

# Test Technique

## DESCRIPTION :

Cet exercice vise à évaluer vos compétences en data science et se compose de 3 étapes : Statistiques Descriptives, Machine Learning, Restitution des résultats. La démarche importe plus que les résultats en eux-mêmes.

Le jeu de données fourni est un ensemble de relevés concernant un système de vélo partagé (type Vélib). On cherche à prédire le nombre de vélos loués par heure dans la ville (variable count).

- datetime - date et heure du relevé
- season - 1 = printemps, 2 = été, 3 = automne, 4 = hiver
- holiday – indique si le jour est un jour de vacances scolaires
- workingday - indique si le jour est travaillé (ni week-end ni vacances)
- weather - 1: Dégagé à nuageux, 2 : Brouillard, 3 : Légère pluie ou neige, 4 : Fortes averses ou neiges
- temp – température en degrés Celsius
- atemp – température ressentie en degrés Celsius
- humidity – taux d’humidité
- windspeed – vitesse du vent
- casual - nombre de locations d’usagers non abonnés
- registered – nombre de locations d’usagers abonnés
- count – nombre total de locations de vélos

Les livrables attendus sont les suivants :

- Les réponses aux questions ci-dessous
- Le code produit durant l’exercice
- Les slides demandés en partie 3

Vous avez le choix entre R et Python.

## QUESTIONS :

Vous souhaitez prévoir l’utilisation du nombre de vélo et pour cela, un projet data science vous semble adapté.

### Partie I – Questions (~20 min)

- Quelles sont les grandes étapes d’un projet comme celui-là ?
- Quelles questions vous paraissent importantes à poser pour répondre au besoin du client ?

### Partie II – Projet Data Science (~55 min)

Réalisez un projet Data science en utilisant le modèle de votre choix. Quelles pistes d’améliorations imaginez-vous ?

### Partie III - Restitution (~ 25 min)

La partie restitution sera livrée sous forme de 4 à 7 slides (questions, graphiques, modèles, performances, etc.).