

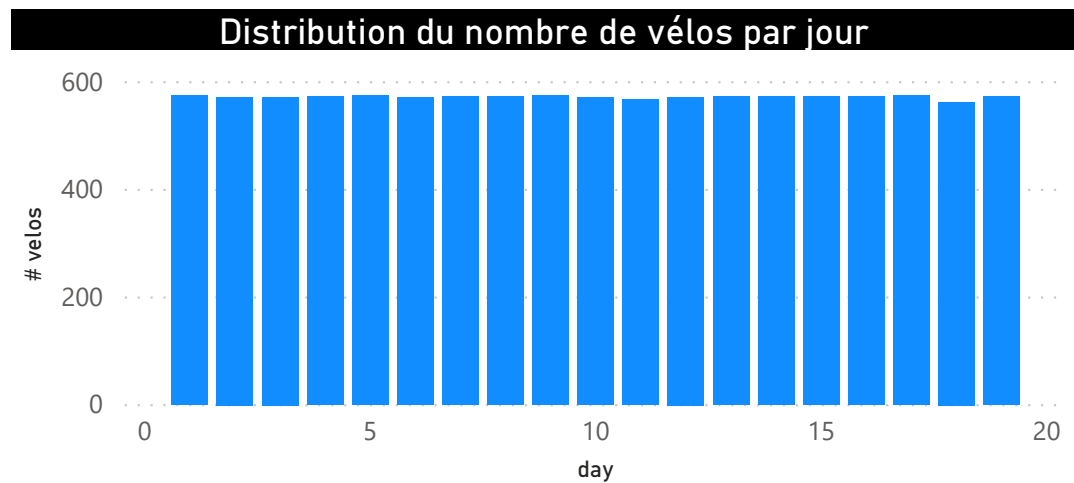
Les objectifs étaient :

- 1) de traiter la problématique de location de vélos à l'heure avec un modèle de Machine Learning (dans le cas présent Random Forest) pour effectuer la prévision.
- 2) On retrouve aussi de l'analyse unidimensionnelle des variables, des analyses multi-variées avec la cible de notre modèle. Pour en savoir plus sur le modèle retenu, vous pouvez regarder le notebook Jupyter avec les différents modèles testés et l'optimisation des hyper-paramètres.
- 3) de voir l'intégration de Power BI avec Python à différents niveaux (pour faire des graphiques avec les librairies Python matplotlib et seaborn) mais aussi pour utiliser l'algorithme de Machine Learning avec la librairie scikit learn

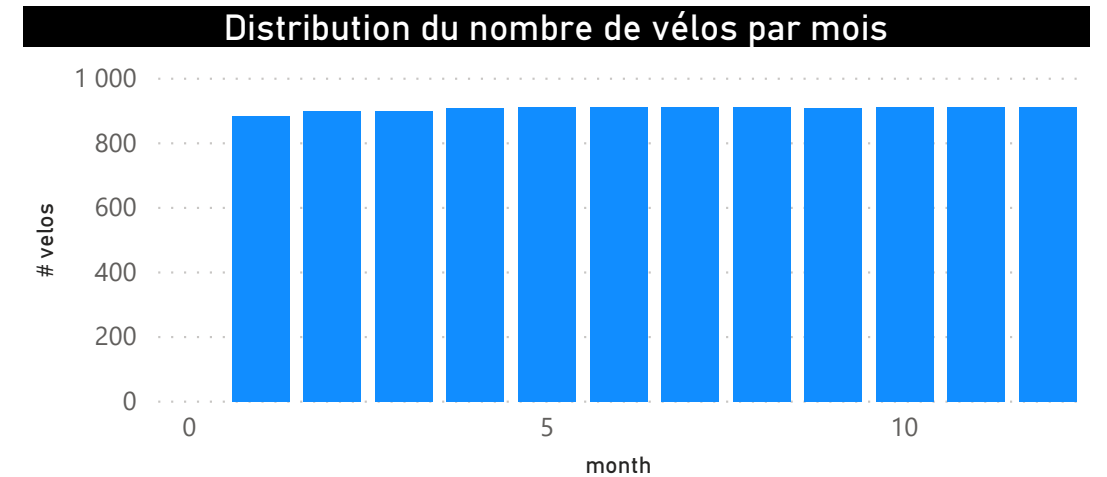
L'objectif est d'analyser les plages de temps de manière unidimensionnelle.

year	Nombre de count
2012	5464
2011	5422
Total	10886

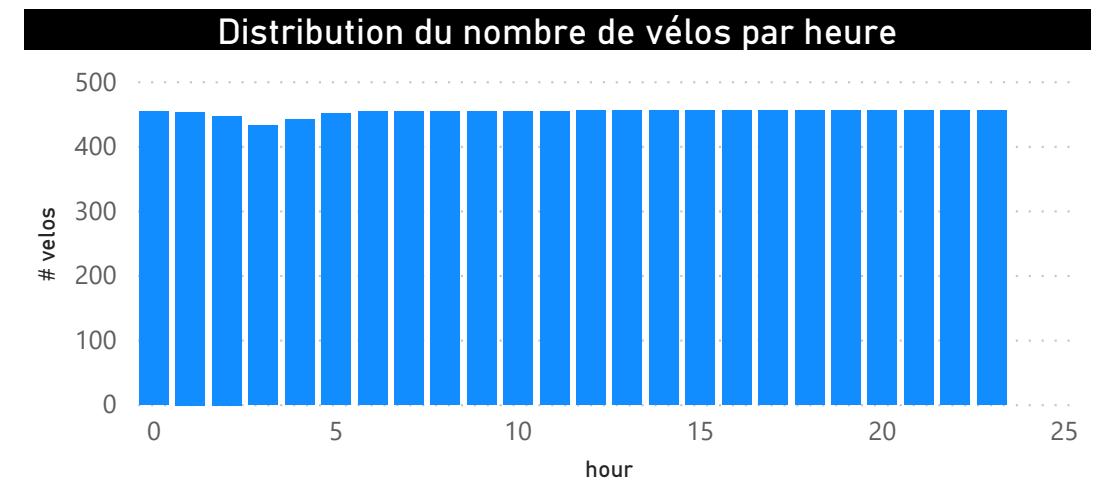
2 années représentées avec le même nombre de valeurs.



Uniquement 19 jours représentés avec le même nombre de valeurs

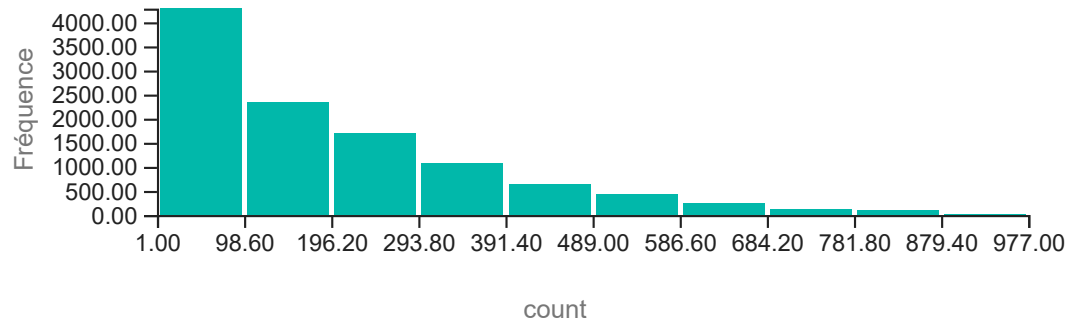


12 mois représenté avec le même nombre de valeurs.

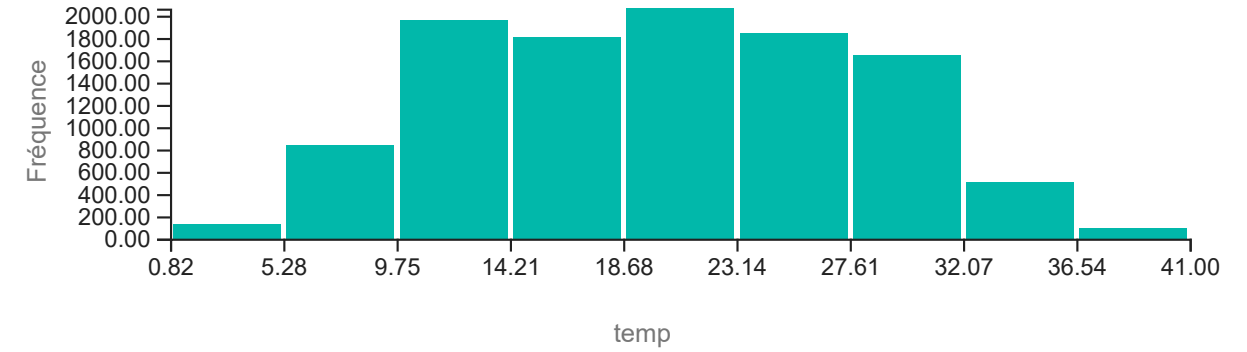


24 heures représenté avec le même nombre de valeurs.

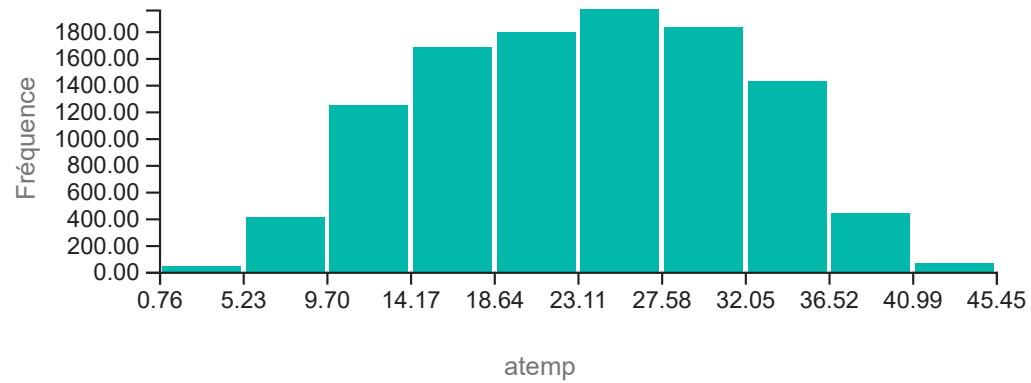
Distribution du nombre de vélos loués



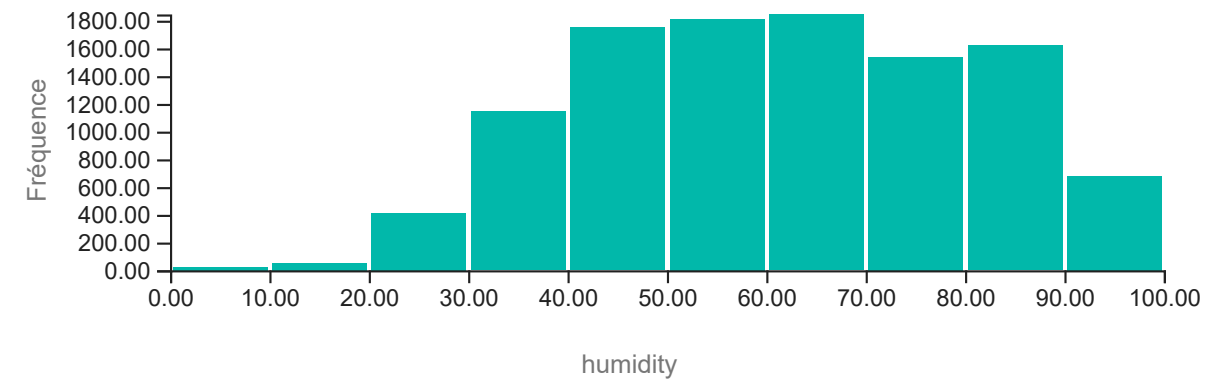
Distribution de la température



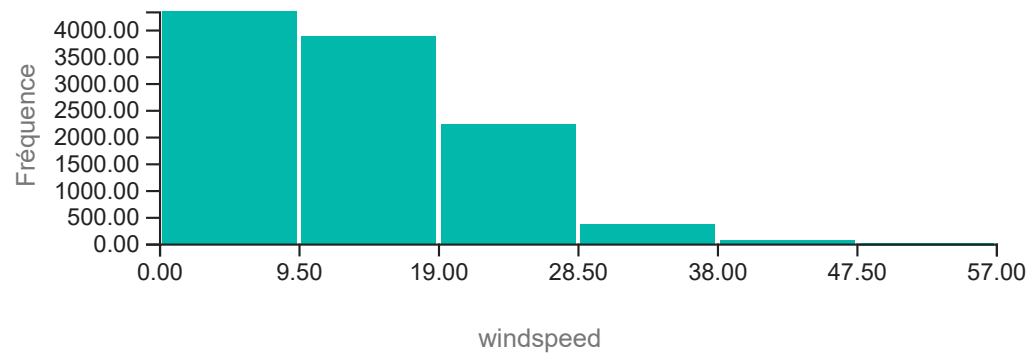
Distribution de la température ressentie



Distribution de l'humidité

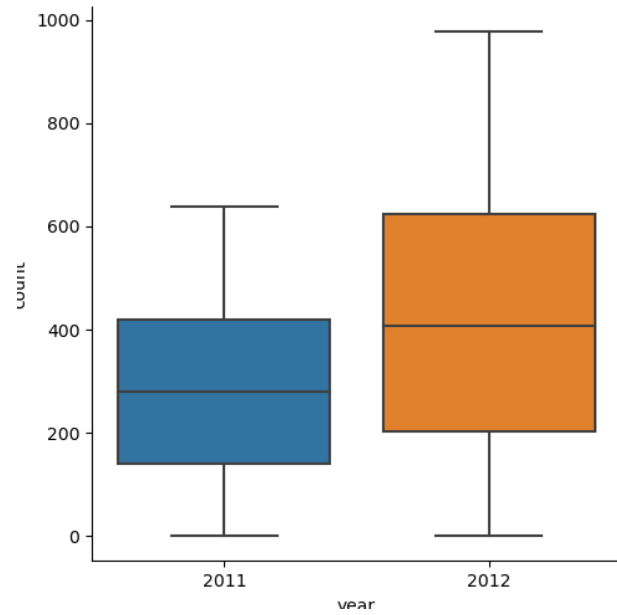


Distribution du vent

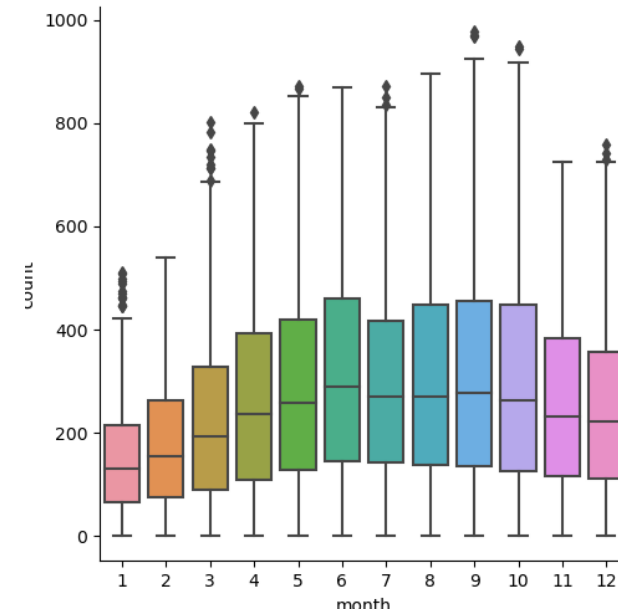


- Distribution de type exponentielle pour la variable cible : nombre de vélos loués
- Distribution de type gaussienne pour la température, température ressentie, humidité
- Distribution de type exponentielle pour la vitesse du vent

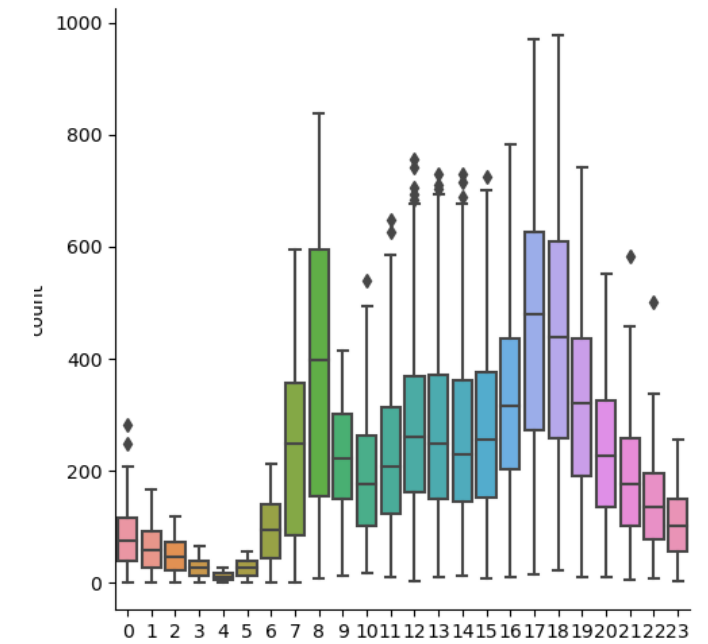
Vélos loués et année



Vélos loués et mois

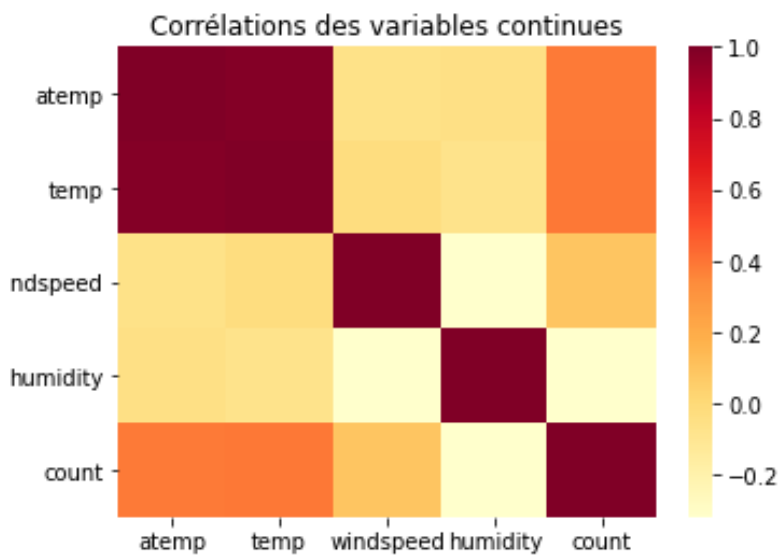


Vélos loués et heures



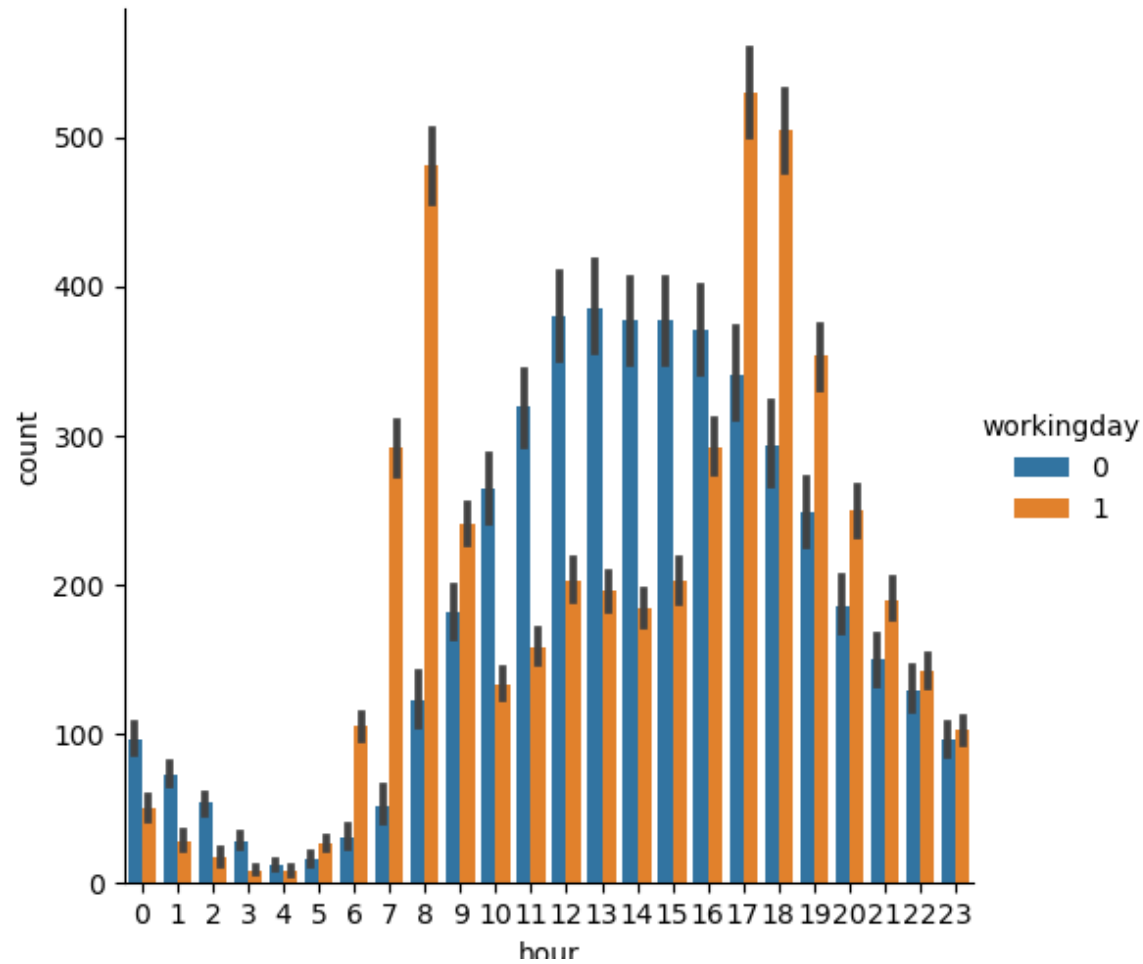
On observe un nombre de vélos loués par année plus forte en 2012 par rapport à 2011. On observe une plus grande utilisation pendant les mois de Mai à Octobre et des différences d'utilisation intra-journalière avec deux pics vers 8h et 18/19h

Heatmap des corrélations



Le nombre de vélos loués dans une heure semble évoluer positivement suivant la température réelle ou ressentie.
Par contre, elle évolue négativement en fonction de l'humidité.
Un élément moins intuitif est qu'elle semble aussi évoluer positivement en fonction de la force du vent.

Location horaire des vélos différenciées par jour travaillé

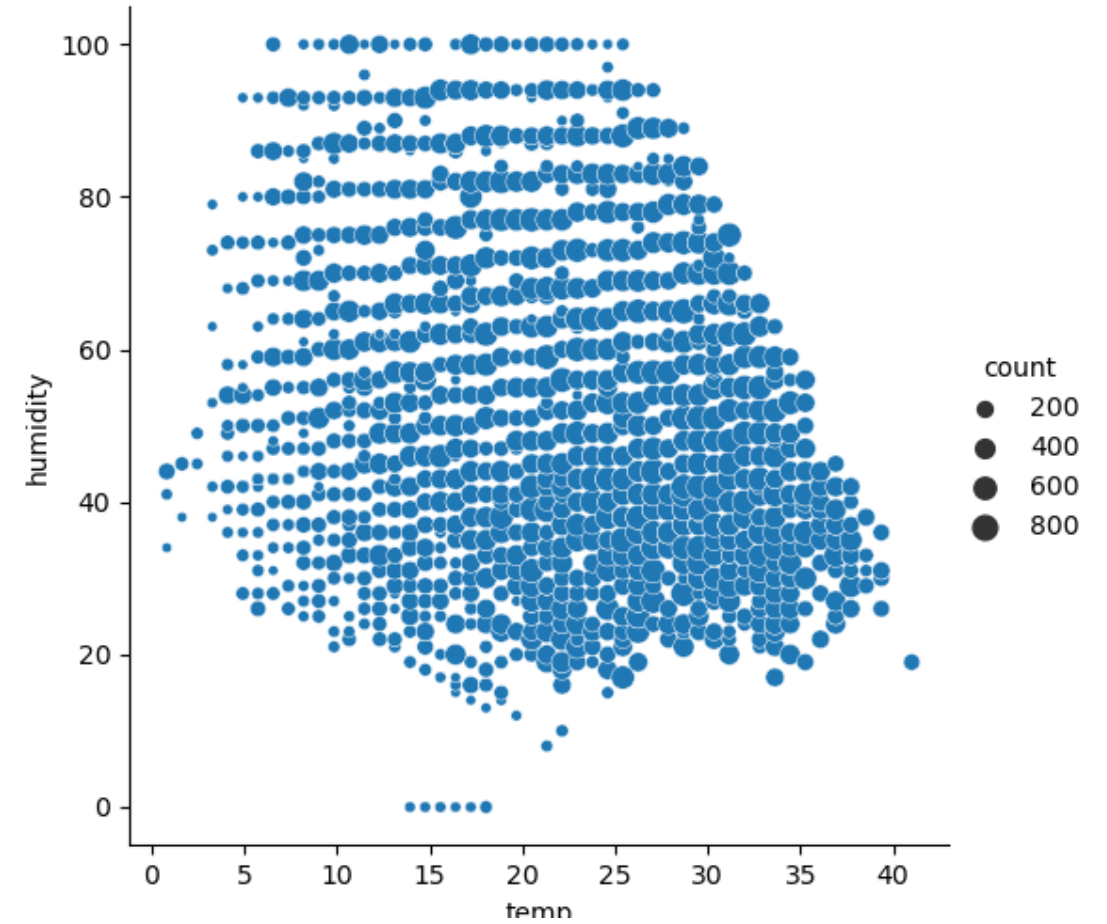


On peut voir l'impact de la variable workingday où on a des comportements bien différents avec des pointes très marquées pour les jours travaillés (7/8h et 17 à 19h).

Par contre pour les jours non travaillés, on a un usage plus important entre 10h et 20h.

On peut aussi voir les intervalles de confiance qui semblent plutôt bons

Location des vélos fonction de la température et de l'humidité



On peut observer que l'usage est dans la majorité plus grande lorsque la température est clémente (supérieure à 15°C) et que le taux d'humidité est inférieure à 60%

10886

Nombre d'enregistrements

8708
Train

43,07
Max Error Train

1,01
MAE Train

5,28
MSE Train

2178
Test

75,39
Max Error Test

2,51
MAE Test

30,26
MSE Test

Pour calibrer le modèle, nous avons séparé l'ensemble de 10886 enregistrements avec 80% constituant la base d'apprentissage (soit 8708 enregistrements) et 20% pour tester le modèle.

Le modèle Random Forest indépendamment de la métrique utilisée (Maximum de l'erreur en absolue, Moyenne des erreurs absolus, ou moyenne des résidus au carré obtient de bons résultats.

Il n'y a pas non plus de sur-apprentissage

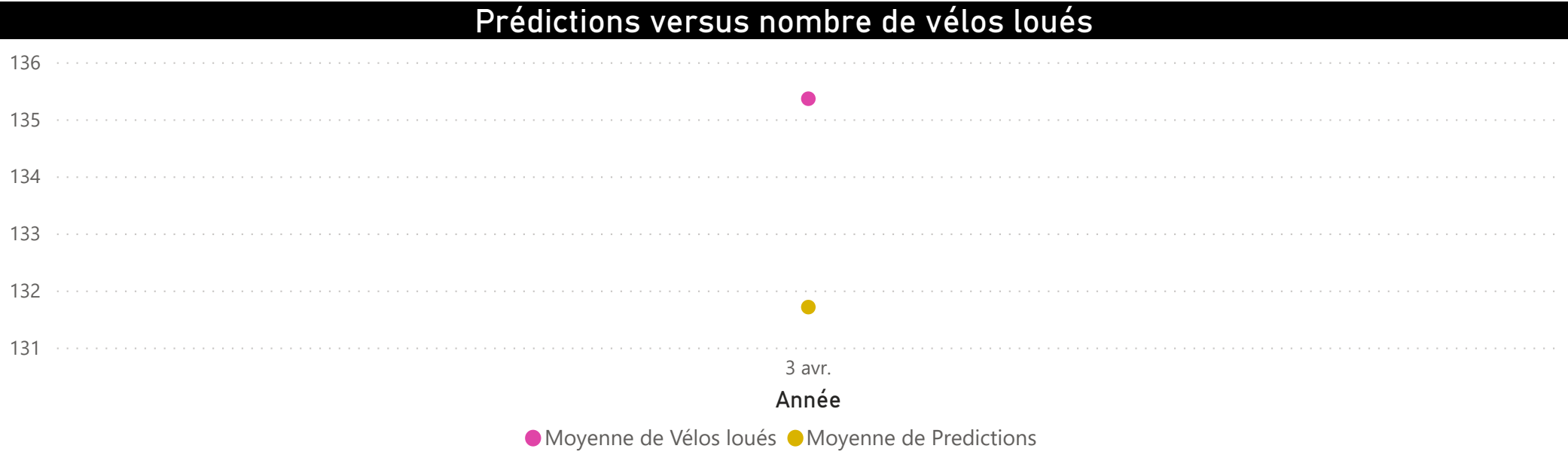
Sélectionner l'ensemble
Train(1) ou Test(0)

Tout

L'erreur entre les prévisions et les prédictions ne se voit pas à l'échelle journalière avec les deux courbes qui se collent. Il faut sélectionner un jour (année / mois / jour) via les filtres. Le tableau ci dessous permet de voir les plages d'heure où les prévisions sont les moins bonnes

Pires prédictions

datetime	Delta
2011-04-03 14:00:00	75,39
2011-04-03 13:00:00	16,68
2011-04-03 15:00:00	9,72
2011-04-03 18:00:00	9,57
2011-04-03 19:00:00	6,71
2011-04-03 12:00:00	2,68
2011-04-03 11:00:00	2,49
2011-04-03 07:00:00	1,52
2011-04-03 04:00:00	0,85
2011-04-03 17:00:00	0,75
2011-04-03 23:00:00	0,59
2011-04-03 21:00:00	0,53



Année

2011

Mois

4

Jour

3

