

KEBE MOUSTAPHA
22117880
M1-MLSD
TIME SERIES
BIKE SHARING



Introduction

Le jeu de données "Bike Sharing Dataset" offre un aperçu détaillé du partage de vélos dans le système Capital à Washington, DC, entre 2011 et 2012. Il rassemble des informations météorologiques et saisonnières, telles que la saison, l'année, le mois, l'heure, les jours de la semaine, les jours fériés, ainsi que les conditions météorologiques comme la température, l'humidité et la vitesse du vent. Il contient également des compteurs horaires et quotidiens de locations de vélos, catégorisés en utilisateurs occasionnels et enregistrés.

Présentation des données

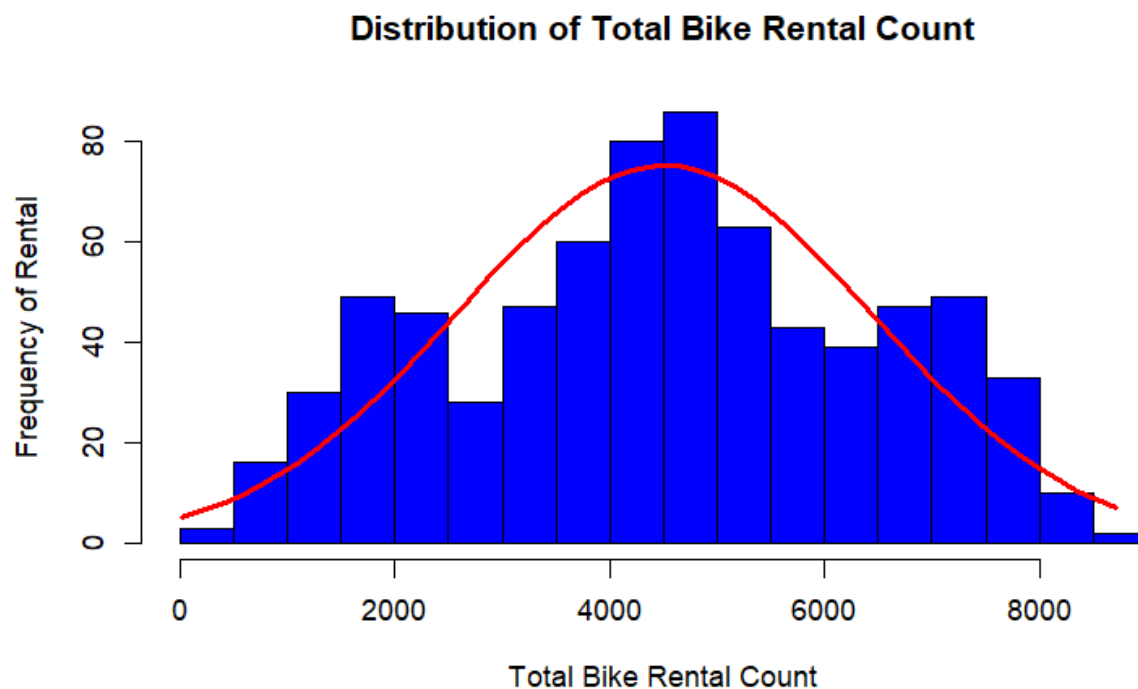
Le jeu de données "Bike Sharing Dataset" capture minutieusement les détails des locations de vélos dans le système de partage Capital à Washington, DC, pour les années 2011 et 2012. Il comporte plusieurs variables décrivant à la fois les caractéristiques météorologiques et les activités de location de vélos.

Variables du jeu de données :

- 1-Instantané (**instant**): Index d'enregistrement unique.
- 2-Jour (**dteday**): Date des observations.
- 3-Saison(**season**) : Catégorisation des saisons (printemps, été, automne, hiver).
- 4-Année(**yr**) : Année des observations (2011 ou 2012).
- 5-Mois(**mnth**) : Numéro du mois (1 à 12).
- 6-Heure (**hr**) : Heure de la journée (0 à 23).
- 7-Jour férié (**holiday**): Indicateur pour les jours fériés.
- 8-Jour de la semaine(**weekday**) : Jour de la semaine (lundi, mardi, etc.).
- 9-Jour ouvrable(**workingday**) : Indicateur si le jour est ouvrable (1) ou non (0).
- 10-Météo (**weathersite**): Description des conditions météorologiques.
- 11-Température (**temp**) : Température normalisée en Celsius.
- 12-Température ressentie (**atemp**) : Température ressentie normalisée en Celsius.
- 13-Humidité (**hum**) : Humidité normalisée.
- 14-Vitesse du vent (**windspeed**) : Vitesse du vent normalisée.
- 15-Utilisateurs occasionnels (**casual**) : Nombre d'utilisateurs occasionnels.
- 16-Utilisateurs enregistrés (**registered**) : Nombre d'utilisateurs enregistrés.
- 17-Total des locations (**cnt**) : Nombre total de locations de vélos (casual + registered).

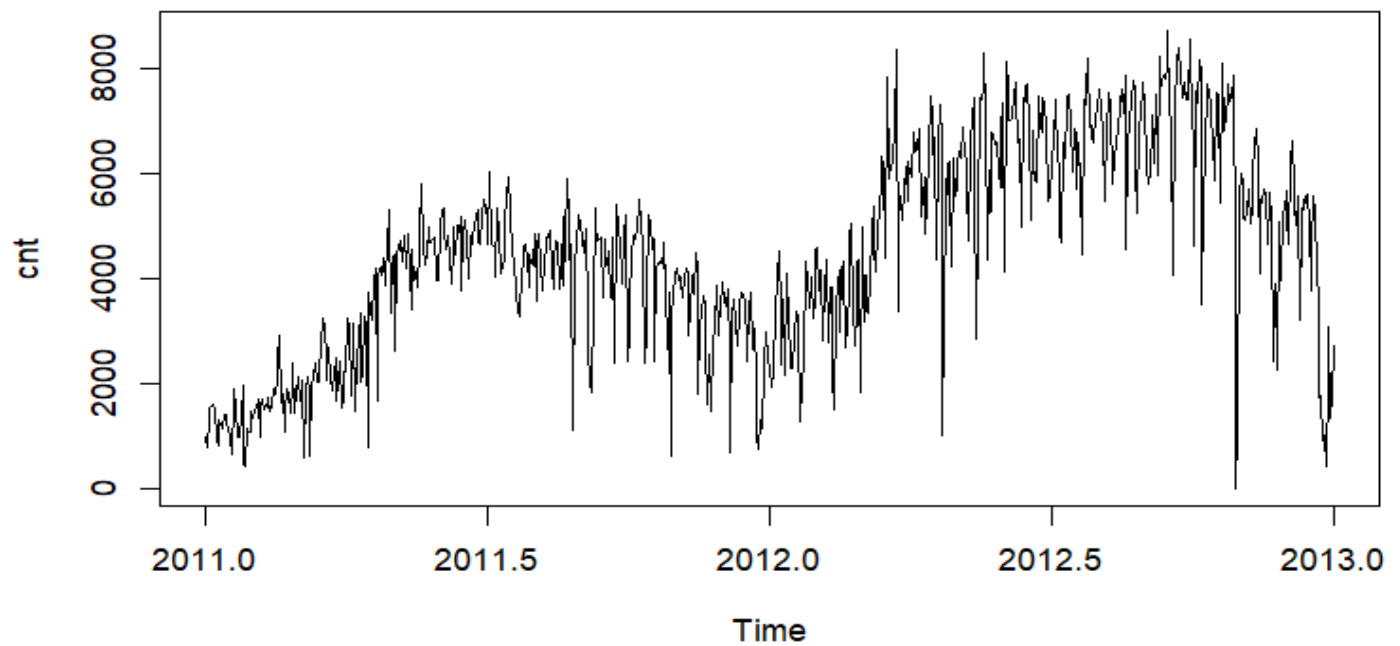
Visualisation initiale

Les graphiques exploratoires sont utilisés pour présenter visuellement les tendances de location de vélos au fil du temps. Des tracés temporels montrant la distribution des décomptes horaires et quotidiens de locations de vélos, ainsi que des histogrammes pour la température, l'humidité, etc., peuvent offrir une première compréhension des données. La série temporelle formée par ces données permet d'observer de fascinantes variations saisonnières. Dans notre problème, on observe une tendance de même qu'une saisonnalité de notre série temporelle.



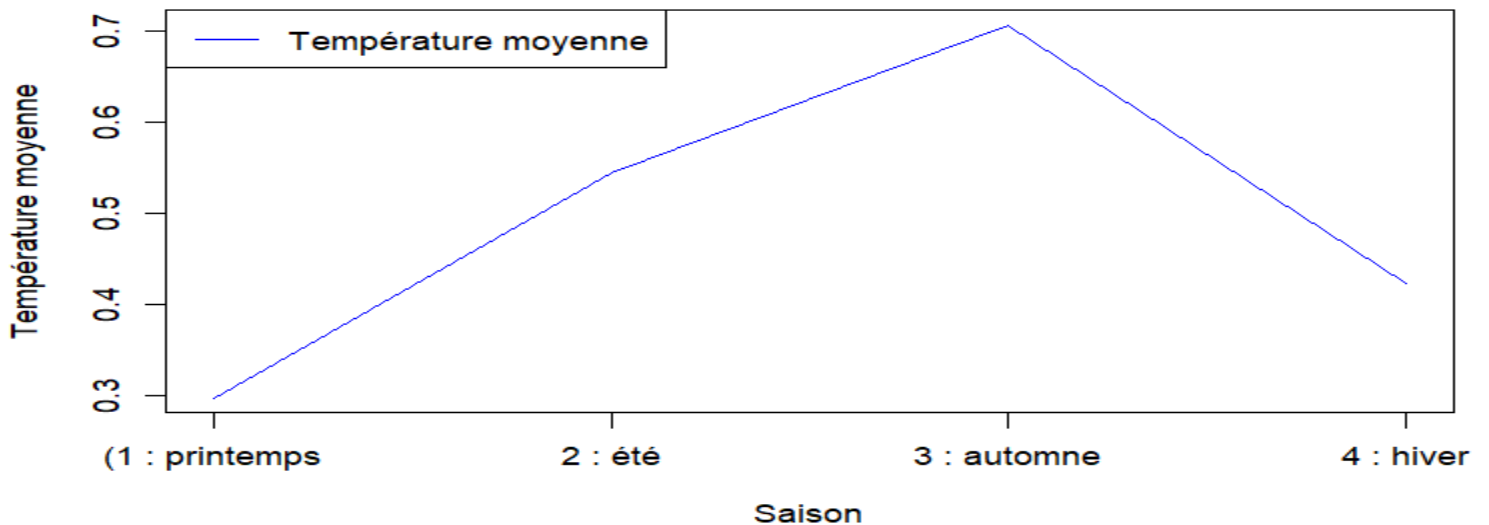
Cette figure ci-dessus montre la distribution de la location de vélos qui peut être assimilée à une loi gaussienne.

Notre série temporelle construit à partir de la variable cnt est le suivant:



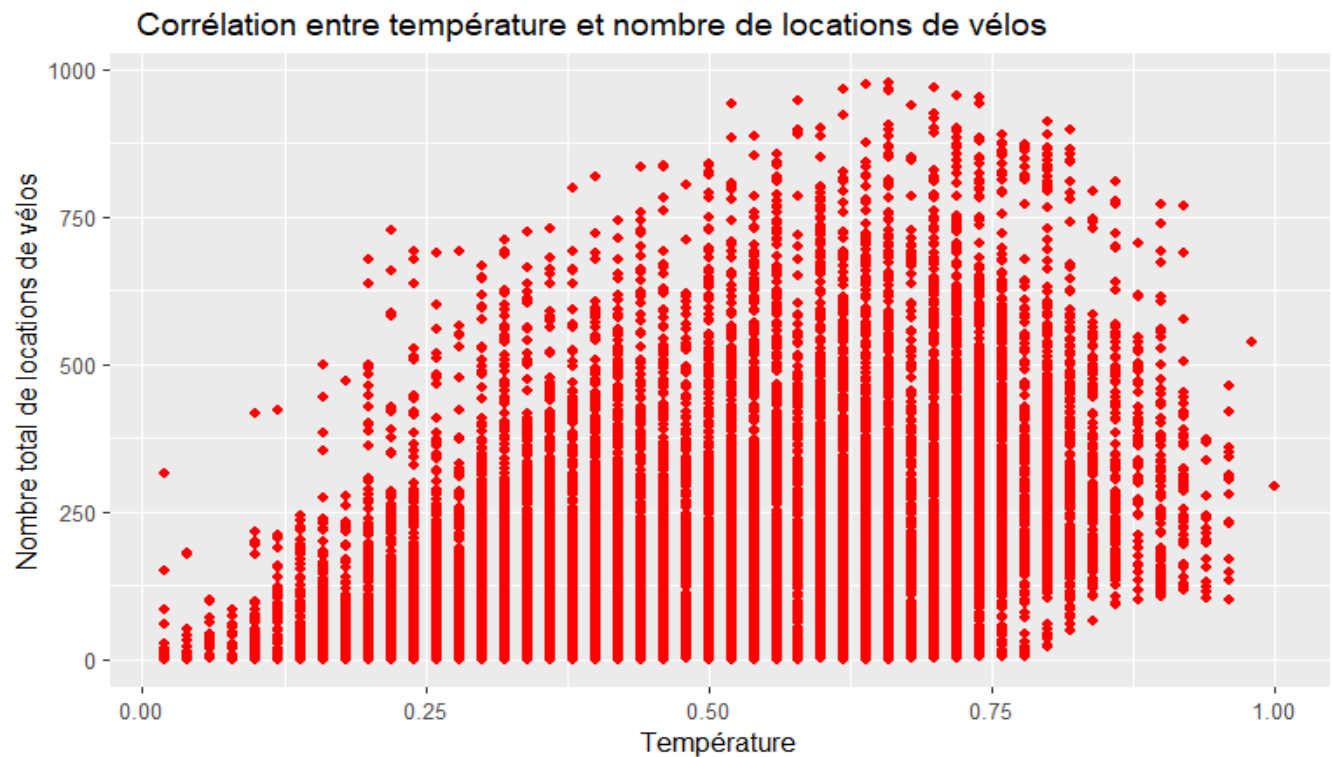
On voit que les pics sur la location de vélos interviennent toujours au milieu de l'année correspondant à l'été et à l'automne.

Ce graphe montre la distribution de la température moyenne normalisée en °C en fonction des différentes saisons (printemps ,été ,automne et l'hiver)



L'analyse des séries temporelles météorologiques montre l'influence significative des conditions climatiques sur la location de vélos. La température, l'humidité et d'autres paramètres ont un impact direct sur les comportements de location. Le nombre de locations de vélos atteint son maximum au milieu de l'année (saison été - automne), il se peut qu'en raison des conditions météorologiques et de la température extérieure, il soit confortable d'utiliser le vélo comme moyen de transport.

Ce nuage de points ci-dessous montre la corrélation entre la température (normalisée en Celsius) et la location de vélos .



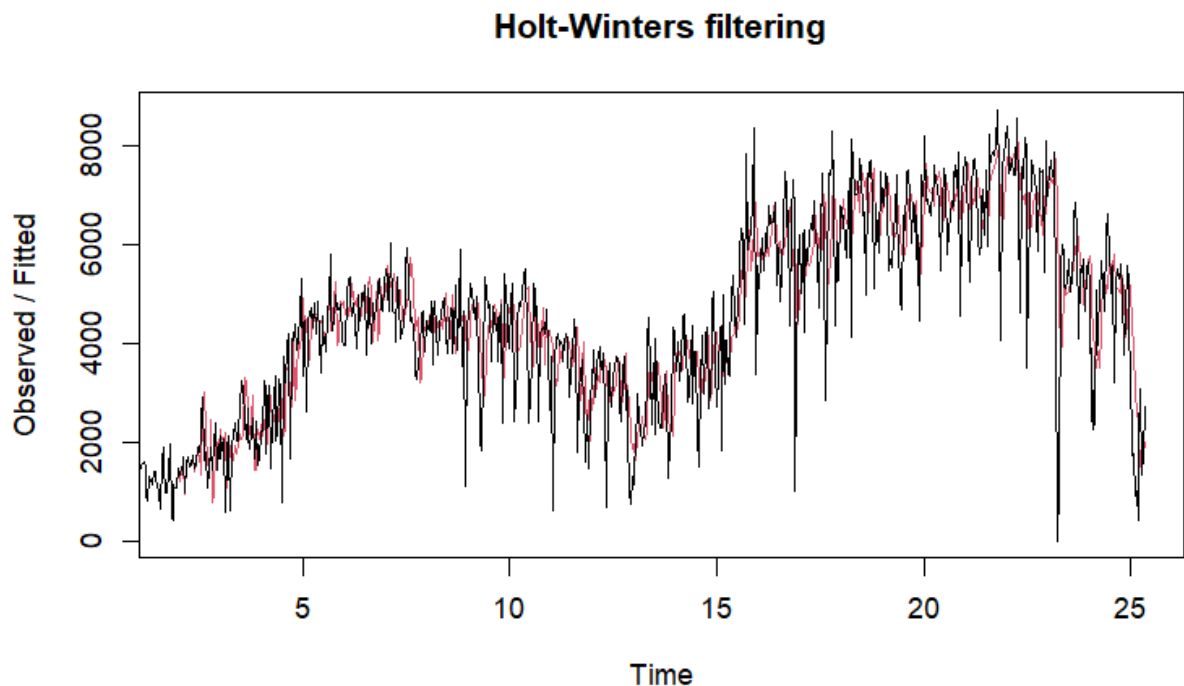
La modélisation des séries temporelles offre des opportunités passionnantes. Les prévisions basées sur ces données méticuleuses peuvent améliorer la gestion des ressources de partage de vélos, permettant aux exploitants de mieux anticiper les pics. Ces analyses peuvent également être étendues pour évaluer les tendances sur des périodes plus longues et aider à planifier des stratégies urbaines durables.

Modélisation et prévisions

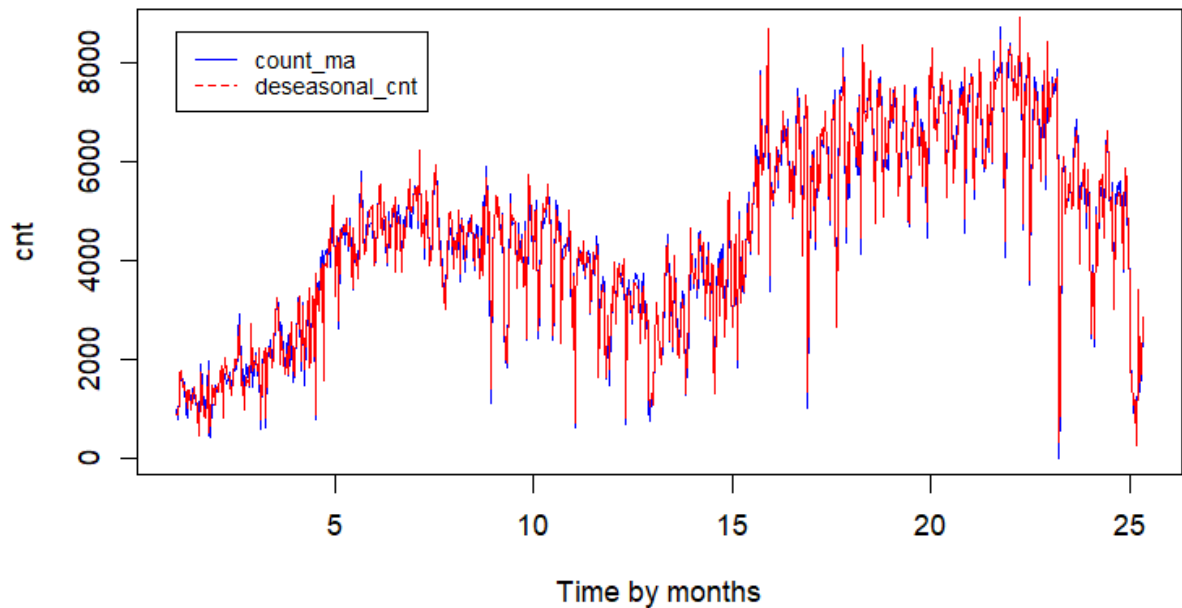
Cette section se concentre sur la modélisation des séries temporelles pour prédire les décomptes de location de vélos et évaluer la performance des modèles.

Choix des modèles :

- Modèles de lissage exponentiel : Expérimentez avec les modèles de lissage exponentiel (Holt-Winters, Simple Exponential Smoothing, etc.) pour évaluer leur efficacité dans la prédiction des tendances de location de vélos. Dans ce projet j'ai utilisé le Holt-Winters pour faire le lissage de ma série temporelle parce que nos données présentent à la fois une tendance et une saisonnalité, le modèle de Holt-Winters peut être efficace pour capturer ces deux composantes (**voire le .Rmd**).



-
- Supprimer la saisonnalité de notre série temporelle



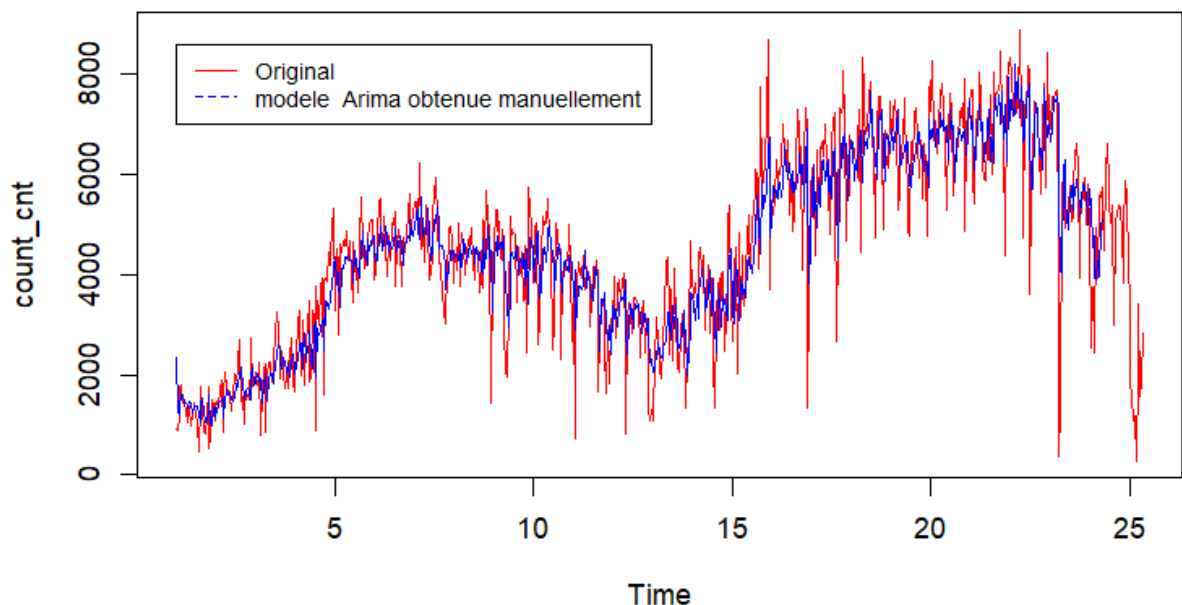
- Pour éliminer les tendances et les effets saisonniers, la différenciation est utilisée pour rendre la série temporelle plus stationnaire, facilitant ainsi l'application de modèles statistiques comme ARIMA. **(voire le .Rmd)**
- Modèles ARIMA : Utilisez les modèles ARIMA (AutoRegressive Integrated Moving Average) pour modéliser les séries temporelles des décomptes horaires ou quotidiens de locations de vélos.
- Explorez différents ordres (p,d,q) grâce à l'ACF ,PACF et l'ordre de différenciation pour identifier le modèle le plus approprié **(voire le .Rmd)**.

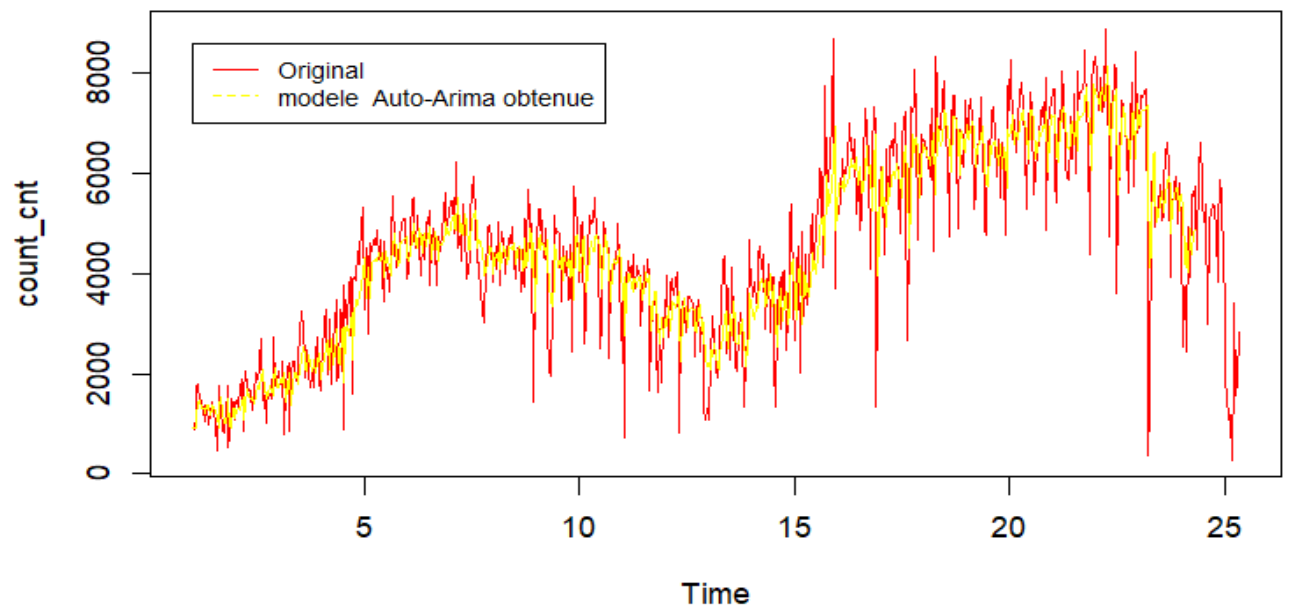
Évaluation des modèles :

- Validation croisée : Divisez les données en ensembles d'entraînement et de test pour évaluer les performances des modèles. Utilisez des métriques telles que ACCURACY ,AIC,..etc (**voire le .Rmd**)
- Comparaison des modèles : Comparez les différents modèles en termes de capacité prédictive et de stabilité pour identifier celui qui offre les prévisions les plus précises et robustes

Prévisions à long terme :

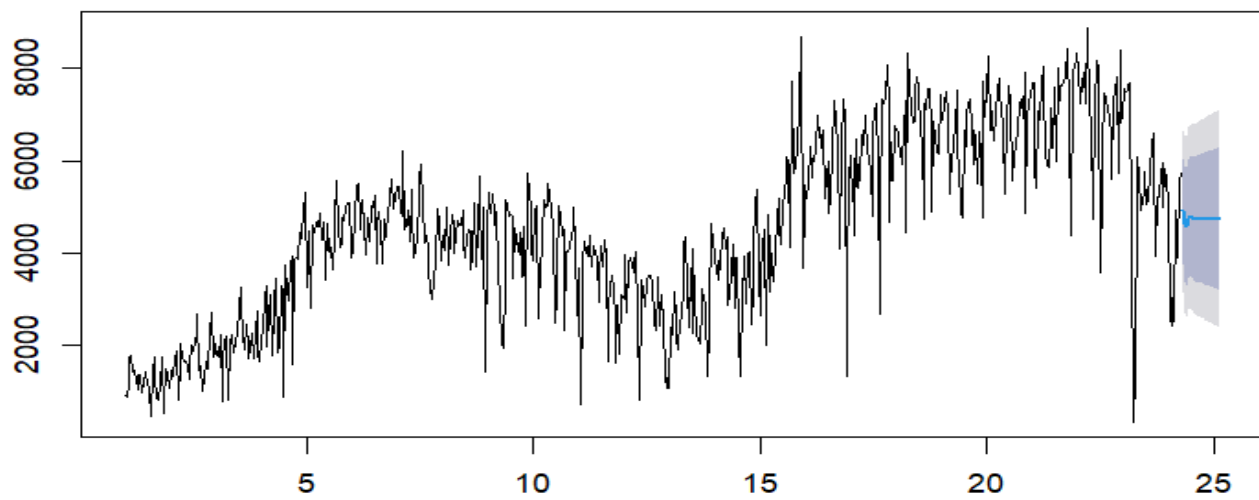
- Prévisions futures : Utilisez le modèle sélectionné manuellement et celui proposer par Auto-ARIMA pour effectuer des prévisions à long terme des décomptes de location de vélos. Visualisez ces prévisions pour comprendre les tendances futures de la demande.



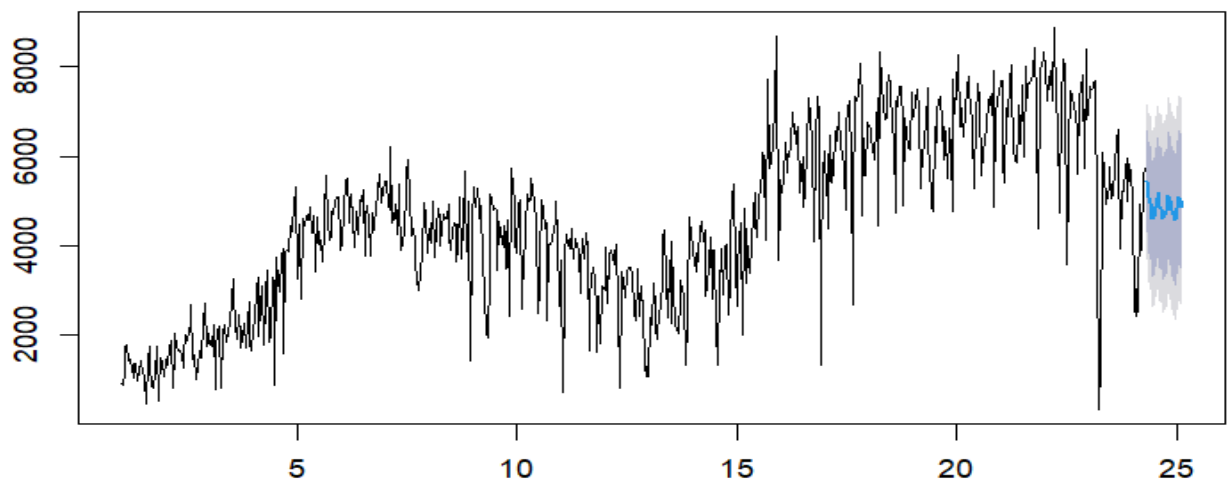


- Comparer les prévisions faites par les deux modèles.

Forecast avec model Auto-Arima obtenue



Forecast avec model Arima obtenue manuellement



- Le modèle Arima obtenu manuellement donne une prévision plus naturelle que le modèle auto-Arima (**voire le .Rmd**).

Cette section démontre l'application pratique de la modélisation des séries temporelles pour prédire la demande de locations de vélos, fournissant des informations précieuses pour la planification et la gestion des ressources dans le système de partage de vélos. Elle offre également un aperçu des performances des différents modèles et de leur capacité à anticiper avec précision les tendances de demande.

Conclusion et recommandations

Récapitulation des principales conclusions :

- Impact des variables météorologiques : Les analyses ont souligné l'influence significative des conditions météorologiques sur la demande de locations de vélos. Des températures plus élevées, une humidité particulière ou des vents forts ont été identifiés comme des facteurs clés influençant les habitudes de location.
- Tendances saisonnières : Les schémas saisonniers ont été observés, avec des pics de demande pendant les mois d'été et automne
- Performance des modèles : Les modèles de séries temporelles (ARIMA, lissage exponentiel, etc.) ont été évalués pour prédire avec précision les décomptes de location de vélos, offrant ainsi des outils efficaces pour anticiper la demande.

Recommandations pour les exploitants de partage de vélos :

- Planification saisonnière : Utiliser les données saisonnières pour anticiper les périodes de demande élevée et ajuster les ressources en conséquence, comme l'offre de vélos supplémentaires pendant les mois d'été et d'automne.
- Amélioration des services : Se servir des analyses météorologiques pour fournir des services améliorés en fonction des conditions, par exemple, des offres spéciales ou des incitations pour les jours de conditions météorologiques favorables.

-
- Optimisation de l'inventaire : Utiliser les modèles prédictifs pour gérer efficacement l'inventaire de vélos en prévoyant les fluctuations de la demande.

Cette section résume les principaux résultats obtenus à partir de l'analyse des séries temporelles et offre des recommandations pratiques pour les opérateurs de partage de vélos, mettant en évidence la valeur des analyses prédictives pour améliorer la planification et la gestion des ressources. Ce jeu de données est un trésor pour les analystes de séries temporelles, offrant une pléthore d'opportunités pour comprendre les dynamiques du partage de vélos et pour élaborer des solutions de mobilité urbaine plus intelligentes et efficaces.