

**Crunching de data**

**Rapport**



Réalisé par : **Mercha Maryem**

**Introduction :**

Ce rapport comporte les différentes étapes de réalisation de l’exercice du Crunching de Data, qui a les objectifs suivants :

* Extraire des infos intéressantes type :
  + Prix min, moyen et max
  + Durée min, max, moyenne par trajet
* Différence de prix moyen et durée selon le train, le bus et le covoit selon la distance du trajet

**Réalisation :**

**Langage de programmation utilisé :**

Dans la programmation de cet exercice j’ai utilisé le langage demandé dans l’énoncé qui est Python, vu sa facilité de manipulation.

Concernant les bibliothèques utilisées dans ce travail j’ai utilisé :

Pandas : manipulation des données sous forme de data frame

Sklearn.tree : contient les différents types d’arbres de décision à utilisé

Matplotlib : contient les outils de construction des graphes

Datetime : consiste à manipuler les données de type date/heure.

Les étapes de réalisation :

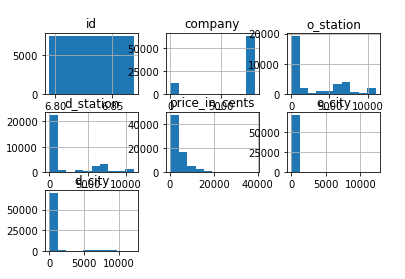
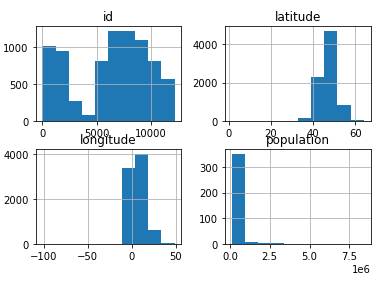
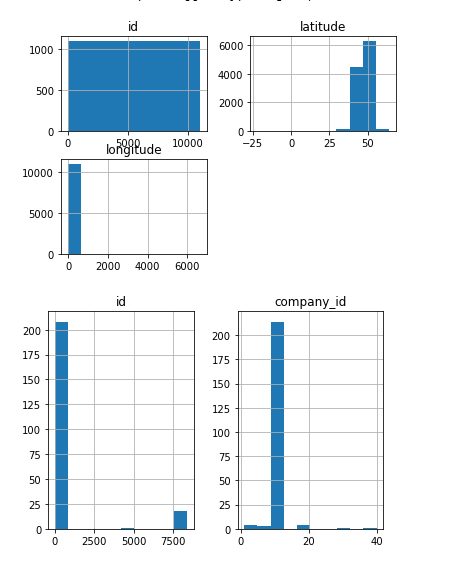
1. **Importation des données des fichiers :**

Dans cette étape j’ai enregistré les données de chaque fichier dans une variable en utilisant la méthode de la bibliothèque pandas read\_csv.

1. **Affichage des informations et des détails de chaque data frame :**

Pour avoir plus d’informations sur les données de chaque fichier, j’ai affiché leurs contenus à l’aide de la fonction head(), et pour afficher les détails des colonnes des fichiers comme le min, le max et la moyenne.

1. **Affichages des histogrammes des colonnes :**



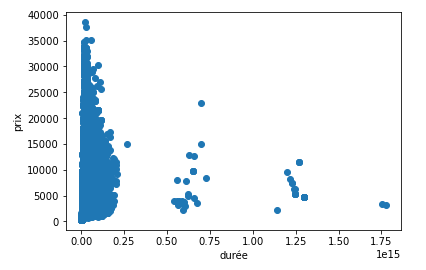
1. **Extraction des informations concernant le prix et la durée :**

Après la réalisation de ces traitements j’ai pu extraire les informations suivantes :

* Le prix max en cents : 38550
* Le prix moyen en cents : 4382.711
* Le prix min en cents : 300
* La durée max : 20 days 12:51:00
* La durée moyenne : 0 days 07:04:37.247600043
* La durée min : 0 days 00:20:00

1. **Affichage du rapport prix/duration sous forme d’un graphe :**

Ce graphe représente la relation entre le prix et la durée des voyages.



1. **Interprétation des résultats :**

On peut constater que plus que la durée du voyage est petite plus que le prix est élevé, mais on peut aussi remarquer qu’il y'a un autre facteur qui impacte le prix appart la durée, notre objectif sera donc de savoir ce facteur*.*

1. **Interprétation des informations du type de transport :**

Les résultats nous montrent que le type de transport a aussi un impact sur les prix des voyages, puisque on peut constater que les voyages réalisés à l'aide des carpooling sont les plus chers, suivis par ceux effectués avec le bus et finalement les moins coûteux sont les voyages effectués par les trains

1. **Création d’un modèle de prédiction des prix :**

Pour la résolution de ce problème j’ai choisi l’algorithme « DecisionTreeRegressor » parce que c’est un algorithme de régression et on général dans ce genre d’exercices on utilise la régression puisque on veut prédire une valeur continue, aussi parce que c’est un des algorithmes qui génèrent des bons résultats.