

# Precios del café a nivel mundial con su relación geográfica

Arguedas Gutiérrez, Josafat., Arias Sanabria, Francisco., Salazar Vargas, Gerardo y Soza Hidalgo, Angie.  
{Josafat.arguedas.gutierrez, Francisco.arias.sanabria,, Gerardo.salazar.vargas, Angie.soza.hidalgo}  
[@est.una.ac.cr](mailto:est.una.ac.cr)

Universidad Nacional de Costa Rica, Heredia, Costa Rica

*Abstract- El siguiente artículo explora las opciones del café alrededor del mundo. Empleando Python para el análisis de datos e integrando el uso de la librería Pandas y Matplotlib con el fin de analizar la relación entre el precio, zona y calidad.*

*Palabras clave, Toaster, rating, Matplotlib*

## I. INTRODUCCION

El café, esa bebida aromática y vigorizante que es disfrutada en todo el mundo, es mucho más que una simple taza de cafeína. Para los amantes del café, cada taza es una experiencia sensorial única que puede variar significativamente según el tipo de café, el grado de tostado y otros factores.

En esta era de datos y análisis, existe la ventaja de contar con poderosas herramientas informáticas, y una de las más destacadas es el lenguaje de programación Python. Python permite no solo recolectar y almacenar datos relacionados con el café, sino también realizar análisis detallados que arrojen luz sobre cómo el tostado y el valor del café pueden influir en la experiencia de los consumidores.

En este contexto, el presente documento constituye un análisis específico que utiliza Python como herramienta principal. Se aborda cómo diferentes grados de tostado, desde el "Medium-Light" hasta el "Dark", afectan la calidad percibida del café. Además, se investiga cómo el valor

promedio de los cafés tanto de Costa Rica y Colombia que son los exportadores principales de café esto expresado en términos de calificaciones y opiniones de los consumidores, influye en la percepción global del producto. Para llevar a cabo este análisis, se hace uso de una base de datos de café y se aplican técnicas estadísticas y de visualización de datos con Python.

Este estudio permitirá obtener valiosas perspectivas sobre cómo los aspectos fundamentales del café, como el tostado y el valor, pueden influir en las preferencias de los consumidores y, en última instancia, ayudar a comprender mejor el mundo del café. Además, ilustrará cómo Python, con su flexibilidad y capacidad para manejar datos complejos, se convierte en una herramienta esencial en el análisis de datos en el contexto del café.

## II. MARCO TEORICO

En el siguiente marco teórico se abordarán los elementos basados en los cuales fue posible realizar la presente investigación.

- ❖ PyISOLVER—A Fast Python OOP Implementation of LRDFIT Model
  - El artículo PySOLVER trata acerca del uso de Python. A sido implementado como una opción alternativa a algunas librerías que están bajo uso comercial, siendo una opción de código

abierto para la programación orientada a objetos. [1]

❖ PyLog: An Algorithm-Centric Python-Based FPGA Programming and Synthesis Flow

- Este artículo se refiere al uso de Python para la optimización de plataformas heterogéneas. Pylog es un compilador que infiere las instrucciones para adaptarse al hardware. Con esta herramienta se disminuye la brecha entre el código y el hardware. [2]

❖ PyOMP: Multithreaded Parallel Programming in Python

- Este artículo se refiere al uso de Python para la implementación de aplicaciones que requieren probar “stress” con multi hilos. Genera un código que se acerca al rendimiento de lenguajes de menor nivel como C. Explica los avances de la herramienta y como instalarla. [3]

❖ Python for Data Analytics, Scientific and Technical Applications

- Este artículo menciona características del lenguaje Python y el porqué se ha convertido en uno de los lenguajes en mayor crecimiento, y así mismo, menciona razones por las cuales Python se ha convertido en el lenguaje referente para aplicaciones de ciencia de datos, además de investigación y desarrollo en este campo. [4]

❖ Sakdas: A Python Package for Data Profiling and Data Quality Auditing

- Este artículo habla sobre el software Sakdas, el cual ha sido desarrollado en Python y está dividido en tres partes. La primera habla sobre como el perfilamiento de datos genera varios análisis de los cuales se sacan perfiles para definir reglas de calidad de datos. La segunda habla sobre la calidad de datos y la tercera de la visualización de datos. [5]

❖ Development of Python-MATLAB Interface Program for Optical Communication System Simulation

- Este artículo habla de un programa que conecta dos programas, es decir Python y MATLAB, juntos simulan como funcionaria un sistema de comunicación inalámbrico óptica, usamos Python para darle información a MATLAB sobre cómo queremos hacer la simulación. Luego, MATLAB hace todos los cálculos necesarios para simular cómo se comportaría el sistema de comunicación y nos da resultados y gráficos que podemos ver en Python. [6]

❖ An Empirical Study on Bugs in Python Interpreters

- El artículo investiga errores en intérpretes de Python, analizando su naturaleza y efectos en la confiabilidad del lenguaje. Los autores exploran tipos de bugs, estadísticas y soluciones para mejorar la calidad y fiabilidad de los intérpretes de Python. Publicado en IEEE Transactions on Reliability en 2022. [7]

❖ Studying Physics, Getting to Know Python: RC Circuit, Simple

#### Experiments, Coding, and Data Analysis With Raspberry Pi

- El artículo se centra en el estudio de la física mediante un circuito RC, experimentos simples, programación en Python y análisis de datos con Raspberry Pi. Este enfoque combina la física experimental con habilidades de programación para analizar datos de un circuito RC utilizando Python y Raspberry Pi. El artículo fue publicado en el año 2021. [8]
- ❖ Pymote 2.0: Development of an Interactive Python Framework for Wireless Network Simulations
- En este artículo se estudia el rendimiento y efectividad de Pymote, una librería de Python de alto nivel para simulación basada en eventos de algoritmos distribuidos en redes móviles de tipo ad-hoc. [9]

necesarias para realizar un análisis de los datos de café y presentar los resultados de una mejor manera.

- B. Influencia de los Aspectos del Café en las Preferencias del Consumidor: Este estudio busca comprender cómo factores como el tostado y la zona de producción, pueden influir en el precio que los consumidores están dispuestos a pagar.
- C. Comparación entre Países Exportadores: El análisis incluye una comparación entre los cafés de Costa Rica y Colombia, dos de los principales países exportadores de café. Esto permite examinar cómo las diferencias en la calidad y el valor de los cafés de estos países pueden afectar la percepción global del producto y las preferencias de los consumidores.

- ❖ Enhancing PowerFactory Dynamic Models with Python for Rapid Prototyping
- Este artículo habla sobre la herramienta PowerFactory, la cual corresponde a una de las herramientas de análisis de sistemas de poder en la industria y en el campo de investigación, y como su uso en Python la hace más efectiva que en otros lenguajes como C o C++.[10]

- A. Herramientas de Análisis de Datos - Pandas y Matplotlib: Pandas es una biblioteca de Python que se utiliza ampliamente para el análisis de datos, mientras que Matplotlib es una biblioteca para la visualización de datos. Estas herramientas son

Los datos se obtuvieron de Kaggle.com, una página que alberga datos para una comunidad de científicos y practicantes de Machine Learning para la compañía Google. El link de descarga de los datos es <https://www.kaggle.com/datasets/schmoyote/coffee-reviews-dataset/download?datasetVersionNumber=6>

### III. DESCRIPCION DE PAGINA

### IV. DESCRIPCION DE PROCESAMIENTO DE LA INFORMACION

El procesamiento de los datos en este estudio se lleva a cabo mediante el uso de Python. Estos datos se organizan en estructuras de datos que permiten una fácil consulta. Se aplican técnicas estadísticas y de visualización de datos para explorar las relaciones entre el precio, la zona de cultivo y la calidad del café. Se calculan

estadísticas descriptivas, como promedios y desviaciones estándar, para comprender la distribución de los precios y las calificaciones. Además, se utilizan gráficos y visualizaciones para representar los patrones y tendencias identificados en los datos. Este proceso de análisis de datos con Python proporciona una base para el estudio y ayuda a obtener perspectivas sobre cómo los diferentes factores afectan la percepción del café por parte de los consumidores.

## V. ANALISIS DE LOS DATOS

### FUNCIONES:

La siguiente es una función que muestra el promedio de tostados de café:

```
def promedio_cafe(df):
    tostados = ['Medium-Light', 'Medium-Dark', 'Medium', 'Light', 'Dark']
    promedios = {}

    total_cafes = len(df) # Total de
    cafés en el DataFrame

    for tostado in tostados:
        df_tostado = df[df['roast'] ==
        tostado]
        if not df_tostado.empty:
            promedio =
            df_tostado['rating'].mean()
            promedios[tostado] =
            round(promedio, 2) # Redondear a 2
            decimales
        else:
            promedios[tostado] = 0
        # Si no hay cafés con este tostado, el
        promedio es cero

        # Calcular la suma de los promedios
        suma_promedios =
        sum(promedios.values())

        # Ajustar los promedios para que
        sumen 100%
        for tostado in tostados:
            if promedios[tostado] != 0:
                promedios[tostado] =
                round((promedios[tostado]
                /
                suma_promedios) * 100, 2)

    return promedios, total_cafes
```

En el siguiente código podremos analizar los cafés provenientes de Costa Rica y

### Colombia:

```
def cafe_origenes(datos):
    # Crear una nueva columna llamada
    "Origen" que combine los valores de
    origin_1 y origin_2
    datos["Origen"] =
    datos["origin_1"].fillna('') +
    datos["origin_2"].fillna('')

    # Contar la cantidad de registros
    que tienen "Costa Rica" en la columna
    "Origen"
    costa_rica_count =
    (datos["Origen"].str.contains("Costa
    Rica")).sum()

    # Contar la cantidad de registros
    que tienen "Colombia" en la columna
    "Origen"
    colombia_count =
    (datos["Origen"].str.contains("Colombia"
    )).sum()

    # Imprimir los resultados
    print("Marcas de Café provenientes
    de Costa Rica:", costa_rica_count)
    print("Marcas de Café provenientes
    de Colombia:", colombia_count)

cafe_origenes(datos)
```

El siguiente código calcula el rating según el origen del café, se puede analizar lo mismo que en el pasado ver el rating de Costa Rica y Colombia:

```
def calcular_media_rating(df, origen1,
origen2):
    # Filtrar los datos por origen
    data_origen1 = df[df['origin_1'] ==
origen1]
    data_origen2 = df[df['origin_1'] ==
origen2]

    # Calcular la media de los valores
    median_rating_origen1 =
data_origen1['rating'].median()
    median_rating_origen2 =
data_origen2['rating'].median()

    # Imprimir el resultado
    print("Media de rating para café de " +
origen1 + ":", median_rating_origen1)
    print("Media de rating para café de " +
origen2 + ":", median_rating_origen2)
```

La salida del código anterior brinda que la media para rating del café de Costa Rica es de: 93.0 y la media de rating para el café de Colombia es de: 92.0

El siguiente código brinda una tabla de cafés con mejores ratings según su origen:

```
def tabla_ratings(df, origen):
    # Filtrar los datos por origen
    data_origen = df[df['origin_1'] ==
origen]

    # Extraer las columnas "rating" y
    "name" de los datos filtrados
    ratings = data_origen['rating']
    names = data_origen['name']

    # Crear un DataFrame con los nombres
    del café y las calificaciones
    dtf = pd.DataFrame({'Coffee Name':
names, 'Rating': ratings})

    # Imprimir un título
    print("Tabla de Calificaciones")
    print(f"Origen: {origen}\n")

    # Imprimir la tabla usando tabulate
    para un formato más bonito
    print(tabulate(dtf, headers='keys',
tablefmt='psql', showindex=False))
```

En este caso lo evaluamos para Costa Rica y estos fueron los resultados:

Coffee Name	Rating
Costa Coast Blend	92
Flight Seasonal Espresso	94
Coffea Diversa Cioiccie	92
Costa Rica Farmer's Project	90
Costa Rica Aris Red Honey Lot 1901 Espresso	95
Costa Rica Yellow Honey Lot Kinka Bear	94
Costa Rica Anaerobic	93
Costa Rica Yellow Honey Kinka Bear Lot 20-02	94
Costa Rica Black Honey	94
Costa Rica Black Honey	94
Chokoreeto Espresso Blend	92
Secret Lovers Blend	93
44°N 68°W Bar Harbor	92

Backwoods Espresso	93
--------------------	----

El siguiente código dice el promedio del precio del café por gramo en Costa Rica:

```
def promedio_precioCafe_CR():
    costa_rica_data =
    datos[datos['origin_1'] == 'Costa Rica']
    median_price_CR =
    costa_rica_data['100g_USD'].median()
    print("Media de precio para café de
    Costa Rica:", median_price_CR)
```

El código anterior indica que el precio promedio es de 4.88 dólares por 100 gramos

El siguiente código brinda una tabla con los precios en Guatemala por 100 gramos según el café:

Nombre	Precio
Dead Reckoning	5.0
Summer Night Blend Espresso	3.54
Cold Brew	0.12
Espresso 1849 Blend	5.29
Cupid Blend	7.2
No. 6 Espresso Blend	3.47
Coffea Diversa Bourbon Rey Guatemala	17.61
La Finca Medina of Guatemala	5.29
Guatemala (from bulk bin)	1.54
100% Guatemalan	0.99
Session Blend	5.14
Amazon Haze	5.0
Heart & Soul Espresso Blend	5.29
RD1 Espresso Blend	4.25
Bull Espresso Blend	3.17
The World Espresso Blend	2.2
Espresso Fino	4.7
Campfire Blend	5.0
Golden Parrot Blend	2.97

Chocolate Lover's Espresso	4.27
Espresso Blend	7.05
Session Blend	5.14
Kasasagi	8.0

El siguiente código muestra el promedio del precio del café por 100g según su país de localización:

```
def
promedio_precio_por_paisDeLocalizacion():
    valor_por_pais =
    datos.groupby("loc_country")["100g_USD"].me
    an()

    # Redondear el valor promedio a 2 dígitos
    después de la coma
    valor_por_pais = valor_por_pais.round(2)

    # Crear un DataFrame con los valores
    promedio
    data_dict = {'País de localización':
    valor_por_pais.index, 'Valor promedio (USD
    por 100g)': valor_por_pais.values}
    df = pd.DataFrame(data_dict)

    # Utilizar tabulate para imprimir la
    tabla de manera bonita
    table = tabulate(df, headers='keys',
    tablefmt='pretty', showindex=False)
    return table
```

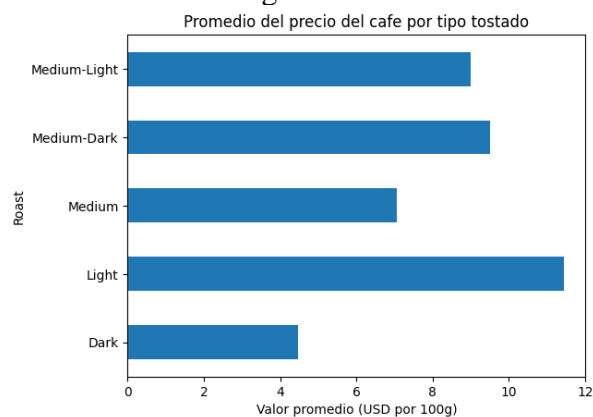
El resultado saliente es el siguiente:

País de localización	Promedio
Australia	42.78
Belgium	3.02
Canada	5.5
China	26.22
England	92.18
Guatemala	3.62
Hawai'i	17.55
Honduras	11.46
Hong Kong	13.32
Japan	13.24
Kenya	6.91
Malaysia	8.11
Mexico	10.58
New Taiwan	9.46

Peru	4.0
Taiwan	9.16
Uganda	3.96
United States	8.46
United States And Floyd	5.88
Australia	42.78
Belgium	3.02
Canada	5.5
China	26.22

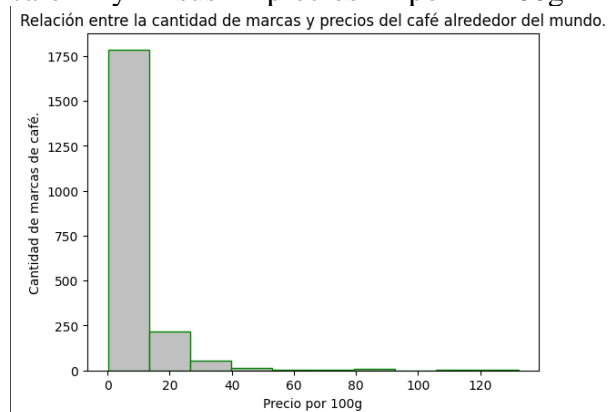
## GRAFICAS:

En el siguiente grafico se observa el precio del café según su tostado



Se puede observar como el tostado “Ligth” tiene a ser más valorado con respecto a los demás tostados

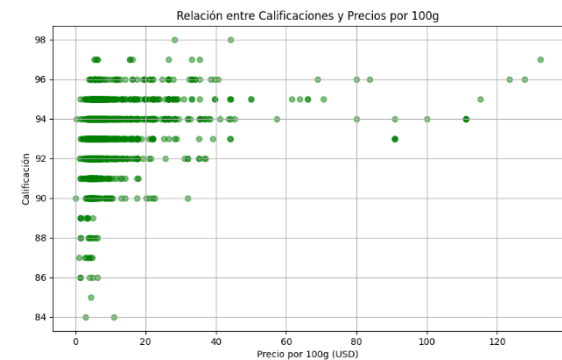
Continuación una gráfica de las marcas de café y sus precios por 100g



Acá se observa como su mayoría de marcas de café van de 0 a los 20\$ por 100g de café

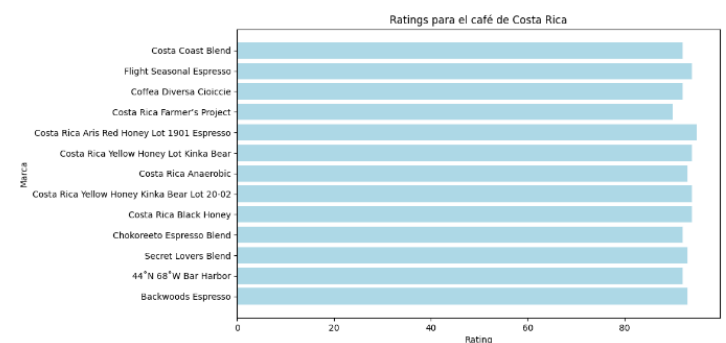
y como solo unas pocas llegan a valer más de 100\$ por 100g de café

La siguiente muestra la relación entre las calificaciones y sus precios por 100g



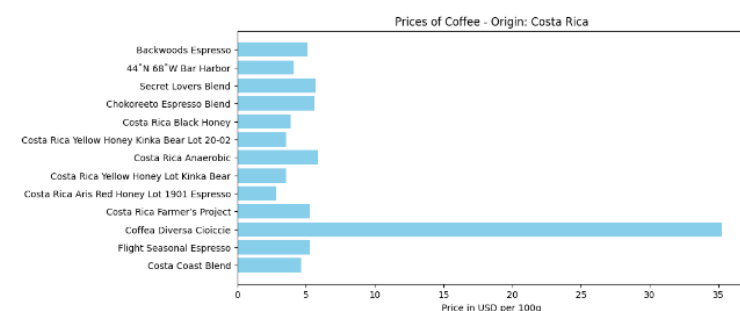
Se puede observar como el café más caro no es el que tiene mejor puntuación, sino que el mejor puntuado se coloca entre los 20\$ y 40\$ por 100g

Ratings de las marcas de café de Costa Rica



En Costa Rica la marca “Costa Rica Aris Red Honey lot 1901 espresso” es quien tiene mejor rating

Precios de café según su marca en Costa Rica



El café “Coffea Diversa ciocci” Tiene el mejor precio superando los 35\$ por 100g donde el origen es Costa Rica.

## VI. CONCLUSIONES Y RECOMENDACIONES

### CONCLUSIONES

- Una conclusión que se puede obtener de la presente investigación que un factor como lo es el tostado (“toaster”) del café posee una gran influencia en la calidad percibida por los consumidores de café. Por la tanto, el tostado de café siempre corresponderá un factor clave a la hora de escogencia de café de preferencia.
- Otra conclusión que puede definirse es como, de forma curiosa e interesante, el precio de café por 100 gramos de este varia, en algunos casos, de forma bastante notable y significativamente, según el país de localización de café. Por ejemplo, es observable que algunos de los precios más barato de café por 100 gramos de este pueden observarse en Guatemala, mientras que la contraparte de estos, algunos de los precios más caros, pueden encontrarse en el país de Australia. Esta significancia en el precio del café según el país de localización podría deberse a una gran variedad de factores ajenos cada uno al país en cuestión.
- También puede concluirse que existen diferencias entre tanto la calidad y el valor de Costa Rica y Colombia, dos países que destacan globalmente por ser de los principales motores exportadores de café, siendo esto algo que afecta tanto la percepción global como la preferencia personal de los consumidores. Por ejemplo, al tomar en cuenta las medias de calificación de ambos países, siendo 93.0 para Costa Rica y 92.0 para Colombia, se da a

entender que en la preferencia personal de los consumidores de café existe un favoritismo hacia el café de Costa Rica en comparación al café colombiano. Así mismo, este es un factor que destaca la fundamental importancia que posee la producción de café en países tanto de Centroamérica como de Sudamérica.

- Otro aspecto más que puede concluirse es que existen países donde el precio promedio por 100 gramos de café es considerablemente alto en comparación con el resto de los países. Ejemplos de esto pueden ser las naciones de Australia e Inglaterra que poseen un valor promedio de café por 100 gramos de 42.78 y 92.18 respectivamente, algo cuya explicación puede deberse a variables como al costo de importancia y el poder adquisitivo promedio de los ciudadanos de estos países, que son ambos de primer mundo
- Otra conclusion consiste en el enomr impacto que el lenguaje de programacion Python ha tenido en la comunidad cientifica global, como se puede constatar en los proyectos y articulos del presente documento, que han influido en la investigacion de diversos ambitos. Esto acentua el impacto positivo de Python en la comunidad cientifica y capacidad para abarcar una gran gama de retos invetigativos-
- Finalmente, se puede concluir que el lenguaje de programación Python constituye una excelente herramienta para el análisis de datos en la industria del café. La capacidad flexibilidad y fácil adaptación de Python ofrecen una herramienta de una utilidad sumamente efectiva para manejar datos complejos, lo que convierte a este lenguaje de programación en un aliado fundamental para la investigación y análisis de datos en



diversos ámbitos, siendo la industria del café uno de estos.

## RECOMENDACIONES

1. Una recomendación consiste en que las empresas productoras de café hagan un uso beneficioso de las opciones de tostado de café utilizando la personalización de estos, o sea, que pueden empezar a lanzar versiones de tostados que sean populares en una localidad específica para poder tener ventas fructíferas en esa área. Para esto, claramente, primero se deberían hacer estudios de las preferencias de los consumidores en determinada área.
2. Otra recomendación consiste en realizar marketing según el país de origen del café. El marketing basado en la identidad nacional puede resultar más atractivo para la población que no es de determinado país, ya que esto permite resaltar características propias del café producido en un determinado país. Naturalmente, muchos consumidores de café buscaran opciones diferentes a las que están acostumbrado los que resultan en estos adquiriendo productos que se pueden publicitar mediante este método.
3. Otra recomendación puede ser el comparar este análisis del mercado de café con el mercado global de otras bebidas populares, como lo pueden ser el té, las gaseosas, las bebidas hidratantes y las bebidas energéticas. El hacer esto podría brindar datos interesantes para realizar comparaciones relativas entre estas 2 bebidas, y así mismo podría mostrar la preferencia de los consumidores de café en el contexto de bebidas que son competencia del mercado del café.
4. Otra recomendación consiste en explorar el variado número de bibliotecas de visualización que posee el lenguaje de programación Python además de Matplotlib, como, por ejemplo, Seaborn o Bokeh, las cuales ofrecen una gran variedad de gráficos y crean efectivas y diferentes visualizaciones.
5. Finalmente, otra recomendación podría ser el observar como la situación de la pandemia empezada en el 2020 por el coronavirus afectó en aspectos precios y consumo al mercado del café en el ámbito global. Hay que recordar que esto fue un evento que afectó de forma negativa la mayoría de los mercados globales dicho año y sería interesante hacer comparaciones en determinadas variables con el fin de visualizarlas de una mejor forma.

## VII. REFERENCIAS

- [1] F. Yang and J. E. Menard, "PyISOLVER—A Fast Python OOP Implementation of LRDFIT Model," *IEEE Trans. Plasma Sci.*, vol. 48, no. 6, pp. 1793–1798, 2020, doi: 10.1109/TPS.2019.2958001.
- [2] S. Huang, K. Wu, H. Jeong, C. Wang, D. Chen, and W.-M. W. Hwu, "PyLog: An Algorithm-Centric Python-Based FPGA Programming and Synthesis Flow," *IEEE Trans. Comput.*, pp. 1–1, 2021, doi: 10.1109/TC.2021.3123465.
- [3] T. G. Mattson, T. A. Anderson, and G. Georgakoudis, "PyOMP: Multithreaded Parallel Programming in Python," *Comput. Sci. Eng.*, vol. 23, no. 6, pp. 77–80, Nov. 2021, doi: 10.1109/MCSE.2021.3128806.
- [4] [1]A. Nagpal and G. Gabrani, "Python for Data Analytics, Scientific and Technical Applications," in 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates: IEEE, 2019, pp. 140–145. doi: 10.1109/AICAI.2019.8701341.
- [5] S. Loetpipatwanich and P. Vichitthamaros, "Sakdas: A Python Package for Data Profiling and Data Quality Auditing," in 2020 1st International Conference on Big Data Analytics

- and Practices (IBDAP), Bangkok, Thailand: IEEE, Sep. 2020, pp. 1–4. doi: 10.1109/IBDAP50342.2020.9245455.
- [6] Y. Hwang, D. Choi, H. An, S. Shin, and C. G. Lee, “Development of Python-MATLAB Interface Program for Optical Communication System Simulation,” in 2019 International Conference on Green and Human Information Technology (ICGHIT), Kuala Lumpur, Malaysia: IEEE, 2019, pp. 46–48. doi: 10.1109/ICGHIT.2019.00018.
- [7] Z. Wang, D. Bu, A. Sun, S. Gou, Y. Wang, and L. Chen, “An Empirical Study on Bugs in Python Interpreters,” *IEEE Trans. Rel.*, vol. 71, no. 2, pp. 716–734, 2022, doi: 10.1109/TR.2022.3159812.
- [8] A. Mandanici, S. Alessandro Sara, G. Fiumara, and G. Mandaglio, “Studying Physics, Getting to Know Python: RC Circuit, Simple Experiments, Coding, and Data Analysis With Raspberry Pi,” *Comput. Sci. Eng.*, vol. 23, no. 1, pp. 93–96, Jan. 2021, doi: 10.1109/MCSE.2020.3037002.
- [9] F. Shahzad, “Pymote 2.0: Development of an Interactive Python Framework for Wireless Network Simulations,” *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1182–1188, 2016, doi: 10.1109/JIOT.2016.2570220.
- [10] C. D. Lopez, M. Cvetkovic, and P. Palensky, “Enhancing PowerFactory Dynamic Models with Python for Rapid Prototyping,” in 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), Vancouver, BC, Canada: IEEE, 2019, pp. 93–99. doi: 10.1109/ISIE.2019.8781432.