Projet Machine Learning : Prédiction du nombre de taxis

dans une position

**Objectif :**

Ayant les coordonnées d’un point dans une map, on prédit le nombre de taxis qu’on peut trouver là-bas.

**Description de la data :**

D’abord on a commencé par choisir 12 endroits se trouvant à Agdal-Irfan pour concevoir une simple map (en utilisant Google Maps) :

1-Centre de radiologie

2-National library

3-Arribat center

4-Lycée

5-Mosquée Badr

6-BMCE Bank

7-McDonald's

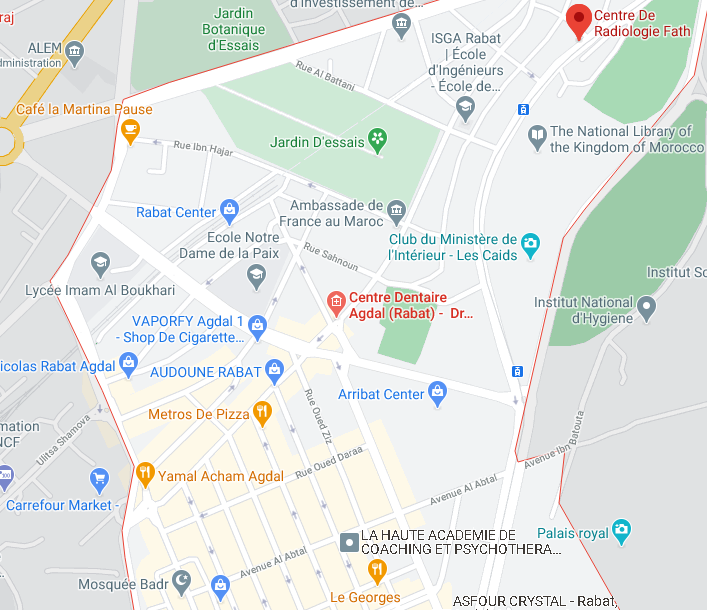
8-Parc Agdal

9-Pharmacie Agdal

10-Campus Irfan

11-INPT

12-Hôpital Militaire







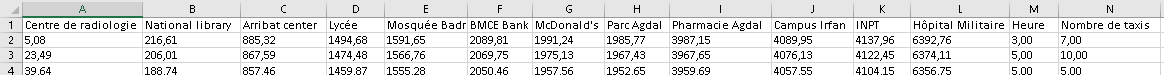


**Map obtenue :**

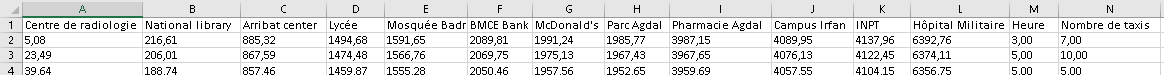


Les coordonnées d’un point dans cette map sont obtenues par rassembler 12 valeurs ; la ième valeur parmi ces 12 représente la distance (en m) entre ce point et le ième endroit parmi les 12 endroits qu’on a cité au-dessus.

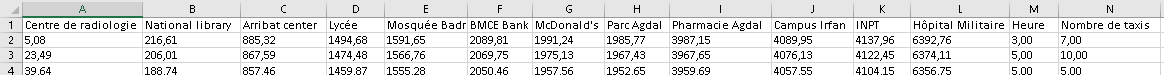
Ces 12 coordonnées ne sont que les 12 premiers features de notre dataset :



Le 13ème feature exprime l’heure où on est à la recherche d’un taxi :



La dernière colonne est la sortie/cible (nombre de taxi trouvés) :



**Data d’entrainement(taxi\_train) (576x14) : X(576x13) et y(576x1)**

**Data de test (taxi\_test) (288x14) : X(288x13) et y(288x1)**

**Approche d’apprentissage adoptée :**

Régression linéaire multi-variables :

* Les valeurs de y ne sont pas discrètes et on ne peut pas les classer dans un nombre raisonnable de classes.
* Plusieurs features.

**Programme 1 :**

Dans le premier programme « taxi1.py » on a suivi les étapes suivantes :

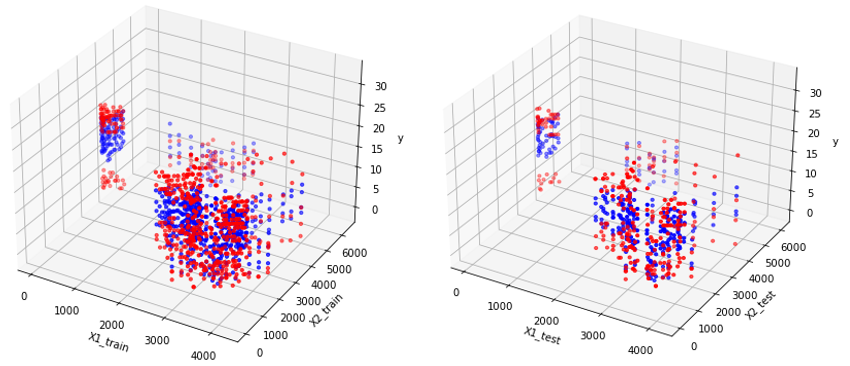
* Importer les datasets et les bibliothèques nécessaires.
* Créer un modèle de régression linéaire multi-variables et l’entrainer sur le dataset d’entrainement.
* Visualisation des résultats.
* Utiliser le dataset de test.
* Visualisation des resultats.
* Évaluation des performances.

**Note :**

* Data originale (rouge)
* Prédiction du modèle (bleu)

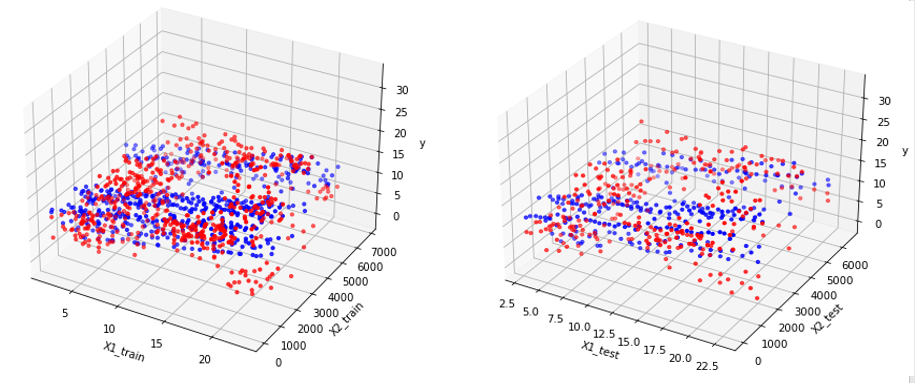
**Exemple 1 :**

**Entrainement** **Test**



**Exemple 2 :**

**Entrainement**  **Test**



**Evaluation des performances :**

Le coefficient de détermination pour les deux datasets est égale à 0.5.

**Programme 2 :**

Afin d’améliorer notre modèle nous avons pensé à utiliser la technique de feature scaling pour normaliser les valeurs de chaque colonne des features. En effet :

Xij = (Xij-mean j) /std j

Où :

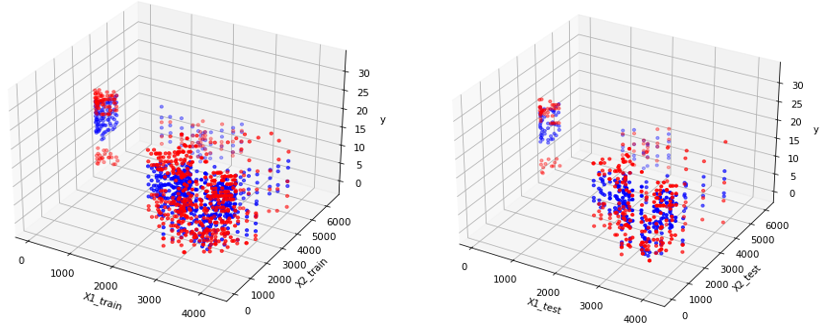
* Mean j : moyenne de la jème colonne
* Std j : écart-type de la jème colonne

On a introduit ces modifications dans le deuxième programme « taxi2.py ».

**Exemples des résultats :**

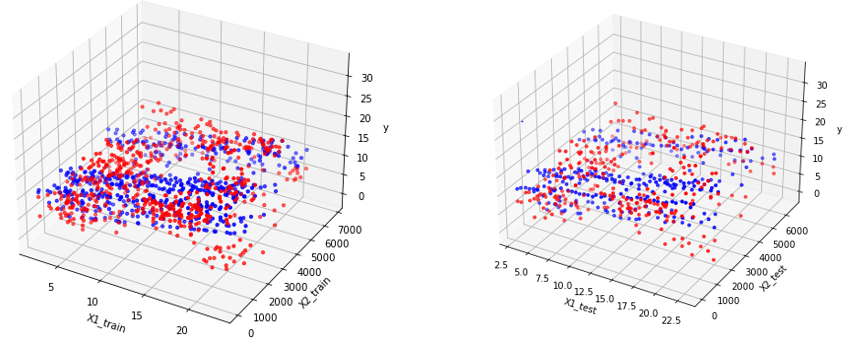
**Exemple1 avec modification :**

**Entrainement** **test**



**Exemple2 avec modification :**

**Entrainement**  **test**



Même après avoir utilisé cette méthode, Le coefficient de détermination pour les deux datasets est encore égale à 0.5.