

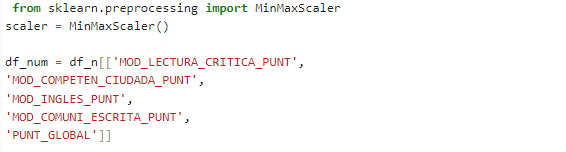


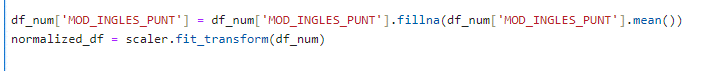
Imagen que contiene Texto

Descripción generada automáticamente

1.2

implementar el algoritmo k-means sin librerías utilizando la distancia de valor absoluto (conocida como la norma 1), crear clase con métodos fit(aprender de los datos) y predict(predecir con los centroides el cluster de un nuevo dato).



Como existen valores nulos , se remplaza estos valores nulos, por la media.  


Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

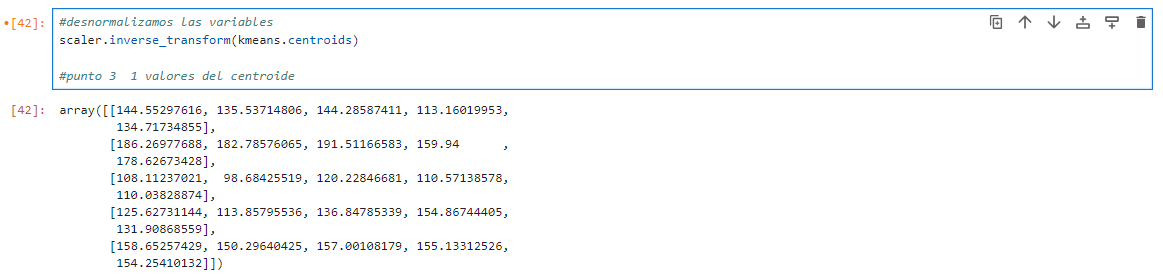
Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

1.3 Crear 5 clusters. Seleccione colocar el centroide de cada cluster, describa que significada cada cluster. Colocar el nombre d ela institucion mas cercana al centroide.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Centroide | 1ra Institución | 2da Institución | ¿que significa el cluster? |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |

Centroide

Para esto de debe de serializa las variables  


1ra Institución y 2da Institución

2 Regresión

¿Qué variables son importantes para predecir el valor?  
Las que tienen correlación alta con la variable objetivo, o inversa mayor a 0,5(directa) y menor a -0,5(inversa)  
  
Existen nulos?, ¿cómo se deben imputar?  
si existen nulos, se imputan dependen si la variable es numérica se imputa por la media y si la variable es categórica por la mas frecuente

Para wuitar valore null

imputer\_categorico = SimpleImputer(strategy='most\_frequent')

datos\_sin\_columnas[columnas\_categoricas] = imputer\_categorico.fit\_transform(datos\_sin\_columnas[columnas\_categoricas])

# Imputar la media en las variables numéricas

imputer\_numerico = SimpleImputer(strategy='mean')

datos\_sin\_columnas[columnas\_numericas] = imputer\_numerico.fit\_transform(datos\_sin\_columnas[columnas\_numericas])

Crear dummy variables para incluirlas en la correlación

# Inicializar el codificador de etiquetas

label\_encoder = LabelEncoder()

# Iterar sobre las columnas categóricas y aplicar Label Encoding

for columna in columnas\_categoricas:

datos\_sin\_columnas[columna] = label\_encoder.fit\_transform(datos\_sin\_columnas[columna])