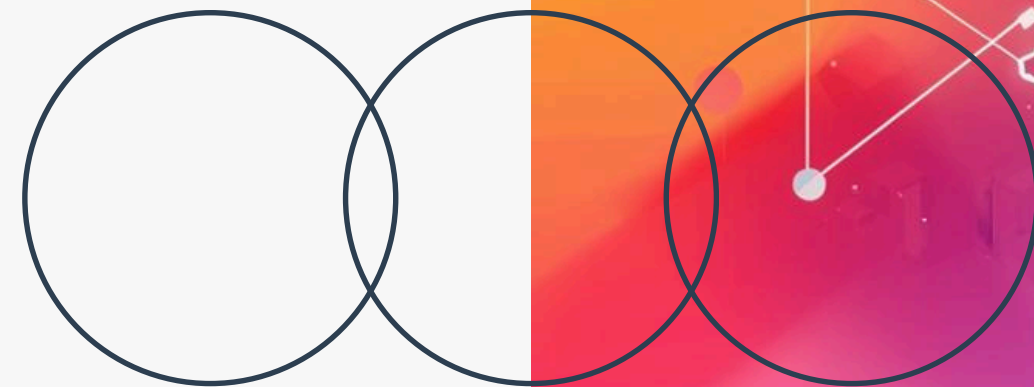
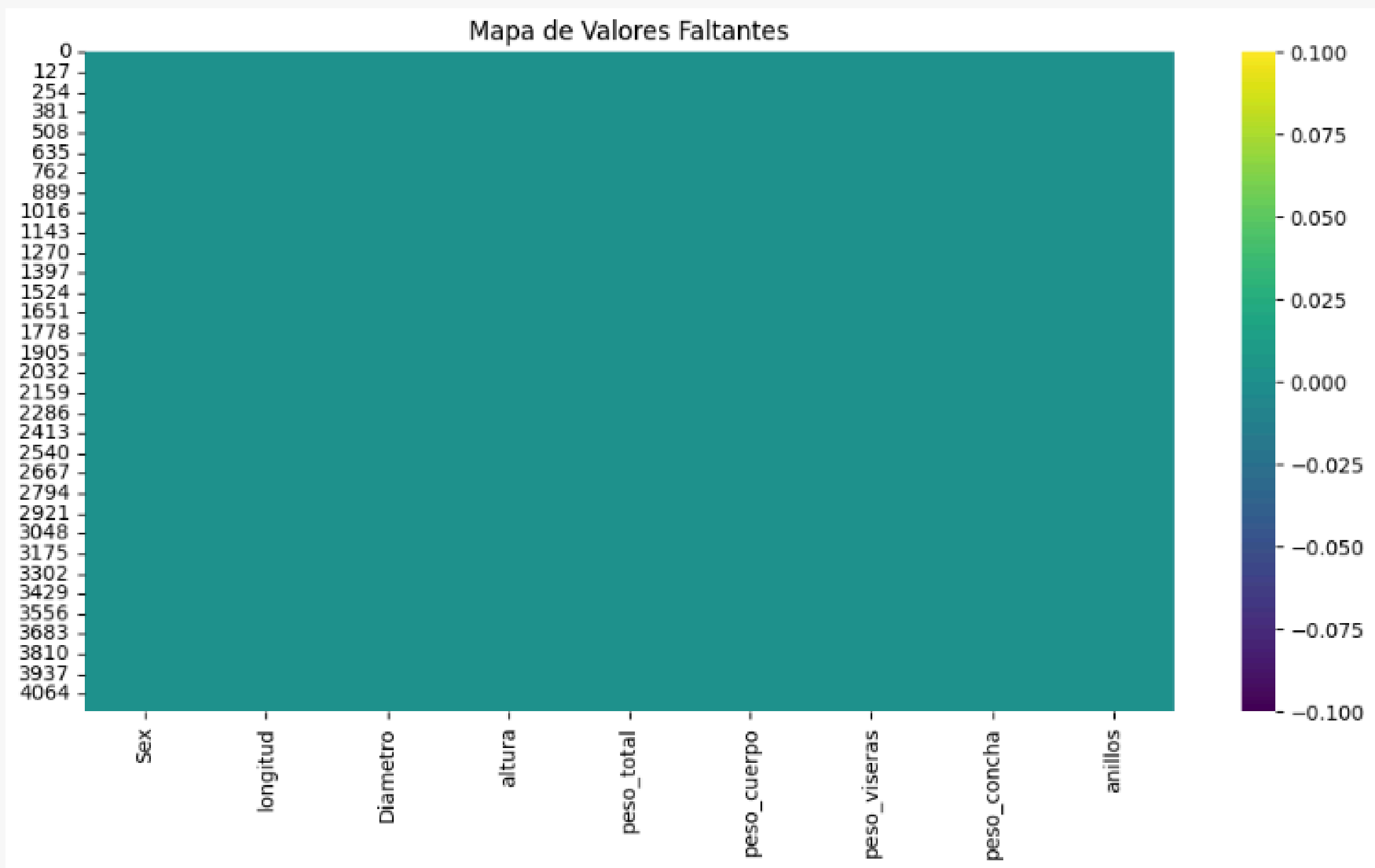


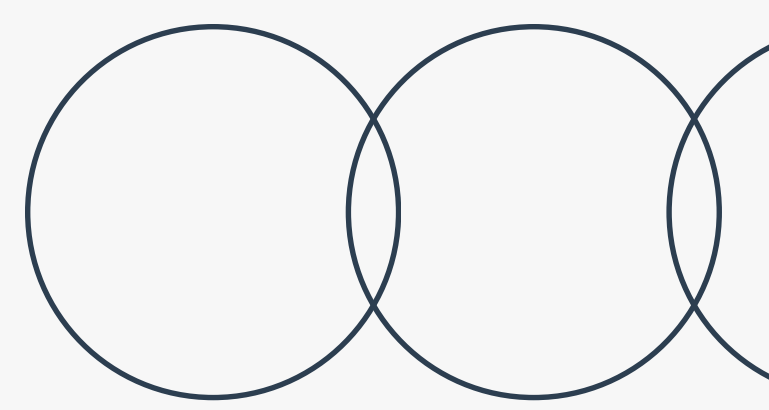
Minería de Datos

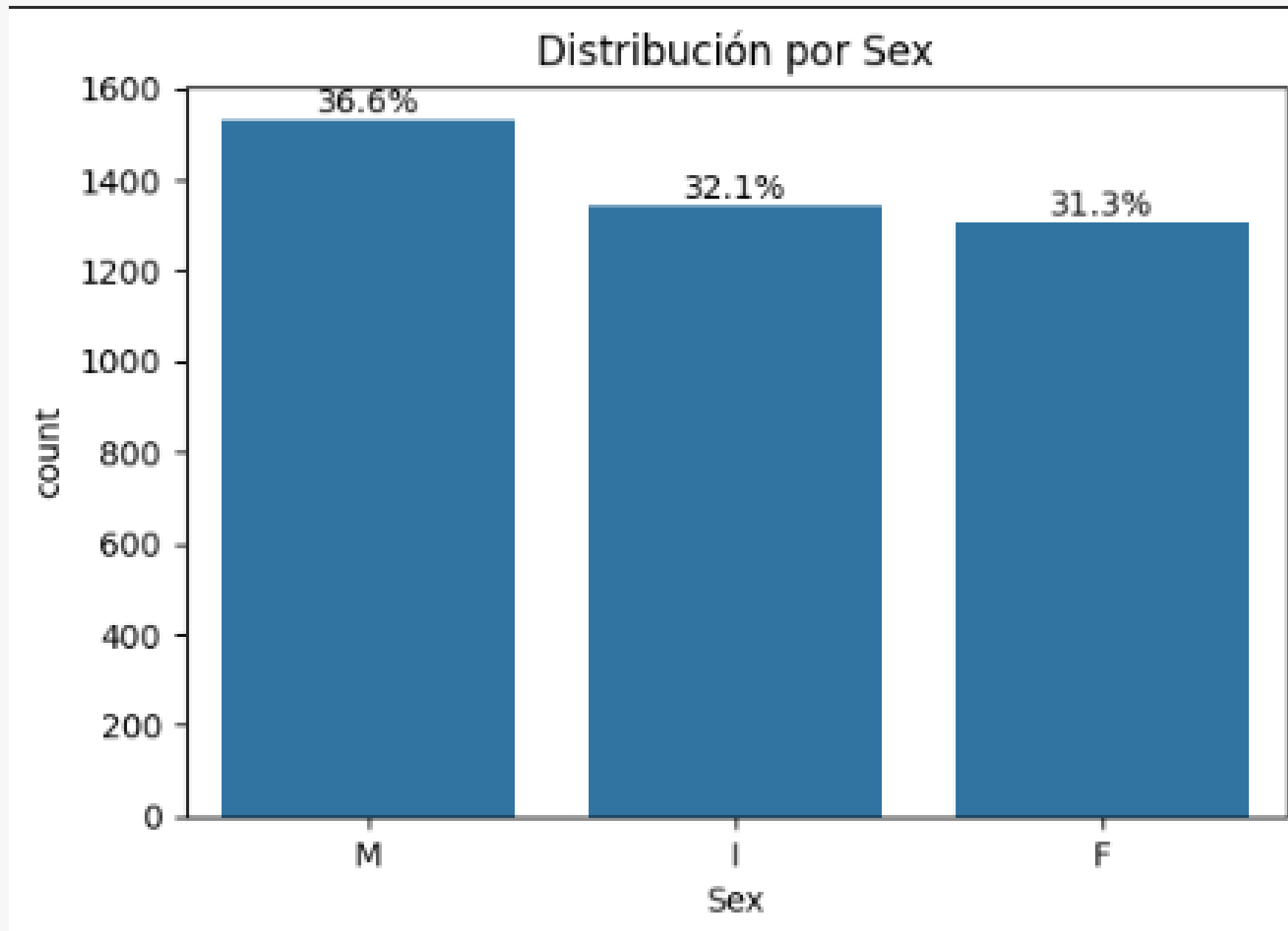
Victor Manuel Brand Cepeda



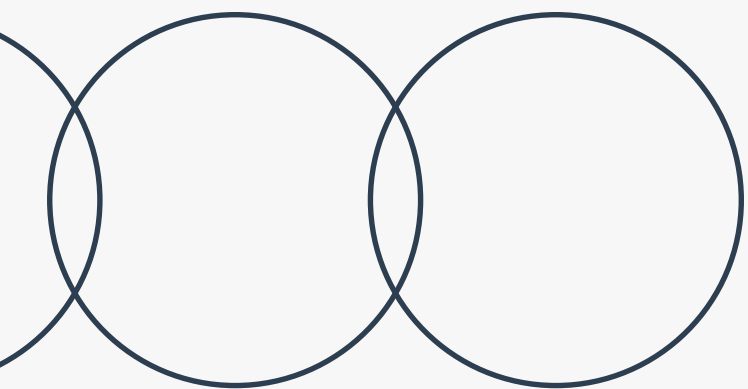


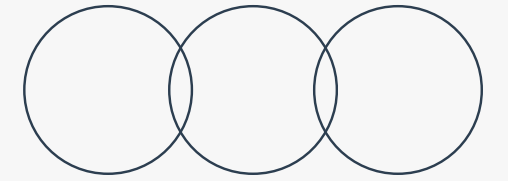
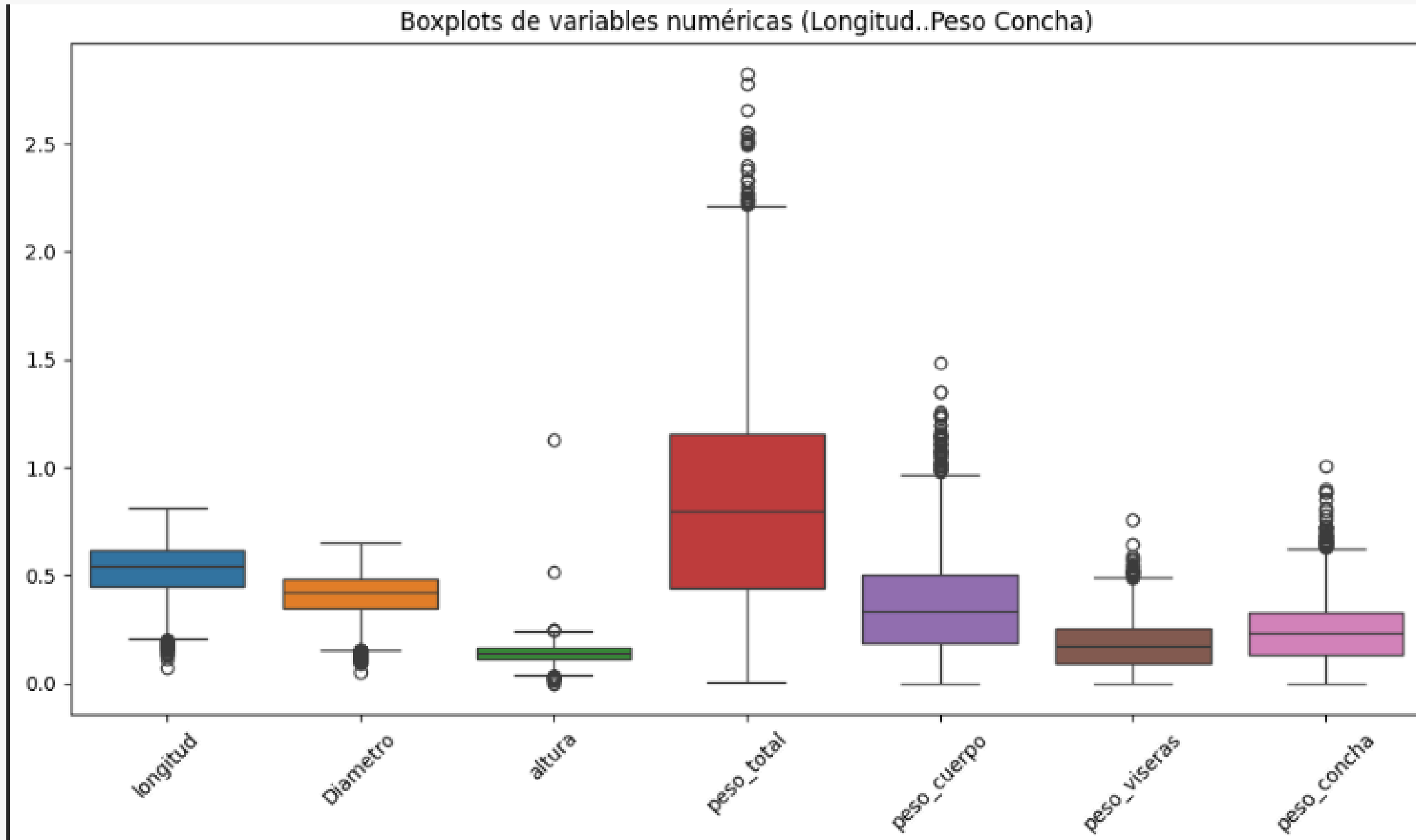
Este diagrama muestra con una marca de calor en los valores faltantes, el mapa de valores faltantes confirma visualmente que el dataset no contiene datos nulos en ninguna de sus variables. Esto simplifica el proceso de limpieza y mejora la confiabilidad del análisis, ya que no se requiere imputación ni eliminación de registros.





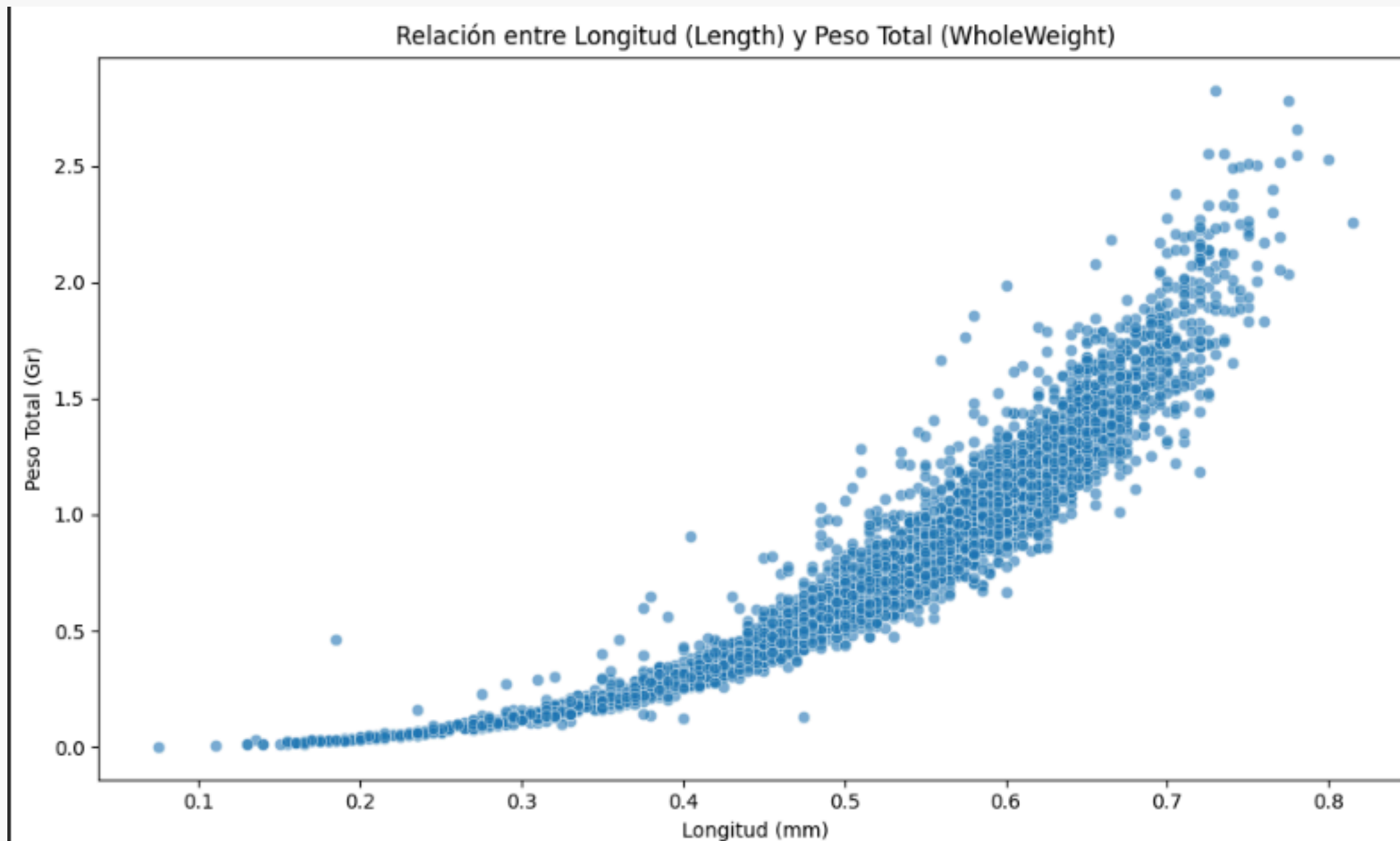
Aquí podemos observar la distribución del sexo de los abulones. Machos y hembras aparecen en proporciones similares, mientras que los infantiles están representados en una cantidad mucho menor. Esto es coherente con la biología de la especie, ya que los juveniles suelen ser difíciles de recolectar por su tamaño pequeño





- La distribución general de cada variable.
- Su mediana y rango intercuartílico (IQR).
- La presencia de valores atípicos (outliers).
- Asimetrías o concentraciones de datos.

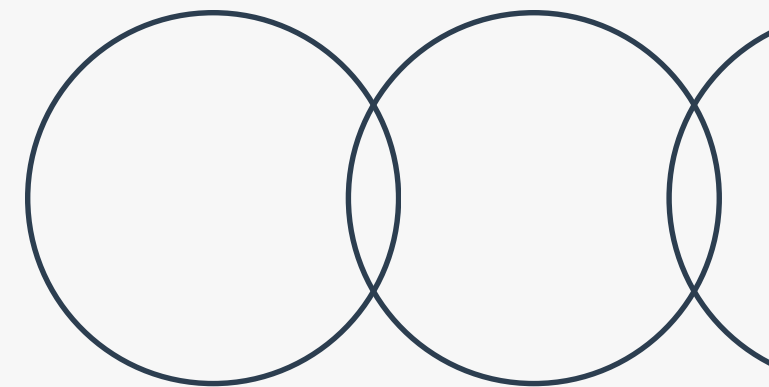
El diagrama boxplot permite identificar la distribución, variabilidad y presencia de valores atípicos. Estos indicadores morfométricos son fundamentales para entender el crecimiento y desarrollo biológico del abulón.

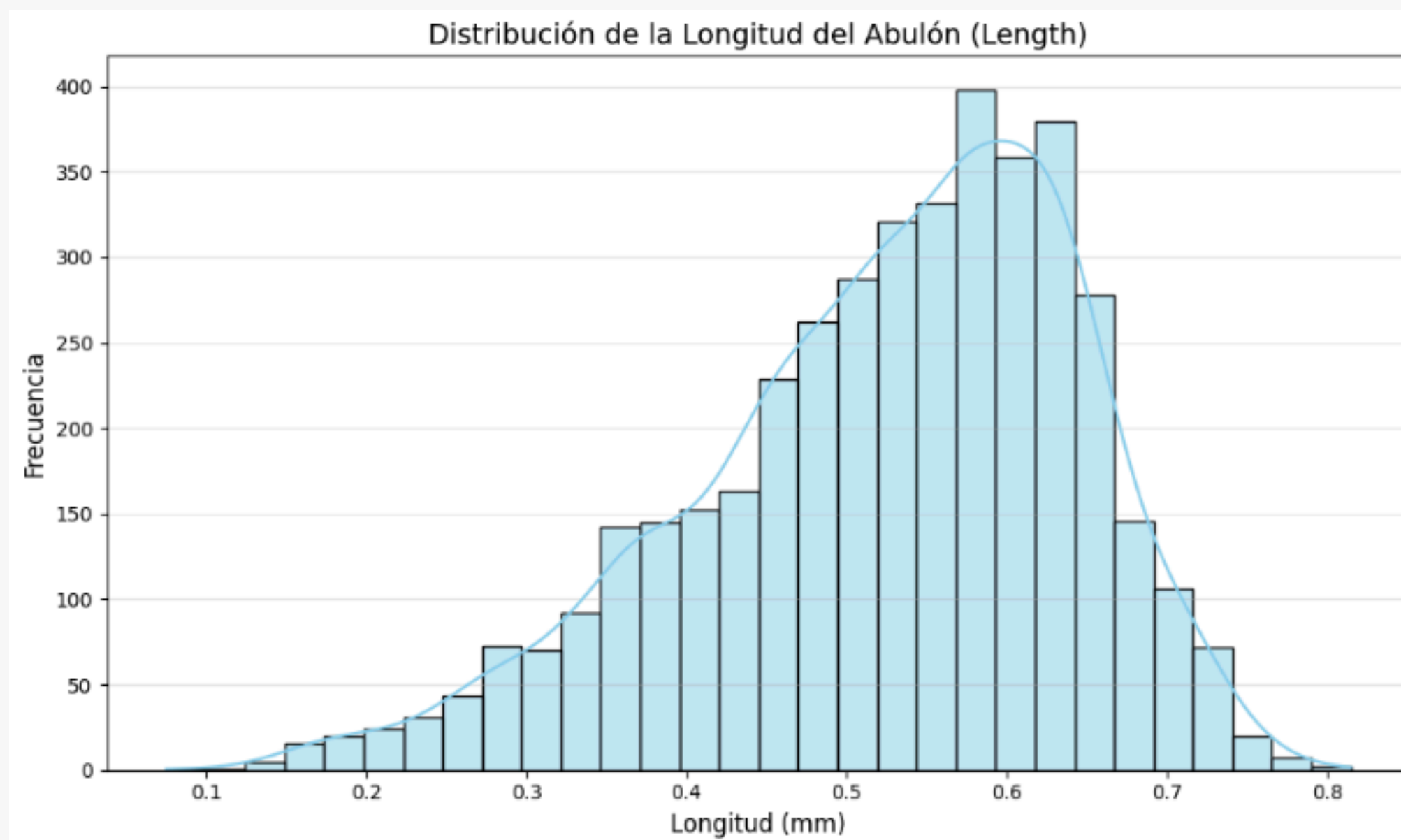
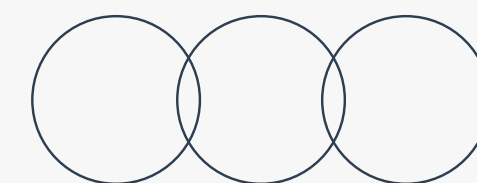


En valores altos de longitud, los puntos están más dispersos:

- Length → Longitud del abulón (mm)
- WholeWeight → Peso total del cuerpo (gramos)

Aunque la relación es positiva, no es perfectamente lineal: hay cierta curvatura que sugiere que el peso aumenta más rápidamente que la longitud cuando el abulón crece.

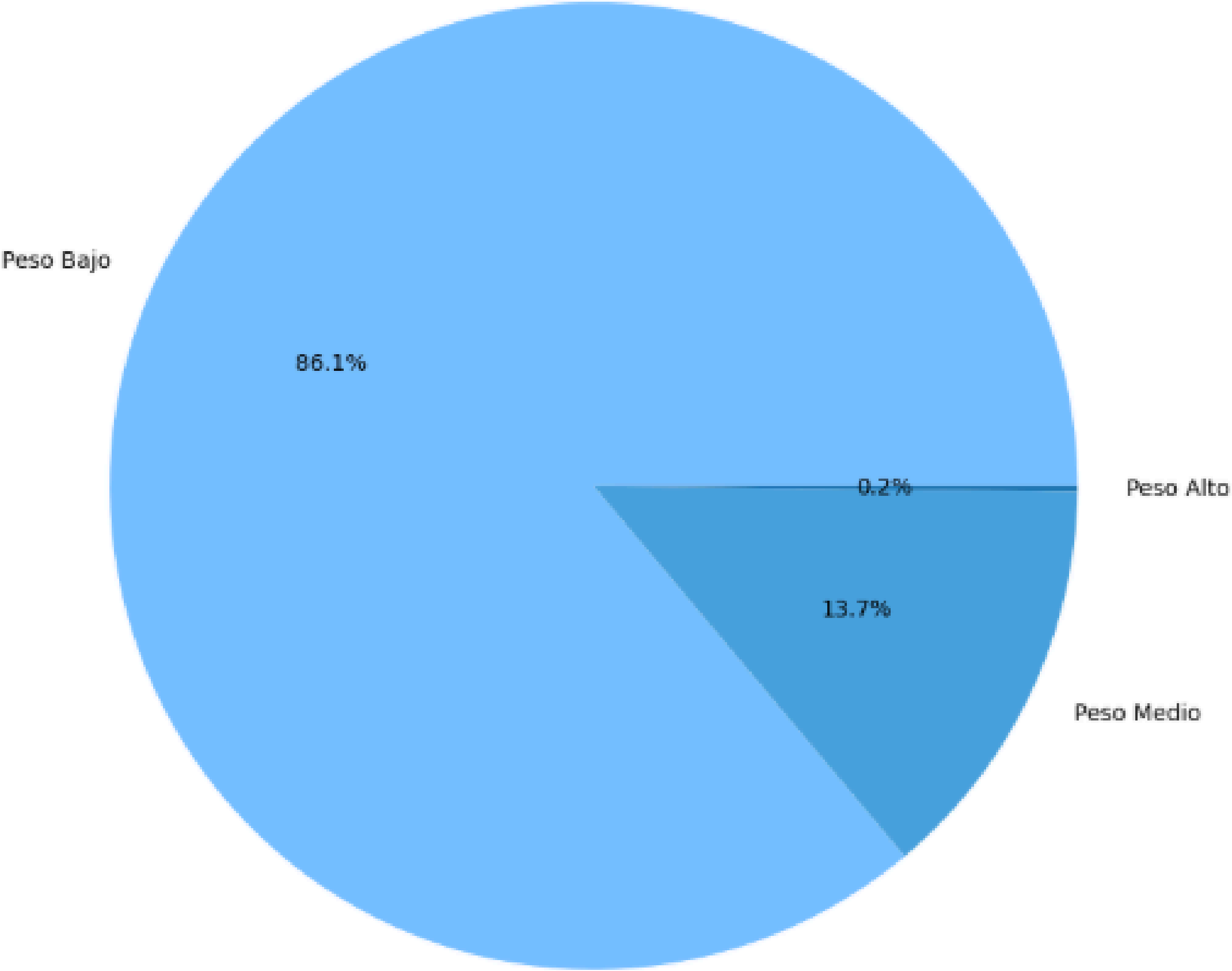
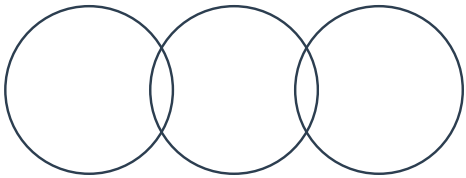




El histograma muestra que la mayor parte de los abulones se concentra en longitudes bajas a moderadas, mientras que los valores altos aparecen con menor frecuencia.

- La forma de la distribución
- La presencia de sesgos
- Los valores más frecuentes
- El comportamiento biológico del tamaño corporal

Proporción de Abulones según Peso Corporal (peso_cuerpo)



Estas categorías permiten visualizar cómo se distribuye la biomasa general de los abulones del conjunto de datos.

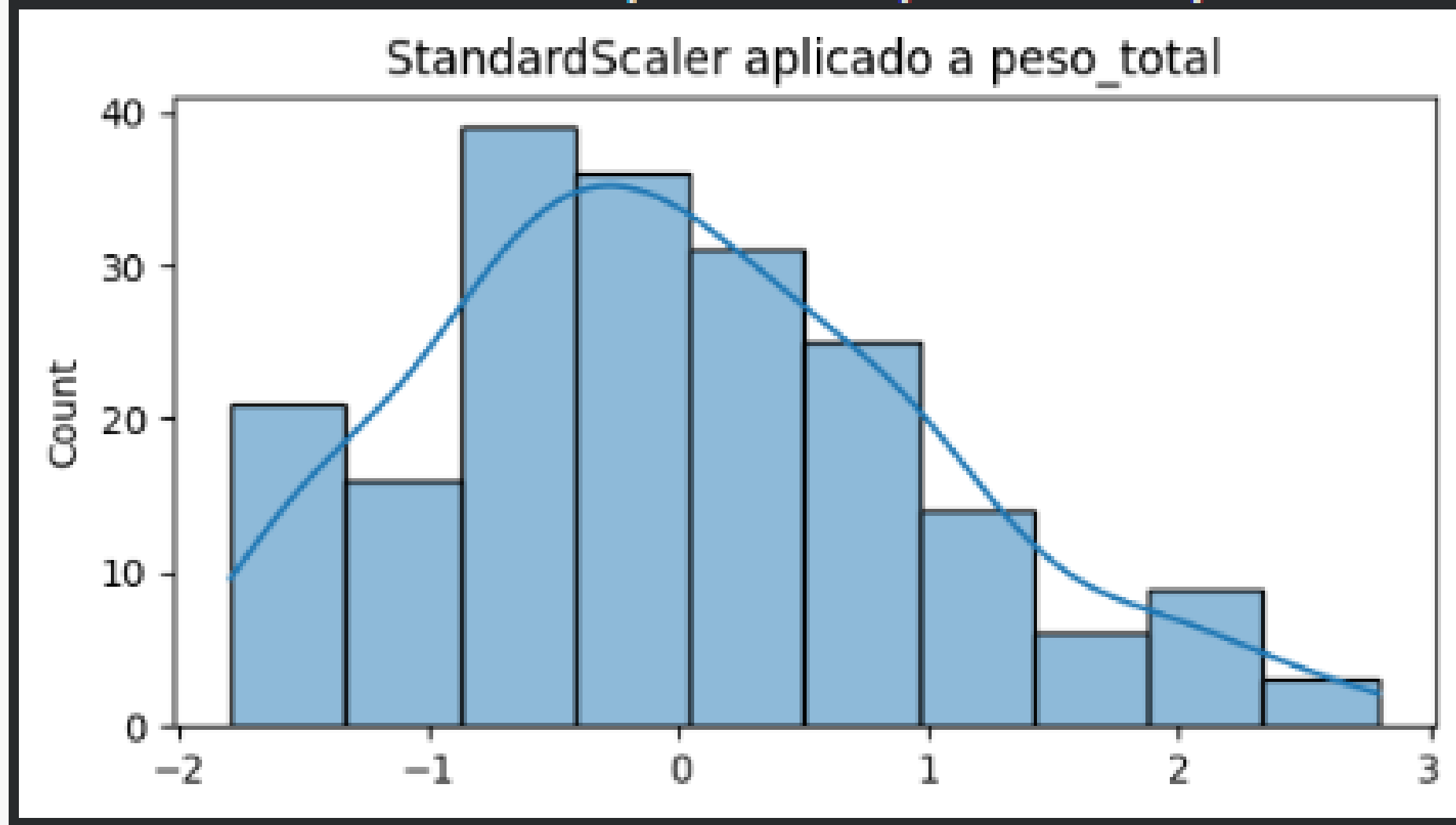
- Peso Bajo
- Peso Medio
- Peso Alto

El segmento más grande corresponde a los abulones con un peso corporal intermedio.

Esto indica que:

- La mayor parte de la población está compuesta por individuos que ya superaron la etapa juvenil.
- Refleja una estructura poblacional típica donde predomina la fase subadulto/adulto joven.

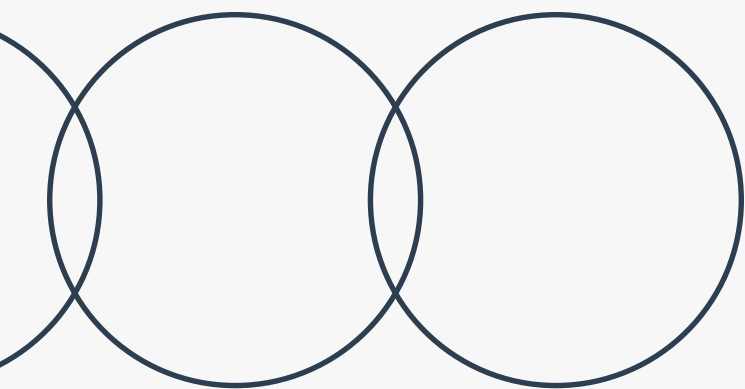
StandardScaler — mean: 0.0000, std: 1.0000, min: -1.7904, max: 2.7952

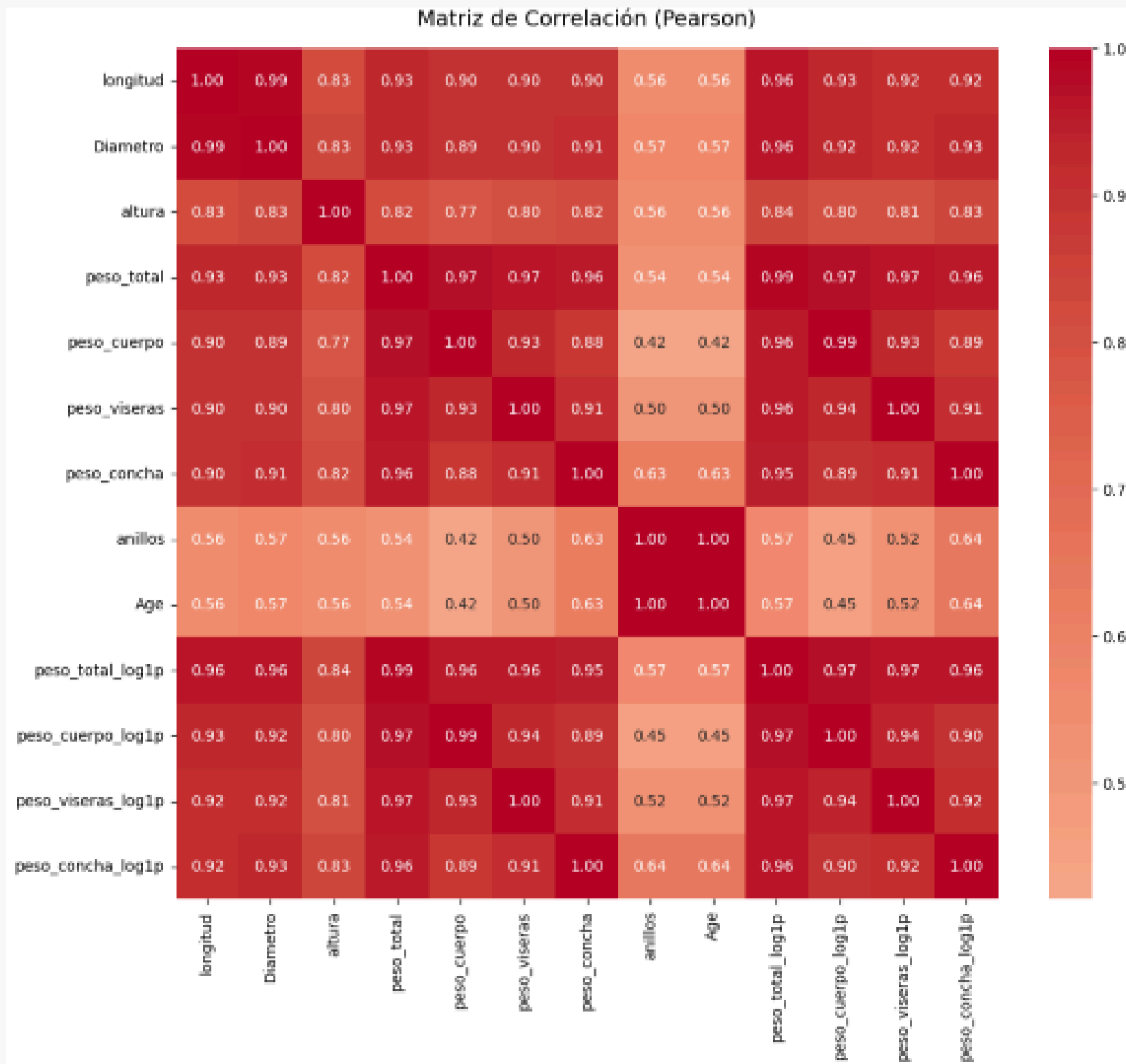


Esto permite observar cómo se redistribuyen los valores del peso total después de eliminar las unidades originales y convertir todo a una escala comparable con otras variables.

- La media se convierte en 0.
- La desviación estándar se convierte en 1.

Por tanto, el uso del StandardScaler mejora la estabilidad y el rendimiento del modelo predictivo.

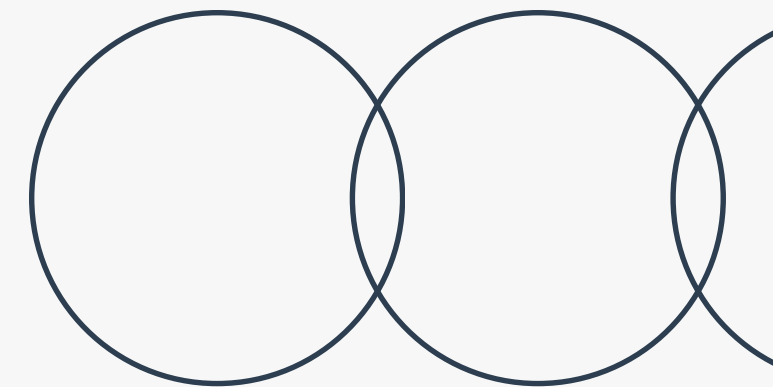


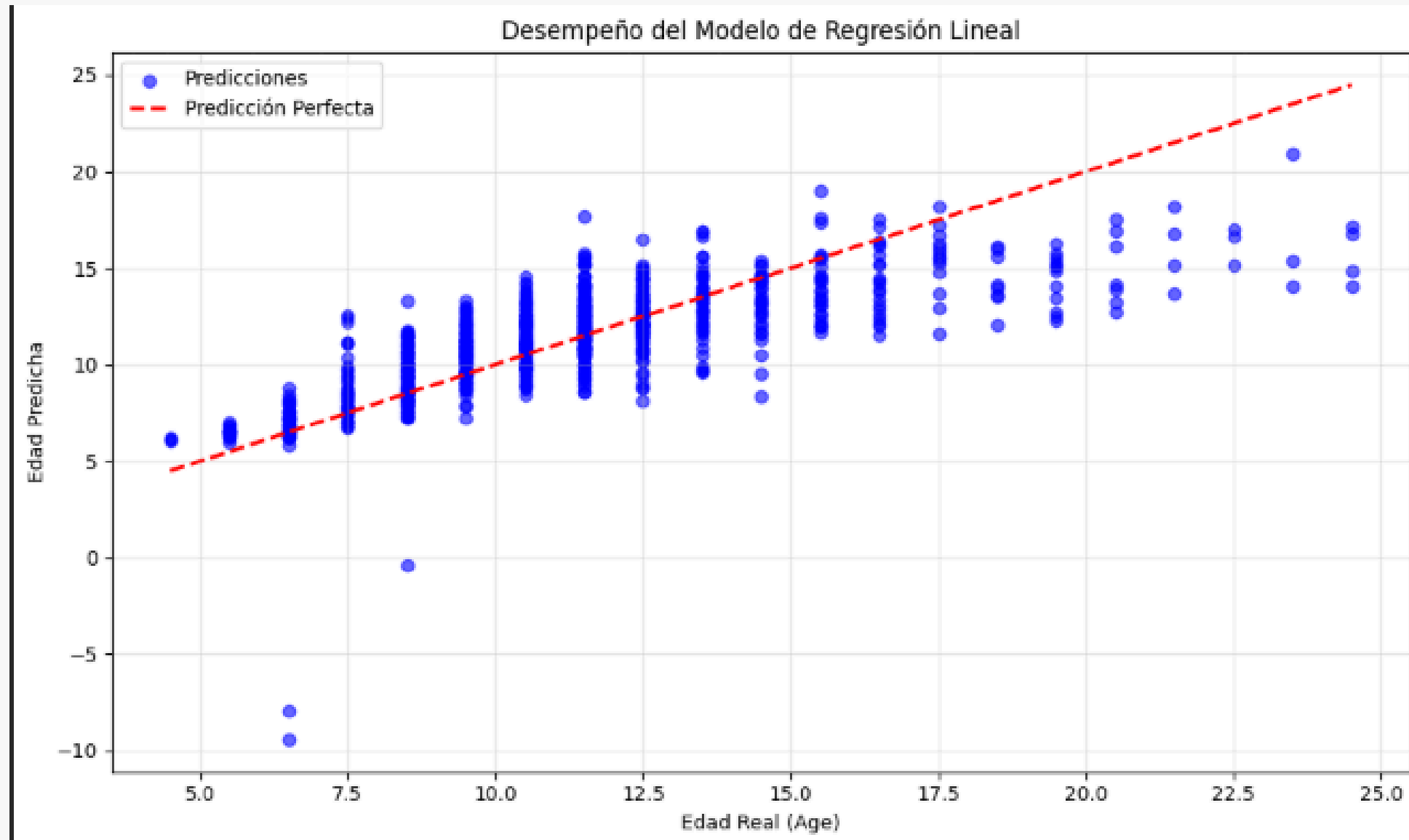


Los colores del mapa de calor permiten identificar rápidamente qué variables están fuertemente relacionadas y cuáles no.

- +1: correlación positiva perfecta
- -1: correlación negativa perfecta
- 0: ausencia de relación lineal

La edad presenta correlaciones moderadas con estas variables, lo que respalda su uso como objetivo del modelo predictivo.





El diagrama de Valores Reales vs. Valores Predichos compara:

- En el eje X → las edades reales de los abulones
- En el eje Y → las edades que el modelo de regresión lineal predice

Además, incluye una línea diagonal punteada ($y = x$) que representa una predicción perfecta.

Mientras más cerca estén los puntos de esta línea, mejor es el desempeño del modelo.

