

Pour ce projet j'ai opté pour la programmation orienté objet pour l'optimisation qu'elle offre par contre le code est en notebook pour faciliter la visualisation des données. S'il était question de machine learning .

1. Description des colonnes

Chaque ligne correspond à un trajet de covoiturage, c'est à dire un couple passager / conducteur. A chaque passager est donc affecté un trajet.

Exemple : un conducteur réalise un déplacement avec deux passagers différents au sein de son véhicule, le nombre de trajets réalisés est de 2. Ceci se traduit par deux lignes.

- **journey_id**
Identifiant unique affecté à un couple passager/conducteur.
- **trip_id**
Identifiant permettant de recouper plusieurs couples passager/conducteur dans un même véhicule.
Exemple : c4124bb1-d8a4-487c-b4d9-367b931ee8ce
- **journey_start_datetime**
Exemple : 2019-10-31T23:00:00.000Z
Date et heure du départ au format ISO 8601 (YYYY-MM-DDThh:mm:ssZ). L'heure est exprimée en UTC (Coordinated Universal Time). UTC n'est pas ajusté sur l'heure d'été et hiver !
- **journey_start_date** **nouveau** **à** **compter** **de** **aout** **2020**
Exemple : **2019-10-31**
Date départ au format YYYY-MM-DD
- **journey_start_time** **nouveau** **à** **compter** **de** **aout** **2020**
Exemple : **02:20:001**
Heure de départ au format hh:mm:ss
- **journey_start_lat**
Exemple : 48.725
Latitude du point de départ (prise en charge passager) comprise entre 90deg et -90deg décimaux en datum WSG-84
- **journey_start_lon**
Exemple : 2.261
Longitude du point de départ (prise en charge passager) comprise entre 180deg et -180deg décimaux en datum WSG-84
- **journey_start_insee**
Exemple : 91377
Code INSEE commune ou arrondissement du point de départ (prise en charge passager).
- **journey_start_town**
Exemple : Massy
Commune du point de départ (prise en charge passager)

- journey_start_department **nouveau à compter de aout 2020** Exemple : 38*
Département du point de départ (prise en charge passager)
- journey_start_country
Exemple : France
Pays du point de départ (prise en charge passager)
 - journey_end_datetime
Exemple : 2019-10-31T23:15:00.000Z
Date et heure de l'arrivée au format ISO 8601 (YYYY-MM-DDThh:mm:ssZ). L'heure est exprimée en UTC (Coordinated Universal Time). UTC n'est pas ajusté sur l'heure d'été et hiver !
 - journey_end_date **nouveau à compter de aout 2020**
Exemple : 2019-10-31
Date d'arrivée au format YYYY-MM-DD
 - journey_end_time **nouveau à compter de aout 2020**
Exemple : 02:20:001
Heure d'arrivée au format hh:mm:ss
 - journey_end_lat
Exemple : 48.695
Latitude du point d'arrivée (dépose passager) comprise entre 90deg et -90deg décimaux en datum WSG-84
 - journey_end_lon
Exemple : 2.162
Longitude du point d'arrivée (dépose passager) comprise entre 180deg et -180deg décimaux en datum WSG-84
 - journey_end_insee
Exemple : 91122
Code INSEE commune ou arrondissement du point d'arrivée (dépose passager).
 - journey_end_department **nouveau à compter de aout 2020**
Exemple : 38
Département du point d'arrivée (dépose passager)
 - journey_end_town
Exemple : Bures-sur-Yvette
Commune du point d'arrivée (dépose passager)
 - journey_end_country
Exemple : France
Classe de preuve
 - operator_class
Exemple : C
Pays du point d'arrivée (dépose passager)

- **passenger_seats** : nouveau à compter de aout 2020
Exemple :
Nombre de sièges réservés par l'occupant passager

2. Analyse

Après avoir effectué une analyse approfondie des données, les constatations suivantes ont été mises en évidence :

- Dans le cadre d'une modélisation de machine learning la colonne "has_incentive" peut être utilisée comme cible.
- **Colonne "passenger_seats"** : Cette colonne présente une unique valeur identique dans toutes les instances. Dans le contexte de la modélisation d'un modèle de machine learning, elle n'ajoutera pas de valeur significative.
- **Heures de pointe** : En analysant les colonnes "journey_start_datetime" et "journey_end_datetime", les heures de pointe pour les départs sont 7h, 8h, 16h, 17h et 18h. Pour les arrivées, les heures de pointe sont 7h, 8h, 17h et 18h. Ces heures de pointe sont définies en utilisant un seuil de 1000 covoiturages par heure correspondante.
- **Colonne "pays"** : Il n'y a pas de trajets internationaux, ni de départs ni d'arrivées de Rouen vers l'international.
- **Départements avec le plus de covoiturages** : En examinant les données, il est apparu que le département 76 est en tête en termes de nombre de covoiturages, suivi du département 27.
- **Groupes de villes ("towngroup")** : Parmi les 259 groupes de villes analysés, la "Métropole Rouen Normandie" détient la première position, suivie de "CA Seine-Eure".
- **Incitations au covoiturage** : Sur l'ensemble des covoiturages répertoriés, seuls 32 d'entre eux ont déclaré être incités.
- **Distance parcourue** : La distance minimale parcourue en une journée est de 1838 unités, tandis que la distance maximale atteint 781647 unités, avec une moyenne de 23725.623 unités.
- **Durée des trajets** : La durée la plus courte d'un trajet est d'une unité de temps, tandis que la plus longue atteint 442 unités de temps, avec une moyenne de 27.429213 unités.

Remarque : Ce document constitue une analyse préliminaire relative à l'exploration et à la visualisation des données. Pour des informations plus détaillées, veuillez-vous référer au cahier (notebook) et au fichier HTML associés qui sont fournis en complément de cette synthèse.