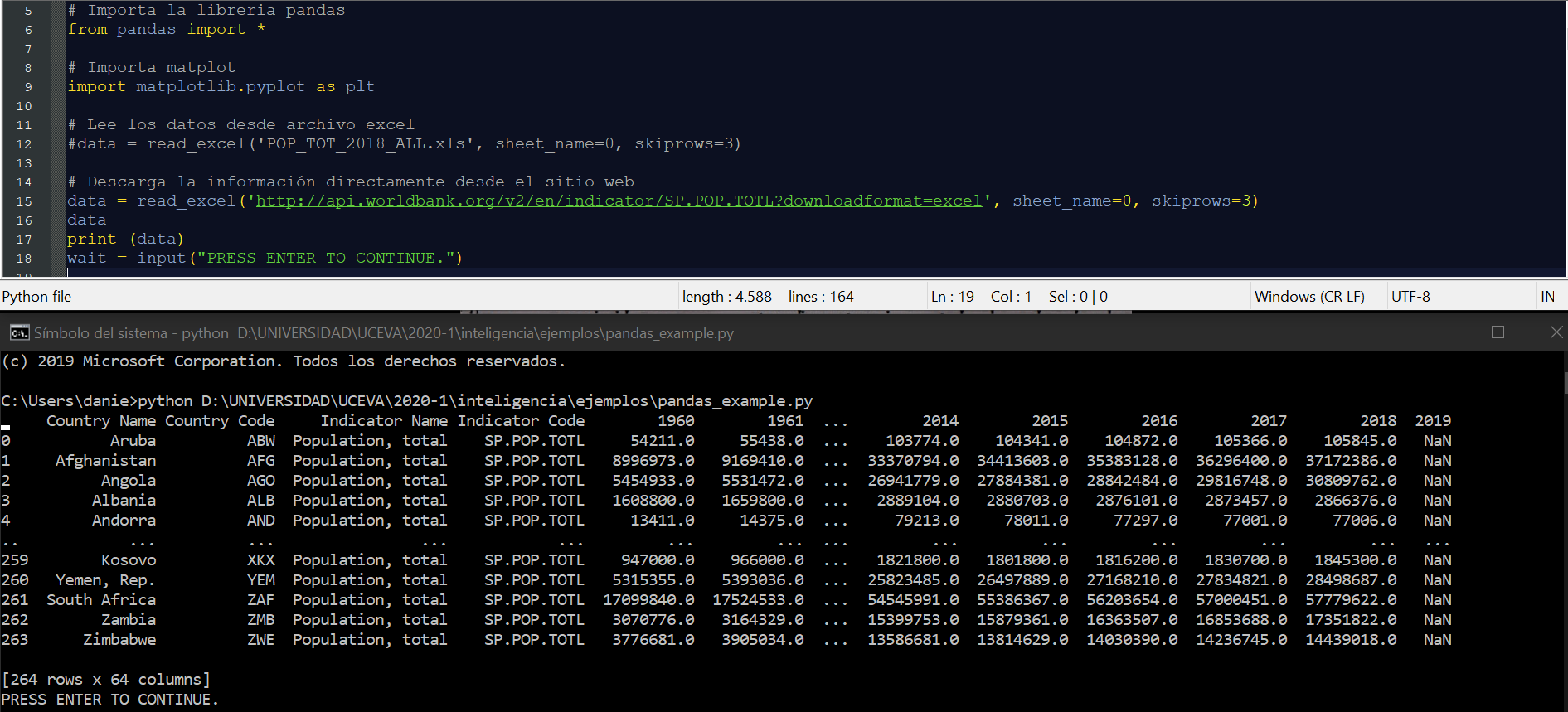
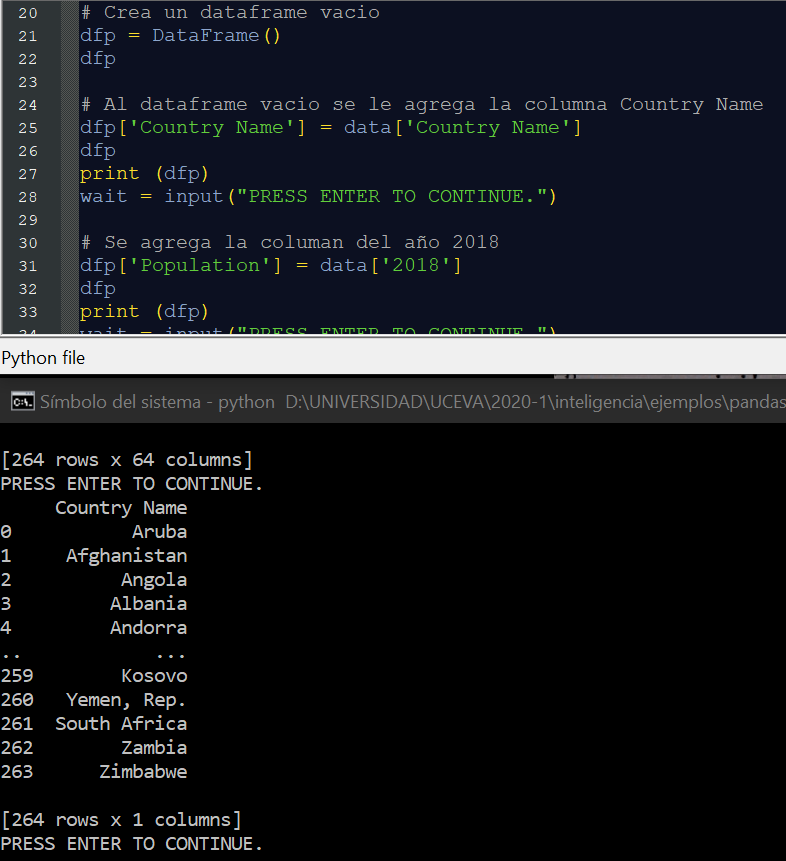
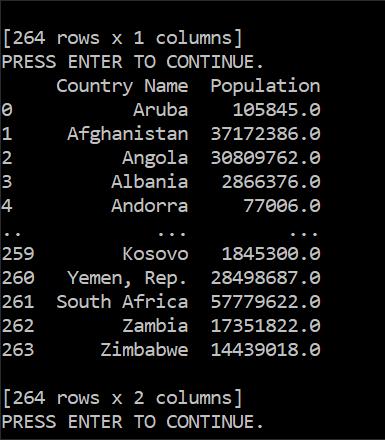
Taller 2

* En las primeras líneas del ejemplo vemos que se importan librerias necesarias para el ejemplo (warnigs, pandas, matplotlib.pyplot).
* Mas adelante vemos que directamente de la pagina del banco mundial obtenemos la información con la función read\_excel en donde vemos 2 atributos. El primero es para indicar que hoja queremos obtener y el segundo es para indicarle que se salte 3 filas antes de empezar a leer en dicha hoja.

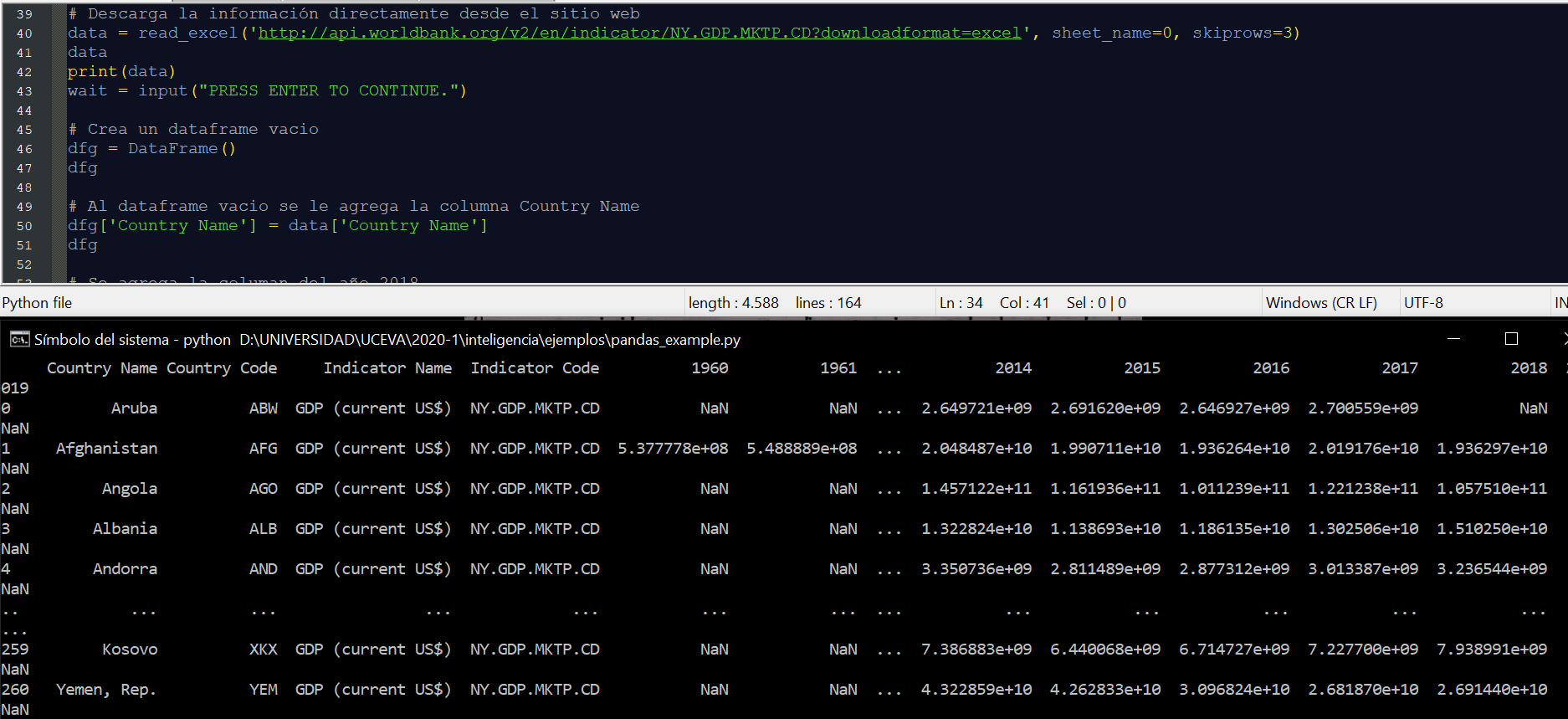


* Vemos que el archivo tiene la función de esperar una entrada y es por ello que el programa no continua hasta dar la tecla enter, después vemos la creación de un dataframe vacío para empezar a organizar la información extrayéndolo de la primer tabla que obtuvimos de internet, se le agrega la columna del nombre del país bajo el titulo de “country name” se muestra luego se le añade como segunda columna “population” la población del 2018 y se muestra de nuevo, haciendo las paradas respectivas de la función “wait”.

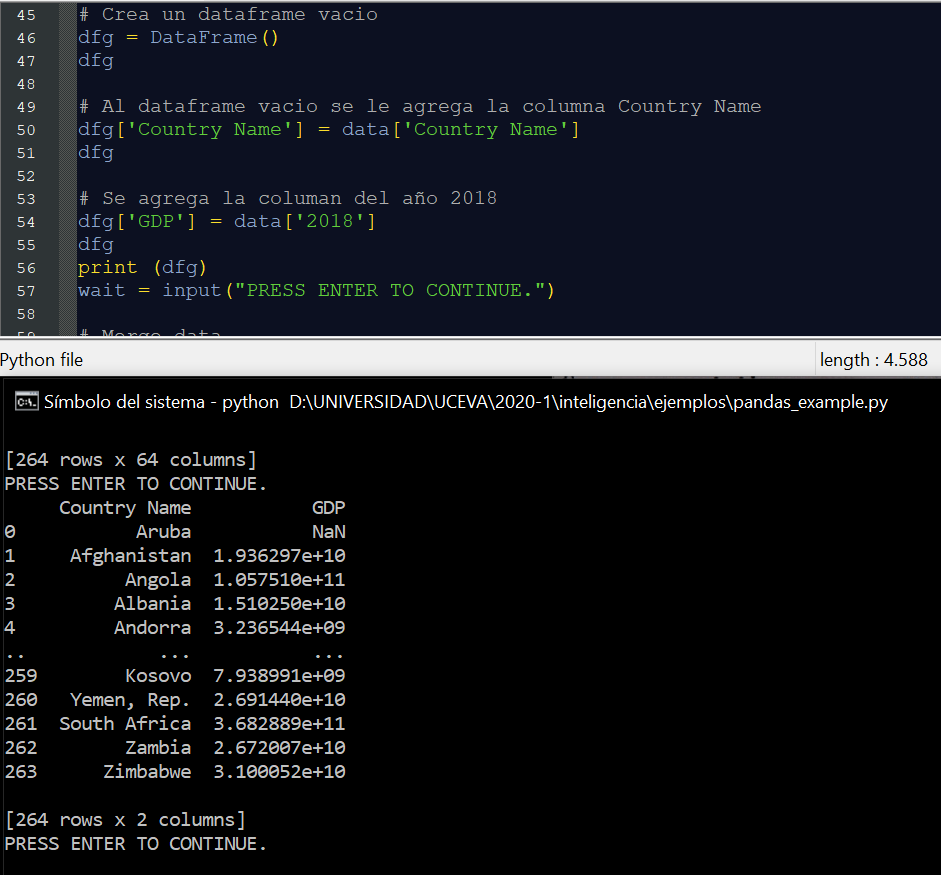




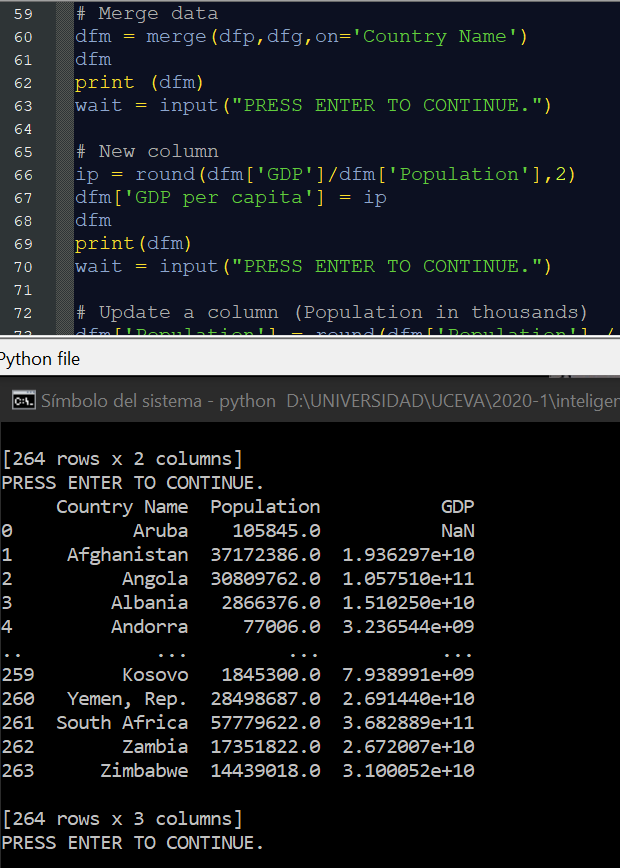
* Se lee un segundo archivo directamente desde internet (cabe aclarar que en los comentarios esta el ejemplo de como obtener datos de un archivo local descargado) y se muestra en las mismas condiciones del archivo anterior (primera hoja dejando 3 filas)



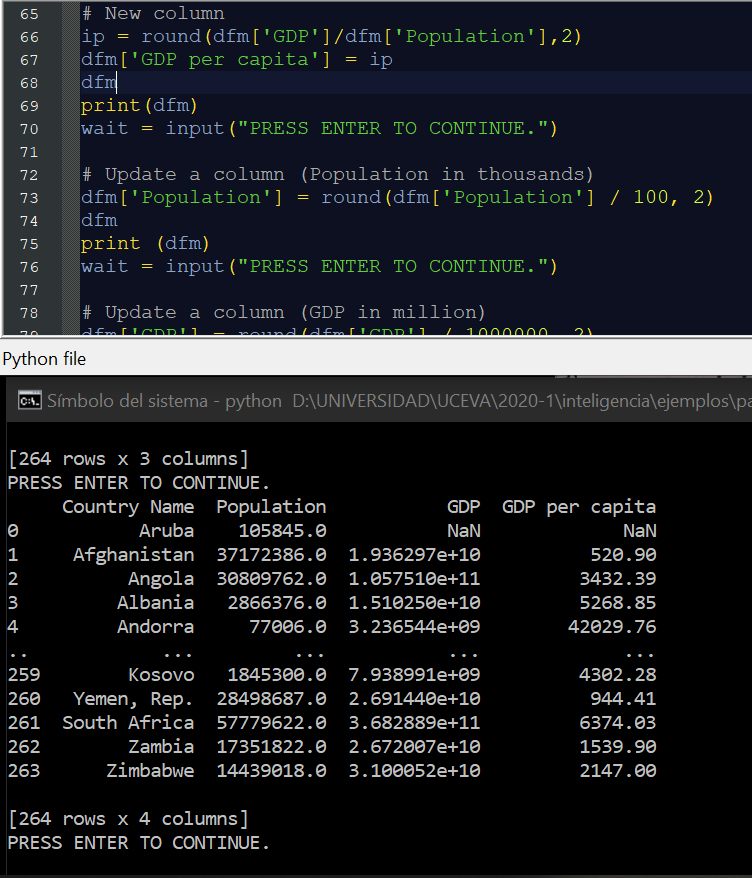
* Nuevamente creamos un dataframe vacio en donde colocamos las columnas “country name” y “GDP” que corresponde al pib del año 2018

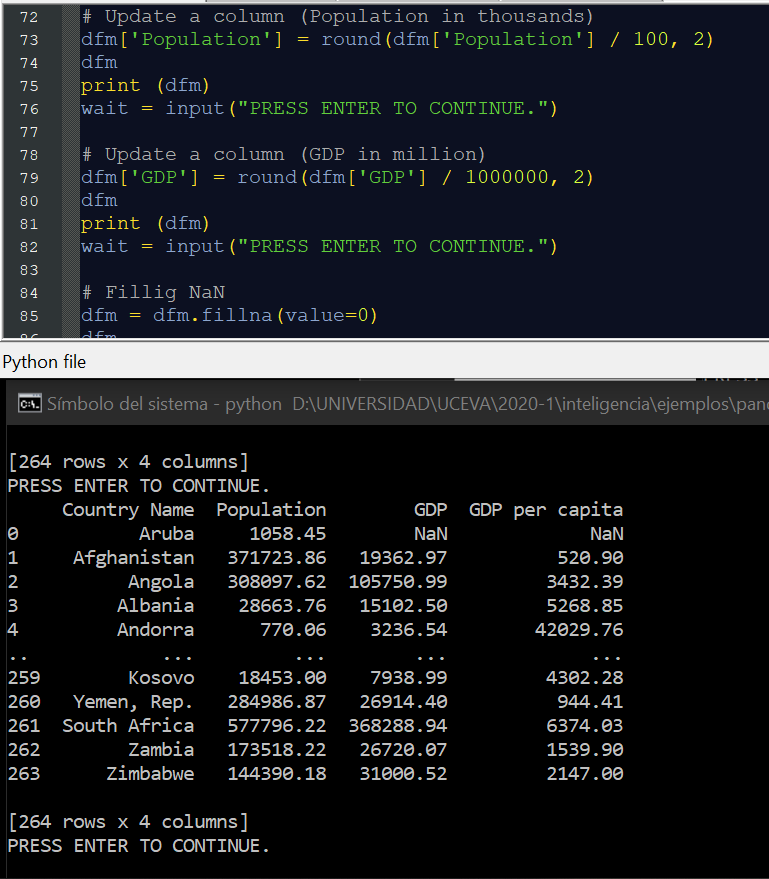


* Ahora con la función merge se crea un dataframe con la información de los dos dataframe anteriormente generados (dfp y dfg) los cuales se van a mezclar en la columna “country name” que da como resultado un dataframe de 3 columnas

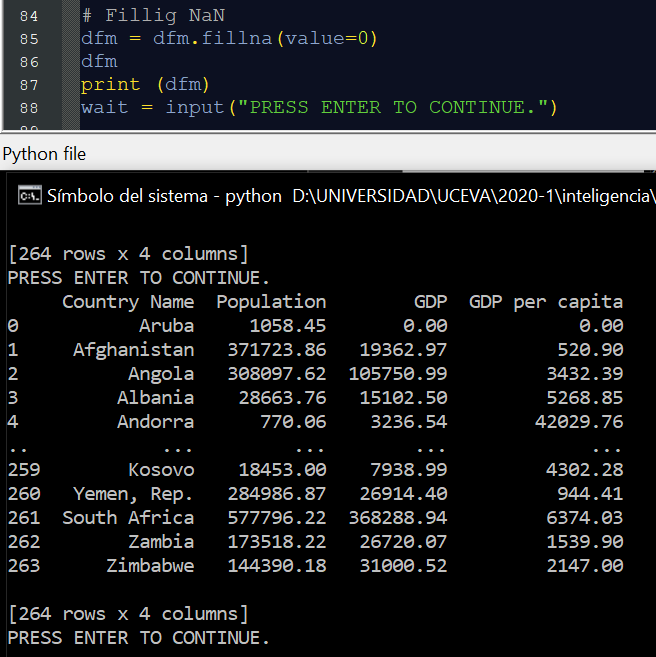


* Con ayuda de una variable adicional creamos el pib per capita simplemente dividiendo el pib entre la población y con la función “round” le ponemos solamente 2 decimales y se la añadimos a nuestro dataframe, luego para que quede mejor presentado actualizamos las columnas de población a cientos y pib a millones y también con la función “round” le ponemos solamente 2 decimales

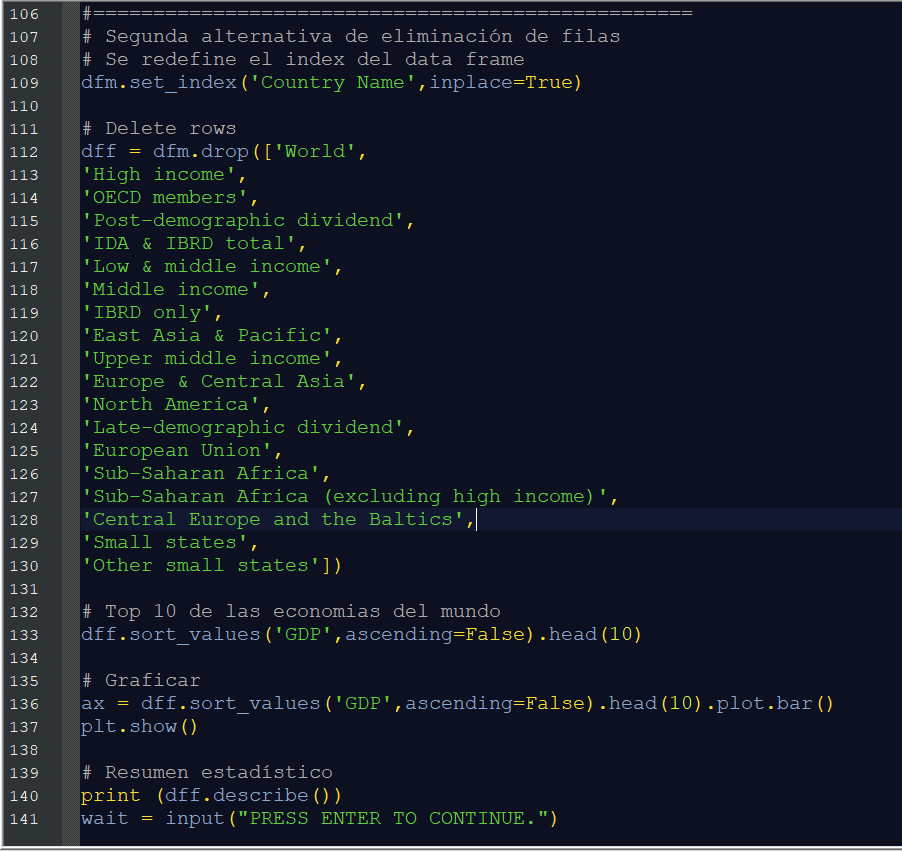


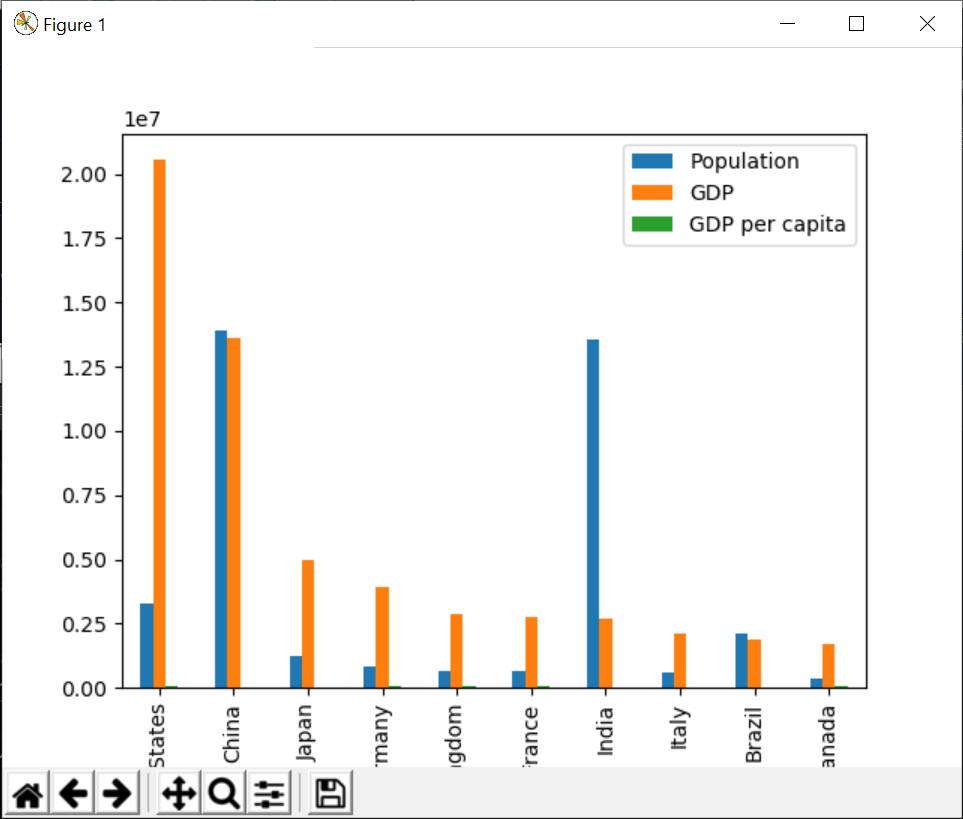


* Para que quede una tabla mas legible y después nos permita graficar debemos remplazar los NaN (not a number) por 0.0 por lo tanto con la función “fillna” le generamos ese valor.



* En lo que sigue del código, se nos explican dos formas de eliminar filas, luego se le define un nuevo index al dataframe, se muestran las 10 mas grandes economías ordenandolas decendentemente y se grafica con ayuda de la librería matplotlib en un plot de barras





* Después con la función describe se genera un resumen estadístico donde encontramos información como: count, mean, std, min, los cuartiles y max

