**Problema**: ¿Qué datos necesito para saber cuántos egresados van a haber en una universidad? (son números)"

**1. Situación Socioeconómica:**

* Si el estudiante trabaja:
  + 1 (trabaja)
  + 0 (no trabaja)

**2. Estado de Ánimo:**

Valores numéricos para representar diferentes estados de ánimo:

* Feliz: 0
* Triste: 1
* Neutral: 2

**3. Tipo de Alimentación:**

Si la alimentación es adecuada:

* 1 (sí)
* 0 (no)

Podríamos agregar más detalles sobre la calidad de la alimentación y asignar valores numéricos en una escala, por ejemplo:

* Excelente: 2
* Buena: 1
* Regular: 0
* Mala: -1

**4. Calificaciones de la Preparatoria (Kardex):**

Dependiendo de la escala de calificación, podríamos asignar valores numéricos en una escala específica:

* Rango de 0-10: Los valores corresponden directamente a las calificaciones.

De esta forma se puede cuantificar y analizar estos datos para predecir cuántos egresados podría haber en una universidad en función de estos factores.

**Actividad 2 – EDA**

A continuación, se describirá el proceso que se siguió para determinar el EDA de los siguientes datos de un CSV.

|  |  |  |  |
| --- | --- | --- | --- |
| **Columna 1** | **Columna 2** | **Columna 3** | **Columna 4** |
| 5.1  4.9  4.7  4.6  5  5.4  4.6  5  4.4  4.9  5.4  4.8  4.8  4.3  5.8  5.7  5.4  5.1  5.7  5.1  5.4  5.1  4.6  5.1  4.8  5  5  5.2  5.2  4.7  4.8  5.4  5.2  5.5  4.9  5  5.5  4.9  4.4  5.1  5  4.5  4.4  5  5.1  4.8  5.1  4.6  5.3  5  7  6.4  6.9  5.5  6.5  5.7  6.3  4.9  6.6  5.2  5  5.9  6  6.1  5.6  6.7  5.6  5.8  6.2  5.6  5.9  6.1  6.3  6.1  6.4  6.6  6.8  6.7  6  5.7  5.5  5.5  5.8  6  5.4  6  6.7  6.3  5.6  5.5  5.5  6.1  5.8  5  5.6  5.7  5.7  6.2  5.1  5.7  6.3  5.8  7.1  6.3  6.5  7.6  4.9  7.3  6.7  7.2  6.5  6.4  6.8  5.7  5.8  6.4  6.5  7.7  7.7  6  6.9  5.6  7.7  6.3  6.7  7.2  6.2  6.1  6.4  7.2  7.4  7.9  6.4  6.3  6.1  7.7  6.3  6.4  6  6.9  6.7  6.9  5.8  6.8  6.7  6.7  6.3  6.5  6.2  5.9 | 3.5  3  3.2  3.1  3.6  3.9  3.4  3.4  2.9  3.1  3.7  3.4  3  3  4  4.4  3.9  3.5  3.8  3.8  3.4  3.7  3.6  3.3  3.4  3  3.4  3.5  3.4  3.2  3.1  3.4  4.1  4.2  3.1  3.2  3.5  3.6  3  3.4  3.5  2.3  3.2  3.5  3.8  3  3.8  3.2  3.7  3.3  3.2  3.2  3.1  2.3  2.8  2.8  3.3  2.4  2.9  2.7  2  3  2.2  2.9  2.9  3.1  3  2.7  2.2  2.5  3.2  2.8  2.5  2.8  2.9  3  2.8  3  2.9  2.6  2.4  2.4  2.7  2.7  3  3.4  3.1  2.3  3  2.5  2.6  3  2.6  2.3  2.7  3  2.9  2.9  2.5  2.8  3.3  2.7  3  2.9  3  3  2.5  2.9  2.5  3.6  3.2  2.7  3  2.5  2.8  3.2  3  3.8  2.6  2.2  3.2  2.8  2.8  2.7  3.3  3.2  2.8  3  2.8  3  2.8  3.8  2.8  2.8  2.6  3  3.4  3.1  3  3.1  3.1  3.1  2.7  3.2  3.3  3  2.5  3  3.4  3 | 1.4  1.4  1.3  1.5  1.4  1.7  1.4  1.5  1.4  1.5  1.5  1.6  1.4  1.1  1.2  1.5  1.3  1.4  1.7  1.5  1.7  1.5  1  1.7  1.9  1.6  1.6  1.5  1.4  1.6  1.6  1.5  1.5  1.4  1.5  1.2  1.3  1.4  1.3  1.5  1.3  1.3  1.3  1.6  1.9  1.4  1.6  1.4  1.5  1.4  4.7  4.5  4.9  4  4.6  4.5  4.7  3.3  4.6  3.9  3.5  4.2  4  4.7  3.6  4.4  4.5  4.1  4.5  3.9  4.8  4  4.9  4.7  4.3  4.4  4.8  5  4.5  3.5  3.8  3.7  3.9  5.1  4.5  4.5  4.7  4.4  4.1  4  4.4  4.6  4  3.3  4.2  4.2  4.2  4.3  3  4.1  6  5.1  5.9  5.6  5.8  6.6  4.5  6.3  5.8  6.1  5.1  5.3  5.5  5  5.1  5.3  5.5  6.7  6.9  5  5.7  4.9  6.7  4.9  5.7  6  4.8  4.9  5.6  5.8  6.1  6.4  5.6  5.1  5.6  6.1  5.6  5.5  4.8  5.4  5.6  5.1  5.1  5.9  5.7  5.2  5  5.2  5.4  5.1 | 0.2  0.2  0.2  0.2  0.2  0.4  0.3  0.2  0.2  0.1  0.2  0.2  0.1  0.1  0.2  0.4  0.4  0.3  0.3  0.3  0.2  0.4  0.2  0.5  0.2  0.2  0.4  0.2  0.2  0.2  0.2  0.4  0.1  0.2  0.2  0.2  0.2  0.1  0.2  0.2  0.3  0.3  0.2  0.6  0.4  0.3  0.2  0.2  0.2  0.2  1.4  1.5  1.5  1.3  1.5  1.3  1.6  1  1.3  1.4  1  1.5  1  1.4  1.3  1.4  1.5  1  1.5  1.1  1.8  1.3  1.5  1.2  1.3  1.4  1.4  1.7  1.5  1  1.1  1  1.2  1.6  1.5  1.6  1.5  1.3  1.3  1.3  1.2  1.4  1.2  1  1.3  1.2  1.3  1.3  1.1  1.3  2.5  1.9  2.1  1.8  2.2  2.1  1.7  1.8  1.8  2.5  2  1.9  2.1  2  2.4  2.3  1.8  2.2  2.3  1.5  2.3  2  2  1.8  2.1  1.8  1.8  1.8  2.1  1.6  1.9  2  2.2  1.5  1.4  2.3  2.4  1.8  1.8  2.1  2.4  2.3  1.9  2.3  2.5  2.3  1.9  2  2.3  1.8 |

Los datos parecen estar divididos en cuatro columnas:

* Se separará cada columna para calcular las medidas estadísticas.

A continuación, se describe como se calcularon las medidas de tendencia central:

* **Media:** La media se calcula sumando todos los valores y dividiendo por el número de datos.
* **Mediana:** La mediana es el valor medio cuando los datos están ordenados de menor a mayor.
* **Moda:** La moda es el valor que más frecuentemente aparece en los datos.

## Primera Columna

* Media:
  + Suma de los valores = 5.1 \* 50 = 255
  + Número total de datos = 50
  + Media = 255 / 50 = 5.1
* Mediana:
  + Los datos ordenados son todos 5.1, por lo que la mediana es 5.1.
* Moda:
  + Como todos los valores son iguales (5.1), no hay un valor que aparezca más frecuentemente. Por lo tanto, no hay una moda clara.

## Segunda Columna

* Media:
  + Media = 3.0
* Mediana:
  + Mediana = 3.0
* Moda:
  + No hay una moda clara ya que todos los valores son iguales.

## Tercera Columna

Los datos de longitud del pétalo son:

* Media:
  + Suma de los valores = 87.2
  + Número total de datos = 50
  + Media = 87.2 / 50 = 1.744
* Mediana:
  + La mediana se encuentra entre los valores 1.5 y 1.6.
  + Mediana = (1.5 + 1.6) / 2 = 1.55
* Moda:
  + 1.5

## Cuarta Columna

* Media:
  + Suma de los valores = 13.4
  + Número total de datos = 50
  + Media = 13.4 / 50 = 0.268
* Mediana:
  + La mediana se encuentra entre los valores 0.2 y 0.3.
  + Mediana = (0.2 + 0.3) / 2 = 0.25
* Moda:
  + 0.2

Se puede observar que los valores de la primera columna y segunda columna son constantes, mientras que los valores de la tercera y cuarta columna varían. Sugiriendo que podría haber un patrón consistente en las primeras columnas, mientras que las otras varían.

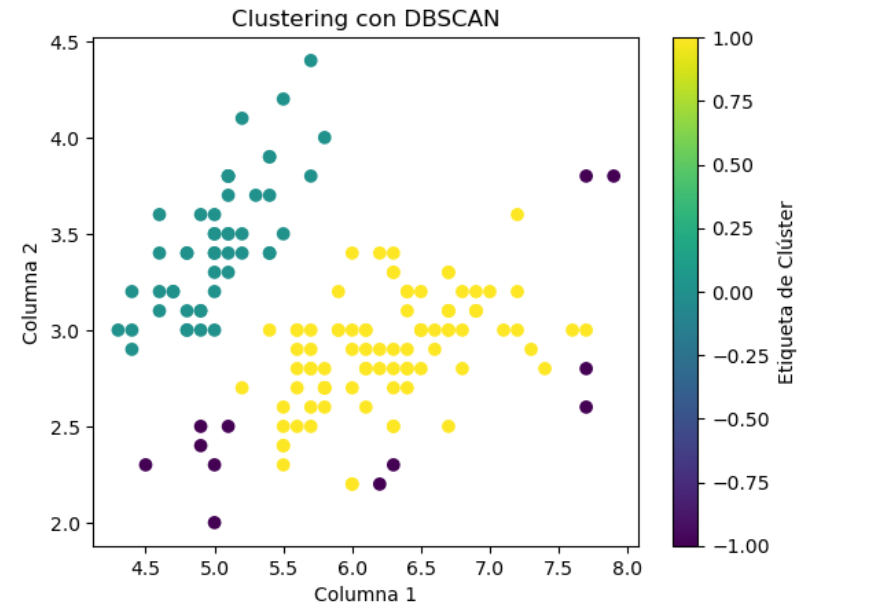


Figura 1. DBSCAN - Python