

# Data'clear

Denis Beauget et Hebiret Hayaat

November 2020

## Présentation

Data'clear est un projet en lien avec la science des données (Data Science). La difficulté de la gestion des données aujourd'hui est leur traitement et leur analyse pour permettre des interprétations.

La création et la récupération de Datasets et de Database est courante et facile d'accès pour les entreprises (leur ventes, la fréquentation de leur magasin..) mais leur traitement et leur interprétation reste difficile et souvent couteux.

Data'clear a pour but de créer une interface (web/logiciel) permettant à une entreprise ou un particulier de faire de l'analyse sur un jeu de données (récupérer des éléments simple comme : la plus grande valeur d'un jeu de données, des graphiques ou des calculs...). Créer un "pont" entre les données (Database/Datasets) et leur interprétation à travers une interface pédagogique.

En bref, Data'clear serait une solution logiciel ressemblant soit un ensemble de script léger permettant des analyses rapides ou une interface plus global prenant en entrée des données (D'une base de donnée SQL par exemple) et permettant des analyses en sortie.

## Technologies

- Python Pandas (traitement des données : <https://pandas.pydata.org/>)
- Machine learning (<https://scikit-learn.org/stable/>)
- Si interface bureau : Java/C++/Python..
- Si interface web : Python (Flask/Django)/Javascript (React..)
- Si gros volume de données (Pyspark pour Hadoop)

## Objectif

- Interprétation de Datasets fixé au début du projet
- Comprendre et utiliser les technologies de Data Science
- Introduction au machine learning si besoin
- Proposer une analyse sur les résultats obtenu

## Cas pratique

Je suis une entreprise ayant une base de données MySQL qui est rempli de mes ventes de 2019 et j'aimerais en faire une analyse complète.

Notre projet y répondra en 3 étapes :

- Récupération des données et pré-traitements (Python Pandas)
- Analyse et machine learning (Python Scikit par exemple)
- Affichage des résultats (interface web/Dataframe/fichier .csv..)