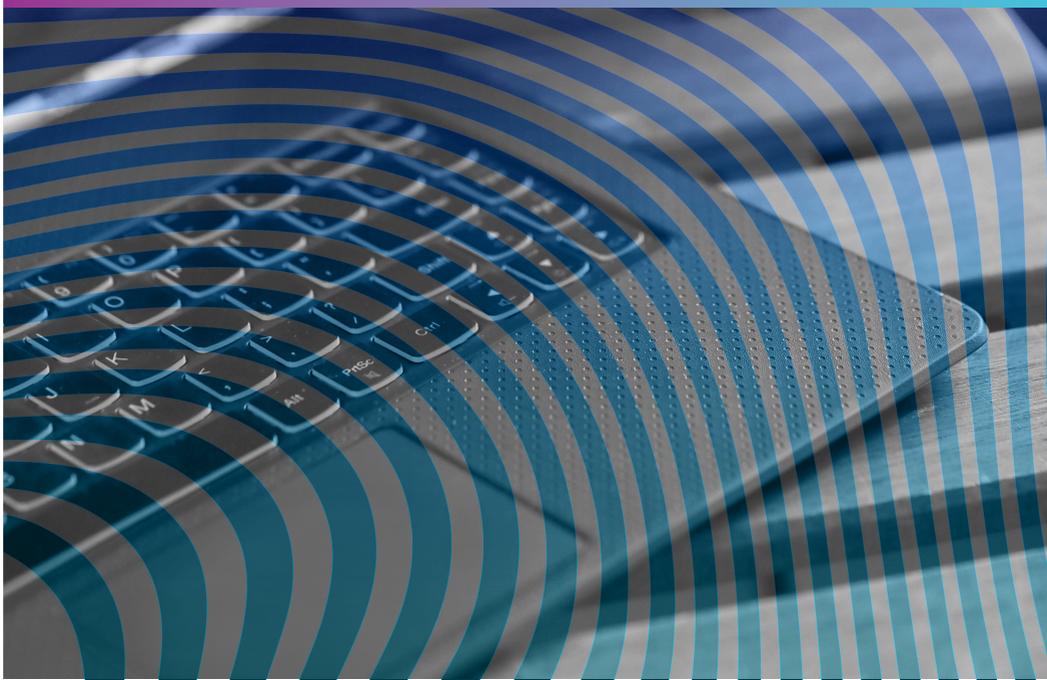


1^{ra} Edición

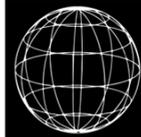


Visualización de Datos Educativos y Sociales con Python

Herramientas para Decisiones Informadas



Sci ELa



AUTORES:

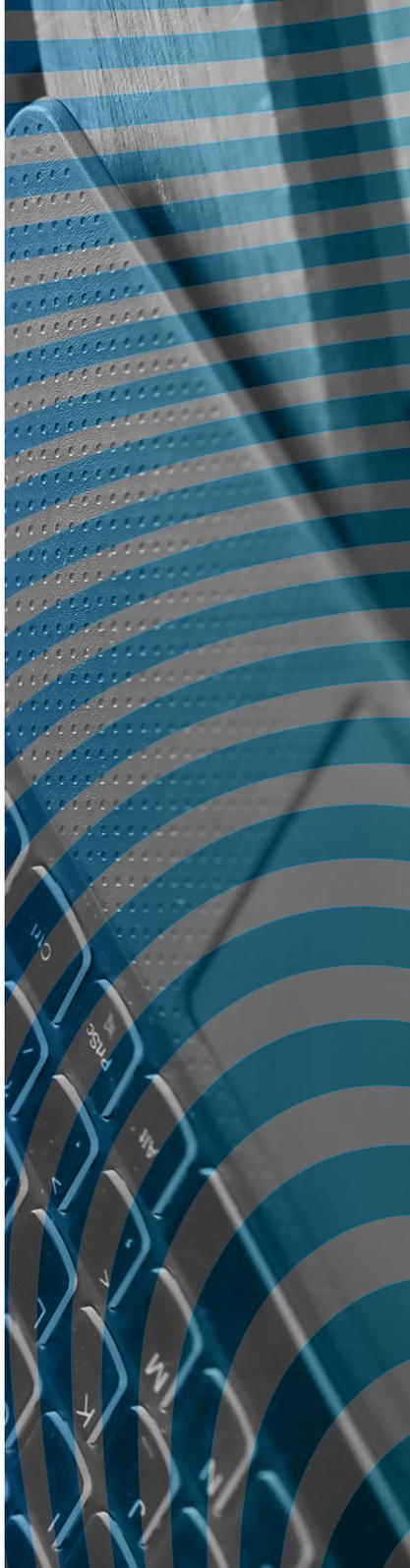
- Toalombo Montero, Marcelo Patricio
- Toalombo Montero, Oscar William
- Ballestero Torres, Francisco Andrés
- Hernández Dávila, Carlos Alfredo
- Ruiz Sarzosa, John Paul





EDUCATIONAL AND SOCIAL DATA VISUALIZATION WITH PYTHON

Tools for Informed Decisions





PRIMERA EDICIÓN, FEBRERO 2024

Visualización de Datos Educativos y Sociales con Python: Herramientas para Decisiones Informadas

ISBN: 978-9942-7173-6-8

DOI: <https://doi.org/10.62131/978-9942-7173-6-8>

Editado por:

Sello editorial:

© Editorial Investigativa
Latinoamericana (SciELa)

Quevedo, Los Rios, Ecuador

E-mail: admin@editorial-sciela.org

Código Postal: 120303

WEB: <https://editorial-sciela.org>

Este libro se sometió a arbitraje bajo el sistema de doble ciego (peer review) y antiplágio. Este producto investigativo cumple con la Declaración de Principios de Budapest, San Francisco, México, Helsinki y Firma del Marco del MIT

Dirección editorial:

Lic. Alexander Fernando Haro, MSI.

Revisor (1):

Ing. María Espinoza Bravo, Mg.

Revisor (2):

Ing. Roberto Cabezas Cabezas, Mg.

Revisor (3):

Ing. Anderson Chilingua García, Mg.

Sistema de clasificación decimal DEWEY

005 - Programación. Programas. Datos de computadores

Clasificación comercial internacional - THEMA

U - Computación y tecnologías de la información

UY - Informática

UYZ - Interacción persona - ordenador

UYZF - Visualización de la Información

Reservados todos los derechos. Está prohibido, bajo las sanciones penales y el resarcimiento civil previstos en las leyes, reproducir, registrar o transmitir esta publicación, íntegra o parcialmente, por cualquier sistema de recuperación y por cualquier medio, sea mecánico, electrónico, magnético, electroóptico, por fotocopia o por cualquiera otro, sin la autorización previa por escrito a la Editorial Investigativa Latinoamericana (SciELa).



| AUTORES |

Marcelo Patricio Toalombo Montero

Universidad Técnica de Cotopaxi



<https://orcid.org/0000-0002-4121-4481>



marcelo.toalombo2639@utc.edu.ec

Magister en Gerencia Informática con Mención en Desarrollo de Software y Redes, e Ingeniero en Sistemas; su experiencia laboral como Primer Administrador General del Área de Tecnologías de la Información del H. Gobierno Provincial de Tungurahua y como Jefe de Tecnologías de la Información y Comunicación de la Empresa Municipal de Agua Potable y Alcantarillado de la Ciudad de Ambato; en experiencia académica como docente de pregrado de la Facultad de Ciencias Administrativas de la Universidad Técnica de Ambato, y docente de pregrado en la Facultad de Ciencias de la Escuela Superior Politécnica de Chimborazo en Riobamba; actualmente es docente de pregrado de la Facultad de Ciencias Agropecuarias y Recursos Naturales de la Universidad Técnica de Cotopaxi, y cursa el Philosophie Doctor en School of Electronic Information and Communications de Huazhong University of Science and Technology de Wuhan-China.

Oscar William Toalombo Montero

Universidad Central del Ecuador



<https://orcid.org/0009-0008-8341-5338>



owtoalombo@uce.edu.ec

Máster Investigación (M2 Matemáticas fundamentales y aplicadas) en la Universidad Paris Sud – Paris XI, Paris – Francia. Ingénieur en Sciences de l’Ecole Polytechnique + Máster en Ciencias (M1) en l’Ecole Polytechnique Paris – Francia. Docente tiempo completo en Facultad de Ingeniería y Ciencias Aplicadas de la Universidad Central del Ecuador.

Francisco Andrés Ballesteros Torres

Universidad Técnica Luis Vargas Torres de Esmeraldas



<https://orcid.org/0000-0002-1862-380X>



andres.ballesteros@utelvt.edu.ec

Ingeniero Agrícola por la Universidad Agraria del Ecuador, Magister en Estadística mención Calidad y Productividad por la ESPOL, actual estudiante del Doctorado en Ciencias Agrarias de la Universidad Nacional de Piura. Docente Investigador de la Facultad de Ciencias Agropecuarias de la Universidad Técnica Luis Vargas Torres de Esmeraldas, en las carreras de Ingeniería Forestal, Agronómica y Zootecnia impartiendo las asignaturas de Bioestadística y Diseño Experimental. Experiencia en Docencia en la Pontificia Universidad Católica sede Esmeraldas en la carrera de Agroindustria. He participado como revisor par externo en diferentes revistas científicas regionales y revisor de Trabajos de Investigación de pregrado y posgrado. Miembro de comités científicos en congresos internacionales. Experiencia en empresas transnacionales como Nestlé y Tesalia CBC en el área de Calidad y Producción.

Carlos Alfredo Hernández Dávila

Universidad Técnica de Ambato



<https://orcid.org/0000-0002-2526-5051>



ca.hernandez@uta.edu.ec

Máster Universitario en Didáctica de las Matemáticas y Lengua en Educación infantil y primaria, como parte de la formación de cuarto nivel están las dos maestrías en la Universidad Internacional de La Rioja, Licenciado en ciencias de la Educación mención Educación Básica en la Universidad Técnica de Ambato y Tecnólogo en Informática en el Instituto Tecnológico Superior Bolívar, su experiencia profesional como técnico en computadoras, asistente administrativo, posteriormente desarrolle experiencia docente en algunas instituciones particulares, profesor de grado y posgrado en la Facultad de Ciencias Humanas y de la Educación de la Universidad Técnica de Ambato, además se ha colaborado en la redacción de artículos científicos indexados y participado en varias capacitaciones en temáticas a fines a las áreas de interés.

John Paul Ruiz Sarzosa

Universidad Técnica de Ambato



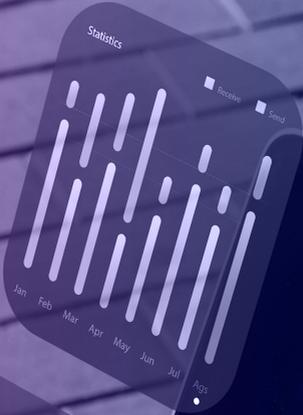
<https://orcid.org/0009-0005-8024-376X>



john.ruiz@educacion.gob.ec

Licenciado en Informática e Ingeniero en Sistemas, docente en la Unidad Educativa Fiscomisional PCEI de Pichincha en el área de matemáticas, en un centro de privación de libertad, con Maestría en Educación mención Enseñanza de la Matemática de la Universidad Técnica de Ambato.





12:30

Welcome
Your task
this week

75%

Task 01

Task 02



ÍNDICE

PREFACIO

DATOS EN EDUCACIÓN Y CIENCIAS SOCIALES

Objetivos y Alcance del Libro	16
-------------------------------------	----

CAPÍTULO I.

INTRODUCCIÓN A LA VISUALIZACIÓN DE DATOS

Definición y Fundamentos de la Visualización de Datos.....	19
Breve Historia de la Visualización de Datos.....	21
La Visualización de Datos en el Contexto Actual	23
Importancia de la Visualización de Datos en el Ámbito Educativo y Social.....	25
Principales Herramientas y Tendencias en Visualización de Datos	27

CAPÍTULO II.

FUNDAMENTOS DE LA VISUALIZACIÓN DE DATOS CON PYTHON

Principios de Diseño en Visualización de Datos	31
Gráficos Estáticos vs. Gráficos Interactivos.....	33
Creación de Gráficos Básicos con Matplotlib.....	35
Introducción a Seaborn para Gráficos Estadísticos.....	37
Visualizaciones Interactivas con Bokeh.....	42

Visualizaciones Interactivas con Plotly.....	47
----------------------------------------------	----

CAPÍTULO III.

DATOS EDUCATIVOS Y SOCIALES

Características de los Datos Educativos y Sociales.....	51
Fuentes Confiables de Datos Educativos	54
Fuentes Confiables de Datos Sociales.....	56
Procesos de Preprocesamiento y Limpieza de Datos.....	59

CAPÍTULO IV.

ANÁLISIS DESCRIPTIVO Y CORRELACIONAL DE DATOS

Técnicas y Herramientas en Contextos Educativos y Sociales...	63
.....	63
Análisis descriptivo.....	63
Matriz de Correlación.....	67
Análisis de Valores Atípicos	70
Gráficos de Tendencia	72
Análisis de Variables Categóricas:	75
Diagramas de Pareto	79

CAPÍTULO V.

VISUALIZACIÓN AVANZADA Y ANÁLISIS EXPLORATORIO DE DATOS EDUCATIVOS

Diseño de Dashboards Educativos.....	83
Análisis de Componentes Principales (PCA):	87
Análisis de Clúster (Clustering).....	91
Mapas de Calor (Heatmaps)	99

CAPÍTULO VI. VISUALIZACIÓN DE DATOS SOCIALES

Análisis de Datos en Medios Sociales.....	103
Visualización de Patrones Demográficos y Sociales.....	105
Herramientas de Visualización Geoespacial.....	110

CAPÍTULO VII. DESAFÍOS Y TENDENCIAS FUTURAS

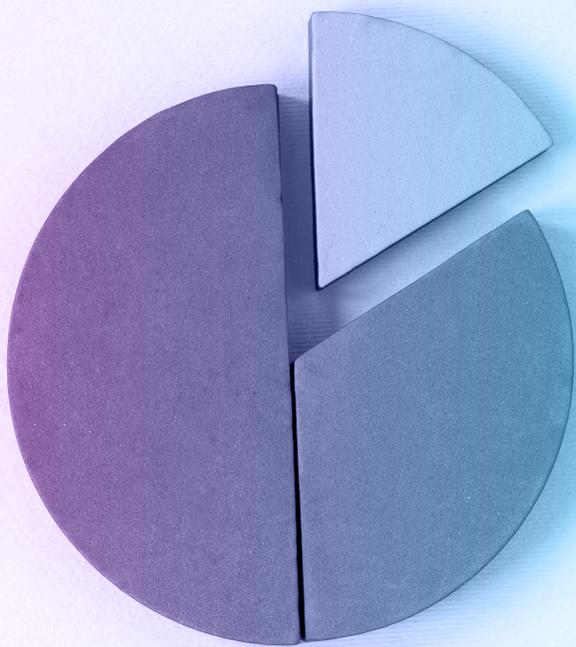
Desafíos Actuales en la Visualización de Datos Educativos y Sociales.....	119
El Rol de la Inteligencia Artificial en la Visualización de Datos.....	121
Tendencias Emergentes y Futuras en Visualización de Datos.....	123
Preparándose para el Futuro de la Visualización de Datos.....	125

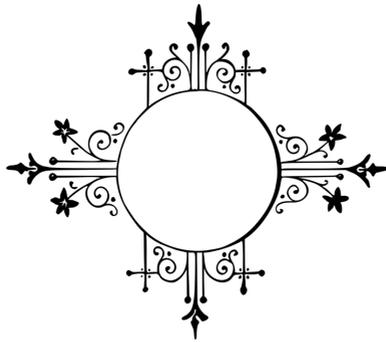
REFUERZO ACADÉMICO CAPÍTULOS TEÓRICOS A EVALUAR

CAPÍTULO I.....	129
CAPITULO II.....	135
CAPITULO III.....	143
CAPÍTULO IV.....	150
CAPITULO VII.....	154

REFERENCIAS BIBLIOGRÁFICAS

CITAS Y REFERENCIAS.....	159
--------------------------	-----





PREFACIO

DATOS EN EDUCACIÓN Y CIENCIAS SOCIALES

La visualización de datos se ha convertido en una herramienta esencial en la sociedad actual debido a su capacidad para abordar la creciente cantidad de información que enfrentamos diariamente. En un mundo inundado de datos, la visualización permite traducir estos números y estadísticas en representaciones visuales comprensibles. Esto facilita la identificación de patrones, tendencias y relaciones que de otra manera podrían pasarnos desapercibidas, ya que la mente humana tiende a procesar mejor la información visual que los datos en forma de números o texto.

Además de simplificar la información, la visualización de datos contribuye significativamente a la toma de decisiones informadas. Los líderes empresariales, políticos y profesionales de la salud, entre otros, pueden utilizar gráficos y visualizaciones para analizar datos complejos y basar sus decisiones en evidencia sólida. Esta capacidad de tomar decisiones respaldadas por datos conduce a una toma de decisiones más precisa y estratégica en diversas áreas de la sociedad.

La comunicación efectiva es otra razón clave por la que la visualización de datos es esencial. Al presentar información en forma de gráficos y visualizaciones, se hace accesible a un público más amplio y diverso. Esto es particularmente importante en un mundo donde la información se comparte constantemente y se consume a gran velocidad. Las visualizaciones de datos ayudan a transmitir mensajes de manera más clara y eficaz, lo que es esencial para la educación, la conciencia pública y la comunicación en general.

La visualización de datos también desempeña un papel fundamental en la identificación de problemas y oportunidades. Al representar datos en gráficos, es más fácil detectar anomalías o áreas en las que se pueden realizar mejoras. Esto es crucial en campos como la gestión empresarial, la ciencia, la salud pública y muchas otras disciplinas donde el análisis de datos desempeña un papel crucial en la resolución de problemas.

En última instancia, la visualización de datos empodera a las personas al proporcionarles acceso a información importante y permitirles tomar decisiones más informadas. Ya sea en contextos políticos, científicos o empresariales, la capacidad de entender y comunicar datos de manera efectiva es esencial para tomar decisiones informadas y avanzar en la sociedad. La visualización de datos no solo simplifica la información, sino que también estimula la innovación, al ayudar a los científicos, investigadores y



empresas a explorar datos de nuevas maneras y descubrir insights que podrían no ser evidentes de otra manera. En resumen, en un mundo cada vez más impulsado por la información, la visualización de datos desempeña un papel vital en la comprensión, la toma de decisiones y la comunicación.

Objetivos y Alcance del Libro

El libro: Visualización de Datos Educativos y Sociales con Python: Herramientas para Decisiones Informadas tiene un conjunto de objetivos claramente definidos que guían su contenido y propósito. En primer lugar, busca proporcionar a los lectores una sólida introducción a la visualización de datos, abarcando desde los conceptos fundamentales hasta técnicas más avanzadas. A lo largo del libro, se centra en el uso de Python como lenguaje de programación principal para crear visualizaciones efectivas, presentando bibliotecas populares como Matplotlib, Seaborn y Plotly.

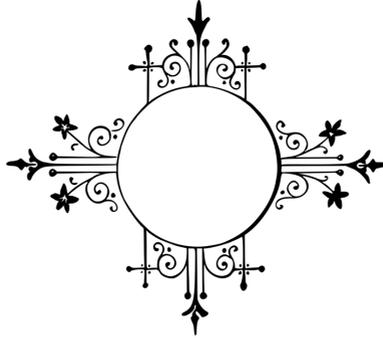
El alcance del libro se extiende a la visualización de datos en dos áreas principales: educación y aspectos sociales. Los ejemplos y ejercicios prácticos se basan en datos reales de estos campos, permitiendo a los lectores aplicar los conocimientos adquiridos a situaciones del mundo real. Además, el libro aborda una variedad de tipos de datos, desde datos tabulares hasta series de tiempo y datos geoespaciales, lo que amplía su aplicabilidad en diversas disciplinas.

Uno de los objetivos más destacados del libro es demostrar cómo la visualización de datos puede contribuir significativamente a la toma de decisiones informadas. Los lectores aprenderán a utilizar las visualizaciones como herramientas poderosas para analizar problemas, identificar patrones y evaluar resultados, tanto en el contexto educativo como en cuestiones sociales más amplias.

El alcance del libro también incluye una serie de casos de estudio y ejemplos concretos que ilustran la aplicación práctica de la visualización de datos en diferentes contextos. Estos casos de estudio permiten a los lectores explorar problemas específicos y entender cómo utilizar las visualizaciones para comunicar información de manera efectiva y respaldar la toma de decisiones.

CAPÍTULO I

INTRODUCCIÓN A LA VISUALIZACIÓN
DE DATOS



CAPÍTULO I.

INTRODUCCIÓN A LA VISUALIZACIÓN DE DATOS

Definición y Fundamentos de la Visualización de Datos

La visualización de datos es una disciplina que se centra en representar información y datos numéricos de manera gráfica y visual, con el objetivo de facilitar la comprensión y el análisis de patrones, tendencias y relaciones en los datos. En esencia, se trata de una forma efectiva de comunicar información compleja a través de imágenes y gráficos visuales¹.

En sus fundamentos, la visualización de datos se basa en conceptos clave. Uno de ellos es la codificación visual, que implica asignar valores numéricos o categorías a propiedades visuales como longitud, color, posición y forma. Esta asignación permite a los espectadores interpretar y comprender la información representada de manera más efectiva².

La percepción visual es otro aspecto fundamental. Se refiere

re a cómo el cerebro humano procesa y organiza la información visual, incluyendo la capacidad de distinguir colores, tamaños y patrones. Comprender los principios de la percepción visual es esencial para diseñar visualizaciones que sean intuitivas y fáciles de entender.

La elección del tipo de gráfico adecuado es esencial en la visualización de datos. Diferentes tipos de gráficos, como barras, líneas, dispersión o mapas, son apropiados para diferentes tipos de datos y objetivos de comunicación. Seleccionar el gráfico correcto es clave para transmitir la información de manera efectiva.

La simplificación y la claridad son principios esenciales. Los datos complejos pueden ser abrumadores, por lo que es importante simplificarlos y presentar solo la información relevante para el mensaje que se quiere transmitir. Además, la claridad en la presentación es esencial para evitar malentendidos y asegurar que los espectadores puedan interpretar fácilmente la visualización³.

La visualización de datos no se trata solo de mostrar números en gráficos, sino de contar una historia. Proporcionar



contexto y narrativa ayuda a los espectadores a comprender el significado detrás de los datos y a conectarlos con el propósito de la visualización.

En el entorno digital actual, la interactividad desempeña un papel importante en la visualización de datos. Las herramientas interactivas permiten a los usuarios explorar los datos por sí mismos y obtener información específica según sus necesidades, lo que mejora la experiencia de usuario y la comprensión de los datos.

Finalmente, la ética y la precisión son fundamentales en la visualización de datos. Los datos deben representarse de manera honesta y precisa, evitando distorsiones o sesgos que puedan conducir a interpretaciones erróneas. La integridad y la ética en la presentación de datos son esenciales para mantener la confianza del público.

Breve Historia de la Visualización de Datos

La historia de la visualización de datos se remonta a siglos atrás, pero su evolución significativa ha ocurrido en los últimos dos siglos, especialmente en el contexto de avances tecnológicos y la creciente cantidad de datos disponibles. Se detalla⁴⁻⁷:

- **Siglo XVII - Siglo XVIII:** Los antecedentes de la visualización de datos se encuentran en los gráficos estadísticos tempranos, como los utilizados por John Graunt en el siglo XVII para representar estadísticas de mortalidad en Londres. Estos gráficos rudimentarios sentaron las bases para la representación gráfica de datos.
- **Siglo XIX:** Vio el desarrollo de técnicas de visualización

más sofisticadas. William Playfair, un estadístico escocés, introdujo gráficos de barras, gráficos de líneas y gráficos circulares en su obra “The Commercial and Political Atlas” (Atlas Comercial y Político) en 1786. Estos gráficos se convirtieron en herramientas valiosas para representar datos económicos y políticos.

- **Siglo XX:** Con el avance de la tecnología de impresión y la estadística, la visualización de datos se volvió más común y sofisticada en el siglo XX. En 1924, Willard C. Brinton publicó “Graphic Methods for Presenting Facts”, que se considera un hito en la historia de la visualización de datos. Durante este siglo, la visualización se utilizó ampliamente en campos como la ciencia, la medicina y los negocios.
- **Década de 1960:** La introducción de las computadoras y el software gráfico en la década de 1960 revolucionó la visualización de datos. Se desarrollaron programas de gráficos por computadora que permitieron a los usuarios crear visualizaciones de datos de manera más eficiente.
- **Década de 1980 y 1990:** La aparición de la World Wide Web en la década de 1990 y el aumento de la disponibilidad de datos en línea dieron lugar a nuevas oportunidades para la visualización de datos. Se crearon herramientas y técnicas específicas para la presentación de



información en línea.

- **Siglo XXI:** En la era de la información digital y el big data, la visualización de datos ha adquirido una importancia aún mayor. La tecnología actual permite crear visualizaciones interactivas y dinámicas, lo que facilita la exploración de datos complejos. Además, se han desarrollado herramientas de software específicas, como Tableau, Power BI y D3.js, para crear visualizaciones avanzadas.
- **Actualidad:** La visualización de datos desempeña un papel crucial en campos como la ciencia de datos, la toma de decisiones empresariales, la comunicación científica y la conciencia pública. Se utiliza para analizar tendencias, comunicar hallazgos, contar historias con datos y empoderar a las personas con información.

La Visualización de Datos en el Contexto Actual

La visualización de datos es una técnica importante y poderosa que se utiliza para representar información de manera gráfica y comprensible. En el contexto actual, la visualización de datos desempeña un papel crucial en diversas áreas, desde la toma de decisiones empresariales hasta la comunicación de información científica y la concienciación pública. Algunos ejemplos⁸⁻¹²:

- **COVID-19 y salud pública:** Durante la pandemia de COVID-19, las visualizaciones de datos se han utilizado ampliamente para mostrar la propagación del virus, las tasas de infección, las tendencias epidemiológicas y la distribución de vacunas. Estos gráficos y mapas ayudaron a los gobiernos y al público en general a compren-

der mejor la situación y tomar decisiones informadas.

- **Ciencia y investigación:** En campos científicos como la biología, la astronomía y la climatología, la visualización de datos es esencial para representar y analizar conjuntos de datos complejos. Las visualizaciones ayudan a los científicos a identificar patrones, tendencias y relaciones en los datos.
- **Negocios y análisis empresarial:** Las empresas utilizan visualizaciones de datos para analizar métricas clave, como ventas, ingresos y rendimiento del mercado. Los gráficos y tablas interactivos permiten a los analistas explorar datos y tomar decisiones informadas.
- **Medios de comunicación y periodismo:** Los medios de comunicación a menudo utilizan infografías y gráficos para explicar temas complejos, como las elecciones, la economía o los problemas sociales. Esto ayuda a que la información sea más accesible para el público.
- **Educación y concienciación:** Las visualizaciones de datos se utilizan para educar al público sobre cuestiones importantes, como el cambio climático, la desigualdad económica y los problemas de salud. Estas representaciones visuales pueden hacer que los datos sean más comprensibles y memorables.
- **Tecnología y análisis de datos:** En el ámbito de la tecnología, la visualización de datos es fundamental para el análisis de big data y la inteligencia empresarial. Las empresas utilizan paneles de control y gráficos interactivos para monitorear el rendimiento de sistemas y aplicaciones en tiempo real.

- **Gobierno y políticas públicas:** Los gobiernos utilizan visualizaciones de datos para comunicar información relevante a los ciudadanos y para respaldar la toma de decisiones en políticas públicas. Esto incluye la presentación de datos demográficos, estadísticas de empleo y resultados de elecciones.
- **Finanzas y mercados:** En el sector financiero, la visualización de datos es esencial para rastrear el rendimiento de las inversiones, analizar tendencias del mercado y tomar decisiones financieras informadas.

Importancia de la Visualización de Datos en el Ámbito Educativo y Social

La visualización de datos desempeña un papel fundamental en el ámbito educativo y social, ya que ayuda a comunicar información de manera efectiva, facilita el aprendizaje y fomenta la comprensión de problemas complejos. Algunos ejemplos¹³⁻¹⁸:

- **Claridad y accesibilidad:** La visualización de datos transforma información abstracta o números en representaciones visuales comprensibles, como gráficos, tablas y diagramas. Esto hace que la información sea más accesible y fácil de entender para personas de diferentes niveles de conocimiento y habilidades.
- **Facilita el aprendizaje:** En



el ámbito educativo, las visualizaciones de datos pueden utilizarse para enseñar conceptos complejos de manera más efectiva. Los estudiantes pueden visualizar tendencias, patrones y relaciones en los datos, lo que facilita la retención y la comprensión.

- **Toma de decisiones informadas:** En el contexto social y educativo, la visualización de datos ayuda a los individuos y las organizaciones a tomar decisiones basadas en evidencia. Al ver datos de manera clara y objetiva, es más probable que se tomen decisiones informadas en temas como políticas educativas, salud pública y desarrollo comunitario.
- **Concienciación y defensa de causas:** Las visualizaciones de datos son poderosas herramientas para crear conciencia sobre cuestiones sociales y promover cambios positivos. Los activistas y las organizaciones sin fines de lucro pueden utilizar gráficos y diagramas para destacar problemas como la desigualdad, el cambio climático o la pobreza, lo que puede movilizar a la sociedad y a los responsables de la toma de decisiones.
- **Comunicación efectiva:** En el ámbito educativo, los educadores pueden utilizar visualizaciones de datos para explicar conceptos, resultados de investigaciones y datos históricos de manera más efectiva. Esto permite una comunicación más clara y memorable en el aula.
- **Análisis de rendimiento:** En la educación, las visualizaciones de datos pueden ayudar a los administradores escolares y a los educadores a realizar un seguimiento del rendimiento académico de los estudiantes y a identificar áreas que requieren atención. Esto es esencial para mejorar la calidad de la educación.

- **Comprensión de problemas complejos:** En el ámbito social, problemas complejos como la migración, la urbanización y la salud pública pueden visualizarse de manera efectiva para ayudar a las personas a comprender las dimensiones y las causas subyacentes de estos problemas.
- **Promoción de la alfabetización de datos:** En un mundo cada vez más impulsado por los datos, la visualización de datos promueve la alfabetización de datos al enseñar a las personas cómo leer y interpretar gráficos y tablas, lo que es una habilidad fundamental en la sociedad actual.

Principales Herramientas y Tendencias en Visualización de Datos

La visualización de datos es una disciplina en constante evolución y su importancia sigue creciendo en diversos ámbitos. En términos de herramientas, existen varias opciones poderosas, como Tableau, Power BI, D3.js, Python con librerías como Matplotlib y Seaborn, R con ggplot2 y plataformas en línea como Datawrapper y Flourish, que permiten crear visualizaciones personalizadas.



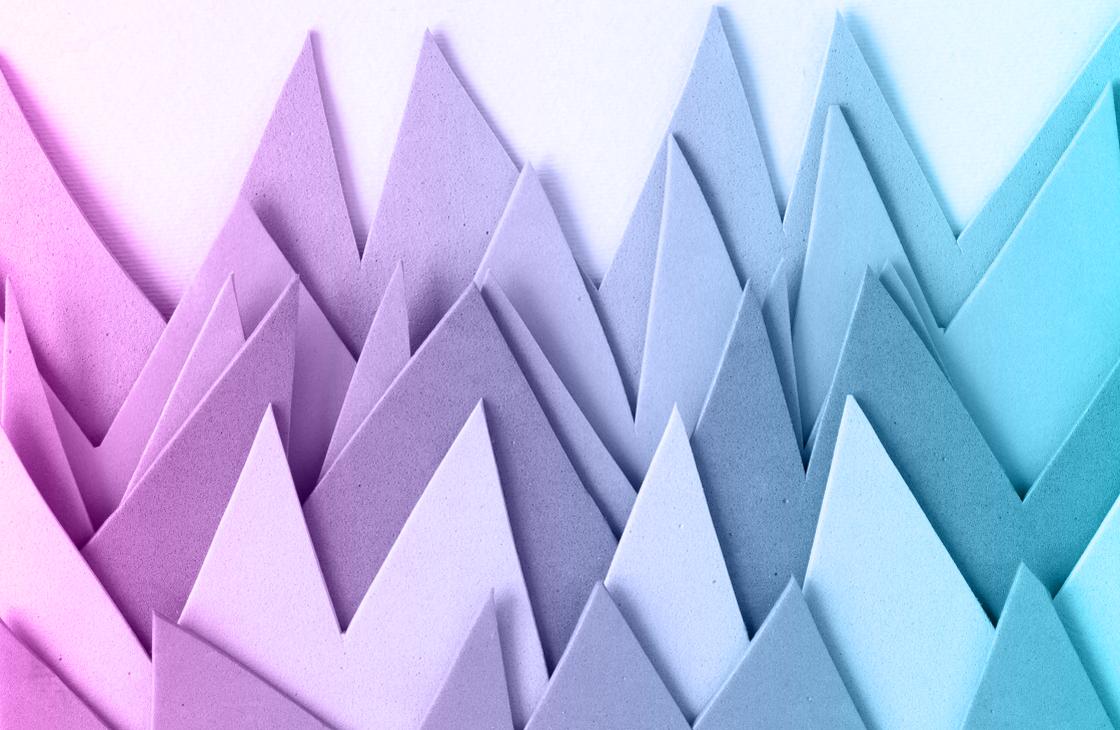
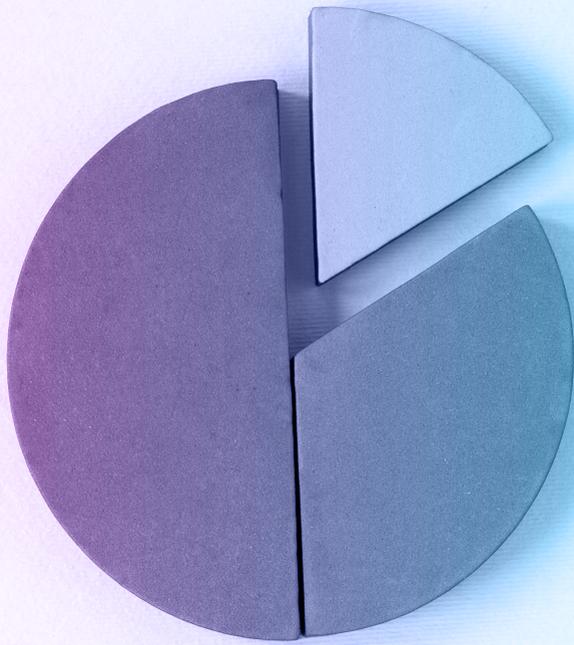
Una tendencia destacada es la creciente demanda de visualizaciones interactivas, que permiten a los usuarios explorar datos por sí mismos y obtener información más detallada al interactuar con gráficos y tablas. Además, la visualización en tiempo real es esencial en áreas como

las finanzas y la ciencia de datos, lo que impulsa la necesidad de herramientas y técnicas que permitan representar datos en tiempo real.

La visualización en 3D y la realidad virtual (RV) están ganando relevancia para representar datos tridimensionales en campos como la ciencia y la medicina. La integración del aprendizaje automático en la visualización de datos también es una tendencia en alza, permitiendo la generación automática de visualizaciones y la detección de patrones en grandes conjuntos de datos.

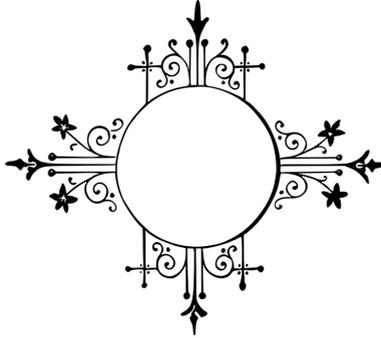
La visualización de datos geoespaciales sigue siendo crucial, con aplicaciones en cartografía y análisis de ubicaciones. Herramientas como ArcGIS y QGIS son ampliamente utilizadas en este campo. Asimismo, las empresas están adoptando la visualización de datos como parte fundamental de la toma de decisiones empresariales, con paneles de control y herramientas de análisis en tiempo real.

En términos de narrativa y accesibilidad, se presta más atención a contar historias con los datos y hacer que las visualizaciones sean comprensibles para personas con discapacidades visuales. En resumen, las herramientas y tendencias en visualización de datos evolucionan para satisfacer las demandas de una sociedad cada vez más orientada hacia los datos, proporcionando opciones más avanzadas y accesibles para representar y comprender la información de manera efectiva.



CAPÍTULO II

FUNDAMENTOS DE LA VISUALIZACIÓN DE
DATOS CON PYTHON



CAPÍTULO II.

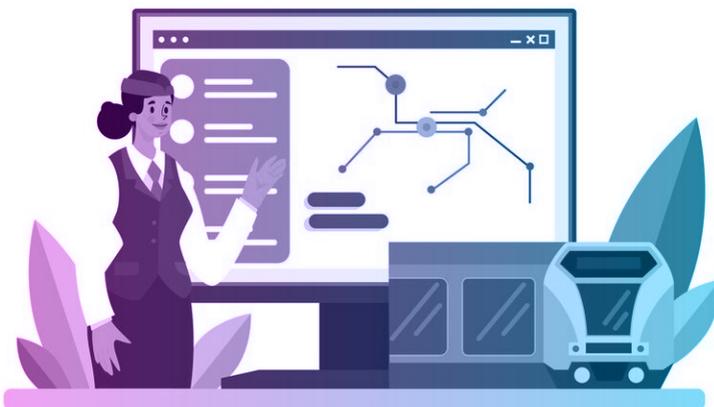
FUNDAMENTOS DE LA VISUALIZACIÓN DE DATOS CON PYTHON

Principios de Diseño en Visualización de Datos

Los principios de diseño en visualización de datos son pautas fundamentales que guían la creación de representaciones visuales efectivas y comprensibles de datos. Estos principios ayudan a garantizar que las visualizaciones sean claras, precisas y capaces de comunicar la información de manera efectiva. Ejemplo¹⁹⁻²²:

- **Claridad y simplicidad:** La visualización debe ser lo más simple y clara posible. Evite el exceso de elementos innecesarios o decorativos que puedan distraer al espectador. La simplicidad facilita la comprensión de la información.
- **Precisión:** Los datos deben representarse de manera precisa y fiel a la realidad. Evite distorsiones, escalas engañosas o malas interpretaciones de los datos.

- **Relevancia:** Muestre solo los datos que son relevantes para la historia que desea contar o la pregunta que desea responder. Evite la sobrecarga de información.
- **Jerarquía visual:** Utilice técnicas de jerarquía visual para destacar la información más importante. Esto puede lograrse mediante el uso de colores, tamaños y ubicaciones estratégicas.
- **Consistencia:** Mantenga una consistencia en la elección de colores, fuentes y estilos a lo largo de la visualización para que sea más fácil de leer y comprender. Esto incluye la consistencia en la forma en que se representan las categorías y los datos.
- **Usabilidad:** Diseñe la visualización pensando en la audiencia. Debe ser fácil de usar y comprender para las personas que la verán. Utilice leyendas, etiquetas y elementos de navegación claros.
- **Narrativa:** Cuente una historia con los datos. La visualización debe tener un propósito claro y comunicar una idea o mensaje específico. Utilice el diseño para guiar a los espectadores a través de la narrativa de la visualización.



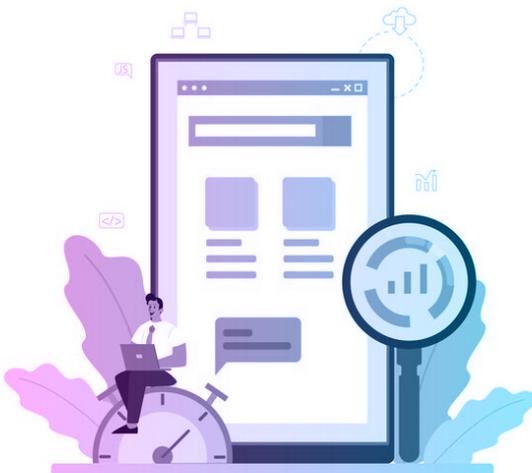
- **Color y contraste:** Utilice el color de manera efectiva para resaltar elementos importantes y crear distinciones visuales. Asegúrese de que los colores sean accesibles para personas con discapacidades visuales y evite el uso excesivo de colores llamativos que puedan dificultar la lectura.
- **Interactividad:** Si es apropiado, agregue elementos interactivos que permitan a los usuarios explorar los datos por sí mismos. Esto puede incluir gráficos desplegable, filtros y herramientas de zoom.
- **Pruebas y retroalimentación:** Antes de publicar una visualización de datos, pruébela con una audiencia representativa y obtenga retroalimentación. Aprender de la retroalimentación ayuda a mejorar la calidad de la visualización.
- **Contexto:** Proporcione contexto y metadatos claros que expliquen la fuente de los datos, la metodología utilizada y cualquier limitación en la interpretación de los datos.
- **Responsividad y accesibilidad:** Asegúrese de que la visualización sea receptiva, es decir, se adapte a diferentes dispositivos y tamaños de pantalla. Además, haga que la visualización sea accesible para personas con discapacidades visuales al proporcionar alternativas textuales y utilizar etiquetas descriptivas.

Gráficos Estáticos vs. Gráficos Interactivos

Los gráficos estáticos y los gráficos interactivos son dos enfoques diferentes para representar y visualizar datos.

Cada uno tiene sus propias ventajas y desventajas, y la elección entre ellos depende de los objetivos y las necesidades específicas de comunicación de datos. Se detalla²³⁻²⁷:

- **Gráficos Estáticos:** Los gráficos estáticos son representaciones visuales de datos que no permiten la interacción del usuario. Son imágenes fijas que muestran información en un formato visual, como barras, líneas, pastel o mapas, pero no permiten que el espectador manipule los datos directamente. Estos gráficos son ideales para presentaciones impresas, informes estáticos o cuando se necesita una representación visual simple y fácil de entender de los datos.
- **Gráficos Interactivos:** Los gráficos interactivos, por otro lado, son herramientas dinámicas que permiten a los usuarios explorar y manipular datos en tiempo real. Estos gráficos suelen estar en formato digital y se pueden utilizar en aplicaciones web, tableros de control interactivos o informes en línea. Los usuarios pueden hacer clic, desplazarse, acercar y alejar, filtrar datos y obtener información detallada al interactuar con el gráfico. Esto facilita la exploración de tendencias, patrones y relaciones ocultas en los datos.



- **Ventajas de los Gráficos Estáticos:** Los gráficos estáticos son eficaces cuando se necesita comunicar una idea clara y sencilla de manera rápida. Son ideales para presentaciones impresas y documentos que no re-

quieren interacción. Además, son más fáciles de crear y pueden ser más adecuados para audiencias que no están familiarizadas con herramientas interactivas.

- **Ventajas de los Gráficos Interactivos:** Los gráficos interactivos son poderosos cuando se necesita explorar datos en profundidad y permitir a los usuarios personalizar su experiencia. Son ideales para presentar grandes conjuntos de datos o datos en constante cambio, ya que se pueden actualizar en tiempo real. Los gráficos interactivos también pueden ser más persuasivos al permitir que los usuarios descubran información por sí mismos, lo que puede llevar a una comprensión más profunda y una toma de decisiones más informada.

Creación de Gráficos Básicos con Matplotlib

La biblioteca Matplotlib es una de las herramientas más populares en Python para la creación de gráficos básicos y avanzados:

Importa la biblioteca Matplotlib: Primero, debes importar la biblioteca Matplotlib en tu código. Normalmente, se importa bajo el alias `plt` para abreviar:

```
import matplotlib.pyplot as plt
```

Prepara tus datos: Antes de crear un gráfico, debes tener tus datos listos en listas o arrays de NumPy, dependiendo de tu preferencia. Por ejemplo:

```
x = [1, 2, 3, 4, 5] y = [10, 15, 13, 18, 22]
```

Crea un gráfico de línea: Un gráfico de línea es uno de los tipos de gráficos más simples que puedes crear con Matplotlib. Utiliza la función plot para esto:

```
plt.plot(x, y)
```

Agrega etiquetas y título: Es importante etiquetar tus ejes y proporcionar un título para tu gráfico. Puedes hacerlo utilizando las funciones xlabel, ylabel, y title:

```
plt.xlabel('Eje X') plt.ylabel('Eje Y') plt.title('Mi Gráfico de Línea')
```

Muestra el gráfico: Para mostrar el gráfico en una ventana emergente, utiliza la función show:

```
plt.show()
```



Introducción a Seaborn para Gráficos Estadísticos

Seaborn es una biblioteca de visualización de datos en Python que se basa en Matplotlib y proporciona una interfaz de alto nivel para crear gráficos estadísticos atractivos y efectivos. Está especialmente diseñada para trabajar con conjuntos de datos complejos y es ampliamente utilizada en el análisis de datos y la exploración de datos estadísticos. Estructuremos un código base para entender mejor:

Instalación de Seaborn: Para comenzar a usar Seaborn, primero debes asegurarte de que esté instalada en tu entorno de Python. Puedes instalarla usando pip:

```
pip install seaborn
```

Importación de Seaborn y de los datos: Después de la instalación, importa la biblioteca Seaborn junto con otras bibliotecas necesarias, como NumPy y Pandas, y carga tus datos en un DataFrame de Pandas. Por ejemplo:

```
import seaborn as sns
import pandas as pd
# Cargar datos en un DataFrame de Pandas
df = pd.read_csv('tu_archivo.csv')
x = [1, 2, 3, 4, 5] y = [10, 15, 13, 18, 22]
```

Creación de gráficos estadísticos: Seaborn ofrece una variedad de funciones para crear gráficos estadísticos de manera sencilla. Algunos ejemplos de tipos de gráficos que puedes crear incluyen:

Gráficos de dispersión (scatter plots):

```
sns.scatterplot(data=df, x='columna_x', y='columna_y')
```

Gráficos de barras (bar plots):

```
sns.barplot(data=df, x='columna_x', y='columna_y')
```

Gráficos de caja (box plots):

```
sns.boxplot(data=df, x='columna_x', y='columna_y')
```

Histogramas y distribuciones (histograms y distributions):

```
sns.histplot(data=df, x='columna_x', kde=True)
```

Mapas de calor (heatmaps):

```
sns.heatmap(data=df.corr(), annot=True,  
            cmap='coolwarm')
```

Personalización y estilización: Seaborn permite personalizar tus gráficos mediante la modificación de colores, estilos, tamaños de fuente, entre otros aspectos. También puedes agregar etiquetas, títulos y leyendas para hacer que tus gráficos sean más informativos y estéticamente agradables.

Exploración y análisis de datos: Una de las ventajas de Seaborn es su capacidad para facilitar la exploración de datos. Puedes utilizar gráficos estadísticos para identificar patrones, tendencias, relaciones y anomalías en tus datos.

Exportación de gráficos: Una vez que hayas creado tus gráficos estadísticos en Seaborn, puedes exportarlos en diversos formatos, como imágenes (PNG, JPEG) o documentos (PDF) para su uso en presentaciones, informes o publicaciones.

Procedamos a ver resultados del código aplicado:

```
# Datos para gráfico de dispersión
np.random.seed(0)
x_scatter = np.random.randn(100)
y_scatter = np.random.randn(100)
sizes = np.random.randint(10, 200, size=100)

# Datos para gráfico de barras
categories = ['A', 'B', 'C', 'D']
values = np.random.randint(10, 50, size=(4, 2))

# Datos para box plot
box_data = np.random.normal(size=(100, 4))

# Datos para histograma
hist_data = np.random.normal(size=1000)

# Datos para mapa de calor
heat_data = np.random.rand(10, 12)

# Creando los gráficos
fig, axes = plt.subplots(3, 2, figsize=(15, 15))
```

```
# Gráfico de dispersión
sns.scatterplot(x=x_scatter, y=y_scatter, size=sizes,
               sizes=(20, 200), ax=axes[0, 0])
axes[0, 0].set_title('Gráfico de Dispersión')

# Gráfico de barras
sns.barplot(x=categories, y=values[:, 0], ax=axes[0, 1])
sns.barplot(x=categories, y=values[:, 1], ax=axes[0, 1],
            color='r', alpha=0.5)
axes[0, 1].set_title('Gráfico de Barras')

# Box plot
sns.boxplot(data=box_data, ax=axes[1, 0])
axes[1, 0].set_title('Box Plot')

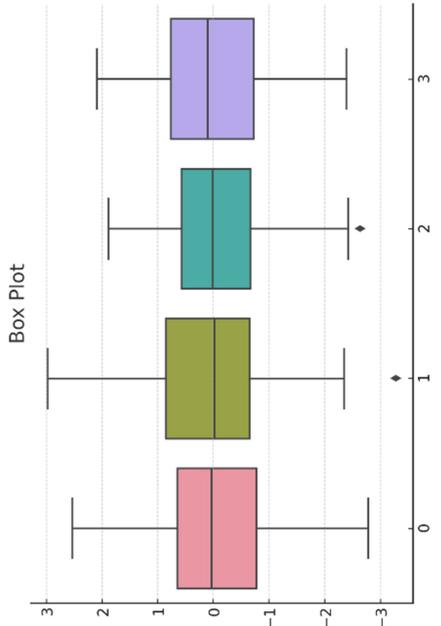
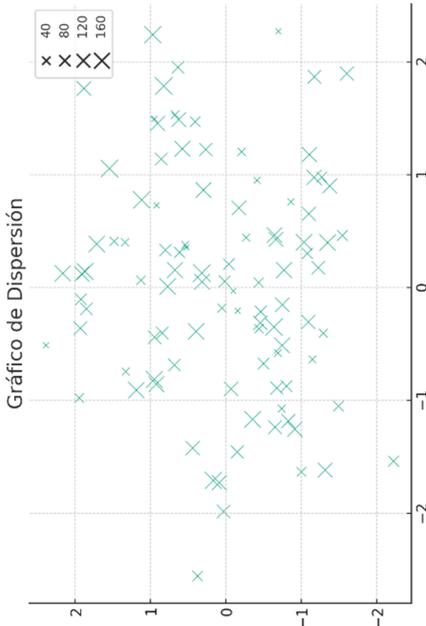
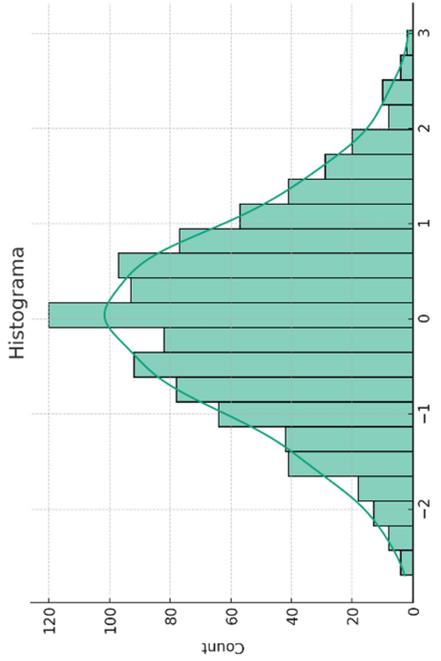
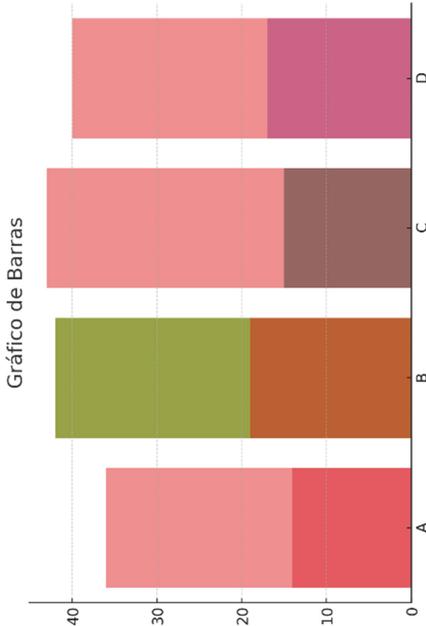
# Histograma
sns.histplot(hist_data, kde=True, ax=axes[1, 1])
axes[1, 1].set_title('Histograma')

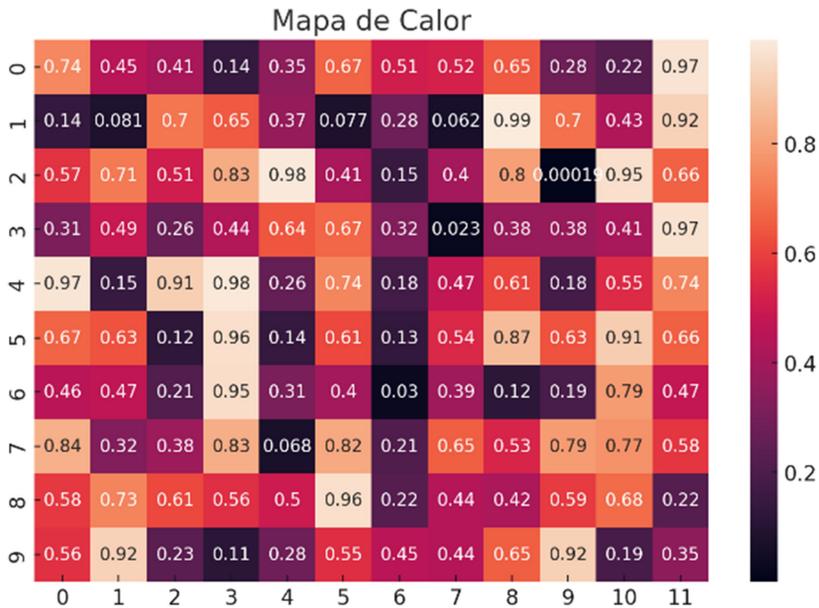
# Mapa de calor
sns.heatmap(heat_data, annot=True, ax=axes[2, 0])
axes[2, 0].set_title('Mapa de Calor')

# Ajustando la disposición de los gráficos
plt.tight_layout()

# Eliminando el gráfico vacío
fig.delaxes(axes[2][1])

plt.show()
```





Visualizaciones Interactivas con Bokeh

Bokeh es una biblioteca de visualización interactiva en Python que se utiliza para crear gráficos y visualizaciones web interactivas de alta calidad. Bokeh es especialmente útil cuando deseas crear visualizaciones que permitan a los usuarios explorar y manipular los datos en tiempo real en aplicaciones web o cuadernos interactivos²⁸. Se genera una estructura:

Instalación de Bokeh: Para comenzar a utilizar Bokeh, primero debes asegurarte de que esté instalado en tu entorno de Python. Puedes instalarlo utilizando pip:

```
pip install bokeh
```

Importación de Bokeh y configuración del entorno: Importa las funciones y clases necesarias de Bokeh y configura el entorno de salida (por ejemplo, si deseas mostrar los gráficos en un cuaderno Jupyter):

```
import seaborn as sns
import pandas as pd
# Cargar datos en un DataFrame de Pandas
df = pd.read_csv('tu_archivo.csv')
x = [1, 2, 3, 4, 5] y = [10, 15, 13, 18, 22]
```

Creación de un gráfico básico: Para crear un gráfico con Bokeh, primero debes crear un objeto figure, que es la lienzo en el que se dibujarán los elementos del gráfico. Luego, puedes agregar diferentes tipos de elementos, como círculos, líneas, barras, etc. Un ejemplo de un gráfico de dispersión básico:

```
# Crear un objeto figure
p = figure(plot_width=400, plot_height=400)
# Agregar un gráfico de dispersión
p.circle([1, 2, 3, 4, 5], [6, 7, 2, 4, 5], size=10, color="navy")
# Mostrar el gráfico
show(p)
```

Personalización de la visualización: Bokeh permite personalizar tus gráficos en detalle. Puedes cambiar colores, tamaños, estilos de línea, agregar etiquetas, títulos y mucho

más. La documentación de Bokeh es una excelente referencia para aprender a personalizar tus gráficos según tus necesidades.

Interactividad: Una de las características distintivas de Bokeh es su capacidad para crear visualizaciones interactivas. Puedes agregar herramientas interactivas a tu gráfico, como zoom, panorámica, selección de datos y más. Por ejemplo, puedes agregar una herramienta de selección de datos con:

```
from bokeh.models import HoverTool
# Crear una herramienta de selección de datos
hover = HoverTool()
hover.tooltips = [{"X", "@x"}, {"Y", "@y"}]

# Agregar la herramienta al objeto figure
p.add_tools(hover)
```

Exportación y uso en aplicaciones web: Una vez que hayas creado tus visualizaciones interactivas con Bokeh, puedes exportarlas como archivos HTML para su uso en aplicaciones web. Bokeh también se puede integrar con otros marcos web como Flask y Django para crear aplicaciones web interactivas más complejas

Se detalla código de implementación interactiva para líneas interactivo:

```
from bokeh.plotting import figure, show, output_notebook
from bokeh.models import HoverTool
import numpy as np
# Preparando los datos para el gráfico de líneas
```

```
x_line = np.arange(1, 11)
y_line = np.random.rand(10) * 100

# Creando el gráfico de líneas interactivo
line_plot = figure(plot_width=600, plot_height=400,
title="Gráfico de Líneas Interactivo",
tools="pan,wheel_zoom,box_zoom,reset,hover,save")
line_plot.line(x_line, y_line, line_width=2, color="navy",
alpha=0.6)
line_plot.circle(x_line, y_line, size=8, color="orange",
alpha=0.9)

# Configurando el hover tool para mostrar los valores
hover = line_plot.select(dict(type=HoverTool))
hover.tooltips = [{"X", "@x"}, {"Y", "@y"}]

# Mostrar el gráfico
output_notebook()

show(line_plot)
```

Se detalla código de implementación interactiva para barras interactivo:

```
# Preparando los datos para el gráfico de barras
categories = ['A', 'B', 'C', 'D', 'E']
values = np.random.randint(10, 100, size=5)

# Creando el gráfico de barras interactivo
```

```
bar_plot = figure(x_range=categories, plot_height=400,
                  title="Gráfico de Barras Interactivo",
                  tools="pan,wheel_zoom,box_zoom,reset,hover,save")
bar_plot.vbar(x=categories, top=values, width=0.5,
              color="green", alpha=0.7)

# Configurando el hover tool para mostrar los valores
hover = bar_plot.select(dict(type=HoverTool))
hover.tooltips = [("Categoría", "@x"), ("Valor", "@top")]

# Mostrar el gráfico
show(bar_plot)
```

Se detalla código de implementación interactiva para dispersión interactivo:

```
# Preparando los datos para el gráfico de dispersión
x_scatter = np.random.rand(50) * 100
y_scatter = np.random.rand(50) * 100
sizes = np.random.randint(10, 30, size=50) # Tamaños
aleatorios para los puntos

# Creando el gráfico de dispersión interactivo
scatter_plot = figure(plot_width=600, plot_height=400,
                      title="Gráfico de Dispersión Interactivo",
                      tools="pan,wheel_zoom,box_zoom,reset,hover,save")
scatter_plot.scatter(x_scatter, y_scatter, size=sizes,
                    color="magenta", alpha=0.6, line_color="navy")

# Configurando el hover tool para mostrar los valores
```

```
hover = scatter_plot.select(dict(type=HoverTool))
hover.tooltips = [{"X", "@x"}, {"Y", "@y"}, {"Tamaño", "@size"}]
# Mostrar el gráfico
show(scatter_plot)
```

Visualizaciones Interactivas con Plotly

Plotly es una biblioteca de visualización en Python que se destaca por la creación de visualizaciones interactivas y dinámicas. Proporciona una variedad de gráficos y herramientas para generar visualizaciones atractivas y personalizables que pueden ser utilizadas en aplicaciones web, cuadernos Jupyter y más²⁹. Se detalla una estructura:

Instalación de Plotly: Antes de comenzar, debes asegurarte de tener Plotly instalado en tu entorno de Python. Puedes instalarlo utilizando pip:

```
pip install plotly
```

Importación de Plotly: Importa las funciones y clases necesarias de Plotly para empezar a crear visualizaciones:

```
import plotly.graph_objects as go
```

Creación de visualizaciones interactivas: Plotly se basa en la creación de figuras que contienen datos y diseño. Puedes utilizar diferentes tipos de figuras, como scatter, bar, heatmap, box. Ejemplo de un gráfico de dispersión (scatter plot):

```
import plotly.graph_objects as go
# Crear un gráfico de dispersión interactivo
fig = go.Figure(data=go.Scatter(x=[1, 2, 3, 4, 5], y=[6, 7, 2, 4, 5], mode='markers'))

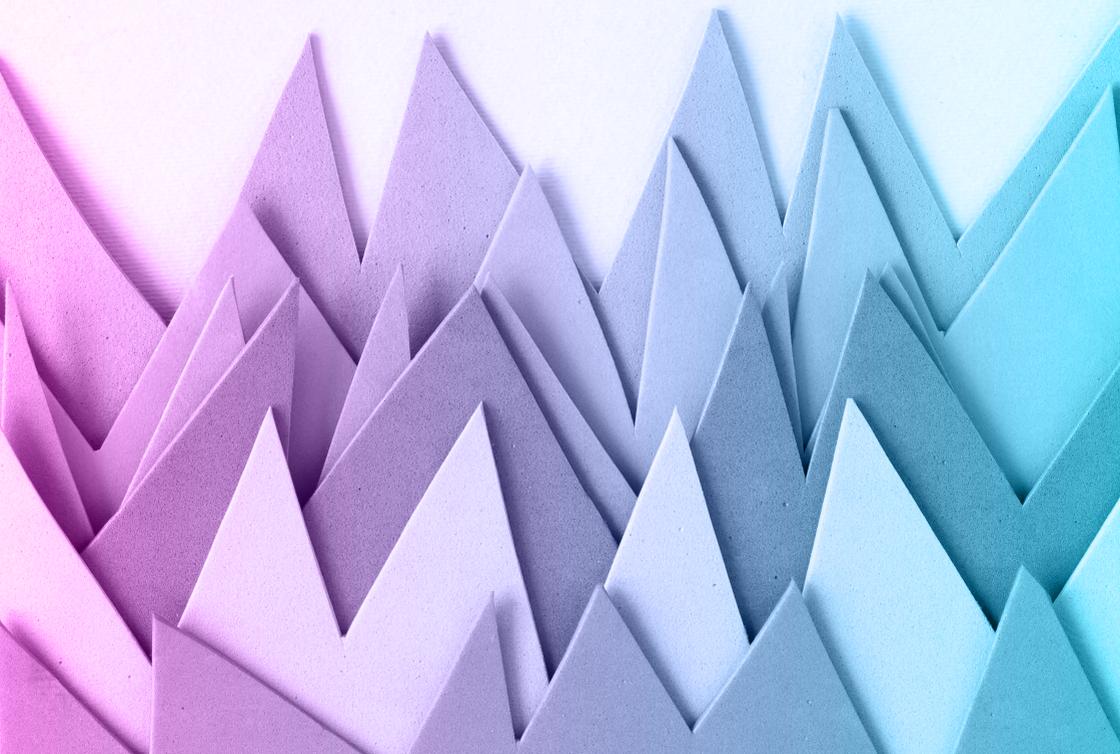
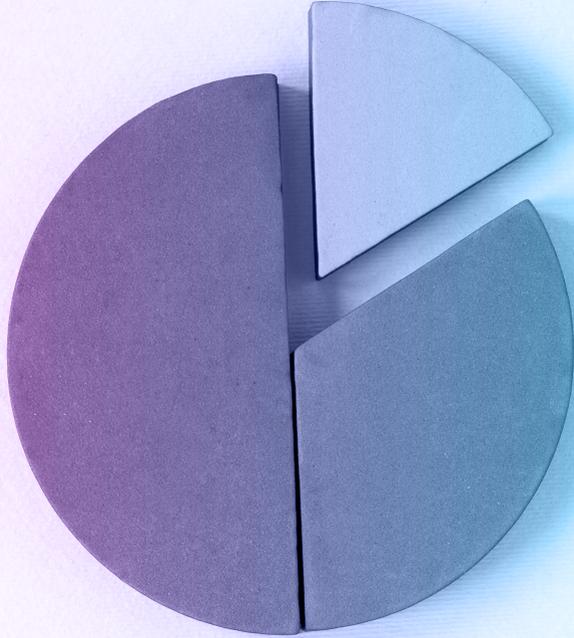
# Personalizar el gráfico
fig.update_layout(
    title='Mi Gráfico de Dispersión',
    xaxis=dict(title='Eje X'),
    yaxis=dict(title='Eje Y')
)

# Mostrar el gráfico interactivo
fig.show()
```

Interactividad y personalización: Plotly proporciona muchas opciones para personalizar tus gráficos y hacerlos interactivos. Puedes agregar herramientas de zoom, selección, información sobre herramientas, botones de exportación y más. Además, puedes ajustar el diseño, los colores, los títulos y las etiquetas según tus preferencias.

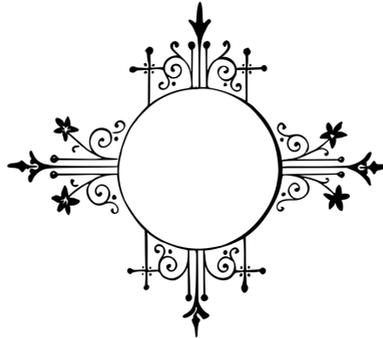
Exportación de visualizaciones: Puedes exportar tus visualizaciones de Plotly en varios formatos, como HTML, imágenes estáticas (PNG, JPEG), PDF y más. Esto facilita la inclusión de tus visualizaciones en informes, presentaciones o aplicaciones web.

Integración con aplicaciones web: Plotly se integra fácilmente con aplicaciones web utilizando frameworks como Dash, que te permiten crear aplicaciones web interactivas basadas en tus visualizaciones de Plotly.



CAPÍTULO III

DATOS EDUCATIVOS Y SOCIALES



CAPÍTULO III.

DATOS EDUCATIVOS Y SOCIALES

Características de los Datos Educativos y Sociales

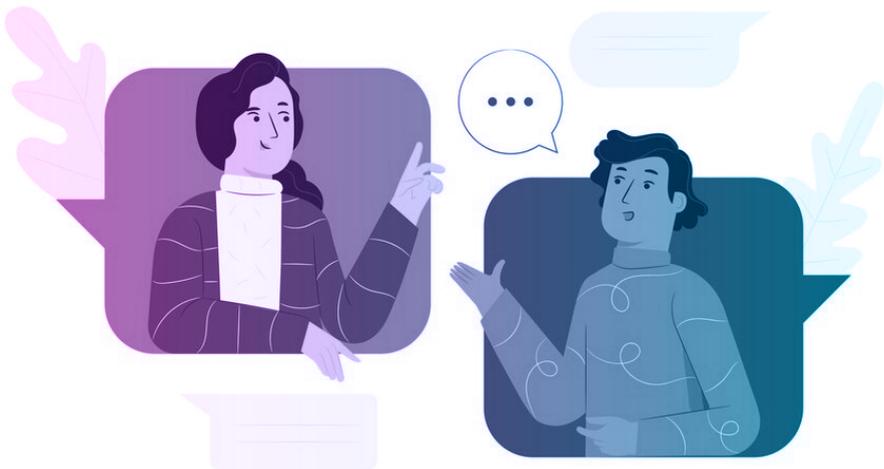
Los datos educativos son un conjunto de información esencial para evaluar, gestionar y mejorar los sistemas educativos en todo el mundo. Una de las principales características de los datos educativos es que son cuantitativos y cualitativos. Esto significa que incluyen tanto datos numéricos, como calificaciones, tasas de graduación y asistencia, como información cualitativa que describe aspectos como la calidad del aprendizaje, la metodología pedagógica y el entorno escolar³⁰.

Otra característica importante de los datos educativos es su periodicidad. Los datos se recopilan y actualizan regularmente para reflejar el progreso de los estudiantes y las tendencias en el sistema educativo a lo largo del tiempo. Esto permite realizar un seguimiento de los resultados a lo largo de las distintas etapas educativas y evaluar la eficacia de las políticas y programas educativos implementados.

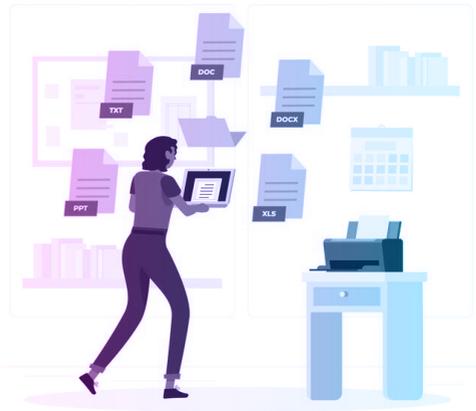
Los datos educativos también son contextualizados, lo que significa que deben estar vinculados a información relevante sobre la ubicación geográfica, el nivel educativo y otros factores que puedan influir en la interpretación de los resultados. Esta contextualización es esencial para comprender mejor las disparidades y desafíos educativos en diferentes regiones o grupos demográficos.

Además, la confidencialidad y la privacidad son preocupaciones fundamentales en la recopilación y el manejo de datos educativos, ya que involucran información personal de estudiantes y educadores. Se deben implementar medidas de seguridad y anonimización para proteger la información sensible de manera adecuada.

En última instancia, los datos educativos son utilizados para la toma de decisiones en el ámbito educativo. Permiten a los responsables de políticas y a los educadores evaluar el desempeño de los estudiantes, identificar áreas de mejora, adaptar estrategias pedagógicas y diseñar políticas educativas efectivas que promuevan un aprendizaje de calidad y la igualdad de oportunidades en el sistema educativo.



Por otro lado; Los datos sociales constituyen un conjunto diverso de información que se relaciona con varios aspectos de la sociedad y la población en general. Una de sus características más destacadas es su diversidad, ya que abarcan una amplia gama de variables, como edad, género, etnia, ingresos, educación, ocupación y más. Esta diversidad refleja la complejidad de la sociedad y permite un análisis detallado de sus múltiples dimensiones, lo que resulta esencial para comprender plenamente la dinámica social³¹.



Además, los datos sociales son inherentemente dinámicos, ya que evolucionan con el tiempo debido a cambios demográficos, económicos, culturales y políticos. La capacidad de mantener los datos actualizados es crucial para capturar tendencias y cambios en la sociedad a lo largo del tiempo y, por lo tanto, para tomar decisiones informadas.

Estos datos son multifacéticos, lo que significa que pueden abordar una amplia variedad de áreas, desde la salud y la educación hasta el empleo y la criminalidad. Esta versatilidad permite un enfoque integral al analizar los problemas y desafíos sociales, permitiendo una comprensión más completa de los factores interconectados que influyen en la sociedad.

Los datos sociales también deben considerar el contexto en



el que se recopilaron, incluyendo factores geográficos, temporales y socioculturales. Este contexto es fundamental para interpretar adecuadamente los datos y comprender cómo influyen en la sociedad en diferentes situaciones y ubicaciones.

La sensibilidad es otra característica clave de los datos sociales, ya que a menudo involucran información personal y sensible, como ingresos, salud mental o antecedentes criminales. Por lo tanto, es esencial proteger la privacidad de los individuos involucrados y garantizar la confidencialidad de los datos.

Finalmente, los datos sociales desempeñan un papel crucial en la formulación de políticas públicas y la toma de decisiones. Ayudan a identificar desigualdades, evaluar la efectividad de programas sociales y orientar la asignación de recursos para abordar problemas específicos en la sociedad, contribuyendo así al mejoramiento del bienestar social y la calidad de vida de las personas.

Fuentes Confiables de Datos Educativos

Fuente	Descripción
UNESCO	La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura recopila y publica datos educativos a nivel mundial.

<p>OECD (Organización para la Cooperación y el Desarrollo Económicos)</p>	<p>La OECD proporciona datos educativos comparativos de sus países miembros.</p>
<p>Instituto Nacional de Estadística y Geografía (INEGI)</p>	<p>En México, el INEGI recopila datos educativos nacionales y regionales.</p>
<p>National Center for Education Statistics (NCES)</p>	<p>En los Estados Unidos, el NCES es una fuente confiable de datos educativos.</p>
<p>Eurostat</p>	<p>La Oficina de Estadísticas de la Unión Europea ofrece datos educativos para países de la UE.</p>
<p>Banco Mundial</p>	<p>El Banco Mundial publica datos relacionados con la educación a nivel global.</p>
<p>Ministerios de Educación</p>	<p>Los ministerios o departamentos de educación de los países suelen proporcionar datos educativos nacionales.</p>
<p>Investigaciones académicas y estudios</p>	<p>Estudios y publicaciones académicas realizadas por instituciones educativas y centros de investigación.</p>

Fuentes Confiables de Datos Sociales

Fuente	Descripción
Instituto Nacional de Estadística y Geografía (INEGI)	El INEGI en México es una fuente confiable de datos demográficos, económicos y sociales que incluye censos, encuestas y estudios diversos.
Instituto Nacional de Estadísticas (INE)	En España, el INE proporciona datos estadísticos oficiales sobre la población, el empleo, la vivienda, la salud y otros aspectos sociales.
Departamento de Censo de los Estados Unidos (US Census Bureau)	El US Census Bureau recopila y publica datos demográficos y sociales sobre la población de los Estados Unidos, incluyendo el censo decenal.
Eurostat	Eurostat, la Oficina de Estadísticas de la Unión Europea, compila y publica datos sociales para los países miembros de la UE.
Organización Mundial de la Salud (OMS)	La OMS es una fuente clave de datos relacionados con la salud a nivel global, incluyendo estadísticas sobre enfermedades, esperanza de vida y más.

Banco Mundial	El Banco Mundial publica datos relacionados con la pobreza, la desigualdad, la economía y otros indicadores sociales en todo el mundo.
Pew Research Center	El Pew Research Center realiza investigaciones sobre temas sociales, políticos y culturales, y publica informes basados en datos sólidos.
Encuestas nacionales	Las encuestas nacionales, como el Barómetro Global de Gallup y la Encuesta de la Comunidad Europea sobre Ingresos y Condiciones de Vida (EU-SILC), recopilan datos sobre la opinión pública y la calidad de vida en diferentes países.
Oficina del Censo de Australia (ABS)	La ABS proporciona datos sociales y económicos para Australia, incluyendo población, empleo, educación, vivienda y más.
Statistics Canada	Statistics Canada es la agencia gubernamental que recopila y publica datos sociales y económicos para Canadá, incluyendo el censo de población.

Office for National Statistics (ONS)	En el Reino Unido, el ONS es la fuente oficial de datos demográficos, económicos y sociales, incluyendo el índice de precios al consumidor y encuestas sociales.
Instituto Brasileiro de Geografia e Estatística (IBGE)	El IBGE en Brasil es responsable de recopilar datos demográficos y económicos, incluyendo el censo de población y encuestas sobre empleo y vivienda.
Estadísticas de la Unión Africana	La Unión Africana recopila y publica datos socioeconómicos para los países africanos, abarcando áreas como salud, educación, economía y desarrollo.
Instituto Nacional de Estadística y Censos (INEC) de Ecuador	El INEC en Ecuador proporciona datos oficiales sobre población, economía, empleo y otros aspectos sociales relevantes para el país.
Departamento de Estadísticas de Nueva Zelanda (Stats NZ)	Stats NZ recopila y publica datos estadísticos sobre la población, la economía y el medio ambiente en Nueva Zelanda.

<p>Instituto Nacional de Estadística y Censos (INEC) de Costa Rica</p>	<p>El INEC en Costa Rica es la fuente oficial de datos estadísticos sobre la población, la economía y la sociedad costarricense.</p>
<p>Instituto Nacional de Estadística y Censos (INEC) de Colombia</p>	<p>El INEC en Colombia proporciona datos confiables sobre población, vivienda, economía y otros aspectos sociales clave para el país.</p>

Procesos de Preprocesamiento y Limpieza de Datos

El proceso de preprocesamiento y limpieza de datos es una etapa esencial en el análisis de datos, la minería de datos y el aprendizaje automático. A medida que se acumulan grandes cantidades de datos, es crucial garantizar que estos sean confiables y adecuados para su uso en análisis posteriores. Se profundizará en los procesos de preprocesamiento y limpieza de datos³²⁻³⁵:

- **Recopilación de Datos:** El primer paso en cualquier proyecto de datos es recopilar los datos relevantes de múltiples fuentes. Estos datos pueden provenir de bases de datos, archivos, sensores, aplicaciones web o cualquier otra fuente de información. Es fundamental asegurarse de que los datos se recojan de manera precisa y completa, ya que cualquier error en esta etapa puede propagarse en todo el proceso.

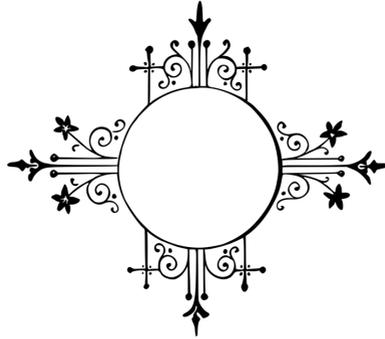
- **Integración de Datos:** En muchos casos, los datos provienen de múltiples fuentes y se deben combinar para formar un conjunto de datos coherente. Durante la integración de datos, se deben abordar problemas como la duplicación de registros, la resolución de conflictos de nombres y la conversión de formatos de datos incompatibles.
- **Lidiar con Valores Faltantes:** Los datos suelen contener valores faltantes, lo que puede deberse a errores de entrada, problemas técnicos o simplemente a la falta de información. Es importante decidir cómo tratar los valores faltantes, ya sea eliminándolos, imputándolos con valores estimados o aplicando métodos más avanzados de imputación.
- **Detección y Tratamiento de Valores Atípicos:** Los valores atípicos, o outliers, son observaciones que se desvían significativamente de la mayoría de los datos. Estos pueden ser errores de medición o eventos inusuales. La identificación y el tratamiento de valores atípicos son esenciales para evitar que distorsionen los resultados del análisis.
- **Normalización y Estandarización:** En algunos casos, es necesario normalizar o estandarizar los datos para que tengan una escala uniforme. Esto es especialmente importante cuando se utilizan algoritmos sensibles a las diferencias de escala, como las redes neuronales o el análisis de componentes principales (PCA).
- **Codificación de Variables Categóricas:** Las variables categóricas deben ser codificadas en una forma numérica para que los algoritmos de aprendizaje automático puedan trabajar con ellas. Se pueden aplicar técnicas

como la codificación one-hot para convertir categorías en variables binarias.

- **Reducción de Dimensionalidad:** En conjuntos de datos de alta dimensionalidad, la reducción de dimensionalidad puede ser beneficiosa para eliminar características irrelevantes o redundantes y mejorar la eficiencia de los modelos.
- **Validación de Datos y Verificación de Consistencia:** Se deben establecer reglas de negocio y realizar validaciones para verificar la coherencia y la calidad de los datos. Esto incluye verificar que los datos cumplan con ciertas condiciones o restricciones específicas del dominio.
- **Documentación y Registro:** Es importante llevar un registro detallado de todas las acciones realizadas durante el preprocesamiento y la limpieza de datos. Esto facilita la reproducibilidad y la trazabilidad de los pasos realizados.
- **Evaluación Continua:** El preprocesamiento y la limpieza de datos son procesos iterativos. A medida que se avanza en el análisis de datos, es posible que se identifiquen nuevos problemas o se requieran ajustes adicionales en el preprocesamiento. Por lo tanto, es importante mantener una evaluación continua de la calidad de los datos.

CAPÍTULO IV

ANÁLISIS DESCRIPTIVO Y CORRELACIONAL DE
DATOS



CAPÍTULO IV.

ANÁLISIS DESCRIPTIVO Y CORRELACIONAL DE DATOS

Técnicas y Herramientas en Contextos Educativos y Sociales

Análisis descriptivo

Estadísticas descriptivas, como la media, la mediana, la desviación estándar y los cuartiles, ayudan a comprender la distribución de las variables numéricas. Se detalla variables para analizar más a profundidad:

- **Años de Educación (Years_of_Education):** Número de años de educación completados.
- **Ingreso del Hogar (Household_Income):** Ingreso anual del hogar en miles.
- **Relación Estudiante-Profesor (Student_to_Teacher_Ratio):** Número de estudiantes por profesor en una escuela.
- **Puntuación de Participación Comunitaria (Communi-**

`ty_Engagement_Score`): Una puntuación que evalúa la participación comunitaria en la educación (sobre 100).

```
np.random.seed(0)

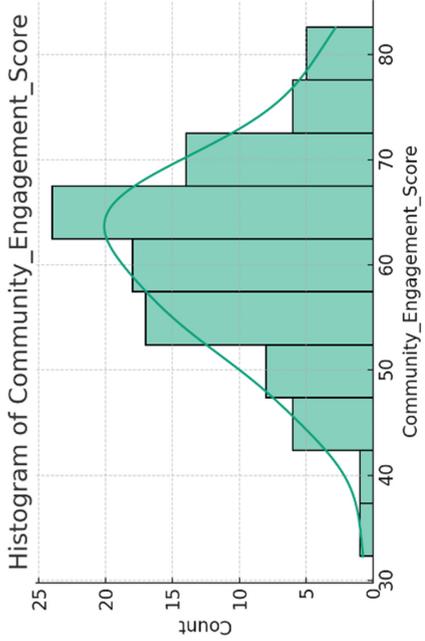
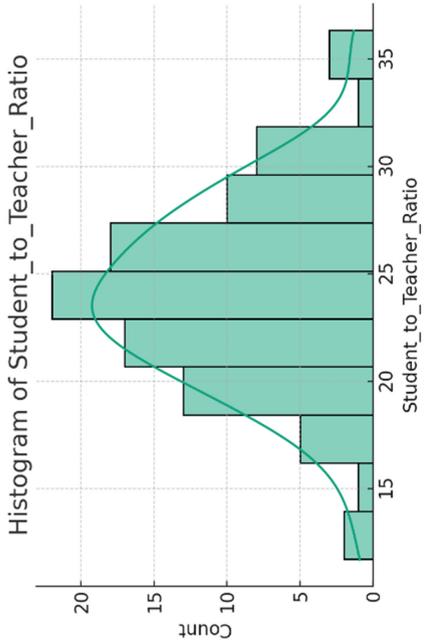
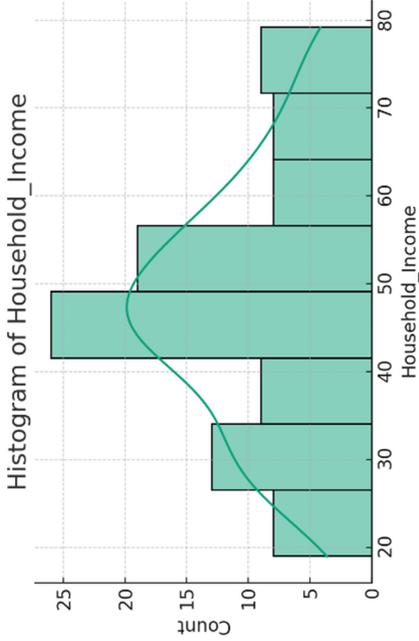
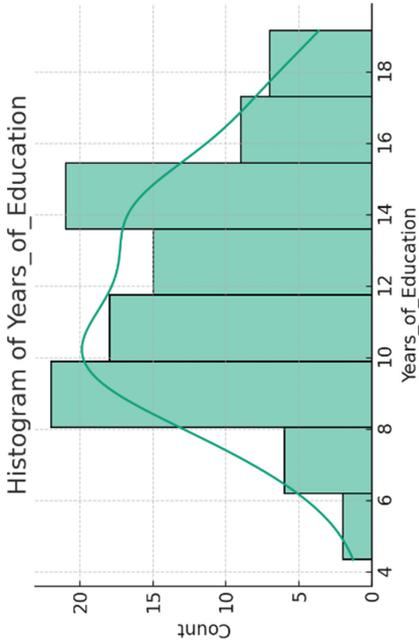
data = np.random.normal(loc=[12, 50, 25, 60], scale=[3, 15, 5, 10], size=(100, 4))
columns = ['Years_of_Education', 'Household_Income',
           'Student_to_Teacher_Ratio', 'Community_Engagement_Score']
educational_df = pd.DataFrame(data, columns=columns)

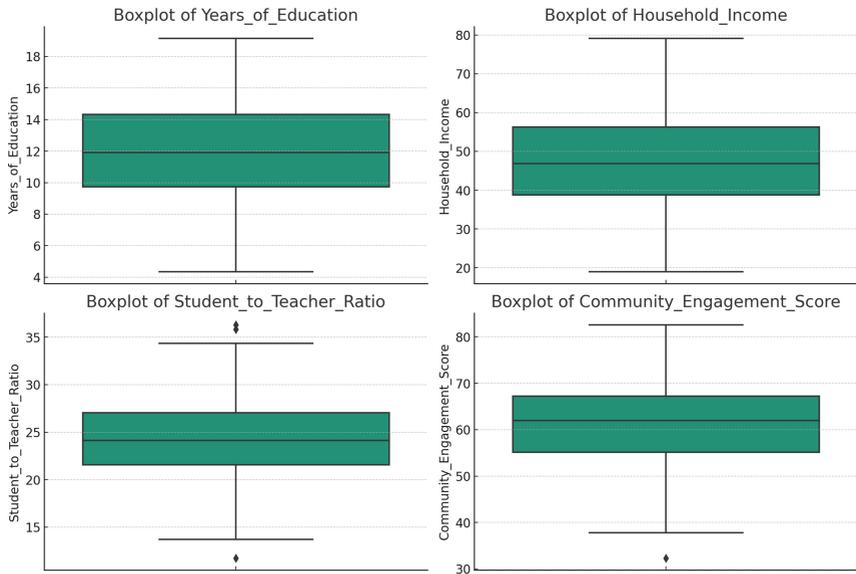
# Applying basic statistical functions to the dataset
educational_desc_stats = educational_df.describe()

# Visualization - Histograms for each variable
plt.figure(figsize=(12, 8))
for i, column in enumerate(educational_df.columns):
    plt.subplot(2, 2, i + 1)
    sns.histplot(educational_df[column], kde=True)
    plt.title(f'Histogram of {column}')
plt.tight_layout()

# Boxplots for each variable
plt.figure(figsize=(12, 8))
for i, column in enumerate(educational_df.columns):
    plt.subplot(2, 2, i + 1)
    sns.boxplot(y=educational_df[column])
    plt.title(f'Boxplot of {column}')
plt.tight_layout()

educational_desc_stats
```





Resultados descriptivos para cada variable:

- **Años de Educación:** En promedio, los individuos han completado unos 12 años de educación, con una desviación estándar de aproximadamente 3,26 años.
- **Ingreso del Hogar:** El ingreso medio del hogar es de aproximadamente 48.323 dólares, con una desviación estándar de 14.660 dólares.
- **Relación Estudiante-Profesor:** En promedio, hay unos 24 estudiantes por profesor, con una desviación estándar de alrededor de 4,48.
- **Puntuación de Participación Comunitaria:** La puntuación media es de aproximadamente 61,42, con una desviación estándar de 9,73.

Matriz de Correlación

Calcula y muestra las correlaciones entre variables numéricas, lo que ayuda a identificar relaciones lineales o asociaciones entre ellas. Se especifica:

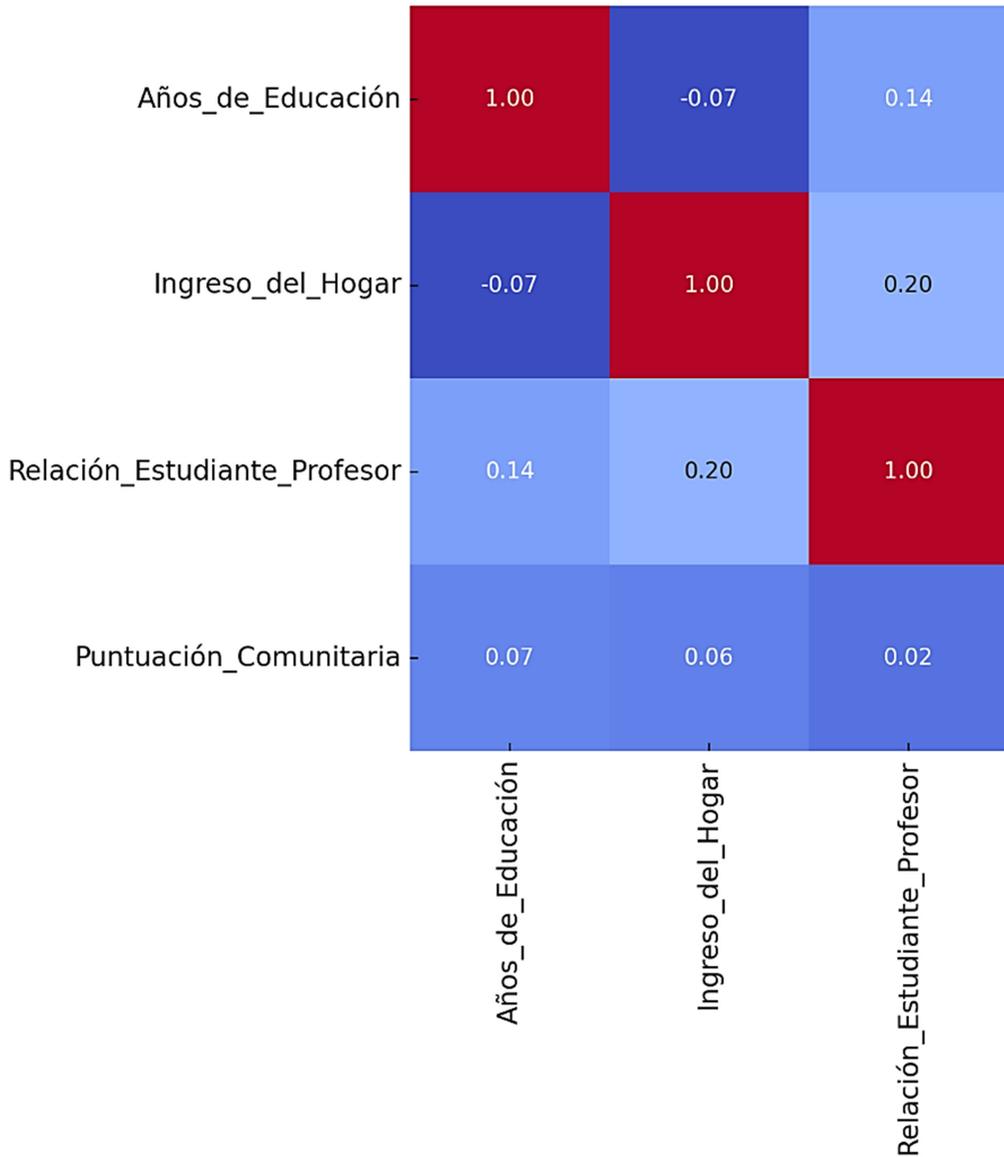
```
# Creando un conjunto de datos ficticio enfocado en
variables educativas y sociales
np.random.seed(0)
data = np.random.normal(loc=[12, 50, 25, 60], scale=[3, 15,
5, 10], size=(100, 4))
columns = ['Años_de_Educación', 'Ingreso_del_
Hogar', 'Relación_Estudiante_Profesor', 'Puntuación_
Comunitaria']
df_educativo_social = pd.DataFrame(data,
columns=columns)

# Calculando la matriz de correlación
matriz_correlacion = df_educativo_social.corr()

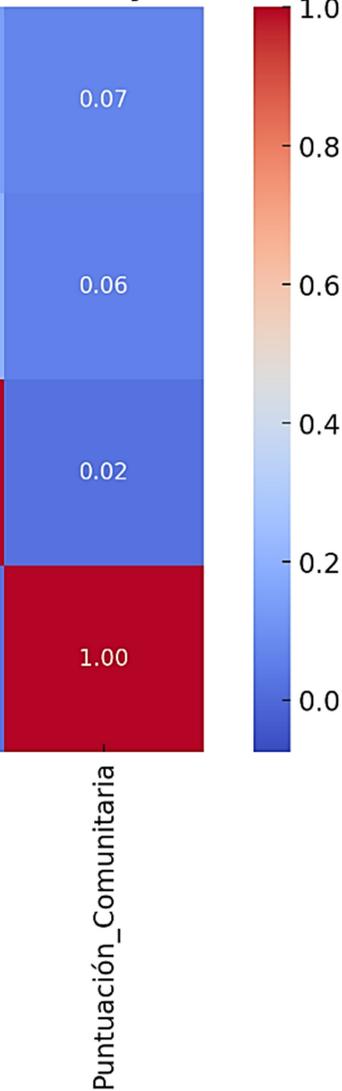
# Visualización - Mapa de calor de la matriz de
correlación
plt.figure(figsize=(8, 6))
sns.heatmap(matriz_correlacion, annot=True,
cmap='coolwarm', fmt=".2f")
plt.title('Matriz de Correlación de Variables Educativas y
Sociales')
plt.show()

matriz_correlacion
```

Matriz de Correlación de Variables Educativas



Actividades y Sociales



El mapa de calor de la matriz de correlación muestra las relaciones entre estas variables. Las correlaciones van de -1 a +1, donde +1 indica una correlación positiva perfecta, -1 indica una correlación negativa perfecta y 0 indica ninguna correlación.

Resultados de la matriz de correlación:

- **Años de Educación e Ingreso del Hogar:** Correlación de -0,0746, indicando una relación negativa muy débil.
- **Años de Educación y Relación Estudiante-Profesor:** Correlación de 0,1398, sugiriendo una relación positiva leve.
- **Años de Educación y Puntuación Comunitaria:** Correlación de 0,0716, una asociación positiva muy débil.
- **Ingreso del Hogar y Relación Estudiante-Profesor:** Correlación de 0,2037, indicando una relación positiva leve.
- **Ingreso del Hogar y Puntuación Comunitaria:** Correlación de 0,0559, muy ligeramente positiva.
- **Relación Estudiante-Profesor y Puntuación Comunitaria:** Correlación de 0,0205, mostrando casi ninguna relación lineal.

Análisis de Valores Atípicos

Identificar y tratar valores atípicos (outliers) que puedan afectar el análisis posterior. Se detalla:

```
# Análisis de Valores Atípicos en un Conjunto de Datos
Educativos Ficticios

# Creación del conjunto de datos
np.random.seed(0)
data = np.random.normal(loc=[15, 60, 30, 70], scale=[5, 20,
10, 15], size=(100, 4))
columns = ['Años_Educación', 'Ingreso_Hogar', 'Ratio_
Estudiante_Profesor', 'Puntuación_Participación']
df_educativo = pd.DataFrame(data, columns=columns)

# Análisis de Valores Atípicos usando diagramas de caja
(Boxplots)
plt.figure(figsize=(12, 8))

for i, column in enumerate(df_educativo.columns):
    plt.subplot(2, 2, i + 1)
    sns.boxplot(data=df_educativo[column])
    plt.title(f'Diagrama de Caja para {column}')

plt.tight_layout()
plt.show()
```

Diagrama de Caja para Ingreso_Hogar

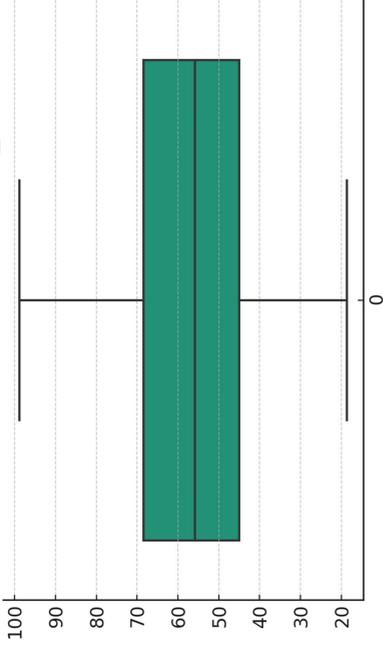


Diagrama de Caja para Puntuación_Participación

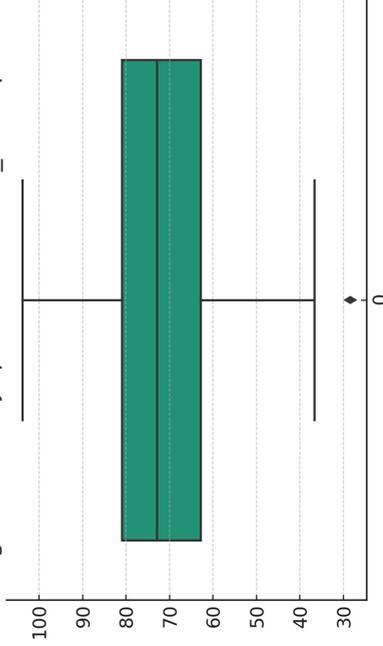


Diagrama de Caja para Años_Educación

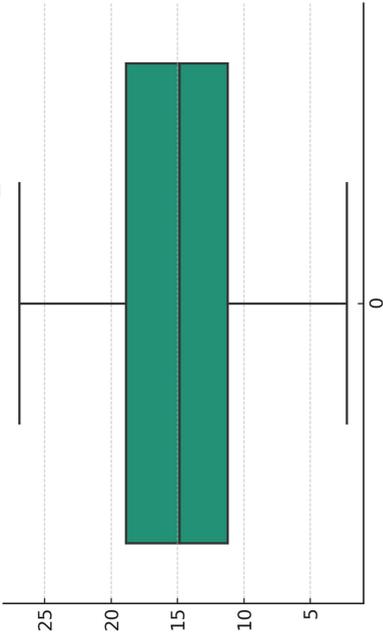
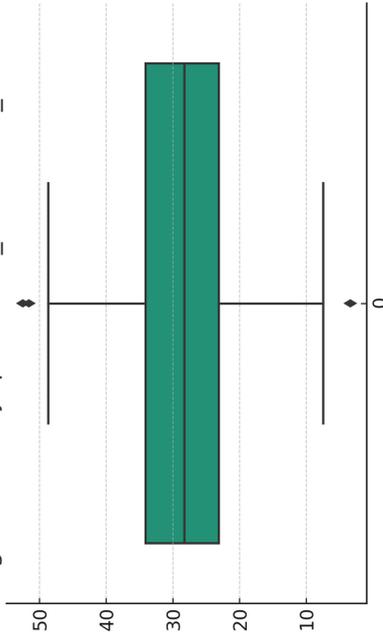


Diagrama de Caja para Ratio_Estudiente_Profesor



En los diagramas de caja generados:

- Los “bigotes” representan el rango intercuartil (IQR), extendiéndose hasta 1,5 veces el IQR desde los cuartiles superior e inferior.
- Los puntos fuera de los bigotes indican valores atípicos, que son aquellos valores que caen fuera del rango de los bigotes.

Gráficos de Tendencia

Gráficos de líneas o series de tiempo para visualizar tendencias a lo largo del tiempo en series temporales. Se detalla:

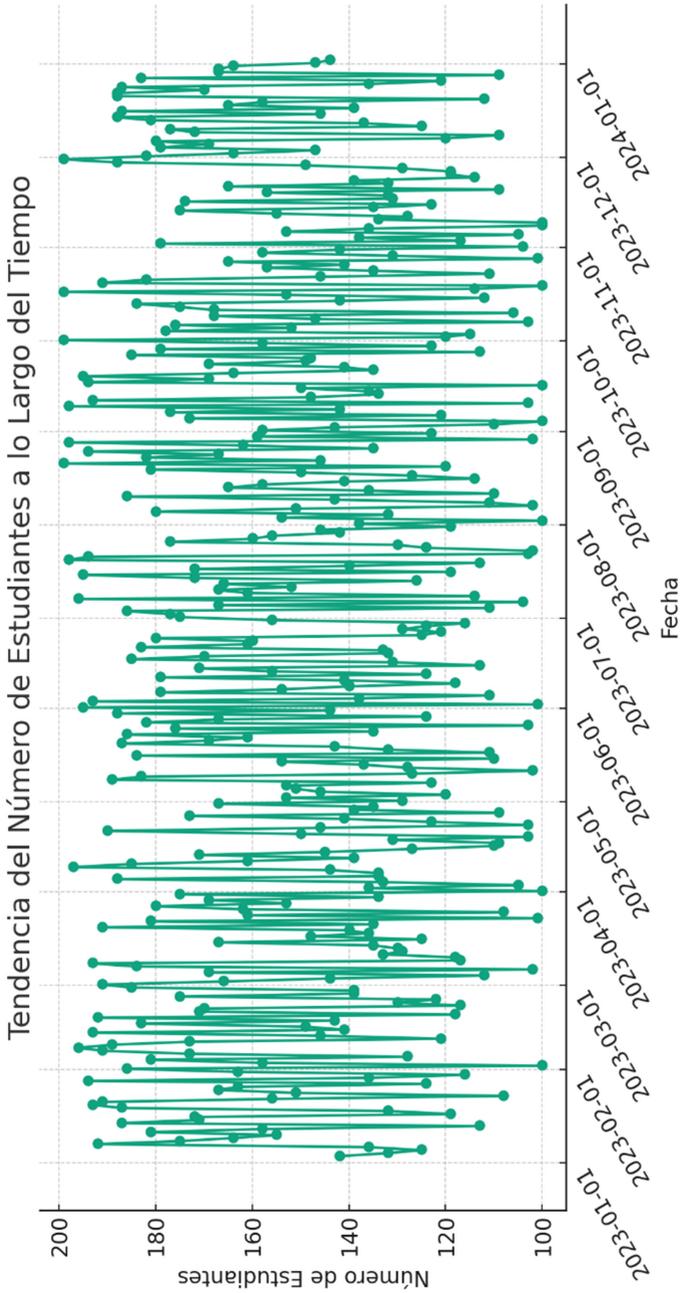
```
import matplotlib.dates as mdates
from datetime import datetime, timedelta

# Creación de un conjunto de datos ficticio para
representar el número de estudiantes a lo largo del
tiempo
np.random.seed(0)
dates = [datetime.now() - timedelta(days=x) for x in
range(0, 365)]
students = np.random.randint(100, 200, size=(365,))

df_students = pd.DataFrame({'Date': dates, 'Number_of_
Students': students})
df_students.sort_values('Date', inplace=True)

# Gráfico de Tendencia
```

```
plt.figure(figsize=(12, 6))
plt.plot(df_students['Date'], df_students['Number_of_
Students'], marker='o', linestyle='-')
plt.title('Tendencia del Número de Estudiantes a lo Largo
del Tiempo')
plt.xlabel('Fecha')
plt.ylabel('Número de Estudiantes')
plt.grid(True)
plt.gca().xaxis.set_major_formatter(mdates.
DateFormatter('%Y-%m-%d'))
plt.gca().xaxis.set_major_locator(mdates.MonthLocator())
plt.gcf().autofmt_xdate() # Rotar las fechas para mejor
visualización
plt.show()
```



El gráfico de tendencia utilizando un conjunto de datos ficticio para representar el número de estudiantes a lo largo del tiempo. Este tipo de gráfico es particularmente útil para visualizar cómo cambia una variable (en este caso, el número de estudiantes) a lo largo de un periodo.

En el gráfico:

- El eje horizontal (X) muestra la fecha.
- El eje vertical (Y) representa el número de estudiantes.
- La línea con marcadores señala la cantidad de estudiantes en cada fecha.

Este gráfico es una herramienta valiosa para identificar patrones, tendencias y fluctuaciones en los datos relacionados con estudiantes, como la inscripción o asistencia a lo largo del tiempo. Por ejemplo, se podría utilizar para examinar cómo varía la inscripción estudiantil a lo largo de un año académico o para identificar periodos de alta o baja asistencia.

Análisis de Variables Categóricas:

Tablas de frecuencia y gráficos de barras para analizar y visualizar variables categóricas. Se detalla:

```
# Creación de un conjunto de datos ficticio con variables  
categóricas  
np.random.seed(0)  
n = 100  
education_levels = ['Primaria', 'Secundaria', 'Universitaria',  
'Posgrado']
```

```
income_brackets = ['Bajo', 'Medio', 'Alto']
regions = ['Norte', 'Sur', 'Este', 'Oeste']

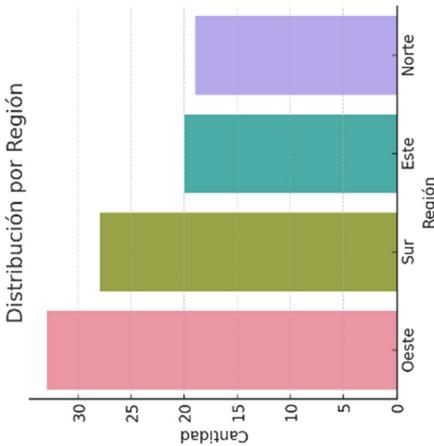
df_categorical = pd.DataFrame({
    'Nivel_Educativo': np.random.choice(education_levels,
n),
    'Rango_Ingresos': np.random.choice(income_brackets,
n),
    'Región': np.random.choice(regions, n)
})

# Análisis de frecuencia de cada categoría
education_count = df_categorical['Nivel_Educativo'].value_counts()
income_count = df_categorical['Rango_Ingresos'].value_counts()
region_count = df_categorical['Región'].value_counts()

# Visualización - Gráficos de barras para cada variable categórica
plt.figure(figsize=(15, 5))

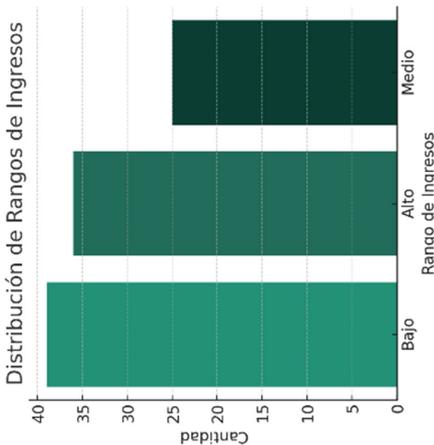
plt.subplot(1, 3, 1)
sns.barplot(x=education_count.index, y=education_count.values)
plt.title('Distribución del Nivel Educativo')
plt.xlabel('Nivel Educativo')
plt.ylabel('Cantidad')
plt.subplot(1, 3, 2)
```

```
sns.barplot(x=income_count.index, y=income_count.  
values)  
plt.title('Distribución de Rangos de Ingresos')  
plt.xlabel('Rango de Ingresos')  
plt.ylabel('Cantidad')  
  
plt.subplot(1, 3, 3)  
sns.barplot(x=region_count.index, y=region_count.  
values)  
plt.title('Distribución por Región')  
plt.xlabel('Región')  
plt.ylabel('Cantidad')  
  
plt.tight_layout()  
plt.show()
```

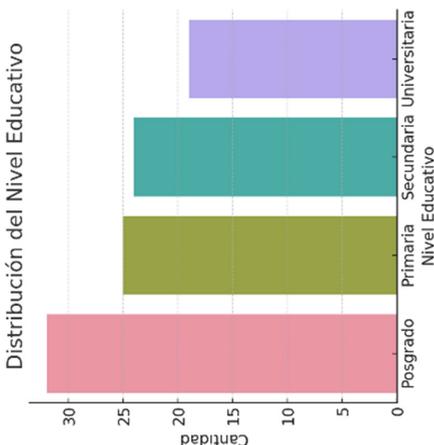


Las variables categóricas analizadas son:

- **Nivel Educativo:** Clasificado en 'Primaria', 'Secundaria', 'Universitaria', 'Posgrado'.
- **Rango de Ingresos:** Clasificado en 'Bajo', 'Medio', 'Alto'.
- **Región:** Clasificado en 'Norte', 'Sur', 'Este', 'Oeste'.



Para cada variable categórica, se ha generado un gráfico de barras que muestra la distribución de frecuencias en cada categoría:



- **Distribución del Nivel Educativo:** Muestra la cantidad de individuos en cada nivel educativo.
- **Distribución de Rangos de Ingresos:** Ilustra la cantidad de individuos en cada rango de ingresos.
- **Distribución por Región:** Representa la cantidad de individuos en cada región geográfica.

Diagramas de Pareto

Identificar las categorías más importantes en una variable categórica ordenando los valores por frecuencia. Se detalla:

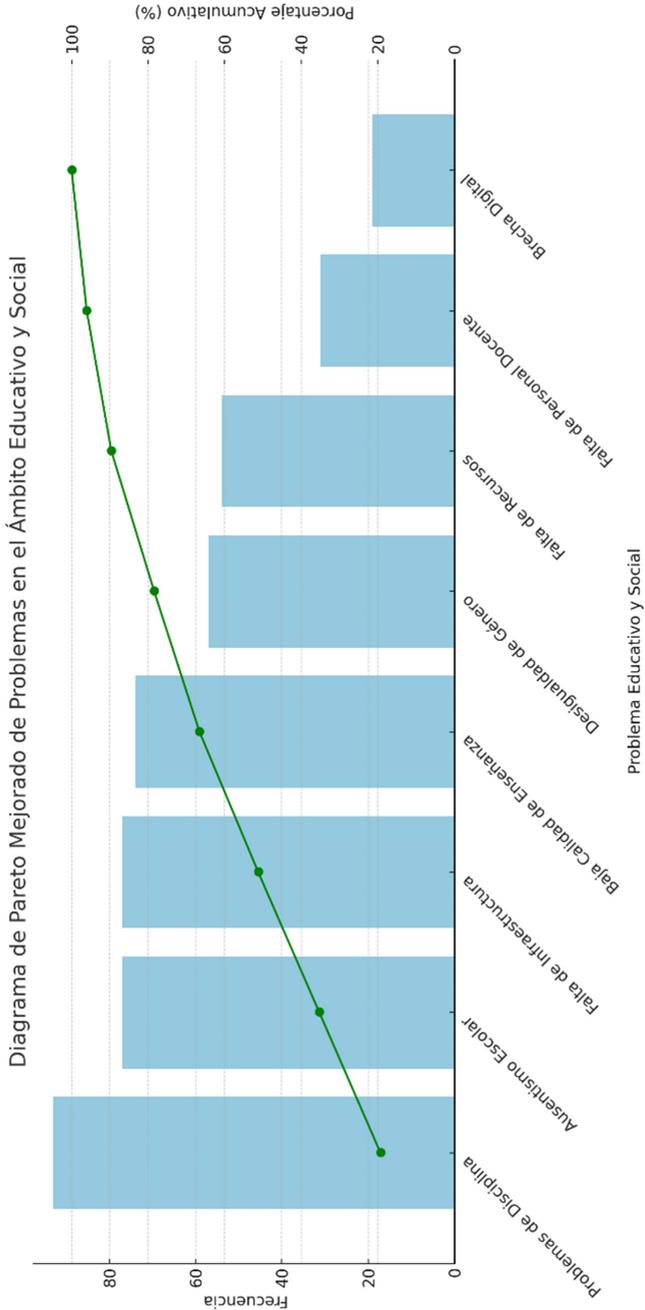
```
# Rotar las etiquetas

plt.figure(figsize=(14, 7))
ax = sns.barplot(x='Problema', y='Frecuencia', data=df_
issues_sorted, color='skyblue')
ax.set_xticklabels(ax.get_xticklabels(), rotation=45,
ha="right") # Rotar las etiquetas

# Línea de acumulación de porcentaje
ax2 = ax.twinx()
ax2.plot(df_issues_sorted['Problema'], df_issues_
sorted['Cumulative_Percentage'], color='green',
marker='o', ms=7)
ax2.set_ylim(0, 110)

# Etiquetas y título
ax.set_xlabel('Problema Educativo y Social')
ax.set_ylabel('Frecuencia')
ax2.set_ylabel('Porcentaje Acumulativo (%)')
plt.title('Diagrama de Pareto Mejorado de Problemas en el
Ámbito Educativo y Social')

plt.tight_layout()
plt.show()
```



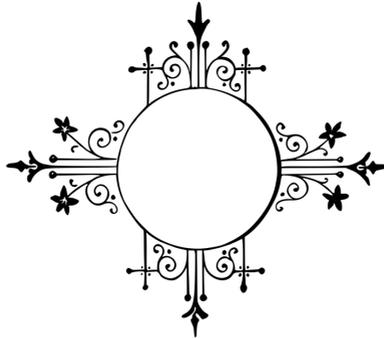
En este diagrama:

- **Eje horizontal (X):** Representa diferentes problemas en el ámbito educativo y social, como 'Falta de Recursos', 'Desigualdad de Género', 'Baja Calidad de Enseñanza', entre otros.
- **Eje vertical izquierdo (Y):** Muestra la frecuencia de cada problema.
- **Línea verde y eje vertical derecho:** Representan el porcentaje acumulativo de los problemas, ayudando a identificar qué problemas contribuyen más significativamente al total.

El diagrama de Pareto es particularmente útil en la gestión y planificación educativa y social, ya que ayuda a focalizar los esfuerzos y recursos en las áreas que tendrán el mayor impacto en la mejora de la situación educativa y social.

CAPÍTULO V

VISUALIZACIÓN AVANZADA Y ANÁLISIS
EXPLORATORIO DE DATOS EDUCATIVOS



CAPÍTULO V.

VISUALIZACIÓN AVANZADA Y ANÁLISIS EXPLORATORIO DE DATOS EDUCATIVOS

Diseño de Dashboards Educativos

El diseño de dashboards educativos es una tarea que implica integrar datos relevantes de manera que sean fácilmente interpretables y útiles para la toma de decisiones. Para diseñar dashboards educativos efectivos³⁶⁻³⁸:

- **Identificación de Usuarios y Necesidades:** Determinar quiénes serán los usuarios principales del dashboard (por ejemplo, administradores, maestros, estudiantes, o padres) y qué información necesitan. Cada grupo de usuarios puede requerir diferentes tipos de datos.
- **Selección de Métricas Clave:** Elegir métricas e indicadores que sean relevantes para los objetivos del dashboard. En un contexto educativo, esto podría incluir tasas de graduación, resultados de exámenes, asistencia, participación en actividades extracurriculares, entre otros.

- **Fuente de Datos:** Asegurar que adjudica acceso a datos confiables y actualizados. Los dashboards suelen alimentarse de sistemas de información escolar, bases de datos de rendimiento académico, encuestas, etc.
- **Diseño Visual:** El diseño debe ser intuitivo y fácil de navegar. Usa colores y gráficos de manera efectiva para resaltar información clave. Evita sobrecargar el dashboard con demasiada información; el enfoque debe estar en la claridad y la facilidad de interpretación.
- **Interactividad y Personalización:** Considerar incluir elementos interactivos, como filtros y opciones de desglose, para permitir a los usuarios explorar los datos de manera más detallada. La personalización puede incluir la capacidad de adaptar el dashboard a las necesidades específicas de diferentes usuarios o grupos.
- **Integración y Accesibilidad:** Asegurar de que el dashboard pueda integrarse con otros sistemas y plataformas utilizados en el entorno educativo. Debe ser accesible desde diferentes dispositivos, incluidos teléfonos móviles y tablets.
- **Actualización de Datos:** Planificar cómo se actualizarán los datos en el dashboard. Algunos dashboards pueden requerir actualizaciones en tiempo real, mientras que otros pueden actualizarse diaria, semanal o mensualmente.
- **Pruebas y Feedback:** Antes de lanzar el dashboard de manera oficial, realiza pruebas con usuarios reales y recoge sus comentarios para hacer ajustes necesarios.
- **Capacitación y Soporte:** Ofrece capacitación y soporte

a los usuarios para asegurar que entienden cómo utilizar el dashboard y sacar el máximo provecho de la información proporcionada.

- **Evaluación y Mejora Continua:** Los dashboards deben ser evaluados y mejorados continuamente para asegurarse de que siguen siendo relevantes y útiles para los usuarios.

Se detalla un ejemplo:

```
import dash
import dash_core_components as dcc
import dash_html_components as html
import plotly.graph_objs as go
import pandas as pd
import numpy as np

# Simulación de datos educativos
np.random.seed(0)
data = {
    'Clase': ['Matemáticas', 'Ciencias', 'Historia', 'Arte',
             'Educación Física'],
    'Asistencia': np.random.randint(70, 100, 5),
    'Calificaciones': np.random.randint(1, 100, 100)
}

df = pd.DataFrame(data)

# Creación de la aplicación Dash
app = dash.Dash(__name__)
```

```
# Definición del layout del dashboard
app.layout = html.Div(children=[
    html.H1(children='Dashboard Educativo'),

    dcc.Graph(
        id='asistencia-clases',
        figure={
            'data': [
                go.Bar(
                    x=df['Clase'],
                    y=df['Asistencia']
                )
            ],
            'layout': go.Layout(
                title='Asistencia por Clase'
            )
        }
    ),

    dcc.Graph(
        id='distribucion-calificaciones',
        figure={
            'data': [
                go.Histogram(
                    x=df['Calificaciones']
                )
            ],
            'layout': go.Layout(
                title='Distribución de Calificaciones'
```

```
    )  
    }  
    )  
])  
  
# Correr el servidor  
if __name__ == '__main__':  
    app.run_server(debug=True)
```

Análisis de Componentes Principales (PCA):

El Análisis de Componentes Principales (PCA, por sus siglas en inglés) es una técnica estadística de reducción de dimensiones que se utiliza ampliamente en diversos campos, incluidos los entornos educativos y sociales. En Python, PCA se puede implementar utilizando bibliotecas como `scikit-learn`^{39,40}. Se detalla cómo aplicar PCA a un conjunto de datos ficticio relacionado con el entorno educativo y social, y se administrará los gráficos correspondientes para visualizar los resultados.

Primero, crearemos un conjunto de datos ficticio que podría representar, por ejemplo, las puntuaciones en diferentes pruebas o encuestas en un entorno educativo y social. Luego, aplicaremos PCA a estos datos y visualizaremos los componentes principales.

Vamos a proceder con el análisis y la generación del gráfico.

```
# Importando librerías necesarias
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

# Creando un conjunto de datos simulado
np.random.seed(0)
n_students = 200
hours_study = np.random.normal(3.5, 1.0, n_students) #
horas de estudio
extracurricular = np.random.normal(5, 1.5, n_students) #
participación en actividades
attendance = np.random.normal(90, 10, n_students) #
asistencia en %
grades = np.random.normal(75, 10, n_students) #
puntuación de los exámenes

# Creando un DataFrame
students_df = pd.DataFrame({
    'Hours of Study': hours_study,
    'Extracurricular Activities': extracurricular,
    'Attendance (%)': attendance,
    'Grades (%)': grades
})

# Estandarizando los datos
scaler = StandardScaler()
```

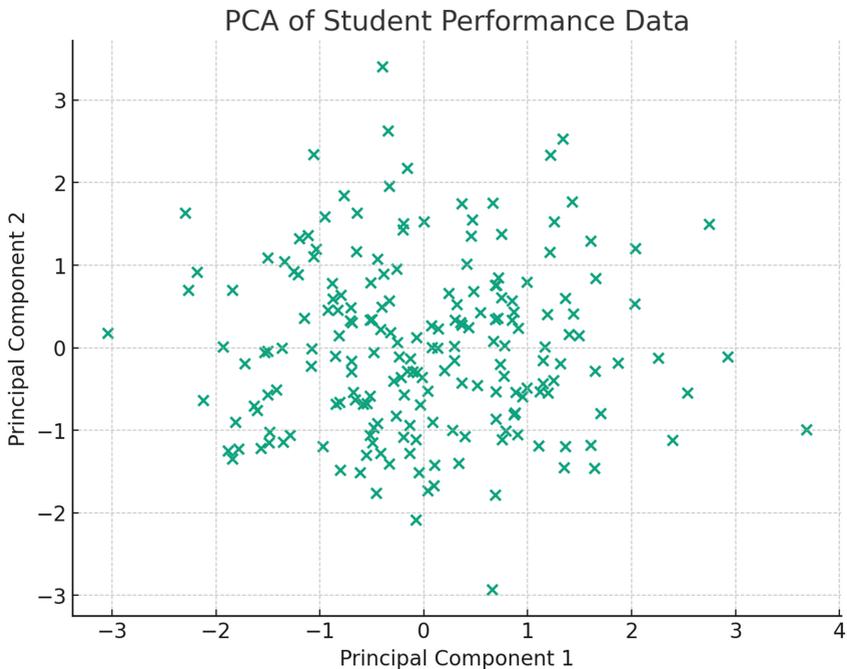
```
students_scaled = scaler.fit_transform(students_df)

# Aplicando PCA
pca = PCA(n_components=2)
students_pca = pca.fit_transform(students_scaled)

# Creando un DataFrame para los resultados de PCA
pca_df = pd.DataFrame(data=students_pca,
                      columns=['Principal Component 1', 'Principal Component 2'])

# Graficando los resultados de PCA
plt.figure(figsize=(8, 6))
plt.scatter(pca_df['Principal Component 1'], pca_df['Principal Component 2'])
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('PCA of Student Performance Data')
plt.show()

# Creando una tabla de varianza explicada
explained_variance = pca.explained_variance_ratio_
explained_variance_df = pd.DataFrame(explained_variance,
                                     index=['PC1', 'PC2'], columns=['Explained Variance'])
explained_variance_df
```



Cada punto en el gráfico representa un estudiante. Los datos se han reducido a dos componentes principales. Este gráfico ayuda a visualizar la relación y distribución de los estudiantes en términos de sus características académicas, como horas de estudio, participación en actividades extracurriculares, asistencia y calificaciones.

Tabla de Varianza Explicada:

- PC1: El primer componente principal explica aproximadamente el 30.52% de la varianza en los datos.
- PC2: El segundo componente principal explica aproximadamente el 26.17% de la varianza.

Estos resultados muestran que, aunque hemos reducido la dimensionalidad de los datos, todavía podemos capturar más de la mitad de la variabilidad total a través de los dos primeros componentes principales. Esto es útil en análisis educativos y sociales para identificar patrones clave y relaciones entre diferentes variables.

Análisis de Clúster (Clustering)

El análisis de clúster, también conocido como clustering, es una técnica de aprendizaje automático no supervisado que se utiliza para agrupar datos en conjuntos o clústeres. En el campo educativo, esta técnica puede ser muy útil para identificar patrones y agrupaciones entre estudiantes, profesores, instituciones, o incluso en datos de rendimiento y comportamiento. Cómo podrías llevar a cabo un análisis de clúster en el campo educativo:

- **Definir el Objetivo:** Primero, es fundamental definir el objetivo del análisis. Por ejemplo, podrías estar interesado en agrupar a los estudiantes según su rendimiento académico, estilos de aprendizaje, o comportamientos en el aula.
- **Recopilar y Preparar Datos:** Los datos pueden provenir de diversas fuentes como registros académicos, encuestas, evaluaciones de desempeño, etc. Es importante limpiar y preprocesar estos datos, lo que incluye: Lidar con valores faltantes, Estandarizar o normalizar los datos, Convertir variables categóricas en numéricas si es necesario.
- **Selección de Características:** Determinar qué características (variables) son relevantes para el análisis. Por ejemplo, si el objetivo es agrupar a los estudiantes por

rendimiento, podrías considerar notas, asistencia, participación en clase, etc.

- **Elección del Método de Clustering:** Algunos métodos populares incluyen K-means, Clustering Jerárquico, DBSCAN, etc. La elección depende del tipo de datos y del objetivo del análisis.
- **Determinar el Número de Clústeres:** En métodos como K-means, es necesario definir el número de clústeres. Herramientas como el método del codo o el índice de Silhouette pueden ayudar en esta elección.
- **Ejecutar el Análisis de Clúster:** Aplicar el método de clustering elegido a los datos.
- **Interpretar los Resultados:** Una vez formados los clústeres, es importante interpretarlos y entender qué caracteriza a cada clúster. Esto puede implicar examinar las características medias de los clústeres o los patrones más comunes dentro de cada grupo.
- **Aplicaciones Prácticas:** Los resultados pueden aplicarse para mejorar estrategias de enseñanza, personalizar el aprendizaje, identificar necesidades de apoyo específicas, entre otros.
- **Visualización:** Usar gráficos como diagramas de dispersión, mapas de calor, o dendrogramas para visualizar los clústeres y sus características.
- **Validación y Revisión:** Finalmente, es importante validar los resultados, posiblemente con la ayuda de expertos en educación, y revisar el análisis si es necesario.

Los datos incluirán las siguientes variables para cada es-

tudiante:

- **Notas Promedio:** Promedio de calificaciones en todas las materias.
- **Asistencia:** Porcentaje de asistencia a clases.
- **Participación en Clases Extraescolares:** Número de actividades extraescolares en las que participa el estudiante.
- **Puntuación en Pruebas Estandarizadas:** Puntuaciones en pruebas estandarizadas como matemáticas, ciencias, etc.
- **Nivel de Satisfacción con la Escuela:** Medido a través de encuestas, en una escala de 1 a 5.
- **Tiempo de Estudio Semanal:** Horas dedicadas al estudio fuera de la escuela por semana.

Vamos a generar datos ficticios para 200 estudiantes con estas variables. Luego aplicaremos un análisis de clúster para identificar patrones y agrupaciones entre los estudiantes.

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import seaborn as sns

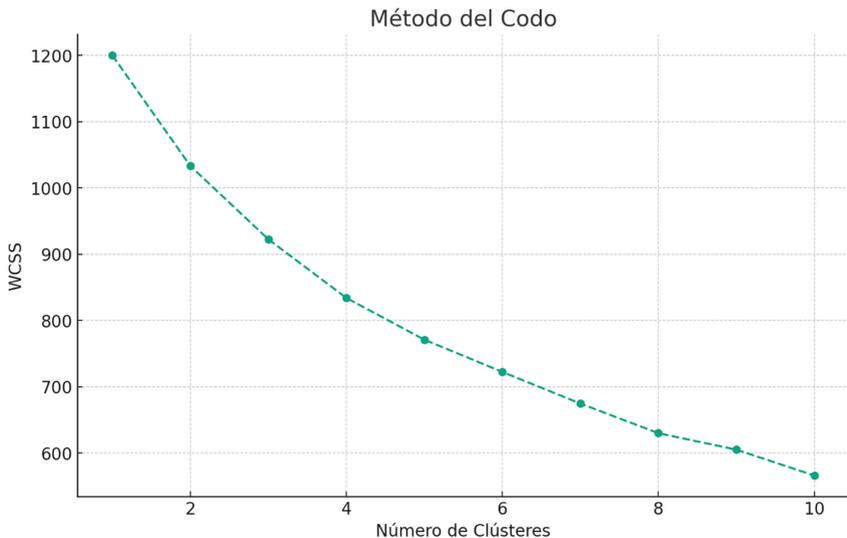
# Generar datos ficticios
```

```
np.random.seed(0)
n_estudiantes = 200
datos = {
    "Notas_Promedio": np.random.uniform(5, 10, n_
estudiantes),
    "Asistencia": np.random.uniform(70, 100, n_
estudiantes),
    "Actividades_Extraescolares": np.random.randint(0, 5,
n_estudiantes),
    "Puntuacion_Pruebas_Estandarizadas": np.random.
uniform(200, 800, n_estudiantes),
    "Satisfaccion_Escuela": np.random.randint(1, 6, n_
estudiantes),
    "Tiempo_Estudio_Semanal": np.random.uniform(0, 20,
n_estudiantes)
}

df = pd.DataFrame(datos)

# Normalizar los datos
scaler = StandardScaler()
df_normalizado = scaler.fit_transform(df)
# Utilizaremos K-means para el análisis de clúster
# Determinar el número óptimo de clústeres
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++',
random_state=42)
    kmeans.fit(df_normalizado)
    wcss.append(kmeans.inertia_)
```

```
# Gráfico del método del codo para determinar el número  
óptimo de clústeres  
plt.figure(figsize=(10, 6))  
plt.plot(range(1, 11), wcss, marker='o', linestyle='--')  
plt.title('Método del Codo')  
plt.xlabel('Número de Clústeres')  
plt.ylabel('WCSS') # WCSS: Within-Cluster Sum of  
Square  
plt.show()
```



El gráfico del método del codo muestra cómo varía la suma de cuadrados dentro del clúster (WCSS) con diferentes números de clústeres. Buscamos un punto donde el descenso en WCSS comienza a disminuir más lentamente, indicando un equilibrio entre el número de clústeres y la cohesión dentro de ellos. Este punto es generalmente considerado

como el número óptimo de clústeres.

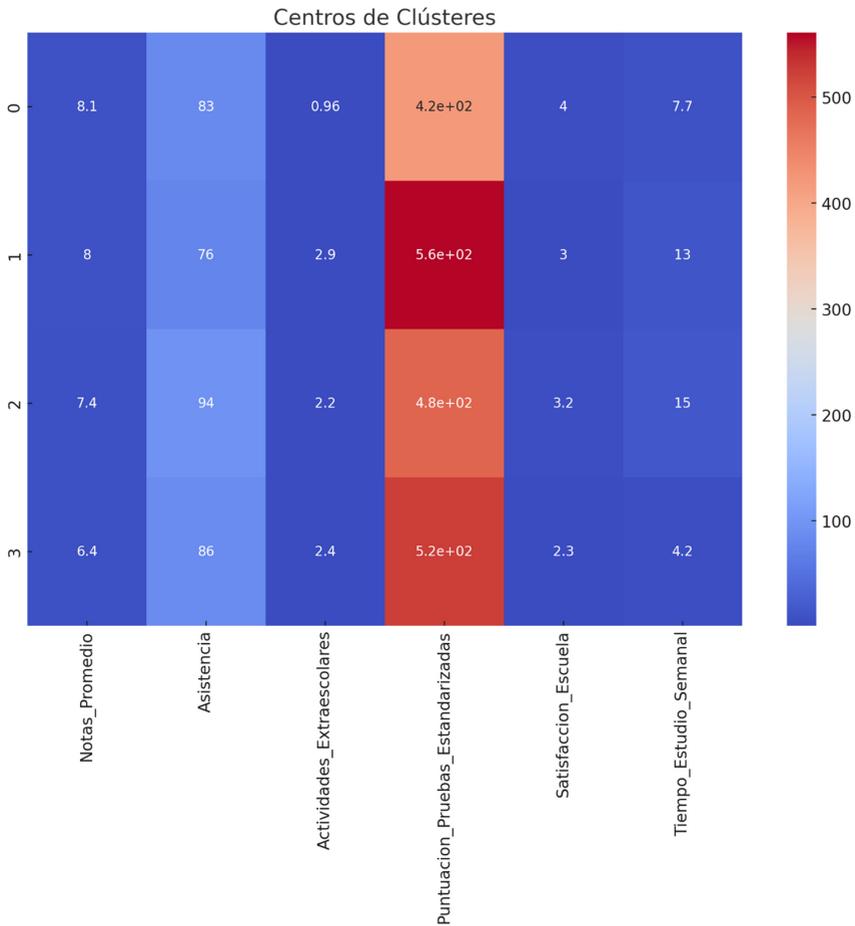
En nuestro gráfico, parece que este punto podría estar alrededor de 3 o 4 clústeres. Para este ejemplo, elegiremos 4 clústeres y procederemos a realizar el análisis de clúster con K-means. Luego, interpretaremos los resultados para entender cómo se agrupan los estudiantes según las variables consideradas.

```
# Ejecutar K-means con 4 clústeres
kmeans = KMeans(n_clusters=4, init='k-means++',
random_state=42)
clústeres = kmeans.fit_predict(df_normalizado)

# Añadir la asignación de clústeres al DataFrame original
df['Clúster'] = clústeres

# Analizar los clústeres formados
centros_clústeres = kmeans.cluster_centers_
df_clústeres_centros = pd.DataFrame(scaler.inverse_
transform(centros_clústeres),
                                columns=df.columns[:-1])
# Visualización de los centros de los clústeres
plt.figure(figsize=(12, 8))
sns.heatmap(df_clústeres_centros, annot=True,
cmap="coolwarm")
plt.title("Centros de Clústeres")
plt.show()

df_clústeres_centros, df['Clúster'].value_counts()
```



Hemos realizado un análisis de clúster K-means en el conjunto de datos ficticio y dividido a los estudiantes en 4 grupos distintos. Se detalla el comportamiento de cada Clúster:

- **Clúster 0 (55 estudiantes):** Este grupo se caracteriza por tener notas promedio altas, asistencia moderada, baja participación en actividades extraescolares, puntuaciones bajas en pruebas estandarizadas, alta satisfacción

con la escuela y tiempo de estudio semanal moderado.

- **Clúster 1 (49 estudiantes):** Estudiantes con notas promedio ligeramente altas, baja asistencia, participación moderada en actividades extraescolares, puntuaciones más altas en pruebas estandarizadas, satisfacción moderada con la escuela y alto tiempo de estudio semanal.
- **Clúster 2 (45 estudiantes):** Este grupo tiene notas promedio moderadas, muy alta asistencia, participación moderada en actividades extraescolares, puntuaciones medias en pruebas estandarizadas, satisfacción moderada con la escuela y muy alto tiempo de estudio semanal.
- **Clúster 3 (51 estudiantes):** Estudiantes con las notas promedio más bajas, asistencia moderada, participación moderada en actividades extraescolares, puntuaciones medias en pruebas estandarizadas, baja satisfacción con la escuela y bajo tiempo de estudio semanal.

¿Qué hacer en cada clúster?

- **Clúster 0** podría representar a estudiantes consistentemente buenos con enfoque académico pero menos involucrados en actividades extracurriculares.
- **Clúster 1** parece indicar estudiantes que, a pesar de su baja asistencia, logran buen rendimiento académico y dedican mucho tiempo al estudio fuera de la escuela.
- **Clúster 2** puede representar a estudiantes muy comprometidos con la escuela en términos de asistencia y estudio, pero con un rendimiento académico y en prue-

bas estandarizadas solo moderado.

- **Clúster 3** incluye a estudiantes que podrían estar luchando académicamente, con menor satisfacción escolar y menor tiempo dedicado al estudio.

Estos insights pueden ser útiles para diseñar estrategias de intervención y apoyo educativo personalizadas, adaptadas a las necesidades y características de cada grupo de estudiantes.

Mapas de Calor (Heatmaps)

Muestran la relación entre múltiples variables a través de colores que representan valores numéricos. Vamos a crear un caso de uso aplicado a la educación utilizando mapas de calor (heatmaps). Imaginemos que tenemos un conjunto de datos que incluye las calificaciones de estudiantes en diferentes materias y queremos visualizar cómo se relacionan estas calificaciones entre sí. Un mapa de calor puede ser una herramienta excelente para identificar patrones, como qué materias tienden a tener rendimientos similares entre los estudiantes.

Generaremos datos ficticios que incluyen las calificaciones de 100 estudiantes en cinco materias diferentes:

- Matemáticas
- Ciencias
- Literatura
- Historia
- Arte

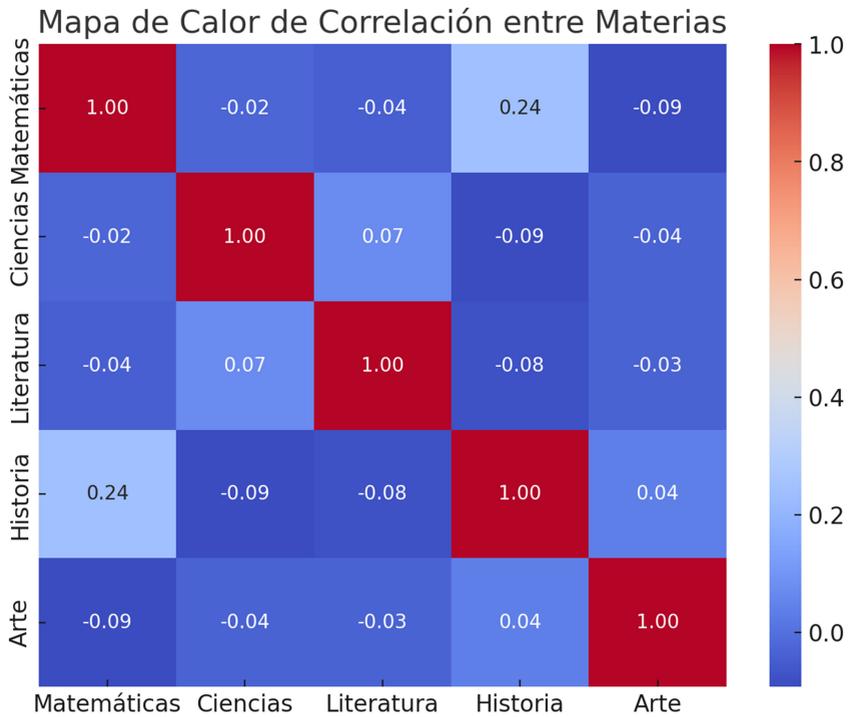
El objetivo es visualizar la correlación entre las calificaciones en estas materias para identificar posibles patrones. Por ejemplo, podríamos descubrir que los estudiantes que obtienen altas calificaciones en Matemáticas también tienden a obtener altas calificaciones en Ciencias, lo que podría sugerir una afinidad en áreas relacionadas con habilidades analíticas.

```
# Generar datos ficticios para las calificaciones
np.random.seed(1)
calificaciones = {
    "Matemáticas": np.random.uniform(0, 10, 100),
    "Ciencias": np.random.uniform(0, 10, 100),
    "Literatura": np.random.uniform(0, 10, 100),
    "Historia": np.random.uniform(0, 10, 100),
    "Arte": np.random.uniform(0, 10, 100)
}

df_calificaciones = pd.DataFrame(calificaciones)

# Calcular la matriz de correlación
matriz_correlacion = df_calificaciones.corr()

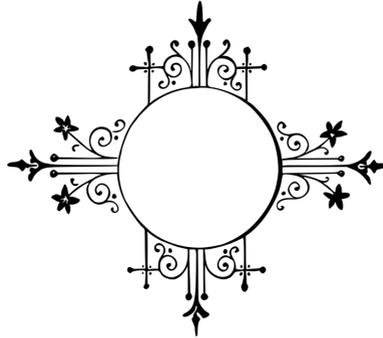
# Crear el mapa de calor
plt.figure(figsize=(8, 6))
sns.heatmap(matriz_correlacion, annot=True,
            cmap='coolwarm', fmt=".2f")
plt.title("Mapa de Calor de Correlación entre Materias")
plt.show()
```



En nuestro ejemplo ficticio, debido a la naturaleza aleatoria de los datos, es posible que no veamos correlaciones fuertes. En un escenario real, este tipo de análisis puede ser muy útil para entender cómo se relacionan diferentes áreas de estudio y podría ayudar a identificar áreas donde los estudiantes podrían necesitar apoyo adicional o áreas donde podrían tener fortalezas ocultas. Por ejemplo, si se descubre una fuerte correlación entre Matemáticas y Ciencias, la escuela podría considerar estrategias de enseñanza integradas para estas materias.

CAPÍTULO VI

VISUALIZACIÓN DE DATOS SOCIALES



CAPÍTULO VI.

VISUALIZACIÓN DE DATOS SOCIALES

Análisis de Datos en Medios Sociales

El análisis de datos en medios sociales, también conocido como análisis de redes sociales, se refiere al proceso de recopilar, procesar y extraer información significativa de las plataformas de medios sociales y las interacciones en línea. Esto es fundamental para comprender y aprovechar la dinámica de las redes sociales y tomar decisiones estratégicas en marketing, investigación de mercado, gestión de la reputación y otras áreas. Técnicas clave para llevar a cabo un análisis efectivo:

- **Definición de objetivos:** Antes de comenzar cualquier análisis, debes establecer tus objetivos claros. ¿Qué quieres lograr con el análisis de medios sociales? Ejemplos de objetivos pueden incluir el seguimiento de la percepción de la marca, la identificación de tendencias, la evaluación del rendimiento de campañas publicitarias, etc.
- **Recopilación de datos:** Determina qué plataformas de

medios sociales son relevantes para tu análisis (por ejemplo, Facebook, Twitter, Instagram, LinkedIn) y utiliza herramientas de monitoreo y análisis para recopilar datos. Estas herramientas pueden ser nativas de las plataformas o de terceros como Hootsuite, Buffer, Sprout Social, Brandwatch, etc.

- **Preprocesamiento de datos:** Limpia los datos eliminando duplicados, corrección de errores y normalización. Estructura los datos en un formato coherente para facilitar su análisis.
- **Análisis descriptivo:** Realiza un análisis descriptivo inicial utilizando estadísticas básicas como recuento de publicaciones, menciones, seguidores, comentarios, compartidos y me gusta. Crea gráficos y visualizaciones para representar la distribución de datos y tendencias.
- **Análisis de sentimiento:** Utiliza técnicas de procesamiento de lenguaje natural (NLP) para determinar el sentimiento detrás de los comentarios y menciones en las redes sociales (positivo, negativo o neutro). Esto puede ayudarte a evaluar la percepción de tu marca.
- **Análisis de tendencias:** Busca palabras clave y temas recurrentes en las conversaciones para identificar tendencias emergentes y temas de interés. Evalúa la efectividad de los hashtags y su uso en tu industria o nicho.
- **Análisis de la competencia:** Compara el rendimiento de tus redes sociales con el de tus competidores para identificar oportunidades y áreas de mejora.
- **Segmentación de la audiencia:** Divide a tu audiencia en grupos demográficos y psicográficos para comprender

mejor a quién te diriges y adaptar tu estrategia de contenido.

- **Medición del rendimiento:** Define y rastrea métricas clave de rendimiento (KPIs) como el alcance, la participación, la conversión y el retorno de la inversión (ROI).
- **Informes y recomendaciones:** Genera informes regulares que resuman los hallazgos y las recomendaciones basadas en el análisis de medios sociales. Comparte estos informes con los equipos relevantes para tomar decisiones estratégicas.

Visualización de Patrones Demográficos y Sociales

La visualización de patrones demográficos y sociales es fundamental para comprender y analizar la composición y las tendencias de una población en una región o área geográfica específica. Estas visualizaciones proporcionan información valiosa sobre diversos aspectos de la sociedad, como la edad, el género, la educación, los ingresos, la distribución geográfica y muchas otras variables clave⁴¹.

Una de las técnicas de visualización más comunes para representar patrones demográficos es el uso de gráficos de barras o gráficos de sectores. Estos gráficos permiten mostrar la distribución de una población en función de una variable específica, como la edad o el género. Por ejemplo, un gráfico de barras puede ilustrar la distribución por edades de una ciudad, resaltando si la población es joven o envejecida.

Además de los gráficos de barras, los mapas temáticos son una herramienta poderosa para visualizar patrones geográficos.

ficos y sociales. Los mapas pueden mostrar la distribución geográfica de variables demográficas y sociales, lo que facilita la identificación de concentraciones o disparidades en diferentes áreas. Por ejemplo, un mapa temático podría mostrar la distribución de ingresos por código postal, resaltando áreas de alta y baja renta.

Otra técnica efectiva es el uso de gráficos de dispersión para analizar la relación entre dos variables demográficas o sociales. Por ejemplo, un gráfico de dispersión podría mostrar la relación entre el nivel de educación y los ingresos, lo que ayudaría a identificar patrones en cómo la educación afecta los ingresos en una población.

Para crear un ejemplo complejo de visualización de patrones demográficos y sociales, imaginemos que tenemos un conjunto de datos que representa una encuesta realizada en una comunidad educativa. Esta encuesta incluye información demográfica y social de los participantes, como edad, género, nivel de educación, ingresos, y satisfacción con los servicios educativos.

Generaremos datos ficticios para 500 personas con las siguientes variables:

- **Edad:** Distribución de edades de los participantes.
- **Género:** Masculino, Femenino, Otro.
- **Nivel de Educación:** Primaria, Secundaria, Universitaria, Posgrado.
- **Ingresos Mensuales:** Categorizados en rangos.
- **Satisfacción con Servicios Educativos:** Puntuación de 1 a 5.

El objetivo es visualizar y analizar los patrones demográficos y sociales en esta comunidad educativa, particularmente en relación con la satisfacción con los servicios educativos. Buscaremos entender cómo la satisfacción se relaciona con otras variables como edad, género, nivel de educación e ingresos.

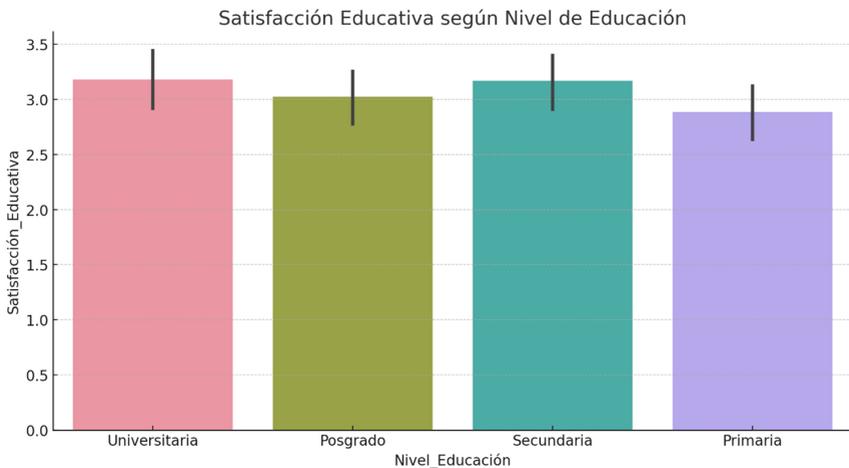
```
# Generar datos ficticios
np.random.seed(2)
datos_demograficos = {
    "Edad": np.random.randint(18, 60, 500),
    "Género": np.random.choice(["Masculino", "Femenino",
    "Otro"], 500),
    "Nivel_Educación": np.random.choice(["Primaria",
    "Secundaria", "Universitaria", "Posgrado"], 500),
    "Ingresos_Mensuales": np.random.choice(["Bajo",
    "Medio", "Alto"], 500),
    "Satisfacción_Educativa": np.random.randint(1, 6, 500)
}

df_demográficos = pd.DataFrame(datos_demograficos)

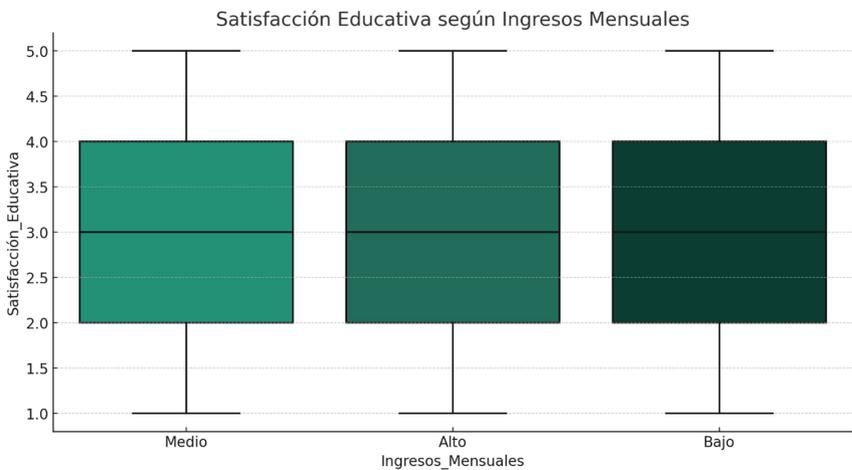
# Crear visualizaciones
# Gráfico de barras para la satisfacción educativa según el
nivel de educación
plt.figure(figsize=(12, 6))
sns.barplot(x="Nivel_Educación", y="Satisfacción_
Educativa", data=df_demográficos)
plt.title("Satisfacción Educativa según Nivel de
Educación")
plt.show()
```

```
# Boxplot para la satisfacción educativa según los ingresos mensuales
plt.figure(figsize=(12, 6))
sns.boxplot(x="Ingresos_Mensuales", y="Satisfacción_Educativa", data=df_demográficos)
plt.title("Satisfacción Educativa según Ingresos Mensuales")
plt.show()

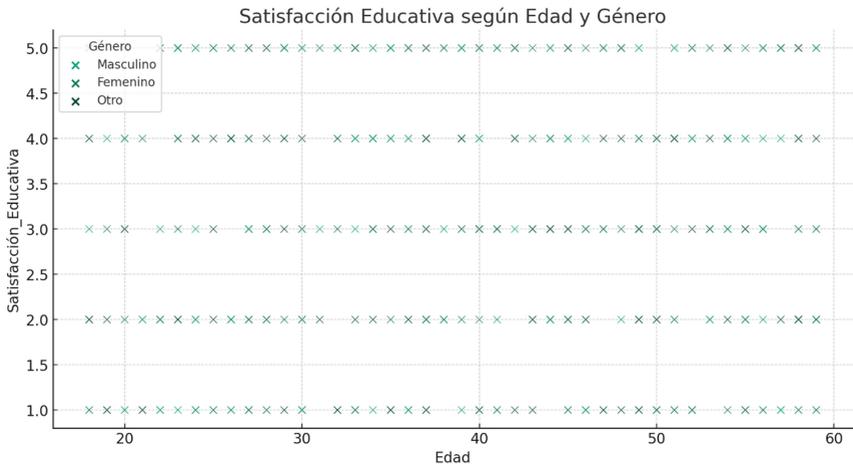
# Diagrama de dispersión para la satisfacción educativa según la edad
plt.figure(figsize=(12, 6))
sns.scatterplot(x="Edad", y="Satisfacción_Educativa", hue="Género", data=df_demográficos)
plt.title("Satisfacción Educativa según Edad y Género")
plt.show()
```



El primer gráfico es un gráfico de barras que muestra la satisfacción educativa media según el nivel de educación. Aquí, podemos observar diferencias en la satisfacción educativa entre los distintos niveles de educación. Por ejemplo, si las barras son significativamente más altas para “Universitaria” y “Posgrado” en comparación con “Primaria” y “Secundaria”, podría indicar una mayor satisfacción con los servicios educativos entre las personas con mayor nivel de educación.



El segundo gráfico es un boxplot que representa la distribución de la satisfacción educativa en diferentes rangos de ingresos mensuales. Esta visualización puede revelar no solo las medianas, sino también la variabilidad y los valores atípicos en la satisfacción entre los distintos grupos de ingresos. Por ejemplo, una mayor dispersión en un grupo de ingresos podría indicar diferencias más significativas en las percepciones o experiencias educativas dentro de ese grupo.



El tercer gráfico es un diagrama de dispersión que muestra la relación entre la edad y la satisfacción educativa, con una distinción de colores por género. Este gráfico puede ayudar a identificar si existen patrones o tendencias en la satisfacción educativa en función de la edad, y si estos patrones varían entre diferentes géneros.

A través de estas visualizaciones, podemos comenzar a entender cómo diferentes factores demográficos y sociales (como el nivel de educación, ingresos, edad y género) pueden influir en la percepción y satisfacción con los servicios educativos. Estos insights son valiosos para diseñar políticas educativas más inclusivas y efectivas, ajustadas a las necesidades y expectativas de diversos grupos dentro de la comunidad educativa. En un escenario real, el análisis de estos patrones podría ser aún más profundo, utilizando métodos estadísticos para confirmar la significancia de las observaciones.

Herramientas de Visualización Geoespacial

Las herramientas de visualización geoespacial son fundamentales para representar datos en un contexto geográfico, lo que facilita la comprensión de patrones y tendencias relacionados con la ubicación. Algunas de las principales herramientas de visualización geoespacial:

- **ArcGIS:** Desarrollado por Esri, ArcGIS es una de las plataformas líderes en el mundo para la visualización y análisis geoespacial. Ofrece una amplia gama de herramientas para crear mapas interactivos, realizar análisis geoespaciales avanzados y compartir información geográfica.
- **QGIS (Quantum GIS):** QGIS es un sistema de información geográfica de código abierto que proporciona una potente capacidad de visualización y análisis geoespacial. Es gratuito y altamente personalizable a través de complementos y extensiones.
- **Google Maps Platform:** Google Maps Platform permite a los desarrolladores integrar mapas y datos geoespaciales en aplicaciones web y móviles. Ofrece APIs para mapas, geocodificación, rutas y más.
- **Mapbox:** Mapbox es una plataforma de cartografía que permite la creación de mapas personalizados y la incorporación de mapas interactivos en aplicaciones. Ofrece opciones de diseño y estilo muy flexibles.
- **Leaflet:** Leaflet es una biblioteca de JavaScript de código abierto que facilita la creación de mapas interactivos en sitios web. Es ligero y altamente personalizable.
- **Tableau:** Tableau es una herramienta de visualización de datos que incluye capacidades geoespaciales. Permi-

te crear gráficos y mapas interactivos utilizando datos geográficos.

- **Power BI:** Power BI es otra herramienta de visualización de datos que admite visualización geoespacial. Permite crear mapas interactivos y paneles de control con datos geográficos.
- **Carto:** Carto es una plataforma de análisis de datos espaciales que facilita la creación de mapas interactivos y análisis espaciales avanzados.
- **OpenStreetMap (OSM):** OpenStreetMap es un proyecto colaborativo de cartografía en línea que ofrece datos geoespaciales de código abierto. Puedes utilizar herramientas y bibliotecas para visualizar y trabajar con datos de OSM.
- **Google Earth:** Google Earth es una aplicación de escritorio y móvil que permite explorar el mundo en 3D y visualizar datos geoespaciales en un entorno virtual.
- **D3.js:** D3.js es una biblioteca de JavaScript que permite crear visualizaciones de datos personalizadas, incluyendo mapas interactivos. Requiere conocimientos de programación para su uso.

Para crear un ejemplo de visualización geoespacial aplicada al campo social, podemos imaginar un escenario donde analizamos la distribución geográfica de ciertas instalaciones o recursos sociales en una ciudad o región. Por ejemplo, podríamos examinar la ubicación de escuelas, hospitales, parques, o centros comunitarios y cómo se relacionan con variables sociodemográficas como la densidad de población, el nivel de ingresos o la distribución por edades.

Supongamos que tenemos un conjunto de datos que incluye la ubicación de varias escuelas en una ciudad ficticia, junto con datos sobre la población en diferentes áreas de la ciudad. Queremos visualizar la ubicación de estas escuelas y analizar su distribución en relación con la densidad de población.

Conjunto de Datos Ficticios

- **Ubicaciones de Escuelas:** Coordenadas (latitud y longitud) de varias escuelas.
- **Datos de Población:** Densidad de población en diferentes áreas de la ciudad.

El objetivo es visualizar la ubicación de las escuelas y analizar si están distribuidas equitativamente en relación con la densidad de población. Procederemos a generar datos ficticios y crear un mapa utilizando matplotlib y geopandas. Vamos a asumir que tenemos los datos de latitud y longitud para simplificar el ejemplo. En un caso real, estos datos podrían obtenerse de fuentes gubernamentales o bases de datos geográficas.

```
import geopandas as gpd
from shapely.geometry import Point
import matplotlib.pyplot as plt

# Generar datos ficticios
# Coordenadas de escuelas (latitud y longitud)
coordenadas_escuelas = [
    (12.9716, 77.5946), # Coordenadas ficticias
```

```
(12.2958, 76.6394),
(13.0827, 80.2707),
(13.6288, 79.4192),
(12.9141, 74.8560)
]

# Densidad de población en áreas de la ciudad (ficticias)
datos_población = {
    "Centro": 10000,
    "Norte": 7000,
    "Sur": 5000,
    "Este": 4000,
    "Oeste": 3000
}

# Crear GeoDataFrame para las escuelas
gdf_escuelas = gpd.GeoDataFrame(
    {"Ubicación": ["Centro", "Norte", "Sur", "Este",
    "Oeste"],
    "geometry": [Point(lon, lat) for lat, lon in coordenadas_
    escuelas]}
)

# Crear un mapa base
world = gpd.read_file(gpd.datasets.get_path('naturalearth_
lowres'))

# Filtrar el mapa al área de interés (ficticia, asumimos que
es India para el ejemplo)
```

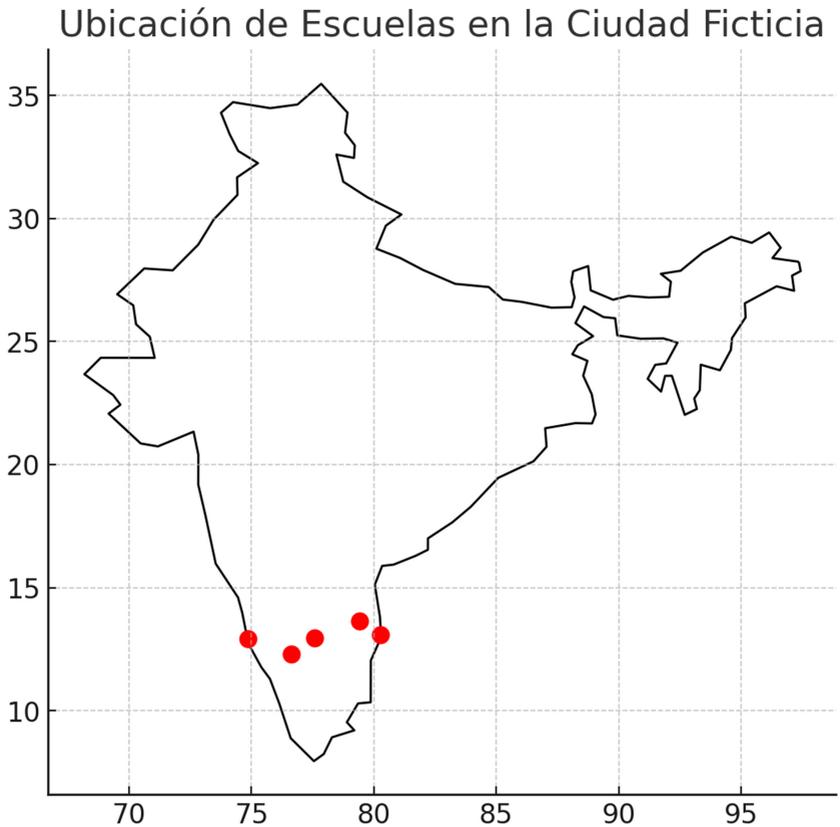
```
base = world[world.name == "India"].plot(color='white',  
edgecolor='black')
```

```
# Plotear las ubicaciones de las escuelas
```

```
gdf_escuelas.plot(ax=base, marker='o', color='red',  
markersize=50)
```

```
plt.title("Ubicación de Escuelas en la Ciudad Ficticia")
```

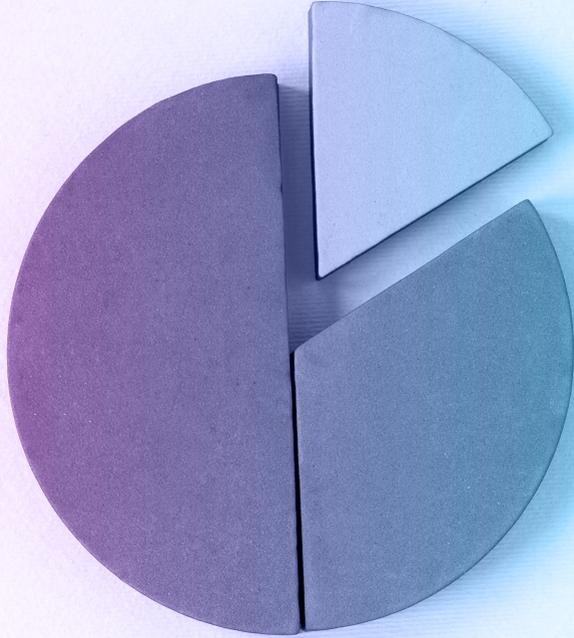
```
plt.show()
```



Hemos creado un mapa que muestra la ubicación de escuelas en una ciudad ficticia. En este mapa, cada punto rojo representa la ubicación de una escuela. Para efectos de este ejemplo, hemos utilizado un mapa base de India, pero en un análisis real, se usaría un mapa que represente la región específica en estudio.

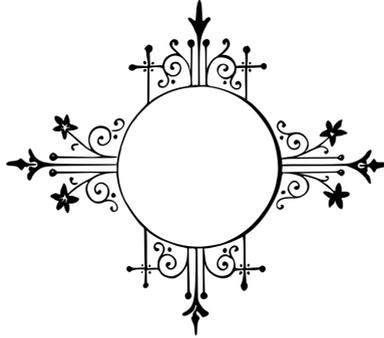
- **Distribución de Escuelas:** Podemos observar cómo están distribuidas las escuelas en la ciudad. Idealmente, las escuelas deben estar distribuidas equitativamente para asegurar un acceso fácil a la educación en todas las áreas de la ciudad.
- **Comparación con Datos de Población:** Aunque no hemos visualizado directamente la densidad de población en este mapa, en un análisis real, podríamos superponer la información de densidad de población o utilizar un mapa de calor para ver cómo la distribución de escuelas se correlaciona con la densidad poblacional. Esto es crucial para entender si las áreas con mayor densidad de población tienen suficiente acceso a instalaciones educativas.

Este tipo de análisis es muy útil en la planificación urbana y la gestión de recursos. Por ejemplo, si se identifica que ciertas áreas con alta densidad de población tienen pocas escuelas, esto podría ser un indicador para la construcción de nuevas instalaciones educativas en esas áreas. Además, este análisis puede ayudar a evaluar la accesibilidad de la educación en diferentes partes de la ciudad y a identificar posibles desigualdades.



CAPÍTULO VII

DESAFÍOS Y TENDENCIAS FUTURAS



CAPÍTULO VII.

DESAFÍOS Y TENDENCIAS FUTURAS

Desafíos Actuales en la Visualización de Datos Educativos y Sociales

La visualización de datos educativos y sociales enfrenta varios desafíos en la actualidad debido a la complejidad de los datos y la necesidad de proporcionar información significativa a una variedad de audiencias. Estos desafíos incluyen⁴²⁻⁴⁶:

- **Privacidad y ética:** Uno de los desafíos más importantes en la visualización de datos educativos y sociales es la protección de la privacidad de los individuos. La recopilación y divulgación de datos personales debe estar sujeta a estrictas normativas y políticas de privacidad para evitar el riesgo de violaciones de datos y el uso indebido de la información.
- **Calidad y consistencia de los datos:** Los datos educativos y sociales provienen de múltiples fuentes y pueden variar en calidad y consistencia. La limpieza y normali-

zación de datos son procesos críticos para garantizar la precisión y la confiabilidad de las visualizaciones.

- **Interpretación adecuada:** La visualización de datos educativos y sociales debe ser accesible y comprensible para una amplia gama de audiencias, incluyendo a educadores, estudiantes, padres y responsables de políticas. Presentar datos de manera clara y evitar sesgos en la interpretación es un desafío constante.
- **Toma de decisiones basada en datos:** Asegurar que las visualizaciones de datos educativos y sociales sean útiles para la toma de decisiones efectivas es esencial. Esto implica identificar las métricas clave y presentar los datos de manera que permita a los responsables de la toma de decisiones actuar de manera informada.
- **Tendencias cambiantes y dinámicas sociales:** Los patrones y tendencias en datos educativos y sociales pueden cambiar con el tiempo debido a factores económicos, sociales y políticos. Mantenerse actualizado con estas dinámicas y ajustar las visualizaciones en consecuencia es un desafío constante.
- **Accesibilidad y diversidad:** Las visualizaciones deben ser accesibles para personas con discapacidades y culturalmente sensibles. Esto implica considerar la diversidad en la audiencia y garantizar que las visualizaciones sean inclusivas.
- **Integración de múltiples fuentes de datos:** A menudo, los datos educativos y sociales provienen de diversas fuentes, como encuestas, registros administrativos y plataformas en línea. Integrar estos datos de manera efectiva para obtener una imagen completa puede ser

un desafío técnico y logístico.

- **Actualización constante:** La visualización de datos educativos y sociales debe mantenerse actualizada para reflejar cambios en el entorno educativo y social. Esto requiere una gestión continua de datos y una adaptación constante de las visualizaciones.

El Rol de la Inteligencia Artificial en la Visualización de Datos

La inteligencia artificial (IA) desempeña un papel cada vez más importante en la visualización de datos, ya que ofrece capacidades avanzadas para analizar, procesar y presentar información de manera más efectiva. Se destacan algunas formas en las que la IA está influyendo en la visualización de datos⁴⁷⁻⁴⁹:

- **Automatización de la visualización:** Los algoritmos de IA pueden generar visualizaciones de datos automáticamente a partir de conjuntos de datos complejos. Esto ahorra tiempo a los analistas y permite explorar visualmente grandes cantidades de datos de manera eficiente.
- **Selección de visualizaciones óptimas:** La IA puede ayudar a elegir el tipo de visualización más adecuado para un conjunto de datos específico. Al analizar la estructura de los datos y los objetivos del usuario, la IA puede recomendar el gráfico o la visualización que mejor resalte los patrones y las relaciones en los datos.
- **Análisis de texto e imágenes:** La IA puede analizar texto no estructurado, como comentarios de redes sociales o informes, y convertirlo en datos estructurados que se

pueden visualizar. También puede analizar imágenes y extraer información relevante para su inclusión en visualizaciones.

- **Detección de anomalías:** Los algoritmos de IA pueden identificar anomalías en los datos y destacarlas en las visualizaciones. Esto es útil para el monitoreo en tiempo real de datos y la detección temprana de problemas o tendencias inusuales.
- **Interacción y personalización:** La IA puede mejorar la interacción del usuario con las visualizaciones, permitiendo búsquedas, filtrado y navegación más avanzados. También puede personalizar las visualizaciones para adaptarse a las preferencias y necesidades individuales de los usuarios.
- **Traducción automática:** La IA puede traducir automáticamente las etiquetas y los datos de las visualizaciones a diferentes idiomas, lo que facilita la comunicación en contextos multilingües.
- **Aprendizaje automático y predicción:** La IA puede utilizar algoritmos de aprendizaje automático para identificar patrones y tendencias en los datos y realizar predicciones futuras. Estas predicciones pueden integrarse en visualizaciones para proporcionar una visión más completa y prospectiva.
- **Procesamiento de lenguaje natural (NLP):** La IA basada en NLP puede ayudar a generar resúmenes y descripciones automáticas de datos y visualizaciones, lo que facilita la comprensión de la información presentada.
- **Automatización de informes:** La IA puede generar infor-

mes automatizados basados en datos y visualizaciones, lo que agiliza la comunicación de resultados y hallazgos a las partes interesadas.

Tendencias Emergentes y Futuras en Visualización de Datos

La visualización de datos es un campo en constante evolución y las tendencias emergentes y futuras reflejan los avances tecnológicos y las necesidades cambiantes de los usuarios. Algunas tendencias son:

- **Visualización de datos en tiempo real:** La demanda de visualizaciones en tiempo real está en aumento, especialmente en áreas como las finanzas, el monitoreo de redes sociales, la salud y la gestión de la cadena de suministro. Las visualizaciones en tiempo real permiten tomar decisiones basadas en datos de manera más rápida y eficiente.
- **Inteligencia artificial y aprendizaje automático:** La integración de la inteligencia artificial y el aprendizaje automático en las herramientas de visualización de datos permite análisis más avanzados y recomendaciones inteligentes. Esto incluye la detección de patrones, la predicción y la automatización de tareas de diseño.
- **Visualización de datos 3D y realidad aumentada (RA):** La visualización de datos en 3D y la RA están ganando terreno en campos como la medicina, la arquitectura y la simulación. Estas tecnologías permiten explorar datos de manera más inmersiva y comprender mejor las relaciones espaciales.
- **Visualización de datos geoespaciales avanzada:** Con la

creciente disponibilidad de datos geoespaciales, se espera un aumento en la visualización de datos relacionados con ubicaciones geográficas. Esto incluye mapas interactivos y visualizaciones que muestran tendencias y patrones en función de la ubicación.

- **Visualización de datos en dispositivos móviles y aplicaciones:** La visualización de datos en dispositivos móviles y aplicaciones se está volviendo más importante a medida que la movilidad se convierte en una parte integral de la vida cotidiana. Las aplicaciones móviles ofrecen experiencias personalizadas y accesibles para los usuarios en movimiento.
- **Énfasis en la narración de datos:** La narración de datos implica contar historias con datos para comunicar de manera efectiva hallazgos y conclusiones. Se espera que esta tendencia continúe, con un enfoque en la comunicación de datos de manera más convincente y significativa.
- **Mayor énfasis en la ética y la privacidad de los datos:** A medida que se recopilan y comparten más datos, la ética y la privacidad de los datos se han vuelto cruciales. Las herramientas de visualización deben abordar estos problemas y garantizar que los datos se manejen de manera responsable.
- **Visualización de datos en realidad virtual (RV):** Aunque aún está en sus primeras etapas, la visualización de datos en RV ofrece la posibilidad de explorar y comprender datos en un entorno completamente inmersivo. Esto puede tener aplicaciones en la investigación, la educación y la toma de decisiones complejas.

- **Automatización de la visualización de datos:** Se espera que la automatización de la creación de visualizaciones continúe evolucionando, lo que permitirá a los usuarios generar rápidamente gráficos y tablas a partir de conjuntos de datos con solo unos pocos clics.
- **Visualización de datos en la atención médica y la biotecnología:** La atención médica y la biotecnología se benefician de la visualización de datos para comprender mejor la genómica, los resultados de los ensayos clínicos y la gestión de la salud. Se prevé que estas aplicaciones sigan creciendo en importancia.

Preparándose para el Futuro de la Visualización de Datos

Implica adoptar enfoques y estrategias que permitan aprovechar al máximo las tendencias emergentes y las tecnologías avanzadas. Algunas recomendaciones para estar listo para el futuro de la visualización de datos:

- **Mantente actualizado con las tendencias:** Invierte tiempo en mantenerte al tanto de las últimas tendencias en visualización de datos, inteligencia artificial, realidad aumentada y otras tecnologías emergentes. Sigue blogs, participa en conferencias y sé parte de comunidades en línea relacionadas con la visualización de datos.
- **Mejora tus habilidades técnicas:** A medida que las herramientas de visualización de datos se vuelven más avanzadas, es importante mejorar tus habilidades técnicas. Aprende a trabajar con herramientas de inteligencia artificial, visualización en 3D y otras tecnologías relevantes para el futuro.

- **Enfócate en la narración de datos:** La habilidad de contar historias efectivas con datos será cada vez más importante. Aprende a narrar tus visualizaciones de manera convincente y efectiva para comunicar hallazgos de manera más clara.
- **Considera la ética y la privacidad de los datos:** A medida que los datos se vuelven más omnipresentes, la ética y la privacidad de los datos son preocupaciones crecientes. Asegúrate de estar al tanto de las regulaciones y mejores prácticas en este ámbito y asegúrate de que tus visualizaciones sean éticas y respeten la privacidad.
- **Invierte en herramientas avanzadas:** Considera la posibilidad de invertir en herramientas de visualización de datos avanzadas que aprovechen la inteligencia artificial y otras tecnologías. Estas herramientas pueden ayudarte a automatizar tareas y crear visualizaciones más sofisticadas.
- **Adapta tu enfoque a las necesidades cambiantes:** Reconoce que las necesidades de visualización de datos pueden cambiar con el tiempo. Mantente flexible y dispuesto a adaptar tus enfoques y estrategias a medida que evolucione el campo.
- **Desarrolla colaboraciones interdisciplinarias:** La visualización de datos a menudo requiere una comprensión profunda de los dominios específicos. Trabaja en colaboración con expertos en esos dominios para comprender mejor los datos y crear visualizaciones más efectivas.
- **Aprende de los líderes del campo:** Observa y aprende de organizaciones y profesionales líderes en la visua-

lización de datos. Estudia sus prácticas, técnicas y enfoques para obtener ideas sobre cómo avanzar en tu propio trabajo.

- **Mantén un enfoque centrado en el usuario:** La usabilidad y la accesibilidad son fundamentales en la visualización de datos. Asegúrate de que tus visualizaciones estén diseñadas pensando en el usuario y sean accesibles para una amplia gama de audiencias.
- **Explora nuevas aplicaciones y sectores:** La visualización de datos tiene aplicaciones en una variedad de campos, desde la atención médica hasta la educación. Explora oportunidades para aplicar tus habilidades en diferentes sectores y contextos.

REFUERZO

ACADÉMICO

CAPÍTULO I.

¿Qué representa la visualización de datos?

- a) Un método para aumentar la cantidad de datos
- b) Una disciplina que se centra en representar información y datos numéricos de manera gráfica y visual
- c) Un proceso de análisis estadístico detallado
- d) Una técnica para recopilar nuevos datos

RESPUESTA: b

¿Qué permite la codificación visual en la visualización de datos?

- a) Aumentar la velocidad de procesamiento de datos
- b) Asignar valores numéricos o categorías a propiedades visuales como longitud, color, posición y forma
- c) Codificar datos en un formato secreto
- d) Crear datos numéricos a partir de imágenes

RESPUESTA: b

¿Cuál es la importancia de la percepción visual en la visualización de datos?

- a) Permite el uso de colores brillantes en los gráficos

- b) Es fundamental para diseñar visualizaciones que sean intuitivas y fáciles de entender
- c) Aumenta la complejidad de los gráficos para expertos
- d) Solo se aplica en el diseño de páginas web

RESPUESTA: b

¿Qué se debe considerar al elegir el tipo de gráfico en la visualización de datos?

- a) La complejidad del gráfico
- b) El costo de producción del gráfico
- c) El tipo de datos y objetivos de comunicación
- d) La preferencia personal del diseñador

RESPUESTA: c

¿Cuál es el propósito de simplificar y aclarar datos en la visualización de datos?

- a) Hacer que los datos sean más complejos y detallados
- b) Presentar solo la información relevante y evitar malentendidos
- c) Hacer que los gráficos sean más coloridos y atractivos
- d) Disminuir el tiempo necesario para crear visualizaciones

RESPUESTA: b

¿Cómo influye la interactividad en la visualización de datos en el entorno digital actual?

- a) Reduce la claridad de los datos presentados
- b) Solo es relevante para los diseñadores gráficos
- c) Hace que los datos sean menos accesibles
- d) Permite a los usuarios explorar los datos por sí mismos

RESPUESTA: d

¿Por qué son fundamentales la ética y la precisión en la visualización de datos?

- a) Para mantener la confianza del público y evitar interpretaciones erróneas
- b) Para aumentar la complejidad de los datos
- c) Para hacer que los gráficos sean más atractivos
- d) Solo importan en el ámbito académico

RESPUESTA: a

¿Quién introdujo gráficos de barras, líneas y circulares en el siglo XIX?

- a) Willard C. Brinton

- b) John Graunt
- c) William Playfair
- d) Un científico anónimo

RESPUESTA: c

¿Cómo afectó la introducción de las computadoras en la década de 1960 a la visualización de datos?

- a) Disminuyó su importancia
- b) La hizo menos accesible
- c) Revolucionó su eficiencia y capacidad
- d) Cambió su enfoque a solo datos financieros

RESPUESTA: c

¿Cuál es el papel de la visualización de datos en la era del big data y la información digital?

- a) Ha disminuido su relevancia
- b) Se centra únicamente en datos financieros
- c) Ha adquirido una importancia aún mayor
- d) Se limita a visualizaciones estáticas

RESPUESTA: c

¿Cómo han utilizado los medios de comunicación las visualizaciones de datos?

- a) Solo para entretenimiento
- b) Para explicar temas complejos como elecciones y economía
- c) Exclusivamente en noticias deportivas
- d) Solo en publicidad

RESPUESTA: b

¿Qué papel juega la visualización de datos en la toma de decisiones empresariales?

- a) Ningún papel importante
- b) Se utiliza principalmente para diseño gráfico
- c) Es fundamental para analizar métricas clave como ventas y rendimiento del mercado
- d) Solo se usa para presentaciones internas

RESPUESTA: c

¿Cómo contribuye la visualización de datos en el ámbito educativo?

- a) Solo se utiliza para decorar las aulas
- b) Facilita la enseñanza de conceptos complejos y mejora la retención y comprensión
- c) Se utiliza para aumentar la carga de trabajo de los estudiantes
- d) No tiene impacto en la educación

RESPUESTA: b

¿Qué herramienta de visualización de datos es conocida por su capacidad para crear visualizaciones interactivas?

- a) Microsoft Word
- b) Tableau
- c) Paint
- d) Adobe Photoshop

RESPUESTA: b

¿Qué tendencia en visualización de datos implica la integración del aprendizaje automático?

- a) Generación automática de visualizaciones y detección de patrones
- b) Uso exclusivo de gráficos estáticos
- c) Enfoque en gráficos impresos
- d) Evitar el uso de cualquier tecnología moderna

RESPUESTA: a

¿Cómo afecta la visualización de datos a la toma de decisiones informadas en el contexto social y educativo?

- a) No tiene ningún efecto
- b) Complica el proceso de toma de decisiones
- c) Facilita la toma de decisiones basadas en evidencia
- d) Solo es relevante para científicos

RESPUESTA: c

¿Qué papel desempeñan las visualizaciones de datos en el análisis de rendimiento académico?

- a) Ninguno, se utilizan exclusivamente en finanzas
- b) Ayudan a identificar áreas que requieren atención para mejorar la calidad de la educación
- c) Solo se usan para calificar exámenes
- d) Se utilizan principalmente para actividades deportivas

RESPUESTA: b

CAPITULO II

¿Cuál es el propósito principal de los principios de diseño en visualización de datos?

- a) Hacer los gráficos más atractivos visualmente.
- b) Garantizar que las visualizaciones sean claras y efectivas en la comunicación de información.

- c) Reducir el tiempo necesario para crear visualizaciones.
- d) Aumentar la complejidad de los gráficos para análisis detallados.

RESPUESTA: b

¿Qué principio de diseño enfatiza la eliminación de elementos innecesarios en las visualizaciones de datos?

- a) Precisión.
- b) Claridad y simplicidad.
- c) Jerarquía visual.
- d) Usabilidad.

RESPUESTA: b

En la visualización de datos, ¿por qué es importante la precisión?

- a) Para hacer los gráficos más interactivos.
- b) Para garantizar la fidelidad a los datos reales.
- c) Para mejorar la estética visual.
- d) Para reducir el tiempo de carga de los gráficos.

RESPUESTA: b

¿Cómo se logra la jerarquía visual en una visualización

de datos?

- a) Mediante la interactividad.
- b) Usando diferentes colores, tamaños y ubicaciones estratégicas.
- c) Reduciendo la cantidad de datos presentados.
- d) Utilizando un único color para toda la visualización.

RESPUESTA: b

¿Qué principio de diseño se relaciona con mantener uniformidad en colores, fuentes y estilos en una visualización?

- a) Narrativa.
- b) Consistencia.
- c) Interactividad.
- d) Claridad y simplicidad.

RESPUESTA: b

¿Por qué es importante considerar la usabilidad en las visualizaciones de datos?

- a) Para asegurar que sean fáciles de usar y entender por la audiencia.
- b) Para hacer que los gráficos sean más atractivos.
- c) Para reducir el tamaño de los archivos de los gráficos.

d) Para incrementar la cantidad de datos mostrados.

RESPUESTA: a

En la visualización de datos, ¿qué rol juega la narrativa?

- a) Hacer los gráficos más interactivos.
- b) Contar una historia y comunicar un mensaje específico.
- c) Aumentar la complejidad de los gráficos.
- d) Reducir la cantidad de datos necesarios.

RESPUESTA: b

¿Cuál es el propósito de utilizar color y contraste en la visualización de datos?

- a) Resaltar elementos importantes y crear distinciones visuales.
- b) Hacer que los gráficos sean más atractivos.
- c) Aumentar la interactividad del gráfico.
- d) Disminuir el tiempo de creación de los gráficos.

RESPUESTA: a

¿Qué papel juegan las pruebas y la retroalimentación en la creación de visualizaciones de datos?

- a) Ayudan a mejorar la calidad de la visualización.
- b) Reducen el tiempo necesario para crear visualizaciones.
- c) Aumentan la interactividad de los gráficos.
- d) Hacen que los gráficos sean más atractivos.

RESPUESTA: a

¿Por qué es importante proporcionar contexto en las visualizaciones de datos?

- a) Para mejorar la estética del gráfico.
- b) Para explicar la fuente y limitaciones de los datos.
- c) Para aumentar la interactividad del gráfico.
- d) Para reducir el tamaño del archivo de la visualización.

RESPUESTA: b

¿Qué se entiende por responsividad en las visualizaciones de datos?

- a) La capacidad del gráfico de adaptarse a diferentes dispositivos y tamaños de pantalla.
- b) La interactividad del gráfico.
- c) La rapidez con la que se carga el gráfico.
- d) La facilidad para exportar el gráfico en diferentes formatos.

RESPUESTA: a

¿Cuál es una ventaja principal de los gráficos estáticos sobre los interactivos?

- a) Permiten una mayor interactividad con los datos.
- b) Son más adecuados para presentaciones impresas y documentos estáticos.
- c) Facilitan la exploración de grandes conjuntos de datos.
- d) Se actualizan automáticamente con datos en tiempo real.

RESPUESTA: b

¿Qué caracteriza principalmente a los gráficos interactivos?

- a) Su capacidad para ser impresos en documentos.
- b) La posibilidad de que los usuarios exploren y manipulen los datos.
- c) Su simplicidad y facilidad de creación.
- d) La limitación en la presentación de grandes conjuntos de datos.

RESPUESTA: b

¿Por qué se recomienda Matplotlib para la creación de

gráficos básicos en Python?

- a) Por su capacidad para crear gráficos interactivos.
- b) Debido a su popularidad y facilidad de uso.
- c) Porque es la única biblioteca disponible para Python.
- d) Porque soporta exclusivamente gráficos estadísticos.

RESPUESTA: b

¿Cuál es una ventaja de Seaborn sobre Matplotlib?

- a) Su capacidad para crear gráficos básicos.
- b) Su enfoque en gráficos estadísticos y conjuntos de datos complejos.
- c) Su limitación a gráficos estáticos.
- d) Su incompatibilidad con Pandas y NumPy.

RESPUESTA: b

¿Cuál es una característica clave de Bokeh?

- a) Su incapacidad para crear visualizaciones interactivas.
- b) La creación de gráficos y visualizaciones web interactivas.
- c) La limitación en la personalización de gráficos.
- d) Su uso exclusivo para gráficos estadísticos.

RESPUESTA: b

¿Qué distingue a Plotly en el ámbito de visualización de datos?

- a) Su enfoque en gráficos estáticos.
- b) La creación de visualizaciones interactivas y dinámicas.
- c) La incapacidad de integrarse con aplicaciones web.
- d) Su limitación a gráficos básicos.

RESPUESTA: b

¿Por qué se prefieren los gráficos interactivos para datos en constante cambio?

- a) Porque son más fáciles de imprimir.
- b) Debido a que pueden actualizarse en tiempo real.
- c) Porque son más fáciles de crear.
- d) Porque requieren menos interacción del usuario.

RESPUESTA: b

¿Cuál es la ventaja principal de la interactividad en las visualizaciones de datos?

- a) Reducir el tamaño del archivo de la visualización.
- b) Aumentar la complejidad visual del gráfico.

- c) Permitir a los usuarios explorar los datos de forma más profunda.
- d) Simplificar el proceso de creación de gráficos.

RESPUESTA: c

CAPITULO III.

¿Cuál de las siguientes opciones NO es una característica de los datos educativos?

- a) Cuantitativos y cualitativos
- b) Periodicidad en su recolección
- c) Exclusivamente numéricos
- d) Contextualización según factores geográficos y educativos

RESPUESTA: c

¿Qué implica la periodicidad en la recolección de datos educativos?

- a) Se recolectan una sola vez
- b) Se recolectan y actualizan irregularmente
- c) Se recolectan y actualizan regularmente
- d) No se actualizan después de la recolección inicial

RESPUESTA: c

¿Qué es fundamental para proteger en la recopilación de datos educativos?

- a) La confidencialidad y la privacidad
- b) La accesibilidad pública de los datos
- c) La exclusión de datos cualitativos
- d) La limitación geográfica de los datos

RESPUESTA: a

¿Para qué se utilizan principalmente los datos educativos?

- a) Solo para reportes estadísticos
- b) Para la toma de decisiones en el ámbito educativo
- c) Únicamente para investigaciones académicas
- d) Solo para comparar sistemas educativos internacionales

RESPUESTA: b

¿Cuál es una característica clave de los datos sociales?

- a) Son estáticos y no cambian con el tiempo
- b) Solo incluyen información demográfica
- c) Su diversidad y amplitud en variables

d) Se enfocan exclusivamente en la educación

RESPUESTA: c

¿Qué reflejan los datos sociales al ser inherentemente dinámicos?

- a) Cambios demográficos, económicos, culturales y políticos
- b) Estabilidad en las tendencias sociales
- c) Datos históricos sin cambios
- d) Información irrelevante para la sociedad actual

RESPUESTA: a

¿Cuál es una consideración importante al tratar con datos sociales?

- a) Ignorar el contexto de recolección de los datos
- b) Enfocarse únicamente en un aspecto social
- c) La protección de la privacidad y la sensibilidad de la información
- d) Utilizar solo datos de fuentes no oficiales

RESPUESTA: c

¿Qué papel desempeñan los datos sociales en la formulación de políticas públicas?

- a) Ningún papel significativo
- b) Solo sirven para análisis históricos
- c) Ayudan a identificar desigualdades y evaluar programas sociales
- d) Se usan exclusivamente para estudios económicos

RESPUESTA: c

¿Cuál de las siguientes NO es una fuente confiable de datos educativos?

- a) UNESCO
- b) OECD
- c) Instituciones de moda
- d) National Center for Education Statistics (NCES)

RESPUESTA: c

¿Qué organismo proporciona datos educativos comparativos de países miembros?

- a) INEGI
- b) Tik Tok
- c) Eurostat
- d) Banco Mundial

RESPUESTA: b

¿Cuál es el primer paso en el proceso de preprocesamiento y limpieza de datos?

- a) Normalización y Estandarización
- b) Recopilación de Datos
- c) Codificación de Variables Categóricas
- d) Evaluación Continua

RESPUESTA: b

¿Qué se debe hacer durante la integración de datos?

- a) Ignorar la duplicación de registros
- b) Resolver conflictos de nombres y convertir formatos
- c) Evitar la combinación de datos de distintas fuentes
- d) Utilizar solo un tipo de formato de datos

RESPUESTA: b

En el preprocesamiento de datos, ¿cómo se deben tratar los valores faltantes?

- a) Deben ser siempre ignorados
- b) Se imputan con valores estimados o se eliminan
- c) Se dejan sin modificar
- d) Se reemplazan con el valor más frecuente en todos los casos

RESPUESTA: b

¿Qué implica la normalización de datos?

- a) Cambiar todos los datos a un idioma común
- b) Ajustar los datos para que tengan una escala uniforme
- c) Eliminar todos los datos numéricos
- d) Hacer que todos los datos sean públicos

RESPUESTA: b

¿Cuál es el objetivo de la codificación de variables categóricas en el análisis de datos?

- a) Hacer que los algoritmos de aprendizaje automático puedan trabajar con ellas
- b) Eliminar estas variables del análisis
- c) Convertir todas las variables en categorías
- d) Reducir la cantidad total de datos

RESPUESTA: a

¿Qué se busca con la reducción de dimensionalidad en los conjuntos de datos?

- a) Aumentar la cantidad de características
- b) Eliminar características irrelevantes o redundantes

- c) Hacer los datos más complejos
- d) Ignorar las características numéricas

RESPUESTA: b

¿Qué se verifica durante la validación de datos y verificación de consistencia?

- a) Que los datos cumplan con condiciones específicas del dominio
- b) Que todos los datos sean idénticos
- c) La eliminación completa de los datos numéricos
- d) La conversión de todos los datos a un formato único

RESPUESTA: a

¿Por qué es importante la documentación y el registro en el preprocesamiento de datos?

- a) Para aumentar la complejidad del proceso
- b) Para facilitar la reproducibilidad y trazabilidad
- c) Solo para cumplir con requisitos legales
- d) Para evitar el análisis de los datos

RESPUESTA: b

¿Cómo se aborda el proceso de preprocesamiento y limpieza de datos?

- a) De manera estática y sin cambios
- b) Como un proceso iterativo con evaluación continua
- c) Únicamente al inicio del análisis de datos
- d) Sin considerar la calidad de los datos

RESPUESTA: b

¿Qué rol juegan las encuestas nacionales en la recolección de datos sociales?

- a) Ningún papel relevante
- b) Solo recogen opiniones políticas
- c) Recolectan datos sobre opinión pública y calidad de vida
- d) Se enfocan únicamente en datos económicos

RESPUESTA: c

CAPÍTULO IV.

¿Cuál es el objetivo principal del análisis de datos en medios sociales?

- a) Entretenimiento personal
- b) Comprender y aprovechar la dinámica de las redes sociales para decisiones estratégicas
- c) Recopilación de datos para estudios académicos

d) Monitorear exclusivamente la actividad de los usuarios individuales

RESPUESTA: b

¿Qué se debe establecer antes de comenzar un análisis de medios sociales?

- a) Una cuenta en todas las redes sociales
- b) Objetivos claros del análisis
- c) Un presupuesto ilimitado para marketing
- d) Una lista de competidores a evitar

RESPUESTA: b

¿Cuál es un paso crucial en el preprocesamiento de datos en análisis de redes sociales?

- a) Aumentar el número de seguidores
- b) Publicar contenido diariamente
- c) Limpieza de datos, incluyendo la eliminación de duplicados y corrección de errores
- d) Ignorar los datos antiguos

RESPUESTA: c

En el análisis de redes sociales, ¿cómo se puede evaluar la percepción de una marca?

- a) Solo a través de las ventas
- b) Centrándose exclusivamente en las menciones positivas
- c) Ignorando los comentarios negativos
- d) Mediante el análisis de sentimiento

RESPUESTA: d

¿Qué técnica es efectiva para identificar tendencias emergentes en análisis de medios sociales?

- a) Evitar el uso de hashtags
- b) Buscar palabras clave y temas recurrentes
- c) Solo enfocarse en los seguidores más activos
- d) Limitar el análisis a una sola red social

RESPUESTA: b

¿Qué permite la segmentación de la audiencia en el análisis de medios sociales?

- a) Reducir el tamaño de la audiencia
- b) Entender mejor a la audiencia y adaptar la estrategia de contenido
- c) Ignorar los datos demográficos
- d) Centrarse únicamente en la audiencia más joven

RESPUESTA: b

¿Cuál es una técnica común para visualizar patrones demográficos?

- a) Gráficos de barras o gráficos de sectores
- b) Uso exclusivo de texto descriptivo
- c) Solo fotografías de las personas involucradas
- d) Evitar cualquier tipo de visualización

RESPUESTA: a

¿Qué muestran los mapas temáticos en la visualización de patrones demográficos y sociales?

- a) Solamente la topografía del terreno
- b) Distribución geográfica de variables demográficas y sociales
- c) Datos irrelevantes para el análisis social
- d) Únicamente la distribución de la fauna y flora

RESPUESTA: b

¿Cómo puede un gráfico de dispersión ser útil en el análisis demográfico y social?

- a) Mostrando la relación entre dos variables demográficas o sociales
- b) Presentando datos de una sola variable
- c) Evitando la comparación de variables

d) Centrándose únicamente en datos históricos

RESPUESTA: a

¿Qué papel juegan los informes y recomendaciones en el análisis de medios sociales?

- a) Ninguna relevancia en el proceso
- b) Sólo para uso interno del equipo de análisis
- c) Resumir hallazgos y recomendaciones para decisiones estratégicas
- d) Únicamente para justificar el gasto en marketing

RESPUESTA: c

CAPITULO VII.

¿Cuál es uno de los desafíos más importantes en la visualización de datos educativos y sociales?

- a) La elección de colores llamativos
- b) La protección de la privacidad de los individuos
- c) El uso de una sola fuente de datos
- d) Ignorar la calidad de los datos

RESPUESTA: b

¿Qué proceso es crítico para garantizar la precisión en la visualización de datos educativos y sociales?

- a) Aumentar el tamaño de los datos
- b) La limpieza y normalización de datos
- c) La selección de una única plataforma de visualización
- d) Enfocarse únicamente en datos cuantitativos

RESPUESTA: b

¿Por qué es un desafío presentar datos educativos y sociales de manera clara y sin sesgos?

- a) Porque deben ser accesibles y comprensibles para una amplia gama de audiencias
- b) Porque siempre deben presentarse en un formato de gráfico de barras
- c) Porque deben centrarse exclusivamente en datos positivos
- d) Porque sólo deben ser comprensibles para expertos en datos

RESPUESTA: a

¿Cuál es el propósito de asegurar que las visualizaciones de datos sean útiles para la toma de decisiones efectivas?

- a) Para aumentar la complejidad de las visualizaciones
- b) Para identificar métricas clave y presentar los datos de manera informativa
- c) Para usar exclusivamente datos históricos
- d) Para enfocarse solo en la estética de la visualización

RESPUESTA: b

¿Cómo afectan las tendencias cambiantes y las dinámicas sociales a la visualización de datos educativos y sociales?

- a) No tienen ningún impacto
- b) Reducen la necesidad de actualización constante
- c) Exigen mantenerse actualizado y ajustar las visualizaciones
- d) Sólo afectan los datos históricos

RESPUESTA: c

¿Qué implica la integración de múltiples fuentes de datos en la visualización?

- a) Ignorar fuentes de datos menores
- b) Unificar todas las fuentes en una sola visualización sin análisis
- c) Combinar efectivamente diversas fuentes para una imagen completa

d) Usar solo la fuente de datos más reciente

RESPUESTA: c

¿Qué papel juega la inteligencia artificial en la automatización de la visualización de datos?

- a) Ninguno, la IA no se utiliza en visualización de datos
- b) Generar visualizaciones automáticamente a partir de conjuntos de datos complejos
- c) Reducir la precisión de las visualizaciones
- d) Limitar el tipo de visualizaciones que se pueden crear

RESPUESTA: b

¿Cómo ayuda la IA en la selección de visualizaciones óptimas?

- a) Recomendando siempre el mismo tipo de gráfico
- b) Analizando la estructura de los datos y objetivos del usuario para recomendar visualizaciones adecuadas
- c) Ignorando las características de los datos
- d) Enfocándose únicamente en visualizaciones en 3D

RESPUESTA: b

¿Qué permite la visualización de datos en tiempo real?

- a) Disminuir la necesidad de análisis de datos
- b) Tomar decisiones basadas en datos de manera más rápida y eficiente
- c) Enfocarse solo en datos históricos
- d) Ignorar las tendencias actuales en los datos

RESPUESTA: b

¿Qué tendencia en la visualización de datos permite explorar datos de manera más inmersiva?

- a) Visualización de datos en dispositivos móviles
- b) Uso de realidad aumentada y visualización 3D
- c) Evitar cualquier visualización interactiva
- d) Limitarse a gráficos de barras tradicionales

RESPUESTA: b

¿Cómo influye la inteligencia artificial en la visualización de datos geoespaciales avanzada?

- a) No tiene ninguna influencia
- b) Complica el proceso de visualización
- c) Permite análisis más avanzados y recomendaciones inteligentes
- d) Limita las visualizaciones a datos no geoespaciales

RESPUESTA: c

¿Cuál es el enfoque de la narración de datos en la visualización?

- a) Comunicar hallazgos y conclusiones de manera efectiva
- b) Presentar datos sin ningún contexto o historia
- c) Usar exclusivamente datos complejos y técnicos
- d) Enfocarse solo en la estética sin contenido

RESPUESTA: a

¿Qué importancia tiene considerar la ética y la privacidad en la visualización de datos?

- a) Ninguna, no afecta la visualización de datos
- b) Es fundamental para manejar los datos de manera responsable
- c) Solo es relevante en ciertos tipos de datos
- d) Debe ser considerado solo en etapas finales del diseño

RESPUESTA: b

¿Cómo se espera que evolucione la automatización en la visualización de datos?

- a) Se reducirá significativamente
- b) Permanecerá sin cambios

- c) Continuará evolucionando para generar rápidamente visualizaciones
- d) Solo se aplicará en campos específicos

RESPUESTA: c)

¿Cuál es una recomendación para prepararse para el futuro de la visualización de datos?

- a) Ignorar las tendencias emergentes
- b) Mantenerse actualizado con las últimas tendencias y tecnologías
- c) Enfocarse únicamente en las habilidades actuales
- d) Evitar la colaboración interdisciplinaria

RESPUESTA: b)

¿Por qué es importante mejorar las habilidades técnicas en visualización de datos?

- a) Para evitar el uso de tecnologías modernas
- b) Las herramientas avanzadas requieren un mayor conocimiento técnico
- c) Solo para utilizar herramientas básicas
- d) Para enfocarse en un solo tipo de visualización

RESPUESTA: b)

¿Cómo puede la narración de datos mejorar la visualización?

- a) Haciéndola más compleja y difícil de entender
- b) Contando historias efectivas para comunicar datos claramente
- c) Eliminando todo contexto narrativo
- d) Centrándose solo en aspectos visuales

RESPUESTA: b

¿Qué se debe considerar al elegir herramientas de visualización de datos avanzadas?

- a) Elegir la herramienta más barata
- b) Ignorar la integración de inteligencia artificial
- c) La capacidad de las herramientas para aprovechar IA y otras tecnologías
- d) Usar solo herramientas tradicionales

RESPUESTA: c

REFERENCIAS

BIBLIOGRÁFICAS

1. Hernández-Vásquez A, Chacón-Torrico H. Manipulación, análisis y visualización de datos de la encuesta demográfica y de salud familiar con el programa R. *Rev Peru Med Exp Salud Publica*. 2019;36:128-133. doi:10.17843/rpmesp.2019.361.4062
2. Fernández Lizana MI. Ventajas de R como herramienta para el Análisis y Visualización de datos en Ciencias Sociales. *Revista Científica de la UCSA*. 2020;7(2):97-111. doi:10.18004/ucsa/2409-8752/2020.007.02.097
3. Morales Vargas A. Datos abiertos y visualización de información en sitios web de universidades chilenas: una asignatura pendiente. In: Brossi L, Dodds T, eds. *Visualización de datos: Periodismo y Comunicación en la era de la información visual*. Editorial Universitaria; 2019:15-26. Accessed January 2, 2024. <http://repositorio.uchile.cl/handle/2250/142788> <http://www.dvine.cl>
4. Connor H. John Graunt F.R.S. (1620-74): The founding father of human demography, epidemiology and vital statistics. *J Med Biogr*. Published online February 15, 2022:09677720221079826. doi:10.1177/09677720221079826
5. Bellhouse DR. *The Flawed Genius of William Playfair: The Story of the Father of Statistical Graphics*. University of Toronto Press; 2023.
6. Millán-Martínez P, Oller R. A Graphical EDA Tool with ggplot2: brinton. *R Journal*. 2020;12(2). Accessed January 2, 2024. <https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=20734859&AN=148792322&h=%2FcS->

R0TZ7Tr%2FgVhjbXwgzBWcgXylyDkIR5Bk7BZK6Y5ldx-7Qh7HACw1cGEv9hDtMPeiipLCoyH35X%2BpvnSkVqo-Q%3D%3D&crl=c

7. Liñán LJ, Mendoza NMG de. Papeles en bits: los orígenes de la difusión digital de arquitectura en las primeras versiones web de Metropolis, Domus, Abitare y El Croquis. *Bitácora Arquitectura*. 2019;(43):4-11. doi:10.22201/fa.14058901p.2020.43.72952
8. Lopezosa C, Pérez-Montoro M, Guallar J. Visualización de datos y medios de comunicación: scoping review. Published online January 2023. Accessed January 2, 2024. <https://diposit.ub.edu/dspace/handle/2445/192602>
9. Casanova H. Graficación Estadística y Visualización de Datos. *Ingeniería*. 2017;21(3):54-75.
10. Valero Sancho JL. La visualización de datos. Published online 2014. Accessed January 2, 2024. <https://idus.us.es/handle/11441/66886>
11. Sancho JLV, Domínguez JC, Ochoa BEM. Aproximación a una taxonomía de la visualización de datos. *Revista Latina de Comunicación Social*. 2014;(69):486-507. doi:10.4185/RLCS-2014-1021
12. Sarango AH, Haro DL, García AC, Sánchez PM, Lapo NL. Estado emocional y desempeño laboral: un estudio asociativo multivariante en profesores universitarios en teletrabajo bajo Covid-19: Emotional state and job performance: a multivariate associative study in teleworking university professors under Covid-19. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*. 2023;4(1):2430-2443.

13. Calderón CA, Ibáñez DB, Castro C. Analítica y visualización de datos en Twitter. Editorial UOC; 2018.
14. Padilla PIC. La visualización de datos a través del motion graphics y el storytelling. *Tsantsa Revista de Investigaciones artísticas*. 2019;(7):59-68.
15. González JS, Arias SM. Periodismo y Nuevos Medios: Perspectivas y Retos. Editorial GEDISA; 2020.
16. Fernández Lizana ML. Ventajas de R como herramienta para el Análisis y Visualización de datos en Ciencias Sociales. *Revista Científica de la UCSA*. 2020;7(2):97-111. doi:10.18004/ucsa/2409-8752/2020.007.02.097
17. Haro Sarango A. Impacto del COVID-19 en el turismo sostenible: un estudio desde la perspectiva estadística financiera en hoteles y alojamientos turísticos. *Revista Cubana de Administración Pública y Empresarial*. 2021;5(3):6.
18. Riva MPP, Sarango AFH, Córdova CAB, Pérez MGA, Mejía CAS, others. Educación financiera basada en los conocimientos financieros: un análisis en los beneficiarios del programa Campo Emprende. *Tesla Revista Científica*. 2023;3(2):e128-e128.
19. Pérez-Montoro M. La disciplina de la visualización de datos en 2017. Published online 2018. Accessed January 2, 2024. <https://diposit.ub.edu/dspace/handle/2445/122161>
20. Castro SM, Larrea ML, Urribarri DK, Ganuza ML, Escarza S. Métricas, técnicas y semántica para la visualización de datos. In: ; 2018. Accessed January 2, 2024.

<http://sedici.unlp.edu.ar/handle/10915/67456>

21. Schab E, Rivera R, Bracco L, et al. Minería de datos y visualización de información. In: ; 2018. Accessed January 2, 2024. <http://sedici.unlp.edu.ar/handle/10915/67382>

22. Jaramillo HAL, Pinos CAE, Sarango AFH, Román HDO. Histograma y distribución normal: Shapiro-Wilk y Kolmogorov Smirnov aplicado en SPSS: Histogram and normal distribution: Shapiro-Wilk and Kolmogorov Smirnov applied in SPSS. LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades. 2023;4(4):596-607. doi:10.56712/latam.v4i4.1242

23. Ladrón-de-Guevara-Alcalá C. Gráficos estadísticos en Educación Primaria: Métodos Interactivos. Published online July 4, 2017. Accessed January 2, 2024. <http://crea.ujaen.es/jspui/handle/10953.1/5731>

24. Díaz Oscullo DC. Los organizadores gráficos interactivos como herramienta didáctica en el proceso enseñanza aprendizaje de lengua y literatura en la unidad educativa del milenio “San Gabriel de Piquiucho.” bachelorThesis. 2023. Accessed January 2, 2024. <http://repositorio.utn.edu.ec/handle/123456789/14981>

25. Font Andreu J, Hernández Abad F, Capdevila Pagés R, Ochoa Vives M. Creación de una base de datos gráficos (BDG) de modelos geométricos estáticos o dinámicos de prototipos virtuales. In: ; 2005:0-1. Accessed January 2, 2024. <https://upcommons.upc.edu/handle/2117/21957>

26. Alcívar SJN, Solórzano RML, Solís OPL, Sarango AFH. Administración de la cartera impaga en la rentabi-

lidad: una aplicación estadística clasificatoria en bancos: Nonperforming portfolio management in profitability: a statistical application to bank rating. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*. 2023;4(3):1603-1616. doi:10.56712/latam.v4i3.1185

27. Sarango AFH, Yacelga APM, Sevilla RMN, Sailema MEC, Lescano JCP. Inteligencia de negocios en la gestión empresarial: un análisis a las investigaciones científicas mundiales: Business intelligence in business management: a review of worldwide scientific research. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*. 2023;4(1):3367-3382.

28. Khan AS. Engineering a Scalable and Interactive Web Tool for JRO Data Exploration Using Python's Bokeh Library. Thesis. University of Illinois at Urbana-Champaign; 2021. Accessed January 2, 2024. <https://hdl.handle.net/2142/113816>

29. Dabbas E. Interactive Dashboards and Data Apps with Plotly and Dash: Harness the Power of a Fully Fledged Frontend Web Framework in Python – No JavaScript Required. Packt Publishing Ltd; 2021.

30. Franco EA, Martínez REL, Domínguez VHM. Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. *Revista de Educación a Distancia (RED)*. 2021;21(66). doi:10.6018/red.463561

31. Torres M, Salazar FG, Paz K. Métodos de recolección de datos para una investigación. Published online September 19, 2019. Accessed January

- 2, 2024. <http://148.202.167.116:8080/xmlui/handle/123456789/2817>
32. Uriz Martín M. Mejora de los algoritmos de minería de datos: combinación de clasificadores, preprocesamiento y sus aplicaciones. <http://purl.org/dc/dcmitype/Text>. Universidad Pública de Navarra; 2021. Accessed January 2, 2024. <https://dialnet.unirioja.es/servlet/tesis?codigo=299441>
33. Sposito OM, Blanco GE, Matteo LR. Técnicas de preprocesamiento de datos en modelos no supervisados aplicados al estudio genético de la raza Aberdeen Angus. ISSN 2525-1333. Published online June 27, 2022. Accessed January 2, 2024. <http://repositoriocyt.unlam.edu.ar/handle/123456789/1210>
34. Castillo Lamas A. Metodologías de aprendizaje basadas en Deep Learning: preprocesamiento de datos y postprocesamiento. Universidad de Granada; 2023. Accessed January 2, 2024. <https://digibug.ugr.es/handle/10481/81261>
35. Loayza Blanco CI. Preprocesamiento de datos mediante herramientas de Big Data. Published online 2023. Accessed January 2, 2024. <https://repositorio.unsaac.edu.pe/handle/20.500.12918/7930>
36. Hidalgo A. Técnicas estadísticas en el análisis cuantitativo de datos. *Revista Sigma*. 2019;15(1):28-44.
37. Sierra JAG, Moheno JM, Hernández BCS. Evaluación de la inclusión financiera en el estado de Hidalgo, México: aplicación de un Análisis Exploratorio de Datos Espaciales. *Mundo FESC*. 2021;11(S3):7-20.

38. Buzai GD. Análisis Exploratorio de Datos Espaciales de los Condicionantes Sociales de la Salud en la cuenca del río Luján, Argentina. Published online June 2019. Accessed January 2, 2024. <https://ri.conicet.gov.ar/handle/11336/130905>
39. Vargas Diaz RE, Galindo Pacheco JR, Henao RG, Vargas Diaz RE, Galindo Pacheco JR, Henao RG. Análisis espacial de propiedades físicas del suelo de la Hoya del río Suárez, Boyacá - Santander (Colombia), utilizando componentes principales. *Investigación Agraria*. 2021;23(1):8-16. doi:10.18004/investig.agrar.2021.junio.2301657
40. Gallego L, Araque O, Gallego L, Araque O. Variables de Influencia en la Capacidad de Aprendizaje. Un Análisis por Conglomerados y Componentes Principales. *Información tecnológica*. 2019;30(2):257-264. doi:10.4067/S0718-07642019000200257
41. Justiz MS, Pérez DJV, Álvarez ENC, Pérez RM. Identificación de patrones socio-demográficos desde una perspectiva espacial en la provincia Cienfuegos. *Universidad y Sociedad*. 2023;15(S1):398-403.
42. Bernate J, Vargas Guativa JA. Desafíos y tendencias del siglo XXI en la educación superior. *Revista de ciencias sociales*. 2020;26(Extra 2):141-154.
43. Salas Bustos DA. Enseñanza remota y redes sociales: estrategias y desafíos para conformar comunidades de aprendizaje. *Revista Andina de Educación*. 2020;4(1):36-42. doi:10.32719/26312816.2021.4.1.5
44. Álvarez-de-Sotomayor ID, Carril PCM, Sanmamed

MG. Hábitos de uso de las redes sociales en la adolescencia: desafíos educativos. *Revista Fuentes*. 2021;23(3):280-295. doi:10.12795/revistafuentes.2021.15691

45. Briceño JCP, Sarango AFH, Fiallos MFN, Jiménez AGM, Molina PGR. Un estudio teórico y bibliométrico sobre la gestión del talento humano en el ámbito empresarial: Tendencias, desafíos y perspectivas actuales: A Theoretical and Bibliometric Study on Human Talent Management in Business: Trends, Challenges and Current Perspectives. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*. 2023;4(3):1059-1075. doi:10.56712/latam.v4i3.1131

46. Haro AF, Vanegas C. Evaluación de las capacidades de innovación: un estudio de caso en una universidad colombiana. *Avances de investigación*. 2021;8(1):7-24.

47. Erazo-Luzuriaga AF, Ramos-Secaira FM, Galarza-Sánchez PC, Boné-Andrade MF. La inteligencia artificial aplicada a la optimización de programas informáticos. *Journal of Economic and Social Science Research*. 2023;3(1):48-63. doi:10.55813/gaea/jessr/v3/n1/61

48. Brenes Carranza JA, Martínez Porras A, Quesada López CU, Jenkins Coronas M. Sistemas de apoyo a la toma de decisiones que usan inteligencia artificial en la agricultura de precisión. *Revista Ibérica de Sistemas y Tecnologías de la Información*, núm E28, pp 217-229. Published online 2020. Accessed January 2, 2024. <https://kerwa.ucr.ac.cr/handle/10669/81549>

49. Pereira Júnior A, Donadon Homem TP, Oliveira Teixeira F, Pereira Júnior A, Donadon Homem TP, Oliveira

Teixeira F. Aplicación de inteligencia artificial para monitorear el uso de mascarillas de protección. *Revista Científica General José María Córdova*. 2021;19(33):205-222. doi:10.21830/19006586.725

ISBN: 978-9942-7173-6-8



9 789942 717368