

Analyser la performance des ventes sur différents produits et clients de 2020 à 2025

Caleb Durandisse

2025-09-01

Problématique

Comment améliorer les performances commerciales en identifiant les produits, catégories ou segments de clients les plus rentables, ainsi que les zones géographiques et périodes favorables pour maximiser les ventes et les bénéfices ?”

Les variables indépendantes de notre dataset sont : Id_commande, catégorie, sous_categorie, quantité, mode_paiement, date_commande, nom_client, ville, etat, annee_mois.

Les variables dépendantes sont : montant, profit.

• **Problème métier** : Identifier les leviers de croissance pour augmenter les ventes et les bénéfices.

source : Kaggle

• **Objectifs** :

- o Identifier les sous_catégories de produit les plus rentables par ville.
- o Analyser la performance commerciale par ville, période et segment de clients.
- o Quels sont les clients les plus rentables ?

Faites une analyse statistique et des visualisations :

- o Analyse des distributions (bénéfices, ventes).
- o Corrélations entre variables (chiffre d'affaires, bénéfices, catégories).
- o Visualisation des tendances temporelles avec des graphiques (ggplot2).

Modélisation et prévision

- o Identifier les sous_catégories homogène
- o Identifier les groupes de clients les plus rentables

- Présentez un rapport détaillé avec R Markdown :

o Insights clés, graphiques, et recommandations.

Télécharger les librairies

```
library(dplyr)
library(tidyverse)
library(ggplot2)
library(lubridate)
library(tidyr)
library(forecast)
library(cluster)
library(factoextra)
library(caret)
library(scales)
library(treemap)
library(tseries)
library(randomForest)
library(lubridate)
library(xgboost)
```

1. Importer le dataframe sales dans R

```
sales <- read.csv("C:/Users/duran/Downloads/Ensemble des données de
vente/Sales Dataset.csv", stringsAsFactors = FALSE)
```

Exploration des données : nous utilisons les fonctions comme head(), summary(), et str() pour la compréhension de la structure des données

```
head(sales)
```

```
##   ID_commande montant Profit Quantite  Categorie  Sous_Categorie
## 1    B-26776   9726   1275         5 Electronics Electronic Games
## 2    B-26776   9726   1275         5 Electronics Electronic Games
## 3    B-26776   9726   1275         5 Electronics Electronic Games
## 4    B-26776   4975   1330        14 Electronics      Printers
## 5    B-26776   4975   1330        14 Electronics      Printers
## 6    B-26776   4975   1330        14 Electronics      Printers
## mode_paiement Date_commande  nom_client  etat  ville annee_mois
## 1          UPI    2023-06-27 David Padilla Florida  Miami    2023-06
## 2          UPI    2024-12-27 Connor Morgan Illinois Chicago  2024-12
## 3          UPI    2021-07-25 Robert Stone New York  Buffalo  2021-07
## 4          UPI    2023-06-27 David Padilla Florida  Miami    2023-06
## 5          UPI    2024-12-27 Connor Morgan Illinois Chicago  2024-12
## 6          UPI    2021-07-25 Robert Stone New York  Buffalo  2021-07
```

```
summary(sales)
```

```
##   ID_commande      montant      Profit      Quantite
## Length:1194      Min.   : 508      Min.   : 50      Min.   : 1.00
## Class :character 1st Qu.:2799 1st Qu.: 410 1st Qu.: 6.00
```

```

## Mode :character      Median :5152      Median :1014      Median :11.00
##                      Mean :5178       Mean :1349       Mean :10.67
##                      3rd Qu.:7626     3rd Qu.:2035     3rd Qu.:16.00
##                      Max. :9992       Max. :4930       Max. :20.00
##      Categorie      Sous_Categorie      mode_paiement      Date_commande
## Length:1194         Length:1194         Length:1194         Length:1194
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##
##
##
##      nom_client      etat      ville      annee_mois
## Length:1194         Length:1194         Length:1194         Length:1194
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##
##
##
str(sales)

## 'data.frame':      1194 obs. of  12 variables:
## $ ID_commande      : chr  "B-26776" "B-26776" "B-26776" "B-26776" ...
## $ montant          : int  9726 9726 9726 4975 4975 4975 1525 1525 883 8127
## ...
## $ Profit           : int  1275 1275 1275 1330 1330 1330 185 185 117 3551 ...
## $ Quantite         : int   5 5 5 14 14 14 12 12 10 16 ...
## $ Categorie        : chr  "Electronics" "Electronics" "Electronics"
## "Electronics" ...
## $ Sous_Categorie: chr  "Electronic Games" "Electronic Games" "Electronic
## Games" "Printers" ...
## $ mode_paiement    : chr  "UPI" "UPI" "UPI" "UPI" ...
## $ Date_commande    : chr  "2023-06-27" "2024-12-27" "2021-07-25" "2023-06-
## 27" ...
## $ nom_client       : chr  "David Padilla" "Connor Morgan" "Robert Stone"
## "David Padilla" ...
## $ etat             : chr  "Florida" "Illinois" "New York" "Florida" ...
## $ ville            : chr  "Miami" "Chicago" "Buffalo" "Miami" ...
## $ annee_mois        : chr  "2023-06" "2024-12" "2021-07" "2023-06" ...

```

Nous disposons d'une base de données qui contient 12 variables et 1194 observations, réparties en 3 numériques, 7 qualitatives et 2 datées.

Le chiffre d'affaires le plus faible est de 508 €, alors que le plus important atteint 9992 €. En moyennel, le chiffre d'affaires relatif aux produits tourne autour de 5178 €. D'un côté, 50 % des produits génèrent un chiffre d'affaires inférieur à 5152 €, tandis que l'autre 50 % enregistrent des chiffres d'affaires qui dépassent ce montant.

Le profit le plus bas s'élève à 50 €, tandis que le plus élevé atteint 4472 €. Le bénéfice moyen s'établit à 1347 €. La moitié des articles génèrent un bénéfice inférieur à 1014 €, tandis que l'autre moitié en produit un bénéfice supérieur à 1014 €.

2. Prétraitement des variables `date_commande` et `annee_mois` en format date

```
sales$annee_mois <- as.Date(paste0(sales$annee_mois, "-01"))
sales$Date_commande <- as.Date(sales$Date_commande)

str(sales)

## 'data.frame':    1194 obs. of  12 variables:
##  $ ID_commande   : chr  "B-26776" "B-26776" "B-26776" "B-26776" ...
##  $ montant       : int   9726 9726 9726 4975 4975 4975 1525 1525 883 8127
##  ...
##  $ Profit        : int   1275 1275 1275 1330 1330 1330 185 185 117 3551 ...
##  $ Quantite      : int     5  5  5 14 14 14 12 12 10 16 ...
##  $ Categorie     : chr   "Electronics" "Electronics" "Electronics"
##  "Electronics" ...
##  $ Sous_Categorie: chr   "Electronic Games" "Electronic Games" "Electronic
##  Games" "Printers" ...
##  $ mode_paiement : chr   "UPI" "UPI" "UPI" "UPI" ...
##  $ Date_commande : Date, format: "2023-06-27" "2024-12-27" ...
##  $ nom_client    : chr   "David Padilla" "Connor Morgan" "Robert Stone"
##  "David Padilla" ...
##  $ etat          : chr   "Florida" "Illinois" "New York" "Florida" ...
##  $ ville         : chr   "Miami" "Chicago" "Buffalo" "Miami" ...
##  $ annee_mois    : Date, format: "2023-06-01" "2024-12-01" ...
```

3. Identifier s'il y a des valeurs manquantes dans le dataframe avec la fonction `colSums()`

```
colSums(is.na(sales))

##      ID_commande      montant      Profit      Quantite      Categorie
##           0           0           0           0           0
## Sous_Categorie mode_paiement Date_commande nom_client      etat
##           0           0           0           0           0
##      ville      annee_mois
##           0           0
```

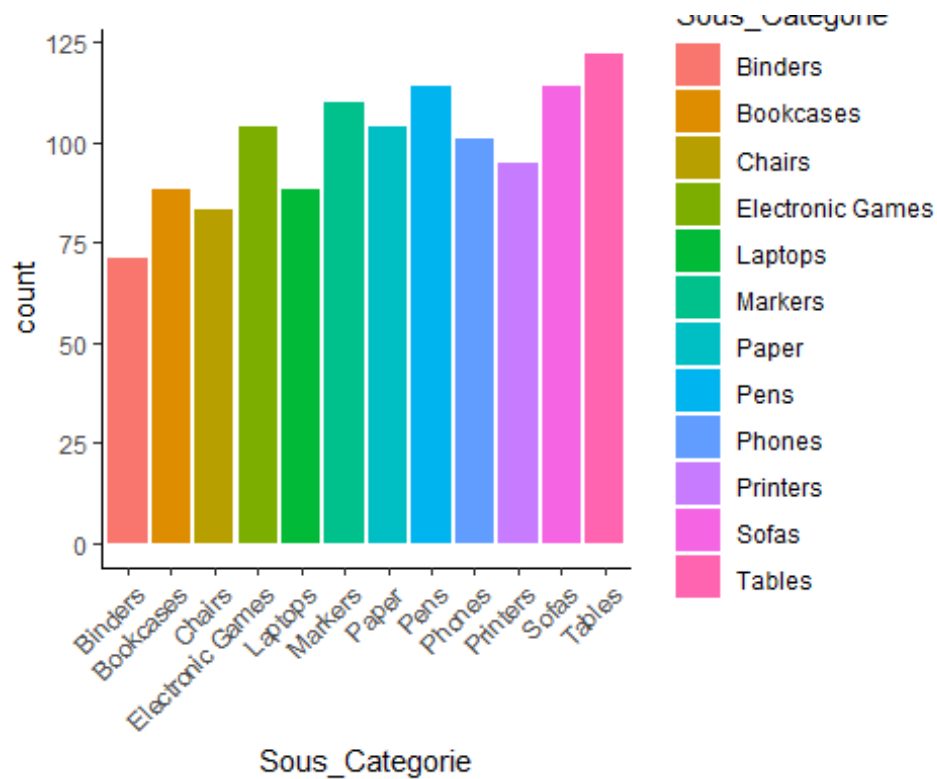
4. Vérifier s'il y a des doublons avec la fonction `"distinct()"`

```
sales <- sales %>% distinct()
```

5. Vérifier la distribution de la variable `Sous_Categorie`

```
sales %>%
  ggplot() +
  aes(x = Sous_Categorie, fill = Sous_Categorie) +
  geom_bar() +
```

```
theme_classic() +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
scale_y_continuous(labels = scales::comma)
```



Ce diagramme révèle des disparités significatives qui pourraient orienter des décisions stratégiques en matière de gestion des stocks, de marketing ou de ventes.

Synthèse Générale Le graphique montre que les sous-catégories “Tables”, “Sofas” et “Pens” sont les plus représentées, avec des comptes supérieurs à 110. À l’opposé, la catégorie “Binders” (Classeurs) est la moins fréquente, avec un compte d’environ 70. La majorité des catégories se situe dans une fourchette de 80 à 110, indiquant une distribution relativement concentrée

Observations Clés

Top 3 des Sous-Catégories (les plus fréquentes) :

Tables : La catégorie avec le plus grand nombre d’occurrences (environ 125).

Sofas : Se situe juste derrière les tables, indiquant une très forte présence.

Pens : Également très élevée, suggérant un volume important.

Bas du Classement (les moins fréquentes) :

Binders (Classeurs) : La catégorie la moins représentée.

Electronic Games (Jeux Électroniques) : Le deuxième compte le plus bas.

Bookcases (Bibliothèques) : Le troisième compte le plus bas.

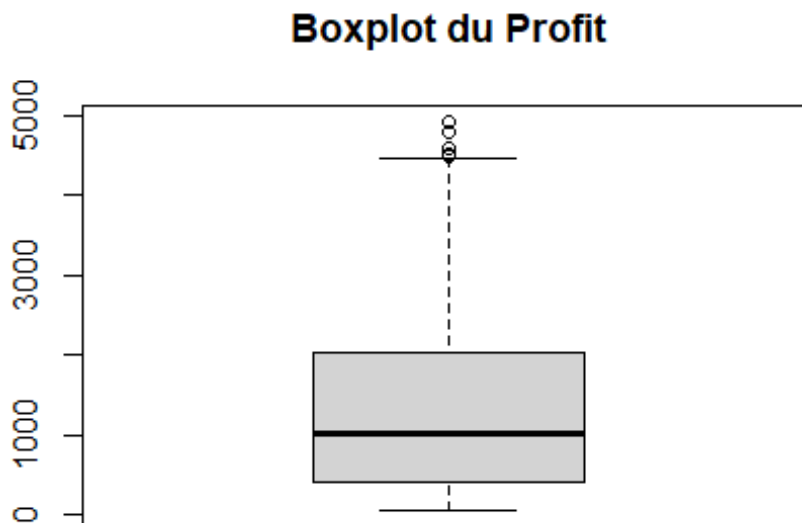
Regroupements d'Intérêt :

Mobilier de bureau et maison ("Tables", "Sofas", "Chairs", "Bookcases") : Ce groupe présente de forts contrastes. Alors que les tables et les canapés dominent, les chaises sont dans la moyenne et les bibliothèques sont en bas de classement.

Fournitures de bureau ("Pens", "Markers", "Paper", "Binders") : Ce segment est également polarisé. Les stylos et marqueurs sont très fréquents, tandis que les classeurs sont la catégorie la plus faible de tout le graphique.

6. distribution de la variable Profit par la visualisation boxplot

```
boxplot(na.omit(sales$Profit), main = "Boxplot du Profit")
```



La distribution du profit est fortement asymétrique vers la droite (asymétrie positive). Cela signifie que la majorité des transactions génère un profit relativement modeste, tandis qu'un petit nombre de transactions génère des profits exceptionnellement élevés. La valeur médiane du profit se situe autour de 1000.

Tendance Centrale (Médiane) :

La ligne épaisse à l'intérieur de la boîte représente la médiane. Elle est située à environ 1000. Cela signifie que 50% des transactions ont généré un profit inférieur à 1000, et 50% un profit supérieur.

Dispersion (Écart Interquartile - IQR) :

La boîte représente l'écart interquartile, qui contient 50% des données centrales.

Le bas de la boîte (1er quartile, Q1) est à environ 500.

Le haut de la boîte (3ème quartile, Q3) est à environ 2000.

Ainsi, la moitié des profits se situe dans une fourchette assez large, entre 500 et 2000.

Asymétrie :

La médiane (1000) est plus proche du bas de la boîte (Q1) que du haut (Q3).

La "moustache" supérieure, qui s'étend jusqu'à environ 4500, est beaucoup plus longue que la moustache inférieure (proche de 0).

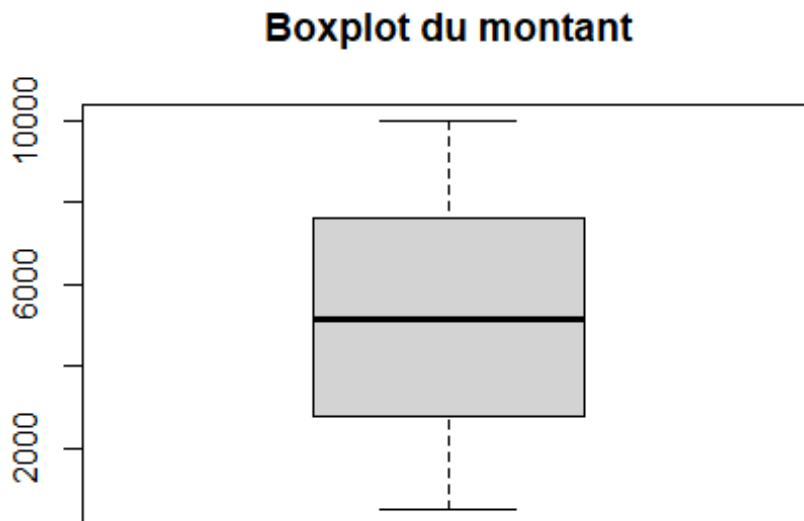
Ces deux éléments confirment une distribution asymétrique à droite.

Valeurs Aberrantes (Outliers) :

Les points situés au-dessus de la moustache supérieure (entre 4500 et 5000) sont des valeurs aberrantes. Ce sont des profits significativement plus élevés que le reste des données.

7. distribution de la variable montant par la visualisation boxplot

```
boxplot(na.omit(sales$montant), main = "Boxplot du montant")
```



La distribution des montants est globalement symétrique et s'étend sur une large plage de valeurs, indiquant une forte variabilité. Contrairement au graphique du profit, celui-ci ne présente aucune valeur aberrante (outlier), ce qui suggère que les montants extrêmes s'inscrivent dans la continuité de la distribution principale.

Tendance Centrale (Médiane) :

La médiane (la ligne épaisse) se situe à environ 5200. Cela signifie que 50% des transactions ont un montant inférieur à 5200 et 50% ont un montant supérieur.

Dispersion (Écart Interquartile - IQR) :

La boîte, qui représente le cœur de la distribution (50% des données), est très étendue :

Le 1er quartile (Q1) est à environ 3000.

Le 3ème quartile (Q3) est proche de 8000 (environ 7800).

Cet écart interquartile de près de 4800 (7800 - 3000) témoigne d'une dispersion très élevée. Les montants des transactions "standard" varient considérablement.

Symétrie et Plage de Valeurs :

La médiane est positionnée presque au centre de la boîte, et les "moustaches" (les lignes s'étendant de la boîte) ont des longueurs comparables. Cela indique une distribution relativement symétrique.

La plage totale des données s'étend d'un peu moins de 1000 jusqu'à un maximum de 10 000.

8. Calculer les KPIs globaux soient le CA_total, le profit_total, la marge_moyenne, la quantité_totale

```
kpi_sales <- sales %>%
  summarise(
    CA_total = sum(montant),
    Profit_total = sum(Profit),
    Marge_moyenne = mean(Profit / montant, na.rm = TRUE) * 100,
    Quantite_totale = sum(Quantite)
  )
print(kpi_sales)

##   CA_total Profit_total Marge_moyenne Quantite_totale
## 1  6182639    1610697    26.05709         12745
```

Synthèse des Indicateurs

Chiffre d'Affaires total (CA_total) : 6 182 639 €

Profit total (Profit_total) : 1 610 697 €

Marge de profit globale (Marge_moyenne) : 26.06 %

Quantité totale d'articles vendus (Quantite_totale) : 12 745

La marge de profit a été calculée sur les totaux (Profit_total / CA_total), confirmant une rentabilité globale solide de 26.06% sur l'ensemble des ventes de la période.

En résumé, l'entreprise dispose d'une base solide avec des transactions de grande valeur et une bonne rentabilité.

9. Calculer le CA, profit, quantité et ticket moyen par ville

```
kpi_ville <- sales %>%
  group_by(ville) %>%
  summarise(
    CA = sum(montant, na.rm = TRUE),
    profit = sum(Profit, na.rm = TRUE),
    taux_profit = mean(Profit / montant, na.rm = TRUE) * 100,
    quantite_totale = sum(Quantite),
    nb_commandes = n_distinct(ID_commande),
    ticket_moyen = CA / nb_commandes) %>%

    arrange(desc(profit))

print(kpi_ville)
```

```
## # A tibble: 18 × 7
##   ville          CA profit taux_profit quantite_totale nb_commandes
##   <chr>      <int> <int>      <dbl>          <int>          <int>
##   <dbl>
## 1 Orlando    452158 128125      28.4            820            45
## 10048.
## 2 Buffalo    418514 111823      26.9            977            57
## 7342.
## 3 Miami      370881 109739      28.9            785            45
## 8242.
## 4 Rochester  407291 109729      26.1            836            45
## 9051.
## 5 San Fran... 440000 108106      25.5            953            51
## 8627.
## 6 Dallas     390144 102483      25.3            762            49
## 7962.
## 7 San Diego  389104 101444      26.0            753            48
## 8106.
## 8 New York... 304243  86954      28.9            666            48
## 6338.
## 9 Springfi... 370131  84205      24.0            682            46
## 8046.
## 10 Austin    335199  82638      25.0            659            45
## 7449.
## 11 Cleveland 334220  81454      23.9            812            46
```

7266.						
## 12 Chicago	344675	80061	23.8	627	39	
8838.						
## 13 Cincinna...	303856	78788	26.2	645	39	
7791.						
## 14 Peoria	263932	76106	28.8	511	39	
6767.						
## 15 Houston	286132	72659	24.9	520	34	
8416.						
## 16 Tampa	268135	70842	26.3	540	41	
6540.						
## 17 Los Ange...	257332	69264	27.2	613	41	
6276.						
## 18 Columbus	246692	56277	22.4	584	35	
7048.						

Orlando est le leader incontesté, dominant le classement en termes de Chiffre d’Affaires (CA), de Profit et, surtout, de Ticket Moyen. À l’opposé, Columbus et Los Angeles apparaissent comme les marchés les moins performants sur presque tous les plans. Entre ces deux extrêmes, des villes comme New York et Peoria présentent des profils atypiques et intéressants.

Les Champions (Top 3)

Orlando :

#1 en CA (452k€) et #1 en Profit (128k€).

#1 en Ticket Moyen (10 048€), loin devant toutes les autres villes. C’est le facteur clé de son succès.

Une rentabilité très solide (28.4%).

San Francisco :

#1 en Nombre de Commandes (51), indiquant une forte activité commerciale.

#2 en CA (440k€) et dans le top 5 en Profit (108k€).

Le ticket moyen (8627€) est élevé, ce qui en fait un marché très performant.

Buffalo :

Un excellent challenger avec le #2 en Profit (111k€) et le #3 en CA (418k€).

Il combine un bon volume de commandes avec une rentabilité solide.

Les Cas Intéressants & Opportunités

New York : Le profil “Haute Fréquence, Faible Panier”

Force : Une des villes les plus rentables avec un taux de profit de 28.9% (#1 ex aequo) et un grand nombre de commandes (48).

Faiblesse : Un ticket moyen très bas (6338€), l'un des plus faibles du classement.

Opportunité : Si New York parvenait à augmenter son ticket moyen pour se rapprocher de la moyenne, son CA et son profit pourraient exploser. C'est une cible prioritaire pour des stratégies de "up-selling" (montée en gamme) ou de "cross-selling" (ventes croisées).

10. Calculer le CA, profit, quantité et ticket moyen par client

```
kpi_client <- sales %>%
  group_by(nom_client) %>%
  summarise(
    profit = sum(Profit),
    CA = sum(montant),
    nb_commandes = n_distinct(ID_commande),
    frequence_achat = n(),
    panier_moyen = sum(montant) / n_distinct(ID_commande)
  ) %>%

  arrange(desc(profit))

print(kpi_client)
```

A tibble: 802 × 6

##	nom_client	profit	CA	nb_commandes	frequence_achat
##	panier_moyen				
##	<chr>	<int>	<int>	<int>	<int>
##	<dbl>				
##	1 Logan Galloway	8840	21116	1	3
##	21116				
##	2 Stephanie Manning	8840	21116	1	3
##	21116				
##	3 Cory Evans	7790	28557	1	4
##	28557				
##	4 Mr. Ralph Garcia Jr.	7275	22589	1	3
##	22589				
##	5 Maria Thomas	7198	15280	1	2
##	15280				
##	6 Nicholas Johnson	7198	15280	1	2
##	15280				
##	7 Kelly Smith	7166	23333	1	4
##	23333				
##	8 Sabrina Hartman	7166	23333	1	4
##	23333				
##	9 Brent Hernandez	7116	19611	1	2
##	19611				
##	10 Patrick Williams	7116	19611	1	2

```
19611
```

```
## # i 792 more rows
```

1. Rentabilité forte chez des clients peu fréquents Tous les clients n'ont passé qu'une seule commande (nb_commandes = 1), ce qui montre que même des clients uniques peuvent générer un bon profit.

Pourtant, la frequence_achat (nombre de lignes/transactions) varie entre 2 et 4, ce qui suggère qu'ils ont commandé plusieurs articles en une seule commande.

2. Le panier_moyen est égal au CA Cela est logique puisque chaque client n'a qu'une seule commande, donc le CA total = panier moyen.

Ça confirme aussi que ces clients ont eu une grosse commande unique, ce qui peut fausser des analyses classiques de "bons clients" : ce sont des "gros coups ponctuels".

3. Valeur client élevée avec une faible fréquence Ces clients sont probablement des clients à haute valeur mais non fidèles.

Exemple : Cory Evans a un CA de 28 557 € avec un profit de 7 790 €, soit un taux de profit environ 27%. Très rentable.

11. Calculer le CA, profit, quantité et ticket moyen par sous_categorie

```
kpi_sous_cat <- sales %>%
  group_by(Sous_Categorie) %>%
  summarise(
    profit = sum(Profit),
    CA = sum(montant),
    nb_commandes = n_distinct(ID_commande),
    frequence_achat = n(),
    marge = mean(Profit / montant, na.rm = TRUE) * 100,
    panier_moyen = CA / nb_commandes
  ) %>%
  arrange(desc(profit))

print(kpi_sous_cat)

## # A tibble: 12 × 7
##   Sous_Categorie  profit      CA nb_commandes frequence_achat marge
##   <chr>          <int> <int>         <int>         <int> <dbl>
##   <dbl>
## 1 Markers      174749 627875           69           110  27.1
## 2 Tables      156796 625177           70           122  24.8
## 3 Paper       149723 524755           67           104  27.4
## 4 Electronic Gam... 148454 565092           70           104  27.3
```

```

8073.
## 5 Printers      146259 566359      59      95 25.6
9599.
## 6 Sofas        142854 568367      75     114 26.1
7578.
## 7 Pens         129846 552269      84     114 23.6
6575.
## 8 Chairs       122892 431964      52      83 28.0
8307
## 9 Bookcases    118000 413165      61      88 27.2
6773.
## 10 Phones      113607 503055      64     101 24.2
7860.
## 11 Laptops     110260 419950      60      88 27.5
6999.
## 12 Binders     97257 384611      49      71 24.5
7849.

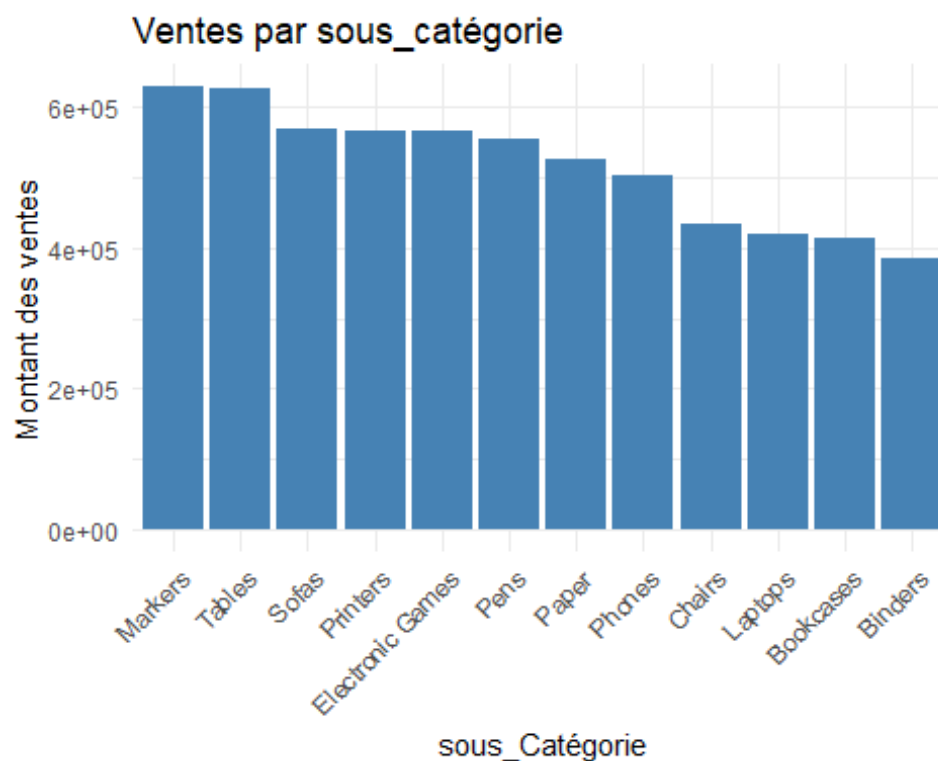
```

#11.1 visualiser Les Ventes par sous_catégorie

```

ggplot(kpi_sous_cat, aes(x = reorder(Sous_Categorie, -CA), y = CA)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Ventes par sous_catégorie", x = "sous_Catégorie", y =
"Montant des ventes") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



Revenus et Rentabilité Bien que les Marqueurs et les Tables soient les plus performants en termes de revenus absolus (CA) et de profit, ils ne sont pas les plus rentables en pourcentage.

Champions de la rentabilité : Les catégories avec les marges de profit (marge) les plus élevées sont les Chaises (28,0 %), les Ordinateurs portables (27,5 %) et le Papier (27,4 %). Ces produits convertissent très efficacement les revenus en profit.

Moteur de volume à faible marge : Les Stylos se distinguent par la marge de profit la plus faible (23,6 %), malgré un volume de ventes élevé et le plus grand nombre de commandes individuelles.

Plus grands bénéficiaires : Les Marqueurs génèrent le profit absolu le plus élevé (174 749 €), suivis de près par les Tables (156 796 €).

Comportement d'achat des clients Les données révèlent deux modèles d'achat distincts : les ventes à haute fréquence et faible valeur contre les ventes à haute valeur et moins fréquentes.

Haute fréquence, faible valeur : Les Stylos enregistrent le plus de commandes (84) mais ont l'une des valeurs moyennes par commande (panier_moyen) les plus basses, à 6 575 €. Cela les identifie comme un produit à fort volume générant du trafic.

Articles à forte valeur : Les Imprimantes ont la valeur moyenne par commande la plus élevée (9 599 €), ce qui indique qu'il s'agit soit d'articles coûteux, soit qu'ils sont généralement achetés en grande quantité par commande.

Résumé stratégique Étoiles (Marqueurs, Tables) : Ils constituent le cœur de l'activité, générant des revenus et des profits élevés. Ils doivent être prioritaires dans les stratégies de marketing et la gestion des stocks.

Pépites cachées (Chaises, Ordinateurs portables) : Avec les marges les plus élevées, ces catégories représentent une opportunité de croissance majeure. Augmenter leur volume de ventes aurait un impact positif significatif sur la rentabilité globale.

Piliers (Stylos) : Cette catégorie est excellente pour attirer des clients mais n'est pas un moteur de profit principal. Les stratégies pourraient inclure leur vente en lots (bundling) avec des articles à plus forte marge ou de légers ajustements de prix.

Spécialistes (Imprimantes) : Leur valeur de commande moyenne élevée les rend idéales pour des campagnes ciblées visant les clients professionnels à forte valeur.

12. visualiser les combinaisons ville et sous_catégorie les plus rentables

```
profit_par_combinaison <- sales %>%
  group_by(Sous_Categorie, ville) %>%
  summarise(
    profit_moyen = mean(Profit, na.rm = TRUE),
    profit_total = sum(Profit, na.rm = TRUE),
```

```

    n_transactions = n()
  ) %>%
  arrange(desc(profit_total)) %>%
  ungroup()

## `summarise()` has grouped output by 'Sous_Categorie'. You can override
## using
## the `groups` argument.

print(profit_par_combinaison)

## # A tibble: 215 × 5
##   Sous_Categorie ville      profit_moyen profit_total n_transactions
##   <chr>          <chr>          <dbl>         <int>         <int>
## 1 Printers      Orlando          2044.         20443          10
## 2 Bookcases     San Francisco    1732          19052          11
## 3 Markers       Orlando          2216.         17732           8
## 4 Tables       San Diego        1773.         17731          10
## 5 Paper        Orlando          1930.         17373           9
## 6 Electronic Games Austin          1447.         17367          12
## 7 Tables       Orlando          1578.         17361          11
## 8 Sofas        Buffalo          1702.         15322           9
## 9 Paper        Miami            2189.         15320           7
## 10 Electronic Games Miami          1884.         15076           8
## # i 205 more rows

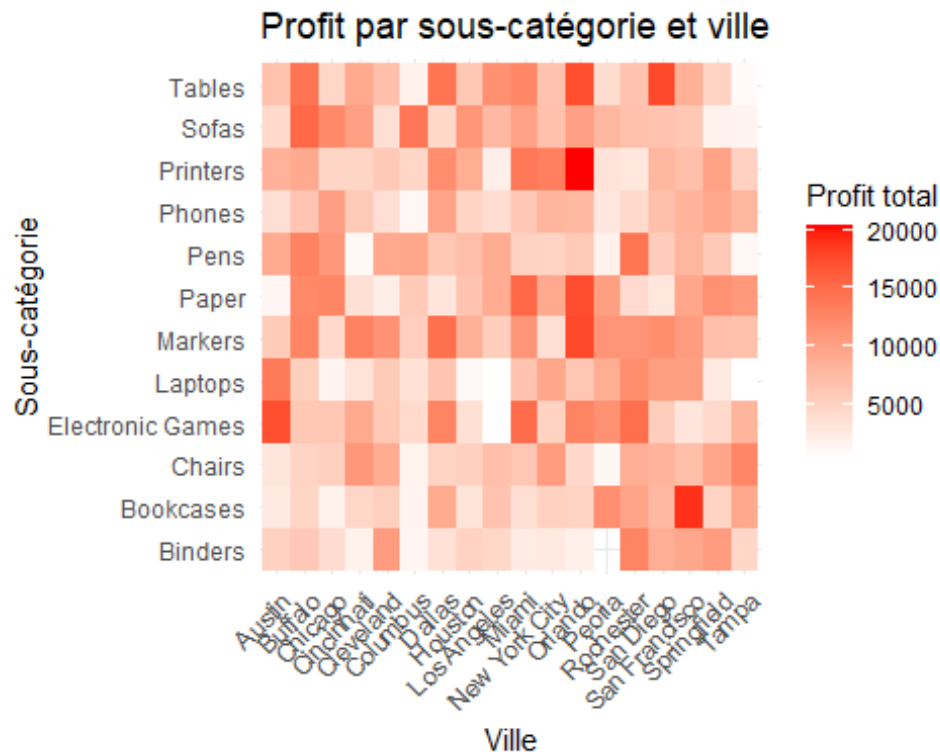
```

#12.1 Visualiser la combinaison entre ville et sous-catégories des produits

```

ggplot(profit_par_combinaison, aes(x = ville, y = Sous_Categorie, fill =
profit_total)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "red") +
  labs(
    title = "Profit par sous-catégorie et ville",
    x = "Ville",
    y = "Sous-catégorie",
    fill = "Profit total"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



Performances exceptionnelles à Orlando :

Les Printers à Orlando montrent le profit total le plus élevé (20,443) avec un profit moyen par transaction de 2,044.

Orlando apparaît également performante pour les Markers et le Paper.

San Francisco et les Bookcases :

La combinaison Bookcases à San Francisco se distingue avec un profit total de 19,052.

Performances modérées à Austin et Miami :

Les Electronic Games fonctionnent bien à Austin et Miami.

Le paper à Miami a un profit moyen très élevé (2,189) mais un volume de transactions plus faible (7).

Buffalo et les canapés :

Une performance notable pour les Sofas à Buffalo avec un profit total de 15,322.

Recommandations stratégiques

Optimisation de l'assortiment :

Concentrer les stocks d'imprimantes et de marqueurs à Orlando.

Développer l'offre de bibliothèques à San Francisco.

13. Etudier la tendance du CA et du profit par année

```
sales_par_annee <- sales %>%
  mutate(annee = year(Date_commande)) %>%
  group_by(annee) %>%
  summarise(
    Total_CA = sum(montant, na.rm = TRUE),
    Total_Profit = sum(Profit, na.rm = TRUE),
    totale_vendue = sum(Quantite),
    .groups = 'drop'
  )

print(sales_par_annee)

## # A tibble: 6 × 4
##   annee Total_CA Total_Profit totale_vendue
##   <dbl>   <int>      <int>      <int>
## 1  2020   859401    224103      1695
## 2  2021  1181446    283231      2358
## 3  2022  1459775    393113      3234
## 4  2023  1229723    321671      2497
## 5  2024  1202478    308336      2523
## 6  2025   249816     80243       438

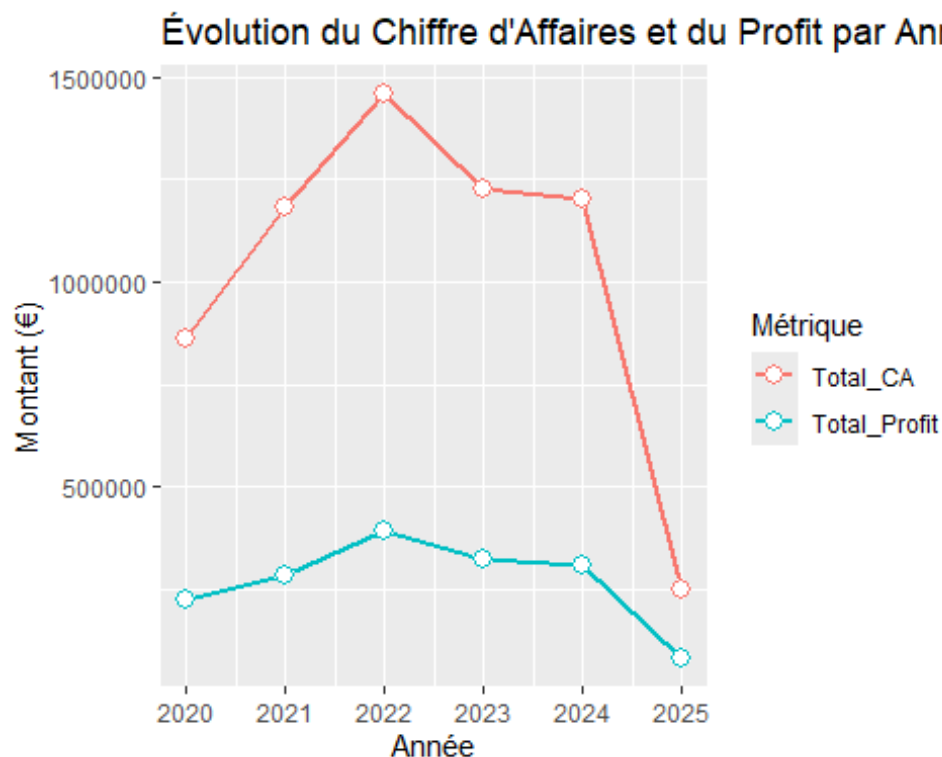
# 13.1 Visualiser l'évolution du CA et du profit par an

sales_par_annee_long <- sales_par_annee %>%
  pivot_longer(
    cols = c(Total_CA, Total_Profit),
    names_to = "Metrique",
    values_to = "Valeur"
  )

ggplot(sales_par_annee_long, aes(x = annee, y = Valeur, color = Metrique,
group = Metrique)) +
  geom_line(size = 1) +
  geom_point(size = 3, shape = 21, fill = "white", stroke = 1) +
  labs(
    title = "Évolution du Chiffre d'Affaires et du Profit par Année",
    x = "Année",
    y = "Montant (€)",
    color = "Métrique"
  )

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Il y a eu une progression notable du bénéfice en 2022, suivie d'un léger recul du bénéfice en 2023 et 2024. Pour 2025, nous ne disposons que de données pour le premier trimestre, ce qui explique la chute significative enregistrée pour cette année.

CA passe de 900k € en 2020 à un pic de près de 1.5 million € en 2022

Profit suit une courbe parallèle, atteignant son maximum (400k €) en 2022

Cette période peut être décrite comme haute croissance, signe :

d'une bonne performance commerciale, d'une gestion saine des coûts (car le profit suit le CA).

Le CA baisse légèrement mais reste élevé (1.25 million €)

Le profit diminue aussi, mais plus modérément

Cela peut refléter : Une saturation du marché ou une hausse des coûts, Une pression sur les marges, Un manque d'innovation ou d'expansion

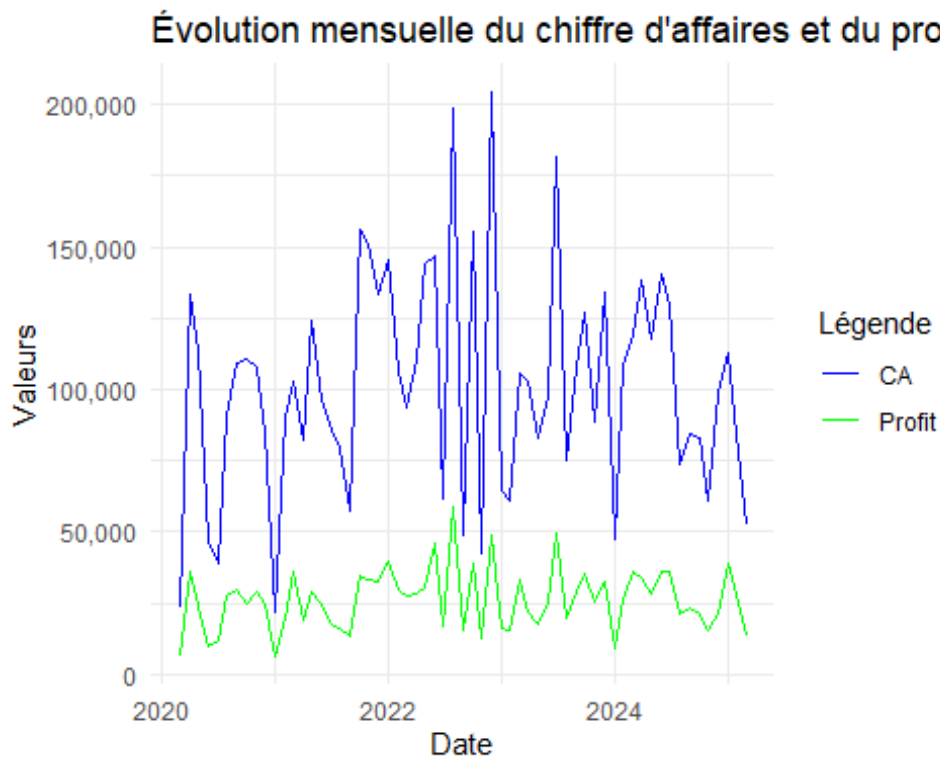
14. Visualisation CA, profit mensuel et le taux de croissance

```
sales %>%
  group_by(annee_mois) %>%
  summarise(
```

```

    CA = sum(montant),
    profit = sum(Profit)
  ) %>%
  ggplot(aes(x = annee_mois)) +
  geom_line(aes(y = CA, color = "CA")) +
  geom_line(aes(y = profit, color = "Profit")) +
  scale_y_continuous(labels = scales::comma) +
  scale_color_manual(values = c("CA" = "blue", "Profit" = "green")) +
  labs(
    title = "Évolution mensuelle du chiffre d'affaires et du profit",
    x = "Date",
    y = "Valeurs",
    color = "Légende"
  ) +
  theme_minimal()

```



```

#Calcul du tx de croissance mensuel
sales_mensuel <- sales %>%
  group_by(annee_mois) %>%
  summarise(
    CA = sum(montant),
    profit = sum(Profit)
  ) %>%
  arrange(annee_mois)
print(sales_mensuel)

```

```
## # A tibble: 61 × 3
##   annee_mois      CA profit
##   <date>      <int> <int>
## 1 2020-03-01   22991   6192
## 2 2020-04-01  133385  36156
## 3 2020-05-01  113287  24294
## 4 2020-06-01   46900   9489
## 5 2020-07-01   38556  12008
## 6 2020-08-01   91117  27824
## 7 2020-09-01  109434  29876
## 8 2020-10-01  110836  24776
## 9 2020-11-01  108170  28979
## 10 2020-12-01   84725  24509
## # i 51 more rows
```

Visualiser le taux du CA et du profit par mois

```
sales_mensuel <- sales_mensuel %>%
  mutate(
    ventes_mois = (CA - lag(CA)) / lag(CA) * 100,
    profit_mois = (profit - lag(profit)) / lag(profit) * 100
  )
print(sales_mensuel)
```

```
## # A tibble: 61 × 5
##   annee_mois      CA profit ventes_mois profit_mois
##   <date>      <int> <int>      <dbl>      <dbl>
## 1 2020-03-01   22991   6192         NA         NA
## 2 2020-04-01  133385  36156       480.       484.
## 3 2020-05-01  113287  24294      -15.1      -32.8
## 4 2020-06-01   46900   9489      -58.6      -60.9
## 5 2020-07-01   38556  12008      -17.8       26.5
## 6 2020-08-01   91117  27824       136.       132.
## 7 2020-09-01  109434  29876        20.1        7.37
## 8 2020-10-01  110836  24776         1.28      -17.1
## 9 2020-11-01  108170  28979        -2.41       17.0
## 10 2020-12-01   84725  24509       -21.7      -15.4
## # i 51 more rows
```

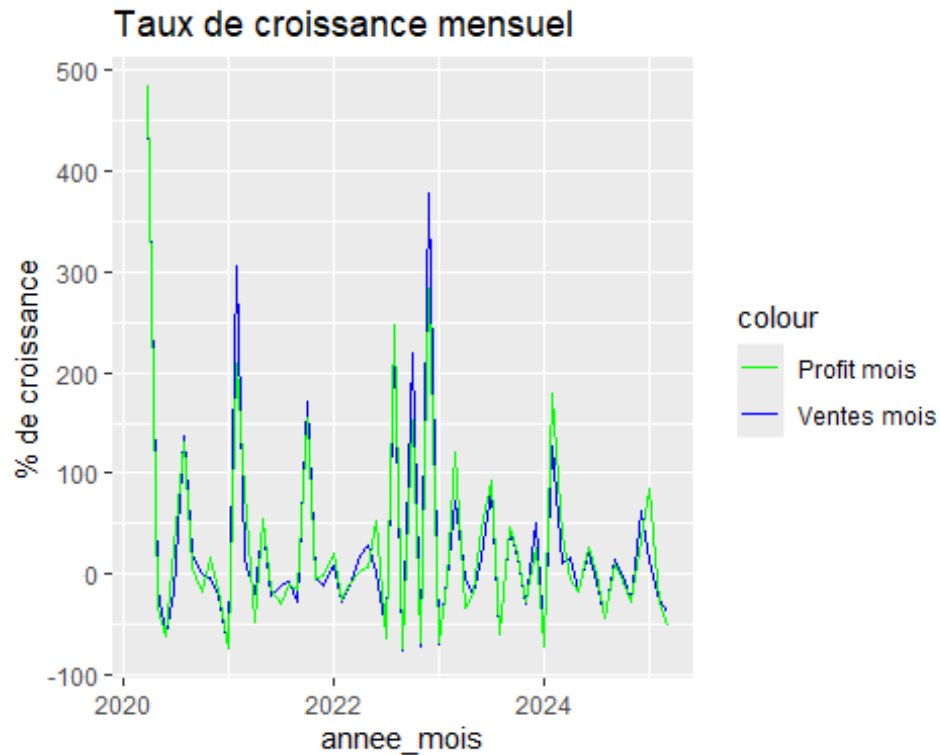
```
ggplot(sales_mensuel, aes(x = annee_mois)) +
  geom_line(aes(y = ventes_mois, color = "Ventes mois")) +
  geom_line(aes(y = profit_mois, color = "Profit mois")) +
  labs(title = "Taux de croissance mensuel", y = "% de croissance") +
  scale_color_manual(values = c("Ventes mois" = "blue", "Profit mois" =
"green"))
```

```
## Warning: Removed 1 row containing missing values or values outside the
scale range
```

```
## (`geom_line()`).
```

```
## Removed 1 row containing missing values or values outside the scale range
```

```
## (`geom_line()`).
```



L'analyse de ces graphiques révèle que l'entreprise connaît une croissance très volatile et instable, caractérisée par des fluctuations mensuelles extrêmes plutôt que par une progression stable et prévisible.

Analyse des Graphiques

1. Une forte volatilité des revenus et des profits Le premier graphique ("Évolution mensuelle") montre que le chiffre d'affaires (CA en bleu) et le profit (en vert) évoluent en "dents de scie". Il n'y a pas de tendance de fond claire à la hausse, mais plutôt une succession de pics et de creux importants d'un mois à l'autre. Cette instabilité rend la prévision financière difficile.
2. Les taux de croissance Le deuxième graphique ("Taux de croissance mensuel") est le plus révélateur. Il confirme l'extrême instabilité de l'activité.

Des pics extrêmes : Les taux de croissance atteignent des sommets (+200%, +300%) et des creux vertigineux (-50%, -80%).

L'effet "yo-yo" : Un pic de croissance massif (ex: +200%) fait souvent suite à un mois de forte baisse. Cela signifie que l'entreprise ne connaît pas une expansion saine, mais plutôt une succession de "rattrapages" après des périodes difficiles. Une forte croissance positive n'est donc pas nécessairement un signe de bonne santé, mais peut simplement indiquer une reprise après un effondrement le mois précédent.

3. Le profit est plus sensible que les ventes (levier d'exploitation) En observant attentivement les deux graphiques, on constate que les variations de profit (en vert) sont souvent plus fortes que celles du chiffre d'affaires (en bleu).

Exemple: En mai 2020, une baisse du CA de 15,1% a entraîné une chute du profit de 32,8%.

Explication : C'est l'effet du levier d'exploitation. Les coûts fixes de l'entreprise (salaires, loyers, etc.) ne baissent pas lorsque les ventes diminuent. Par conséquent, toute baisse de revenu a un impact amplifié sur le profit. Inversement, une forte hausse des ventes peut faire exploser le profit, car les coûts fixes sont déjà couverts.

14.1 Visualiser les mois pics et creux

```
# Agréger les ventes par mois
df_mensuel <- sales %>%
  group_by(annee_mois) %>%
  summarise(
    ventes = sum(montant),
    profit = sum(Profit)
  ) %>%
  arrange(annee_mois)

# Mois avec le chiffre d'affaires le plus élevé
mois_pic <- df_mensuel %>%
  filter(ventes == max(ventes))

# Mois avec le chiffre d'affaires le plus bas
mois_creux <- df_mensuel %>%
  filter(ventes == min(ventes))

print("Mois de pic de ventes :")
## [1] "Mois de pic de ventes :"

print(mois_pic)

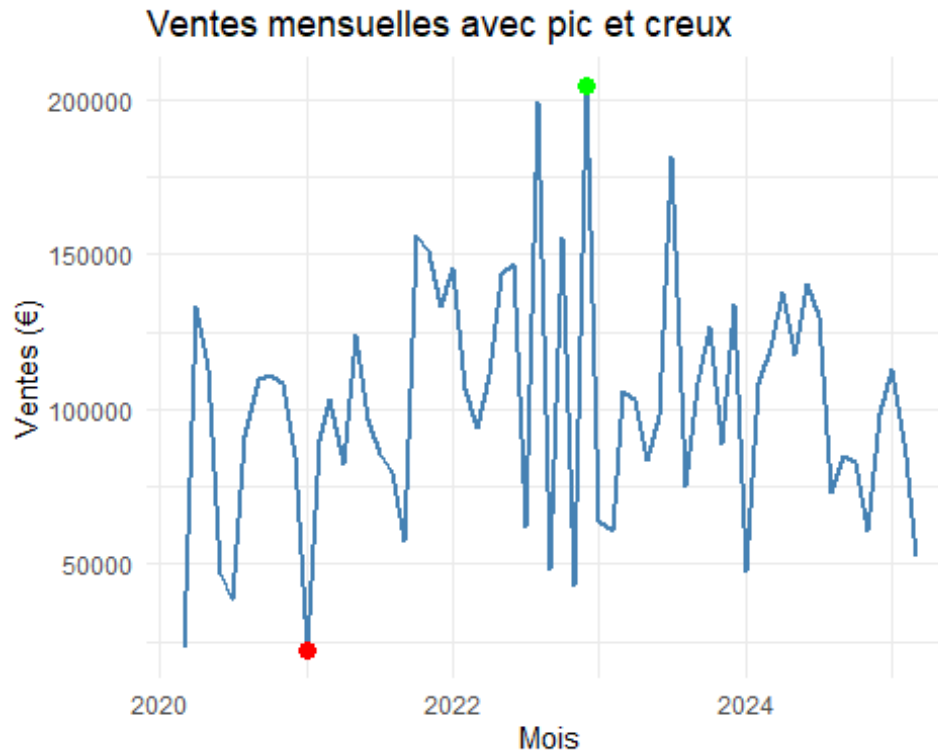
## # A tibble: 1 × 3
##   annee_mois ventes profit
##   <date>      <int> <int>
## 1 2022-12-01 204413  48915

print("Mois de creux de ventes :")
## [1] "Mois de creux de ventes :"

print(mois_creux)

## # A tibble: 1 × 3
##   annee_mois ventes profit
##   <date>      <int> <int>
## 1 2021-01-01  22187   6449
```

```
ggplot(df_mensuel, aes(x = annee_mois, y = ventes)) +
  geom_line(color = "steelblue", size = 1) +
  geom_point(data = mois_pic, aes(x = annee_mois, y = ventes), color =
"green", size = 3) +
  geom_point(data = mois_creux, aes(x = annee_mois, y = ventes), color =
"red", size = 3) +
  labs(title = "Ventes mensuelles avec pic et creux",
x = "Mois", y = "Ventes (€)") +
  theme_minimal()
```



```
# Top 3 mois de ventes les plus élevées
top_pics <- df_mensuel %>%
  arrange(desc(ventes)) %>%
  head(3)

# Bottom 3 mois
bottom_creux <- df_mensuel %>%
  arrange(ventes) %>%
  head(3)

print("Top 3 pics de ventes:")
## [1] "Top 3 pics de ventes:"
print(top_pics)
```

```
## # A tibble: 3 × 3
##   annee_mois ventes profit
##   <date>      <int> <int>
## 1 2022-12-01 204413  48915
## 2 2022-08-01 198965  58918
## 3 2023-07-01 181554  49391

print("Top 3 creux de ventes:")

## [1] "Top 3 creux de ventes:"

print(bottom_creux)

## # A tibble: 3 × 3
##   annee_mois ventes profit
##   <date>      <int> <int>
## 1 2021-01-01  22187   6449
## 2 2020-03-01  22991   6192
## 3 2020-07-01  38556  12008
```

Observations :

Deux pics majeurs en 2022, ce qui peut indiquer une année exceptionnelle (campagnes, produits populaires, marchés ouverts post-COVID ?).

Décembre en tête, ce qui est typique si l'activité est B2C (effet de fin d'année).

Les profits sont également élevés sur ces mois, donc ce sont des mois rentables, pas juste du volume.

Janvier est systématiquement faible → après les achats de fin d'année, les ventes chutent : effet classique de saisonnalité.

Moins d'activité aussi en juillet 2020, possiblement à cause de la pandémie ou creux estival (vacances).

Profits bas corrélés au CA bas = il ne s'agit pas de ventes à faible marge, mais bien d'un creux de demande.

Le graphe montre :

Des fluctuations nettes entre les pics et les creux (de 20K € à +200K € de CA)

Une saisonnalité visuelle, notamment autour de été/fin d'année

Un pic très net en Décembre 2022, suggérant une anomalie positive à étudier

15. Tests statistiques

15.1 Etudier la relation entre le profit et les villes

```
anova_villes <- aov(Profit ~ ville, data = sales)
summary(anova_villes)

##              Df      Sum Sq Mean Sq F value Pr(>F)
## ville         17 2.873e+07 1689950   1.359  0.148
## Residuals    1176 1.462e+09 1243546
```

Hypothèses :

H_0 : Les moyennes des profits sont égales entre les villes.

H_1 : Au moins une ville a une moyenne de profit différente.

Statistiques clés :

F value = 1.368 : Rapport entre la variance expliquée par le facteur (ville) et la variance résiduelle.

Pr(>F) = 0.148 : p-value associée au test F.

Conclusion : La p-value (0.148) est supérieure au seuil de signification usuel (0.05).

On ne rejette donc pas l'hypothèse nulle. Il n'y a pas de différence statistiquement significative entre les profits moyens selon les villes.

En d'autres termes, la variable "ville" n'explique pas significativement la variation des profits dans les données.

15.2 Etudier la relation entre le profit et les sous_categories de produits

```
anova_sous_cat <- aov(Profit ~ Sous_Categorie , data = sales)
summary(anova_sous_cat)

##              Df      Sum Sq Mean Sq F value Pr(>F)
## Sous_Categorie  11 2.520e+07 2290464   1.847 0.0424 *
## Residuals      1182 1.466e+09 1240224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interprétation : p-value = 0.0421 : Cette valeur est inférieure à 0.05, ce qui signifie que le résultat est statistiquement significatif au seuil de 5%.

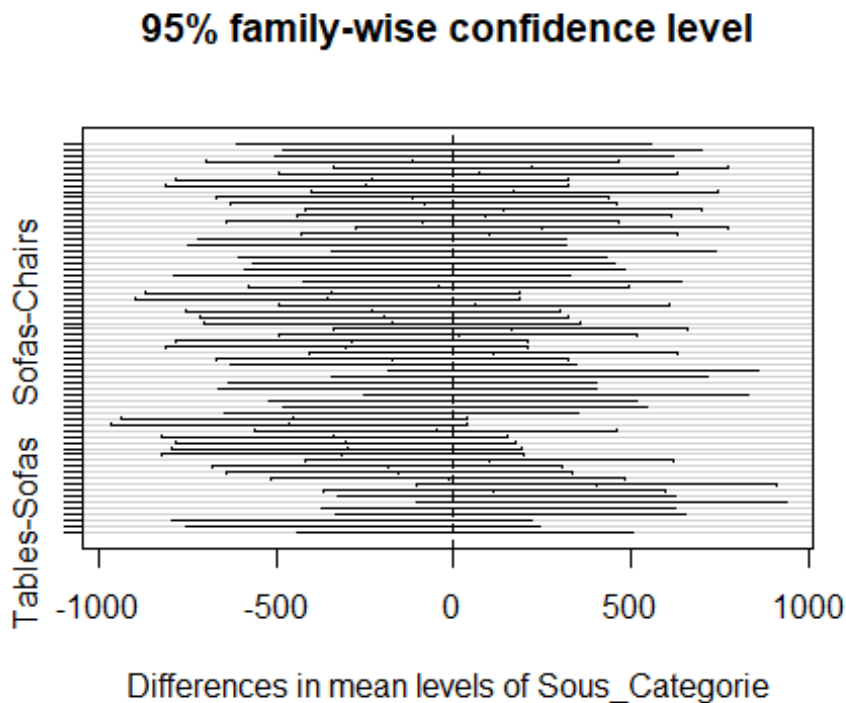
F value = 1.847 : Indique que la variation expliquée par les sous-catégories est environ 1.85 fois plus grande que la variation résiduelle (aléatoire), ce qui est modérément significatif.

Conclusion : Il existe une différence significative dans les profits moyens entre certaines sous-catégories. La variable Sous_Categorie a un effet significatif sur le profit. Certaines sous-catégories sont plus ou moins rentables que d'autres.

15.3 identifier quelles sous-catégories diffèrent significativement les unes des autres.

```
tukey_resultat <- TukeyHSD(anova_sous_cat)
```

```
plot(tukey_resultat)
```



Les résultats de p adj ne sont pas inférieurs à 0.05, indique une différence non statistiquement significative entre les sous-catégories.

15.4 Etudier la Corrélation entre quantité et CA/profit, montant et profit

```
cor(sales$Quantite, sales$montant, use = "complete.obs")
```

```
## [1] 0.04463148
```

```
cor(sales$Quantite, sales$Profit, use = "complete.obs")
```

```
## [1] 0.06608766
```

```
cor(sales$montant, sales$Profit, method = "pearson")
```

```
## [1] 0.6752854
```

Les deux coefficients de corrélation sont proches de zéro, ce qui indique qu'il n'y a pas de relation significative entre la quantité vendue et les deux mesures de performance (CA et profit).

```
client_sales$cluster <- km$cluster

# Calculer le profit moyen par cluster
prof_par_cluster <- client_sales %>%
  group_by(cluster) %>%
  summarise(profit_moyen = mean(profit))
```

```

# Identifier le cluster le plus rentable
cluster_rentable <- prof_par_cluster %>%
  filter(profit_moyen == max(profit_moyen))

# Récupérer les clients du cluster le plus rentable
clients_rentables <- client_sales %>%
  filter(cluster == cluster_rentable$cluster)

print(clients_rentables)

## # A tibble: 76 × 5
##   nom_client      CA profit  freq cluster
##   <chr>      <int> <int> <int>   <int>
## 1 Aaron Johnson  19018  6161     2       3
## 2 Andrew Allen   18675  6345     4       3
## 3 Andrew Fernandez 15121  4533     3       3
## 4 Andrew Griffin 13028  4432     3       3
## 5 Andrew Krueger  16880  4812     3       3
## 6 Anna Blackburn  21254  5595     4       3
## 7 Anna Ferguson  22530  3793     4       3
## 8 Benjamin Meadows 14319  5334     3       3
## 9 Brenda Murphy   15081  3508     3       3
## 10 Brent Hernandez 19611  7116     2       3
## # i 66 more rows

```

Profil Détaillé des Segments de Clients L'axe principal Dim1 (qui explique 80% de la différence entre les clients) peut être interprété comme un axe de "valeur client" globale. Plus un client est à droite sur le graphique, plus son CA, son profit et sa freq sont élevés.

Cluster 1 (Rouge) : Les "Clients Occasionnels" ou "Petits Clients" Profil : Ce groupe, situé à gauche (valeurs Dim1 négatives), est composé des clients avec le CA, le profit et la fréquence d'achat les plus faibles. Le fait qu'ils soient très groupés (denses) suggère qu'ils se ressemblent beaucoup : ce sont des acheteurs sporadiques ou à faible panier.

Stratégie :

Objectif : Augmenter la fréquence d'achat.

Actions : Campagnes d'e-mailing automatisées, promotions pour les inciter à un deuxième achat, newsletters pour garder la marque à l'esprit. L'investissement marketing par client doit rester faible.

Cluster 2 (Vert) : Les "Clients Réguliers" ou "Bons Clients" Profil : Ce sont vos clients intermédiaires. Ils ont un CA, un profit et une fréquence corrects et constituent probablement le cœur de votre base de clients. Ils se situent entre les clients occasionnels et les clients VIP.

Stratégie :

Objectif : Fidéliser et augmenter la valeur (upsell/cross-sell).

Actions : Programmes de fidélité, offres personnalisées basées sur leur historique, communication régulière pour les valoriser. Ils ont le potentiel de devenir des clients VIP.

Cluster 3 (Bleu) : Les “Clients VIP” ou “Clients Clés” Profil : Ce groupe à droite (valeurs Dim1 élevées) contient vos meilleurs clients. Ils génèrent le CA et le profit les plus élevés et achètent le plus fréquemment. Ils sont moins nombreux mais extrêmement précieux. La dispersion du groupe (hétérogénéité) montre que même parmi vos VIPs, les comportements peuvent varier (certains ont un CA énorme en peu de commandes, d’autres ont un CA élevé via de très nombreuses commandes).

Stratégie :

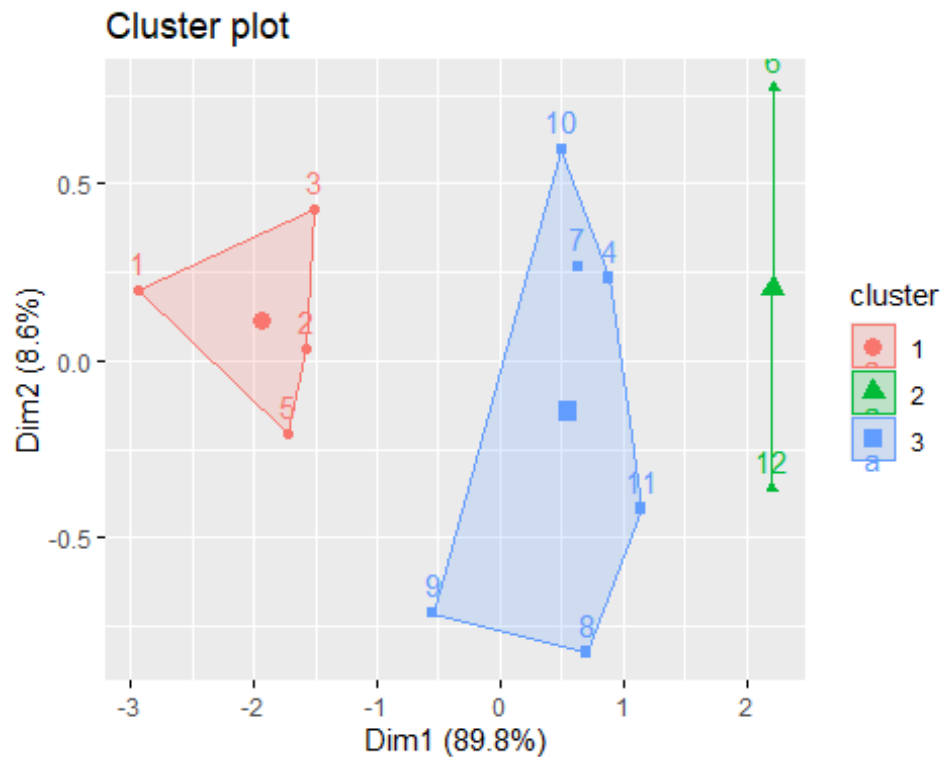
Objectif : Rétention à tout prix.

Actions : Service client prioritaire, offres exclusives, accès en avant-première aux nouveautés, contact personnalisé. Il est crucial de comprendre pourquoi ils sont si rentables et de s’assurer de leur satisfaction.

17. Identifier les sous_categorie qui sont rentables par la méthode de Clustering

```
sous_cat_sales <- sales %>%
  group_by(Sous_Categorie) %>%
  summarise(CA = sum(montant), profit = sum(Profit), freq = n())

kmeans_data_1 <- scale(sous_cat_sales[, c("CA", "profit", "freq")])
set.seed(123)
km <- kmeans(kmeans_data_1, centers = 3)
fviz_cluster(km, data = kmeans_data_1)
```



```
sous_cat_sales$cluster <- km$cluster

# Calculer le profit moyen par cluster
prof_par_cluster_1 <- sous_cat_sales %>%
  group_by(cluster) %>%
  summarise(profit_moyen = mean(profit))

# Identifier le cluster le plus rentable
cluster_rentable_1 <- prof_par_cluster_1 %>%
  filter(profit_moyen == max(profit_moyen))

# Récupérer les clients du cluster le plus rentable
sous_categorie_rentables <- sous_cat_sales %>%
  filter(cluster == cluster_rentable_1$cluster)

print(sous_categorie_rentables)

## # A tibble: 2 × 5
##   Sous_Categorie    CA profit  freq cluster
##   <chr>          <int> <int> <int>   <int>
## 1 Markers      627875 174749   110     2
## 2 Tables      625177 156796   122     2
```

Le Graphique : Le graphique est une “carte” en 2D qui positionne chaque sous-catégorie (représentée par un numéro).

L'axe horizontal (Dim1) explique à lui seul 89.8% des différences entre les produits. C'est l'axe de la performance globale. Plus un produit est à droite, plus son CA, son profit et son volume de vente sont élevés.

L'axe vertical (Dim2) explique les 8.6% restants des différences.

Au total, ce graphique représente 98.4% de l'information, ce qui le rend extrêmement fiable pour comprendre les relations entre les produits.

Les 3 Segments Révélés L'algorithme a identifié 3 segments distincts et stratégiquement pertinents :

Cluster 2 (Vert, Triangles) : Les “Étoiles” Position : Extrême droite sur l'axe de la performance (Dim1).

Composition probable : Ce segment ne contient que deux produits, qui sont les champions incontestés. En se basant sur les analyses précédentes, il s'agit très certainement des Tables et des Markers.

Rôle Stratégique : Ce sont les moteurs de l'entreprise. Ils génèrent le plus de revenus, de profits et de volume. Ils méritent une attention maximale.

Cluster 3 (Bleu, Carrés) : Le “Cœur de Gamme” Position : Au centre du graphique.

Composition probable : C'est le plus grand groupe, contenant la majorité des produits : Chaises, Papier, Stylos, Téléphones, Imprimantes, etc.

Rôle Stratégique : Ce sont les “bons élèves”. Ils constituent la base stable et solide de l'entreprise. Individuellement, ils ne sont pas des superstars, mais collectivement, ils sont absolument essentiels au chiffre d'affaires et au profit globaux.

Cluster 1 (Rouge, Cercles) : Les “Produits de Niche” ou “Défis” Position : À gauche sur l'axe de la performance (Dim1).

Composition probable : Ce groupe contient les produits les moins performants en termes de volume et de CA. On y retrouve très probablement les Classeurs (Binders), les Bibliothèques (Bookcases) et les Jeux Électroniques (Electronic Games).

Résumé du Projet : Analyse des Performances Commerciales

Ce projet a été mené pour répondre à une problématique métier centrale : comment améliorer les performances commerciales en identifiant les leviers de croissance les plus efficaces ? L'objectif était d'analyser une base de données de ventes pour identifier les produits, segments de clients et zones géographiques les plus rentables afin de maximiser les ventes et les bénéfices.

Méthodologie : Basé sur un jeu de données de 1194 observations, le projet a suivi une démarche analytique complète.

Exploration et Prétraitement : Analyse de la distribution des variables clés comme le chiffre d'affaires (allant de 508 € à 9 992 €) et le profit (médiane à 1014 €), et préparation des données (gestion des dates, vérification des doublons et valeurs manquantes).

Analyse Statistique et Visualisation : Calcul de KPIs par sous_catégorie, ville et client. Des visualisations ont été créées pour illustrer les tendances, notamment la performance des sous-catégories et les évolutions temporelles.

Modélisation : Des techniques de clustering (k-means) ont été appliquées pour segmenter les clients et les produits en groupes homogènes, révélant ainsi des profils distincts.

Principaux Résultats :

Performance Produits : L'analyse a identifié des "Étoiles" (Marqueurs, Tables) générant le plus de revenus , des "Pépites cachées" (Chaises, Ordinateurs portables) avec les meilleures marges , et des "Piliers" (Stylos) servant de produits d'appel à fort volume mais faible marge.

Performance Géographique : Orlando s'est révélée être la ville la plus performante sur tous les plans (CA, profit, panier moyen) , tandis que New York présente un profil atypique de "haute fréquence, faible panier".

Performance Temporelle : L'activité de l'entreprise est marquée par une forte volatilité mensuelle plutôt que par une croissance stable, avec un pic de performance notable en 2022.

Segmentation Client : Trois segments de clients ont été identifiés : les "Clients VIP" à haute valeur, les "Clients Réguliers" formant le cœur de l'activité, et les "Clients Occasionnels" à faible panier.

Recommandations Stratégiques

Basées sur les analyses, les actions suivantes sont recommandées pour stimuler la croissance et la rentabilité :

1. Optimisation du Portefeuille Produits

Protéger les "Étoiles" (Marqueurs, Tables) : Prioriser la gestion des stocks et les efforts marketing sur ces produits qui sont le cœur de l'activité. **Développer les "Pépites" (Chaises, Ordinateurs portables) :** Lancer des campagnes ciblées pour augmenter le volume de ventes de ces produits à haute marge afin de maximiser la rentabilité globale. **Rentabiliser les "Piliers" (Stylos) :** Utiliser ces produits pour attirer les clients, mais les inclure dans des offres groupées ("bundles") avec des articles à plus forte marge pour améliorer leur contribution au profit.

2. Stratégie Géographique Ciblée

Capitaliser sur Orlando : Renforcer la présence des produits les plus performants (Imprimantes, Marqueurs) dans cette ville leader pour maximiser les retours. Spécialiser l'offre à San Francisco : Développer le catalogue de Bibliothèques, une combinaison produit-ville qui a prouvé sa rentabilité.

Augmenter la valeur à New York : Mettre en place des stratégies de "up-selling" et "cross-selling" pour augmenter le panier moyen, qui est anormalement bas malgré un fort volume de commandes.

3. Gestion de la Relation Client (CRM)

Choyer les "Clients VIP" : Mettre en place un programme de rétention premium avec un service client prioritaire et des offres exclusives. Fidéliser les "Clients Réguliers" : Déployer des programmes de fidélité et des communications personnalisées pour les faire évoluer vers le segment VIP. Activer les "Clients Occasionnels" : Lancer des campagnes de réactivation à faible coût (emailing, promotions ciblées) pour augmenter leur fréquence d'achat.

4. Stabilisation de la Performance

Analyser la volatilité : Investiguer les causes des pics et creux mensuels (contrats ponctuels, saisonnalité, campagnes marketing) afin de lisser les revenus et d'améliorer la prévisibilité financière.

Sauvegarde des résultats

```
write.csv(kpi_ville, "kpi_par_ville.csv", row.names = FALSE)
write.csv(kpi_sous_cat, "kpi_sous_cat.csv", row.names = FALSE)
write.csv(kpi_client, "clients_rentables.csv", row.names = FALSE)
write.csv(client_sales, "segmentation_clients.csv", row.names = FALSE)
write.csv(sous_cat_sales, "sous_cat_sales.csv", row.names = FALSE)
```