Analyse des Clients de Cartes de Crédit

Chargement des Données

```
# Lire le jeu de données
tab <- read.csv(file = "./data/credit-card-customers/BankChurners.csv", sep = ',', dec = '.', header = T
```

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), ]
```

Statistiques Sommaires

Variables Quantitatives

```
# Sommaire pour les variables quantitatives
quantitative_vars <- c("CLIENTNUM", "Customer_Age", "Dependent_count", "Months_on_book", "Total_Relation
                       "Months_Inactive_12_mon", "Contacts_Count_12_mon", "Credit_Limit", "Total_Revolvi
                       "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])
summaries_data
```

```
##    CLIENTNUM          Customer_Age   Dependent_count Months_on_book
##  Min.   :708082083   Min.   :26.00   Min.   :0.000   Min.   :13.00
##  1st Qu.:713036770   1st Qu.:41.00   1st Qu.:1.000   1st Qu.:31.00
##  Median :717926358   Median :46.00   Median :2.000   Median :36.00
##  Mean   :739177606   Mean   :46.33   Mean   :2.346   Mean   :35.93
##  3rd Qu.:773143533   3rd Qu.:52.00   3rd Qu.:3.000   3rd Qu.:40.00
##  Max.   :828343083   Max.   :73.00   Max.   :5.000   Max.   :56.00
##  Total_Relationship_Count Months_Inactive_12_mon Contacts_Count_12_mon
##  Min.   :1.000            Min.   :0.000          Min.   :0.000
```

```
##  1st Qu.:3.000            1st Qu.:2.000          1st Qu.:2.000
##  Median :4.000            Median :2.000          Median :2.000
##  Mean   :3.813            Mean   :2.341          Mean   :2.455
##  3rd Qu.:5.000            3rd Qu.:3.000          3rd Qu.:3.000
##  Max.   :6.000            Max.   :6.000          Max.   :6.000
##   Credit_Limit   Total_Revolving_Bal Avg_Open_To_Buy Total_Amt_Chng_Q4_Q1
##  Min.   : 1438   Min.   :   0        Min.   :    3   Min.   :0.0000
##  1st Qu.: 2555   1st Qu.: 359        1st Qu.: 1324   1st Qu.:0.6310
##  Median : 4549   Median :1276        Median : 3474   Median :0.7360
##  Mean   : 8632   Mean   :1163        Mean   : 7469   Mean   :0.7599
##  3rd Qu.:11068   3rd Qu.:1784        3rd Qu.: 9859   3rd Qu.:0.8590
##  Max.   :34516   Max.   :2517        Max.   :34516   Max.   :3.3970
##  Total_Trans_Amt Total_Trans_Ct   Total_Ct_Chng_Q4_Q1 Avg_Utilization_Ratio
##  Min.   :  510   Min.   : 10.00   Min.   :0.0000      Min.   :0.0000
##  1st Qu.: 2156   1st Qu.: 45.00   1st Qu.:0.5820      1st Qu.:0.0230
##  Median : 3899   Median : 67.00   Median :0.7020      Median :0.1760
##  Mean   : 4404   Mean   : 64.86   Mean   :0.7122      Mean   :0.2749
##  3rd Qu.: 4741   3rd Qu.: 81.00   3rd Qu.:0.8180      3rd Qu.:0.5030
##  Max.   :18484   Max.   :139.00   Max.   :3.7140      Max.   :0.9990
```

Variables Catégoriques

```r
# 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
```

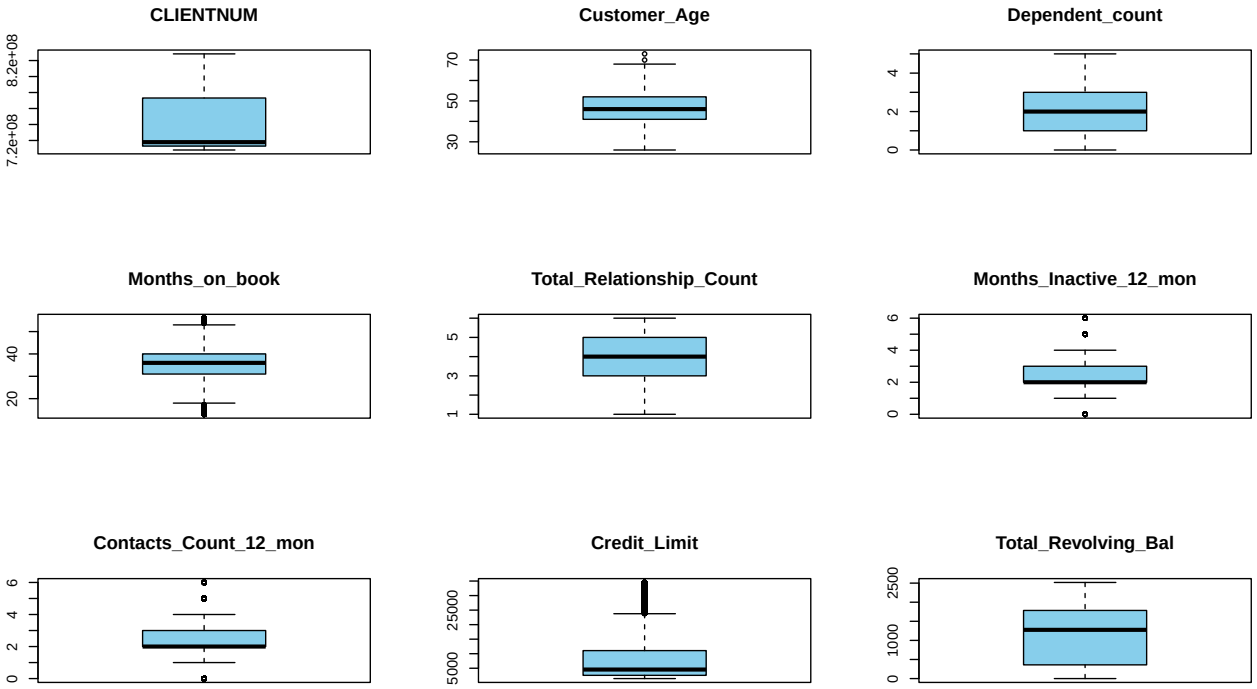
```
## $Attrition_Flag
##
## Attrited Customer Existing Customer
##              1627              8500
##
## $Gender
##
##    F    M
## 5358 4769
##
## $Education_Level
##
##       College     Doctorate      Graduate   High School Post-Graduate
##          1013           451          3128          2013           516
##    Uneducated       Unknown
##          1487          1519
##
## $Marital_Status
##
```

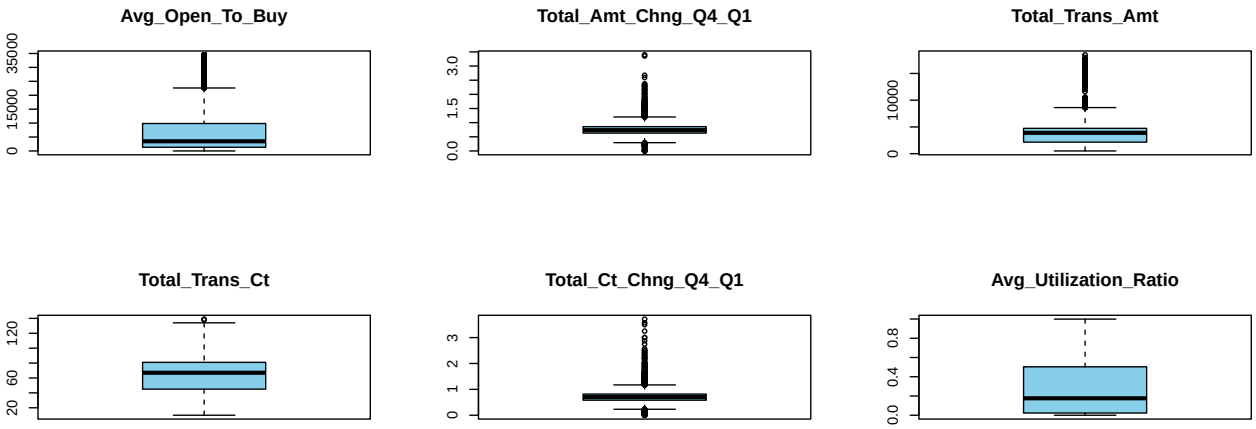
```
## Divorced  Married   Single  Unknown
##      748     4687     3943      749
##
## $Income_Category
##
##        $120K +    $40K - $60K    $60K - $80K   $80K - $120K Less than $40K
##            727           1790           1402           1535           3561
##        Unknown
##           1112
##
## $Card_Category
##
##     Blue     Gold Platinum   Silver
##     9436      116       20      555
```

Visualisation des Données

Boxplots pour les Variables Quantitatives

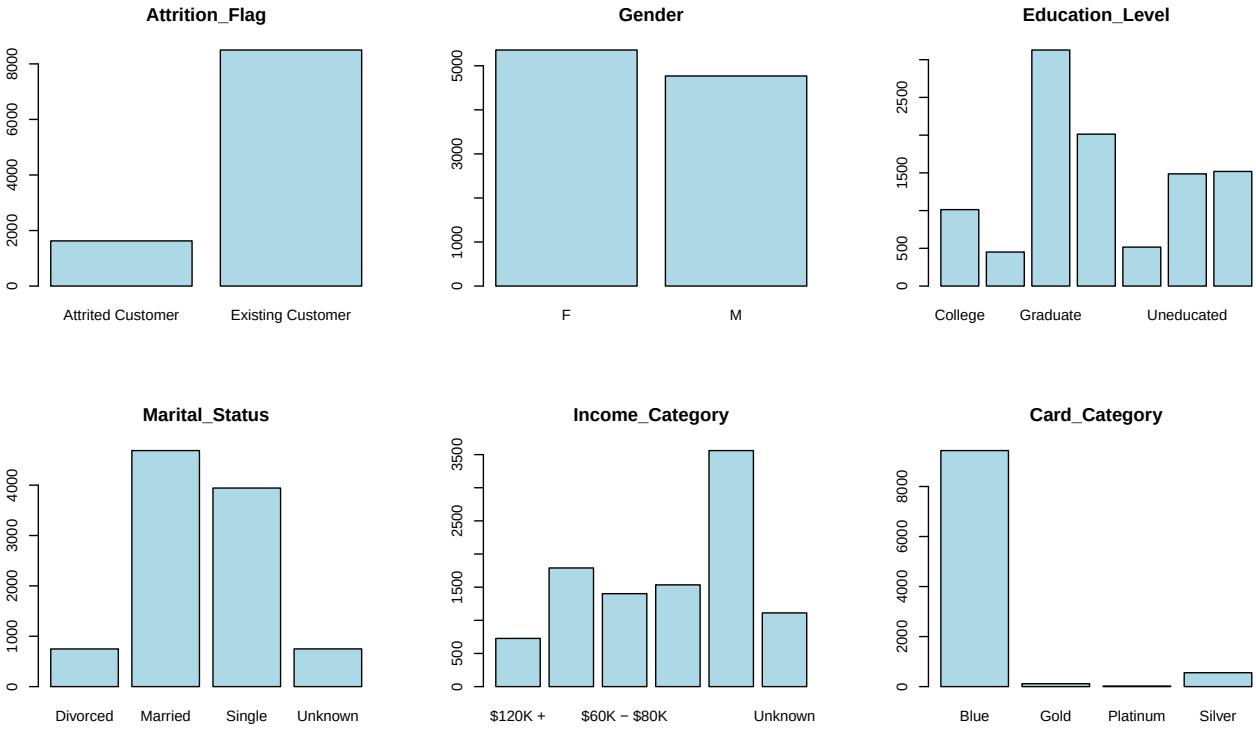
```r
# Boxplots pour les Variables Quantitatives
par(mfrow = c(3, 3))  # Ajustez la grille selon vos préférences

for (i in 1:length(quantitative_vars)) {
  boxplot(tab[, quantitative_vars[i]], main = quantitative_vars[i], col = "skyblue", border = "black")
}
```





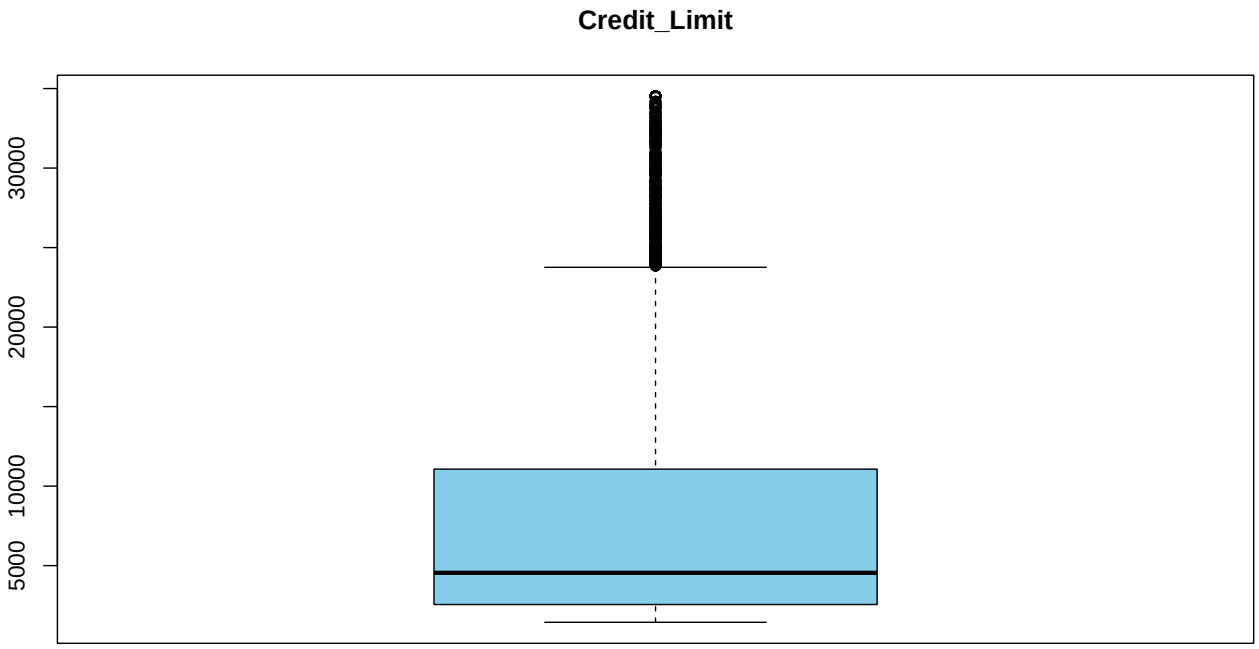
Histogrammes pour les Variables Catégoriques

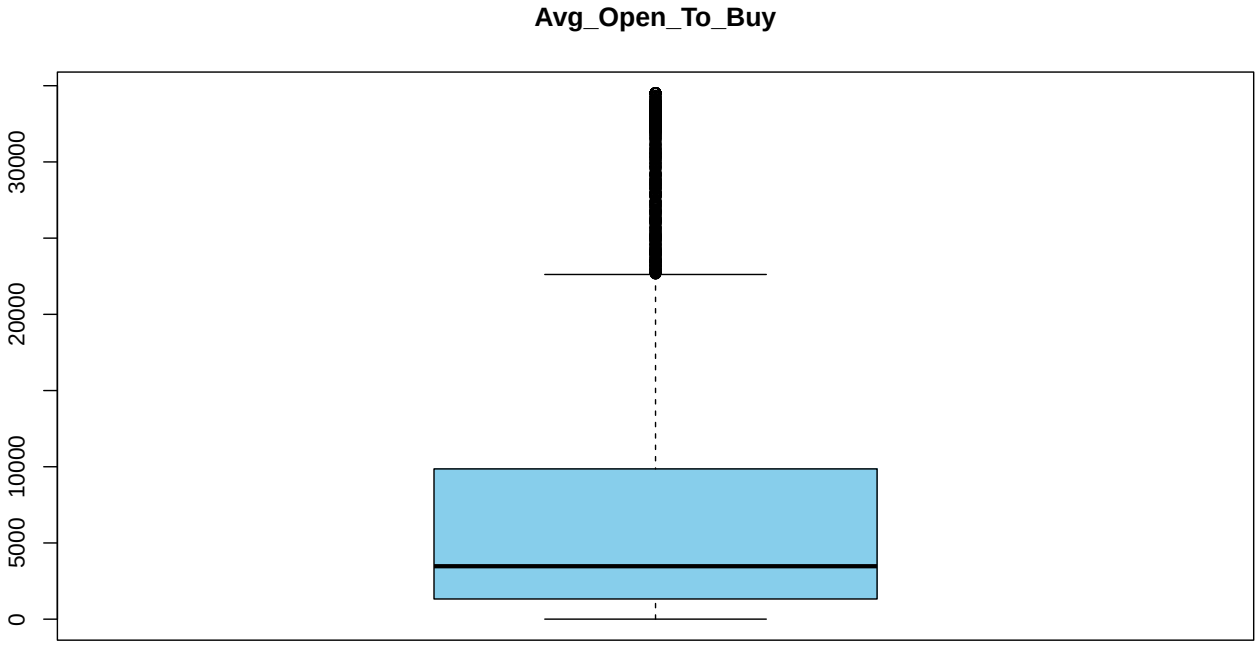
```r
# Histogrammes pour les Variables Catégoriques
par(mfrow = c(2, 3))  # Ajustez la grille selon vos préférences

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

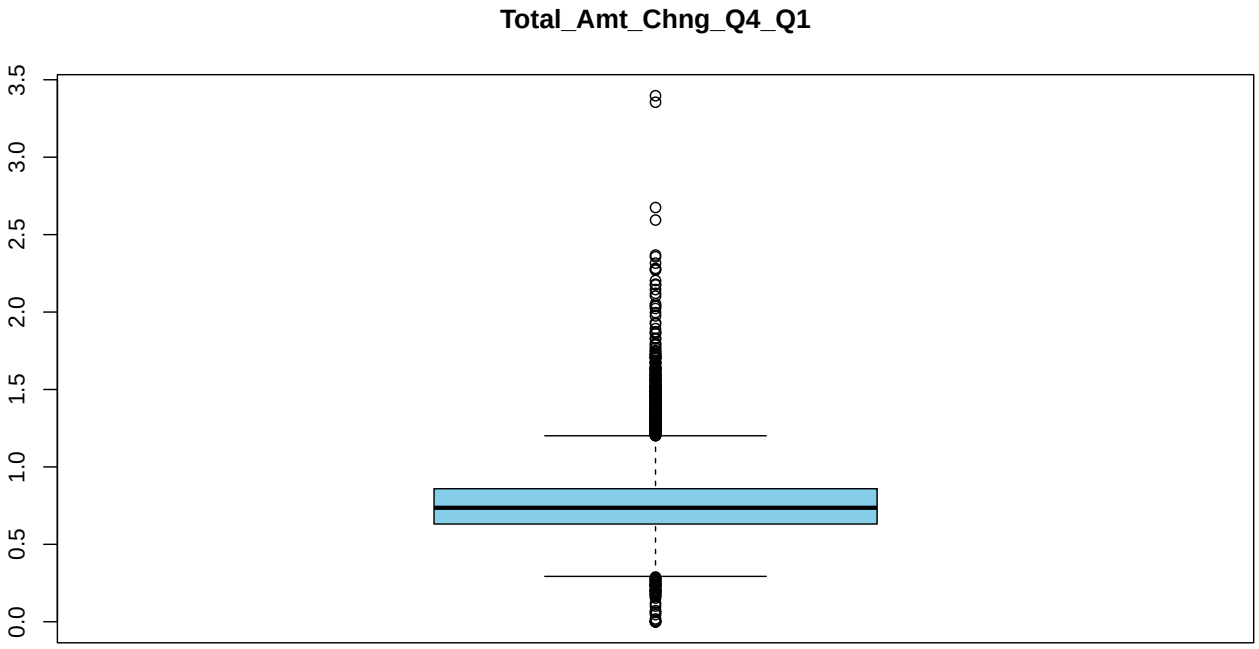


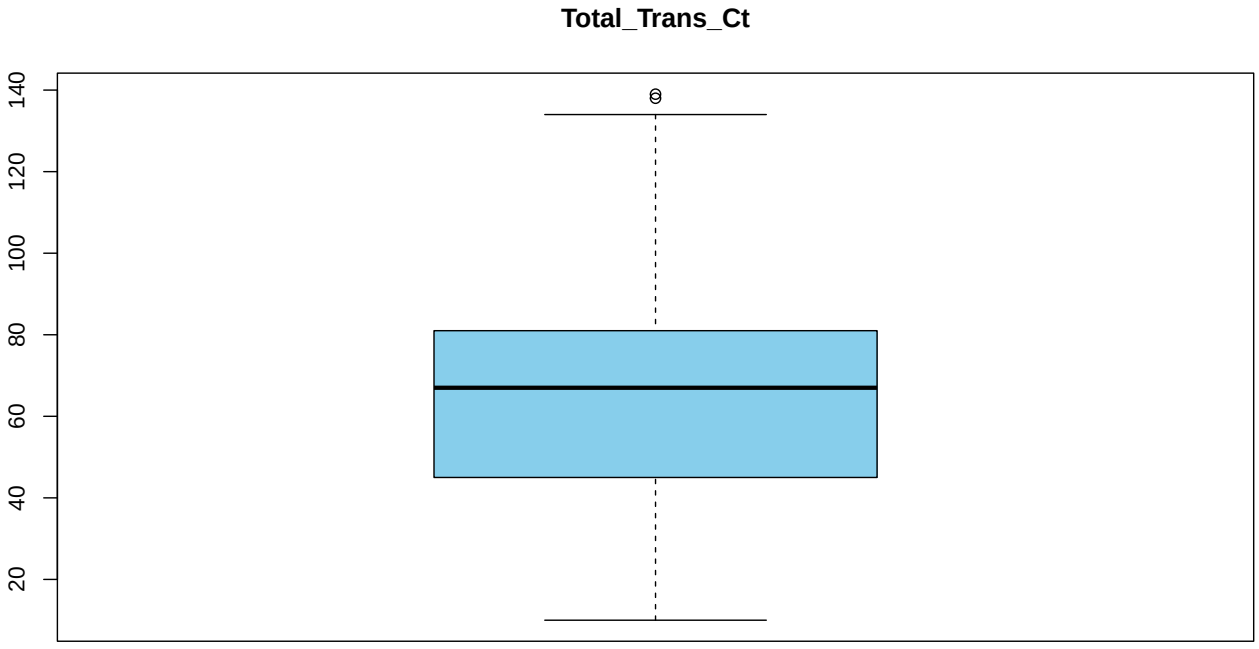
Vérification et Gestion des Valeurs Aberrantes

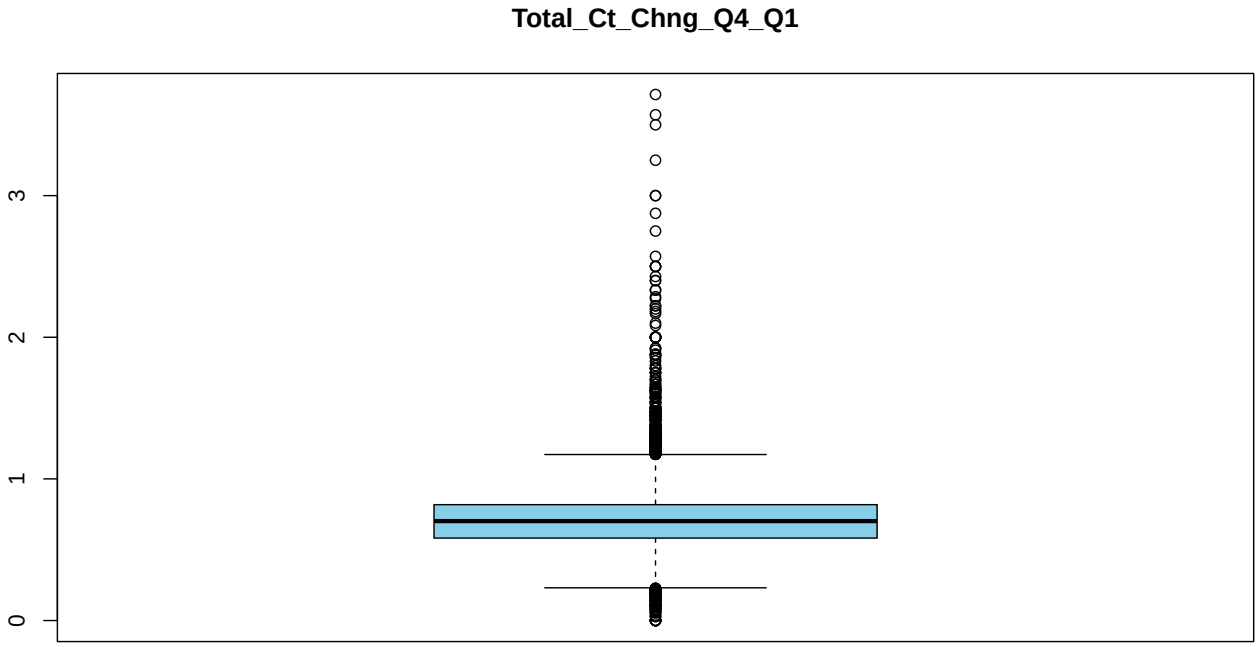
```r
# Vérification des valeurs aberrantes
quantitative_vars_to_check <- c("Credit_Limit", "Avg_Open_To_Buy", "Total_Amt_Chng_Q4_Q1", "Total_Trans_

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
}
```






Total_Amt_Chng_Q4_Q1
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5


Total_Trans_Ct
20
40
60
80
100
120
140



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.