

Carrera:

Ingeniería en Software

Asignatura:

Minería de Datos

Proyecto 1. Conociendo tus datos

Alumno:

Brandon Iván Márquez Morales - 220030140

Profesor:

Dr. Sergio Valadez Godínez

Fecha de entrega: 19/06/2021

Instrucciones:

A partir de un caso práctico sobre minería de datos, elaborar un reporte que incluya:

- a) Descripción de la base de datos
- b) Tipos de atributos
- c) Medidas de tendencia central
- d) Medidas de dispersión
- e) Diagramas de caja y valores atípicos
- f) Regresión lineal
- g) Análisis de Componentes Principales (PCA)
- h) Reglas de separación de patrones
- i) Conclusiones
- j) Bibliografía

a) Descripción de la base de datos

La base de datos "Identificación de Vidrios" es un conjunto de datos multivariante en el área temática de física, específicamente en la identificación de diferentes tipos de vidrio. Este conjunto de datos ha sido recopilado por el Servicio de Ciencias Forenses de los Estados Unidos y consta de 214 instancias.

El conjunto de datos contiene un total de 9 atributos para cada instancia. Estos atributos son de tipo real y se utilizan para describir las características de los diferentes tipos de vidrio en función de su contenido en óxidos, como el sodio (Na), hierro (Fe), potasio (K), entre otros. Los atributos proporcionan información cuantitativa sobre la composición química de los vidrios.

El objetivo principal de este conjunto de datos es la clasificación de los diferentes tipos de vidrio. Se proporcionan 6 tipos de vidrio distintos, cada uno de los cuales ha sido definido en función de su contenido en óxidos. Utilizando las características proporcionadas por los atributos, se pueden aplicar técnicas de clasificación para identificar y distinguir los diferentes tipos de vidrio presentes en la base de datos.

b) Tipos de atributos

Los atributos presentes en la base de datos "Identificación de Vidrios" se describen de la siguiente manera:

Id number: Un número de identificación único asignado a cada instancia de vidrio en el rango de 1 a 214.

RI: Índice de refracción. Es una medida que indica la velocidad a la que la luz se propaga a través del vidrio. Se trata de un atributo numérico continuo.

Na: Sodio. Representa el porcentaje en peso del sodio presente en el vidrio en forma de óxido de sodio (Na2O). Es un atributo numérico continuo.

Mg: Magnesio. Indica el porcentaje en peso de magnesio en el vidrio, en forma de óxido de magnesio (MgO). Es un atributo numérico continuo.

Al: Aluminio. Muestra el porcentaje en peso de aluminio en el vidrio, en forma de óxido de aluminio (Al2O3). Es un atributo numérico continuo.

Si: Silicio. Indica el porcentaje en peso de silicio en el vidrio, en forma de óxido de silicio (SiO2). Es un atributo numérico continuo.

K: Potasio. Representa el porcentaje en peso de potasio presente en el vidrio, en forma de óxido de potasio (K2O). Es un atributo numérico continuo.

Ca: Calcio. Muestra el porcentaje en peso de calcio en el vidrio, en forma de óxido de calcio (CaO). Es un atributo numérico continuo.

Ba: Bario. Indica el porcentaje en peso de bario en el vidrio, en forma de óxido de bario (BaO). Es un atributo numérico continuo.

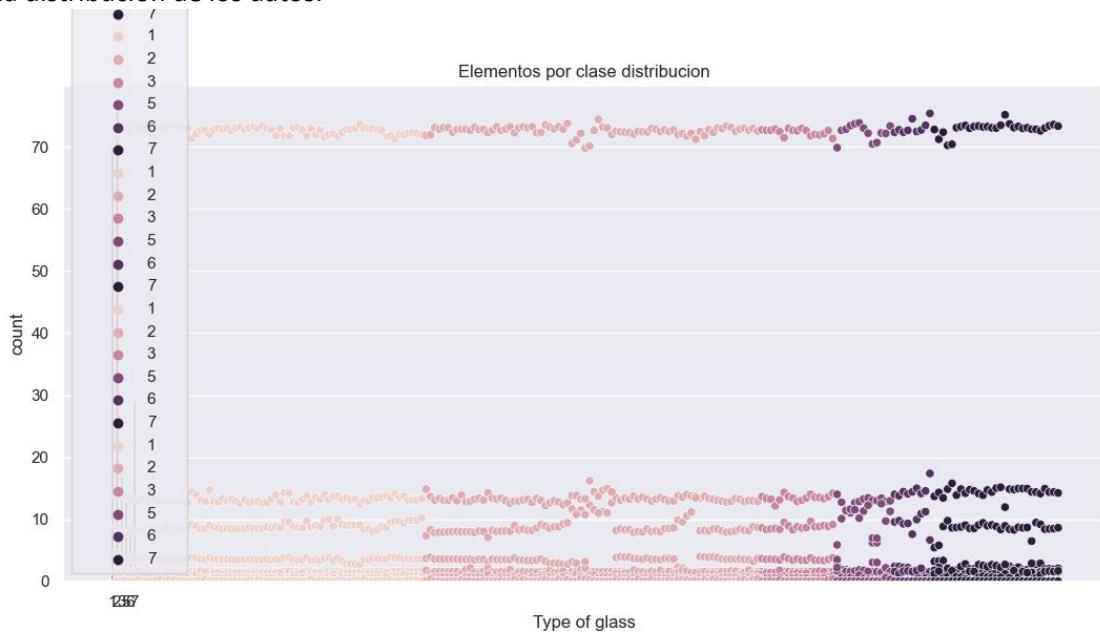
Fe: Hierro. Representa el porcentaje en peso de hierro presente en el vidrio, en forma de óxido de hierro (Fe2O3). Es un atributo numérico continuo.

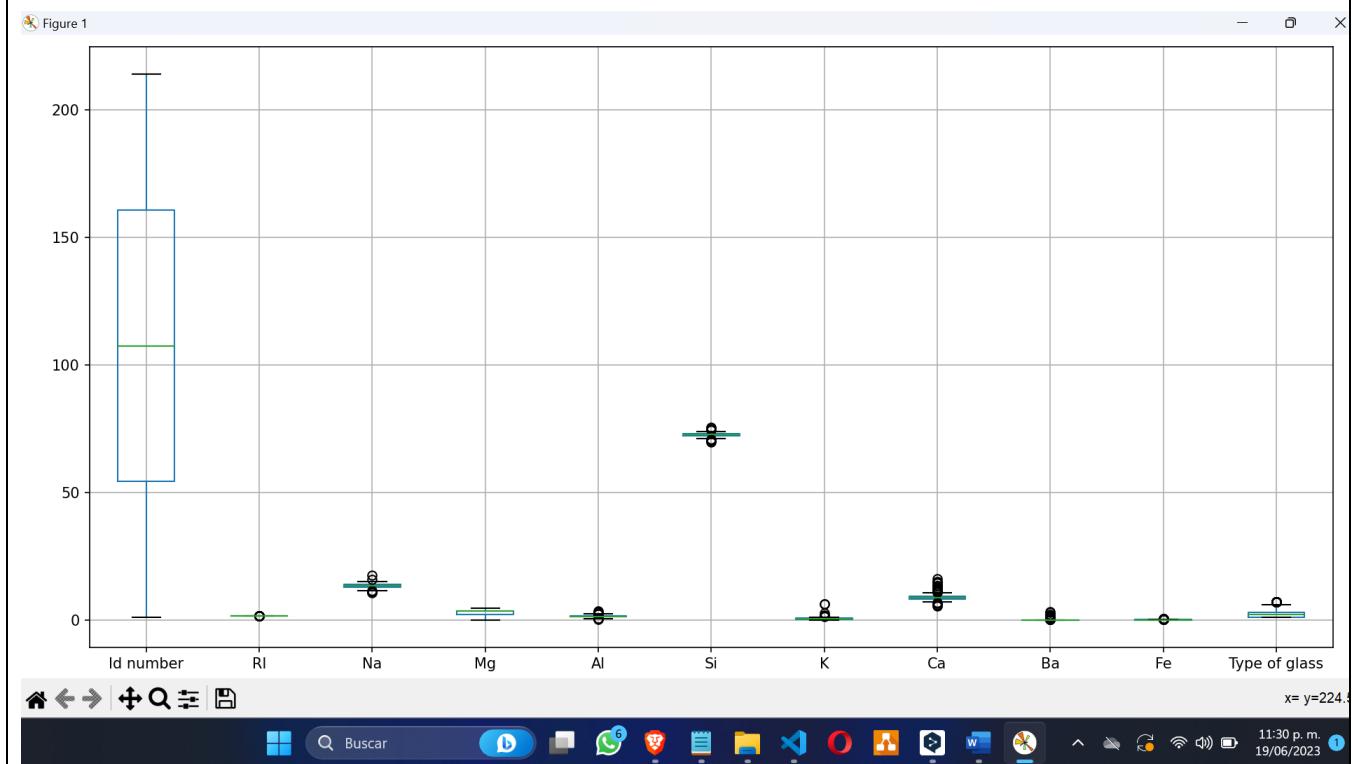
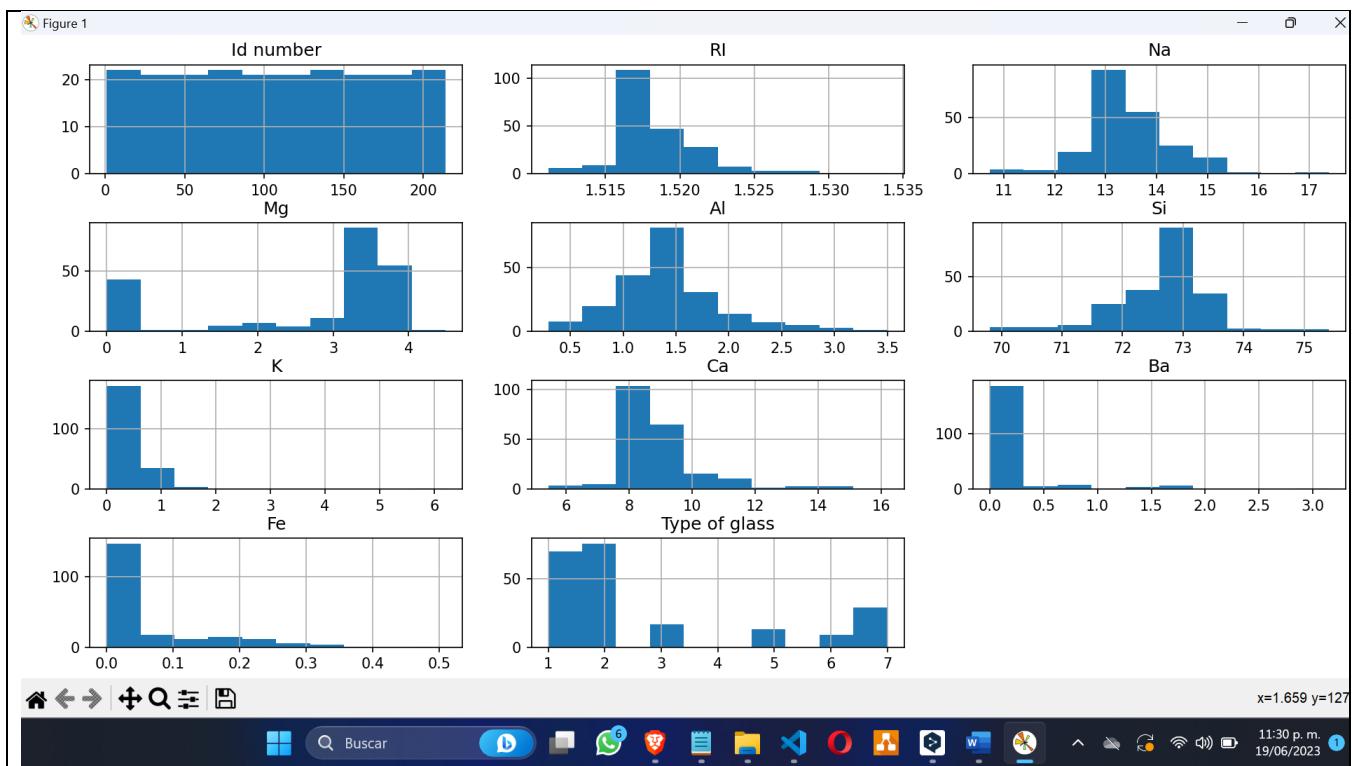
Type of glass: (atributo de clase) Representa el tipo de vidrio y se utiliza como atributo objetivo en la clasificación. Tiene los siguientes valores categóricos:

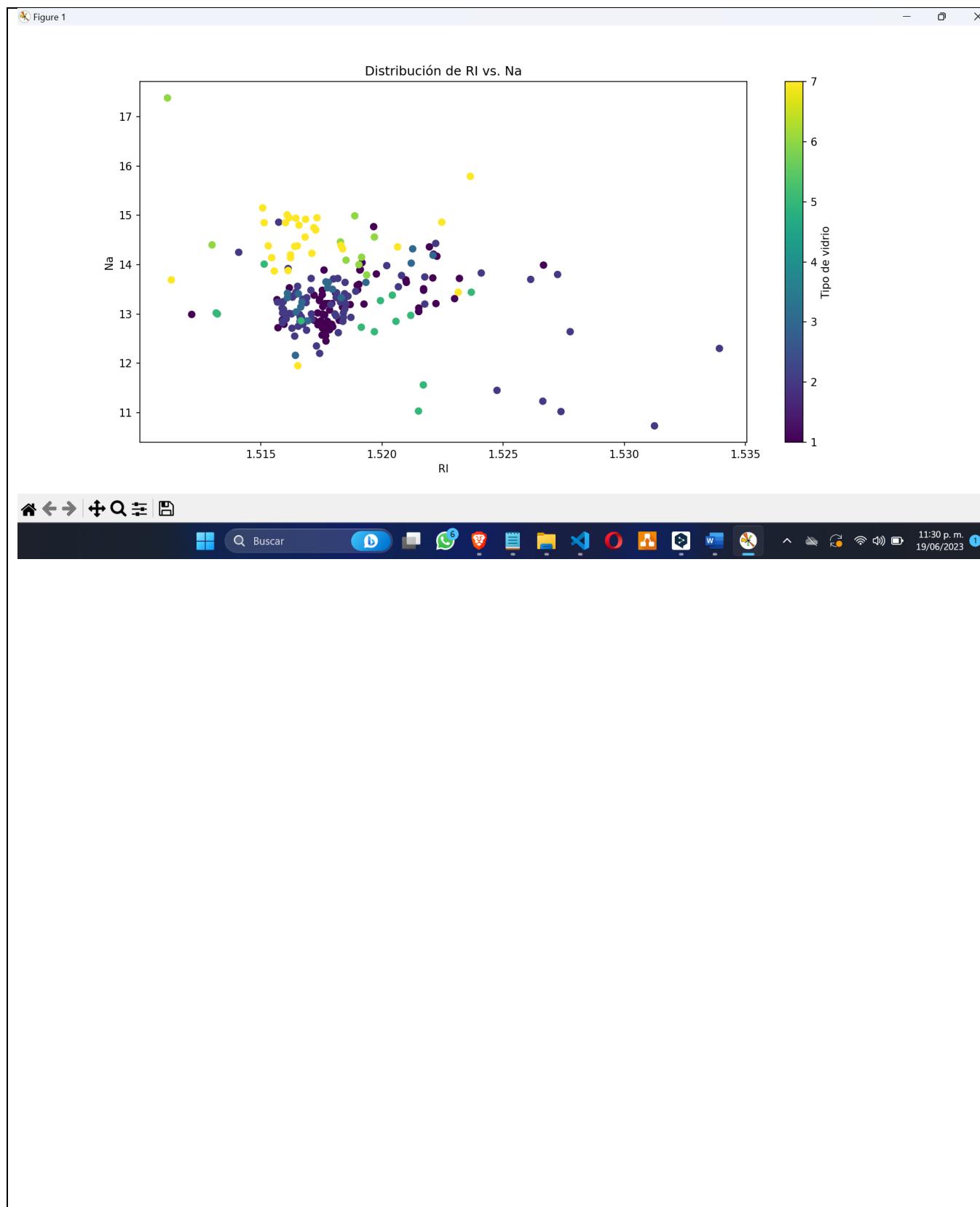
- 1: Building Windows Float Processed (Ventanas de construcción procesadas por flotación).
- 2: Building Windows Non-Float Processed (Ventanas de construcción no procesadas por flotación).
- 3: Vehicle Windows Float Processed (Ventanas de vehículos procesadas por flotación).
- 4: Vehicle Windows Non-Float Processed (Ventanas de vehículos no procesadas por flotación, no presente en esta base de datos).
- 5: Containers (Contenedores).
- 6: Tableware (Vajilla).
- 7: Headlamps (Faros).

Estos atributos proporcionan información sobre la composición química y las propiedades ópticas de los diferentes tipos de vidrio, y el atributo "Type of glass" es el objetivo de clasificación que indica la categoría a la que pertenece cada instancia de vidrio en particular.

Graficar la distribución de los datos.



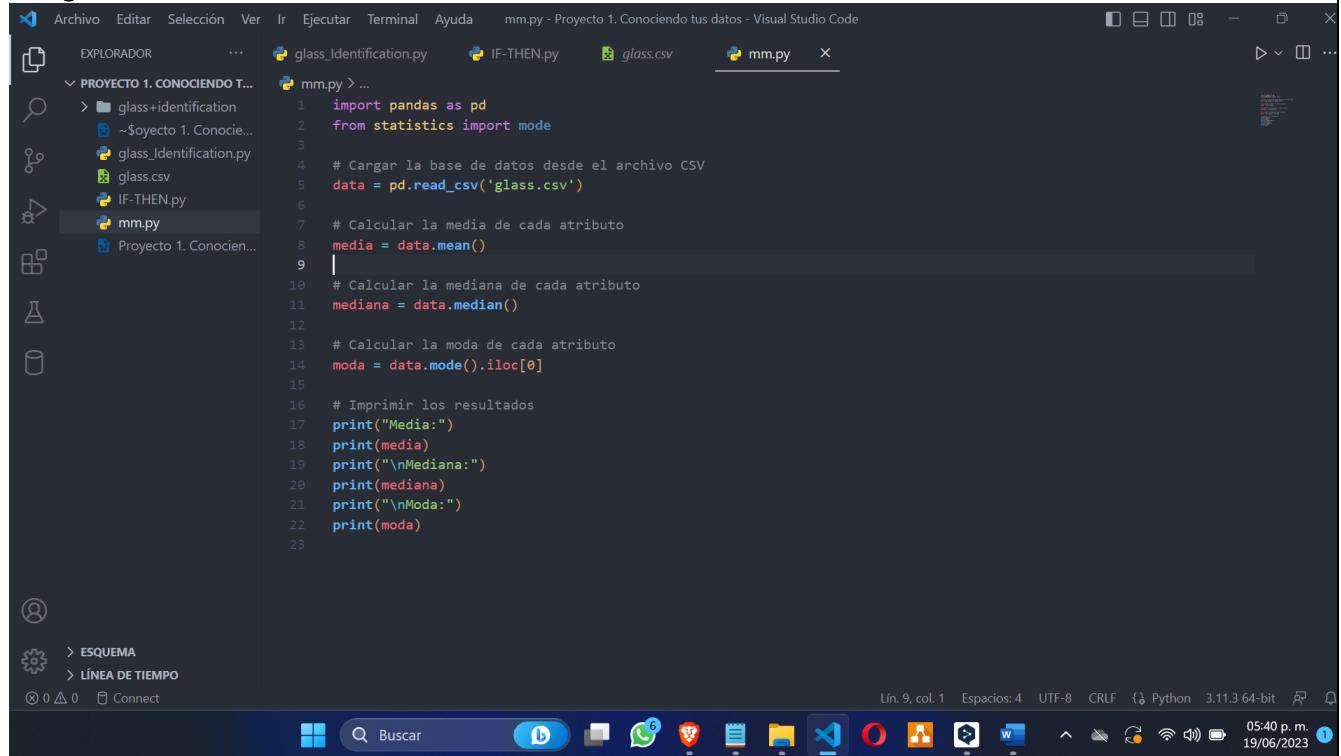




c) Medidas de tendencia central

Calcular la media, mediana y moda para cada uno de los atributos de la base de datos.
 Poner código fuente del cálculo y el resultado.

Código fuente:



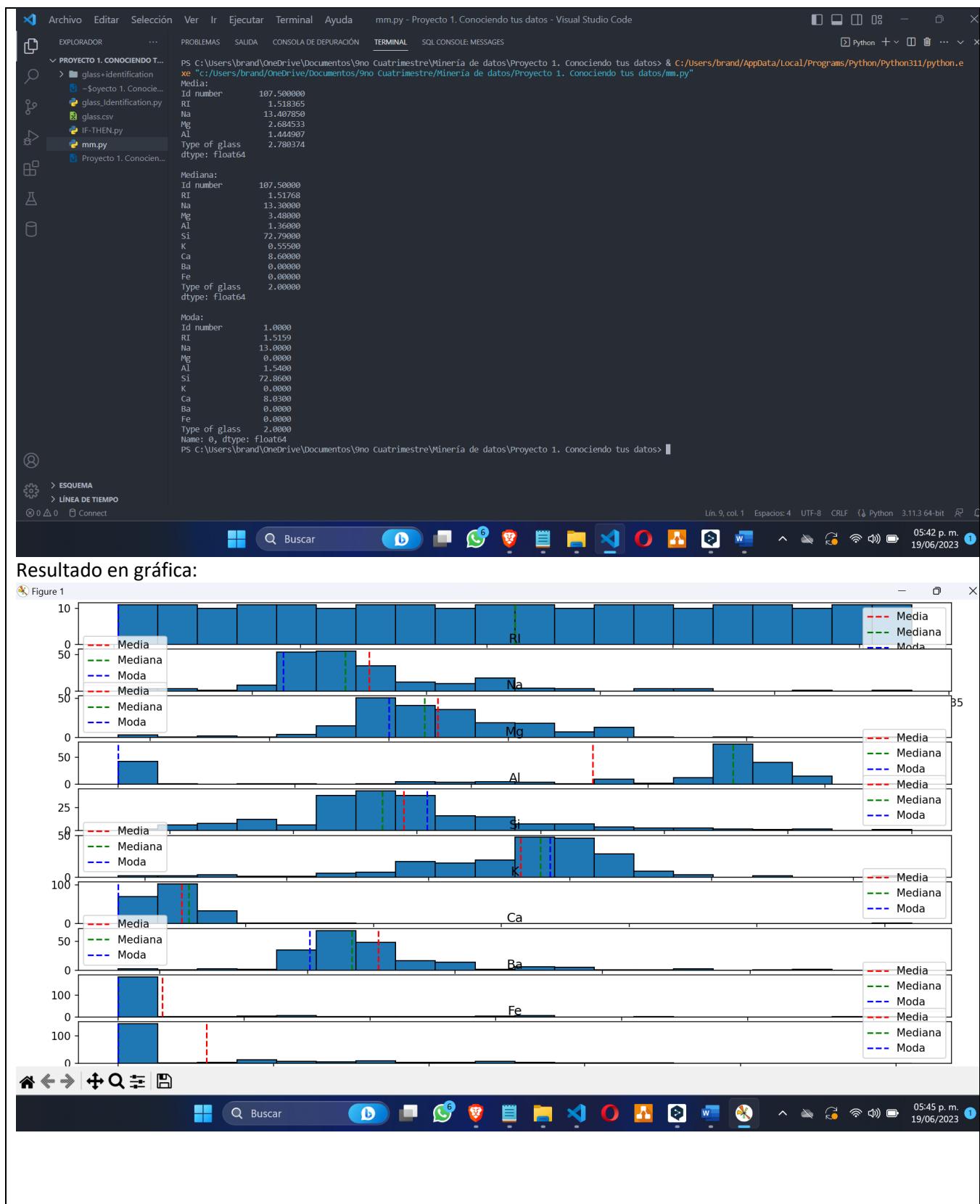
```

Archivo Editar Selección Ver Ir Ejecutar Terminal Ayuda mm.py - Proyecto 1. Conociendo tus datos - Visual Studio Code

EXPLORADOR ...
PROYECTO 1. CONOCIENDO T...
  > glass+identification
  ~$oyecto 1. Conocie...
  glass_Identification.py
  glass.csv
  IF-THEN.py
  mm.py
  Proyecto 1. Conocien...
mm.py ...
mm.py ...
1 import pandas as pd
2 from statistics import mode
3
4 # Cargar la base de datos desde el archivo CSV
5 data = pd.read_csv('glass.csv')
6
7 # Calcular la media de cada atributo
8 media = data.mean()
9
10 # Calcular la mediana de cada atributo
11 mediana = data.median()
12
13 # Calcular la moda de cada atributo
14 moda = data.mode().iloc[0]
15
16 # Imprimir los resultados
17 print("Media:")
18 print(media)
19 print("\nMediana:")
20 print(mediana)
21 print("\nModa:")
22 print(moda)
23
  
```

Lín. 9, col. 1 Espacios:4 UTF-8 CRLF Python 3.11.3 64-bit 05:40 p. m. 19/06/2023

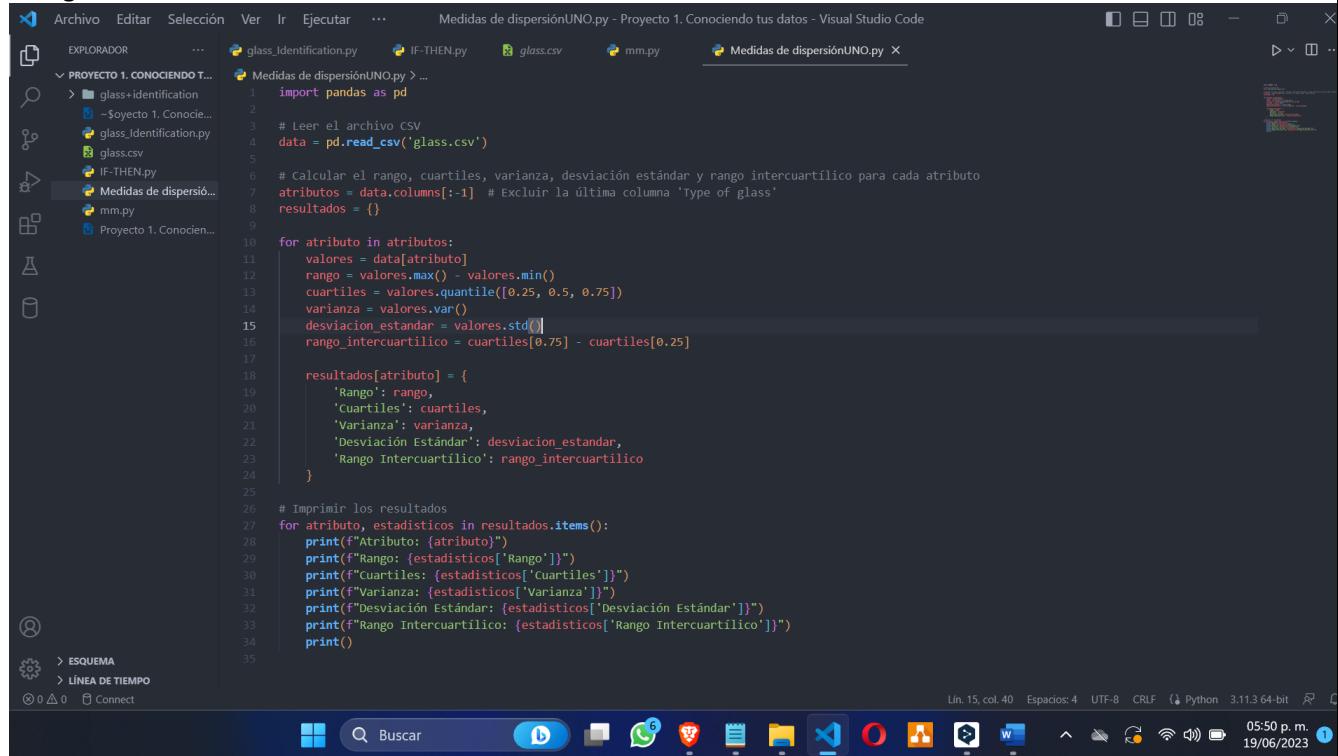
Resultado en consola:



d) Medidas de dispersión

Calcular el rango, los cuartiles, la varianza, la desviación estándar y el rango intercuartílico para cada uno de los atributos de la base de datos.

Código fuente:



```

Archivo Editar Selección Ver Ir Ejecutar ... Medidas de dispersiónUNO.py - Proyecto 1. Conociendo tus datos - Visual Studio Code
EXPLORADOR ...
PROYECTO 1. CONOCIENDO T...
glass+identification IF-THEN.py glass.csv mm.py Medidas de dispersiónUNO.py ...
glass_Identification.py ...
Medidas de dispersiónUNO.py ...
1 import pandas as pd
2
3 # Leer el archivo CSV
4 data = pd.read_csv('glass.csv')
5
6 # Calcular el rango, cuartiles, varianza, desviación estándar y rango intercuartílico para cada atributo
7 atributos = data.columns[:-1] # Excluir la última columna 'Type of glass'
8 resultados = {}
9
10 for atributo in atributos:
11     valores = data[atributo]
12     rango = valores.max() - valores.min()
13     cuartiles = valores.quantile([0.25, 0.5, 0.75])
14     varianza = valores.var()
15     desviacion_estandar = valores.std()
16     rango_intercuartilico = cuartiles[0.75] - cuartiles[0.25]
17
18     resultados[atributo] = {
19         'Rango': rango,
20         'Cuartiles': cuartiles,
21         'Varianza': varianza,
22         'Desviación Estándar': desviacion_estandar,
23         'Rango Intercuartílico': rango_intercuartilico
24     }
25
26 # Imprimir los resultados
27 for atributo, estadisticos in resultados.items():
28     print(f"Atributo: {atributo}")
29     print(f"Rango: {estadisticos['Rango'])")
30     print(f"Cuartiles: {estadisticos['Cuartiles'])")
31     print(f"Varianza: {estadisticos['Varianza'])")
32     print(f"Desviación Estándar: {estadisticos['Desviación Estándar'])")
33     print(f"Rango Intercuartílico: {estadisticos['Rango Intercuartílico'])")
34     print()
35
  
```

Lín. 15, col. 40 Espacios: 4 UTF-8 CRLF Python 3.11.3 64-bit 05:50 p. m. 19/06/2023

Resultado en consola:

Archivo Editar Selección Ver Ir Ejecutar ... Medidas de dispersiónUNO.py - Proyecto 1. Conociendo tus datos - Visual Studio Code

EXPLORADOR PROBLEMAS SALIDA CONSOLA DE DEPURACIÓN TERMINAL SQL CONSOLE: MESSAGES

> PROYECTO 1. CONOCENDO T... glass+identification ~\$oyecto 1. Conocie... glass_Identification.py glass.csv IF-THEN.py Medidas de dispersió... mm.py Proyecto 1. Conocien...

```
PS C:\Users\brand\OneDrive\Documents\9mo Cuatrimestre\Minería de datos\Proyecto 1. Conociendo tus datos> & C:/Users/brand/AppData/Local/Programs/Python/Python311/python.exe "c:/Users/brand/OneDrive/Documentos/9mo Cuatrimestre/Mineria de datos/Proyecto 1. Conociendo tus datos/~/Medidas de dispersiónUNO.py"
Atributo: Id number
Rango: 213
Cuartiles: 0.25 54.25
0.50 107.90
0.75 150.50
Name: Id, dtype: float64
Varianza: 3834.1666666666665
Desviación Estándar: 61.92064814475594
Rango Intercuartílico: 106.5

Atributo: RI
Rango: 0.022780000000000022
Cuartiles: 0.25 1.516522
0.50 1.530000
0.75 1.530157
Name: RI, dtype: float64
Varianza: 9.222541371594076e-06
Desviación Estándar: 0.0030368637393854334
Rango Intercuartílico: 0.002634999999999943

Atributo: Na
Rango: 6.649999999999999
Cuartiles: 0.25 12.9075
0.50 13.0000
0.75 13.8250
Name: Na, dtype: float64
Varianza: 0.6668413672063533
Desviación Estándar: 0.81660355714983
Rango Intercuartílico: 0.9174999999999986

Atributo: Mg
Rango: 4.49
Cuartiles: 0.25 2.115
0.50 3.480
0.75 4.680
Name: Mg, dtype: float64
Varianza: 2.08954039943793
Desviación Estándar: 1.442407844870442
Rango Intercuartílico: 1.4850000000000003

Atributo: Al
Rango: 3.21
Cuartiles: 0.25 1.19
0.50 3.36
0.75 1.63
Name: Al, dtype: float64
Varianza: 0.24927017901803336

ESQUEMA LÍNEA DE TIEMPO
```

Lín. 15, col. 40 Espacios: 4 UTF-8 CRLF ↵ Python 3.11.3 64-bit 05:51 p. m. 19/06/2023

Archivo Editar Selección Ver Ir Ejecutar ... Medidas de dispersiónUNO.py - Proyecto 1. Conociendo tus datos - Visual Studio Code

EXPLORADOR PROBLEMAS SALIDA CONSOLA DE DEPURACIÓN TERMINAL SQL CONSOLE: MESSAGES

> PROYECTO 1. CONOCENDO T... glass+identification ~\$oyecto 1. Conocie... glass_Identification.py glass.csv IF-THEN.py Medidas de dispersió... mm.py Proyecto 1. Conocien...

```
Atributo: Si
Rango: 5.599999999999994
Cuartiles: 0.25 72.2800
0.50 72.7900
0.75 73.75
Name: Si, dtype: float64
Varianza: 0.5999211881883197
Desviación Estándar: 0.774547947651124
Rango Intercuartílico: 0.8075000000000045

Atributo: K
Rango: 6.21
Atributo: Ca
Rango: 10.760000000000002
Cuartiles: 0.25 8.2400
0.50 8.6000
0.75 9.1725
Name: Ca, dtype: float64
Varianza: 2.0253658483611954
Desviación Estándar: 1.42153487281395
Rango Intercuartílico: 0.9324999999999992

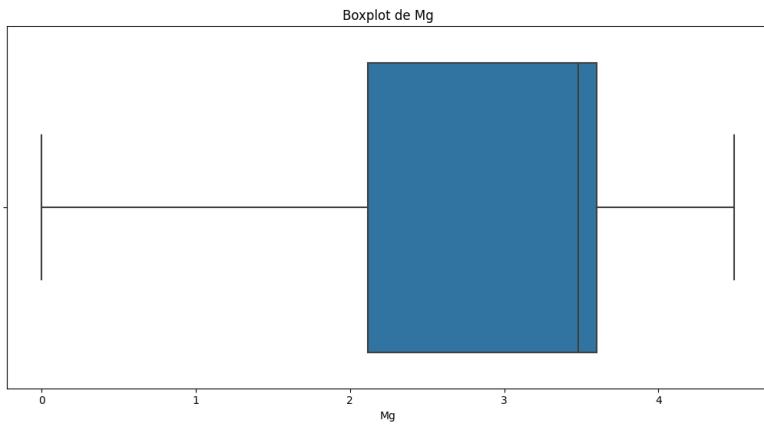
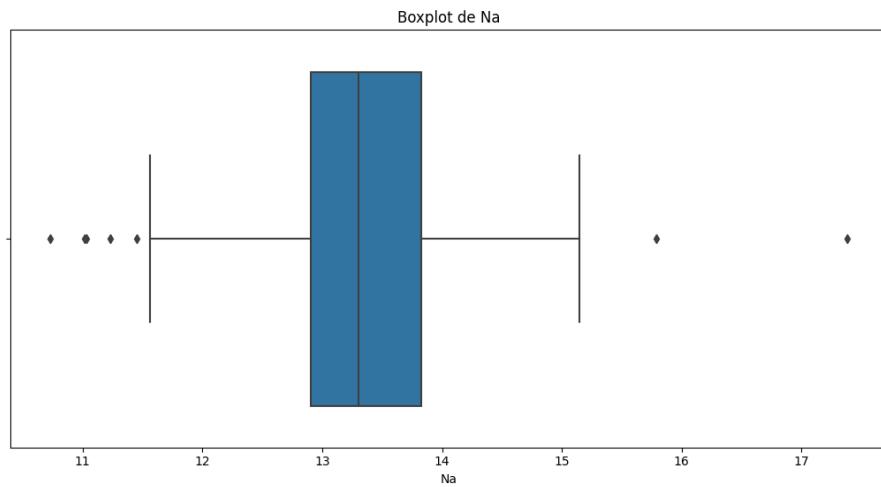
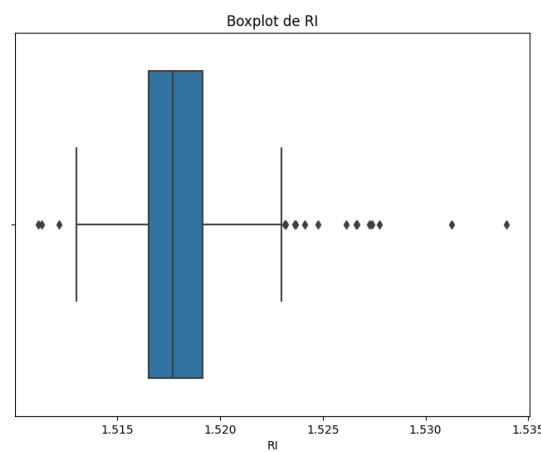
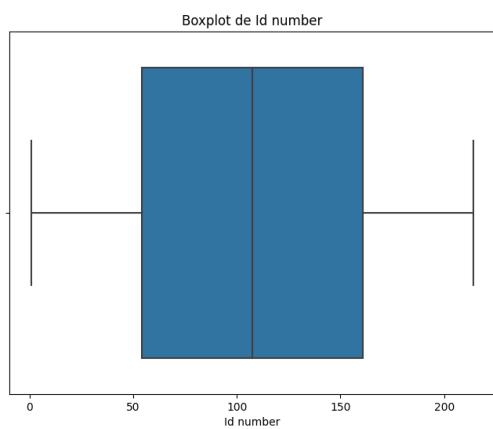
Atributo: Ba
Rango: 3.15
Cuartiles: 0.25 0.0
0.50 0.0
0.75 0.0
Name: Ba, dtype: float64
Varianza: 0.2472269911131159
Desviación Estándar: 0.49721926059970356
Rango Intercuartílico: 0.0

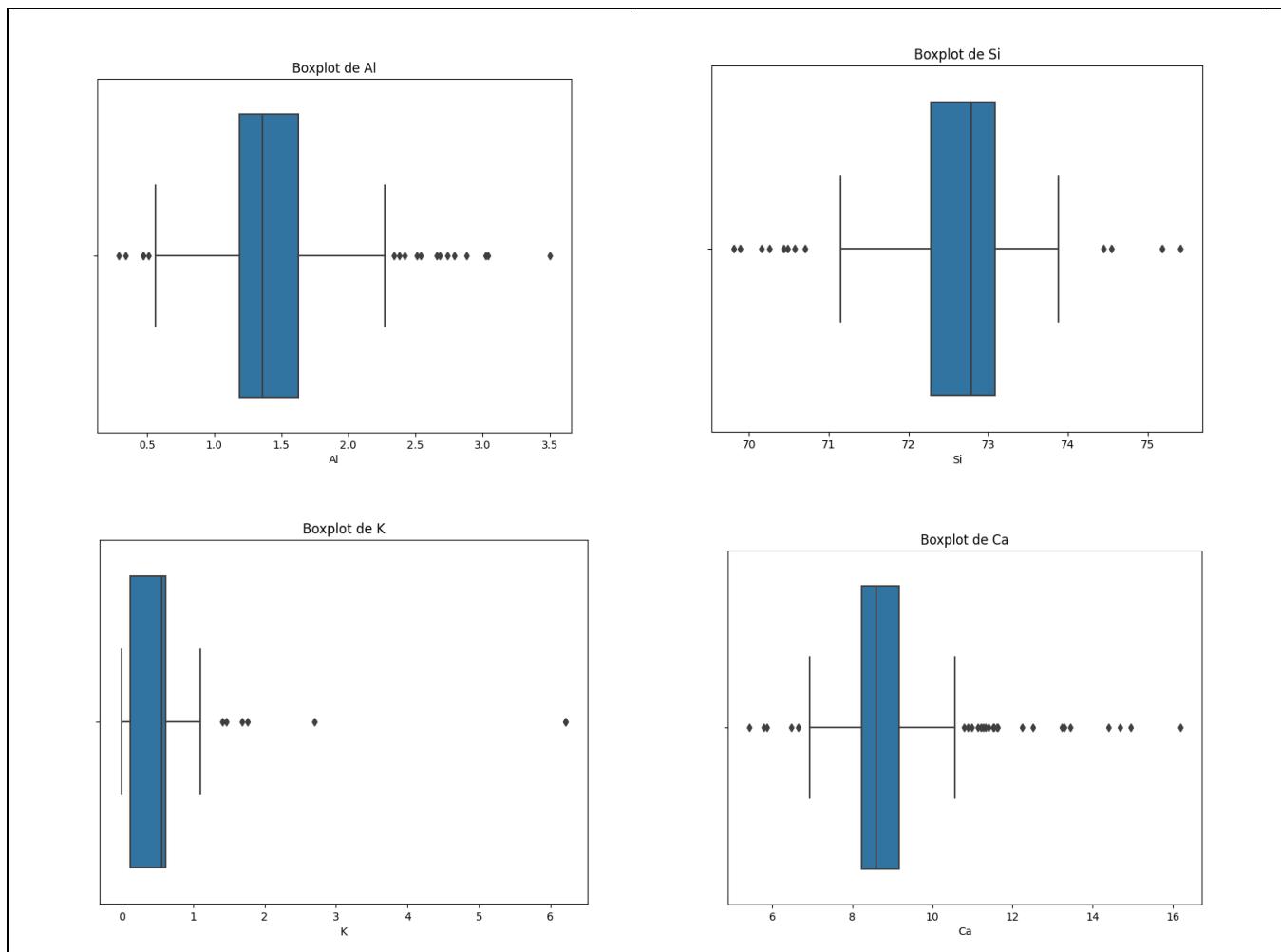
Atributo: Fe
Rango: 0.51
Cuartiles: 0.25 0.0
0.50 0.0
0.75 0.1
Name: Fe, dtype: float64
Varianza: 0.00949430038172963
Desviación Estándar: 0.09743670063650084
Rango Intercuartílico: 0.1

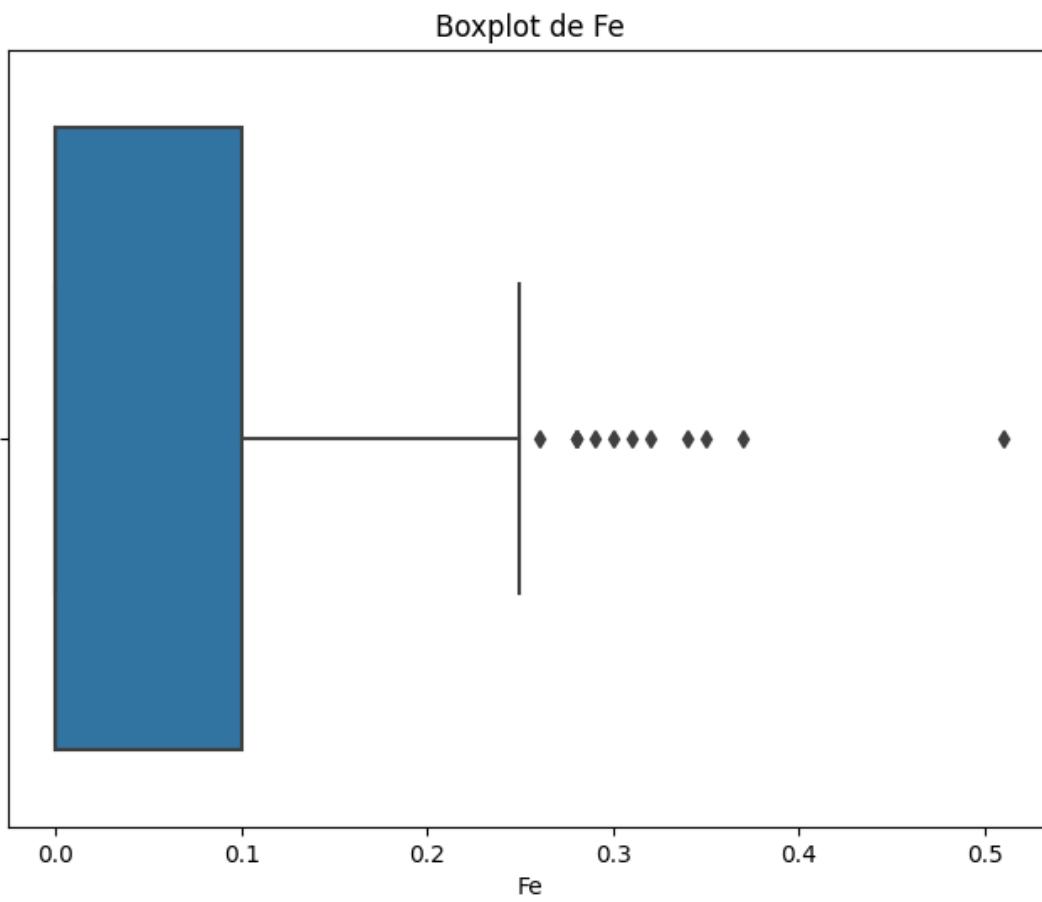
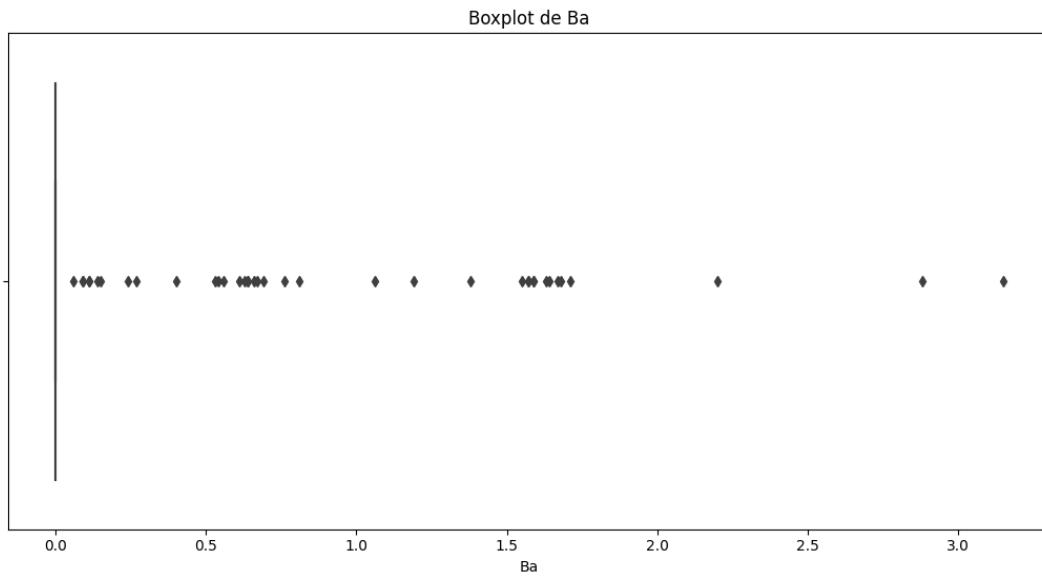
PS C:\Users\brand\OneDrive\Documents\9mo Cuatrimestre\Minería de datos\Proyecto 1. Conociendo tus datos>
```

Lín. 15, col. 40 Espacios: 4 UTF-8 CRLF ↵ Python 3.11.3 64-bit 05:52 p. m. 19/06/2023

Resultado gráficamente de cada uno de los atributos:

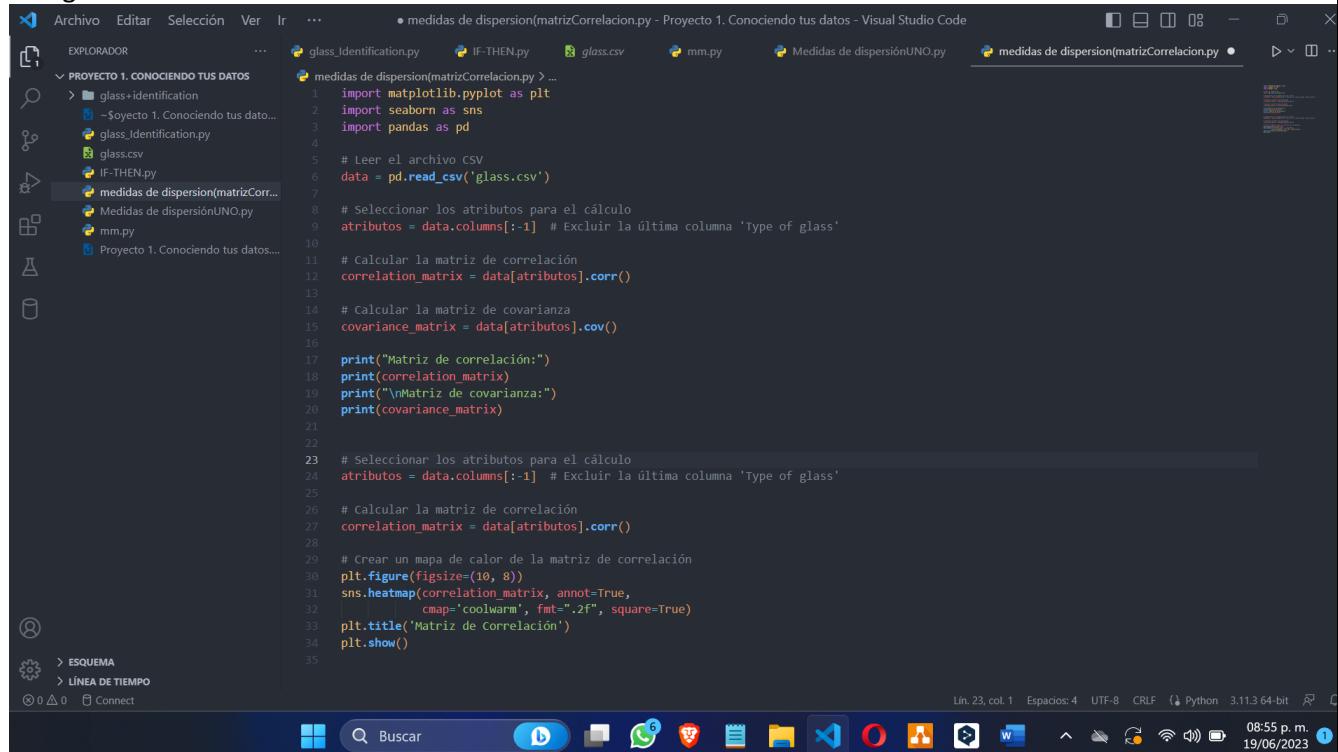






Calcular la matriz de correlación y covarianza de la base de datos.

Código fuente:



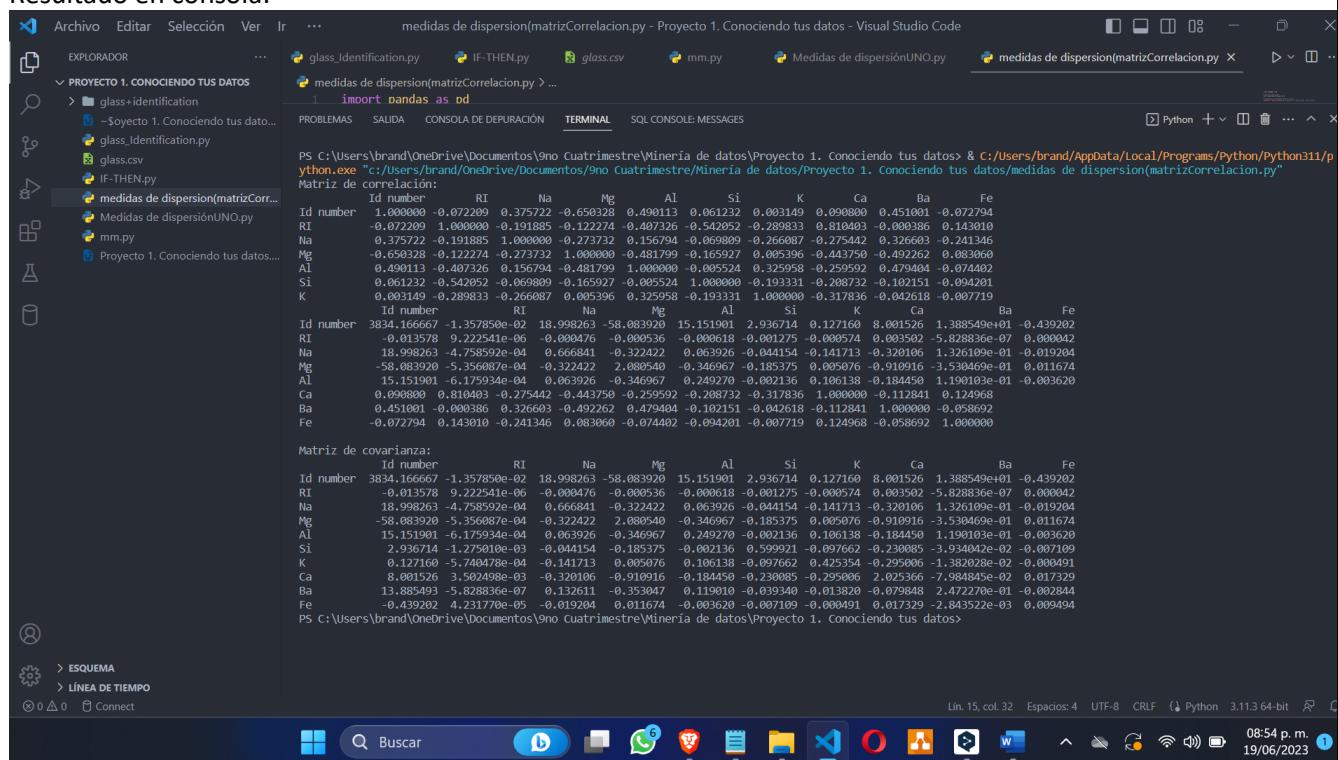
```

Archivo Editar Selección Ver Ir ...
• medidas de dispersion(matrizCorrelacion.py) - Proyecto 1. Conociendo tus datos - Visual Studio Code
EXPLORADOR ... medidas de dispersion(matrizCorrelacion.py) > ...
PROYECTO 1. CONOCIENDO TUS DATOS
> glass+identification
-> $oyecto 1. Conociendo tus dato...
glass_identification.py IF-THEN.py glass.csv mm.py Medidas de dispersiónUNO.py
medidas de dispersion(matrizCorr... medidas de dispersion(matrizCorrelacion.py) ...
1 import matplotlib.pyplot as plt
2 import seaborn as sns
3 import pandas as pd
4
5 # Leer el archivo CSV
6 data = pd.read_csv('glass.csv')
7
8 # Seleccionar los atributos para el cálculo
9 atributos = data.columns[:-1] # Excluir la última columna 'Type of glass'
10
11 # Calcular la matriz de correlación
12 correlation_matrix = data[atributos].corr()
13
14 # Calcular la matriz de covarianza
15 covariance_matrix = data[atributos].cov()
16
17 print("Matriz de correlación:")
18 print(correlation_matrix)
19 print("\nMatriz de covarianza:")
20 print(covariance_matrix)
21
22
23 # Seleccionar los atributos para el cálculo
24 atributos = data.columns[:-1] # Excluir la última columna 'Type of glass'
25
26 # Calcular la matriz de correlación
27 correlation_matrix = data[atributos].corr()
28
29 # Crear un mapa de calor de la matriz de correlación
30 plt.figure(figsize=(10, 8))
31 sns.heatmap(correlation_matrix, annot=True,
32             cmap='coolwarm', fmt=".2f", square=True)
33 plt.title('Matriz de Correlación')
34 plt.show()
35

```

Lín. 23, col. 1 Espacios: 4 UTF-8 CRLF Python 3.11.3 64-bit 08:55 p.m. 19/06/2023

Resultado en consola:



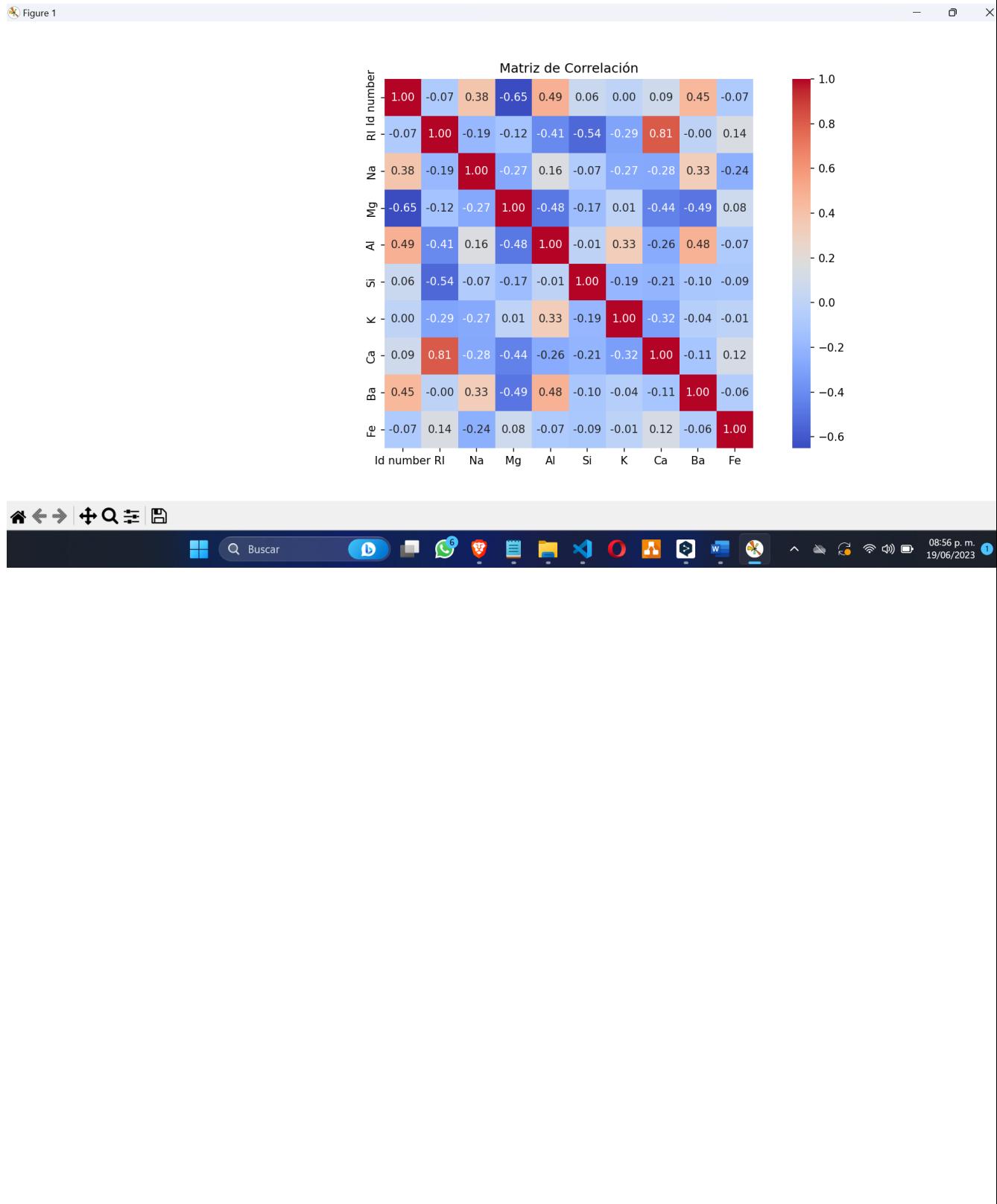
```

Archivo Editar Selección Ver Ir ...
medidas de dispersion(matrizCorrelacion.py) - Proyecto 1. Conociendo tus datos - Visual Studio Code
EXPLORADOR ... medidas de dispersion(matrizCorrelacion.py) > ...
PROBLEMAS SALIDA CONSOLA DE DEPURACIÓN TERMINAL SQL CONSOLE MESSAGES
medidas de dispersion(matrizCorrelacion.py) ...
1 import pandas as pd
Matriz de correlación:
Id number RI Na Mg Al Si K Ca Ba Fe
Id number 1.000000 -0.072209 0.375722 -0.650328 0.490113 0.061232 0.003149 0.090800 0.451001 -0.072794
RI -0.072209 1.000000 -0.191885 -0.122274 -0.407326 -0.542052 -0.289833 0.810403 -0.000386 0.143046
Na 0.375722 -0.191885 1.000000 -0.273732 0.156794 -0.069809 -0.266087 -0.275442 0.326603 -0.241346
Mg -0.650328 -0.122274 0.156794 1.000000 -0.481799 -0.165927 0.005396 -0.443750 -0.492262 0.083060
Al 0.490113 -0.407326 0.156794 -0.481799 1.000000 -0.005524 0.325958 -0.259592 0.479404 -0.074402
Si 0.061232 -0.542052 -0.069809 -0.165927 0.005524 1.000000 -0.193331 -0.208732 -0.102151 -0.094201
K 0.003149 -0.289833 -0.266087 0.005396 0.325958 -0.193331 1.000000 -0.317836 -0.042618 -0.007719
Id number RI Na Mg Al Si K Ca Ba Fe
Id number 3834.166667 -1.357850e-02 18.998263 -58.083920 15.151901 2.936714 0.127160 8.001526 1.388549e+01 -0.439202
RI -0.013578 9.222541e-06 -0.000476 -0.000536 -0.000618 -0.000574 0.003502 -5.828836e-07 0.000042
Na 18.998263 -4.758592e-04 0.666841 -0.322422 0.063926 -0.044154 -0.141713 -0.320106 1.326109e-01 -0.019204
Mg -58.083920 -5.356087e-04 -0.322422 2.080540 -0.346967 -0.185375 0.005076 -0.910916 -3.530469e-01 0.011674
Al 15.151901 -6.175934e-04 0.063926 -0.346967 0.249270 -0.002136 0.106138 -0.184450 1.190103e-01 -0.003620
Ca 0.090800 0.810403 -0.275442 -0.443750 -0.259592 -0.208732 -0.317836 1.000000 -0.112841 0.124968
Ba 0.451001 -0.000386 0.326603 -0.492262 0.479404 -0.102151 -0.042618 -0.112841 1.000000 -0.058692
Fe -0.072794 0.143010 -0.241346 0.083060 -0.074402 -0.094201 -0.007719 0.124968 -0.058692 1.000000
Matriz de covarianza:
Id number RI Na Mg Al Si K Ca Ba Fe
Id number 3834.166667 -1.357850e-02 18.998263 -58.083920 15.151901 2.936714 0.127160 8.001526 1.388549e+01 -0.439202
RI -0.013578 9.222541e-06 -0.000476 -0.000536 -0.000618 -0.000574 0.003502 -5.828836e-07 0.000042
Na 18.998263 -4.758592e-04 0.666841 -0.322422 0.063926 -0.044154 -0.141713 -0.320106 1.326109e-01 -0.019204
Mg -58.083920 -5.356087e-04 -0.322422 2.080540 -0.346967 -0.185375 0.005076 -0.910916 -3.530469e-01 0.011674
Al 15.151901 -6.175934e-04 0.063926 -0.346967 0.249270 -0.002136 0.106138 -0.184450 1.190103e-01 -0.003620
Ca 0.2936714 -1.275010e-03 -0.044154 -0.185375 -0.002136 0.599921 -0.097662 -0.230085 -3.934042e-02 -0.007109
Ba 0.127160 -5.740479e-04 -0.141713 0.005076 0.106138 -0.097662 0.425354 -0.295006 -1.382028e-02 -0.000491
Fe 8.001526 3.502498e-03 -0.320106 -0.910916 -0.184450 -0.230085 -0.295006 2.025366 -7.984845e-02 0.017329
K 13.885493 -5.828836e-07 0.132611 -0.353047 0.119010 -0.039340 -0.013820 -0.079848 2.472270e-01 -0.002844
Ba 0.451001 -0.000386 0.326603 -0.492262 0.479404 -0.102151 -0.042618 -0.112841 1.000000 -0.058692
Fe -0.439202 4.231770e-05 -0.019204 0.011674 -0.003620 -0.007109 -0.000491 0.017329 -2.843522e-03 0.009494

```

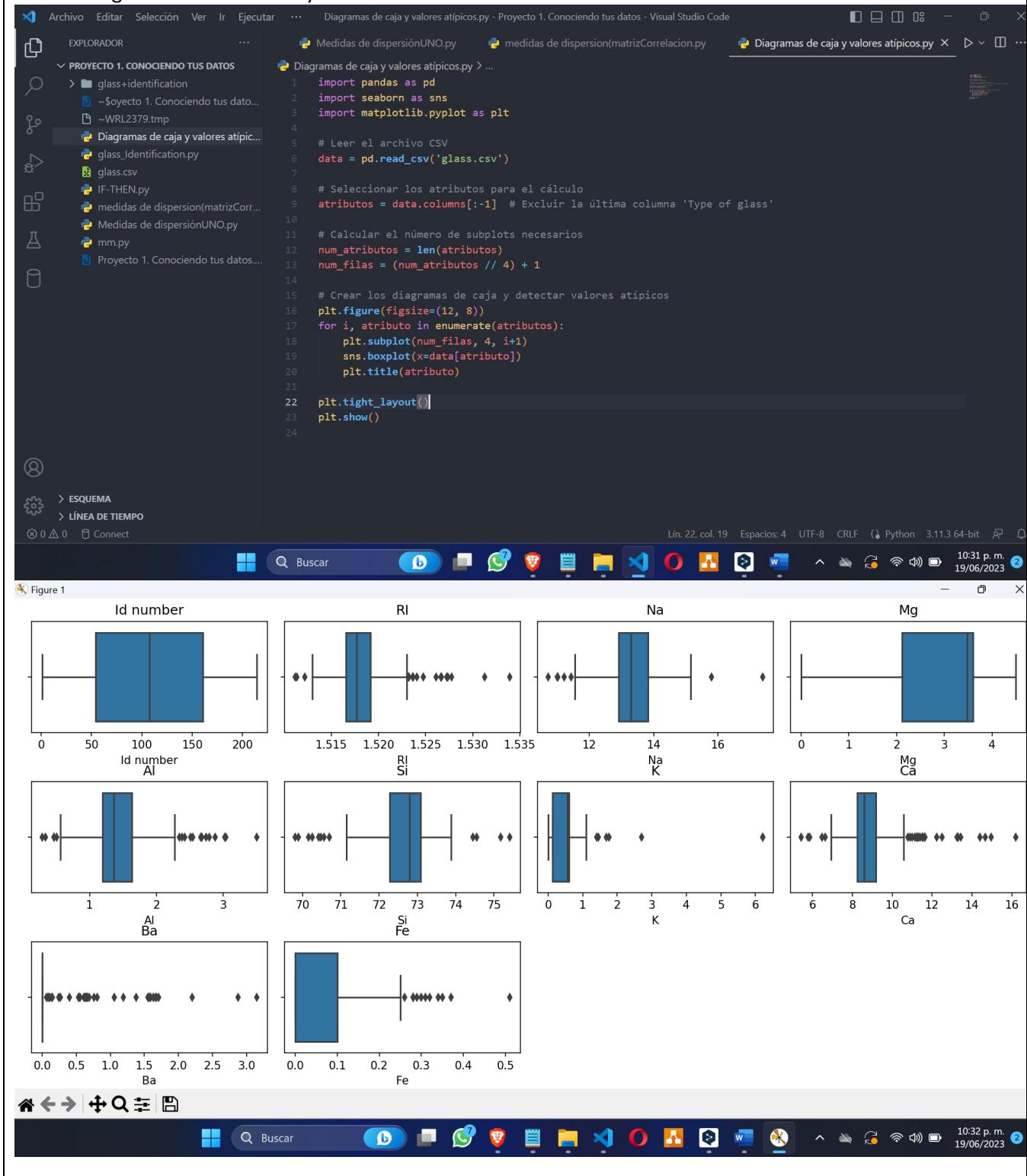
Lín. 15, col. 32 Espacios: 4 UTF-8 CRLF Python 3.11.3 64-bit 08:54 p.m. 19/06/2023

Resultado en Gráfica:



e) Diagramas de caja y valores atípicos

Calcular los diagramas de caja y los valores atípicos para cada uno de los atributos de la base de datos.
Poner código fuente del cálculo y el resultado.



The screenshot shows a Visual Studio Code interface with the following details:

- File Explorer:** Shows a project named "PROYECTO 1. CONOCIENDO TUS DATOS" containing several files including "Diagramas de caja y valores atípicos.py", "glass_Identification.py", and "glass.csv".
- Code Editor:** Displays the Python script "Diagramas de caja y valores atípicos.py" which uses pandas, seaborn, and matplotlib.pyplot to load a CSV file, select attributes, calculate subplot counts, and generate box plots for each attribute.
- Terminal:** Shows the command "python Diagramas de caja y valores atípicos.py" being run, with output indicating the script is running.
- Output:** Shows the execution results of the script, including the generated box plots.
- Status Bar:** Provides information like line number (Lín. 22), column number (col. 19), and file encoding (UTF-8).
- Taskbar:** Shows various system icons and the current date and time (10:31 p.m., 19/06/2023).
- Figure 1:** A grid of 12 box plots showing the distribution of attributes from the "glass" dataset. The attributes are: Id number, RI, Na, Mg, Al, Si, K, Ca, Ba, Fe, and Ba again. Each plot includes whiskers, a central box, and individual data points (outliers) marked with diamonds.

Especificar las tuplas o registros que son considerados valores atípicos.

Este código mostrará los diagramas de caja para cada atributo y, al mismo tiempo, imprimirá los registros considerados como valores atípicos para cada atributo en la consola.

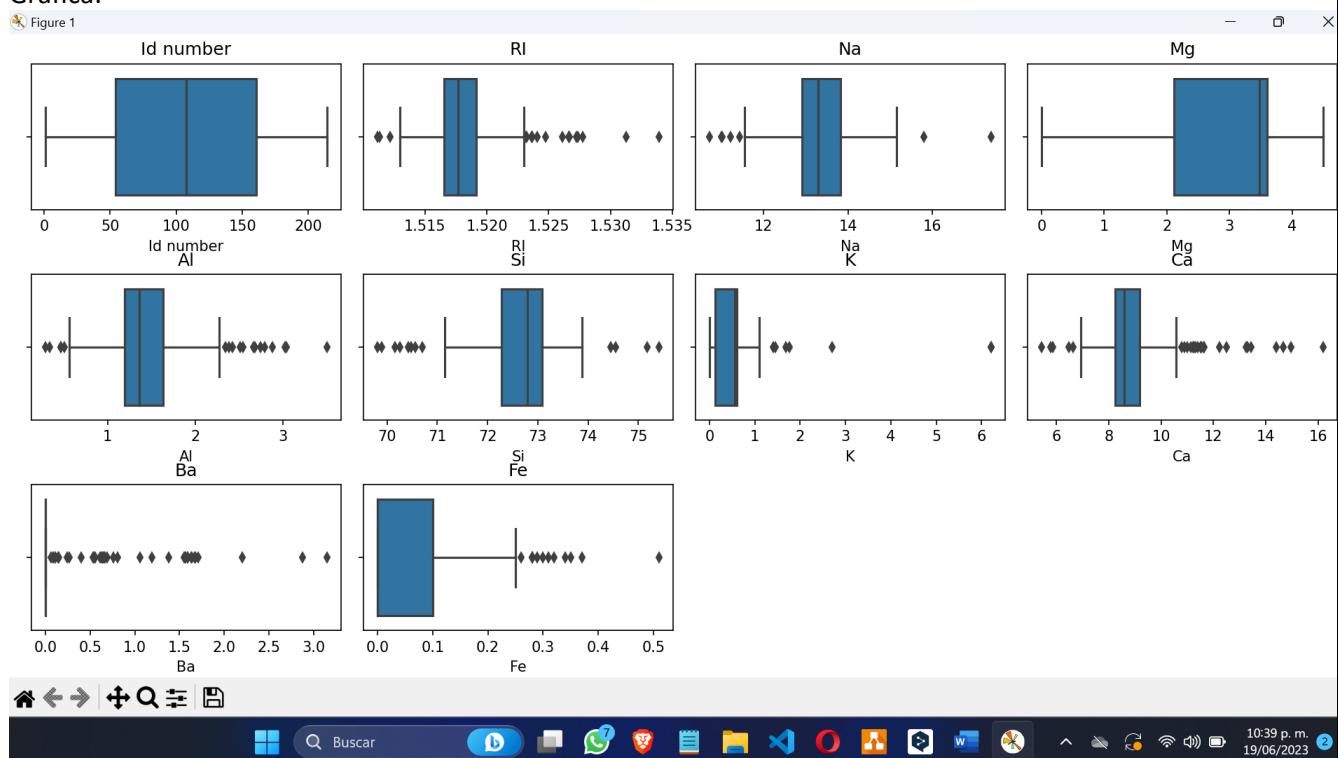
```

Archivo Editar Selección Ver Ir Ejecutar ...
PROYECTO 1. CONOCIENDO TUS DATOS
    glass_Identification.py
    ~-Soyecto 1. Conociendo tus datos...
    -WRL2379.tmp
    Diagramas de caja y valores atípicos.py
    glass_Identification.py
    glass.csv
    IF-THEN.py
    medidas de dispersion(matrizCorrelacion.py)
    medidas de dispersionUNO.py
    mm.py
    Proyecto 1. Conociendo tus datos...
    valoresAtipicos.py

valoresAtipicos.py - Proyecto 1. Conociendo tus datos - Visual Studio Code
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4
5 # Leer el archivo CSV
6 data = pd.read_csv('glass.csv')
7
8 # Seleccionar los atributos para el cálculo
9 atributos = data.columns[:-1] # Excluir la última columna 'Type of glass'
10
11 # Calcular el número de subplots necesarios
12 num_atributos = len(atributos)
13 num_filas = (num_atributos // 4) + 1
14
15 # Crear los diagramas de caja y detectar valores atípicos
16 plt.figure(figsize=(12, 8))
17 outliers = {} # Diccionario para almacenar los valores atípicos por atributo
18
19 for i, atributo in enumerate(atributos):
20     plt.subplot(num_filas, 4, i+1)
21     sns.boxplot(x=data[atributo])
22     plt.title(atributo)
23
24     # Detectar valores atípicos utilizando el rango intercuartilico (IQR)
25     q1 = data[atributo].quantile(0.25)
26     q3 = data[atributo].quantile(0.75)
27     iqr = q3 - q1
28     lower_threshold = q1 - 1.5 * iqr
29     upper_threshold = q3 + 1.5 * iqr
30
31     outliers[atributo] = data[(data[atributo] < lower_threshold) | (data[atributo] > upper_threshold)]
32
33 plt.tight_layout()
34 plt.show()
35
36 # Mostrar los registros considerados como valores atípicos por atributo
37 for atributo, df_outliers in outliers.items():
38     print(f"Valores atípicos en el atributo '{atributo}':")
39     print(df_outliers)
40     print()
41
42 Lin. 11, col. 44 Espacios: 4 UTF-8 CR/LF Python 3.11.3 64-bit

```

Gráfica:



Resultado en consola:

```

Archivo Editar Selección Ver Ir Ejecutar ...
valoresAtípicos.py - Proyecto 1. Conociendo tus datos - Visual Studio Code
PROYECTO 1. CONOCIENDO TUS DATOS
PROBLEMAS SALIDA CONSOLA DE DEPURACIÓN TERMINAL SQL CONSOLE
Python + v ll ... > Python ll ... >
PS C:\Users\brand\OneDrive\Documentos\Uno Cuatrimestre\Minería de datos\Proyecto 1. Conociendo tus datos & C:/Users/brand/AppData/Local/Programs/Python/Python311/python.exe "c:/Users/brand/OneDrive/Documentos/Uno Cuatrimestre/Minería de datos/Proyecto 1. Conociendo tus datos/valoresAtípicos.py"
Valores atípicos en el atributo 'Id number':
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Id number | RI | Na | Mg | Al | Si | K | Ca | Ba | Fe | Type of glass |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Valores atípicos en el atributo 'RI':
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Id number | RI | Na | Mg | Al | Si | K | Ca | Ba | Fe | Type of glass |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 47 | 48 | 1.52667 | 13.99 | 3.7 | 0.71 | 71.57 | 0.02 | 9.82 | 0 | 0.1 | 1 |
| 50 | 51 | 1.5232 | 13.72 | 3.72 | 0.51 | 71.75 | 0.09 | 10.06 | 0 | 0.16 | 1 |
| 56 | 57 | 1.52125 | 12.99 | 3.47 | 1.12 | 72.98 | 0.62 | 8.35 | 0 | 0.31 | 1 |
| 103 | 104 | 1.52725 | 13.8 | 3.15 | 0.66 | 70.57 | 0.08 | 11.64 | 0 | 0 | 2 |
| 104 | 106 | 1.52624 | 13.83 | 2.9 | 1.17 | 71.15 | 0.08 | 10.79 | 0 | 0 | 2 |
| 106 | 107 | 1.52745 | 13.75 | 0 | 1.08 | 72.0 | 0.01 | 13.0 | 0 | 0.34 | 2 |
| 106 | 107 | 1.53125 | 10.72 | 0 | 0.71 | 69.81 | 0.58 | 13.3 | 3.15 | 0.28 | 2 |
| 107 | 108 | 1.53393 | 12.3 | 0 | 1 | 70.16 | 0.12 | 16.19 | 0 | 0.24 | 2 |
| 108 | 110 | 1.52664 | 11.23 | 0 | 0.77 | 73.21 | 0.12 | 14.68 | 0 | 0 | 2 |
| 110 | 111 | 1.52739 | 11.02 | 0 | 0.75 | 73.08 | 0 | 14.96 | 0 | 0 | 2 |
| 111 | 112 | 1.52779 | 12.64 | 0 | 0.67 | 72.02 | 0.06 | 14.4 | 0 | 0 | 2 |
| 112 | 113 | 1.52614 | 13.7 | 0 | 1.36 | 71.24 | 0.19 | 13.44 | 0 | 0.1 | 2 |
| 131 | 132 | 1.52369 | 13.44 | 0 | 1.58 | 72.22 | 0.32 | 12.24 | 0 | 0 | 5 |
| 170 | 171 | 1.52369 | 13.44 | 0 | 1.58 | 72.22 | 0.32 | 12.24 | 0 | 0 | 5 |
| 184 | 185 | 1.51115 | 17.38 | 0 | 0.34 | 75.41 | 0 | 6.65 | 0 | 0 | 6 |
| 185 | 186 | 1.51131 | 13.69 | 3.2 | 1.81 | 72.81 | 1.76 | 5.43 | 1.19 | 0 | 7 |
| 187 | 188 | 1.52315 | 13.44 | 3.34 | 1.23 | 72.38 | 0.6 | 8.83 | 0 | 0 | 7 |
| 189 | 190 | 1.52365 | 15.79 | 1.83 | 1.31 | 70.43 | 0.31 | 8.61 | 1.68 | 0 | 7 |

Valores atípicos en el atributo 'Na':
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Id number | RI | Na | Mg | Al | Si | K | Ca | Ba | Fe | Type of glass |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Valores atípicos en el atributo 'Mg':
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Id number | RI | Na | Mg | Al | Si | K | Ca | Ba | Fe | Type of glass |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 105 | 106 | 1.52976 | 13.45 | 0 | 1.88 | 72.19 | 0.81 | 13.24 | 0 | 0.34 | 2 |
| 106 | 107 | 1.52125 | 13.4 | 0 | 1.88 | 72.19 | 0.81 | 13.24 | 0 | 0.34 | 2 |
| 110 | 111 | 1.52664 | 11.23 | 0 | 0.77 | 73.21 | 0 | 14.68 | 0 | 0 | 2 |
| 111 | 112 | 1.52739 | 11.02 | 0 | 0.75 | 73.08 | 0 | 14.96 | 0 | 0 | 2 |
| 166 | 167 | 1.52151 | 11.03 | 1.71 | 1.56 | 73.44 | 0.58 | 11.62 | 0 | 0 | 5 |
| 184 | 185 | 1.51115 | 17.38 | 0 | 0.34 | 75.41 | 0 | 6.65 | 0 | 0 | 6 |
| 189 | 190 | 1.52365 | 15.79 | 1.83 | 1.31 | 70.43 | 0.31 | 8.61 | 1.68 | 0 | 7 |

Valores atípicos en el atributo 'Al':
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Id number | RI | Na | Mg | Al | Si | K | Ca | Ba | Fe | Type of glass |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 21 | 22 | 1.51966 | 14.77 | 3.75 | 0.29 | 72.02 | 0.03 | 9 | 0 | 0 | 1 | |
| 38 | 39 | 1.52213 | 14.21 | 3.82 | 0.47 | 71.77 | 0.11 | 9.57 | 0 | 0 | 1 |
| 39 | 40 | 1.52213 | 14.21 | 3.82 | 0.47 | 71.77 | 0.11 | 9.57 | 0 | 0 | 1 |
| 50 | 51 | 1.52315 | 13.72 | 0 | 3.72 | 0.76 | 74.75 | 0.09 | 10.86 | 0 | 0.16 | 1 |
| 164 | 165 | 1.51514 | 14.08 | 2.68 | 3.5 | 69.1 | 1.68 | 5.87 | 2.2 | 0 | 5 |
| 171 | 172 | 1.51316 | 13.02 | 0 | 3.04 | 70.48 | 6.1 | 6.82 | 1 | 0 | 5 |
| 172 | 173 | 1.51321 | 13 | 0 | 3.02 | 70.7 | 6.21 | 6.93 | 0 | 0 | 5 |
| 184 | 185 | 1.51115 | 17.38 | 0 | 0.34 | 75.41 | 0 | 6.65 | 0 | 0 | 6 |
| 191 | 192 | 1.51602 | 14.85 | 0 | 2.38 | 73.28 | 0 | 8.76 | 0.64 | 0.09 | 7 |
| 192 | 193 | 1.51623 | 14.2 | 0 | 2.79 | 73.46 | 0.04 | 9.04 | 0.4 | 0.09 | 7 |
| 195 | 196 | 1.51545 | 14.14 | 0 | 2.68 | 73.39 | 0.08 | 9.07 | 0.61 | 0.05 | 7 |
| 196 | 197 | 1.51556 | 13.87 | 0 | 2.54 | 73.23 | 0.14 | 9.41 | 0.81 | 0.01 | 7 |
| 197 | 198 | 1.51727 | 14.7 | 0 | 2.34 | 73.28 | 0 | 8.95 | 0.66 | 0 | 7 |
| 198 | 199 | 1.51531 | 14.38 | 0 | 2.66 | 73.1 | 0.84 | 9.08 | 0.64 | 0 | 7 |
| 200 | 201 | 1.51514 | 14.08 | 0 | 2.01 | 73.08 | 1.05 | 1 | 0 | 0 | 7 |
| 202 | 203 | 1.51514 | 14.85 | 0 | 1.42 | 73.72 | 0 | 8.39 | 0.56 | 0 | 7 |
| 208 | 209 | 1.5164 | 14.37 | 0 | 2.74 | 72.85 | 0 | 9.45 | 0.54 | 0 | 7 |
| 209 | 210 | 1.51623 | 14.14 | 0 | 2.88 | 72.61 | 0.08 | 9.18 | 1.06 | 0 | 7 |

Valores atípicos en el atributo 'Si':
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Id number | RI | Na | Mg | Al | Si | K | Ca | Ba | Fe | Type of glass |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Valores atípicos en el atributo 'K':
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Id number | RI | Na | Mg | Al | Si | K | Ca | Ba | Fe | Type of glass |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 103 | 104 | 1.52725 | 13.8 | 3.15 | 0.66 | 70.57 | 0.08 | 11.64 | 0 | 0 | 2 |
| 106 | 107 | 1.53125 | 10.73 | 0 | 2.1 | 69.81 | 0.58 | 13.3 | 3.15 | 0.28 | 2 |
| 107 | 108 | 1.53393 | 12.3 | 0 | 1 | 70.16 | 0.12 | 16.19 | 0 | 0.24 | 2 |
| 109 | 110 | 1.51818 | 13.72 | 0 | 0.56 | 74.45 | 0 | 10.99 | 0 | 0 | 2 |
| 163 | 164 | 1.51514 | 14.01 | 2.68 | 3.5 | 69.89 | 1.68 | 5.87 | 2.2 | 0 | 5 |
| 171 | 172 | 1.51316 | 13.02 | 0 | 3.04 | 70.48 | 6.21 | 6.93 | 0 | 0 | 5 |
| 172 | 173 | 1.51321 | 13 | 0 | 3.02 | 70.7 | 6.21 | 6.93 | 0 | 0 | 5 |
| 180 | 181 | 1.51299 | 14.4 | 1.74 | 1.54 | 74.55 | 0 | 7.59 | 0 | 0 | 6 |
| 184 | 185 | 1.51115 | 17.38 | 0 | 0.34 | 75.41 | 0 | 6.65 | 0 | 0 | 6 |
| 188 | 189 | 1.52247 | 14.86 | 2.2 | 2.06 | 70.26 | 0.76 | 9.76 | 0 | 0 | 7 |
| 189 | 190 | 1.52365 | 15.79 | 1.83 | 1.31 | 70.43 | 0.31 | 8.61 | 1.68 | 0 | 7 |

```

Valores atípicos en el atributo 'K':

	Id number	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type of glass
163	164	1.51514	14.01	2.68	3.5	69.89	1.68	5.87	2.2	0	5
171	172	1.51516	13.02	0	3.04	70.48	6.21	6.96	0	0	5
172	173	1.51521	13	0	3.02	70.7	6.21	6.93	0	0	5
185	186	1.51511	13.69	3.2	1.81	72.81	1.76	5.43	1.19	0	7
186	187	1.51538	14.32	3.26	2.22	71.25	1.46	5.79	1.63	0	7
201	202	1.51653	11.95	0	1.19	75.18	2.7	8.93	0	0	7
207	208	1.51831	14.39	0	1.82	72.86	1.41	6.47	2.88	0	7

Valores atípicos en el atributo 'Ca':

	Id number	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type of glass
103	104	1.52725	13.8	3.15	0.66	57.57	0.08	11.64	0	0	2
104	105	1.52541	13.83	2.9	1.17	71.15	0.08	10.79	0	0	2
106	107	1.52475	11.45	0	1.88	72.19	0.81	13.24	0	0.34	2
106	107	1.52471	11.45	0	2.1	69.89	0.58	10.15	0.12	0.28	2
107	108	1.53393	12.3	0	1	70.16	1.02	16.19	0	0.24	2
108	109	1.52222	14.43	0	1	72.67	0.1	11.52	0	0.08	2
109	110	1.51818	13.72	0	0.56	74.45	0	10.99	0	0	2
110	111	1.52664	11.23	0	0.77	73.21	0	14.68	0	0	2
111	112	1.52739	11.02	0	0.75	73.08	0	14.96	0	0	2
112	113	1.52777	12.64	0	0.67	72.02	0.06	14.4	0	0	2
130	131	1.52177	13.75	1.01	1.36	72.19	0.33	11.14	0	0	2
131	132	1.52614	13.7	0	1.36	71.24	0.19	13.44	0	0.1	2
163	164	1.51511	14.01	2.68	3.1	69.89	1.08	8.87	2.2	0	5
164	165	1.51517	11.56	0.88	0.56	72.02	0.07	11.01	0	0	5
165	166	1.52151	11.01	1.71	1.56	72.92	0.6	11.62	0	0	5
166	167	1.52151	11.01	1.71	1.56	72.92	0.58	11.62	0	0	5
167	168	1.51969	12.64	0	1.65	73.75	0.38	11.53	0	0	5
169	170	1.51994	13.27	0	1.76	73.03	0.47	11.32	0	0	5
170	171	1.52369	13.44	0	1.58	72.22	0.32	12.24	0	0	5
173	174	1.52043	13.38	0	1.4	72.25	0.33	12.5	0	0	5
175	176	1.52119	12.97	0.33	1.51	73.39	0.13	11.27	0	0.28	5
182	183	1.51916	14.15	0	2.09	72.74	0	10.88	0	0	6
183	184	1.51969	14.56	0	0.56	73.48	0	11.22	0	0	6
184	185	1.51131	17.38	0	0.56	75.41	0	10.08	0	0	6
185	186	1.51131	11.69	3.2	1.81	72.05	1.76	5.43	1.19	0	7
186	187	1.51538	14.32	0.26	2.23	71.25	1.46	5.79	1.63	0	7
187	188	1.51831	14.39	0	1.82	72.86	1.41	6.47	2.88	0	7
188	189	1.51531	14.38	0	2.66	73.1	0.04	9.08	0.64	0	7
189	190	1.51613	13.88	1.78	1.27	72.18	0.53	9.57	0.27	0.16	2
190	191	1.51613	13.88	1.78	1.27	72.18	0.53	9.57	0.27	0.16	2
191	192	1.51602	14.85	0	3.63	72.83	0.57	8.41	0.89	0.17	2
192	193	1.51623	14.2	0	2.79	73.46	0.04	9.04	0.4	0.09	7
193	194	1.51719	14.75	0	2	73.02	0	8.53	1.59	0.08	7
194	195	1.51683	14.56	0	1.98	73.29	0	8.52	1.57	0.07	7
195	196	1.51545	14.14	0	2.68	73.39	0.08	9.07	0.61	0.05	7
196	197	1.51556	13.87	0	2.54	73.23	0.14	9.41	0.81	0.01	7
197	198	1.51727	14.7	0	2.34	73.28	0	8.95	0.66	0	7
198	199	1.51531	14.38	0	2.66	73.1	0.04	9.08	0.64	0	7
199	200	1.51613	14.38	0	2.66	73.1	0.04	9.08	0.64	0	7

Valores atípicos en el atributo 'Ba':

	Id number	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type of glass
32	33	1.51775	12.85	3.48	1.23	72.97	0.61	8.56	0.09	0.22	1
36	37	1.51909	13.89	3.53	1.32	71.81	0.51	8.78	0.11	0	1
61	62	1.51977	13.81	3.58	1.32	71.72	0.12	8.67	0.69	0	1
99	100	1.51811	12.96	2.96	1.43	72.92	0.6	8.79	0.14	0	2
100	101	1.51655	12.75	2.85	1.44	73.27	0.57	8.79	0.11	0.22	2
101	102	1.51655	12.75	2.85	1.44	73.27	0.57	8.79	0.11	0.22	2
102	103	1.51655	12.75	2.85	1.44	73.27	0.57	8.79	0.11	0.22	2
103	104	1.51655	12.75	2.85	1.44	73.27	0.57	8.79	0.11	0.22	2
104	105	1.51655	12.75	2.85	1.44	73.27	0.57	8.79	0.11	0.22	2
105	106	1.51655	12.75	2.85	1.44	73.27	0.57	8.79	0.11	0.22	2
106	107	1.51655	12.75	2.85	1.44	73.27	0.57	8.79	0.11	0.22	2
107	108	1.51325	10.75	0	1.24	69.81	0.58	13.3	3.15	0.28	2
108	109	1.51325	10.75	0	1.24	69.81	0.58	13.3	3.15	0.28	2
109	110	1.51673	13.3	3.64	1.53	72.53	0.65	8.03	0	0.29	2
110	111	1.51789	13.19	3.9	1.3	72.33	0.55	8.44	0	0.28	2
111	112	1.51839	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
112	113	1.52211	14.19	3.78	0.91	71.36	0.23	9.14	0	0.37	3
113	114	1.51908	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
114	115	1.52058	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
115	116	1.51831	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
116	117	1.51621	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
117	118	1.51673	13.3	3.64	1.53	72.53	0.65	8.03	0	0.29	2
118	119	1.51789	13.19	3.9	1.3	72.33	0.55	8.44	0	0.28	2
119	120	1.51839	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
120	121	1.52211	14.19	3.78	0.91	71.36	0.23	9.14	0	0.37	3
121	122	1.51908	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
122	123	1.52058	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
123	124	1.51831	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
124	125	1.51621	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
125	126	1.51673	13.3	3.64	1.53	72.53	0.65	8.03	0	0.29	2
126	127	1.51789	13.19	3.9	1.3	72.33	0.55	8.44	0	0.28	2
127	128	1.51839	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
128	129	1.52211	14.19	3.78	0.91	71.36	0.23	9.14	0	0.37	3
129	130	1.51908	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
130	131	1.52058	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
131	132	1.51831	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
132	133	1.51621	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
133	134	1.51673	13.3	3.64	1.53	72.53	0.65	8.03	0	0.29	2
134	135	1.51789	13.19	3.9	1.3	72.33	0.55	8.44	0	0.28	2
135	136	1.51839	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
136	137	1.52211	14.19	3.78	0.91	71.36	0.23	9.14	0	0.37	3
137	138	1.51908	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
138	139	1.52058	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
139	140	1.51831	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
140	141	1.51621	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
141	142	1.51673	13.3	3.64	1.53	72.53	0.65	8.03	0	0.29	2
142	143	1.51789	13.19	3.9	1.3	72.33	0.55	8.44	0	0.28	2
143	144	1.51839	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
144	145	1.52211	14.19	3.78	0.91	71.36	0.23	9.14	0	0.37	3
145	146	1.51908	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
146	147	1.52058	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
147	148	1.51831	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
148	149	1.51621	12.85	3.67	1.24	72.57	0.62	8.68	0	0.35	2
149	150	1.51673	13.3	3.64	1.53	72.53	0.65	8.03	0	0.29	2
150	151	1.51789	13.19	3.9	1.3	72.33	0.55	8.44	0	0.28	2
151	152	1.51839	12.85	3.							

The screenshot shows a Visual Studio Code interface with the following details:

- File Explorer:** Shows files related to the project "CONOCIENDO TUS DATOS", including "valoresAtipicos.py", "glass.csv", and various Python scripts like "glass_identification.py", "IF-THEN.py", and "medidas de dispersion/matrixCorrelacion.py".
- Terminal:** Displays the output of the Python script "valoresAtipicos.py" which prints a table of outliers for the 'Fe' attribute.
- Output:** Shows the printed table of outliers.
- Status Bar:** Includes file path (C:\Users\bbrand\OneDrive\Documentos\9mo Cuatrimestre\Mineria de datos\Proyecto 1. Conociendo tus datos>), line count (Lin. 19, col. 1), character count (Espacios: 4), encoding (UTF-8), and system information (CRLF, Python 3.11.3 64-bit).

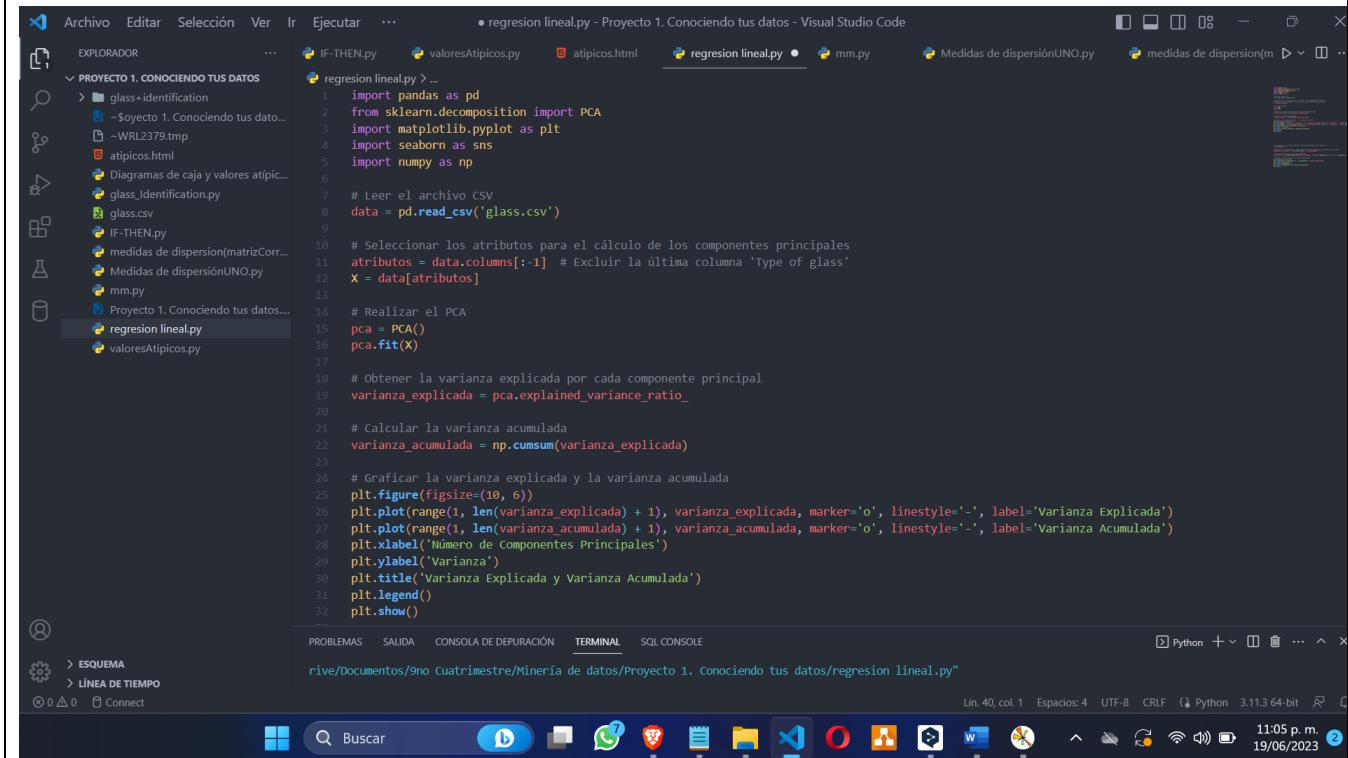
	Id number	Ri	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type of glass
5	6	1.51596	12.79	3.61	1.62	72.97	0.64	8.87	0	0.26	1
44	45	1.51786	12.73	3.43	1.19	72.95	0.62	8.76	0	0.3	1
56	57	1.51215	12.99	3.47	1.12	72.98	0.62	8.35	0	0.31	1
71	72	1.51848	13.64	3.87	1.27	71.96	0.54	8.32	0	0.32	2
105	106	1.52475	11.45	0	1.88	72.19	0.81	13.24	0	0.34	2
106	107	1.53125	10.73	0	2.1	69.81	0.58	13.3	3.15	0.28	2
118	119	1.51673	13.3	3.64	1.53	72.53	0.65	8.03	0	0.29	2
135	136	1.51789	13.19	3.9	1.3	72.33	0.55	8.44	0	0.28	2
145	146	1.51801	12.89	3.67	1.24	72.57	0.62	8.68	0	0.35	2
163	164	1.52021	10.19	3.16	0.96	71.38	0.23	9.14	0	0.3	3
174	175	1.52058	12.89	3.61	1.17	71.38	0.16	0.24	0.51	5	5
175	176	1.52119	12.97	0.33	1.51	73.39	0.13	11.27	0	0.28	5

f) Regresión lineal

Calcular los valores de la regresión lineal (coeficientes y cruces de las líneas, coeficiente de determinación y correlación) para los atributos de la base de datos.

Graficar la distribución de los datos con su respectiva línea de regresión.

Poner código fuente del cálculo y el resultado.



```

Archivo Editar Selección Ver Ir Ejecutar ...
• regresion lineal.py - Proyecto 1. Conociendo tus datos - Visual Studio Code
EXPLORADOR ... regresion lineal.py valoresAtipicos.py atipicos.html regresion lineal.py mm.py Medidas de dispersiónUNO.py medidas de dispersion(m) ..
PROYECTO 1. CONOCIENDO TUS DATOS
> glass+identification
-> -$oyecto 1. Conociendo tus dato...
-> -WRL2379.tmp
-> atipicos.html
-> Diagramas de caja y valores atípic...
-> glass_Identification.py
-> glass.csv
-> IF-THEN.py
-> medidas de dispersion(matrizCorr...
-> Medidas de dispersiónUNO.py
-> mm.py
Proyecto 1. Conociendo tus datos...
-> regresion lineal.py
-> valoresAtipicos.py

regresion lineal.py > ...
1 import pandas as pd
2 from sklearn.decomposition import PCA
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 import numpy as np
6
7 # Leer el archivo CSV
8 data = pd.read_csv('glass.csv')
9
10 # Seleccionar los atributos para el cálculo de los componentes principales
11 atributos = data.columns[:-1] # Excluir la última columna 'Type of glass'
12 X = data[atributos]
13
14 # Realizar el PCA
15 pca = PCA()
16 pca.fit(X)
17
18 # Obtener la varianza explicada por cada componente principal
19 varianza_explicada = pca.explained_variance_ratio_
20
21 # Calcular la varianza acumulada
22 varianza_acumulada = np.cumsum(varianza_explicada)
23
24 # Graficar la varianza explicada y la varianza acumulada
25 plt.figure(figsize=(10, 6))
26 plt.plot(range(1, len(varianza_explicada) + 1), varianza_explicada, marker='o', linestyle='-', label='Varianza Explicada')
27 plt.plot(range(1, len(varianza_acumulada) + 1), varianza_acumulada, marker='o', linestyle='--', label='Varianza Acumulada')
28 plt.xlabel('Número de Componentes Principales')
29 plt.ylabel('Varianza')
30 plt.title('Varianza Explicada y Varianza Acumulada')
31 plt.legend()
32 plt.show()

PROBLEMAS SALIDA CONSOLA DE DEDUPLICACIÓN TERMINAL SQL CONSOLE
rive/Documentos/9no Cuatrimestre/Minería de datos/Proyecto 1. Conociendo tus datos/regresion lineal.py"
Lin. 40, col. 1 Espacios: 4 UTF-8 CRLF Python 3.11.3 64-bit 11:05 p.m. 19/06/2023
0 0 △ 0 Connect
Buscar

```

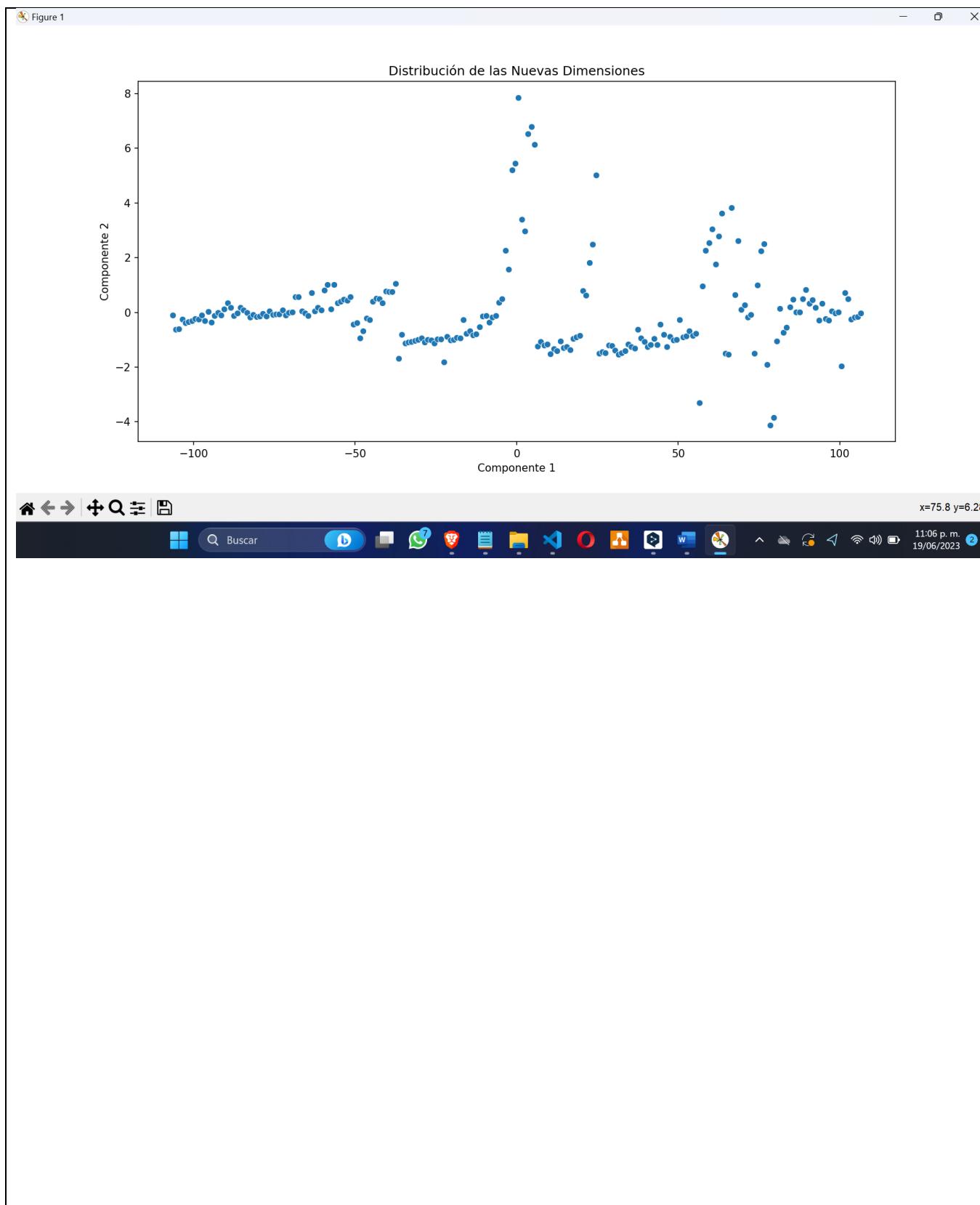
The screenshot shows a Visual Studio Code interface with the following details:

- File Explorer:** Shows a project named "PROYECTO 1. CONOCIENDO TUS DATOS" containing files like "regresion lineal.py", "IF-THEN.py", and "glass.csv".
- Code Editor:** Displays the content of "regresion lineal.py":

```

37
38
39
40
41
42 # Seleccionar el número de componentes principales deseados (por ejemplo, 2)
43 num_componentes = 2
44
45 # Realizar la proyección de los datos originales en los componentes principales seleccionados
46 componentes_principales = pca.transform(X)[:, :num_componentes]
47
48 # Crear un DataFrame con las nuevas dimensiones
49 df_componentes = pd.DataFrame(componentes_principales, columns=[f'Componente {i+1}' for i in range(num_componentes)])
50
51 # Graficar la distribución de las nuevas dimensiones
52 plt.figure(figsize=(8, 6))
53 sns.scatterplot(x='Componente 1', y='Componente 2', data=df_componentes)
54 plt.xlabel('Componente 1')
55 plt.ylabel('Componente 2')
56 plt.title('Distribución de las Nuevas Dimensiones')
57 plt.show()
58

```
- Bottom Status Bar:** Shows "Lin. 40, col. 1 Espacios: 4 UTF-8 Python 3.11.3 64-bit" and the date/time "11:07 p.m. 19/06/2023".
- Figure Window:** A scatter plot titled "Varianza Explicada y Varianza Acumulada" showing the relationship between the number of principal components and explained variance. The x-axis is "Número de Componentes Principales" (2 to 10) and the y-axis is "Varianza" (0.0 to 1.0). The plot shows a sharp drop from 1.0 at 1 component to near 0.0 at 2 components, followed by a flat line at 0.0 for all other components.
- Taskbar:** Shows the Windows taskbar with various pinned icons and the date/time "11:06 p.m. 19/06/2023".



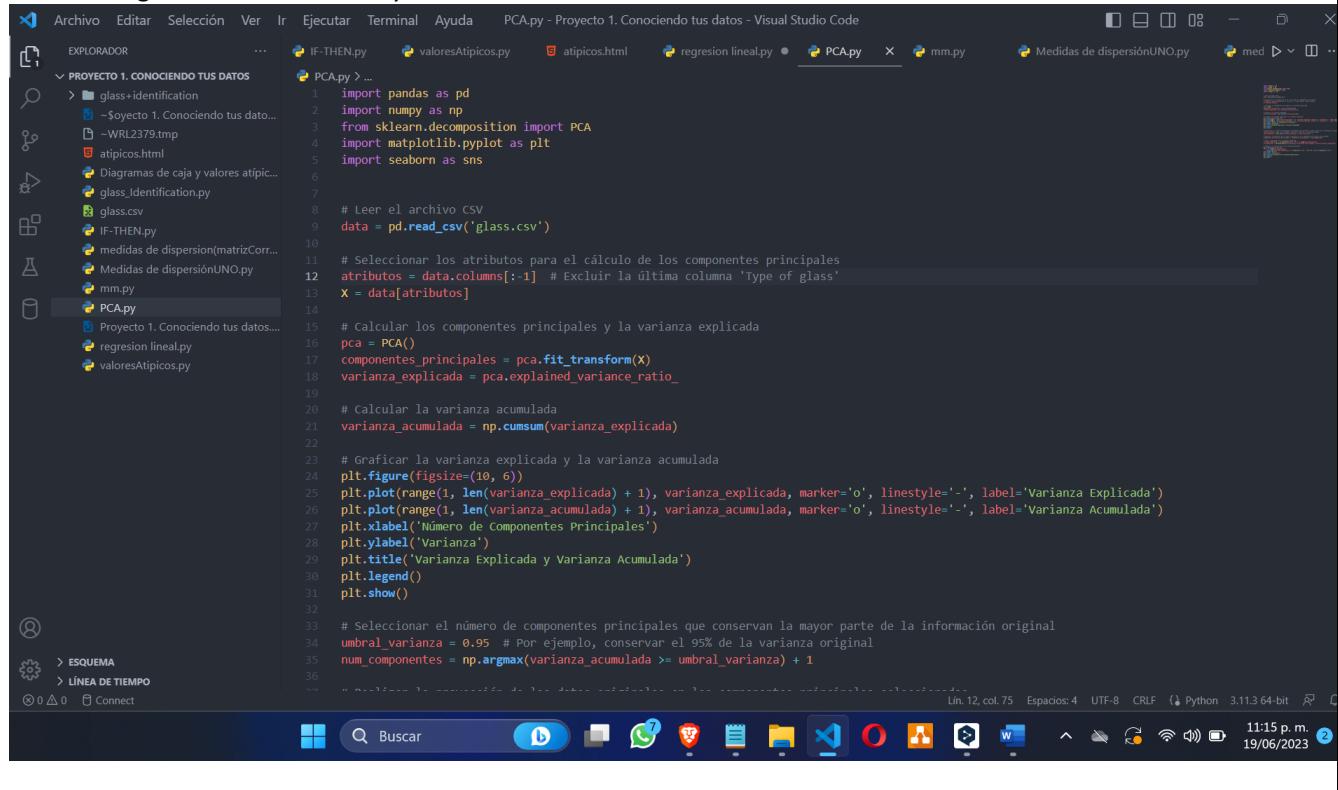
g) Análisis de Componentes Principales (PCA)

Calcular los componentes principales de la base de datos y su nivel de varianza.

Seleccionar el número de componentes que conservan la mayor parte de información original.

Graficar la distribución de las nuevas dimensiones de la base de datos.

Poner código fuente del cálculo y el resultado.



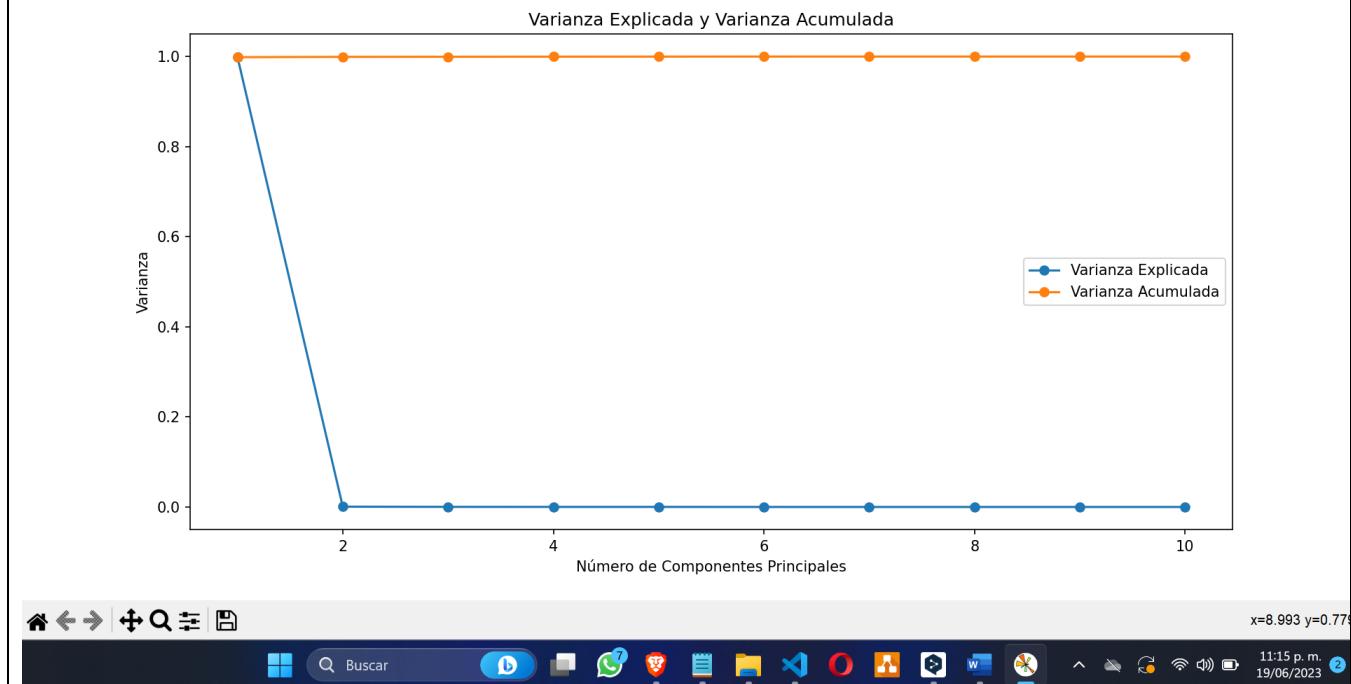
The screenshot shows the Visual Studio Code interface with the following details:

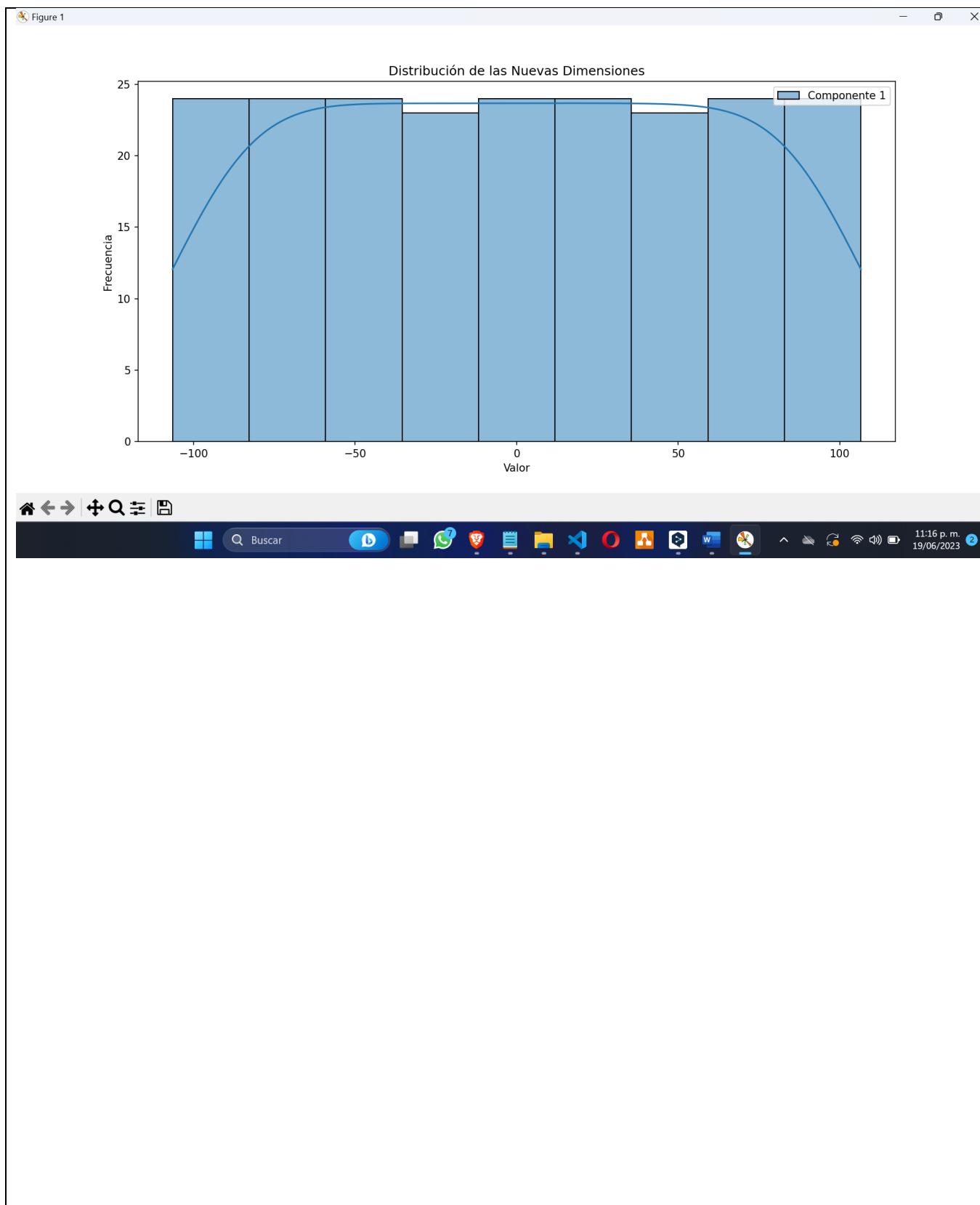
- File Explorer:** Shows the project structure "PROYECTO 1. CONOCIENDO TUS DATOS" containing files like glass+identification, IF-THEN.py, PCA.py, etc.
- Code Editor:** Displays the content of the PCA.py script:

```

1 import pandas as pd
2 import numpy as np
3 from sklearn.decomposition import PCA
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6
7
8 # Leer el archivo CSV
9 data = pd.read_csv('glass.csv')
10
11 # Seleccionar los atributos para el cálculo de los componentes principales
12 atributos = data.columns[:-1] # Excluir la última columna 'Type of glass'
13 X = data[atributos]
14
15 # Calcular los componentes principales y la varianza explicada
16 pca = PCA()
17 componentes_principales = pca.fit_transform(X)
18 varianza_explizada = pca.explained_variance_ratio_
19
20 # Calcular la varianza acumulada
21 varianza_acumulada = np.cumsum(varianza_explizada)
22
23 # Graficar la varianza explicada y la varianza acumulada
24 plt.figure(figsize=(10, 6))
25 plt.plot(range(1, len(varianza_explizada) + 1), varianza_explizada, marker='o', linestyle='--', label='Varianza Explizada')
26 plt.plot(range(1, len(varianza_acumulada) + 1), varianza_acumulada, marker='o', linestyle='--', label='Varianza Acumulada')
27 plt.xlabel('Número de Componentes Principales')
28 plt.ylabel('Varianza')
29 plt.title('Varianza Explizada y Varianza Acumulada')
30 plt.legend()
31 plt.show()
32
33 # Seleccionar el número de componentes principales que conservan la mayor parte de la información original
34 umbral_varianza = 0.95 # Por ejemplo, conservar el 95% de la varianza original
35 num_componentes = np.argmax(varianza_acumulada >= umbral_varianza) + 1
36

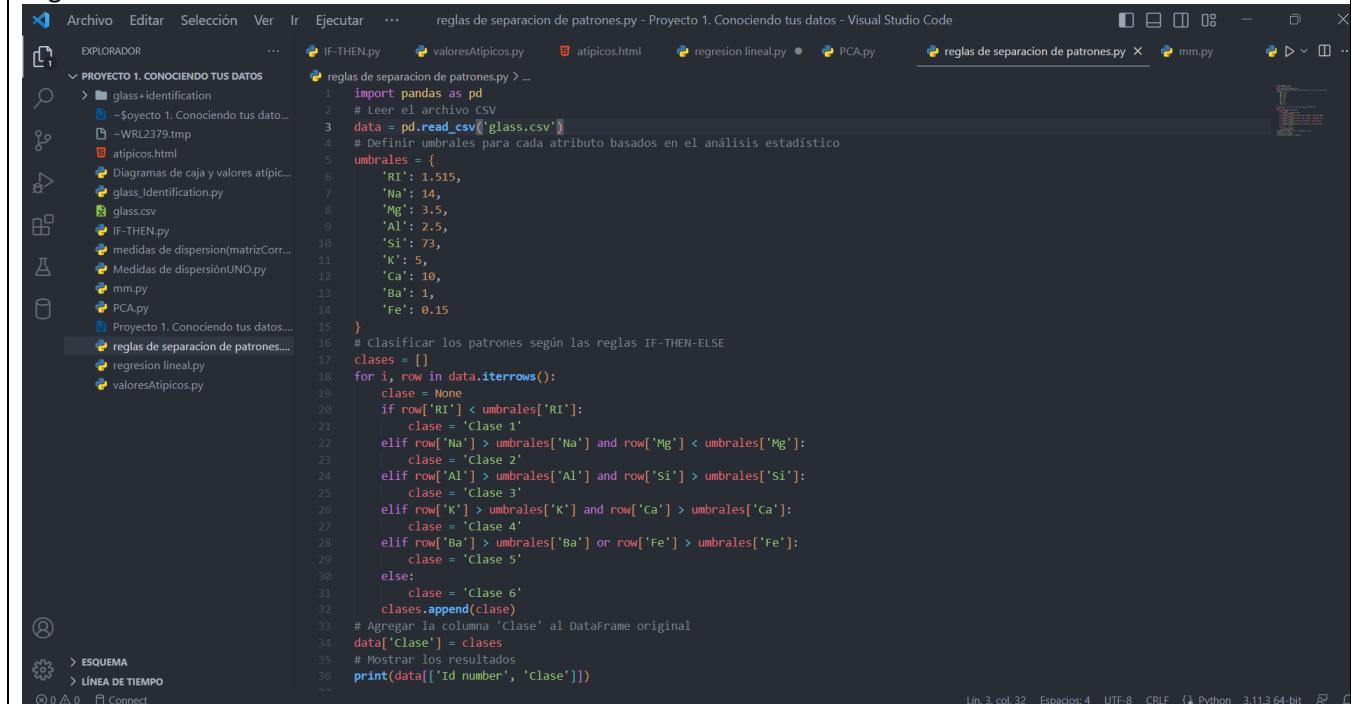
```
- Bottom Status Bar:** Shows file path "C:\Users\...\\Proyecto 1. Conociendo tus datos\\PCA.py - Proyecto 1. Conociendo tus datos - Visual Studio Code", line count "Lín. 12, col. 75", character count "Espacios: 4", encoding "UTF-8", and date/time "11:15 p.m. 19/06/2023".





h) Reglas de separación de patrones

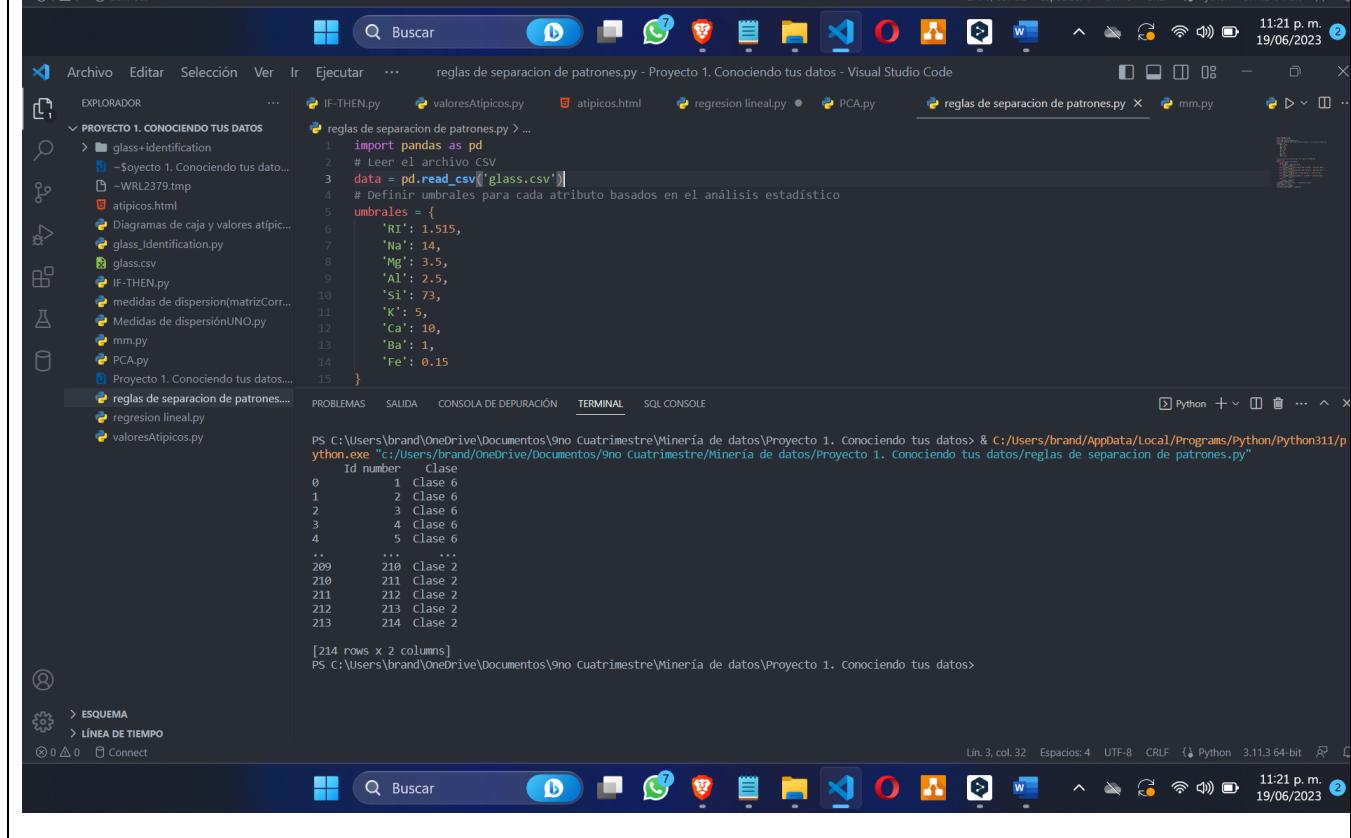
Crear un algoritmo con reglas IF-THEN-ELSE que permita separar lo mejor posible las clases de los patrones según el análisis estadístico realizado.



```

    Archivo Editar Selección Ver Ir Ejecutar ... reglas de separacion de patrones.py - Proyecto 1. Conociendo tus datos - Visual Studio Code
    EXPLORADOR PROYECTO 1. CONOCIENDO TUS DATOS
    > glass+identification
    > -$oyecto 1. Conociendo tus dato...
    > atipicos.html
    > Diagramas de caja y valores atípic...
    > glass_identification.py
    > glass.csv
    > IF-THEN.py
    > medidas de dispersion(matrizCorr...
    > Medidas de dispersionUNO.py
    > mm.py
    > PCA.py
    > Proyecto 1. Conociendo tus datos...
    > reglas de separacion de patrones...
    > regresion lineal.py
    > valoresAtipicos.py
    reglas de separacion de patrones.py > ...
    1 import pandas as pd
    2 # Leer el archivo CSV
    3 data = pd.read_csv('glass.csv')
    4 # Definir umbrales para cada atributo basados en el análisis estadístico
    5 umbrales = {
    6     'RI': 1.515,
    7     'Na': 14,
    8     'Mg': 3.5,
    9     'Al': 2.5,
    10    'Si': 73,
    11    'K': 5,
    12    'Ca': 10,
    13    'Ba': 1,
    14    'Fe': 0.15
    15 }
    16 # Clasificar los patrones según las reglas IF-THEN-ELSE
    17 clases = []
    18 for i, row in data.iterrows():
    19     clase = None
    20     if row['RI'] < umbrales['RI']:
    21         clase = 'Clase 1'
    22     elif row['Na'] > umbrales['Na'] and row['Mg'] < umbrales['Mg']:
    23         clase = 'Clase 2'
    24     elif row['Al'] > umbrales['Al'] and row['Si'] > umbrales['Si']:
    25         clase = 'Clase 3'
    26     elif row['K'] > umbrales['K'] and row['Ca'] > umbrales['Ca']:
    27         clase = 'Clase 4'
    28     elif row['Ba'] > umbrales['Ba'] or row['Fe'] > umbrales['Fe']:
    29         clase = 'Clase 5'
    30     else:
    31         clase = 'Clase 6'
    32     clases.append(clase)
    33 # Agregar la columna 'Clase' al DataFrame original
    34 data['Clase'] = clases
    35 # Mostrar los resultados
    36 print(data[['Id number', 'Clase']])
    ~
  
```

Lin. 3, col. 32 Espacios: 4 UTF-8 CRLF Python 3.11.3 64-bit



```

    Buscar
    Ejecutar ... reglas de separacion de patrones.py - Proyecto 1. Conociendo tus datos - Visual Studio Code
    EXPLORADOR PROYECTO 1. CONOCIENDO TUS DATOS
    > glass+identification
    > -$oyecto 1. Conociendo tus dato...
    > atipicos.html
    > Diagramas de caja y valores atípic...
    > glass_identification.py
    > glass.csv
    > IF-THEN.py
    > medidas de dispersion(matrizCorr...
    > Medidas de dispersionUNO.py
    > mm.py
    > PCA.py
    > Proyecto 1. Conociendo tus datos...
    > reglas de separacion de patrones...
    > regresion lineal.py
    > valoresAtipicos.py
    reglas de separacion de patrones.py > ...
    1 import pandas as pd
    2 # Leer el archivo CSV
    3 data = pd.read_csv('glass.csv')
    4 # Definir umbrales para cada atributo basados en el análisis estadístico
    5 umbrales = {
    6     'RI': 1.515,
    7     'Na': 14,
    8     'Mg': 3.5,
    9     'Al': 2.5,
    10    'Si': 73,
    11    'K': 5,
    12    'Ca': 10,
    13    'Ba': 1,
    14    'Fe': 0.15
    15 }
    16 # Clasificar los patrones según las reglas IF-THEN-ELSE
    17 clases = []
    18 for i, row in data.iterrows():
    19     clase = None
    20     if row['RI'] < umbrales['RI']:
    21         clase = 'Clase 1'
    22     elif row['Na'] > umbrales['Na'] and row['Mg'] < umbrales['Mg']:
    23         clase = 'Clase 2'
    24     elif row['Al'] > umbrales['Al'] and row['Si'] > umbrales['Si']:
    25         clase = 'Clase 3'
    26     elif row['K'] > umbrales['K'] and row['Ca'] > umbrales['Ca']:
    27         clase = 'Clase 4'
    28     elif row['Ba'] > umbrales['Ba'] or row['Fe'] > umbrales['Fe']:
    29         clase = 'Clase 5'
    30     else:
    31         clase = 'Clase 6'
    32     clases.append(clase)
    33 # Agregar la columna 'Clase' al DataFrame original
    34 data['Clase'] = clases
    35 # Mostrar los resultados
    36 print(data[['Id number', 'Clase']])
    ~
  
```

Lin. 3, col. 32 Espacios: 4 UTF-8 CRLF Python 3.11.3 64-bit

PROBLEMAS SALIDA CONSOLA DE DEPURACIÓN TERMINAL SQL CONSOLE

```

    PS C:\Users\brand\OneDrive\Documentos\9no Cuatrimestre\Minería de datos\Proyecto 1. Conociendo tus datos & C:/Users/brand/AppData/Local/Programs/Python/Python311/python.exe "c:/Users/brand/OneDrive/Documentos/9no Cuatrimestre/Minería de datos/Proyecto 1. Conociendo tus datos/reglas de separacion de patrones.py"
    Id number Clase
    0 1 Clase 6
    1 2 Clase 6
    2 3 Clase 6
    3 4 Clase 6
    4 5 Clase 6
    ...
    209 ...
    210 ...
    211 ...
    212 ...
    213 ...
    214 ...
  
```

[214 rows x 2 columns]

PS C:\Users\brand\OneDrive\Documentos\9no Cuatrimestre\Minería de datos\Proyecto 1. Conociendo tus datos>

Lin. 3, col. 32 Espacios: 4 UTF-8 CRLF Python 3.11.3 64-bit

i) Conclusiones

En conclusión, los elementos analizados, como la descripción de la base de datos, los tipos de atributos, las medidas de tendencia central, las medidas de dispersión, los diagramas de caja y valores atípicos, la regresión lineal, el análisis de componentes principales (PCA) y las reglas de separación de patrones, son herramientas esenciales en el análisis de datos. Ya que nos permiten comprender la estructura, características y patrones de los datos, lo que a su vez nos ayuda a tomar decisiones informadas y extraer conocimiento significativo.

Además, al describir y comprender la base de datos, podemos establecer un contexto adecuado para su análisis. Al identificar los tipos de atributos presentes, podemos seleccionar las técnicas apropiadas para su manipulación y modelado. Las medidas de tendencia central y dispersión nos dan una idea de la ubicación y variabilidad de los datos, mientras que los diagramas de caja nos ayudan a visualizar su distribución y detectar valores atípicos.

Y finalmente la regresión lineal nos permite modelar relaciones entre variables, lo que es útil para predecir valores o entender la dirección de una asociación. El PCA nos ayuda a reducir la dimensionalidad de los datos y descubrir patrones, mientras que las reglas de separación de patrones nos permiten clasificar y distinguir diferentes clases o grupos.

j) Bibliografía

Referencias

UCI Machine Learning Repository. (s. f.). <https://archive.ics.uci.edu/dataset/42/glass+identification>