



Universidad Técnica Particular de Loja
Departamento de Ciencias de la Computación
Maestría en Inteligencia Artificial Aplicada
Herramientas Para Inteligencia Artificial

Tarea: Trabajo Final Escrito

Maestranes:

- Alvarez Aguilar Antonio Alfredo
- Mogrovejo Loyola Jose Daniel
- Vivanco Gualan Ramiro Israel

Objetivo: Generar una aplicación de inteligencia artificial que use librerías de software libre a través de herramientas colaborativas.

1 Introducción

En los últimos años el crecimiento acelerado de la inteligencia artificial (IA) y ha transformado diversos sectores de la sociedad. El presente proyecto busca explorar la aplicación de técnicas de aprendizaje automático en el ámbito educativo, específicamente en el análisis a nivel mundial del acceso de las personas a internet y a nuevas tecnologías utilizando herramientas y librerías de software libre.

2 Marco Teórico de las Tecnologías/Librerías Usadas

En nuestro proyecto, hemos recurrido a una serie de tecnologías y librerías que facilitan la manipulación de grandes volúmenes de datos y la representación gráfica. A continuación, se detallan las herramientas utilizadas

2.1 Pandas:

Pandas es una librería de código abierto que proporciona estructuras y herramientas de análisis de datos para Python. Pandas facilita tareas como la manipulación de datos, la limpieza, la exploración y el análisis de grandes volúmenes de información. (McKinney, 2010).

2.2 NumPy:

NumPy es una librería de Python que soporta la creación y manipulación de vectores y matrices ofreciendo una colección de funciones matemáticas para operar con estas estructuras de datos (Harris et al., 2020).



2.3 Matplotlib:

Matplotlib es una librería que permite crear una gran variedad de gráficos y figuras estadísticas en Python. La librería puede generar diagramas de líneas, barras, errores, dispersión, entre otros (Hunter, J. D., 2007).

2.4 Seaborn

Seaborn es otra librería de visualización de datos en Python basada en Matplotlib para la creación de gráficos estadísticos. Además, simplifica la generación de ciertos tipos de gráficos que manejan directamente al ejecutar DataFrames de pandas (Waskom, 2021).

2.5 SQLite3

SQLite3 es un módulo que proporciona una base de datos SQL ligera que no requiere un proceso de servidor y permite el acceso a la base de datos utilizando una variante del lenguaje de consulta SQL. (SQLite, 2022).

2.6 Skimpy

Skimpy es una herramienta de Python para imprimir resúmenes de DataFrames, está diseñada para proporcionar un resumen rápido de los datos en un formato que sea fácil de leer (Skimpy, 2022).

2.7 Google Colab

Google Colab es un entorno de notebook que permite escribir y ejecutar Python en el navegador con configuraciones que facilitan el uso compartido de proyectos y el acceso a recursos de computación sin necesidad de configuración (Bisong, 2019).

2.8 Google Colab

Bokeh es una librería para realizar visualizaciones interactivas de alta calidad. Los gráficos se generan utilizando HTML, CSS y JavaScript lo que permite que sean interactivos, personalizables y puedan adaptarse a diferentes dispositivos (DataScienTest, 2023).

2.9 Google Colab

Pygwalger es una librería que permite realizar gráficos por medio de una interfaz de exploración interactiva. Se caracteriza por su agil integración a los notebooks.

3 Descripción del Dataset Usado

Los conjuntos de datos descritos a continuación proporcionan información sobre aspectos sociodemográficos, económicos, tecnológicos y ambientales de todas las naciones.



3.1 Global Country Information Dataset 2023

Enciclopedia estadística de todos los países, Cubre una variedad de indicadores que van desde la demografía y la economía hasta la salud y la educación, lo que permite una evaluar el desarrollo de cada nación. Global Country Information Dataset 2023. (2023).

- Facilita estudios comparativos entre naciones.
- Proporciona datos para análisis del acceso a Internet de las personas en cada país.

3.2 International Telecommunication Union (ITU) Statistics

Este conjunto de datos provee métricas sobre la expansión de la tecnología digital y su accesibilidad en diferentes regiones del mundo. Presenta cómo las diferencias en el acceso a la tecnología afectan el desarrollo económico y social y para identificar las brechas digitales que persisten globalmente. International Telecommunication Union. (2023).

- Ofrece estadísticas para evaluar la penetración de las TIC en los mercados emergentes y desarrollados.
- Presenta información sobre el nivel de inclusión digital global.
- Analiza información sobre la relación entre tecnología y varios factores de desarrollo, incluyendo educación, economía y acceso a servicios.

4 Descripción de los pasos realizados en el proyecto.

Inicialmente el proyecto fue realizado en la plataforma de Google Colab, plataforma que permite realizar notebooks de trabajo de manera directa y que además permite colaboración directa entre los usuarios que tengan acceso.

La primera parte se realizan las instalaciones e importaciones de librerías importantes como Pandas, NumPy, Matplotlib, SQLite, entre otras; a continuación, se realiza el consumo de la primera base de datos a través de la librería Pandas; posteriormente se realiza la lectura de un segundo Dataset que se realizó a través de un SQL e importado al cuaderno de trabajo por la misma librería Pandas.

La siguiente parte realizamos un acondicionamiento de los Datasets; en el primero se realiza una limpieza de columnas innecesarias, se pasan indicadores a columnas, eliminamos filas con valores nulos y con el 50% de valores faltantes, se renombran columnas a nombres en español y se realiza una muestra con la librería Skimpy como se observa en la **Figura 1**.

Data Summary		Data Types	
dataframe	Values	Column Type	Count
Number of rows	35	float64	18
Number of columns	20	string	2




number										
column_name	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
Active mobile-broadband and subscriptions per 100 inhabitants	11	31.43	82.25	29.14	42.2	62.85	75.55	96.58	173.5	
Fixed broadband basket as a % of GNI p.c.	0	0	5.454	7.072	0	2.5	3.7	6	42.3	
Fixed broadband subscriptions per 100 inhabitants	10	28.57	19.27	10.71	2.9	10.7	20.2	24.6	42.6	

Figura 1. Visualización de primer Dataset.

En el segundo Dataset mostrado en la **Figura 2** se realizan procesos similares como eliminar caracteres específicos, espacios en blancos, asignamos “NaN” en las filas con denominación “None”, y realizamos una muestra del Dataset final.

Data Summary		Data Types	
dataframe	Values	Column Type	Count
Number of rows	35	float64	46
Number of columns	48	string	2



number										
column_name	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
Active mobile-broadband subscriptions per 100 inhabitants	11	31.43	82.25	29.14	42.2	62.85	75.55	96.58	173.5	
Fixed broadband basket as a % of GNI p.c.	0	0	5.454	7.072	0	2.5	3.7	6	42.3	

Figura 2. Presentación de segundo Dataset.

En el siguiente paso se realizó la unión de los dos Datasets esto se realizó por las columnas “País”, se realizan correcciones en nombre de ciertos países y se procede a eliminar columnas que se refieren como innecesarias y se presenta una muestra del Dataset final.

Adicionalmente se crearon 5 columnas para el Dataset final en donde se consideró los datos ya establecidos y tras una investigación se correlacionaron; primero se realizó una adición entre las suscripciones activas de banda ancha fija y móvil, obteniendo una Tasa de penetración de banda ancha y así con las otras asignaciones obtenemos proporción de banda ancha de alta velocidad, índices de costos de suscripción móvil, relación de banda ancha con respecto a la población y porcentajes de población urbana.

4.1 Breve descripción de las visualizaciones generadas

Una vez determinada la base de datos final procedemos a realizar las siguientes visualizaciones:

La primera visualización mostrada en la **Figura 3** fue realizada con la librería Matplotlib; este cuadro de dispersión nos permite tener una relación entre el Índice de costos de suscripción móvil y una proyección logarítmica de la población de diferentes países de la región de América.

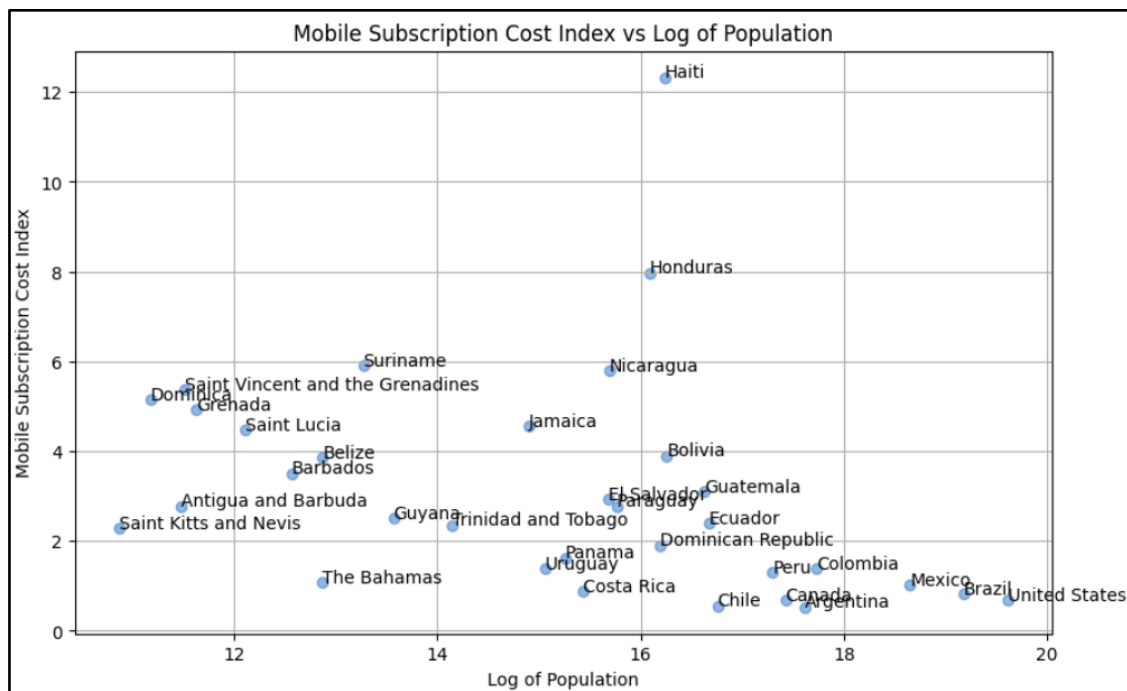


Figura 3. Primera Visualización con Librería Matplotlib.

La **Figura 4** fue realizada con la librería Matplotlib; en este caso en específico es un cuadro estadístico de barras; el cual refleja una relación entre banda ancha y la población de diferentes países de la región de América.

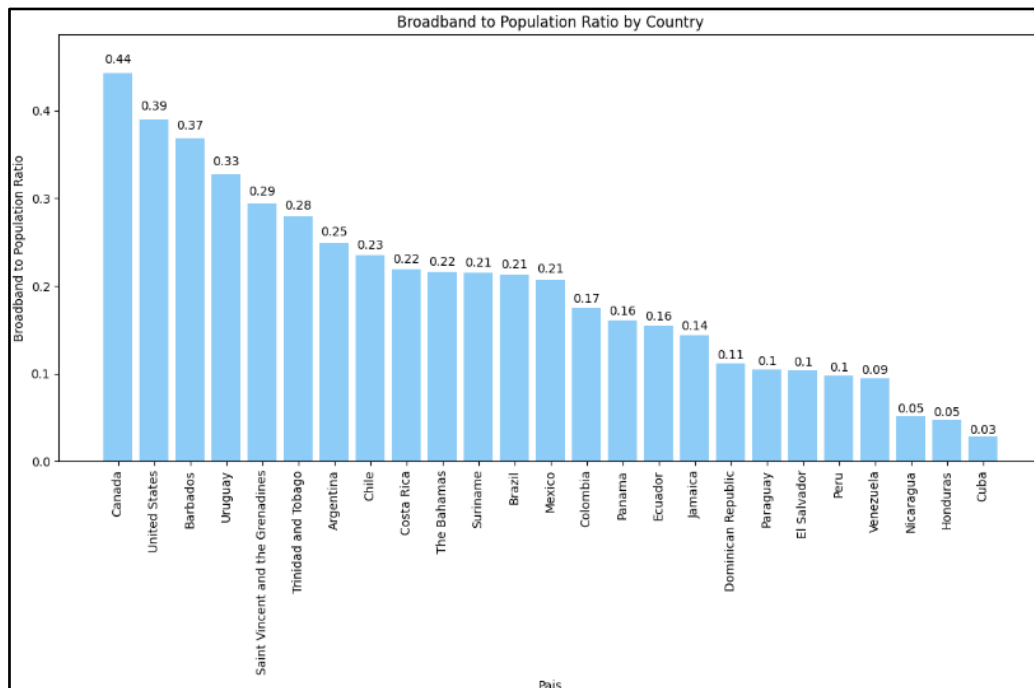


Figura 4. Segunda Visualización con Librería Matplotlib.

La **Figura 5** se realizó con la Librería Bokeh, este grafico más interactivo que los anteriores, consta de funciones para descargar directamente, ampliar la imagen y seleccionar zonas más determinadas por el usuario; en el caso específico se tiene una relación entre la cantidad de individuos que usan internet en una población urbana, todo esto visto desde contextos de diferentes países de América.

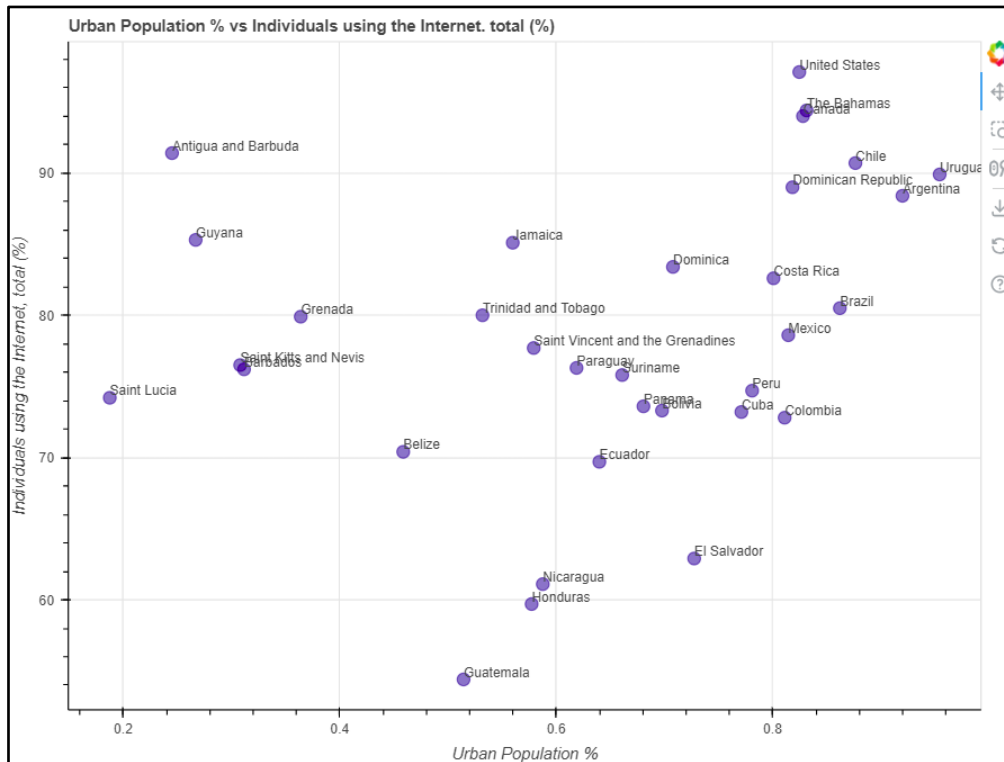


Figura 5. Primera visualización con Librería Bokeh.

La **Figura 6** se realizó de igual manera con la Librería Bokeh, en este gráfico se tiene una relación entre el porcentaje de matriculación bruta en educación terciaria y la tasa de penetración de banda ancha y de igual manera todo visto desde de diferentes países de América.

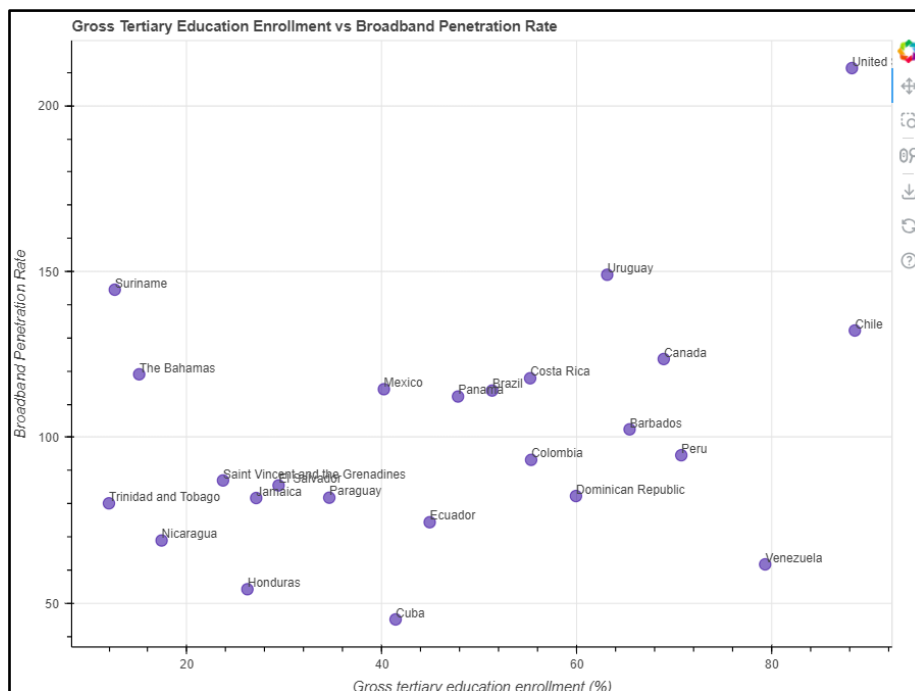


Figura 6. Segunda visualización con Librería Bokeh.

Las **Figuras 7 y 8** se realizaron con la Librería Pygwalker, nos permiten agregar en un simple paso toda la base de datos y representa un gráfico completo, donde nos indican a través de la longitud y latitud las distintas características que se incluyan se pueden representar al igual que las gráficas anteriores distintas métricas y verlas reflejas con un punto específico dentro del mapa mundial.

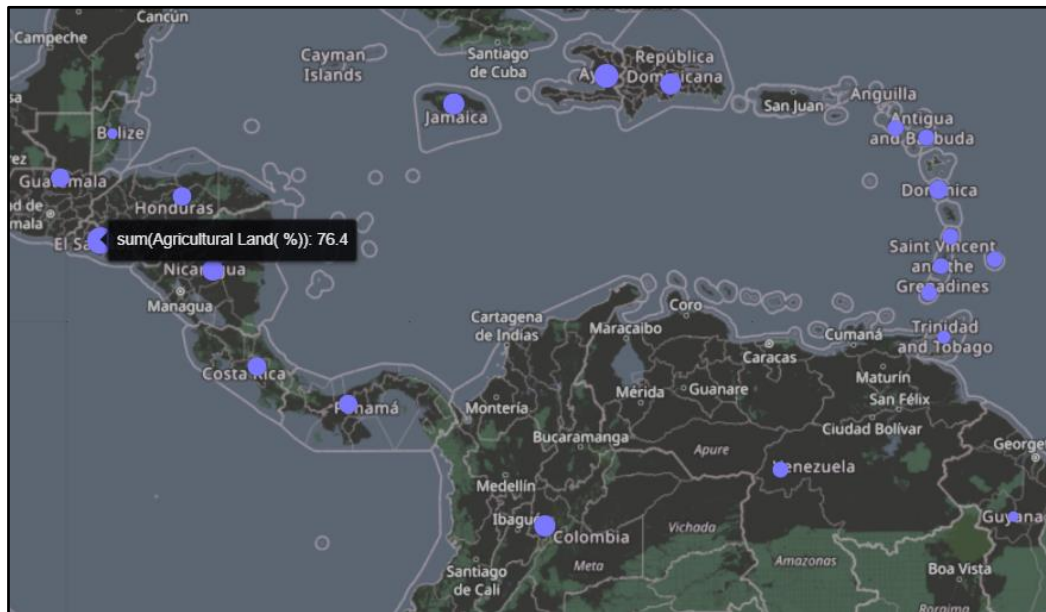


Figura 7. Primera visualización con Librería Pygwalker.

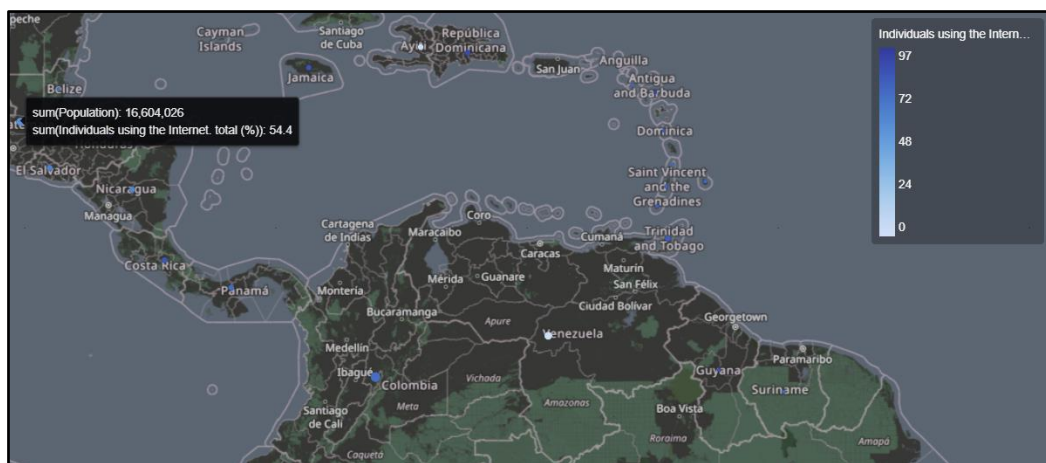


Figura 8. Segunda visualización con Librería Pygwalker.

5 Conclusiones.

Las conclusiones sobre este proyecto final nos describen inicialmente el uso de metodologías de investigación referidas al uso de bases de datos de todo tipo; en el caso puntual se ocuparon bases datos con situaciones demográficas



implicadas en el uso de tecnologías actuales en aspectos de Telecomunicaciones.

Estas bases de datos pueden referir no solo para uso netamente estudiantil, si no en aspectos profesionales para determinar estadísticas y predecir comportamientos futuros, como se logra esto, con las distintas herramientas y librerías implementadas dentro del proyecto, estas han sido de gran utilidad y no solo representan cuadros estadísticos varios si no implementaciones en tiempo real que nos permiten interactuar con cada aspecto de la base de datos inicial.

Se puede destacar que cada base de datos siempre implica grandes cantidades de información que son ocupadas dependiendo los diferentes casos de estudio, en muchos de estos la información debe ser depurada, esto con el fin de obtener solo la necesaria y que permita a las herramientas puntuales definirse de la mejor manera para llegar a mostrar las gráficas estadísticas y ser el punto de partida para futuros implementaciones de Inteligencia Artificial (IA).

Por otra parte, todo el proyecto ha permitido revisar y estudiar adecuadamente las librerías y herramientas de IA, apreciando la facilidad de uso de cada una de ellas y como simplifica la codificación en pro mostrar los resultados adecuadamente; en este proyecto se ocuparon 3 específicamente para gráficos, pero existen distintas herramientas por explorar; sin embargo al entender la base general, permite seguir adentrando en este vasto mundo de las tecnologías informáticas y darles un uso justificado.

Como conclusión final, el proyecto permite interacción entre varias tecnologías, librerías, códigos y plataformas que sirven para el uso y desarrollo de IA adecuado, destacando esencialmente que el uso es netamente práctico y que va definiendo el modelo de estudio propuesto siendo esencialmente aplicado a casos de estudio y posteriormente a casos profesionales.

6 Bibliografía

1. Reback, J., McKinney, W., et al., (2020). pandas-dev/pandas: Pandas. Zenodo. <https://doi.org/10.5281/zenodo.3509134>
2. Harris, C. R., Millman, K. J., van der Walt, S. J., et al., (2020). Array programming with NumPy. Nature, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
3. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95. <https://doi.org/10.1109/MCSE.2007.55>
4. Waskom, M. (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021. <https://doi.org/10.21105/joss.03021>
5. SQLite. (2022). SQLite Home Page. <https://www.sqlite.org/index.html>
6. Skimpy. (2022). PyPI. <https://pypi.org/project/skimpy/>
7. Bisong, E. (2019). Building machine learning and deep learning models on Google Cloud Platform: A comprehensive guide for beginners. Apress. <https://doi.org/10.1007/978-1-4842-4470-8>



8. Global Country Information Dataset 2023: Datahub. (2023). Global Country Information Dataset 2023. Recuperado de <https://datahub.io/global-country-information-2023>
9. International Telecommunication Union (ITU) Statistics: International Telecommunication Union. (2023). ITU Statistics. Recuperado de <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>
10. DataScienTest (2023). Bokeh: la biblioteca de visualización Python de nueva generación. <https://datascientest.com/es/bokeh-la-biblioteca-de-visualizacion-python>
11. Nicolás Urrego (2023). Visualización avanzada de datos en Python: Guía para configurar y ejecutar PyGWalker. Medium. <https://nicolasurrego.medium.com/visualización-avanzada-de-datos-descubre-el-poder-de-pygwalker-en-python-9d708724ddfb>