

ESCUELA POLITÉCNICA NACIONAL

ESCUELA DE FORMACIÓN DE TECNÓLOGOS **DESARROLLO DE SOFTWARE**

INFORME DE ANÁLISIS DE DATOS

Limpieza, Visualización e Interpretación de Datos. Sobre el Índice de Desarrollo Humano (IDH)

Elaborado por: Jimy Calvo M.

Martin Rosero B.

Fecha de Entrega: 12-Sept-2022

Periodo Académico2022-A.



Resumen

En este informe se presentará una de las diversas formas de extracción, manejo y limpieza de datos obtenido desde un documento (en el caso presentado en este informe, será desde un documento en formato Excel). Siendo el caso de estudio el "Índice de Desarrollo Humano conocido también como (IDH)" donde se extraerá, limpiará, analizará los datos obtenidos sobre este. Cuya finalidad es aplicar los conocimientos de sobre la extracción y almacenamiento de una base de datos.

Para el proceso de extracción y limpieza de los datos se lo realizara mediante el entorno de trabajo interactivo conocido como "Jupyter", entorno donde se emplear el lenguaje de programación "Python". Además, del uso de librerías que gestiona conjuntos de datos como lo son Pandas, Numpy y otras librerías para la realización de gráficos. Donde dichos gráficos serán de utilidad para la interpretación de los datos y su posterior análisis.

"Sin análisis de grandes volúmenes de datos, las empresas son ciegos y sordos, vagando hacia fuera sobre la web como ciervos en una autopista."

- Geoffrey Moore

Tabla de contenido

Resumen	1
Tabla de Ilustraciones	3
1. Introducción	4
1.1. Objetivos	4
1.1.1. General:	4
1.1.2. Específicos:	4
1.2. Planteamiento del problema	4
2. Marco Teórico	5
2.1. IDH – Índice de desarrollo Humano	5
2.2. Python	5
2.3. Scripting	5
2.4. Jupyter	6
Librerías:	6
2.5. Pandas	6
2.6. Numpy	6
2.7. Matplotlib	6
2.8. Seaborn	7
2.9. Statsmodels	7
2.10. Py Mosaic	7
Comandos de las librerías empleadas	8
3. DESARROLLO	8
3.1. Instalación e Importación de Comandos	8
3.2 Obtención de los Datos	9
3.2. Integridad de Datos	.2
3.3. Limpieza	.5
3.4. Visualización de los datos	.8
Graficas en Mosaico	.9
Graficas en Barras	.9
4. Conclusiones	<u>!</u> 1
Bibliografía2	22

Tabla de Ilustraciones

FIG. 1. CAPTURA DE LA EJECUCIÓN DE LOS COMANDOS DE EXTRACCIÓN DE DATOS	10
FIG. 2. CAPTURA DE LA EJECUCIÓN DE LOS COMANDOS DE EXTRACCIÓN DE DATOS	11
FIG. 3. GRAFICA DE UN MAPA DE CALOR DE LOS VALORES VACÍOS POR COLUMNA	11
FIG. 4. CAPTURA DE LA EJECUCIÓN DE LOS COMANDOS DE LA VARIABLE QUE CONTIENE LOS DATOS	12
FIG. 5. CAPTURA DE LA EJECUCIÓN DE LOS COMANDOS DE LA VARIABLE QUE CONTIENE LOS DATOS.	13
FIG. 6. CAPTURA DE LA EJECUCIÓN DE LOS COMANDOS DE LA VARIABLE QUE CONTIENE LOS DATOS	14
FIG. 7. CAPTURA DE LA EJECUCIÓN DE LOS COMANDOS .DROPNA() DE LA VARIABLE QUE CONTIENE LOS DATOS	15
FIG. 8. CAPTURA DE LA EJECUCIÓN DE LOS COMANDOS .DROPNA() DE LA VARIABLE QUE CONTIENE LOS DATOS	15
FIG. 9. CAPTURA DE LA EJECUCIÓN DE .SORT_VALUES() DE LA VARIABLE QUE CONTIENE LOS DATOS	16
FIG. 10. GRAFICA DE UN MAPA DE CALOR DE LOS VALORES DE LAS COLUMNAS CON DATOS	17
FIG. 11. CAPTURA DE LA EJECUCIÓN DEL DATAFRAME QUE CONTIENE LOS DATOS.	17
FIG. 12. CAPTURA DE LA EJECUCIÓN DEL DATAFRAME QUE CONTIENE LOS DATOS.	18
FIG. 13. GRAFICA TIPO MOSAICO DEL TOP 10 DEL RANKIG DEL IDH	19
FIG. 14. GRAFICA TIPO BARRA DEL TOP 10 DEL RANKIG DE PAÍSES Y EL IDH	19
FIG. 15. GRAFICA TIPO BARRA DE LOS ÚLTIMOS 10 PAISES DEL RANKIG DEL IDH	20
FIG. 16. GRAFICA DE ÁREA DE LOS ÚLTIMOS 10 PAÍSES DEL RANKIG DEL IDH	20

1. Introducción

El proyecto consiste en el empleo de los conocimientos sobre los procesos de extracción, limpieza, graficación e identificación de datos obtenidos desde un documento digital o sitio web, empleando los conocimientos adquiridos en la asignatura de Análisis de Datos de la carrera Técnica de Desarrollo de Software. Dichos procesos se lo aplicarán a datos reales, obtenidos desde un documento digital, siendo el caso de estudio de este informe, los datos obtenidos sobre el Índice de Desarrollo Humano [1]. A estos datos se aplicará los procesos anteriormente mencionados, para su posterior análisis de estos. Durante estos procedimientos se emplear herramientas (software como Jupyter) y librerías de Python de gestión y vista de datos, con la finalidad de facilitar el estudio de los datos recopilados y llegar a convertir los datos en ideas y posteriormente en conocimiento.

1.1.Objetivos

1.1.1. General:

 Analizar la información disponible del archivo sobre el Índice de Desarrollo Humano a su elección, para la limpieza, visualización e interpretación de datos.

1.1.2. Específicos:

- Aplicar los conceptos de Análisis de datos en un caso real con la finalidad de llegar a la interpretación.
- Identificar las funciones y librería a utilizar para conversión de datos en conocimiento.
- Interpretar la información obtenida una vez realizada la limpieza,
 preparación y visualización de datos

1.2. Planteamiento del problema

Se busca la extracción de datos mediante el uso de herramientas de análisis de

datos empleado el lenguaje de programan de alto nivel Python junto a las librerías y gestión y graficacion de datos, y todo esto dentro del entono de Júpiter. Y para esto se ha planteado la extracción de los datos del IDH el cual obtener desde el sitio web de Human Develoment Reports. Acto seguido, empleando Jupiter se extraerá los datos de la Tabla sobre Human Development Index and its components. A dichos datos se creará el código correspondiente parala limpieza, y visualización de datos empleado las respectivas librerías de Python. Y para finalizar, se llevará a cabo la interpretación los datos

2. Marco Teórico

2.1. IDH – Índice de desarrollo Humano

Según el sitio web de Human Development Reports es una media que resume el crecimiento de una nación bajo ciertos criterios como son el PIB, nivel de educación, esperanza de vida, entre otros parámetros cuya finalidad es estimular al debate sobre las necesidades de cada país, en el ámbito de fomentar el desarrollo de este. A demás de servir como medidor sobre qué tan desarrollado se encuentra un país con respecto a los demás [2].

2.2. Python

Python es un lenguaje de programación que es utilizado en varios sectores, este lenguaje se lo puede encontrar en aplicaciones web, en el desarrollo de software, la ciencia de datos (análisis de Datos) y el machine learning (ML) [3]. Python también es un lenguaje de alto nivel orientado a objetos es decir que es un lenguaje ideal para el desarrollo de aplicaciones, así como scripting. Es el leguaje que emplea Jupyter dentro de su entorno por defecto. Este lenguaje incluye librerías que permite la gestión de datos, Dicho lenguaje será empleado en el proceso de extracción de aquí su importancia de conocer sobre este.

2.3. Scripting

Es un lenguaje de programación que permite ejecutar funciones.

Relaciona los objetos de resultados y la interfaz de usuario, además de poder ejecutar sintaxis de comandos. En Python es empleado para personalizar una tabla dinámica [4].

2.4. Jupyter

Es una base de un entorno informático interactivo donde se puede procesar datos como en el caso de la computación científica, la ciencia de datos y el análisis el cual emplea la tecnología de los navegadores web para ejecutar. Este entorno presenta la función de cuaderno el cual permite la creación de documentos que combinan código en vivo con texto narrativo, ecuaciones y visualizaciones [5].

Librerías:

2.5. Pandas

Es una herramienta de análisis y manipulación de datos de código especializada en el manejo y análisis de estructuras de datos. El cual permite leer y escribir fácilmente archivos en distintos formatos como son CSV, Excel, y bases de datos en SQL [6].

2.6. Numpy

Es una librería que permite el cálculo numérico y análisis de datos el cual es ideal para manejar datos de gran columna ya que permite la manipulación de datos mediante arreglos los cuales son colecciones de datos de un mismo tipo lo que permite la rapidez y eficiencia de la manipulación de datos [7].

2.7. Matplotlib

Es una herramienta que permite la creación de varios tipos de gráficos como también la personalización de estos entre los distintos tipos de graficas tenemos:

- Diagramas de áreas
- Diagramas de barras
- Diagramas de caja y bigotes
- Diagramas de contorno
- Diagramas de dispersión o puntos
- Diagramas de líneas
- Diagramas de sectores

- Diagramas de violín
- Histograma
- Mapas de color

Al utilizar esta librería, creará objetos Figure y Axes los cuales pueden llamará a sus métodos para agregar contenido y modificar la apariencia [8].

2.7.1. Pylab

Es un modulo que incluye algunas funciones que permite manipular la interfaz vasado en los distintos estados de la figura generada por matplotlib.

2.8. Seaborn

Librería que permite la visualización de datos mediante gráficos estadísticos de un conjunto de datos, el cual está basado en matplotlib con una estética mejora al momento de generar las gráficas, que a su vez proporciona más información [9].

2.9. Statsmodels

Es un módulo de Python que permite crear modelos estadísticos de diferentes formas y lo realiza mediante la proporción de clase y funciones que admite datos desde una lista de datos las cuales pueden ser de tipo DataFrames. Los resultados se prueban con los paquetes estadísticos existentes para garantizar que sean correctos [10].

2.9.1. `Graphics.Mosaicplot

Permite la creación de un gráfico de tipo mosaico a partir de una tabla de contingencia. Esta a su vez permite la visualizar datos en diversas categorías multivariados de forma rigurosa e informativa.

2.10. Py Mosaic

Es un modulo de datos que permite representar los datos mediante una simulación molecular, el cual permite también la realización de cálculos, y este se puede combinar con códigos, datos, y documentación [11].

Comandos de las librerías empleadas

Tabla 1. Módulos Empleado Para extracción, limpieza y grafica de datos

	Funciones Usadas
Comando	Funciones
pd.read_excel	Función para leer un archivo.xlsx
df.info()	Función para mostrar la información de un DataFrame
df.head()	Función que imprime los primeros registros de un DataFrame
df.tail()	Función que imprime los últimos registros de un DataFrame
df.isna().sum().sort_values()	Función conjunta que imprime la cantidad total de valores nulos
sns.heatmap(df.isna())	Función que imprime los valores nulos de manera grafica
df.dropna()	Función que elimina los registros que contengan algún valor nulo
df.rename(columns={})	Función que renombra las columnas seleccionadas por medio de un diccionario
df.reset_index(drop=True, inplace=True)	Función que crea un Índex nuevo reemplazando al anterior
df.set_index()	Función que te permite establecer una columna existente como Índex
mosaic(df,[])	Función para graficas de Mosaico
df.plot(kind='barch')	Función para graficas de filas
df.plot.area()	Función para graficas de líneas con área

Resumen de los comandos empleados de cada Liberia de Python y su funcionamiento, el cual será empleado para la extracción de datos limpieza y su posterior visualización

3. DESARROLLO

3.1.Instalación e Importación de Comandos

Para la extracción, limpieza y visualización de los datos Jupyter requiere que se instale e importe los módulos y librerías par ello se lo realiza mediante las siguientes líneas de comando

```
%pip install matplotlib
%pip install seaborn
%pip install modelo
%pip install pyMosaic
%pip install statsmodels
```

```
#IMPORTACION DE LIBRERIAS
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.graphics.mosaicplot import Mosaic
```

3.2 Obtención de los Datos

Para el proceso de obtención de datos, sea empleado la librería panda junto al comando read el cual permite leer el archivo obtenido de una ruta.

```
#FUNCION PARA LEER ARCHIVO EXCEL

def leer_datos(ruta):
    df = pd.read_excel(ruta)
    return df
```

Dichos datos serna almacenados en una variable df que permite la extracción de los datos.

```
Output exceeds the size limit. Open the full output data in a text editor
-----Informacion de la data-----
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 266 entries, 0 to 265
Data columns (total 9 columns):
# Column
                                      Non-Null Count Dtype
                                      190 non-null object
0
  Unnamed: 0
                                     255 non-null object
1 Unnamed: 1
2 Human Development Index (HDI)
                                    212 non-null object
3 Life expectancy at birth
                                     212 non-null object
4 Expected years of schooling
                                    212 non-null object
5 Mean years of schooling
                                     212 non-null object
6 Gross national income (GNI) per capita 212 non-null object
    GNI per capita rank minus HDI rank
                                    211 non-null object
                                      211 non-null
   HDI rank
                                                   object
dtypes: object(9)
memory usage: 18.8+ KB
-----Encabezado de los data-----
 Unnamed: 0
                          Unnamed: 1 Human Development Index (HDI)
0 HDI rank
                             Country
1
                                                          2019
       NaN VERY HIGH HUMAN DEVELOPMENT
                                                          NaN
 GNI per capita rank minus HDI rank HDI rank
                            NaN
                                   NaN
1
                           2019
                                   2018
2
                            NaN
                                   NaN
```

Fig. 1. Captura de la ejecución de los comandos de extracción de datos. (Fuente: Propia)

Para el proceso de limpieza se realiza la búsqueda de valores nulos para ello se ejecuta los siguientes comandos.

```
#CONTAVILIZAMOS LOS VALORES NULOS DEL DATAFRAME

df_Original.isna().sum().sort_values()

#Esto nos denota que existen dos columnas sin identificar
#Ademas de que existe datos nulos constantes en todas las demás columnas
```

Unnamed: 1	11
Human Development Index (HDI)	54
Life expectancy at birth	54
Expected years of schooling	54
Mean years of schooling	54
Gross national income (GNI) per capita	54
GNI per capita rank minus HDI rank	55
HDI rank	55
Unnamed: 0	76
dtype: int64	

Fig. 2. Captura de la ejecución de los comandos de extracción de datos. (Fuente: Propia)

Con el fin de representar dichos valores se emplea la librería **Seaborn** el cual permitirá representar dichos valores en una grafica en este caso de tipo heatmap.



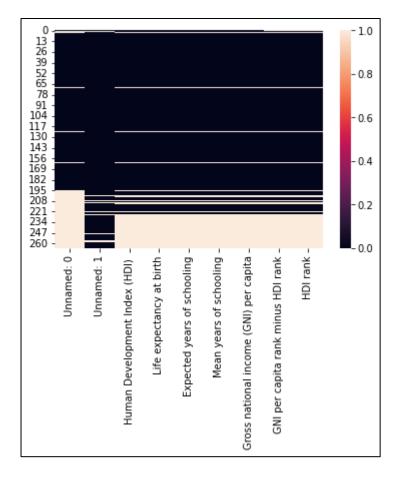


Fig. 3. Grafica de un mapa de calor de los valores vacíos por columna (Fuente: Propia)

Para una mayor visualización de los datos se imprime los DataFrame que contiene la variable.

#IMPRIMIREMOS EL DATAFRAME PARA LOGRAR IDENTIFICAR LA
UBICACIÓN DELOS DATOS NULOS

df_Original

#Podemos observar cómo los datos NaN se ubican especialmente
en la última sección del DATAFRAME

#Se denota que existen además datos no correspondientes en
varias columnas

#Recordaremos los años en los que se hicieron estos datos ya
que es importante para su análisis

	Unnamed: 0	Unnamed: 1	Human Development Index (HDI)	Life expectancy at birth	Expected years of schooling	Mean years of schooling	Gross national income (GNI) per capita	GNI per capita rank minus HDI rank	HDI rank
0	HDI rank	Country	Value	(years)	(years)	(years)	(2017 PPP \$)	NaN	NaN
1	NaN	NaN	2019	2019	2019	2019	2019	2019	2018
2	NaN	VERY HIGH HUMAN DEVELOPMENT	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3		Norway	0.957	82.4	18.06615	12.89775	66494.25217		1
4		Ireland	0.955	82.31	18.70529	12.666331	68370.58737		3
61	NaN	Column 2: UNDESA (2019a).	NaN	NaN	NaN	NaN	NaN	NaN	NaN
62	NaN	Column 3: UNESCO Institute for Statistics (202	NaN	NaN	NaN	NaN	NaN	NaN	NaN
63	NaN	Column 4: UNESCO Institute for Statistics (202	NaN	NaN	NaN	NaN	NaN	NaN	NaN
64	NaN	Column 5: World Bank (2020a), IMF (2020) and U	NaN	NaN	NaN	NaN	NaN	NaN	NaN
65	NaN	Column 6: Calculated based on data in columns	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6 row	s × 9 columns								

Fig. 4. Captura de la ejecución de los comandos de la variable que contiene los datos. (Fuente: Propia)

3.2.Integridad de Datos

Para este proceso primero se verificara los datos obtenidos para ello se ejecutara el comando head de la Liberia panda el cual permite visualizar los datos de un determinado número de filas.

2 NaN VERY HIGH LIMAN DUYLOMENT Nan Nan Nan Nan Nan Nan Nan 3 1 Norway 0.957 82.4 18.06615 12.89775 66494.22217 7 1 4 2 Netseland 0.955 82.31 18.70629 12.666331 68370.93737 4 3 3 2 5 2 Switzerland 0.955 83.78 16.28947 12.27996 6393.2376 3 2 6 4 Hong King China (SAR) 0.949 84.66 16.92947 12.27996 6290.47553 7 4 4 6 6 64.4 19.0819 12.772787 54602.38057 14 4		Unnamed: 0	Unnamed: 1	Human Development Index (HDI)	Life expectancy at birth	Expected years of schooling	Mean years of schooling	Gross national income (GNI) per capita	GNI per capita rank minus HDI rank	HDI rank
1	0	HDI rank	Country	Value	(years)	(years)	(years)	(2017 PPP \$)	NaN	
Section Content	1	NaN		2019						
4 2 between the section of the section	2	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN
5 2 Sementener 0.958 0.878 1.520 0.050 0.	3		Norway		82.4	18.06615	12.89775	66494.25217		
6 4 Phong Group Chenk 3.098 M.56 1.02279 Cale (Section) COMINGATION A PART OF CHENK 7 4 6 Germany 0.047 8.13 1.9779 1.41146 55143355 1.1 4.2 8 6 Germany 0.045 8.13 1.97179 1.45146 55143555 1.1 4.2 1.0 <	4		Ireland	0.955		18.70529	12.666331	68370.58737		
7 4 Solend 30,86 5.00 7.00 7	5		Switzerland	0.955		16.32844	13.380812	69393.52076		
6 Genmany CMR CMR </td <td>6</td> <td></td> <td>Hong Kong, China (SAR)</td> <td>0.949</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	6		Hong Kong, China (SAR)	0.949						
1	7		Iceland	0.949	82.99	19.08309		54682.38057		
10	8		Germany	0.947	81.33	16.97719	14.15168	55314.35355		
1	9		Sweden	0.945		19.48234	12.54847	54507.80504		
12 10 Operandar 0.98 8.98 8.98 1.98 <t< td=""><td>10</td><td></td><td>Australia</td><td>0.944</td><td>83.44</td><td>21.95433</td><td>12.724691</td><td>48084.84207</td><td></td><td></td></t<>	10		Australia	0.944	83.44	21.95433	12.724691	48084.84207		
13 In Palame 0.038 8.19 9.038772 2.02413 4.0513378 11 1.1 16 11 Sergence 0.038 0.032 0.	11		Netherlands	0.944		18.48513	12.4148	57707.06867		
14 11 Grappore 0.038 CRAS 16.4008 11.2461 381521431 2.8 1.2 16 13 United Kingdom 0.931 0.13 17.7488 12.05007 250445009 0.8 1.8 16 14 Meyor Zealard 0.931 0.22 18.08307 12.78007 40761749 0.18 0.18 16 14 Meyor Zealard 0.931 0.22 18.08307 12.78007 40761749 0.18 0.18 0.18 0.18 0.18 0.18 0.18 0.18 0.18 0.18 0.18 0.18 0.18 0.00 0.93 0.23 0.18 0.14 0.00 0.18 0.23 0.14 0.14 0.00 0.18 0.23 0.14 0.14 0.00	12		Denmark	0.94	80.9	18.89342	12.613803	58661.87084		
15 13 United Englane 6.93 6.1.22 17.6868 13.24007 2000 (4481) 3 1.8 1.9			Finland	0.938			12.82413			
16 14 Melgium 0.931 0.140 19.77430 12.25077 2004,5000 6 17.07 17 14 New Zealand 0.931 0.222 18.8887 12.72077 40.981790 15 44 18 15 Canada 0.929 0.243 16.15789 13.431841 6825.53748 .7 .71 19 17 United States 0.929 0.248 16.10027 13.41344 6818.55748 .7 .71 21 19 Austral 0.919 0.237 16.16283 10.04854 4018.65853 14 .2 21 19 Austral 0.919 0.432 15.2057 12.8 4.931.6573 .9 .0 22 19 Depart 0.919 0.823 16.16283 12.03444 30016.5583 14 .2 24 22 Soveren Specific of 0.919 0.823 14.52057 12.8814 30016.5583 .1 .2 .2 .2 <td></td>										
17			United Kingdom							
18 16 Canada 0.92 0.243 16.15789 13.366105 48.527.05751 5 14 19 17 United States 0.026 7.886 18.1019 13.1414 68282.56744 7.77 17.87 21 19 Anatria 0.092 16.164 16.00007 12.264614 6916,00060 3 18 21 19 Anapan 0.019 0.627 16.16283 13.2857 12.28 4.9818,0693 14 2.72 21 19 Lucchtenstein 0.019 0.667 14.9023 12.5354 131011509 15 2.42 24 22 Soonen 0.017 61.32 17.5766 12.25176 43817107 41 2.22 12.23765 43817107 41 2.22 2.22 12.23765 43817107 41 2.22 2.22 12.23765 43817107 41 2.22 2.22 2.22 12.23755 42.22 12.23735 43817107 41 2.22	16		Belgium	0.931	81.63	19.77438	12.05027	52084.59209		
19 17 United States 0.926 78.86 16.3109 13.41344 0.525.65748 7.77 7.77 20 18 Austra 0.922 81.54 16.09267 12.46444 551.08988 3 18 21 19 Bayan 0.919 0.297 16.16283 13.3384854 4.911.69755 9 22 22 19 Bayan 0.919 0.663 15.2967 12.85 4.291.16975 9 20 24 29 Shorena 0.917 61.32 17.57206 12.58584 3007.953399 16 22 25 23 Kores (Republic of) 0.916 8.03 16.68165 12.23166 4004371077 4 22 26 23 Lucembourg 0.916 6.25 14.25562 12.3116 4097452400 6 2.25 27 2 Farus 0.90 0.57 17.6161 10.25178 407715248 1.1 2.22 28	17		New Zealand	0.931		18.83857	12.782077	40798.7196		
20 18 Austral 0.922 0.154 16.09207 12.546144 6616,68969 18 21 19 Irasel 0.919 0.297 16.16283 13.43844 4018,684383 14 22 22 19 Japan 0.919 0.843 15.23977 12.85 4.931,8575 9 20 23 19 Uschimettini 0.919 0.843 15.23975 12.81844 13101,3586 16 18 19 24 22 Kors Sepulation 0.916 0.23 16,48165 12,211876 40043,71057 4 22 25 23 Lusembourg 0.916 0.25 14,25582 12,211876 40043,71057 4 22 26 23 Lusembourg 0.916 0.25 14,25582 12,211876 40774,53482 19 2 26 23 Lusembourg 0.916 0.25 14,25582 12,21186 47772,53484 4778 14,2518 <	18		Canada	0.929	82.43	16.15789	13.366105	48527.03573		
21 19 krant 0.919 0.297 16.16203 13.034854 4.0186.84583 1.2 2.2 19 kapan 0.919 0.843 15.23057 12.85 4.291.69758 9 2.2 2.2 19 kinchtenten 0.919 0.867 14.00233 12.53854 4.3101.15388 1.8 19 2.4 1.2 2.000min 0.919 0.812 17.75706 12.68874 30.0753.399 15 2.2 2.2 30.000min 0.916 0.825 14.25582 12.213165 40.9753.399 15 2.2 2.2 12.213765 40.911.67412 1.9 2.2 12.213765 40.911.67412 1.9 2.2 12.213765 40.211.67412 1.9 2.2 1.2 </td <td>19</td> <td></td> <td>United States</td> <td>0.926</td> <td>78.86</td> <td>16.31039</td> <td>13.41344</td> <td>63825.65748</td> <td></td> <td></td>	19		United States	0.926	78.86	16.31039	13.41344	63825.65748		
22 19 Japan 0.919 84.63 152.2057 12.85 42931.69575 9 20 23 19 Lischtersten 0.919 0.867 14.400233 12.233844 131031.5986 1-8 19 24 22 Soveria 0.916 83.03 16.48165 12.213765 43043.71057 14 22 25 23 Lusembourg 0.916 82.25 14.25582 12.211765 43043.71057 1 2 26 23 Lusembourg 0.916 82.25 14.25582 12.91146 4971.7253748 -1 22 26 23 France 0.901 82.66 15.64217 11.4795 47172.53748 -1 26 29 27 Crecks 0.90 78.38 16.78211 12.772382 3510.55662 9 2 31 29 Estoria 0.99 78.75 15.999 13.14001 36019.26692 -1 3 32			Austria	0.922	81.54			56196.89869		
23 19 Liechtentein 0.919 0.067 1.40023 12.53854 13101.5898 -18 19 24 22 Soowea 0.917 0.912 17.57706 12.65874 38079.53399 15 24 25 23 Korea Republic of 0.916 82.35 14.2582 12.31145 72711.67412 4 22 26 23 Lucembourg 0.916 82.25 14.2582 12.31145 72711.67412 4 22 27 25 Span 0.904 83.57 17.51613 10.25176 49974.52408 6 6 22 28 27 Cachela 0.91 82.66 15.64217 11.272382 38108.5662 6 6 2 30 28 Cachela 0.99 75.35 16.10407 11.32613 39554.52076 6 2 31 29 Estons 0.99 28.51 16.0407 11.32613 30154.6438 12.4 2			Israel							
24 22 Soverein Republic of 0.917 0.132 17.57206 12.65874 30.079.53399 15 24.24 25 23 Korea Republic of 0.916 0.823 16.48165 12.213765 43043.71677 4 22 26 23 Lumembourg 0.916 0.825 14.2582 12.31145 17.1167412 0.919 2.52 27 25 Spain 0.904 0.826 15.64171 11.4795 47172.5348 .1 2.62 28 25 France 0.901 0.826 15.64171 11.4795 47172.5348 .1 2.62 30 27 Chechia 0.901 82.53 16.169107 11.22613 39054.5278 .0 2.62 31 27 Chechia 0.905 78.25 15.909 13.143061 39054.5278 .0 2.0 31 23 Chall 0.909 77.77 14.3441 12.1112 67462.0953 .0 2.0				0.919	84.63			42931.69575		
25 23 Korea (Nepublic of) 0.916 83.03 16.48165 12.213765 43043.71057 4 22 26 23 Luxembourg 0.916 82.25 14.25882 12.31145 72711.67412 1.9 23 28 25 Spain 0.901 82.66 15.64217 11.4795 47172.53746 -1 26 28 26 France 0.901 82.66 15.64217 11.4795 47172.53746 -1 26 29 27 Czechia 0.93 79.38 16.78511 12.722382 38108.56662 9 26 30 28 Malta 0.995 82.33 16.0407 11.22613 39594.52078 6 28 30 29 Estoria 0.992 78.75 15.98 13.14061 3619.56622 9 20 31 29 Estoria 0.999 77.77 14.3441 12.1122 6762.0933 -24 30 32 <										
26 23 Lucembourg 0.916 62.25 14.2582 12.31145 72711.67412 1.9 2.3 27 28 Spain 0.904 0.837 17.61613 10.25178 4.09745,24008 6.6 25.2 29 27 Cachcia 0.91 79.38 16.167511 12.272382 38108,56662 9.9 2.6 30 28 Malta 0.95 26.33 16.10407 11.32613 39554,52078 6.6 2.8 31 29 Estoria 0.95 28.31 16.00407 11.32613 39554,52078 6.6 2.8 31 29 Batul 0.895 83.51 16.0089 10.37955 42776,55002 0.6 2.9 33 31 United Arab Emirates 0.99 77.97 13.3441 12.1112 67462,0953 -24 3.0 34 32 Greece 0.888 22.24 17.9079 12.7112 357642,0953 -24 3.0			Slovenia	0.917	81.32		12.65874	38079.53399		
27 25 Spain 0.904 83.57 17.61913 10.25178 49974,52408 6 25 28 26 France 0.901 82.66 15.64217 11.4795 47172,33748 -1 26 29 27 Czechia 0.9 79.38 16.78111 12.22282 31300,66662 9 26 31 28 Malta 0.995 6.23 16.10407 11.32613 36954,52978 6 26 31 29 Estonia 0.98 78.75 15.989 13.143061 36019,5692 9 30 32 29 Haly 0.89 83.51 16.0889 10.37955 42776,5692 0 6 29 34 32 Greece 0.88 82.24 17.90792 10.55207 30154,62438 14 33 35 33 Cyprus 0.887 80.34 15.0897 12.17122 3806,64049 2 2 45 33 <td>25</td> <td></td> <td>Korea (Republic of)</td> <td>0.916</td> <td>83.03</td> <td></td> <td></td> <td>43043.71057</td> <td></td> <td></td>	25		Korea (Republic of)	0.916	83.03			43043.71057		
28 26 France 0.901 82.66 15.64217 11.4795 47172.53748 1 26 29 27 Czechia 0.9 79.38 16.78511 12.722382 38108.05662 9 26 30 28 Malta 0.095 82.33 16.10407 11.32613 39554.52078 6 28 31 29 Estonia 0.095 83.51 15.099 13.14061 36019.26692 9 30 32 29 Italy 0.09 77.97 14.3441 12.11122 67.62.0953 24 30 33 31 United Arabimiates 0.09 77.97 14.3441 12.11122 67.62.0953 24 30 34 32 Gerece 0.088 62.24 17.9092 10.55207 30154.63483 14 33 35 33 Cyprus 0.087 0.09 15.16867 12.17123 3506.64049 2 32 36 34			Luxembourg	0.916	82.25	14.25582	12.31145	72711.67412		
29 27 Cxx-chia 0.9 79.38 16.78511 12.722382 38108.56662 9 26 30 28 Malta 0.895 82.33 16.10407 11.32613 39554.52078 6 28 31 29 Estonia 0.895 78.75 15.989 31.43001 3619.26692 6 28 32 29 haly 0.992 83.51 16.0889 10.37955 42776.55692 0 29 33 31 United Arab Emirates 0.89 77.97 14.3441 12.11122 67462.0953 -24 30 34 32 Greece 0.88 82.24 17.90792 10.55207 30154.6438 2.2 32 35 33 Cyprus 0.88 82.24 17.90792 10.55207 30154.6438 2.2 32 36 34 Uthania 0.88 75.33 16.6731 13.0771 35786.64414 5 35 35 37 <td>27</td> <td></td> <td>Spain</td> <td>0.904</td> <td>83.57</td> <td></td> <td></td> <td>40974.52408</td> <td></td> <td></td>	27		Spain	0.904	83.57			40974.52408		
30 28 Maita 0.095 62.33 16.10407 11.32613 39554.52078 6 28 31 29 Statonia 0.092 78.75 15.969 13.143061 36019.6662 9 20 32 29 Italy 0.892 83.1 16.089 10.37955 42776.55662 0 22 33 13 United Arab Emirates 0.89 77.97 14.3441 12.11122 67462.0953 -24 30 34 32 Greece 0.888 82.24 17.59072 10.55207 301546548 14 33 35 33 Cypris 0.887 80.24 17.59072 10.55207 301546548 14 33 37 35 Opeland 0.08 78.73 16.69731 13.07721 35708.67041 5 35 37 35 Opland 0.08 78.73 16.69731 13.07721 35708.67041 5 36 37 Latvia	28		France	0.901	82.66	15.64217	11.4795	47172.53748		
31 29 Estonia 0.092 78.75 15.989 13.143061 36019.26692 9 30 32 29 taty 0.892 83.51 16.0889 10.37955 4.276.55692 0.9 29 34 31 United Arab Emirates 0.88 82.24 17.90792 10.55207 30154.64348 14 33 35 32 Gerece 0.88 82.24 17.90792 10.55207 30154.6438 14 33 35 33 Cyprus 0.887 80.88 15.16887 12.17123 38206.64049 2 22 36 34 Ultuaria 0.88 75.33 16.90171 12.47232 38206.64049 2 2 36 34 Ultuaria 0.88 75.33 16.90171 12.47232 31622.55392 8 3 39 37 Latvia 0.866 75.23 16.16769 13.0098 30202.33333 8 3 41 3	29		Czechia			16.78511	12.722382	38108.56662		
32 29 Italy 0.092 8.351 16,0089 10,37955 42776,35692 0 29 33 31 United Arabi Emirates 0.09 77.97 14,3441 12,11122 6746,20933 -24 30 34 32 Greece 0.888 82,24 17,90792 10,55207 30154,63438 14 32 35 33 Cyprus 0.887 80.98 15,16887 12,17123 38206,64049 2 32 36 34 Lithuania 0.882 75.93 16,66731 13,0771 379,0676741 5 35 37 35 Pollular 0.88 75.93 16,6731 13,07721 379,0676741 5 35 38 36 Andorra 0.88 75.93 16,16791 13,00239 10,50176 5600,03036 -20 36 49 37 Latvia 0.866 75.29 16,16799 13,0005 3022,39353 8 37	30		Malta	0.895		16.10407	11.32613	39554.52078		
33 31 United Arab Eminates 0.89 77.97 14.3441 12.11122 67462.0953 -24 30 34 32 Greece 0.888 82.24 17.90792 10.55207 301546348 14 33 35 33 Cyprus 0.887 80.08 15.16087 12.17123 38206.64049 2 22 36 34 Uthuania 0.882 75.93 16.63731 13.07721 35796.67041 5 35 37 35 Poland 0.88 78.73 16.63731 13.07721 35796.67041 5 35 37 35 Poland 0.88 78.73 16.30171 12.47232 3162255302 8 34 39 37 Lutvia 0.866 75.29 16.16769 13.03085 30282.39353 8 37 40 38 Portugal 0.864 82.05 16.53799 9.6847 3396.682756 2 38 42	31		Estonia	0.892		15.989	13.143061	36019.26692		
34 32 Greece 0.888 6.224 17.90792 10.55207 30154.63438 14 33 35 33 Gyprus 0.887 80.98 15.16887 12.17123 38206.60409 2 32 36 34 Lithuania 0.882 75.33 166.301171 12.47232 316225.5302 8 3 37 35 Poland 0.88 78.73 16.301171 12.47232 316225.5302 8 3 38 36 Andorra 0.868 78.23 16.16769 13.0005 5000.30336 -20 36 39 37 Latvia 0.866 75.29 16.16769 13.0005 30282.39333 8 2 3 41 39 Slovakia 0.866 75.29 16.16769 13.00055 30282.39333 8 3 3 42 40 Hungary 0.864 75.23 14.48704 12.693157 32113.04569 4 4			Italy	0.892	83.51	16.0889	10.37955	42776.35692		
35 33 Cyprus 0.887 0.908 15.16887 12.17123 38206.84049 2 32 36 34 Lithuania 0.882 75.33 16.67371 13.07721 3579.67041 5 35 38 35 Poland 0.88 78.73 16.0731 12.47232 3162.55302 8 36 38 36 Andorra 0.868 81.91 13.300239 10.50176 56000.03136 -20 36 39 37 Latvia 0.866 75.29 16.16769 13.08085 30282.39353 8 37 41 39 Slovaka 0.864 82.05 15.5759 9.26347 32913.04569 3 39 42 40 Hungary 0.854 76.88 15.20418 11.96329 31328.0599 4 4 43 40 Saud Arabia 0.854 75.13 16.13575 10.22517 47495.42697 -16 40 45 42<			United Arab Emirates	0.89				67462.0953		
36 34 Lithuania 0.882 75.33 16.63731 13.07721 35786.67041 5 35 37 35 Poland 0.08 78.73 16.301711 12.47232 316225.5302 8 34 38 36 Andorna 0.086 6191 13.30029 10.50176 56000.3036 -20 36 39 37 Lutvia 0.066 75.29 16.16769 13.03085 30282.39353 8 37 40 38 Portugal 0.864 82.05 16.53579 9.5647 3396.62756 2 38 41 39 Slovaka 0.66 77.54 14.4874 12.693157 31328.0599 4 4 43 40 Hungary 0.084 75.13 16.33575 10.2517 47495.42697 -16 40 43 40 Sauth Arabia 0.854 75.13 16.33575 10.2517 47495.42697 -16 40 45 -	34		Greece	0.888	82.24	17.90792	10.55207	30154.63438		
37 35 Poland 0.88 78.73 16.301171 12.47232 316225302 8 34 38 36 Andorra 0.868 81.91 13.300239 10.50176 5600.03336 -20 36 39 37 Lutvia 0.866 75.29 16.16769 13.03085 3028.239353 8 37 40 38 Portugal 0.864 82.05 16.167597 9.26347 3396.632756 2 38 41 39 Slovakia 0.86 77.54 14.48704 12.693157 32113.04569 3 39 42 40 Hungany 0.854 76.88 15.20418 11.96329 3132.800599 4 42 43 40 Saudi Arabia 0.854 75.13 16.3575 10.22517 4749.42697 -16 40 44 42 Bahrain 0.854 75.13 16.25462 35192 42521.70451 -12 41 45 <td< td=""><td></td><td></td><td>Cyprus</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>			Cyprus							
38 36 Andorra 0.868 8.191 13.300239 10.50176 56000.3036 -20 36 39 37 Latvia 0.866 75.29 16.16769 13.08085 30202.39333 8 37 41 38 Portugal 0.864 82.05 16.35579 9.26347 3396.82756 2.8 3 41 39 Slovaka 0.66 77.54 14.46704 12.693157 32113.04599 3 3 42 40 Hungary 0.854 76.88 15.20416 11.96329 31328.0059 4 4 4 43 40 Saudi Arabia 0.854 75.13 161575 10.22517 4749.5667 1-16 4 4 44 42 Bahrain 0.854 77.39 162.5862 9.5192 4252.17041 -12 41 45 43 Chile 0.851 0.818 16.2482 9.5192 4252.17041 -12 41										
39 37 Lutvia 0.866 75.29 16.16769 13.03085 30282.39353 8 37 40 38 Portugal 0.864 82.05 16.57579 9.26447 3396.662756 2 38 41 39 Slowaka 0.66 77.54 14.48704 12.693157 3213.04569 3 38 42 40 Hungary 0.854 76.8 15.2018 11.96329 31328.0599 4 4 43 40 Saudi Arabia 0.854 75.13 16.13575 10.22517 47495.42697 1.6 40 44 4.2 Bahrain 0.852 77.3 16.13575 10.22517 47495.42697 1.6 40 45 43 Chile 0.851 80.18 16.14755 10.58295 2.2261.29726 1.6 43 46 43 Croatia 0.851 78.49 15.22672 11.44932 2009.984837 6 44 47 <th< td=""><td></td><td></td><td>Poland</td><td>0.88</td><td></td><td></td><td>12.47232</td><td></td><td></td><td></td></th<>			Poland	0.88			12.47232			
40 38 Portugal 0.864 8.205 16.53579 9.26347 33966.82756 2 38 41 39 Slowakia 0.86 77.54 14.48704 12.693157 3213.04569 3 39 42 40 Hungary 0.854 76.88 15.2018 11.96329 3132.00599 4 42 43 40 Sauch Arabia 0.854 75.13 16.13575 10.22517 47495.42697 -1.6 46 44 42 Bahrain 0.852 77.29 16.25482 9.5192 42521.70451 -12 41 45 43 Chile 0.851 80.18 16.44755 10.5295 23261.39726 16 43 46 43 Croatia 0.851 78.49 15.22672 11.44932 2069.9M837 6 4 47 45 Catar 0.848 80.23 12.0575 9.73144 92418.23056 -43 4 48 46 </td <td></td> <td></td> <td>Andorra</td> <td>0.868</td> <td>81.91</td> <td></td> <td></td> <td></td> <td></td> <td></td>			Andorra	0.868	81.91					
41 39 Slovakia 0.86 77.54 14.48704 12.693157 32113.04569 3 39 42 40 Hungany 0.854 76.88 15.20418 11.96329 3132.880599 4 42 43 40 Saudi Arabia 0.854 75.13 16.15575 10.2517 4795.42667 -16 40 44 42 Bahrain 0.852 77.29 16.25482 85192 4252.17041 -12 41 45 43 Chole 0.851 80.18 16.44755 10.58295 23261.29726 16 43 46 43 Croatia 0.851 78.49 15.22672 11.44932 2009.984837 6 44 47 45 Qata 7.248 80.23 12.0395 9.73144 92418.23086 4.3 4.3 48 46 Argentina 0.845 76.67 17.6542 10.94661 21190.17661 116 46			Latvia	0.866			13.03085	30282.39353		
42 40 Hungary 0.814 76.88 15.20418 11.96329 31328.00599 4 42 43 40 Suduk Arabia 0.854 75.13 16.13575 10.22517 47495.42697 -16 40 44 42 Bahrain 0.852 77.29 16.2842 9.5192 4251.70451 -1.2 41 45 43 Chile 0.851 80.18 16.44755 10.58295 2.2261.29726 16 43 46 43 Croatia 0.851 78.49 15.22672 11.44932 2009.984837 6 44 47 45 Qatric 0.848 80.23 12.03595 9.73144 9248.23056 -43 -43 48 46 Argentria 0.845 76.67 17.6542 10.940501 2119.07661 16 46										
43 40 Saudi Arabia 0.854 75.13 16.13575 10.22517 47495.42697 -16 40 44 42 Bahrain 0.852 77.29 16.25482 95.192 42521.70451 -12 41 45 43 Chile 0.815 80.18 16.4755 10.5295 22261.29726 16 43 46 43 Croatia 0.816 78.49 15.22672 11.44932 2009.94837 6 43 47 45 Qatar 0.848 80.23 12.23595 9.73144 92418.2306 -43 45 48 46 Argentina 0.845 76.67 17.6542 10.940601 21190.77661 16 46			Slovakia							
44 42 Bahrain 0.852 77.29 16.25482 9.5192 42521,70451 -12 41 45 43 Chile 0.851 80.18 16.44755 10.58295 23261,3976 16 43 46 43 Crountia 0.8151 78.49 15.2672 11.44932 20069,94837 6 44 47 45 Catar 0.848 80.23 12.0395 9.73144 92418,2305 -43 -45 48 46 Argentina 0.845 76.67 17.6542 10.940601 21190,17661 16 46										
45 43 Chile 0.851 80.18 16.44755 10.58295 23261,29726 16 43 46 43 Crostia 0.851 78.49 15.22672 11.44932 28069,98437 6 44 47 45 Qatar 0.848 80.23 12.03595 9.73144 9248.23036 -4.3 45 48 46 Argentina 0.845 76.67 17.6542 10.940501 21190.17651 16 46			Saudi Arabia							
46 43 Croatia 0.851 78.49 15.2672 11.44932 28069.84837 6 44 47 45 Qatar 0.848 80.23 12.03595 9.73144 92418.23036 -43 45 48 46 Argentina 0.845 76.67 17.6842 10.940601 21190.17661 16 46										
47 45 Qatar 0.848 80.23 12.03595 9.73144 92418.23036 43 45 48 46 Argentina 0.845 76.67 17.6542 10.940601 21190.17661 16 46	45		Chile		80.18	16.44755	10.58295			
48 46 Argentina 0.845 76.67 17.6542 10.940601 21190.17661 16 46			Croatia							
49 47 Brunei Darussalam 0.838 75.86 14.31391 9.14 63965.09279 -38 47										
	49		Brunei Darussalam	0.838	75.86			63965.09279		

Fig. 5. Captura de la ejecución de los comandos de la variable que contiene los datos. (Fuente: Propia)

De la misma forma se imprime los datos de las ultimas filas para verificar si los datos a utilizar son de utilidad para ello recursos el comando del módulo de panda Tail.

#IMPRIMIMOS LOS ULTIMOS 50 DATOS PARA VERIFICAR SI SON UTILES O NO LAS SECCIONES CON DATOS NULOS

df_Original.tail(50)

#Existe una gran cantidad de datos adicionales provenientes del archivo original que no sirve

238	NaN	I. Estimated using the PPP rate and projected	NaN						
239	NaN	m. Updated by HDRO based on data from United N	NaN						
240	NaN	n. Based on cross-country regression.	NaN						
241	NaN	o. Updated by HDRO using projections from Rarr	NaN						
242	NaN	p. Updated by HDRO based on data from ICF Macr	NaN						
243	NaN	q. Based on cross-country regression and	NaN						
		the p r. Updated by HDRO based on data from							
244	NaN	CEDLAS a	NaN						
245	NaN	s. HDRO estimate based on data from World Bank	NaN						
246	NaN	t. HDRO estimate based on data from World Bank	NaN						
247	NaN	u. HDRO estimate based on data from World Bank	NaN						
248	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
249	NaN	Definitions Human Development Index (HDI): A	NaN						
250	NaN	composite ind	NaN						
251	NaN	Life expectancy at birth: Number of years a ne	NaN						
252	NaN	Expected years of schooling: Number of years o	NaN						
253	NaN	Mean years of schooling: Average number of yea	NaN						
254	NaN	Gross national income (GNI) per capita:	NaN						
255	NaN	Aggreg GNI per capita rank minus HDI rank: Difference	NaN						
		Difference HDI rank for 2018: Ranking by HDI value for							
256	NaN	20	NaN	NaN	NaN	NaN	NaN	NaN	NaN NaN
257 258	NaN NaN	NaN NaN	NaN NaN	NaN NaN	NaN NaN	NaN NaN	NaN NaN	NaN NaN	NaN NaN
259			NaN	NaN		NaN		NaN	NaN
238	NaN	I. Estimated using the PPP rate and projected	NaN						
239	NaN	m. Updated by HDRO based on data from United N	NaN						
240	NaN	n. Based on cross-country regression.	NaN						
241	NaN	o. Updated by HDRO using projections from Barr	NaN						
242	NaN	p. Updated by HDRO based on data from ICF Macr	NaN						
243	NaN	q. Based on cross-country regression and the p	NaN						
244	NaN	r. Updated by HDRO based on data from	NaN						
		CEDLAS a s. HDRO estimate based on data from							
245	NaN	World Bank t. HDRO estimate based on data from World	NaN						
246	NaN	Bank	NaN						
247	NaN	u. HDRO estimate based on data from World Bank	NaN						
248 249	NaN NaN	NaN Definitions	NaN NaN						
250	NaN	Human Development Index (HDI): A	NaN						
251	NaN	composite ind Life expectancy at birth: Number of years a	NaN						
		ne Expected years of schooling: Number of							
252	NaN	years o	NaN						
253	NaN	Mean years of schooling: Average number of yea	NaN						
254	NaN	Gross national income (GNI) per capita: Aggreg	NaN						
255	NaN	GNI per capita rank minus HDI rank: Difference	NaN						
256	NaN	HDI rank for 2018: Ranking by HDI value for 20	NaN						
257									
258 259	NaN NaN	NaN Main data sources	NaN NaN						
238		I. Estimated using the PPP rate and							
	NaN	projected m. Updated by HDRO based on data from	NaN						
239	NaN	United N	NaN						
240	NaN	n. Based on cross-country regression. o. Updated by HDRO using projections from	NaN						
241	14014	Barr	NaN						
242	NaN	p. Updated by HDRO based on data from ICF Macr	NaN						
243	NaN	q. Based on cross-country regression and the p	NaN						
244	NaN	r. Updated by HDRO based on data from CEDLAS a	NaN						
245	NaN	s. HDRO estimate based on data from	NaN						
246		World Bank t. HDRO estimate based on data from World	NaN	NaN	NaN	NaN	NaN	NaN	
	IValv	Bank u. HDRO estimate based on data from							NaN
247	NaN	World Bank	NaN						
248 249	NaN NaN	NaN Definitions	NaN NaN						
250	NaN	Human Development Index (HDI): A	NaN						
251	NaN	composite ind Life expectancy at birth: Number of years a	NaN						
		ne Expected years of schooling: Number of							
252	NaN	years o	NaN						
253	NaN	Mean years of schooling: Average number of yea	NaN						
254	NaN	Gross national income (GNI) per capita: Aggreg	NaN						
255	NaN	GNI per capita rank minus HDI rank: Difference	NaN						
256	NaN	HDI rank for 2018: Ranking by HDI value for	NaN						
256	NaN NaN	20 NaN	NaN NaN	NaN NaN	NaN NaN	NaN NaN	NaN NaN	NaN	NaN
258	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
259	NaN	Main data sources	NaN						

Fig. 6. Captura de la ejecución de los comandos de la variable que contiene los datos. (Fuente: Propia)

3.3.Limpieza

Para este procedimiento se empleara la librería de panda y se invocara la función .dropna el cual eliminara estos espacios , y dicho resultado será enviado una nueva variable pero sin las filas que no contienen datos.

```
#CREACION DE UN DATAFRAME EXTRA PARA NO ALTERAR EL ORIGINAL,
EL CUAL CONTIENE LOS DATOS YA FILTRANDO LOS 'NaN'
df = df_Original.dropna()
df.head(25)
```

	0	Onnamed, 1	(HDI)	birth	schooling	schooling	capita	rank	rank
3		Norway	0.957	82.4	18.06615	12.89775	66494.25217	7	
-		Ireland	0.955	82.31	18.70529	12.666331	68370.58737		
-		Switzerland	0.955	83.78	16.32844	13.380812	69393.52076		
,		Hong Kong, China	0.953	63.76	10.32044	13,300012	69393.32076		
6		Hong Kong, China (SAR)					62984.76553		4
7		Iceland	0.949		19.08309		54682.38057		4
8									4
9		Sweden			19.48234	12.54847			7
10							48084.84207		7
11		Netherlands			18.48513		57707.06867		9
12									10
13									11
14		Singapore			16.44088				12
15		United Kingdom			17.49843		46070.64481		14
16		Belgium							13
17		New Zealand			18.83857		40798.7196		14
18									14
19		United States							17
20							56196.89869		18
21						13.034854	40186.84583		21
22									20
23		Liechtenstein							19
24									24
25		Korea (Republic of)			16.48165		43043.71057		22
26		Luxembourg							23
27		Spain					40974.52408		25

Fig. 7. Captura de la ejecución de los comandos .dropna() de la variable que contiene los datos. (Fuente: Propia)

También se comprará los valores de las filas finales empleando los comandos anteriores

```
#COMPROBAMOS LA INTEGRIDAD DE LA DATA EN SUS ULTIMAS SECCIONES

df = df_Original.dropna()

df.tail(25)
```

Unnamed: 0	Unnamed: 1	Human Development Index (HDI)	Life expectancy at birth	Expected years of schooling	Mean years of schooling	Gross national income (GNI) per capita	GNI per capita rank minus HDI rank	HDI rank
	Lesotho							
	Djibouti					5689.348966		
				12.66498	4.949926	1602.34934		168
	Senegal			8.58486				
	Afghanistan							
						1706.809043		
				7.884416		3828.660219		
	Ethiopia							
	Malawi							
	Congo (Democratic Republic of the)			9.729687	6.76008	1062.543004		
	Liberia							
						1593.704031		
	Mozambique				3.540698	1250.405688		
						1667.844447		
	Central African Republic					993.008842		
	Niger				2.079049	1200.898463		

Fig. 8. Captura de la ejecución de los comandos .dropna() de la variable que contiene los datos. (Fuente: Propia)

Y por último se comprobara de forma general mediante la generación de un informe

```
#COMPROBAMOS QUE YA NO EXISTEN VALORES NAN EN EL DATAFRAME

df.isna().sum().sort_values()
```

```
Unnamed: 0
                                           0
Unnamed: 1
                                          0
Human Development Index (HDI)
                                           0
Life expectancy at birth
                                          0
Expected years of schooling
                                          0
Mean years of schooling
                                          0
Gross national income (GNI) per capita
                                          0
GNI per capita rank minus HDI rank
                                          0
HDI rank
                                           0
dtype: int64
```

Fig. 9. Captura de la ejecución de .sort_values() de la variable que contiene los datos. (Fuente: Propia)

Una vez verificado que toda las filas se realizar el mismo proceso a las columnas pero ante de eso se visualizar los datos dentro de una grafica de mapa de calor

```
#VERIFICAMOS LOS DATOS COMPLETADOS

print()
    sns.heatmap(df.isna())

#Aun podemos ver que existen 2 columnas sin nombre del cual una es de los
nombres de los
#paises y otra que representa al ranking de los países según vemos
inferimos de los datos
```

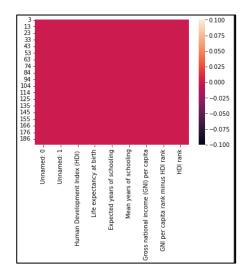


Fig. 10. Grafica de un mapa de calor de los valores de las columnas con datos (Fuente: Propia)

Finalmente se cambiará los valores de las columnas con los valores correspondiente a cada campo aplicando el siguiente comando.

```
#COLOCAREMOS LOS NOMBRES A LAS COLUMNAS QUE NO LO TIENEN Y LE
AGREGAMOS EL AÑO (DEL DATAFRAME ORIGINAL) A LAS COLUMNAS
NECESARIAS

df = df.rename(columns={'Unnamed: 0':'RANK (2019)','Unnamed:
1':'CONTRY','HDI rank':'HDI rank (2018)'})
```

Y finalmente verificamos los resultados

									Python
	RANK (2019)	CONTRY	Human Development Index (HDI)	Life expectancy at birth	Expected years of schooling	Mean years of schooling	Gross national income (GNI) per capita	GNI per capita rank minus HDI rank	HDI rank (2018)
		Norway	0.957	82.4	18.06615	12.89775	66494.25217		
		Ireland			18.70529	12.666331	68370.58737		
		Switzerland	0.955		16.32844	13.380812	69393.52076		
		Hong Kong, China (SAR)	0.949	84.86	16.92947	12.27996	62984.76553		
		Iceland	0.949	82.99	19.08309	12.772787	54682.38057		
184		Burundi	0.433		11.06933	3.287983	753.908748		184
185		South Sudan	0.433		5.296258		2003.318894		
186		Chad	0.398	54.24	7.34935	2.52368	1555.373575		
		Central African Republic			7.56836		993.008842		
188		Niger	0.394	62.42	6.47145	2.079049	1200.898463		189

Fig. 11. Captura de la ejecución del dataframe que contiene los datos. (Fuente: Propia)

Y para finalizar el proceso de limpieza y ajuste de la tabla de datos se actulizara los índices y se comprobara

```
#REAPRACION DEL INDEX YA QUE FALTAN LOS QUE SE ELIMINARON

df.reset_index(drop=True, inplace=True)

#COMPROBAMOS QUE EL CAMBIO EN SU INDEX SE EFECTUO

df
```

#COMI	#COMPROBANOS QUE EL CAMBIO EN SU IMBEX SE EFECTUO df											
✓ 0.1s									Python			
	RANK (2019)	CONTRY	Human Development Index (HDI)	Life expectancy at birth	Expected years of schooling	Mean years of schooling	Gross national income (GNI) per capita	GNI per capita rank minus HDI rank	HDI rank (2018)			
0		Norway	0.957	82.4	18.06615	12.89775	66494.25217					
1		Ireland				12.666331	68370.58737					
2		Switzerland	0.955		16.32844	13.380812	69393.52076					
3		Hong Kong, China (SAR)	0.949	84.86	16.92947	12.27996	62984.76553					
4		Iceland	0.949	82.99	19.08309		54682.38057					
184		Burundi	0.433		11.06933	3.287983	753.908748					
185		South Sudan	0.433		5.296258		2003.318894					
186		Chad	0.398	54.24	7.34935	2.52368						
187		Central African Republic			7.56836	4.282	993.008842					
		Niger	0.394	62.42	6.47145	2.079049	1200.898463					
189 rows	× 9 columns											

Fig. 12. Captura de la ejecución del dataframe que contiene los datos. (Fuente: Propia)

3.4. Visualización de los datos

Para este pardo se empleara los datos de top ten delos países que ms desarrollados entre los anos del 2019 y 2018.

```
#APARTIR DEL DATAFRAME YA COMPLETO CREAMOS OTROS PARA EL ANALISIS DEL IDH
EN LOS 10 RANKIS SUPERIORES E INFERIORES en el 2019 y 2018

df_Superior = df.head(10).drop(['Human Development Index (HDI) ', 'Life
expectancy at birth', 'Expected years of schooling', 'Mean years of
schooling', 'Gross national income (GNI) per capita', 'GNI per capita
rank minus HDI rank'], axis=1)
df_Superior = df_Superior.set_index('CONTRY')
df_Superior

df_Inferior = df.tail(10).drop(['Human Development Index (HDI) ', 'Life
expectancy at birth', 'Expected years of schooling', 'Mean years of
schooling', 'Gross national income (GNI) per capita', 'GNI per capita
rank minus HDI rank'], axis=1)
df_Inferior = df_Inferior.set_index('CONTRY')
df_Inferior
```

```
#USAREMOS UN GRAFICO DE MOSAICO PARA REALIZAR LA COMPARACION DEL RANKIN
DE LOS 10 PRIEMROS PAISES EN 2019 Y 2018
mosaic(df_Superior, ['RANK (2019)', 'HDI rank (2018)'])
```

Graficas en Mosaico

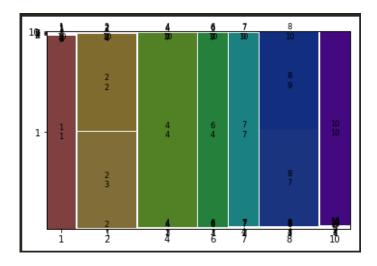


Fig. 13. Grafica tipo mosaico del top 10 del rankig del IDH (Fuente: Propia)

En este gráfico de mosaico permite visualizar los datos dentro de un diagrama de barras apiladas que muestra los porcentajes de datos en distintos grupos en este caso seria los valores de los países que ocupan el top 10 de los datos de año 2018-2019 está grafica nos permite compara entre distintos grupos.

Graficas en Barras

#CON LA AYUDA DE UNA GRAFICO DOBLE BARRA HORIZONTAL DENOTAREMOS LA
VARIACION DEL RANKING DE LOS 10 PRIMEROS PAISES

df_Superior.plot(kind = 'barh')

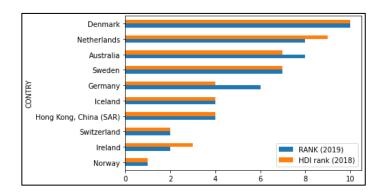


Fig. 14. Grafica tipo barra del top 10 del rankig de países y el IDH (Fuente: Propia)

Gráfico en barras el cual indica el incremento o deceso del índice de desarrollo humano de cada país con respeto a los anos 2018 y 2019 donde la mayoría de países se han mantenido pero los existe casos como el de Alemania ha incrementado pero también existe caso donde el IDH de países como Australia, Noruega e Irlanda han decrecido.

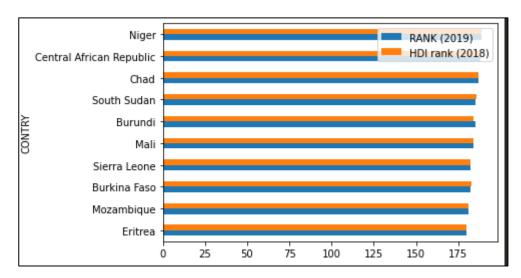


Fig. 15. Grafica tipo barra de los últimos 10 paises del rankig del IDH (Fuente: Propia)

df_Superior.plot.area(alpha=0.4);

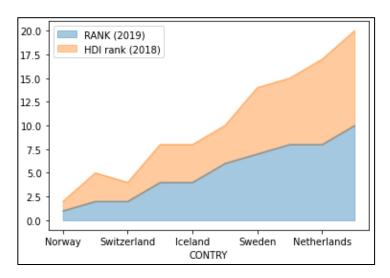


Fig. 16. Grafica de área de los últimos 10 países del rankig del IDH (Fuente: Propia)

Para este ultimo grafico se aprecia la diferencia entre el año 2018 y el 2019. Siendo este ultimo el cual indica que el crecimiento del año 2018 fue mayor que del año 2019.

4. Conclusiones

El uso de herramientas de informática ayuda a la agilización del manejo, análisis y visualización de datos, los cuales por su gran cantidad serían muy extensos de analizarlos de manera manual.

El procesamiento de datos de archivos digitales a través de software permite la recopilación de datos de manera automática pero dicha información puede venir con datos vacíos. De ahí la importancia de realizar los procesos de limpieza de datos y verificación de la integridad de estos.

Las variaciones de IDH en los países seleccionados a través de un año son pocas, muchos países conservan su ranking, las variaciones más marcadas se encuentran en los últimos 10 países del top

Bibliografía

- [1] Z. Yanchun, «Human Development Reports,» Table Human Development Index and its components, 15 07 2022. [En línea]. Available: https://hdr.undp.org/sites/default/files/2021-22_HDR/HDR21-22_Statistical_Annex_HDI_Table.xlsx. [Último acceso: 10 09 2022].
- [2] Human Development Reports, «Human development index (HDI),» Human Development Reports, 10 07 2022. [En línea]. Available: https://hdr.undp.org/. [Último acceso: 11 09 2022].
- [3] Python.org, «What is python?,» Python.org, [En línea]. Available: https://www.python.org/doc/essays/blurb/. [Último acceso: 11 09 2022].
- [4] IBM, «Creación de scripts en lenguaje de programación Python,» IBM Deutschland | IBM, [En línea]. Available: https://www.ibm.com/docs/es/spss-statistics/25.0.0?topic=SSLVMB_25.0.0/spss/base/scripts_python.html. [Último acceso: 11 9 2022].
- [5] Jupyter Team, «Architecture jupyter documentation,» Jupyter Project Documentation, 30 9 2019. [En línea]. Available: https://docs.jupyter.org/en/latest/projects/architecture/content-architecture.html. [Último acceso: 11 09 2022].
- [6] Pandas, «Pandas python data analysis library,» pandas Python Data Analysis Library, 31 8 2022. [En línea]. Available: https://pandas.pydata.org/. [Último acceso: 11 09 2022].
- [7] Aprende con Alf, «La librería Numpy,» Aprende con Alf, 12 04 2022. [En línea]. Available: https://aprendeconalf.es/docencia/python/manual/numpy/. [Último acceso: 11 9 2022].
- [8] J. Hunter, D. Dale, E. Firing y M. Droettboom, «Matplotlib Visualization with Python,» Matplotlib, 2012. [En línea]. Available: https://matplotlib.org/stable/devel/documenting_mpl.html. [Último acceso: 11 09 2022].
- [9] M. Waskom, «Seaborn: Statistical data visualization,» Seabor, 2012. [En línea]. Available: https://seaborn.pydata.org/. [Último acceso: 09 11 2022].
- [10] J. Perktold, S. Seabold y J. Taylor, «Introduction Statsmodels,» Statsmodels-developers, 27 8 2022. [En línea]. Available: https://www.statsmodels.org/dev/index.html. [Último acceso: 11 09 2022].
- [11] PyPI, «PyMosaic,» PyPI, [En línea]. Available: https://pypi.org/project/pyMosaic/. [Último acceso: 11 09 2022].
- [12] Aprende con Alf, «La librería Pandas,» Aprende con Alf, 12 4 2022. [En línea]. Available: https://aprendeconalf.es/docencia/python/manual/pandas/. [Último acceso: 11 09 202].