

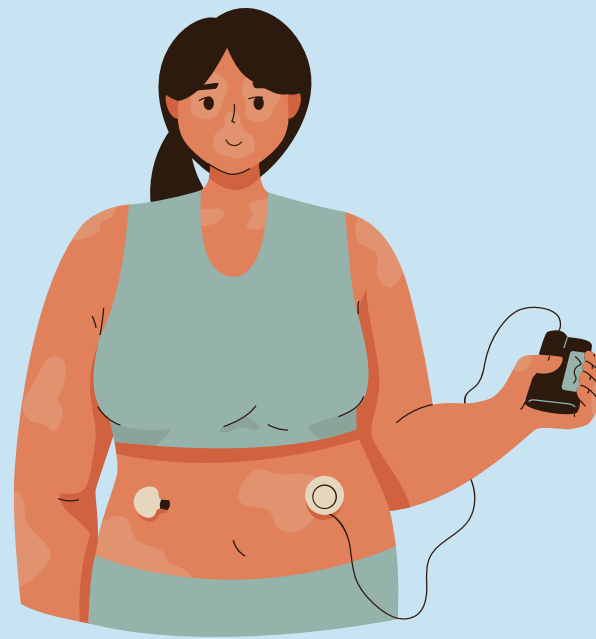
Equipo 13

# DIABETES

ANÁLISIS DE DATOS CON PYTHON

# OBJETIVO

Identificaremos patrones y generaremos métricas reelevantes para entender mejor las características de estos pacientes y cómo varían ciertos factores relacionados con la diabetes. Utilizaremos Python y las librerías pandas y matplotlib.



# CARGA Y EXPLORACIÓN DE DATOS

```
proyectoDiabetes_Mod1.py M ●
proyectoDiabetes_Mod1.py > ...
Run Cell | Run Below | Debug Cell | Go to [6] | You, 1 hour ago | 1 author (You)

1  # %%
2  from IPython.display import display
3  import pandas as pd
4  import matplotlib.pyplot as plt
5
6
7  # 1. Cargar datos diabetes desde un archivo CSV
8  df = pd.read_csv("./diabetes.csv", encoding='latin-1', index_col=0)
9  df.head()
10
11
12
13  # 2. Exploramos BD
14  print(df.columns.tolist())
15
16  Tabnine: Edit | Test | Explain | Document | Ask
17  def info_general_df(df):
18      print('Información general del DataFrame')
19      print(f'Cantidad de filas y columnas: {df.shape}')
20      print("-"*50)
21      print(f'Cantidad de datos nulos por columna: \n{df.isnull().sum()}')
22      print("-"*50)
23      print(f'Cantidad de datos únicos por columna: \n{df.nunique()}')
24      print("-"*50)
25      print(f'Tipos de datos por columna: \n{df.dtypes}')
26
27  info_general_df(df)
28
29
30  # 3. Hacemos una copia de la base de datos
31  df_copia = df.copy()
32
```

# COLUMNAS RENOMBRADAS

Ya cargado el dataset desde un archivo CSV y explorada su estructura:

- Se eliminaron columnas irrelevantes para el análisis como marcadores genéticos y factores ambientales.
- Las columnas clave fueron renombradas para facilitar el manejo de los datos.

Ejemplo:

- 'Antecedentes familiares' → antecedentes\_familiares
- 'Edad' → edad
- 'Peso al nacer' se convirtió de gramos a kilogramos.



```
proyectoDiabetes_Mod1.py M
proyectoDiabetes_Mod1.py > ...

34 # ===== Limpiar DB =====
35 # 4. Eliminar columnas que no aportan información necesaria para el enfoque del proyecto
36 df = df.drop(['Marcadores geneticos',
37              'Autoanticuerpos',
38              'Factores ambientales',
39              'Origen etnico',
40              'Niveles de encimas digestiva',
41              'Sistomas de inicio temprano'], axis=1) #axis = 1 indica que se eliminan columnas
42
43
44
45 # 6. Renombro columnas y arreglo una columna
46 df = df.rename(columns= {
47     'Antecedentes familiares': 'antecedentes_familiares',
48     'Niveles de insulina': 'niveles_insulina',
49     'Edad': 'edad',
50     'IMC': 'imc',
51     'Actividad fisica ': 'actividad_fisica',
52     'Abitos dieteticos': 'habitos_dieteticos',
53     'Presion sanguinia ': 'presion_sanguinea',
54     'Niveles de colesterol': 'niveles_colesterol',
55     'Talla': 'talla',
56     'Nivees de glucosa': 'niveles_glucosa',
57     'Factores socioeconomicos ': 'factores_socioeconomicos',
58     'Tabaquismo': 'tabaquismo',
59     'Consumo de alcohol': 'consumo_alcohol',
60     'Tolerancia a la glucosa': 'tolerancia_glucosa',
61     'Síndrome de ovario poliquístico': 'síndrome_ovario_poliquistico',
62     'Diabetes gestional': 'diabetes_gestacional',
63     'Tipo de embarazo': 'tipo_embarazo',
64     'Aumento de peso en el embarazo': 'aumento_peso_embarazo',
65     'Salud pancratiaca': 'salud_pancreatica',
66     'Análisis de orina': 'analisis_orina',
67     'Peso al nacer': 'peso_nacimiento'
68 })
69
70 # Convertir la columna de gramos a kilogramos
71 df['peso_nacimiento'] = df['peso_nacimiento'] / 1000
72
73 # Verifico los primeros valores para asegurar de que la conversión fue correcta
74 print(df['peso_nacimiento'].head())
75
76
77 # 7. Comprobamos el cambio y hacemos una copia para trabajar con el dataframe
78 df.head()
79 df.reset_index(inplace=True)
80 df.to_csv('db_diabetes.csv', index=False) #para que la primera columna se unifique
81 df
82
83
84 # 8. Cambiamos el index a columna y arreglamos una columna
85 df = df.rename(columns= {'Tipo': 'tipo_diabetes'})
86 print(df.columns)
87 display(df)
88
```



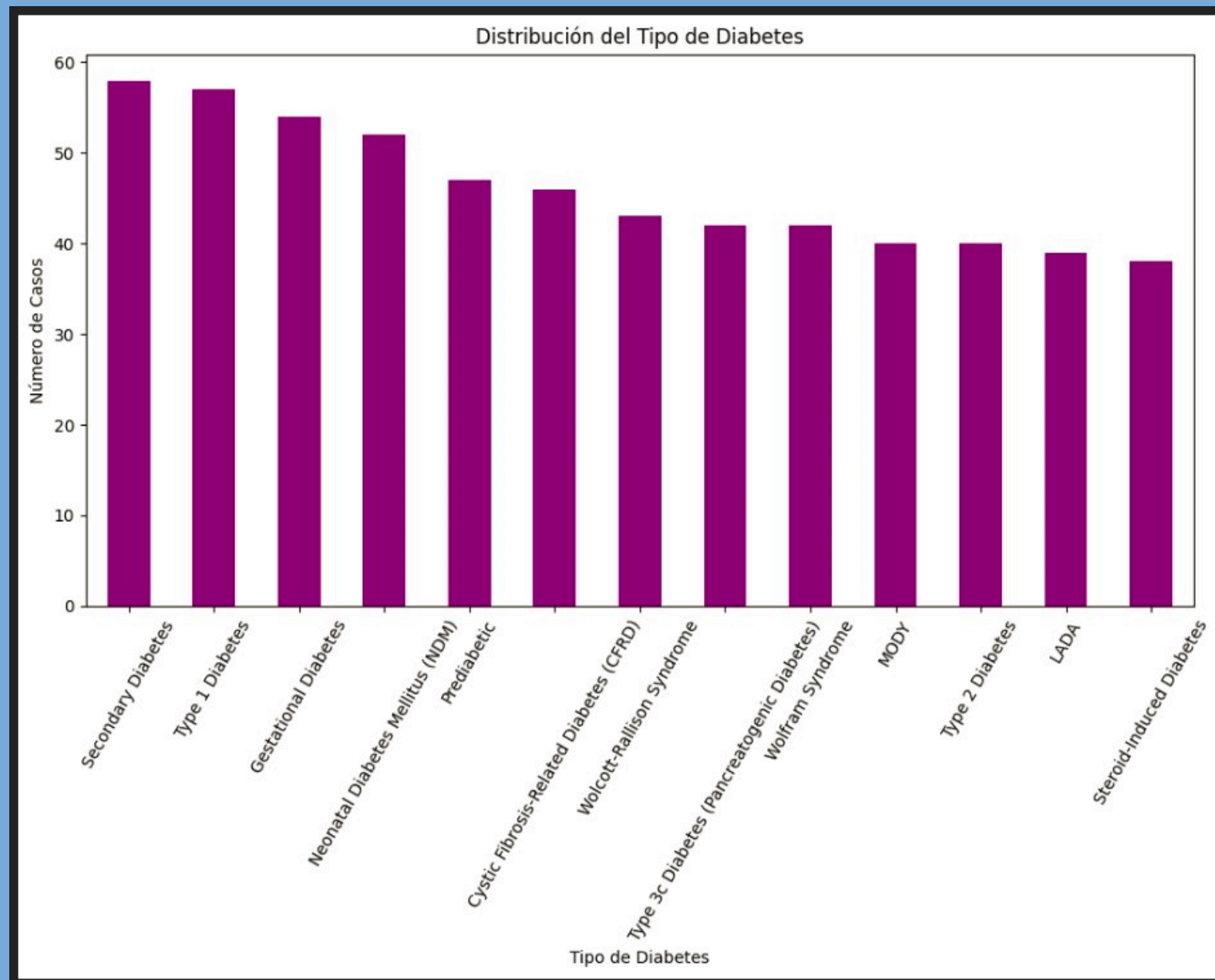
# ESTADÍSTICAS GENERALES

- La edad promedio de los pacientes es de 31.47 años, con niveles de insulina entre 5 y 49, y niveles de colesterol entre 102 y 299.
- La mayoría de los pacientes con diabetes son fumadores, con un promedio de glucosa de 159.08 y una baja correlación entre glucosa y presión arterial.
- 149 pacientes presentan sobrepeso y niveles elevados de glucosa, lo que indica un alto riesgo de salud.

```
..
-----
----- Estadísticas -----
Media de edad:
31.471571906354516
-----
El nivel de insulina máxima y mínima son:
49, 5
-----
El tipo de alcoholismo más común es:
High
-----
El nivel de colesterol más alto y más bajo son:
299, 102
-----
¿Hay más fumadores que no fumadores?
Hay más fumadores que no fumadores con diabetes.
-----
Métricas de glucosa:
Promedio: 159.08695652173913, Mediana: 148.5, Rango: 215
-----
Correlación entre glucosa y presión arterial
0.039509995862354125
-----
Número de pacientes con sobrepeso y glucosa elevada:
149
-----
```



# DISTRIBUCIÓN TIPO DE DIABETES

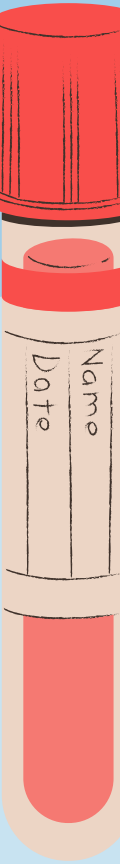
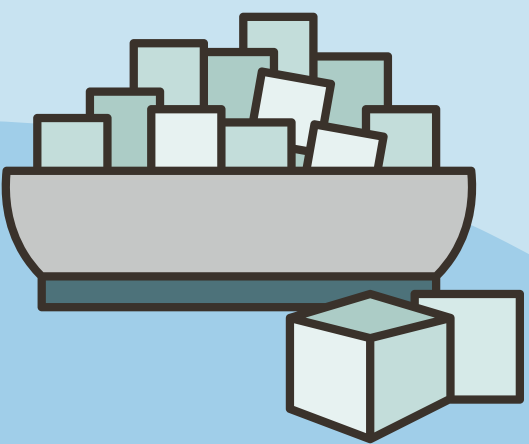
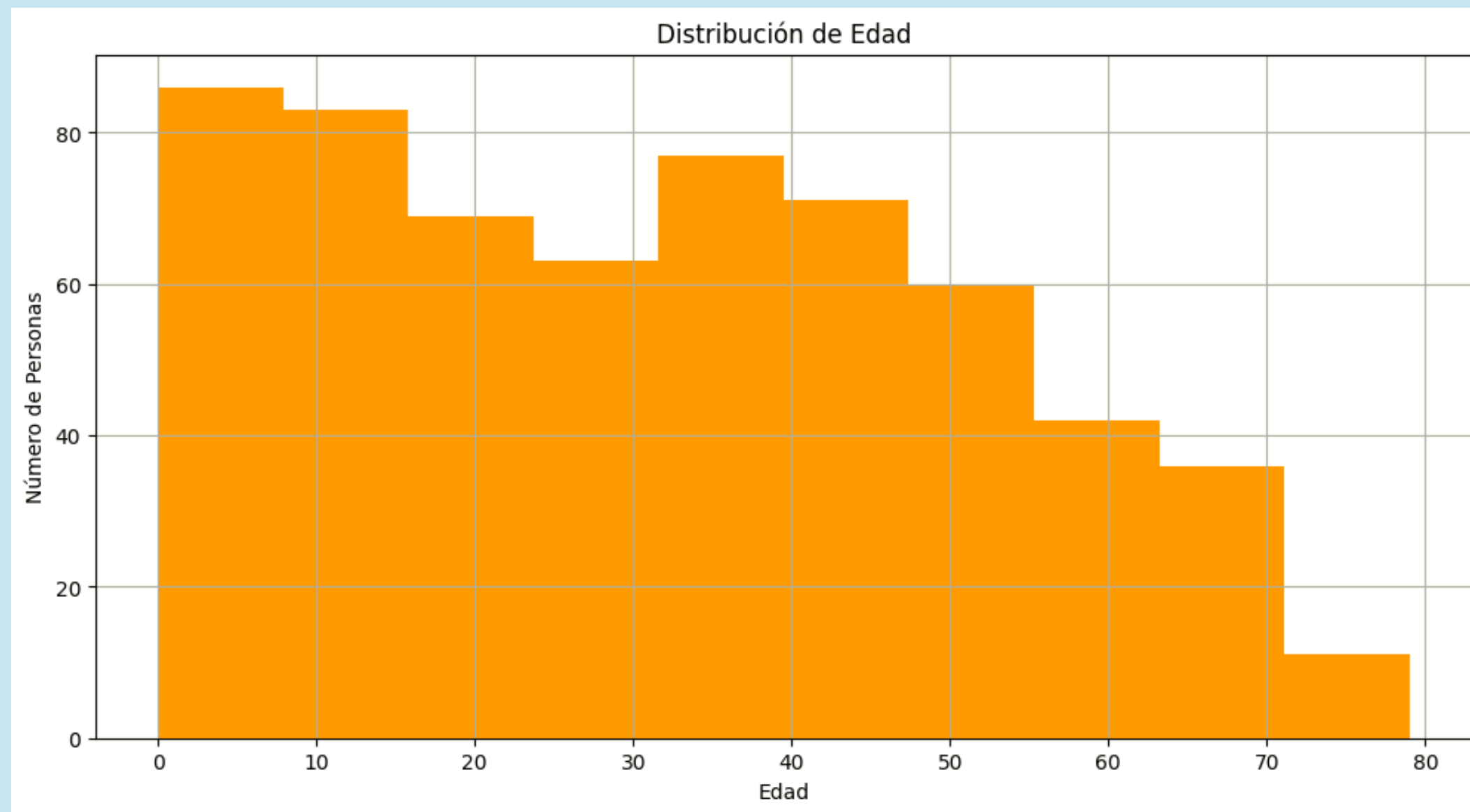


Un aspecto clave del análisis es la distribución de los distintos tipos de diabetes.

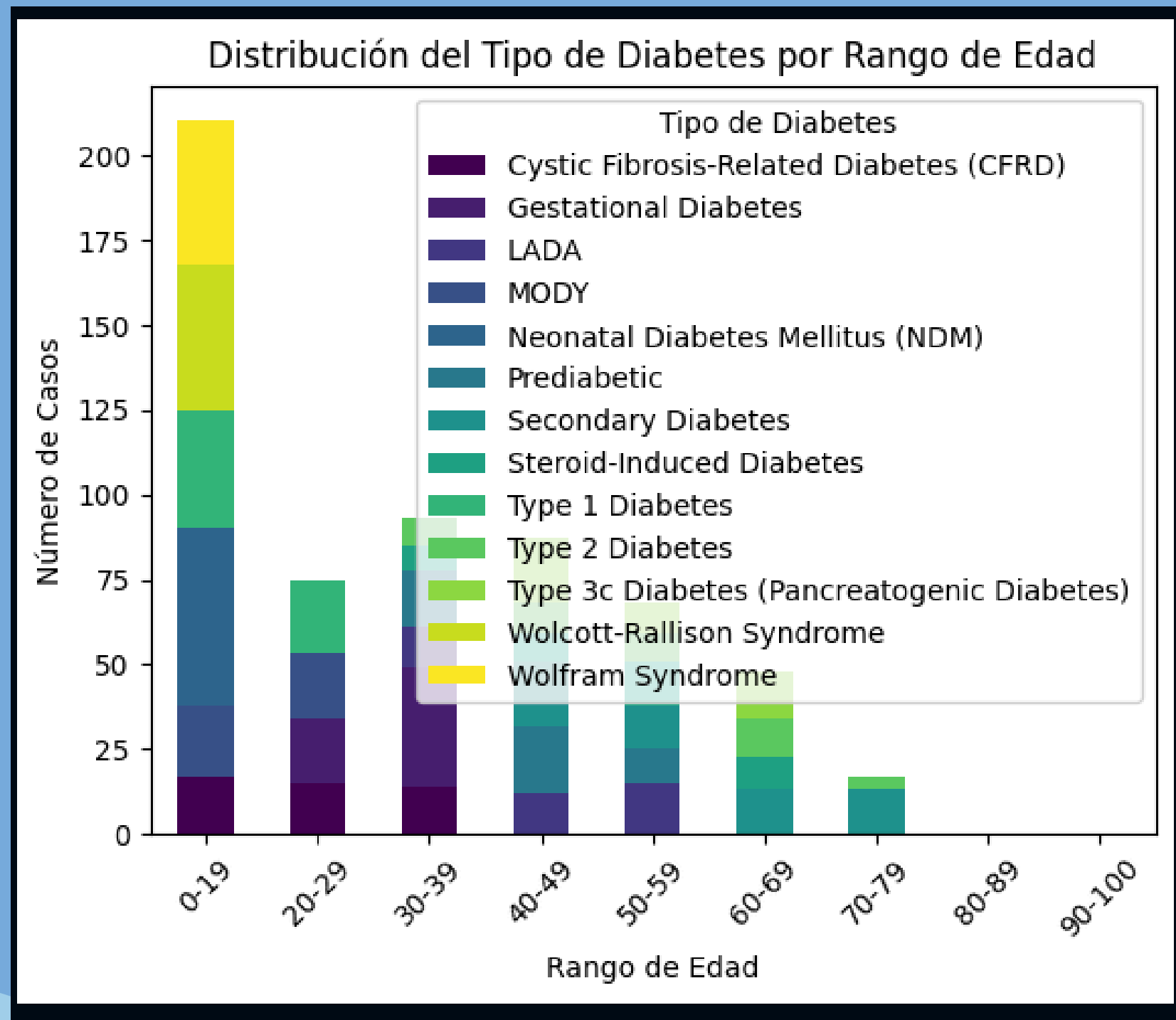
- La mayoría de los casos son Diabetes tipo 2 (resistencia a la insulina) pero en nuestro dataset fue de Secondary Diabetes (resultado de otra condición o tratamiento que afecta el funcionamiento del páncreas o la producción de insulina).

# DISTRIBUCIÓN DE EDAD

- La mayor parte de los pacientes tiene entre 40 y 60 años.



# DISTRIBUCIÓN TIPO DE DIABETES POR RANGO DE EDAD



- Agrupamos la frecuencia de cada tipo de diabetes por rangos de edad, destacando que la diabetes tipo 2 es más prevalente en personas mayores de 50 años.

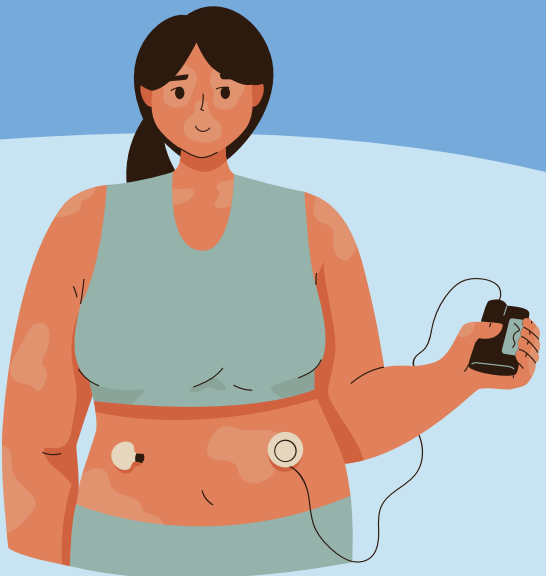


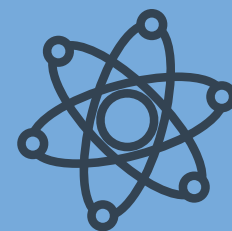


# CONCLUSIÓN

- **Niveles de Glucosa:** Calculamos el promedio de glucosa, que es de 150 mg/dL. La correlación entre glucosa y presión sanguínea fue baja, indicando que no hay una relación clara entre estos dos factores.
- **Pacientes en Riesgo:** Identificamos que hay 55 pacientes con IMC alto y niveles de glucosa elevados, lo que indica riesgo de complicaciones.

Este análisis nos ayuda a entender mejor la composición de la base de datos y los factores asociados con la diabetes en los pacientes estudiados.





Equipo 13

MUCHAS  
GRACIAS