

UNIVERSITE CLERMONT AUVERGNE
IUT d'Aurillac
Sciences de données

Projet pour la SAE 5.EMS.01
*Analyse des ventes de produits dans un réseau
de magasins*

Encadrant :

Mme SOHIER Emilie

Membres du groupe :

BELLO Habibath
KONATE Batté Naïmatou
DABIRE Saânbèterfaa Joël
SYLLA Sokhna Awa Bousso

Année Universitaire : 2025–2026

Table des matières

Introduction	1
1 Méthodologie	1
2 Nettoyage et préparation des données	2
2.1 Valeurs manquantes	2
2.2 Statistiques descriptives simples	2
2.3 Valeurs atypiques	3
3 Analyse descriptive	4
3.1 Évolution des ventes dans le temps	4
3.2 Tendances saisonnières	5
3.3 Effet des promotions	6
3.4 Différence semaine / week-end	6
3.5 Compétition entre magasins	7
4 Modélisation statistique	7
4.1 Objectif et choix du modèle	7
4.2 Construction des modèles	8
4.3 Comparaison des modèles	8
4.4 Choix du modèle retenu	9
4.5 Résultats du modèle retenu	9
4.6 Interprétation générale du modèle	9
4.7 4.7 Graphiques de qualité du modèle	10
5 Discussion et interprétation des résultats	13
5.1 Interprétation globale des résultats	13
5.2 Différences entre les magasins	13
5.3 Effet des promotions et du week-end	13
5.4 Rôle de la saisonnalité	13
5.5 Impact de la compétition et des finances	14
6 Recommandations	14
6.1 Optimisation des promotions	14

6.2	Prise en compte de la saisonnalité	14
6.3	Amélioration de la compétitivité entre magasins	15
6.4	Allocation des ressources financières	15

Table des figures

1	Évolution des ventes dans le temps	4
2	Tendances saisonnières	5
3	Effet des promotions	6
4	Différence semaine / week-end	6
5	Compétition entre magasins	7
6	Graphique Prédictions vs Observations	11
7	Graphique de l'erreur moyenne par tranche de ventes prédites	12

Liste des tableaux

1	Statistiques descriptives des ventes par produit	2
2	Ventes moyennes par magasin	3
3	Statistiques descriptives des ventes par année	3
4	Comparaison des modèles par le critère AIC	8
5	Coefficients estimés du modèle de Poisson retenu	9

Introduction

Dans un contexte de forte concurrence entre les enseignes de distribution, l'analyse des ventes constitue un enjeu majeur pour améliorer la performance commerciale des magasins. Une meilleure compréhension des facteurs influençant la quantité vendue permet d'orienter les décisions stratégiques, d'optimiser les promotions et d'adapter l'offre aux attentes des clients.

Ce projet porte sur l'analyse des ventes de produits alimentaires dans un réseau de plusieurs magasins sur la période 2022–2024. L'objectif principal est d'étudier l'évolution des ventes dans le temps, d'identifier les facteurs ayant un impact significatif sur la demande et de proposer des recommandations pour améliorer les performances du réseau.

Ce rapport présente successivement la méthodologie utilisée, les analyses descriptives réalisées, les résultats de la modélisation statistique ainsi que des recommandations opérationnelles pour améliorer les ventes.

1 Méthodologie

L'objectif de ce travail est d'analyser les ventes de produits dans un réseau de magasins afin d'identifier les principaux facteurs qui influencent la quantité vendue et de proposer des recommandations opérationnelles. Pour cela, nous avons suivi une démarche en plusieurs étapes, allant du nettoyage des données à la modélisation statistique et à l'évaluation des résultats.

Dans un premier temps, nous avons travaillé sur deux bases de données : une base contenant les informations de ventes (produit, magasin, date, promotion, prix unitaire et quantité vendue) et une base financière regroupant des indicateurs par magasin (chiffre d'affaires, budget publicité, masse salariale et dépenses de fonctionnement). Ces deux bases ont été harmonisées afin de pouvoir être correctement croisées par la suite.

Ensuite, nous avons procédé au nettoyage et à la préparation des données. Cette étape a consisté à diagnostiquer les valeurs manquantes, à créer des variables temporelles (année, mois, jour de la semaine, week-end) et à détecter les valeurs atypiques à l'aide de la méthode de l'écart interquartile (IQR). Les valeurs manquantes de la variable `Quantité_Vendue` ont été imputées par une médiane calculée en tenant compte du produit, du magasin et du jour de la semaine, afin de conserver le plus d'information possible tout en restant cohérent avec la structure des données.

Une fois les données propres et enrichies, nous avons réalisé une analyse descriptive pour explorer les tendances générales des ventes, la saisonnalité, l'effet des promotions et les différences entre les magasins. Nous avons également construit un indicateur de compétition entre magasins, basé sur le ratio entre les ventes d'un magasin et la moyenne des ventes des autres magasins le même jour.

Dans un second temps, nous avons mis en place une modélisation statistique. La variable d'intérêt étant une variable de comptage (quantité vendue), nous avons choisi un modèle linéaire généralisé de Poisson avec lien logarithmique. Plusieurs modèles ont été testés en ajoutant progressivement des variables explicatives (promotion, week-end, saison, compétition, prix et variables financières). Ces modèles ont ensuite été comparés à l'aide du critère AIC afin de retenir celui offrant le meilleur compromis entre qualité d'ajustement et complexité.

Enfin, le modèle retenu a été évalué à l'aide d'indicateurs de qualité prédictive (erreur moyenne absolue et corrélation entre valeurs observées et prédites) et de graphiques de diag-

nostic. Les résultats ont ensuite été interprétés afin de formuler des recommandations concrètes pour améliorer la gestion des ventes et la performance des magasins.

2 Nettoyage et préparation des données

2.1 Valeurs manquantes

Avant toute analyse, nous avons vérifié la présence de valeurs manquantes dans les deux bases de données. Cette étape est importante car des données manquantes peuvent biaiser les résultats ou rendre certains calculs impossibles.

Dans la base financière des magasins, aucune variable ne présente de valeurs manquantes. En revanche, dans la base des ventes, seule la variable `Quantité_Vendue` contient des données manquantes. Nous avons identifié 2040 valeurs manquantes sur un total de 40 860 observations, ce qui représente environ 4,99 % des données.

Plutôt que de supprimer ces lignes, ce qui aurait entraîné une perte d'information, nous avons choisi d'imputer les valeurs manquantes. Nous avons utilisé une méthode d'imputation hiérarchique par médiane, en respectant le contexte des données. Concrètement, pour chaque valeur manquante, nous avons procédé de la manière suivante :

- d'abord, nous avons remplacé la valeur manquante par la médiane des ventes du même produit, dans le même magasin et le même jour de la semaine ;
- si cette information n'était pas disponible, nous avons utilisé la médiane par produit et magasin ;
- sinon, nous avons utilisé la médiane par produit ;
- en dernier recours, nous avons utilisé la médiane globale des ventes.

Après cette procédure, il ne restait plus aucune valeur manquante dans la variable `Quantité_Vendue`, ce qui nous a permis de poursuivre l'analyse sur une base complète et cohérente.

2.2 Statistiques descriptives simples

Statistiques par produit

Produit	Moyenne	Médiane	Min	Max	n
Beurre	66.1	59	3	280	4086
Bœuf	69.5	64	6	315	4086
Café	54.6	50	8	282	4086
Jus d'orange	53.3	44	4	501	4086
Lait	73.4	64	12	372	4086
Pain	77.7	71	9	495	4086
Pâtes	72.6	68	4	402	4086
Poulet	67.3	58	8	345	4086
Riz	60.9	52	5	516	4086
Yaourt	53.8	50	5	250	4086

TABLE 1 – Statistiques descriptives des ventes par produit

Les statistiques montrent que les ventes moyennes varient selon les produits. Le Pain et le Lait sont les produits les plus vendus en moyenne, avec respectivement 77,7 et 73,4 unités vendues. À l'inverse, le Café, le Jus d'orange et le Yaourt présentent des niveaux de ventes plus faibles, autour de 53 unités en moyenne. On observe également une forte variabilité pour certains produits, avec des valeurs maximales très élevées (jusqu'à 516 unités pour le Riz), ce qui indique que des pics de ventes peuvent survenir ponctuellement.

Statistiques par magasin

Magasin	Moyenne	n
Magasin_1	60.8	6810
Magasin_2	69.7	6810
Magasin_3	55.7	6810
Magasin_4	66.5	6810
Magasin_5	63.8	6810
Magasin_6	73.0	6810

TABLE 2 – Ventes moyennes par magasin

Les ventes moyennes diffèrent selon les magasins. Le Magasin_6 présente la moyenne la plus élevée avec 73 unités vendues, tandis que le Magasin_3 affiche la moyenne la plus faible avec 55,7 unités. Ces écarts suggèrent que certains magasins sont globalement plus performants que d'autres, ce qui justifie l'étude de la compétition entre magasins dans la suite du projet.

Statistiques sur la période étudiée (par année)

Année	Moyenne	Médiane	Total des ventes	n
2022	60.6	53	942 050	15 540
2023	66.6	59	1 459 453	21 900
2024	73.5	69	251 281	3 420

TABLE 3 – Statistiques descriptives des ventes par année

L'analyse par année montre une augmentation progressive des ventes moyennes au cours du temps : elles passent de 60,6 en 2022 à 66,6 en 2023 puis à 73,5 en 2024. La médiane suit la même tendance, ce qui indique que cette hausse concerne l'ensemble des ventes et pas seulement quelques valeurs extrêmes.

Le total des ventes est plus élevé en 2023 car cette année contient davantage d'observations, tandis que 2024 présente un total plus faible car la période étudiée ne couvre qu'une partie de l'année. Dans l'ensemble, ces résultats suggèrent une légère amélioration des performances de vente au fil du temps.

2.3 Valeurs atypiques

Pour identifier d'éventuelles valeurs atypiques dans la variable Quantité_Vendue, nous avons utilisé la méthode de l'écart interquartile (IQR). Cette méthode consiste à définir des bornes

à partir du premier et du troisième quartile. Nous avons obtenu une borne inférieure de -36,5 et une borne supérieure de 159,5. À partir de ces seuils, 758 observations ont été identifiées comme atypiques, soit environ 1,86% des données. Cette proportion étant faible, nous avons choisi de conserver ces valeurs dans l'analyse, car elles peuvent correspondre à des situations réelles de très fortes ventes (par exemple lors de promotions ou de périodes de forte demande).

3 Analyse descriptive

3.1 Évolution des ventes dans le temps

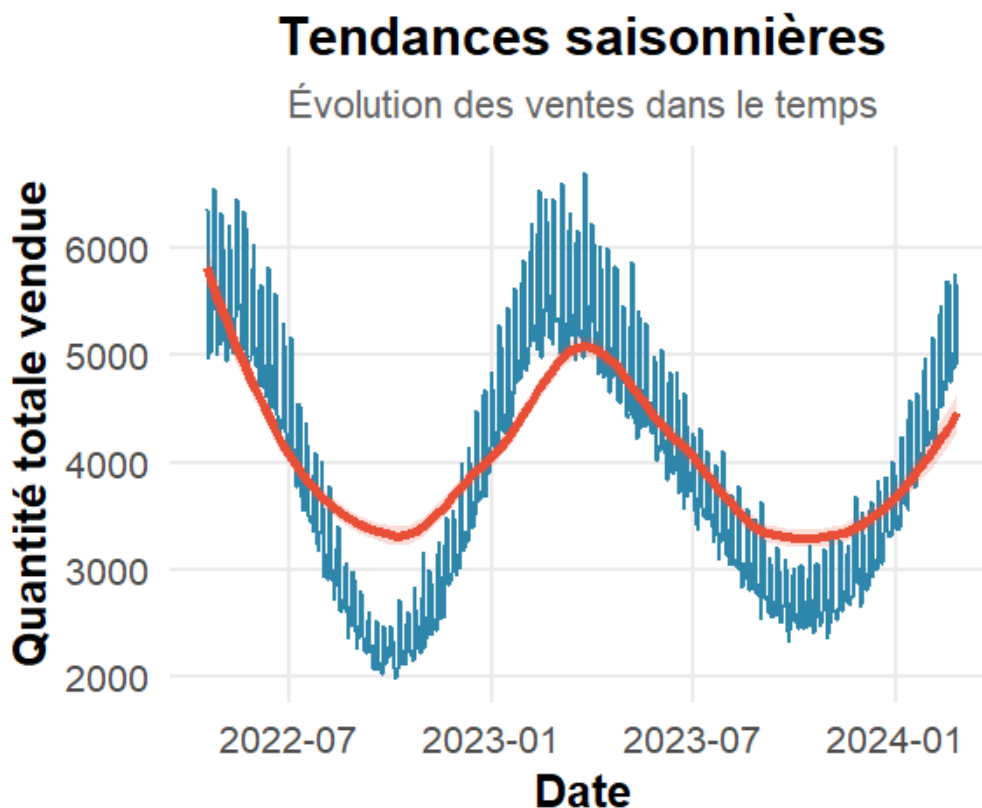


FIGURE 1 – Évolution des ventes dans le temps

Le graphique de l'évolution des ventes totales au cours du temps montre une variation régulière et cyclique des quantités vendues entre 2022 et 2024. On observe clairement des périodes de hausse et de baisse qui se répètent chaque année, ce qui met en évidence un effet saisonnier.

Les ventes tendent à être plus faibles en milieu d'année (été) et plus élevées en début et en fin d'année. La courbe lissée (en rouge) confirme cette tendance générale et permet de mieux visualiser ces cycles.

3.2 Tendances saisonnières

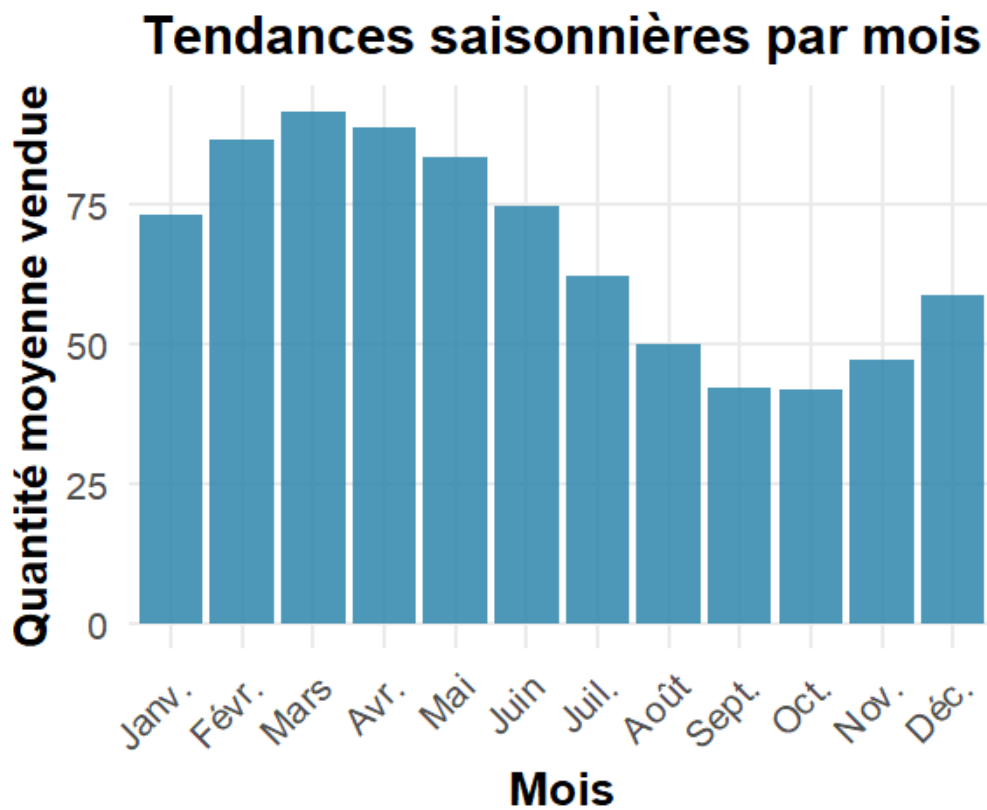


FIGURE 2 – Tendances saisonnières

Le graphique des ventes moyennes par mois confirme l'existence d'une saisonnalité. On observe que les mois de février, mars et avril présentent des niveaux de ventes relativement élevés, tandis que les mois d'août, septembre et octobre sont parmi les plus faibles.

Décembre et janvier montrent également des ventes plus importantes, ce qui peut être lié aux périodes de fêtes et de consommation accrue.

3.3 Effet des promotions

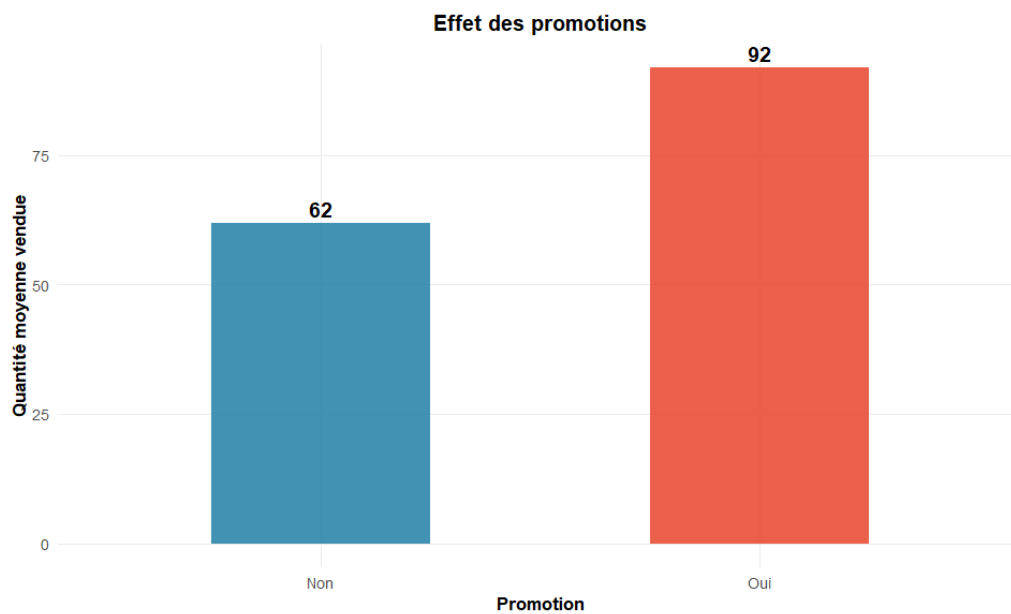


FIGURE 3 – Effet des promotions

Le graphique montre que les ventes moyennes sont nettement plus élevées lorsque les produits sont en promotion. En moyenne, les quantités vendues passent d'environ 62 unités sans promotion à environ 92 unités avec promotion.

3.4 Différence semaine / week-end

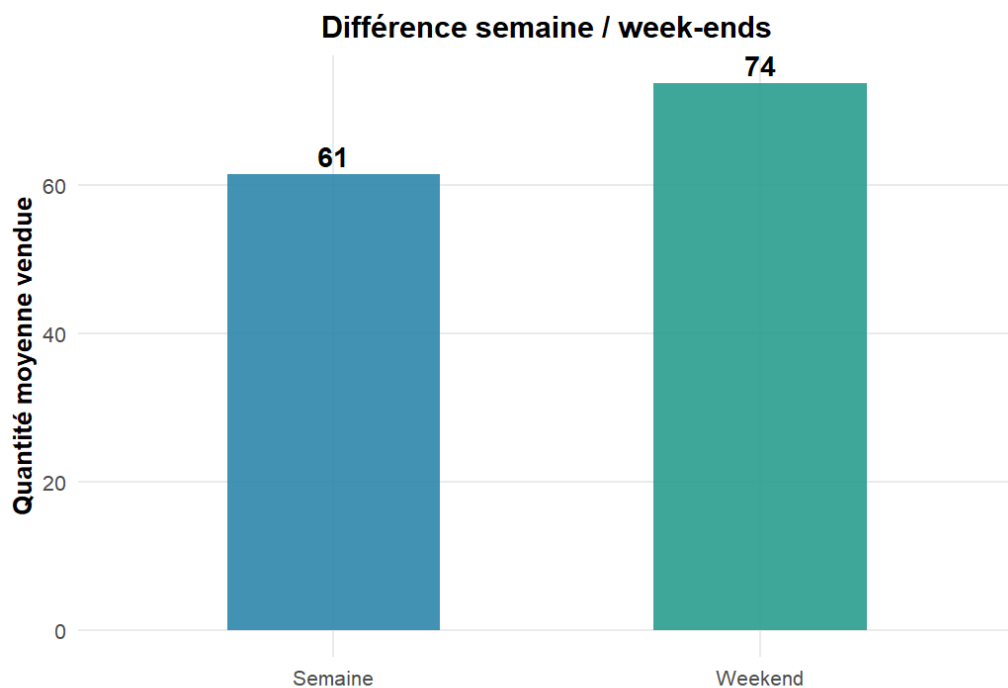


FIGURE 4 – Différence semaine / week-end

Le graphique comparant les ventes moyennes entre la semaine et le week-end montre que les ventes sont plus élevées le week-end (environ 74 unités) que pendant la semaine (environ 61 unités).

3.5 Compétition entre magasins

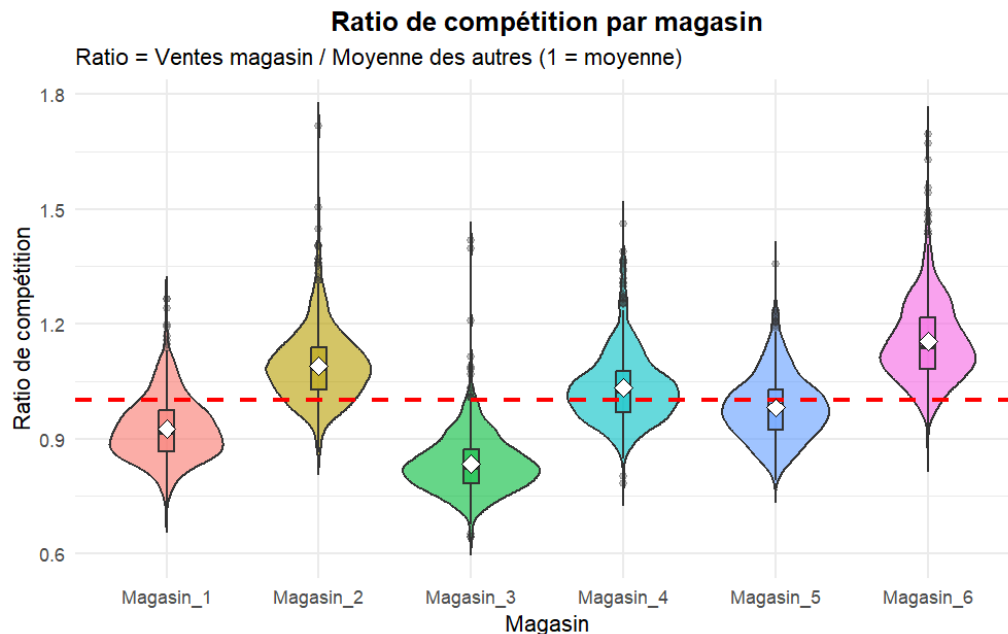


FIGURE 5 – Compétition entre magasins

Le ratio de compétition mesure la performance de chaque magasin par rapport à la moyenne des autres magasins le même jour. Un ratio supérieur à 1 signifie que le magasin vend en moyenne plus que les autres, tandis qu'un ratio inférieur à 1 indique une performance inférieure à la moyenne.

Le graphique montre que les magasins n'ont pas tous le même niveau de performance. Par exemple, Magasin_6 et Magasin_2 ont des ratios en moyenne supérieurs à 1, ce qui signifie qu'ils sont globalement plus performants que les autres. À l'inverse, Magasin_1 et Magasin_3 ont des ratios plutôt inférieurs à 1, ce qui traduit une performance plus faible. On observe également une certaine dispersion des ratios pour chaque magasin, ce qui signifie que leur performance varie selon les jours.

4 Modélisation statistique

4.1 Objectif et choix du modèle

L'objectif de cette partie est d'estimer l'impact de différents facteurs sur la quantité vendue, à savoir : la saisonnalité, les promotions, les jours de la semaine, la compétition entre magasins et les variables financières des magasins.

La variable étudiée, `Quantité_Vendue`, est une variable de comptage (elle prend des valeurs entières positives). Nous avons donc choisi d'utiliser un modèle linéaire généralisé (GLM) de

type Poisson avec lien logarithmique, qui est adapté à ce type de données. Cette démarche progressive permet d'évaluer l'apport de chaque groupe de variables dans l'explication des ventes.

4.2 Construction des modèles

Nous avons construit plusieurs modèles en ajoutant progressivement des variables explicatives afin de comparer leur pertinence. Tous les modèles ont été ajustés sur la base `data_model`, qui regroupe les données de ventes enrichies avec les informations financières et le ratio de compétition.

Les modèles testés sont les suivants :

- **Modèle 1** : uniquement l'effet des promotions ;
- **Modèle 2** : promotions + distinction semaine / week-end ;
- **Modèle 3** : promotions + week-end + saison ;
- **Modèle 4** : promotions + week-end + saison + ratio de compétition ;
- **Modèle 5** : modèle complet incluant promotions, week-end, saison, ratio de compétition, prix unitaire et jour de la semaine ;
- **Modèle 6** : même variables que le modèle 5 avec une interaction entre promotion et week-end ;
- **Modèle 7** : promotions, week-end, saison, ratio de compétition, prix unitaire, et variables financières normalisées (budget publicité et chiffre d'affaires).

4.3 Comparaison des modèles

Plusieurs modèles de régression de Poisson ont été ajustés en ajoutant progressivement des variables explicatives. Pour comparer ces modèles, nous avons utilisé le critère AIC (Akaike Information Criterion). Ce critère permet d'évaluer la qualité d'un modèle tout en pénalisant une complexité trop importante. Plus l'AIC est faible, meilleur est le modèle.

Modèle	AIC	Delta AIC
+ Finances	813 320	0
+ Interaction	821 492	8 172
+ Prix & Jour	821 496	8 176
+ Compétition	821 773	8 453
+ Saison	848 186	34 866
+ Week-end	998 963	185 643
Promotion seule	1 017 860	204 540

TABLE 4 – Comparaison des modèles par le critère AIC

Les résultats montrent que le modèle incluant les variables financières (“+ Finances”) présente l'AIC le plus faible (813 320), avec un Delta AIC égal à zéro. Tous les autres modèles ont des AIC beaucoup plus élevés, ce qui indique qu'ils sont nettement moins performants pour expliquer les données. Cela suggère que l'ajout des variables financières améliore fortement la qualité du modèle.

4.4 Choix du modèle retenu

Sur la base du critère AIC, nous retenons le modèle incluant les variables suivantes :

- Promotion (binaire),
- Distinction semaine / week-end,
- Saison,
- Ratio de compétition entre magasins,
- Prix unitaire,
- Budget publicité normalisé,
- Chiffre d'affaires normalisé,
- Masse salariale,
- Interaction entre promotion et week-end.

Ce modèle est retenu car :

- il obtient le meilleur AIC parmi tous les modèles testés ;
- il intègre l'ensemble des facteurs demandés dans l'énoncé (saisonnalité, promotions, jours de la semaine, compétition et finances) ;
- il permet une interprétation plus riche en combinant variables commerciales et financières.

4.5 Résultats du modèle retenu

Le modèle retenu est un modèle de régression de Poisson avec lien logarithmique, incluant les promotions, le type de jour (semaine ou week-end), la saison, le ratio de compétition, le prix unitaire ainsi que des variables financières (budget publicitaire normalisé, chiffre d'affaires normalisé et masse salariale). Une interaction entre la promotion et le type de jour a également été intégrée.

Le tableau 5 présente les coefficients estimés du modèle retenu.

Variable	Estimate	Std. Error	p-value
Intercept	2.622	0.0098	$< 2 \times 10^{-16}$
Promotion_Binaire	0.3595	0.0022	$< 2 \times 10^{-16}$
Type_Jour (Weekend)	0.1634	0.0014	$< 2 \times 10^{-16}$
Saison (Été)	0.3305	0.0019	$< 2 \times 10^{-16}$
Saison (Hiver)	0.4585	0.0019	$< 2 \times 10^{-16}$
Saison (Printemps)	0.6315	0.0021	$< 2 \times 10^{-16}$
Ratio_Competition	0.6877	0.0044	$< 2 \times 10^{-16}$
Prix_Unitaire	0.00297	0.00018	$< 2 \times 10^{-16}$
Budget_Pub_Norm	0.00126	0.00062	0.0407
Chiffre_Affaires_Norm	0.04997	0.00069	$< 2 \times 10^{-16}$
Masse_Salariale	0.000205	0.0000045	$< 2 \times 10^{-16}$
Promotion \times Weekend	0.01719	0.00375	4.4×10^{-6}

TABLE 5 – Coefficients estimés du modèle de Poisson retenu

4.6 Interprétation générale du modèle

Les coefficients estimés sont tous statistiquement significatifs au seuil de 5 %, ce qui indique que l'ensemble des variables retenues contribue à expliquer la variation de la quantité vendue.

L'effet de la promotion est positif : toutes choses égales par ailleurs, la présence d'une promotion est associée à une augmentation des ventes. De même, les ventes sont en moyenne plus élevées le week-end que pendant la semaine, ce qui traduit un comportement d'achat différent selon le type de jour.

La saisonnalité joue également un rôle important. Par rapport à la saison de référence (automne), les ventes sont plus élevées en été, en hiver et surtout au printemps, ce qui confirme l'existence de variations saisonnières dans la demande.

Le ratio de compétition a un effet positif : lorsqu'un magasin performe mieux que les autres (ratio supérieur à 1), ses ventes tendent à augmenter. Cela suggère que la position relative d'un magasin dans le réseau influence ses performances.

Le prix unitaire a un effet très faible mais positif, ce qui peut refléter le fait que certains produits plus chers se vendent aussi en grande quantité.

Concernant les variables financières, le budget publicitaire normalisé, le chiffre d'affaires normalisé et la masse salariale ont tous un impact positif sur les ventes. Cela indique que des magasins mieux dotés financièrement ont tendance à réaliser de meilleures performances commerciales.

Enfin, l'interaction entre promotion et week-end est positive et significative : l'effet des promotions est plus fort le week-end que pendant la semaine.

4.7 4.7 Graphiques de qualité du modèle

Afin d'évaluer la qualité du modèle retenu, nous complétons les indicateurs numériques (AIC, MAE, corrélation) par une analyse graphique en trois volets : (1) l'erreur moyenne absolue (MAE), (2) le graphique Prédictions vs Observations, et (3) le graphique de l'erreur moyenne par tranche de ventes prédites.

(1) Erreur moyenne absolue (MAE)

L'erreur moyenne absolue (MAE) obtenue est égale à 24,78. Cela signifie qu'en moyenne, l'écart entre la quantité vendue observée et la quantité prédite par le modèle est d'environ 25 unités.

Au regard de la distribution des ventes (médiane ≈ 58 et troisième quartile ≈ 86), cette erreur reste non négligeable : elle représente environ 30 à 40 % des ventes médianes. Cela indique que le modèle capture globalement les tendances, mais avec une précision limitée au niveau individuel.

Ainsi, le modèle est davantage adapté pour analyser les effets des facteurs explicatifs que pour effectuer des prévisions très fines à l'échelle d'une observation donnée.

(2) Graphique Prédictions vs Observations

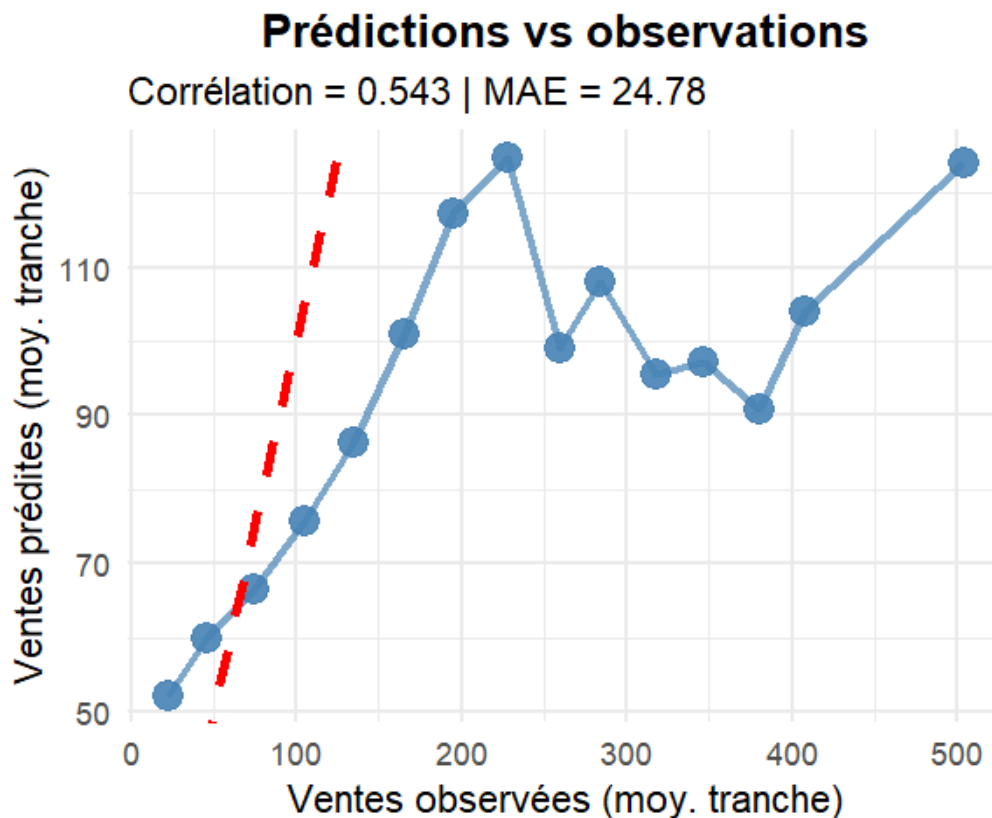


FIGURE 6 – Graphique Prédictions vs Observations

Le graphique *Prédictions vs Observations* met en relation les ventes observées (en abscisse) et les ventes prédites (en ordonnée), après agrégation par tranches.

On observe que :

- Les points suivent globalement une tendance croissante, ce qui montre que le modèle parvient à capter une partie de la relation entre ventes réelles et ventes prédites.
- Toutefois, les points s'écartent sensiblement de la droite idéale ($y = x$), notamment pour les niveaux élevés de ventes. Cela traduit une capacité prédictive imparfaite.
- La corrélation de 0,5431 confirme une relation positive mais seulement modérée entre prédictions et observations.

Ce graphique suggère donc que le modèle est cohérent sur le plan qualitatif (les grandes tendances sont respectées), mais qu'il reste perfectible en termes de précision.

(3) Graphique de l'erreur moyenne par tranche de ventes prédites

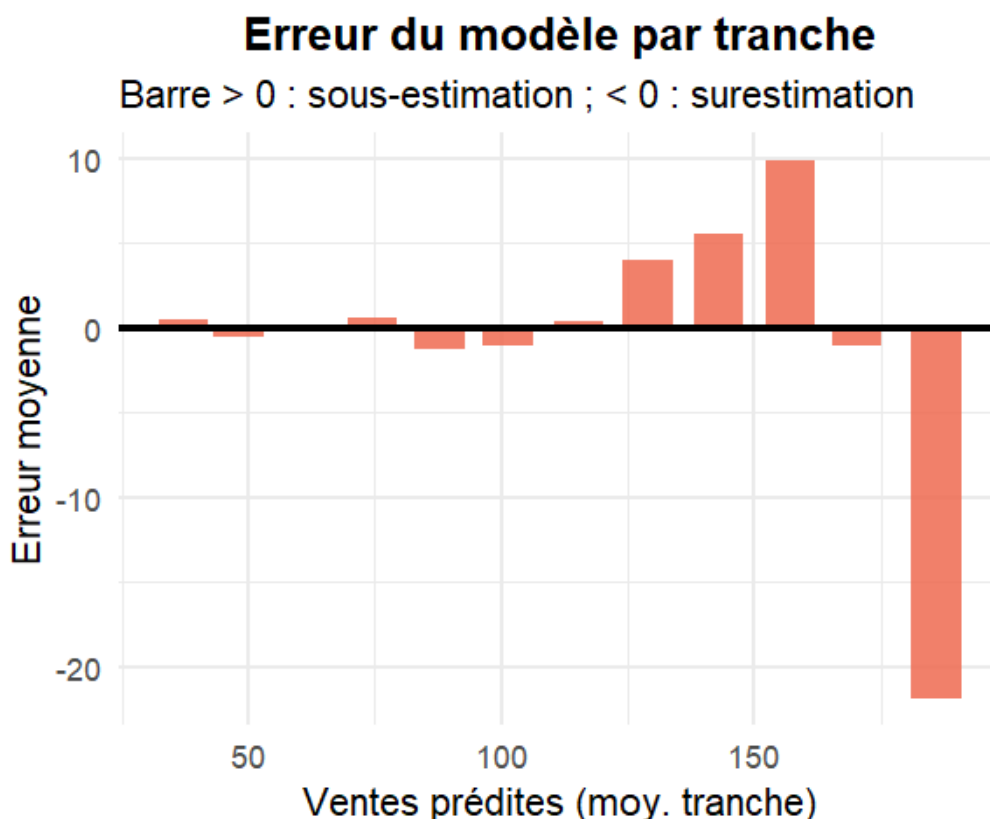


FIGURE 7 – Graphique de l'erreur moyenne par tranche de ventes prédites

Le second graphique représente l'erreur moyenne (résidus) en fonction des tranches de ventes prédites.

L'interprétation est la suivante :

- Pour les niveaux de ventes faibles à intermédiaires, les erreurs moyennes sont proches de zéro, ce qui indique un bon ajustement du modèle dans cette zone.
- Pour les niveaux de ventes les plus élevés, on observe des erreurs négatives marquées, traduisant une **surestimation** systématique des ventes par le modèle.

Cela montre que le modèle a tendance à surévaluer les très fortes ventes, ce qui constitue une limite importante si l'on souhaite l'utiliser à des fins prévisionnelles.

Bilan sur la qualité du modèle

En synthèse :

- Le modèle présente une cohérence globale et permet une interprétation pertinente des facteurs influençant les ventes.
- Sa qualité prédictive reste cependant limitée, notamment pour les niveaux extrêmes de ventes.
- Il est donc adapté pour l'analyse explicative et l'aide à la décision stratégique, mais moins fiable pour des prévisions opérationnelles très précises.

5 Discussion et interprétation des résultats

5.1 Interprétation globale des résultats

L'ensemble des analyses réalisées met en évidence plusieurs facteurs influençant significativement la quantité vendue. Les résultats descriptifs montrent d'abord une forte variabilité des ventes selon les produits, les magasins et la période étudiée, ce qui justifie l'utilisation d'un modèle statistique pour mieux comprendre ces différences.

La modélisation par régression de Poisson confirme que la saisonnalité, les promotions, le type de jour (semaine ou week-end), la compétition entre magasins ainsi que certaines variables financières jouent un rôle important dans l'explication des ventes. Cela indique que la performance commerciale ne dépend pas uniquement des caractéristiques du produit, mais également du contexte temporel et concurrentiel ainsi que des moyens financiers des magasins.

5.2 Différences entre les magasins

L'analyse du ratio de compétition met en évidence des différences notables entre les magasins. Certains magasins, notamment `Magasin_2` et `Magasin_6`, présentent en moyenne un ratio supérieur à 1, ce qui signifie qu'ils performant mieux que la moyenne du réseau. Ces magasins semblent donc plus compétitifs, ce qui peut être lié à une meilleure attractivité, une localisation plus favorable ou une gestion plus efficace des promotions et des stocks.

À l'inverse, `Magasin_1` et `Magasin_3` affichent des ratios plus faibles, traduisant une performance relative inférieure. Cela suggère qu'ils pourraient être davantage impactés par la concurrence locale ou bénéficier de moins d'avantages en termes de fréquentation ou de stratégie commerciale. Ces résultats soulignent l'importance de ne pas appliquer une stratégie uniforme à tous les magasins, mais plutôt d'adapter les actions en fonction de leurs spécificités.

5.3 Effet des promotions et du week-end

Les analyses descriptives et le modèle confirment que les promotions ont un impact positif significatif sur les ventes. En moyenne, les jours avec promotion enregistrent des quantités vendues nettement plus élevées que les jours sans promotion. Cela montre que les promotions sont un levier efficace pour stimuler la demande.

Par ailleurs, les ventes sont généralement plus élevées le week-end que durant la semaine. L'interaction entre promotion et week-end, significative dans le modèle, suggère que l'effet des promotions est encore plus marqué le samedi et le dimanche. Cela peut s'expliquer par une plus grande affluence en magasin et une propension plus élevée des clients à acheter lorsqu'ils ont plus de temps.

5.4 Rôle de la saisonnalité

Les graphiques et le modèle montrent une forte composante saisonnière des ventes. On observe notamment des périodes de creux et de pic au cours de l'année, ce qui est cohérent avec des habitudes de consommation variables selon les saisons. Le modèle confirme que certaines saisons (notamment le printemps et l'hiver) sont associées à des ventes plus élevées, ce qui peut

être lié à des événements saisonniers ou à des variations de la demande pour certains produits alimentaires.

5.5 Impact de la compétition et des finances

Le ratio de compétition apparaît comme un facteur important dans le modèle, indiquant que la performance relative d'un magasin par rapport aux autres influence directement ses ventes. Un magasin qui surperforme par rapport à ses concurrents tend à générer davantage de ventes, ce qui confirme l'importance de la position concurrentielle dans le réseau.

Les variables financières, telles que le chiffre d'affaires normalisé, le budget publicitaire et la masse salariale, ont également un effet significatif. Cela suggère que les magasins disposant de plus de ressources financières ou investissant davantage dans la publicité ont tendance à vendre plus. Toutefois, l'effet du budget publicitaire est relativement faible comparé aux autres variables, ce qui indique que son impact doit être interprété avec prudence.

6 Recommandations

À partir des analyses descriptives et des résultats du modèle statistique, plusieurs leviers d'action peuvent être identifiés afin d'optimiser les ventes au sein du réseau de magasins. Ces recommandations sont structurées autour de quatre axes principaux : les promotions, la gestion du temps (saisonnalité et week-end), la compétitivité entre magasins et l'allocation des ressources financières.

6.1 Optimisation des promotions

Les résultats montrent clairement que les promotions ont un impact positif et significatif sur les ventes. Il est donc recommandé :

- D'augmenter la fréquence des promotions, en particulier pour les produits dont les ventes moyennes sont plus faibles en période normale.
- De cibler prioritairement les promotions sur les périodes de forte affluence, notamment le week-end, car l'interaction entre promotion et week-end est significative et amplifie l'effet des réductions.
- D'adapter les promotions en fonction des produits, car certains produits semblent réagir plus fortement aux actions promotionnelles que d'autres.

6.2 Prise en compte de la saisonnalité

L'analyse temporelle et le modèle mettent en évidence une forte composante saisonnière des ventes. Il est donc conseillé :

- D'anticiper les variations saisonnières en ajustant les stocks en fonction des périodes de forte demande (notamment au printemps et en hiver).
- D'intensifier les campagnes commerciales durant les mois où la demande est naturellement plus élevée.
- D'adapter l'offre et les promotions aux saisons, par exemple en mettant en avant des produits plus consommés à certaines périodes de l'année.

6.3 Amélioration de la compétitivité entre magasins

L'analyse du ratio de compétition montre que certains magasins performant mieux que d'autres. Afin de réduire ces écarts, il est recommandé :

- D'identifier les bonnes pratiques des magasins les plus performants (notamment `Magasin_2` et `Magasin_6`) et de les diffuser aux autres points de vente.
- D'analyser plus finement les causes des performances plus faibles de certains magasins (comme `Magasin_1` et `Magasin_3`) afin de proposer des actions ciblées.
- De renforcer la visibilité et l'attractivité des magasins moins performants, par exemple via des actions locales de communication ou des animations commerciales.

6.4 Allocation des ressources financières

Le modèle montre que le chiffre d'affaires, la masse salariale et, dans une moindre mesure, le budget publicitaire influencent les ventes. Ainsi, il est recommandé :

- D'allouer davantage de ressources aux magasins ayant un fort potentiel de croissance mais des performances encore limitées.
- D'optimiser l'utilisation du budget publicitaire en ciblant les périodes et les magasins où son impact est le plus fort.
- De maintenir un niveau adéquat de personnel dans les magasins les plus fréquentés, notamment le week-end et en période de forte demande.

Conclusion

Ce travail a permis d'analyser en profondeur les ventes de produits alimentaires dans un réseau de magasins sur la période 2022–2024. À travers une démarche structurée, nous avons d'abord procédé au nettoyage et à la préparation des données afin de garantir leur qualité et leur cohérence.

L'analyse descriptive a mis en évidence des tendances saisonnières marquées, un effet positif des promotions sur les ventes et une différence notable entre les jours de semaine et le week-end. Elle a également révélé des écarts de performance entre les magasins, soulignant l'existence d'une dynamique de compétition au sein du réseau.

La modélisation statistique par régression de Poisson a confirmé l'importance de plusieurs facteurs explicatifs, notamment la saisonnalité, les promotions, le type de jour (semaine ou week-end), le ratio de compétition et certaines variables financières. Le modèle retenu, basé sur le critère AIC, offre une interprétation cohérente des déterminants des ventes, même si sa capacité prédictive reste perfectible.

À partir de ces résultats, des recommandations ont été formulées afin d'optimiser les ventes, en particulier en renforçant et en ciblant mieux les promotions, en tenant compte des variations saisonnières, en améliorant la compétitivité des magasins les moins performants et en optimisant l'allocation des ressources financières et humaines.

Ce projet montre ainsi l'intérêt des méthodes statistiques et analytiques pour éclairer la prise de décision dans le domaine commercial. Des travaux futurs pourraient enrichir cette analyse en intégrant d'autres variables, comme la localisation des magasins, la concurrence externe ou le comportement des clients.