

# I. Introduction

## Contexte du projet

L'analyse porte sur un portefeuille de ventes comportant 5 catégories principales (Sport, Électronique, Jouets, Maison, Vêtements) et identifie les tendances de chiffre d'affaires (CA) sur la période allant de janvier 2023 à décembre 2024.

## Objectif business

L'objectif est d'identifier les leviers de croissance et les zones de risques du business afin d'optimiser l'allocation des ressources (stocks, marketing) pour l'année 2025.

# II. Données

## Source

Les données ont toutes été générées aléatoirement et intentionnellement avec des erreurs par intelligence artificielle.

## Description rapide des colonnes

1. Ventes (Numérique) : C'est l'indicateur de performance (KPI). Elle représente généralement le montant monétaire (Chiffre d'Affaires) ou la quantité d'unités vendues pour une transaction donnée.
2. Produits (Catégorique) : C'est l'identifiant de l'objet vendu (nom, SKU ou code barre). Elle permet de distinguer ce qui a été acheté.
3. Dates (Temporelle) : C'est le marqueur chronologique. Elle indique le moment précis (jour, mois, année, voire heure) où la transaction a eu lieu.
4. Catégories (Catégorique de haut niveau) : C'est le groupe auquel appartient le produit (ex: "Électronique", "Vêtements"). Elle sert à organiser les produits en familles pour une analyse globale.

# III. Nettoyage

## Problèmes rencontrés

1. Données manquantes: Plus de 3 % de données manquantes par colonne. Cela fausse les calculs statistiques et empêche les analyses globales.
2. Types de données incorrects: C'est le problème le plus critique.

Ventes : Doit avoir des valeurs numériques pour permettre des sommes ou moyennes.

Dates : Doit être au format datetime pour extraire les mois, années ou saisonnalités.

3. Doublons (213 lignes) : Ils gonflent artificiellement le chiffre d'affaires et faussent la réalité du business.

4. Inccohérence potentielle : Il peut souvent y avoir des erreurs de saisie (ex: "100€" au lieu de "100", ou des formats de date mixtes).

### Actions effectuées

1. Suppression des doublons: On ne peut pas prendre de décisions sur des données répétées par erreur technique
2. Conversion des dates: On ne peut pas tracer de graphique temporel si la colonne Dates est considérée comme du texte
3. Transformation de Ventes en valeurs numériques, en gérant les symboles monétaires si nécessaire: On ne peut pas additionner du texte
4. Traitement des valeurs manquantes : Supprimer 300 lignes sur 10 200 (environ 3 %) est acceptable si cela garantit la précision. Cependant, imputer (remplacer) par la médiane ou le mode permet de conserver le volume de données sans trop biaiser les statistiques

Ventes : Remplacer par la médiane (plus robuste que la moyenne).

Catégories/Produits : Utiliser la valeur la plus fréquente (mode).

Dates : Suppression de la ligne car si la date est manquante, la ligne est souvent inutilisable pour une analyse temporelle.

5. Normalisation : Remplacer chaque première lettre d'un Produit et d'une Catégorie en majuscule pour éviter que "lego" et "Lego" soient comptés séparément par exemple. Cela garantit que les regroupements par produit ou catégorie soient exacts et ne créent pas de catégories fantômes à cause d'une majuscule.

## IV. KPI & Analyses

### Chiffre d'affaires

1. La catégorie "Jouets": Portée par un "Produit Star", c'est l'exemple le plus frappant de dépendance. CA Catégorie : 455 731 €. Produit Star (Lego) : 240 502 €. Analyse : Le produit Lego représente à lui seul 53 % du chiffre d'affaires de sa catégorie (en supposant que Poupée, Puzzle et Voiture rc complètent le reste). Risque : S'il y a un problème d'approvisionnement sur les Lego, c'est la moitié de la santé de ce département qui s'effondre.
2. La catégorie "Sport": Une catégorie homogène. Bien qu'elle soit la première catégorie en volume (608 792 €), elle est beaucoup plus équilibrée. Composants : Tapis de yoga (119k), Ballon (115k), Raquette (110k), Haltères (106k)... Analyse : Aucun produit ne domine outrageusement les autres. Chaque référence contribue de manière équivalente (environ 18-20 % chacune). Avantage : C'est une catégorie robuste. La perte d'un produit ne met pas en péril l'ensemble de la catégorie.
3. La catégorie "Électronique": Une tête de liste marquée. CA Catégorie : 458 998 € Leaders : Ordinateur (122k) et Casque (122k) sont au coude-à-coude. Analyse : Ces deux produits portent plus de 53 % de la catégorie, tandis que le Smartphone (107k) et la Montre connectée (119k) complètent bien l'offre. C'est un mix sain, tiré par un duo de tête.

4. La catégorie "Vêtements" : Très homogène. Composants : Veste (120k), Robe (116k), Jean (108k), T-shirt (107k). Analyse : L'écart entre le meilleur produit (Veste) et le moins bon (T-shirt) est faible (environ 13k€). Les ventes sont très bien réparties.
5. La catégorie "Maison" : Un ensemble très équilibré. CA Total de la catégorie : 454 444 € Produits identifiés : Coussin (118k), Chaise (116k), Table (111k), Lampe (109k). Répartition : Comme pour la catégorie "Vêtements", les revenus sont extrêmement bien répartis. Le produit leader (Coussin) ne représente que 26 % du chiffre d'affaires total de la catégorie. Stabilité : L'écart entre le produit le plus vendu (Coussin) et le moins vendu (Lampe) n'est que de 9 %. Cela indique une gamme cohérente où chaque meuble ou accessoire trouve son public de manière équivalente.

### Top produits

Produit Star (Lego) : Le produit "Lego" est le moteur principal de la croissance, générant près du double de CA que ses concurrents directs (environ 240 000 €). Il attire la majorité des revenus et prouve une forte demande.

Plateau pour les autres produits : Les quatre autres produits ("Ordinateur", "Casque", "Veste", "Voiture rc") ont des niveaux de ventes presque identiques, se situant autour de 120 000 € de CA.

Risque de dépendance : L'entreprise est extrêmement dépendante du produit "Lego". Si ce produit venait à manquer en stock, ou si un concurrent proposait une alternative, la moitié des revenus générés par ces cinq produits serait immédiatement menacée.

### Evolution temporelle

Saisonnalité marquée : L'activité est cyclique, avec des fluctuations régulières. Des pics de ventes très nets apparaissent, notamment autour des périodes de rentrée (août-septembre) et en fin d'année (octobre-novembre 2023), suivis de creux significatifs (avril-mai).

Croissance de la performance : Le niveau général du chiffre d'affaires en 2024 est visiblement plus élevé que celui de 2023. Par exemple, le creux d'avril 2024 est plus haut que le pic de mai 2023.

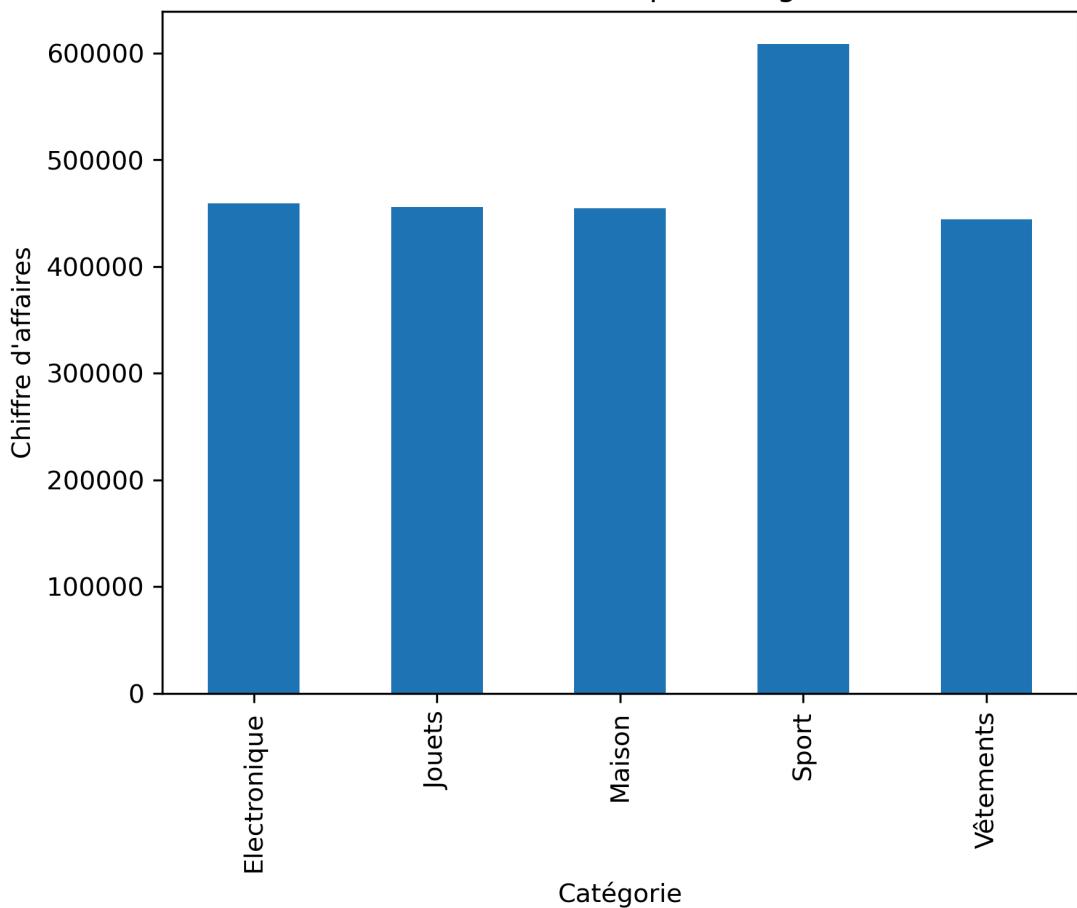
Pic historique : Le record de ventes a été atteint en janvier 2024 (plus de 110 000 €).

Tendance récente préoccupante : On observe une forte chute de l'activité entre août et décembre 2024, le dernier point étant le plus bas de toute la période observée.

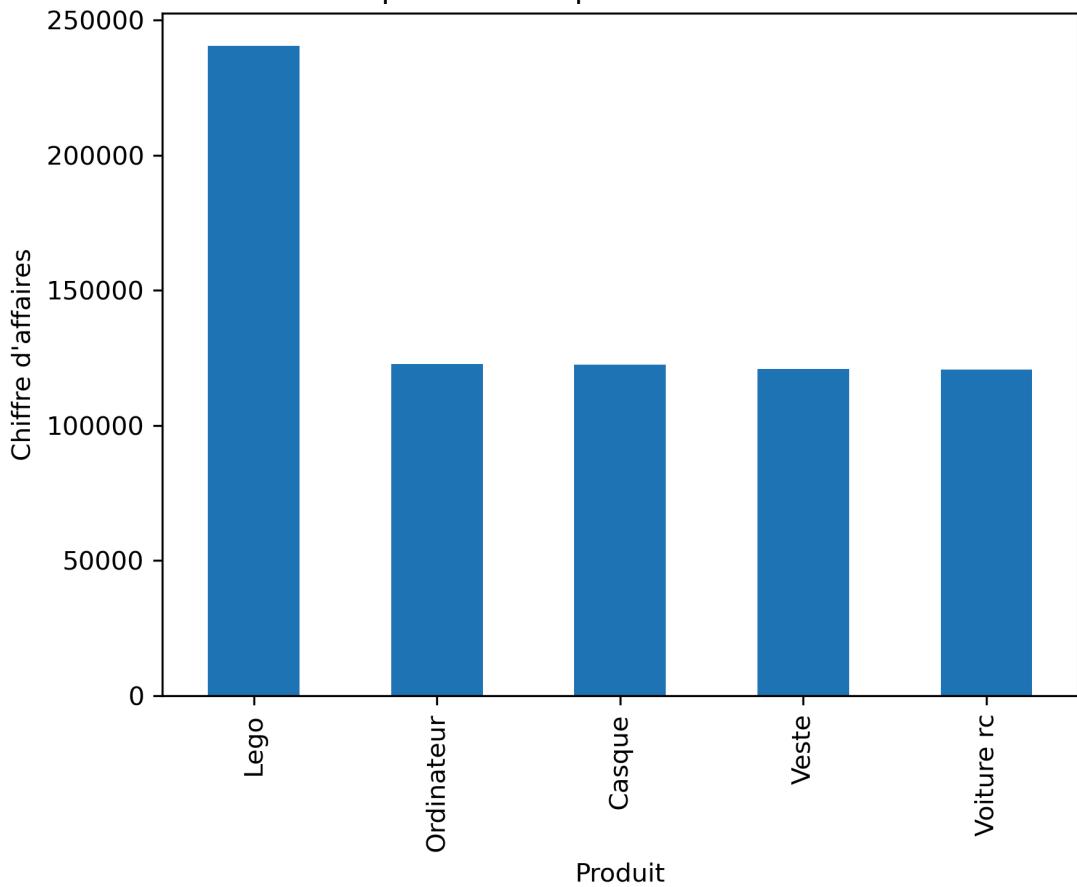
## V. Visualisations

### Graphiques clés

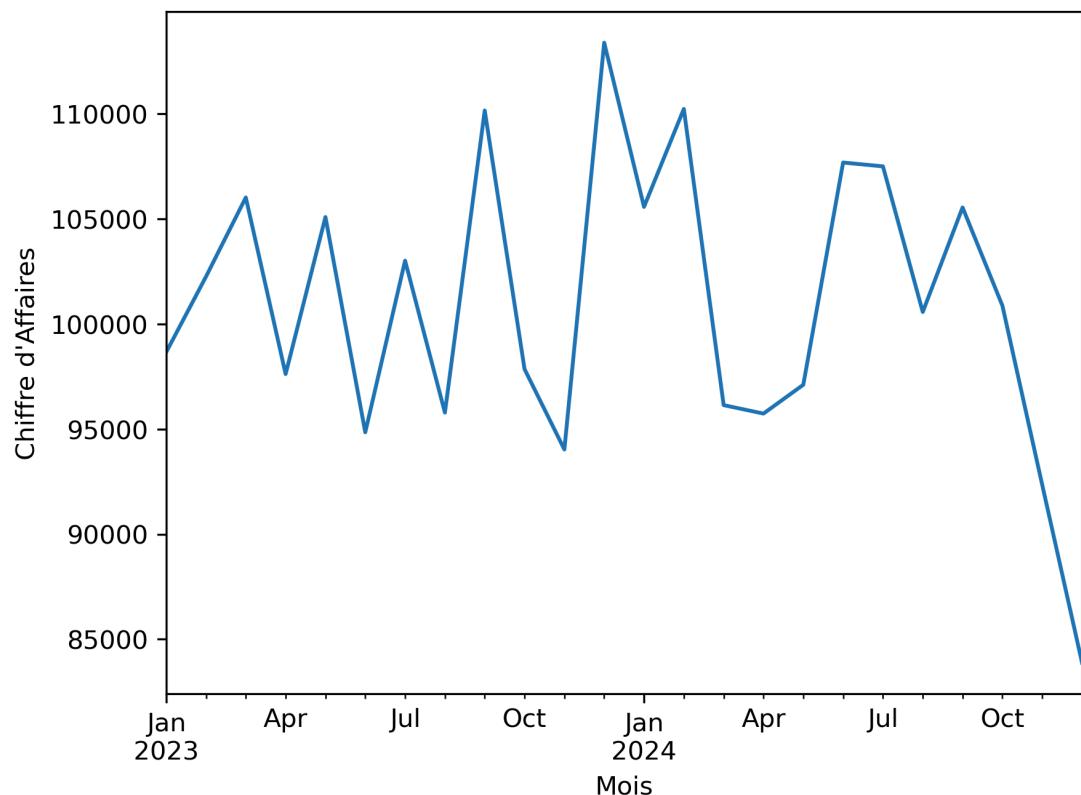
Chiffre d'affaires par catégorie



Top 5 Produits par Chiffre d'Affaires



## Evolution mensuelle du chiffre d'affaires



# VI. Insights & Recommandations

## Résumé exécutif

Le secteur Sport est le moteur principal de l'entreprise : C'est la catégorie la plus performante, générant 30% de revenus de plus que la moyenne des autres catégories. Une croissance annuelle positive, mais vulnérable : L'entreprise a vu son CA augmenter entre 2023 et 2024, mais une chute brutale et inexplicable a eu lieu en décembre 2024 (chute de 28% par rapport à 2023).

Risque de concentration majeur sur un produit unique : Le produit "Lego" génère à lui seul plus de la moitié du CA de la catégorie "Jouets". Si ce produit est indisponible, la catégorie s'effondre.

Saisonnalité identifiable : Des pics de vente sont récurrents en début d'année (janvier-février) et en été (juillet-août), avec des creux au printemps (avril-mai).

## Actions proposées

1. Sécuriser le produit "Lego" immédiatement : Négocier des garanties de stock avec le fournisseur pour éviter une nouvelle rupture et garantir 53 % des revenus de la catégorie "Jouets".
2. Booster les catégories "Maison" et "Électronique" : Ces secteurs sont stables et homogènes. Des investissements marketing ciblés pourraient les éléver au niveau de la catégorie "Sport" et ainsi diversifier les sources de revenus.
3. Investiguer et corriger la chute de Décembre 2024 : Trouver la cause exacte de la contre-performance de fin d'année 2024. L'objectif 2025 est de revenir au niveau de 2023 (104k€ de CA) sur cette période cruciale.