**Cargue de datos**

# importar las librerias

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib as mpl

import matplotlib.pyplot as plt

%matplotlib inline

# Cargar la base de datos en excel

data = 'guatemala\_cos\_fy19\_eng\_final.xlsx'

df1 = pd.read\_excel(data)

**Inspección**

# Buscar la cantidad de registros que hay, identificar que tipo de dato es, si hay valores unicos y su hay valores nulos (en blanco).

print('Valores nulos por columna: shape', df1.shape)

print('Columnas \t tipo\túnicos\tnulos')

print('--'\*20)

for col in df1.columns.to\_list():

    print('%s:    \t%s\t%3d/%d\t%2d\t'%(col, df1[col].dtype, len(df1[col].unique()),len(df1),len(df1[df1[col].isnull()])))

**Limpieza**

# Cambio del tipo de dato, todas las variables a float64

df1['id']= df1['id'].astype('float64')

df1['a1']= df1['a1'].astype('float64')

df1['a2']= df1['a2'].astype('float64')

df1['h1']= df1['h1'].astype('float64')

df1['h7']= df1['h7'].astype('float64')

df1['h8']= df1['h8'].astype('float64')

df1['h1r']= df1['h1r'].astype('float64')

# Reemplazar valores nulos por cero

df1 = df1.replace(np.nan, 0)

# Exportar a CSV

df1.to\_csv('guatemala\_cos\_fy19.csv', index=False)