



UNIVERSIDAD TECNOLÓGICA DE PANAMÁ
FACULTAD DE INGENIERÍA DE SISTEMAS
COMPUTACIONALES



DEPARTAMENTO DE INGENIERIA DE SOFTWARE
CARRERA LICENCIATURA EN INGENIERÍA DE SOFTWARE

GESTIÓN DE LA INFORMACIÓN

PROYECTO FINAL

PROFESORA: CARMEN ORTEGA

GRUPO 1SF-142

INTEGRANTES:

IAN BECKFORD	8-967-1300
REGINO CORNEJO	8-954-1692
ROBERTO CORTES	8-974-1253
SAMUEL NUÑEZ	8-854-1791

FECHA DE ENTREGA: 19 DE JULIO DE 2023

ÍNDICE DE CONTENIDO

INTRODUCCIÓN	2
DESCRIPCIÓN DE LOS DATASETS DE CASOS DE ESTUDIOS.....	3
DESARROLLOS DE LOS CASOS DE ESTUDIOS	4
CASO DE ESTUDIO 1: DRUGS	4
CASO DE ESTUDIO 2: TITANIC	15
CASO DE ESTUDIO 3: VEHICLES	23
CONCLUSIONES Y RECOMENDACIONES PARA EL CASO DE ESTUDIO 3	31
CONCLUSIONES	41
BIBLIOGRAFÍA	42

INTRODUCCIÓN

En este examen semestral, exploraremos el análisis de tres casos de estudio que ejemplifican la importancia y desafíos asociados con la gestión de la información en diferentes empresariales y organizacionales. Sabemos que la gestión de la información es una disciplina enorme y vital en el mundo actual, donde la demanda y complejidad de este, crece exponencialmente en cada sector.

Durante el semestre, hemos abordado una variedad de temas, desde la recopilación y almacenamiento de datos hasta la transformación de la información en conocimiento útil y accionable. Ahora en este examen, aplicaremos esos conocimientos a casos concretos, enfrentándonos a situaciones reales que requieren soluciones inteligentes y estratégicas para optimizar la gestión de la información.

En cada caso de estudio, nos adentraremos en la identificación de los desafíos específicos que enfrentó cada problema u organización, comprendiendo sus necesidades y objetivos. A través del análisis detallado de la información disponible, aplicaremos herramientas y técnicas de gestión de datos con modelos adecuados para proponer soluciones y/o encontrar el origen de los problemas a resolver.

En resumen, este examen semestral de Gestión de la Información nos brindará una oportunidad única para aplicar nuestros conocimientos en un entorno práctico, preparándonos para enfrentar los desafíos reales que la gestión y análisis de datos presentan en el mundo empresarial actual.

DESCRIPCIÓN DE LOS DATASETS DE CASOS DE ESTUDIOS

Caso de Estudio 1: Drugs

Las variables que se recogen en los historiales clínicos son:

Age: Edad.

Sex: Sexo.

BP: Tensión sanguínea.

Cholesterol: nivel de colesterol.

Na: Nivel de sodio en la sangre.

K: Nivel de potasio en la sangre.

Hay cinco fármacos posibles: DrugA, DrugB, DrugC, DrugX, DrugY

Se trata de predecir el tipo de fármaco (drug) que se debe administrar a un paciente afectado de rinitis alérgica según distintos parámetros/variables.

Caso de Estudio 2: Titanic

Este segundo caso corresponde a las características de los 2201 pasajeros del Titanic. Estos datos son reales, obtenidos de: "Report on the Loss of the 'Titanic' (S.S.)" (1990), British Board of Trade Inquiry Report_ (reprint), Gloucester, UK: Allan Sutton Publishing.

Los atributos de estudio son los siguientes:

Clase: 0=tripulación, 1=primera, 2=segunda, 3=tercera

Edad: 1=adulto, 0=niño

Sexo :1=hombre, 0=mujer

Sobrevivió: 1=si, 0=no

El dataset contiene el registro de las personas que sobrevivieron y las que no sobrevivieron al hundimiento del Titanic.

Caso de Estudio 3: Vehicles

Este grupo de datos proviene de una investigación llevada a cabo por los doctores Pete Mowforth y Barry Shepherd del TuringInstitute. Basados en data originada en un estudio previo, realizado en 1986-87 por JP Siebert.

El propósito de la investigación es realizar un estudio por medio de computación, sobre las siluetas pertenecientes a fotografías realizadas en varios ángulos y distancias, a modelos específicos de vehículos, tomando en cuenta tanto atributos conocidos del vehículo, como métricas obtenidas operando con valores deducibles de las siluetas.

Los parámetros fueron extraídos de las siluetas mediante HIPS (Hierarchical Image Processing System), que extrae una combinación de medidas independientes de la escala utilizando tanto las medidas basadas en los momentos clásicos como una varianza escalada y otros parámetros estadísticos sobre los ejes mayor y menor del vehículo, así como medidas heurísticas de oquedades, circularidad, vehículo compacto, etc.

Los cuatro modelos tomados como referencia son: un bus de dos plantas, VanChevrolet, Saab 9000 y Opel Manta 400. Esta elección se dio porque las siluetas del bus la van y cualquiera de los coches son claramente diferentes. Las imágenes se tomaron con una cámara apuntando hacia abajo con un ángulo de elevación fijo ($34,2^\circ$ sobre la horizontal). Los vehículos se situaron sobre un fondo claro y se pintaron de negro mate para minimizar reflejos. Todas las imágenes tenían la misma resolución (128x128), se tomaron con la misma cámara, en escala de 64 tonos de gris. Las imágenes fueron filtradas para producir siluetas binarias de los vehículos, hechas negativo, limpiadas y almacenadas. Se realizaron más fotos rotando los vehículos, almacenándose al final un total de 2 juegos de 60 fotos por vehículo, cubriendo los 360° de rotación posibles. Después se realizaron dos tandas más de 2 juegos de 60 fotos, salvo la van que tuvo 2 juegos de 46, a dos elevaciones diferentes ($37,5^\circ$ y $30,8^\circ$).

DESARROLLOS DE LOS CASOS DE ESTUDIOS

CASO DE ESTUDIO 1: DRUGS

En primer lugar, cargamos los datos del fichero "drug1n.arff" que pertenece al primer hospital, asumiendo que contiene 200 registros, lo que nos permite realizar pruebas iniciales con mayor facilidad.

Para obtener una idea de los fármacos más comunes en general y su efectividad, analizamos el atributo "drug" para observar su distribución por clases. Se concluye que el fármaco más efectivo es el "Y", el cual ha tenido éxito en casi la mitad de los pacientes. Una regla simple podría ser administrar el fármaco "Y" en primer lugar, y si no funciona, probar con el fármaco "X" y así sucesivamente, siguiendo las frecuencias de éxito.

En Weka, existe un método llamado ZeroR4 en la familia de reglas, que permite generar este modelo simple asignando a todos los ejemplos la clase mayoritaria. Si lo aplicamos y evaluamos sobre todos los datos de entrenamiento, obtendremos un clasificador con una precisión del 45.5%, como era de esperar debido a su simplicidad.

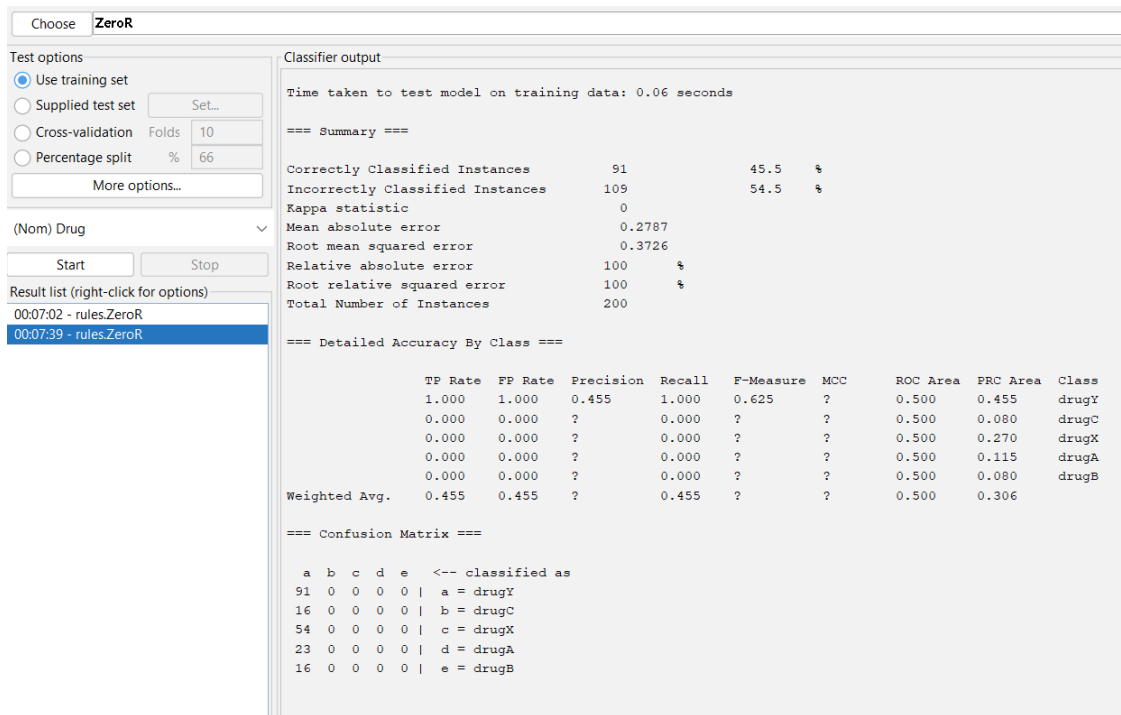


Figura 1. Clúster con filtro ZeroR

Por otro lado, podemos utilizar métodos más complejos, como el J48, que genera un árbol de decisión y logra una precisión del 97% al clasificar los datos.

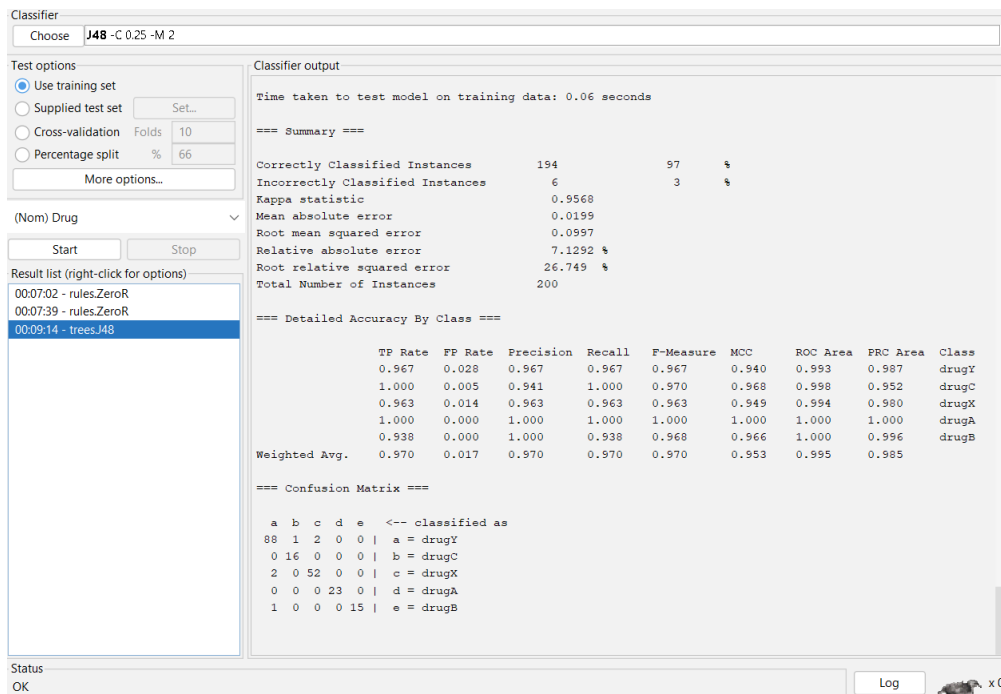


Figura 2. Clúster con filtro J48

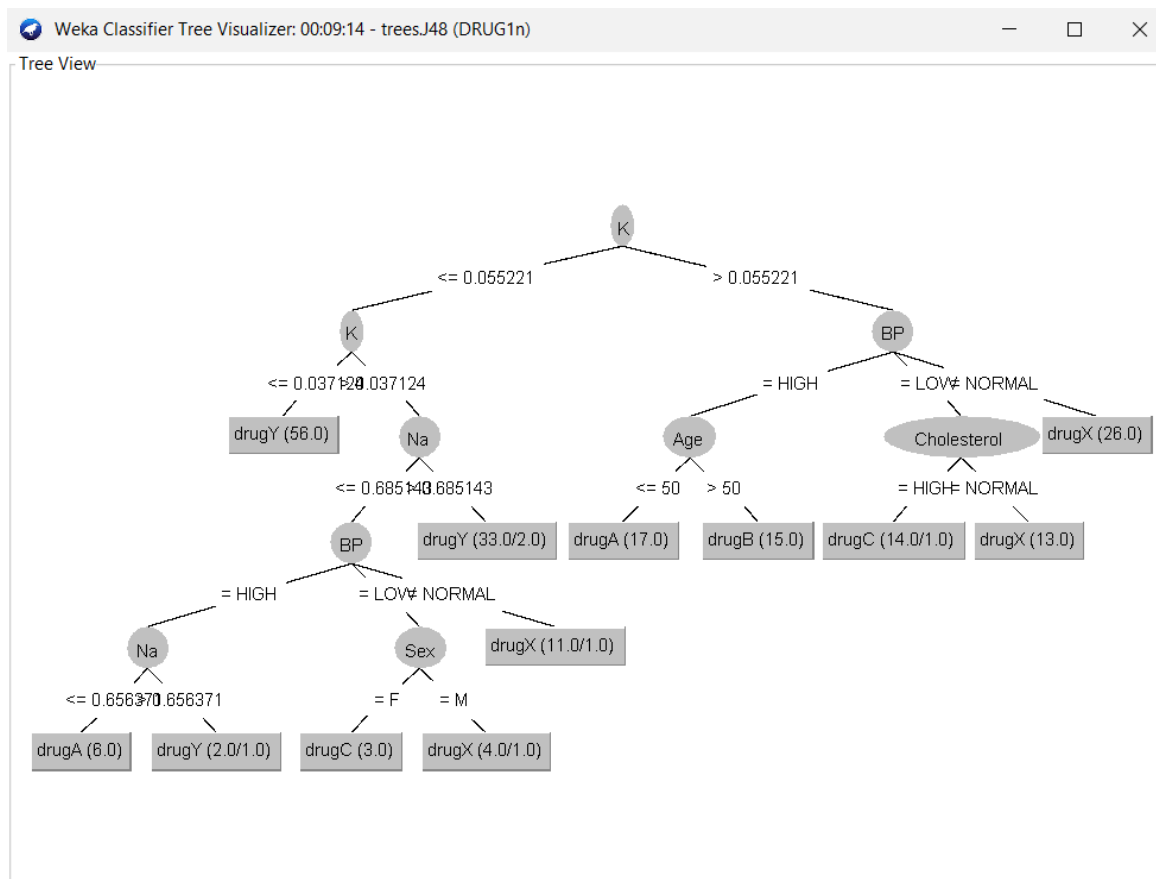


Figura 3. Árbol de decisión J48

Es importante destacar que al usar un método más sofisticado como J48, se obtiene una mejora significativa en la precisión del clasificador en comparación con el enfoque simple de ZeroR4. El árbol de decisión generado contiene un total de 12 reglas que determinan la clasificación de los datos. Al evaluar el árbol sobre los datos de entrenamiento, observamos que su precisión, es decir, el acierto en la clasificación es del 96.5%. Esto significa que el modelo tiene un error de solo el 3.5% al predecir correctamente la clase de los ejemplos en el conjunto de entrenamiento. Este rendimiento es significativamente mejor que simplemente guiarnos por la distribución de los fármacos. Si nos basamos en la distribución, cometeríamos un error del 54.5% al considerar el medicamento "DRUGY" como el adecuado para todos los pacientes, lo que claramente sería un enfoque muy ineficiente.

En resumen, el árbol de decisión demostró ser una herramienta mucho más efectiva para clasificar correctamente los datos en comparación con la aproximación basada en la distribución general de los fármacos. Su precisión del 96.5% indica que es un modelo altamente confiable para la tarea de clasificación.

Para comparar la relación entre atributos en Weka, podemos acceder al entorno "Visualize". En esta sección, tenemos la posibilidad de crear gráficas entre pares de atributos y observar si existe alguna relación entre ellos y las clases a las que pertenecen. Entre todas las combinaciones posibles, se destaca la que utiliza los niveles de sodio (Na) y potasio (K) como atributos.

Una vez seleccionada esta combinación de atributos, podemos mejorar la visualización del gráfico al cambiar los colores asignados a las clases, lo que facilita la interpretación de los datos. Además, podemos ampliar y observar con mayor detalle una zona específica del gráfico utilizando la opción "select instance" y luego haciendo clic en "submit".

En resumen, el entorno "Visualize" en Weka nos permite realizar un análisis comparativo de atributos al crear gráficas entre pares de ellos y observar su relación con las clases. La opción de cambiar colores y utilizar el "select instance" nos ayuda a visualizar de manera más clara y detallada los patrones de los datos.

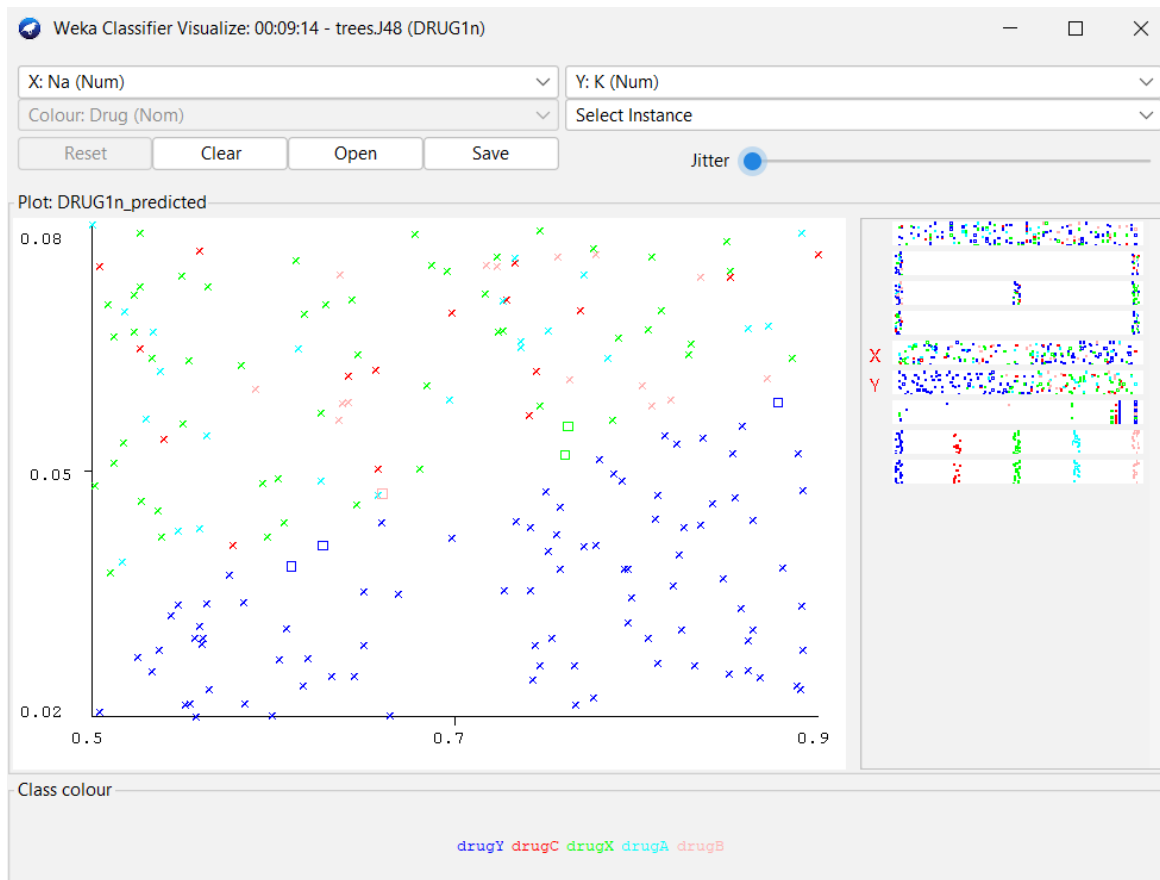


Figura 4. Visualización de datos referente a K y NA y su eficacia

En el gráfico observado, se pueden identificar características muy significativas. Existe una clara separación lineal entre una relación alta de K/Na y una relación baja de K/Na. En particular, para concentraciones bajas de K/Na, el fármaco "Y" muestra una efectividad destacada, mientras que por encima de cierto cociente de K/Na, este medicamento deja de ser efectivo y se deben considerar los otros cuatro fármacos.

El conocimiento obtenido a partir de este gráfico puede ser utilizado para mejorar nuestros modelos de clasificación.

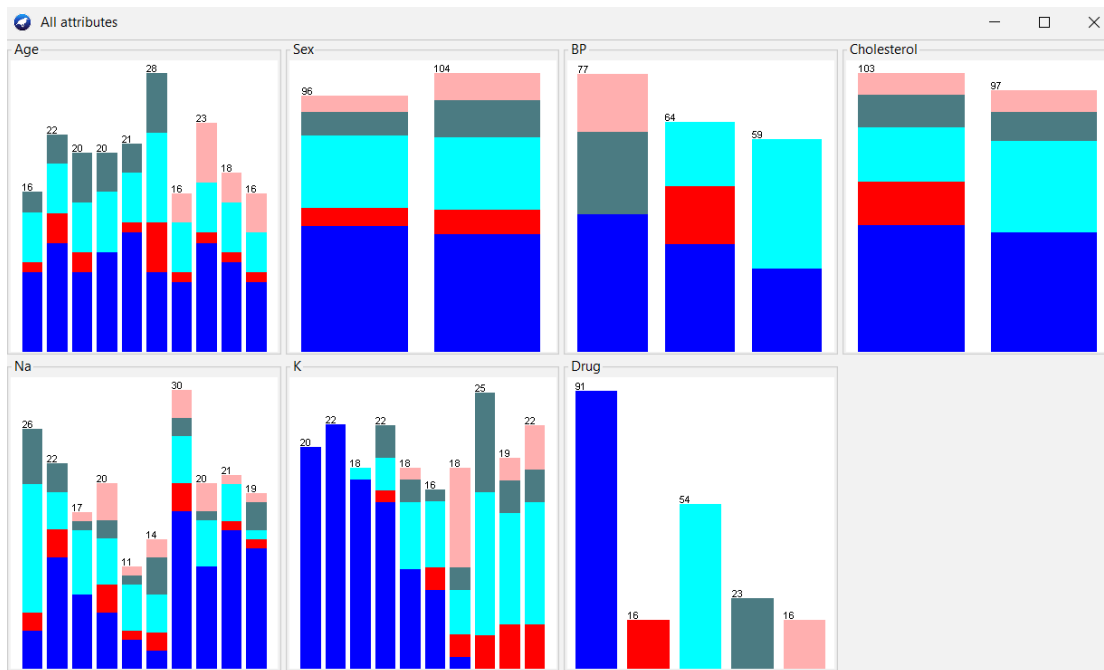


Figura 5. Tabla general de datos

No.	Label	Count	Weight
1	drugY	91	91
2	drugC	16	16
3	drugX	54	54
4	drugA	23	23
5	drugB	16	16

Class: Drug (Nom) Visualize All

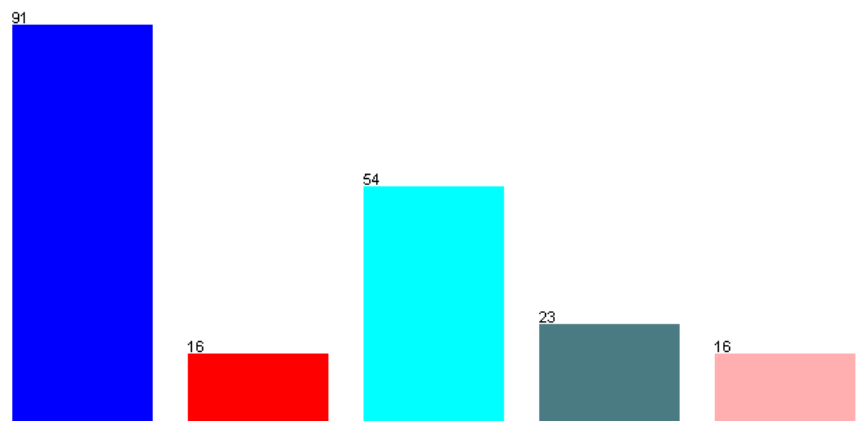


Figura 6. Visualización de tipos de droga y la cantidad usada

De los tipos de drogas, el tipo Y es el más eficaz en los pacientes, seguido de la droga X, droga A, Droga B y por último C.

Para tener más detalles de los datos utilizaremos Python para esta parte:

```
import pandas as pd

# Ruta local del archivo CSV
ruta_archivo_csv = r'C:\Users\Lenovo\Documents\1er Semestre 2023\Gestion de la informacion\Panda\Drug1n.csv'

# Cargar el archivo CSV en un DataFrame de pandas
df = pd.read_csv(ruta_archivo_csv)

# Ver las primeras filas del DataFrame
print(df.head())

# Obtener información sobre el DataFrame
print(df.info())

# Realizar análisis exploratorio de los datos, gráficos, etc.
# Por ejemplo, podemos calcular el rango de edades:
age_range = df['Age'].max() - df['Age'].min()
print("Rango de edades:", age_range)
```

Figura 7. Código principal de python para el análisis de drugs1n.csv

Teniendo todos los datos visualmente sencillos:

```
...   Age Sex   BP Cholesterol   Na   K   Drug
0    23  F   HIGH          HIGH  0.792535  0.031258 drugY
1    47  M   LOW           HIGH  0.739309  0.056468 drugC
2    47  M   LOW           HIGH  0.697269  0.068944 drugC
3    28  F  NORMAL          HIGH  0.563682  0.072289 drugX
4    61  F   LOW           HIGH  0.559294  0.030998 drugY
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Age             200 non-null   int64
1   Sex             200 non-null   object
2   BP              200 non-null   object
3   Cholesterol      200 non-null   object
4   Na              200 non-null   float64
5   K               200 non-null   float64
6   Drug            200 non-null   object
dtypes: float64(2), int64(1), object(4)
memory usage: 11.1+ KB
None
Rango de edades: 59
```

Figura 8. Análisis de datos en panda

Basándonos en los datos se obtiene el siguiente análisis:

- **Age:** La columna "Age" representa la edad de los individuos en el conjunto de datos. El rango de edades va desde 23 hasta 82 años, con un promedio aproximado de 44 años y una mediana (valor medio) alrededor de 45 años. La edad máxima registrada es de 82 años y la edad mínima es de 23 años.
- **Sex:** La columna "Sex" indica el género de los individuos, donde "F" representa a los individuos de género femenino y "M" a los individuos de género masculino.
- **BP:** La columna "BP" representa la presión arterial de los individuos y tiene tres categorías posibles: "HIGH" (alta presión arterial), "LOW" (baja presión arterial) y "NORMAL" (presión arterial normal).
- **Cholesterol:** La columna "Cholesterol" indica el nivel de colesterol de los individuos, con dos categorías posibles: "HIGH" (alto colesterol) y "NORMAL" (colesterol normal).
- **Na:** La columna "Na" representa el nivel de sodio en el cuerpo de los individuos, con valores numéricos continuos que varían dentro del rango de 0.56 a 0.84 aproximadamente.
- **K:** La columna "K" representa el nivel de potasio en el cuerpo de los individuos, con valores numéricos continuos que varían dentro del rango de 0.03 a 0.08 aproximadamente.
- **Drug:** La columna "Drug" indica el tipo de medicamento recetado a los individuos. Existen varias drogas diferentes: "drugY", "drugX", "drugC", "drugA" y "drugB". Cada droga es un tratamiento médico diferente y se receta en función de las condiciones y características de cada individuo.

Haciendo pruebas:

```
count_unique_drugs = df['Drug'].nunique()
print("Cantidad de drogas únicas:", count_unique_drugs)
```

[13] ✓ 0.0s

... Cantidad de drogas únicas: 5

Figura 9

Tenemos un total de 5 drogas anteriormente mencionadas

```
df_describe = df.describe()
print(df_describe)
```

[18] ✓ 0.0s

...	Age	Na	K
count	200.000000	200.000000	200.000000
mean	44.315000	0.697095	0.050174
std	16.544315	0.118907	0.017611
min	15.000000	0.500169	0.020022
25%	31.000000	0.583887	0.035054
50%	45.000000	0.721853	0.049663
75%	58.000000	0.801494	0.066000
max	74.000000	0.896056	0.079788

Figura 10

Estos datos son un resumen estadístico de las columnas "Age", "Na" y "K" del DataFrame. El resumen proporciona información sobre la distribución y características principales de estos datos numéricos. Aquí está lo que significa cada parte:

- **count:** Representa el número de elementos no nulos en cada columna. En este caso, todos los elementos tienen valores no nulos, por lo que el recuento es igual al número total de filas en el DataFrame (200).
- **mean:** Es el valor promedio de cada columna, que representa el centro de la distribución de los datos. Por ejemplo, el promedio de "Age" es aproximadamente 44.32, el promedio de "Na" es aproximadamente 0.70 y el promedio de "K" es aproximadamente 0.05.
- **std:** Es la desviación estándar, que mide la dispersión o variabilidad de los datos alrededor del promedio. Cuanto mayor sea la desviación estándar, mayor será la variabilidad de los datos. Por ejemplo, la desviación estándar de "Age" es aproximadamente 16.54, de "Na" es aproximadamente 0.12 y de "K" es aproximadamente 0.02.
- **min:** Es el valor mínimo de cada columna, que representa el valor más pequeño presente en los datos. Por ejemplo, el valor mínimo de "Age" es 15, de "Na" es aproximadamente 0.50 y de "K" es aproximadamente 0.02.
- **25%:** Es el percentil 25 (primer cuartil), que es el valor que divide los datos en cuatro partes iguales, dejando el 25% de los datos por debajo de este valor. Por ejemplo, el percentil 25 de "Age" es 31, de "Na" es aproximadamente 0.58 y de "K" es aproximadamente 0.04.
- **50%:** Es el percentil 50 (mediana), que es el valor que divide los datos en dos partes iguales, dejando el 50% de los datos por debajo y el 50% por encima de este valor. Por

ejemplo, el percentil 50 de "Age" es 45, de "Na" es aproximadamente 0.72 y de "K" es aproximadamente 0.05.

- **75%:** Es el percentil 75 (tercer cuartil), que es el valor que divide los datos en cuatro partes iguales, dejando el 75% de los datos por debajo de este valor. Por ejemplo, el percentil 75 de "Age" es 58, de "Na" es aproximadamente 0.80 y de "K" es aproximadamente 0.07.
- **max:** Es el valor máximo de cada columna, que representa el valor más grande presente en los datos. Por ejemplo, el valor máximo de "Age" es 74, de "Na" es aproximadamente 0.90 y de "K" es aproximadamente 0.08.

Para comprender mejor, se utiliza una librería que nos permite ver por medio de graficas mejor la edad de los pacientes:

```
import matplotlib.pyplot as plt

plt.hist(df['Age'], bins=10, edgecolor='black')
plt.xlabel('Edad')
plt.ylabel('Frecuencia')
plt.title('Histograma de Edades')
plt.show()
```

✓ 1.7s

Figura 11

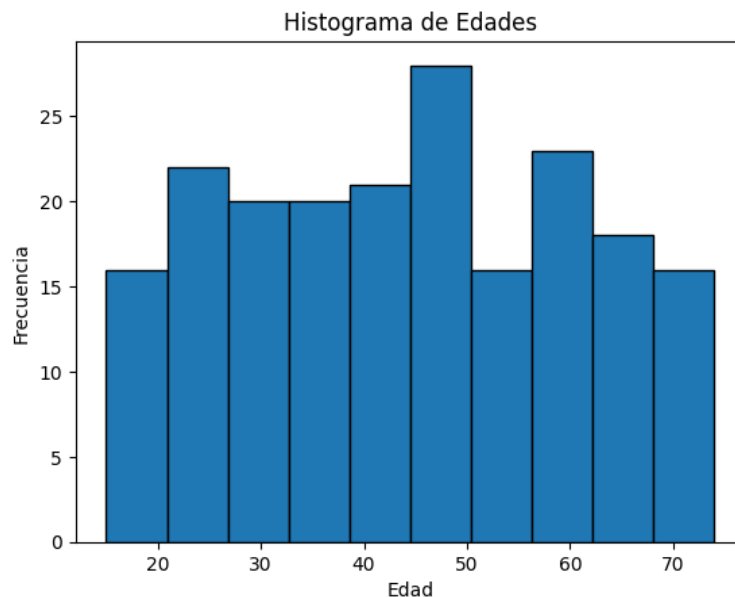


Figura 12

```
import matplotlib.pyplot as plt

gender_counts = df['Sex'].value_counts()
plt.bar(gender_counts.index, gender_counts.values)
plt.xlabel('Género')
plt.ylabel('Cantidad')
plt.title('Cantidad de individuos por género')
plt.show()
```

✓ 0.0s

Figura 13

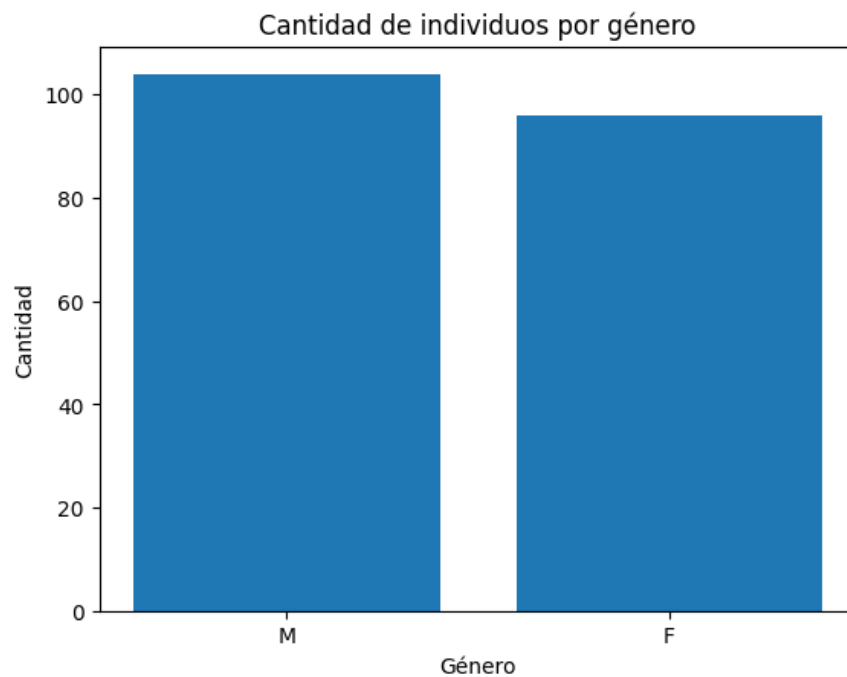


Figura 14

Representamos la cantidad de personas por genero siendo el masculino el predominante.

```
import matplotlib.pyplot as plt

drug_counts = df['Drug'].value_counts()
plt.pie(drug_counts.values, labels=drug_counts.index, autopct='%1.1f%%')
plt.title('Distribución de Drogas Recetadas')
plt.show()
```

20] ✓ 0.0s

Figura 15

Distribución de Drogas Recetadas

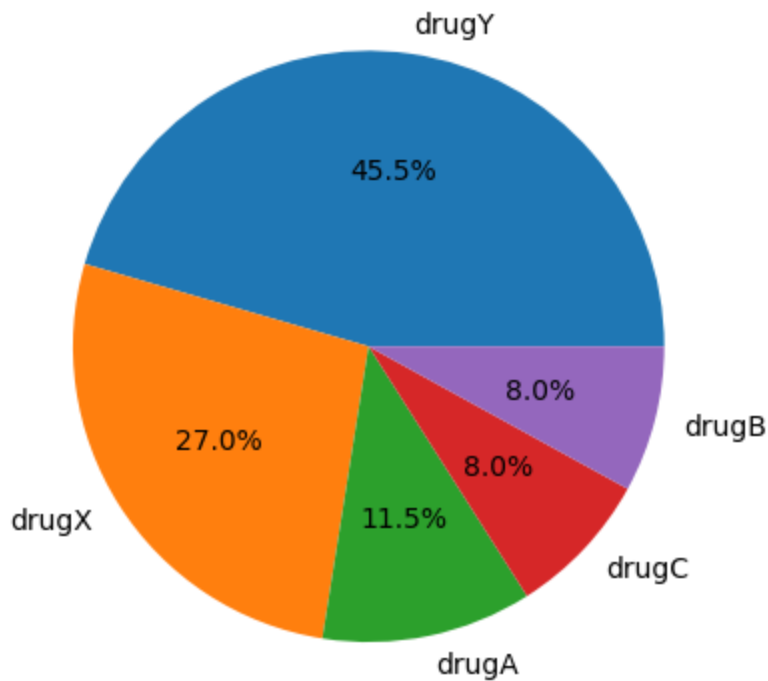


Figura 16

Si vemos por medio de gráficas, los tipos de drogas utilizados en pacientes logramos determinar que drugY es la más eficaz seguido por drug X y los respectivos.

Para finalizar con una vista general de drogas por genero tenemos lo siguiente:

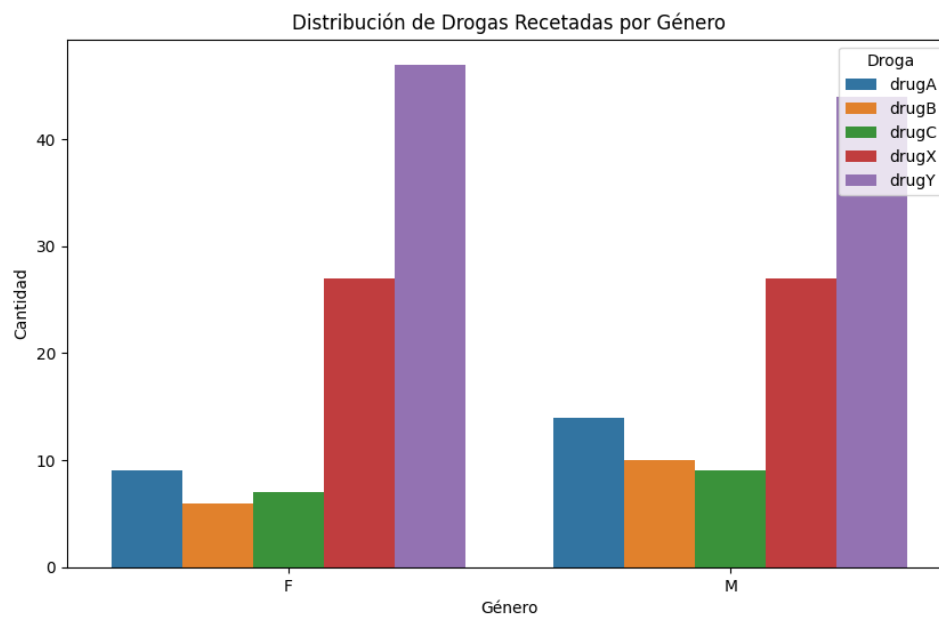


Figura 17

CASO DE ESTUDIO 2: TITANIC

Los datos, contenidos en el fichero titanic.arff, recogen, para cada persona a bordo del transatlántico Titanic durante el fatal viaje inaugural, el sexo, la edad (adulto o niño), el estatus económico (primera, segunda o tercera clase, o tripulación) y si la persona sobrevivió o no al naufragio del transatlántico.

El objetivo de este caso de estudio es analizar, por una parte, la composición de pasajeros y tripulación del Titanic, y, por otra, relación de supervivencia según diversos factores: género, edad y clase social.

Finalmente, se pretende comprobar si la famosa regla “los niños y las mujeres primero” tuvo algún efecto significativo.

- **Preparación de datos**

Viewer

Relation: titanic.txt

No.	1: Clase	2: Edad	3: Sexo	4: Sobrevivió?
	Nominal	Nominal	Nominal	Nominal
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	1	1	1	1
8	1	1	1	1
9	1	1	1	1
10	1	1	1	1
11	1	1	1	1
12	1	1	1	1

Figura 18. Lectura de datos en Weka

Selected attribute				
Name: Clase		Type: Nominal		
Missing: 0 (0%)		Distinct: 4		Unique: 0 (0%)
No.	Label	Count	Weight	
1	0	885	885.0	
2	1	325	325.0	
3	2	285	285.0	
4	3	706	706.0	

Selected attribute				
Name: Edad		Type: Nominal		
Missing: 0 (0%)		Distinct: 2		Unique: 0 (0%)
No.	Label	Count	Weight	
1	0	109	109.0	
2	1	2092	2092.0	

Selected attribute				
Name: Sexo		Type: Nominal		
Missing: 0 (0%)		Distinct: 2		Unique: 0 (0%)
No.	Label	Count	Weight	
1	0	470	470.0	
2	1	1731	1731.0	

Selected attribute				
Name: Sobrevivió?		Type: Nominal		
Missing: 0 (0%)		Distinct: 2		Unique: 0 (0%)
No.	Label	Count	Weight	
1	0	1490	1490.0	
2	1	711	711.0	

Figura 19. Revisión de parámetros de los atributos

Primero se realiza la preparación de datos para el análisis de este caso de uso de Titanic, el cual tiene 4 atributos que son “Clase”, “Edad”, “Sexo”, “¿Sobrevivió?”. Para la preparación de datos se hacen las siguientes revisiones con la herramienta **Weka** (Figura 18 y figura 19) y se llegan a los siguientes resultados:

- El dataset no contiene datos faltantes.
- Los atributos son de tipo “Nominal”.

Luego de estas revisiones nos encontramos en la posición de poder iniciar el manejo de datos.

```
import pandas as pd

titanic = pd.read_csv('titanic.csv', sep=',', encoding='latin-1')
print(titanic)
```

Figura 20. Se lee en Python el archivo “titanic.csv”

	Clase	Edad	Sexo	Sobrevivió?
0	1	1	1	1
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
...
2196	0	1	0	1
2197	0	1	0	1
2198	0	1	0	0
2199	0	1	0	0
2200	0	1	0	0

[2201 rows x 4 columns]

Figura 21. Se imprime el archivo “titanic” para verificar los datos

Segundo, pasamos a reemplazar los datos para ser visualizados de forma que no se muestren en 1 y 0, sino por el nombre de cada dato. Para esto tomamos el archivo, anteriormente convertido, “titanic.csv” para cambiar los datos por medio de **Python** (figura 20), y verificamos que se lean los 2201 registros y las columnas, Clase, Edad, Sexo, Sobrevivió (figura 21).

Luego de esto, transformamos los tipos de datos a “string” y reemplazamos los valores (figura 22):

Clase: 0=tripulación, 1=primera, 2=segunda, 3=tercera

Edad: 1=adulto, 0=niño

Sexo :1=hombre, 0=mujer

Sobrevivió: 1=si, 0=no

```
titanic['Edad'] = titanic['Edad'].astype(str).str.replace('1', 'adulto').str.replace('0', 'niño')
titanic['Clase'] = titanic['Clase'].astype(str).str.replace('0', 'tripulacion')
titanic['Clase'] = titanic['Clase'].astype(str).str.replace('1', 'primera').str.replace('2', 'segunda').str.replace('3', 'tercera')
titanic['Sexo'] = titanic['Sexo'].astype(str).str.replace('1', 'hombre').str.replace('0', 'mujer')
titanic['Sobrevivió?'] = titanic['Sobrevivió?'].astype(str).str.replace('1', 'si').str.replace('0', 'no')

print(titanic)
```

Figura 22. Reemplazo de datos

```

      Clase  Edad  Sexo  Sobrevivió?
0      primera  adulto  hombre      si
1      primera  adulto  hombre      si
2      primera  adulto  hombre      si
3      primera  adulto  hombre      si
4      primera  adulto  hombre      si
...      ...      ...      ...      ...
2196  tripulacion  adulto  mujer      si
2197  tripulacion  adulto  mujer      si
2198  tripulacion  adulto  mujer      no
2199  tripulacion  adulto  mujer      no
2200  tripulacion  adulto  mujer      no

[2201 rows x 4 columns]
```

Figura 23. Se imprime el archivo “titanic” para verificar los datos

```
titanic.to_csv('titanic2.csv', sep=',', encoding='latin-1', index=False)
```

Figura 24. creación de archivo “titanic2.csv”

Podemos ver que se reemplaza los datos (figura 23), por el nombre de los datos, y que en Clase sea tripulación, primera, segunda, tercera; en Edad, adulto, niño; en Sexo, hombre, mujer; y en Sobrevivió, sí y no.

Luego de esto creamos un archivo con nombre “titanic2.csv” para guardar estos cambios que hemos realizados (figura 24).

- **Modelado y análisis de datos**

Para el modelado y análisis de datos utilizaremos la herramienta **Tableau**.

titanic2

Conexión
☒ En tiempo real

titanic2.csv



¿Necesita más datos?

Arrastre tablas para establecer relaciones entre ellas. [Más información](#)

titanic2.csv 8 campos 2201 filas			
Nombre titanic2.csv			
Campos			
Tipo	Nombre de campo	Tabla física	Nombre de ...
Abc	Clase	titanic2.csv	Clase
Abc	Edad	titanic2.csv	Edad
Abc	Sexo	titanic2.csv	Sexo
Abc	Sobrevivió?	titanic2.csv	Sobrevivió?

Abc titanic2.csv	Abc titanic2.csv	Abc titanic2.csv	Abc titanic2.csv
Clase	Edad	Sexo	Sobrevivió?
primera	adulto	hombre	si
primera	adulto	hombre	si
primera	adulto	hombre	si
primera	adulto	hombre	si
primera	adulto	hombre	si
primera	adulto	hombre	si
primera	adulto	hombre	si
primera	adulto	hombre	si

Figura 25. Ingreso de dataset "titanic2.csv" en Tableau

Como primer paso agregamos el archivo "titanic2.csv" como fuente de datos en Tableau. Vemos en la Figura 25 que se carga correctamente los campos y los registros de los datos, por lo que podemos empezar con la visualización por medio de gráficas.

Pasajeros y Tripulantes

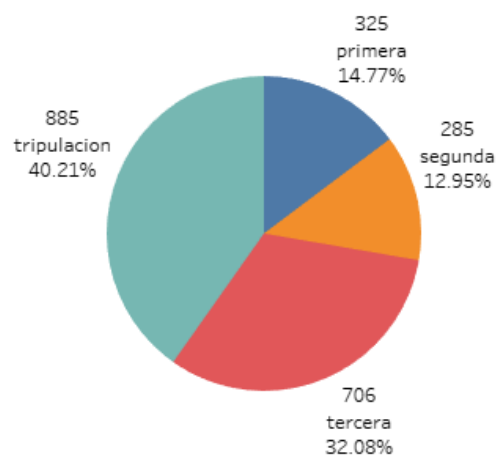


Figura 26. Gráfica de pie de cantidad de pasajeros y tripulantes

En la Figura 26 podemos ver que el número total de personas a bordo del Titanic durante la travesía inaugural era de 2201, divididas en 885 miembros de la tripulación y 1316 pasajeros, los cuales se dividen 706 en tercera, 285 en segunda y 325 en primera clase.

Edad-Clase

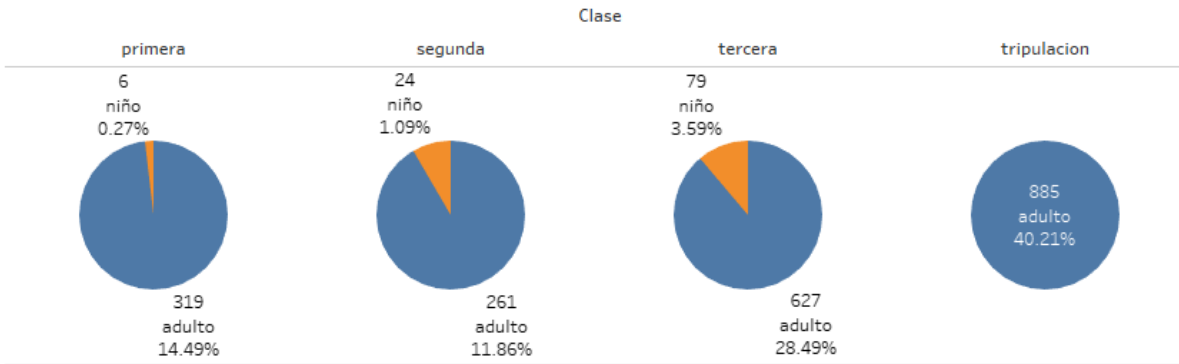


Figura 27. gráfico de pie de clase vs edad

En la Figura 27 se muestra la distribución de cuantos adultos y niños hay por clase, se puede ver en la tripulación son todos adultos, en primera clase hay menos niños con un total de 6, y en tercera clase hay más niños con un total de 79.

Sexo-Clase

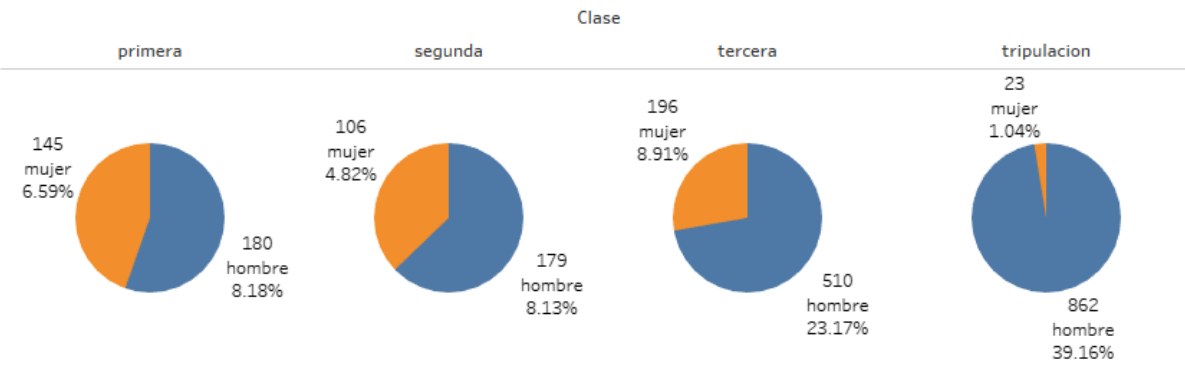


Figura 28. gráfico de pie de clase vs sexo

En la Figura 28 se muestra la distribución de cuantas mujeres y hombres hay por clase, dentro de los miembros de la tripulación se puede mencionar que 23 de ellos mujeres y el resto, 862, hombres. En primera clase encontramos 145 mujeres y hombres mientras que en tercera clase tenemos 196 mujeres y

Por su parte, la distribución de pasajeros se muestra en las siguientes tablas:

Sexo-Edad-Clase

Edad	Clase	Sexo		Total general
		hombre	mujer	
adulto	primera	175	144	319
niño	primera	5	1	6
Total general		180	145	325

Figura 29. Pasajeros de primera clase por sexo y edad

Sexo-Edad-Clase

Edad	Clase	Sexo		Total general
		hombre	mujer	
adulto	segunda	168	93	261
niño	segunda	11	13	24
Total general		179	106	285

Figura 30. Pasajeros de segunda clase por sexo y edad

Sexo-Edad-Clase

Edad	Clase	Sexo		Total general
		hombre	mujer	
adulto	tercera	462	165	627
niño	tercera	48	31	79
Total general		510	196	706

Figura 31. Pasajeros de tercera clase por sexo y edad

El gran impacto del naufragio del Titanic se tuvo (y tiene), se debe entre otras cosas a una comparación del número de sobrevivientes con el número total de personas a bordo. Concretamente, fallecieron 1490 de 2201, mientras que solo sobrevivieron 711, o el 32,3% del total.

En base a los datos que tenemos a mano podemos construir las siguientes tablas que muestran el número de sobrevivientes del naufragio primero con relación a la clase y la edad, segundo con relación la clase y el sexo, por último, los sobrevivientes por cada clase.

%Sobrevivio-Edad

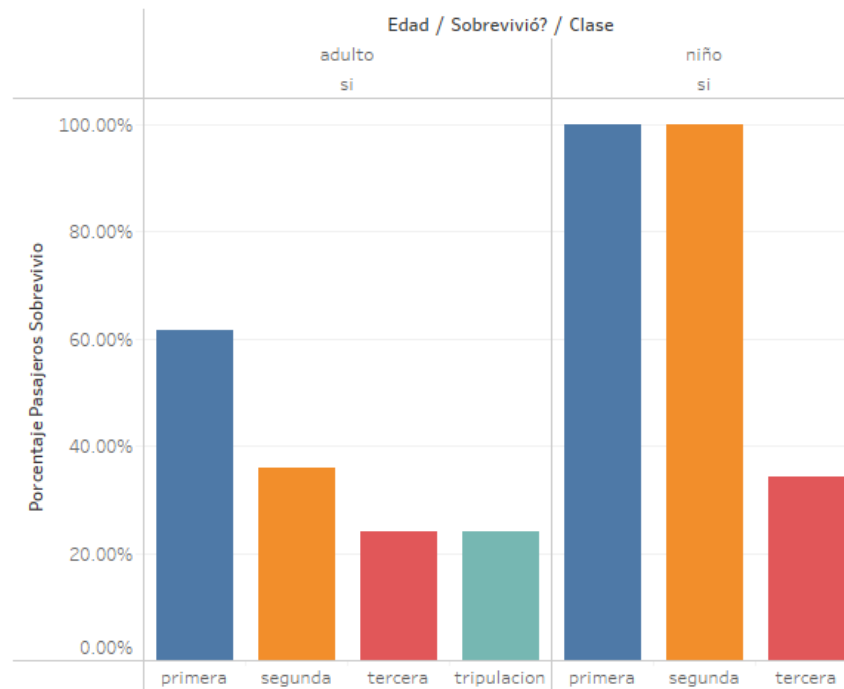


Figura 32. Sobrevivientes del naufragio primero con relación a la clase y la edad

En la figura 32 se puede observar que los niños de la primera y segunda clase cuentan con un porcentaje de 100% que sobrevivieron.

Sobrevivio-Sexo

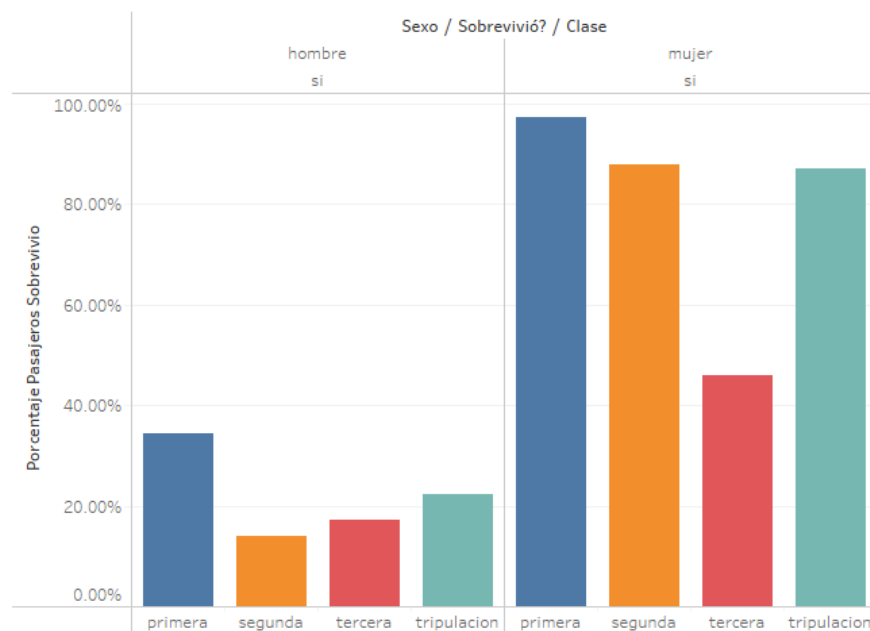


Figura 33. Sobrevivientes del naufragio primero con relación a la clase y el sexo

En la figura 33 se puede observar que las mujeres de todas las clases tuvieron un porcentaje mayor de sobrevivencia en comparación con los hombres.

Sobrevivio-Clase2

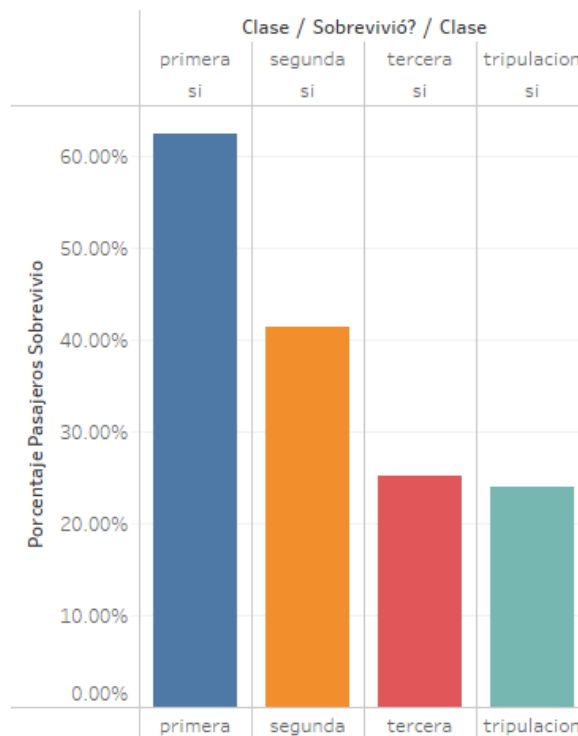


Figura 34. Sobrevivientes del naufragio primero con relación a la clase

En la figura 34 se puede observar que la primera clase fue donde un mayor porcentaje de sobrevivientes seguido por la segunda clase, luego la tercera y por último la tripulación.

Según lo visto podemos mencionar las siguientes conclusiones de este caso:

- La frase “las mujeres y los niños primero” parece verificarse en su mayor parte.
- Por otra parte, el estatus económico también tuvo un efecto importante en la relación de supervivencia. Los pasajeros de primera clase fueron los que más probabilidad de sobrevivir tuvieron.
- Dentro de cada clase, la relación de supervivencia para mujeres y niños fueron mayores que para hombres, pero la relación de supervivencia para los hombres de primera clase fue igual que el de los niños de tercera clase y casi tan alto como el de las mujeres de tercera clase.

CASO DE ESTUDIO 3: VEHICLES

Una descripción breve sobre este caso de estudio es que los datos contienen características extraídas de la silueta de los carros en diferentes ángulos. Para el experimento se utilizaron cuatro modelos de vehículos; un bus de dos pisos, una Van Chevrolet y dos carros; un Saab 9000 y un Opel Manta 400. Estas elecciones de vehículos fueron con el propósito de que el bus, la van y cualquiera de los carros fuera claramente de distinguir el uno con el otro, sin embargo, será más difícil distinguir los dos carros entre sí ya que son físicamente parecidos.



Figura 35. Van Chevrolet. Chevrolet Express



Figura 36. Opel Manta



Figura 37. Saab 9000



Figura 38. Bus de dos pisos. B430R Marcopolo DD1800

Tenemos como objetivo clasificar una silueta dada como uno de los cuatro tipos de vehículos utilizando un conjunto de características extraídas de la silueta. El vehículo en cuestión se puede ver desde uno de los muchos ángulos diferentes.


```

COMPACTNESS      (average perim)**2/area
CIRCULARITY      (average radius)**2/area
DISTANCE CIRCULARITY  area/(av.distance from border)**2
RADIUS RATIO     (max.rad-min.rad)/av.radius
PR.AXIS ASPECT RATIO  (minor axis)/(major axis)
MAX.LENGTH ASPECT RATIO (length perp. max length)/(max length)
SCATTER RATIO    (inertia about minor axis)/(inertia about major axis)
ELONGATEDNESS    area/(shrink width)**2
PR.AXIS RECTANGULARITY area/(pr.axis length*pr.axis width)
MAX.LENGTH RECTANGULARITY area/(max.length*length perp. to this)
SCALED VARIANCE  (2nd order moment about minor axis)/area
ALONG MAJOR AXIS
SCALED VARIANCE  (2nd order moment about major axis)/area
ALONG MINOR AXIS
SCALED RADIUS OF GYRATION  (mavar+mivar)/area
SKEWNESS ABOUT  (3rd order moment about major axis)/sigma_min**3
MAJOR AXIS
SKEWNESS ABOUT  (3rd order moment about minor axis)/sigma_maj**3
MINOR AXIS
KURTOSIS ABOUT  (4th order moment about major axis)/sigma_min**4
MINOR AXIS
KURTOSIS ABOUT  (4th order moment about minor axis)/sigma_maj**4
MAJOR AXIS

```

Figura 39. Cálculos de los atributos.

Podemos ver también, que las personas que realizaron este estudio, debieron sacar y calcular los datos, pero nos hicimos la pregunta ¿Cómo sacaron estos resultados o que formulas usaron? Al ver la imagen de arriba, abriendo el archivo .arff en un bloc de notas, nos encontramos con una descripción breve de los cálculos y formulas utilizados para sacar los números y resultados plasmados en los atributos de la tabla del dataset.

vehicles.csv 19 campos 846 filas				100 → filas	
<div>Nombre</div> <div>vehicles.csv</div>				#	#
				vehicles.csv	vehicles.csv
				Compactness	Circularity
				95	48
				91	41
				104	50
				93	41
				85	44
				107	57
				97	43
				90	43
				86	34
				93	44
				86	36
				90	34
				83	178
				84	141
				106	209
				82	159
				70	205
				106	172
				73	173
				66	157
				62	140
				98	197
				70	143
				66	136
				72	
				57	
				66	
				63	
				103	
				50	
				65	
				65	
				61	
				62	
				61	
				55	

Figura 40. Vista Atributos Tableau Vehicles

Implementando los datos en Tableau, se obtiene una vista de los atributos de esta manera. Tenemos 846 filas (en realidad eran 946 pero 100 ejemplos fueron conservados por la universidad de Strathclyde para su validación), y 19 columnas o atributos. Todos los campos son numéricos excepto el “Class”.

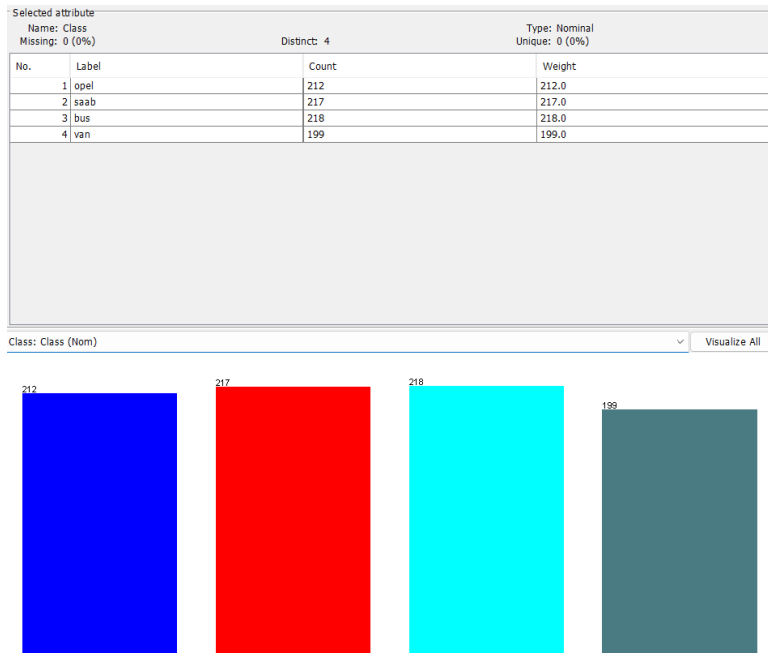


Figura 41. Vista de cantidad de clases en Weka

Existen 429 datos que son “carros”, 218 datos que son tipo “bus” y 199 datos que son tipo “van”

Para dar una pequeña explicación sobre lo que es cada atributo, aquí mostraremos una pequeña traducción de cada uno:

- 'COMPACTNESS' real----->compacidad, dato tipo real
- 'CIRCULARITY' real----->circularidad, dato tipo real
- 'DISTANCE CIRCULARITY' real----->distancia de la circularidad, dato tipo real
- 'RADIUS RATIO' real----->ratio del radio, dato tipo real
- 'PR. AXIS ASPECT RATIO' real----->aspecto del eje, dato tipo real
- 'MAX.LENGTH ASPECT RATIO' real----->aspecto maximo del largo, dato tipo real
- 'SCATTER RATIO' real----->dispersión, dato tipo real
- 'ELONGATEDNESS' real----->alargamiento, dato tipo real
- 'PR.AXIS RECTANGULARITY' real----->rectangularidad del eje, dato tipo real
- 'MAX.LENGTH RECTANGULARITY' real----->rectangularidad maxima del largo, dato tipo real
- 'SCALED VARIANCE_MAJOR' real----->varianza escalada a lo largo del eje mayor, dato tipo real

- 'SCALED VARIANCE_MINOR' real----->varianza escalada a lo largo del eje menor, dato tipo real
- 'SCALED RADIUS OF GYRATION' real---->radio de giro escalado, dato tipo real
- 'SKEWNESS ABOUT_MAJOR' real----->oblicuidad a lo largo del eje mayor, dato tipo real
- 'SKEWNESS ABOUT_MINOR' real----->oblicuidad a lo largo del eje menor, dato tipo real
- 'KURTOSIS ABOUT_MAJOR' real---->curtosis a lo largo del eje mayor, dato tipo real
- 'KURTOSIS ABOUT_MINOR' real----->curtosis a lo largo del eje menor, dato tipo real
- 'HOLLOWS RATIO' real----->proporcion de huecos, dato tipo real
- 'Class' {opel,saab,bus,van}

Para proporcionar una manera mas detallada de cada fila, es necesario que cada una tenga su propio "ID" único y así poder distinguir de manera mas exacta si queremos usar herramientas propias en Tableau o Weka o cualquier herramienta.

```
@relation vehicle
```

```
@attribute 'ID' real
@attribute 'COMPACTNESS' real
@attribute 'CIRCULARITY' real
@attribute 'DISTANCE CIRCULARITY' real
@attribute 'RADIUS RATIO' real
@attribute 'PR.AXIS ASPECT RATIO' real
@attribute 'MAX.LENGTH ASPECT RATIO' real
@attribute 'SCATTER RATIO' real
@attribute 'ELONGATEDNESS' real
@attribute 'PR.AXIS RECTANGULARITY' real
@attribute 'MAX.LENGTH RECTANGULARITY' real
@attribute 'SCALED VARIANCE_MAJOR' real
@attribute 'SCALED VARIANCE_MINOR' real
@attribute 'SCALED RADIUS OF GYRATION' real
@attribute 'SKEWNESS ABOUT_MAJOR' real
@attribute 'SKEWNESS ABOUT_MINOR' real
@attribute 'KURTOSIS ABOUT_MAJOR' real
@attribute 'KURTOSIS ABOUT_MINOR' real
@attribute 'HOLLOWS RATIO' real
@attribute 'Class' {opel,saab,bus,van}
```

```
@data
```

```
1,95,48,83,178,72,10,162,42,20,159,176,379,184,70,6,16,187,197,van
2,91,41,84,141,57,9,149,45,19,143,170,330,158,72,9,14,189,199,van
3,104,50,106,209,66,10,207,32,23,158,223,635,220,73,14,9,188,196,saab
4,93,41,82,159,63,9,144,46,19,143,160,309,127,63,6,10,199,207,van
5,85,44,70,205,103,52,149,45,19,144,241,325,188,127,9,11,180,183,bus
6,107,57,106,172,50,6,255,26,28,169,280,957,264,85,5,9,181,183,bus
7,97,43,73,173,65,6,153,42,19,143,176,361,172,66,13,1,200,204,bus
8,90,43,66,157,65,9,137,48,18,146,162,281,164,67,3,3,193,202,van
9,86,34,62,140,61,7,122,54,17,127,141,223,112,64,2,14,200,208,van
10,93,44,98,197,62,11,183,36,22,146,202,505,152,64,4,14,195,204,saab
11,86,36,70,143,61,9,133,50,18,130,153,266,127,66,2,10,194,202,van
12,90,34,66,136,55,6,123,54,17,118,148,224,118,65,5,26,196,202,saab
13,88,46,74,171,68,6,152,43,19,148,180,349,192,71,5,11,189,195,bus
14,89,47,85,144,58,10,157,44,19,144,173,345,161,72,8,13,187,197,van
```

Figura 42. ID de cada fila vista en bloc de notas

Otro aspecto muy importante, es revisar valores faltantes. Una manera muy fácil y compleja de ver, buscar y encontrar datos faltantes en filas o columnas puede ser gracias a la herramienta de Weka, que automáticamente, nos proporciona. Al encontrar los datos faltantes, es necesario tomar acciones pertinentes para rellenar estos “espacios” con datos artificiales o no.

Selected attribute	
Name: RADIUS RATIO	
Missing: 0 (0%)	Distinct: 134
Type: Numeric	
Unique: 14 (2%)	
Statistic	Value
Minimum	104
Maximum	333
Mean	168.941
StdDev	33.472

Figura 43. Datos Faltantes Atributo “Radius Ratio”

Selected attribute	
Name: DISTANCE CIRCULARITY	
Missing: 0 (0%)	Distinct: 63
Type: Numeric	
Unique: 7 (1%)	
Statistic	Value
Minimum	40
Maximum	112
Mean	82.089
StdDev	15.772

Figura 44. Datos Faltantes Atributo “Distance Circularity”

Selected attribute	
Name: CIRCULARITY	
Missing: 0 (0%)	Distinct: 27
Type: Numeric	
Unique: 1 (0%)	
Statistic	Value
Minimum	33
Maximum	59
Mean	44.862
StdDev	6.17

Figura 45. Datos Faltantes Atributo “Circularity”

Como podemos apreciar gracias a estas imágenes, Weka nos proporciona de manera automática si hay datos faltantes o no en cada atributo. Sin embargo, no encontramos ningún dato faltante en ninguno de los otros atributos (no se colocaron las imágenes para ahorrar espacio).

A continuación, mostraremos las correlaciones de los atributos gracias a la propiedad de Weka de Visualización. Pero, antes que nada, ¿para que sirven las correlaciones? Son herramientas poderosas para medir la asociación entre dos variables y poder proporcionar valiosa información sobre como se comportan conjuntamente. Tiene varios usos útiles como, por ejemplo, identificar si ambas variables tienden a aumentar o disminuir juntas, también existe la correlación negativa que indica cuando una variable tiende a aumentar cuando la otra disminuye. Con esto también se puede medir la fuerza de la relación, se puede predecir valores basándose en los valores de otra variable correlacionada, se pueden identificar patrones ocultos e incluso validar hipótesis sobre la relación de dos variables.

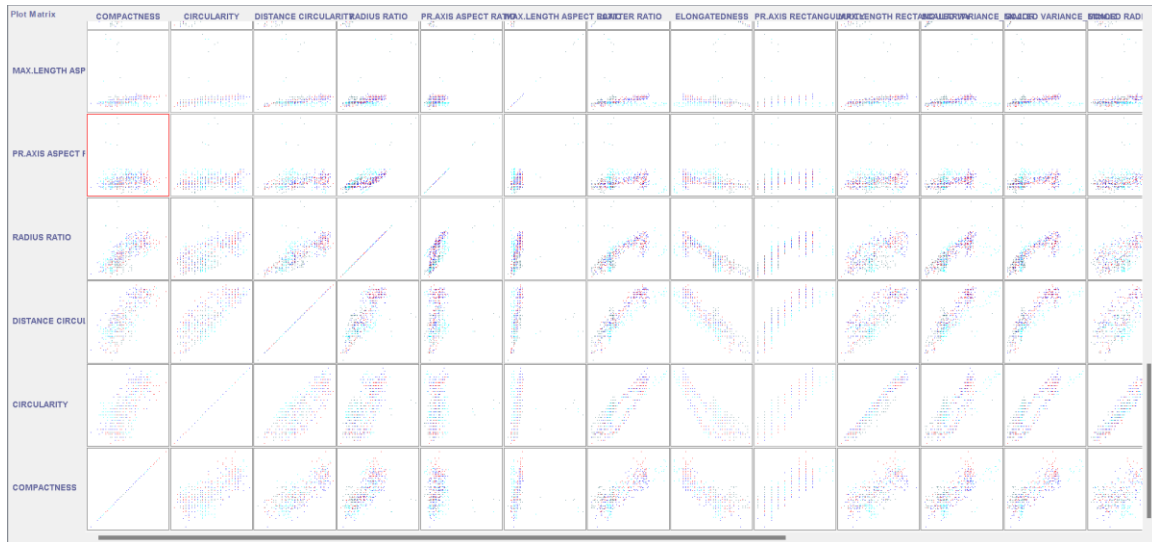


Figura 46. Correlación atributos Vehicles

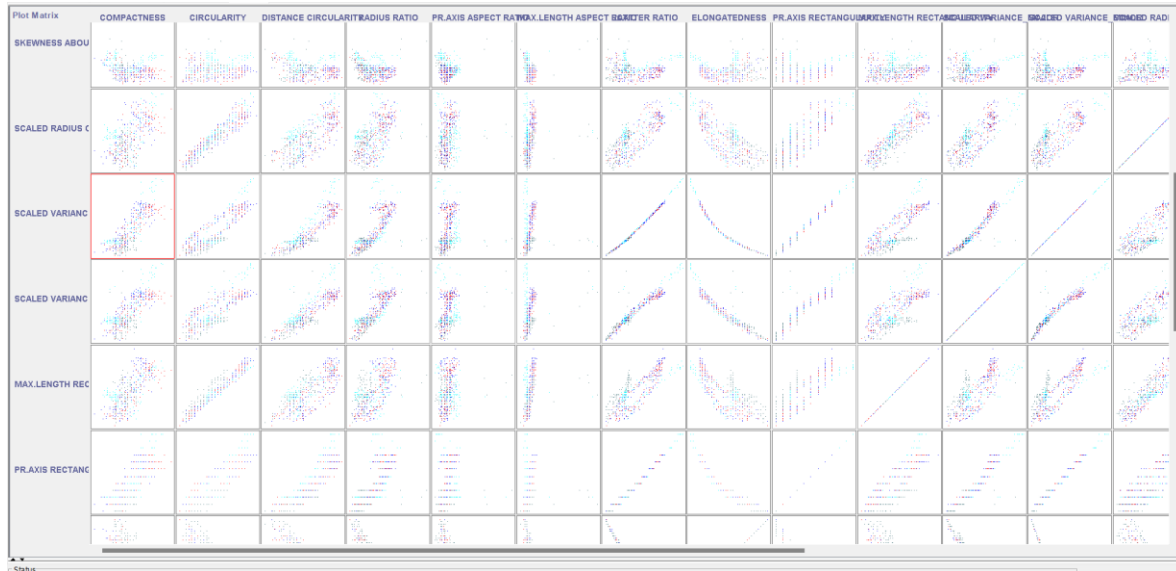


Figura 47. Correlación atributos Vehicles



Figura 48. Correlación atributos Vehicles

Pero, las correlaciones no solamente se pueden “calcular” a simple vista. A continuación, mostraremos un método muy útil para calcularlas con la herramienta de Tableau. Tomaremos de ejemplo los atributos de “Scatter Ratio” y “Scaled Variance Minor”, ya que a simple vista se puede ver una estrecha correlación.

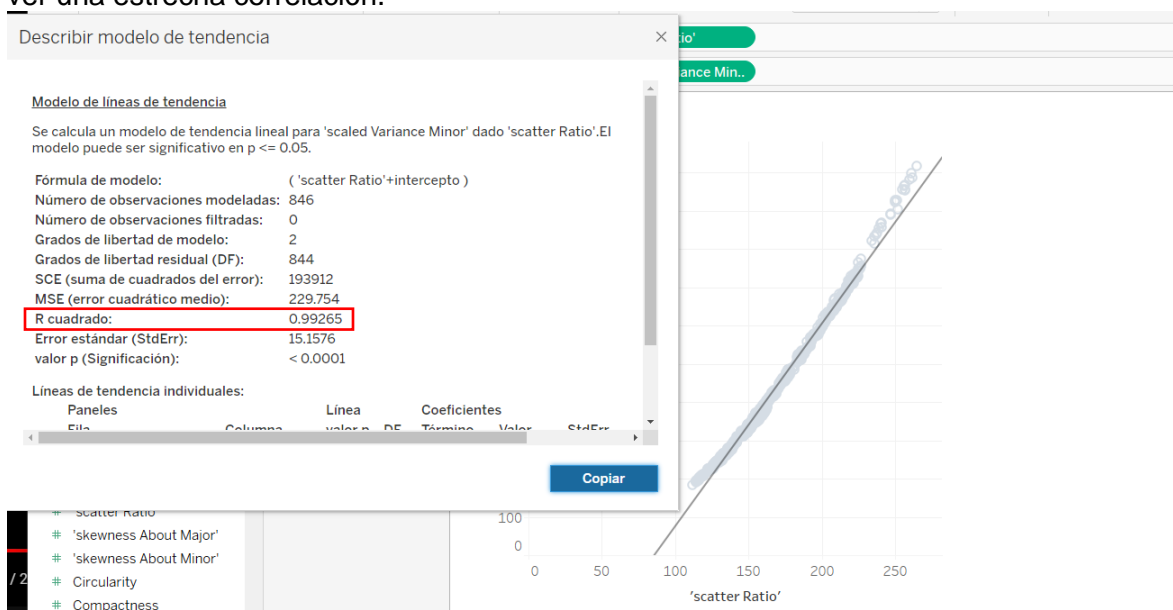


Figura 49. Correlación Tableau

Nosotros al relacionar los dos atributos y activar la línea de tendencia, podemos darle click derecho a la línea de tendencia para así mostrarnos datos importantes. Pero el que nos interesa por ahora, es el R cuadrado, que es la correlación que tienen los dos atributos. Según la documentación oficial de Tableau; cuanto más próxima sea la correlación (r) a -1 o 1, más fuerte

será la relación entre x e y. Si R es próximo o igual a 0, existe una relación débil entre las medidas o no hay relación.

Y siendo más exactos, por regla general, se puede interpretar los valores de R de esta forma:

- +0,70 o más indica una relación positiva muy fuerte;
- +0,40 hasta +0,69 indica una relación positiva fuerte;
- +0,20 hasta +0,39 indica una relación positiva moderada;
- -0,19 hasta +0,19 indica ausencia de relación o una relación débil;
- -0,20 hasta -0,39 indica una relación negativa moderada;
- -0,40 hasta -0,69 indica una relación negativa fuerte;
- -0,70 o inferior indica una relación negativa muy fuerte.

El único “inconveniente con esto” es que es una correlación promedio, si queremos saber una correlación mas estrecha y exacta entre cada fila, existe una manera de hacerlo.

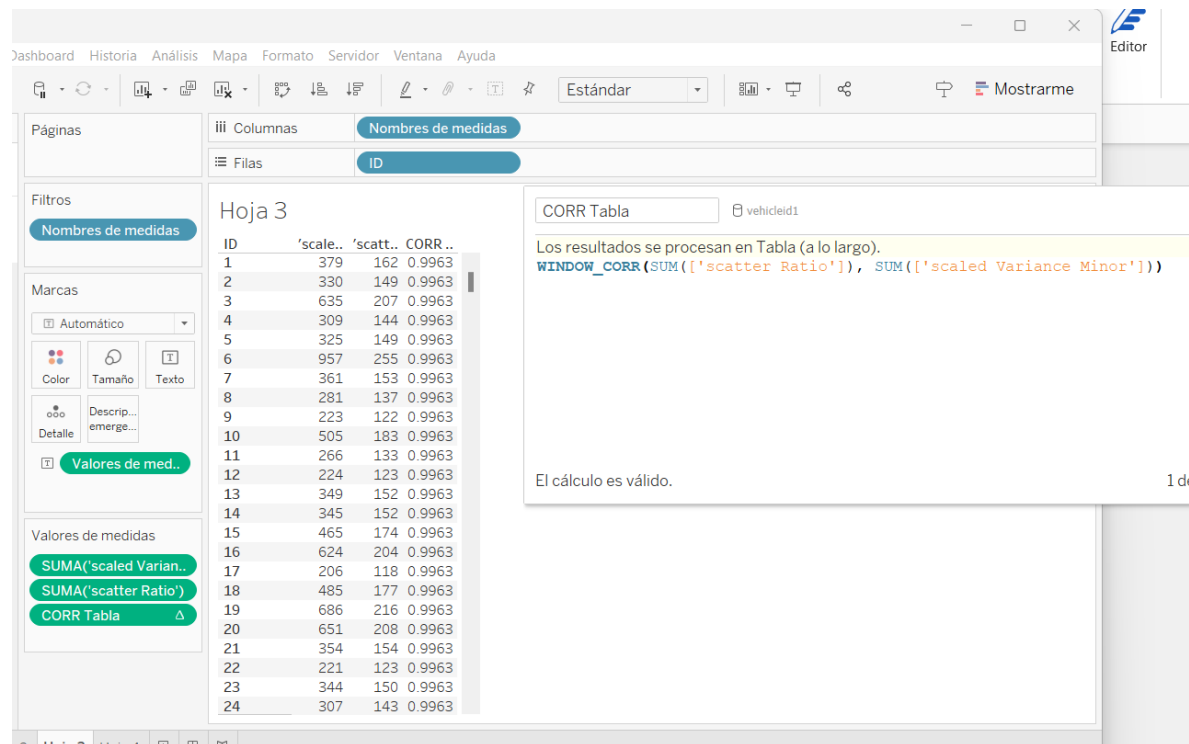


Figura 50. Correlación por fila Tableau

Como se ve en la Figura 15, podemos ver que hemos creado una correlación entre fila de los atributos anteriormente mencionados, simplemente creamos un campo calculado como se ve en la figura que es la correlación entre dos expresiones dentro de la ventana. Y lo implementamos en los valores de las medidas de las tablas.

Sabiendo esto, que ya sabemos calcular la correlación y revisando los demás atributos podemos decir y concluir que los atributos que tienen una correlación muy fuerte son; “Scaled Variance

Minor” y “Scatter Ratio” que tienen una correlación casi perfecta de 1. También están las correlaciones de “Scaled Variance Major” y “Scaled Variance Minor”, “Elongatedness” y “Scaled Variance Minor”, “Elongatedness” y “Scaled Variance Major” con una correlación muy fuerte. Existen también muchas más correlaciones por arriba de los ± 0.7 .

CONCLUSIONES Y RECOMENDACIONES PARA EL CASO DE ESTUDIO 3

- El promedio del “Compactness” es menor para las van. El promedio del “Compactness” es mucho mayor para los carros. Para los buses el “Compactness” indica que hay menos cantidad de buses con alta compacidad.

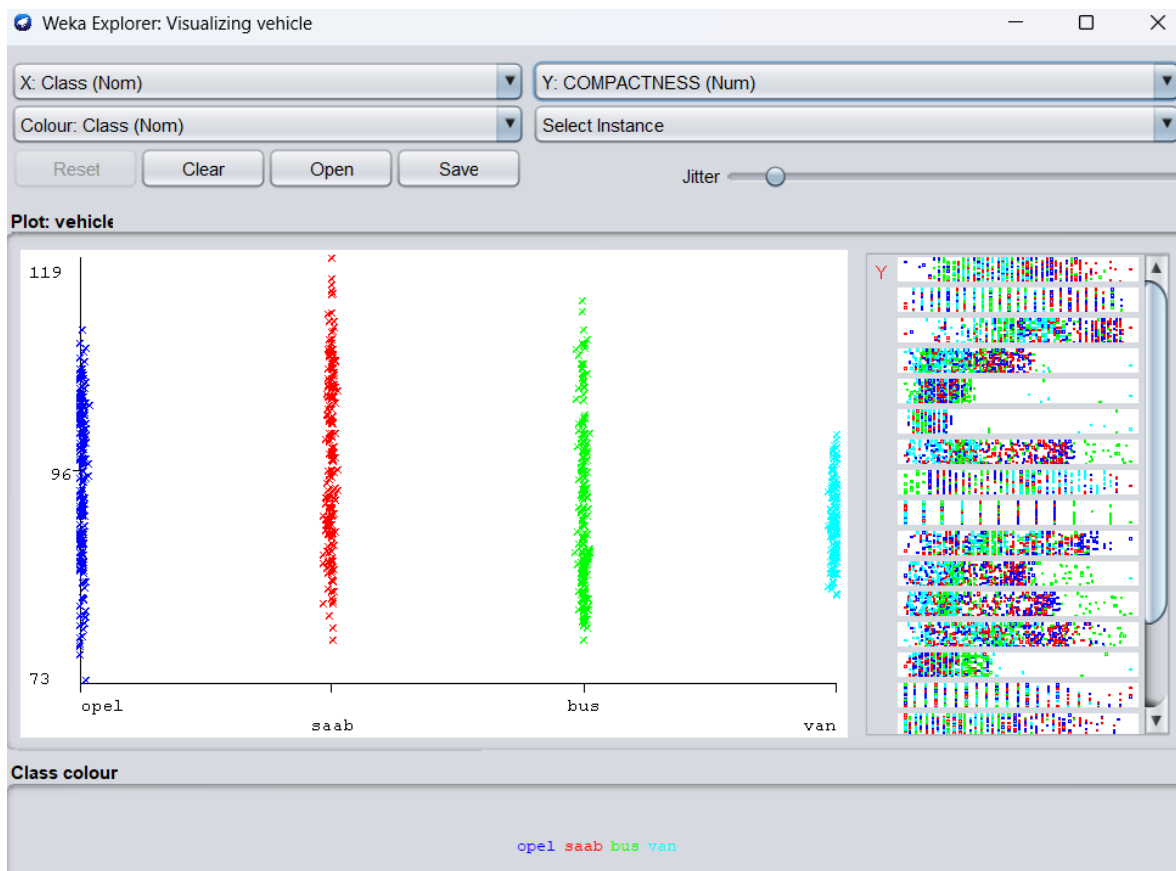


Figura 51. Gráfico de dispersión. Promedio del atributo “Compactness” para el conjunto de vehículos de cada clase.

- La “Circularity” media es mayor para los carros.

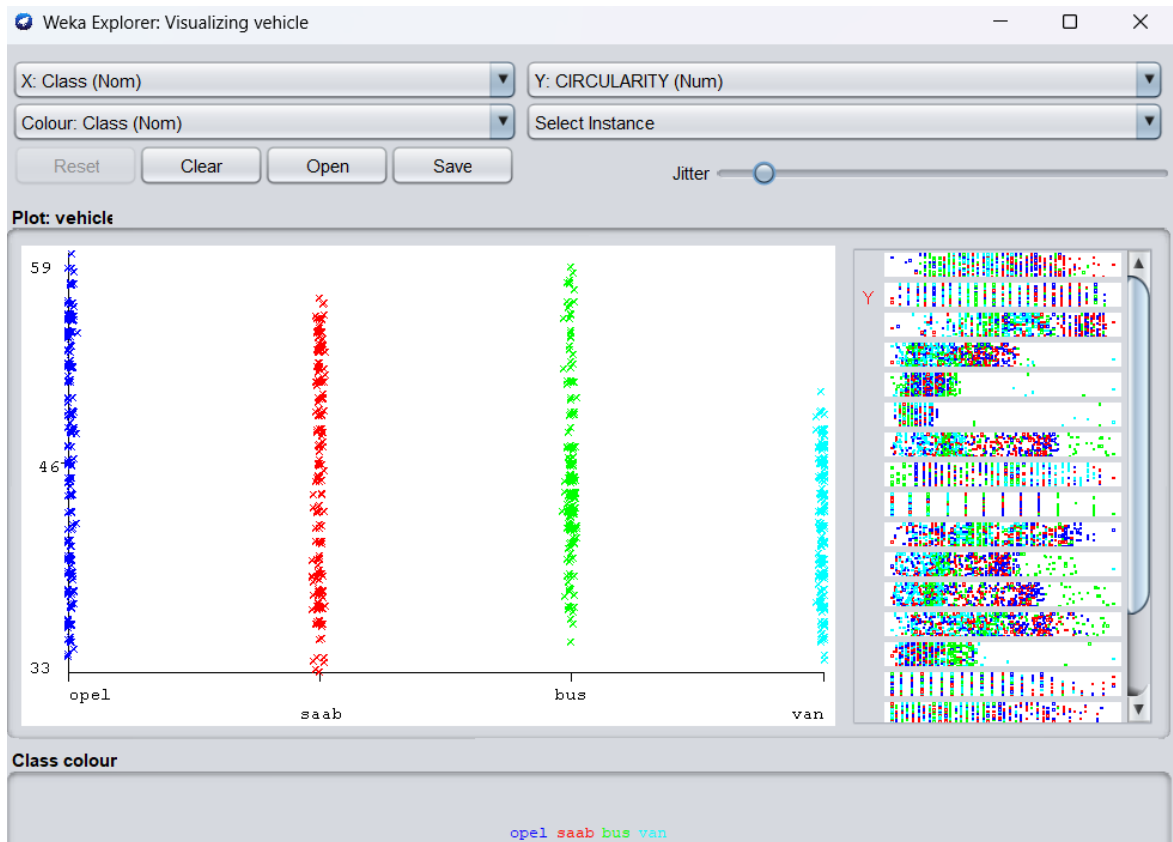


Figura 52. Gráfico de dispersión. Media del atributo “Circularity” para el conjunto de vehículos de cada clase.

- La “Distance Circularity” promedio es también mayor para los carros y le siguen los buses.

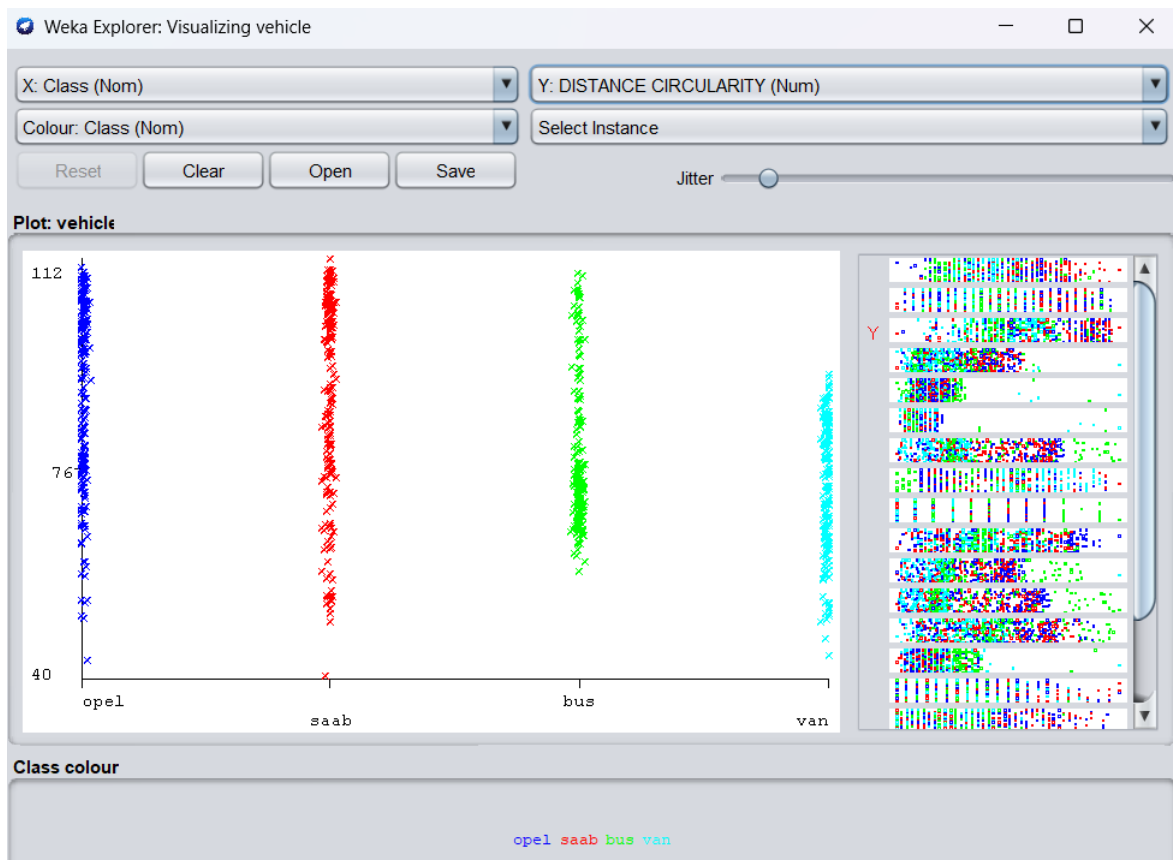


Figura 53. Gráfico de dispersión. Promedio del atributo “Distance Circularity” para el conjunto de vehículos de cada clase.

- El “Radius Ratio” promedio es mayor para los carros y le siguen los buses. Pero en las van es mucho menor.

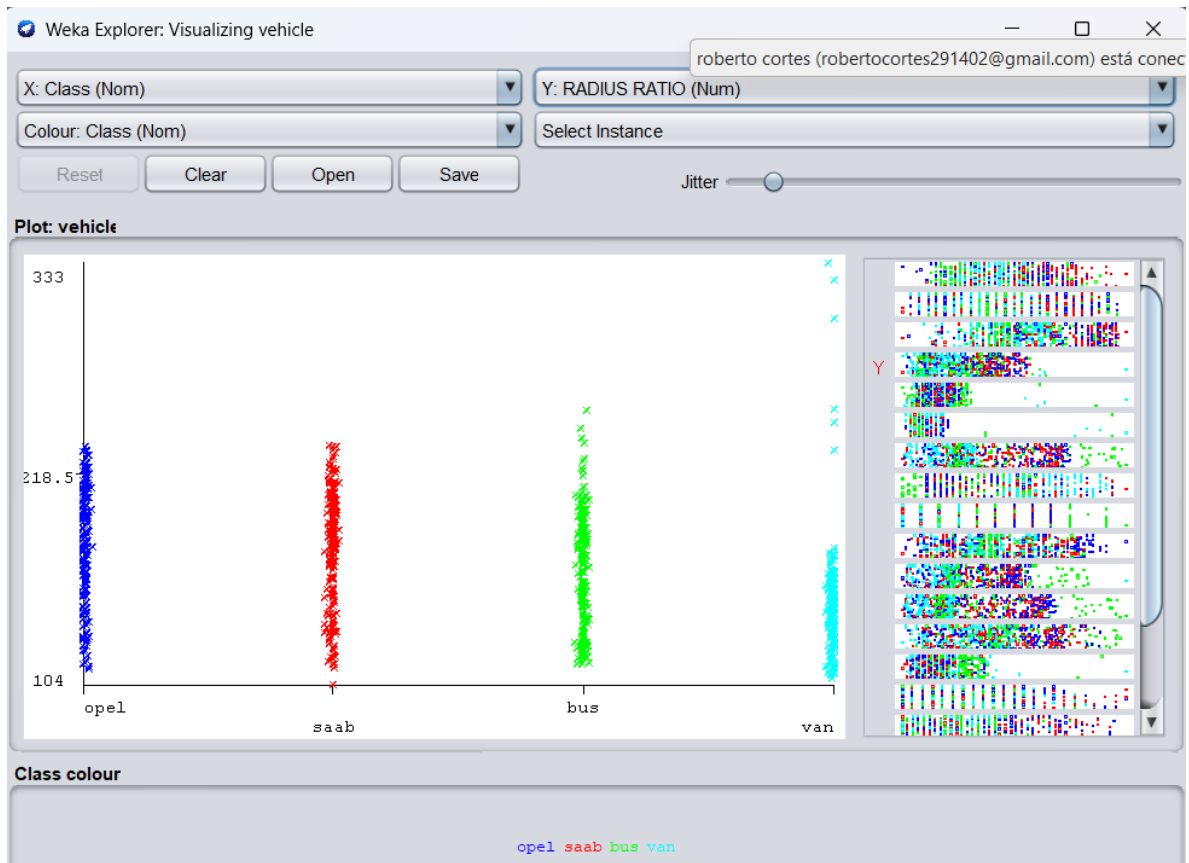


Figura 54. Gráfico de dispersión. Promedio del atributo “Radius Ratio” para el conjunto de vehículos de cada clase.

- El “Pr. Axis Aspect Ratio” tiene casi la misma distribución para automóviles, vans y autobuses.

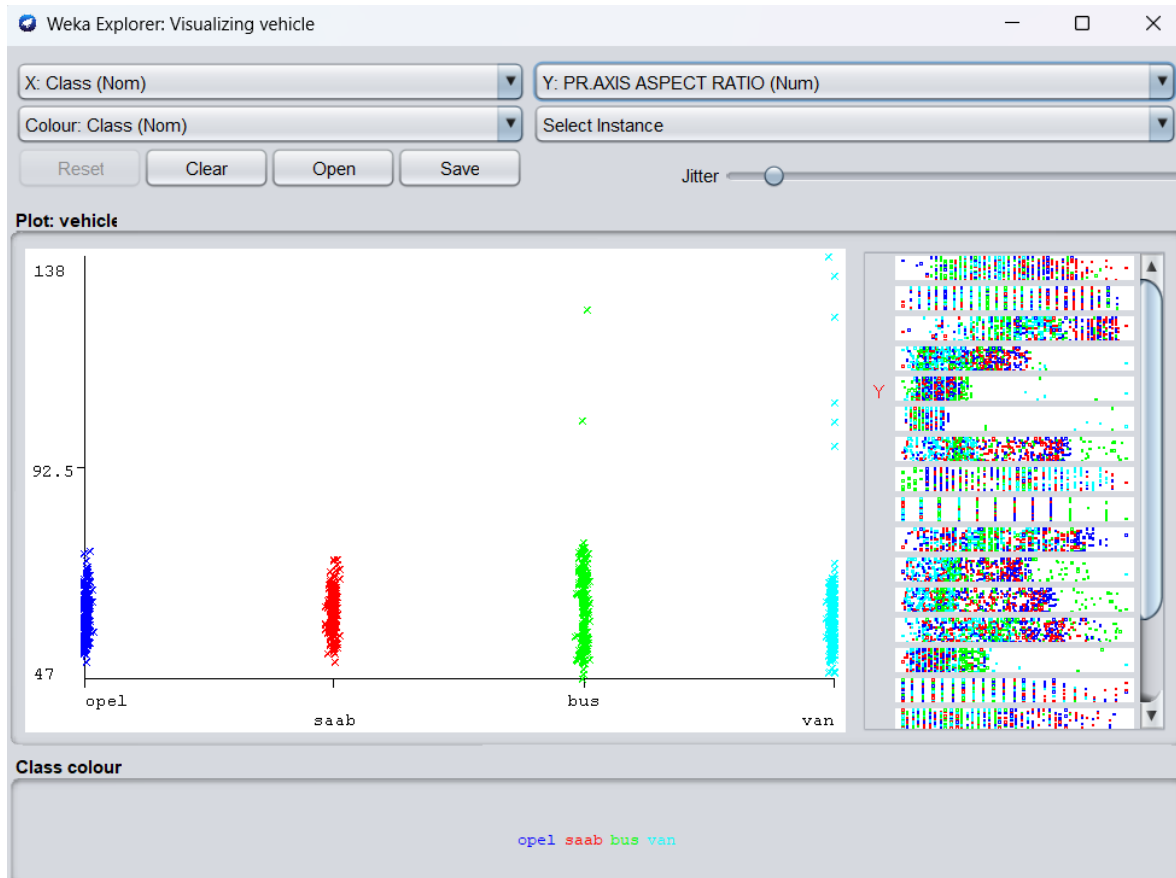


Figura 55. Gráfico de dispersión. Distribución de cada conjunto de clase para el atributo “Pr. Axis Aspect”

- El “Max. Lenght Aspect Ratio” es casi igual para automóviles y vans, pero es menor para autobuses.

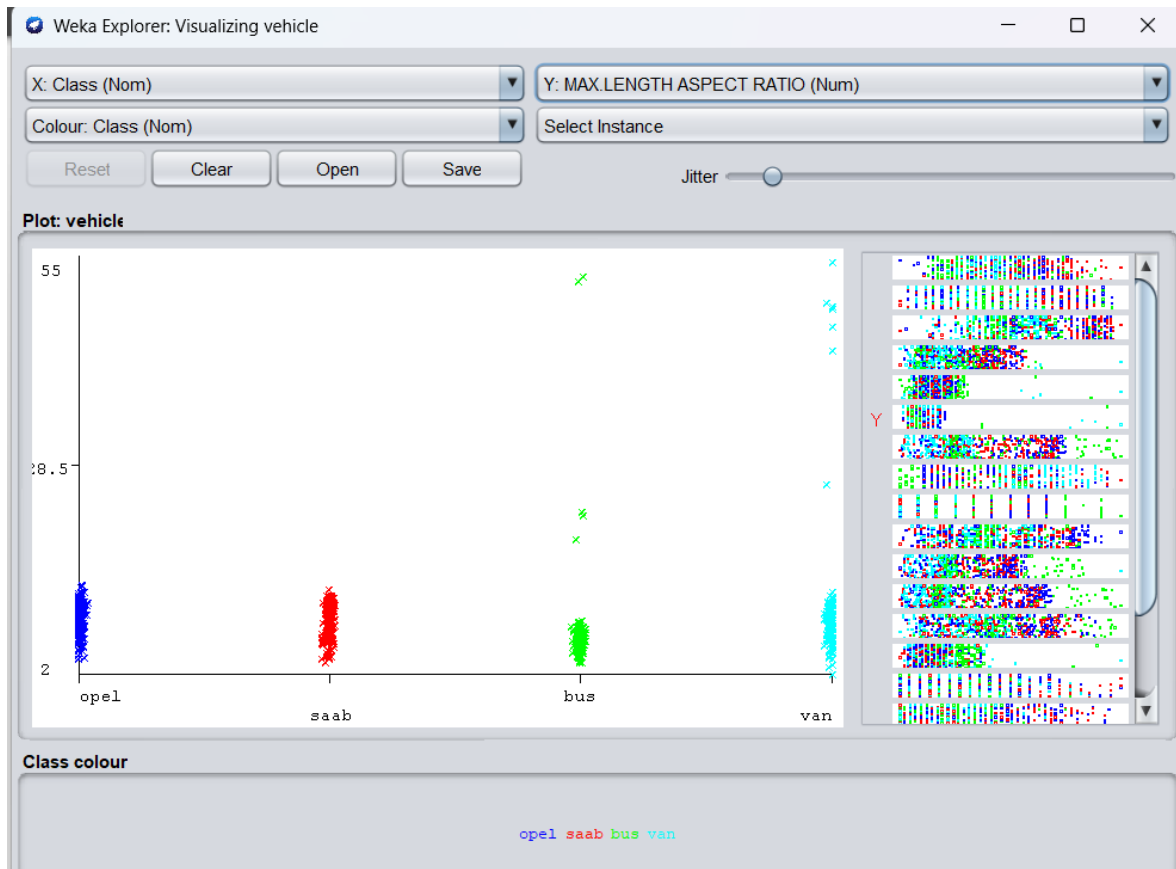


Figura 56. Gráfico de dispersión. Distribución de cada conjunto de clase para el atributo “Max. Lenght Aspect Ratio”

- El “Scatter Ratio” promedio es más alta para los automóviles, seguidos de los autobuses y las vans.

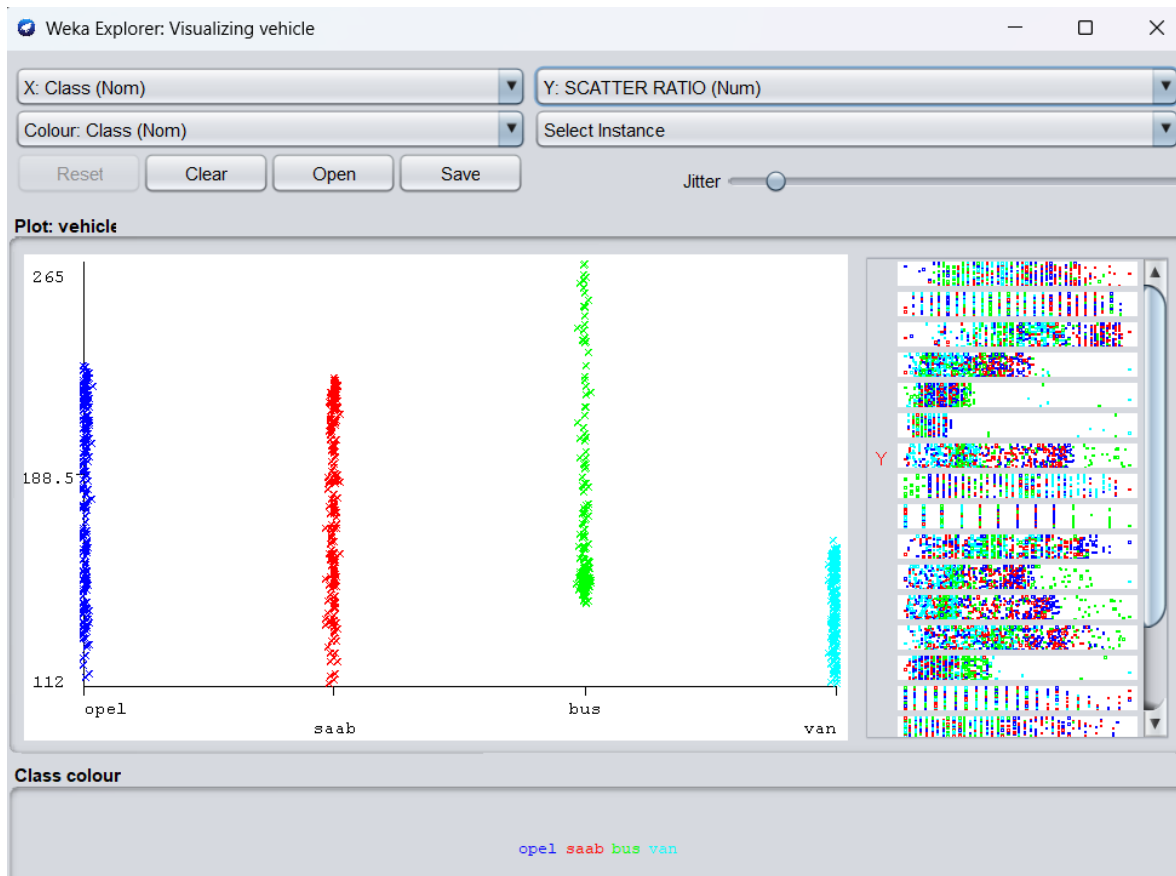


Figura 57. Gráfico de dispersión. Promedio del atributo “Scatter Ratio” para el conjunto de vehículos de cada clase.

- El “Elongatedness” promedio es más alta para vans, le prosigue autobuses y por ultimo los automóviles.

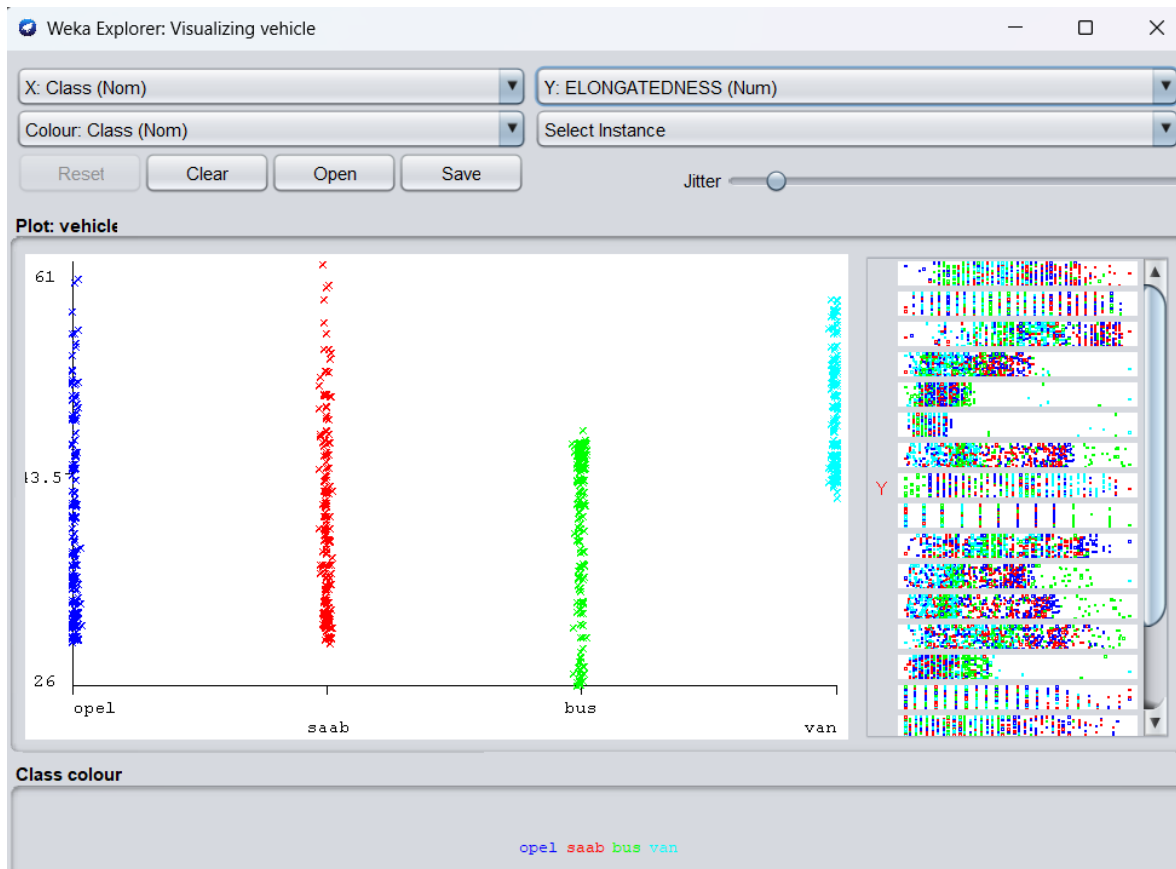


Figura 58. Gráfico de dispersión. Promedio del atributo “Elongatedness” para el conjunto de vehículos de cada clase.

- La distribución de “MAX. Lenght Rectangularity” es casi la misma para automóviles, autobuses y vans.

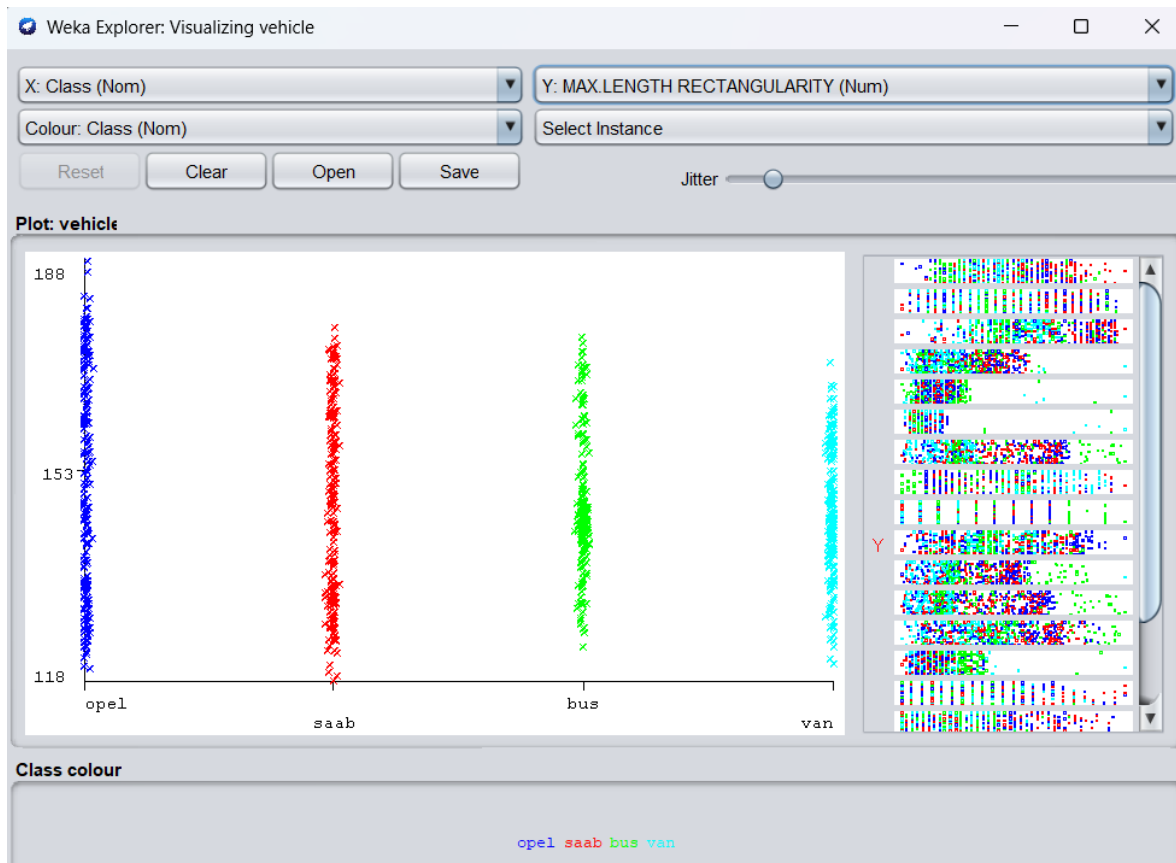


Figura 59. Gráfico de dispersión. Distribución de cada conjunto de clase para el atributo “Max. Lenght Rectangularity”

- La relación de “Scatter Ratio” y “Scaled Variance Minor” tiene una correlación lineal positiva casi perfecta.

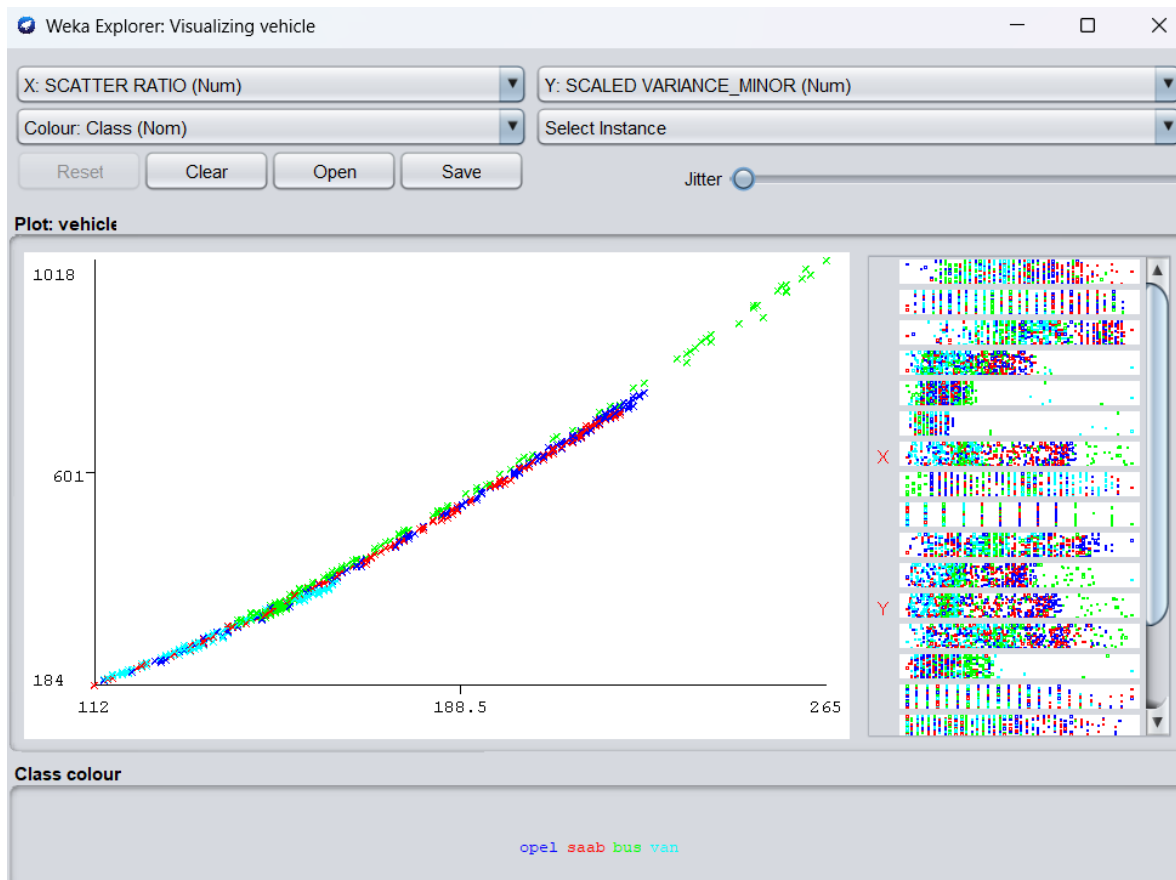


Figura 60. Gráfico de dispersión. Relación entre los atributos “Scatter Ratio” y “Scaled Variance Minor”

CONCLUSIONES

Ian Beckford: Este examen semestral ha sido una experiencia enriquecedora que nos ha permitido aplicar nuestros conocimientos teóricos en situaciones prácticas estudiadas en el mundo real. Gracias a este semestral y documento, nos ha proporcionado una base sólida en Gestión de la Información y análisis de datos, lo que nos permitirá tomar decisiones informadas y estratégicas, y contribuir al éxito y sostenibilidad de las organizaciones en un mundo en constante evolución, esto contando sobre los datos interactivos del mundo real. Al concluir este semestre, estamos preparados con bases sólidas y con las habilidades y el conocimiento necesario para enfrentar los desafíos futuros en el ámbito de los análisis de datos. La capacidad de gestionar y analizar datos se ha convertido en una ventaja competitiva esencial en cualquier organización, y estamos listos para aportar nuestros conocimientos en el desarrollo y crecimiento de futuros proyectos empresariales y universitarios

Regino Cornejo: Después de haber realizado varios laboratorios con diferentes herramientas, es posible desarrollar este semestral con conocimientos aprendidos de múltiples formas, teniendo la facilidad de manejar los datos con decisiones que tomamos como mejor forma de analizarlos, decidimos utilizar herramientas de nuestra preferencia como weka y tableau. Al terminar el semestre, nos damos cuenta de la capacidad de analizar información, las dificultades y los logros que podemos obtener, no fue fácil encontrar lo que se necesitaba para conseguir ese análisis específico de cada caso suministrado, pero lo encontramos y supimos como seguir en grupo, sencillamente con tener las bases para el análisis de información es suficiente para continuar y defendernos en este ámbito de la carrera.

Roberto Cortes: Las herramientas como Tableau, Weka y Notebooks de Python son fundamentales para el análisis y la minería de datos en diversos casos de estudio. Tableau destaca por su capacidad de visualización interactiva y presentación de datos, Weka es valioso para el aprendizaje automático y análisis predictivo, y los Notebooks de Python ofrecen flexibilidad y poder de programación para tareas complejas. Combinar estas herramientas proporciona un enfoque completo y efectivo para obtener perspectivas valiosas, facilitar la toma de decisiones informadas y realizar análisis profundos en diferentes campos y sectores.

Samuel Nuñez: Para este proyecto final se ha usado diferentes herramientas estudiadas, tales como Weka, Python, y Tableau, para desarrollar el análisis de tres casos de estudios. En el primer caso se trata de predecir un tipo de fármaco (drug) que se debe administrar a un paciente afectado de rinitis alérgica según distintos parámetros/variables. Para el segundo caso se desea analizar la composición de pasajeros y tripulación del Titanic, y también, la relación de supervivencia según diversos factores: género, edad y clase social. Y en el tercer se busca clasificar la silueta dada como uno de los cuatro tipos de vehículos utilizando un conjunto de características extraídas de la silueta.

BIBLIOGRAFÍA

- *Category:Opel manta 400*. (s/f). Wikimedia.org. Recuperado el 20 de julio de 2023, de https://commons.wikimedia.org/wiki/Category:Opel_Manta_400
- Wikipedia contributors. (s/f). *Saab 9000*. Wikipedia, The Free Encyclopedia. https://es.wikipedia.org/w/index.php?title=Saab_9000&oldid=132543905
- *Inicio*. (s/f). Chevrolet Panamá. Recuperado el 20 de julio de 2023, de <https://www.chevrolet.com.pa/comerciales/express-passenger>
- Portafolio. (s/f). *Volvo y Super Polo lanzaron primer bus de dos pisos*. Portafolio.co. Recuperado el 20 de julio de 2023, de <https://www.portafolio.co/negocios/empresas/volvo-super-polo-lanzaron-primer-bus-dos-pisos-101076>
- *Buscar la correlación de Pearson*. (s/f). Tableau.com. Recuperado el 20 de julio de 2023, de <https://kb.tableau.com/articles/howto/finding-the-pearson-correlation?lang=es-es>
- de Tejada, P. S. (2020, abril 5). *¿Cómo identificar y filtrar outliers (valores atípicos) en Tableau?* The Information Lab; The Information Lab Spain. <https://www.theinformationlab.es/como-identificar-y-filtrar-outliers-valores-atipicos-en-tableau/>
- *Muestre rápidamente valores atípicos y patrones sospechosos en sus datos*. (s/f). Tableau. Recuperado el 20 de julio de 2023, de <https://www.tableau.com/es-es/solutions/workbook/quickly-show-outliers-and-suspicious-patterns-your-data>