



OBJECTIFS

Objectifs pédagogiques

Bienvenue dans ce projet qui va s'étaler sur 5 jours de formation.

Vous allez avoir l'honneur de travailler sur des données issues de la Métropole Grenobloise qui n'ont jamais été mises à la disposition du grand public. Ces données traitent d'une partie de la gestion des déchets au sein de la Métro.

Vous allez pouvoir travailler sur une problématique assez large de Business Intelligence. Voilà une trame simplifiée afin que vous puissiez mieux comprendre le déroulé de votre travail.

Démarche pédagogique

- Gestion de projet
- Découverte et compréhension des données
- Détermination des problématiques auxquelles vous souhaitez répondre
- Mise en place d'un Datawarehouse
- Réalisation d'un ou plusieurs dashboards
- Présentation de vos travaux

Compétences développées :

Vous allez être confronté à une vraie problématique de Business Intelligence. Seule la partie de récupération des données est simplifiée pour vous car nous vous fournissons les données "en dur" via des fichiers csv.

Au cours de ce module vous allez apprendre à :

- Gérer un projet en vue de livrer un produit BI
- Analyser des données et en faire émaner des problématiques
- Importer des données au format .csv dans une base de données
- Créer un modèle avec vos données afin d'enrichir un Datawarehouse
- Transformer des données en SQL
- Créer un ou plusieurs dashboard pérennes afin de répondre aux problématiques choisies.
- Restituer votre travail et l'expliquer intelligiblement.

MODALITÉS

Livrables

- 1 repository GIT contenant votre code (EDA, dbt)
- 1 Dashboard
- 1 support de présentation (format de votre choix).
- 1 documentation (format de votre choix).

Une présentation orale en binôme d'une durée de 15 min suivie d'une dizaine de minutes de questions.

Durée

Module : 5 jours au total

Lancement le 05/08/2024 et clotûre le 09/08/2024.

Formateur(s)

Mickaël Vitry

TRAME

		Planning	Activités
05/08	MODULE	Formateur 1	Démarrage du module. Analyse exploratoire, définition des objectifs et modélisation.
06/08	MODULE	Formateur 1	Suite de la modélisation. Prise en main des outils.
07/08	MODULE	Autonomie	Construction du DWH, connection à l'outil de Dataviz
08/08	MODULE	Formateur 1	Création Dashboard.
09/08	MODULE	Formateur 1	Finalisation projet et documentation. 15h: Présentation des travaux.

Jour 1&2

Modalités

- Travail en binôme
- 2 jours en présentiel

Livrables

Un repo GIT avec le code de l'EDA.

Les problématiques auxquelles vous souhaitez répondre: le début de votre documentation.

Les schémas (DWH) et maquettes (Dashboards)

Objectifs

1. Avoir une démarche produit.
2. Comprendre et exploiter les données.
3. Définir les problématiques à résoudre.
4. Formaliser sous forme de schémas le DWH et modéliser les dashboards
5. Les outils sont installés et opérationnels

L'importance du travail en binôme va prendre tout son sens ici. Alors avant de foncer tête baissée : on s'organise, on se répartit les tâches et surtout on garde du temps chaque jour pour partager les avancées de chacun. N'hésitez pas à utiliser les outils de votre choix pour le suivi et la répartition des tâches.

Au bout de ces 2 jours les outils devront être installés et opérationnels, n'hésitez pas à me solliciter pour remplir cet objectif, le 3ème jour vous êtes en autonomie.

1.0 – Démarrage du module

0.5h – Présentiel

Pré-requis: [git](#), [docker](#) et [docker-compose](#)

1.1 – Analyse exploratoire des données

3 h – Présentiel

Nous avons mis à votre disposition des fichiers .csv sur le drive. Téléchargez les et analysez- les en binôme avec l'outil de votre choix.

Prenez le temps de comprendre, faites des manipulations, des tests pour vous assurer d'avoir bien saisi la substantifique moelle des données.

A l'issue de cette analyse, vous devez être capable de formuler des problématiques auxquelles vous souhaiteriez répondre.

RESSOURCES

- Les fichiers csv

COMPÉTENCES ASSOCIÉES

- Comprendre des données ainsi que leur contexte associé.
- Réaliser un schéma UML récapitulant les relations entre les 6 fichiers CSV.

1.2 – Définir les modèles pour construire votre DWH.

2-4h – Présentiel

Avant de vous lancer dans le grand bain, il va falloir prendre du recul sur vos problématiques et comprendre comment organiser vos données au sein de votre BDD afin d'optimiser leur utilisation dans l'outil de Dataviz.

La base d'un ou plusieurs dashboard bien réalisés réside dans la construction de votre DWH qui est la source des données montrées dans vos Dashboard.

RESSOURCES

- Rappels sur la modélisation dimensionnelle (schéma en étoile, ...) :
 - <https://www.cartelis.com/blog/data-warehouse-modelisation-etoile/#essentiel-modelisation-etoile-data-warehouse>
 - <https://www.techtarget.com/searchdatamanagement/definition/star-schema>

COMPÉTENCES ASSOCIÉES

- Etablir un modèle (sur papier) afin de construire un DWH

2.1 – Définir la maquette de vos Dashboards.

2h – Présentiel

Avant de se lancer dans le dashboarding, une étape importante est celle de définir une ébauche de ce que l'on veut faire. Cela permet d'éviter de se perdre dans ses idées et de partir dans tous les sens. Ce qu'on attend de vous : vous avez fixé une ou plusieurs problématiques en début de module. Le dashboard est censé être le support qui vous permettra d'y répondre. Encore faut-il que tout soit clair dans votre tête pour ne pas faire quelque chose de contre productif. Alors on reprend un stylo et on se pose des questions :

- A qui je veux que ce dashboard s'adresse ?
- Plusieurs onglets ? Si oui, un onglet = une problématique ? Autre ?
- Comment je vais remplir la page de chaque onglet ? Quel KPI ?
- Sentez vous libre de brainstormer sur d'autres points qui vous paraissent utiles à anticiper.

RESSOURCES

- Internet

COMPÉTENCES ASSOCIÉES

- Faire une ébauche d'un Dashboard

2.2 — Intégration des données dans votre Base de Données**2h — Présentiel**

Vous allez maintenant pouvoir vous connecter à votre BDD.
Installer PostGreSQL sur une de vos machines.

! Lors du lancement du container, penser à mapper les ports pour accéder à la base et à monter un volume pour garder les données en local.

Afin de vous connecter, nous vous conseillons d'installer DBeaver community.
Créer une base de données et un nouveau schéma pour votre projet.
Enfin, trouvez le moyen d'injecter vos fichiers csv sous forme de table dans le schéma.
Prenez soin de vérifier le typage, l'encodage et autre de vos données, lors de la création des tables dans votre BDD.

RESSOURCES

- [Image docker de PostgreSQL](#)
- DBeaver : <https://dbeaver.io/>

COMPÉTENCES ASSOCIÉES

- Importer des données au format .csv dans une base de données

2.2 — Installer et configurer DBT**1h — Présentiel**

DBT sera votre outil principal qui vous permettra de construire votre DWH. Il est maintenant venu l'heure de l'installer et de le configurer. Référez-vous à la documentation (je vous laisse lire le tuto DBT préparé par Martin : DBT_tuto.pptx).

RESSOURCES

- <https://docs.getdbt.com/docs/introduction>
- DBT_tuto.pptx

COMPÉTENCES ASSOCIÉES

- Installer et configurer DBT

2.3 — Réaliser votre premier modèle DBT

2h — Présentiel

On y est ! Les choses sérieuses commencent... Si vous avez bien pensé votre modèle, vous aurez compris que vos données “raw” dans le schéma “ods” vont avoir besoin d’être transformées. C’est maintenant que la puissance de DBT va prendre tout son sens. Avant de passer à la construction du DWH, il est important de s’échauffer en réalisant un ou plusieurs premiers modèles très simples afin de comprendre la mécanique DBT. Sans surprise, qui dit base de données dit ... SQL.

RESSOURCES

- <https://docs.getdbt.com/docs/>

COMPÉTENCES ASSOCIÉES

- Réaliser un premier modèle DBT

2.4 — Connecter l’outil de Dataviz à votre DWH

2h — Présentiel

Installer votre outil de Dataviz sur de vos machines. Ouvrez votre outil de Dataviz et trouvez un moyen de connecter votre DWH.

RESSOURCES

- Internet

COMPÉTENCES ASSOCIÉES

- [Installer Superset](#)
- Connecter un outil de Dataviz à un Datawarehouse.

Jour 3

Modalités

- Travail en binôme
- 1 jour en présentiel

Livrables

Le code du projet dbt dans le repo GIT

Le DWH est alimenté par dbt

La documentation produit

Objectifs

- Création d'un DWH
- Création d'une documentation produit

3 – Construction du DWH + documentation

Journée – Présentiel

L'objectif du jour est clair : le DWH que l'on avait imaginé sur papier doit prendre forme grâce à dbt.

Il est temps aussi d'alimenter la documentation du produit.

Attention ! Cette journée remplie de PostgreSQL doit s'achever avec l'utilisation d'une ou plusieurs fonctions intégrées à vos modèles DBT. Un point sera fait pour que vous montriez qu'une ou plusieurs fonctions sont présentes et qu'elles fonctionnent correctement dans vos modèles.

RESSOURCES

- Internet

COMPÉTENCES ASSOCIÉES

- Création d'un Datawarehouse
- Utiliser une ou plusieurs fonctions postgresQL dans un modèle DBT
- Création d'une documentation produit

Jour 4

Modalités

- Travail en binôme
- 1 jour en présentiel

Livrables

Les dashboards dans l'outil de Dataviz

La documentation produit

Objectifs

- Création de dashboards
- Création d'une documentation produit

4 — Construction des Dashboards + documentation

Journée — Présentiel

L'objectif du jour est de créer vos dashboards et de compléter votre documentation. Votre travail doit être compréhensible par les gens qui l'utiliseront. Pour ce faire, ne sous estimez pas l'intérêt des démarcations, des titres, de la disposition de vos visualisations...

RESSOURCES

- Internet

COMPÉTENCES ASSOCIÉES

- Création de dashboards
- Création d'une documentation produit

Jour 5

Modalités

- Travail en binôme
- 1 jour en présentiel

Livrables

Le code du projet dans le repo GIT

Le dashboard

La documentation produit

Le support de présentation

Objectifs

- Création d'un support de présentation
- Création d'une documentation produit

5.1 — Finalisation travaux et documentation

½ Journée — Présentiel

L'objectif du jour est de finaliser vos travaux, la documentation et de créer le support de présentation.

COMPÉTENCES ASSOCIÉES

- Créer un support de présentation

5.2 — Présentation des travaux

½ Journée — Présentiel

C'est la clôture de ce module, vous présentez vos travaux et rendez les livrables (Dashboard et documentation).

COMPÉTENCES ASSOCIÉES

- Communication des résultats

Bonus - jour ∞

∞.1 – Découvrir, transformer et exploiter de nouvelles données.

NOUVELLES DONNÉES

La Métro a pu voir que vous avancez bien et perçoit en vous tout le potentiel afin de répondre à une problématique majeure en interne.

Vous avez pu remarquer que le sujet des circuits réalisés est une véritable problématique pour la Métro. Vous allez avoir la tâche complexe de trouver une solution qui permette d'attribuer le bon circuit aux véhicules ayant réalisés leurs tournées.

Pour ce faire, rendez vous sur votre drive où vous trouverez :

1. L'intégralité des circuits théoriques.
2. Une semaine de données de tous les circuits réalisés sur cette période.

Prenez le temps de regarder les données. Ensuite votre première mission est de concaténer les données afin de terminer avec **deux tables dans votre base de données**. Une table qui recense tous les circuits théoriques, une autre de tous les circuits réalisés.

⚠ ⚠ ⚠ Les coordonnées de la table des circuits théoriques doivent être au même format que celles des circuits réalisés fournis en .csv.

Cette phase sera réalisée en python.

RESSOURCES

- Votre drive

COMPÉTENCES ASSOCIÉES

- Transformer des données géographiques en des tables exploitables dans votre outil de dataviz.

∞.2 — Enrichissement du Dashboard de base

— Présentiel

Maintenant que vous avez réalisé un premier dashboard à l'aide des données de la Métro. Il est temps de savoir comment vous pouvez l'enrichir à votre manière en ajoutant des données qui vous semblent pertinentes. C'est une phase très intéressante dans la BI car c'est à ce moment précis que vous amenez une véritable valeur ajoutée. C'est aussi ce qui vous permet de laisser libre cours à votre créativité en prenant toujours en compte les problématiques auxquelles vous souhaitez répondre.

ATTENTION : tout ajout de données doit être sourcé et documenté. Lorsque l'on prend le parti de rajouter des données car on estime qu'elles ont de la valeur pour un projet, le gros risque est d'utiliser des données totalement ou partiellement fausses. Dans tous les cas, il est de votre responsabilité de donner l'intégralité des informations disponibles sur les données ajoutées.

RESSOURCES

- Internet

COMPÉTENCES ASSOCIÉES

- Enrichir un Datawarehouse avec des nouvelles données

∞.3 — Nouveau Dashboard sur les nouvelles données

Maintenant que les données des circuits théoriques et réalisés sont en base sous la forme de deux tables et que vous avez développé un algorithme qui permet de faire une association entre un circuit réalisé et un circuit théorique, il faut prendre le temps de créer un datawarehouse avec un schéma en étoile en DBT. Puis vous pourrez construire un dashboard sur Power BI pour mettre en valeur vos travaux.

RESSOURCES

- Internet

COMPÉTENCES ASSOCIÉES

- Enrichir un Datawarehouse avec des nouvelles données