

Atelier 1 - Sources d'un modèle prédictif de trafic

Pour établir un modèle prédictif de trafic sur la rocade bordelaise, vous devez constituer un jeu de données d'entraînement. L'atelier consiste à identifier, qualifier et ingérer les sources d'information qui composeront ce jeu de données dans un datalake.

Étape 1 - Périmètre d'analyse

Faire un brainsorming pour établir une liste des facteurs influant plus ou moins directement sur le trafic routier de la rocade bordelaise. En déduire une liste par ordre décroissant d'importance de ces facteurs.

Étape 2 - Sources de données

Pour les informations de trafic ainsi que chaque facteur identifié dans l'étape précédente, proposer une source de données gratuite (Open Data, ...) ou payante. Préciser également si un travail spécifique (recoupement, analyse, ...) et ou des ressources (application, stockage, ...) sont nécessaires pour l'exploitation de ces données.

Étape 3 - Ingestion dans un Datalake

Installer le module python duckdb avec pip .

Ouvrir une connexion persistente (<https://duckdb.org/docs/api/python/overview.html#persistent-storage>)

Charger les 2 ou 3 sources principales et facilement disponibles de données en vous inspirant éventuellement de ce code :

```
CREATE OR REPLACE TABLE nom_table AS
SELECT * FROM read_csv_auto(
    'fichier.csv',
    normalize_names=True
)
```

Remise

Pdf contenant le nom des membres du groupe et le détail de chaque étape avant la fin de semaine à sylvain.labasse@mail-formateur.net.