Projet Entrepôts de données

Pierrick Moreau, Nicolas Marechal

# Introduction

* Explication de l’objectif du projet

# 1ère étape : choix des données

* <https://www.data.gouv.fr/fr/datasets/base-de-donnees-accidents-corporels-de-la-circulation-sur-6-annees/>
* Déjà deux sources de données différentes : accidents et véhicules impliqués
* **Données gouvernementales** : données intègres et complètes puisque ces données sont mises à jour directement par les forces de l’ordre lors de la constatation de l’accident.
* **Aucunes données inutiles**: chaque accident à au moins un blessé.
* **Données récentes** (2006-2011) et incluant tous les départements français (DOM-TOM inclus)
* Données complètes -> de nombreuses problématiques pertinentes possibles

SUJETS CHOISIS :

* Quel est nombre d’accidents par an en France ? Quels départements sont les plus accidentées ?
* Quel est le nombre de morts sur la route ? Quels départements sont les plus meurtriers ?
* Corrélation avec le type de véhicule (moto, scooter, voiture, camion).

# 2ème étape : création de la base de production

1. Insertion des données dans la base de production
2. Filtrage des données afin de supprimer les informations inutiles ou incomplètes
   1. Suppression de colonne (V1, V2, gps, lat, long, coderivoli, distancemetre, PR, PR1)

# 3ème étape : mise en place de l’ETL (Talend)

1. Pourquoi Talend ?
2. …

# 4ème étape : modélisation des cubes de données

1. Sujet 1
2. Sujet 2
3. Sujet 3

# 5ème étape : exploitation des cubes et mise en place du reporting

1. Sujet 1
2. Sujet 2
3. Sujet 3

# Conclusion