INTELIGENCIA ARTIFICIAL

WORKSHOP3

PRESENTADO A:

GONZALO GOMEZ MILLAN

PRESENTADO POR:

JEISON ANDRES FUENTES ORTEGA

230172007

UNIDAD CENTRAL DEL VALLE DEL CAUCA

ING SISTEMAS

SEMESTRE VIII

2021

**ANALISIS CODIGO WORKSHOP3**



En esta primera parte del código se importa las librerías que se van a usar en el ejercicio

Librería Warnings -> Entrega advertencias no fatales al usuario sobre los problemas encontrados al ejecutar un programa.

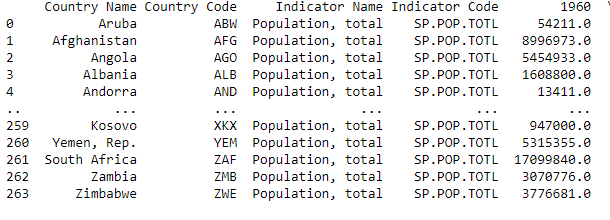
Librería Pandas -> Pandas es una librería de Python especializada en el manejo y análisis de estructuras de datos.

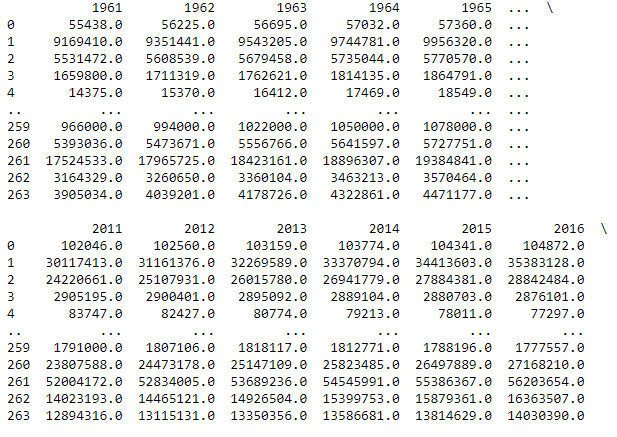
Librería Matplotlib -> una biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays

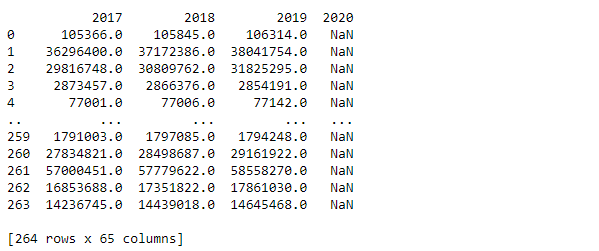
Después de importar las librerías

En una variable llamada data se le asigna un read\_excel que permite leer un archivo de Excel en un DataFrame de pandas por medio de una URL y para extraer información del dataset o archivo Excel se debe instalar la librería xlrd.

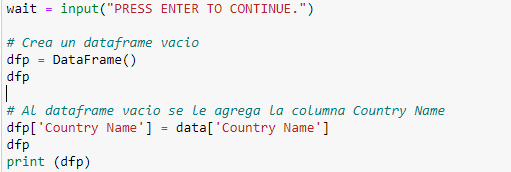
Al correr esta parte del código nos mostraría la información de la URL



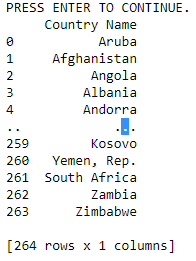




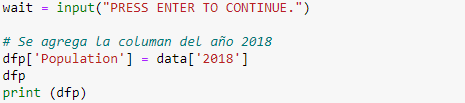
En las siguientes líneas de códigos tenemos



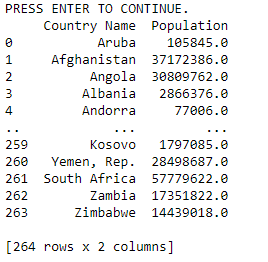
Encontramos wait que funciona para detener el programa y en este caso para que pueda continuar se debe presionar enter, luego se crea un dataframe vacío, un dataframe es una estructura de datos con dos dimensiones en la cual se puede guardar datos de distintos tipos en columnas, al dataframe creado se le agregara la columna de country name (nombre del pais) del dataset y posteriormente mostrarlo.



Siguiendo el código



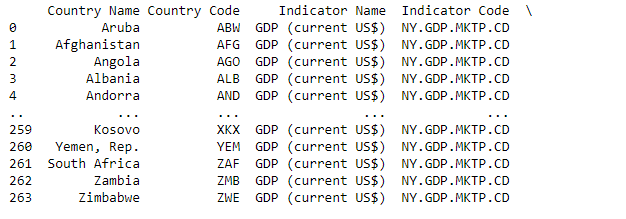
Volvemos a tener el wait, agregamos al dataframe, creado anteriormente y que ya contiene la columna country name, la columna population con la del año 2018 del dataset y la mostramos al presionar enter.

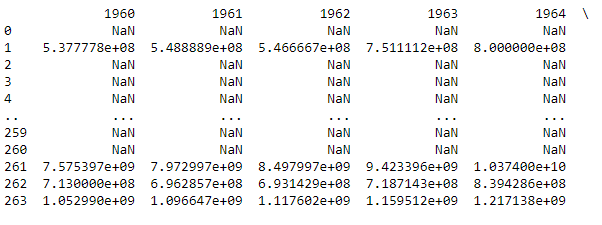


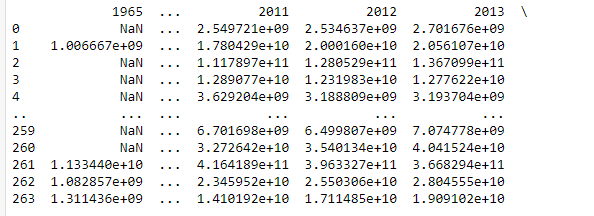
Siguiendo el código

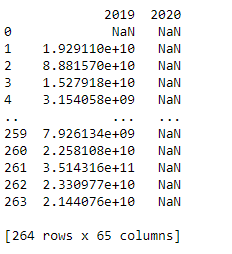


Volvemos a tener el wait, a la variable data le asignamos información de otro dataset, y se muestra al presionar enter.

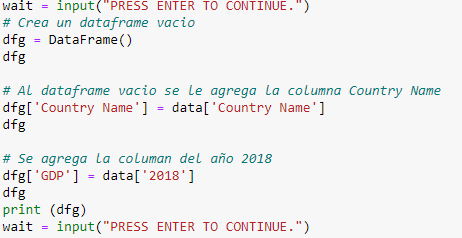








Siguiendo el código

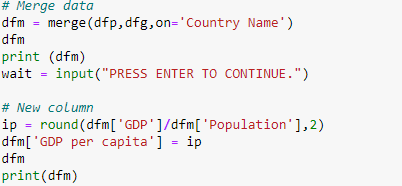


Se vuelve a crear un dataframe vacío con la variable dfg

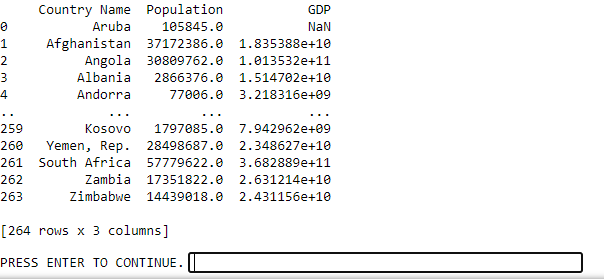
A la cual le asignamos las columnas de country name y GDP de la información del dataset nuevo, la información del GDP es de la columna del año 2018 del dataset, al presionar enter nos mostrara lo que se realizo



Al continuar con el código tenemos

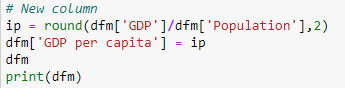


Se crea una variable dfm en la cual se inserta un merge entre los dataframe, dfp y dfg, el merge es un join que une las tablas, al correrlo nos arroja lo siguiente.

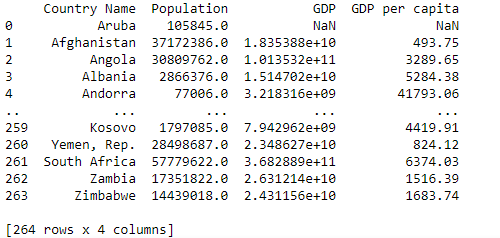


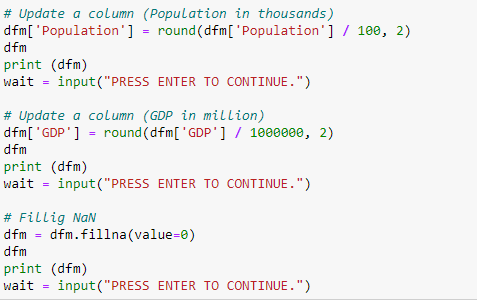
Se puede observar que nos muestra 3 columnas country name, population y GDP trayendo los datos por medio del country name, una especie de llave primaria, lo cual genera un tercer dataframe.

Siguiendo el código.

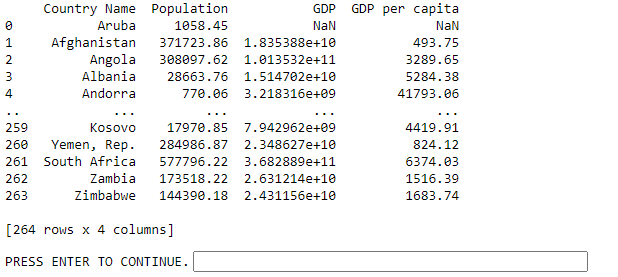


Se crea una variable llamada ip en la cual se encuentran los datos de la división entre la columna GDP y population del dataframe dfm redondeado a dos decimas, y al dataframe dfm se le crea una nueva columna llamada GDP per capita y se le agrega ip, después se realiza un print para mostrar lo realizado.

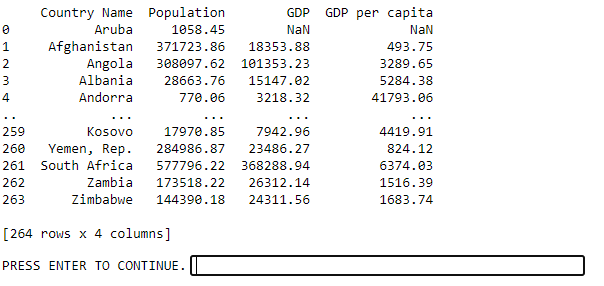




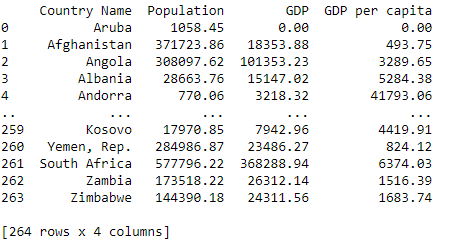
Actualizamos la columna population del dataframe dfm dividiendo cada fila de este en cien y redondeándolo a dos decimales, al presionar enter nos arroja la actualización.



Luego se modifica la columna GDP deividiendo cada fila en un millón con un redondeo de 2 decimales arrojando los siguientes datos.

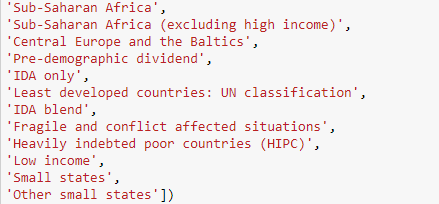


Los valores nulos que se encuentren en el dataframe dfm se quedaran en 0 gracias al método fillna como se muestra a continuación:

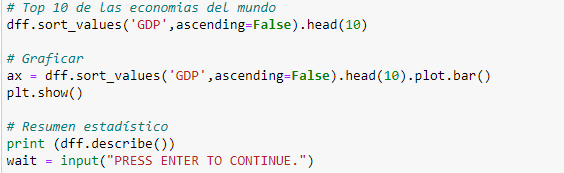


Siguiendo con el código, hay dos formas de eliminar datos de las filas, una seria eliminando por código y otra forma es eliminando por palabras que contengan en la columna country name que en este caso seria, con los métodos set\_index para escoger la columna en la cual quiera buscar y el drop para eliminar ya sea por código o por palabras que contengan las filas.



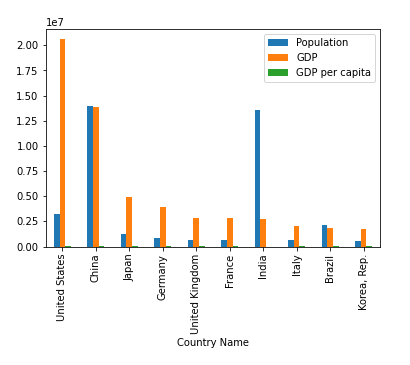


Siguiendo el código



Tenemos el método sort\_values que nos sirve para ordenar ya sea de forma descendente o ascendente en este caso sería la columna GDP

Luego vemos que utilizamos el método plot.bar para generar una gráfica de forma descendente y mostrándola posteriormente con el método plt.show().



Y por ultimo vemos el método describe() el cual devuelve información estadística de los datos del dataframe. Esta información incluye el número de muestras, el valor medio, la desviación estándar, el valor mínimo, máximo, la mediana y los valores correspondientes a los percentiles 25% y 75%.

