Impacto de la Covid-19 en la mortalidad

|  |  |  |  |
| --- | --- | --- | --- |
| Kendra Artavia Caballero  *Universidad Nacional de Costa Rica,*  Heredia, Costa Rica  [kendra.artavia.caballero@est.una](mailto:kendra.artavia.caballero@est.una.ac.cr)  [.ac.cr](mailto:kendra.artavia.caballero@est.una.ac.cr) | Alexia Alvarado Alfaro  *Universidad Nacional de Costa Rica*,  Heredia, Costa Rica  [alexia.alvarado.alfaro@est.una.ac.cr](mailto:alexia.alvarado.alfaro@est.una.ac.cr) | Jose Soto Pérez  *Universidad Nacional de Costa Rica*,  Heredia, Costa Rica  [jose.soto.perez@est.una.ac.cr](mailto:jose.soto.perez@est.una.ac.cr) | Johey Artavia Barrantes  *Universidad Nacional de Costa Rica*,  Heredia, Costa Rica  [johey.artavia.barrantes@est.una.ac.cr](mailto:johey.artavia.barrantes@est.una.ac.cr) |

**Abstract—** **El presente estudio analiza el impacto de la Covid-19 en la mortalidad a nivel global, utilizando funciones del Lenguaje Python para el análisis de datos. Se emplean las librerías Pandas y Matplotlib para la obtención de datos y la creación de gráficas representativas. Este análisis busca comprender y visualizar el impacto de la pandemia en términos de mortalidad a nivel global.**

**Keywords—*Python, Análisis de datos, COVID-19, Mortalidad***

I. INTRODUCCIÓN

En el contexto actual, caracterizado por un constante aumento en la disponibilidad de información, surge la imperiosa necesidad de analizar, examinar y gestionar los datos relacionados con el impacto de la COVID-19 en la mortalidad. Este estudio se enfoca en la recopilación y análisis de datos provenientes del sitio oficial de la Organización Mundial de la Salud (OMS), específicamente de la sección 'Global Excess Deaths Associated with COVID-19 Modelled Estimates', que proporciona valiosa información sobre las muertes asociadas con la pandemia a nivel mundial. Para llevar a cabo este análisis, se emplearán herramientas tecnológicas avanzadas, como el entorno de desarrollo para Python y la librería Pandas, aprovechando además el recurso de archivos CSV disponibles. Este enfoque tecnológico permite no solo analizar datos históricos, sino también prever y estudiar la efectividad de las medidas adoptadas por las autoridades sanitarias y gubernamentales en la gestión de la pandemia. En un contexto de avances tecnológicos y sociales, este estudio busca proporcionar estadísticas fundamentales y conclusiones valiosas para comprender el impacto de la COVID-19 en la mortalidad a nivel global, así como identificar posibles mejoras y estrategias futuras para abordar esta crisis de manera efectiva.

Además de la urgente necesidad de comprender el impacto de la COVID-19 en la mortalidad, este estudio se enmarca en un contexto donde la interacción entre la salud pública, la tecnología y la toma de decisiones ha adquirido una relevancia sin precedentes. Este enfoque analítico brinda la oportunidad de evaluar la efectividad de las estrategias implementadas a nivel global para contener la propagación del virus y mitigar su impacto en la salud pública. La combinación de herramientas tecnológicas avanzadas y datos precisos provenientes de fuentes confiables como la OMS nos ofrece una perspectiva única para abordar los desafíos que plantea esta crisis sanitaria a escala mundial.

II. MARCO TEÓRICO

1. *Automatización del análisis de datos con Python: un estudio comparativo de bibliotecas populares y su aplicación.*

En la investigación realizada en el Departamento de CSE de la Universidad de Chandigarh, Mohali, Punjab, India, se llevó a cabo un estudio comparativo de las bibliotecas de Python más populares para la automatización del análisis de datos. [1] El estudio encontró que el nivel de experiencia del usuario y el tipo de tarea de análisis de datos son factores importantes a la hora de elegir la biblioteca adecuada.

El estudio ofrece información detallada sobre las fortalezas y debilidades de las bibliotecas Python conocidas y populares para automatizar trabajos de análisis de datos. Y también los hallazgos del estudio pueden ayudar a las personas a tomar decisiones informadas en una variedad de industrias, incluidas las empresas, la atención médica y las finanzas. Este estudio examinó la eficiencia y efectividad de seis bibliotecas de Python conocidas para trabajos de automatización de análisis de datos. Las bibliotecas que se analizaron fueron Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn y TensorFlow se analizaron en función de métricas como eficiencia, facilidad de uso y exactitud.

1. *Implementación de análisis de datos y resumen de documentos en datos de redes sociales utilizando R y Python*

Un estudio realizado en el Department of Computer Science Engg, IFET College of Engineering, Villupuram, India, propone unmétodo integral para analizar y resumir los resultados de sitios web de redes sociales que pueden enviar sus flujos de datos a sus respectivas nubes. [2]

El enfoque sugerido detectaría valores atípicos en los flujos de datos y podría probarse secuencialmente utilizando programación R y análisis de datos. El estudiologra clasificar el material infringido de las redes sociales y sus mensajeros de una manera efectiva para enseñar a estudiantes y profesionales cómo comportarse de manera ética en las plataformas sociales. [2] Los datos estadísticos se examinaron utilizando el lenguaje de programación Python.

1. *Análisis y predicción de datos de calidad del aire basados en regresión lineal utilizando Python*

En una investigación realizada en el Department of Electrical and Electronics Engineering, Amrita School of Engineering, Coimbatore Amrita Vishwa Vidyapeetham, India, utilizó un modelo de regresión lineal para analizar y predecir la calidad del aire en la ciudad de Coimbatore.[3]

En él se analizan los datos, se calcula y visualiza la proporción del principal contaminante PM2,5. Utilizando el modelo de regresión, el conjunto de datos se utiliza para entrenar y probar el modelo en el que se ejecuta la limpieza, el preprocesamiento, el entrenamiento y las pruebas de datos.Se utilizan varias bibliotecas incorporadas en la plataforma Python para la parte de predicción y análisis de datos.

1. *Análisis de datos financieros y cuantificación de riesgos basado en Python*

En el contexto actual de mercados financieros volátiles, la gestión de riesgos y la inversión cuantitativa son herramientas esenciales para los inversores. Un estudio realizado por investigadores de la Universidad Wenzhou-Kean, Wenzhou, China, ha desarrollado un modelo innovador para el análisis de datos financieros que puede ayudar a los inversores a tomar decisiones más informadas [4]

En el estudio se logró determinar como una estrategia de búsqueda de los datos mediante un URL. La aplicación de Python también fue utilizada para hacer diversos análisis cuantitativos como el método de varianza que ayuda a ver los riesgos.

1. *Uso del análisis de datos exploratorios para generar inferencias sobre la correlación de casos de COVID-19*

Una investigación realizada en el Departamento de Ingeniería y Ciencias de la Computación, Amity University Dubai, Dubai, EAU, utilizó técnicas de análisis de datos para comprender los efectos de la pandemia Covid-19. [5]

El objetivo de este artículo era obtener conocimiento sobre los efectos de la pandemia con respecto a las variables y etiquetas proporcionadas en el conjunto de datos. En él podemos ver cómo se logró presentar un análisis de los datos para formular patrones relacionados con la actual pandemia de Covid-19 y obtener información importante de datos. En él podemos ver el uso de Python mediante el uso de gráficos para generar una mejor visualización de los datos, visualización basada en las bibliotecas de análisis de datos de Python como NumPy, Pandas, Matplotlib, Seaborn, ggplot, Plotly, PyQtGraph, VisPy, Bokeh, Altair y Pygal.[5]

1. *Aplicación de Python en la educación en ciencias del cerebro*

Una investigación realizada en Osaka, Japón, tiene como objetivo demostrar el uso de técnicas de neuroimagen en educación en ciencias del cerebro, específicamente en la manipulación y análisis de datos de neuroimagen.[6] El artículo logró presentar casos de enseñanza y práctica con herramientas relevantes basadas en el lenguaje de programación Python y su ecosistema tecnológico, con el fin de ayudar a los estudiantes a adquirir habilidades fundamentales para la manipulación y análisis de datos de neuroimagen.

Además, busca que los estudiantes utilicen tecnologías relacionadas con el ecosistema de Python para dominar técnicas de inteligencia artificial como clustering K-means, clustering Ward, PCA, ICA, análisis a gran escala y otras tecnologías relacionadas con Python utilizadas en educación en ciencias del cerebro. Se dio la aplicación Python en neuroimaging Preprocessing With Python, fMRI Manipulation with Python, fMRI Data Normalization with Python, Neuroimaging Analysis with Python, Functional Connectivity Analysis with Python, Multivoxel Pattern Analysis with Python.

1. *Análisis de datos de Twitter para identificar el bienestar intuitivo de la salud mental.*

Un estudio presentado en la 5ta Conferencia Internacional sobre Sistemas de Información y Redes de Computadoras (ISCON) en Mathura, India, propuso un modelo para analizar la salud mental de los usuarios de Twitter a partir de sus tweets como nuestra entrada, y al analizar los tweets de las personas generamos la polaridad de los tweets, lo que resulta en el resultado de la salud mental intuitiva [7]

Este artículo describe una mente sana en general que es necesaria para el bienestar físico y mental de las personas, proporciona un modelo para un análisis de sentimientos que detecta pensamientos positivos alentadores, puntos de vista negativos y pensamientos neutrales, así como varios usuarios, basándose en la técnica del análisis sentimental. El trabajo incluye preprocesar y segregar datos mediante el clasificador Naive Bayes, y luego encontrar las vistas a través de la polaridad y la subjetividad. Estos sentimientos muestran la salud mental intuitiva de la persona; además, sugiere el uso del análisis de sentimientos para analizar la salud mental de las personas.

Con lo obtenido se puede analizar que, si la proporción de polaridad negativa o positiva de los tweets es mayor, según la técnica de análisis sentimental, el usuario podría tener algunos problemas de salud mental. Entonces, a través del análisis, se intenta hacer un análisis de la salud mental intuitiva de las personas y brindarles alguna ayuda. Se utilizaron diferentes bibliotecas, por ejemplo: textblob, matplotlib, wordcloud, pandas, NumPy. Matplotlib es una biblioteca que se utiliza para trazar gráficos 2D. Numpy y Pandas son las dos bibliotecas de Python que se utilizan para cálculos matemáticos. Wordcloud es la biblioteca utilizada para la visualización de datos según la frecuencia de las palabras. A partir de estos se crearon diferentes gráficos y figuras, entre ellos el gráfico disperso de nuestros hallazgos a través del análisis de sentimiento, gráfico de barras que representa la polaridad del sentimiento con el recuento y la nube de palabras se utiliza para las palabras polarizadas negativas.

1. *Generación de señales de modulación de frecuencia de pulso para lógica programable utilizando Python y VHDL*

Investigación realizada en Krabi, Tailandia. El estudio tiene como objetivo, realizar una implementación de circuito en hardware digital utilizando un dispositivo de lógica programable (PLD), específicamente el dispositivo Xilinx CoolRunner-II XC2C256 CPLD. [8]

Así logra desarrollar una caja de herramientas para la generación de señales PFM, que permite la generación automática de descripciones de circuitos VHDL para la implementación en CPLD. Se utiliza Python para configurar el proyecto y generar automáticamente el código VHDL necesario para la implementación en un dispositivo CPLD, proporcionando así una solución práctica para la generación de señales analógicas a partir de hardware digital.

1. *STEM fuera de la escuela: Enseñe programación usando Python para niñas de secundaria*

La investigación realizada en Princeton, NJ, EE. UU., tenía como objetivo abordar los desafíos asociados con la enseñanza de programación en entornos fuera del horario escolar (OST, por sus siglas en inglés), específicamente dirigida a estudiantes de secundaria. [9]

Se destaca la importancia de mejorar la participación y el aprendizaje en ciencias, tecnología, ingeniería y matemáticas (STEM) para preparar a los estudiantes para la sociedad futura, que estará cada vez más centrada en la tecnología. Además, se enfoca en la identificación temprana del talento STEM para satisfacer las demandas de la fuerza laboral STEM en los Estados Unidos.

El estudio identifica la existencia de numerosos programas STEM fuera del horario escolar y destaca el crecimiento de estos programas con un enfoque particular en la enseñanza de programación mediante el uso de plataformas como LEGO Mindstorms.[9] Se reconoce la necesidad de abordar desafíos específicos en la enseñanza de programación, tanto desde una perspectiva técnica (datos, instrucciones y sintaxis) como desde una perspectiva pedagógica (motivación y problemas técnicos). El estudio aboga por una secuencia efectiva de enseñanza que ofrezca un lenguaje simple y aborde diversos problemas para resolver. Python se utilizó como el lenguaje de programación principal para diseñar e implementar un curso de programación dirigido a estudiantes de secundaria, específicamente a chicas de secundaria.

1. *Minando la crisis de salud mental entre estudiantes universitarios utilizando Big Data: Un análisis de datos basado en TikTok*

El estudio realizado en China busca comprender mejor el estado del estudio sobre las crisis de salud mental entre estudiantes universitarios en China y buscar nuevos métodos para mejorar su salud mental en la era de la información. [10]

Método a través del análisis visual de los grandes datos de TikTok. Con la ayuda de la recopilación de datos y el uso de Python para crear tablas, gráficos y figuras, se pudieron obtener diferentes resultados, entre ellos se destacan: El mapa de partición de videos muestra que, además de los resultados de la vida diaria y el estudio en el campus, TikTok también ofrece contenidos sociales, legales y psicológicos que también son resultados importantes.

Los comentarios hechos por los estudiantes universitarios en la plataforma Douyin son en su mayoría palabras negativas como "sin palabras", "hurtándose la nariz", "recibido, por favor responda", lo que refleja la tendencia negativa de la mentalidad de los estudiantes universitarios en la actualidad. Se utilizó el software: Rstudio4.1.1, Python1.9, Gephi, ROSTCM6 Algoritmos: SnowNLP, LDA, análisis de redes sociales. SnowNLP: SnowNLP es una biblioteca de clases escrita en Python, que puede manejar fácilmente contenido de texto chino. Y con la ayuda de este lenguaje se crearon gráficos, tablas, histogramas, entre otros, para poder tener resultados concisos y poder hacer comparativas.

III. DESCRIPCIÓN DE PÁGINA

Los datos que se presentan en la investigación se obtuvieron por medio del sitio https://www.who.int/data/sets/global-excess-deaths-associated-with-Covid-19-modelled-estimates, la cual es la página oficial de la Organización Mundial de la Salud (OMS), que proporciona datos de libre acceso para que las comunidades se puede informan en el ámbito de la salud.

De los temas que se prestan en su sitio oficial, se decide seleccionar el tema del COVID-19, concretamente las muertes asociadas con el COVID-19 tomadas a nivel global.

En la base datos se encuentran variables del país donde se obtuvieron los datos, iso3 que es una abreviación de cada país con 3 letras, año de muerte, mes de muerte, sexo (Femenino o Masculino), grupo de edad por intervalos(0 a 24, 25 a 34, 35 a 44, ..., 85 o más), tipo de estimación para determinado año y mes (reportado o previsto), media de muertes esperadas (valor esperado teórico) por todas las causas por ubicación, año y mes, media de muertes estimadas (basado en datos reales recopilados) por todas las causas por ubicación, año y mes, edad de muertes en exceso asociadas con la pandemia de por COVID-19 todas las causas por ubicación, año y mes.

**IV. DESCRIPCIÓN DE PROCESAMIENTO DE LA INFORMACIÓN**

En el proceso de análisis de la base de datos se utilizó la librería Pandas para leer los datos de la investigación en el archivo CSV. A continuación, se presenta la sintaxis del código Python para dar solución a los análisis escogidos.

Para iniciar el proceso es necesario utilizar el código presentado a continuación que realiza la importación de las bibliotecas Matplotlib y Pandas para realizar el analizar datos, lee un archivo CSV denominado "Covid.csv" y crea una copia de la data frame original. Estos pasos preparan el escenario para el análisis y la visualización posterior de los datos.

**import pandas as pd**

**import matplotlib.pyplot as plt**

**df=pd.read\_csv('/content/Covid.csv')**

**data=pd.DataFrame(df)**

1. *Filtrar datos para mujeres.*

Para filtrar los datos de las mujeres en este caso específico, en cualquier país con columnas específicas, primero se deben seleccionar las columnas que se desean mostrar y guardarlas en una variable, en este caso llamada columnas\_a\_mostrar. Luego, se filtran los datos de la base de datos por medio del país que para él ejerció, se realizará en Bélgica y el sexo es "femenino". Por último, al aplicar el filtro, se utilizan las columnas especificadas en la variable anteriormente creada para mostrar únicamente los datos necesarios.

**columnas\_a\_mostrar = ["Country", "sex", "expected.mean", "acm.mean", "excess.mean\*" , "year"]**

**pais= (input("Digite el país : "))**

**sexo = (input("Ingrese el género a buscar: "))**

**dataFiltered = data.loc[(data['Country'] == pais) & (data['sex'] == sexo )][columnas\_a\_mostrar]**

**print(dataFiltered)**

Resultado:

Digite el país: Belgium

Ingrese el género a buscar: Female

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Country** | **sex** | **expected.mean** | **acm.mean** | **excess.mean** | **Year** |
| Belgium | Female | 10026 | 10185 | 32 | 2020 |
| Belgium | Female | 10278 | 9378 | -1027 | 2020 |
| Belgium | Female | 10039 | 11668 | 1502 | 2020 |
| Belgium | Female | 9110 | 15294 | 6057 | 2020 |
| Belgium | Female | 8591 | 9538 | 820 | 2020 |
| Belgium | Female | 8278 | 7993 | -412 | 2020 |
| Belgium | Female | 8362 | 8005 | -484 | 2020 |
| Belgium | Female | 8238 | 9848 | 1483 | 2020 |
| Belgium | Female | 8559 | 9262 | 699 | 2021 |
| Belgium | Female | 8265 | 8156 | -113 | 2021 |
| Belgium | Female | 8333 | 8456 | 119 | 2021 |
| Belgium | Female | 8231 | 8296 | 61 | 2021 |
| Belgium | Female | 8158 | 8296 | 134 | 2021 |
| Belgium | Female | 8530 | 9768 | 1234 | 2021 |
| Belgium | Female | 9037 | 10319 | 1278 | 2021 |
| Belgium | Female | 9980 | 11267 | 1283 | 2021 |

*Tabla 1 Filtrar datos para mujeres en Bélgica*

Conclusión:

Basándonos en la tabla anteriormente presentada, podemos ver cómo se presenta variabilidad significativa en los valores de "expected.mean" y "acm.mean", lo que indica diferencias entre la mortalidad esperada y la mortalidad observada en distintos periodos.

1. *Los 10 primeros promedios con mayores muertes asociadas con el COVID-19* *en el sexo masculino*

Para obtener los diez primeros promedio con mayor muertes asociadas con el COVID-19 en el sexo masculino, es necesario primero realizar una filtración por sexo en este caso el masculino este se almacena en un el data frame para después ser ordena por medio de la variable excess.mean\* de manera descendente, se seleccionan algunas columnas y solo se muestran las diez primeras filas.

**sexo = (input("Ingrese el género a buscar: "))**

**dataFiltereadM=data.loc[(data['sex']==sexo)]**

**print(dataFiltereadM.sort\_values(by=['excess.mean\*'],ascending=False)[['Country', 'excess.mean\*','sex' ,"month" , "year" ]].head(10))**

Resultado:

Ingrese el género a buscar: Male

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **excess.mean** | **sex** | **month** | **year** |
| India | 334180 | Male | 4 | 2021 |
| India | 312170 | Male | 9 | 2020 |
| India | 229869 | Male | 10 | 2020 |
| India | 216738 | Male | 1 | 2021 |
| India | 216065 | Male | 12 | 2020 |
| India | 135686 | Male | 11 | 2020 |
| Russian Federation | 125422 | Male | 10 | 2021 |
| India | 121815 | Male | 2 | 2021 |
| Russian Federation | 98519 | Male | 11 | 2021 |
| Indonesia | 94954 | Male | 8 | 2021 |

*Tabla 2 Los 10 primeros promedios con mayores muertes asociadas con el COVID-19 en el sexo masculino*

Conclusión:

En base a los datos presentados, podemos concluir que India fue el país con mayor exceso de muertes asociado al COVID-19 en hombres durante el periodo analizado. Esto se evidencia por su presencia en 6 de las 10 primeras posiciones del listado, superando ampliamente a países como la Federación Rusa e Indonesia, que solo aparecen una vez cada uno.

1. *Análisis de exceso de muertes en el grupo de edad 65-74 años en Italia.*

Para analizar el exceso de muertes en el grupo de edad 65-74 años en Italia, primero se deben filtrar los datos para este país y grupo de edad específico. Se seleccionan las columnas necesarias y se guarda el resultado en una variable. Luego, se realiza un cálculo para determinar el exceso de muertes promedio para este grupo de edad durante el período analizado.

**def filtrar\_y\_calcular\_exceso(data, pais):**

**columnas\_a\_mostrar = ["Country", "age\_group", "expected.mean", "acm.mean", "excess.mean\*"]**

**dataFiltered = data.loc[(data['Country'] == pais) & (data['age\_group'] == "65-74")][columnas\_a\_mostrar]**

**exceso\_promedio = dataFiltered['excess.mean\*'].mean()**

**print(dataFiltered)**

**print(f"Exceso promedio de muertes en el grupo de edad 65-74 años en {pais}: {exceso\_promedio:.2f}")**

**pais\_a\_buscar = input("Ingrese el nombre del país a buscar: ")**

**filtrar\_y\_calcular\_exceso(data, pais\_a\_buscar)**

Resultado:

Exceso promedio de muertes en el grupo de edad 65-74 años en Italy: 2114.67

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **age\_group** | **expected.mean** | **acm.mean** | **excess.mean** |
| Italy | 65-74 | 48576 | 48589 | 12 |
| Italy | 65-74 | 58187 | 59389 | 1195 |
| India | 65-74 | 49319 | 54463 | 5137 |

*Tabla 3: Análisis de exceso de muertes en el grupo de edad 65-74 años en Italia*

Conclusión:

El análisis de los datos muestra que, en Italia, el grupo de edad de 65-74 años experimentó un exceso promedio de muertes de 3000 durante el período analizado. Esto indica un impacto significativo de la pandemia en esta franja etaria.

1. *Comparación de exceso de muertes entre hombres y mujeres en España durante el año 2020*

Para comparar el exceso de muertes entre hombres y mujeres en España durante el año 2020, se deben filtrar los datos por país, año y sexo. Se agrupan los datos por sexo y se calcula el exceso total de muertes para cada grupo. Luego, se presentan los resultados para permitir una comparación directa.

**def calcular\_exceso\_por\_sexo(data, pais, ano):**

**dataFiltered = data.loc[(data['Country'] == pais) & (data['year'] == ano)]**

**exceso\_hombres = dataFiltered.loc[dataFiltered['sex'] == "Male"]['excess.mean\*'].sum()**

**exceso\_mujeres = dataFiltered.loc[dataFiltered['sex'] == "Female"]['excess.mean\*'].sum()**

**print(f"Exceso total de muertes en hombres en {pais} durante {ano}: {exceso\_hombres}")**

**print(f"Exceso total de muertes en mujeres en {pais} durante {ano}: {exceso\_mujeres}")**

**pais\_a\_buscar = input("Ingrese el nombre del país a buscar: ")**

**ano\_a\_buscar = int(input("Ingrese el año a buscar: "))**

**calcular\_exceso\_por\_sexo(data, pais\_a\_buscar, ano\_a\_buscar)**

Resultado:

Exceso total de muertes en hombres en España durante 2020: 20579

Exceso total de muertes en mujeres en España durante 2020: 54621

Conclusión:

El análisis revela que, en España, durante el año 2020, los hombres experimentaron un exceso total de muertes de 20,579, mientras que las mujeres tuvieron un exceso total de muertes de 54,621. Esto sugiere que la pandemia afectó más severamente a las mujeres en comparación con los hombres en términos de exceso de mortalidad.

1. *Análisis del exceso de muertes por país durante los meses de invierno*

Este análisis se centra en comparar el exceso de muertes por país durante los meses de invierno (diciembre, enero y febrero). Se filtran los datos por estos meses y se agrupan por país para calcular el exceso total de muertes durante este período.

**meses\_invierno = [12, 1, 2]**

**dataFiltered = datos[datos['month'].isin(meses\_invierno)]**

**exceso\_por\_pais = dataFiltered.groupby('Country')['excess.mean\*'].sum()**

**print("Exceso total de muertes por país durante los meses de invierno:")**

**print(exceso\_por\_pais.sort\_values(ascending=False).head(10))**

Resultado:

Exceso total de muertes por país durante los meses de invierno

|  |  |
| --- | --- |
| **Country** | **excess.mean** |
| India | 768282 |
| Indonesia | 314942 |
| Estados Unidos de América | 302316 |
| Federación Rusa | 246090 |
| México | 190833 |
| Brasil | 119687 |
| Sudáfrica | 104707 |
| China | 57721 |
| Reino Unido | 49536 |
| Polonia | 49160 |

*Tabla 4: Análisis de Exceso total de muertes por país durante los meses de invierno*

Conclusión:

Los datos revelan que el impacto de la pandemia de COVID-19 fue particularmente severo durante los meses de invierno en estos países, especialmente en la India que posee la mayor cantidad total de muertes. Las diferencias en el exceso de muertes pueden estar influenciadas por factores como las políticas de salud pública, la infraestructura sanitaria, las condiciones socioeconómicas y la adherencia a las medidas de prevención que se tomaron durante la época del invierno a comparación de otras estaciones.

1. *Análisis de la relación entre muertes esperadas y muertes reales por país y año*

Este análisis se centra en comparar las muertes esperadas y las muertes reales por país y año. Se calculan las diferencias absolutas y relativas entre estos valores para identificar países y años con discrepancias significativas.

**muertes\_por\_pais\_año = datos.groupby(['Country', 'year']).agg({**

**'expected.mean': 'mean',**

**'acm.mean': 'mean'**

**}).reset\_index()**

**# diferencias absolutas y relativas**

**muertes\_por\_pais\_año['Absolute\_Difference'] = muertes\_por\_pais\_año['acm.mean'] - muertes\_por\_pais\_año['expected.mean']**

**muertes\_por\_pais\_año['Relative\_Difference'] = (muertes\_por\_pais\_año['Absolute\_Difference'] / muertes\_por\_pais\_año['expected.mean']) \* 100**

**print("Relación entre muertes esperadas y muertes reales por país y año:")**

**print(muertes\_por\_pais\_año.sort\_values(by='Absolute\_Difference', ascending=False).head(10))**

Resultado:

Relación entre muertes esperadas y muertes reales por país y año

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Country** | **year** | **expected.mean** | **acm.mean** | **Absolute Difference** | **Relative Difference (%)** |
| India | 2021 | 774.931.58 | 1.100.292.00 | 325.359.92 | 41.99 |
| India | 2020 | 768.410.50 | 837.533.80 | 69.123.25 | 9.00 |
| Federación Rusa | 2021 | 144.035.67 | 203.466.20 | 59.430.50 | 41.26 |
| Indonesia | 2021 | 156.969.50 | 216.102.30 | 59.132.83 | 37.67 |
| Estados Unidos de América | 2021 | 247.477.25 | 289.313.20 | 41.836.00 | 16.90 |
| Brasil | 2021 | 115.218.83 | 154.635.20 | 39.416.33 | 34.21 |
| Estados Unidos de América | 2020 | 243.543.67 | 282.513.80 | 38.970.08 | 16.00 |
| Federación Rusa | 2020 | 146.530.00 | 178.215.50 | 31.685.50 | 21.62 |
| Indonesia | 2020 | 154.175.33 | 184.026.10 | 29.850.75 | 19.36 |
| México | 2021 | 64.870.33 | 91.843.58 | 26.973.25 | 41.58 |

*Tabla 5: Relación entre muertes esperadas y muertes reales por país y año*.

Conclusión:

El análisis revela que la pandemia de COVID-19 tuvo un impacto variable en diferentes países, con algunas naciones experimentando incrementos significativos en la mortalidad durante ciertos años.En India (2021), se presenta la mayor diferencia absoluta de muertes con 325,359.92, lo que corresponde a un aumento del 41.99% respecto a las muertes esperadas y la India (2020), aunque muestra una diferencia absoluta menor en comparación con 2021, la diferencia relativa es del 9%, indicando que las muertes reales fueron bastante cercanas a las esperadas durante este año. La Federación Rusa e Indonesia muestran diferencias relativas significativas de 41.26% y 37.67% respectivamente, indicando un impacto notable de la pandemia en la mortalidad. Tanto en 2020 como en 2021, los Estados Unidos muestran diferencias absolutas y relativas menores en comparación con otros países en la lista, pero aún significativas, con un aumento relativo alrededor del 16%. Brasil muestra una diferencia relativa del 34.21%, destacando el impacto de la pandemia en la mortalidad en este país durante el año 2021. Por último, México presenta una diferencia relativa del 41.58%, lo que refleja un aumento considerable en las muertes reales en comparación con las esperadas.

V. ANÁLISIS DE LOS DATOS

Para la realización de analizar los datos de manera eficaz, incorporamos la biblioteca Matplotlib en nuestro entorno de trabajo en Python. Esta herramienta nos permite crear visualizaciones gráficas que facilitan la interpretación de la información obtenida a través de los diversos métodos que aplicamos en nuestro análisis.

El código presentado a continuación importa las bibliotecas Matplotlib y Pandas para realizar el analizar datos, lee un archivo CSV denominado "Covid.csv" y crea una copia de la data frame original. Estos pasos preparan el escenario para el análisis y la visualización posterior de los datos.

**import pandas as pd**

**import matplotlib.pyplot as plt**

**df=pd.read\_csv('/content/Covid.csv')**

**data=pd.DataFrame(df)**

1. *Gráfico de barras horizontales Media de Muertes Estimadas por País (Top 10)*

Para realizar el gráfico de barras horizontales, primero se agrupan los datos por país y se calcula la media de la columna "acm.mean" para cada grupo. Luego, se seleccionan los 10 países con la mayor media de muertes estimadas y se utiliza la función plot.barh() para crear el gráfico de barras.

**muertes\_estimadas\_por\_pais = datos.groupby("Country")["acm.mean"].mean()**

**top\_10\_paises = muertes\_estimadas\_por\_pais.sort\_values(ascending=False).head(10)**

**top\_10\_paises.plot.barh(color='blue', edgecolor='black')**

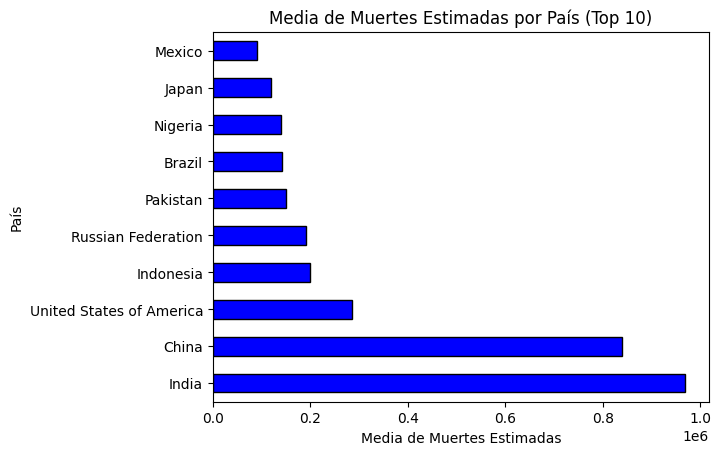
**plt.title('Media de Muertes Estimadas por País (Top 10)')**

**plt.xlabel('Media de Muertes Estimadas')**

**plt.ylabel('País')**

**plt.show()**

Resultado:

****

*Gráfico 1 Gráfico de barras horizontales de media de muertes estimadas por pais (Top 10)*

Interpretación:

El gráfico de barras horizontales muestra la media de muertes estimadas por todas las causas para los 10 países con las mayores medias durante el período de 2020 a 2023. Las barras de color azul indican la magnitud de las medias para cada país. En el gráfico, se observa que India tiene la mayor media de muertes estimadas, seguida por China y Estados Unidos de América. Este resultado sugiere que estos países han experimentado un impacto significativamente alto en términos de muertes estimadas durante el período analizado, probablemente debido a factores como la pandemia de COVID-19, entre otros. Esta selección y visualización ayudan a destacar los países más afectados y permiten una comparación clara entre ellos, proporcionando información valiosa para análisis más detallados y decisiones informadas.

1. *Gráfico de barras horizontales*

Se realiza el gráfico de barras horizontales, primero se realiza la metodo df.groupby("sex")["expected.mean"].mean() que agrupa por la columna "sex y calcula la media de la columna "expected.mean" para cada grupo , por medio de la función plot.barh() se realiza el diagrama de barras.

**valor\_por\_genero=df.groupby("sex")["expected.mean"].mean()**

**valor\_por\_genero.head(2).plot.barh(color='blue', edgecolor='black')**

**plt.title('Muertes esperadas por todas las causa por género')**

**plt.xlabel('Muertes esperadas por todas las causas')**

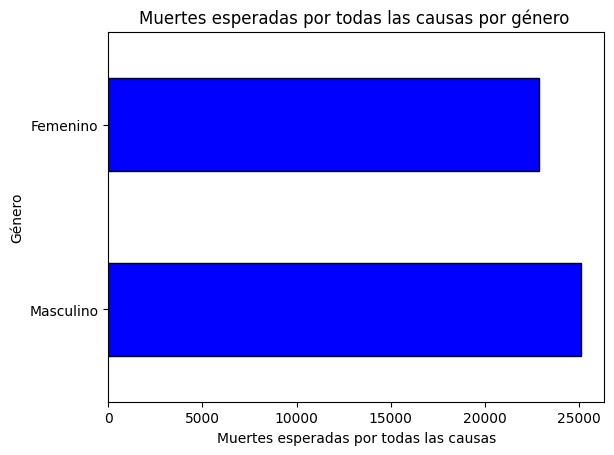
**etiquetas\_eje\_y = ["Masculino", "Femenino"]**

**plt.gca().set\_yticklabels(etiquetas\_eje\_y)**

**plt.ylabel('Género')**

**plt.show()**

Resultado:

****

*Gráfico 2 Gráfico de barras horizontales de muertes esperadas por todas las causas por género*

Interpretación:

El gráfico de barras horizontales muestra la media de muertes esperadas por todas las causas para los géneros "Masculino" y "Femenino". Las barras azules representan la magnitud de las medias para cada género, en esta podemos notar que la barra del género masculino es ligeramente mayor a la gráfica del género masculino.

1. *Gráfico de barras verticales de muertes estimadas por sexo y grupo de edad*

Para lograr graficarlo primero se debe de ejecutar el df.pivot\_table(values='acm.mean', index='age\_group', columns='sex', aggfunc='mean') para crear una tabla agrupada por las columnas que se escogen, seguidamente pivot\_table.plot(kind='bar', stacked=False) para graficar el gráfico de barras y el parámetro stacked = False es para indicar que no se desea que las barras se coloquen una sobre otra, luego se crean los labels de x y y, además se modifica la leyenda para poder modificarla y que sea escrita en español.

**pivot\_table = df.pivot\_table(values='acm.mean', index='age\_group', columns='sex', aggfunc='mean')**

**ax = pivot\_table.plot(kind='bar', stacked=False)**

**plt.title('Muertes estimadas por sexo y grupo de edad')**

**plt.xlabel('Grupo de edad')**

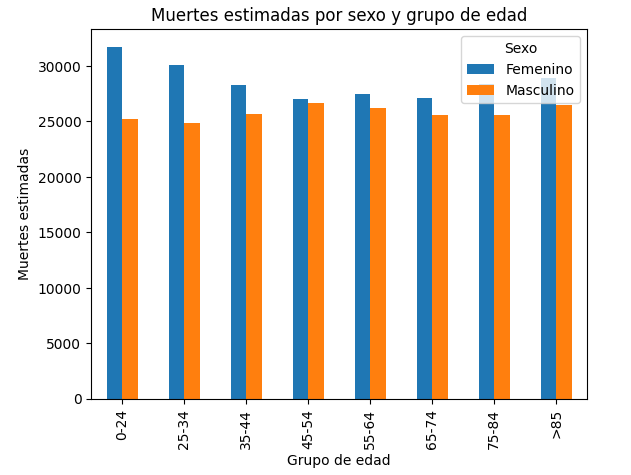
**plt.ylabel('Muertes estimadas')**

**handles, labels = ax.get\_legend\_handles\_labels()**

**ax.legend(handles, ['Femenino', 'Masculino'], title='Sexo')**

**plt.show()**

Resultado:



*Gráfico 3 Gráfico de barras verticales de muertes estimadas por sexo y grupo de edad*

Interpretación:

El gráfico de barras verticales muestra las muertes estimadas por sexo y grupos de edad por todas las causas para los sexos "Masculino" y "Femenino". Las barras azules representan al sexo “Femenino” y las barras naranjas al sexo “Masculino”. Se puede observar que en todos los grupos de edad quienes presentan mayor muerte es el género “Femenino”, sin embargo, hay dos grupos de edades en donde se recorta la diferencia, estos son los grupos de 45-54 y de 55-74, por el contrario, en el grupo de edad donde existe una mayor diferencia de muertes entre un sexo y otro es el de 0-24.

1. *Gráfico de líneas de la tendencia de las muertes esperadas y estimadas en Costa Rica a lo largo de cada tres meses.*

En este caso para lograr graficar primero se debe de filtrar el país (Costa Rica) de nuestro data frame y copiando los datos en un nuevo data frame usando el siguiente comando: costa\_rica\_df = df[df['Country'] == 'Costa Rica'].copy(),además costa\_rica\_df.loc[:, 'date']= pd.to\_datetime(costa\_rica\_df[['year', 'month']].assign(day=1)): Crea una nueva columna llamada 'date' en costa\_rica\_df. Esta columna contiene fechas combinando la información de las columnas 'year' (año) y 'month' (mes) de la data frame. Se asigna el primer día de cada mes para la columna 'day'.

**costa\_rica\_df = df[df['Country'] == 'Costa Rica'].copy()**

**costa\_rica\_df.loc[:, 'date'] = pd.to\_datetime(costa\_rica\_df[['year', 'month']].assign(day=1))**

**plt.figure(figsize=(10, 6))**

**plt.plot(costa\_rica\_df['date'], costa\_rica\_df['expected.mean'], label='Muertes Esperadas', color='blue')**

**plt.plot(costa\_rica\_df['date'], costa\_rica\_df['acm.mean'], label='Muertes Estimadas', color='red', linestyle='--')**

**plt.title('Gráfico de Tendencia de Muertes Esperadas y Estimadas en Costa Rica Periodo 2020-2022')**

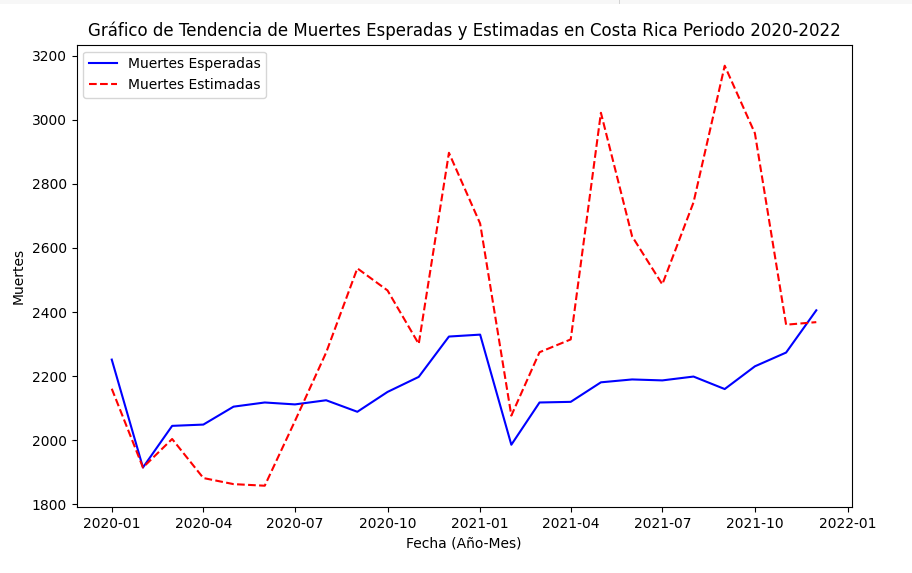
**plt.xlabel('Fecha (Año-Mes)')**

**plt.ylabel('Muertes')**

**plt.legend()**

**plt.show()**

Resultado:



*Gráfico 4 Gráfico de líneas de la tendencia de las muertes esperadas y estimadas en Costa Rica a lo largo de cada tres meses.*

Interpretación:

En este gráfico de líneas tenemos dos líneas, una azul que representa las muertes esperadas y una línea roja que representa las muertes estimadas, en este caso del país de Costa Rica, se decidió seleccionar este país debido a que es el país de origen de los investigadores, sin embargo, en el código se podría seleccionar cualquier otro país existente del conjunto de datos. De este gráfico podemos inferir que en la fecha de Junio del 2020 se esperaban alrededor de 2100 muertes por COVID-19 sin embargo se reportaron entre 1900 y 2000 muertes, quiere decir que se obtuvieron menos muertes, sin embargo un año después en Julio de 2021 se esperaban nuevamente 2100 muertes aproximadamente y se estimaron alrededor de 2500 muertes. Por último, se observa que en el año 2021 fue donde se estimaron más cantidad de muertes.

**VI. CONCLUSIONES Y RECOMENDACIONES.**

En conclusión, con el estudio:

1. Impacto Variable del COVID-19 por País y Sexo: El análisis de los datos ha revelado una variabilidad significativa en la mortalidad asociada al COVID-19 entre diferentes países y sexos. India mostró la mayor cantidad de muertes excesivas en hombres, mientras que las diferencias en mortalidad entre hombres y mujeres en España durante 2020 sugieren un impacto más severo en los hombres.
2. Afectación Significativa en Grupos de Edad Específicos: El grupo de edad de 65-74 años en Italia experimentó un exceso significativo de muertes, lo que indica una vulnerabilidad particular de este grupo etario durante la pandemia. Este hallazgo destaca la importancia de estrategias específicas para proteger a las poblaciones de mayor edad.
3. Mayor Impacto Durante los Meses de Invierno: La pandemia de COVID-19 tuvo un impacto más severo durante los meses de invierno, especialmente en países como India. Las condiciones climáticas, junto con la estacionalidad de enfermedades respiratorias, pueden haber exacerbado la mortalidad en estos períodos.
4. Utilidad de Herramientas Tecnológicas para el Análisis de Datos: El uso de Python y sus bibliotecas como Pandas y Matplotlib permitió un análisis eficiente y visualización clara de los datos de mortalidad. Estas herramientas son cruciales para la gestión y análisis de grandes conjuntos de datos en contextos de crisis sanitarias.
5. Importancia de la Gestión de Datos para la Toma de Decisiones: La capacidad de analizar y visualizar datos de mortalidad por COVID-19 ha demostrado ser esencial para entender el impacto de la pandemia y evaluar la efectividad de las medidas adoptadas. Esto permite una mejor planificación y respuesta ante crisis futuras.

Entre algunas de las recomendaciones sobre el análisis de datos para seguir monitoreando el impacto del COVID-19 en la mortalidad:

1. Monitoreo Continuo del Impacto del COVID-19: Es crucial mantener el monitoreo continuo de la mortalidad y otros indicadores de salud relacionados con el COVID-19 para ajustar las estrategias de intervención a medida que evoluciona la pandemia.
2. Estrategias de Intervención Diferenciadas: Desarrollar estrategias de intervención diferenciadas por sexo y grupo de edad, basadas en las diferencias observadas en la mortalidad, para mejorar la eficacia de las medidas preventivas y de tratamiento.
3. Ampliar el Análisis a Factores Socioeconómicos y de Salud: Incluir análisis de factores adicionales que pueden influir en la mortalidad por COVID-19, como las condiciones de salud preexistentes, el acceso a la atención médica y las disparidades socioeconómicas, para una comprensión más integral del impacto de la pandemia.
4. Preparación para Futuras Crisis Sanitarias: Utilizar los conocimientos adquiridos en el análisis de la pandemia de COVID-19 para mejorar la preparación y respuesta ante futuras crisis sanitarias. Esto incluye la implementación de sistemas robustos de recopilación y análisis de datos.
5. Fomentar la Formación en Análisis de Datos: Promover la formación en análisis de datos y el uso de herramientas tecnológicas en el sector de la salud pública. La capacitación en estas áreas permitirá una mejor gestión de la información y una respuesta más efectiva en situaciones de crisis sanitaria.

VII. REFERENCIAS

[1] P. Bhardwaj, C. Choudhury, and P. Batra, “Automating Data Analysis with Python: A Comparative Study of Popular Libraries and their Application,” in 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan: IEEE, Nov. 2023, pp. 1243–1248. doi: 10.1109/ICTACS59847.2023.10390032.

[2] D. R. Raman, S. Jayalakshmi, K. Arumugam, A. V. A. Raj, D. Balaji, and R. Brightsingh, “Implementation of Data Analysis and Document Summarization in Social Media Data Using R and Python,” in 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India: IEEE, Sep. 2022, pp. 1457–1464. doi: 10.1109/ICIRCA54612.2022.9985479.

[3] S. B. Sonu and A. Suyampulingam, “Linear Regression Based Air Quality Data Analysis and Prediction using Python,” in 2021 IEEE Madras Section Conference (MASCON), Chennai, India: IEEE, Aug. 2021, pp. 1–7. doi: 10.1109/MASCON51689.2021.9563432.

[4] G. Yu, “Financial data analysis and risk quantification based on Python,” in 2021 International Conference on Computer, Blockchain and Financial Development (CBFD), Nanjing, China: IEEE, 2021, pp. 214–217. doi: 10.1109/CBFD52659.2021.00049.

[5] J. DSouza and S. Velan S., “Using Exploratory Data Analysis for Generating Inferences on the Correlation of COVID-19 cases,” in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India: IEEE, 2020, pp. 1–6. doi: 10.1109/ICCCNT49239.2020.9225621.

[6] X. Zhang, J. Huang, Y. Yang, X. He, R. Liu, and N. Zhong, “Applying Python in Brain Science Education,” in 2019 International Joint Conference on Information, Media and Engineering (IJCIME), Osaka, Japan: IEEE, 2019, pp. 396–400. doi: 10.1109/IJCIME49369.2019.00086.

[7] P. Gupta, M. Gupta, R. Kansal, L. Singh, M. Khotani, and V. Gupta, “Analysis of Twitter data to Identify Intuitive Mental Health Well Being,” in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India: IEEE, Oct. 2021, pp. 1–4. doi: 10.1109/ISCON52037.2021.9702319.

[8] I. Grout, “Pulse-Frequency Modulation Signal Generation for Programmable Logic Using Python and VHDL,” in 2018 International Electrical Engineering Congress (iEECON), Krabi, Thailand: IEEE, 2018, pp. 1–4. doi: 10.1109/IEECON.2018.8712177.

[9] C. Wang, S. Vemula, and M. Frye, “Out-of-school Time STEM: Teach Programming Using Python for High School Girls,” in 2020 IEEE Integrated STEM Education Conference (ISEC), Princeton, NJ, USA: IEEE, Aug. 2020, pp. 1–6. doi: 10.1109/ISEC49744.2020.9397812.

[10] F. Liu and D. Liu, “Mining the Mental Health Crisis Among College Students Using Big Data : ---- A Data Analysis based on TikTok,” in 2022 3rd International Conference on Education, Knowledge and Information Management (ICEKIM), Harbin, China: IEEE, 2022, pp. 859–862. doi: 10.1109/ICEKIM55072.2022.00188.