**Rapport**

**ANALYSE UNIVARIEE**

**Code univariéé**

**Vérifier la structure des données**

str(bbd) summary(bbd) # Statistiques descriptives générales

**Histogrammes pour chaque variable numérique**

for (col in colnames(bbd)) { if (is.numeric(bbd[[col]])) { print(ggplot(bbd, aes(x = .data[[col]])) + geom\_histogram(bins = 30, fill = "blue", alpha = 0.5) + ggtitle(paste("Histogramme de", col))) } }

**Boxplots pour détecter les valeurs extrêmes (outliers)**

for (col in colnames(bbd)) { if (is.numeric(bbd[[col]])) { print(ggplot(bbd, aes(y = .data[[col]])) + geom\_boxplot(fill = "red", alpha = 0.5) + ggtitle(paste("Boxplot de", col))) } }

**Mesures de dispersion et de forme**

for (col in colnames(bbd)) { if (is.numeric(bbd[[col]])) { cat("\nVariable:", col) cat("\nMoyenne:", mean(bbd[[col]], na.rm = TRUE)) cat("\nMédiane:", median(bbd[[col]], na.rm = TRUE)) cat("\nVariance:", var(bbd[[col]], na.rm = TRUE)) cat("\nÉcart-type:", sd(bbd[[col]], na.rm = TRUE)) cat("\nSkewness (Asymétrie):", skewness(bbd[[col]], na.rm = TRUE)) cat("\nKurtosis:", kurtosis(bbd[[col]], na.rm = TRUE)) cat("\n----------------------------\n") } }

**Test de normalité de Shapiro-Wilk pour chaque variable numérique**

for (col in colnames(bbd)) { if (is.numeric(bbd[[col]])) { test <- shapiro.test(bbd[[col]]) cat("\nTest de normalité pour", col, ": p-value =", test$p.value, "\n") } } le code R

**ANALYSE BIVARIEE ( les variables explicatives en fonction de la variable cible)**

les variables explicatives (8) sont les suivantes:  
-Cement(kg/m³) : Quantité de ciment utilisée par mètre cube de mélange.

Il existe une corrélation de entre le ciment et la résistance du béton avec une valeur de 0.498. cela implique qu’elle n’est pas trop significative.

-Blast Furnace Slag(kg/m³) : Quantité de laitier de haut fourneau utilisé comme substitut partiel du ciment.

-Fly Ash(kg/m³) : Quantité de cendres volantes utilisées comme substitut partiel du ciment.

Il existe une corrélation négative entre cette variable et la résistance du béton ce qui peut signifier une réduction de 0.10 de la résistance à la compression du béton quand cette variable augmente de 1.

-Water(kg/m³) : Quantité d’eau utilisée pour le mélange du béton.

La correlation est de – 0.290. la relation est faible entre l’eau et la résistance du béton. Cette variable n’est pas trop significative.

-Superplasticizer(kg/m³) : Additif chimique améliorant la fluidité et la maniabilité du béton.

La correlation est de 0.366 ce qui signifie qu’une augmentation engendre une augmentation de la resistance de 30%.

-Coarse Aggregate(kg/m³) : Quantité de granulats grossiers (graviers) utilisés dans le mélange.

-Fine Aggregate(kg/m³) : Quantité de granulats fins (sable) utilisés dans le mélange.

-Age : Nombre de jours depuis le coulage du béton avant la mesure de la résistance.

La correlation est de 0.329. c’est -à-dire la résistance du béton augmente avec les années.  
  
  
La variable cible est :  
-Concrete Compressive Strength (MPa) : Résistance à la compression du béton mesurée en mégapascals (MPa).

**Code analyse bivarieé**

# Installer et charger le package corrplot  
install.packages("corrplot")  
library(corrplot)  
  
# Charger tes données (assure-toi que Concrete\_Data est un data frame)  
bdd <- Concrete\_Data  
  
# Vérifier les premières lignes des données  
head(bdd)  
  
# Vérifier si toutes les colonnes sont numériques  
str(bdd)  
  
bdd\_numeric <- bdd[, sapply(bdd, is.numeric)]  
  
# Calcul de la matrice de corrélation  
cor\_matrix <- cor(bdd\_numeric, use = "complete.obs")  
  
#matrice de corrélation  
print(cor\_matrix)