



Projets de analyse de séries temporelles Vélo en libre-service

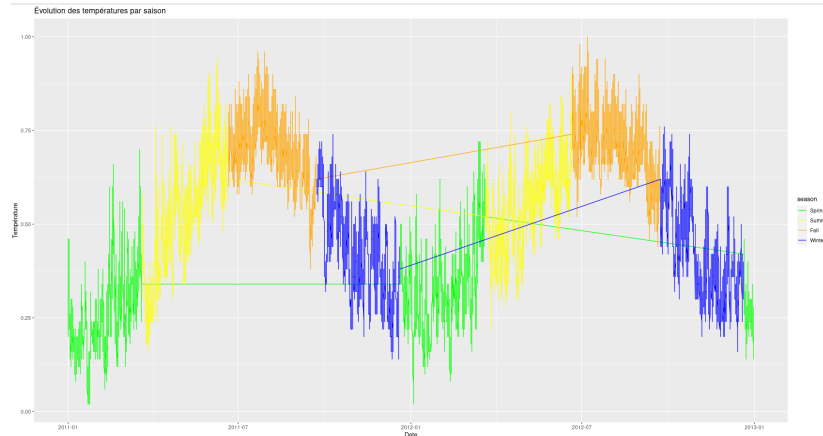
Auteur:
Samb MOUHAMED

Année : 2023/2024

1 Examinez vos données

Évolution des Températures au Fil des Saisons

Les saisons apportent des variations distinctes dans les températures, créant un cycle saisonnier captivant. Voici un aperçu de la façon dont les températures évoluent au fil des saisons :



- **Printemps** : Les températures augmentent progressivement, marquant la fin de l'hiver.
- **Été** : Les températures atteignent généralement leur point culminant, caractérisant la période la plus chaude de l'année.
- **Automne** : Les températures commencent à diminuer, créant des conditions plus fraîches et agréables.
- **Hiver** : Les températures diminuent progressivement, atteignant leurs valeurs les plus basses et marquant la saison froide.

Les fluctuations de température dans chaque saison semblent suivre un schéma de taille relativement constante au fil du temps.

Analyses des Températures

Les données suivantes présentent les températures moyennes et médianes pour chaque saison, initialement en échelle normalisée et ensuite converties en Celsius :

En convertissant ces valeurs normalisées en Celsius, les températures moyennes et médianes en Celsius sont les suivantes :

Ces analyses visuelles et numériques mettent en lumière les variations saisonnières, fournissant une compréhension approfondie des changements climatiques au fil de l'année.

Saison	Moyenne (Normalisée)	Médiane (Normalisée)
Spring	0.299	0.28
Summer	0.545	0.56
Fall	0.706	0.7
Winter	0.423	0.42

Table 1: Températures moyennes et médianes par saison (échelle normalisée).

Saison	Moyenne (Celsius)	Médiane (Celsius)
Spring	6.06	5.16
Summer	17.6	18.3
Fall	25.2	24.9
Winter	11.9	11.7

Table 2: Températures moyennes et médianes par saison (en Celsius).

Corrélation entre la Température et le Nombre Total de Locations

La corrélation entre la température (`temp`) et le nombre total de locations (`cnt`) est d'environ 0.4048. Une corrélation positive indique une tendance où des températures plus élevées sont associées à un nombre total de locations plus élevé. Cependant, la force de cette corrélation est modérée (0.4048), suggérant que d'autres facteurs peuvent également influencer le nombre total de locations.

Corrélation entre la Température Ressentie et le Nombre Total de Locations

La corrélation entre la température ressentie (`atemp`) et le nombre total de locations (`cnt`) est d'environ 0.4009. La température ressentie semble avoir une corrélation similaire à la température réelle avec le nombre total de locations. Les deux variables (`temp` et `atemp`) sont fortement corrélées entre elles (0.9877), ce qui est attendu car la température ressentie est souvent dérivée de la température réelle.

Ces résultats soulignent l'impact positif de la température, ainsi que de la température ressentie, sur le nombre total de locations de vélos. La corrélation significative suggère que des conditions météorologiques plus favorables sont associées à une demande plus élevée pour les services de location de vélos.

Moyennes par Mois

Les moyennes mensuelles de la température, de l'humidité, de la vitesse du vent et du nombre total de locations sont présentées ci-dessous :

Ces données fournissent un aperçu des variations mensuelles dans les conditions météorologiques et l'utilisation du service de location de vélos.

Année	Mois	Temp. Moyenne	Humidité Moyenne	Vitesse Vent Moyenne	Location Moyenne
2011	1	0.197	0.574	0.197	55.5
2011	2	0.284	0.560	0.229	74.3
2011	3	0.331	0.570	0.232	87.7
2011	4	0.471	0.668	0.244	132.0
2011	5	0.577	0.713	0.181	183.0
2011	6	0.693	0.593	0.178	199.0
2011	7	0.759	0.590	0.172	190.0
2011	8	0.706	0.626	0.188	187.0
2011	9	0.613	0.784	0.153	178.0
2011	10	0.470	0.707	0.176	166.0

Table 3: Moyennes mensuelles de la température, de l'humidité, de la vitesse du vent et des locations.

Association entre la Température et les Locations de Vélos

La température est associée aux locations de vélos pour les utilisateurs enregistrés (**registered**) et occasionnels (**casual**). Les résultats des analyses de corrélation révèlent les relations suivantes :

Corrélation avec utilisateurs enregistrés (**registered**) :

- La corrélation entre la température et les locations de vélos pour les utilisateurs enregistrés est d'environ 0.46.
- Une corrélation positive de 0.46 suggère une relation modérée entre la température et les locations de vélos pour les utilisateurs enregistrés.
- En moyenne, les utilisateurs enregistrés ont tendance à louer plus de vélos lorsque la température est plus élevée.

Corrélation avec utilisateurs occasionnels (**casual**) :

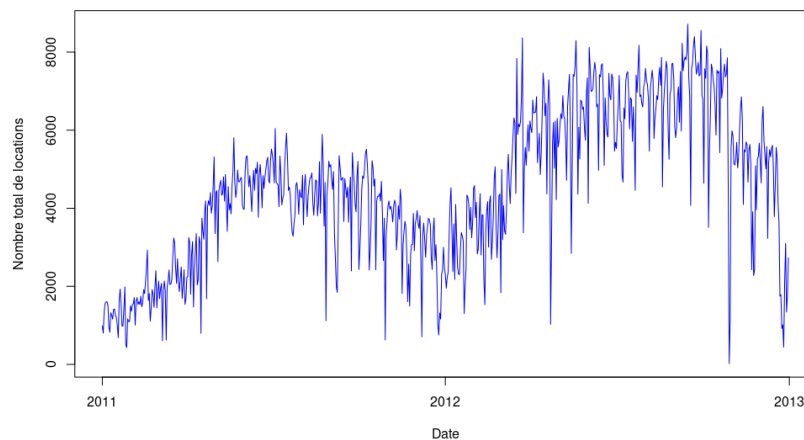
- La corrélation entre la température et les locations de vélos pour les utilisateurs occasionnels est d'environ 0.34.
- Une corrélation positive de 0.34 suggère également une relation positive, mais légèrement plus faible, entre la température et les locations de vélos pour les utilisateurs occasionnels.

En conclusion, les résultats indiquent que la température a une influence positive sur les locations de vélos, aussi bien pour les utilisateurs enregistrés que pour les utilisateurs occasionnels. La corrélation positive suggère que, en général, les locations de vélos augmentent avec une augmentation de la température.

Cela peut être attribué à une préférence générale pour le vélo par temps plus chaud, observée chez les deux groupes d'utilisateurs.

Tracé du Nombre Total de Locations en Fonction du Jour

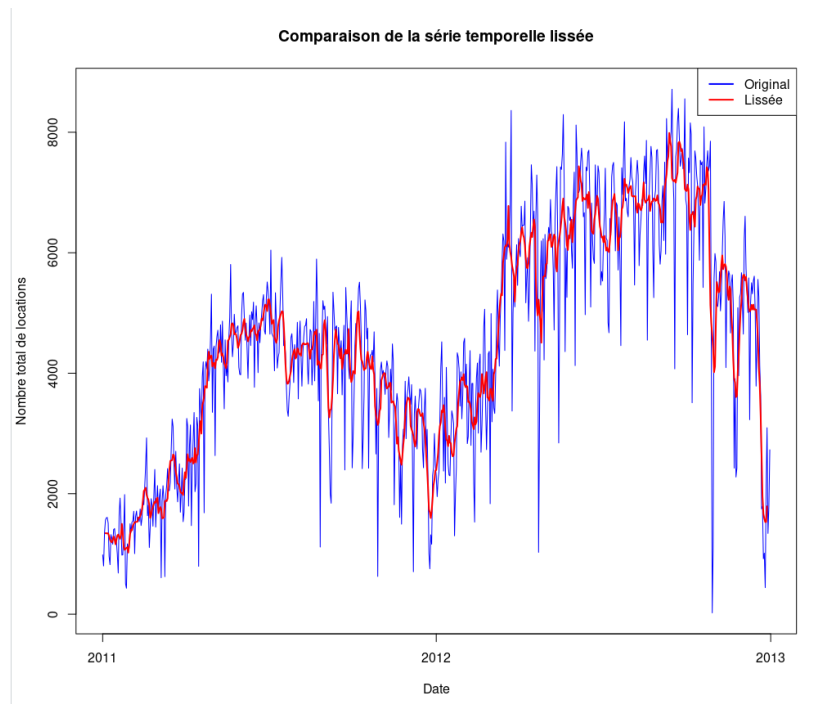
Le graphique suivant représente le nombre total de locations (`cnt`) en fonction du jour, permettant d'examiner les tendances et les irrégularités.



Le graphique montre clairement une tendance saisonnière, avec un pic des locations en été et un creux en hiver. On observe une irrégularité marquée, illustrée par l'augmentation significative du nombre de locations en été 2012 par rapport à l'été 2011, ainsi que la diminution notable du nombre de locations en hiver 2012 par rapport à l'hiver 2011.

Série Temporelle Lissée

La nouvelle série temporelle lissée montre une tendance à la hausse plus régulière que la série temporelle originale. Les irrégularités sont également moins visibles.

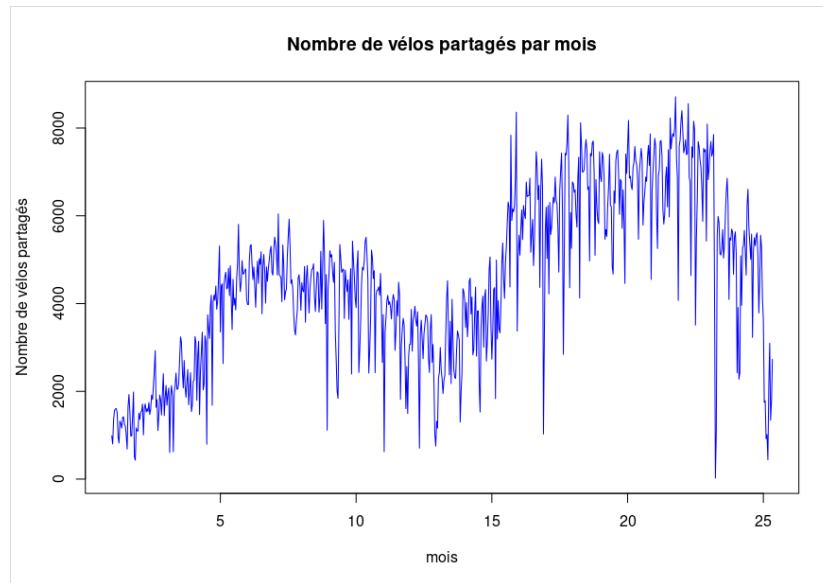


2 Utilisation de la Série Temporelle Lissée du Nombre Total de Locations

Dans cette section, nous explorerons l'utilisation de la série temporelle lissée du nombre total de locations (`cnt_lissé`). La méthode de lissage utilisée est la moyenne mobile. La moyenne mobile aide à réduire le bruit dans une série temporelle en lissant les fluctuations à court terme. Cela peut rendre les tendances sous-jacentes plus visibles.

Série Temporelle avec Fréquence de 30 Observations par ans

Le graphique ci-dessous représente une série temporelle avec une fréquence de 30 observations par Années entre janvier 2011 et janvier 2013.



La fréquence de 30 observations par années a été choisie pour capturer les variations mensuelle sur une période de deux ans, offrant ainsi une résolution fine et détaillée de la série temporelle.

La nouvelle série temporelle, avec une fréquence de 30, est plus stationnaire que la série temporelle originale, avec une fréquence de 1. Cela est dû au fait que la nouvelle série temporelle est moyenne sur une période plus longue, ce qui atténue les variations saisonnières.

3 Modélisation de la Série Temporelle Lissée avec le Modèle ARIMA

Graphiques ACF et PACF

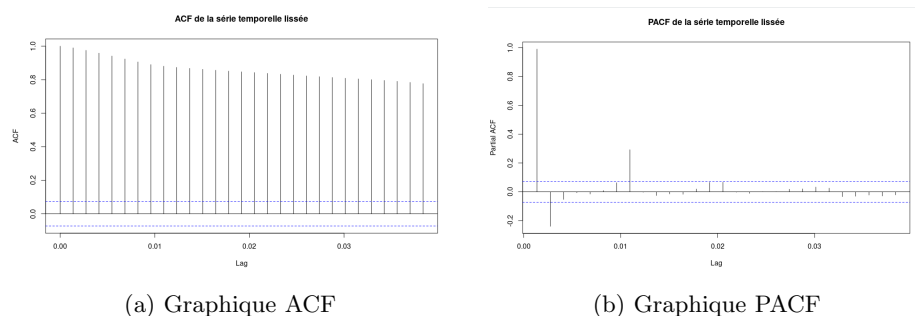


Figure 1: Graphiques ACF et PACF pour la série temporelle lissée.

Sur la base de l'ACF de la série temporelle lissée, il y a plusieurs retards significatifs qui peuvent nous donner des indications sur les ordres du modèle ARIMA. Sur la base du PACF, il y a un seul retard significatif, qui est le retard 1. Cela suggère qu'il existe une corrélation entre les valeurs actuelles et les valeurs précédentes de la série temporelle, mais que cette corrélation disparaît après une période. Sur la base de ce résultat, un modèle ARIMA(1,0,0) serait un bon point de départ pour la modélisation de la série temporelle. Ce modèle est un modèle autorégressif d'ordre 1, ce qui signifie qu'il utilise la valeur précédente de la série temporelle pour prédire la valeur actuelle.

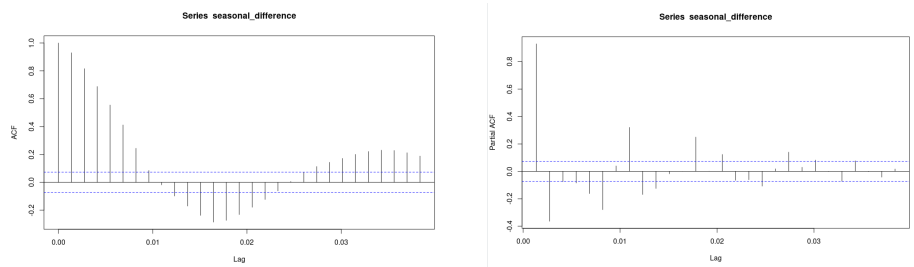
4 Prévision avec les Modèles ARIMA

Ajustement du Modèle ARIMA sur la Série Temporelle Désaisonnalisée de `cnt`

Nous allons maintenant ajuster un modèle ARIMA à la série temporelle désaisonnalisée de `cnt`. Cette série temporelle a été ajustée pour éliminer les variations saisonnières, permettant ainsi de se concentrer sur les tendances sous-jacentes.

Graphiques ACF et PACF de la Série Temporelle Désaisonnalisée

Les graphiques ACF (AutoCorrelation Function) et PACF (Partial AutoCorrelation Function) de la série temporelle désaisonnalisée peuvent fournir des indications sur les ordres appropriés pour le modèle ARIMA.



(a) Graphique ACF de la série temporelle désaisonnalisée de `cnt`. (b) Graphique PACF de la série temporelle désaisonnalisée de `cnt`.

Figure 2: Graphiques ACF et PACF de la série temporelle désaisonnalisée de `cnt`.

Sur la base de ces graphiques, nous pouvons identifier les retards significatifs pour l'ajustement du modèle ARIMA. Grâce à cela nous avons choisi d'ajuster un modèle ARIMA(1,1,1) à la série temporelle désaisonnalisée de `cnt`. Ce modèle est spécifié par un composant autorégressif (AR) d'ordre 1, un composant de différenciation (I) d'ordre 1, et un composant de moyenne mobile (MA) d'ordre 1.

Ajustement ARIMA avec Auto-ARIMA

Après utilisation de la fonction `auto.arima()` pour ajuster un modèle ARIMA à la série temporelle désaisonnalisée de `cnt`. Cette fonction cherche automatiquement le modèle qui minimise un critère d'information comme le critère d'information bayésien (BIC) ou le critère d'information akaike (AIC).

Après l'ajustement du modèle, nous avons examiné les résidus du modèle en utilisant la fonction `checkresiduals`, ce qui nous a permis de valider la qualité du modèle. Voici le graphique des résidus obtenus :

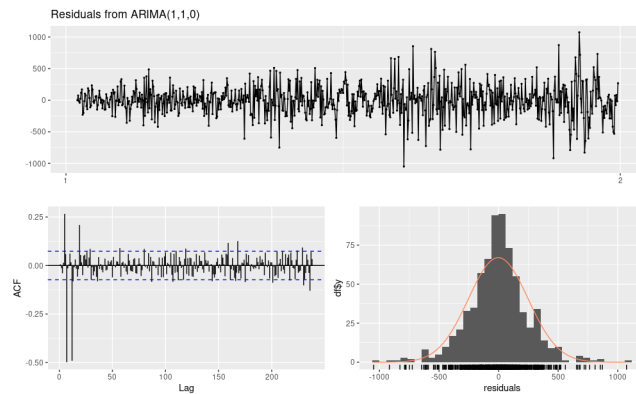


Figure 3: Graphique des résidus du modèle ARIMA ajusté avec la fonction `auto.arima()`.

5 Prévisions avec les modèles ARIMA

Après avoir ajusté le modèle ARIMA sur la série temporelle lissée, nous avons procédé à la prévision des 25 prochaines observations. Nous avons également comparé ces prévisions avec un modèle ajusté manuellement pour évaluer la performance de notre modèle automatique.

Prévision Automatique

Les prévisions automatiques ont été générées en utilisant le modèle ARIMA ajusté sur la série temporelle lissée. Le graphique suivant compare la série temporelle originale avec les prévisions automatiques :

Prévision Manuelle

En parallèle, nous avons ajusté manuellement un modèle ARIMA sur la partie d'entraînement de la série temporelle lissée. Les prévisions manuelles ont été comparées avec la série temporelle originale et les prévisions automatiques pour évaluer la performance du modèle manuel :

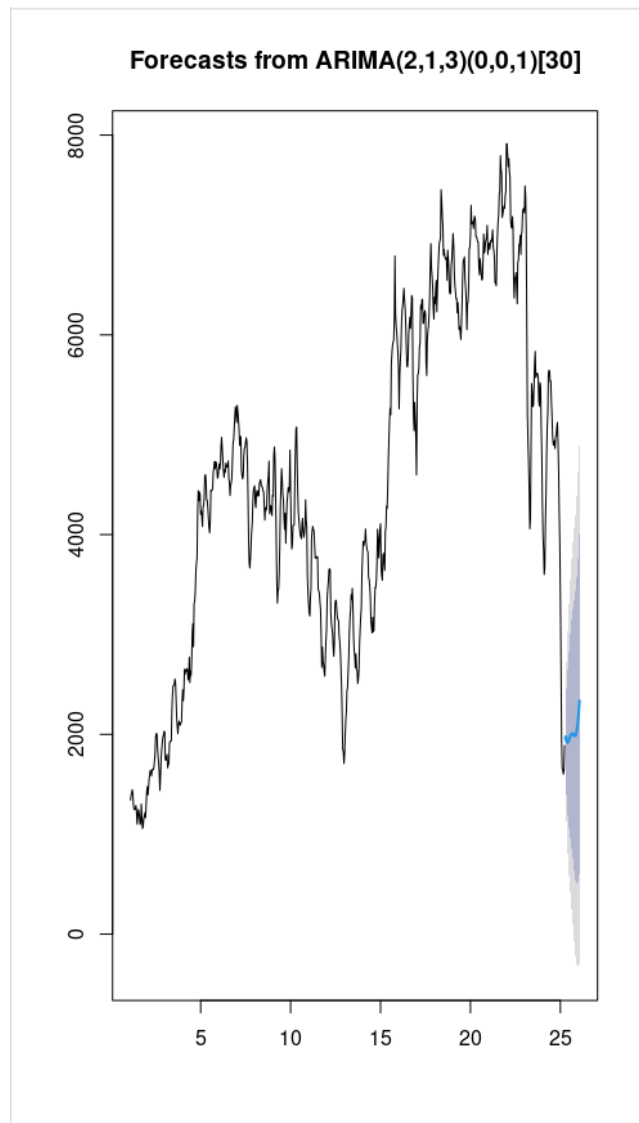


Figure 4: Prévisions automatiques avec le modèle ARIMA.

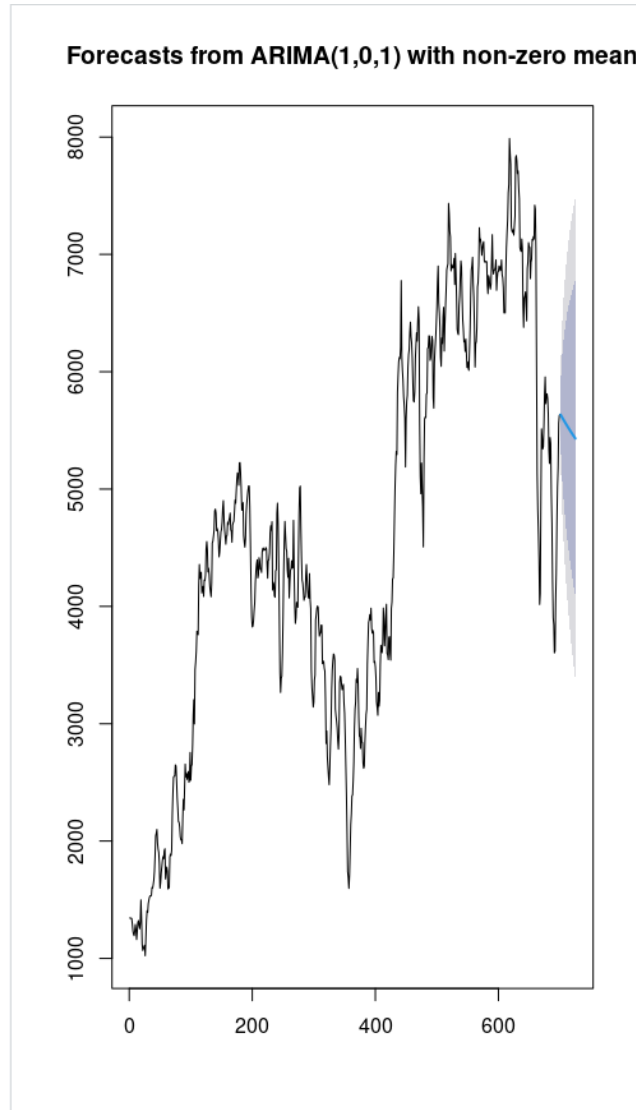


Figure 5: Prévisions manuelles avec le modèle ARIMA ajusté manuellement.

Ces comparaisons visuelles permettent d'évaluer l'efficacité des modèles ARIMA dans la prévision des locations de vélos.

6 Conclusion

En résumé, il apparaît que les prévisions générées par le modèle ARIMA ajusté manuellement présentent une cohérence et une adéquation plus naturelles par rapport à celles issues du modèle ARIMA ajusté automatiquement.