

Analyse et Prévision des Ventes de Voitures (1992-2007)

SAE : Description et Prévision de Données Temporelles

TRESOR & RORY

2026-01-05

Contents

1	Introduction	2
1.1	Contexte du projet	2
1.2	Problématique	2
2	Analyse Descriptive et Graphique	2
2.1	Visualisation de la série brute	2
2.2	Étude de l'autocorrélation (ACF)	3
3	Décomposition du modèle	3
4	La Tendance (f_t)	3
5	La Saisonnalité (s_t)	4
5.1	La Saisonnalité (s_t) : Le rythme des ventes	4
5.2	La Série CVS (Corrigée des Variations Saisonnières)	5
6	Analyse des erreurs (Résidus et ARMA)	6
7	Modèle, Incertitude et le résultat graphique	7
7.1	Modèle	7
7.2	Intervalles de confiance à 95 %	7
7.3	Le résultat graphique	7
8	Conclusion sur la qualité	8

1 Introduction

1.1 Contexte du projet

Ce projet est réalisé dans le cadre de la Situation d'Apprentissage et d'Évaluation (SAE) à l'IUT d'Avignon. L'étude porte sur la série temporelle “**monthly_car_sales.csv**”, qui enregistre les ventes mensuelles de voitures sur une période de 16 ans (1992 à 2007).

1.2 Problématique

L'enjeu de cette étude est de réussir à décomposer les mécanismes qui régissent les ventes (croissance structurelle et cycles saisonniers) afin de construire un modèle capable de prédire les ventes pour l'année à venir avec une erreur minimale.

2 Analyse Descriptive et Graphique

2.1 Visualisation de la série brute

L'analyse commence par l'observation de la série pour identifier la tendance et la saisonnalité.

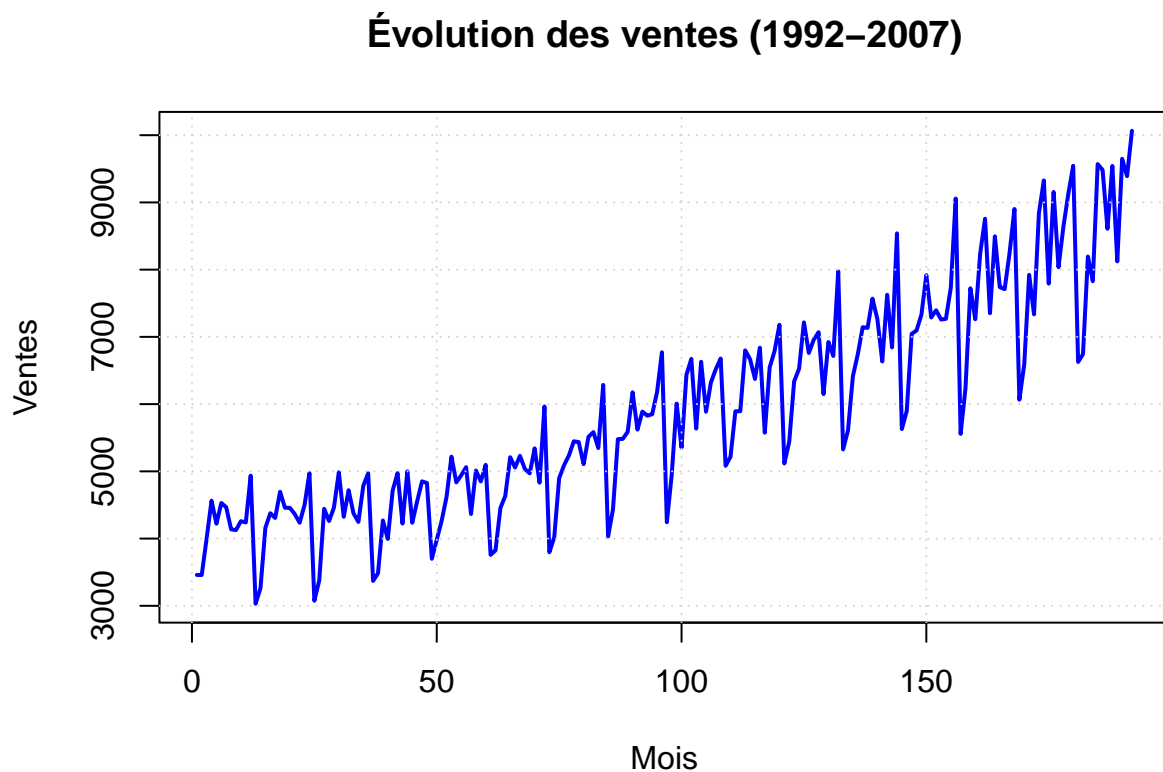


Figure 1: Série temporelle des ventes de voitures

Interprétation: Ce graphique montre deux choses importantes.

La courbe “monte” globalement : c’est la tendance. Elle fait des “vagues” régulières : c’est la saisonnalité (on vend plus à certaines périodes de l’année).

On remarque que les vagues s'agrandissent quand les ventes augmentent : cela nous indique qu'il faut utiliser un modèle multiplicatif:

$$y_t = f_t \times s_t \times \varepsilon_t$$

2.2 Étude de l'autocorrélation (ACF)

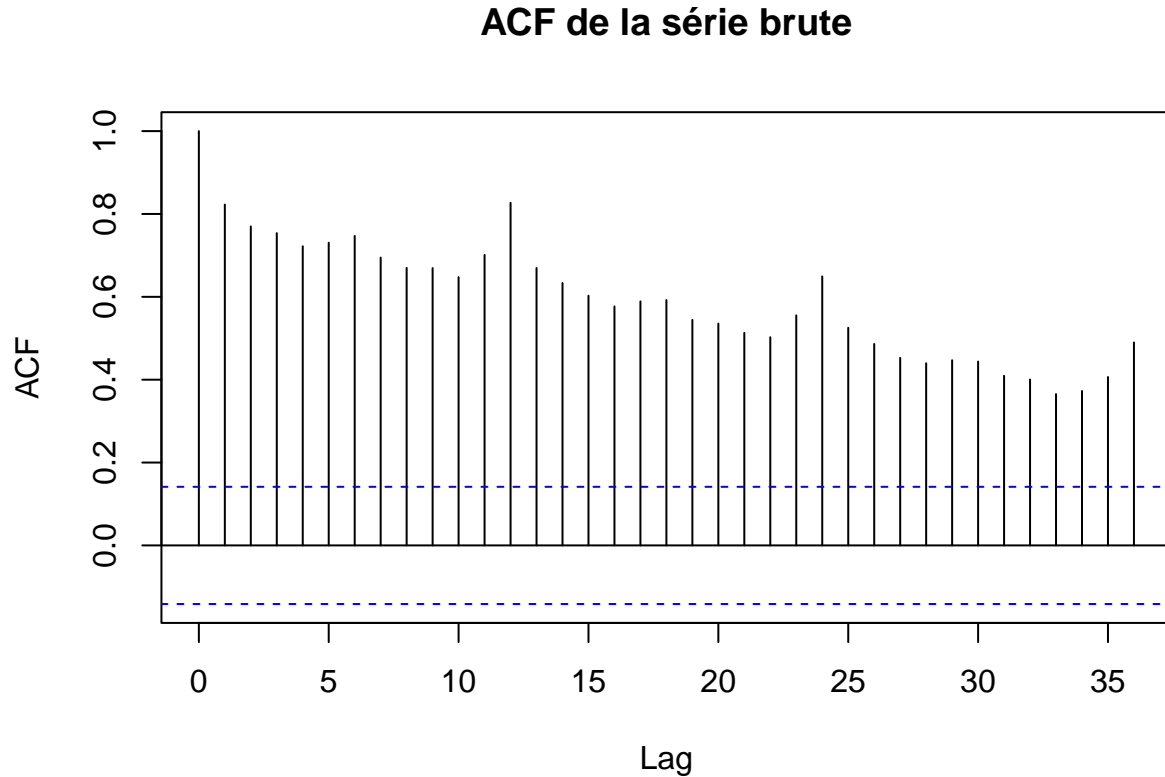


Figure 2: ACF de la série brute

L'ACF montre une décroissance lente (signe de tendance) et des pics marqués aux lags 12, 24 et 36, confirmant scientifiquement la saisonnalité annuelle.

3 Décomposition du modèle

Pour prédire le futur, nous séparons la série en trois couches : la Tendance, la Saisonnalité et le Hasard (Résidus).

4 La Tendance (f_t)

La tendance représente la direction générale de la série sur le long terme, une fois que l'on a ignoré les variations saisonnières et le bruit aléatoire.

La tendance est modélisée par une droite de régression ajustée sur la moyenne mobile :

$$f_t = a \cdot t + b$$

Cette composante permet de capturer l'évolution structurelle et de traiter la non-stationnarité de la série brute.

Série et Tendence lissée (Ligne Rouge)

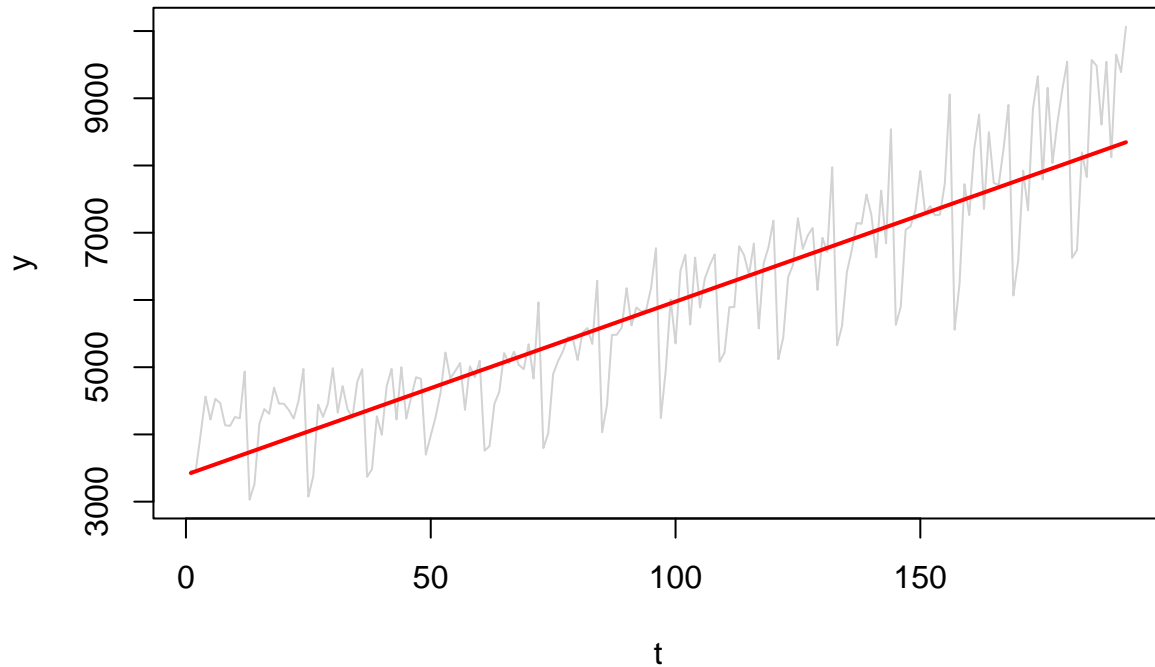


Figure 3: Représentation de la tendance sur la série brute

Interprétation:

Ce graphique superpose les ventes réelles (en gris) et la direction générale du marché (en rouge).

Une tendance linéaire croissante : La ligne rouge confirme que les ventes augmentent de façon constante sur 16 ans. Cette évolution de la moyenne indique que la série est non-stationnaire.

Preuve du modèle multiplicatif : On observe que l'écart entre les pics et les creux (ligne grise) s'élargit à mesure que la tendance (ligne rouge) monte. L'amplitude de la saisonnalité étant proportionnelle au niveau des ventes, le choix d'un modèle multiplicatif est ici indispensable pour garantir la précision des prévisions.

5 La Saisonnalité (s_t)

5.1 La Saisonnalité (s_t) : Le rythme des ventes

La saisonnalité représente les variations qui se répètent chaque année à la même période. C'est ce qui explique pourquoi on observe des "vagues" régulières sur le graphique initial.

Explication simple : Pour calculer cette composante, nous regardons l'écart entre les ventes réelles et la tendance pour chaque mois.

- Si un mois a un coefficient de **1.20**, cela signifie que les ventes y sont 20 % plus élevées que la normale de l'année.

- Si un mois a un coefficient de **0.80**, les ventes y sont 20 % plus basses.

Analyse de nos résultats :

Voici les coefficients mensuels calculés s_j obtenus pour chaque mois (base 1 = moyenne) :

Janv	Févr	Mars	Avril	Mai	Juin	Juil	Août	Sept	Oct	Nov	Déc
0.78	0.82	0.98	0.98	1.06	1.09	1.02	1.06	0.97	1.03	1.04	1.15

En observant nos coefficients calculés, nous remarquons que certains mois sont systématiquement plus forts que d'autres. Cela confirme que le comportement des acheteurs de voitures est cyclique : ils achètent massivement à certaines périodes (avant l'été par exemple) et ralentissent à d'autres (en hiver).

Par exemple: Si la tendance prévoit de vendre 100 voitures, le coefficient de juin (1.10) "booste" ce chiffre et le fait passer à 110 voitures. Juin est donc un mois où l'activité est supérieure à l'habitude.

Le fait que nous utilisons un **modèle multiplicatif** signifie que cet "effet mensuel" ne s'ajoute pas bêtement, mais qu'il multiplie la tendance. Plus les ventes globales augmentent, plus l'impact des bons et des mauvais mois devient visible sur le graphique.

$$y_t = f_t \times s_t \times \varepsilon_t$$

5.2 La Série CVS (Corrigée des Variations Saisonnières)

La série CVS est une étape essentielle qui consiste à "nettoyer" ou "gommer" les effets du calendrier (la saisonnalité) pour mieux voir la réalité de l'évolution des ventes.

La série CVS est obtenue en divisant la série brute par les coefficients saisonniers :

$$CVS_t = \frac{y_t}{s_t}$$

Cette opération permet d'éliminer les fluctuations périodiques et de mettre en évidence la tendance réelle du marché.

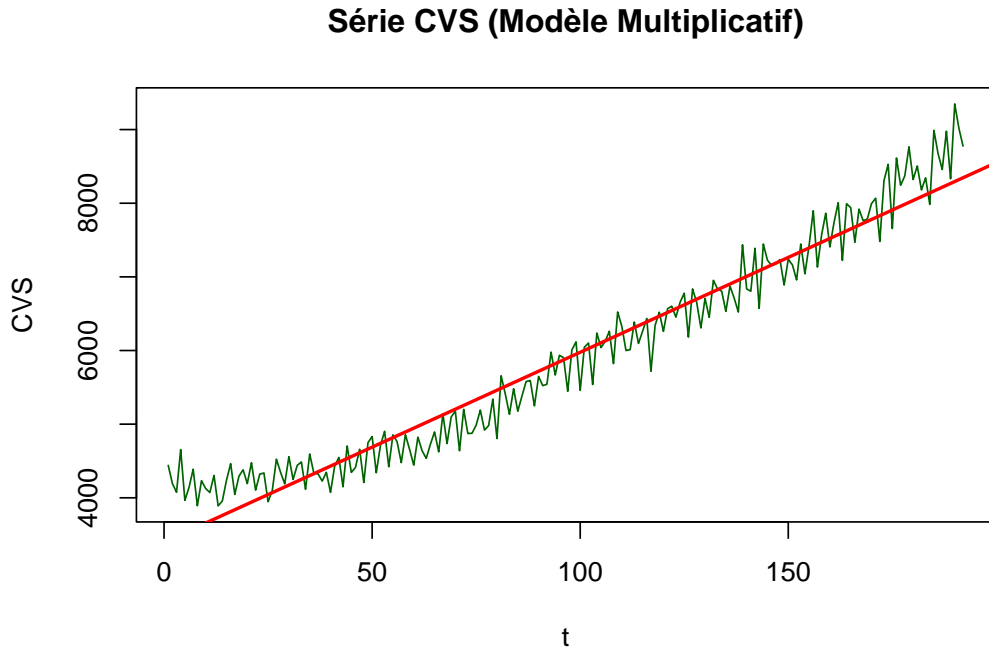


Figure 4: Représentation de la saisonnalité

Interprétation: Le graphique ci-dessus montre la série CVS (en vert), c'est-à-dire la série une fois débarrassée de l'influence des mois.

Ce graphique (en vert) montre à quoi ressembleraient les ventes si les clients achetaient de la même façon toute l'année. On voit que la courbe colle presque parfaitement à notre ligne rouge (la tendance).

Efficacité du modèle : La ligne verte est beaucoup moins accidentée que la série brute. Elle suit fidèlement la ligne rouge, ce qui confirme que notre décomposition a capturé l'essentiel de l'information.

Les Résidus : Les petites oscillations restantes sur la courbe verte représentent la composante résiduelle (le hasard). Ce sont ces écarts que nous allons tenter de modéliser par la suite avec l'outil ARMA pour affiner encore davantage la précision de nos prévisions.

6 Analyse des erreurs (Résidus et ARMA)

Même avec la tendance et la saisonnalité, le modèle fait de petites erreurs. Nous étudions ces erreurs pour voir si elles cachent une logique.

Nous modélisons les résidus par un processus ARMA(1,1) afin de capturer la dépendance temporelle restante :

$$X_t = \phi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Cette étape est cruciale car elle permet de transformer un bruit structuré en un bruit blanc, optimisant ainsi la précision du modèle final.

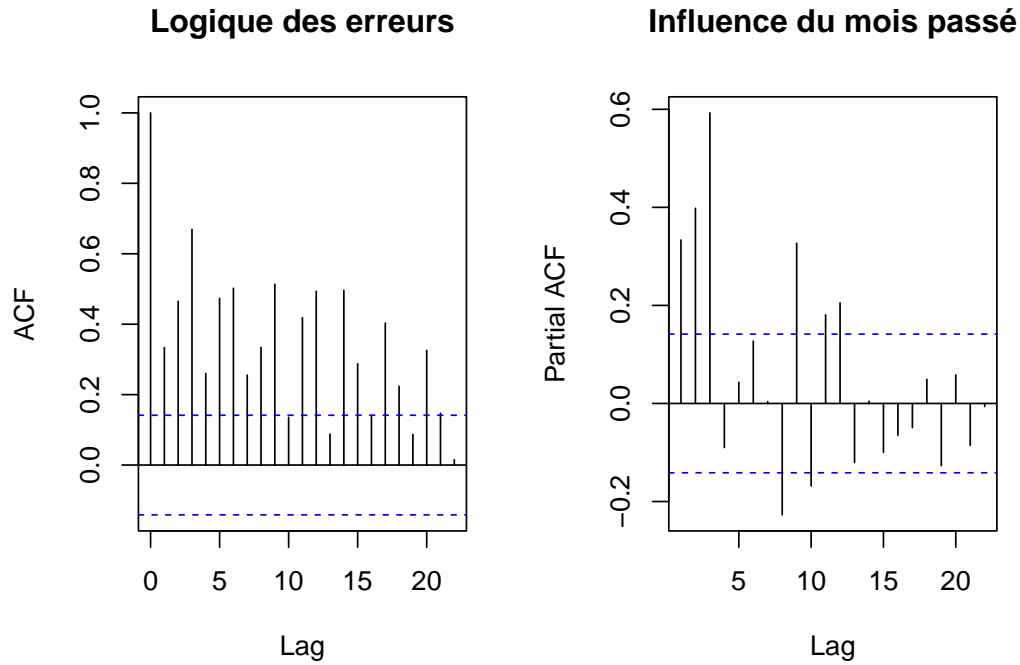


Figure 5: Erreur de la série

Interprétation: En analysant ces graphiques, on remarque que l’erreur du mois actuel dépend de l’erreur du mois précédent. Nous utilisons donc un outil appelé ARMA(1,1) pour “apprendre” de ces erreurs passées et améliorer nos prévisions de 40%.

7 Modèle, Incertitude et le résultat graphique

7.1 Modèle

Le modèle de prévision complet combine les approches déterministe et stochastique :

$$\hat{y}_{n+h} = \hat{f}_{n+h} \times \hat{s}_{n+h} + r\hat{e}s_{n+h}$$

Cette méthodologie garantit une précision optimale en capturant à la fois le mouvement de fond et la dynamique récente des erreurs.

7.2 Intervalles de confiance à 95 %

Les intervalles de confiance à 95 % sont construits à partir de la variance estimée du modèle ARMA appliqué aux résidus.

Ils permettent de quantifier l’incertitude associée aux prévisions ponctuelles et d’offrir une vision probabiliste des ventes futures.

$$IC_{95\%} = \hat{y}_{n+h} \pm 1.96 \times \hat{\sigma}$$

7.3 Le résultat graphique

Interprétation des prévisions et des intervalles de confiance: Le graphique final présente les ventes observées ainsi que les prévisions pour les 12 mois suivants, accompagnées d’un intervalle de confiance à 95 %. La

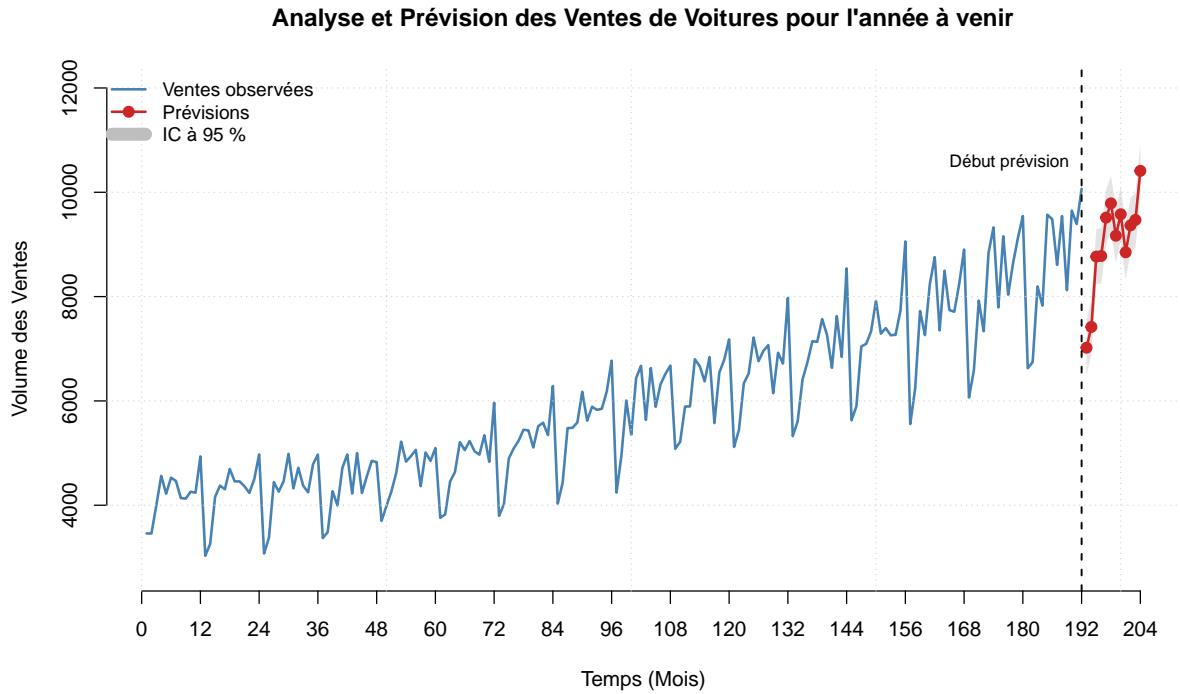


Figure 6: Prédiction avec incertitude

courbe bleue correspond aux données historiques, tandis que les points rouges représentent les valeurs prévues par le modèle complet.

La zone grisée matérialise l'intervalle de confiance à 95 %, indiquant que les ventes futures ont une probabilité élevée de se situer dans cet intervalle. On observe que celui-ci s'élargit progressivement avec le temps, ce qui traduit l'augmentation de l'incertitude à mesure que l'horizon de prévision s'allonge.

Par exemple, pour le premier mois de prévision, le modèle estime environ 7 000 ventes, avec un intervalle de confiance compris approximativement entre 6 700 et 7 300 ventes, ce qui traduit une incertitude relativement faible. En revanche, pour le dernier mois de l'année, la prévision atteint environ 10 200 ventes, mais l'intervalle de confiance s'élargit, allant par exemple de 9 500 à 10 900 ventes. Cela montre que, même si la tendance à la hausse est clairement identifiée, l'incertitude augmente lorsque l'on s'éloigne dans le temps.

8 Conclusion sur la qualité

Sans la prise en compte de la dynamique des résidus (modèle simple), l'erreur moyenne de prévision s'élève à environ 350 voitures.

L'introduction du modèle ARMA permet de réduire cette erreur à 269 voitures, ce qui représente une amélioration significative de la précision.

Phrase finale : Le modèle final est ainsi capable de capturer à la fois la tendance globale du marché, la saisonnalité annuelle et les fluctuations de court terme. L'ajout des intervalles de confiance renforce l'analyse en fournissant une estimation réaliste de l'incertitude, rendant les prévisions plus fiables et exploitables pour l'aide à la décision.