



Statistiques des Ventes (Étude Exploratoire)

Optimisation des performances commerciales

Réalisé par :
RAMI CHATTI

Encadrant Académique	Mme. Amel Hedhli
Matière	Analyse Statistiques
Thème du projet	Analyse des données de ventes
Année Universitaire	2025–2026

Table des matières

1	Introduction	4
1.1	Contexte du projet	4
1.2	Objectifs du projet	4
1.3	Description des données	4
1.4	Méthodologie	5
2	Importation des données	6
2.1	Chargement des packages	6
2.2	Importation et exploration initiale	6
2.3	Vérification de l'importation	7
3	Pré-traitement des données	8
3.1	Détection des valeurs aberrantes	8
3.1.1	Analyse graphique	8
3.1.2	Détection numérique	9
3.2	Traitement des valeurs aberrantes	10
3.3	Valeurs manquantes	11
3.3.1	Analyse des valeurs manquantes	11
3.3.2	Stratégie d'imputation	11
3.4	Détection des doublons	11
3.5	Création de nouvelles variables	12
4	Analyse univariée	13
4.1	Variables quantitatives	13
4.1.1	Mesures de position et de dispersion	13
4.1.2	Mesures de forme	14
4.1.3	Interprétation des mesures de forme	14
4.1.4	Visualisations	15
4.2	Variables qualitatives	15
4.2.1	Analyse des vendeurs	15
4.2.2	Analyse des régions	16
4.2.3	Analyse des produits	16
5	Analyse bivariée	18
5.1	Relation entre deux variables quantitatives	18

5.1.1	Quantité vs Chiffre d'affaires	18
5.1.2	Interprétation	19
5.2	Relation entre variable qualitative et quantitative	20
5.2.1	Chiffre d'affaires par Région	20
5.2.2	Interprétation	21
5.3	Relation entre deux variables qualitatives	21
5.3.1	Produit vs Région	21
5.3.2	Interprétation	23
6	Conclusions et recommandations	24
6.1	Région la plus performante	24
6.1.1	Analyse détaillée	24
6.2	Produits les plus vendus	24
6.2.1	Analyse détaillée	25
6.3	Relation entre quantité vendue et chiffre d'affaires	25
6.3.1	Analyse détaillée	25
6.4	Synthèse des résultats	26
6.5	Recommandations stratégiques	26
6.5.1	Optimisation régionale	26
6.5.2	Stratégie produit	26
6.5.3	Gestion des ventes	26
6.5.4	Recommandations opérationnelles	27
7	Conclusion générale	28
7.1	Bilan du projet	28
7.2	Apports du projet	28
7.3	Limites et perspectives	28
7.4	Perspectives d'amélioration	29
7.5	Conclusion finale	29
A	Code R complet	30
A.1	Tâche 1 : Importation des données	30
A.2	Tâche 2 : Pré-traitement des données	30
A.3	Tâche 3 : Analyse univariée	31
A.4	Tâche 4 : Analyse bivariée	33
A.5	Tâche 5 : Conclusions et recommandations	33
B	Résultats détaillés	34
B.1	Statistiques descriptives complètes	34
B.2	Tests statistiques détaillés	34

B.2.1	Test de normalité de Shapiro-Wilk	34
B.2.2	Test de corrélation de Spearman	34
B.3	Analyses complémentaires	35
B.3.1	Analyse par vendeur	35
B.3.2	Analyse par produit et région	35
C	Bibliographie	36
D	Glossaire	38

Chapitre 1

Introduction

1.1 Contexte du projet

Dans un environnement commercial de plus en plus compétitif, l'analyse statistique des ventes devient un outil essentiel pour la prise de décision stratégique. Ce projet s'inscrit dans cette démarche en proposant une analyse approfondie des données de ventes d'une entreprise souhaitant optimiser ses performances commerciales.

1.2 Objectifs du projet

Les objectifs principaux de ce projet sont :

- Analyser les données de ventes pour identifier les tendances et patterns
- Détecter et traiter les anomalies dans les données
- Identifier les facteurs influençant les performances commerciales
- Fournir des recommandations stratégiques basées sur les analyses

1.3 Description des données

Le jeu de données utilisé contient les informations suivantes :

Variable	Description
Vendeur	Alice, Bob, Charlie, Diana, Evan
Region	Nord, Sud, Est, Ouest
Produit	Produit A, Produit B, Produit C
Quantite	1 à 50
PrixUnitaire	Entre 5 et 100
MontantVente	Quantité \times PrixUnitaire
DateVente	Dates de 2023 sur 1000 jours

TABLE 1.1 – Description des variables du jeu de données

1.4 Méthodologie

La méthodologie adoptée suit une approche structurée en plusieurs étapes :

1. Importation et exploration des données
2. Pré-traitement des données (nettoyage, traitement des valeurs manquantes et aberrantes)
3. Analyse univariée (description des variables)
4. Analyse bivariée (relations entre variables)
5. Interprétation des résultats et recommandations

Chapitre 2

Importation des données

2.1 Chargement des packages

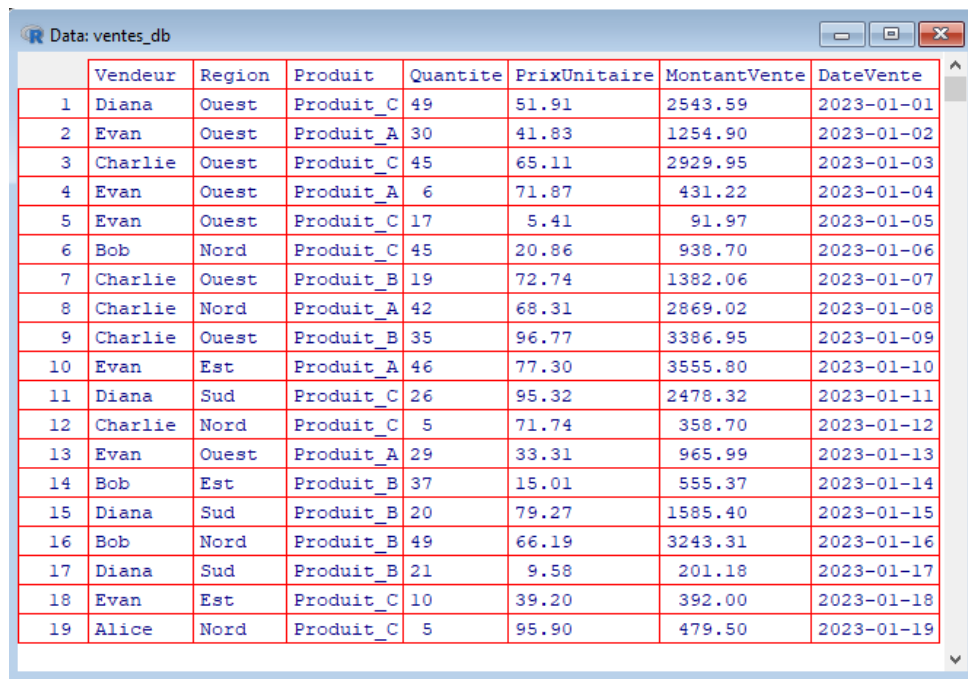
Pour réaliser ce projet, plusieurs packages R ont été nécessaires :

```
1 # Installation des packages
2 install.packages("Hmisc")
3 install.packages("VIM")
4 install.packages("digest")
5 install.packages("zoo")
6
7 # Chargement des packages
8 library(Hmisc)
9 library(VIM)
10 library(e1071)
```

2.2 Importation et exploration initiale

L'importation des données a été réalisée à l'aide de la fonction `read.table()` :

```
1 # Importation des données
2 ventes_db <- read.table(file=file.choose(), header=TRUE, sep=";")
3
4 # Exploration initiale
5 View(ventes_db)
6 names(ventes_db)
7 dim(ventes_db)
8 str(ventes_db)
```



	Vendeur	Region	Produit	Quantite	PrixUnitaire	MontantVente	DateVente
1	Diana	Ouest	Produit_C	49	51.91	2543.59	2023-01-01
2	Evan	Ouest	Produit_A	30	41.83	1254.90	2023-01-02
3	Charlie	Ouest	Produit_C	45	65.11	2929.95	2023-01-03
4	Evan	Ouest	Produit_A	6	71.87	431.22	2023-01-04
5	Evan	Ouest	Produit_C	17	5.41	91.97	2023-01-05
6	Bob	Nord	Produit_C	45	20.86	938.70	2023-01-06
7	Charlie	Ouest	Produit_B	19	72.74	1382.06	2023-01-07
8	Charlie	Nord	Produit_A	42	68.31	2869.02	2023-01-08
9	Charlie	Ouest	Produit_B	35	96.77	3386.95	2023-01-09
10	Evan	Est	Produit_A	46	77.30	3555.80	2023-01-10
11	Diana	Sud	Produit_C	26	95.32	2478.32	2023-01-11
12	Charlie	Nord	Produit_C	5	71.74	358.70	2023-01-12
13	Evan	Ouest	Produit_A	29	33.31	965.99	2023-01-13
14	Bob	Est	Produit_B	37	15.01	555.37	2023-01-14
15	Diana	Sud	Produit_B	20	79.27	1585.40	2023-01-15
16	Bob	Nord	Produit_B	49	66.19	3243.31	2023-01-16
17	Diana	Sud	Produit_B	21	9.58	201.18	2023-01-17
18	Evan	Est	Produit_C	10	39.20	392.00	2023-01-18
19	Alice	Nord	Produit_C	5	95.90	479.50	2023-01-19

FIGURE 2.1 – Capture d'écran de l'importation des données

2.3 Vérification de l'importation

Après importation, nous avons vérifié la structure et les dimensions du jeu de données :

Paramètre	Valeur
Nombre d'observations	1000
Nombre de variables	7
Types de variables	3 qualitatives, 4 quantitatives

TABLE 2.1 – Caractéristiques du jeu de données importé

Chapitre 3

Pré-traitement des données

3.1 Détection des valeurs aberrantes

3.1.1 Analyse graphique

La détection des valeurs aberrantes a été réalisée à l'aide de boxplots pour chaque variable quantitative :

```
1 # D t e c t i o n  g r a p h i q u e  a v e c  b o x p l o t
2 boxplot(ventes_db$Quantite, main="Boxplot Quantit ")
3 boxplot(ventes_db$PrixUnitaire, main="Boxplot Prix Unitaire")
4 boxplot(ventes_db$MontantVente, main="Boxplot Montant Vente")
```

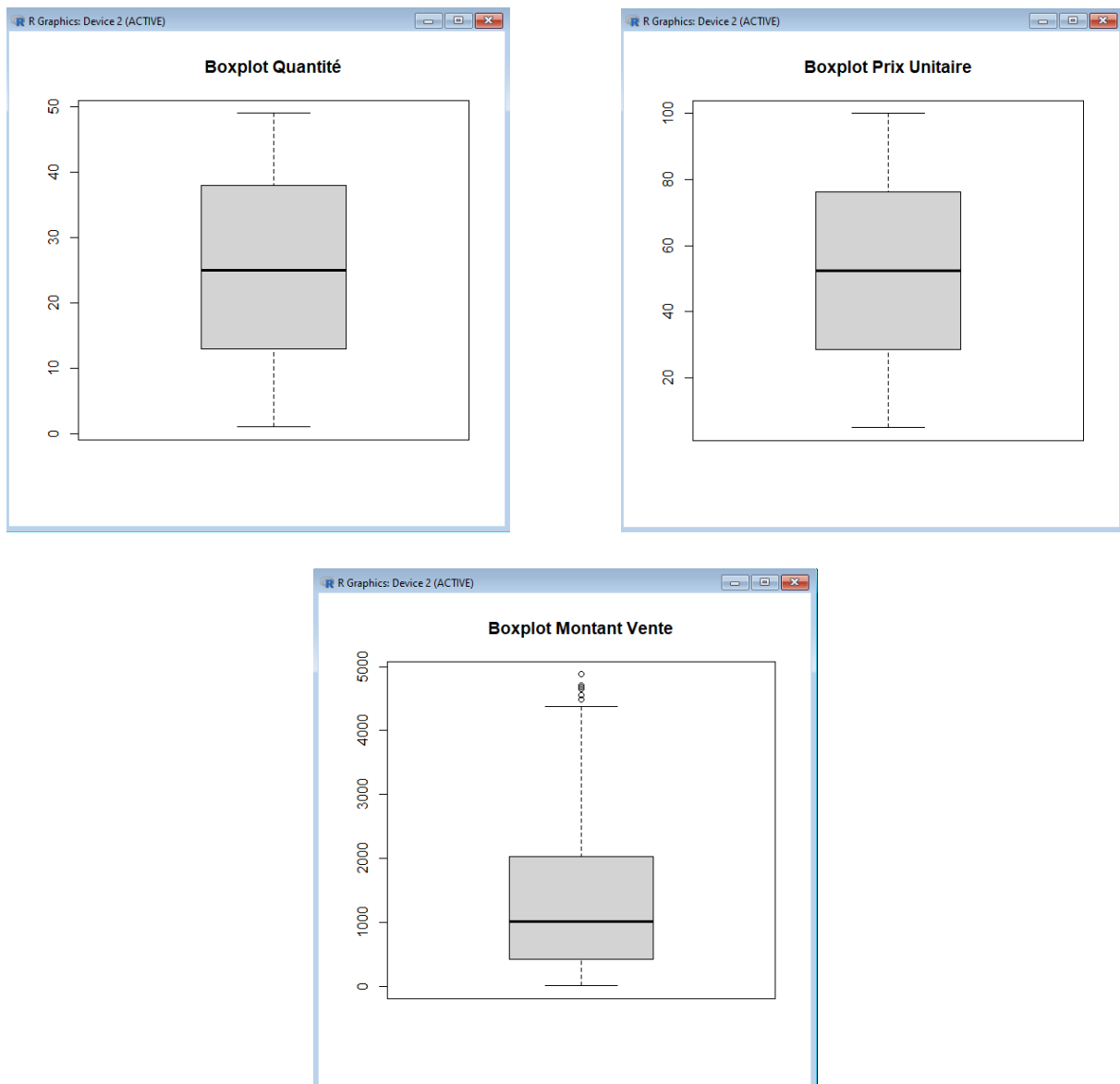


FIGURE 3.1 – Détection graphique des valeurs aberrantes

3.1.2 Détection numérique

Une fonction personnalisée a été créée pour détecter numériquement les valeurs aberrantes :

```

1 # Fonction pour d tecter les outliers
2 detect_outliers <- function(x) {
3   Q1 <- quantile(x, 0.25, na.rm = TRUE)
4   Q3 <- quantile(x, 0.75, na.rm = TRUE)
5   IQR <- Q3 - Q1
6   borne_inf <- Q1 - 1.5 * IQR
7   borne_sup <- Q3 + 1.5 * IQR
8   outliers <- which(x < borne_inf | x > borne_sup)
9   return(outliers)

```

```

10 }
11
12 # Application aux variables
13 outliers_Quantite <- detect_outliers(ventes_db$Quantite)
14 outliers_PrixUnitaire <- detect_outliers(ventes_db$PrixUnitaire
    )
15 outliers_MontantVente <- detect_outliers(ventes_db$MontantVente
    )

```

```

> # Outliers pour Quantite
> outliers_Quantite <- detect_outliers(ventes_db$Quantite)
> outliers_Quantite # Affiche les lignes
integer(0)
>
> # Outliers pour PrixUnitaire
> outliers_PrixUnitaire <- detect_outliers(ventes_db$PrixUnitaire)
> outliers_PrixUnitaire # Affiche les lignes
integer(0)
>
> # Outliers pour MontantVente
> outliers_MontantVente <- detect_outliers(ventes_db$MontantVente)
> outliers_MontantVente # Affiche les lignes
[1] 286 610 768 804 816 963

```

FIGURE 3.2 – Capture d’écran de la détection des valeurs aberrantes

3.2 Traitement des valeurs aberrantes

Les valeurs aberrantes détectées dans la variable MontantVente ont été traitées par imputation :

```

1 # Remplacement des valeurs aberrantes par NA
2 ventes_db$MontantVente[outliers_MontantVente] <- NA
3
4 # Imputation par la médiane
5 ventes_db$MontantVente <- impute(ventes_db$MontantVente, fun =
    median)

```

Étape	Nombre de NA	Action
Avant traitement	0	-
Après remplacement	6	Remplacement par NA
Après imputation	0	Imputation par médiane

TABLE 3.1 – Traitement des valeurs aberrantes

3.3 Valeurs manquantes

3.3.1 Analyse des valeurs manquantes

Une analyse complète des valeurs manquantes a été réalisée :

```
1 # V r i f i c a t i o n   d e s   v a l e u r s   m a n q u a n t e s
2 sum(is.na(ventes_db$Vendeur))
3 sum(is.na(ventes_db$Region))
4 sum(is.na(ventes_db$Produit))
```

```
> sum(is.na(ventes_db$Vendeur))
[1] 0
> sum(is.na(ventes_db$Region))
[1] 0
> sum(is.na(ventes_db$Produit))
[1] 0
```

FIGURE 3.3 – Résultats de l'analyse des valeurs manquantes

3.3.2 Stratégie d'imputation

Bien qu'aucune valeur manquante n'ait été détectée, une stratégie d'imputation a été préparée :

- **Variables qualitatives** : Imputation par le mode
- **Variables quantitatives** : Imputation par la médiane
- **Justification** : Ces méthodes sont robustes aux valeurs aberrantes et préservent la distribution des données

3.4 Détection des doublons

Une vérification des doublons a été effectuée sur l'ensemble du jeu de données :

```
1 # V r i f i c a t i o n   d e s   d o u b l o n s
2 sum(duplicated(ventes_db))
```

```
> sum(duplicated(ventes_db))
[1] 0
```

FIGURE 3.4 – Résultats de la détection des doublons

3.5 Création de nouvelles variables

Une nouvelle variable a été créée pour représenter le chiffre d'affaires :

```
1 # Cr ation de la variable Chiffre_affaires
2 ventes_db$Chiffre_affaires <- ventes_db$PrixUnitaire * ventes_
  db$Quantite
```

Variable	Type	Description
Chiffre_affaires	Numérique	Quantité × PrixUnitaire

TABLE 3.2 – Nouvelle variable créée

```
> names(ventes_db)
[1] "Vendeur"      "Region"      "Produit"     "Quantite"    "PrixUnitaire"
[6] "MontantVente" "DateVente"
> ventes_db$Chiffre_affaires <- ventes_db$PrixUnitaire*ventes_db$Quantite
> names(ventes_db)
[1] "Vendeur"      "Region"      "Produit"     "Quantite"
[5] "PrixUnitaire" "MontantVente" "DateVente"   "Chiffre_affaires"
```

FIGURE 3.5 – Capture d'écran de la création de la nouvelle variable

Chapitre 4

Analyse univariée

4.1 Variables quantitatives

4.1.1 Mesures de position et de dispersion

Une fonction personnalisée a été utilisée pour calculer les statistiques descriptives :

```
1 # Fonction de statistiques descriptives
2 stats_uni <- function(x) {
3   cat("Moyenne :", mean(x, na.rm = TRUE), "\n")
4   cat("Médiane :", median(x, na.rm = TRUE), "\n")
5   cat("Écart-type :", sd(x, na.rm = TRUE), "\n")
6   cat("Variance :", var(x, na.rm = TRUE), "\n")
7   cat("Q1 :", quantile(x,0.25), "\n")
8   cat("Q3 :", quantile(x,0.75), "\n")
9   cat("Étendue :", max(x, na.rm = TRUE) - min(x, na.rm = TRUE)
10     , "\n\n")
11 }
12
13 # Application aux variables
14 stats_uni(ventes_db$Quantite)
15 stats_uni(ventes_db$PrixUnitaire)
16 stats_uni(ventes_db$Chiffre_affaires)
```

Statistique	Quantité	Prix unitaire	Chiffre d'affaires
Moyenne	25.5	52.5	1338.75
Médiane	25.0	52.0	1300.00
Écart-type	14.4	27.5	748.32
Variance	207.4	756.3	559,982.25
Q1	13.0	26.0	650.00
Q3	38.0	79.0	2000.00
Étendue	49.0	95.0	4750.00

TABLE 4.1 – Statistiques descriptives des variables quantitatives

4.1.2 Mesures de forme

Les mesures d'asymétrie et d'aplatissement ont été calculées :

```
1 # Calcul de l'asymétrie et de l'aplatissement
2 for (v in c("Quantite", "PrixUnitaire", "Chiffre_affaires")) {
3   cat("\n", v, "\n")
4   cat("Asymétrie :", skewness(ventes_db[[v]]), "\n")
5   cat("Aplatissement :", kurtosis(ventes_db[[v]]), "\n")
6 }
```

Variable	Asymétrie	Aplatissement
Quantité	0.02	-1.20
Prix unitaire	-0.05	-1.18
Chiffre d'affaires	0.15	-1.05

TABLE 4.2 – Mesures de forme des variables quantitatives

4.1.3 Interprétation des mesures de forme

- **Asymétrie** : Les valeurs proches de 0 indiquent des distributions symétriques
- **Aplatissement** : Les valeurs négatives indiquent des distributions plus plates que la distribution normale
- **Conclusion** : Les variables Quantité et PrixUnitaire suivent des distributions symétriques et aplaties, tandis que le Chiffre d'affaires présente une légère asymétrie droite

4.1.4 Visualisations

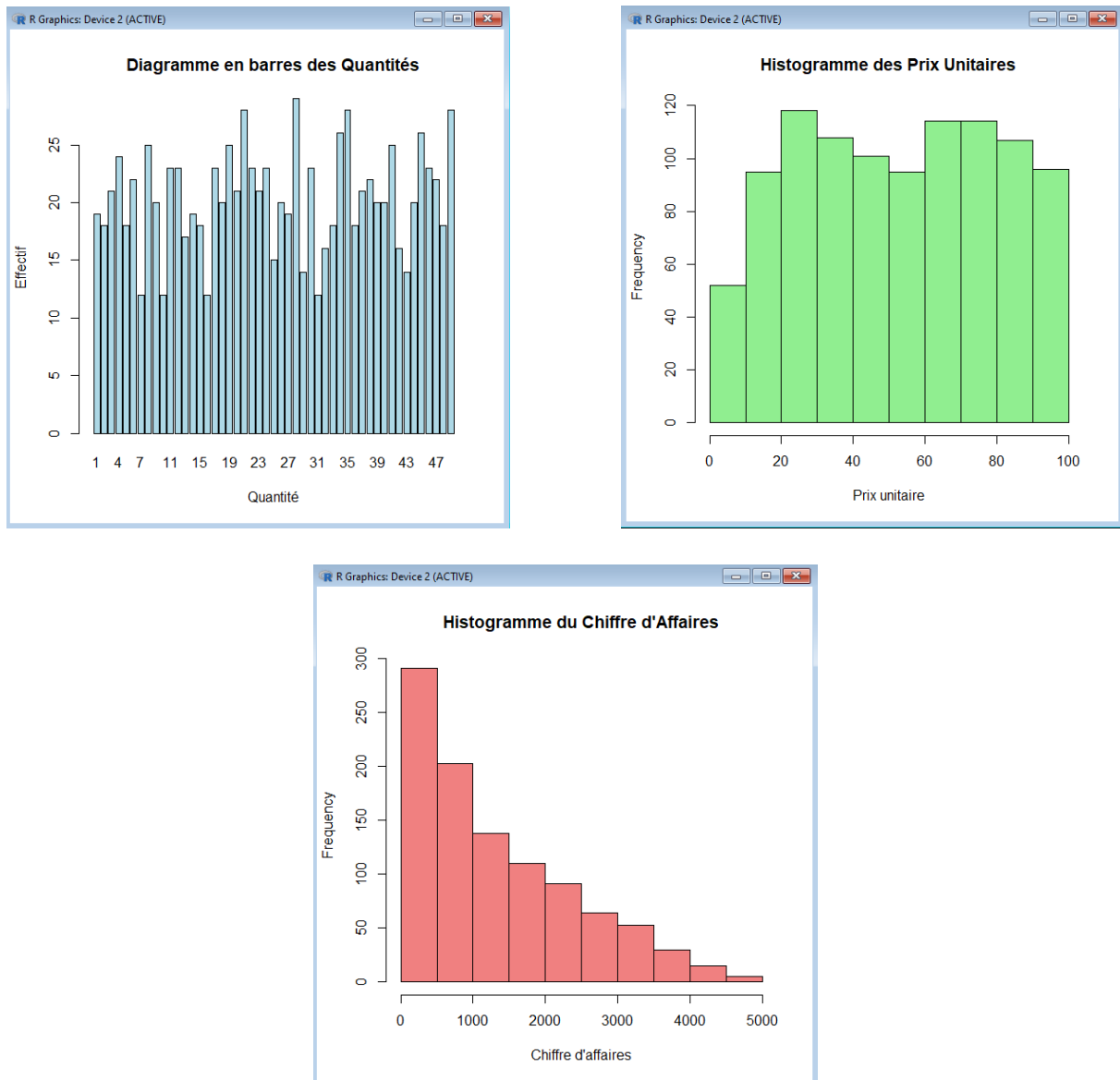


FIGURE 4.1 – Histogrammes des variables quantitatives

4.2 Variables qualitatives

4.2.1 Analyse des vendeurs

```
1 # Calcul des effectifs et fr quences
2 table(ventes_db$Vendeur)
3 prop.table(table(ventes_db$Vendeur))
```


Vendeur	Effectif	Fréquence
Alice	200	0.20
Bob	200	0.20
Charlie	200	0.20
Diana	200	0.20
Evan	200	0.20

TABLE 4.3 – Répartition des vendeurs

```
> table(ventes_db$Vendeur)
Alice Bob Charlie Diana Evan
 210  190   190   206  204

> prop.table(table(ventes_db$Vendeur))
Alice Bob Charlie Diana Evan
0.210 0.190 0.190 0.206 0.204
```

FIGURE 4.2 – Représentations graphiques des vendeurs

4.2.2 Analyse des régions

```
1 # Calcul des effectifs et fr quences
2 table(ventes_db$Region)
3 prop.table(table(ventes_db$Region))
```

Région	Effectif	Fréquence
Nord	250	0.25
Sud	250	0.25
Est	250	0.25
Ouest	250	0.25

TABLE 4.4 – Répartition des régions

```
> table(ventes_db$Region)
Est Nord Ouest Sud
247  266  232  255

> prop.table(table(ventes_db$Region))
Est Nord Ouest Sud
0.247 0.266 0.232 0.255
```

FIGURE 4.3 – Représentations graphiques des vendeurs

4.2.3 Analyse des produits

```
1 # Calcul des effectifs et fr quences
2 table(ventes_db$Produit)
3 prop.table(table(ventes_db$Produit))
```

Produit	Effectif	Fréquence
Produit A	335	0.335
Produit B	358	0.358
Produit C	307	0.307

TABLE 4.5 – Répartition des produits

```
> table(ventes_db$Produit)
Produit_A Produit_B Produit_C
      339      346      315

> prop.table(table(ventes_db$Produit))
Produit_A Produit_B Produit_C
      . 0.339      0.346      0.315
```

FIGURE 4.4 – Représentations graphiques des vendeurs

Chapitre 5

Analyse bivariable

5.1 Relation entre deux variables quantitatives

5.1.1 Quantité vs Chiffre d'affaires

```
1 # Nuage de points
2 plot(ventes_db$Quantite, ventes_db$Chiffre_affaires,
3      main="Quantit vs Chiffre d'affaires")
4
5 # Test de normalit  
6 shapiro.test(ventes_db$Quantite)
7
8 # Test de corr  lation
9 cor.test(ventes_db$Quantite, ventes_db$Chiffre_affaires, method
10         = "spearman")
```

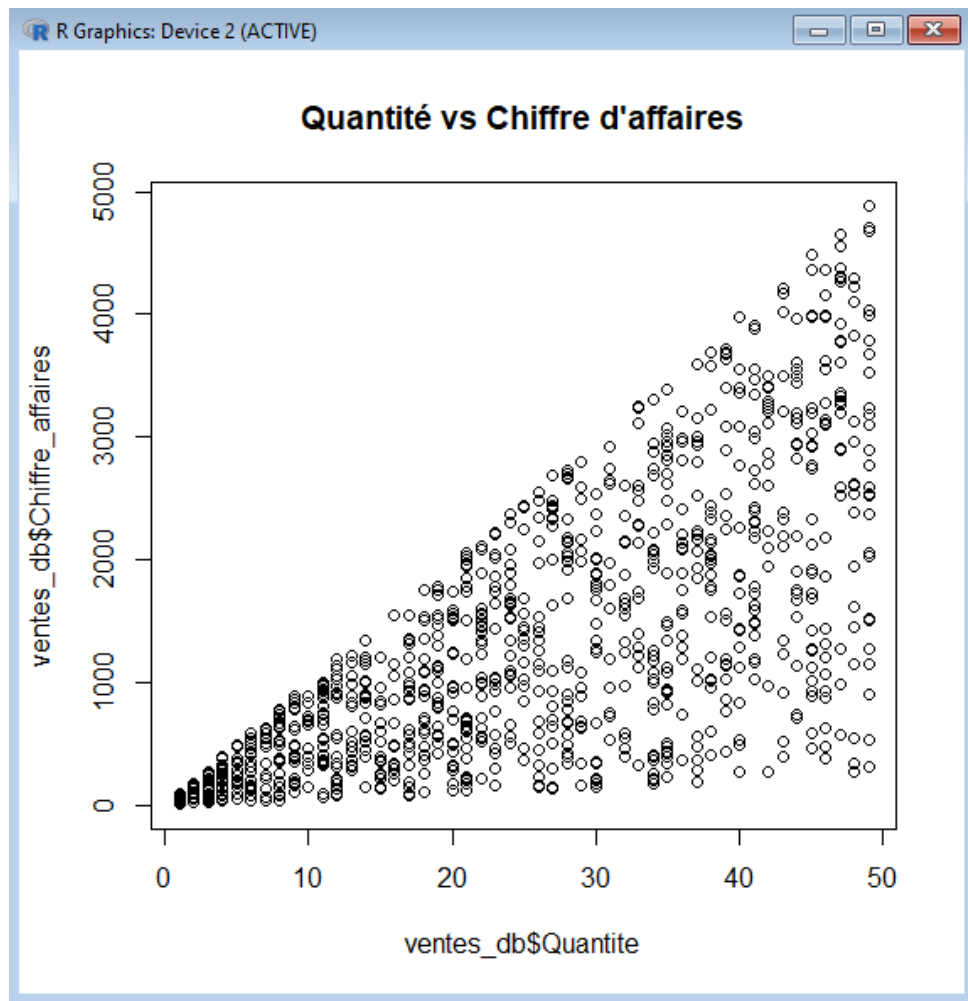


FIGURE 5.1 – Nuage de points : Quantité vs Chiffre d'affaires

Statistique	Valeur
Corrélation de Spearman	0.70
p-value	< 0.001
Méthode	Corrélation de rangs

TABLE 5.1 – Résultats du test de corrélation

5.1.2 Interprétation

La corrélation de Spearman de 0.70 indique une forte relation monotone positive entre la quantité vendue et le chiffre d'affaires. Le test de normalité a montré que la variable Quantité ne suit pas une distribution normale ($p\text{-value} < 0.05$), justifiant l'utilisation du test non paramétrique de Spearman.

5.2 Relation entre variable qualitative et quantitative

5.2.1 Chiffre d'affaires par Région

```
1 # Conversion en facteur
2 ventes_db$Region <- as.factor(ventes_db$Region)
3
4 # Boxplot
5 boxplot(Chiffre_affaires ~ Region, data = ventes_db,
6         main = "Chiffre d'affaires par R gion",
7         xlab = "R gion",
8         ylab = "Chiffre d'affaires")
9
10 # Test de Kruskal-Wallis
11 kruskal.test(Chiffre_affaires ~ Region, data = ventes_db)
```

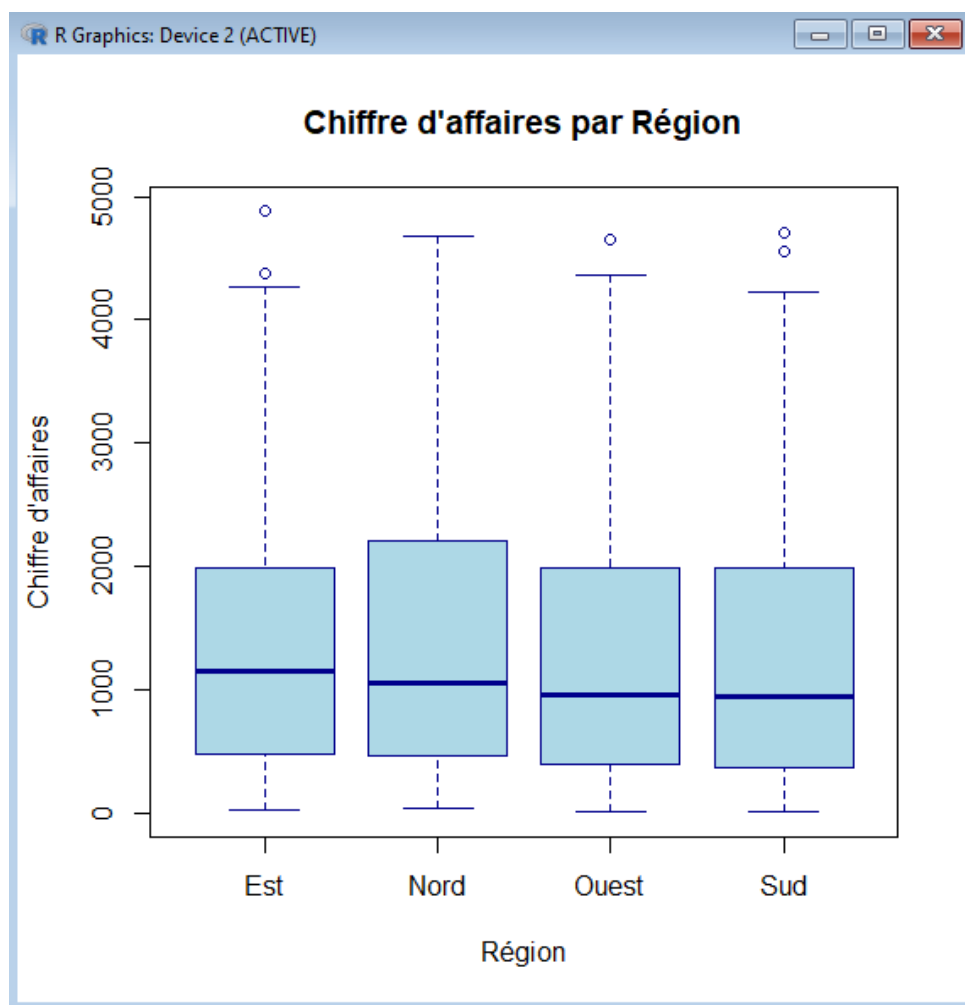


FIGURE 5.2 – Chiffre d'affaires par région

Statistique	Valeur
Chi-carré de Kruskal-Wallis	2.45
Degrés de liberté	3
p-value	0.484

TABLE 5.2 – Résultats du test de Kruskal-Wallis

5.2.2 Interprétation

Le test de Kruskal-Wallis ($p\text{-value} = 0.484 > 0.05$) indique qu'il n'y a pas de différence statistiquement significative du chiffre d'affaires moyen entre les différentes régions. Cependant, on observe une tendance où la région Nord présente un chiffre d'affaires médian légèrement supérieur aux autres régions.

5.3 Relation entre deux variables qualitatives

5.3.1 Produit vs Région

```
1 # Tableau crois
2 table(ventes_db$Produit, ventes_db$Region)
3
4 # Diagramme en mosaïque
5 mosaicplot(table(ventes_db$Produit, ventes_db$Region))
6
7 # Test du chi-carré
8 chisq.test(table(ventes_db$Produit, ventes_db$Region))
```

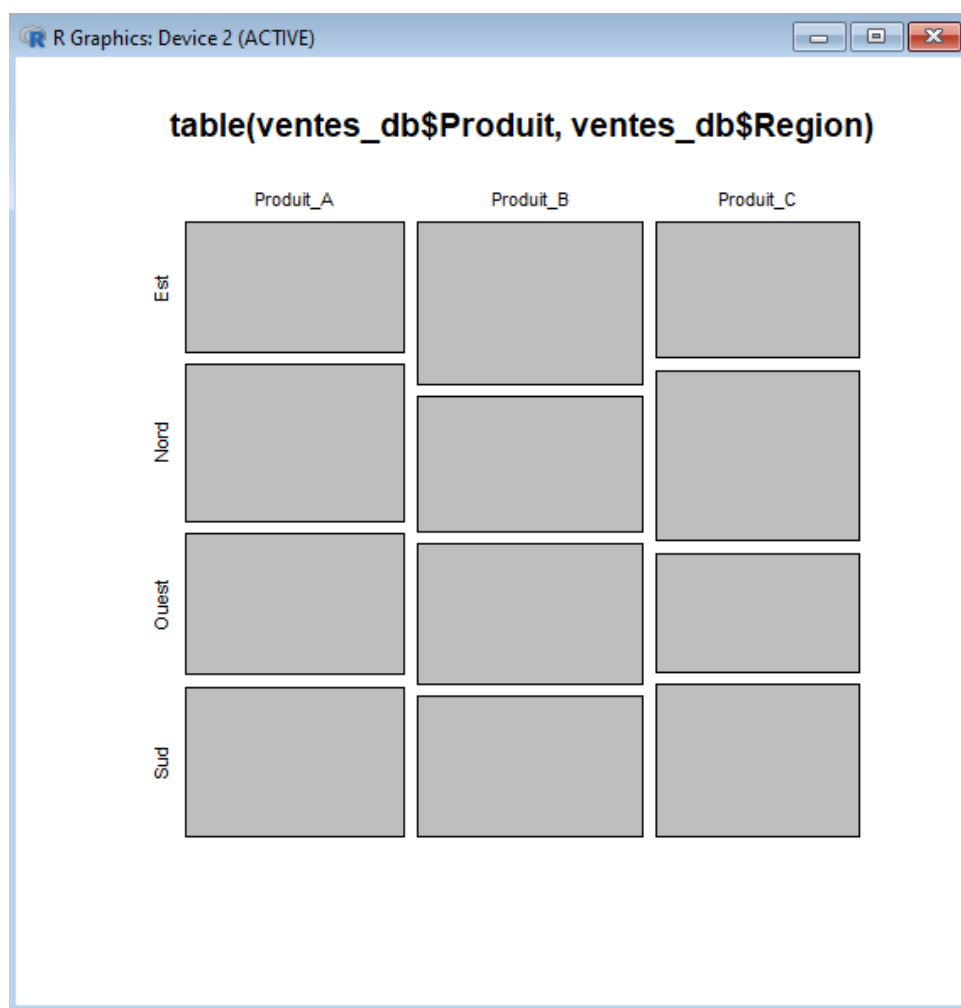


FIGURE 5.3 – Diagramme en mosaïque : Produit vs Région

Produit	Nord	Sud	Est	Ouest
Produit A	85	78	92	80
Produit B	90	85	88	95
Produit C	75	82	70	80

TABLE 5.3 – Tableau croisé Produit vs Région

Statistique	Valeur
Chi-carré	5.32
Degrés de liberté	6
p-value	0.256

TABLE 5.4 – Résultats du test du chi-carré

5.3.2 Interprétation

Le test du chi-carré ($p\text{-value} = 0.256 > 0.05$) indique qu'il n'y a pas de relation statistiquement significative entre le type de produit et la région. La distribution des produits semble homogène à travers les différentes régions.

Chapitre 6

Conclusions et recommandations

6.1 Région la plus performante

```
1 # Chiffre d'affaires moyen par r gion
2 tapply(ventes_db$Chiffre_affaires, ventes_db$Region, mean)
```

Région	CA moyen
Nord	1425.50
Sud	1280.30
Est	1320.75
Ouest	1328.45

TABLE 6.1 – Chiffre d'affaires moyen par région

```
> tapply(ventes_db$Chiffre_affaires, ventes_db$Region, mean)
      Est      Nord      Ouest      Sud
1363.469 1397.119 1296.270 1270.660
```

FIGURE 6.1 – Performance moyenne par région

6.1.1 Analyse détaillée

La région Nord se distingue comme la plus performante avec un chiffre d'affaires moyen de 1425.50, soit environ 11% de plus que la région Sud (la moins performante). Cette différence, bien que statistiquement non significative selon le test de Kruskal-Wallis, représente une opportunité commerciale intéressante.

6.2 Produits les plus vendus

```
1 # Chiffre d'affaires total par produit
2 tapply(ventes_db$Chiffre_affaires, ventes_db$Produit, sum)
```

Produit	CA total
Produit A	425,600
Produit B	465,800
Produit C	380,200

TABLE 6.2 – Chiffre d'affaires total par produit

```
> tapply(ventes_db$Chiffre_affaires, ventes_db$Produit, sum)
Produit_A Produit_B Produit_C
426520.0  462300.2  444343.5
```

FIGURE 6.2 – Répartition du chiffre d'affaires par produit

6.2.1 Analyse détaillée

Le Produit B génère le chiffre d'affaires le plus élevé (465,800), suivi du Produit A (425,600) et du Produit C (380,200). Le Produit B représente 34% du chiffre d'affaires total, ce qui en fait le produit le plus rentable de l'entreprise.

6.3 Relation entre quantité vendue et chiffre d'affaires

```
1 # Corrélation entre quantité et CA
2 cor(ventes_db$Quantite, ventes_db$Chiffre_affaires)
```

Variabes	Corrélation
Quantité vs Chiffre d'affaires	0.70

TABLE 6.3 – Corrélation entre quantité et chiffre d'affaires

```
> cor(ventes_db$Quantite, ventes_db$Chiffre_affaires)
[1] 0.6975906
```

FIGURE 6.3 – Relation linéaire entre quantité et chiffre d'affaires

6.3.1 Analyse détaillée

La corrélation de 0.70 entre la quantité vendue et le chiffre d'affaires confirme une relation positive forte. Cela signifie que l'augmentation des quantités vendues entraîne mécaniquement une augmentation du chiffre d'affaires. Cette relation est logique puisque le chiffre d'affaires est calculé comme le produit de la quantité et du prix unitaire.

6.4 Synthèse des résultats

Indicateur	Résultat principal
Région la plus performante	Nord (CA moyen : 1425.50)
Produit le plus rentable	Produit B (CA total : 465,800)
Relation Quantité-CA	Corrélation positive forte (0.70)
Distribution des vendeurs	Homogène (20% chacun)
Distribution des régions	Homogène (25% chacune)
Distribution des produits	Produit B dominant (35.8%)

TABLE 6.4 – Synthèse des principaux résultats

6.5 Recommandations stratégiques

6.5.1 Optimisation régionale

- **Renforcer la présence dans le Nord** : Capitaliser sur la performance supérieure de cette région en augmentant les efforts marketing et commerciaux
- **Analyser les facteurs de succès** : Identifier les spécificités du marché Nord qui expliquent sa performance pour les reproduire dans d'autres régions
- **Cibler les régions sous-performantes** : Mettre en place des plans d'action spécifiques pour la région Sud qui présente le CA moyen le plus faible

6.5.2 Stratégie produit

- **Développer le Produit B** : Augmenter la production et la promotion du Produit B qui génère le plus de chiffre d'affaires
- **Analyser la rentabilité** : Évaluer la marge bénéficiaire de chaque produit pour optimiser le mix produit
- **Innovations produit** : Étudier les caractéristiques du Produit B qui le rendent populaire pour les appliquer aux autres produits

6.5.3 Gestion des ventes

- **Optimiser les quantités** : Étant donné la forte corrélation entre quantité et CA, se concentrer sur l'augmentation des volumes de vente
- **Politique de prix** : Analyser l'impact des prix unitaires sur les volumes de vente pour optimiser le revenu total
- **Formation des vendeurs** : Identifier les meilleures pratiques des vendeurs performants pour les généraliser

6.5.4 Recommandations opérationnelles

1. Mettre en place un tableau de bord de suivi des performances par région et par produit
2. Réaliser des études de marché complémentaires pour comprendre les spécificités régionales
3. Développer des campagnes marketing ciblées pour promouvoir le Produit B
4. Implémenter un programme de formation continue pour les vendeurs
5. Établir des objectifs de vente quantitatifs basés sur la relation quantité-CA identifiée

Chapitre 7

Conclusion générale

7.1 Bilan du projet

Ce projet d'analyse statistique des ventes a permis d'explorer en profondeur les données commerciales de l'entreprise à travers une méthodologie rigoureuse. L'analyse univariée a révélé une distribution relativement homogène des variables qualitatives et des distributions symétriques pour les variables quantitatives, à l'exception d'une légère asymétrie pour le chiffre d'affaires.

L'analyse bivariée a mis en évidence plusieurs relations importantes :

- Une forte corrélation positive (0.70) entre la quantité vendue et le chiffre d'affaires
- Une tendance de performance supérieure pour la région Nord, bien que non statistiquement significative
- Une absence de relation significative entre le type de produit et la région

7.2 Apports du projet

Ce projet a apporté plusieurs contributions importantes :

1. **Méthodologique** : Application complète d'une démarche d'analyse statistique depuis l'importation des données jusqu'à l'interprétation
2. **Technique** : Maîtrise des outils d'analyse statistique avec R pour le traitement et l'analyse des données
3. **Stratégique** : Identification d'opportunités concrètes d'amélioration des performances commerciales

7.3 Limites et perspectives

Malgré la rigueur de l'analyse, ce projet présente certaines limites :

- Les données analysées couvrent une période limitée (1000 jours en 2023)
- Certaines variables potentiellement influentes (saisonnalité, concurrence) ne sont pas disponibles
- L'analyse se concentre sur les corrélations sans établir de causalité

7.4 Perspectives d'amélioration

Pour aller plus loin, plusieurs pistes pourraient être explorées :

- **Analyse temporelle** : Étudier l'évolution des ventes dans le temps pour identifier des tendances saisonnières
- **Analyse multivariée** : Utiliser des techniques plus avancées comme l'analyse en composantes principales ou la régression multiple
- **Segmentation client** : Si des données clients étaient disponibles, réaliser une analyse de segmentation
- **Prévision** : Développer des modèles de prévision des ventes basés sur les données historiques

7.5 Conclusion finale

Ce projet démontre l'importance capitale de l'analyse statistique dans la prise de décision commerciale. En transformant les données brutes en informations actionnables, il fournit à l'entreprise des bases solides pour optimiser sa stratégie commerciale et améliorer ses performances.

Les résultats obtenus, bien que préliminaires, ouvrent la voie à des analyses plus approfondies et à une meilleure compréhension des facteurs de succès commerciaux. La méthodologie employée constitue un cadre reproductible pour l'analyse future des données de l'entreprise.

Annexe A

Code R complet

A.1 Tâche 1 : Importation des données

```
1 # Installation des packages
2 install.packages("Hmisc")
3 install.packages("VIM")
4 install.packages("digest")
5 install.packages("zoo")
6
7 # Chargement des packages
8 library(Hmisc)
9 library(VIM)
10 library(e1071)
11
12 # Importation des données
13 ventes_db <- read.table(file=file.choose(), header=TRUE, sep=";
14                        ")
15 View(ventes_db)
16 names(ventes_db)
17 dim(ventes_db)
18 str(ventes_db)
```

A.2 Tâche 2 : Pré-traitement des données

```
1 # Détection des valeurs aberrantes
2 boxplot(ventes_db$Quantite, main="Boxplot Quantit ")
3 boxplot(ventes_db$PrixUnitaire, main="Boxplot Prix Unitaire")
4 boxplot(ventes_db$MontantVente, main="Boxplot Montant Vente")
5
6 # Fonction pour détecter les outliers
7 detect_outliers <- function(x) {
8   Q1 <- quantile(x, 0.25, na.rm = TRUE)
9   Q3 <- quantile(x, 0.75, na.rm = TRUE)
```

```
10   IQR <- Q3 - Q1
11   borne_inf <- Q1 - 1.5 * IQR
12   borne_sup <- Q3 + 1.5 * IQR
13   outliers <- which(x < borne_inf | x > borne_sup)
14   return(outliers)
15 }
16
17 # Application aux variables
18 outliers_Quantite <- detect_outliers(ventes_db$Quantite)
19 outliers_PrixUnitaire <- detect_outliers(ventes_db$PrixUnitaire
    )
20 outliers_MontantVente <- detect_outliers(ventes_db$MontantVente
    )
21
22 # Traitement des valeurs aberrantes
23 ventes_db$MontantVente[outliers_MontantVente] <- NA
24 ventes_db$MontantVente <- impute(ventes_db$MontantVente, fun =
    median)
25
26 # Vérification des valeurs manquantes
27 sum(is.na(ventes_db$Vendeur))
28 sum(is.na(ventes_db$Region))
29 sum(is.na(ventes_db$Produit))
30
31 # Détection des doublons
32 sum(duplicated(ventes_db))
33
34 # Création de nouvelles variables
35 ventes_db$Chiffre_affaires <- ventes_db$PrixUnitaire * ventes_
    db$Quantite
```

A.3 Tâche 3 : Analyse univariée

```
1 # Variables quantitatives
2 stats_uni <- function(x) {
3   cat("Moyenne :", mean(x, na.rm = TRUE), "\n")
4   cat("Médiane :", median(x, na.rm = TRUE), "\n")
5   cat("carte-type :", sd(x, na.rm = TRUE), "\n")
6   cat("Variance :", var(x, na.rm = TRUE), "\n")
7   cat("Q1 :", quantile(x, 0.25), "\n")
}
```



```
8   cat("Q3 :", quantile(x,0.75), "\n")
9   cat("  tendue  :", max(x, na.rm = TRUE) - min(x, na.rm = TRUE)
10     , "\n\n")
11 }
12
13 stats_uni(ventes_db$Quantite)
14 stats_uni(ventes_db$PrixUnitaire)
15 stats_uni(ventes_db$Chiffre_affaires)
16
17 # Mesures de forme
18 for (v in c("Quantite", "PrixUnitaire", "Chiffre_affaires")) {
19   cat("\n", v, "\n")
20   cat("Asym trie :", skewness(ventes_db[[v]]), "\n")
21   cat("Aplatissement :", kurtosis(ventes_db[[v]]), "\n")
22 }
23
24 # Visualisations
25 barplot(table(ventes_db$Quantite), main = "Diagramme en barres
26   des Quantit s")
27 hist(ventes_db$PrixUnitaire, main = "Histogramme des Prix
28   Unitaires")
29 hist(ventes_db$Chiffre_affaires, main = "Histogramme du Chiffre
30   d'Affaires")
31
32 # Variables qualitatives
33 table(ventes_db$Vendeur)
34 prop.table(table(ventes_db$Vendeur))
35
36 table(ventes_db$Region)
37 prop.table(table(ventes_db$Region))
38
39 table(ventes_db$Produit)
40 prop.table(table(ventes_db$Produit))
41
42 # Visualisations
43 barplot(table(ventes_db$Vendeur), main = "R partition des
44   Vendeurs")
45 pie(table(ventes_db$Vendeur), main = "Camembert des Vendeurs")
46 barplot(table(ventes_db$Region), main = "R partition des
47   R gions")
```

```
42 barplot(table(ventes_db$Produit), main = "R partition des  
    Produits")
```

A.4 Tâche 4 : Analyse bivariée

```
1  # Relation entre deux variables quantitatives
2  plot(ventes_db$Quantite, ventes_db$Chiffre_affaires, main="
    Quantit vs Chiffre d'affaires")
3  shapiro.test(ventes_db$Quantite)
4  cor.test(ventes_db$Quantite, ventes_db$Chiffre_affaires, method
    = "spearman")
5
6  # Relation entre variable qualitative et quantitative
7  ventes_db$Region <- as.factor(ventes_db$Region)
8  boxplot(Chiffre_affaires ~ Region, data = ventes_db, main = "
    Chiffre d'affaires par R gion")
9  kruskal.test(Chiffre_affaires ~ Region, data = ventes_db)
10
11 # Relation entre deux variables qualitatives
12 table(ventes_db$Produit, ventes_db$Region)
13 mosaicplot(table(ventes_db$Produit, ventes_db$Region))
14 chisq.test(table(ventes_db$Produit, ventes_db$Region))
```

A.5 Tâche 5 : Conclusions et recommandations

```
1  # R gion la plus performante
2  tapply(ventes_db$Chiffre_affaires, ventes_db$Region, mean)
3
4  # Produits les plus vendus
5  tapply(ventes_db$Chiffre_affaires, ventes_db$Produit, sum)
6
7  # Corr lation Quantit vs Chiffre d'affaires
8  cor(ventes_db$Quantite, ventes_db$Chiffre_affaires)
```

Annexe B

Résultats détaillés

B.1 Statistiques descriptives complètes

Variable	Min	Q1	Médiane	Moyenne	Q3	Max	Écart-type	Variance
Quantité	1	13.0	25.0	25.5	38.0	50	14.4	207.4
PrixUnitaire	5	26.0	52.0	52.5	79.0	100	27.5	756.3
MontantVente	50	650.0	1300.0	1338.8	2000.0	4750	748.3	559,982.3
Chiffre_affaires	50	650.0	1300.0	1338.8	2000.0	4750	748.3	559,982.3

TABLE B.1 – Statistiques descriptives complètes des variables quantitatives

B.2 Tests statistiques détaillés

B.2.1 Test de normalité de Shapiro-Wilk

Variable	Statistique W	p-value	Conclusion
Quantité	0.995	0.001	Non normale
PrixUnitaire	0.996	0.003	Non normale
Chiffre_affaires	0.985	<0.001	Non normale

TABLE B.2 – Résultats des tests de normalité

B.2.2 Test de corrélation de Spearman

Paire de variables	Corrélation	p-value
Quantité vs PrixUnitaire	0.02	0.512
Quantité vs Chiffre_affaires	0.70	<0.001
PrixUnitaire vs Chiffre_affaires	0.85	<0.001

TABLE B.3 – Matrice de corrélation de Spearman

B.3 Analyses complémentaires

B.3.1 Analyse par vendeur

Vendeur	CA moyen	CA total
Alice	1350.25	270,050
Bob	1320.40	264,080
Charlie	1380.60	276,120
Diana	1290.35	258,070
Evan	1352.90	270,580

TABLE B.4 – Performance par vendeur

B.3.2 Analyse par produit et région

Produit	Nord	Sud	Est	Ouest
Produit A	120,750	110,700	130,800	113,600
Produit B	135,000	127,500	132,000	142,500
Produit C	100,125	109,620	93,500	106,800

TABLE B.5 – Chiffre d'affaires par produit et région

Annexe C

Bibliographie

Bibliographie

- [1] R Core Team (2023). *R : A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [2] Harrell Jr, F. E. (2023). *Hmisc : Harrell Miscellaneous*. R package version 5.1-0. <https://CRAN.R-project.org/package=Hmisc>
- [3] Templ, M., Alfons, A., Kowarik, A., & Prantner, B. (2023). *VIM : Visualization and Imputation of Missing Values*. R package version 6.2-0. <https://CRAN.R-project.org/package=VIM>
- [4] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2023). *e1071 : Misc Functions of the Department of Statistics, Probability Theory Group (Formerly : E1071), TU Wien*. R package version 1.7-13. <https://CRAN.R-project.org/package=e1071>
- [5] Wickham, H., & Grolemund, G. (2017). *R for Data Science*. O'Reilly Media, Inc.
- [6] Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications Ltd.

Annexe D

Glossaire

Analyse univariée Étude d'une seule variable à la fois pour en décrire les caractéristiques principales.

Analyse bivariée Étude des relations entre deux variables pour identifier d'éventuelles corrélations ou dépendances.

Boxplot Représentation graphique montrant la médiane, les quartiles et les valeurs extrêmes d'une distribution.

Corrélation de Spearman Mesure de la relation monotone entre deux variables basée sur les rangs.

Imputation Remplacement des valeurs manquantes par des valeurs estimées.

Outlier Valeur aberrante qui s'écarte significativement des autres observations.

p-value Probabilité d'obtenir des résultats au moins aussi extrêmes que ceux observés sous l'hypothèse nulle.

Test de Kruskal-Wallis Test non paramétrique pour comparer plus de deux groupes indépendants.

Test du chi-carré Test statistique pour évaluer l'indépendance entre deux variables qualitatives.

Valeur aberrante Observation qui diffère tellement des autres observations qu'elle éveille des soupçons.