

Projet DataWarehouse – Thématique : Analyse mondiale

Le projet BIUM vise à mettre bout-à-bout les différentes notions vues dans l'UE (ETL, DatawareHouse, OLAP, Reporting, Visualisation) et utiliser également des concepts vus dans d'autres cours (ML, RITAL, IAMSI, ...)

Le projet doit analyser une problématique de votre choix au niveau mondial (santé, éducation, économie, politique, militaire, ...). Vous êtes libres de la problématique. *Elle doit cependant être assez "difficile" à résoudre : par exemple, un seul histogramme ne permet pas de répondre à la problématique.*

Exercice 1 – Choix du sujet et collecte des données

Des exemples de problématiques (anciens projets d'étudiants) :

- Où dois-je implanter mon entreprise (activité à définir) ?
- Quel est le lien entre discours politiques et situation économique du pays ?
- Quel est le pays que vous conseilleriez pour vos prochaines vacances ?
- Comment se définit la notion de liberté dans chaque pays ? Quel est le lien avec l'économie ?
- Attentats et niveau de démocratie ? Quel est le lien ?
- Y a-t-il une corrélation entre les violences commises dans le pays, et le niveau d'éducation et du niveau de vie de ses habitants ?
- Comment rendre mes posts visibles sur Twitter ?
- Quel est le lien entre violence, jeux vidéos et situation économique ?
- Sommes nous tous égaux en termes de prise en charge médicale ?

Les contraintes du projet sont les suivantes :

- Le projet est organisé en binôme (un ou deux groupes de 3 sont tolérés, avec une exigence plus conséquente).
- Le projet doit être centré sur une problématique claire d'étude (que vous devez définir)
- Le projet doit être construit à partir de la base de données relationnelle "world" <http://downloads.mysql.com/docs/world.sql.gz>
- Une partie des données doit être issue d'une capture Web (API, Open Data, extraction pages HTML (<http://www.garethjames.net/a-guide-to-web-scrapping-tools/>)). Vous pouvez aussi exploiter des portal open data des organisations internationales :
 - <http://www.who.int/gho/database/en/>
 - <https://data.world/datasets/>
 - <https://datacatalog.worldbank.org/dataset/global-economic-monitor>
 -
- Une attention particulière doit être portée à la dimension Localisation (et éventuellement Temps) qui doit être la plus fine possible

Exercice 2 – Les étapes du projet

- Data Scientist : Définir les axes d'analyse et les méthodes/techniques à utiliser pour réaliser ces analyses. Ces étape est totalement décorrélée de la qualité des données. Réfléchir "dans un monde idéal" ce que vous feriez.
- Business Intelligence : A partir de ces besoins, réaliser la modélisation du schéma en constellation de votre base de données multidimensionnelle. Une modélisation conceptuelle et logique (normalisée et dénormalisée) de schémas en étoile. Il sera apprécié d'avoir éventuellement plusieurs schémas et/ou un schéma en constellation agrégeant plusieurs schémas en étoile
- Data Ingeneer : Définir les pré-traitements à effectuer et les réaliser avec Dataiku. Importez vos

données dans la base de données en suivant la modélisation du schéma en constellation.

- Data Analyst : Réaliser les analyses à effectuer pour répondre à la problématique. On vous demandera de générer, entre autres, un dashboard complet (titres/axes d'analyse et graphiques). Vous devez également inclure a minima une carte géographique dans vos analyses.
- Data storytelling : Racontez votre projet... N'oubliez pas que vous vous adressez à un décideur d'entreprise!

Exercice 3 – Ce qu'il faut rendre

A la fin du projet, vous devez restituer :

- Un rapport technique destiné à un Responsable de Projet afin de faire valider vos axes d'analyse, votre modélisation, les transformations et les différentes analyses/commentaires.
- Une soutenance orale avec slides, destinée à un Dirigeant d'entreprise. Prendre en compte ses objectifs décisionnels et son niveau de technicité, tout en lui démontrant que vous avez fait correctement les choses.
- Un dashboard restituant la synthèse de vos analyses.