

Dev Front

Projet Dataviz 2

Projet

Création d'une plateforme de visualisation de données

- Objectifs :
 - Dégager une **problématique** sur un thème choisi
 - Créer une **pipeline ETL automatisé** pour exploiter un jeu de données
 - Afficher les données sortie de l'ETL sur un site web
 - Rendre ces données **compréhensible** et **lisible** pour n'importe qui grâce à la *data visualisation interactive*
 - Présenter les résultats en expliquant leurs significations
 - Répondre à la **problématique** annoncée

Problématique

- **Problématique** = question visant à **comprendre** quelque chose
 - *Comment ?*
 - Comment faire quelque chose ?
 - Comment évolue tel indicateur depuis tel événement ?
 - *Quoi ?*
 - Quel sont les causes de tel constat
 - Quel est l'impact de tel donnée sur tel autre donnée
- Vous pouvez définir des **sous questions**
- Vous pouvez proposer une **hypothèse** avant étude

Problématique

- Toutes les réponses à votre problématique ne se trouveront pas dans le dataset
- Il faudra probablement **faire des recherches** pour **comprendre** les raisons de l'évolution des données
- Utilisez tous les outils à votre disposition pour comprendre vos données et **expliquer les résultats**

Exemple : *Le nombre de cas COVID en france baisse après le 17 mars 2020*

Explication : *C'est la date du premier confinement*

Thèmes - émissions de CO₂ par pays

Jeu de données :

- [Our World in Data - CO₂ Emissions](#)
- [Banque Mondiale - Données climatiques](#)

Possibilités :

- Visualiser l'évolution des émissions de CO₂ par pays ou région.
- Comparer les émissions par habitant et par secteur (énergie, transport, etc.).
- Identifier les pays les plus polluants et ceux qui réduisent leurs émissions.

Thèmes - Trafic routier

Jeu de données :

- [OpenStreetMap - Données de trafic](#)
- [Data.gouv.fr - Trafic routier en France](#)

Possibilités :

- Analyser les flux de trafic en temps réel ou historique.
- Identifier les heures de pointe et les axes les plus fréquentés.
- Corréler le trafic avec des événements (vacances, grèves, etc.).

Thèmes - Football

Jeu de données :

- Kaggle - European football 2008 - 2016

Possibilités :

- Analyser les performances des joueurs/équipes.
- Visualiser les statistiques (buts, passes, victoires).
- Identifier des tendances (ex : impact de l'âge sur les performances).

Thèmes - Données de santé publique (COVID)

Jeu de données :

- [WHO - Données COVID-19](#)
- [Santé Publique France - Données épidémiologiques](#)

Possibilités :

- Visualiser la propagation de la maladie par région/pays.
- Analyser l'impact des mesures sanitaires (confinement, vaccination).
- Comparer les taux de mortalité et de guérison.

Thèmes - Transport en commun

Jeu de données :

- [RATP - Données Open Data](#)
- [Data.gouv.fr - Transports en commun](#)

Possibilités :

- Visualiser la fréquentation des lignes de métro/bus.
- Identifier les heures de pointe et les retards.
- Optimiser les trajets

Thèmes - consommation énergétique

Jeu de données :

- [ENEDIS - Données de consommation électrique](#)
- [Eurostat - Énergie](#)

Possibilités :

- Visualiser la consommation par région ou type de bâtiment.
- Identifier les pics de consommation et leur cause (météo, heure).
- Comparer les sources d'énergie (renouvelable vs fossile).

Thèmes - films et séries (IMDb, Netflix)

Jeu de données :

- [Kaggle - IMDb Movies Dataset](#)
- [Netflix - Données publiques](#)

Possibilités :

- Analyser les tendances de production (genres, pays).
- Visualiser les notes et popularité des films/séries.
- Identifier les acteurs/réaliseurs les plus prolifiques.

Thèmes - pollution de l'air

Jeu de données :

- [AirParif - Qualité de l'air en Île-de-France](#) (clé API OK)
- [OpenAQ - Données mondiales](#)

Possibilités :

- Visualiser la qualité de l'air par ville/région.
- Corréler la pollution avec des facteurs (trafic, industrie, météo).
- Identifier les périodes de pic de pollution.

Projet

- Etapes:
 - Formation des groupes (**3-4 obligatoire** 7 groupes max)
 - Choix d'un sujet et de la problématique
 - Création de l'ETL
 - Création du site web
 - Interprétation et explication des résultats et réponse à la problématique
 - Présentation du sujet à l'oral vendredi

Projet

- Choix des graphes :
 - Trouvez les plus pertinent pour expliquer votre sujet
 - Expliquez pourquoi avoir choisi tel type de graph
 - Cherchez **d'autre graphs** que ceux vu en cours
 - Utilisez **d'autre librairie** que *chart.js*
 - <https://datavizcatalogue.com/>

Projet

- Création du site :
 - Doit être **clair**, bien **organisé**, et **bien construit**
 - Utilisez plusieurs pages pour montrer plusieurs sujets
 - Rendez les graphs clair et **lisible** facilement pour pouvoir les présenter à l'oral
 - Donnez une **réflexion cohérente** à partir des data montrées

Projet

- Oral :
 - 10-15 min présentation
 - 20-30 min de questions réponses **techniques ET fonctionnelles**
 - **Préparez vous** en amont
 - Répartissez bien le temps de parole
 - **Expliquez et justifiez** toutes les étapes de votre projet :
 - Choix du sujet, problématique que vous allez expliquer
 - Choix de la / des sources
 - Choix des graphs
 - Résultats

Projet

- Conseils :
 - Je suis là pour vous guider, **utilisez cet aide** à toute les étapes
 - Affiner le sujet, trouver des données, choisir les graphes ...
 - Ne confiez pas tout à un *LLM*, il ne vous aidera pas lors de l'oral
 - **Communiquez** entre vous, sachez qui fait quoi et pourquoi
 - **Comprenez l'ensemble du projet**
 - Je ne veux pas entendre "c'est pas moi qui ai fait cette partie" à l'oral
 - C'est un projet de groupe, vous devez tous le maîtriser entièrement
 - Je poserai des **questions ciblées** lors de l'oral

GO!