## 1. Càrrega de dades:

**df <- read.csv("C:/Users/pere amat/CE/PIA/PRACTICA1/salary\_prediction\_data.csv")**

## 2. Exploració inicial:

# Verificar la estructura del conjunto de datos

**str(df)**

# Resumen estadístico de las columnas numéricas

**summary(df)**

**# Ver los nombres de las columnas**

**colnames(df)**

**# Verificar cuántos valores faltantes hay por columna**

**colSums(is.na(df))**

## 3. Visualización univariada:

### histogramas

**hist(df$Salary, main="Histograma de Salarios",**

**xlab="Salarios", col="blue", breaks=10)**

### GRÁFICO DE BARRAS

# Gráfico de barras de location

**barplot(table(df$Location),**

**main="Frecuencia de Location",**

**xlab="Localización", ylab="Frecuencia",**

**col="red", las=2)**

### boxplot

**# Crear Boxplot de Salarios**

**boxplot(df$Salary, main="Boxplot de Salarios",**

**ylab="Salarios", col="lightblue",**

**outline = TRUE) # Asegúrate de que los outliers se dibujen**

**# Obtener los valores de los outliers**

**outliers <- boxplot.stats(df$Salary)$out**

**cat("Outliers encontrados en el Salario:", outliers, "\n")**

## 

## 4 Visualización bivariada:

### Gràfico de dispersión

**plot(df$Experience, df$Salary,**

**main="Años de Experiencia vs Salario",**

**xlab="Años de Experiencia",**

**ylab="Salario",**

**col="blue",**

**pch=16, # Cambia el tipo de punto a un círculo sólido**

**cex=1.2) # Tamaño de los puntos ligeramente aumentado**

## 5. Analisi correlación

**plot(df$Experience, df$Salary,**

**main="Años de Experiencia vs Salario",**

**xlab="Años de Experiencia",**

**ylab="Salario",**

**col="blue",**

**pch=16, # Cambia el tipo de punto a un círculo sólido**

**cex=1.2) # Tamaño de los puntos ligeramente aumentado**

## 6. codificación

**# Convertir la columna "Education" a numérico**

**df$Education <- as.numeric(factor(df$Education,**

**levels = unique(df$Education),**

**labels = 0:(length(unique(df$Education)) - 1)))**

**# Convertir la columna "Location" a numérico**

**df$Location <- as.numeric(factor(df$Location,**

**levels = unique(df$Location),**

**labels = 0:(length(unique(df$Location)) - 1)))**

**# Convertir la columna "Gender" a numérico**

**df$Gender <- as.numeric(factor(df$Gender,**

**levels = unique(df$Gender),**

**labels = 0:(length(unique(df$Gender)) - 1)))**

**# Verificar la conversión**

**head(df)**