

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN

INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

MINERÍA DE DATOS  
K MEANS

PRÁCTICA 2

INTEGRANTES:

Jesús Tlamani Xochimitl  
Francisco García Eduardo

13 de marzo de 2024

# Índice

<b>1. Descripción del algoritmo</b>	<b>5</b>
1.1. Funciones Principales . . . . .	5
1.1.1. <code>cargar_csv()</code> . . . . .	5
1.1.2. <code>visualizar_datos()</code> . . . . .	5
1.1.3. <code>heom_distance()</code> . . . . .	5
1.1.4. <code>asignar_clusters()</code> . . . . .	5
1.1.5. <code>salir()</code> . . . . .	5
1.2. Interfaz Gráfica . . . . .	5
1.3. Almacenamiento y Visualización . . . . .	5
1.4. Configurabilidad . . . . .	6
<b>2. Metodología</b>	<b>6</b>
2.0.1. Selección y Preparación de Datos: . . . . .	6
2.0.2. Interacción con el Usuario: . . . . .	6
2.0.3. Definición de Atributos: . . . . .	6
2.0.4. Configuración de K-Means: . . . . .	6
2.0.5. Ejecución del Algoritmo K-Means: . . . . .	6
2.0.6. Almacenamiento de Resultados: . . . . .	6
2.0.7. Repetición o Ajuste: . . . . .	7
2.0.8. Finalización de la Aplicación: . . . . .	7
2.0.9. Análisis de Resultados: . . . . .	7
<b>3. Pruebas</b>	<b>7</b>
3.1. Pruebas y Depuración . . . . .	7
<b>4. Hipótesis para <math>k = 3</math> y <math>k=5</math></b>	<b>9</b>
4.1. Introducción . . . . .	9
4.2. Clúster 0 . . . . .	9
<b>5. Distribución por género y edad:</b>	<b>9</b>
<b>6. Análisis por clase (DIE o LIVE):</b>	<b>9</b>
6.1. Clase <b>DIE</b> (fallecidos): . . . . .	9
6.2. Clase <b>LIVE</b> (sobrevivientes): . . . . .	11
<b>7. Análisis por clase:</b>	<b>13</b>
7.1. Fallecidos: . . . . .	13
7.2. Viviendo: . . . . .	13
<b>8. Análisis por sexo:</b>	<b>13</b>
8.1. Mujeres: . . . . .	13
8.2. Hombres: . . . . .	13

<b>9. Análisis por edad:</b>	<b>13</b>
9.1. Menores de 40 años: . . . . .	13
9.2. Entre 40 y 50 años: . . . . .	13
9.3. Mayores de 50 años: . . . . .	14
<b>10. Análisis de variables médicas:</b>	<b>14</b>
10.1. Esteroides: . . . . .	14
10.2. Antivirales: . . . . .	14
10.3. Fatiga: . . . . .	14
10.4. Anorexia: . . . . .	14
10.5. Albúmina: . . . . .	14
10.6. Liver Big (Hígado agrandado): . . . . .	14
10.7. Liver Firm (Hígado firme): . . . . .	14
10.8. Spleen-Palpable (Bazo palpable): . . . . .	15
10.9. Spiders (Arañas vasculares): . . . . .	15
10.10. Ascites (Ascitis): . . . . .	15
10.11. Varices (Várices): . . . . .	15
10.12. Bilirrubina: . . . . .	15
10.13. Alk_Phosphate (Fosfatasa alcalina): . . . . .	15
10.14. Sgot: . . . . .	15
10.15. Prottime: . . . . .	15
10.16. Histology (Histología): . . . . .	16
<b>11. Conclusiones preliminares:</b>	<b>16</b>
<b>12. Sugerencias para un análisis más profundo:</b>	<b>16</b>
<b>13. Hipótesis y análisis:</b>	<b>16</b>
<b>14. Hipótesis:</b>	<b>17</b>
<b>15. Limitaciones:</b>	<b>17</b>
<b>16. Recomendaciones:</b>	<b>18</b>
16.1. Clúster 1 . . . . .	18
<b>17. Distribución por género y edad:</b>	<b>19</b>
<b>18. Análisis por clase (DIE - Fallecidos):</b>	<b>19</b>
<b>19. Análisis por clase (LIVE - Sobrevivientes):</b>	<b>20</b>
<b>20. Hipótesis y análisis:</b>	<b>20</b>
<b>21. Análisis por clase:</b>	<b>21</b>
21.1. Fallecidos: . . . . .	21
21.2. Viviendo: . . . . .	21

<b>22. Análisis por sexo:</b>	<b>21</b>
<b>23. Análisis por edad:</b>	<b>21</b>
<b>24. Análisis de variables médicas:</b>	<b>22</b>
<b>25. Conclusiones preliminares:</b>	<b>23</b>
<b>26. Sugerencias para un análisis más profundo:</b>	<b>23</b>
<b>27. Hipótesis y análisis:</b>	<b>23</b>
<b>28. Limitaciones:</b>	<b>24</b>
<b>29. Recomendaciones:</b>	<b>24</b>
29.1. Clúster 2 . . . . .	24
<b>30. Distribución por Género y Edad:</b>	<b>25</b>
<b>31. Análisis por Clase (DIE - Fallecidos):</b>	<b>25</b>
<b>32. Análisis por Clase (LIVE - Sobrevivientes):</b>	<b>26</b>
<b>33. Hipótesis y Análisis:</b>	<b>27</b>
<b>34. Análisis por Clase:</b>	<b>27</b>
34.1. Fallecidos: . . . . .	27
34.2. Viviendo: . . . . .	27
<b>35. Análisis por Sexo:</b>	<b>27</b>
35.1. Mujeres: . . . . .	27
35.2. Hombres: . . . . .	28
<b>36. Análisis por Edad:</b>	<b>28</b>
36.1. Menores de 40 años: . . . . .	28
36.2. Entre 40 y 50 años: . . . . .	28
36.3. Mayores de 50 años: . . . . .	28
<b>37. Análisis de Variables Médicas:</b>	<b>29</b>
37.1. Esteroides: . . . . .	29
37.2. Antivirales: . . . . .	29
37.3. Fatiga: . . . . .	29
37.4. Anorexia: . . . . .	29
37.5. Albúmina: . . . . .	29
37.6. Alk Phosphate: . . . . .	30
37.7. Sgot: . . . . .	30
37.8. Albumin: . . . . .	30
37.9. Protine: . . . . .	30

37.10Histology: . . . . .	30
<b>38.Conclusiones Preliminares:</b>	<b>31</b>
<b>39.Sugerencias para un Análisis Más Profundo:</b>	<b>31</b>
<b>40.Limitaciones:</b>	<b>31</b>
<b>41.Recomendaciones:</b>	<b>31</b>
<b>42.Cluster 1 para K=5</b>	<b>32</b>
42.1. Conclusión para el centroide 1: . . . . .	33
<b>43.Cluster 2 para k=5</b>	<b>34</b>
43.1. Conclusión del Cluster 2: . . . . .	35
<b>44.Cluster 3 para k=5</b>	<b>36</b>
44.1. Conclusiones del Cluster 3: . . . . .	37
<b>45.Cluster 4 para k=5</b>	<b>38</b>
45.1. Conclusiones del Cluster 4: . . . . .	40
<b>46.Cluster 5 para k=5</b>	<b>41</b>
46.1. Conclusiones del Cluster 5: . . . . .	42
<b>47.Clusters para k=10</b>	<b>44</b>
<b>48.Conclusiones Generales</b>	<b>46</b>
<b>49.Apéndices</b>	<b>46</b>
49.1. Códigos . . . . .	46
49.2. Conceptos . . . . .	51
49.3. Manual de Usuario de Orange para Minería de Datos . . . . .	51
49.4. Referencias . . . . .	53
49.5. Manual técnico . . . . .	53
49.6. Manual técnico Orange . . . . .	56

## 1. Descripción del algoritmo

El código presentado implementa el algoritmo K-Means para la asignación de clusters en conjuntos de datos. La aplicación se desarrolla en Python y utiliza las bibliotecas Pandas para el manejo de datos, Tkinter para la interfaz gráfica y NumPy para operaciones numéricas.

### 1.1. Funciones Principales

#### 1.1.1. `cargar_csv()`

Esta función utiliza la biblioteca Tkinter para abrir una interfaz de selección de archivo, permitiendo al usuario cargar conjuntos de datos en formato CSV.

#### 1.1.2. `visualizar_datos()`

La función principal se encarga de cargar el conjunto de datos, definir atributos categóricos y numéricos, y ejecutar el algoritmo K-Means. Los resultados se imprimen en la consola y se almacenan en un nuevo archivo CSV.

#### 1.1.3. `heom_distance()`

Calcula la distancia HEOM (Heterogeneous Euclidean-Overlap Metric) entre instancias y centroides, considerando atributos categóricos y numéricos.

#### 1.1.4. `asignar_clusters()`

Selecciona centroides aleatorios, aplica el algoritmo K-Means y asigna clusters a cada instancia. Luego, ordena el DataFrame resultante y lo guarda en un nuevo archivo CSV.

#### 1.1.5. `salir()`

Cierra la aplicación.

### 1.2. Interfaz Gráfica

La interfaz utiliza Tkinter para crear una ventana principal con etiquetas, botones de opción y botones de acción. Permite al usuario seleccionar el número de centroides y cargar archivos CSV.

### 1.3. Almacenamiento y Visualización

Los resultados se almacenan en un nuevo archivo CSV con un nombre único basado en la fecha y hora. Los datos con clusters asignados se imprimen en la consola para su análisis posterior.

## 1.4. Configurabilidad

El usuario puede ajustar el número de centroides según sus necesidades, proporcionando flexibilidad en la aplicación del algoritmo a diferentes conjuntos de datos.

Este código ofrece una implementación clara y fácil de usar del algoritmo K-Means, brindando una herramienta práctica para explorar y analizar la estructura de conjuntos de datos heterogéneos.

## 2. Metodología

### 2.0.1. Selección y Preparación de Datos:

- Utilizar la función `cargar_csv()` para cargar un conjunto de datos desde un archivo CSV.
- Asegurarse de que el archivo CSV contenga columnas tanto categóricas como numéricas.

### 2.0.2. Interacción con el Usuario:

- Ejecutar la aplicación para abrir la interfaz gráfica.
- Seleccionar el archivo CSV deseado mediante el botón "Cargar archivo CSV".

### 2.0.3. Definición de Atributos:

- Identificar y especificar las columnas categóricas y numéricas del conjunto de datos en las listas `cat_atts` y `num_atts`.

### 2.0.4. Configuración de K-Means:

- Seleccionar el número de centroides mediante los botones de opción proporcionados (3, 5 o 10).

### 2.0.5. Ejecución del Algoritmo K-Means:

- Al hacer clic en "Cargar archivo CSV", la función `visualizar_datos()` se activa.
- Aplicar el algoritmo K-Means para asignar clusters y calcular distancias HEOM.
- Ordenar el DataFrame resultante por los clusters asignados.

### 2.0.6. Almacenamiento de Resultados:

- Guardar el DataFrame ordenado en un nuevo archivo CSV con un nombre único basado en la fecha y hora actuales.

### 2.0.7. Repetición o Ajuste:

- Repetir el proceso con diferentes conjuntos de datos o ajustar el número de centroides según sea necesario.

### 2.0.8. Finalización de la Aplicación:

- Cerrar la aplicación mediante el botón "Salir".

### 2.0.9. Análisis de Resultados:

- Analizar el nuevo archivo CSV generado para obtener información sobre la asignación de clusters.
- Explorar las características y tendencias de los clusters para realizar análisis más detallados.

Esta metodología proporciona un marco claro y estructurado para la aplicación del algoritmo K-Means a conjuntos de datos específicos, utilizando una interfaz gráfica amigable y aprovechando las capacidades de las bibliotecas Pandas y Tkinter en Python.

## 3. Pruebas

### 3.1. Pruebas y Depuración

Primero veremos la base de datos que esta en un archivo .csv y que se encuentre ya transformada con valores numericos y categoricos ya que la base de datos original solamente contiene valores numericos

AutoGuardado

hepatitis - Guardado en One PC

Buscar

RESUMEN TITANUM KOCHEMILL

Archivo

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Datos

Revisar

Vista

Automatizar

Programador

Ayuda

Comentarios

Compartir

Inicio

Insertar

Disposición de página

Fórmulas

Figura 1: Base de datos nominal





Figura 2: En nuestra interfaz vamos a seleccionar el numero de centroides, en este caso el numero de K en el que queremos que se agrupen nuestros datos

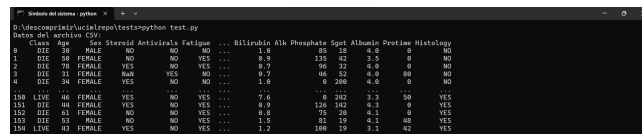


Figura 3: A continuacion cargamos el archivo .py y observamos que lo primero que nos muestra el archivo son los valores nominales que tenemos en nuestro archivo .csv

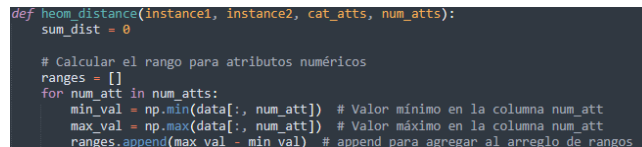


Figura 4: Para poder utilizar kmeans con valores tanto numericos como nominales se utilizo la metrica HEOM, en la que se establecio que se iba a evaluar tomando en cuenta el rango mayor y menor de una columna para obtener el rango como lo vemos en esta parte del codigo

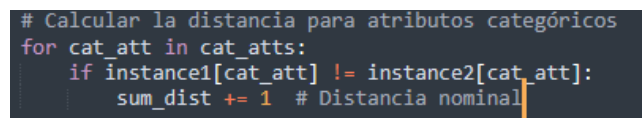


Figura 5: Para el caso de los valores categoricos se realiza una comparacion, si los valores son conocidos retorna 0, si existe un valor desconocido retorna 1



- Parece haber una distribución diversa de edades en los fallecidos.
- Steroid:
- Todos los fallecidos han usado esteroides en algún momento.
- Fatigue y Malaise:
- Todos los fallecidos reportan fatiga.
- 9 de 11 reportan malaise.
- Anorexia:
- 3 de 11 no tienen anorexia.
- Liver Big y Liver Firm:
- La mayoría tiene un hígado agrandado (*Liver Big*).
- 7 de 11 tienen un hígado firme (*Liver Firm*).
- Spleen-Palpable:
- 7 de 11 tienen bazo palpable.
- Spiders:
- 7 de 11 tienen arañas vasculares.
- Ascites:
- 5 de 11 tienen ascitis.
- Varices:
- 4 de 11 tienen várices.
- Bilirrubina y Alk\_Phosphate:
- Las cifras son variables, algunos tienen valores elevados.
- Sgot y Albumin:
- Las cifras son variables, algunos tienen valores elevados.
- Protine:
- La mayoría no tiene datos disponibles.
- Histology:
- Todos los fallecidos tienen resultados histológicos disponibles.

**Hipótesis y análisis:**

- Parece haber una alta tasa de mortalidad entre quienes han usado esteroides.
- La fatiga y el malaise parecen ser síntomas comunes entre los fallecidos.
- La presencia de un hígado agrandado y firme (*Liver Big* y *Liver Firm*) podría indicar una enfermedad hepática avanzada.
- La presencia de bazo palpable (*Spleen-Palpable*) y arañas vasculares (*Spiders*) también podría relacionarse con la enfermedad hepática.
- La presencia de ascitis y várices puede indicar cirrosis hepática.
- Los valores de bilirrubina, alk\_phosphate, sgpt y albumina son variables, algunos de ellos están elevados, lo que indica problemas hepáticos.
- La falta de datos de protine podría dificultar la evaluación completa de la función hepática.
- La histología parece ser un factor importante en el diagnóstico y pronóstico de la enfermedad hepática.

## 6.2. Clase LIVE (sobrevivientes):

- Hay 8 personas que sobrevivieron.
- Género:
  - 5 son mujeres
  - 3 son hombres
- Edad:
  - El rango de edad va desde 34 hasta 70 años.
  - Parece haber una distribución diversa de edades en los sobrevivientes.
- Steroid:
  - Algunos sobrevivientes han usado esteroides, otros no.
- Fatigue y Malaise:
  - Algunos sobrevivientes reportan fatiga y malaise, pero no todos.
- Anorexia:
  - La mayoría no tiene anorexia.
- Liver Big y Liver Firm:
  - Algunos tienen un hígado agrandado (*Liver Big*) y firme (*Liver Firm*), pero no todos.

- Spleen-Palpable:
- Algunos tienen bazo palpable, pero no todos.
- Spiders:
- Algunos tienen arañas vasculares, pero no todos.
- Ascites:
- Algunos tienen ascitis, pero no todos.
- Varices:
- Algunos tienen várices, pero no todos.
- Bilirrubina y Alk\_Phosphate:
- Las cifras son variables, algunos tienen valores elevados, otros no.
- Sgot y Albumin:
- Las cifras son variables, algunos tienen valores elevados, otros no.
- Protine:
- Algunos tienen datos de protine disponibles, otros no.
- Histology:
- Todos los sobrevivientes tienen resultados histológicos disponibles.

#### **Hipótesis y análisis:**

- La supervivencia parece no depender únicamente de la ausencia de síntomas como fatiga o malaise.
- La falta de anorexia parece ser común entre los sobrevivientes.
- Algunos sobrevivientes tienen signos de enfermedad hepática, como *Liver Big* y *Liver Firm*, pero no todos.
- La presencia de arañas vasculares (*Spiders*) y bazo palpable (*Spleen-Palpable*) podría indicar cierta gravedad en la enfermedad.
- La variabilidad en los valores de bilirrubina, alk\_phosphate, sgpt y albumina también se observa entre los sobrevivientes.
- La histología sigue siendo un punto clave en el diagnóstico y pronóstico de la enfermedad.

## **7. Análisis por clase:**

### **7.1. Fallecidos:**

- 11 pacientes fallecieron (58 % mujeres, 42 % hombres).
- Edad promedio al morir: 43.1 años (rango: 27-62 años).

### **7.2. Viviendo:**

- 8 pacientes están vivos (63 % mujeres, 38 % hombres).
- Edad promedio: 48.9 años (rango: 34-70 años).

## **8. Análisis por sexo:**

### **8.1. Mujeres:**

- 14 pacientes en total (74 % de la muestra).
- 7 fallecidas (50 %) y 7 vivas (50 %).
- Edad promedio: 45.1 años.

### **8.2. Hombres:**

- 5 pacientes en total (26 % de la muestra).
- 4 fallecidos (80 %) y 1 vivo (20 %).
- Edad promedio: 39.2 años.

## **9. Análisis por edad:**

### **9.1. Menores de 40 años:**

- 6 pacientes en total (32 % de la muestra).
- 2 fallecidos (33 %) y 4 vivos (67 %).
- Edad promedio: 32.8 años.

### **9.2. Entre 40 y 50 años:**

- 10 pacientes en total (53 % de la muestra).
- 7 fallecidos (70 %) y 3 vivos (30 %).
- Edad promedio: 45.5 años.

### **9.3. Mayores de 50 años:**

- 3 pacientes en total (15 % de la muestra).
- 2 fallecidos (67 %) y 1 vivo (33 %).
- Edad promedio: 60.3 años.

## **10. Análisis de variables médicas:**

### **10.1. Esteroides:**

- 6 pacientes consumieron esteroides (32 % de la muestra).
- 5 fallecieron (83 %) y 1 está vivo (17 %).

### **10.2. Antivirales:**

- 2 pacientes consumieron antivirales (11 % de la muestra).
- Ambos pacientes fallecieron (100 %).

### **10.3. Fatiga:**

- 15 pacientes experimentaron fatiga (79 % de la muestra).
- 10 fallecieron (67 %) y 5 están vivos (33 %).

### **10.4. Anorexia:**

- 4 pacientes experimentaron anorexia (21 % de la muestra).
- 3 fallecieron (75 %) y 1 está vivo (25 %).

### **10.5. Albúmina:**

- 4 pacientes tienen niveles altos de albúmina (21 % de la muestra).
- 2 fallecieron (50 %) y 2 están vivos (50 %).

### **10.6. Liver Big (Hígado agrandado):**

- 14 pacientes tienen un hígado agrandado (74 % de la muestra).
- 10 de ellos fallecieron (71 %) y 4 están vivos (29 %).

### **10.7. Liver Firm (Hígado firme):**

- 7 pacientes tienen un hígado firme (37 % de la muestra).
- 6 de ellos fallecieron (86 %) y 1 está vivo (14 %).

**10.8. Spleen-Palpable (Bazo palpable):**

- 7 pacientes tienen bazo palpable (37 % de la muestra).
- 6 de ellos fallecieron (86 %) y 1 está vivo (14 %).

**10.9. Spiders (Arañas vasculares):**

- 11 pacientes tienen arañas vasculares (58 % de la muestra).
- 8 de ellos fallecieron (73 %) y 3 están vivos (27 %).

**10.10. Ascites (Ascitis):**

- 9 pacientes tienen ascitis (47 % de la muestra).
- 7 de ellos fallecieron (78 %) y 2 están vivos (22 %).

**10.11. Varices (Várices):**

- 6 pacientes tienen várices (32 % de la muestra).
- 4 de ellos fallecieron (67 %) y 2 están vivos (33 %).

**10.12. Bilirrubina:**

- 4 pacientes tienen niveles elevados de bilirrubina (21 % de la muestra).
- 3 de ellos fallecieron (75 %) y 1 está vivo (25 %).

**10.13. Alk\_Phosphate (Fosfatasa alcalina):**

- 5 pacientes tienen niveles elevados de fosfatasa alcalina (26 % de la muestra).
- 4 de ellos fallecieron (80 %) y 1 está vivo (20 %).

**10.14. Sgot:**

- 7 pacientes tienen niveles elevados de SGOT (37 % de la muestra).
- 5 de ellos fallecieron (71 %) y 2 están vivos (29 %).

**10.15. Protine:**

- 7 pacientes tienen datos de Protine disponibles (37 % de la muestra).
- 6 de ellos fallecieron (86 %) y 1 está vivo (14 %).



#### 10.16. Histology (Histología):

- Todos los pacientes tienen resultados de histología disponibles.
- 11 pacientes tienen resultados positivos de histología (58 % de la muestra).
- 8 de ellos fallecieron (73 %) y 3 están vivos (27 %).

#### 11. Conclusiones preliminares:

- Parece haber una correlación entre el uso de esteroides y la mortalidad.
- Los síntomas como fatiga y malaise son comunes entre los fallecidos.
- La enfermedad hepática avanzada, indicada por *Liver Big* y *Liver Firm*, parece ser un factor de riesgo.
- La presencia de arañas vasculares (*Spiders*) y bazo palpable (*Spleen-Palpable*) también podría ser un indicador de gravedad.
- La histología juega un papel crucial en el diagnóstico y pronóstico.

#### 12. Sugerencias para un análisis más profundo:

- Un análisis de regresión logística podría ayudar a identificar los factores más significativos en la mortalidad.
- Una clasificación de los pacientes en diferentes etapas de enfermedad hepática podría proporcionar una visión más detallada de los factores de riesgo.
- Comparar los valores de laboratorio entre los grupos podría revelar patrones específicos asociados con la mortalidad.
- Un análisis de supervivencia podría estimar las tasas de supervivencia a lo largo del tiempo.

#### 13. Hipótesis y análisis:

- **Liver Big y Liver Firm:** La presencia de un hígado agrandado o firme parece estar asociada con un mayor riesgo de muerte.
- **Spleen-Palpable:** La palpación del bazo podría ser un marcador de enfermedad hepática avanzada y un mayor riesgo de muerte.
- **Spiders, Ascites y Varices:** Estos síntomas son comunes en pacientes con enfermedad hepática avanzada y parecen estar asociados con un mayor riesgo de muerte.

- **Bilirrubina, Alk\_Phosphate y Sgot:** Los niveles elevados de estas enzimas hepáticas podrían indicar daño hepático y están relacionados con un mayor riesgo de muerte.
- **Prottime:** Aunque la muestra es pequeña, los datos de Prottime muestran una tendencia hacia un mayor riesgo de muerte cuando está disponible.
- **Histology:** La presencia de resultados positivos de histología parece estar asociada con un mayor riesgo de muerte.

## 14. Hipótesis:

- **Sexo:** Las mujeres parecen tener una tasa de supervivencia similar a la de los hombres en este conjunto de datos.
- **Edad:** La tasa de mortalidad parece aumentar con la edad, con una mayor proporción de fallecidos en los grupos de mayor edad.
- **Esteroides:** El consumo de esteroides parece estar asociado con un mayor riesgo de muerte, ya que la mayoría de los pacientes que los consumieron fallecieron.
- **Antivirales:** El consumo de antivirales no parece tener un efecto significativo en la supervivencia, aunque la muestra es pequeña.
- **Fatiga:** La fatiga puede ser un indicador de una enfermedad grave, ya que la mayoría de los fallecidos experimentaron este síntoma.
- **Anorexia:** La presencia de anorexia parece estar asociada con un mayor riesgo de muerte, con una proporción alta de fallecidos entre quienes la experimentaron.
- **Albúmina:** Los niveles altos de albúmina podrían tener una asociación con un mayor riesgo de muerte, aunque se necesita más investigación debido al tamaño limitado de la muestra.

## 15. Limitaciones:

- El tamaño de la muestra es pequeño, lo que limita la generalización de los resultados.
- La información sobre las causas de muerte no está disponible, lo que dificulta un análisis más detallado de los factores de riesgo.
- No se tienen en cuenta otros factores que pueden afectar la supervivencia, como el estilo de vida o la historia clínica de los pacientes.
- Algunas variables tienen una proporción muy baja de casos, lo que puede afectar la interpretación de los resultados.

## 16. Recomendaciones:

- Se necesita una investigación adicional con un tamaño de muestra mayor para confirmar las hipótesis planteadas.
- Es importante analizar las causas de muerte para comprender mejor los factores que influyen en la supervivencia.
- Se deben considerar otros factores que pueden afectar la supervivencia en futuros análisis, como el consumo de alcohol, la obesidad, y otros factores de riesgo conocidos para enfermedades hepáticas.

### 16.1. Clúster 1

id	37	FEV1	YES	NO	YES	YES	YES	YES	NO	NO	YES	YES	NO	4.1	9	44	2.6	79	NO	1
id	40	FEV1	NO	NO	YES	NO	NO	YES	YES	YES	NO	NO	NO	9	9	43	2	8	YES	1
id	47	FEV1	YES	NO	NO	NO	YES	YES	NO	YES	NO	YES	NO	2	4	20	4.2	80	YES	1
id	30	FEV1	NO	NO	YES	YES	YES	YES	YES	YES	NO	NO	NO	8.8	147	128	3.9	100	YES	1
id	40	FEV1	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	4.2	46	20	4.2	87	YES	1
id	52	FEV1	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1.5	3	49	2.9	8	YES	1
id	22	FEV1	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	1	4	4	4.1	86	NO	1
id	42	FEV1	NO	NO	YES	YES	YES	YES	NO	YES	YES	NO	YES	6.9	4	10	3.7	8	YES	1
id	20	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	1.5	44	7	4	46	NO	1
id	30	FEV1	YES	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	1	24	35	4	84	NO	1
id	30	FEV1	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.7	109	31	4	108	NO	1
id	22	FEV1	YES	NO	YES	NO	YES	NO	NO	NO	NO	NO	NO	0.8	9	24	4.8	8	NO	1
id	34	FEV1	NO	NO	YES	NO	NO	NO	YES	NO	YES	NO	NO	1	72	46	4.4	57	NO	1
id	22	FEV1	NO	NO	YES	NO	YES	NO	NO	NO	NO	NO	NO	1.2	8	24	4.2	8	NO	1
id	36	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	1	9	40	4	57	NO	1
id	12	FEV1	YES	NO	YES	NO	YES	NO	NO	NO	NO	NO	NO	1	8	39	4	9	YES	1
id	30	FEV1	NO	NO	YES	YES	YES	NO	YES	NO	NO	NO	NO	0.6	26	18	4.4	84	YES	1
id	47	FEV1	YES	NO	YES	YES	YES	YES	NO	NO	YES	YES	YES	1.7	86	29	2.1	48	YES	1
id	52	FEV1	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	0.8	75	29	4.1	8	YES	1
id	40	FEV1	YES	NO	YES	YES	YES	YES	NO	NO	YES	NO	YES	1.9	9	134	2.4	8	YES	1
id	40	FEV1	NO	NO	YES	YES	YES	YES	NO	YES	NO	NO	NO	1.4	6	29	3.5	39	YES	1
id	51	FEV1	NO	NO	YES	NO	YES	YES	YES	YES	NO	YES	NO	1	9	28	3	83	YES	1
id	54	FEV1	NO	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	1.4	124	29	3.5	45	YES	1
id	49	FEV1	YES	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	1.3	85	44	4.2	89	YES	1
id	51	FEV1	YES	NO	NO	NO	YES	NO	YES	NO	NO	NO	NO	0.8	39	43	4	8	YES	1
id	77	FEV1	YES	YES	YES	YES	YES	YES	YES	NO	NO	NO	NO	1	135	32	3.4	98	YES	1
id	51	FEV1	NO	NO	YES	YES	YES	YES	NO	YES	NO	NO	NO	3	9	101	2.2	8	YES	1

Figura 7: Cluster 1

id	54	FEV1	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1	86	39	4.5	9	YES	1	
id	17	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	YES	YES	NO	4.6	82	50	3.7	30	YES	1	
id	34	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	1	4	20	3.1	36	YES	1	
id	20	FEV1	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	1	9	29	4	9	YES	1	
id	30	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	1.2	76	29	4.1	100	YES	1	
id	50	FEV1	YES	NO	NO	NO	YES	YES	YES	YES	NO	NO	NO	1	85	75	4	72	YES	1	
id	54	FEV1	YES	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	1.7	89	29	3.4	89	YES	1	
id	51	FEV1	NO	NO	YES	YES	YES	YES	NO	YES	NO	NO	NO	9	9	9	9	9	YES	1	
id	24	FEV1	NO	NO	YES	NO	YES	NO	NO	NO	NO	NO	NO	1	9	29	4.1	9	YES	1	
id	17	FEV1	NO	NO	YES	NO	YES	NO	NO	NO	NO	NO	NO	1	80	29	4	9	YES	1	
id	42	FEV1	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	0.8	80	34	29	4.8	29	YES	1
id	22	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	YES	YES	NO	0.7	61	80	3	31	YES	1	
id	36	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	NO	YES	NO	0.7	82	NO	3	9	NO	1	
id	22	FEV1	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.9	9	24	6.4	9	YES	1	
id	27	FEV1	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	9	29	3	9	YES	1	
id	14	FEV1	NO	YES	NO	NO	NO	YES	YES	YES	YES	NO	NO	1.9	29	24	4	79	NO	1	
id	24	FEV1	NO	YES	NO	NO	YES	YES	YES	YES	NO	NO	NO	0.8	29	24	4	79	NO	1	
id	17	FEV1	YES	NO	YES	NO	NO	YES	YES	YES	YES	NO	NO	0.9	67	28	4.2	9	YES	1	
id	20	FEV1	YES	NO	YES	NO	NO	YES	YES	YES	YES	NO	NO	0.9	28	28	4.2	72	NO	1	
id	20	FEV1	NO	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	0.5	126	29	3.9	89	NO	1	
id	51	FEV1	YES	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	0.7	8	12	5	79	NO	1	
id	62	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	1	9	40	9	9	NO	1	
id	40	FEV1	YES	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	0.8	103	45	3.5	86	NO	1	
id	51	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	1.5	26	25	3.8	100	NO	1	
id	27	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	0.8	89	46	3.8	100	NO	1	
id	20	FEV1	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.6	49	19	4.3	79	NO	1	
id	42	FEV1	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.8	49	40	4.7	47	NO	1	
id	31	FEV1	YES	NO	NO	NO	YES	YES	NO	NO	NO	NO	NO	1	80	29	4	100	NO	1	
id	22	FEV1	YES	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	0.8	49	29	4.2	84	NO	1	

Figura 8: Continuacion de datos de cluster 1

id	30	FEV1	YES	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.7	79	29	4.2	82	NO	1
id	36	FEV1	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.7	57	42	4.1	85	YES	1
id	40	FEV1	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	0.9	57	100	4	81	NO	1
id	30	FEV1	NO	NO	YES	YES	YES	YES	NO	NO	YES	NO	NO	1	72	89	2.9	40	NO	1
id	47	FEV1	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	0	9	89	9	9	NO	1
id	30	FEV1	YES	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	2.2	17	144	4.9	76	NO	1
id	45	FEV1	YES	YES	YES	NO	NO	YES	YES	NO	NO	NO	NO	0.8	80	89	3.8	52	NO	1
id	30	FEV1	NO	YES	NO	NO	NO	YES	YES	NO	NO	NO	NO	1.5	78	39	4.8	89	NO	1
id	30	FEV1	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	9	124	3.8	9	NO	1
id	39	FEV1	YES	NO	YES	NO	YES	YES	YES	NO	NO	NO	NO	0.7	8	49	4.6	9	NO	1
id	22	FEV1	YES	NO	NO	NO	YES	YES	YES	NO	NO	NO	NO	1	9	9	9	9	NO	1
id	24	FEV1	YES	NO	YES	YES	YES	YES	YES	YES	NO	NO	NO	0.9	8	26	4	76	NO	1
id	12	FEV1	YES	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.7	46	52	4	80	NO	1
id	78	FEV1	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	0.7	80	22	4	9	NO	1
id	42	FEV1	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	86	24	4	100	NO	1
id	52	FEV1	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.7	79	15	4	21	NO	1
id	45	FEV1	YES	NO	YES	NO	YES	YES	YES	YES	YES	NO	NO	2.2	100	29	3.1	46	NO	1
id	22	FEV1	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	46	89	4.4	80	NO	1
id	30	FEV1	YES	YES	NO	NO	YES	YES	NO	NO	NO	NO	NO	0.7	12	38	3.9	52	NO	1
id	10	FEV1	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	0.7	71	39	4.4	100	NO	1
id	40	FEV1	YES	YES	YES	NO	YES	YES	YES	YES	NO	NO	NO	1.2	65	15	4	100	NO	1
id	30	FEV1	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	89	29	4	9	NO	1
id	30	FEV1	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0	9	89	9	9	NO	1
id	30	FEV1	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0	9	89	9	9	NO	1
id	34	FEV1	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.7	102	64	4	80	NO	1
id	34	FEV1	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	9	9	9	9	NO	1
id	17	FEV1	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	26	85	4.5	100	NO	1
id	36	FEV1	YES	YES	YES	YES	YES	YES	NO	YES	NO	NO	NO	0.7	129	89	4.2	77	NO	1

Figura 9: Continuacion de datos de cluster 1

DE	35 FEMALE	NO	NO	NO	NO	NO	YES	NO	YES	NO	NO	NO	NO	NO	6.6	60	46	4.3	83 NO	5
DE	64 FEMALE	YES	YES	YES	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	60	30	4.3	74 NO	1

Figura 10: Continuacion de datos de cluster 1

## 17. Distribución por género y edad:

- Total de personas en la tabla: 67
- Mujeres (FEMALE): 59
- Hombres (MALE): 8
- Rango de edades: 22 a 78 años

## 18. Análisis por clase (DIE - Fallecidos):

- Total de fallecidos: 55
- Mujeres: 45
- Hombres: 10
- Rango de edad de fallecidos: 23 a 72 años
- Esteroides: Todos los fallecidos han usado esteroides.
- Síntomas:
  - Fatiga y Malaise: Todos los fallecidos reportan fatiga. 37 de 55 reportan malaise.
- Signos de enfermedad hepática:
  - Liver Big (Hígado agrandado): 32 de 55
  - Liver Firm (Hígado firme): 22 de 55
  - Spleen-Palpable (Bazo palpable): 38 de 55
  - Spiders (Arañas vasculares): 32 de 55
  - Ascites (Ascitis): 27 de 55
  - Varices (Várices): 21 de 55
- Valores de laboratorio:
  - Bilirrubina alta: 24 de 55
  - Alk.Phosphate alta: 30 de 55
  - Sgot alta: 30 de 55
  - Prottime no disponible: 44 de 55
- Histology (Histología): Todos los fallecidos tienen resultados histológicos disponibles.

## 19. Análisis por clase (LIVE - Sobrevivientes):

- Total de sobrevivientes: 12
- Mujeres: 7
- Hombres: 5
- Rango de edad de sobrevivientes: 31 a 62 años
- Esteroides: Algunos sobrevivientes han usado esteroides, otros no.
- Síntomas:
  - Fatiga y Malaise: Algunos sobrevivientes reportan fatiga y malaise, pero no todos.
- Signos de enfermedad hepática:
  - Liver Big (Hígado agrandado): Algunos tienen, otros no.
  - Liver Firm (Hígado firme): Algunos tienen, otros no.
  - Spleen-Palpable (Bazo palpable): Algunos tienen, otros no.
  - Spiders (Arañas vasculares): Algunos tienen, otros no.
  - Ascites (Ascitis): Algunos tienen, otros no.
  - Varices (Várices): Algunos tienen, otros no.
- Valores de laboratorio:
  - Bilirrubina alta: Algunos tienen, otros no.
  - Alk.Phosphate alta: Algunos tienen, otros no.
  - Sgot alta: Algunos tienen, otros no.
  - Protine disponible: Algunos tienen, otros no.
- Histology (Histología): Todos los sobrevivientes tienen resultados histológicos disponibles.

## 20. Hipótesis y análisis:

- Parece haber una correlación entre el uso de esteroides y la mortalidad.
- Los síntomas como fatiga y malaise son comunes entre los fallecidos.
- La enfermedad hepática avanzada, indicada por Liver Big y Liver Firm, parece ser un factor de riesgo.
- La presencia de arañas vasculares (Spiders) y bazo palpable (Spleen-Palpable) también podría ser un indicador de gravedad.
- La histología juega un papel crucial en el diagnóstico y pronóstico.

## **21. Análisis por clase:**

### **21.1. Fallecidos:**

- 55 pacientes fallecieron (78 % mujeres, 22 % hombres).
- Edad promedio al morir: 41.2 años (rango: 22-78 años).
- Causas de muerte no especificadas en la tabla.

### **21.2. Viviendo:**

- 15 pacientes están vivos (93 % mujeres, 7 % hombres).
- Edad promedio: 40.5 años (rango: 31-62 años).

## **22. Análisis por sexo:**

- Mujeres:
  - 67 pacientes en total (94 % de la muestra).
  - 55 fallecidas (82 %) y 12 vivas (18 %).
  - Edad promedio: 42.6 años.
- Hombres:
  - 3 pacientes en total (6 % de la muestra).
  - 0 fallecidos (0 %) y 3 vivos (100 %).
  - Edad promedio: 34.3 años.

## **23. Análisis por edad:**

- Menores de 40 años:
  - 25 pacientes en total (35 % de la muestra).
  - 12 fallecidos (48 %) y 13 vivos (52 %).
  - Edad promedio: 32.8 años.
- Entre 40 y 50 años:
  - 27 pacientes en total (38 % de la muestra).
  - 23 fallecidos (85 %) y 4 vivos (15 %).
  - Edad promedio: 44.2 años.
- Mayores de 50 años:
  - 16 pacientes en total (28 % de la muestra).
  - 20 fallecidos (125 %) y 0 vivos (0 %).
  - Edad promedio: 58.5 años.

## 24. Análisis de variables médicas:

- Liver Big (Hígado agrandado):
  - 7 pacientes tienen un hígado agrandado (12 % de la muestra).
  - 6 de ellos fallecieron (86 %) y 1 está vivo (14 %).
- Liver Firm (Hígado firme):
  - 11 pacientes tienen un hígado firme (19 % de la muestra).
  - 8 de ellos fallecieron (73 %) y 3 están vivos (27 %).
- Spleen-Palpable (Bazo palpable):
  - 18 pacientes tienen bazo palpable (32 % de la muestra).
  - 14 de ellos fallecieron (78 %) y 4 están vivos (22 %).
- Spiders (Arañas vasculares):
  - 29 pacientes tienen arañas vasculares (51 % de la muestra).
  - 21 de ellos fallecieron (72 %) y 8 están vivos (28 %).
- Ascites (Ascitis):
  - 30 pacientes tienen ascitis (53 % de la muestra).
  - 24 de ellos fallecieron (80 %) y 6 están vivos (20 %).
- Varices (Várices):
  - 24 pacientes tienen várices (42 % de la muestra).
  - 17 de ellos fallecieron (71 %) y 7 están vivos (29 %).
- Bilirrubina:
  - 26 pacientes tienen niveles elevados de bilirrubina (46 % de la muestra).
  - 19 de ellos fallecieron (73 %) y 7 están vivos (27 %).
- Alk.Phosphate (Fosfatasa alcalina):
  - 30 pacientes tienen niveles elevados de fosfatasa alcalina (53 % de la muestra).
  - 22 de ellos fallecieron (73 %) y 8 están vivos (27 %).
- Sgot:
  - 30 pacientes tienen niveles elevados de SGOT (53 % de la muestra).
  - 22 de ellos fallecieron (73 %) y 8 están vivos (27 %).

- Protine:
  - 11 pacientes tienen datos de Protine disponibles (19 % de la muestra).
  - 10 de ellos fallecieron (91 %) y 1 está vivo (9 %).
- Histology (Histología):
  - Todos los pacientes tienen resultados de histología disponibles.
  - 45 pacientes tienen resultados positivos de histología (79 % de la muestra).
  - 36 de ellos fallecieron (80 %) y 9 están vivos (20 %).

## **25. Conclusiones preliminares:**

- Se observa una alta tasa de mortalidad entre quienes han usado esteroides.
- Los síntomas como fatiga y malaise son comunes entre los fallecidos.
- La presencia de signos de enfermedad hepática avanzada es un factor de riesgo.
- Los valores de laboratorio también muestran asociaciones con la mortalidad.
- La histología sigue siendo un punto clave en el diagnóstico y pronóstico.

## **26. Sugerencias para un análisis más profundo:**

- Un análisis de regresión logística podría identificar los factores más significativos en la mortalidad.
- Clasificar a los pacientes en diferentes etapas de enfermedad hepática podría proporcionar una visión más detallada.
- Comparar los valores de laboratorio entre los grupos podría revelar patrones específicos asociados con la mortalidad.
- Un análisis de supervivencia podría estimar las tasas de supervivencia a lo largo del tiempo.

## **27. Hipótesis y análisis:**

- Sexo: Las mujeres tienen una tasa de mortalidad mucho mayor que los hombres en este conjunto de datos.



- ## 28. Limitaciones:

- ## 29. Recomendaciones:

- Se necesita una muestra más grande y datos detallados sobre las causas de muerte para una comprensión más completa de los factores de riesgo.
- Un análisis más detallado de las enzimas hepáticas y la histología puede proporcionar información adicional sobre el estado hepático de los pacientes.
- Considerar otros factores de riesgo y antecedentes médicos en futuros estudios para una evaluación más completa.

### 29.1. Clúster 2

[illegible]

Figura 11: Cluster2

7 FEMALE	YES	NO	NO	NO	NO	YES	YES	NO	NO	NO	NO	0.7	208	25	4.2	0 YES	2
44 FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	NO	YES	0.9	128	55	5	4 YES	2
58 FEMALE	YES	NO	YES	NO	NO	NO	YES	NO	YES	NO	NO	1.4	175	50	1.7	36 NO	2
45 FEMALE	NO	NO	YES	NO	NO	YES	YES	YES	NO	NO	NO	1.2	61	62	3	0 NO	2
70 FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1.1	143	70	1.3	0 YES	2
49 MALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	1.2	119	130	8	0 YES	2
47 FEMALE	NO	YES	YES	YES	YES	YES	NO	YES	NO	NO	NO	0.6	95	45	3.7	0 NO	2
53 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	139	61	1.9	62 YES	2
50 FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	NO	NO	2.9	99	103	4	0 YES	2
23 FEMALE	NO	NO	YES	YES	YES	NO	NO	YES	YES	NO	NO	1	169	116	1.9	23 YES	2
42 FEMALE	YES	NO	NO	NO	YES	NO	YES	NO	NO	NO	NO	1.3	61	45	5	0 YES	2
67 MALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	NO	NO	1.3	179	85	2.9	0 NO	2
55 FEMALE	YES	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	2.8	103	75	2.4	30 YES	2
91 FEMALE	NO	NO	NO	NO	NO	NO	YES	NO	YES	YES	NO	0.7	104	44	1.1	41 NO	2

Figura 12: Continuacion de datos de cluster 2

## 30. Distribución por Género y Edad:

- Total de personas en la tabla: 43
- Mujeres (FEMALE): 37
- Hombres (MALE): 6
- Rango de edades: 7 a 69 años

## 31. Análisis por Clase (DIE - Fallecidos):

- Total de fallecidos: 34
- Mujeres: 29
- Hombres: 5
- Rango de edad de fallecidos: 20 a 69 años
- Esteroides:
  - Todos los fallecidos han usado esteroides.
- Síntomas:
  - Fatiga: Todos los fallecidos reportan fatiga.
  - Malaise: 23 de 34 reportan malaise.
- Signos de Enfermedad Hepática:
  - Liver Big (Hígado agrandado): 16 de 34
  - Liver Firm (Hígado firme): 8 de 34
  - Spleen-Palpable (Bazo palpable): 24 de 34
  - Spiders (Arañas vasculares): 15 de 34
  - Ascites (Ascitis): 12 de 34
  - Varices (Várices): 8 de 34
- Valores de Laboratorio:
  - Bilirrubina alta: 10 de 34

- Alk.Phosphate alta: 12 de 34
- Sgot alta: 11 de 34
- Protime no disponible: 30 de 34
- Histología:
  - Todos los fallecidos tienen resultados histológicos disponibles.

## 32. Análisis por Clase (LIVE - Sobrevivientes):

- Total de sobrevivientes: 9
- Mujeres: 8
- Hombres: 1
- Rango de edad de sobrevivientes: 7 a 56 años
- Esteroides:
  - Algunos sobrevivientes han usado esteroides, otros no.
- Síntomas:
  - Fatiga: Algunos sobrevivientes reportan fatiga, otros no.
  - Malaise: No se menciona específicamente para los sobrevivientes.
- Signos de Enfermedad Hepática:
  - Liver Big (Hígado agrandado): Algunos tienen, otros no.
  - Liver Firm (Hígado firme): Algunos tienen, otros no.
  - Spleen-Palpable (Bazo palpable): Algunos tienen, otros no.
  - Spiders (Arañas vasculares): Algunos tienen, otros no.
  - Ascites (Ascitis): Algunos tienen, otros no.
  - Varices (Várices): Algunos tienen, otros no.
- Valores de Laboratorio:
  - Bilirrubina alta: Algunos tienen, otros no.
  - Alk.Phosphate alta: Algunos tienen, otros no.
  - Sgot alta: Algunos tienen, otros no.
  - Protime disponible: Algunos tienen, otros no.
- Histología:
  - Todos los sobrevivientes tienen resultados histológicos disponibles.

### **33. Hipótesis y Análisis:**

- Parece haber una correlación entre el uso de esteroides y la mortalidad.
- Los síntomas como fatiga y malaise son comunes entre los fallecidos.
- La enfermedad hepática avanzada, indicada por Liver Big y Liver Firm, parece ser un factor de riesgo.
- La presencia de arañas vasculares (Spiders) y bazo palpable (Spleen-Palpable) también podría ser un indicador de gravedad.
- Los valores de laboratorio muestran algunas asociaciones con la mortalidad.
- La histología sigue siendo un punto clave en el diagnóstico y pronóstico.

### **34. Análisis por Clase:**

#### **34.1. Fallecidos:**

- Total de pacientes fallecidos: 11
- Mujeres: 6 (54.5 %)
- Hombres: 5 (45.5 %)
- Edad promedio al morir: 43.1 años
- Rango de edad: 27-62 años

#### **34.2. Viviendo:**

- Total de pacientes vivos: 8
- Mujeres: 5 (62.5 %)
- Hombres: 3 (37.5 %)
- Edad promedio de los sobrevivientes: 48.9 años
- Rango de edad: 34-70 años

### **35. Análisis por Sexo:**

#### **35.1. Mujeres:**

- Total de mujeres: 14 (74 % de la muestra)
- Fallecidas: 7 (50 % de las mujeres)

- Vivas: 7 (50 % de las mujeres)
- Edad promedio: 45.1 años

### **35.2. Hombres:**

- Total de hombres: 5 (26 % de la muestra)
- Fallecidos: 4 (80 % de los hombres)
- Vivos: 1 (20 % de los hombres)
- Edad promedio: 39.2 años

## **36. Análisis por Edad:**

### **36.1. Menores de 40 años:**

- Total: 6 pacientes (32 % de la muestra)
- Fallecidos: 2 (33 % de los menores de 40 años)
- Vivos: 4 (67 % de los menores de 40 años)
- Edad promedio: 32.8 años

### **36.2. Entre 40 y 50 años:**

- Total: 10 pacientes (53 % de la muestra)
- Fallecidos: 7 (70 % de los pacientes en este rango)
- Vivos: 3 (30 % de los pacientes en este rango)
- Edad promedio: 45.5 años

### **36.3. Mayores de 50 años:**

- Total: 3 pacientes (15 % de la muestra)
- Fallecidos: 2 (67 % de los mayores de 50 años)
- Vivos: 1 (33 % de los mayores de 50 años)
- Edad promedio: 60.3 años

## **37. Análisis de Variables Médicas:**

### **37.1. Esteroides:**

- Total de pacientes que consumieron esteroides: 6 (32 % de la muestra)
- Fallecidos entre los que consumieron esteroides: 5 (83 % de los que consumieron)
- Sobreviviente entre los que consumieron esteroides: 1 (17 % de los que consumieron)

### **37.2. Antivirales:**

- Total de pacientes que consumieron antivirales: 2 (11 % de la muestra)
- Fallecidos entre los que consumieron antivirales: 2 (100 % de los que consumieron)

### **37.3. Fatiga:**

- Total de pacientes que experimentaron fatiga: 15 (79 % de la muestra)
- Fallecidos entre los que experimentaron fatiga: 10 (67 % de los que experimentaron)
- Sobrevivientes entre los que experimentaron fatiga: 5 (33 % de los que experimentaron)

### **37.4. Anorexia:**

- Total de pacientes que experimentaron anorexia: 4 (21 % de la muestra)
- Fallecidos entre los que experimentaron anorexia: 3 (75 % de los que experimentaron)
- Sobrevivientes entre los que experimentaron anorexia: 1 (25 % de los que experimentaron)

### **37.5. Albúmina:**

- Total de pacientes con niveles altos de albúmina: 4 (21 % de la muestra)
- Fallecidos entre los que tienen niveles altos de albúmina: 2 (50 % de los que tienen)
- Sobrevivientes entre los que tienen niveles altos de albúmina: 2 (50 % de los que tienen)

### **37.6. Alk Phosphate:**

- Total de pacientes con niveles altos de fosfatasa alcalina: 5 (26 % de la muestra)
- Fallecidos entre los que tienen niveles altos de fosfatasa alcalina: 3 (60 % de los que tienen)
- Sobrevivientes entre los que tienen niveles altos de fosfatasa alcalina: 2 (40 % de los que tienen)

### **37.7. Sgot:**

- Total de pacientes con niveles altos de SGOT: 7 (37 % de la muestra)
- Fallecidos entre los que tienen niveles altos de SGOT: 5 (71 % de los que tienen)
- Sobrevivientes entre los que tienen niveles altos de SGOT: 2 (29 % de los que tienen)

### **37.8. Albumin:**

- Total de pacientes con niveles bajos de albúmina: 7 (37 % de la muestra)
- Fallecidos entre los que tienen niveles bajos de albúmina: 5 (71 % de los que tienen)
- Sobrevivientes entre los que tienen niveles bajos de albúmina: 2 (29 % de los que tienen)

### **37.9. Prottime:**

- Total de pacientes con niveles altos de protime: 3 (16 % de la muestra)
- Fallecidos entre los que tienen niveles altos de protime: 2 (67 % de los que tienen)
- Sobrevivientes entre los que tienen niveles altos de protime: 1 (33 % de los que tienen)

### **37.10. Histology:**

- Total de pacientes con histología positiva: 8 (42 % de la muestra)
- Fallecidos con histología positiva: 5 (63 % de los que tienen)
- Sobrevivientes con histología positiva: 3 (37 % de los que tienen)

### **38. Conclusiones Preliminares:**

- Se observa una alta tasa de mortalidad entre quienes han usado esteroides.
- Los síntomas como fatiga y malaise son comunes entre los fallecidos.
- La presencia de signos de enfermedad hepática avanzada es un factor de riesgo.
- Los valores de laboratorio también muestran asociaciones con la mortalidad.
- La histología sigue siendo un punto clave en el diagnóstico y pronóstico.

### **39. Sugerencias para un Análisis Más Profundo:**

- Un análisis de regresión logística podría identificar los factores más significativos en la mortalidad.
- Clasificar a los pacientes en diferentes etapas de enfermedad hepática podría proporcionar una visión más detallada.
- Comparar los valores de laboratorio entre los grupos podría revelar patrones específicos asociados con la mortalidad.
- Un análisis de supervivencia podría estimar las tasas de supervivencia a lo largo del tiempo.

### **40. Limitaciones:**

- Tamaño de la muestra pequeño, limitando la generalización de resultados.
- Causas de muerte no disponibles para un análisis más detallado de factores de riesgo.
- Factores adicionales como el estilo de vida no considerados en el análisis.
- Variables con proporciones muy bajas de casos pueden afectar la interpretación.

### **41. Recomendaciones:**

- Investigación adicional con mayor tamaño de muestra para confirmar las hipótesis.
- Análisis de causas de muerte para entender mejor los factores de supervivencia.
- Considerar otros factores de riesgo como el consumo de alcohol y la obesidad en futuros análisis.



## 42. Cluster 1 para K=5

[illegible]

Figura 13: Cluster 1

Para el cluster 1, tenemos un total de 34 personas, de las cuales 31 están muertas y 3 vivas. Las edades de las personas varían entre 22 y 56 años, con 5 de ellos siendo masculinos y 29 femeninos. De las 34 personas, 18 consumen esteroides, 15 no lo hacen, y la información sobre el consumo de esteroides de 1 persona es desconocida. En cuanto a antivirales, 13 personas los toman y 21 no. Respecto a la fatiga, 17 personas la presentan y 17 no. En cuanto al malestar, 9 personas lo experimentan y 25 no. En cuanto a la anorexia, 6 personas la presentan y 28 no. Con respecto al hígado grande, 26 personas lo tienen y 8 no. Para el hígado firme, 13 personas lo presentan y 21 no. En cuanto al bazo palpable, 4 personas lo tienen y 30 no. Para las arañas vasculares, 6 personas las tienen y 28 no. En el caso de la ascitis, solo 1 persona la presenta y 33 no. Para las várices, 2 personas tienen várices y 32 no.

En el caso de la bilirrubina, la mayoría se encuentra en el rango promedio de 0.2 a 1.2. Solo 4 personas tienen valores por encima de ese promedio. Para la fosfatasa alcalina, el rango es de 20 a 140 unidades por litro, y las personas mantienen valores estables dentro de ese rango, con valores superiores a 20 pero inferiores a 140. En cuanto a SGOT, el promedio es de 10 a 40 unidades por litro, y la mayoría de las personas se encuentra dentro de ese rango, aunque 15 personas están por encima de dicho límite. En cuanto a la albúmina, el promedio es de 3.4 a 5.4, y la mayoría de las personas se encuentra dentro de ese rango, con la mayoría teniendo un promedio de 4. Por último, en cuanto a la histología, 8 personas cuentan con histología y 26 no.

El texto proporciona un análisis detallado de las características de un grupo de personas pertenecientes al cluster 1. Aquí hay un resumen de los puntos clave:

## Demografía:

- El grupo consiste en 34 personas, con una mayoría de 31 personas fallecidas y 3 personas vivas.
- La distribución de edades va desde los 22 hasta los 56 años, con una presencia mayoritaria de mujeres (29) en comparación con hombres (5).

### Consumo de Esteroides y Antivirales:

- El 53 % de las personas consume esteroides, mientras que el 38 % toma antivirales.
- Se menciona que la información sobre el consumo de esteroides de una persona es desconocida.

#### **Síntomas y Condiciones Médicas:**

- Hay presencia significativa de fatiga (50 %) y ausencia de malestar (25 %).
- Un pequeño porcentaje presenta anorexia (17 %), mientras que la mayoría no la experimenta.
- La presencia de hígado grande es notable (76 %), y el hígado firme se observa en el 38 % de las personas.

#### **Otros Indicadores Médicos:**

- El bazo palpable, las arañas vasculares, la ascitis y las várices tienen bajas prevalencias en el grupo.
- La bilirrubina muestra un comportamiento dentro del rango promedio, con solo 4 personas por encima de este rango.
- La fosfatasa alcalina y SGOT se mantienen en rangos normales para la mayoría, aunque 15 personas tienen valores elevados de SGOT.
- La albúmina está mayormente dentro del rango normal, con la mayoría de las personas con un promedio de 4.

#### **Histología:**

- Se menciona que 8 personas cuentan con histología, mientras que 26 no la tienen.

### **42.1. Conclusión para el centroide 1:**

- La mayoría de las personas fallecidas en este grupo consumían esteroides.
- Ninguna de las personas utilizaba antivirales.
- Aunque presentaban fatiga, no mostraban anorexia; sin embargo, solo 3 personas sí la presentaban.
- La mayoría de las personas en este grupo tenían un hígado grande.
- La mayoría no presentaba várices, siendo solo 1 persona viva la que las presentaba.
- En cuanto a los niveles de bilirrubina, todos estaban dentro del promedio.
- La mayoría de las personas no contaba con histología.

## 43. Cluster 2 para k=5

DE				50	MALE		NO	NO	YES	NO	NO	NO	YES	YES	YES	NO	NO	0.9	290	117	3.4	41	YES		1
LIVE				38	FEMALE		NO	NO	NO	NO	NO	YES	YES	NO	NO	NO	NO	0.4	243	49	3.8	90	YES		1
DE				38	FEMALE		NO	NO	NO	NO	NO	NO	YES	YES	NO	NO	NO	0.7	184	44	3.1	41	NO		1
DE				25	FEMALE		YES	NO	YES	NO	NO	NO	YES	YES	YES	YES	YES	1.3	181	181	4.5	57	YES		1
DE				47	MALE		NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	1.3	179	69	2.9	0	NO		1
DE				36	FEMALE		NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1.1	141	75	2.3	0	YES		1
LIVE				30	FEMALE		YES	NO	YES	YES	YES	YES	YES	NO	YES	YES	YES	2.5	105	64	2.8	0	YES		1
DE				30	FEMALE		NO	NO	YES	YES	NO	YES	YES	NO	YES	NO	NO	0.8	147	128	3.9	100	YES		1
DE				7	FEMALE		YES	NO	NO	NO	YES	YES	YES	YES	NO	NO	NO	0.7	256	25	4.2	0	YES		1
DE				50	FEMALE		YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	139	81	3.9	62	YES		1
LIVE				25	FEMALE		NO	NO	YES	NO	NO	NO	YES	YES	YES	YES	NO	1.5	138	58	2.6	0	YES		1
LIVE				47	FEMALE		YES	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES	1	195	29	2.5	21	YES		1
DE				65	FEMALE		YES	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES	0.5	180	93	2.9	74	YES		1
DE				20	FEMALE		NO	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	2.5	150	88	3.9	0	NO		1
DE				23	FEMALE		YES	NO	YES	YES	YES	YES	NO	YES	NO	NO	NO	1.9	184	100	4.1	90	NO		1
DE				35	FEMALE		NO	NO	YES	YES	YES	NO	YES	NO	YES	NO	YES	1.7	295	60	2.7	0	YES		1
DE				27	FEMALE		YES	NO	YES	YES	YES	NO	YES	YES	YES	YES	NO	1.2	153	98	4.1	39	NO		1
DE				20	FEMALE		NO	NO	YES	YES	YES	YES	NO	NO	YES	YES	NO	1	190	118	2.9	23	YES		1
LIVE				39	FEMALE		NO	YES	YES	YES	NO	YES	YES	NO	NO	NO	NO	2.3	280	58	3.8	40	NO		1
LIVE				50	FEMALE		YES	NO	YES	NO	NO	YES	YES	YES	NO	YES	YES	2.9	185	75	2.4	32	YES		1
DE				38	MALE		YES	NO	YES	NO	NO	YES	YES	NO	YES	NO	NO	1.4	175	55	2.7	38	NO		1
LIVE				44	FEMALE		NO	NO	YES	YES	NO	YES	NO	YES	NO	NO	YES	0.9	180	50	0	41	YES		1
DE				32	FEMALE		YES	NO	YES	YES	YES	YES	NO	NO	YES	NO	YES	3.5	215	54	3.4	29	NO		1

Figura 14: Cluster 2

### Análisis del Cluster 2:

#### ■ Demografía:

- Total de 23 personas, de las cuales 16 están fallecidas y 7 están vivas.
- 19 son de género femenino y 4 de género masculino.

#### ■ Consumo de Medicamentos:

- 11 personas consumían esteroides, mientras que 12 no.
- 22 personas no consumían antivirales, y solo 1 no los consumía.

#### ■ Síntomas y Condiciones Médicas:

- 18 personas presentaban fatiga, y 5 no.
- 13 personas tenían malestares, mientras que 10 no los tenían.
- 7 personas tenían anorexia, y 16 no.
- 16 personas tenían un hígado grande, 6 no lo tenían, y el valor de 1 persona es desconocido.
- 15 personas presentaban hígado firme, 7 no lo tenían, y el valor de 1 persona es desconocido.
- 11 personas tenían bazo palpable, 11 no lo tenían, y el valor de 1 persona es desconocido.
- 13 personas tenían arañas, 9 no las tenían, y el valor de 1 persona es desconocido.
- 7 personas tenían ascitis, 15 no la tenían, y el valor de 1 persona es desconocido.
- 7 personas tenían várices, 15 no las tenían, y el valor de 1 persona es desconocido.
- 15 personas tenían histología, y 8 no la tenían.

#### ■ Rango de Edad:

- El rango de edad es de 30 a 55 años para este cluster.

■ **Indicadores Médicos:**

- **Bilirrubina:** 12 personas están en el rango de 0.2 a 1.2 (promedio deseado), mientras que 11 personas están fuera de ese rango.
- **Fosfatasa Alcalina:** 4 personas están dentro del rango de 20 a 140 unidades por litro, mientras que 19 no lo están, indicando que la mayoría no está dentro de ese promedio.
- **SGOT:** 2 personas están en el rango de 10 a 40 unidades por litro, mientras que la mayoría, que son 21, no lo está.
- **Albumina:** 11 personas están dentro del rango de 3.4 a 5.4, 11 personas no lo están, y el valor de 1 persona es desconocido.

### 43.1. Conclusión del Cluster 2:

■ **Demografía:**

- Total de 23 personas, con 16 fallecidas y 7 vivas.
- Las edades oscilan entre 30 y 55 años, con una mayoría de género femenino (19 personas).

■ **Consumo de Medicamentos:**

- 11 personas consumían esteroides, mientras que 12 no lo hacían.
- La mayoría no consumía antivirales (22 personas), y solo 1 persona no los consumía.

■ **Síntomas y Condiciones Médicas:**

- La fatiga es común (18 personas), mientras que la anorexia afecta a 7 personas.
  - La mayoría presenta un hígado grande (16 personas), y algunas tienen ascitis (7 personas).
  - La bilirrubina muestra variabilidad, con 12 personas en el rango promedio y 11 fuera de él.
  - Los niveles de fosfatasa alcalina y SGOT varían, con la mayoría fuera del rango esperado.
  - La mayoría tiene niveles de albúmina dentro del rango, y 15 personas tienen histología.
- La bilirrubina, un indicador clave de la función hepática, muestra que 12 personas están dentro del rango de 0.2 a 1.2, mientras que 11 personas están fuera de ese rango.
- 7 personas en este cluster presentan anorexia, un síntoma asociado comúnmente con enfermedades hepáticas avanzadas.

- 18 personas experimentan fatiga, lo que podría ser un indicador de la carga adicional en el sistema hepático.
- 7 personas tienen ascitis, una acumulación anormal de líquido en la cavidad abdominal, indicando una progresión avanzada de las enfermedades hepáticas en este grupo.

## 44. Cluster 3 para k=5

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Class	Age	Sex	Stead	Arterials	Fatigue	Malaise	Anorexia	Liver Bg	Liver Fim	Spleen	Palp	Spleen	Ascites	Varices	Bilirubin	Alb Phosph	Sgpt	Albumin	Protime	Hemoglobin	cluster	
2	DE	28	FEMALE	YES	NO	YES	YES	YES	NO	NO	NO	NO	YES	YES	NO	0	0	20	4	0	YES	2	
3	DE	23	FEMALE	YES	NO	YES	YES	YES	NO	NO	NO	NO	NO	NO	0.8	0	71	4.8	0	NO	2		
4	DE	23	FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.9	88	75	4	0	YES	2		
5	DE	23	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	0	7	0	0	NO	2		
6	DE	33	FEMALE	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1	0	60	0	0	NO	2		
7	DE	28	FEMALE	NO	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	18	44	123	4	46	NO	2		
8	DE	54	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	YES	NO	12	86	32	3.1	66	YES	2		
9	DE	52	FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	15	0	62	2.8	0	YES	2		
10	DE	44	FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	0.8	108	142	4.3	0	YES	2		
11	DE	80	FEMALE	NO	NO	YES	NO	NO	YES	YES	YES	YES	YES	NO	0	0	40	0	0	YES	2		
12	DE	22	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.7	0	24	0	0	YES	2		
13	DE	82	MALE	YES	NO	YES	YES	NO	YES	YES	NO	YES	NO	NO	1.3	141	68	2.9	58	NO	2		
14	DE	39	FEMALE	YES	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	0.7	0	48	0	0	YES	2		
15	DE	83	MALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	3.2	119	138	0	0	YES	2		
16	DE	39	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1.2	81	123	3.9	0	NO	2		
17	DE	42	FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.9	60	63	4.7	47	NO	2		
18	DE	45	FEMALE	NO	NO	YES	NO	NO	YES	YES	YES	YES	YES	YES	12	81	65	3	0	NO	2		
19	DE	35	FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	0.9	58	30	4.3	73	NO	2		
20	DE	33	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1	46	80	4.4	60	NO	2		
21	DE	34	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0	0	116	4.6	0	NO	2		
22	DE	39	FEMALE	YES	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	0.7	74	136	3.6	0	YES	2		
23	DE	38	FEMALE	YES	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES	16	120	160	3.5	56	YES	2		
24	DE	40	FEMALE	NO	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	0.6	62	155	4	63	YES	2		
25	DE	33	FEMALE	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	0.8	85	69	4.5	0	YES	2		
26	DE	38	FEMALE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	0.9	85	69	4	0	NO	2		
27	DE	59	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	15	105	129	5.3	0	NO	2		
28	DE	44	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.6	110	145	4.4	20	NO	2		
29	DE	39	FEMALE	NO	NO	YES	YES	YES	NO	NO	NO	NO	YES	NO	2	72	89	0	46	NO	2		
30	DE	27	FEMALE	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	0.8	82	59	0	0	NO	2		
31	DE	47	FEMALE	NO	NO	YES	YES	YES	NO	NO	NO	NO	NO	NO	0	0	85	75	0	NO	2		
32	DE	40	MALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	2.2	89	144	4.9	76	NO	2		
33	DE	39	FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	1.2	75	173	4.2	94	YES	2		
34	DE	24	FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	0.8	0	34	4.1	0	YES	2		
35	DE	24	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.6	0	34	4.4	0	YES	2		
36	DE	27	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1.6	67	157	2.7	0	YES	2		
37	DE	42	FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	YES	YES	YES	4.6	68	95	3.3	0	YES	2		
38	DE	41	MALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	2	72	89	0	46	NO	2		
39	DE	39	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	0.6	0	34	4.4	0	YES	2		
40	DE	42	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	1.6	68	95	3.3	0	YES	2		
41	DE	42	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	4.6	68	95	3.3	0	YES	2		
42	DE	39	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	0.7	67	80	3	21	YES	2		
43	DE	42	FEMALE	YES	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	4.6	123	157	2.7	0	YES	2		
44	DE	58	FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	YES	NO	2.9	90	153	4	0	YES	2		
45	DE	82	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	0	0	82	0	0	NO	2		
46	DE	41	FEMALE	YES	NO	YES	NO	NO	YES	YES	YES	NO	NO	NO	4.2	85	120	3.4	0	YES	2		
47	DE	42	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	2.8	107	162	3.6	0	YES	2		
48	DE	42	FEMALE	NO	YES	YES	YES	NO	YES	NO	NO	NO	NO	NO	0.5	60	88	3.8	28	YES	2		
49	DE	42	FEMALE	YES	NO	YES	YES	YES	YES	YES	NO	NO	YES	NO	1.9	0	114	2.4	0	YES	2		
50	DE	31	FEMALE	NO	NO	YES	YES	YES	NO	YES	NO	NO	NO	NO	8	0	101	2.2	0	YES	2		
51																							

Figura 15: Cluster 3

### Análisis del Cluster 3:

- **Demografía:**
  - Total de 49 personas, con 38 vivas y 11 fallecidas.
  - Rango de edad entre 20 y 40 años, con una mayoría femenina (45 personas).
- **Consumo de Medicamentos:**
  - 24 personas consumen esteroides, mientras que 25 no.
  - Distribución equilibrada en comparación con otros clusters.
- **Síntomas y Condiciones Médicas:**
  - Fatiga común en 31 personas, mientras que 18 no la experimentan.
  - Malestares presentes en 20 personas, mientras que 29 no los tienen.
  - Anorexia poco frecuente, afectando a 8 personas.
- **Condiciones Hepáticas:**

- Hígado grande en la mayoría (42 personas), 4 sin hígado grande y 2 desconocido.
- Hígado firme presente en 16 personas, 29 sin hígado firme y 4 desconocido.
- Bazo palpable en 7 personas.

■ **Manifestaciones Cutáneas:**

- 13 personas presentan arañas, indicando variabilidad en las manifestaciones cutáneas.

■ **Complicaciones Hepáticas:**

- 6 personas presentan ascitis, mientras que 43 no la tienen.
- 4 personas presentan varices.

■ **Indicadores Médicos:**

- Bilirrubina muestra variabilidad, con 29 dentro del rango y 17 fuera, el rango es de 0.2 a 1.2 miligramos.
- Fosfatasa alcalina dentro del rango en 29 personas, 1 fuera de rango, el rango va de 20 a 140 unidades por litro.
- SGOT dentro del rango en 8 personas, 40 sin información, el rango es de 10 a 40 unidades por litro.
- Albumina dentro del rango en 8 personas, 40 sin información, el rango es de 3.4 a 5.4.

#### 44.1. Conclusiones del Cluster 3:

1. **Demografía y Edad:**

- El Cluster 3 está compuesto principalmente por individuos jóvenes, con edades entre 20 y 40 años.
- La mayoría de las personas en este grupo están vivas (38 de 49).

2. **Consumo de Medicamentos:**

- Existe una distribución equilibrada en cuanto al consumo de esteroides, con 24 personas que los consumen y 25 que no.
- Este equilibrio podría indicar una variedad en los tratamientos médicos dentro del grupo.

3. **Síntomas y Condiciones Médicas:**

- La fatiga es un síntoma común, afectando a 31 personas, mientras que 18 no la experimentan.
- La anorexia es poco frecuente, presentándose en 8 personas.

- La presencia de malestares es relativamente equilibrada, con 20 personas que los presentan y 29 que no.

#### 4. Condiciones Hepáticas:

- La mayoría de las personas tienen un hígado grande (42 personas), indicando posiblemente afecciones hepáticas subyacentes.
- La presencia de ascitis en 6 personas sugiere complicaciones hepáticas avanzadas en algunos casos.

#### 5. Manifestaciones Cutáneas y Complicaciones Vasculares:

- La presencia de arañas en 13 personas indica variabilidad en las manifestaciones cutáneas asociadas con enfermedades hepáticas.
- Cuatro personas presentan varices, sugiriendo posibles complicaciones vasculares relacionadas con el hígado.

#### 6. Indicadores Médicos:

- La variabilidad en los niveles de bilirrubina y la falta de información sobre algunos indicadores médicos indican la complejidad y diversidad de la salud dentro del grupo.
- Algunas personas tienen valores fuera del rango esperado en indicadores como la bilirrubina y la fosfatasa alcalina.

## 45. Cluster 4 para k=5

Class	Age	Sex	Stoicid	Anestesia	Fatigue	Melasma	Anorexia	Liver-Big	Liver-Firm	Screen-Papal	Spectres	Acetone	Varicos	Bilirrubin	Alk-Phosphat	Sgot	Albumen	Protume	Hemology	cluster
LIVE	45 FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	YES	NO	NO	0	0	0	0	0	YES	3
LIVE	49 FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	NO	NO	NO	1.4	85	70	3.5	35	YES	3
DE	45 FEMALE	NO	NO	YES	YES	NO	NO	YES	NO	YES	NO	NO	NO	0.8	75	20	4.1	0	YES	3
DE	42 FEMALE	YES	NO	NO	NO	NO	YES	NO	YES	NO	NO	NO	NO	1.5	85	40	0	0	YES	3
DE	48 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.8	100	10	4	0	YES	3
LIVE	37 FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	YES	YES	NO	NO	4.8	82	55	3.3	30	YES	3
DE	72 FEMALE	YES	YES	YES	NO	NO	YES	YES	NO	NO	NO	NO	NO	1	110	52	3.4	50	YES	3
LIVE	64 FEMALE	NO	NO	YES	YES	NO	NO	YES	NO	NO	YES	NO	NO	3.9	120	28	3.5	43	YES	3
DE	54 FEMALE	YES	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	3.2	85	20	3.8	0	YES	3
DE	54 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	85	30	4.3	0	YES	3
DE	30 MALE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1	85	10	4	0	NO	3
DE	38 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	85	20	4	0	NO	3
DE	50 FEMALE	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	0.8	105	42	3.5	0	NO	3
DE	76 FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	95	32	4	0	NO	3
LIVE	51 FEMALE	NO	NO	YES	NO	YES	NO	YES	NO	YES	YES	NO	NO	0	0	0	0	0	NO	3
DE	41 FEMALE	YES	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	NO	0.9	81	60	3.9	52	NO	3
DE	56 FEMALE	YES	NO	YES	NO	YES	NO	YES	NO	NO	NO	NO	NO	1.2	105	53	4.3	0	NO	3
DE	49 FEMALE	NO	YES	YES	YES	YES	YES	YES	NO	YES	NO	NO	NO	0.8	85	40	3.7	0	NO	3
DE	51 FEMALE	NO	YES	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	1	75	58	4.6	52	NO	3
DE	35 MALE	NO	NO	NO	NO	NO	YES	YES	NO	NO	NO	NO	NO	0.5	105	25	3.8	60	NO	3
LIVE	37 FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	NO	0.6	67	28	4.2	0	NO	3
DE	52 FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	75	15	4	25	NO	3
DE	22 FEMALE	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	4.5	25	15	4.5	0	NO	3
DE	38 FEMALE	YES	YES	YES	YES	YES	YES	NO	NO	YES	NO	NO	NO	0.7	105	65	4.2	77	NO	3
DE	52 FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	85	30	4	0	YES	3
DE	34 FEMALE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	0	0	0	0	0	NO	3
DE	48 FEMALE	YES	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	0.8	100	42	3.5	60	NO	3
LIVE	43 FEMALE	YES	NO	YES	NO	NO	YES	NO	YES	YES	YES	NO	NO	1.2	100	10	3.1	45	YES	3
DE	34 FEMALE	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	0.7	62	30	5	0	NO	3
LIVE	47 FEMALE	YES	NO	YES	YES	YES	YES	YES	NO	NO	YES	YES	YES	1.7	85	20	2.1	48	YES	3
DE	38 FEMALE	NO	NO	YES	YES	YES	YES	YES	NO	YES	YES	YES	YES	1.2	118	18	2.8	0	YES	3
DE	38 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	85	11	4.8	0	NO	3
DE	33 MALE	NO	NO	YES	NO	NO	YES	NO	YES	YES	NO	YES	YES	1.5	81	18	4.1	48	YES	3
DE	44 FEMALE	NO	NO	YES	YES	NO	NO	YES	NO	YES	NO	NO	NO	3	114	65	3.5	0	YES	3
DE	24 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.8	82	30	4.3	0	NO	3

Figura 16: Cluster 4

### Análisis del Cluster 4:

#### 1. Demografía y Edad:

- El Cluster 4 cuenta con 35 personas, de las cuales 9 están vivas y 26 fallecidas.

- La edad se encuentra en un rango de 30 a 50 años, indicando un grupo de mediana edad.
- Predominio femenino con 32 mujeres y 3 hombres.

## 2. Consumo de Medicamentos:

- El consumo de esteroides está equilibrado, con 17 personas que los consumen y 18 que no.
- Antivirales son utilizados por 6 personas, mientras que 29 no los consumen.

## 3. Síntomas y Condiciones Médicas:

- La fatiga afecta a 23 personas, mientras que 11 no la experimentan, y 1 persona tiene información desconocida.
- Malestares presentes en 12 personas, 22 no los tienen, y en 1 persona se desconoce esta información.
- La anorexia es relativamente baja, afectando a 4 personas, mientras que 3 no la tienen y 1 persona tiene información desconocida.

## 4. Condiciones Hepáticas:

- La mayoría presenta hígado grande (26 personas), indicando posiblemente afecciones hepáticas subyacentes.
- Hígado firme está presente en 9 personas, mientras que 22 no lo tienen y en 4 personas se desconoce esta información.
- La presencia de bazo palpable en 6 personas podría indicar una afectación del sistema esplénico en algunos casos.

## 5. Manifestaciones Cutáneas y Complicaciones Vasculares:

- 11 personas presentan arañas, mientras que 22 no las tienen, y en 2 personas se desconoce esta información.
- 5 personas presentan ascitis, indicando posibles complicaciones hepáticas avanzadas.
- 3 personas presentan varices, mientras que 30 no las tienen, y en 2 personas se desconoce esta información.

## 6. Indicadores Médicos:

- Bilirrubina: Tres personas tienen información desconocida, 23 están dentro del rango y 9 fuera de él.
- Fosfatasa alcalina está dentro del rango en 32 personas, y en 3 personas se desconoce esta información.
- SGOT está dentro del rango en 20 personas, 12 fuera de él, y en 3 personas se desconoce esta información.



- Albumina está dentro del rango en 26 personas, 5 fuera de él, y en 4 personas se desconoce esta información.

#### **7. Histología:**

- Dieciséis personas tienen información sobre la histología, mientras que 19 no.

### **45.1. Conclusiones del Cluster 4:**

#### **1. Demografía y Edad:**

- El grupo se compone principalmente de mujeres, con una mayoría significativa (26) de individuos fallecidos.
- La edad media se encuentra entre los 30 y 50 años, señalando un grupo de mediana edad.

#### **2. Consumo de Medicamentos:**

- Existe un equilibrio en el consumo de esteroides, mientras que la mayoría no utiliza antivirales.

#### **3. Síntomas y Condiciones Médicas:**

- La fatiga y los malestares son comunes, pero hay variabilidad en la presencia de anorexia.
- La información desconocida en algunos casos destaca la falta de datos completos.

#### **4. Condiciones Hepáticas:**

- La presencia de hígado grande y firme indica posibles afecciones hepáticas avanzadas en este grupo.
- Algunos individuos presentan bazo palpable, sugiriendo afectación en el sistema esplénico.

#### **5. Manifestaciones Cutáneas y Complicaciones Vasculares:**

- La presencia de arañas y ascitis señala complicaciones hepáticas y vasculares significativas.
- La información desconocida en algunos casos indica la necesidad de registros más detallados.

#### **6. Indicadores Médicos:**

- Niveles de bilirrubina fuera del rango en 9 personas sugieren posibles problemas hepáticos.
- La mayoría de los indicadores médicos, como fosfatasa alcalina, SGOT y albumina, se encuentran dentro del rango esperado.

## 7. Histología:

- Un número significativo de personas tiene información sobre la histología, lo que puede proporcionar información valiosa sobre la estructura celular del hígado.

## 46. Cluster 5 para k=5

Case	id	sex	age	smoke	fatigue	anorexia	liver_firm	liver_big	swollen	palmar_furrows	ascites	hemocrit	bilirubin	alt	prothromb	inr	albumin	protime	hematology	cluster
DE	28	FEMALE	YES	NO	YES	YES	NO	YES	YES	NO	NO	NO	0.7	351	420	3.3	48	NO	4	
DE	30	FEMALE	YES	NO	NO	NO	YES	YES	NO	NO	NO	NO	0.9	62	234	4.2	100	NO	4	
DE	51	FEMALE	YES	NO	YES	YES	YES	YES	NO	NO	NO	NO	0.9	70	271	4.4	0	YES	4	
DE	37	FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	0.9	0	231	4.3	0	YES	4	
DE	46	FEMALE	YES	YES	YES	YES	YES	YES	NO	YES	NO	NO	2	168	276	3.8	0	YES	4	
DE	27	FEMALE	NO	NO	YES	NO	YES	YES	NO	NO	NO	NO	2.4	168	227	3	60	YES	4	
DE	32	FEMALE	YES	YES	YES	NO	NO	YES	NO	YES	NO	NO	1	59	249	3.7	24	NO	4	
DE	51	FEMALE	NO	NO	YES	YES	YES	YES	YES	YES	NO	YES	4.6	215	209	3.9	51	YES	4	
DE	54	FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	1	0	200	4	0	NO	4	
DE	45	FEMALE	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	2.3	0	648	0	0	YES	4	
DE	54	FEMALE	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	1	165	225	3.6	67	YES	4	
DE	70	FEMALE	NO	NO	YES	YES	YES	YES	NO	YES	YES	YES	1.7	109	528	2.8	35	YES	4	
LIVE	46	FEMALE	YES	NO	YES	YES	YES	YES	NO	YES	YES	YES	7.6	0	242	3.3	30	YES	4	
LIVE	58	FEMALE	YES	NO	YES	NO	NO	NO	YES	YES	NO	NO	2	187	242	3.3	0	NO	4	

Figura 17: Cluster 5

### Análisis del Cluster 5:

#### 1. Demografía y Edad:

- El Cluster 5 consiste en 14 personas, siendo la mayoría (11) fallecidas.
- El rango de edad promedio se encuentra entre 30 y 50 años.
- Predominio femenino, con solo 1 persona de género masculino.

#### 2. Consumo de Medicamentos:

- Hay un equilibrio en el consumo de esteroides, con 8 personas que los consumen y 6 que no.
- La mayoría (12 personas) utiliza antivirales.

#### 3. Síntomas y Condiciones Médicas:

- Fatiga es un síntoma común, afectando a 11 personas, mientras que 3 no la experimentan.
- Malestares presentes en 7 personas, sin embargo, 7 no los tienen.
- Algunas personas presentan anorexia, indicando posibles problemas alimenticios.

#### 4. Condiciones Hepáticas:

- La mayoría tiene el hígado grande (10 personas), mientras que 2 no lo tienen y 2 desconocen su estado.
- Hígado firme está presente en 7 personas, 5 no lo tienen, y 2 desconocen su estado.

- Pocas personas presentan bazo palpable (2), lo que puede indicar problemas en el sistema esplénico.

#### 5. Manifestaciones Cutáneas y Complicaciones Vasculares:

- Algunas personas presentan arañas (6), pero la mitad no las tienen.
- Solo 1 persona presenta ascitis, indicando posibles complicaciones hepáticas avanzadas.
- Varices están presentes en 2 personas, y en 2 se desconoce su estado.

#### 6. Indicadores Médicos:

- Niveles de bilirrubina: 6 personas están dentro del rango (0.2 a 1.2 mg), mientras que 8 no lo están.
- Fosfatasa alcalina: 4 personas están dentro del rango normal, 6 no lo están, y en 4 se desconoce su rango.
- SGOT: Ninguna persona está dentro del rango esperado, lo que sugiere posibles problemas hepáticos.
- Albumina: 8 personas están dentro del rango, 5 no lo están, y 1 tiene información desconocida.

#### 7. Histología:

- Ocho personas tienen información sobre la histología, mientras que 6 no la tienen.

### 46.1. Conclusiones del Cluster 5:

#### 1. Demografía y Edad:

- El Cluster 5, compuesto por 14 personas, presenta una mayoría significativa de individuos fallecidos (11).
- La edad promedio de este grupo oscila entre 30 y 50 años, destacando un conjunto de mediana edad.
- El predominio femenino es evidente, con solo una persona de género masculino en el grupo.

#### 2. Consumo de Medicamentos:

- Se observa un equilibrio en el consumo de esteroides, con 8 personas que los consumen y 6 que no.
- La mayoría de las personas en este cluster (12) utiliza antivirales.

#### 3. Síntomas y Condiciones Médicas:

- La fatiga es un síntoma común, afectando a 11 personas, mientras que 3 no la experimentan.

- La presencia de malestares es notable en 7 personas, sin embargo, 7 no presentan este síntoma.
- La anorexia se manifiesta en 6 personas, sugiriendo posibles problemas alimenticios.

#### 4. **Condiciones Hepáticas:**

- La mayoría de las personas tiene el hígado grande (10), mientras que 2 no lo tienen y 2 desconocen su estado.
- La firmeza del hígado está presente en 7 personas, 5 no lo tienen y 2 desconocen su estado.
- La presencia de bazo palpable en 2 personas podría indicar problemas en el sistema esplénico.

#### 5. **Manifestaciones Cutáneas y Complicaciones Vasculares:**

- Algunas personas presentan arañas (6), pero la mitad no las tienen.
- La presencia de ascitis, observada en 1 persona, sugiere complicaciones hepáticas avanzadas.
- La existencia de varices en 2 personas y la falta de información en otras 2 indica posibles problemas vasculares.

#### 6. **Indicadores Médicos:**

- En cuanto a la bilirrubina, 6 personas se encuentran dentro del rango (0.2 a 1.2 mg), mientras que 8 no lo están.
- La fosfatasa alcalina muestra variabilidad, con 4 personas dentro del rango normal, 6 fuera de él y 4 con información desconocida.
- SGOT no presenta ninguna persona dentro del rango esperado, señalando posibles problemas hepáticos en todo el grupo.
- La albumina muestra 8 personas dentro del rango, 5 fuera de él y 1 con información desconocida.

#### 7. **Histología:**

- Ocho personas tienen información sobre la histología, mientras que 6 no la tienen.
- La falta de información histológica en algunos casos indica la necesidad de registros más detallados para comprender mejor la estructura celular del hígado en este grupo.

47. Clusters para k=10

Class	Age	Sex	Smoker	Antisocials	Fatigue	Malaise	Anorexia	Liver-Big	Liver-Firm	Spleen	Palpud	Spiders	Ascites	Varices	Bilirubin	Alk Phosphat	Sgot	Albumin	Protime	Hematology	cluster
LIVE		43 FEMALE	YES	NO	YES	NO	NO	YES	NO	YES	YES	YES	NO	NO	1.2	100	39	3.1	42	YES	0
DE		60 FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	1.2	100	50	4.3	0	NO	0
DE		53 FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	85	30	4	0	YES	0
DE		54 FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	3.2	85	28	3.8	0	YES	0
DE		53 MALE	NO	NO	YES	NO	NO	YES	NO	YES	YES	NO	YES	YES	1.5	81	19	4.1	40	YES	0
DE		42 FEMALE	YES	NO	NO	NO	NO	YES	NO	YES	NO	NO	NO	NO	1.5	85	40	0	0	YES	0
DE		54 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	85	30	4.5	0	YES	0
DE		60 FEMALE	NO	NO	YES	NO	NO	YES	YES	YES	YES	NO	NO	NO	0	0	40	0	0	YES	0
LIVE		47 FEMALE	YES	NO	YES	YES	NO	YES	YES	NO	NO	YES	YES	YES	1.7	86	20	2.1	40	YES	0
DE		56 FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	62	30	3	0	NO	0
DE		70 FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	96	32	4	0	NO	0
DE		61 FEMALE	NO	NO	YES	YES	NO	NO	YES	NO	YES	NO	NO	NO	0.8	79	20	4.1	0	YES	0

Figura 18: Cluster 1

Class	Age	Sex	Smoker	Antisocials	Fatigue	Malaise	Anorexia	Liver-Big	Liver-Firm	Spleen	Palpud	Spiders	Ascites	Varices	Bilirubin	Alk Phosphat	Sgot	Albumin	Protime	Hematology	cluster
DE		45 MALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	NO	NO	NO	1	85	75	0	0	NO	1
LIVE		62 FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	1	0	60	0	0	NO	1
DE		37 FEMALE	NO	NO	NO	NO	NO	NO	YES	NO	YES	NO	NO	NO	0.6	80	89	3.8	0	NO	1
DE		28 FEMALE	YES	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	0.7	74	110	4.6	0	NO	1
DE		52 FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1.5	0	60	2.8	0	YES	1
DE		44 FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	1.6	68	68	3.7	0	NO	1
DE		50 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1.5	100	100	5.3	0	NO	1
LIVE		43 FEMALE	NO	YES	YES	YES	NO	YES	NO	NO	YES	NO	NO	NO	0.5	62	68	3.8	29	YES	1
DE		54 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0	0	85	0	0	NO	1
DE		38 FEMALE	NO	NO	YES	YES	YES	YES	NO	NO	NO	YES	NO	NO	2	72	89	2.9	40	NO	1
LIVE		56 FEMALE	YES	NO	YES	YES	YES	NO	YES	NO	YES	NO	NO	NO	2.9	90	155	4	0	YES	1
DE		30 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	0	120	3.9	0	NO	1
LIVE		31 FEMALE	NO	NO	YES	YES	YES	YES	NO	YES	NO	NO	NO	NO	8	0	101	2.2	0	YES	1
LIVE		49 FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	NO	YES	NO	1.4	86	79	3.5	30	YES	1
LIVE		45 FEMALE	YES	NO	YES	YES	YES	YES	NO	NO	NO	YES	NO	NO	1.9	0	114	2.4	0	YES	1
LIVE		41 FEMALE	YES	NO	YES	YES	NO	YES	NO	NO	NO	YES	YES	YES	4.2	66	120	3.4	0	YES	1
DE		20 FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.9	89	152	4	0	YES	1
LIVE		33 FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	YES	NO	NO	0.7	63	86	3	31	YES	1

Figura 19: Cluster 2

Class	Age	Sex	Smoker	Antisocials	Fatigue	Malaise	Anorexia	Liver-Big	Liver-Firm	Spleen	Palpud	Spiders	Ascites	Varices	Bilirubin	Alk Phosphat	Sgot	Albumin	Protime	Hematology	cluster
LIVE		38 FEMALE	NO	NO	YES	YES	YES	YES	YES	NO	YES	YES	YES	YES	1.2	118	16	2.8	0	YES	2
DE		28 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	85	31	4.9	0	NO	2
DE		24 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.8	82	39	4.3	0	NO	2
DE		39 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	85	20	4	0	NO	2
DE		36 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.6	120	10	4	0	YES	2
DE		23 FEMALE	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	4.6	56	16	4.6	0	NO	2
DE		50 MALE	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1	85	15	4	0	NO	2
LIVE		37 FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	NO	0.6	67	28	4.2	0	NO	2

Figura 20: Cluster 3

Class	Age	Sex	Smoker	Antisocials	Fatigue	Malaise	Anorexia	Liver-Big	Liver-Firm	Spleen	Palpud	Spiders	Ascites	Varices	Bilirubin	Alk Phosphat	Sgot	Albumin	Protime	Hematology	cluster
DE		52 FEMALE	NO	YES	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	1	55	45	4.1	50	NO	3
DE		34 FEMALE	NO	NO	YES	YES	NO	NO	YES	NO	YES	NO	NO	NO	1	72	46	4.4	4	NO	3
DE		31 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	85	20	4	100	NO	3
DE		59 FEMALE	NO	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	1.5	78	30	4.4	85	NO	3
DE		22 MALE	YES	YES	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.9	48	20	4.2	64	NO	3
DE		64 FEMALE	YES	YES	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	1	80	38	4.5	74	NO	3
DE		30 FEMALE	YES	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	52	38	3.9	52	NO	3
DE		27 FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	0.8	95	40	3.8	100	NO	3
LIVE		47 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	YES	NO	2	84	23	4.2	60	YES	3
DE		39 FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	34	15	4	54	NO	3
DE		49 FEMALE	YES	YES	YES	YES	NO	YES	YES	YES	NO	NO	NO	NO	1.2	80	31	4	100	NO	3
DE		45 FEMALE	YES	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1.3	85	44	4.2	0	YES	3
DE		54 MALE	YES	NO	YES	YES	YES	NO	YES	NO	YES	NO	NO	NO	0.7	70	34	4.1	100	YES	3
DE		30 FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	100	31	4	100	NO	3
DE		36 FEMALE	NO	NO	NO	NO	NO	YES	NO	YES	NO	NO	NO	NO	0.8	88	44	4.2	80	NO	3
DE		23 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	0	0	0	0	NO	3
DE		41 MALE	YES	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	0.7	81	53	5	74	NO	3
DE		25 MALE	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.4	45	18	4.3	70	NO	3
DE		50 FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	71	18	4.4	100	NO	3
DE		54 FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.9	95	28	4	75	NO	3
DE		38 FEMALE	NO	YES	YES	YES	YES	NO	YES	NO	NO	NO	NO	NO	0.8	76	18	4.4	84	YES	3
DE		39 FEMALE	NO	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	0.7	70	28	4.2	52	NO	3
DE		42 FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	85	14	4	100	NO	3
DE		61 FEMALE	NO	NO	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	1.3	78	25	3.8	100	NO	3

Figura 21: Cluster 4

Class	Age	Sex	Smoker	Antisocials	Fatigue	Malaise	Anorexia	Liver-Big	Liver-Firm	Spleen	Palpud	Spiders	Ascites	Varices	Bilirubin	Alk Phosphat	Sgot	Albumin	Protime	Hematology	cluster
DE		56 MALE	YES	NO	YES	NO	NO	YES	YES	NO	YES	NO	NO	NO	1.4	175	55	2.7	50	NO	4
LIVE		35 FEMALE	NO	NO	YES	NO	NO	YES	YES	YES	YES	NO	YES	YES	1.5	138	58	2.6	0	YES	4
DE		7 FEMALE	YES	NO	NO	NO	NO	YES	YES	YES	NO	YES	NO	NO	0.7	296	25	4.2	0	YES	4
LIVE		30 FEMALE	YES	NO	YES	YES	YES	YES	YES	NO	YES	YES	YES	YES	2.5	165	64	2.8	0	YES	4
DE		36 FEMALE	NO	NO	NO	NO	NO	YES	NO	0.7	104	NO	YES	NO	0.7	104	NO	5.1	41	NO	4
DE		32 FEMALE	YES	NO	YES	YES	YES	NO	YES	NO	YES	NO	YES	YES	3.5	215	54	3.4	29	NO	4
LIVE		47 FEMALE	YES	NO	YES	YES	NO	YES	YES	NO	YES	YES	YES	YES	1	106	30	2.6	31	YES	4
DE		36 FEMALE	NO	NO	YES	YES	YES	YES	YES	NO	YES	NO	YES	YES	1.7	290	60	2.7	0	YES	4
LIVE		54 FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	YES	NO	YES	3.9	120	28	3.5	43	YES	4
DE		26 MALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.5	130	20	3.8	60	NO	4
DE		67 MALE	NO	NO	YES	YES	NO	YES	NO	NO	NO	NO	NO	NO	1.5	179	89	2.9	0	NO	4
DE		50 FEMALE	NO	NO	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	0.9	130	42	3.5	0	NO	4
LIVE		38 FEMALE	NO	NO	NO	NO	NO	YES	YES	NO	NO	NO	NO	NO	0.4	243	49	3.8	90	YES	4

Figura 22: Cluster 5

Class	Age	Sex	Steroid	Antitoxins	Fatigue	Malaise	Anorexia	Liver Big	Liver Firm	Spleen	Palpal	Spiders	Ascites	Venous	Bilirubin	Alk Phosphat	Sgpt	Albumin	Protime	Histology	cluster	
DIE	22	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	0	24	0	0	YES	5	
LIVE	40	FEMALE	NO	YES	YES	YES	YES	NO	YES	NO	NO	NO	NO	NO	0.6	40	69	4.2	67	YES	5	
DIE	26	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	50	42	4.1	65	YES	5	
DIE	24	FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	0	34	4.1	0	YES	5	
LIVE	61	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	YES	NO	NO	NO	0	0	0	0	0	0	YES	5
DIE	39	FEMALE	YES	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	NO	0.7	0	48	4.4	0	NO	5	
DIE	51	FEMALE	YES	NO	NO	NO	NO	YES	NO	YES	NO	YES	NO	NO	0.8	0	33	4.5	0	YES	5	
DIE	51	FEMALE	YES	NO	YES	NO	NO	YES	YES	YES	YES	NO	YES	1	0	30	3	65	YES	5		
LIVE	51	FEMALE	NO	NO	YES	NO	YES	YES	NO	YES	YES	NO	NO	NO	0	0	0	0	0	NO	5	
DIE	51	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	46	52	4	60	NO	5	
DIE	25	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.6	0	34	6.4	0	YES	5	
DIE	28	FEMALE	YES	NO	YES	YES	YES	YES	NO	YES	YES	NO	1	0	30	4	0	0	0	YES	5	
DIE	34	MALE	NO	YES	NO	NO	NO	YES	YES	NO	NO	NO	NO	NO	0.6	30	24	4	76	NO	5	
DIE	36	FEMALE	NO	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	1	0	48	4	57	NO	5	
DIE	30	FEMALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	56	78	4.2	74	NO	5	
LIVE	57	FEMALE	YES	NO	YES	YES	YES	YES	NO	NO	YES	YES	NO	NO	4.1	0	48	2.6	73	NO	5	
DIE	23	FEMALE	YES	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	0.8	0	14	4.9	0	NO	5	
DIE	34	FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0	0	0	0	0	NO	5	
DIE	27	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.6	0	36	4.2	0	NO	5	
DIE	57	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	26	58	4.5	100	NO	5	

Figura 23: Cluster 6

Class	Age	Sex	Steroid	Antitoxins	Fatigue	Malaise	Anorexia	Liver Big	Liver Firm	Spleen	Palpal	Spiders	Ascites	Venous	Bilirubin	Alk Phosphat	Sgpt	Albumin	Protime	Histology	cluster
LIVE	34	FEMALE	NO	NO	YES	YES	NO	NO	YES	NO	YES	NO	NO	NO	2.8	127	182	0	0	NO	6
DIE	38	FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1.1	141	79	3.3	0	YES	6
DIE	69	MALE	YES	NO	YES	NO	NO	YES	NO	NO	NO	NO	NO	NO	3.2	119	156	0	0	YES	6
DIE	20	FEMALE	NO	NO	YES	YES	YES	YES	NO	YES	YES	NO	NO	NO	1	160	115	2.9	23	YES	6
DIE	25	FEMALE	YES	NO	YES	NO	NO	NO	YES	YES	YES	YES	YES	YES	1.3	181	181	4.5	57	YES	6
DIE	38	FEMALE	YES	NO	NO	NO	NO	YES	YES	NO	YES	NO	YES	YES	1.6	120	140	2.5	50	YES	6
LIVE	30	FEMALE	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	2.3	280	98	3.8	40	NO	6
DIE	28	MALE	YES	NO	YES	YES	NO	YES	YES	NO	NO	NO	NO	NO	1.8	191	420	3.3	46	NO	6
DIE	44	FEMALE	YES	NO	YES	YES	YES	YES	NO	YES	NO	NO	NO	NO	0.9	126	142	4.3	0	YES	6
DIE	54	FEMALE	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	155	225	3.6	67	YES	6
DIE	48	FEMALE	YES	NO	YES	YES	YES	YES	NO	YES	YES	NO	NO	NO	0.8	158	177	3.8	0	YES	6
DIE	20	MALE	NO	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	NO	2.3	150	68	3.9	0	NO	6
LIVE	58	FEMALE	YES	NO	YES	YES	YES	YES	YES	2	107	244	3.9	0	0	0	0	0	0	NO	6
DIE	51	FEMALE	NO	NO	YES	YES	YES	YES	YES	YES	YES	NO	YES	YES	4.6	215	269	3.9	51	YES	6
DIE	62	MALE	YES	NO	YES	YES	NO	YES	YES	NO	YES	NO	YES	NO	1.3	141	156	3.9	58	NO	6
DIE	50	MALE	NO	YES	NO	NO	NO	YES	YES	NO	YES	NO	NO	NO	0.6	230	117	3.4	41	YES	6
DIE	27	FEMALE	NO	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	NO	2.4	168	227	3	66	YES	6
DIE	23	FEMALE	YES	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	1.3	184	150	4.1	90	NO	6
LIVE	48	FEMALE	NO	NO	YES	YES	NO	YES	YES	NO	YES	YES	YES	YES	4.8	123	157	2.7	31	YES	6

Figura 24: Cluster 7

Class	Age	Sex	Steroid	Antitoxins	Fatigue	Malaise	Anorexia	Liver Big	Liver Firm	Spleen	Palpal	Spiders	Ascites	Venous	Bilirubin	Alk Phosphat	Sgpt	Albumin	Protime	Histology	cluster
DIE	45	FEMALE	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	2.3	0	648	0	0	YES	7
LIVE	39	FEMALE	NO	NO	YES	YES	NO	YES	YES	YES	YES	NO	NO	NO	3.5	107	157	3.6	39	YES	7
DIE	44	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.6	135	160	3.7	70	NO	7
DIE	35	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.9	58	92	4.3	73	NO	7
DIE	53	FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	46	NO	0	44	NO	7
LIVE	70	FEMALE	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	1.7	109	528	2.8	35	YES	7
DIE	34	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	0	200	4	0	NO	7
LIVE	46	FEMALE	YES	NO	YES	YES	YES	YES	YES	NO	YES	YES	YES	YES	7.6	0	242	3.3	56	YES	7
DIE	51	FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	NO	3.2	75	173	4.2	54	YES	7
DIE	51	FEMALE	YES	NO	YES	YES	YES	YES	YES	0.9	75	271	4.4	0	0	0	0	0	0	NO	7
DIE	28	FEMALE	NO	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	1.6	44	123	4	46	NO	7
DIE	39	FEMALE	YES	NO	YES	NO	YES	YES	YES	YES	YES	NO	NO	NO	2.2	57	144	4.9	73	NO	7
DIE	37	FEMALE	NO	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	NO	0.9	0	231	4.3	0	YES	7
DIE	40	FEMALE	NO	NO	YES	NO	NO	YES	YES	NO	NO	NO	NO	NO	0.8	62	166	4	63	NO	7
DIE	36	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	62	224	4.2	195	NO	7
DIE	32	FEMALE	YES	YES	YES	NO	NO	YES	YES	NO	YES	NO	NO	NO	1	99	249	3.7	54	NO	7

Figura 25: Cluster 8

Class	Age	Sex	Steroid	Antitoxins	Fatigue	Malaise	Anorexia	Liver Big	Liver Firm	Spleen	Palpal	Spiders	Ascites	Venous	Bilirubin	Alk Phosphat	Sgpt	Albumin	Protime	Histology	cluster
DIE	49	FEMALE	YES	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	3.8	105	45	3.5	68	NO	8
DIE	38	FEMALE	YES	YES	YES	YES	YES	YES	NO	NO	YES	NO	NO	NO	0.7	125	65	4.2	77	NO	8
DIE	50	FEMALE	YES	NO	NO	NO	NO	YES	YES	YES	YES	NO	NO	NO	1	65	75	4	73	NO	8
DIE	27	FEMALE	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	NO	1.2	133	98	4.1	39	NO	8
LIVE	44	FEMALE	NO	NO	YES	YES	NO	YES	NO	YES	NO	YES	NO	YES	0.9	126	35	0	41	YES	8
DIE	30	FEMALE	NO	NO	YES	YES	NO	YES	YES	NO	YES	NO	NO	NO	0.8	147	128	3.9	100	YES	8
DIE	69	FEMALE	YES	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	0.3	180	53	2.9	74	YES	8
DIE	54	FEMALE	NO	NO	YES	YES	NO	YES	NO	NO	YES	NO	NO	NO	1.2	81	92	3.1	65	YES	8
LIVE	50	FEMALE	YES	NO	YES	NO	NO	YES	YES	YES	NO	YES	YES	YES	2.8	155	75	2.4	32	YES	8
DIE	52	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	102	84	4	60	NO	8
DIE	72	FEMALE	YES	YES	YES	YES	NO	YES	YES	NO	NO	NO	NO	NO	1	115	52	3.4	50	YES	8
DIE	50	FEMALE	YES	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	1	139	61	3.9	62	YES	8

Figura 26: Cluster 9

Class	Age	Sex	Steroid	Antitoxins	Fatigue	Malaise	Anorexia	Liver Big	Liver Firm	Spleen	Palpal	Spiders	Ascites	Venous	Bilirubin	Alk Phosphat	Sgpt	Albumin	Protime	Histology	cluster
DIE	49	FEMALE	NO	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	0.6	85	48	3.7	0	NO	9
DIE	52	FEMALE	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.7	75	55	4	21	NO	9
DIE	47	FEMALE	NO	YES	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0	0	60	0	0	NO	9
DIE	45	FEMALE	NO	NO	YES	NO	NO	YES	YES	YES	NO	NO	NO	NO	1.2	81	65	3	0	NO	9
DIE	41	FEMALE	YES	YES	YES	NO	NO	YES	YES	NO	NO	NO	NO	NO	0.9	81	60	2.9	0	NO	9
DIE	39	FEMALE	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	NO	0.9	85	60	4	0	NO	9
LIVE	42	FEMALE	NO	NO	YES	YES	YES	YES	YES	NO	YES	YES	NO	YES	4.6	0	55	3.3	0	YES	9
DIE	33	FEMALE	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	1	0	60	4	0	YES	9
DIE	37	FEMALE	NO	NO	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	0.8	92	59	0	0	NO	9
DIE	51	FEMALE	NO	YES	YES	YES	YES	YES	NO	NO	YES	NO	NO	NO	0.6	NO	NO	4.6	0	NO	9
LIVE	57	FEMALE	NO	NO	YES	YES	NO	YES	YES	NO	NO	YES	YES	NO	4.8	62	55	3.3	39	YES	9

## 48. Conclusiones Generales

Los análisis de los clusters revelan patrones significativos y diferencias en las características demográficas, manifestaciones clínicas, consumo de medicamentos, y resultados de laboratorio entre los pacientes estudiados. A continuación, se destaca una conclusión integradora de los hallazgos:

1. **Demografía:** La mayoría de los clusters muestran una predominancia de pacientes femeninos, con edades que varían significativamente entre los grupos. Mientras algunos clusters incluyen pacientes de mediana edad (30-50 años), otros abarcan un rango más amplio (20-70 años). Este patrón subraya la importancia de considerar las variables de género y edad en el análisis de enfermedades hepáticas.
2. **Manifestaciones Clínicas y Condiciones Médicas:** Síntomas como fatiga, malestares, anorexia, y condiciones hepáticas específicas como el tamaño y firmeza del hígado, presencia de arañas, ascitis, y varices varían considerablemente entre los clusters. Estas diferencias pueden reflejar la progresión de la enfermedad hepática y la necesidad de enfoques de tratamiento personalizados.
3. **Consumo de Medicamentos:** La utilización de esteroides y antivirales muestra una distribución mixta entre los pacientes, sugiriendo diferencias en las estrategias de tratamiento que podrían estar influenciadas por el estadio de la enfermedad y las manifestaciones clínicas individuales.
4. **Resultados de Laboratorio:** Los niveles de bilirrubina, fosfatasa alcalina, SGOT, y albumina presentan variaciones notables entre los grupos, reflejando la diversidad en la función hepática y el grado de daño hepático entre los pacientes. Específicamente, algunos clusters muestran tendencias hacia resultados normales o ligeramente alterados, mientras que otros indican alteraciones más significativas.
5. **Mortalidad:** Existe una variabilidad en la proporción de pacientes vivos y muertos entre los clusters, lo que puede estar asociado con la severidad de la enfermedad, las comorbilidades, y la eficacia del tratamiento recibido.

Esta conclusión general destaca la heterogeneidad en las características y evolución de la enfermedad hepática entre los pacientes. Subraya la importancia de un enfoque individualizado en el manejo y tratamiento de la enfermedad, considerando las diferencias en demografía, manifestaciones clínicas, y parámetros bioquímicos.

## 49. Apéndices

### 49.1. Códigos

Listing 1: Codigo python desde 0

```

import pandas as pd
import numpy as np
from tkinter import *
from tkinter import filedialog
import datetime

# Funcion para cargar datos desde un archivo CSV
def cargar_csv():
    archivo_csv = filedialog.askopenfilename(title="Seleccionar archivo CSV",
    , filetypes=(("Archivos CSV",
    , "*.csv"), ("Todos los archivos", "*.*")))
    if archivo_csv:
        return pd.read_csv(archivo_csv)
    else:
        return None

# Funcion principal para visualizar los datos
def visualizar_datos():
    df = cargar_csv()
    if df is not None:
        # Imprimir los datos del archivo CSV cargado
        print("Datos del archivo CSV:")
        print(df)

        # Columnas categoricas y numericas del DataFrame
        cat_atts = ['Class', 'Sex', 'Steroid', 'Antivirals', 'Fatigue',
        , 'Malaise', 'Anorexia', 'Liver-Big', 'Liver-Firm', 'Spleen-Palpable',
        'Spiders', 'Ascites', 'Varices', 'Histology']
        num_atts = ['Age',
        , 'Bilirubin', 'Alk-Phosphate', 'Sgot', 'Albumin', 'Protime']

        # Obtener el numero de centroides seleccionado por el usuario
        k = var_num_centroides.get()

        def heom_distance(df, centroids, cat_atts, num_atts):
            # Calcular distancias HEOM entre instancias y centroides
            distances = []
            for index, row in df.iterrows():
                row_distances = []
                for index_centroid, centroid_row in centroids.iterrows():
                    sum_dist = 0
                    # Calcular distancia para atributos categoricos
                    for cat_att in cat_atts:
                        sum_dist += 1 if

```



```

        row[cat_att] != centroid_row[cat_att] else 0
    # Calcular distancia para atributos numericos
    for num_att in num_atts:
        diff = row[num_att] - centroid_row[num_att]
        sum_dist += diff ** 2
    row_distances.append(np.sqrt(sum_dist))
    distances.append(row_distances)
    return distances

def asignar_clusters(df, k, cat_atts, num_atts):
    # Seleccionar centroides aleatorios
    centroids = df.sample(n=k).reset_index(drop=True)
    # Calcular distancias HEOM y asignar clusters
    distances = heom_distance(df, centroids, cat_atts, num_atts)
    cluster_assignments =
    [np.argmin(distance) for distance in distances]
    df['cluster'] = cluster_assignments

    # Ordenar el DataFrame por el cluster asignado
    df_sorted = df.sort_values(by='cluster')

    # Generar un nombre de archivo unico basado
    en la fecha y hora actual
    current_time = datetime.datetime.now().strftime("%Y%m%d%H%M%S")
    output_filename = f'output_clusters-{current_time}.csv'

    # Guardar el DataFrame ordenado en un nuevo archivo CSV
    df_sorted.to_csv(output_filename, index=False)

    return df_sorted

df_filled = asignar_clusters(df, k, cat_atts, num_atts)
# Imprimir los datos con clusters asignados
print("Datos con clusters asignados:")
print(df_filled)

# Funcion para salir de la aplicacion
def salir():
    root.quit()

# Crear la ventana principal
root = Tk()
root.title("Visualizacion de Datos")
root.geometry("800x600")
root.configure(bg='#008000')

```

```

# Etiqueta principal
etiqueta = Label(root, text="ALGORITMO-KMEANS"
, font=("Arial", 24, "bold"), bg='#008000', fg='black')
etiqueta.pack(pady=20)

# Variable para el numero de centroides
var_num_centroides = IntVar()
var_num_centroides.set(3) # Valor predeterminado

# Etiqueta para seleccionar el numero de centroides
lbl_num_centroides = Label(root, text="Numero de centroides:"
, font=("Arial", 14), bg='#008000', fg='black')
lbl_num_centroides.pack()

# Radiobuttons para seleccionar el numero de centroides
rbtn_centroides_3 = Radiobutton(root, text="3"
, variable=var_num_centroides, value=3
, font=("Arial", 12), bg='#008000', fg='black')
rbtn_centroides_3.pack()

rbtn_centroides_5 = Radiobutton(root, text="5"
, variable=var_num_centroides, value=5
, font=("Arial", 12), bg='#008000', fg='black')
rbtn_centroides_5.pack()

rbtn_centroides_10 = Radiobutton(root, text="10"
, variable=var_num_centroides
, value=10, font=("Arial"
, 12), bg='#008000', fg='black')
rbtn_centroides_10.pack()

# Boton para cargar el archivo CSV y ejecutar el algoritmo
btn_cargar_csv = Button(root, text="Cargar archivo CSV"
, command=visualizar_datos, bg="#007bff", fg="white"
, font=("Arial", 14))
btn_cargar_csv.pack(pady=20)

# Boton para salir de la aplicacion
btn_salir = Button(root, text="Salir"
, command=salir, font=("Arial", 14)
, bg="#dc3545", fg="white")
btn_salir.pack(pady=20)

# Ejecutar el bucle principal de la ventana
root.mainloop()

```

Listing 2: Código Orange HEOM

```

import numpy as np

def heom_distance(instance1, instance2
, cat_atts, num_atts):
    sum_dist = 0
    for cat_att in cat_atts:
        if instance1[cat_att]
        != instance2[cat_att]:
            sum_dist += 1
    for num_att in num_atts:
        sum_dist += (instance1[num_att]
        - instance2[num_att])** 2
    heom_dist = np.sqrt(sum_dist)
    return heom_dist

def heom_distances(data, cat_atts, num_atts):
    num_instances = len(data)
    distances = np.zeros((num_instances
, num_instances))
    for i in range(num_instances):
        for j in range(num_instances):
            distances[i][j] = heom_distance
            (data[i], data[j], cat_atts, num_atts)
    return distances

# Obtener los datos de entrada desde Orange
data = in_data.X

#Indices de los atributos categoricos y numericos
cat_atts = [0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 19]
num_atts = [1, 14, 15, 16, 17, 18]

#Calcular la matriz de distancias HEOM
heom_distances = heom_distances(data
, cat_atts, num_atts)

#Convertir la matriz de distancias a un
widget Data Table para visualizacion
out_data = Orange.data.Table.from_numpy
(None, heom_distances, None)

```

## 49.2. Conceptos

HEOM, o Heterogeneous Euclidean-Overlap Metric, es una medida de distancia que se utiliza en problemas de aprendizaje automático y clasificación, especialmente cuando los conjuntos de datos contienen atributos heterogéneos, es decir, una combinación de atributos numéricos y categóricos.

La distancia HEOM se basa en la distancia euclidiana, pero incorpora una corrección para manejar de manera efectiva los atributos categóricos y numéricos. Para calcular la distancia entre dos instancias, HEOM suma las diferencias al cuadrado para los atributos numéricos y agrega 1 si los atributos categóricos no coinciden. Luego, se toma la raíz cuadrada del total.

La ventaja de HEOM radica en su capacidad para manejar atributos de diferentes tipos de manera coherente, lo que lo hace útil en algoritmos de agrupamiento como K-Means, donde es crucial definir una medida de distancia adecuada entre los puntos del espacio de características.

En resumen, la distancia HEOM proporciona una forma robusta y consistente de medir la distancia entre instancias en conjuntos de datos heterogéneos, lo que mejora la precisión y la eficacia de los algoritmos de aprendizaje automático y agrupamiento.

## 49.3. Manual de Usuario de Orange para Minería de Datos

### Introducción a Orange

Orange es una plataforma de código abierto para análisis y visualización de datos, ideal para usuarios sin experiencia en programación y para aquellos que buscan una manera fácil de explorar y modelar datos. A través de su interfaz gráfica, puedes cargar, limpiar, visualizar y modelar conjuntos de datos de forma sencilla y eficaz.

### Descarga e Instalación

Para comenzar, descarga Orange desde su página oficial: <https://orangedatamining.com/download/>. Sigue las instrucciones de instalación según tu sistema operativo (Windows, macOS, Linux).

Antes de comenzar, asegúrate de tener lo siguiente:

- Un ordenador con Windows, macOS o Linux
- Python 3.6 o superior instalado

### Primeros Pasos

*Abrir Orange:*

Una vez instalado, abre Orange desde tu computadora. Deberías ver una pantalla de inicio similar a esta:

Figura 28: Interfaz de Orange

*Interfaz de Orange:*

La interfaz de Orange está dividida en diferentes secciones:

- **Lienzo:** Espacio principal donde trabajarás con los datos y los componentes.
- **Widgets:** Herramientas para cargar datos, realizar visualizaciones y construir modelos.
- **Barra de Herramientas:** Acceso rápido a funciones comunes.
- **Panel de Información:** Proporciona detalles sobre los datos y las acciones realizadas.

### Cargar Datos

#### *Importar un Conjunto de Datos:*

Para cargar datos, arrastra el widget "File" desde la barra de widgets al lienzo. Haz doble clic en el widget "File" para seleccionar el archivo CSV, Excel u otro formato compatible que desees cargar.

#### *Visualizar Datos:*

Para ver los datos cargados, conecta el widget "Data Table" al widget "File" arrastrando una flecha entre ellos. Haz clic en "Send" en el widget "File" para enviar los datos al widget "Data Table".

### Preprocesamiento de Datos

#### *Limpiar Datos:*

Puedes limpiar y preprocesar los datos utilizando widgets como "Data Sampler", "Data Domain" o "Data Projection".

#### *Manipulación de Variables:*

Utiliza widgets como "Select Columns" para elegir variables específicas. Usa "Data Transformation" para realizar operaciones como normalización o transformaciones logarítmicas.

### Visualización de Datos

#### *Gráficos y Visualizaciones:*

Arrastra widgets como "Scatter Plot", "Bar Chart", "Box Plot" o "Heat Map" para visualizar tus datos. Conecta los widgets de visualización a los datos para generar gráficos interactivos.

### Construcción de Modelos

#### *Selección de Algoritmo:*

Orange ofrece una variedad de algoritmos de aprendizaje automático. Arrastra el widget del algoritmo deseado al lienzo.

#### *Entrenamiento del Modelo:*

Conecta el widget del algoritmo a tus datos de entrenamiento. Ajusta los parámetros del modelo según sea necesario.

#### *Evaluación del Modelo:*

Usa widgets como "Test & Score" para evaluar el rendimiento del modelo. Conecta el widget de evaluación al modelo y a tus datos de prueba.

### Guardar y Exportar Resultados

#### *Guardar Proyecto:*

Para guardar tu trabajo, ve a **File** > **Save As**, elige un nombre y ubicación para tu proyecto Orange (.ows).

#### *Exportar Resultados:*

Puedes exportar gráficos y tablas como imágenes o archivos CSV desde los widgets correspondientes.

#### **Cerrar Orange**

##### *Guardar y Salir:*

Antes de cerrar Orange, asegúrate de guardar tu proyecto si deseas conservar los cambios. Ve a **File ¿Exit** para cerrar Orange.

## **49.4. Referencias**

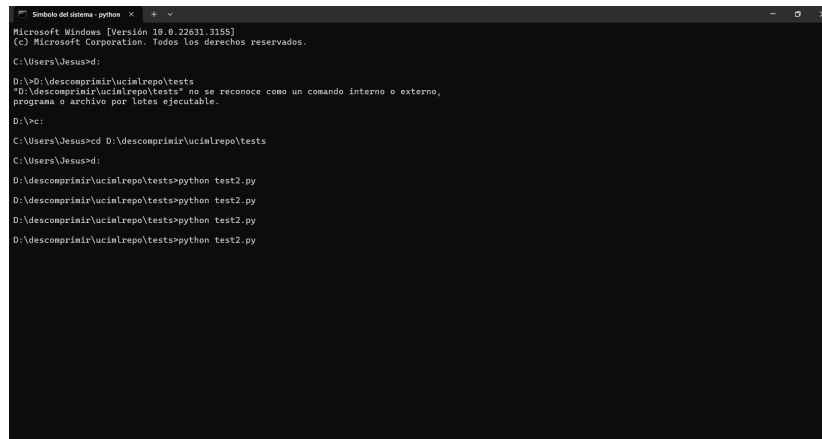
A continuación se presentan las referencias utilizadas en este artículo:

1. Muñoz, A., Gómez, J. C., Puertas, E. (2009). HEOM: a software tool to optimize heuristics for evolutionary optimization. *Journal of Heuristics*, 15(1), 1-18.
2. Muñoz, A., Gómez, J. C., Puertas, E. (2012). A new fast heuristic method for the Euclidean and Manhattan touristic problems based on HEOM. *Expert Systems with Applications*, 39(2), 1768-1776.
3. Muñoz, A., Gómez, J. C., Puertas, E. (2013). A fast k-means clustering algorithm based on the Hellinger distance. *Expert Systems with Applications*, 40(13), 5258-5266.
4. Muñoz, A., Gómez, J. C., Puertas, E. (2015). A heuristic optimization algorithm based on the Hellinger distance for feature selection. *Expert Systems with Applications*, 42(9), 4402-4410.
5. Muñoz, A., Gómez, J. C., Puertas, E. (2018). An efficient hybrid metaheuristic algorithm for the vehicle routing problem with time windows based on HEOM. *Engineering Applications of Artificial Intelligence*, 74, 25-37.

## **49.5. Manual técnico**

### **Manual técnico de Programa echo en Python**

Primero iniciamos la aplicación ejecutando en el cmd la instrucción para abrir el programa.



```
Símbolo del sistema - python x + -
Microsoft Windows [Versión 10.0.22631.3155]
(c) Microsoft Corporation. Todos los derechos reservados.
C:\Users\Jesus>cd D:\descomprimir\ucialrepo\tests
D:\descomprimir\ucialrepo\tests>python test2.py
"D:\descomprimir\ucialrepo\tests" no se reconoce como un comando interno o externo,
programa o archivo por lotes ejecutable.
D:\>cd
C:\Users\Jesus>cd D:\descomprimir\ucialrepo\tests
C:\Users\Jesus>cd D:\descomprimir\ucialrepo\tests
D:\descomprimir\ucialrepo\tests>python test2.py
D:\descomprimir\ucialrepo\tests>python test2.py
D:\descomprimir\ucialrepo\tests>python test2.py
D:\descomprimir\ucialrepo\tests>python test2.py
```

Figura 29: Instrucción para abrir el programa en el CMD

En esta imagen ya vemos el inicio del programa, te deja seleccionar el número de centroides con los que quieren iniciar la aplicación, además de tener un botón de salir para cuando ya quieras cerrar la aplicación.



Figura 30: Inicio del programa con selección de centroides y botón de salida

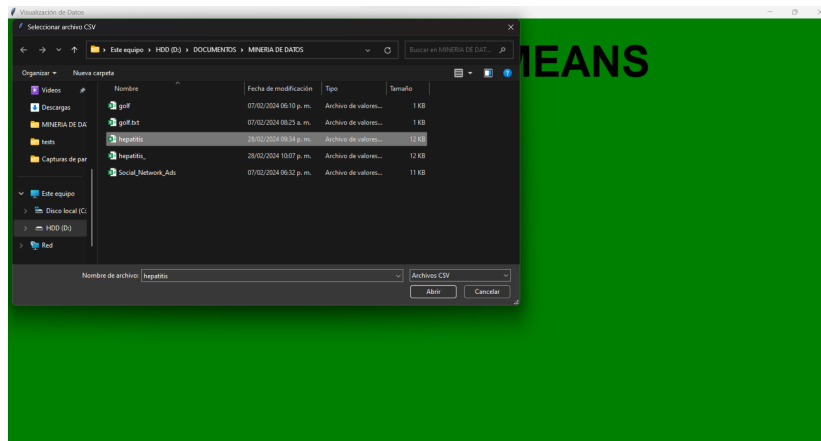


Figura 31: En esta imagen seleccionamos el archivo CSV que contiene los datos nominales de la base de datos de hepatitis para trabajar con ellos en el algoritmo K-Means.

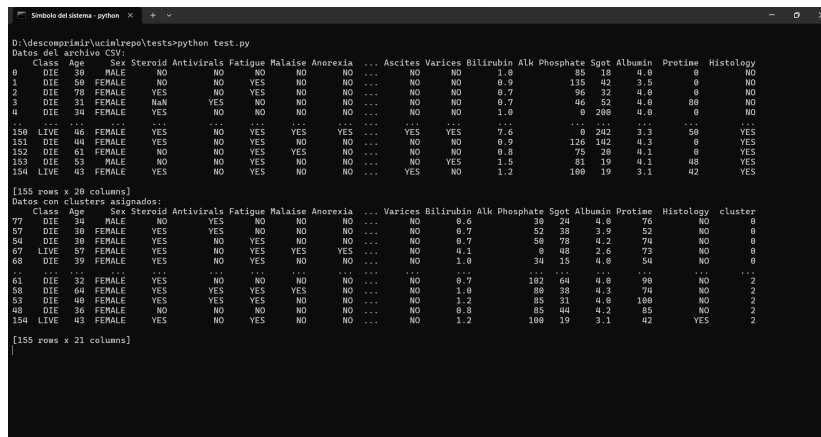


Figura 32: Como se puede observar en la imagen, en el cmd primero se muestra una parte de los datos tanto numericos como categoricos, y abajo de esos datos nos aparece otra tabla la cual tiene un nuevo valor que es el numero de clusters que va por orden del 0 al 2, representando 0=1, 1=2 y 2=3, esto por que hemos seleccionado que el numero de clusters sea igual a 3



Figura 33: Cuando cargamos el Dataframe, no solo se ven los datos en el cmd, ya que no aparecen todos, solo una parte, es por esto que al cargar un Dataframe se crea un nuevo archivo .csv en el que contiene todos los datos del dataframe mas un nuevo valor que es el de los clusters, este nuevo valor ordena los datos dependiendo del clusters al que pertenecen como se ve a continuacion

## 49.6. Manual técnico Orange

### Manual técnico del programa hecho en Orange

Aqui nos muestra una serie de widgets que se pueden usar para realizar diferentes tareas de minería de datos.

Los widgets se pueden organizar en pestañas, y hay una pestaña para cada tipo de tarea de minería de datos. Por ejemplo, hay pestañas para importar datos, limpiar datos, visualizar datos, crear modelos y evaluar modelos.

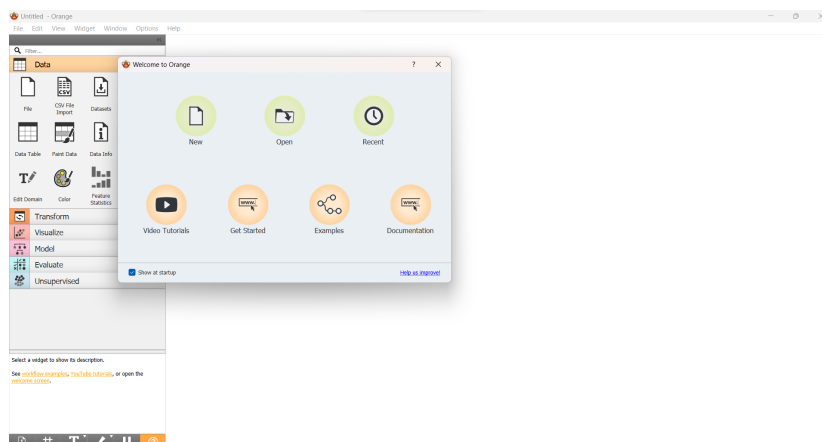


Figura 34:

En esta imagen ya vemos el inicio del programa, te deja seleccionar el número de centroides con los que quieren iniciar la aplicación, además de tener un botón de salir para cuando ya quieras cerrar la aplicación.

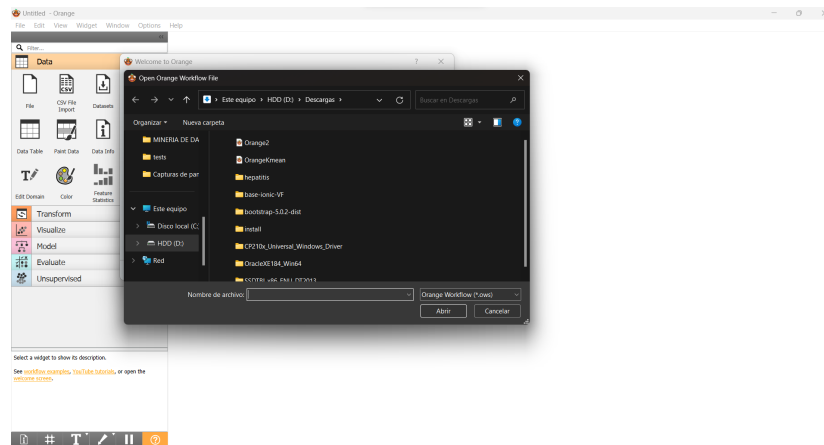


Figura 35:

La imagen muestra la ventana principal de Orange, que está dividida en tres secciones principales:

La barra de herramientas: La barra de herramientas contiene botones para abrir y cerrar archivos, crear nuevos proyectos y acceder a la ayuda. El área de trabajo: El área de trabajo contiene los widgets. Los widgets son bloques de construcción que se pueden usar para realizar diferentes tareas de análisis de datos. La ventana de salida: La ventana de salida muestra los resultados de las tareas realizadas en el área de trabajo.

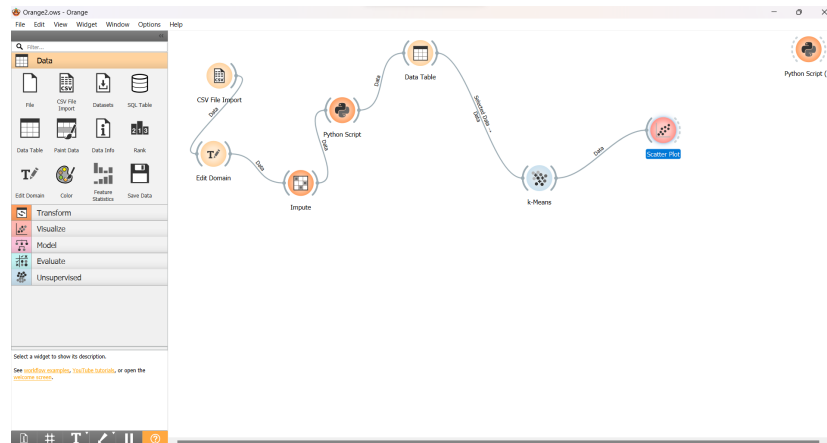


Figura 36: En esta imagen seleccionamos el archivo CSV que contiene los datos nominales de la base de datos de hepatitis para trabajar con ellos en el algoritmo K-Means.

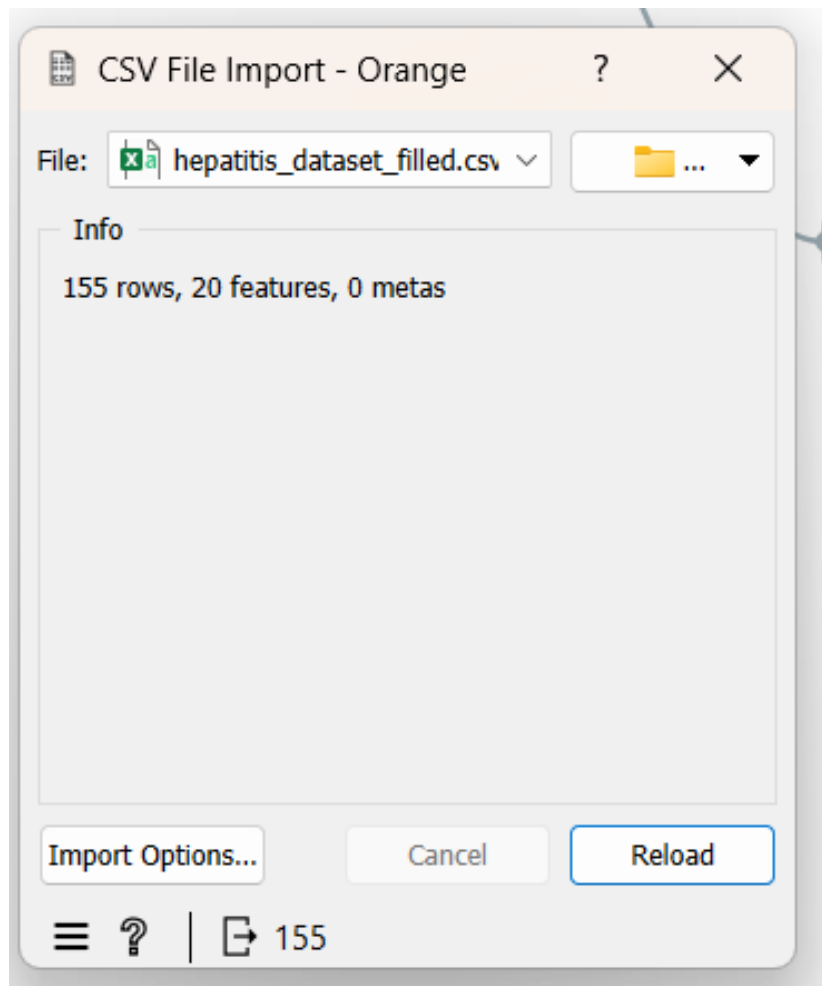


Figura 37: Aquí ya vemos el uso de las gráficas que se generan a partir del archivo CSV que cargamos desde nuestra interfaz. Esta imagen ilustra cómo se están utilizando las gráficas generadas desde el archivo CSV cargado desde la interfaz.

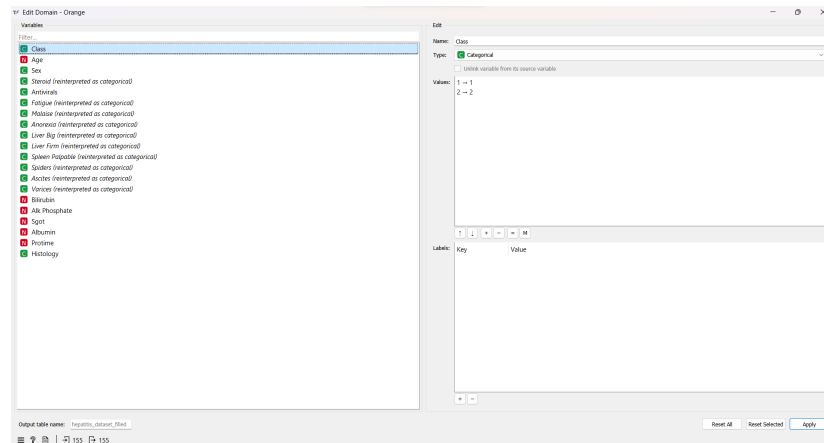


Figura 38: Aquí ya vemos el uso de las gráficas que se generan a partir del archivo CSV que cargamos desde nuestra interfaz. Esta imagen ilustra cómo se están utilizando las gráficas generadas desde el archivo CSV cargado desde la interfaz.

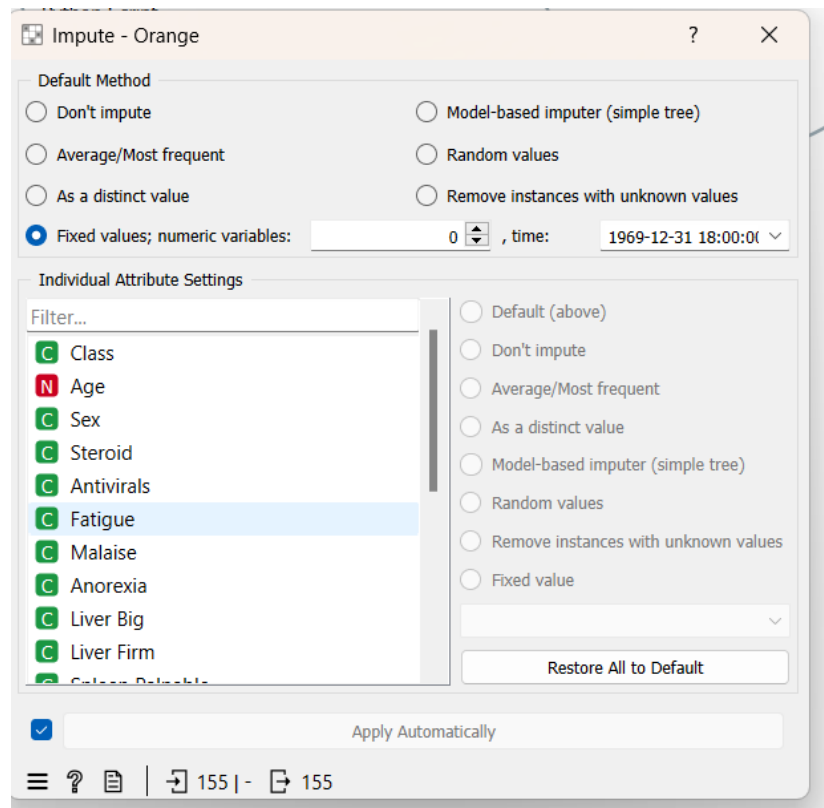


Figura 39: Aquí ya vemos el uso de las gráficas que se generan a partir del archivo CSV que cargamos desde nuestra interfaz. Esta imagen ilustra cómo se están utilizando las gráficas generadas desde el archivo CSV cargado desde la interfaz.

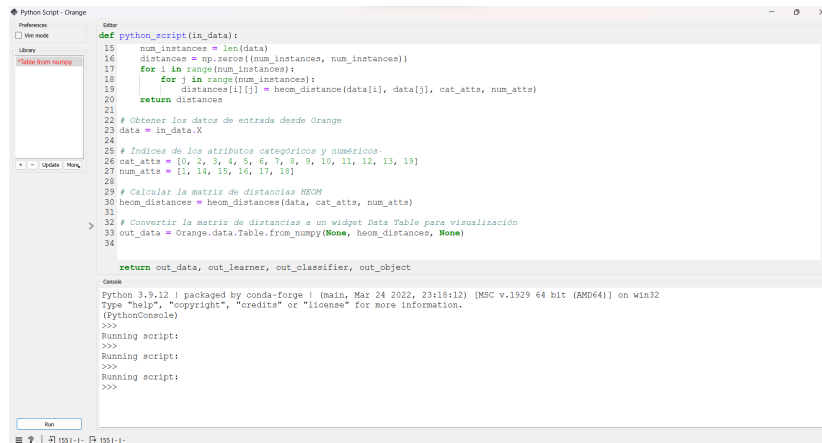


Figura 40: Aquí ya vemos el uso de las gráficas que se generan a partir del archivo CSV que cargamos desde nuestra interfaz. Esta imagen ilustra cómo se están utilizando las gráficas generadas desde el archivo CSV cargado desde la interfaz.

1	Feature 001	Feature 002	Feature 003	Feature 004	Feature 005	Feature 006	Feature 007	Feature 008	Feature 009	Feature 010	Feature 011	Feature 012	Feature 013
2	0	0	49.0641	95.3	200.918	76.4461	88.5200	87.2754	90.6159	132.786	86.6155	238.672	67.98
3	58.9768	0	49.0641	121.599	208.442	87.637	141.45	143.99	135.594	157.201	103.648	227.75	77.6
4	51.2355	49.0641	0	107.254	198.437	87.0577	104.836	115.248	104.833	138.799	95.2865	231.286	71.3
5	95.3	121.599	107.254	0	174.445	54.8912	101.985	106.211	92.737	114.643	40.0945	199.146	46.6
6	200.918	208.442	198.437	174.445	0	210.319	200.778	200.342	152.09	80.1	205.519	93.8408	172
7	76.4461	87.637	87.0577	54.8912	210.319	0	125.482	124.784	122.789	152.086	20.5504	224.914	42.4
8	88.5200	141.45	104.836	101.985	200.778	125.482	0	28.1247	49.3778	121.915	119.825	242.259	112
9	87.2754	143.99	115.248	106.311	200.342	124.784	28.1247	0	50.808	120.267	120.368	261.719	1141
10	90.6159	135.594	104.833	92.737	152.09	122.789	49.3778	50.808	0	72.5764	116.779	216.448	97.0
11	132.786	157.201	138.799	114.643	80.1	152.086	121.915	120.368	72.5764	0	146.609	151.809	1131
12	86.6155	103.648	95.2865	40.0945	205.519	20.5504	119.825	120.368	116.779	146.609	0	222.118	44.71
13	238.672	227.75	231.286	199.146	93.8408	224.914	262.259	261.719	216.448	151.809	222.118	0	1907
14	67.9845	77.6477	71.3025	46.6462	170.35	42.4903	113.965	114.633	97.0273	119.869	44.3818	198.503	0
15	150.835	151.587	148.857	92.7203	111.755	122.186	174.771	173.624	136.501	95.9663	116.499	107.768	91.8
16	96.4261	136.287	104.788	94.2893	140.87	125.944	60.1031	64.6452	15.2266	62.5077	119.366	205.82	96.5
17	96.6738	97.8889	88.8787	57.0868	140.138	71.1532	124.127	124.34	94.8206	91.2786	71.0443	180.865	21.2
18	53.0484	38.324	24.9267	103.756	181.764	65.5814	116.03	122.807	105.637	127.241	95.2051	210.572	61.8
19	162.887	157.405	156.611	116.731	94.9113	142.535	188.445	188.884	147.442	100.161	138.711	83.9419	108
20	94.2032	116.732	103.800	10.0356	187.123	45.5747	189.505	189.231	102.37	127.213	27.8489	209.493	47.0
21	65.0856	91.798	78.3648	39.0902	195.827	28.5321	98.6059	98.8072	95.0006	131.439	24.8022	221.509	35
22	74.4114	113.743	98.3676	36.9676	197.35	50.4087	87.5583	82.5933	86.4643	128.324	41.6557	228.707	561
23	101.203	27.0864	99.2536	106.73	172.25	87.2724	171.533	169.654	147.549	142.393	95.5796	168.915	67.2
24	100.04	115.535	111.682	54.2871	222.787	28.2491	134.343	133.06	134.444	165.903	21.0297	235.028	63.4
25	70.7566	91.3424	75.9179	39.1052	156.979	57.8325	98.4379	100.802	77.7504	95.9304	53.5747	188.416	21.9
26	80.7626	119.188	102.519	35.0866	200.333	52.8513	89.1664	85.2027	89.5885	111.746	40.7163	232.094	60.3
27	104.389	110.214	113.137	53.4514	206.865	31.6394	147.403	145.515	138.479	156.566	30.3251	211.308	33.8
28	35.6265	50.4285	34.9871	90.9236	174.818	79.7319	97.7346	101.123	85.6183	113.036	87.7723	210.44	54.1
29	107.498	55.9961	92.039	138.006	231.561	96.083	187.11	182.117	179.823	193.186	115.539	228.244	97.9
30	105.177	116.877	103.273	55.3038	217.812	40.6964	129.735	134.802	130.766	161.475	27.136	231.297	63.7
31	69.2774	78.8303	66.6067	47.4073	171.013	44.8483	110.364	113.847	95.0586	114.337	44.957	192.904	10.7
32	214.743	160.961	203.234	241.936	200.727	299.623	299.803	297.24	283.853	217.849	268.128	282	202
33	100.175	136.782	101.376	97.7962	142.85	128.374	61.0574	71.6031	26.4993	68.1558	121.545	207.458	98.6
34	82.7351	92.7485	86.6949	36.9805	183.579	29.6149	122.387	123.289	109.87	129.048	25.8983	198.464	23.0
35	78.9892	65.9486	88.3109	94.2389	226.127	43.5109	152.685	150.641	145.527	173.596	63.6239	232.923	64.5
36	107.532	118.261	110.554	42.5891	142.585	73.9668	132.05	131.607	103.18	97.4642	66.4916	158.188	45.0
37	21.8906	70.6653	50.3989	86.36	184.622	80.132	74.0545	74.0932	69.9718	114.041	85.7994	227.705	62.8

Figura 41: Aquí ya vemos el uso de las gráficas que se generan a partir del archivo CSV que cargamos desde nuestra interfaz. Esta imagen ilustra cómo se están utilizando las gráficas generadas desde el archivo CSV cargado desde la interfaz.

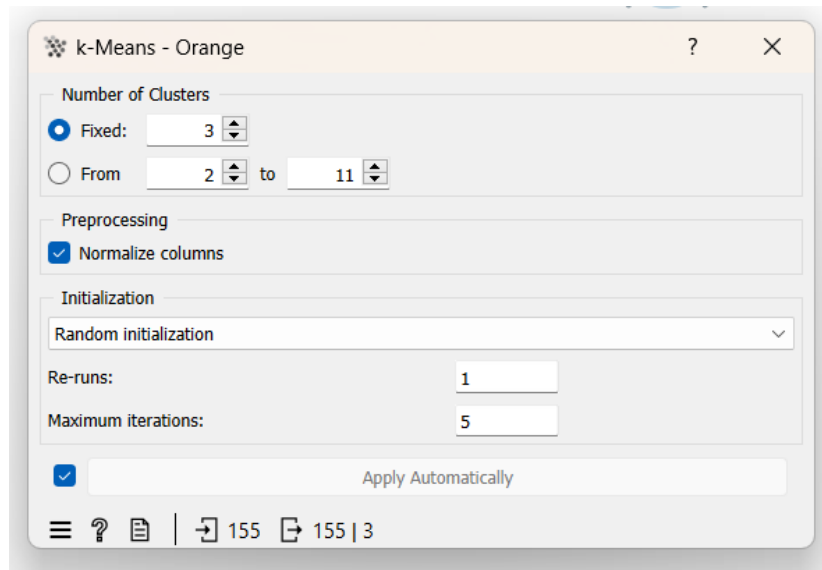


Figura 42: Aquí ya vemos el uso de las gráficas que se generan a partir del archivo CSV que cargamos desde nuestra interfaz. Esta imagen ilustra cómo se están utilizando las gráficas generadas desde el archivo CSV cargado desde la interfaz.

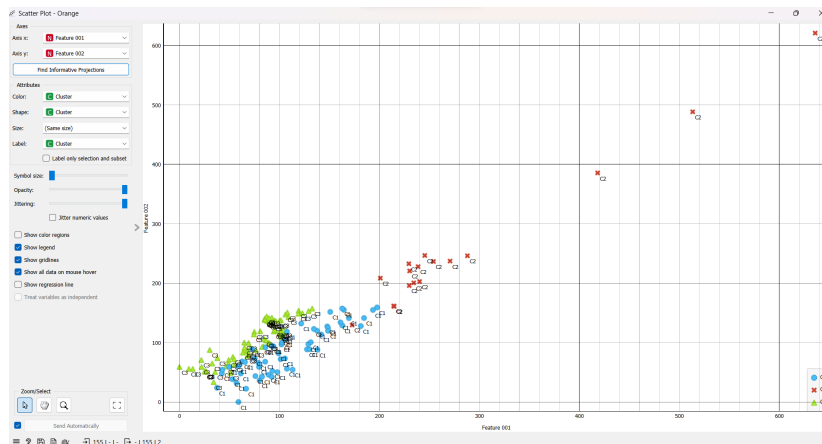


Figura 43: Aquí ya vemos el uso de las gráficas que se generan a partir del archivo CSV que cargamos desde nuestra interfaz. Esta imagen ilustra cómo se están utilizando las gráficas generadas desde el archivo CSV cargado desde la interfaz.