

UNIVERSIDAD DE PIURA



CURSO: PYTHON PARA EL ANÁLIS DE DATOS

PROFESOR: ING. PEDRO ROTTA

GRUPO: 7

INTEGRANTES:

JIMÉNEZ FARFÁN ALEJANDRA

MONDRAGON PARDO JACKSON

PRADO CHAFLOQUE SARAI ELIZABETH

SIANCAS HUERTA MANUEL REYNALDO

ZATTA RUIZ GERSÓN ANDRÉ

PIURA, 22 DE ENERO DE 2022

Implementación de una herramienta de sistematización en Python para el análisis del gasto público en educación en el Perú

Mondragon, J., Jiménez, A., Siancas, M., Prado, S., Zatta, G.

*Universidad de Piura
Piura, Peru*

Abstract - In this article, a program developed with the Python programming language is presented. It allows to systematise the analysis process of public spending in education in Perú. The aim is for it to become a useful tool in the execution of future economic or social studies, as well as have an impact in the decision-making in the education field in the country since in the last decades peruvian education shows a decline in its development ("Perú entre los países que menos invierten en educación, por debajo de los US\$ 50,000", 2018) when it comes down to education compared to other countries in Latin America ("Perú es el país que menos invierte en Educación en Sudamérica, según su PBI", 2017). In regards to the program, its inputs are the year, public education spending (in millions of euros), public education spending percentage, GDB percentage assigned to education spending, education spending per capita as well as new variables that are created by the user. The outputs are the dataframe containing the inputs, plots and statistical data. On the subject of the program code, it has been written in the Python programming language (programming language that has become very well-known during the last decade and that is currently being applied in the data science world) due to the easy syntax as well as the possibility of running the program either in the command prompt of a computer or in Google Colaboratory.

Key words - Python, Public spending, Education, Systematization, Numpy, Matplotlib, Pandas.

Resumen - En este trabajo se ha desarrollado un programa elaborado en el lenguaje de programación Python que permite sistematizar el proceso de análisis del gasto público en educación en el Perú, con el objetivo de que dicha herramienta favorezca la ejecución de futuros estudios económicos o sociales en el tema, así como también el constituir una herramienta para la toma de decisiones en el ámbito de la educación en nuestro país, que durante las últimas décadas muestra un menor desarrollo ("Perú entre los países que menos invierten en educación, por debajo de los US\$ 50,000", 2018) en este sector con respecto a otros países de Latinoamérica ("Perú es el país que menos invierte en Educación en Sudamérica, según su PBI", 2017). En cuanto refiere al análisis del sistema se ha tomado en cuenta como variables de entrada: la fecha, el gasto en educación (en millones de euros), el porcentaje de gasto público en educación(en %), el porcentaje de PIB destinado al gasto en educación(en %), el gasto de educación per cápita (en euros/persona) así como nuevas variables que se puedan considerar y ser ingresadas; por otro lado las salidas del sistema

son además de las variables de entrada: los gráficos de análisis entre variables y datos estadísticos. Finalmente, respecto a la implementación del código del programa, se ha elaborado en Python (lenguaje de programación que se ha popularizado durante la última década y que en la actualidad viene siendo utilizado tanto en el mundo de la ciencia de datos como en aplicaciones laborales) debido a la facilidad de la sintaxis, así como la posibilidad de ejecutar dicho programa desde la terminal o desde Google Colaboratory.

Palabras clave - Python, Gasto público, Educación, Sistematización, Numpy, Matplotlib, Pandas.

I. INTRODUCTION

La educación es uno de los componentes más influyentes para el avance de los seres humanos, las sociedades, es por ello que es de importancia para la economía. Las inversiones realizadas en este ámbito han tenido un impacto positivo en cierta medida, evidenciándose en el comportamiento de variables económicas como el PIB. Desafortunadamente, en 2020, plena pandemia, se dedicó tan solo el 15,96% del gasto público, de ahí sus deficientes resultados. Ello puso en evidencia la importancia de una buena inversión en la educación de nuestro país. A través de una herramienta de sistematización en Python se analizará el gasto público en educación del Perú desde 1999 hasta el 2020.

II. ANÁLISIS DEL SISTEMA

Se ha desarrollado en la herramienta Google Colaboratory un programa en lenguaje Python que permite, dado un archivo xlsx conteniendo datos y convertido en un dataframe, ingresar nuevos datos, filtrar datos, observar gráficos que relacionan los datos, observar datos estadísticos y agregar columnas de nuevos datos. Esto se logra a través de un bucle while infinito con 6 estados.

El primer estado permite ingresar un nuevo dato, es decir, generar una nueva fila en el dataframe. El segundo estado permite filtrar datos por medio de características: por el año o escoger una columna y obtener los datos menores o iguales a un valor contenidos en esa columna.

El tercer estado permite solicitar gráficos de análisis entre algunas variables: Fecha vs. Gasto en educación (M. €), Fecha vs. Porcentaje de gasto público en educación (%Gto Pub), Fecha vs. Porcentaje de PIB destinado al gasto en educación (%PIB), Fecha vs Gasto de Educación Per Cápita. Este estado también permite realizar una comparación entre más de dos variables: entre Fecha, Gasto en educación (M.€), Porcentaje de gasto público en educación (%Gto Pub) y Porcentaje de PIB destinado al gasto en educación (%PIB). El cuarto estado permite observar datos estadísticos mediante el método describe de la librería Pandas. Entre los datos estadísticos que se muestran están la media, los cuartiles y la desviación estándar. El quinto estado permite generar nuevas columnas en el dataframe. El sexto estado permite cerrar el programa.

Se le presentan al usuario las opciones: Ingresar un dato nuevo, filtrar datos, solicitar gráficos, solicitar datos estadísticos, generar nuevas características o finalizar el programa.

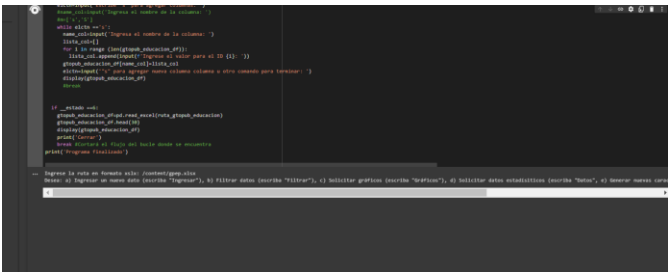


Fig. 2 Ingreso de la ruta del archivo xlsx y selección del estado para el bucle while.

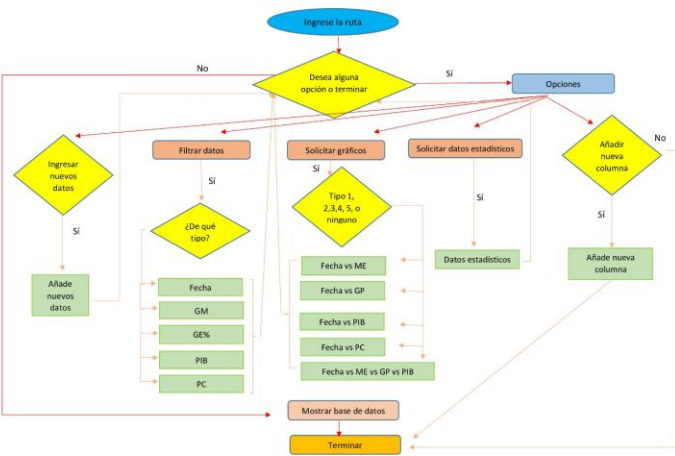


Fig. 1 Diagrama de flujo del programa.

A. Entradas

Son los datos contenidos en el archivo xlsx. Los nombres de las columnas se enuncian a continuación.

- Gasto en educación (M. €): Gasto en educación en millones de euros correspondiente a un año.
- Porcentaje de gasto público en educación (%Gto Pub): Es el porcentaje del total del gasto público en un año que se destinó a la educación.
- Porcentaje de PIB destinado al gasto en educación (%PIB).
- Gasto de educación Per Cápita: Se refiere al gasto promedio de un individuo de la población peruana en educación en un año. Esta en euros.
- Fecha: Se refiere al año.

B. Salidas

Son el dataframe, los datos filtrados en el dataframe, gráficos de análisis y datos estadísticos.

III. PRUEBAS

A. Ruta y selección del estado

B. Estado 1

Para agregar un nuevo dato, se ingresa toda una fila. Es decir, se ingresa un nuevo valor en cada columna. Finalmente, el programa devuelve el dataframe actualizado.

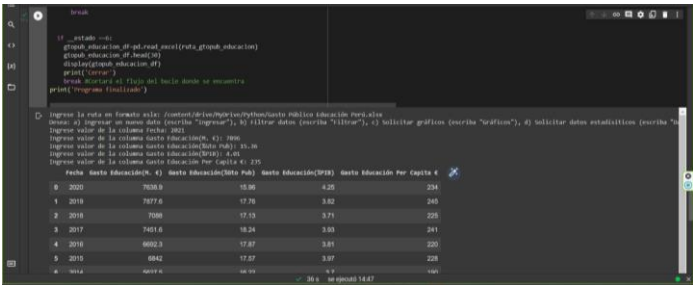


Fig. 3 Ejecución del estado 1.

C. Estado 2

Se ha escogido filtrar datos mayores o iguales a 50 en la columna Gasto Educación Per Cápita.

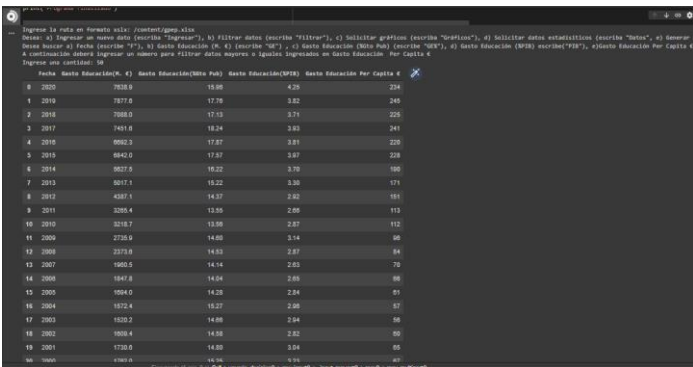


Fig. 4 Ejecución del estado 2.

D. Estado 3

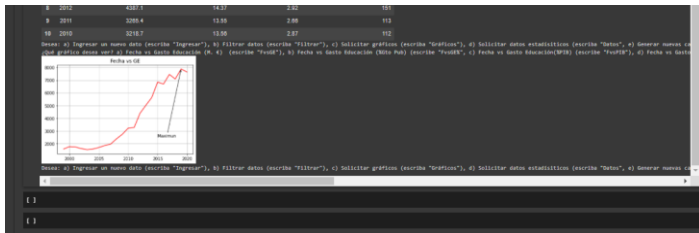


Fig. 5 Ejecución del estado 3. Se escogió graficar Fecha vs. Gasto en educación (M.€).

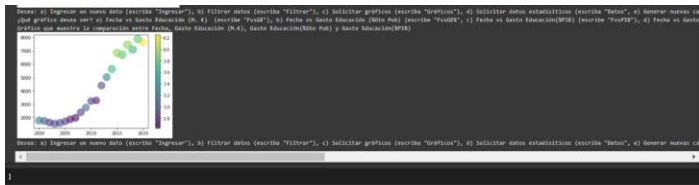


Fig. 6 Gráfico comparativo entre Fecha, Gasto en educación (M.€), Porcentaje de gasto público en educación (%Gto Pub) y Porcentaje de PIB destinado al gasto en educación (%PIB).

E. Estado 4

El programa devuelve los valores estadísticos de media (mean), desviación estándar (std), el valor máximo (max), el valor mínimo (min) y los cuartiles para cada columna de datos.

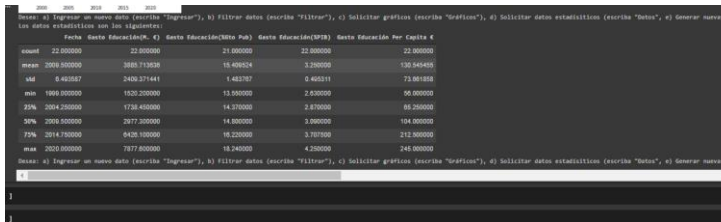


Fig. 7 Ejecución del estado 4.

F. Estado 5

Al seleccionar la opción de ‘generar nuevas características’, el usuario puede añadir una nueva columna. Para ello, primera ingresa el nombre de ella, luego sus elementos y finalmente el programa por defecto devuelve el dataframe actualizado.

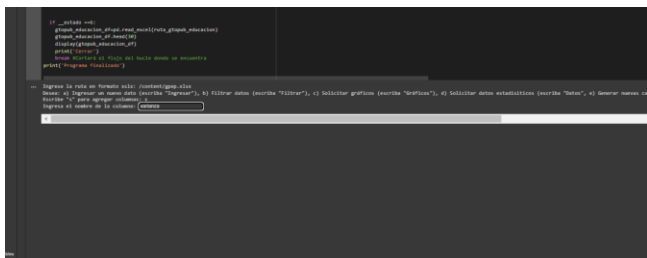


Fig. 7 Ingreso del nombre de la nueva columna ‘varianza’.

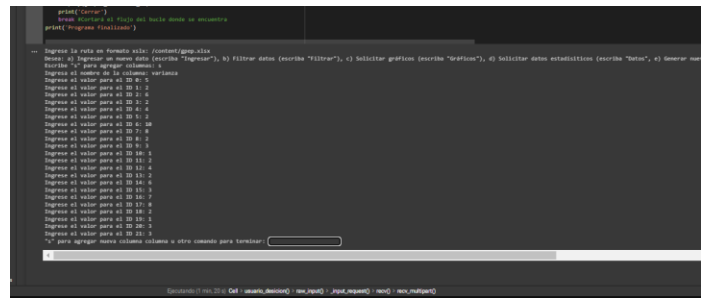


Fig. 8 Ingreso de los datos de la nueva columna ‘varianza’.

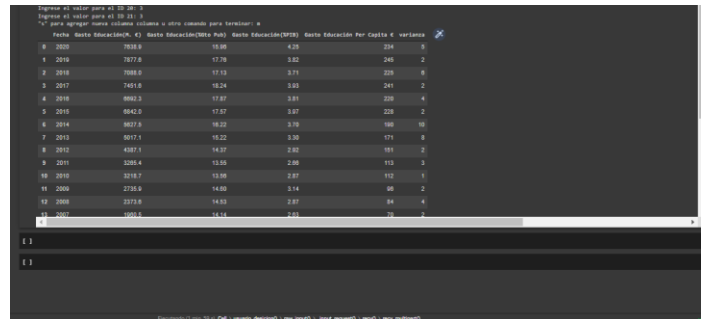


Fig. 9 Dataframe conteniendo la nueva columna.

G. Problemas presentados

En el estado 5, la opción de añadir columnas mostraba una falla debido a que había un error de sintaxis. Al utilizar el comando display se había ingresado el nombre del dataframe como gtopub_educacion, cuando debía ser gtopub_educacion_df.

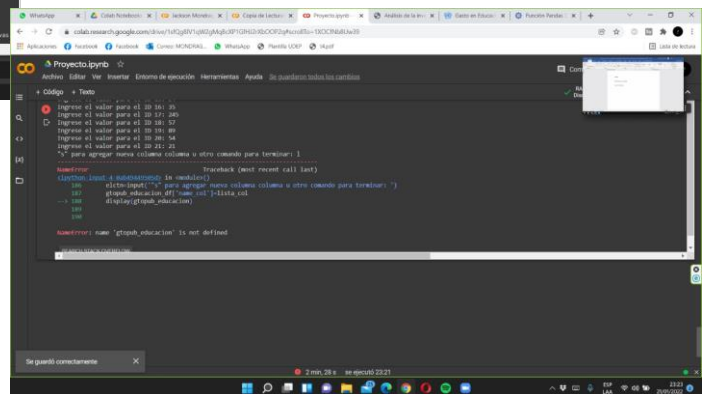


Fig. 10 Primer error en el estado 5.

El gráfico de dispersión no se mostraba después de haber seleccionado el estado 3 sino que se mostraba una vez terminado el programa desde el menú principal, esto fue debido a que se había omitido el comando plt. show().

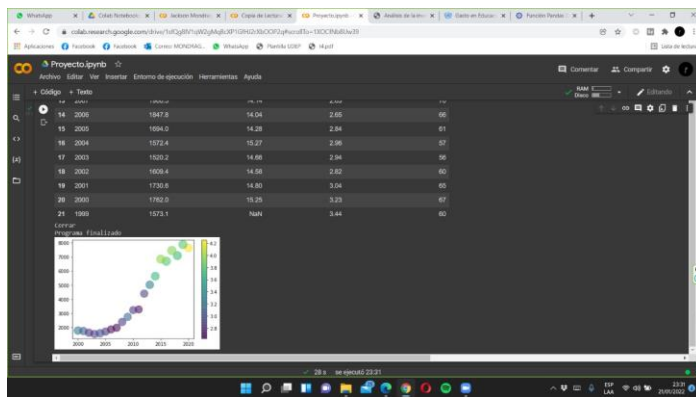


Fig. 11 Error en el estado 3.

Durante la prueba del estado 5 se presentó también un error de sintaxis. Se le había asignado previamente el código el nombre 'name_col' a la columna nueva, por lo que esta no se guardaba con el nombre que ingresaba el usuario.

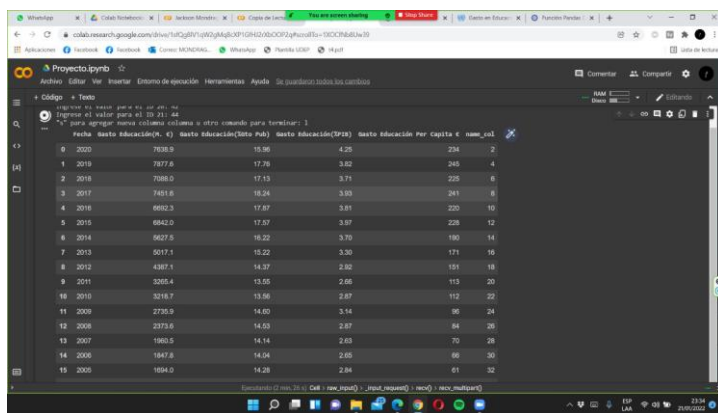


Fig. 12 Segundo error en el estado 5.

IV. CONCLUSIONES

Programas como el desarrollado en el presente trabajo son muy útiles para digitalizar datos que solían guardarse en papel. Un ejemplo de ello son el inventario y datos económicos como gastos, ingresos y egresos de un emprendimiento. Dicho ello, el programa presentado podría modificarse y a partir de él desarrollar otro programa que permita ayudar a emprendedores.

La herramienta de sistematización en Python es de gran utilidad para el análisis de datos del gasto público en educación del Perú, de manera que nos permite la comparativa entre variables, análisis de características, utilizando diferentes librerías.

Python es un lenguaje de programación que se adecua significativamente para el desarrollo de nuevas herramientas que faciliten el estudio de problemas dentro del campo de las ciencias sociales y la economía.

La herramienta desarrollada permite un análisis más ágil y versátil de la problemática actual de nuestro país relacionada a la inversión de gasto público en educación.

Este programa es muy útil porque nos brinda una serie de herramientas con las cuales podemos analizar cada una de las variables que intervienen en el gasto público destinado a la educación en el Perú de tal modo que podamos tomar decisiones para girar el rumbo que lleva la educación peruana hoy en día de manera positiva, por su parte el programa python es un muy buen instrumento que nos permitió de manera fácil y sencilla poder crear este código que realiza además otras funciones, como el filtrado de datos, análisis de variables con gráficos y el ingreso de datos.

La pandemia de COVID 19 ha evidenciado la necesidad de darle importancia al gasto público en educación en el Perú y en muchos países de Latinoamérica.

AGRADECIMIENTOS

Este artículo fue hecho gracias a las clases y la asesoría del Ing. Pedro Rotta. Asimismo, se agradece a Vida Universitaria de la Universidad de Piura por haber hecho posible la realización del curso Python para el análisis de datos.

REFERENCIAS

- [1] *Perú - Gasto público Educación 2020*. datosmacro.com. (2020). Recuperado de <https://datosmacro.expansion.com/estado/gasto/educacion/peru>.
- [2] *Perú entre los países que menos invierten en educación, por debajo de los US\$ 50,000*. Gestión. (2018). Recuperado de <https://gestion.pe/economia/peru-paises-invierten-educacion-debajo-us-50-000-229121-noticia/>.
- [3] *Perú es el país que menos invierte en Educación en Sudamérica, según su PBI*. RPP. (2017). Recuperado de <https://rpp.pe/politica/estado/peru-es-el-pais-que-menos-invierte-en-educacion-en-america-latina-noticia-1071004>.
- [4] Rotta, P. (2022). Lectura 1. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [5] Rotta, P. (2022). Lectura 10. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [6] Rotta, P. (2022). Lectura 2. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [7] Rotta, P. (2022). Lectura 3. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [8] Rotta, P. (2022). Lectura 4. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [9] Rotta, P. (2022). Lectura 5. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [10] Rotta, P. (2022). Lectura 6. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [11] Rotta, P. (2022). Lectura 7. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [12] Rotta, P. (2022). Lectura 8. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.
- [13] Rotta, P. (2022). Lectura 9. En P. Rotta (Comp.), *Python para el análisis de datos*. Universidad de Piura.