

## Diego Emilio Muñiz Ramirez #621784 - A1.1 Aprendizaje estadístico-automático

### ✓ 1. Importación del conjunto de datos

Se importó la base de datos la cual es "A1.1 Obesidad.csv" utilizando Python y la biblioteca pandas como bien se ve en el código. La base de datos contiene información demográfica, hábitos alimenticios y variables relacionadas con la salud de individuos así como su nivel de obesidad.

```
import pandas as pd
df = pd.read_csv("A1.1_Obesidad.csv")
```

### ✓ 2. Exploración inicial de los datos

Se realizó una exploración inicial del conjunto de datos para comprender su estructura, tamaño y tipos de variables incluidas.

```
df.head()
```

	Sexo	Edad	Estatura	Peso	FamiliarConSobrepeso	ComeMuchasCalorias	ComeVegetales	Fumador	ConsumoDeAgua	NivelDeObesidad
0	Female	21.0	1.62	64.0		yes	no	2.0	no	2.0
1	Female	21.0	1.52	56.0		yes	no	3.0	yes	3.0
2	Male	23.0	1.80	77.0		yes	no	2.0	no	2.0
3	Male	27.0	1.80	87.0		no	no	3.0	no	2.0
4	Male	22.0	1.78	89.8		no	no	2.0	no	2.0

```
df.shape
```

```
(2111, 10)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   Sexo              2111 non-null   object  
 1   Edad              2111 non-null   float64 
 2   Estatura          2111 non-null   float64 
 3   Peso              2111 non-null   float64 
 4   FamiliarConSobrepeso  2111 non-null   object  
 5   ComeMuchasCalorias  2111 non-null   object  
 6   ComeVegetales      2111 non-null   float64 
 7   Fumador            2111 non-null   object  
 8   ConsumoDeAgua      2111 non-null   float64 
 9   NivelDeObesidad     2111 non-null   object  
dtypes: float64(5), object(5)
memory usage: 165.1+ KB
```

El conjunto de datos incluye variables cuantitativas como edad, estatura y peso, así como variables cualitativas relacionadas con hábitos alimenticios, consumo de agua, tabaquismo y nivel de obesidad. La base de datos presenta información consistente y no se observan valores faltantes relevantes.

### ✓ 3. Análisis descriptivo de variables numéricas

Aquí se calcularon estadísticas descriptivas para las variables numéricas con la finalidad de identificar sus valores promedio, rangos y dispersión. Estas medidas permiten obtener una visión general de las características físicas de la población analizada.

```
df.describe()
```

	Edad	Estatura	Peso	ComeVegetales	ConsumoDeAgua
<b>count</b>	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000
<b>mean</b>	24.312600	1.701677	86.586058	2.419043	2.008011
<b>std</b>	6.345968	0.093305	26.191172	0.533927	0.612953
<b>min</b>	14.000000	1.450000	39.000000	1.000000	1.000000
<b>25%</b>	19.947192	1.630000	65.473343	2.000000	1.584812
<b>50%</b>	22.777890	1.700499	83.000000	2.385502	2.000000
<b>75%</b>	26.000000	1.768464	107.430682	3.000000	2.477420
<b>max</b>	61.000000	1.980000	173.000000	3.000000	3.000000

```
df.describe(include="object")
```

	Sexo	FamiliarConSobrepeso	ComeMuchasCalorías	Fumador	NivelDeObesidad
<b>count</b>	2111	2111	2111	2111	2111
<b>unique</b>	2	2	2	2	7
<b>top</b>	Male	yes	yes	no	Obesity_Type_I
<b>freq</b>	1068	1726	1866	2067	351

#### ▼ 4. Variable de interés para inferencia

Seleccioné el nivel de obesidad como variable de salida del estudio ya que representa el principal fenómeno de interés. Esta variable es categórica y clasifica a los individuos en distintos niveles de peso corporal.

```
df["NivelDeObesidad"].value_counts()
```

	count
<b>NivelDeObesidad</b>	
Obesity_Type_I	351
Obesity_Type_III	324
Obesity_Type_II	297
Overweight_Level_I	290
Overweight_Level_II	290
Normal_Weight	287
Insufficient_Weight	272

```
dtype: int64
```

La distribución del nivel de obesidad muestra que una proporción importante de los individuos se encuentra en la categoría de peso normal, mientras que otros presentan sobrepeso u obesidad en distintos niveles.

#### ▼ 5. Análisis gráfico por subpoblación

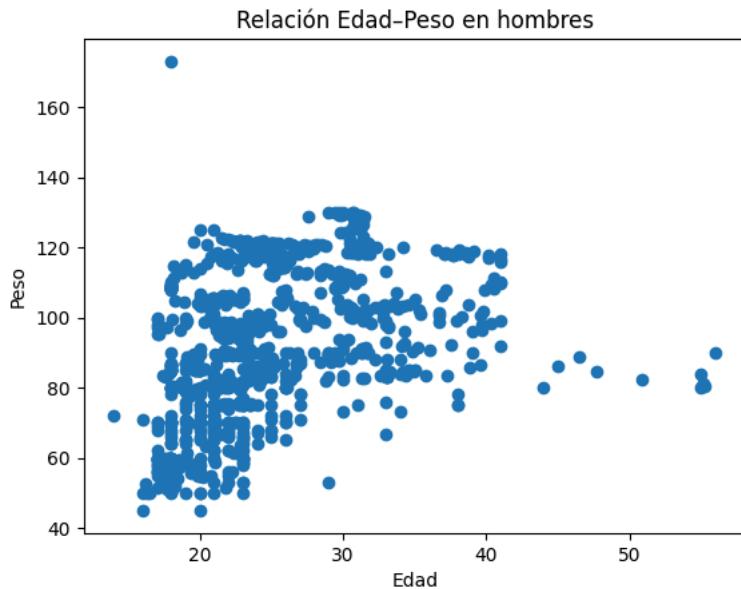
Se seleccionó la subpoblación masculina para analizar la relación entre la edad y el peso corporal mediante una gráfica de dispersión.

```
import matplotlib.pyplot as plt

hombres = df[df["Sexo"] == "Male"]

plt.scatter(hombres["Edad"], hombres["Peso"])
plt.xlabel("Edad")
plt.ylabel("Peso")
plt.title("Relación Edad-Peso en hombres")
```

```
plt.show()
```



La gráfica muestra una alta dispersión del peso corporal a diferentes edades lo que indica que el peso no depende únicamente de la edad. Esto sugiere que otros factores como hábitos alimenticios y estilo de vida, influyen en el nivel de obesidad.

## ▼ Conclusión

Para finalizar esta actividad puedo decir que el análisis exploratorio permitió comprender la estructura del conjunto de datos y analizar la relación entre distintas variables asociadas a la obesidad. Los resultados sugieren que el nivel de obesidad está influenciado por múltiples factores por lo que un análisis inferencial más profundo podría aportar conclusiones relevantes.

Start coding or [generate](#) with AI.