

PREGUNTA 1. Seleccione un data set de datos tabulares (UCI, KAAGLE). Realice lo siguiente:

Conjunto de datos (Data Set):

PREDICCIÓN DE OBESIDAD



Descripción: El conjunto de datos de predicción de la obesidad proporciona una colección completa de atributos relacionados con la demografía, los hábitos de estilo de vida y los indicadores de salud de los individuos, con el objetivo de facilitar la predicción de la prevalencia de la obesidad. Este conjunto de datos ofrece un recurso valioso para investigadores, científicos de datos y profesionales de la salud interesados en explorar la compleja interacción de los factores que contribuyen a la obesidad y desarrollar estrategias de intervención efectivas.

El conjunto de datos incluye variables como edad, sexo, altura, peso, índice de masa corporal (IMC), nivel de actividad física y categoría de obesidad, lo que ofrece información valiosa sobre los factores que influyen en los resultados de la obesidad.

Edad: La edad del individuo, expresada en años.

Sexo: El género del individuo, categorizado como masculino o femenino.

Altura: La altura del individuo, generalmente medida en centímetros o pulgadas.

Peso: El peso del individuo, generalmente medido en kilogramos o libras.

Índice de masa corporal (IMC): Una métrica calculada derivada del peso y la altura del individuo.

Nivel de actividad física: Esta variable cuantifica el nivel de actividad física del individuo por semana.

Categoría de obesidad: Categorización de individuos según su IMC en diferentes categorías de obesidad peso normal, obeso, exceso peso y bajo peso.

a) Con Python sin uso de librerías, calcule del ultimo cuartil, percentil 80 por columna; explique qué significa en cada caso.

<pre>Edad Percentil 80% = 69 Ultimo Cuartil 75% = 66 -----</pre>	<p>El percentil del 80% de la primera columna (Edad) es 69, lo que significa que el 80% de las personas tienen una edad igual o menor a 69 años.</p> <p>El ultimo cuartil (75%) de la primera columna (Edad) es 66, lo que significa que el 75% de las personas tienen una edad igual o menor a 66 años.</p>
---	--

<pre> Altura Percentil 80% = 178.8029624925325 Ultimo Cuartil 75% = 177.36757755441087 ----- </pre>	<p>El percentil del 80% de la tercera columna (Altura) es 178.803, lo que significa que el 80% de las personas tienen una altura igual o menor a 178.803 cm.</p> <p>El ultimo cuartil (75%) de la tercera columna (Altura) es 177.368, lo que significa que el 75% de las personas tienen una altura igual o menor a 177.368 cm.</p>
<pre> Peso Percentil 80% = 84.29536220717854 Ultimo Cuartil 75% = 81.14124942680387 ----- </pre>	<p>El percentil del 80% de la cuarta columna (Peso) es 84.295, lo que significa que el 80% de las personas tienen un peso igual o menor a 84.295 kg.</p> <p>El ultimo cuartil (75%) de la cuarta columna (Peso) es 81.141, lo que significa que el 75% de las personas tienen un peso igual o menor a 81.141 kg.</p>
<pre> IMC Percentil 80% = 29.75623218349644 Ultimo Cuartil 75% = 28.73432471904472 ----- </pre>	<p>El percentil del 80% de la quinta columna (IMC) es 29.756, lo que significa que el 80% de las personas tienen un IMC igual o menor a 29.756.</p> <p>El ultimo cuartil (75%) de la quinta columna (ICM) es 28,734, lo que significa que el 75% de las personas tienen un IMC igual o menor a 28.734.</p>
<pre> ActividadFisica Percentil 80% = 4 Ultimo Cuartil 75% = 4 ----- </pre>	<p>El percentil del 80% de la sexta columna (Actividad física) es 4, lo que significa que el 80% de las personas realiza 4 veces o menos alguna actividad física por semana.</p> <p>El ultimo cuartil (75%) de la sexta columna (Actividad física) es 4, lo que significa que el 75% de las personas realiza 4 veces o menos alguna actividad física por semana</p>

El percentil y cuartil me ayudaran a comprender la distribución de las variables en un conjunto de datos.

```

Edad
Percentil      80% = 69
Ultimo Cuartil 75% = 66
-----
Altura
Percentil      80% = 178.8029624925325
Ultimo Cuartil 75% = 177.36757755441087
-----
Peso
Percentil      80% = 84.29536220717854
Ultimo Cuartil 75% = 81.14124942680387
-----
IMC
Percentil      80% = 29.75623218349644
Ultimo Cuartil 75% = 28.73432471904472
-----
ActividadFisica
Percentil      80% = 4
Ultimo Cuartil 75% = 4
-----

```

b) Realice lo mismo del inciso a) con el uso de numpy y pandas.

```
Edad
Percentil      80% = 69.0
Ultimo Cuartil 75% = 66.0
-----
Altura
Percentil      80% = 178.77740832879684
Ultimo Cuartil 75% = 177.35359566466025
-----
Peso
Percentil      80% = 84.1177364133045
Ultimo Cuartil 75% = 81.13374576773099
-----
IMC
Percentil      80% = 29.751750539831033
Ultimo Cuartil 75% = 28.73213190988069
-----
ActividadFisica
Percentil      80% = 4.0
Ultimo Cuartil 75% = 4.0
-----
```

Se logra observar que los resultados del percentil 80% y el ultimo cuartil 75% son iguales a los resultados del inciso a) y la explicación de dichos resultados serían prácticamente los mismos.

c) Obtenga la media, mediana, moda y geométrica; explique la diferencia de los resultados y cuál de ellas se puede utilizar en un artículo científico.

Primera Columna (Edad)

```
Edad
->Media      = 49.857
->Mediana     = 50.0
->Media Geometrica = 46.13269783308886
->Moda        = 79
-----
```

El promedio de edades de las personas están alrededor de 49 años. La mediana edad de las personas es de 50, esto significa que el 50% de las edades son menores o iguales a 50 años y el otro 50% son mayores a 50 años. La moda de edad es de 79 años, lo que significa que la edad más común o la edad que más se repite es de 79 años. Y la media geométrica de edad es de aproximadamente 46 años que representa una tendencia central.

Segunda Columna (Sexo)

```
Sexo
->Moda      = Masculino
-----
```

Dado que no se proporcionaron datos numéricos para calcular la media, mediana o media geométrica en relación con la variable "Sexo", solo se calculó la moda en este caso. La moda indica que la mayoría de las observaciones realizadas en el conjunto de datos sobre personas obesas corresponden al sexo masculino. Esto sugiere que hay una predominancia de hombres en las observaciones registradas en relación con la obesidad.

Tercera Columna (Altura)

```
Altura
->Media      = 170.05241692306717
->Mediana     = 169.80166537248084
->Media Geometrica = 169.73886007129428
->Moda        = 173.5752624383722
-----
```

El promedio de altura de las personas están alrededor de 170.05 cm. La mediana altura de las personas es de 169.80 cm, esto significa que el 50% de las estaturas son menores o iguales a 169.80 cm y el otro 50% son mayores a 169.80 cm. La moda de altura es de 173.57 cm, lo que significa que la estatura más común o la altura que más se repite es de 173.57 cm. Y la media geométrica de la altura es de aproximadamente 169.73 cm que representa una tendencia central.

Cuarta Columna (Peso)

```
Peso
->Media           = 71.20576903878789
->Mediana          = 71.9290717998748
->Media Geometrica = 69.37236501378806
->Moda             = 71.98205082003972
-----
```

El promedio de pesos de las personas están alrededor de 71.20 kg. La mediana de los pesos de las personas es de 71.92 kg, esto significa que el 50% de los pesos son menores o iguales a 71.92 kg y el otro 50% son mayores a 71.92 kg. La moda del peso es de 71.98 kg, lo que significa que el peso más común o el peso que más se repite es de 71.98 kg. Y la media geométrica del peso es de aproximadamente 69.37 kg que representa una tendencia central.

Quinta Columna (IMC)

```
IMC
->Media           = 24.888316768452263
->Mediana          = 24.69864656152483
->Media Geometrica = 24.078195597358576
->Moda             = 23.89178262396797
-----
```

El promedio de IMC (Índice de masa corporal) de las personas están alrededor de 24.88. La mediana del IMC de las personas es de 24.69, esto significa que el 50% de los IMC son menores o iguales a 24.69 y el otro 50% son mayores a 24.69. La moda del IMC es de 23.89, lo que significa que el IMC más común o el índice de masa corporal que más se repite es de 23.89. Y la media geométrica del IMC es de aproximadamente 24.07 que representa una tendencia central.

Sexta Columna (Actividad Física)

```
ActividadFisica
->Media           = 2.534
->Mediana          = 3.0
->Media Geometrica = 2.248815710460743
->Moda             = 4
-----
```

El promedio de actividad física de las personas están alrededor de 2 veces por semana. La mediana de la actividad física es de 3, esto significa que el 50% de las personas realizan 3 o menos veces actividades físicas por semana y el otro 50% realizan más de 3 veces por semana actividades físicas. La moda de la actividad física es de 4, lo que significa que la cantidad más común de veces que se realiza actividades físicas por semana es de 4. Y la media geométrica de la actividad física es de aproximadamente 2 que representa una tendencia central.

Séptima Columna (Categoría Obesidad)

```
CategoriaObesidad
->Moda             = Peso normal
-----
```

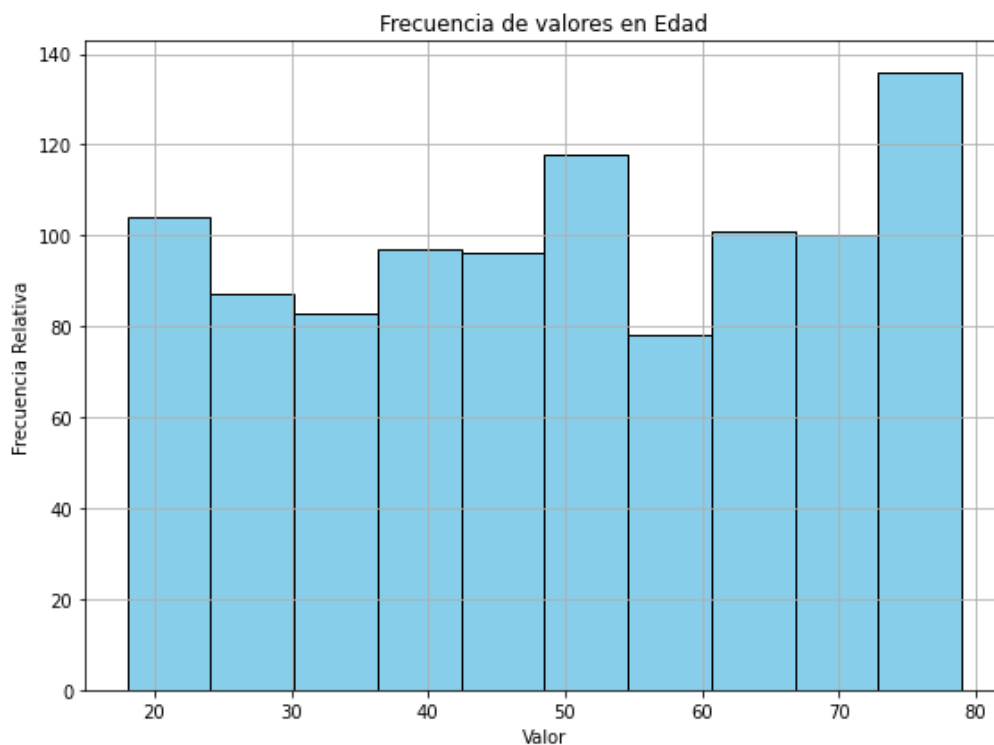
Dado que no se proporcionaron datos numéricos para calcular la media, mediana o media geométrica en relación con la variable clase o variable objetivo "Categoría obesidad", solo se calculó la moda en este caso. La moda indica que la mayoría de las observaciones realizadas en el conjunto de datos sobre personas obesas corresponden a que se encuentran en su peso normal. Esto sugiere que hay una predominancia de individuos con un peso normal en las observaciones registradas en relación con la obesidad.

¿Cuál de ellas se puede utilizar en un artículo científico?

La mediana de la variable IMC (índice de masa corporal) se puede utilizar en un artículo científico porque es una medida comúnmente utilizada para evaluar el estado nutricional y el riesgo de enfermedades relacionadas con el peso. Podría ser relevante para investigaciones sobre obesidad, salud metabólica, enfermedades cardiovasculares y otros temas relacionados con la salud pública.

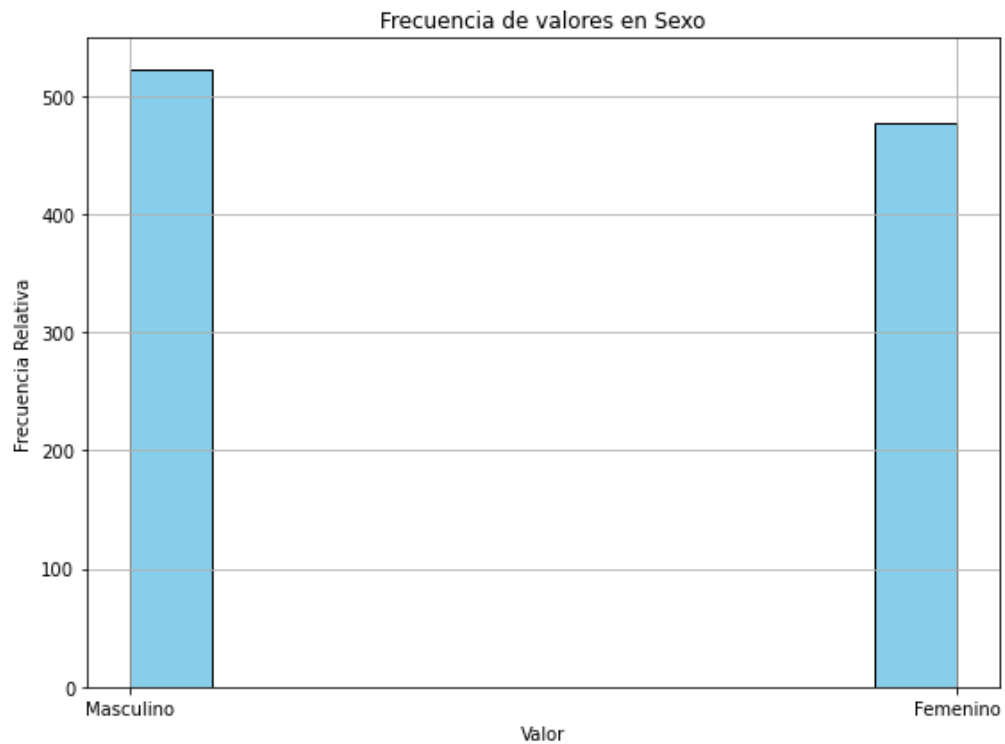
d) Grafique los datos y explique su comportamiento (PYTHON).

Grafica 1: de la variable edad



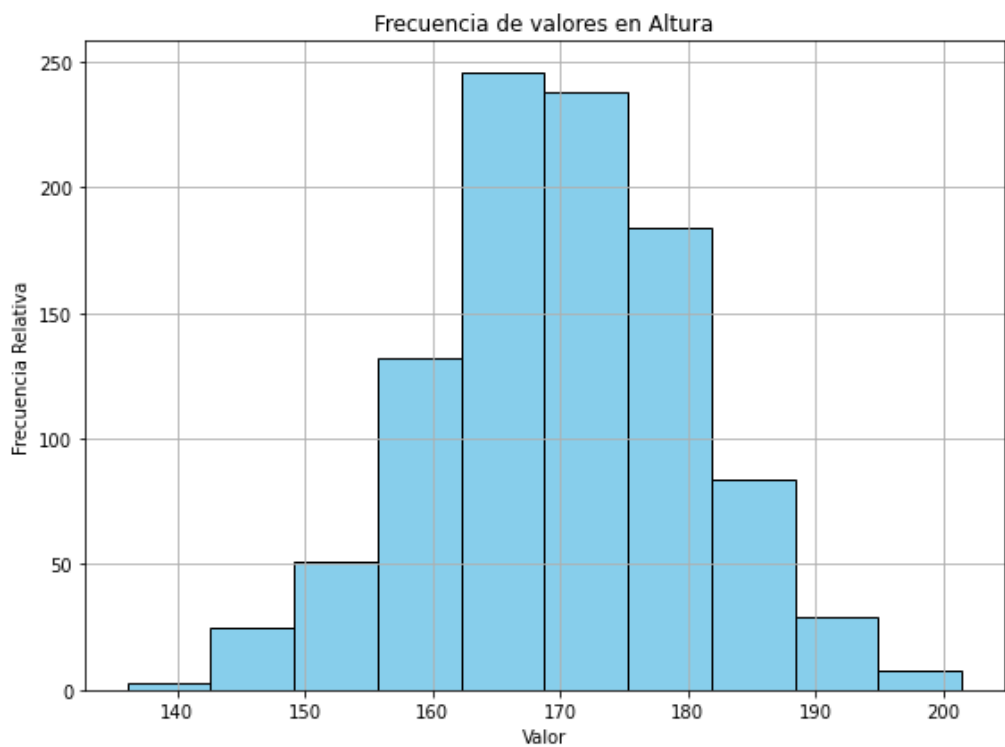
En el grafico se observa cómo se distribuyen las diferentes edades en grupos o segmentos. Cada segmento en el histograma representa un rango específico de edades, y la altura de cada segmento indica la proporción de la muestra que cae dentro de ese rango de edades en particular, es decir nos indica cuántos individuos de nuestro conjunto datos se encuentran dentro de ese rango específico de edades.

Grafica 2: de la variable sexo



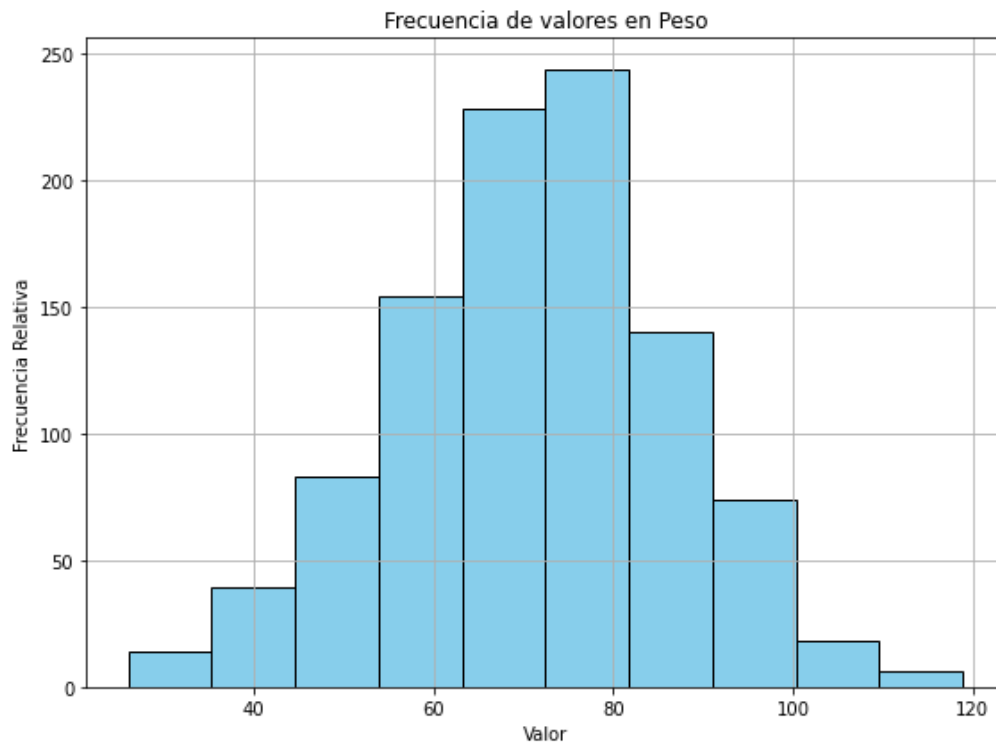
La grafica nos permite ver la proporción de hombres y mujeres en nuestro conjunto de datos.

Grafica 3: de la variable altura



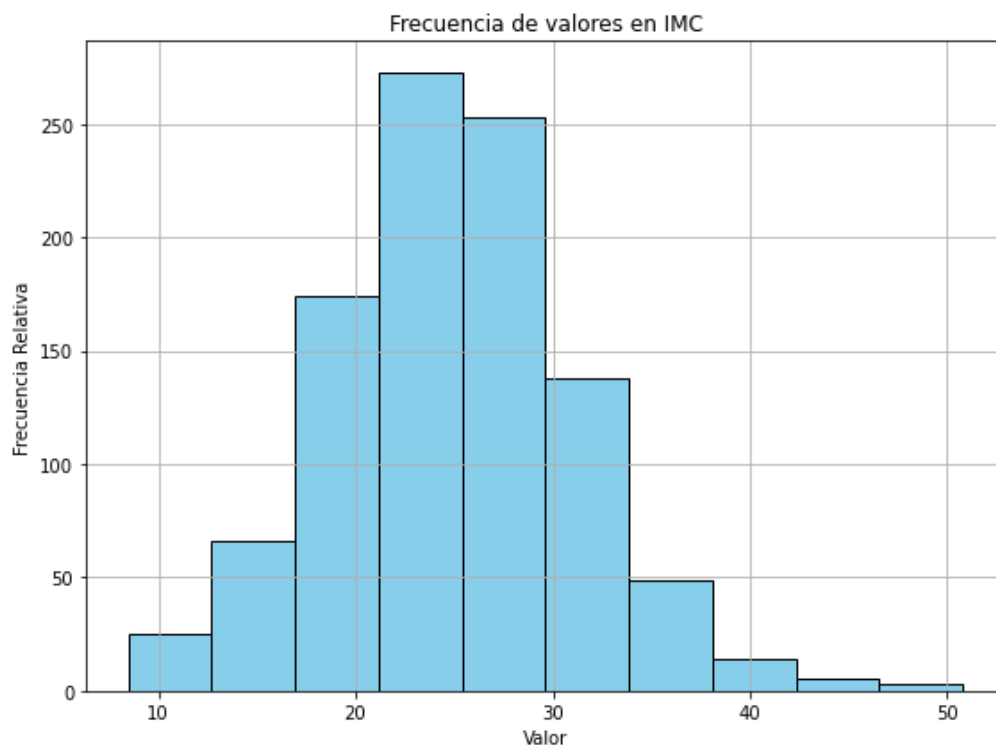
En el grafico se observa cómo se distribuyen las diferentes alturas de los individuos en segmentos. Cada segmento en el histograma representa un rango específico de alturas, y la altura de cada segmento nos indica cuántos individuos de nuestro conjunto de datos se encuentran dentro de ese rango específico de alturas.

Grafica 4: de la variable peso



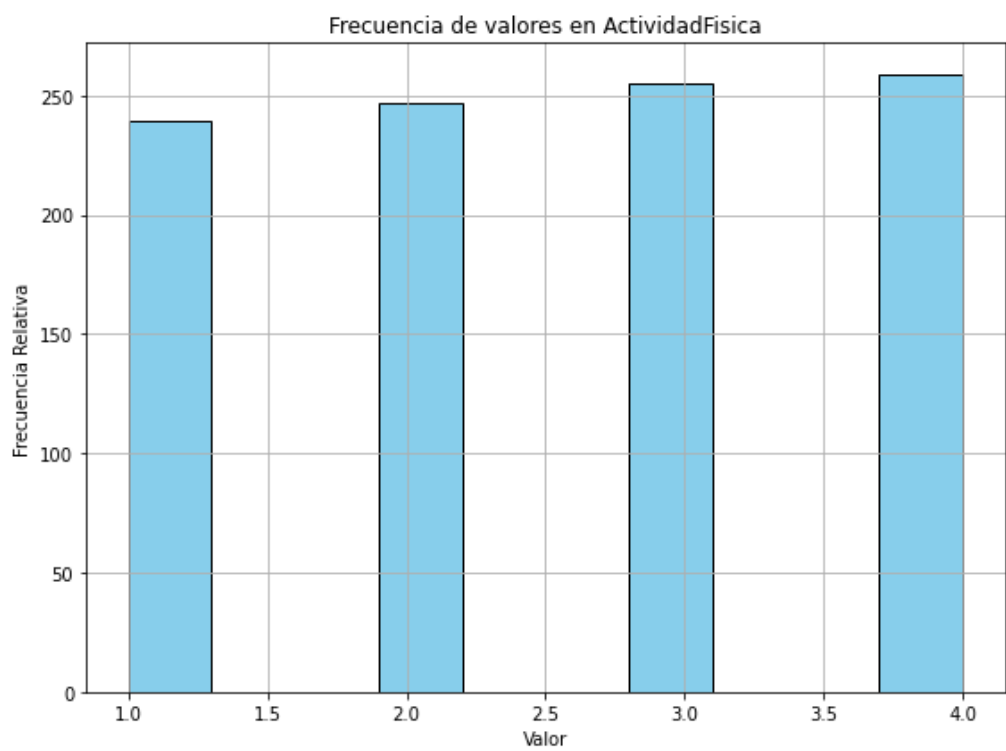
En el grafico se observa cómo se distribuyen los diferentes pesos de los individuos en segmentos. Cada segmento en el histograma representa un rango específico de pesos, y la altura de cada segmento nos indica cuántos individuos de nuestro conjunto datos se encuentran dentro de ese rango específico de pesos.

Grafica 5: de la variable IMC



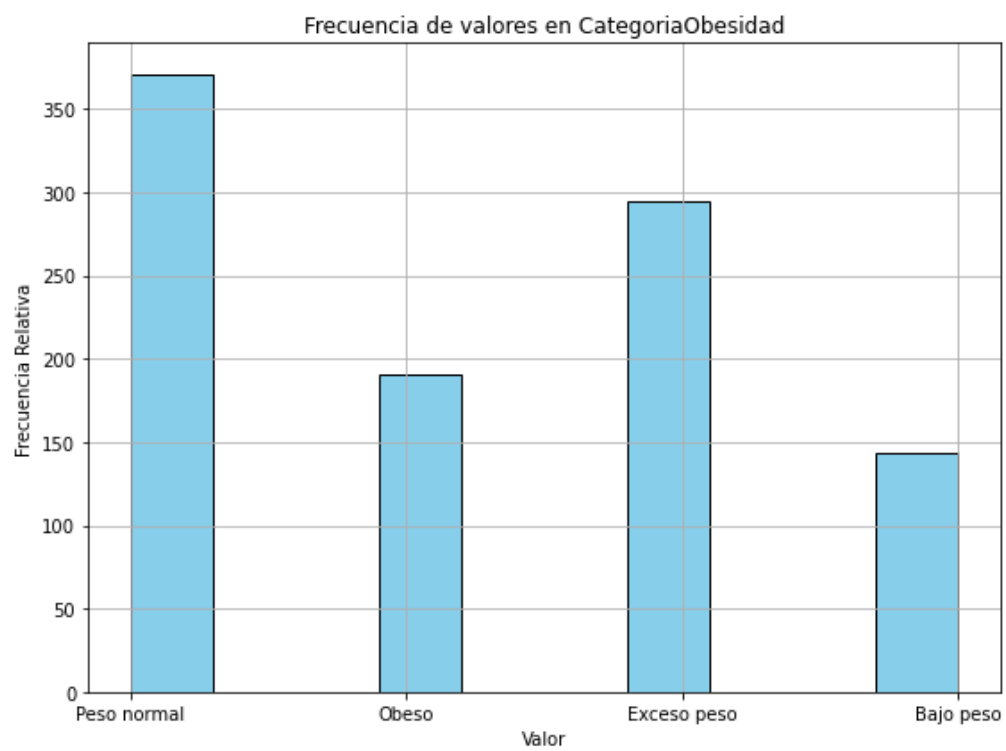
En el grafico se observa cómo se distribuyen los diferentes IMC (Índice de masa corporal) de los individuos en segmentos. Cada segmento en el histograma representa un rango específico de índices de masas corporales, y la altura de cada segmento nos indica cuántos individuos de nuestros conjunto datos se encuentran dentro de ese rango específico de índices de masas corporales.

Grafica 6: de la variable actividad física



En el grafico se observa cómo se distribuyen el número de veces que una persona realiza actividades físicas por semana en segmentos. Cada segmento en el histograma representa un rango específico de cantidad de veces que realiza alguna actividad física, y la altura de cada segmento nos indica cuántos individuos de nuestro conjunto de datos se encuentran dentro de ese rango específico actividades físicas.

Grafica 7: de la variable objetivo categoría de obesidad



La grafica nos permite ver la proporción de individuos que pertenecen al grupo de personas con un peso normal, grupo de personas que son obesos, grupo de personas con un exceso de peso o al grupo de personas con un bajo peso.