In [13]:
```python
#Cargar Librerias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [11]:
```python
# Cargar la base de datos
import kagglehub

# Download latest version
path = kagglehub.dataset_download("uciml/pima-indians-diabetes-datal

print("Path to dataset files:", path)
```

Using Colab cache for faster access to the 'pima-indians-diabetes-da
tabase' dataset.
Path to dataset files: /kaggle/input/pima-indians-diabetes-database

In [14]:
```python
# Crear un Dataframe del Archivo
df = pd.read_csv(f"{path}/diabetes.csv")
```

In [17]:
```python
plt.figure(figsize=(5,4))
df['Outcome'].value_counts().plot(kind='bar', color=['green', 'red'
plt.title('Distribución de personas con y sin diabetes')
plt.xlabel('Outcome (0 = No, 1 = Sí)')
plt.ylabel('Cantidad de personas')
plt.show()
```
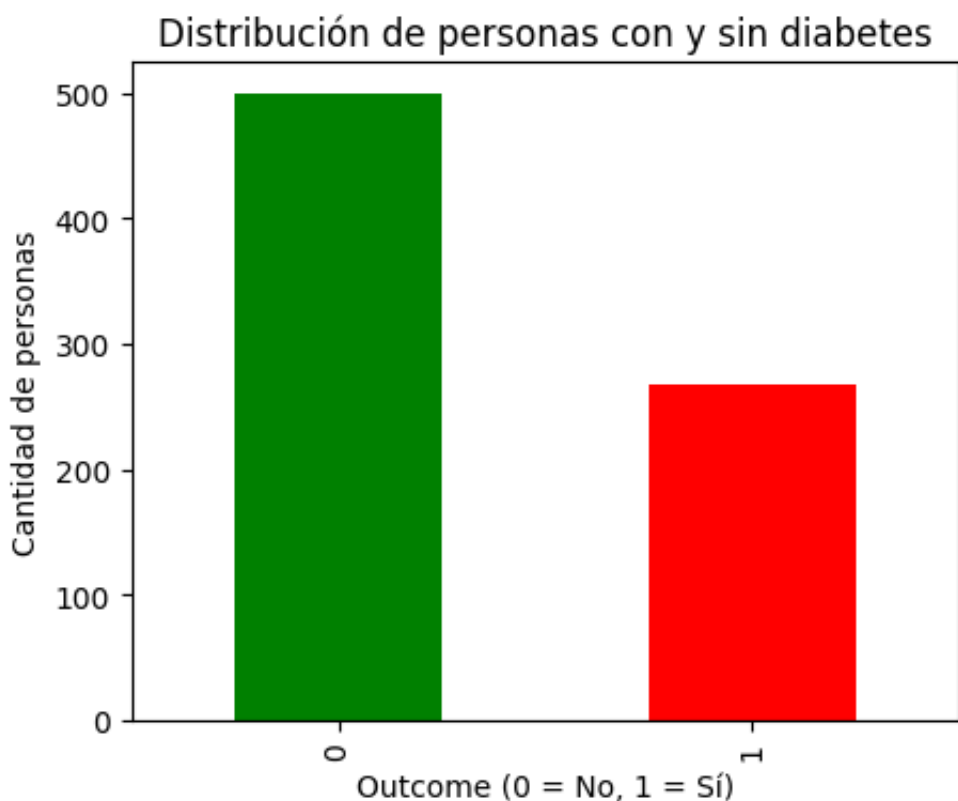

In [26]:
```python
df.groupby('Outcome')['Glucose'].mean().plot(kind='bar', color=['gr
plt.title('Promedio de Glucosa según condición de diabetes')
plt.xlabel('Outcome (0 = No, 1 = Sí)')
```

```python
plt.ylabel('Promedio de Glucosa')
plt.show()
```


Promedio de Glucosa según condición de diabetes
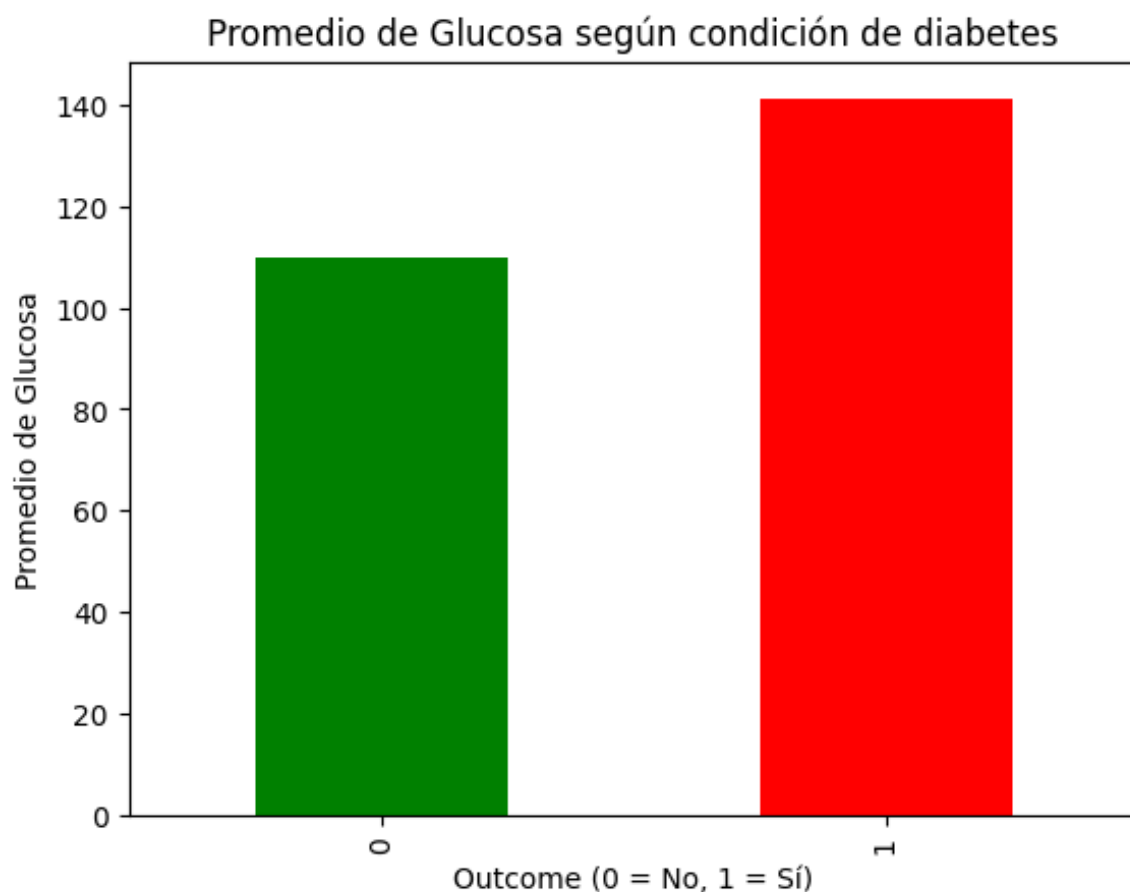
In [20]:
```python
num_cols = ['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickne
            'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age']

for col in num_cols:
    plt.figure(figsize=(6,4))
    plt.hist(df[col].dropna(), bins=10, color='skyblue', edgecolor=
    plt.title(f'Histograma de {col}')
    plt.xlabel(col)
    plt.ylabel('Frecuencia')
    plt.show()
```
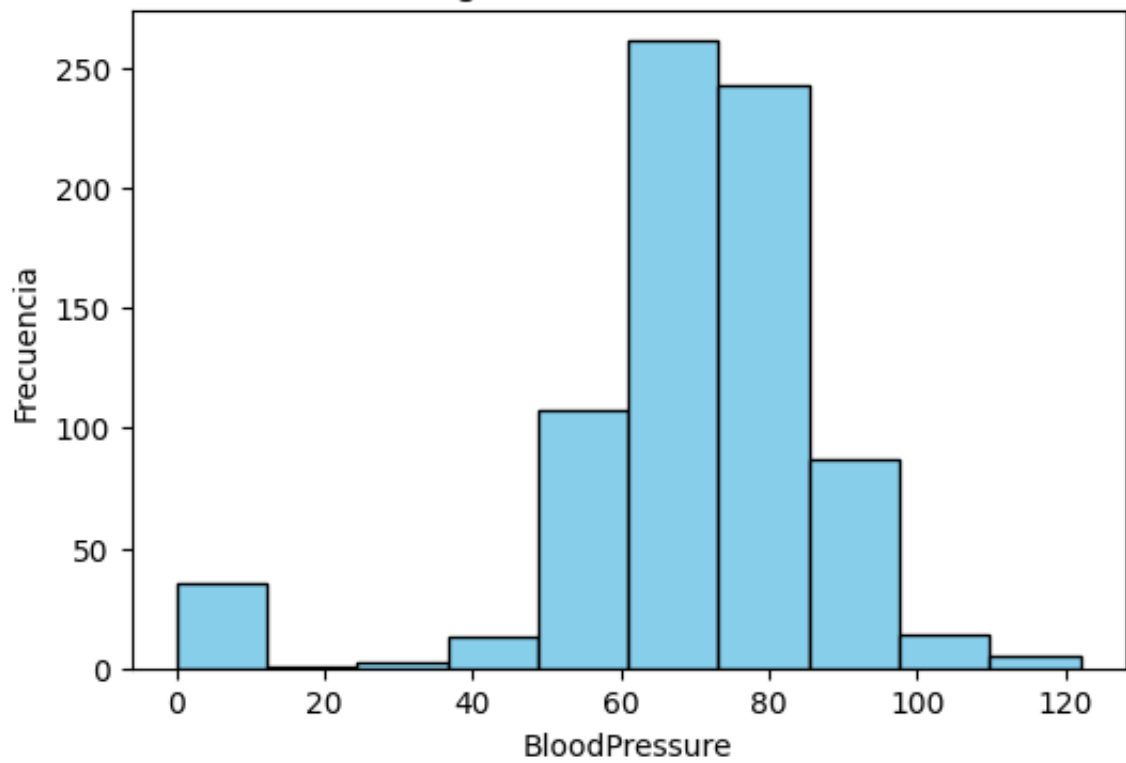
Histograma de Pregnancies
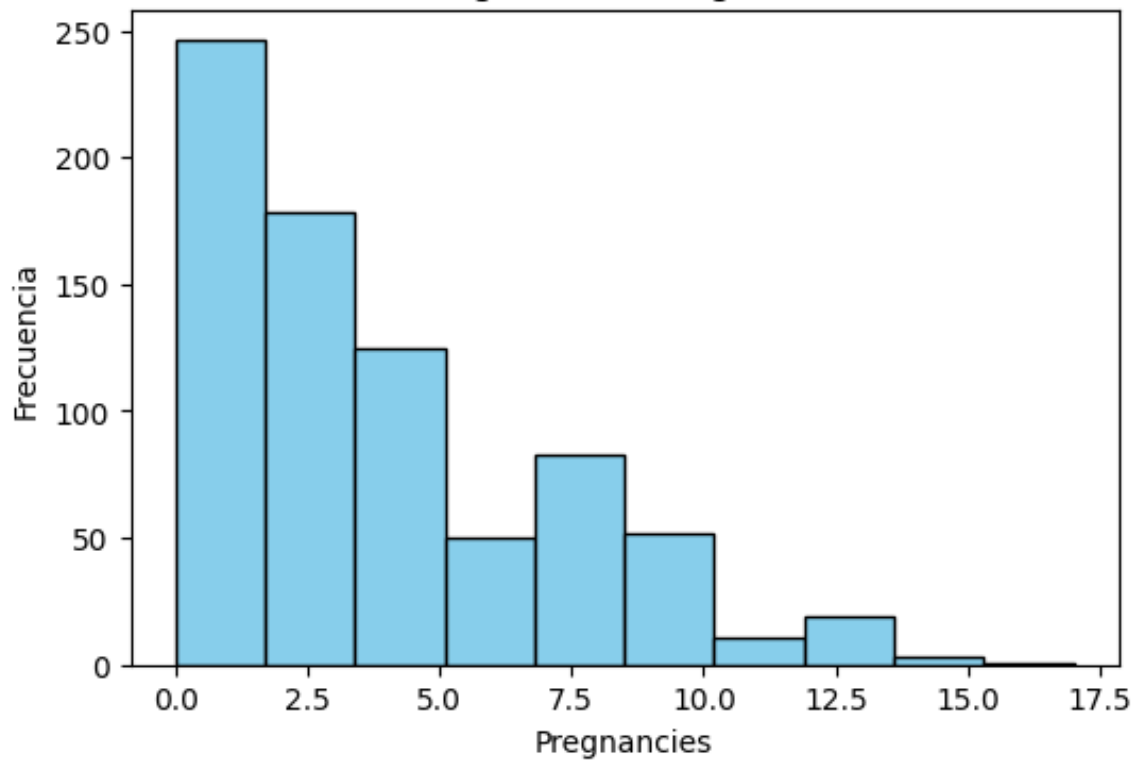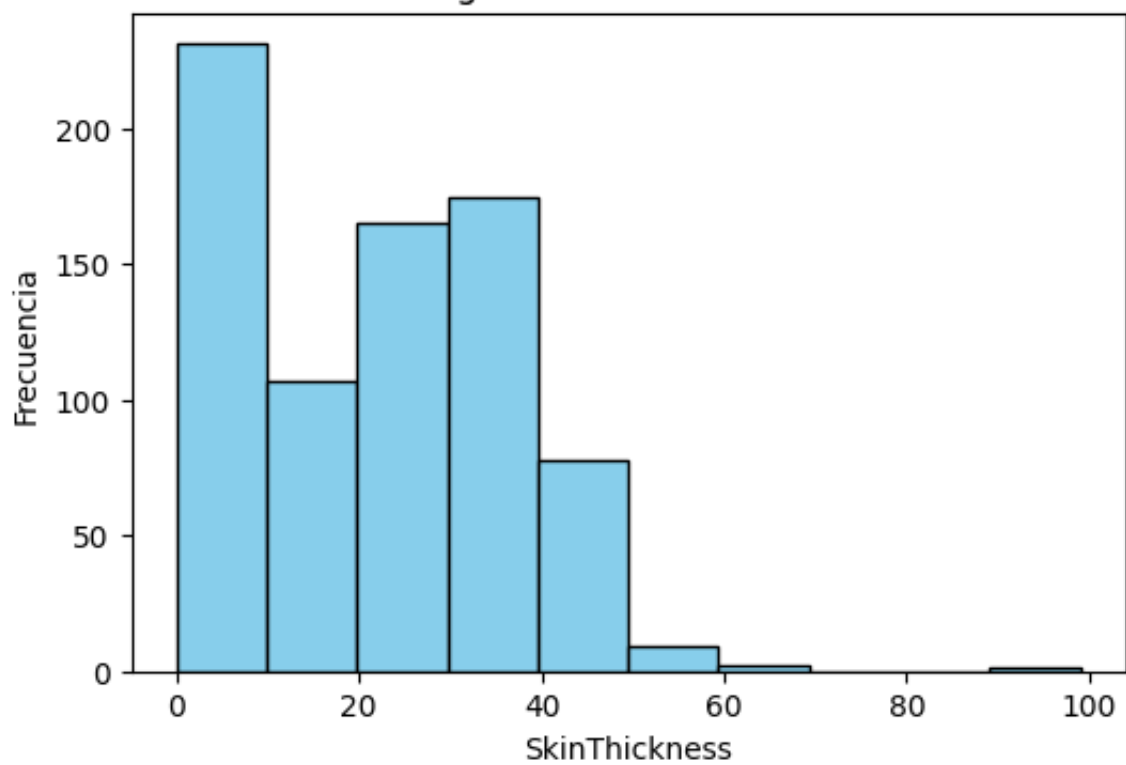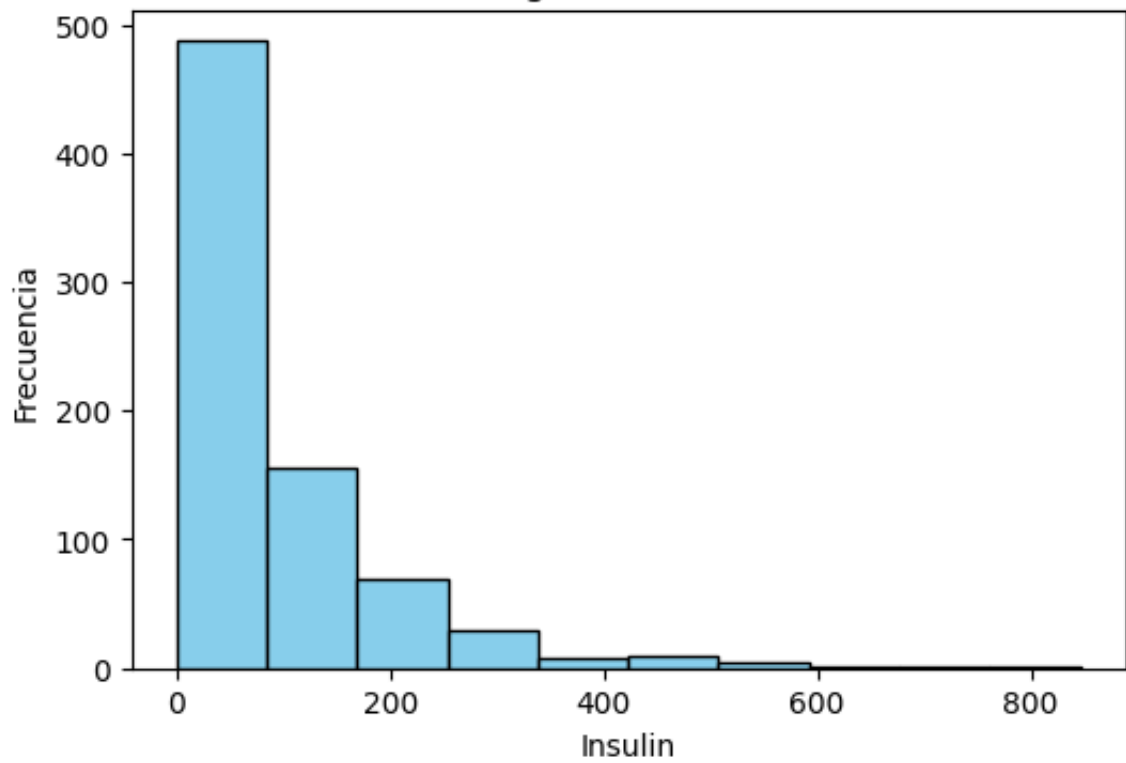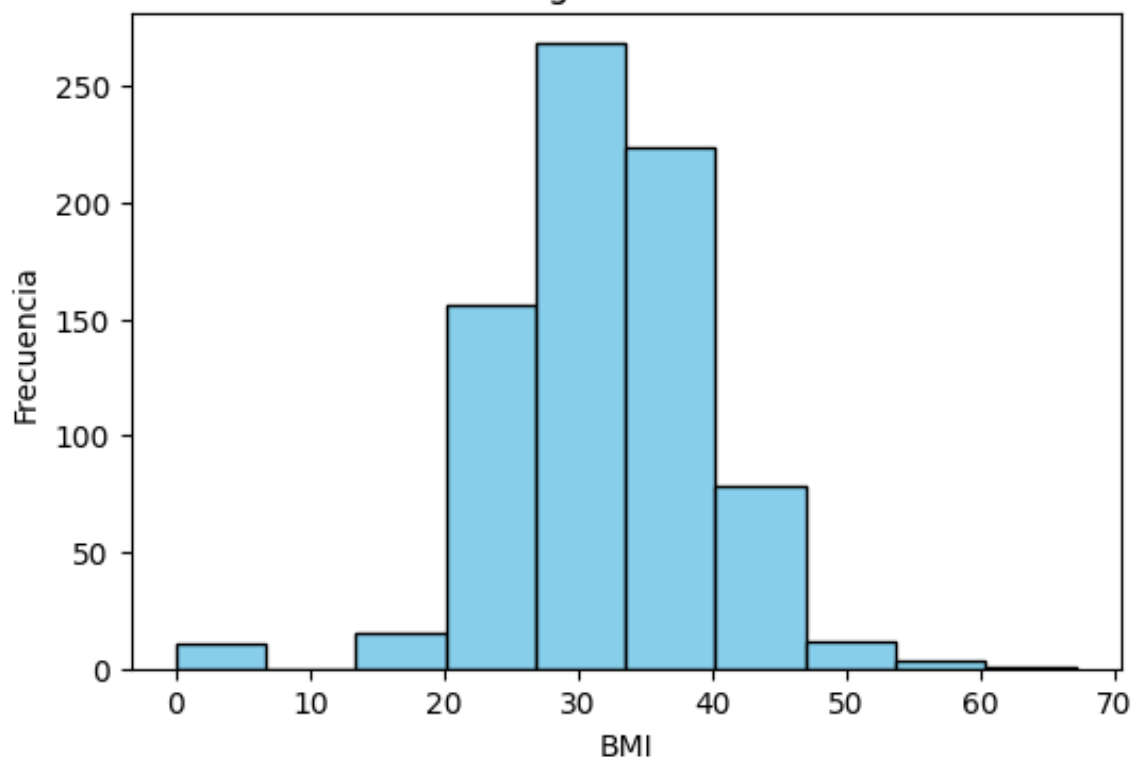

Histograma de Glucose
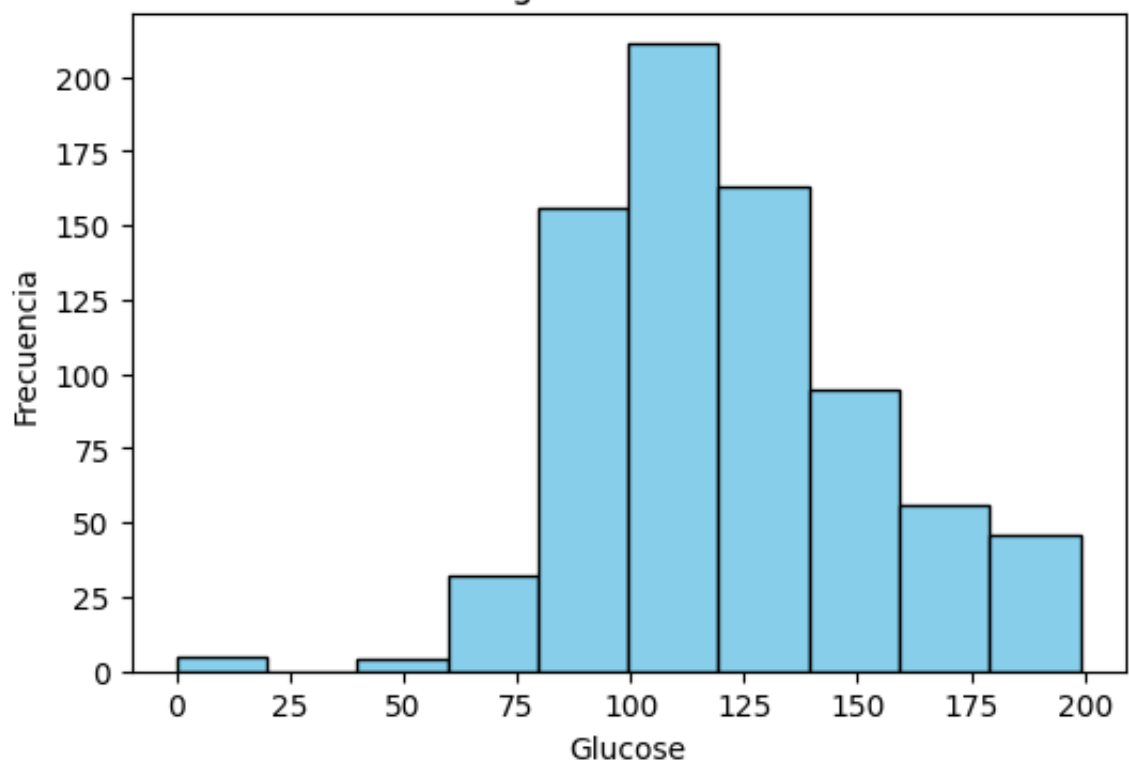
Histograma de BloodPressure

Histograma de SkinThickness

Histograma de Insulin

Histograma de BMI

Histograma de DiabetesPedigreeFunction

Histograma de Age
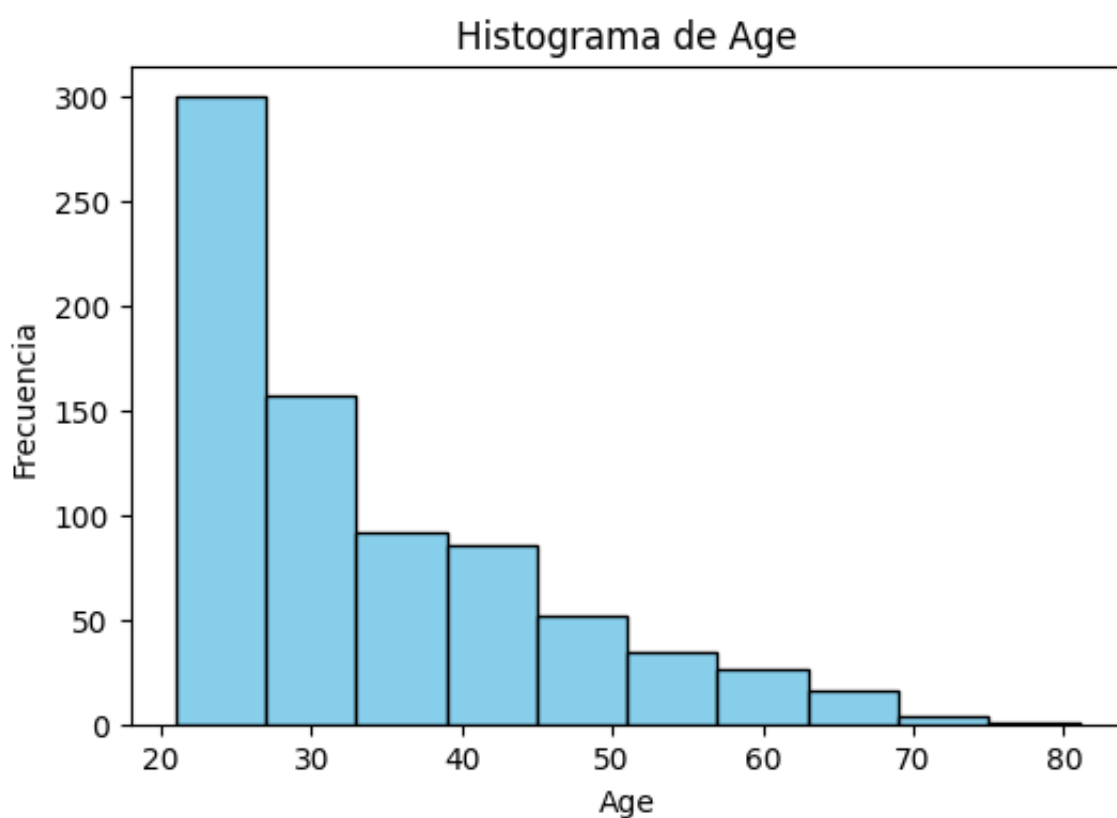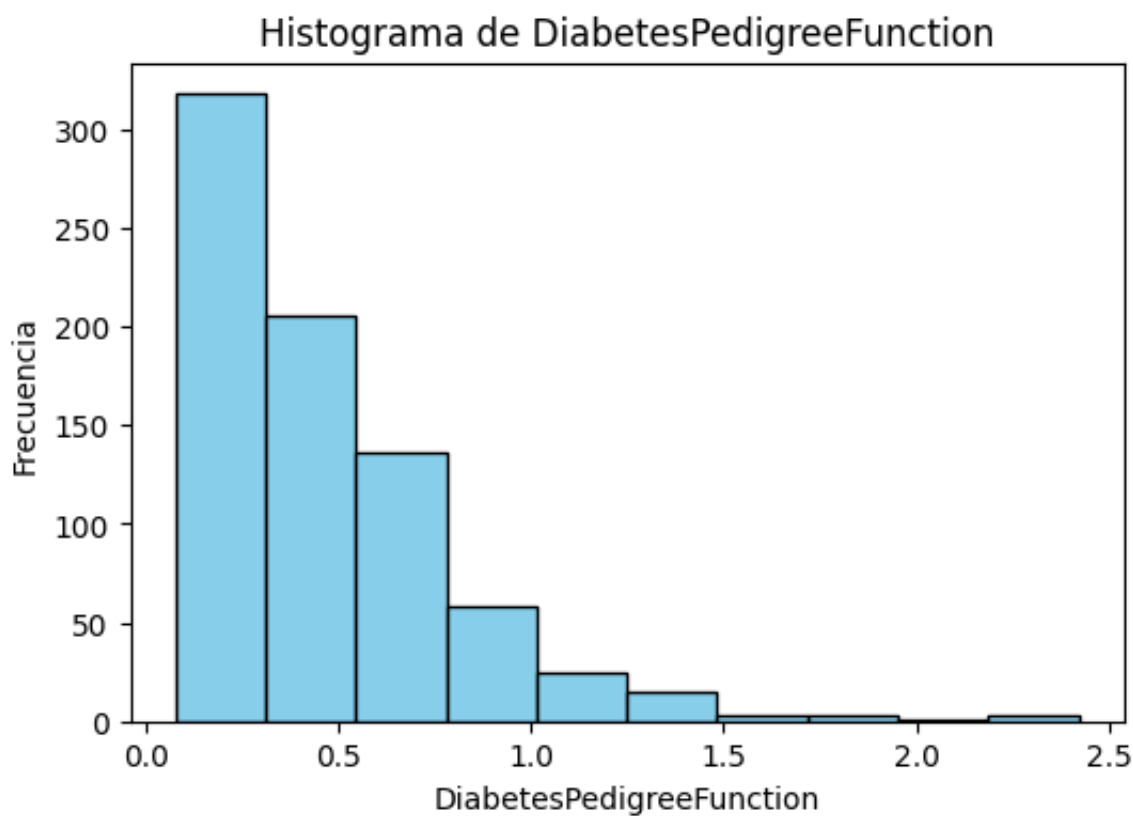
```
In [21]:  for col in num_cols:
              plt.figure(figsize=(6,4))
              plt.boxplot(df[col].dropna())
              plt.title(f'Boxplot de {col}')
              plt.ylabel(col)
              plt.show()
```

Boxplot de Pregnancies

Boxplot de Glucose

Boxplot de BloodPressure

Boxplot de SkinThickness
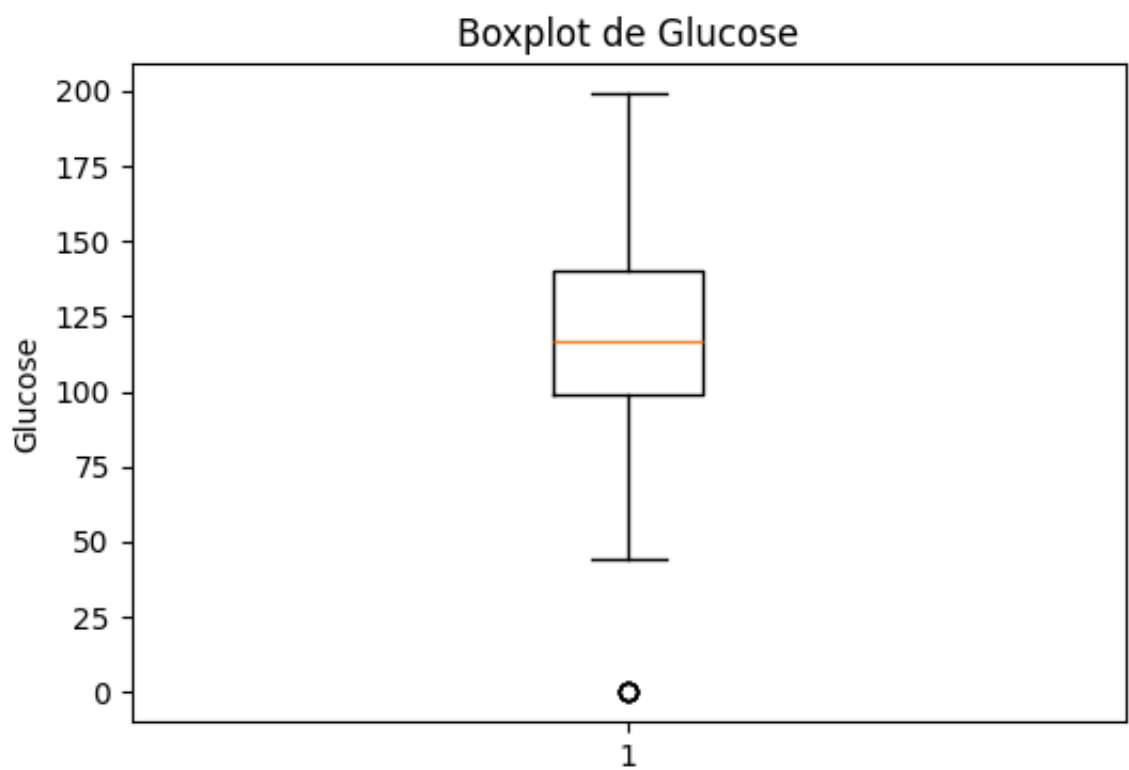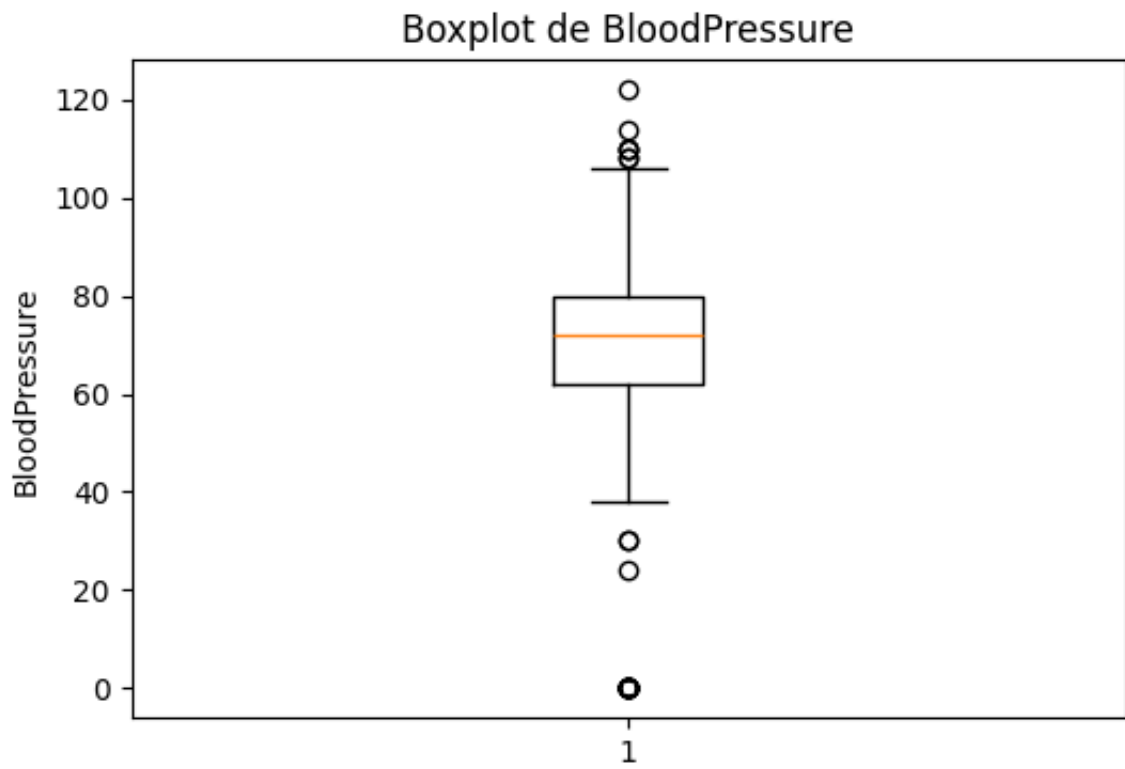
Boxplot de Insulin

Boxplot de BMI

Boxplot de DiabetesPedigreeFunction
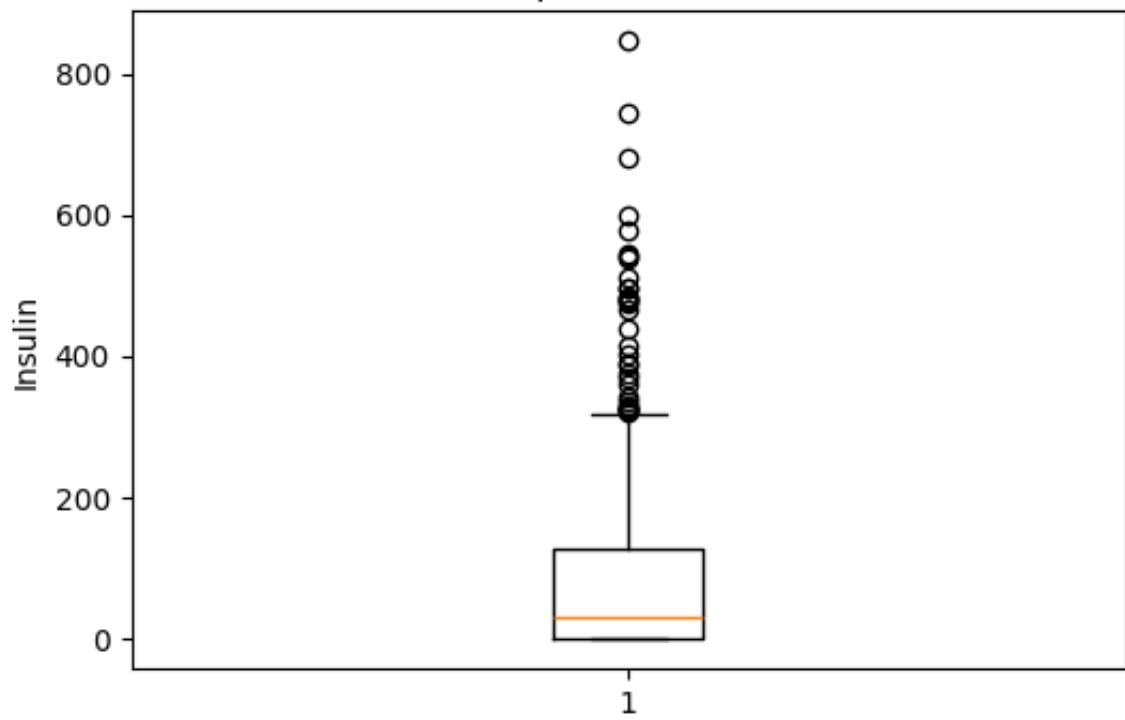
Boxplot de Age


```
In [25]:   corr = df.corr(numeric_only=True)
           plt.figure(figsize=(10,8))
           sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
           plt.title('Mapa de calor de correlaciones (todas las variables)')
           plt.show()
```
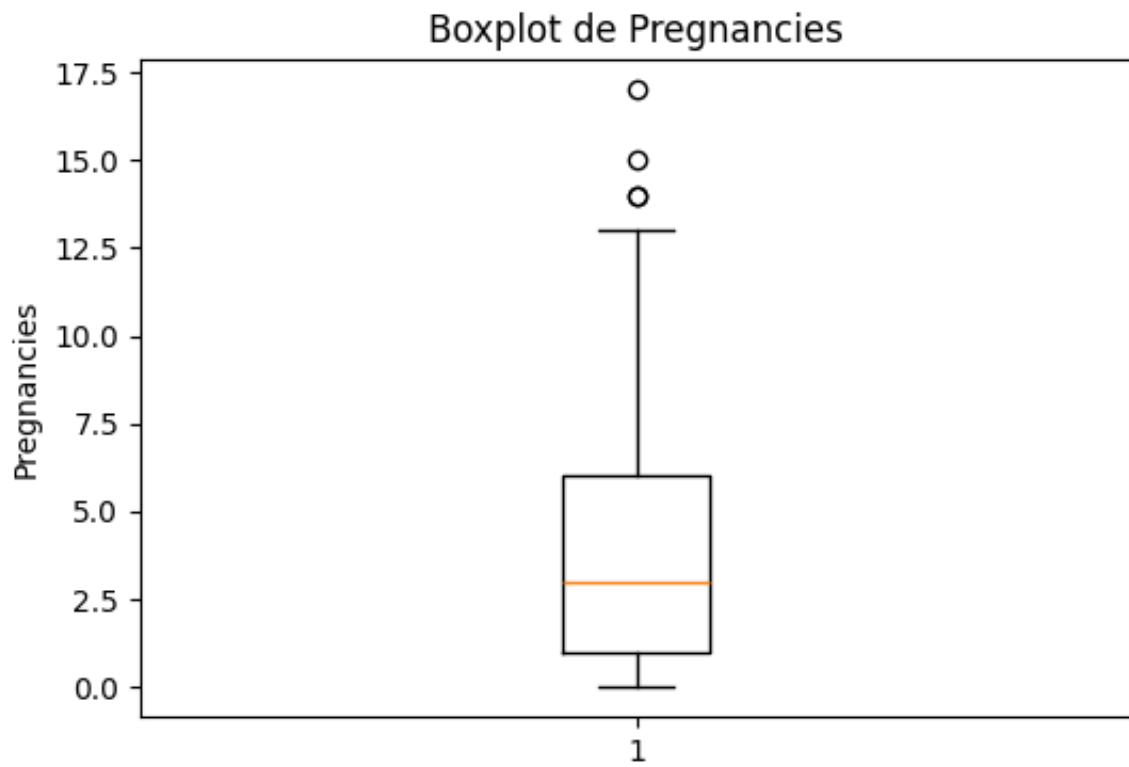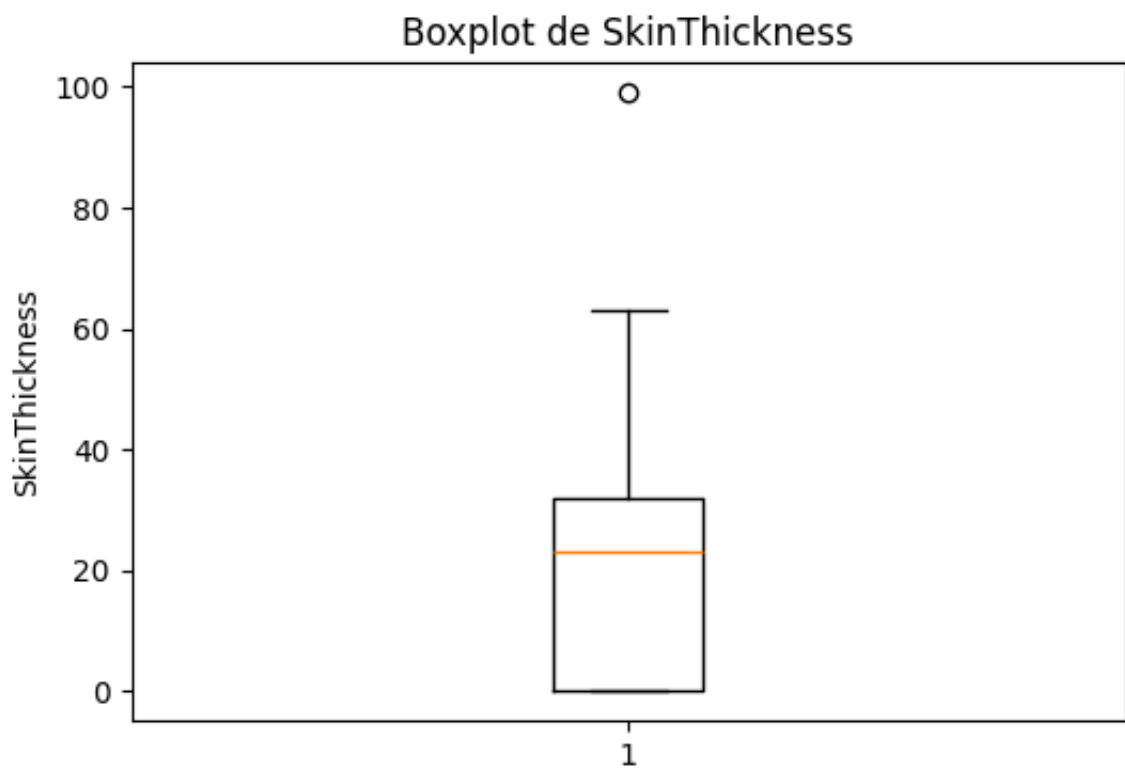
Mapa de calor de correlaciones (todas las variables)

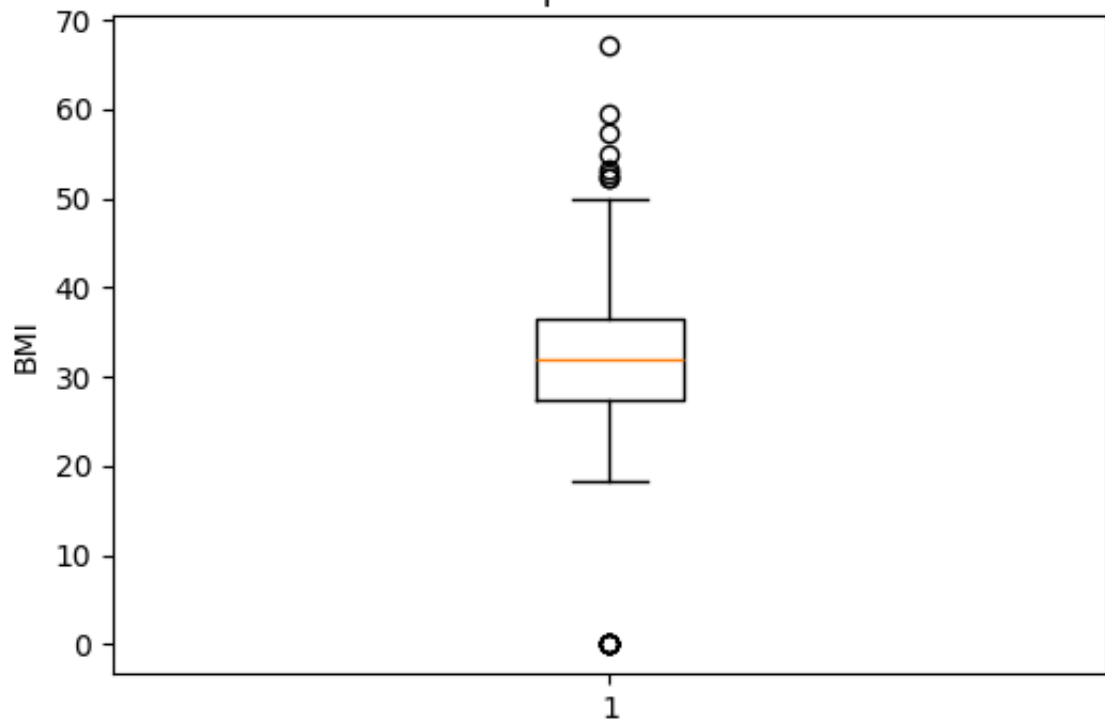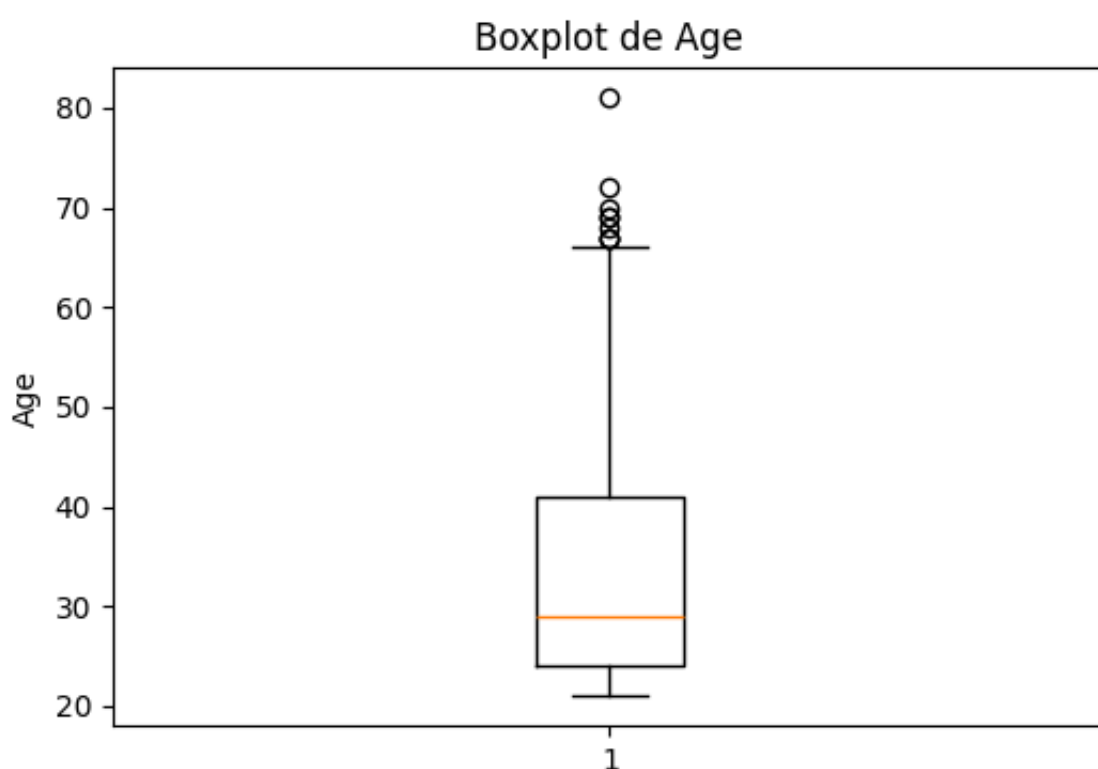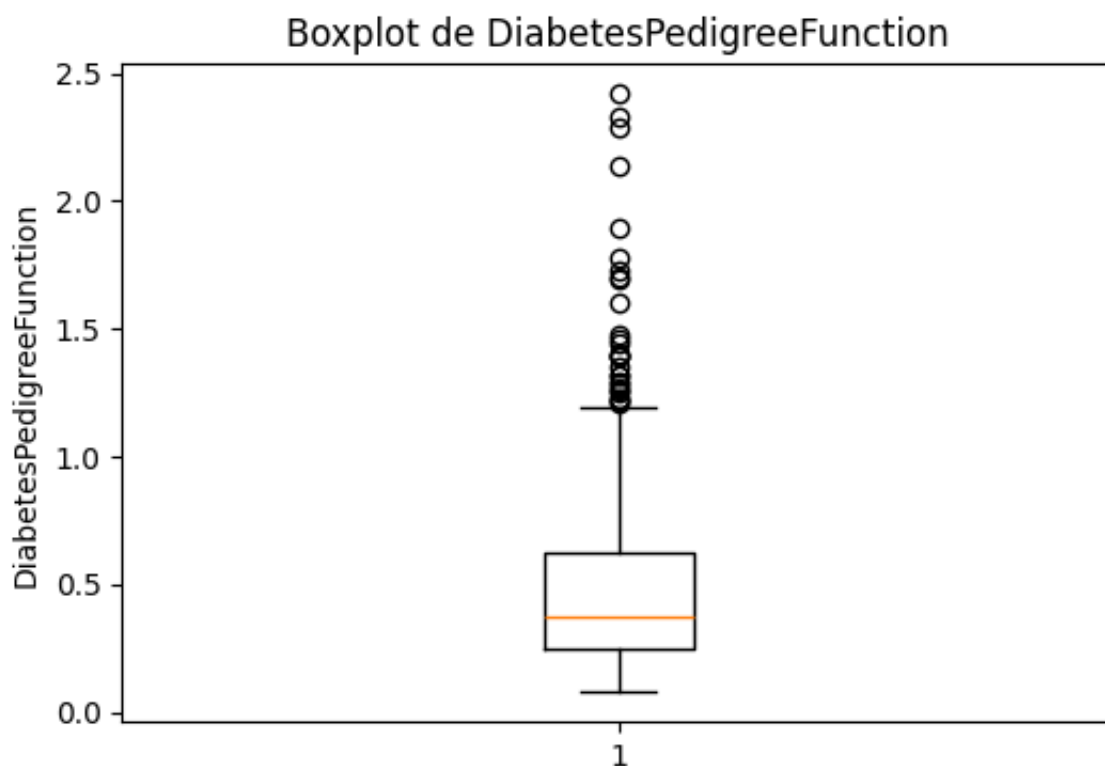|	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
Pregnancies	1.00	0.13	0.14	-0.08	-0.07	0.02	-0.03	0.54	0.22
Glucose	0.13	1.00	0.15	0.06	0.33	0.22	0.14	0.26	0.47
BloodPressure	0.14	0.15	1.00	0.21	0.09	0.28	0.04	0.24	0.07
SkinThickness	-0.08	0.06	0.21	1.00	0.44	0.39	0.18	-0.11	0.07
Insulin	-0.07	0.33	0.09	0.44	1.00	0.20	0.19	-0.04	0.13
BMI	0.02	0.22	0.28	0.39	0.20	1.00	0.14	0.04	0.29
DiabetesPedigreeFunction	-0.03	0.14	0.04	0.18	0.19	0.14	1.00	0.03	0.17
Age	0.54	0.26	0.24	-0.11	-0.04	0.04	0.03	1.00	0.24
Outcome	0.22	0.47	0.07	0.07	0.13	0.29	0.17	0.24	1.00