

Analyse des Clients de Cartes de Crédit

Chargement des Données

```
tab <- read.csv(file = "../data/credit-card-customers/BankChurners.csv",  
               sep = ',',  
               dec = '.',  
               header = TRUE)
```

Nettoyage des Données

```
# Vérifier les valeurs NaN  
if (sum(is.na(tab)) > 0) {  
  print(tab[is.na(tab)])  
  stop("Des valeurs NaN ont été détectées. Veuillez nettoyer les données avant de continuer.")  
}  
  
# Supprimer les lignes avec des valeurs inconnues  
tab <- tab[complete.cases(tab), ]
```

- Il n'y a pas de données NaN dans la variable tab.

Statistiques Sommaires

Variables Quantitatives

```
# Sommaire pour les variables quantitatives  
quantitative_vars <- c("CLIENTNUM", "Customer_Age", "Dependent_count",  
                      "Months_on_book", "Total_Relationship_Count", "Months_Inactive_12_mon",  
                      "Contacts_Count_12_mon", "Credit_Limit", "Total_Revolving_Bal",  
                      "Avg_Open_To_Buy", "Total_Amt_Chng_Q4_Q1", "Total_Trans_Amt",  
                      "Total_Trans_Ct", "Total_Ct_Chng_Q4_Q1", "Avg_Utilization_Ratio")  
  
summaries_data <- summary(tab[, quantitative_vars])  
  
kable(summaries_data, "latex", booktabs = TRUE) %>%  
  kable_styling(  
    full_width = FALSE,  
    latex_options = c("striped", "scale_down")
```

```
) %>%
row_spec(0, background = "#3498db")
```

```
# Tableau de fréquence pour les variables catégoriques
tables_data <- list(
  Attrition_Flag = table(tab$Attrition_Flag),
  Gender = table(tab$Gender),
  Education_Level = table(tab$Education_Level),
  Marital_Status = table(tab$Marital_Status),
  Income_Category = table(tab$Income_Category),
  Card_Category = table(tab$Card_Category)
)
tables_data
```

2

Var1	Freq						
Attrited Customer	1627						
Existing Customer	8500						

Var1	Freq	Var1	Freq	Var1	Freq	Var1	Freq
F	5358	College	1013	Divorced	748	\$120K +	727
M	4769	Doctorate	451	Married	4687	\$40K - \$60K	1790
		Graduate	3128	Single	3943	\$60K - \$80K	1402
		High School	2013	Unknown	749	\$80K - \$120K	1535
		Post-Graduate	516			Less than \$40K	3561
		Uneducated	1487			Unknown	1112
		Unknown	1519				

Var1	Freq
Blue	9436
Gold	116
Platinum	20
Silver	555

```
##
##      Blue      Gold Platinum      Silver
##      9436       116        20       555
```

```
kable(tables_data, "latex", booktabs = TRUE) %>%
  kable_styling(
    full_width = FALSE,
    latex_options = c("striped", "scale_down")
  ) %>%
  row_spec(0, background = "#3498db")
```

Visualisation des Données

Boxplots pour les Variables Quantitatives

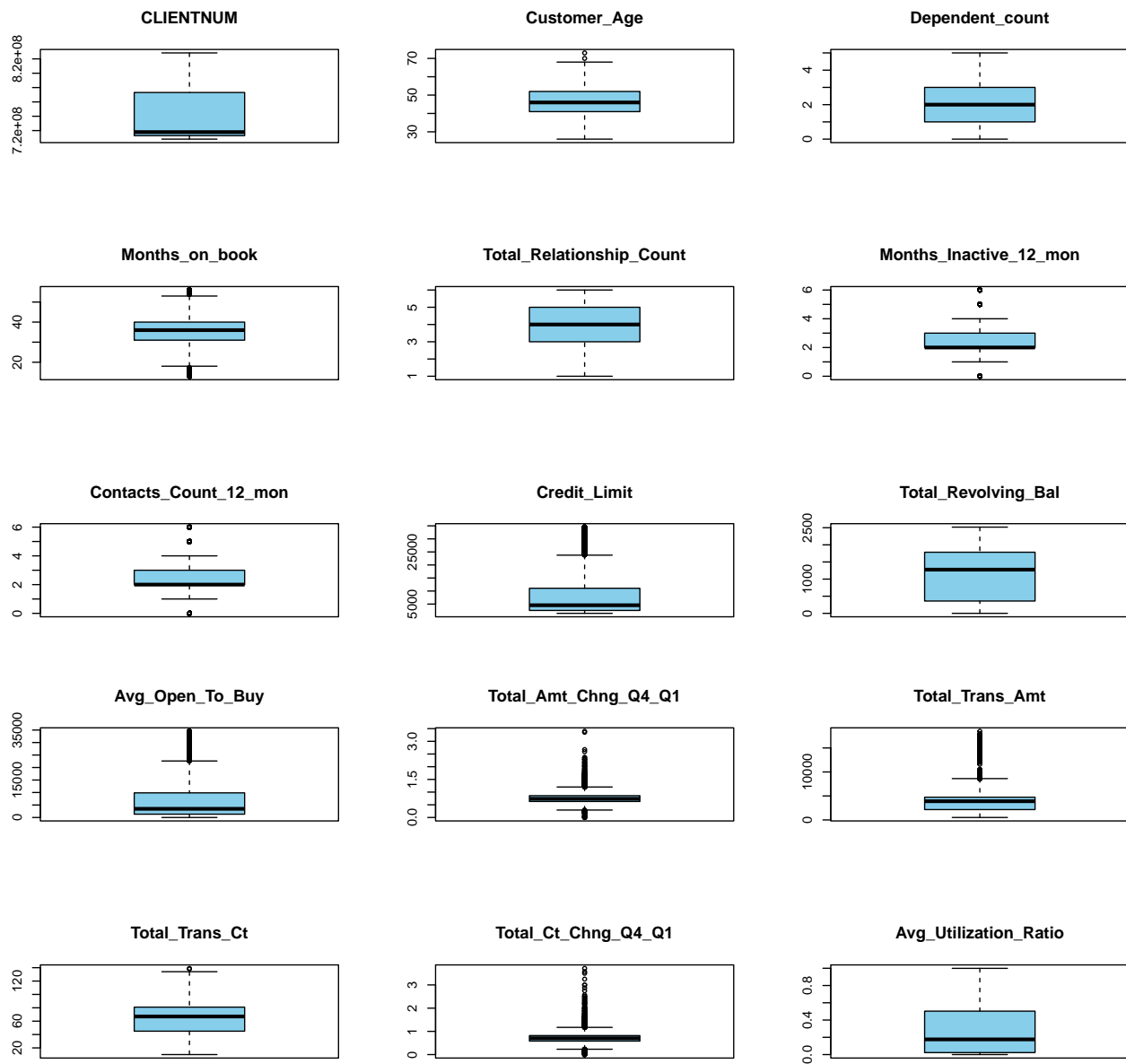
```
# Boxplots pour les Variables Quantitatives
par(mfrow = c(3, 3)) # Ajustez la grille selon vos préférences

for (i in 1:length(quantitative_vars)) {
```

```

boxplot(tab[, quantitative_vars[i]], main = quantitative_vars[i], col = "skyblue", border = "black")
}

```



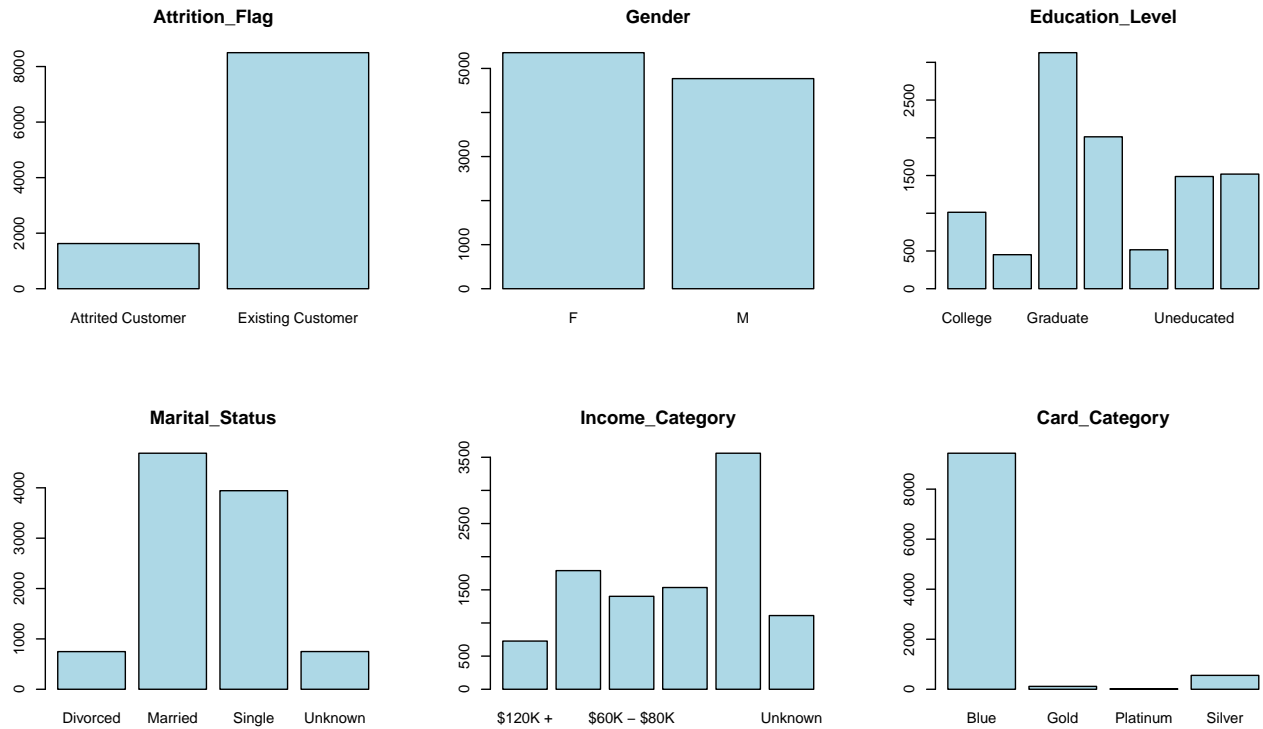
Histogrammes pour les Variables Catégoriques

```

# Histogrammes pour les Variables Catégoriques
par(mfrow = c(2, 3)) # Ajustez la grille selon vos préférences

for (j in 1:length(names(tables_data))) {
  barplot(tables_data[[j]], main = names(tables_data)[j], col = "lightblue")
}

```

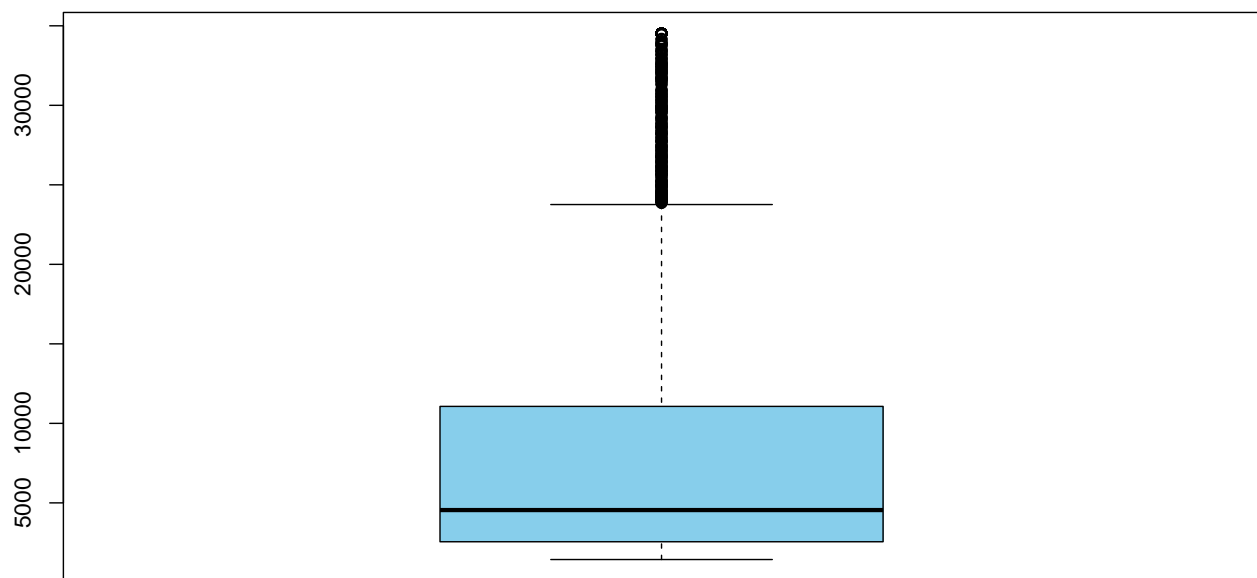


Vérification et Gestion des Valeurs Aberrantes

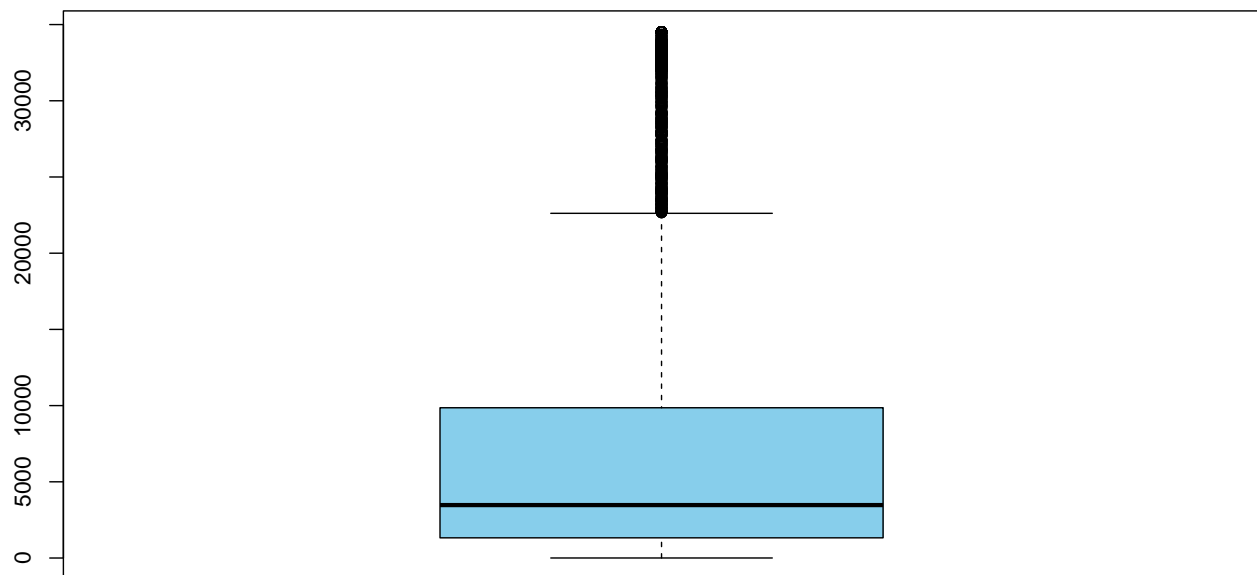
```
# Vérification des valeurs aberrantes
quantitative_vars_to_check <- c("Credit_Limit", "Avg_Open_To_Buy", "Total_Amt_Chng_Q4_Q1", "Total_Trans")

for (var in quantitative_vars_to_check) {
  boxplot(tab[, var], main = var, col = "skyblue", border = "black")
  # Ajouter le code pour gérer les valeurs aberrantes si nécessaire
}
```

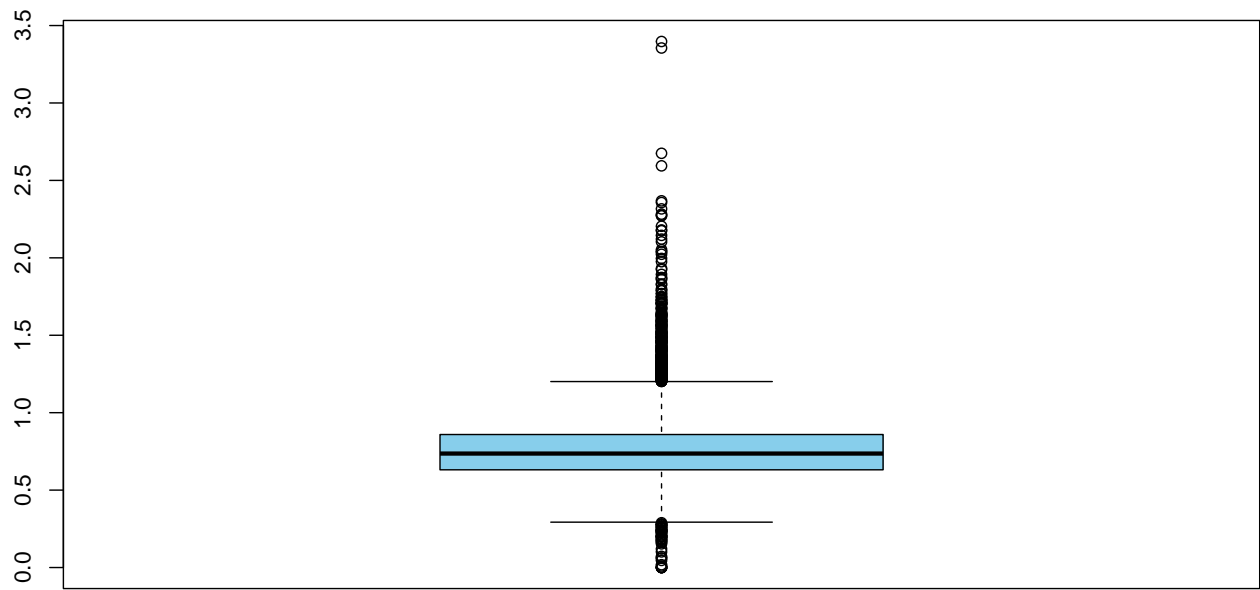
Credit_Limit



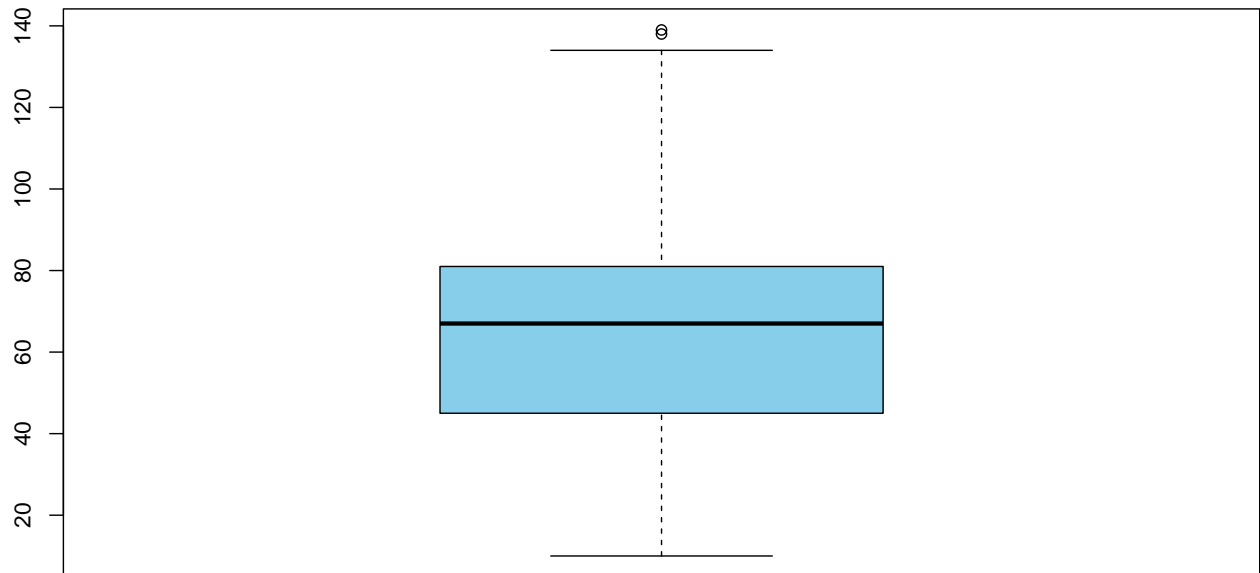
Avg_Open_To_Buy

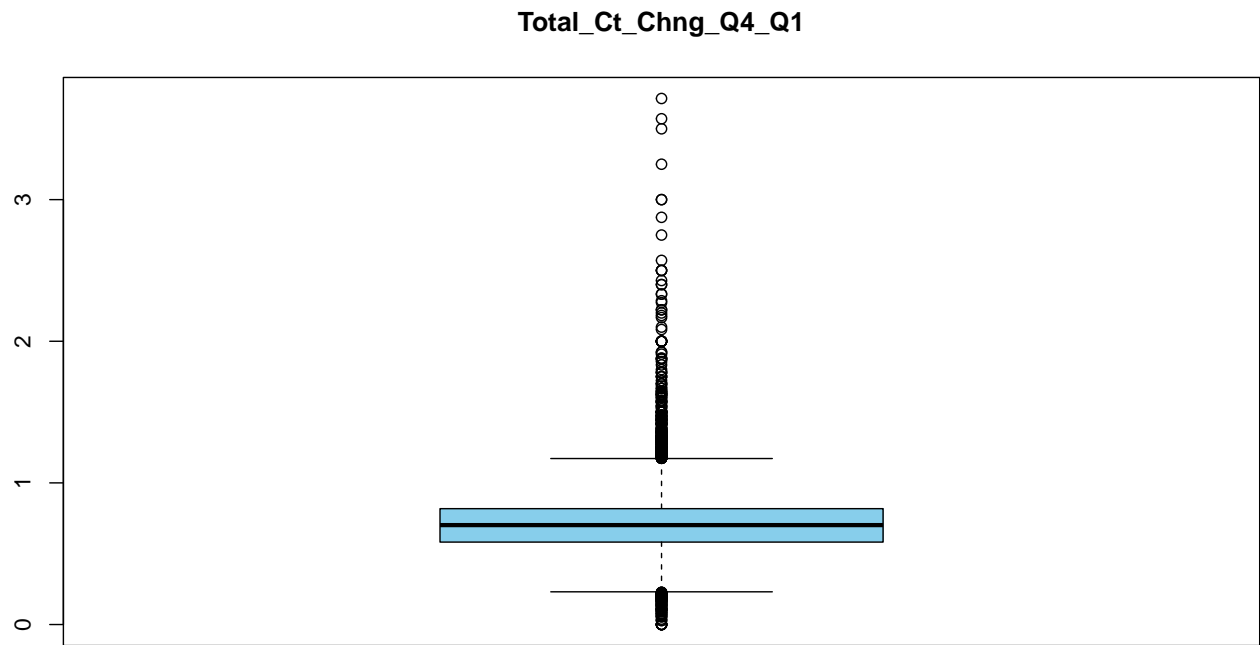


Total_Amt_Chng_Q4_Q1



Total_Trans_Ct





Conclusion

Ce document fournit une analyse initiale du jeu de données des clients de cartes de crédit. Une exploration et une analyse plus approfondies peuvent être nécessaires en fonction des objectifs spécifiques de votre projet.