In [1]:
```python
import pandas as pd

# Cargar datos
df = pd.read_csv("diabetes.csv")
print(df.head())

# Número de filas y columnas
print("Número de filas:", df.shape[0])
print("Número de columnas:", df.shape[1])

# Información general
print(df.info())

# Valores nulos
print(df.isnull().sum())

# Estadísticas descriptivas (variables seleccionadas)
print(df[["Glucose", "BMI", "Outcome"]].describe())

# Medias, medianas y desviaciones estándar
glucose_mean = df["Glucose"].mean()
glucose_median = df["Glucose"].median()
glucose_std = df["Glucose"].std()

bmi_mean = df["BMI"].mean()
bmi_median = df["BMI"].median()
bmi_std = df["BMI"].std()

outcome_mean = df["Outcome"].mean()
outcome_median = df["Outcome"].median()
outcome_std = df["Outcome"].std()

print("Glucose - mean, median, std:", glucose_mean, glucose_median, glucose_std)
print("BMI - mean, median, std:", bmi_mean, bmi_median, bmi_std)
print("Outcome - mean, median, std:", outcome_mean, outcome_median, outcome_std)

# Consultas
print(df[df["Glucose"] > 120].head())  # Glucosa >120
print(df.groupby("Outcome")["BMI"].mean())  # Promedio BMI por outcome
print(df[(df["Outcome"] == 0) & (df["BMI"] > 30)].head())  # No diabéticos BMI >30

# Exportar dataset filtrado (opcional)
df[["Glucose", "BMI", "Outcome"]].to_csv("mi_dataset_filtrado.csv", index=False)
```

```
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0            6      148             72             35        0  33.6
1            1       85             66             29        0  26.6
2            8      183             64              0        0  23.3
3            1       89             66             23       94  28.1
4            0      137             40             35      168  43.1

   DiabetesPedigreeFunction  Age  Outcome
0                     0.627   50        1
1                     0.351   31        0
2                     0.672   32        1
3                     0.167   21        0
4                     2.288   33        1
Número de filas: 768
Número de columnas: 9
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               768 non-null    int64
 1   Glucose                   768 non-null    int64
 2   BloodPressure             768 non-null    int64
 3   SkinThickness             768 non-null    int64
 4   Insulin                   768 non-null    int64
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64
 8   Outcome                   768 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
None
Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
          Glucose         BMI     Outcome
count  768.000000  768.000000  768.000000
mean   120.894531   31.992578    0.348958
std     31.972618    7.884160    0.476951
min      0.000000    0.000000    0.000000
25%     99.000000   27.300000    0.000000
50%    117.000000   32.000000    0.000000
75%    140.250000   36.600000    1.000000
max    199.000000   67.100000    1.000000
Glucose - mean, median, std: 120.89453125 117.0 31.97261819513622
BMI - mean, median, std: 31.992578124999998 32.0 7.884160320375446
Outcome - mean, median, std: 0.3489583333333333 0.0 0.47695137724279896
   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0            6      148             72             35        0  33.6
```

```
2            8      183             64              0        0  23.3
4            0      137             40             35      168  43.1
8            2      197             70             45      543  30.5
9            8      125             96              0        0   0.0

   DiabetesPedigreeFunction  Age  Outcome
0                     0.627   50        1
2                     0.672   32        1
4                     2.288   33        1
8                     0.158   53        1
9                     0.232   54        1
Outcome
0    30.304200
1    35.142537
Name: BMI, dtype: float64
    Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
7            10      115              0              0        0  35.3
10            4      110             92              0        0  37.6
18            1      103             30             38       83  43.3
20            3      126             88             41      235  39.3
21            8       99             84              0        0  35.4

    DiabetesPedigreeFunction  Age  Outcome
7                      0.134   29        0
10                     0.191   30        0
18                     0.183   33        0
20                     0.704   27        0
21                     0.388   50        0
```

In [ ]: