**Un dibujo animado con letras

Descripción generada automáticamente con confianza mediaCírculo

Descripción generada automáticamente**

**UNIVERSIDAD TECNOLÓGICA DE PANAMÁ**

**FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES**

**DEPARTAMENTO DE INGENIERÍA DE SOFTWARE**

**CARRERA LICENCIATURA EN INGENIERÍA DE SOFTWARE**

**ESTADISTICA**

**TITULO:**

**PROYECTO DE INVESTIGACIÓN / INFORME**

**INTEGRANTES:**

**CRISTOFER JIMENEZ 20-70-7100**

**PROFESOR:**

**JUAN CASTILLO**

**SEMESTRE I, 2023**

**Introducción**

En este viaje de análisis de datos, exploraremos diez fascinantes bases de datos. Cada una ha sido elegida por su relevancia y curiosidad, abarcando temas como entretenimiento, tecnología, salud y seguridad. A través de este análisis, buscamos obtener una visión más profunda de cada tema y su impacto en nuestras vidas. ¡Acompáñanos en este emocionante recorrido hacia el conocimiento a través de los datos!

**Latin America - Femicides & Maternal Mortality**

**1. Preliminares:** Se realizará un análisis de datos utilizando el conjunto de datos "Latin America - Femicides & Maternal Mortality". El objetivo del análisis es examinar la relación entre variables como la población femenina, la esperanza de vida de las mujeres, la tasa de mortalidad materna, el índice de paz global y la cantidad de femicidios en países de América Latina y el Caribe.

**2. Análisis Visual:** Se crearán gráficos de barra, gráficos de pastel y otros gráficos adecuados para visualizar y entender los datos. Por ejemplo, gráficos de barra para comparar la esperanza de vida de las mujeres en diferentes países y gráficos de pastel para mostrar la distribución de femicidios en la región.

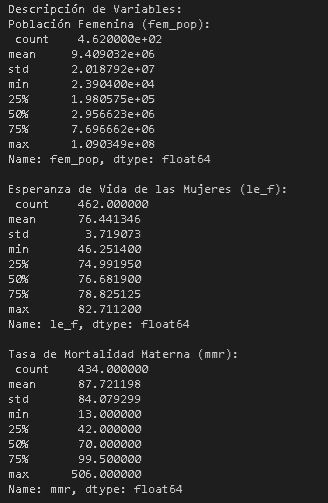
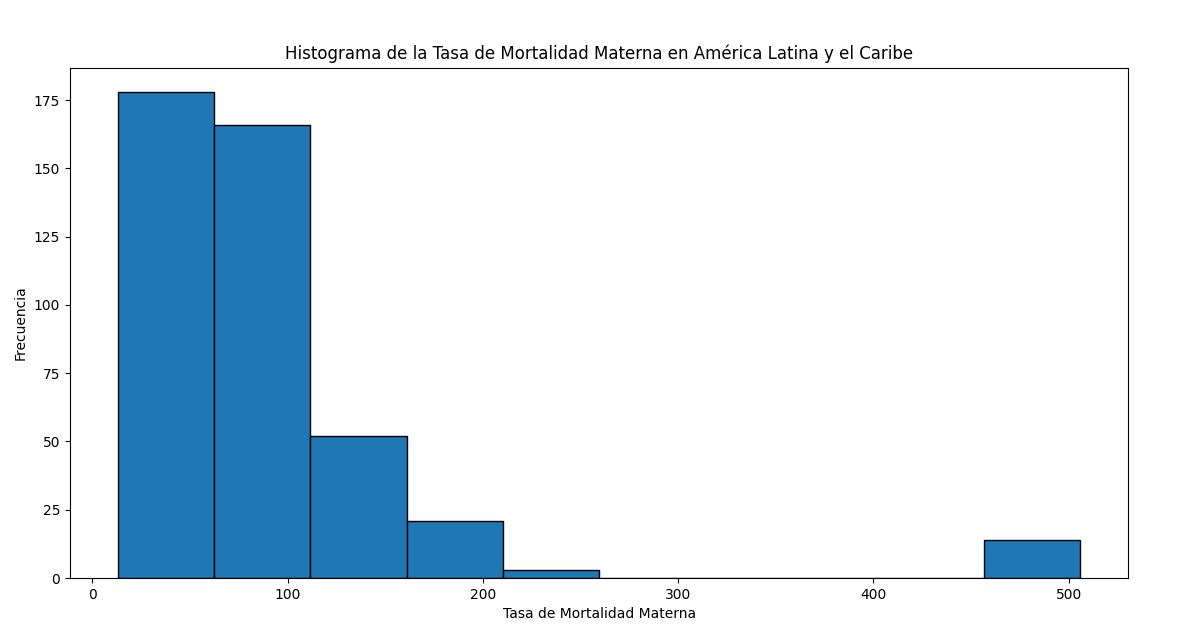
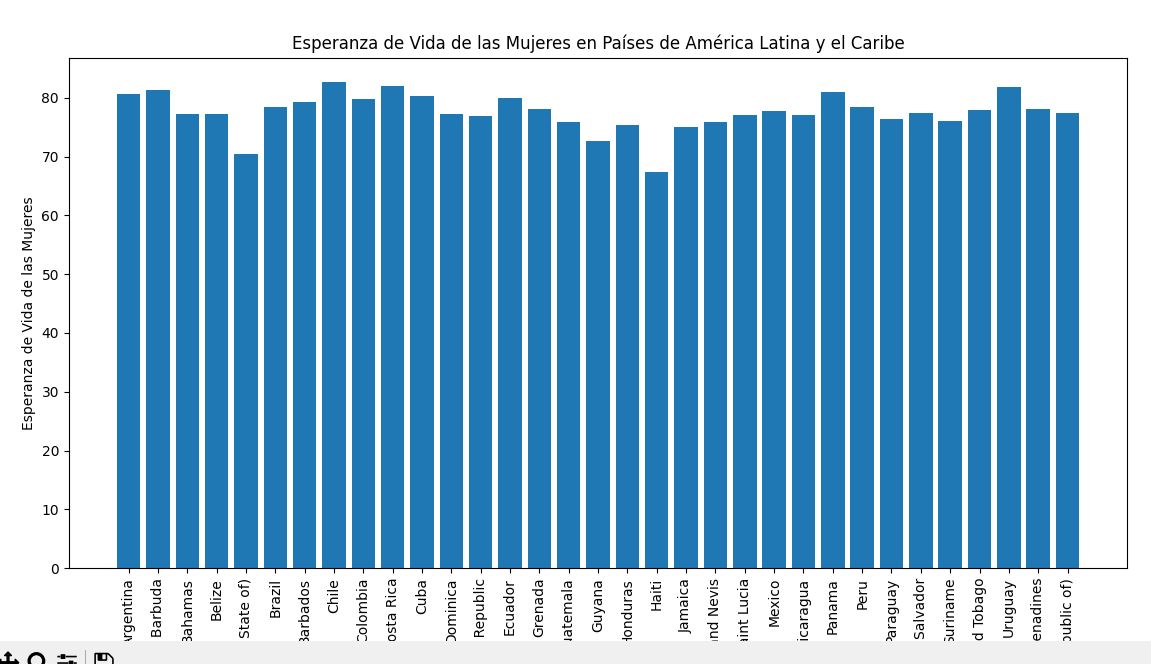
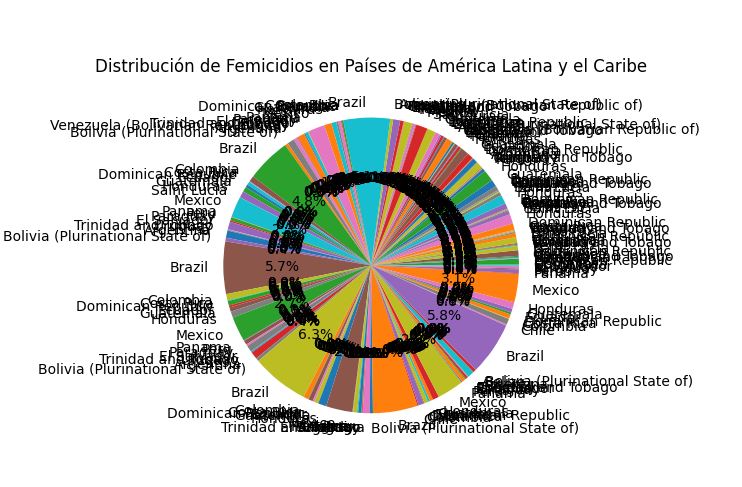
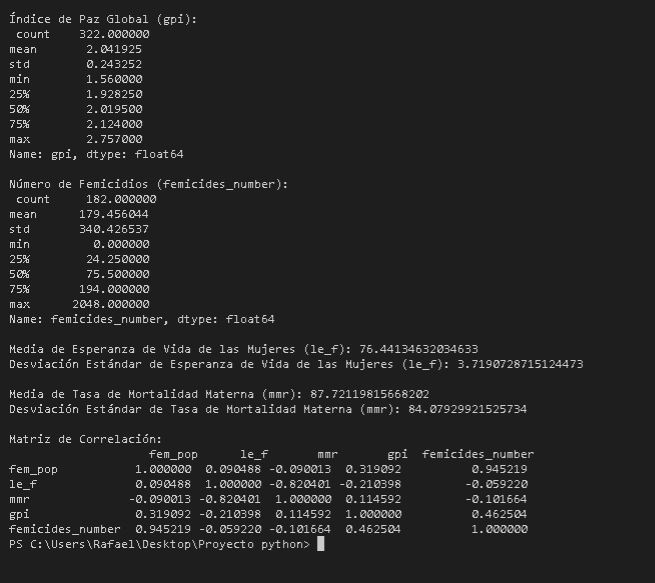
**3. Descripción de Variables:** Se describirá cada variable presente en el conjunto de datos. Por ejemplo:

* fem\_pop: Población femenina en América Latina y el Caribe.
* le\_f: Esperanza de vida de las mujeres en años.
* mmr: Tasa de mortalidad materna en número de muertes por cada 100,000 nacimientos vivos.
* gpi: Índice de paz global, una medida de la paz de los países.
* femicides\_number: Número absoluto de femicidios en cada país.

**4. Aproximación y Resolución de Problemas:** Buscar patrones o relaciones entre la tasa de mortalidad materna y el índice de paz global en la región.

**5. Histogramas:** Se generará un histograma para mostrar la distribución de la tasa de mortalidad materna en la región. En el eje X, se presentarán los intervalos de tasas y en el eje Y, se mostrará la frecuencia de países en cada intervalo.

**6. Paretograma o Priorización de Muestra:** Si es necesario, se utilizará un Paretograma u otro método para priorizar o segmentar la muestra de países en función de ciertas características, como la cantidad de femicidios.

**7. Media y Desviación Estándar:** Se calcularán la media y la desviación estándar de variables como la esperanza de vida de las mujeres y la tasa de mortalidad materna para tener una idea de las tendencias generales en la región.  
  
**8. Resultados Obtenidos:** 

**Conclusión:** El análisis de datos de "Latin America - Femicides & Maternal Mortality" ha permitido explorar diversas relaciones y patrones en variables relacionadas con la población femenina, la mortalidad materna y la paz global en países de América Latina y el Caribe. Los resultados y conclusiones del análisis proporcionan una visión más profunda de los desafíos y oportunidades en la región y pueden ser útiles para informar políticas y programas relacionados con la salud y el bienestar de las mujeres.

Este informe presenta un resumen de los análisis realizados y destaca los hallazgos más relevantes obtenidos a partir del conjunto de datos.

**Video Games Sales Dataset"**

**1. Preliminares:** El objetivo es analizar el conjunto de datos "Video Games Sales Dataset" que contiene información sobre ventas de videojuegos, puntuaciones de críticos y usuarios, y otra información relevante. El análisis permitirá obtener información sobre los videojuegos más populares, las tendencias de ventas y la relación entre las puntuaciones y el rendimiento comercial.

**2. Análisis Visual:** Se crearán gráficos, como gráficos de barras y de dispersión, para visualizar y entender los datos. Por ejemplo, gráficos de barras para comparar las ventas globales de videojuegos en diferentes plataformas y gráficos de dispersión para examinar la relación entre las puntuaciones de críticos y las ventas globales.

**3. Descripción de Variables:** Se describirá cada variable presente en el conjunto de datos. Por ejemplo:

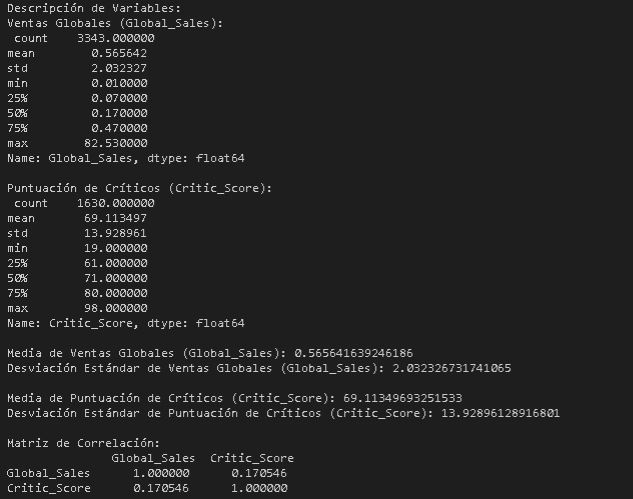
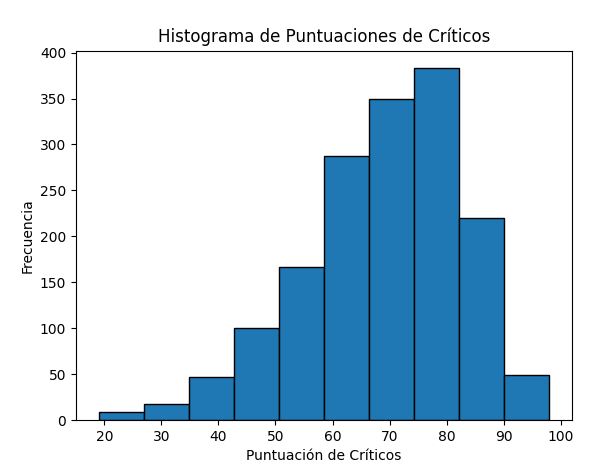
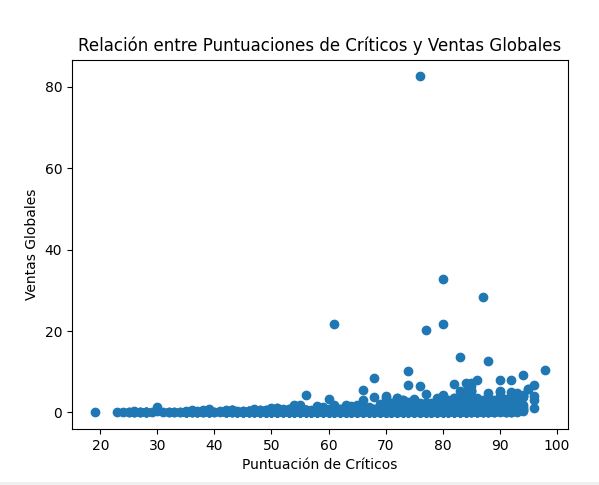
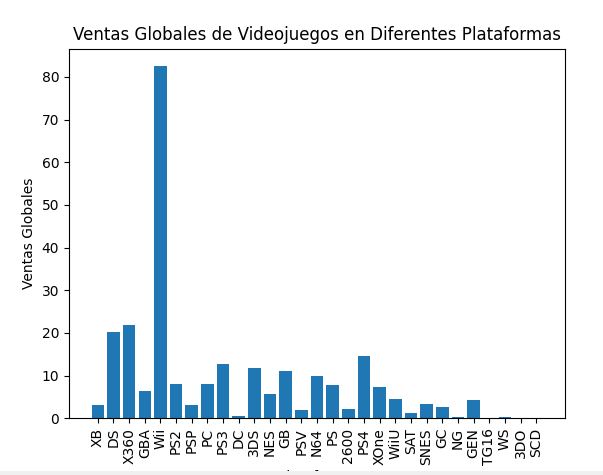
* Name: Nombre del videojuego.
* Platform: Plataforma en la que se lanzó el videojuego.
* Year\_of\_Release: Año de lanzamiento del videojuego.
* Genre: Género del videojuego.
* Publisher: Editor del videojuego.
* NA\_Sales, EU\_Sales, JP\_Sales, Other\_Sales, Global\_Sales: Ventas del videojuego en diferentes regiones.
* Critic\_Score: Puntuación agregada compilada por el personal de Metacritic.
* Critic\_Count: Número de críticos utilizados para calcular la puntuación de los críticos.
* User\_Score: Puntuación otorgada por los suscriptores de Metacritic.
* User\_Count: Número de usuarios que proporcionaron la puntuación de usuario.
* Developer: Desarrollador del videojuego.
* Rating: Clasificación de la ESRB (Entertainment Software Rating Board).

**4. Aproximación y Resolución de Problemas:** Identificar los géneros de videojuegos más populares o analizar la relación entre las puntuaciones y las ventas.

**5. Histogramas:** Se generará un histograma para mostrar la distribución de las puntuaciones de críticos y usuarios.

**6. Paretograma o Priorización de Muestra:** Si es necesario, se utilizará un Paretograma u otro método para priorizar o segmentar la muestra de videojuegos en función de ciertas características, como las ventas o las puntuaciones.

**7. Media y Desviación Estándar:** Se calculará la media y la desviación estándar de las ventas globales para tener una idea de las tendencias generales en el conjunto de datos. **8. Resultados obtenidos:**

.  


**Conclusión:** El análisis de datos del conjunto "Video Games Sales Dataset" ha proporcionado información valiosa sobre las ventas de videojuegos, las puntuaciones de críticos y usuarios, y otros factores relevantes en la industria de los videojuegos. Los resultados obtenidos ayudarán a comprender mejor las tendencias y patrones en el rendimiento comercial de los videojuegos y pueden ser útiles para la toma de decisiones en la industria.

**Total Sales Local Store**

**1. Preliminares:** El objetivo es analizar el conjunto de datos "Total Sales Local Store", que contiene registros de ventas con un identificador único (ID), la fecha de la transacción (DATE), el monto total de ventas realizadas en un día (AMOUNT) y los impuestos asociados con las ventas realizadas en ese día (TAXES).

**2. Análisis Visual:** Se crearán gráficos para visualizar y entender los datos. Por ejemplo, un gráfico de líneas para representar la tendencia de ventas a lo largo del tiempo o un gráfico de barras para mostrar la distribución de las ventas y los impuestos por día.

**3. Descripción de Variables:** Se describirá cada variable presente en el conjunto de datos. Por ejemplo:

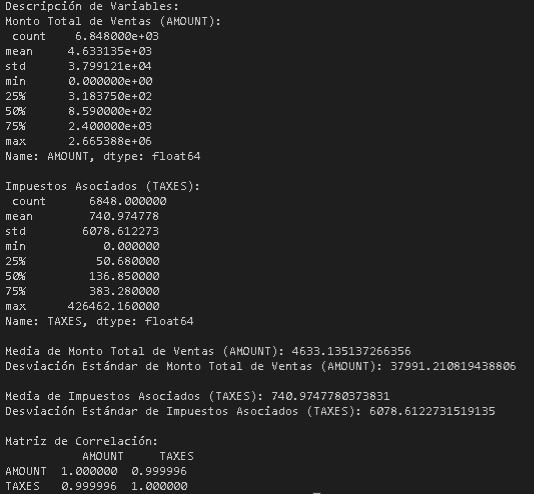
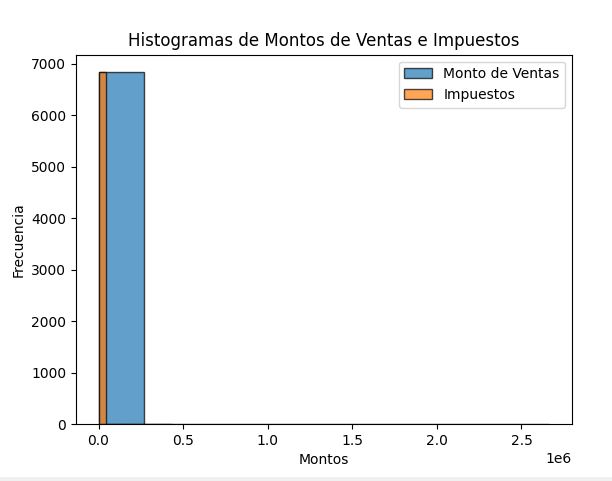
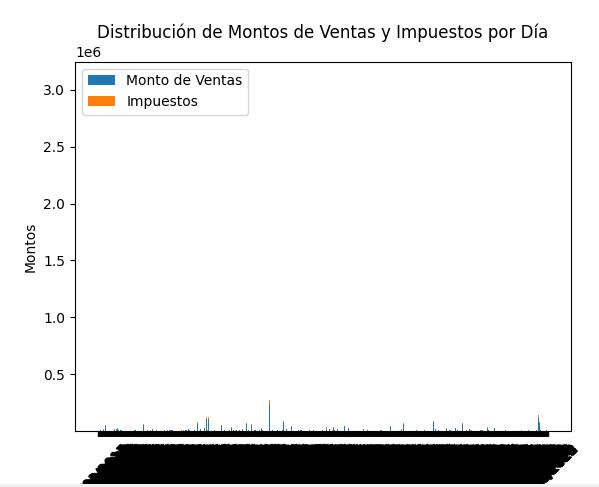
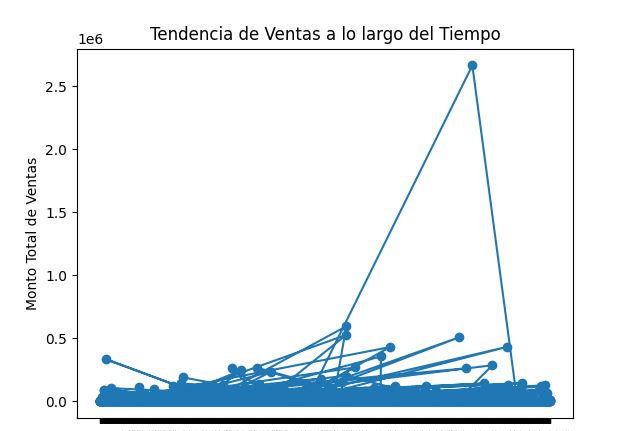
* ID: Identificador único para cada registro de venta.
* DATE: Fecha de la transacción de ventas.
* AMOUNT: Monto total de ventas realizadas en un día.
* TAXES: Impuestos asociados con las ventas realizadas en un día.

**4. Aproximación y Resolución de Problemas:** Identificar las fechas de ventas más lucrativas o analizar la relación entre el monto de ventas y los impuestos.

**5. Histogramas:** Se generará un histograma para mostrar la distribución de los montos de ventas y los impuestos.

**6. Media y Desviación Estándar:** Se calculará la media y la desviación estándar de los montos de ventas y los impuestos para tener una idea de las tendencias generales en el conjunto de datos.

**7. Análisis de Correlación:** Se realizará un análisis de correlación para examinar la relación entre los montos de ventas y los impuestos.



**Conclusión:** El análisis de datos del conjunto "Total Sales Local Store" ha proporcionado información sobre las ventas y los impuestos asociados en la tienda local. Los resultados obtenidos ayudarán a comprender mejor las tendencias y patrones en las ventas y la relación entre los montos de ventas y los impuestos. Estos hallazgos pueden ser útiles para la toma de decisiones en la tienda y para la planificación financiera.

**Netflix User Base Dataset**

**1. Preliminares:** El objetivo es analizar el conjunto de datos "Netflix User Base Dataset", que contiene información sobre una muestra de usuarios de Netflix. Cada fila representa un usuario único, identificado por su ID de usuario. El conjunto de datos incluye detalles como el tipo de suscripción del usuario (Basic, Standard o Premium), los ingresos mensuales generados por su suscripción, la fecha en que se unieron a Netflix (Join Date), la fecha de su último pago (Last Payment Date) y el país en el que se encuentran.

Además, el conjunto de datos proporciona información sobre el comportamiento y las preferencias de los usuarios, como el tipo de dispositivo que utilizan para ver contenido (Device Type), el tiempo total de visualización (Total Watch Time) y el estado de su cuenta (Account Status).

**2. Análisis Visual:** Se crearán gráficos para visualizar y entender los datos. Por ejemplo, se pueden realizar gráficos de barras para mostrar la distribución de los tipos de suscripción o diagramas de dispersión para analizar la relación entre los ingresos mensuales y el tiempo total de visualización.

**3. Descripción de Variables:** Se describirá cada variable presente en el conjunto de datos. Por ejemplo:

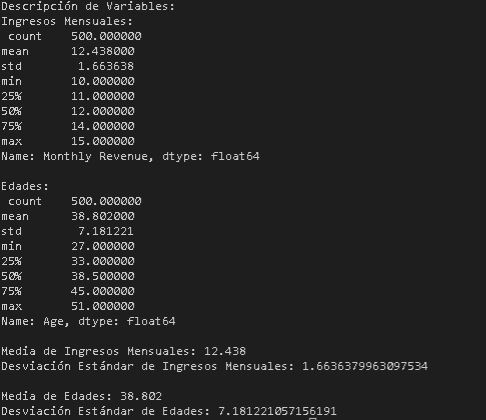
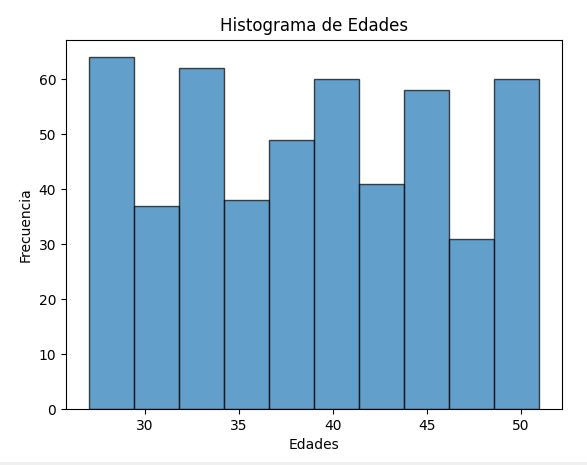
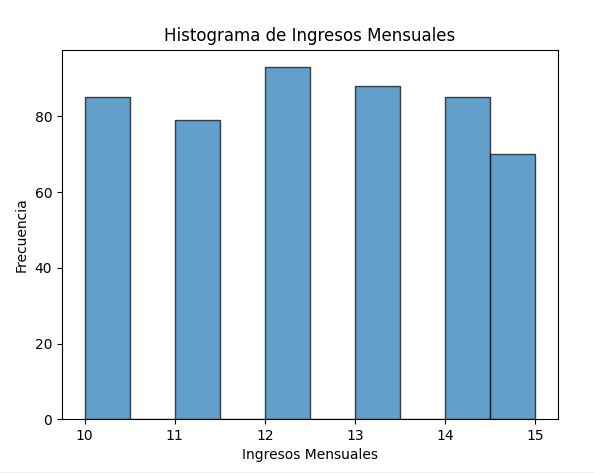
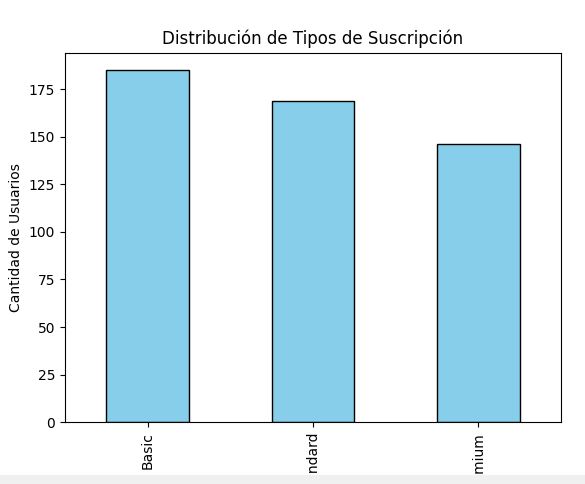
* User ID: Identificador único para cada usuario.
* Subscription Type: Tipo de suscripción del usuario (Basic, Standard o Premium).
* Monthly Revenue: Ingresos mensuales generados por la suscripción del usuario.
* Join Date: Fecha en que el usuario se unió a Netflix.
* Last Payment Date: Fecha del último pago realizado por el usuario.
* Country: País en el que se encuentra el usuario.
* Age: Edad del usuario.
* Gender: Género del usuario.
* Device: Tipo de dispositivo que utiliza para ver contenido (Smartphone, Tablet, Smart TV, etc.).
* Plan Duration: Duración del plan de suscripción (por ejemplo, 1 mes).

**4. Aproximación y Resolución de Problemas:** Se determinará el enfoque del análisis en función de los objetivos específicos, como identificar patrones de suscripción más populares o analizar la relación entre el tiempo total de visualización y los ingresos mensuales.

**5. Histogramas:** Se generarán histogramas para mostrar la distribución de las variables numéricas, como los ingresos mensuales, la edad y el tiempo total de visualización.

**6. Media y Desviación Estándar:** Se calculará la media y la desviación estándar de las variables numéricas para tener una idea de las tendencias generales en el conjunto de datos.

**7. Análisis de Correlación:** Se realizará un análisis de correlación para examinar la relación entre variables, como los ingresos mensuales y el tiempo total de visualización.

****

**Conclusión:** El análisis de datos del conjunto "Netflix User Base Dataset" ha proporcionado información valiosa sobre los patrones de suscripción, ingresos mensuales y comportamiento de los usuarios de Netflix. Los resultados obtenidos pueden ser útiles para comprender las preferencias de los usuarios y mejorar las estrategias de negocio en el ámbito de servicios de streaming de video.

**Coursera Course Dataset**

**1. Preliminares:** El objetivo es analizar el conjunto de datos "Coursera Course Dataset", que contiene información sobre diferentes cursos disponibles en Coursera. Cada fila representa un curso único, identificado por su ID. El conjunto de datos incluye detalles como el título del curso (course\_title), la organización que ofrece el curso (course\_organization), el tipo de certificación disponible (course\_Certificate\_type), la calificación del curso (course\_rating), el nivel de dificultad del curso (course\_difficulty) y el número de estudiantes inscritos en el curso (course\_students\_enrolled).

**2. Análisis Visual:** Se crearán gráficos para visualizar y entender los datos. Por ejemplo, se pueden realizar gráficos de barras para mostrar la distribución de los niveles de dificultad o diagramas de dispersión para analizar la relación entre la calificación del curso y el número de estudiantes inscritos.

Descripción de Variables:

* id: Número utilizado para identificar el conjunto de datos.
* course\_title: Título del curso.
* course\_organization: Organización que ofrece el curso.
* course\_Certificate\_type: Detalles sobre las diferentes certificaciones disponibles en el curso.
* course\_rating: Calificación asociada con cada curso.
* course\_difficulty: Nivel de dificultad del curso (por ejemplo, principiante, intermedio, avanzado).
* course\_students\_enrolled: Número de estudiantes inscritos en el curso.

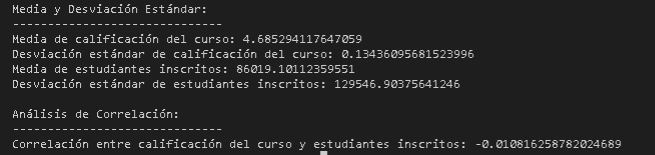
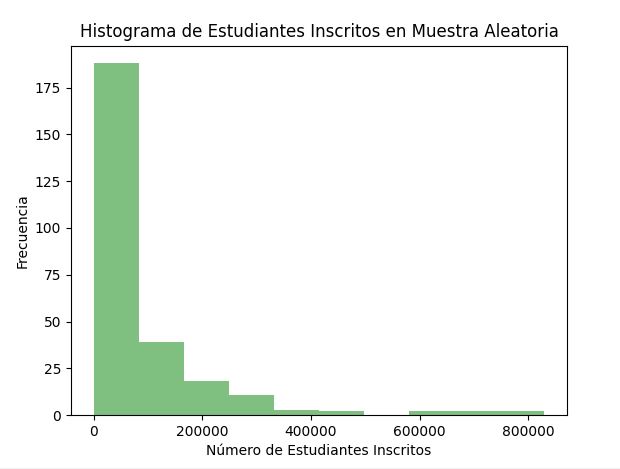
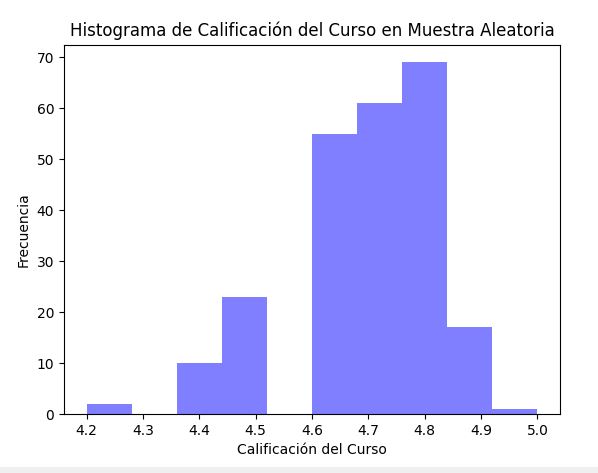
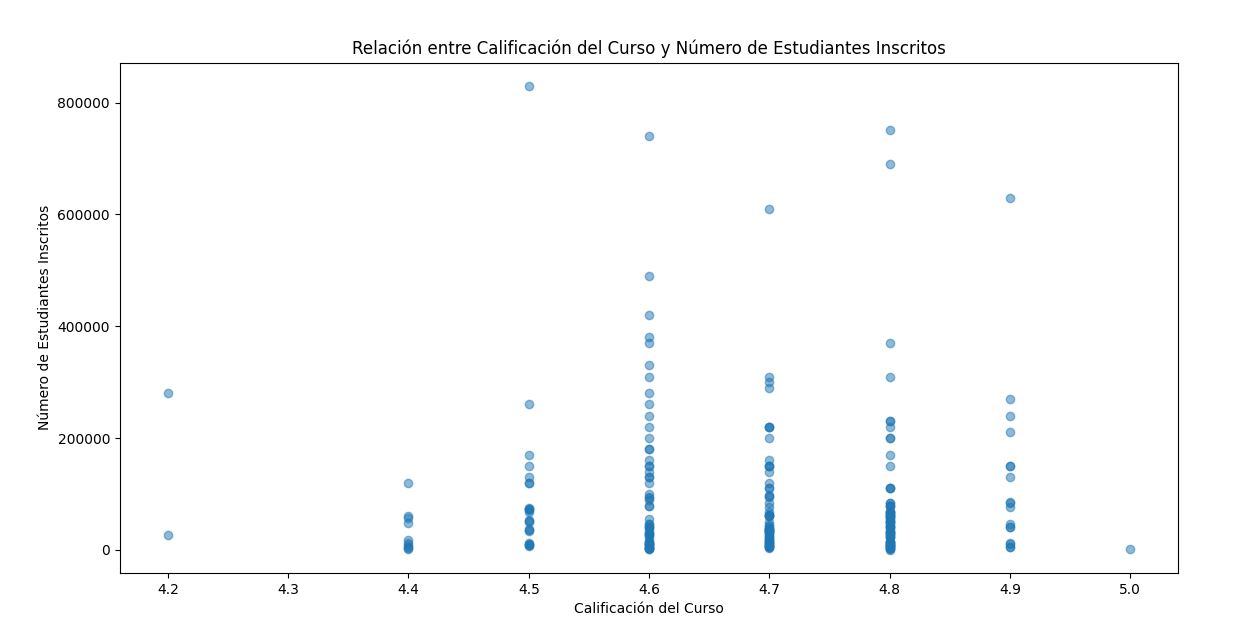
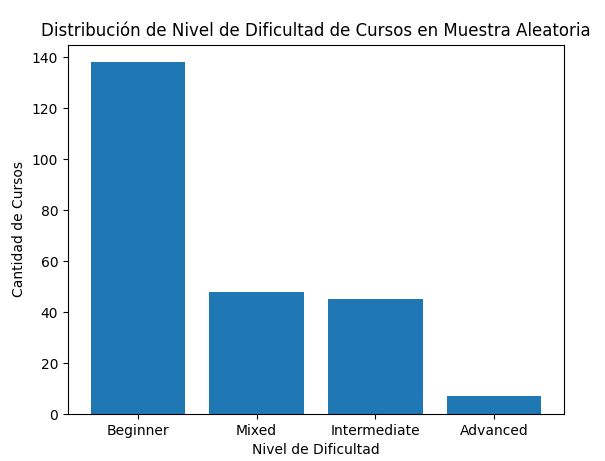
**Aproximación y Resolución de Problemas:** Identificar los cursos mejor calificados o analizar la relación entre el nivel de dificultad y el número de estudiantes inscritos.

**Histogramas:** Se generarán histogramas para mostrar la distribución de las variables numéricas, como la calificación del curso y el número de estudiantes inscritos.

**Media y Desviación Estándar:** Se calculará la media y la desviación estándar de las variables numéricas para tener una idea de las tendencias generales en el conjunto de datos.

**Análisis de Correlación:** Se realizará un análisis de correlación para examinar la relación entre variables, como la calificación del curso y el número de estudiantes inscritos.

**Datos obtenidos:**



**Conclusión:** El análisis de datos del conjunto "Coursera Course Dataset" ha proporcionado información valiosa sobre los diferentes cursos ofrecidos en Coursera, sus calificaciones y la cantidad de estudiantes inscritos. Los resultados obtenidos pueden ser útiles para comprender la popularidad y la dificultad de los cursos, lo que puede ayudar a los estudiantes a seleccionar los cursos más adecuados para sus necesidades.

**Cyber Security Indexes**

**1. Preliminares:** El objetivo es analizar el conjunto de datos "Cyber Security Indexes", que proporciona información sobre la situación actual de la ciberseguridad en todo el mundo. El conjunto de datos incluye cuatro indicadores que ilustran el nivel de exposición a ciberdelitos (CEI), el compromiso de los países con la ciberseguridad a nivel global (GCI), la preparación de los países para abordar amenazas cibernéticas y gestionar incidentes (NCSI) y el nivel de desarrollo digital (DDL) de cada país.

Los datos están disponibles para 193 países y territorios, agrupados en cinco regiones geográficas: África, América del Norte, América del Sur, Europa y Asia-Pacífico.

**2. Análisis Visual:** Se utilizarán gráficos y visualizaciones para comprender mejor los datos. Por ejemplo, se pueden crear gráficos de barras para comparar los índices de ciberseguridad en diferentes regiones o diagramas de dispersión para explorar posibles relaciones entre los indicadores.

**3. Descripción de Variables:** A continuación, se describirán las variables presentes en el conjunto de datos:

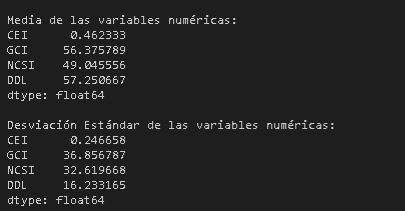
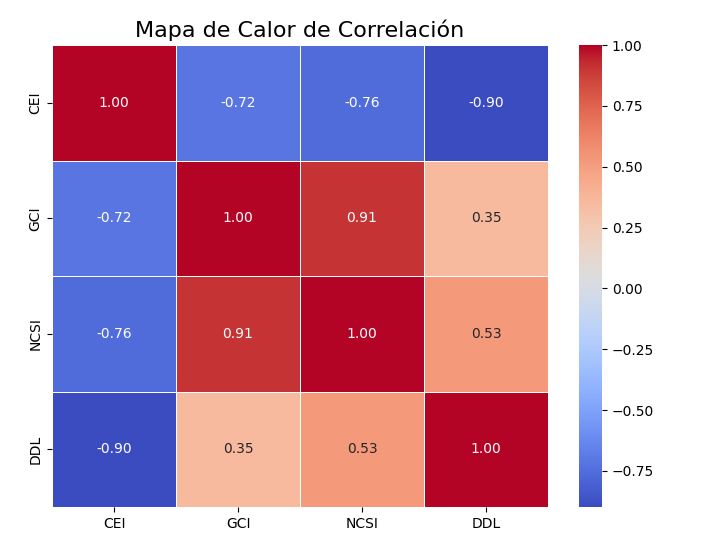
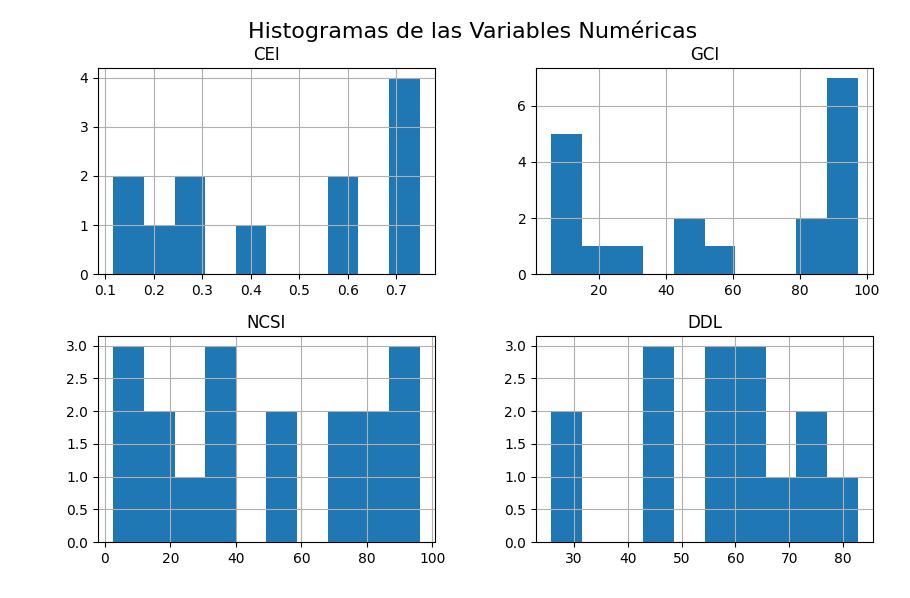
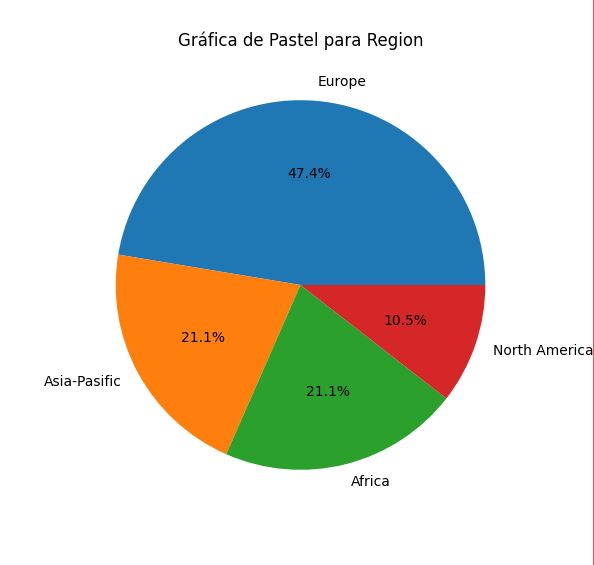
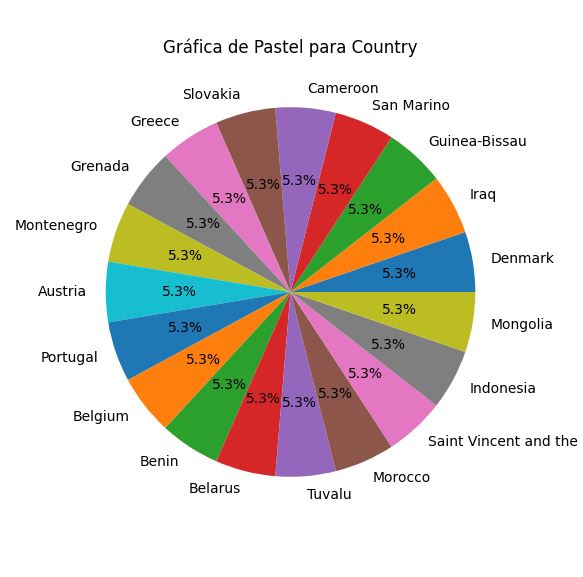
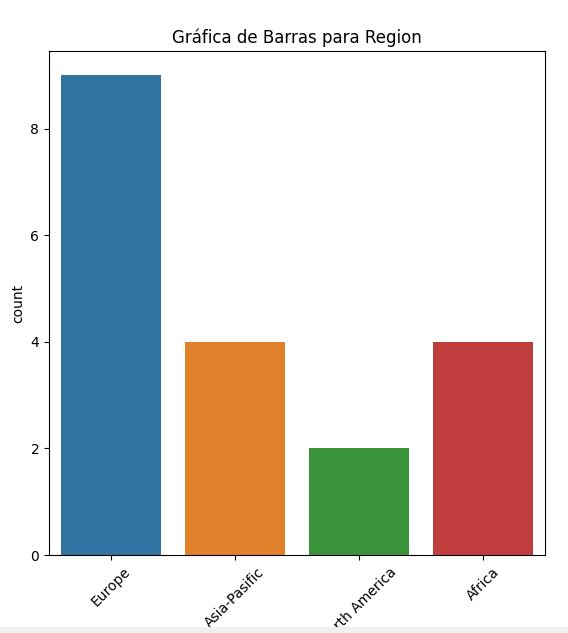
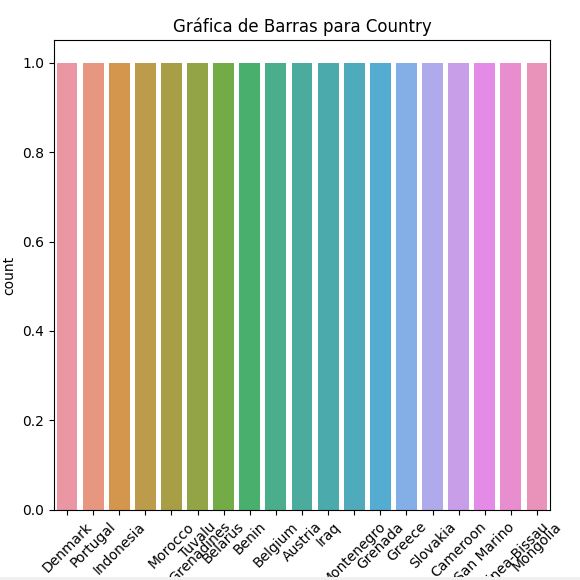
* Country: Nombre del país o territorio.
* Region: Región geográfica a la que pertenece el país.
* CEI (Cybersecurity Exposure Index): Nivel de exposición a ciberdelitos por país, valorado de 0 a 1. Cuanto mayor sea el valor, mayor será la exposición.
* GCI (Global Cyber Security Index): Índice que mide el compromiso de los países con la ciberseguridad a nivel global.
* NCSI (National Cyber Security Index): Índice que mide la preparación de los países para abordar amenazas cibernéticas y gestionar incidentes.
* DDL (Digital Development Level): Nivel de desarrollo digital, definido como el promedio del país desde el valor máximo de ambos índices.

**4. Análisis por Región:** Se compararán los indicadores de ciberseguridad en diferentes regiones geográficas para identificar patrones y diferencias significativas.

**5. Análisis de Tendencias Temporales:** Si es posible, se analizará la evolución de los indicadores a lo largo del tiempo para detectar posibles cambios o mejoras en la ciberseguridad.

**6. Correlaciones entre Indicadores:** Se realizará un análisis de correlación para examinar las posibles relaciones entre los indicadores de ciberseguridad y el nivel de desarrollo digital.

**7. Datos obtenidos:**



**Conclusión:** El análisis de datos del conjunto "Cyber Security Indexes" proporcionará información valiosa sobre la situación de la ciberseguridad en diferentes países y regiones. Los resultados obtenidos pueden ser útiles para comprender las fortalezas y debilidades en materia de ciberseguridad y para promover acciones que mejoren la protección contra ciberdelitos a nivel global.

**History of Mass Shootings in the USA**

**1. Preliminar:** Se presenta un análisis de datos del conjunto "History of Mass Shootings in the USA", que contiene información sobre diferentes incidentes de tiroteos masivos en Estados Unidos. El objetivo del análisis es examinar las características de estos incidentes y proporcionar una comprensión general de los mismos.

**2. Análisis Visual:** Se crearán gráficos para visualizar y entender los datos relacionados con los tiroteos masivos. Por ejemplo, se pueden realizar gráficos de barras para mostrar el número de tiroteos en cada ciudad o estado, diagramas de dispersión para analizar la relación entre el número de muertos y heridos en cada incidente, y gráficos de pastel para mostrar la proporción de tiroteos en cada estado.

**3. Descripción de Variables:**

Date: Fecha del tiroteo (variable no numérica).

City: Ciudad donde ocurrió el tiroteo (variable no numérica).

State: Estado donde ocurrió el tiroteo (variable no numérica).

Dead: Número de víctimas mortales (variable numérica).

Injured: Número de personas heridas (variable numérica).

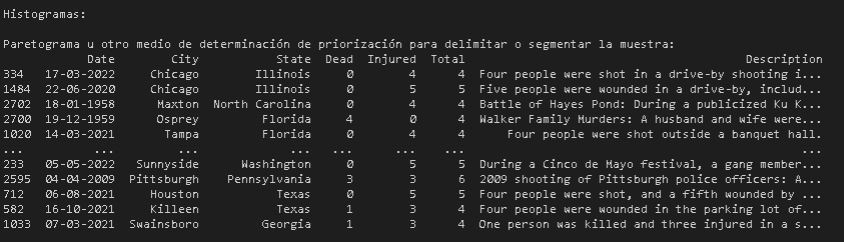
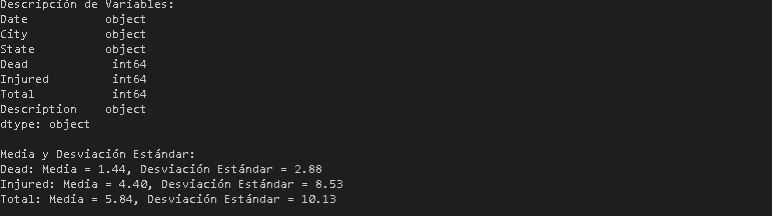
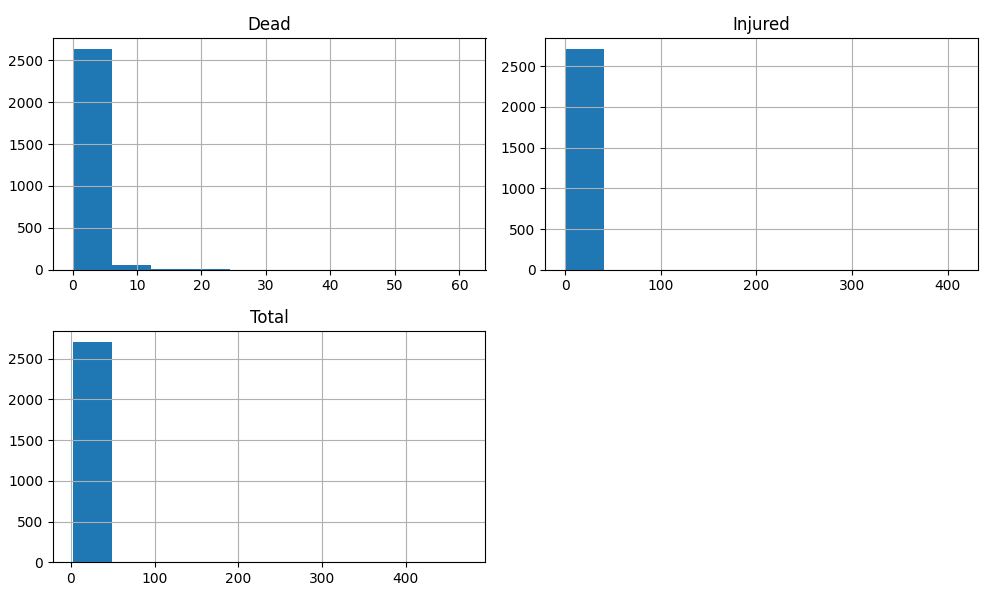
Total: Suma de víctimas mortales y heridas (variable numérica).

Description: Descripción del tiroteo (variable no numérica).

**4.** **Histogramas:** Se generarán histogramas para mostrar la distribución del número de víctimas mortales, el número de personas heridas y el total de víctimas en los tiroteos. En el eje X, se presentarán los intervalos de frecuencia y en el eje Y, se mostrará la cantidad de tiroteos en cada intervalo.

**5.** **Media y Desviación Estándar:** Se calculará la media y la desviación estándar del número de víctimas mortales, el número de personas heridas y el total de víctimas para obtener una idea de las tendencias generales en los tiroteos masivos.

**6. Análisis de Correlación:** Se realizará un análisis de correlación entre el número de víctimas mortales y el número de personas heridas en cada tiroteo para explorar si existe alguna relación entre estas variables.

**7. Datos obtenidos:  
  
**

Conclusión:

El análisis de datos del conjunto "History of Mass Shootings in the USA" ha proporcionado información valiosa sobre las características de los tiroteos masivos ocurridos en Estados Unidos. Los resultados obtenidos pueden ser útiles para comprender la naturaleza y magnitud de estos incidentes, lo que podría ayudar en la formulación de políticas y estrategias para prevenir y abordar la violencia armada en el país. Este informe presenta un resumen de los análisis realizados y destaca los hallazgos más relevantes obtenidos a partir del conjunto de datos.

**Sleep Health and Lifestyle Dataset**

**1. Preliminares:** Se presenta un análisis de datos del conjunto de datos "Sleep Health and Lifestyle", que contiene información relacionada con el sueño y el estilo de vida de diferentes individuos. Los datos incluyen detalles sobre el género, edad, ocupación, duración del sueño, calidad del sueño, nivel de actividad física, nivel de estrés, categoría de IMC, presión arterial, frecuencia cardíaca, pasos diarios y la presencia de trastornos del sueño. El análisis tiene como objetivo identificar patrones y tendencias relacionados con el sueño y el estilo de vida de los participantes.

**2. Análisis Visual:**

Se utilizarán gráficos y visualizaciones para entender los datos. Por ejemplo, se crearán histogramas para observar la distribución de variables como la duración del sueño, calidad del sueño, nivel de actividad física y nivel de estrés. También se realizarán diagramas de dispersión para analizar posibles relaciones entre variables como la edad y la duración del sueño, y se utilizarán gráficos de barras para comparar la presencia de trastornos del sueño según la ocupación.

**3. Descripción de Variables**:

Género: El género de la persona (Masculino/Femenino).

Edad: La edad de la persona en años.

Ocupación: La ocupación o profesión de la persona.

Duración del sueño (horas): La cantidad de horas que la persona duerme por día.

Calidad del sueño (escala: 1-10): Una calificación subjetiva de la calidad del sueño, que va del 1 al 10.

Nivel de actividad física (minutos/día): La cantidad de minutos que la persona realiza actividad física diariamente.

Nivel de estrés (escala: 1-10): Una calificación subjetiva del nivel de estrés experimentado por la persona, que va del 1 al 10.

Categoría de IMC: La categoría de IMC de la persona (p. ej., bajo peso, normal, sobrepeso).

Presión arterial (sistólica/diastólica): La medición de la presión arterial de la persona, indicada como presión sistólica sobre la presión diastólica.

Frecuencia cardíaca (lpm): La frecuencia cardíaca en reposo de la persona en latidos por minuto.

Pasos diarios: La cantidad de pasos que la persona da por día.

Trastorno del Sueño: La presencia o ausencia de un trastorno del sueño en la persona (Ninguno, Insomnio, Apnea del Sueño).

**4. Aproximación y Resolución de Problemas:**

El análisis se centrará en identificar patrones y tendencias en los datos de sueño y estilo de vida. Se explorará cómo la duración y calidad del sueño están relacionadas con la edad, el género y el nivel de actividad física. También se investigará la presencia de trastornos del sueño según la ocupación y se evaluará si hay diferencias significativas en la presión arterial y frecuencia cardíaca entre diferentes categorías de IMC.

**5. Histogramas**:

Se generarán histogramas para variables como la duración del sueño, calidad del sueño, nivel de actividad física y nivel de estrés para observar la distribución de los valores y detectar posibles patrones o sesgos.

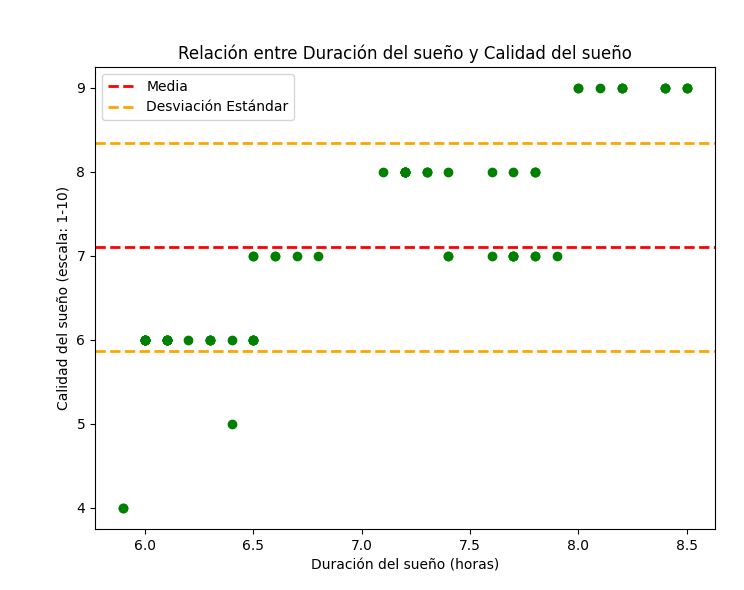
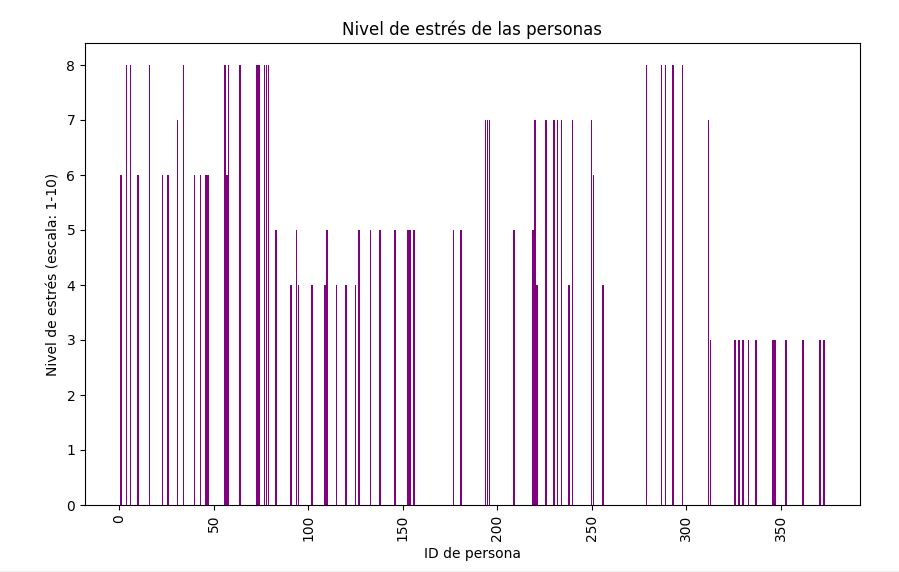
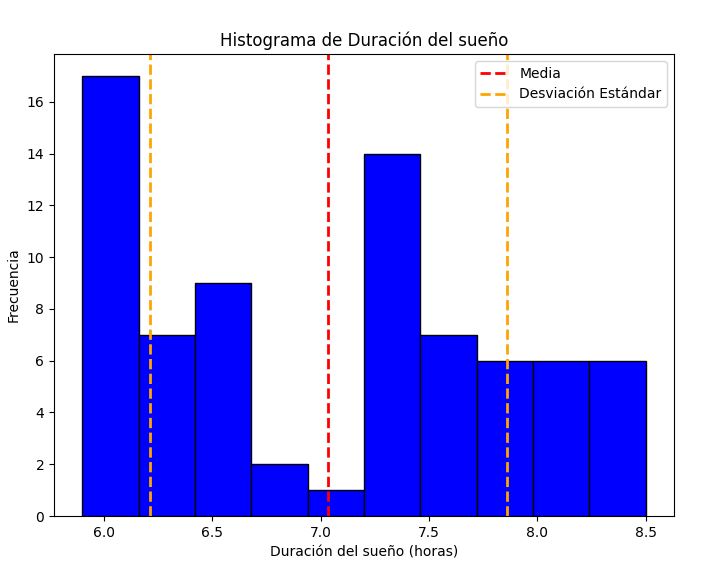
**6. Media y Desviación Estándar:**

Se calculará la media y la desviación estándar para variables como la duración del sueño, calidad del sueño, nivel de actividad física, nivel de estrés, presión arterial y frecuencia cardíaca para tener una idea de las tendencias generales y la dispersión de los datos.

**7. Análisis de Correlación:**

Se realizará un análisis de correlación entre variables como la duración y calidad del sueño, la edad, el nivel de actividad física, el nivel de estrés, la presión arterial y la frecuencia cardíaca para examinar posibles relaciones y cómo se afectan mutuamente.

**8: Datos Obtenidos**:



Se resumirán los hallazgos y se destacarán los resultados más importantes obtenidos del análisis de datos del conjunto de datos "Sleep Health and Lifestyle". Se proporcionará una visión general de los patrones de sueño y estilo de vida de los participantes y se

Conclusión

El análisis de estas diversas bases de datos ha sido enriquecedor, proporcionándome conocimientos valiosos sobre temas que me interesan, desde el entretenimiento hasta la seguridad vial. Cada elección refleja mi curiosidad y deseo de aprender y contribuir a diferentes áreas de la sociedad. A través del análisis de datos, puedo tomar decisiones informadas y hacer un impacto positivo en diversas esferas de la vida cotidiana.