qui ont du SP95:

1. Définir les métriques

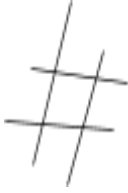
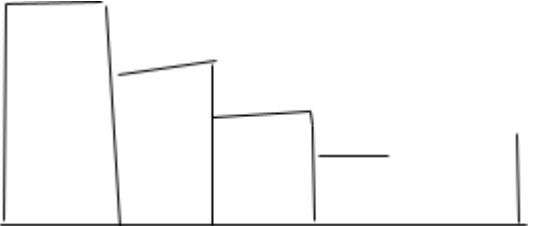
2. Dessiner le dashboard

MÉTRIQUE 1:

Nombre de stations essence

MÉTRIQUE 2:

Répartition des prix


 <p>Stations essence 50</p>	<p>Répartition des prix du SP95</p> 
--	--

MÉTRIQUE 3:

Tableau des stations essence

MÉTRIQUE 3:

Carte des stations

<p>Tableau des</p>  <p>stations (adresse, ville, prix sp95, Distance (km))</p>	<p>Carte des stations essences</p> <p>Report carte GPS</p>
--	--

3. Préparer les données

D'après notre croquis, on a besoin:

- ❖ L'adresse complète des stations (pour le tableau) et la ville
- ❖ Le prix SP95 (arrondi au centime près)
- ❖ La distance entre chez moi et la station essence en km

Les données ce qu'on avait précédemment, voici nos données préparées:

#ROUND() recréer une nouvelle table pour la visualisation (dashboard), dans bigquery

```
SELECT
  adresse,
  ville,
  CONCAT (adresse, ' ', code_postal, ' ', ville) AS adresse_complete,
  ROUND (prix_sp95, 2) AS prix_sp95,
  ROUND (ST_DISTANCE(ST_GEOGPOINT(2.3157953, 48.8115313),
ST_GEOGPOINT(double_field_6, latitude))/1000, 2) AS distance_in_kms,
FROM `fr_carburants.fr_carburants`
WHERE code_departement='83'
AND prix_sp95 IS NOT NULL
ORDER BY distance_in_kms ASC;
```

LE TABLEAU FINAL

Le projet est enregistré sur une nouvelle table: / (JE LE REND PUBLIC voir chatgpt)

SELECT

```
CONCAT (adresse, "", code_postal, "", ville) AS adresse_complete,
ville,
latitude,
double_field_6,
ROUND (prix_sp95, 2 ) AS prix_sp95 ,
ROUND (ST_DISTANCE(ST_GEOGPOINT(2.3157953, 48.8115313),
ST_GEOGPOINT(double_field_6, latitude))/1000, 2) AS distance_in_kms,
FROM `fr_carburants.fr_carburants`
WHERE code_departement='83'
AND prix_sp95 IS NOT NULL ;
```

4. Créer le dashboard

❖ Ajouter en source de données la requête SQL précédente

Je vais partir d'un template de dashboard que j'ai fait (sinon je crée sur looker directement);

❖ Ajouter la source de données au dashboard

<https://lookerstudio.google.com/navigation/reporting> → +(click) → source (il faut connecter une base de données) → requête personnalisée → click sur le projet initial → sélectionner la dernière requête du tableau à analyser → click sur associé, renommer le projet sur looker (au dessus) → Créer un rapport → ajouter des données → mes sources de données (click sur le projet)

❖ Créer les charts 1 par 1 (dans le dashboard)

➤ Créer la station à essence (métrique 1) :

Click sur créer un graphique, choisir nombre → click sur le graph sélectionné, choisir la configuration et le style, pour donner un titre à chaque graphique: métrique, puis renommer le graphique.

➤ Création de la carte de station (Graphique Map) :

Click créer un graphique, choisir map → click sur le graph → configurer champs (introduire adresse complète) / configuration → emplacement → adresse complète / style → standard / Configuration → info-bulle (prix SP-95) → Ajouter une légende intensité couleur : configuration → métrique couleur (prix sp-95) / Varier l'intensité selon les kms: style → couleurs (maximale, moyenne, minimale) / style → définir comme vue par défaut.

➤

❖ **Ajouter les filtres**

Ajouter les filtres:

- Filtres distances : Ajouter un selecteur → curseur (Déposer ss station essence)
- Filtres prix SP95 : Pareil → on peut styliser les filtres

❖ **Enjoy**

➤