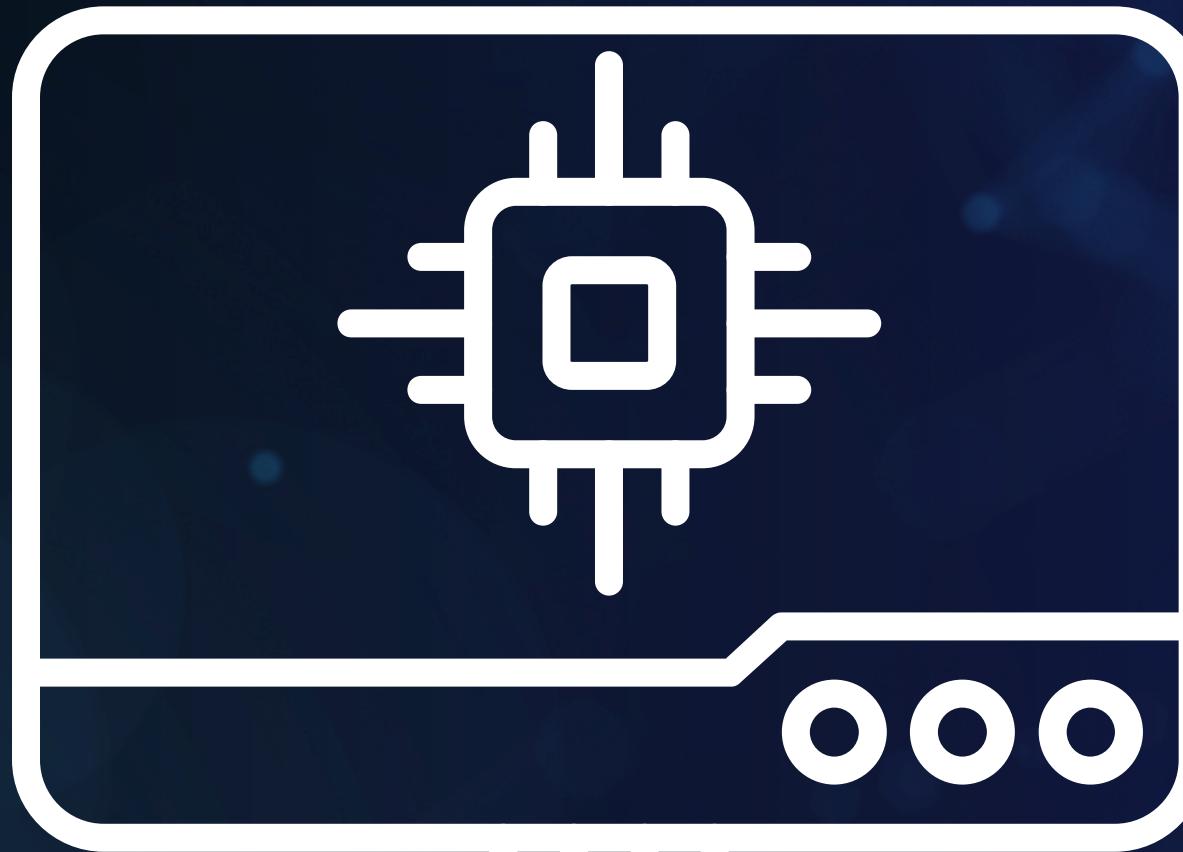


# GRÁFICOS Y VISUALIZACIONES MÚLTIPLES

- ALVARO JOSE FLOREZ
- JOSE MANUEL SALAS
- SANTIAGO FELIPE DIAZ

# PROMPT



- Tema: Gráficos y visualizaciones múltiples
- Enfoque: Utilizar diversos tipos de gráficos (barras, tortas, líneas, boxplots, etc.) para representar visualmente los hallazgos del análisis.

# PASO 1 : DATA SET

## CARACTERÍSTICAS:

- ColumnaDescripción
- Id
- Identificador único de cada persona (1 a 500)
- Age
- Edad de la persona (entre 18 y 79 años)
- Sex
- Sexo de la persona (Male o Female)
- Ethnicity
- Grupo étnico (Hispanic, Asian, Black, White, Other)
- BMI
- Índice de Masa Corporal (valores típicos entre 15 y 50)
- Diabetes Probability
- Probabilidad de tener diabetes (valor entre 0 y 1, calculado sintéticamente)

Id	Age	Sex	Ethnicity	BMI	Diabetes Probability
1	56	Female	Asian	24,3	0,16
2	69	Male	Asian	29	0,35
3	46	Female	Asian	26,8	0,15
4	32	Male	Black	32,5	0,22
5	60	Male	Other	27,6	0,21
6	25	Female	Other	27,8	0,09
7	78	Female	Hispanic	25,2	0,27
8	38	Female	Hispanic	26,7	0,23
9	56	Male	Asian	28,5	0,18
10	75	Female	Hispanic	18,4	0,24
11	36	Male	Black	20,3	0,08
12	40	Male	Other	30,7	0,2
13	28	Male	Asian	27,9	0,08
14	28	Female	Hispanic	26,1	0,06

- pandas
- Para cargar, manipular y analizar datos en estructuras tipo tabla (DataFrames).
- matplotlib.pyplot
- Para crear gráficos básicos (líneas, barras, histogramas, etc.).
- seaborn
- Es una biblioteca basada en matplotlib que permite crear gráficos estadísticos más atractivos y con menos código.

## PASO 2: FUNCIONES

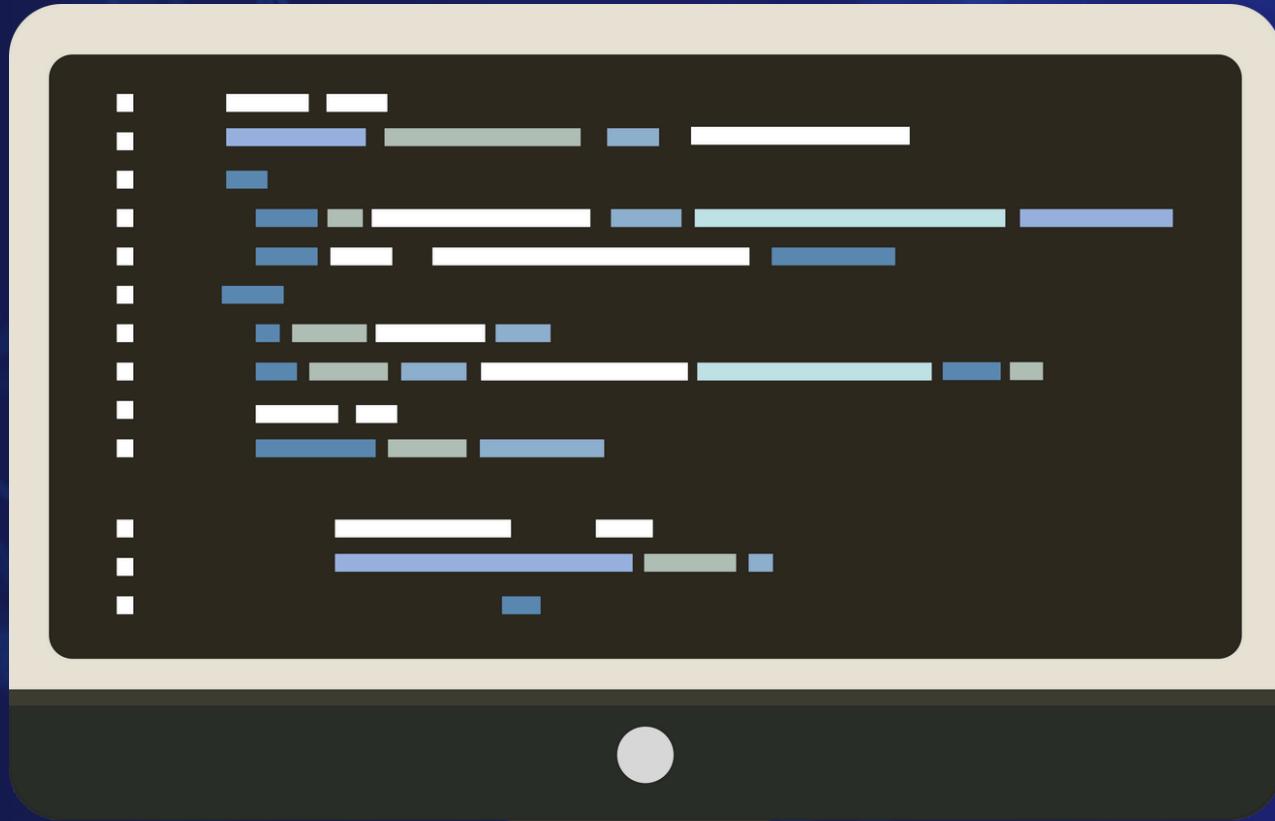
```
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns
```

# PASO 3

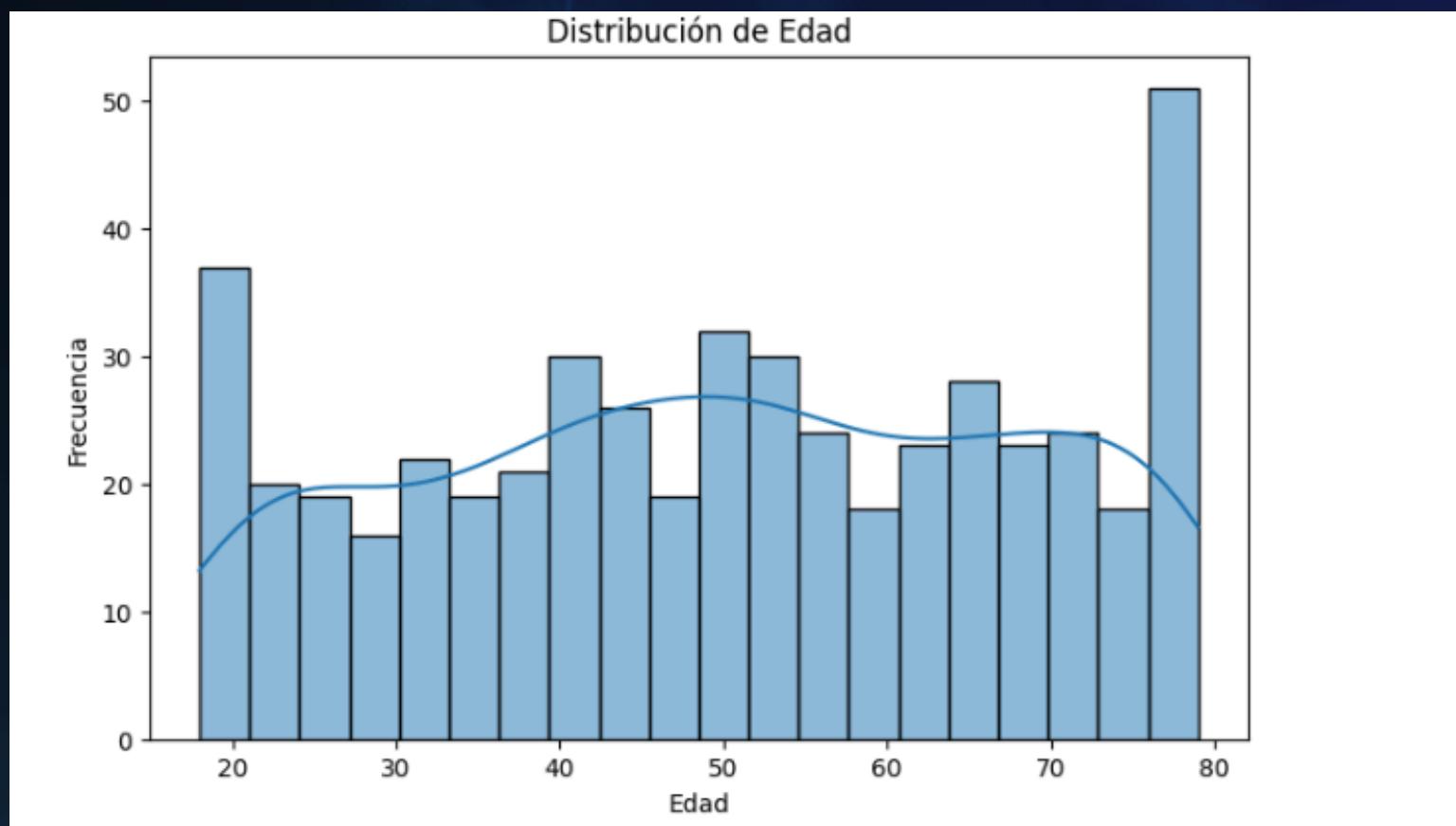
- Nuestro siguiente paso es importar nuestro data set desde excel para que Colab añada los datos al código

```
from google.colab import files  
uploaded = files.upload() |
```

```
df = pd.read_excel("diabetes_probability_dataset_500.xlsx")
```



# PASO 4: (HISTOGRAMA)



- Ahora que tenemos nuestro data set en nuestro libro nos podemos sentir en la libertad de pedirle nuestros graficos que coincidan con los datos que tenemos.
- Primero usamos el histograma de las edades de las personas censadas y comparamos

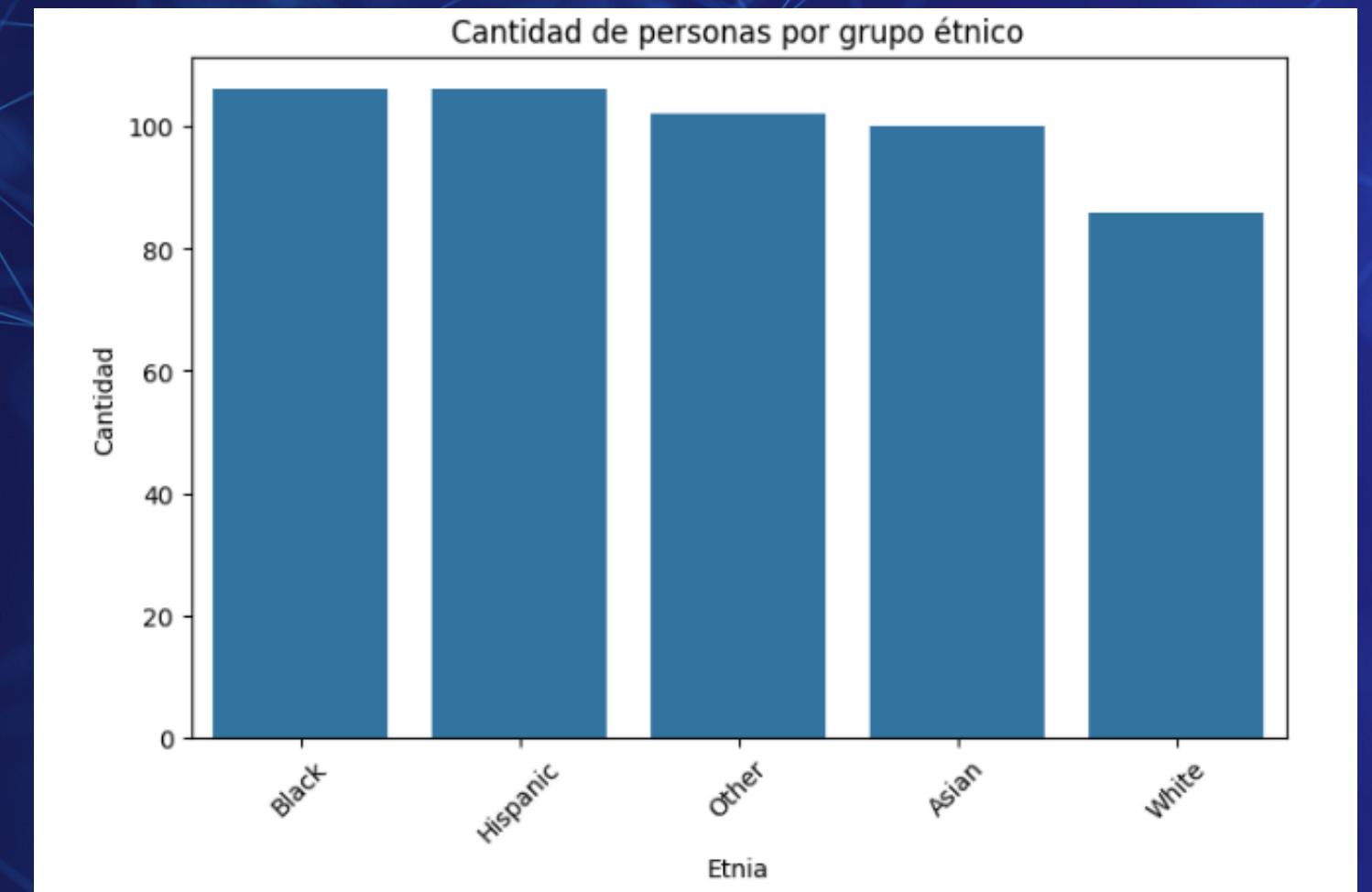
```
plt.figure(figsize=(8, 5))
sns.histplot(df["Age"], bins=20, kde=True)
plt.title("Distribución de Edad")
plt.xlabel("Edad")
plt.ylabel("Frecuencia")
plt.show()
```

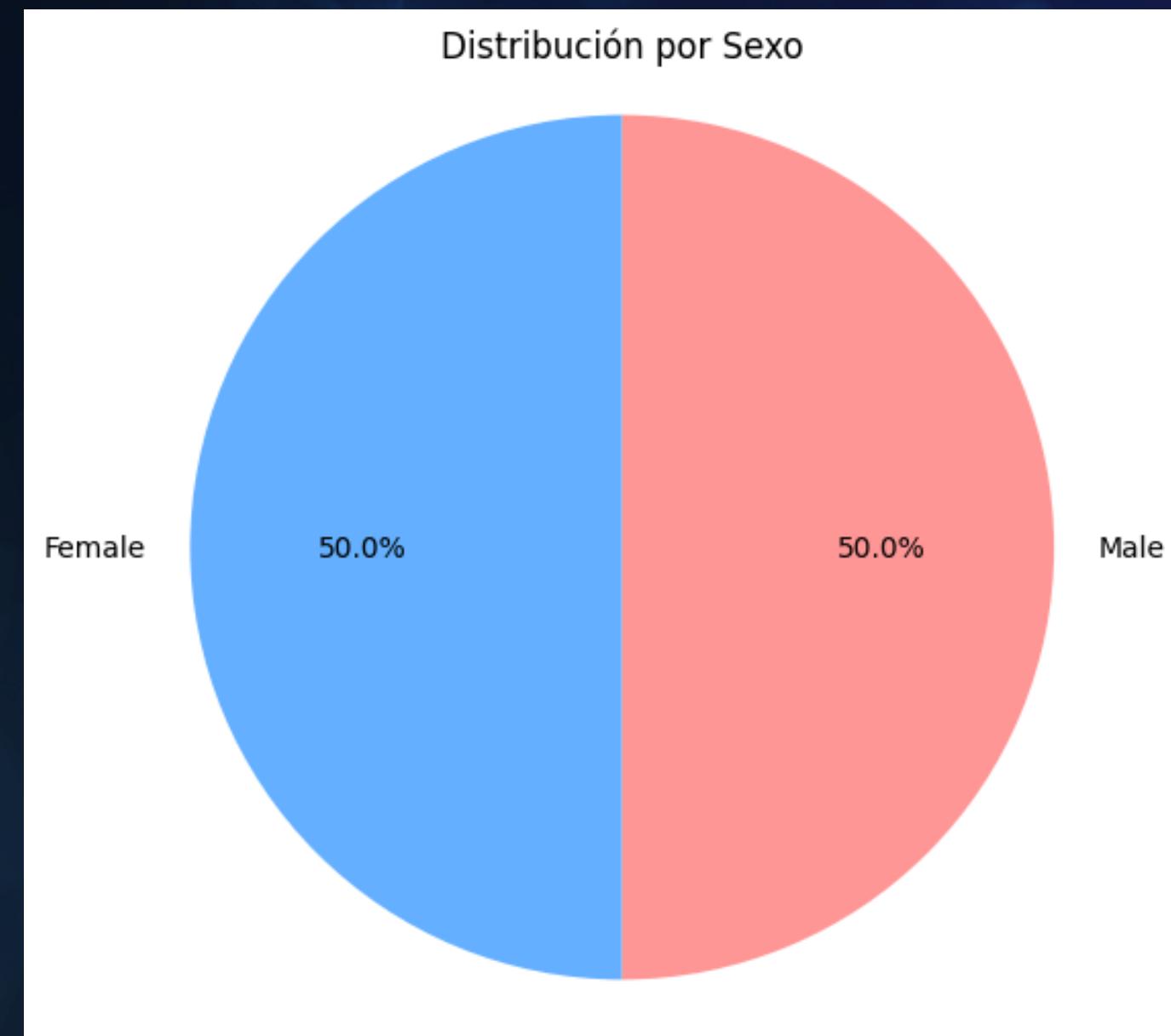
# PASO 5 (GRÁFICO DE BARRAS)



- Ahora que tenemos nuestro data set en nuestro libro nos podemos sentir en la libertad de pedirle nuestros graficos que coincidan con los datos que tenemos.
- Segundo usamos el grafico de barras segun el tipo de etnias de las personas sensadas

```
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x="Ethnicity", order=df["Ethnicity"].value_counts().index)
plt.title("Cantidad de personas por grupo étnico")
plt.xlabel("Etnia")
plt.ylabel("Cantidad")
plt.xticks(rotation=45)
plt.show()
```





```
plt.figure(figsize=(6, 6))
sex_counts = df['Sex'].value_counts()
plt.pie(sex_counts, labels=sex_counts.index, autopct='%1.1f%%', startangle=90, colors=['#66b3ff', '#ff9999'])
plt.title("Distribución por Sexo")
plt.axis('equal')
plt.show()
```

## PASO 6 (GRÁFICO DE TORTA)

- Ahora que tenemos nuestro data set en nuestro libro nos podemos sentir en la libertad de pedirle nuestros graficos que coincidan con los datos que tenemos.
- Tercero usamos el grafico de torta según el genero de las personas censadas



# PASO 7 (DIAGRAMA DE CAJA )



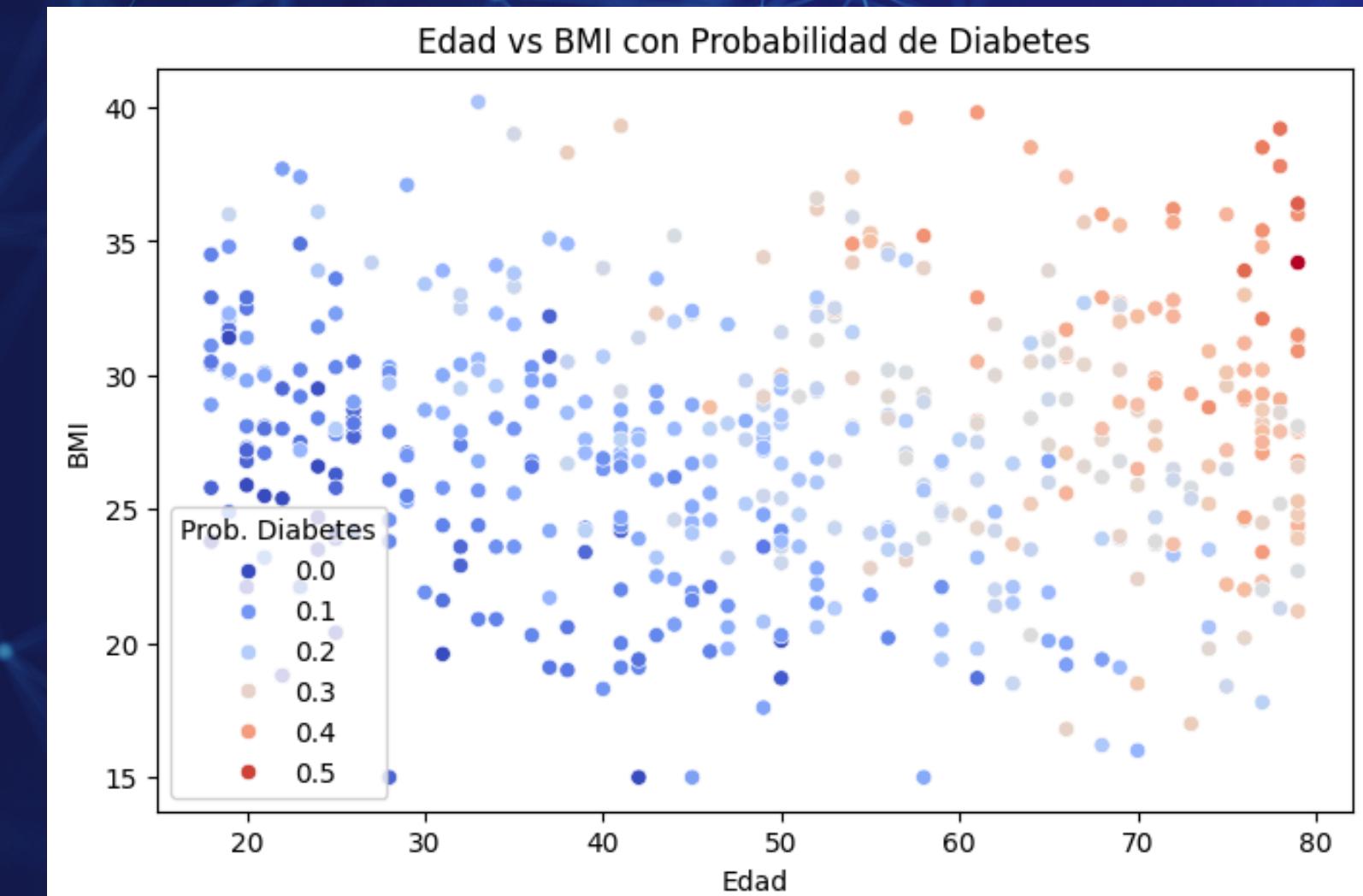
- Ahora que tenemos nuestro data set en nuestro libro nos podemos sentir en la libertad de pedirle nuestros graficos que coincidan con los datos que tenemos.
- Cuando usamos el grafico de caja según el BMI por sexo de las personas censadas

```
plt.figure(figsize=(8, 5))
sns.boxplot(data=df, x="Sex", y="BMI")
plt.title("Distribución del BMI por Sexo")
plt.xlabel("Sexo")
plt.ylabel("BMI")
plt.show()
```

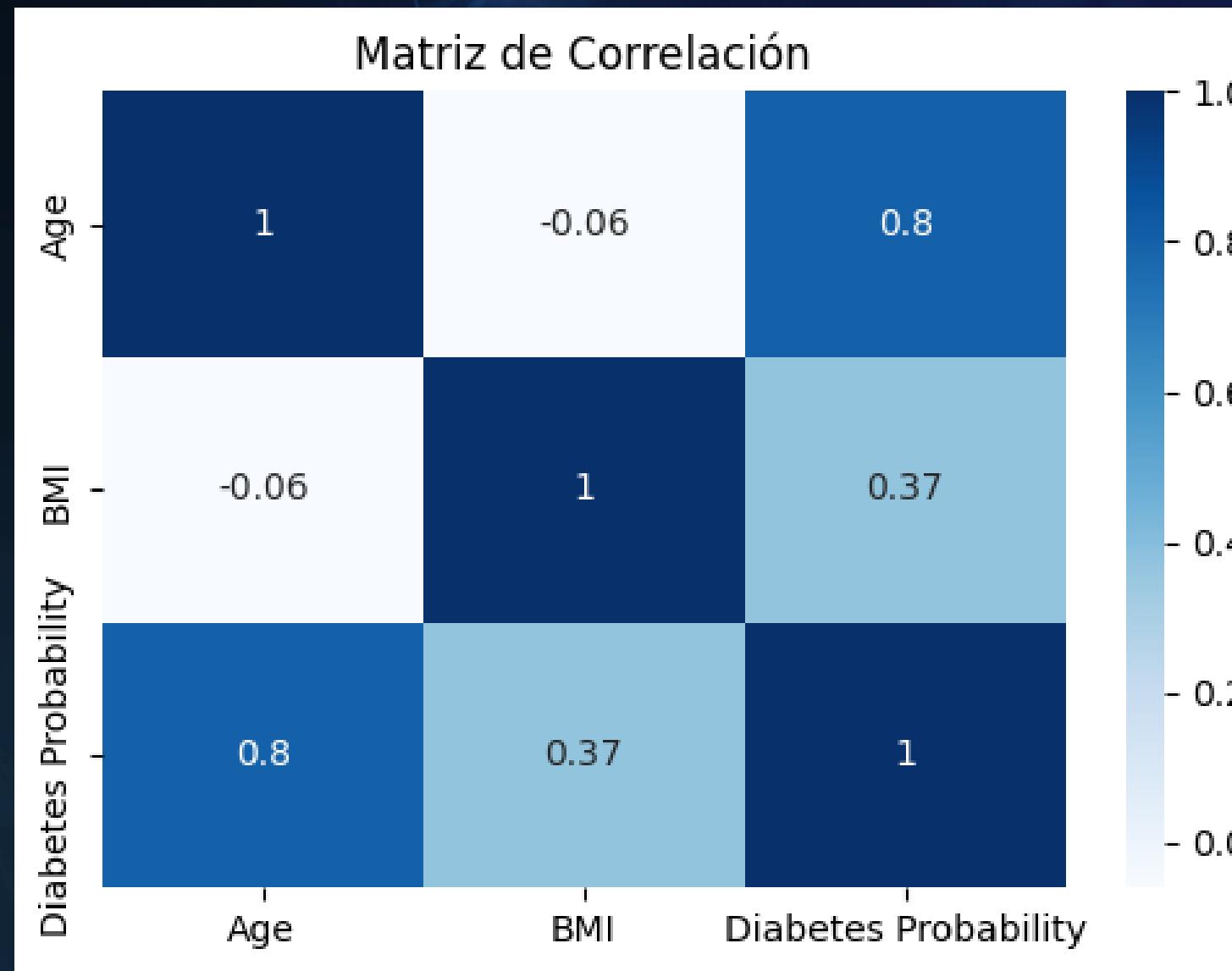
# PASO 8 (DIAGRAMA DE DISPERSIÓN)

- Ahora que tenemos nuestro data set en nuestro libro nos podemos sentir en la libertad de pedirle nuestros gráficos que coincidan con los datos que tenemos.
- Quito usamos el diagrama de dispersión comparando el índice de masa corporal (IMC) vs la edad de las personas

```
plt.figure(figsize=(8, 5))
sns.scatterplot(data=df, x="Age", y="BMI", hue="Diabetes Probability", palette="coolwarm")
plt.title("Edad vs BMI con Probabilidad de Diabetes")
plt.xlabel("Edad")
plt.ylabel("BMI")
plt.legend(title="Prob. Diabetes")
plt.show()
```



## PASO 9: (MATRIZ DE CORRELACIÓN )



- Ahora que tenemos nuestro data set en nuestro libro nos podemos sentir en la libertad de pedirle nuestros graficos que coincidan con los datos que tenemos.
- Sexto usamos el diagrama de mapa de calor comparando Edad (Age), Índice de Masa Corporal (BMI) y Probabilidad de Diabetes (Diabetes Probability).

```
plt.figure(figsize=(6, 4))
sns.heatmap(df[["Age", "BMI", "Diabetes Probability"]].corr(), annot=True, cmap="Blues")
plt.title("Matriz de Correlación")
plt.show()
```

# THANK YOU!

