



**L3 Informatique, SD**

**Projet annuel**

**Tableau de bord de l'open-data des vélos en location libre-service de la  
Métropole RouenNormandie**



## Réalisé par:

- Aguida Yahia
- Mejdoub Yanis
- Bafdel Moufdi Zakaria

## Supervisé par:

- Mr. Pierre Heroux



# Introduction

**Dans le cadre de notre projet annuel en science des données, nous avons conçu un tableau de bord interactif pour analyser l'utilisation des vélos en libre-service à Rouen. Grace aux données publiques disponibles entre février et juillet 2024, notre objectif était de mieux comprendre les comportements des usagers, de détecter les tensions sur certaines stations, et de proposer des pistes concrètes d'amélioration du service.**



# Technologies utilisées

L'outil a été conçu avec Python, en utilisant Dash et Plotly pour les visualisations, Pandas pour le traitement des données, et Scikit-learn pour l'analyse par clustering. Une carte interactive a également été créée à l'aide de Folium



# Problème : données locales avec des "trous" temporels

**OBJECTIF : RECONSTRUIRE CES DONNÉES MANQUANTES POUR LES COMPARER AU DATASET COMPLET**

## ● 2. VISUALISATION DES DONNÉES

**Affichage des gaps (valeurs manquantes) sur les courbes de temps**

**Taux de données manquantes : >99% sur certains cas !**

**Résolution : resampling toutes les 15 minutes**



# Objectif de la présentation

**Comparer deux jeux de données de stations de vélos :**

**Un jeu complet et fiable, enregistré automatiquement par un serveur.**

**Un jeu local, sauvegardé de manière irrégulière (avec des coupures dues à l'extinction du PC).**



### ● 3. MÉTHODES DE RECONSTRUCTION

Interpolation avant/après

Moyenne glissante (rolling)

Interpolation linéaire ✓

Interpolation polynomiale

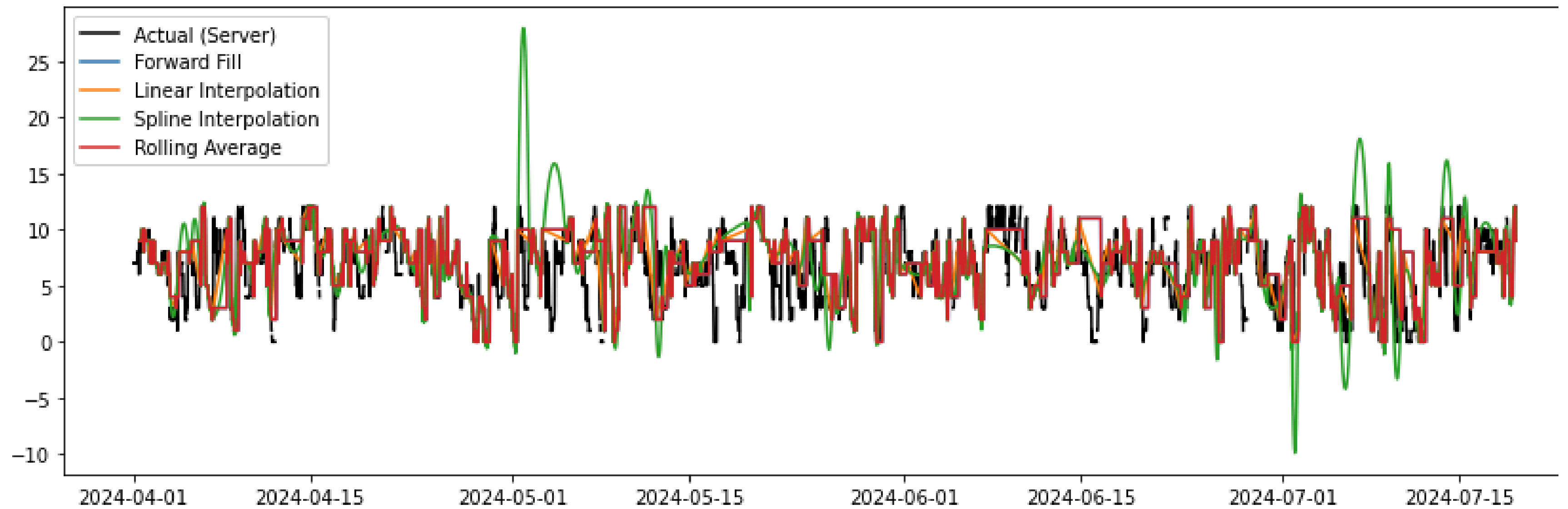
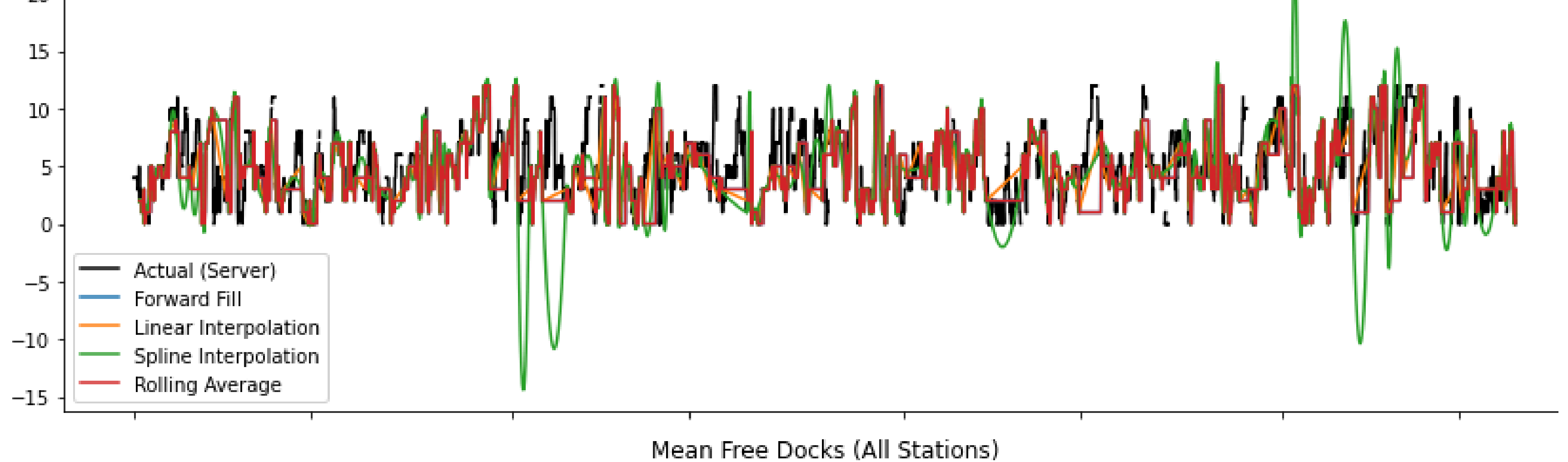


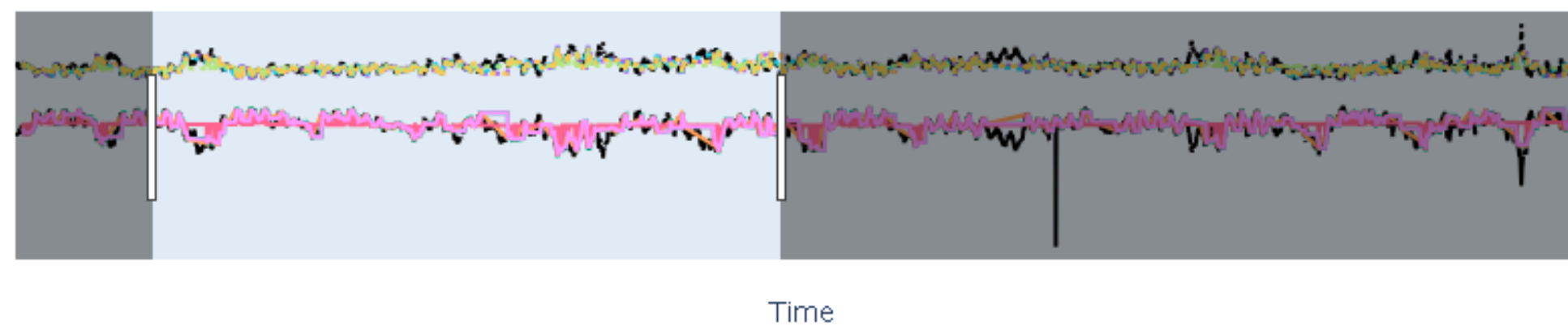
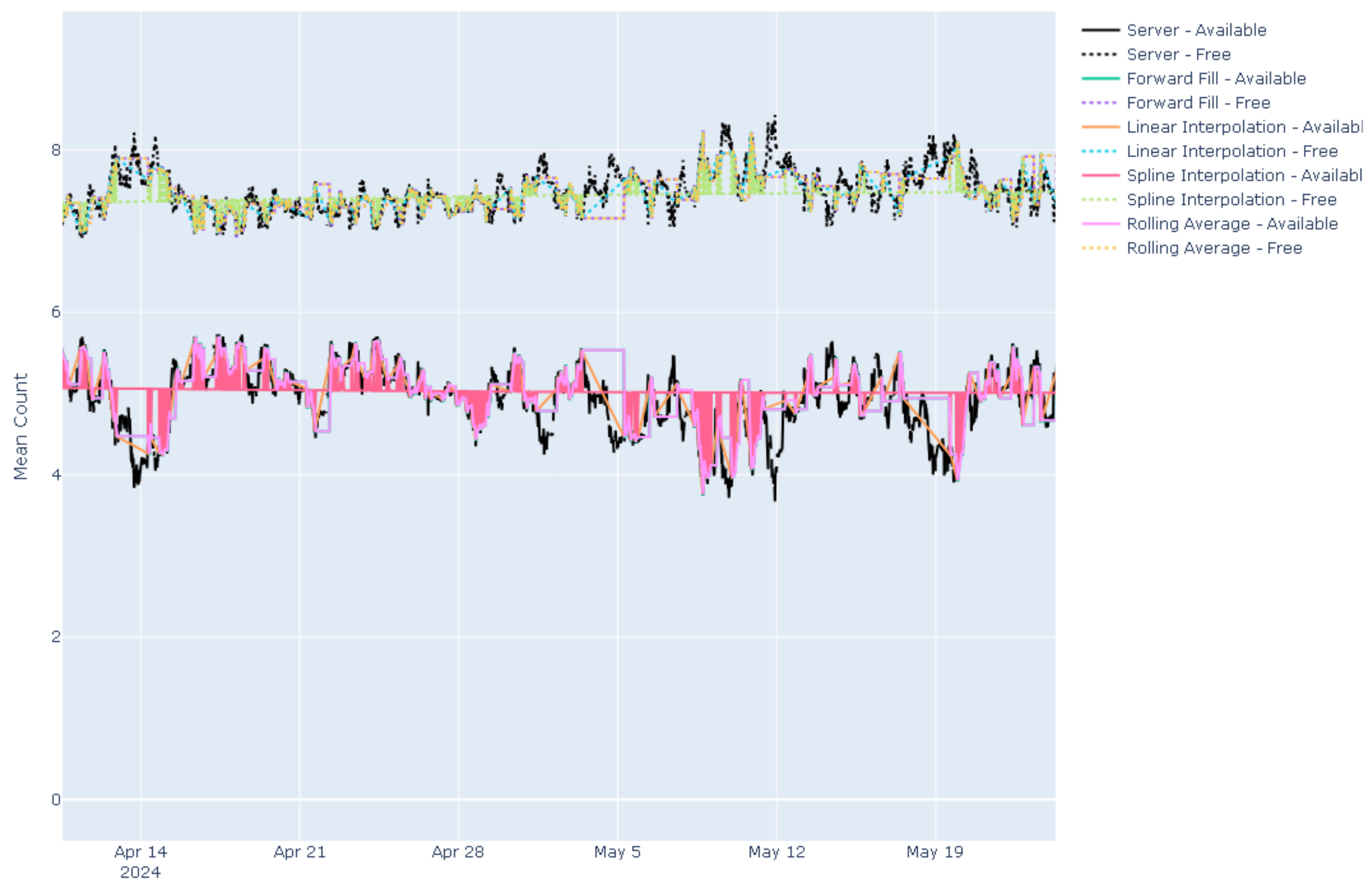


# UTILISATION DE PANDAS ET PLOTLY POUR VISUALISER LES INTERPOLATIONS SUR LES SÉRIES TEMPORELLES





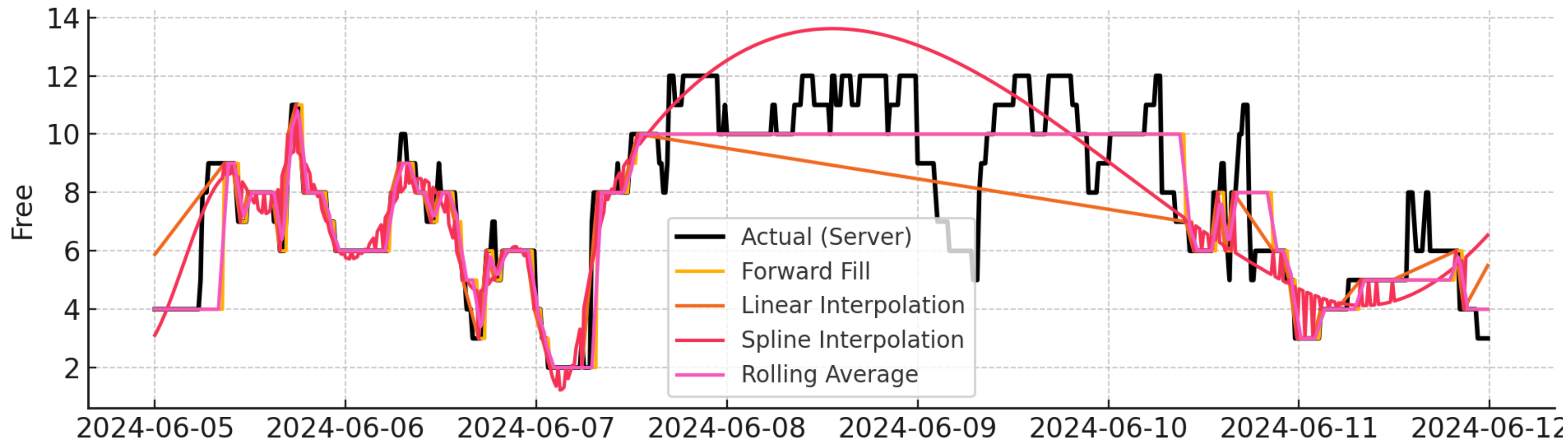




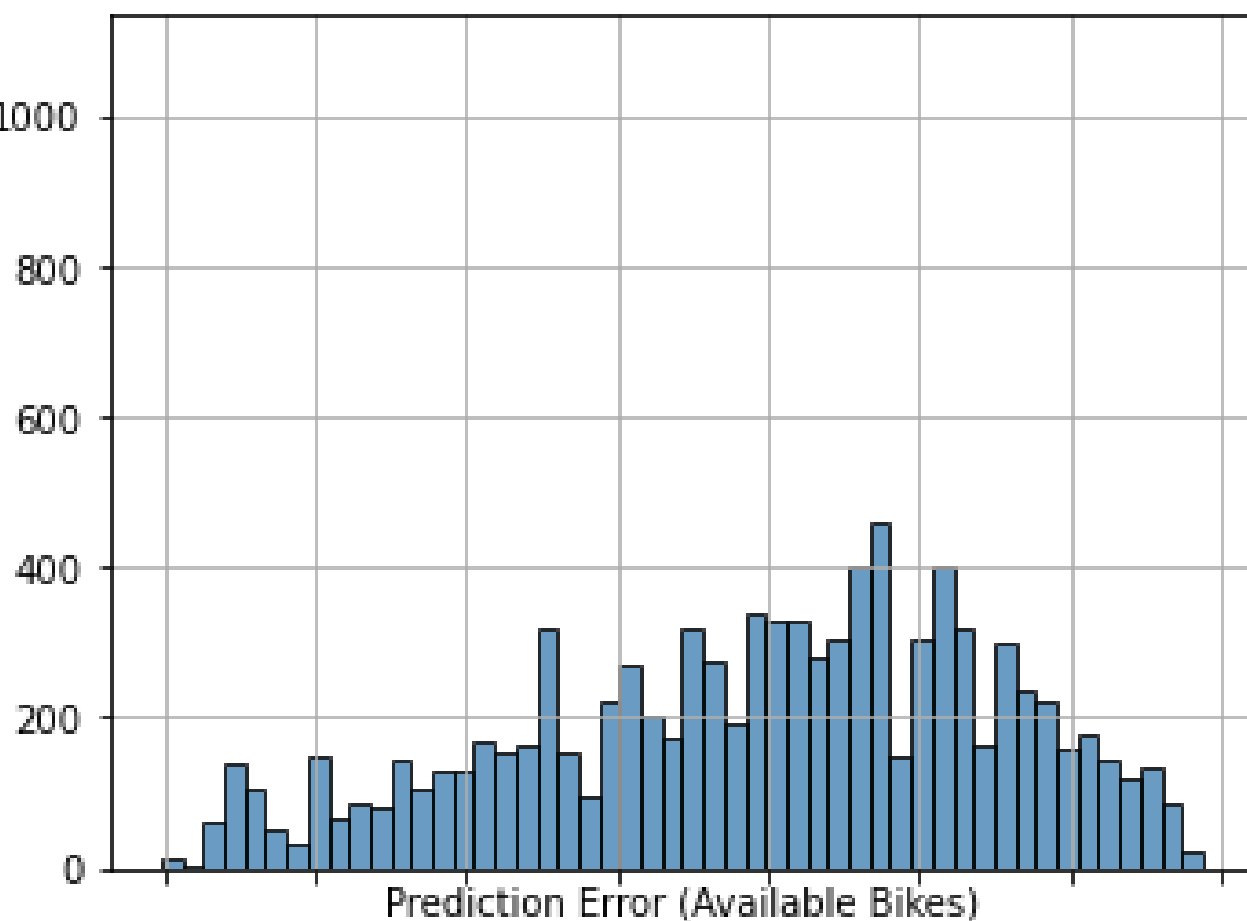
### Station 3680 - Available Bikes



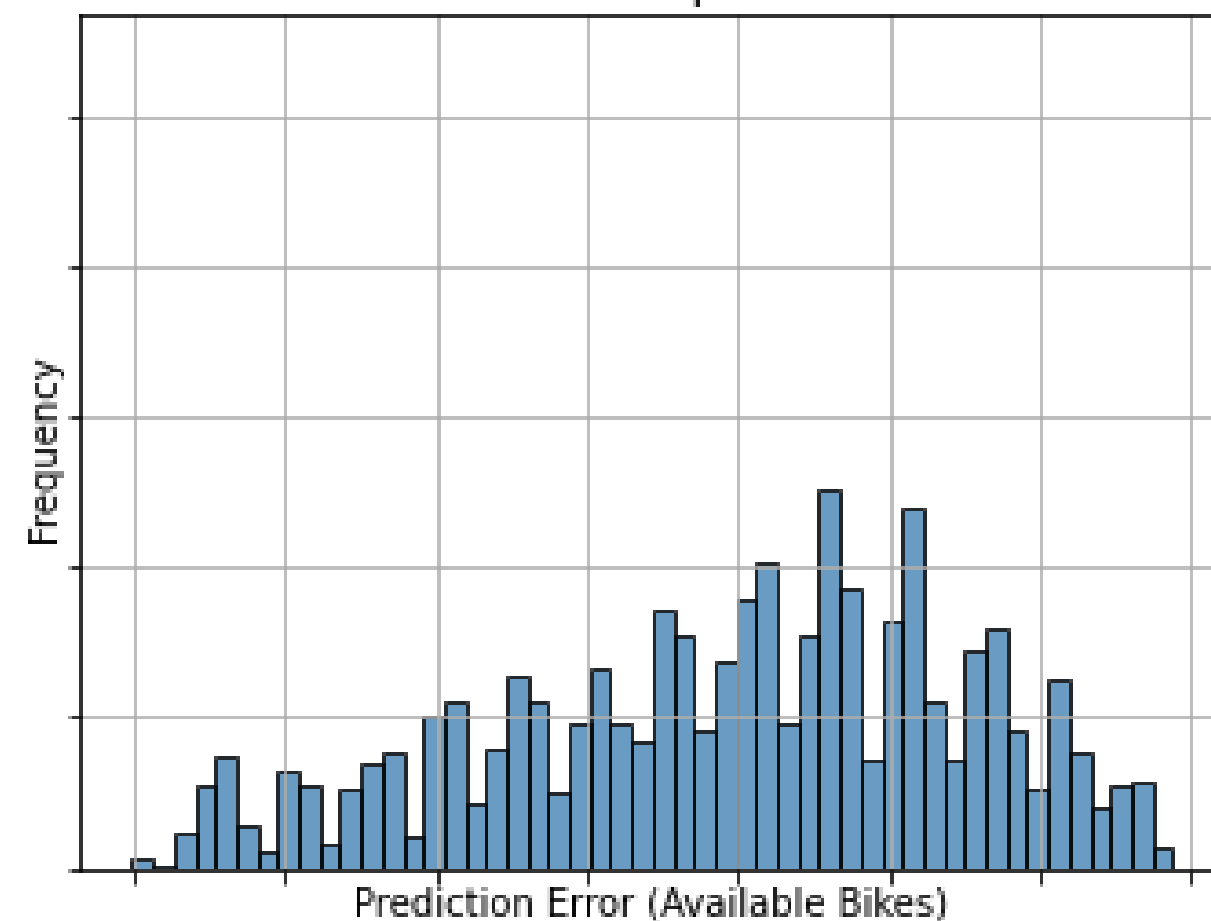
### Station 3680 - Free Docks



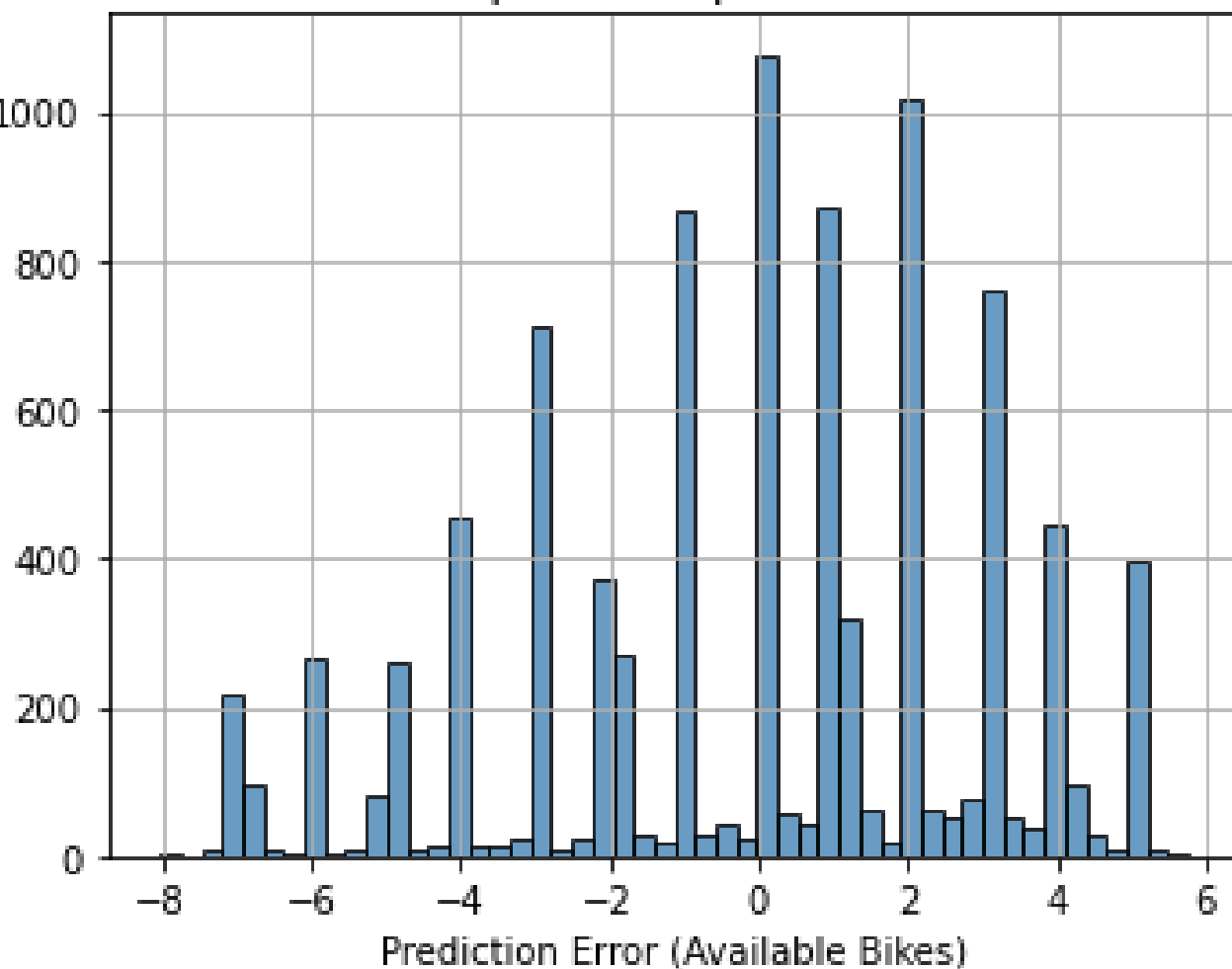
Forward Fill



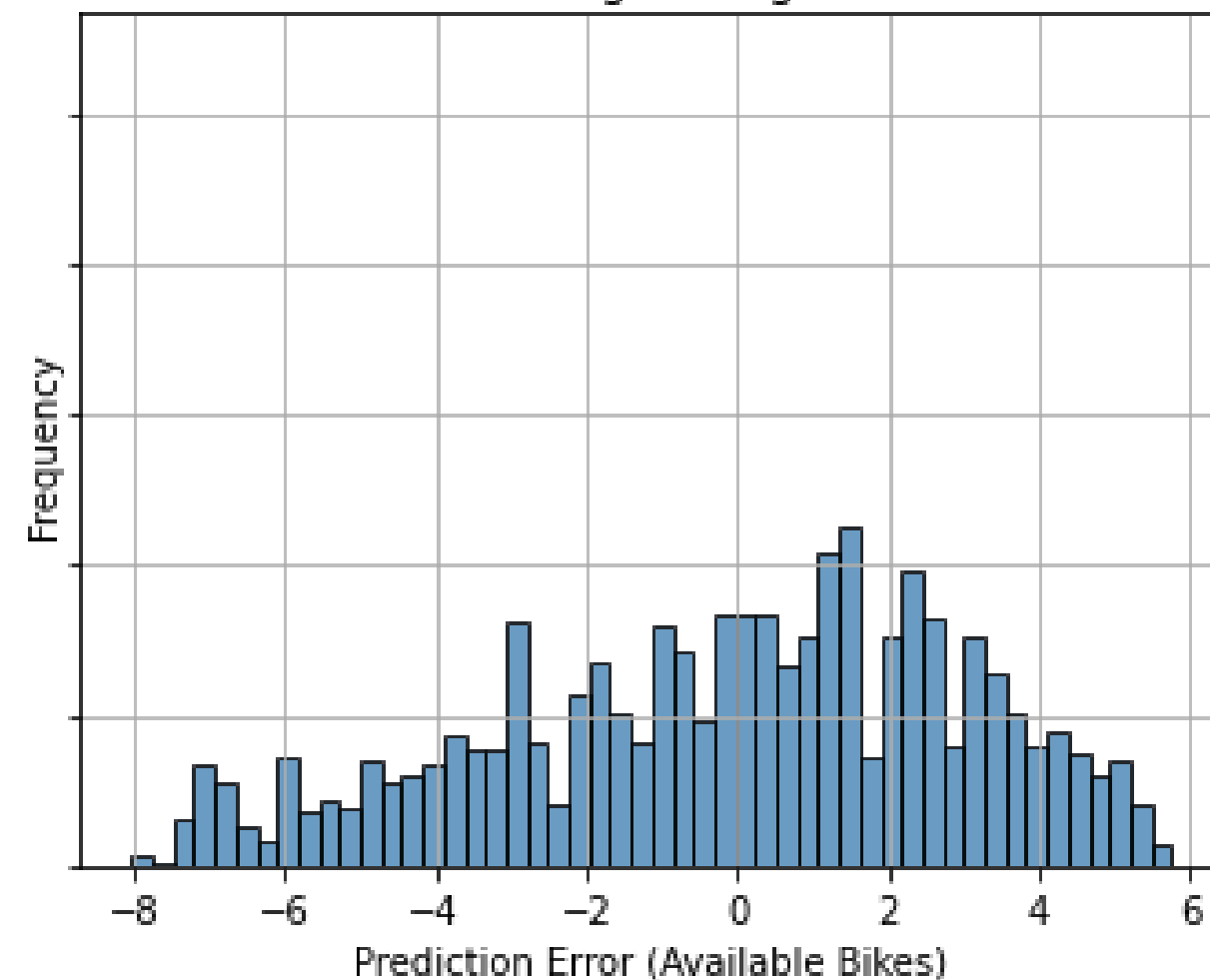
Linear Interpolation



Spline Interpolation



Rolling Average



## MAE ET RMSE PAR MÉTHODE :

**MÉTHODE MAE RMSE**  
**◆ FORWARD FILL 0.192 0.291**

**◆ LINEAR INTERPOLATION**  
**✓ 0.136 ✓ 0.228**

**◆ SPLINE INTERPOLATION**  
**0.294 0.402**

**◆ ROLLING AVERAGE**  
**0.201 0.296**

## 1. Interpolation Avant/Après (Forward Fill / Backward Fill)

- ◆ On recopie la dernière valeur connue (ou la première) pour combler les trous.

 Idée :

“Si je ne sais pas combien il y avait de vélos à 14h15, je prends le même nombre qu’à 14h00.”

 Avantages :

Très simple

Ne modifie pas brutalement la courbe

 Inconvénients :

Suppose qu’il n’y a aucun changement pendant la coupure → pas réaliste sur de longues périodes



## 2. Moyenne Glissante (Rolling Average)

- ◆ On fait la moyenne des dernières valeurs disponibles dans une fenêtre temporelle (ex : dernière heure)

 Idée :

“Pour 14h15, je prends la moyenne des valeurs de 13h30, 13h45, 14h00 et 14h15.”

 Avantages :

Lisse les courbes

Réduit l'impact des pics anormaux

 Inconvénients :

Ralentit les variations → moins réactif aux vrais changements rapides



### 3. Interpolation Linéaire

- ◆ On trace une ligne droite entre deux points connus pour remplir les valeurs manquantes.

 Idée :

“Il y avait 2 vélos à 14h00 et 6 à 14h30 → on suppose une progression régulière : 3 à 14h10, 4 à 14h20, etc.”

 Avantages :

Très précis sur les courtes périodes

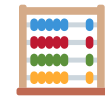
Rapide à calculer

 Inconvénients :

Suppose une évolution constante, ce qui peut être faux si la réalité est plus chaotique







#### 4. Interpolation polynomiale (ordre 2 ou 3)

- ◆ Méthode plus avancée : on trace une courbe fluide (polynôme) qui passe par les points connus.

#### Interpolation Polynomiale



#### Avantages :

Produit des courbes plus naturelles

Meilleure continuité dans les dérivées (la pente évolue doucement)



#### Inconvénients :

Peut être instable si les points sont très espacés

Plus coûteux en calcul





## ● 4. ÉVALUATION DE L'EXACTITUDE

**Comparaison avec les vraies valeurs serveur sur la période Avril–Juillet**

**Métriques utilisées :**

**MAE (Mean Absolute Error)**

**RMSE (Root Mean Square Error)**

**Résultats : Linear Interpolation = meilleur score**



## ● 5. VISUALISATION INTERACTIVE

**Utilisation de Plotly avec un menu déroulant pour sélectionner une station**

**Zoom, scroll, légende dynamique**

**Courbes : vélos disponibles / bornes libres**



## ● 6. CONCLUSION

**Linear Interpolation est la méthode la plus fiable dans ce contexte**  
**Visualisation interactive utile pour l'analyse station par station**  
**Ce travail peut être appliqué à d'autres types de capteurs ou de séries temporelles**



# CONCLUSION

**Notre tableau de bord permet aujourd'hui de mieux comprendre la dynamique de mobilité urbaine à Rouen à travers le service de vélos libre-service. Il a permis d'identifier les stations critiques, les horaires de saturation et les profils de fréquentation.**

**En plus de fournir une base solide pour la prise de décision, ce travail ouvre la voie à des améliorations futures, comme l'intégration de données météo, la prédiction de disponibilités via des modèles de machine learning, ou encore la comparaison avec d'autres villes.**

**Nous avons réussi à transformer un volume de données brutes en un outil visuel, pertinent, et facilement exploitable par les décideurs ou les citoyens soucieux d'une mobilité plus intelligente**

