

PROYECTO ANÁLISIS DE DATOS

EMISIONES DE CO₂

JULIO ANTHONY ENGELS
RUIZ COTO - 1284719

EDDIE ALEJANDRO GIRÓN
CARRANZA - 1307419

UNIVERSIDAD RAFAEL LANDIVAR GUATEMALA
JULIO 2022

ANÁLISIS PRELIMINAR

Determinar la ruta que toma la generación del dióxido de carbono para su posterior predicción. Tomando las emisiones de dióxido de carbono de empresas y vehículos como un subsistema del sistema de gases de efecto invernadero.

OBJETIVO DEL ESTUDIO DEL SISTEMA

Determinar predicciones de las emisiones de CO₂.

CONCEPTUALIZACIÓN DEL SISTEMA

Medición de las emisiones de CO₂ producidas por vehículos y empresas generadoras de gases. Las empresas y vehículos emiten CO₂, las plantas por medio de la fotosíntesis toman el CO₂ del ambiente y lo convierten en carbón vegetal, este ayuda al crecimiento de la flora que posteriormente sirve de alimento para la fauna, la fauna genera desechos orgánicos, los cuales con el pasar de los años se transforman en petróleo, y es utilizado nuevamente por la industria y la sociedad.

ANÁLISIS FUNCIONAL

Entradas:

- Combustibles fósiles
- Transporte
- Electricidad y calor
- Industria
- Agricultura y ganadería

Procesos:

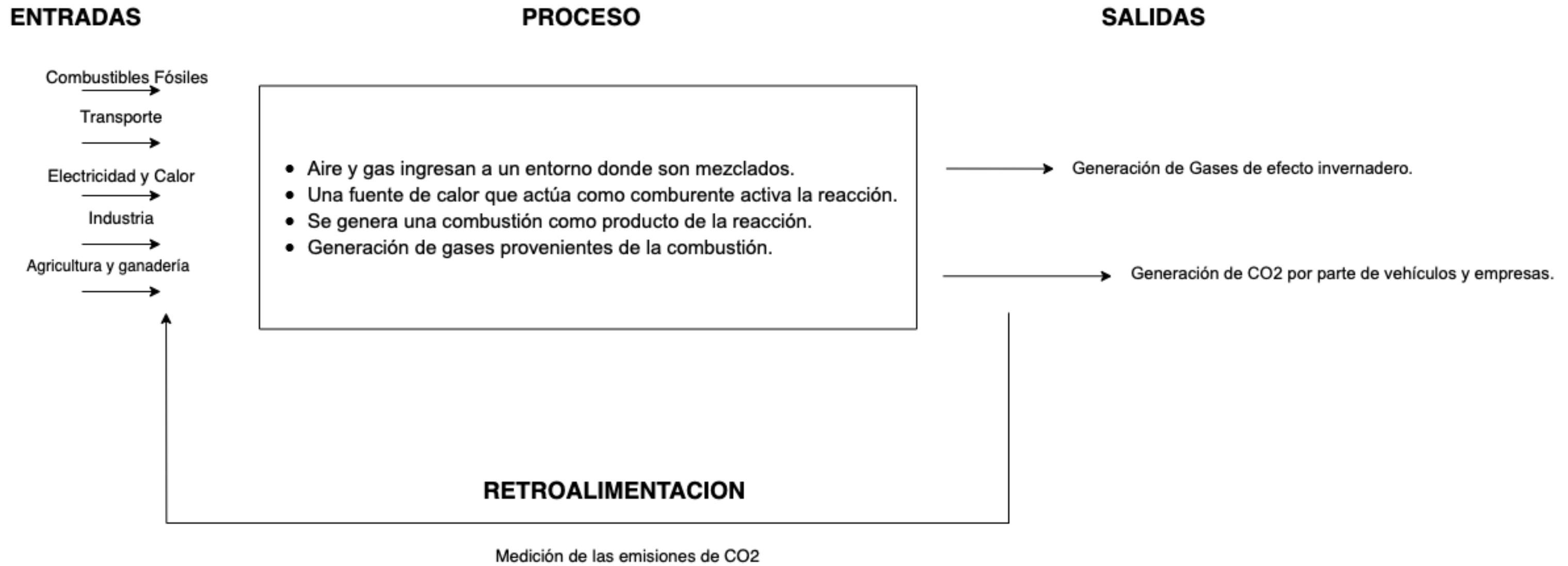
- Aire y gas ingresan a un entorno donde son mezclados
- Una fuente de calor que actúa como comburente activa la reacción
- Se genera una combustión como producto de la reacción
- Generación de gases provenientes de la combustión

Salidas:

- Generación de gases de efecto invernadero
- Generación de CO₂ por parte de empresas y vehículos

Retroalimentación:

- La medición de las emisiones de CO₂



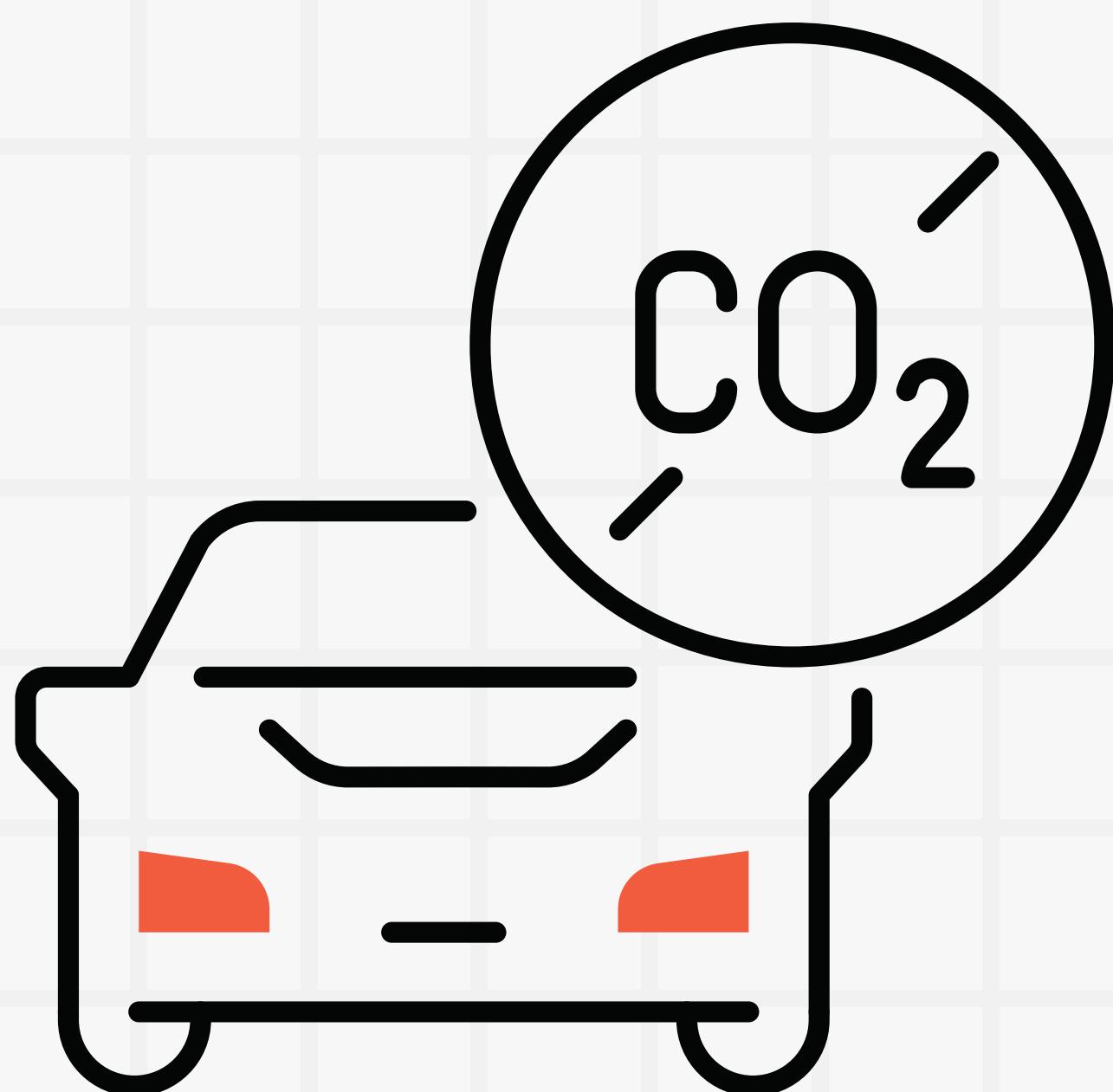
ANÁLISIS DE CONDICIONES

- El tiempo: Ya que el proceso de generación de CO₂ con resultados notables puede tomar bastante tiempo.
- El medio ambiente: Dependiendo de los elementos que se encuentren en este, las emisiones pueden variar en gran o menor medida.
- Las necesidades humanas: Debido a que estas no poseen un estándar fijo o una ruta, según cambian las necesidades humanas las emisiones de CO₂ pueden variar también.

ESTUDIO DEL SISTEMA

Variables:

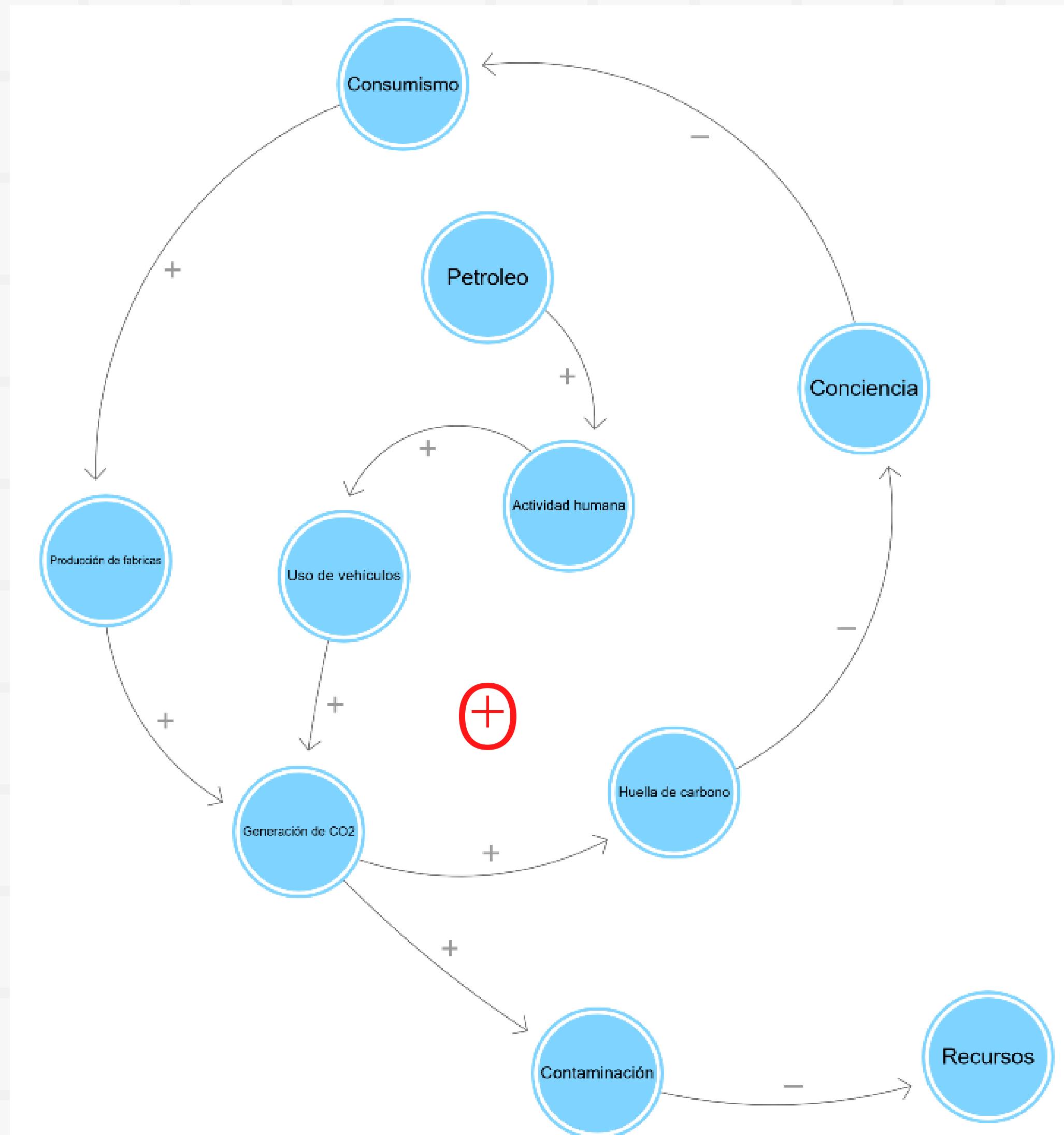
- Producción de fábricas
- Uso de vehículos
- Actividad humana
- Generación de CO₂
- Contaminación
- Recursos
- Uso de carbón
- Petróleo
- Industria



Efectos del sistema:

- Contaminación
- Incendios forestales
- Desertificación
- Contaminación de mares
- Problemas de salud

Diagrama causal:



NOTA: la huella de carbono es una métrica ambiental que calcula la totalidad de las emisiones de gases de efecto invernadero GEI generadas, directa e indirectamente, por una persona, un grupo, una organización, empresa o incluso un producto o servicio.

DETALLE DE HERRAMIENTA ESCOGIDA

Pandas es una muy popular librería de código abierto dentro de los desarrolladores de Python, y sobre todo dentro del ámbito de Data Science y Machine Learning, ya que ofrece unas estructuras muy poderosas y flexibles que facilitan la manipulación y tratamiento de datos. Las dos estructuras de datos principales dentro del paquete Pandas son: Series: array unidimensional etiquetado capaz de almacenar cualquier tipo de dato. DataFrame: estructura bidimensional con columnas que pueden ser también de cualquier tipo. Estas columnas son a su vez Series. Pandas surgió como necesidad de aunar en una única librería todo lo necesario para que un analista de datos pudiese tener en una misma herramienta todas las funcionalidades que necesitaba en su día a día, como son: cargar datos, modelar, analizar, manipular y prepararlos. Todo esto trabaja sobre una plataforma abierta “Colaboratory”, este se encuentra disponible en internet, es una herramienta en la cual se puede desarrollar machine learning, data Mining, Deep learning todo bajo el entorno de Python. Para este proyecto el tipo de modelo que utilizamos fue el modelo de regresión lineal múltiple multivariable usando la poderosa librería sklearn para obtener el algoritmo de aprendizaje automático, con este algoritmo se entrenó con el DataFrame para crear el modelo de pronostico para un conjunto de datos específico.

JUSTIFICACIÓN DE LA HERRAMIENTA

Como pareja de trabajo escogimos la herramienta Pandas para Python porque permite, de forma fácil e intuitiva realizar operaciones capaces de gestionar, manipular y predecir una variable objetivo que este caso es la carga de CO₂ en Guatemala (MTm CO₂), cualquier tipo de información (análisis estadístico, visualización de datos) sin importar el formato, y sobre todo de una forma rápida y eficaz.

ALIMENTACIÓN DE LA HERRAMIENTA

Para la alimentación de la herramienta se descargo el Data Frame de la pagina Global Carbono Atlas en formato .CSV, este Data Frame contiene la información de las emisiones de CO₂ de 221 países de todo el mundo este cuenta con 61 filas y 222 columnas (contiene una de más en filas y columnas por el encabezado de los datos), esta información fue de mucha utilidad ya que con ella realizamos el modelo de predicción para Guatemala.

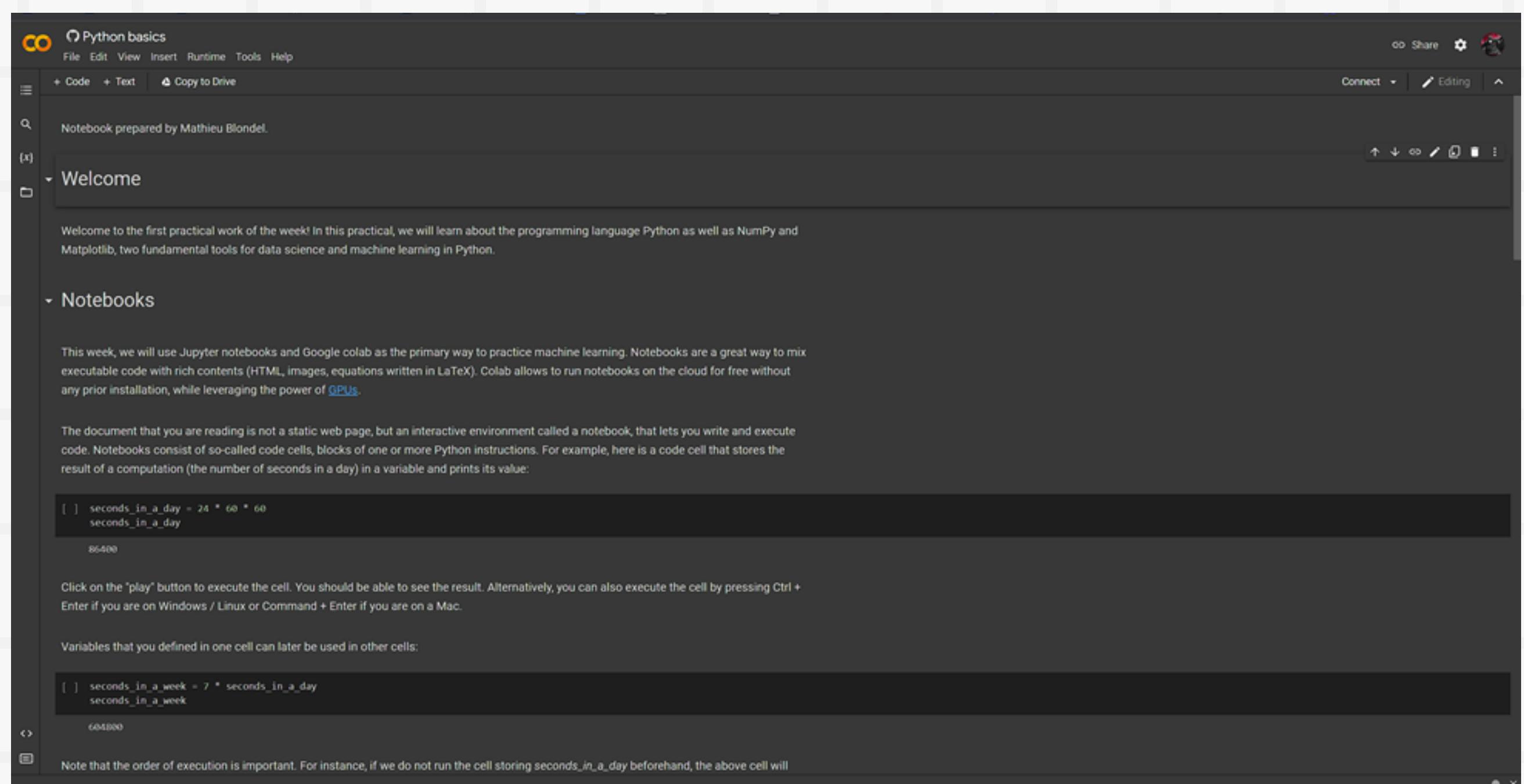
DETALLE DE INSTALACIÓN DE LA HERRAMIENTA

1. Como primer paso buscamos en nuestro motor de búsqueda favorito el siguiente enlace:

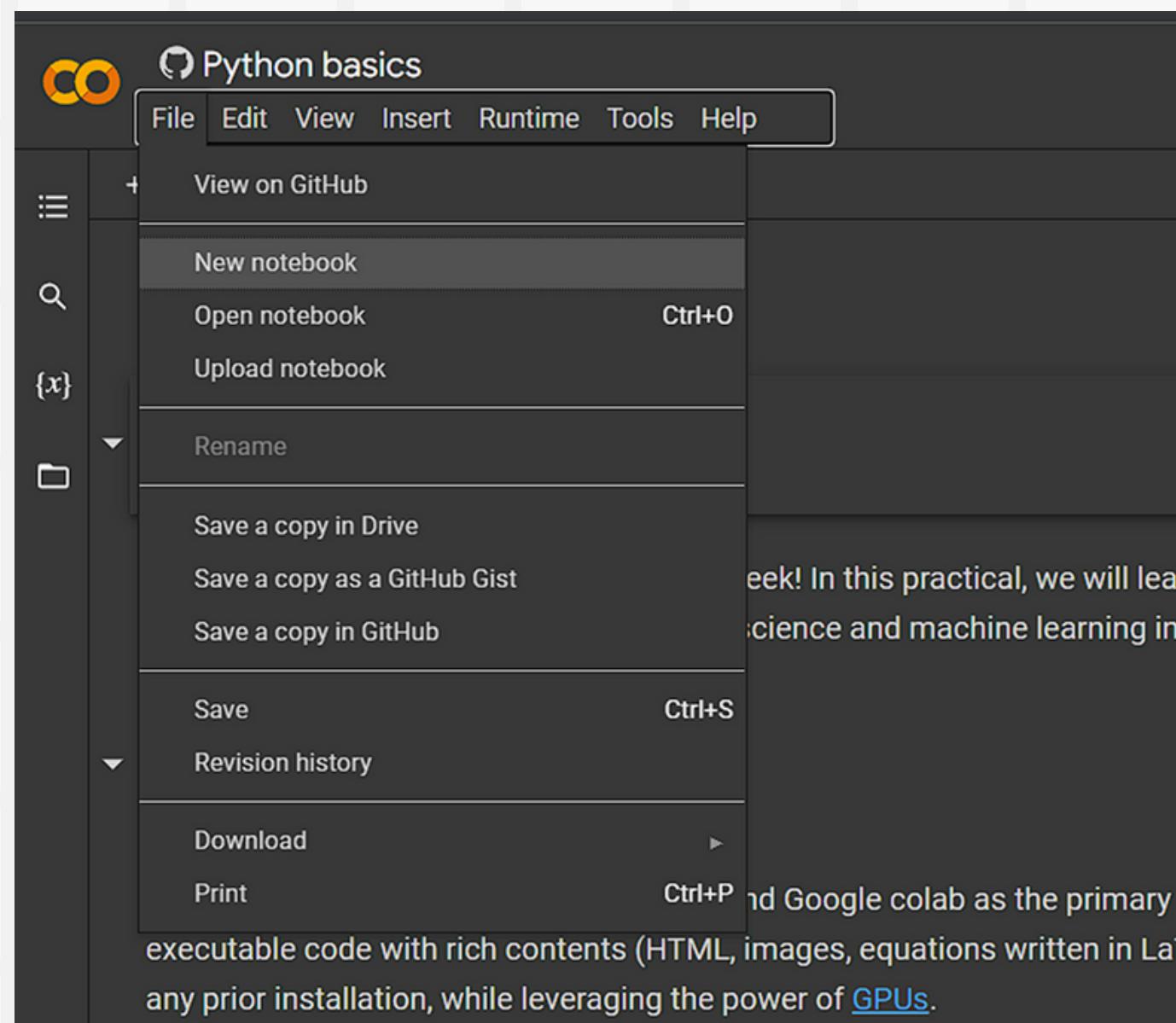
https://colab.research.google.com/github/data-psl/lectures2020/blob/master/notebooks/01_python_basics.ipynb#scrollTo=2HuuwOcQEYQI

https://colab.research.google.com/github/data-psl/lectures2020/blob/master/notebooks/01_python_basics.ipynb#scrollTo=JbVT11Y8CbAu

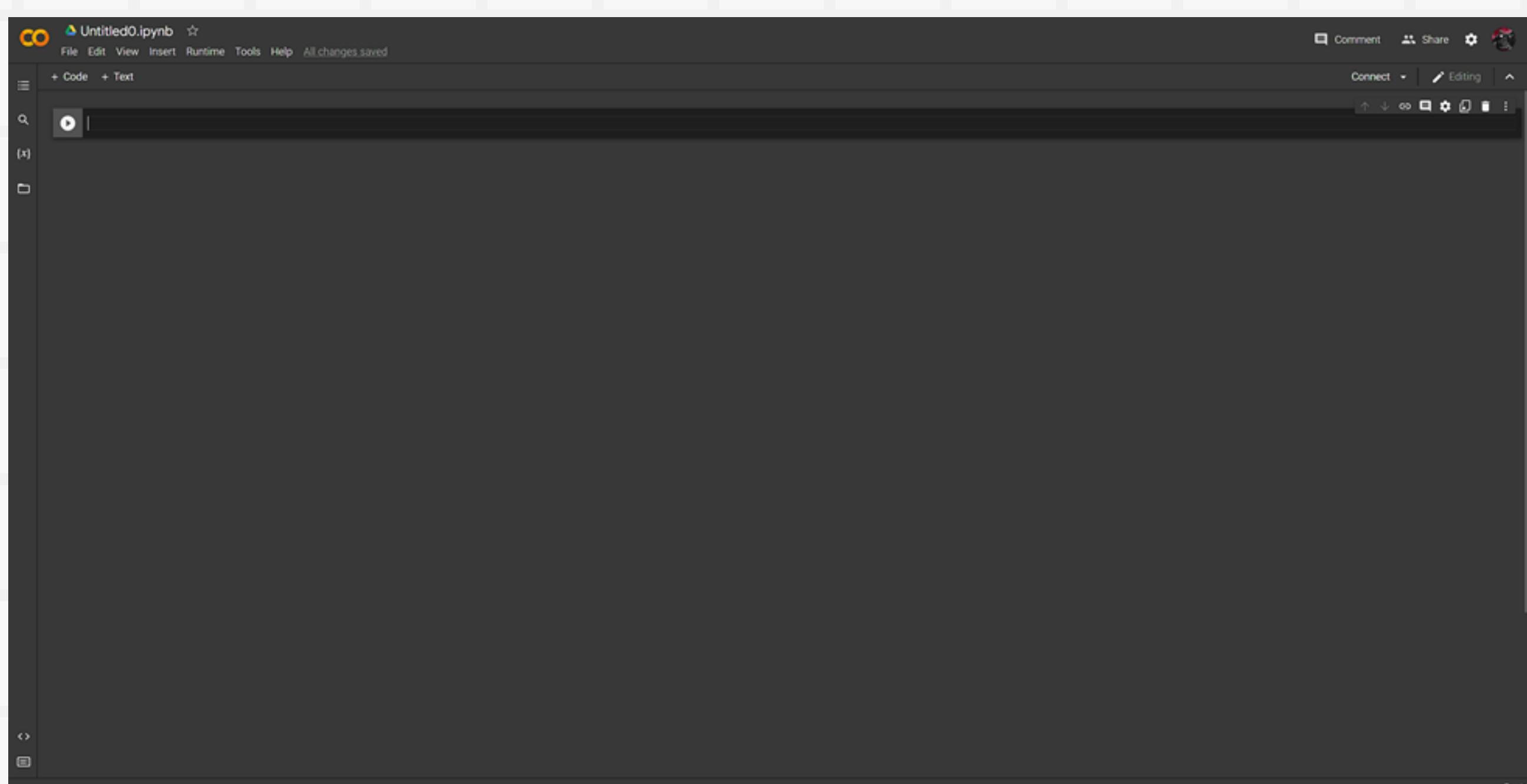
2. Luego aparecerá la siguiente pagina web que corresponde a Python Colaboratory.



3. Con nuestro mouse nos posicionamos donde dice File y le damos clic encima de él y luego buscamos la opción que diga New notebook (nuevo notebook).



4. Después que cargue la página, aparecerá nuestro ambiente de trabajo para poder codificar nuestro modelo de predicción, usando la herramienta Pandas para Python.



MODELOS PARA PROCESAR LA INFORMACIÓN

Como primer procesamiento se tiene la conversión del CSV a DataFrame, como segunda conversión se tiene el tratamiento de los datos faltantes, se eliminaron los países que no tenían ni un solo dato y así también los países que tuvieron el 40% de datos faltantes, y se tomo el criterio de llenar los datos faltantes con el valor promedio de cada columna, como tercera conversión se analizaron valores outliers del conjunto de datos, estos valores son los que están fuera del rango percentil upper y lower, estos ocasionan desviaciones muy grandes en las métricas de validación, como cuarta conversión se entrenó el modelo de aprendizaje machine learning de regresión lineal múltiple multivariable con el DataFrame limpio, luego de entrenar el modelo se obtienen las predicciones y seguidamente se evalúan las predicciones con las métricas establecidas para regresión lineal y por ultimo se presentan los resultados de las predicciones.

RESULTADOS

Como resultados de nuestro modelo de predicción se tiene lo siguiente:

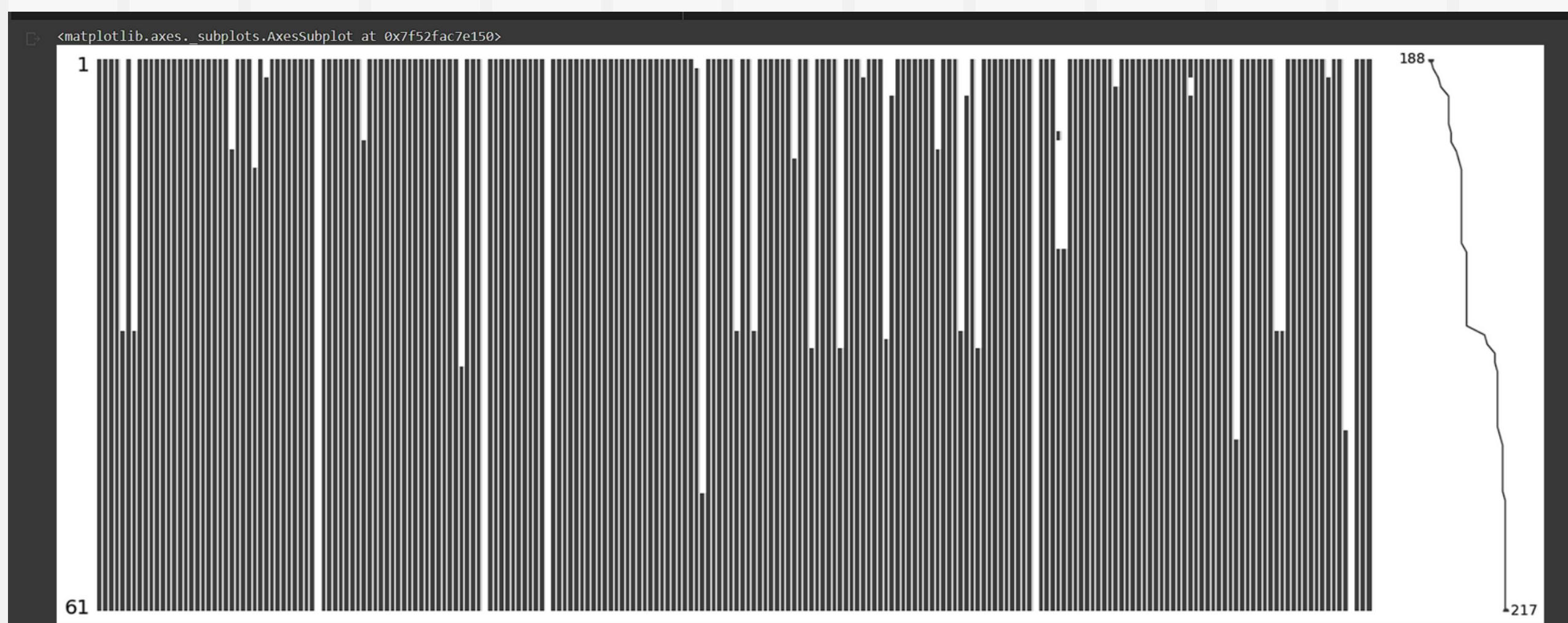


FIGURA NO.1 Visualización por medio de matriz de los datos faltantes (NAN), estos se representan por los rectángulos de color blanco.

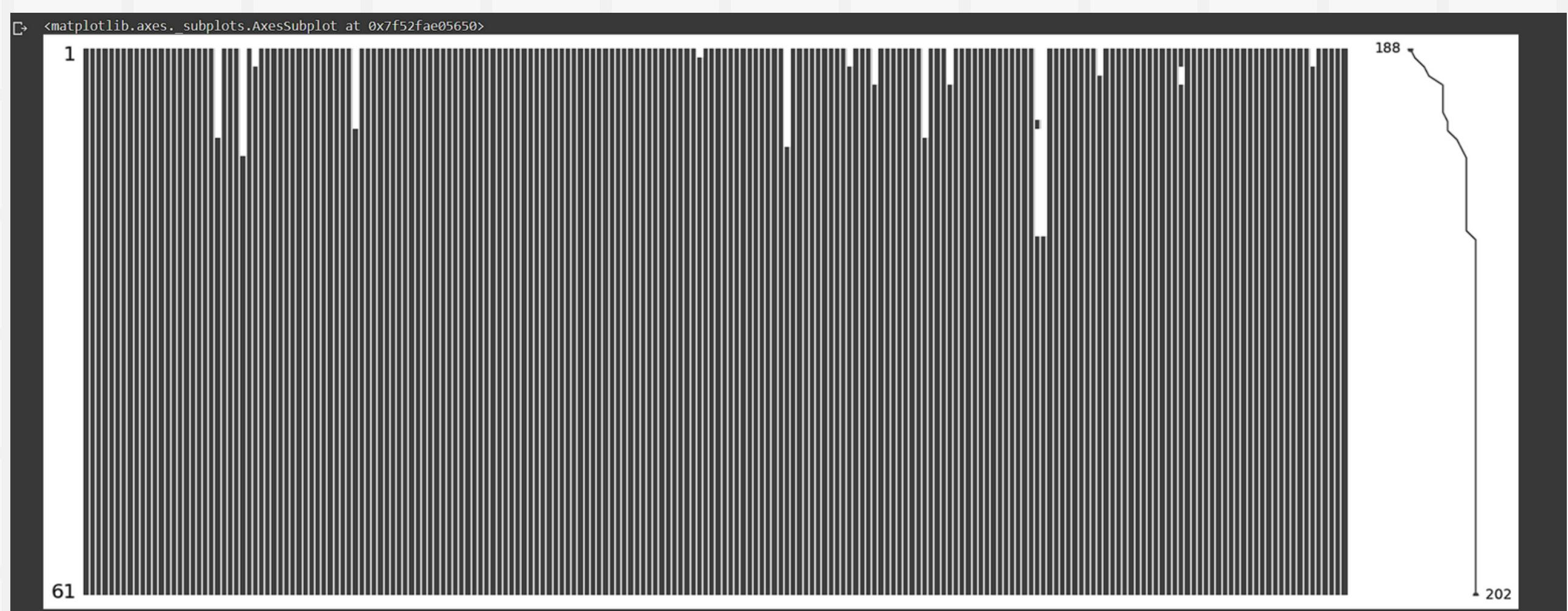


FIGURA NO.2 Matriz resultante luego de eliminar las columnas con más del 40% de datos faltantes.

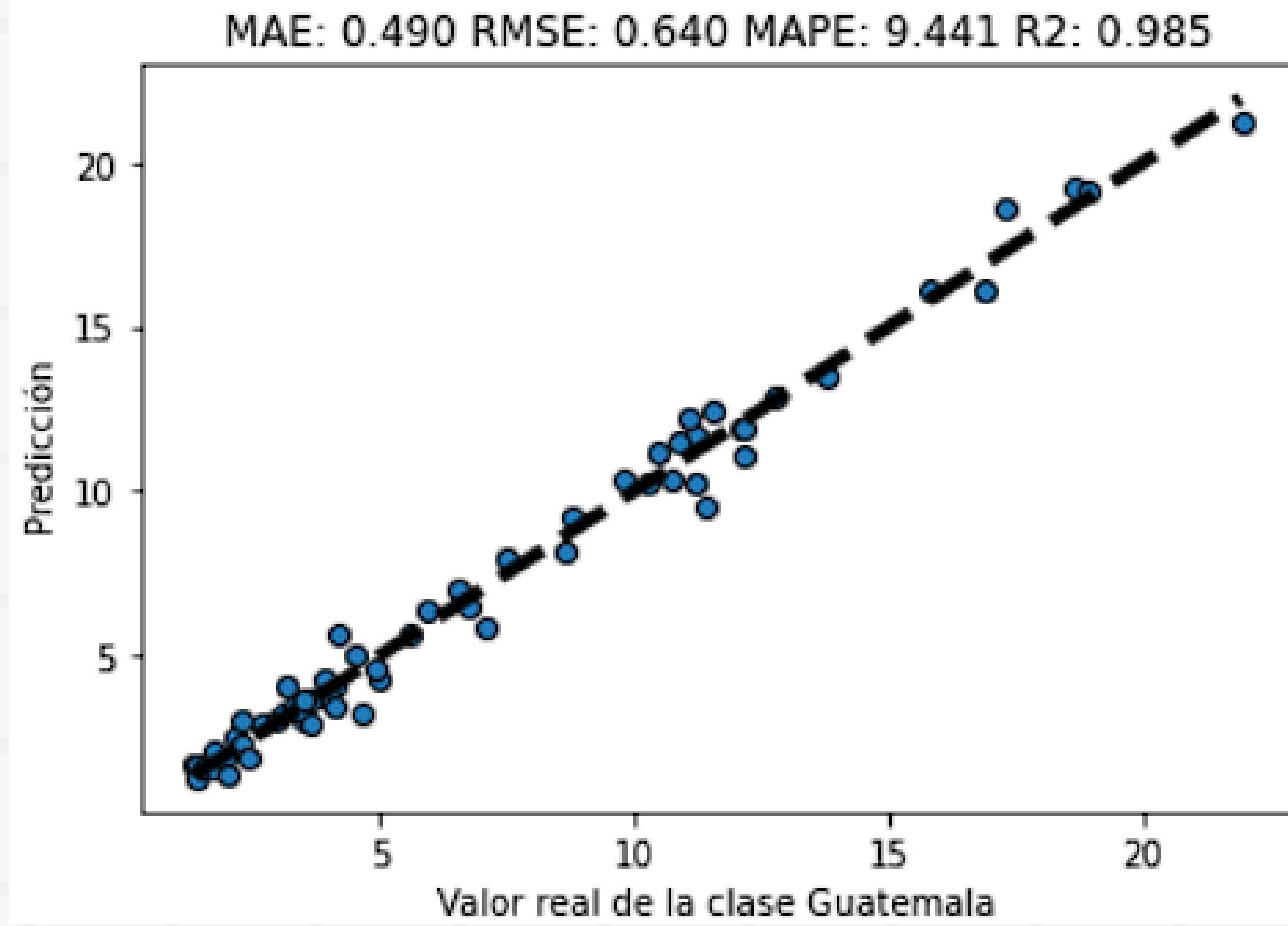


FIGURA NO.3 Representación de las predicciones (los puntos de color azul) mediante un grafica de dispersión y su respectiva línea de bondad y sus respectivas métricas.

```
[27] print(export_df[['Guatemala','prediccion']].head(61))

   Guatemala  prediccion
0      1.3442  1.664374
1      1.4065  1.661828
2      1.3809  1.278098
3      1.5199  1.623234
4      1.7690  1.630302
..        ...
56     16.9130  16.091568
57     17.2919  18.618947
58     18.6464  19.276437
59     21.9341  21.201215
60     18.9379  19.173307

[61 rows x 2 columns]
```

FIGURA NO.4 se muestran los valores predecidos vrs los valores reales guardados en el dataset.

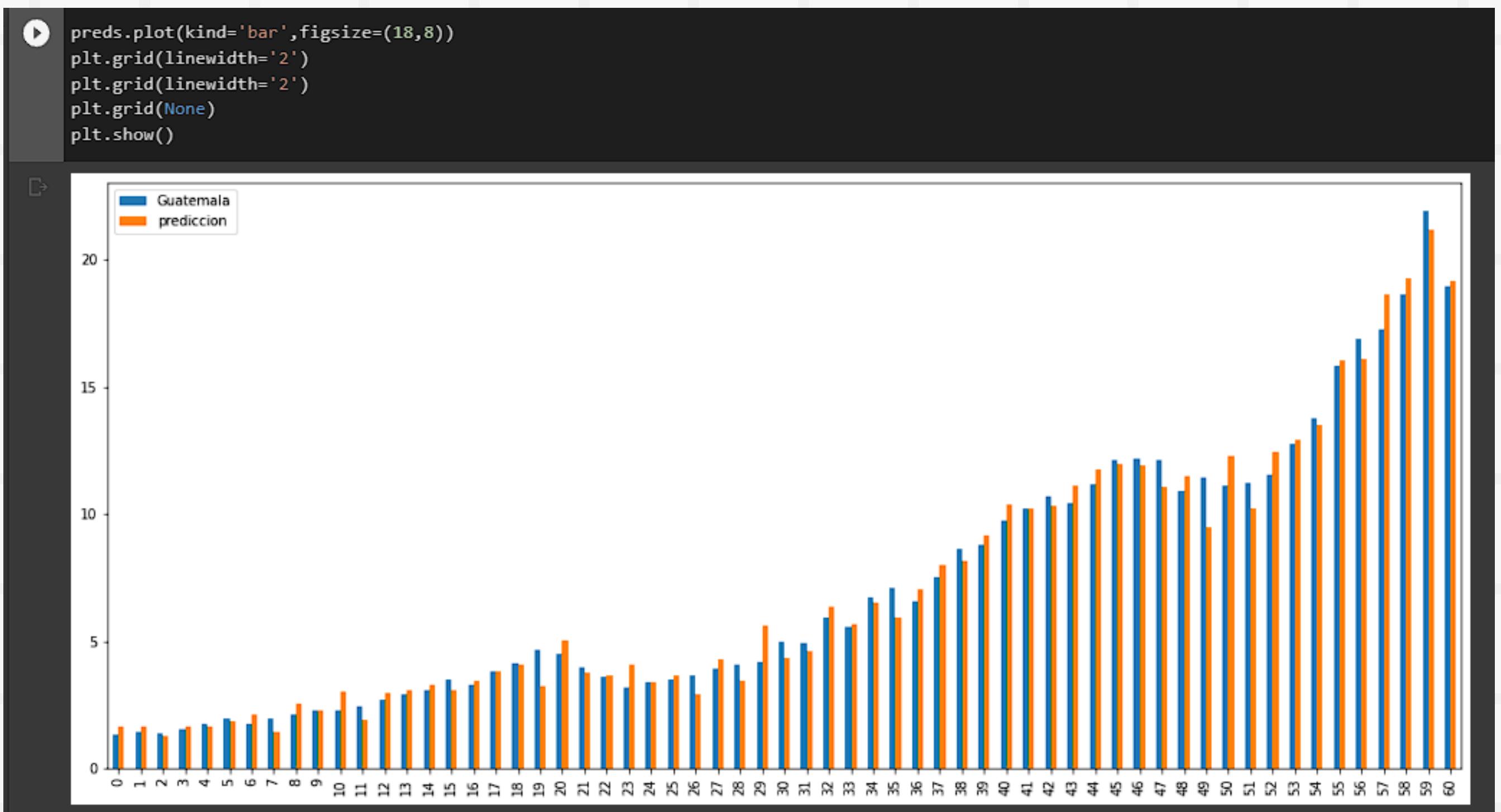


FIGURA NO.5 Representación en un histograma de barras en color azul los datos de Guatemala y de color anaranjado la predicción de nuestro modelo.

CONCLUSIONES

- Mientras exista consumismo las personas no dejarán de utilizar adquirir productos, lo que provocará que las fabricas continuen produciendo y lo que a su vez, segurá generando CO2.
- Mientras el uso del petroleo sea mayor, la actividad humana seguirá creciendo, y por ende también crecerá el uso de transportes, lo que terminará provocando una mayor generación de CO2.
- Cualquier tipo de generación de CO2 terminará dañando el ambiente.
- Nuestro modelo puede predecir en base a una serie de datos los valores de una variable en particular en este caso predecir la carga de CO2 de la ciudad de guatemala.
- Para poder predecir se puede aplicar inteligencia artificial usando técnicas de machine learning.

Entropía/Homeostasis

- Debido a que el sistema no posee un agente regulador que haga que disminuya el uso del petroleo o el consumismo humano, el sistema tiende a la entropía.

REFERENCIAS

- Chacón, J. L. (2022, 18 abril). Introducción a Pandas, la librería de Python para trabajar con datos. Profile Software Services. <https://profile.es/blog/pandas-python/#:%7E:text=%C2%BFQu%C3%A9%20es%20Pandas%3F,manipulaci%C3%B3n%20y%20tratamiento%20de%20datos>
- Greenpeace México. (2007). Huella de carbono: aprende a calcular tu impacto ambiental. <https://www.greenpeace.org/mexico/blog/9386/huella-de-carbono/>
- ¿Qué es la regresión lineal? (2002). MATLAB & Simulink. <https://la.mathworks.com/discovery/linear-regression.html>
- CO2 Emissions | Global Carbon Atlas. (2004). Global Carbon Atlas. <http://www.globalcarbonatlas.org/en/CO2-emissions>
- N. (2020, 19 diciembre). Detección de outliers en Python. Aprende Machine Learning. <https://www.aprendemachinelearning.com/deteccion-de-outliers-en-python-anomalia/>
- Cómo eliminar valores atípicos en Python en 2022 → STATOLOGOS®. (2021, 7 mayo). Statologos: El sitio web para que aprendas estadística en Stata, R y Phyton. <https://statologos.com/remove-outliers-python/>
- Kleppen, E. (2022, 24 febrero). How To Find Outliers in Data Using Python (and How To Handle Them). CareerFoundry. <https://careerfoundry.com/en/blog/data-analytics/how-to-find-outliers/>
- D. (2021, 21 enero). La regresión se refiere a problemas de modelado predictivo que implican predecir. Top Big Data. <https://topbigdata.es/metricas-de-regresion-para-el-aprendizaje-automatico/>