# Le projet

Le projet que nous avons choisi est intitulé Bike Sharing Demand et proposé par Kaggle.

Le système de partage de vélos au sein d’une ville est très simple : plusieurs kiosques sont réparties dans la ville, où des personnes (abonnés ou non), peuvent louer un vélo et faire le trajet qu’elle souhaite jusqu’à un autre kiosque.

Les données qui sont réparties comme ceci :

* Le set d’entrainement contient deux ans de données, où sont relevés toutes les heures différentes informations, dont le nombre total de location. Ce set ne contient que les données des 19 premiers jours de chaque fois.
* Le set de test contient des données exactement similaires au set d’entraînement, mais pour tous les jours après le 20 du mois inclus.

L’objectif ici est de prédire le nombre de total de vélos loués à chaque heure pour les périodes couvertes par le set de test, en utilisant seulement les informations du set d’entraînement.

# Première approche

Après une première visualisation des données proposées par le set d’entraînement, on constate qu’il contient 10 886 points de données, répartis sur 12 attributs.

Étant donné le grand nombre d’attribut, la première étape pour comprendre les relations qui les relient seraient de calculer la corrélation entre chacun d’entre eux. On pourra donc « éliminer », ou ou moins attribuer une importance moindre au attribut qui ne corrèlent pas avec le *bike count*.