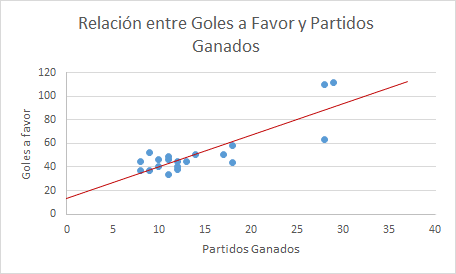
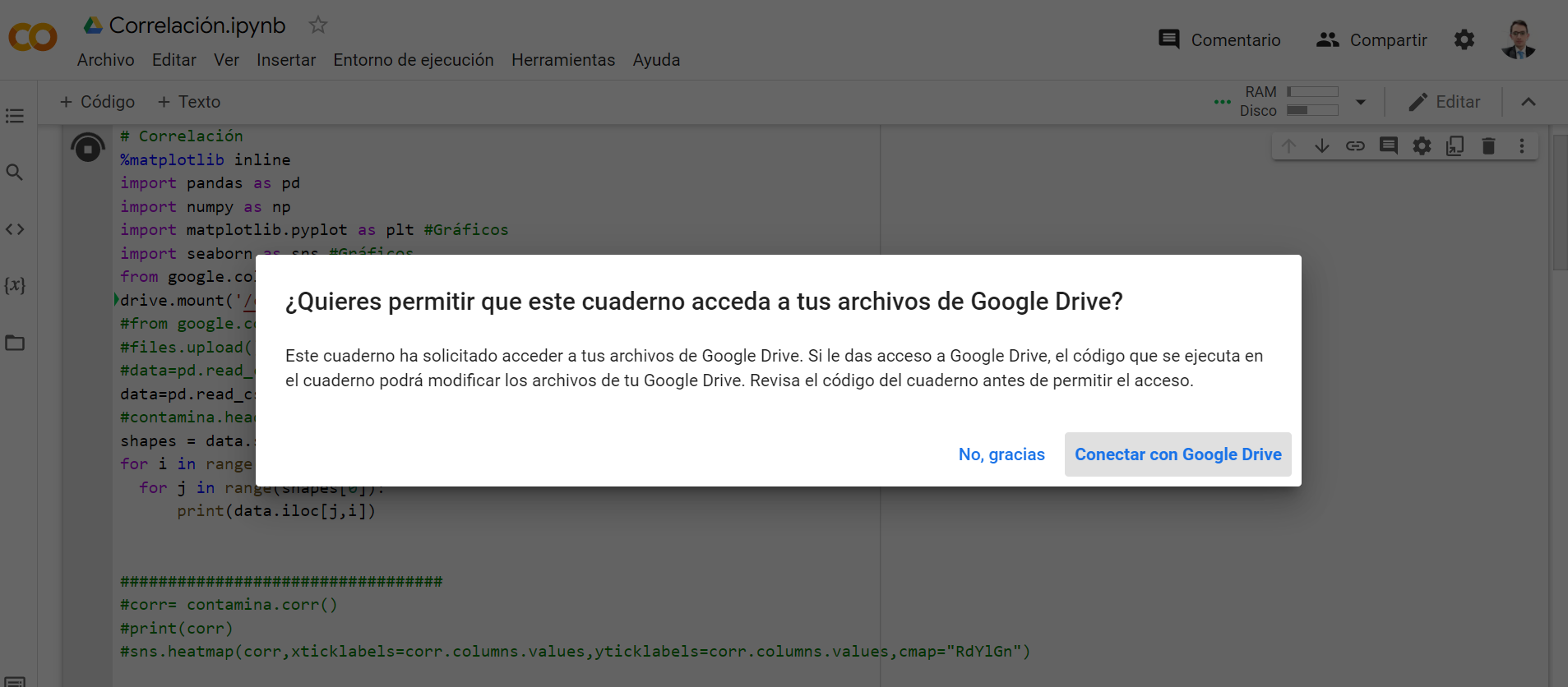
¿Es posible que el tiempo practicando deporte tenga que ver con una vida más longeva? ¿Hay alguna relación entre la cantidad de pan comido al día y el nivel de azúcar medio en sangre? ¿Puedes afirmar que estudiando más horas sacarás mejor notas en el examen de mates? Este tipo de preguntas se pueden responder con correlación lineal teniendo en cuenta que tenemos dos o más **variables numéricas ,** en el que se puede determinar en un plano cartesiano si en el crecimiento de una variable la otra paralelamente también lo hace.



1. Descargar e Instalar [**python**](https://www.python.org/downloads/)
2. Buscar en el menú inicio cmd y ejecutar como administrador y posterior a ello escribir “cd C:\Users\handres\AppData\Local\Programs\Python\PythonX” donde X es la versión de python(ejemplo:39 o 310) donde handres es el nombre de usuario de su computadora. en caso de no existir la ruta verificar la ruta en power bi archivo->opciones y python y en la consola de comandos ejecutar los siguientes comandos (**Nota:** si pip install librería no funciona probar usando python -m pip install librería o en caso de solicitar actualizar pip usamos el comando python -m pip install --upgrade pip)
   1. python -m pip install pandas
   2. python -m pip install numpy
   3. python -m pip install matplotlib
   4. python -m pip install seaborn
3. Para conectarnos a una hoja de cálculo de google lo que debemos hacer es sobre la hoja de cálculo debemos ir a archivo y publicar en la web y seleccionamos todo el documento y en el tipo xls. El enlace generado lo copiamos (https://docs.google.com/spreadsheets/d/e/2PACX-1vR6yxL8Qy7ZaGhe4y2pE0WFLtsfd6NqMSq403tFFONVM0R1z3FBWUpyyhY2QDJCXS926Cx\_EAe9vZR7/pub?output=csv)
4. En power Bi vamos a obtener desde una web y pegamos el enlace y nos conectamos para posteriormente transformar datos
5. Si tenemos filas iniciales en blanco podemos usar la primera fila como encabezado o quitar dichas filas
6. A la columna que tiene el texto le llamaremos contenido y en la pestaña transformar seleccionamos la opción ejecutar script de python
7. Pegamos el script :



criterios=['op','co','ex'] #Columnas de criterios a analizar

cantidad\_cluster=5 #cantidad de conjuntos

nombre\_columna\_analizar="usuario" #columna a usar

nombre\_nueva\_columna="cluester" # Columna nueva

#####

import pandas as pd #

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sb

from sklearn.cluster import KMeans

from sklearn.metrics import pairwise\_distances\_argmin\_min

rango\_criterios = np.array(dataset[criterios])

kmeans = KMeans(n\_clusters=cantidad\_cluster).fit(rango\_criterios)

centroids = kmeans.cluster\_centers\_

closest, \_ = pairwise\_distances\_argmin\_min(kmeans.cluster\_centers\_, rango\_criterios)

closest

total=[centroids]

columnas={}

contadortotal=0

nombrecategoria=''

for n in centroids:

columnas[contadortotal]=n

contadortotal=contadortotal+1

clusters\_finales = pd.DataFrame(columnas)

res=np.transpose(clusters\_finales)

nombrescolumnas=list()

for columna in criterios:

nombrescolumnas.append(columna)

res.columns = [nombrescolumnas]

##########

columna\_de\_coeficientes=list()

for i in dataset[nombre\_columna\_analizar].index:

coeficientes=list()

for criterio in criterios:

coeficientes.append(dataset.loc[i, criterio])

conjunto\_coeficientes= np.array([coeficientes])

coeficiente\_por\_registro = kmeans.predict(conjunto\_coeficientes)

columna\_de\_coeficientes.append(coeficiente\_por\_registro)

dataset[nombre\_nueva\_columna]= pd.DataFrame(columna\_de\_coeficientes,columns=[nombre\_nueva\_columna],dtype ='string')

print(dataset)

1. Si notifica sobre privacidad de los datos pulsamos sobre continuar y sobre las listas de niveles de publicidad en todas dejamos público y continuar
2. Desplegamos sobre el botón junto a la columna de value las columnas que presentaran los sentimientos, verificamos que este las 3 seleccionadas y aceptar
3. Cambiamos el nombre de la columna y en la pestaña de transformar tipo de datos decimal de punto fijo debido a que se encuentra como texto para manipularlo de mejor forma y no podríamos comparar texto con valores para el siguiente paso
4. Agregamos una nueva columna personalizada con el nombre categoria sentimiento y usamos la fórmula

if [sentimiento] < 50 then "NEGATIVO" else if [sentimiento] >50 then "POSITIVO" else if [sentimiento]=50 then "NEUTRAL" else null

**Nota:** No olvidar que entre los corchetes va el nombre del campo que contiene el coeficiente de sentimientos

**Nota:** si deseamos leer estos numeros en notacion cientifica podemos acceder a https://tulengua.es/numeros-texto/