5 de agosto de 2025

NICOLAS MAURICIO CHIGUANO MEZA -EMILY JULIET PAZMIÑO MORA -MELANY BRILLITH PERUGACHI TOAPANTA

nicolas.chiguano@epn.edu.ec-emily.pazmino02@epn.edu.ec – melany.perugachi@epn.edu.ec

PROYECTO FINAL – BIMESTRE 2

Tabla de contenido

[**Descripción:** 1](#_Toc205266263)

[**Objetivo:** 1](#_Toc205266264)

[**Objetivos Específicos:** 1](#_Toc205266265)

**Descripción:**

En este proyecto se desarrollará un dashboard interactivo que permitirá visualizar y analizar datos provenientes de diversas fuentes, como XML, JSON, CSV y bases de datos relacionales (SQLite3, MySQL, MongoDB, etc.). Los datos serán recopilados y procesados siguiendo • Elaborar un cronograma detallado de actividades, estableciendo tiempos y responsables.

• Asignar herramientas y actividades específicas a los integrantes del equipo según la arquitectura definida.

• Recopilar y organizar datos de las fuentes establecidas, asegurando su calidad y consistencia.

• Aplicar técnicas de análisis gráfico para generar visualizaciones significativas utilizando herramientas adecuadas.

Arquitectura para la realización del Dashboarduna arquitectura previamente diseñada, que garantizará la integración eficiente de las fuentes y su preparación para el análisis gráfico.

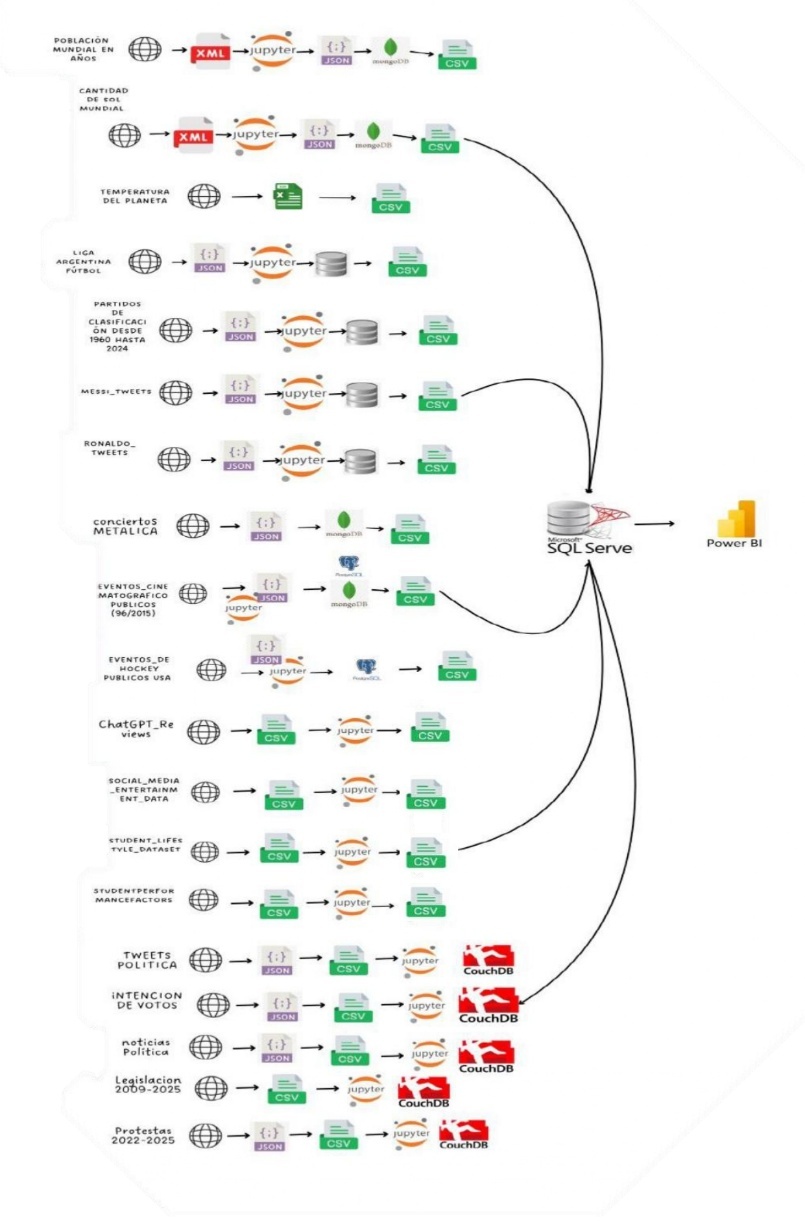
Se seleccionarán 5 temas propuestos para poder elaborar estudios de caso en base al tratamiento de los datos, los cuales serán seleccionados con base en su relevancia y potencial para generar información significativa. El trabajo será organizado mediante un cronograma que distribuirá las actividades entre los integrantes del equipo, quienes serán responsables de utilizar las herramientas asignadas para cumplir con las tareas correspondientes.

Los resultados serán representados gráficamente utilizando Power BI para la visualización, lo que permitirá una interpretación clara y efectiva de la información. Finalmente, se documentará de manera exhaustiva cada etapa del proceso, incluyendo las metodologías empleadas, los resultados obtenidos y las conclusiones alcanzadas.

**Objetivo:**

Aplicar los conocimientos adquiridos sobre análisis de datos, extracción, limpieza, transformación, visualización de datos mediante la aplicación Power BI.

**Objetivos Específicos:**

* Seleccionar 5 temas propuestos para generar casos de estudio más relevantes para el análisis de datos.
* Diseñar una arquitectura sólida para la recopilación, procesamiento y almacenamiento de datos. 

**Desarrollo:**

En este proyecto se abordó la creación de un dashboard interactivo mediante recopilación, transformación y análisis de datos obtenidos de diversas fuentes. Inicialmente, se utilizaron plataformas en línea para descargar bases de datos en formatos como XML, JSON, XLMS, entre otros. Las cuales fueron seleccionadas por su relevancia y relación con los temas propuestos.

Estas bases de datos fueron procesadas y transformadas en archivos CSV utilizandoherramientas como Jupyter Notebook, MongoDB, SQLite3, CouchDB Capella empleando bibliotecas como pandas, SO, json y exportaciones que permiten estas bases de datos y teniendo en cuenta que se tomó en cuenta limpieza estructuración de los datos. De lo anterior mencionado, se hizo uso de diferentes sistemas de gestión de bases de datos para demostrar su aplicación en escenarios diversos:

MySQL: Un sistema de gestión de bases de datos relacional ampliamente utilizado en aplicaciones web y empresariales por su capacidad para manejar grandes volúmenes de datos y realizar consultas SQL complejas [1].

MongoDB: Una base de datos NoSQL orientada a documentos que se caracteriza por su flexibilidad y escalabilidad, ideal para trabajar con datos no estructurados o semi- estructurados [2].

SQLite3: Una base de datos relacional ligera que se integra fácilmente en proyectos pequeños o aplicaciones móviles, destacada por no requerir un servidor independiente para su funcionamiento [3].

Jupyter Notebook: Se utilizaron diversas bibliotecas de Python para el procesamiento y análisis de los datos como pandas, JSON, tabalas SQL, numpy [4].

Durante la fase de integración, los datos fueron normalizados y almacenados en los sistemas mencionados, permitiendo proponer los estudios de caso con la ayuda de una herramienta de análisis de datos e inteligencia la cual permite conectar, transformar y visualizar datos de diversas fuentes para recopilación, transformación y análisis de datos obtenidos de diversas fuentes para generar informes y paneles interactivos llamada Power BI [5].

Por cierto, quiero agregar que

Esta herramienta facilitó la creación de gráficos claros y descriptivos que reflejan patrones, tendencias y correlaciones presentes en los datos.

Por último, se documentaron detalladamente todos los procedimientos realizados por cada uno de los compañeros que conforman el grupo, desde la selección de las fuentes de datos hasta la generación del dashboard final. **Temas Generales de casos de estudio**

Se obtendrá dashboards de los siguientes temas:

**Política en Ecuador 2024-2025: Intención de Voto y Opinión Pública**

En 2025, Ecuador enfrentará nuevas elecciones en un escenario polarizado, donde el oficialismo se medirá nuevamente con el correísmo. Twitter jugará un papel clave en la difusión de posturas políticas e intención de voto.

El análisis de redes sociales permite medir la intención de voto en las cinco principales ciudades del país, identificando tendencias y preferencias electorales. Aunque Twitter no es la red más usada en Ecuador, su impacto en el debate político es significativo.

Además, el monitoreo de noticias, opiniones y retweets revela la percepción ciudadana sobre temas políticos, midiendo el alcance de diferentes posturas. Paralelamente, se han recopilado datos sobre los proyectos presentados en la Asamblea Nacional y su estado actual, permitiendo analizar qué iniciativas avanzan y cuáles quedan archivadas.

Finalmente, el estudio abarca las protestas organizadas por distintos partidos políticos, identificando sus motivaciones y su impacto en la opinión pública. Estos datos ofrecen una visión integral del panorama político ecuatoriano en la antesala de las elecciones.

**Eventos deportivos a nivel mundial. (Estadísticas de partidos de clasificación, liga argentina, estadística de tweets de Messi y Ronaldo)**

Este estudio se centra en el análisis de eventos deportivos a nivel mundial, con un enfoque específico en las estadísticas de partidos de fútbol correspondientes a los años de 1960 a 2024, partidos y torneos de la liga argentina, y los tweets de los dos jugadores más importantes de todos los tiempos.

El propósito principal es identificar patrones de rendimiento, tendencias competitivas que definen el éxito de los equipos en estas competiciones. A través del análisis de datos, se busca comprender las dinámicas del juego, como resultados de los partidos, y tweets más relevantes de los dos jugadores.

**Conciertos y eventos públicos**

**(Conciertos de Metalica , Eventos públicos cinematográficos y eventos hockey)**

Aquí se encuentra el primer análisis que trata de conciertos de Metallica una banda estadounidense de heavy metal formada en 1981 en Los Ángeles, Estados Unidos. Analiza los eventos de cine y sus ratings en el período comprendido entre 1996 y 2016.

Este informe analiza también los eventos de hockey, enfocándose en la distribución de eventos por ciudad y la sumatoria de puntos anotados en cada uno. El análisis de datos, se busca comprender la evolución de sus giras, también se busca comprender la distribución de los ratings, las tendencias en la popularidad de los géneros cinematográficos y cómo han evolucionado las preferencias del público en este período.

**Actividades y Hobbies**

**Nivel de estrés y estilo de vida de estudiantes, como comentarios al**

**momento de utilizar Chatgpt , tiempo de utilización de redes sociales**

**1.- Definición del Caso de Estudio**

El presente estudio analiza la relación entre las actividades diarias, el nivel de estrés y el estilo de vida de los estudiantes, enfocándose en el uso de herramientas tecnológicas como ChatGPT y las redes sociales. En un contexto donde la carga académica y las responsabilidades personales pueden generar altos niveles de estrés, el uso de estas plataformas digitales se ha vuelto parte fundamental de la rutina estudiantil, ya sea como apoyo académico, entretenimiento o socialización. Sin embargo, el tiempo dedicado a estas herramientas puede influir tanto positiva como negativamente en la salud mental, la gestión del tiempo y el rendimiento académico de los estudiantes. Este estudio busca comprender cómo la interacción con ChatGPT y el uso de redes sociales impactan en su estilo de vida, analizando hábitos, motivaciones y percepciones para determinar si funcionan como mecanismos de afrontamiento o si, por el contrario, contribuyen al aumento de la ansiedad y la distracción.

**Eventos o noticias mundiales**

Se obtuvieron datos de UnData relacionados con la temperatura mundial a través de los años,la cantidad de sol que se a registrado en los diferentes años en todos los países y el crecimiento poblacional de los países del mundo con el fin de obtener un estudio de caso de cada una de estas bases de datos obtenida de la fuente mencionada anteriormente. Todo esto con el fin de proporcionar un estudio de caso que sea de interés.

**Casos de Estudios**

**Análisis del Pulso Político en Ecuador**

El análisis de datos es clave para comprender el pulso político en Ecuador. Para ello, recopilamos información de redes sociales y encuestas, utilizando Twarc2 para extraer tweets a través de la API de Twitter. Esto nos permitió analizar interacciones y discursos de figuras políticas relevantes.

Además, trabajamos con un CSV proporcionado por Cedatos, que contiene datos de encuestas sobre la percepción ciudadana. Complementamos esto con web scraping para extraer noticias de medios digitales, obteniendo información sobre el tratamiento mediático de distintos actores políticos.

También utilizamos un CSV disponible en la página de la Asamblea Nacional, que contiene información sobre legisladores y su actividad parlamentaria.

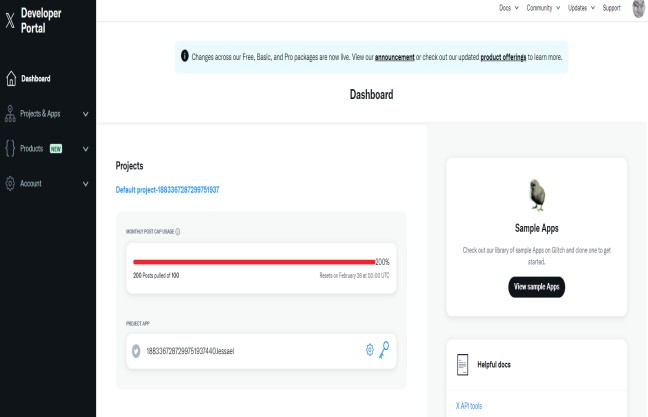
Para almacenar y gestionar los datos, empleamos Couchbase, una base de datos NoSQL que facilita la organización y consulta eficiente de grandes volúmenes de información.

Con estos datos, es posible analizar tendencias y sentimientos, identificando patrones en el discurso público y evaluando la percepción ciudadana sobre distintos temas y figuras políticas.

1. **Tweets y Retweets en Ecuador Objetivos**
   * Identificar las tendencias
   * políticas y opiniones expresadas en Twitter en Ecuador.
   * Analizar el sentimiento de los tweets en relación con candidatos y partidos.
   * Determinar la influencia de redes sociales en la percepción política.

**Métodos y Herramientas Utilizadas**

* + Twarc para recopilación de tweets.



* + NLTK para análisis de sentimiento.
  + Pandas para limpieza y manipulación de datos.
  + Power BI para reperesnetar las visualizaciones

## **Datos Recopilados**

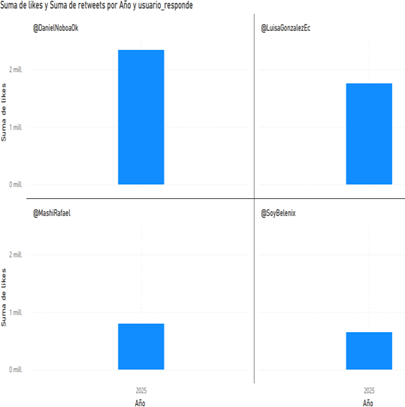
* + 11,000 tweets y retweets de usuarios de todo el Ecuador.
  + Hashtags más utilizados sobre política en Ecuador.
  + Usuarios más influyentes en la discusión política.

## **Análisis de Datos**

* + Limpieza de tweets (eliminación de ruido y datos irrelevantes).
  + Clasificación del sentimiento (positivo, negativo, neutral).
  + Identificación de tendencias y temas más discutidos.

## **Visualizaciones**

* + Tenemos una suma comparativa del número de likes de los tweets y retweetss de los usuarios más influyentes en las redes sociales de ecuador en este caso twitter



* + - Como se puede observar en este caso Daniel Noboa es el usuario al que tanto mensajes y retweets tiene dado su posición como presidente de la república, la siguiente es Luisa Gonzales contendiente electoral en estas elecciones 2025.

**Resultados Obtenidos Conclusiones:**

1. La conversación política en Twitter está altamente polarizada, con picos de actividad en momentos clave (elecciones, protestas).
2. Quito y Guayaquil presentan mayor volumen de discusión en comparación con otras ciudades.
3. Los tweets negativos tienen mayor alcance y viralización.

**Recomendaciones:**

* + Ampliar la recopilación de datos incluyendo Facebook y TikTok.
  + Aplicar modelos de análisis de redes para identificar influencia política.
  + Incorporar machine learning para predicción de tendencias futuras.

Desafíos y Problemas Encontrados

**Desafíos y Problemas Encontrados**

1. Ruido en los datos debido a la presencia de bots y spam.
2. Limitaciones en el acceso a la API de Twitter por restricciones de uso.
3. Dificultad para validar información debido a la subjetividad del lenguaje.
4. **Noticias en Ecuador y el Extranjero Objetivos**
   * Comparar la cobertura de noticias nacionales e internacionales sobre política ecuatoriana.
   * Evaluar el sesgo mediático en la información presentada.
   * Identificar los temas más reportados en medios.

**Métodos y Herramientas Utilizadas**

* + BeautifulSoup para web scraping.
  + Pandas para organización y estructuración de datos.
  + Power BI para visualización.

**Datos Recopilados**

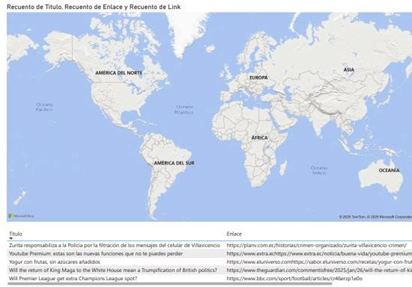
* + 700 noticias de medios ecuatorianos e internacionales.
  + Fecha, titular, Enlace y contenido de cada noticia.

**Análisis de Datos**

* + Comparación de términos usados por medios nacionales e internacionales.
  + Análisis de similitud de noticias por diferentes medios.
  + Cuáles son los puntos más importantes geopolíticos a inicios de 2025

**Visualizaciones**

**Al usar** BeautifulSoup para el web scraping en varias fuentes informáticas tenemos las siguientes visualizaciones donde se prefiere buscar que país esta más en tendencia en las noticas de enero de 2025.



* + - Donde podemos ver que las noticas más en tendencias son tano de parte de la politica en ecuador y sobre el nuevo presidente electo Donald Trump.

**Resultados Obtenidos Conclusiones:**

* + Los medios nacionales enfatizan más en la figura de los candidatos, mientras que los internacionales destacan el contexto socioeconómico.
  + Existen diferencias en la cobertura dependiendo de la afiliación política de cada medio.

**Recomendaciones:**

1. Analizar más fuentes para reducir posibles sesgos.
2. Utilizar modelos de procesamiento de lenguaje natural (NLP) para detección de tendencias.

**Desafíos y Problemas Encontrados**

* + Restricciones en algunas páginas para realizar web scraping.
  + Posible sesgo en los medios seleccionados.

1. **Intención de Voto en las Cinco Ciudades Objetivos**
   * Identificar tendencias en la intención de voto en Quito, Guayaquil, Cuenca, Manta y Ambato.
   * Analizar cambios en la preferencia electoral en distintos períodos.

**Métodos y Herramientas Utilizadas**

* + **Base de datos de Cedatos (200,000 registros)**.
  + **Power BI** para visualización de tendencias.

**Datos Recopilados**

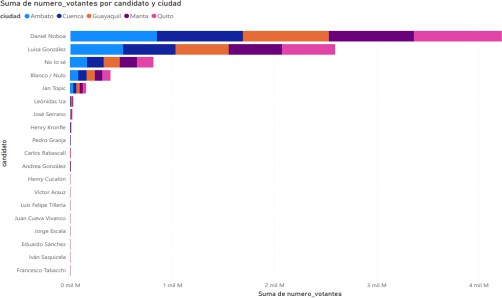
* + Registros de intención de voto en **Quito, Guayaquil, Cuenca, Manta y Ambato**.
  + Segmentación por preferencia política.

**Análisis de Datos**

* + Evaluación de cambios en la intención de voto a lo largo del tiempo.
  + Comparación de preferencias por región.

**Visualizaciones**

* + Bajo un CSV proporcionado por cedatos donde se nos muestra cual es el número de ciudadanos de las 5 ciudades más influyentes del país que son Quito, Guayaquil, Cuenca, Ambato y Manta.



* + - Donde podemos ver que la intención de voto está encabezada por Daniel Noboa y Luisa Gonzales

**Resultados Obtenidos Conclusiones:**

* + Existe una diferencia significativa entre la intención de voto en la Sierra y la Costa.
  + Los jóvenes muestran una mayor indecisión en comparación con otros grupos de edad.

**Recomendaciones:**

1. Aplicar modelos predictivos para estimar posibles resultados electorales.
2. **Leyes en Debate y su Relación con las Protestas en Ecuador (2022-2025) Objetivos**
   * Analizar la frecuencia, causas y repercusiones de las protestas en Ecuador entre 2022 y 2025 en relación con leyes en discusión o aprobación.
   * Identificar patrones de movilización vinculados a proyectos de ley y respuestas gubernamentales.

**Métodos y Herramientas Utilizadas**

* + Web scraping para recopilar datos de noticias, redes sociales y documentos legislativos.
  + Uso de Power BI para visualizar tendencias en la aprobación de leyes y su impacto en la conflictividad social.

**Datos Recopilados**

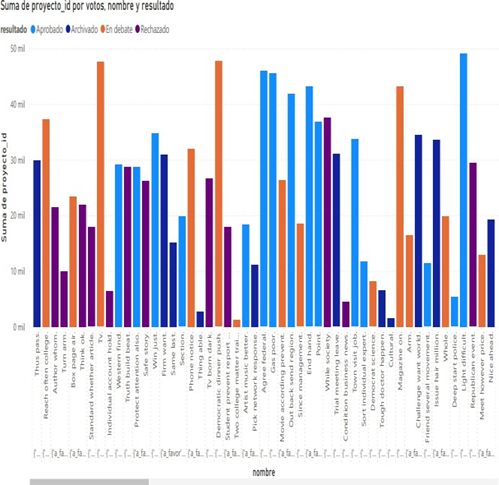
* + Reportes de protestas de 2022 a 2025.
  + Listado de leyes en debate o aprobadas que generaron movilizaciones.
  + Datos sobre enfrentamientos políticos relacionados con reformas legislativas.

**Análisis de Datos**

* + Categorización de leyes según su impacto en la ciudadanía (económicas, laborales, sociales, ambientales).
  + Análisis de discurso en redes sociales sobre leyes polémicas.
  + Evaluación de la respuesta del gobierno y actores políticos ante las manifestaciones.

**Visualizaciones**

* + Gráficos de correlación entre debates legislativos y como las protestas han generado leyes que están en estado de Stand By o ya han sido aprobadas.



**Resultados Obtenidos**

* + Se identificó un aumento en las protestas en torno a leyes económicas y laborales.
  + Los años electorales mostraron mayor debate legislativo y movilización social.

**Conclusiones**

* + Las leyes que afectan derechos laborales y subsidios han sido los principales detonantes de protestas.
  + La falta de consulta ciudadana en ciertos proyectos de ley ha intensificado la conflictividad social.

**Recomendaciones**

* + Desarrollar modelos predictivos sobre conflictos sociales a partir del análisis de propuestas legislativas.
  + Implementar mecanismos de diálogo temprano en el proceso de aprobación de leyes.

**Desafíos y Problemas Encontrados**

* + Falta de datos estructurados sobre el impacto de leyes en la conflictividad social.
  + Acceso limitado a información legislativa en tiempo real.

1. **Legislación en Ecuador (2009-2025) Objetivos**
   * Analizar la evolución de las leyes en Ecuador entre 2009 y 2025.

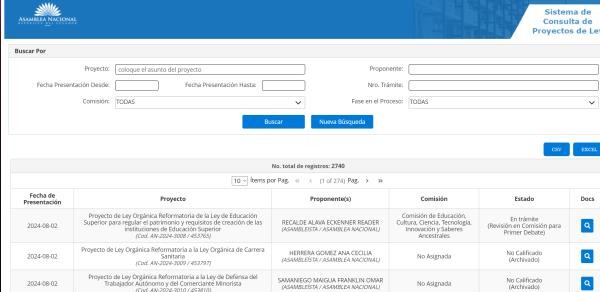
**Métodos y Herramientas Utilizadas**

* + **Pandas** para estructuración de datos.

**Datos Recopilados**

* + **2,800 leyes y reformas** desde 2009 hasta 2025.
  + Proponentes de las leyes y su función.
  + Mediante un CSV que se encuentra disponible en las páginas de la asamblea nacional

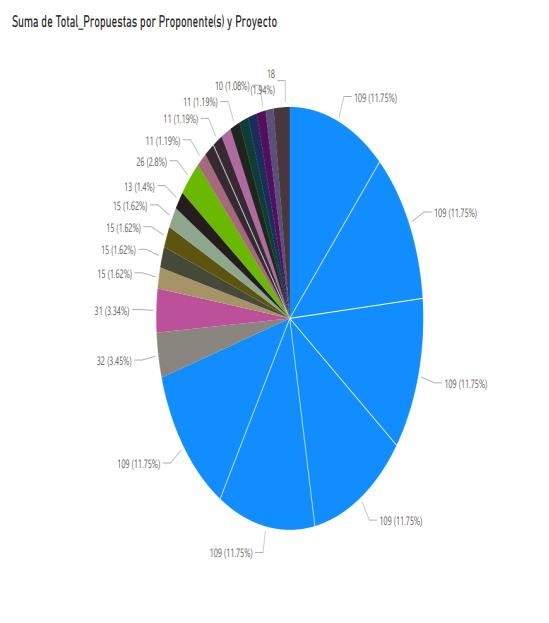
https://leyes.asambleanacional.gob.ec



**Análisis de Datos**

* + Clasificación por temática y año.

**Visualizaciones**

* + - **M**ediante el CSV proporcionado hicimos un análisis para ver quien era el proponente de leyes con mayor influencia de aprobación en la asamblea nacional además de como estas han sido aprobadas y seleccionadas para reformas. 
    - Como se puede observar Rafael Correa Delgado a sido el mayor proponente en los periodos que estuvo al poder observando así una dominancia total en estos periodos de 2009 a 2025.

**Resultados Obtenidos Conclusiones:**

* + Que el mayor número de leyes y proyectos aprobados fueron en el Peridio de Rafael Correa

**Recomendaciones:**

1. Implementar análisis de impacto legislativo.

**Desafíos y Problemas Encontrados**

* + Dificultad en el acceso a datos legislativos completos.

1. **ENFOQUE**

El análisis de datos políticos permite comprender tendencias y percepciones ciudadanas en Ecuador. Para ello, recopilamos información de **redes sociales, encuestas, noticias y registros oficiales**. Usamos **Twarc2** para extraer tweets de figuras políticas y analizar interacciones. Cedatos nos proporcionó un **CSV con encuestas**, reflejando la opinión pública. Mediante **web scraping**, recopilamos noticias de medios digitales para evaluar la cobertura mediática. Además, utilizamos un **CSV de la Asamblea Nacional**, con datos sobre legisladores y su actividad.

Toda la información se almacenó en **Couchbase**, facilitando su consulta y análisis. Estos datos permiten estudiar discursos, identificar patrones en el debate público y evaluar la percepción de diferentes actores políticos.

1. **INTERPRETACION DE TODOS LOS DATOS OBTENIDOS**

La interpretación de los datos recopilados revela tendencias significativas en distintos ámbitos políticos y sociales de Ecuador.

**Análisis de Twitter**: Los tweets extraídos con **Twarc2** muestran cómo las interacciones en redes reflejan la polarización política del país. Rafael Correa tiene una alta presencia digital con mensajes críticos hacia el gobierno, mientras que Daniel Noboa y Luisa González generan tanto apoyo como rechazo en diferentes sectores.

1. **Noticias Globales**: A través de **web scraping**, se recopilaron noticias internacionales, permitiendo analizar cómo los medios extranjeros perciben la situación política y económica de Ecuador.
2. **Intención de voto**: Los datos de intención de voto en **Quito, Guayaquil, Cuenca, Manta y Ambato** muestran la preferencia electoral en las ciudades más influyentes del país.
3. **Reformas legales tras protestas**: Se analizaron las leyes en estado **pendiente, reformada o archivada** desde las protestas de 2022 hasta 2025, evidenciando el impacto de la movilización social en la legislación.
4. **Leyes aprobadas (2009-2025)**: La cantidad de proyectos y leyes aprobados en este período muestra la evolución de la actividad legislativa y su respuesta a las demandas ciudadanas.

En general, estos datos permiten entender la relación entre redes sociales, medios, procesos electorales y cambios en la legislación ecuatoriana

1. **CONCLUCION**

El análisis de datos políticos nos ha permitido comprender mejor la dinámica del debate público en Ecuador. A través de la recopilación de información de redes sociales, encuestas, noticias y registros oficiales, logramos construir una base sólida para evaluar tendencias y opiniones. El uso de Twarc2 para extraer tweets, junto con el CSV de Cedatos y los datos de la Asamblea Nacional, nos brindó una visión amplia de la percepción ciudadana y la actividad política.

El almacenamiento en Couchbase facilitó la organización y consulta de la información, permitiendo identificar patrones en el discurso político y en la interacción entre ciudadanos y líderes. Este enfoque demuestra el valor del análisis de datos en la toma de decisiones y en la comprensión de los cambios sociales, contribuyendo a estudios más precisos sobre la opinión pública y la comunicación política.

1. **Link del grafico en Línea**

[**https://app.powerbi.com/groups/me/repor**](https://app.powerbi.com/groups/me/reports/3af71b77-72bc-4ecb-a169-58cb8033aa1d/9ca6c88e6b002d9ac278?experience=power-bi)[**ts/3af71b77-72bc-4ecb-a169-**](https://app.powerbi.com/groups/me/reports/3af71b77-72bc-4ecb-a169-58cb8033aa1d/9ca6c88e6b002d9ac278?experience=power-bi)[**58cb8033aa1d/9ca6c88e6b002d9ac278?ex**](https://app.powerbi.com/groups/me/reports/3af71b77-72bc-4ecb-a169-58cb8033aa1d/9ca6c88e6b002d9ac278?experience=power-bi)[**perience=power-bi**](https://app.powerbi.com/groups/me/reports/3af71b77-72bc-4ecb-a169-58cb8033aa1d/9ca6c88e6b002d9ac278?experience=power-bi)

**Eventos deportivos a nivel mundial. (Estadísticas de partidos de clasificación, liga argentina, estadística de tweets de Messi y Ronaldo)**

Este estudio se centra en el análisis de eventos deportivos a nivel mundial, con un enfoque específico en las estadísticas de partidos de fútbol correspondientes a los años de 1960 a 2024, partidos y torneos de la liga argentina, y los tweets de los dos jugadores más importantes de todos los tiempos.

El propósito principal es identificar patrones de rendimiento, tendencias competitivas que definen el éxito de los equipos en estas competiciones. A través del análisis de datos, se busca comprender las dinámicas del juego, como resultados de los partidos, y tweets más relevantes de los dos jugadores.

1. **Objetivo general**

Analizar el desempeño del fútbol argentino, la influencia de Lionel Messi y Cristiano Ronaldo en el deporte y la evolución de los partidos de clasificación desde 1960 hasta 2024, utilizando datos estadísticos y redes sociales para identificar patrones, tendencias e impactos en el ámbito futbolístico.

1. **Objetivos específicos**

Evaluar el rendimiento de los equipos en la Liga Argentina de Fútbol, identificando tendencias de desempeño, campeonatos ganados, goleadores destacados y otros factores clave.

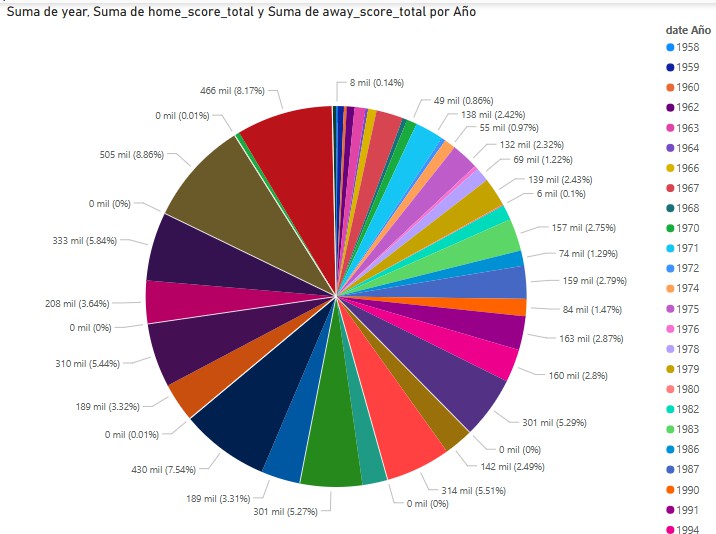
* + Analizar la influencia de Lionel Messi en las redes sociales, explorando la reacción del público en Twitter, las palabras clave más utilizadas y la relación entre su actividad futbolística y la interacción digital.
  + Estudiar la evolución de los partidos de clasificación desde 1960 hasta 2024, examinando cambios en formatos, desempeño de selecciones y tendencias en clasificaciones a torneos internacionales.

1. **Enfoque**

El enfoque para esta sección de deportes es la visualización y análisis de los datos relacionados con la liga argentina, la clasificación de equipos desde el año 1940 hasta el 2024 y los tweets de los jugadores más importante del futbol mundial Leonel Messi Cristiano Ronaldo. Para ello, se recolectaron datos relevantes desde diferentes fuentes, incluyendo archivos JASON, y bases de datos públicas, y se transformaron a formatos compatibles como CSV mediante herramientas como Jupyter Notebook. Se utilizaron plataformas como Power BI para crear gráficos y tablas que facilitan la visualización de estadísticas clave, como los torneo con más goles de la liga argentina, quien tiene el mayor número de tweets, y equipos con más goles en la clasificación mundial. Finalmente, se interpretaron los resultados obtenidos, proporcionando hallazgos significativos que destacan relaciones entre los distintos conjuntos de datos analizados.

1. **Métodos y Herramientas Utilizadas**
   * **Extracción de Datos**: Se recopiló información de cada partido y los tweets, se analizó cada resultado resaltando el marcador, torneos goles y marcadores de mayor importancia
   * **Almacenamiento de Datos**: Los datos se gestionaron en bases de datos estructuradas para su análisis.
   * **Visualización**: Se utilizaron herramientas de visualización de datos como Power BI para representar gráficamente la información sobre estadísticas y porcentajes más relevantes.
2. **Extracción de datos**

Los datos fueron extraídos de UnData y Bases de datos abiertas, centrados en 3 casos de estudio clave:

* + **Partidos de clasificación desde 1960 hasta 2024:** Se trata de un archivo JASON que contiene información histórica sobre partidos de clasificación, permitiendo identificar tendencias de goles.
  + **Liga Argentina Fútbol:** Un archivo JASON que proporciona datos detallados sobre todas las competiciones de la liga argentina, lo que facilita analizar estadísticas de goles partidos y competencias con mayor aceptación del público.
  + **Tweets de Ronaldo y Messi:** Mediante un archivo JASON proporciono datos detallado de todos los tweets que califican el desempeño de los dos mejores jugadores, lo que facilita hacer un análisis sobre cual jugador es mejor clasificado y cual jugadores es más aceptado por los fans.

1. **Análisis y Visualización de Información**

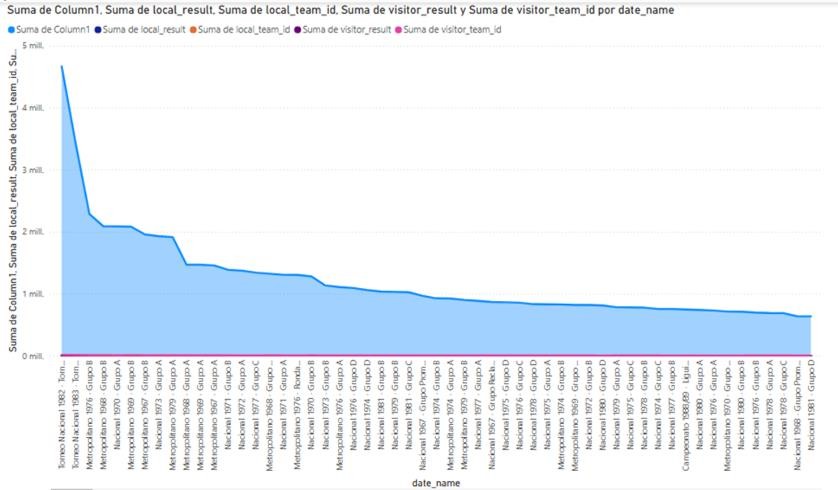
**Partidos de clasificación desde 1960 hasta 2024.**

La imagen muestra un gráfico de pastel que representa la suma de partidos jugados como local y visitante por año en los partidos de clasificación desde 1960 hasta 2024. Los años con mayor cantidad de goles anotados por los equipos visitantes y locales tienen porciones más grandes. Se observa una variabilidad en los datos, con algunos años dominando más la suma total de goles.

**Liga Argentina Fútbol**

El gráfico muestra la suma de varias métricas (goles, IDs de equipos, etc.) agrupadas por date\_name (torneo y fecha). La distribución sugiere que algunos torneos tuvieron más goles que otros.

La parte pintada en azul representa la suma acumulada de los valores en el eje vertical (goles ) para cada date\_name (torneo o temporada).



**Tweets de Ronaldo y Messi:**

Las imágenes muestran gráficos de pastel con la distribución de diversas métricas en tweets sobre Messi y Cristiano Ronaldo.

**Imagen 1: Tweets sobre Messi**

Color morado representa la mayoría de los datos, lo que indica que la cantidad de tweets sobre Messi domina la información. Color azul representan un pequeño porcentaje de la distribución.

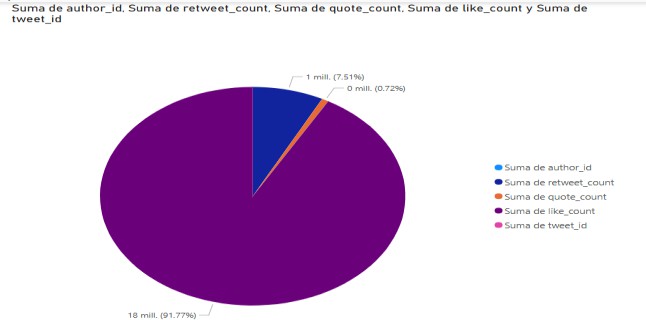
Color naranja las citas en tweets tienen una participación mínima.

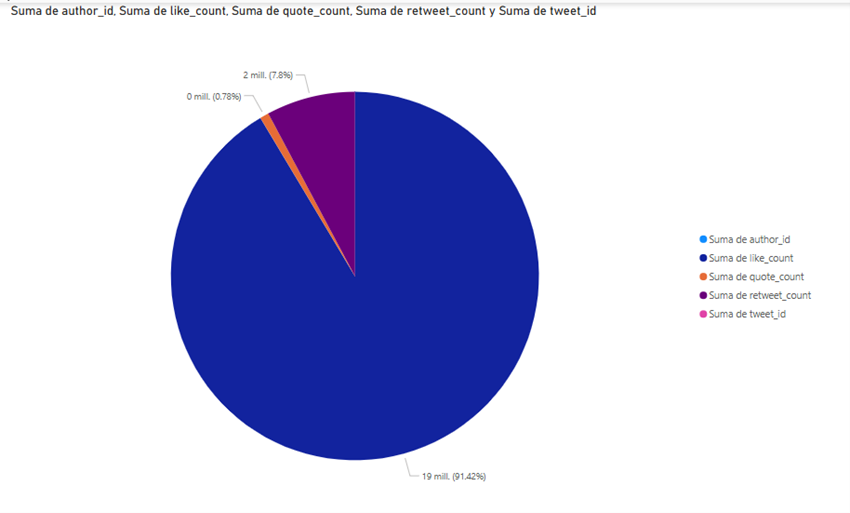
**Imagen 2: Tweets sobre Ronaldo**

Color azul en este caso, el número de tweets es la métrica dominante.

Color morado indica que los likes tienen una presencia significativa.

Color naranja las citas son una fracción muy pequeña del total.





1. **Interpretaciones de los datos obtenidos**

**Interpretación Gráfico 1 (Partidos de clasificación desde 1960 hasta 2024).**

Se observa una distribución de goles por año con muchas diferencias, indicando que en ciertos años se jugaron más partidos o hubo una mayor cantidad de goles anotados. Algunas secciones del gráfico indican "0 mil (0%)", lo que sugiere que en esos años no se registraron partidos, los años con mayor cantidad de goles son representados con secciones más grandes.

Este gráfico refleja cómo ha evolucionado el número de goles en los partidos de clasificación a lo largo de los años, con picos en ciertos períodos y bajas en otros.

**Interpretación Gráfico 2(Liga Argentina Fútbol).**

La interpretación de los datos en el eje:

* **Horizontal (X)**: Se listan los distintos torneos y temporadas en la historia del fútbol argentino (ejemplo: "Torneo 1983",

"Metropolitano 1985", "Nacional 1976 - Grupo C", etc.).

* **Vertical (Y)**: Se muestra una suma acumulada de las métricas mencionadas, representadas en millones.

La zona azul rellena representa la suma total de estas métricas para cada torneo/temporada, ordenadas de mayor a menor.

Los primeros torneos tienen los valores más altos, a medida que avanzamos hacia la derecha, la acumulación de métricas disminuye gradualmente, indicando que esos torneos o temporadas tuvieron menor cantidad de partidos o goles en comparación con los primeros.

**Interpretación Gráfico 3 y 4 (Tweets de Messi y Ronaldo).**

Los tweets sobre Lionel Messi y Cristiano Ronaldo en Twitter siguen un patrón muy similar, donde la mayoría de la interacción proviene de publicaciones originales, seguidas por retweets y, en menor medida, citas.

Messi tiene un porcentaje ligeramente mayor de retweets, lo que sugiere que sus tweets son compartidos con más frecuencia.

Ronaldo tiene una ligera ventaja en la cantidad total de tweets, lo que indica que genera un volumen de conversación ligeramente mayor en Twitter.

Las citas representan un porcentaje muy bajo en ambos casos, lo que sugiere que los usuarios prefieren retuitear en lugar de comentar citando el tweet.

Ambos jugadores generan una enorme presencia en Twitter con una distribución de interacciones muy similar. Sin embargo, Messi parece provocar más retweets, mientras que Ronaldo genera un mayor volumen de conversación en términos absolutos.

1. **Conclusión**

El análisis de los gráficos permite observar tendencias y patrones en distintos ámbitos del fútbol. En cuanto a los partidos de clasificación desde 1960 hasta 2024, se evidencia una variabilidad en la cantidad de goles anotados, reflejando diferencias en el número de encuentros disputados a lo largo de los años.

El gráfico de la Liga Argentina muestra cómo la acumulación de métricas ha variado a través de las distintas temporadas y torneos, destacando que los primeros torneos presentan los valores más altos, mientras que en las más recientes la acumulación disminuye gradualmente.

Finalmente, el análisis de los tweets sobre Lionel Messi y Cristiano Ronaldo indica que ambos generan una gran presencia en Twitter, con interacciones similares en términos de retweets y citas. Sin embargo, Messi parece generar más retweets, mientras que Ronaldo recibe un mayor volumen total de menciones.

Estos gráficos evidencian la evolución del fútbol en distintos aspectos, desde el rendimiento en clasificación hasta la influencia digital de las grandes figuras del deporte.

1. **Link del grafico en Línea**

[**https://app.powerbi.com/view?r=eyJrIjoi**](https://app.powerbi.com/view?r=eyJrIjoiYTZiODA1YmYtZmU3Zi00YmU3LTllY2UtZmZiOTNkOTFhMTJjIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9)[**YTZiODA1YmYtZmU3Zi00YmU3LTllY**](https://app.powerbi.com/view?r=eyJrIjoiYTZiODA1YmYtZmU3Zi00YmU3LTllY2UtZmZiOTNkOTFhMTJjIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9)

[**2UtZmZiOTNkOTFhMTJjIiwidCI6IjY4**](https://app.powerbi.com/view?r=eyJrIjoiYTZiODA1YmYtZmU3Zi00YmU3LTllY2UtZmZiOTNkOTFhMTJjIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9)[**MmE0ZTZhLWE3N2YtNDk1OC1hM2F**](https://app.powerbi.com/view?r=eyJrIjoiYTZiODA1YmYtZmU3Zi00YmU3LTllY2UtZmZiOTNkOTFhMTJjIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9)

[**jLTllMjY2ZDE4YWEzNyIsImMiOjR9**](https://app.powerbi.com/view?r=eyJrIjoiYTZiODA1YmYtZmU3Zi00YmU3LTllY2UtZmZiOTNkOTFhMTJjIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9)

**Conciertos y eventos públicos**

**(Conciertos de Metalica , Eventos públicos cinematográficos y eventos hockey)**

Aquí se encuentra el primer análisis que trata de conciertos de Metallica una banda estadounidense de heavy metal formada en 1981 en Los Ángeles, Estados Unidos. Analiza los eventos de cine y sus ratings en el período comprendido entre 1996 y 2016.

Este informe analiza también los eventos de hockey, enfocándose en la distribución de eventos por ciudad y la sumatoria de puntos anotados en cada uno. El análisis de datos, se busca comprender la evolución de sus giras, también se busca comprender la distribución de los ratings, las tendencias en la popularidad de los géneros cinematográficos y cómo han evolucionado las preferencias del público en este período.

Este informe analiza los eventos de hockey, enfocándose en la distribución de eventos por ciudad y la sumatoria de puntos anotados en cada uno.

1. **Objetivo general**

Indagar y sacar conclusiones sobre la trayectoria de conciertos de Metallica desde 1986 hasta 2024, identificando patrones en la frecuencia de presentaciones, los lugares más visitados y la evolución de sus giras, utilizando datos estadísticos y visualización de datos para destacar tendencias clave.

Analizar la evolución de los eventos de cine y sus ratings desde 1996 hasta 2016, identificando patrones en la calificación depelículas, los géneros más populares y las tendencias en la industria del cine.

Analizar los eventos de hockey a nivel estadístico, considerando la cantidad de puntos anotados en cada partido y la frecuencia de eventos en distintas ciudades.

1. **Objetivos específicos**
   * Evaluar la distribución geográfica de los conciertos de Metallica y los recintos más utilizados.
   * Evaluar la distribución de ratings en las películas analizadas.
   * Analizar la sumatoria de puntos anotados en cada evento y su variación en distintas ciudades.
   * Crear datos clave a través de herramientas de análisis como Power BI para facilitar la interpretación de la información.

**Enfoque**

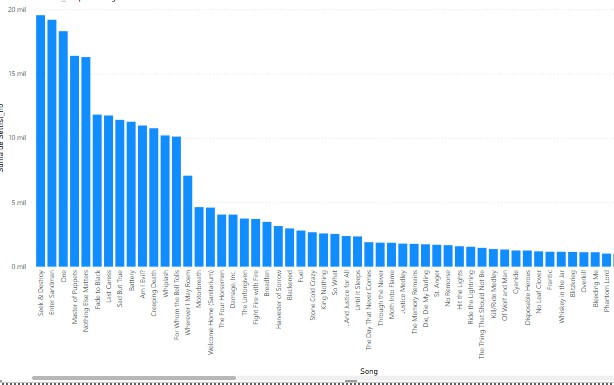
El análisis se los realiza a través de la recopilación de datos registrados, incluyendo fechas, lugares y asistencia, datos sobre películas estrenadas entre 1994 hasta 2008 considerando visualizaciones de estas películas en años posteriores como 2015. Se utilizaron fuentes como bases de datos públicas y archivos JSON convertidos a formatos compatibles (CSV) mediante herramientas como Jupyter Notebook.

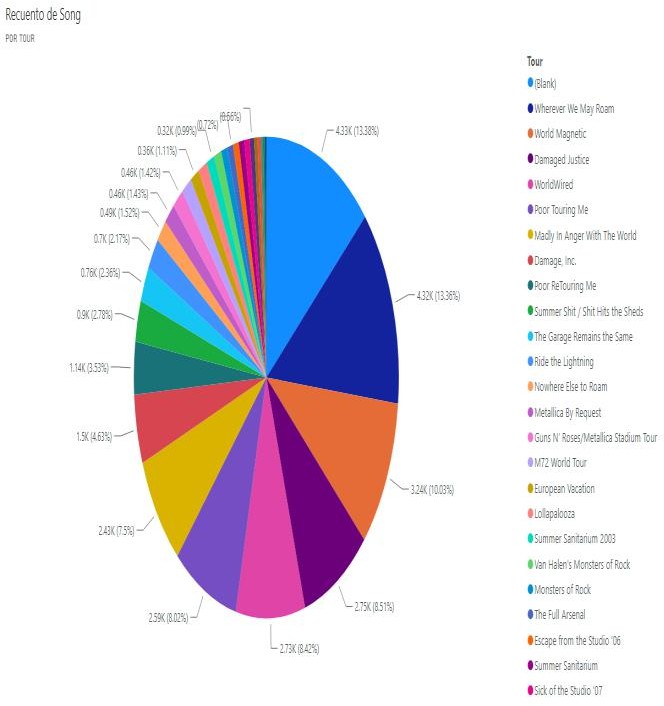
Posteriormente, los datos fueron visualizados utilizando Power BI para representar de manera efectiva las estadísticas más relevantes.

1. **Métodos y Herramientas Utilizadas**
   * **Extracción de Datos:** Se recopilaron datos de cada concierto de Metallica, incluyendo fecha, lugar y asistencia, datos de películas, incluyendo año de estreno, género y calificación, datos de eventos de hockey con información sobre la ciudad y los puntos anotados.
   * **Almacenamiento de Datos:** Los datos fueron estructurados en bases de datos para su análisis.
   * **Visualización:** Se utilizaron herramientas como Power BI para representar la información de manera gráfica y facilitar la interpretación de los patrones identificados.
2. **Extracción de datos**

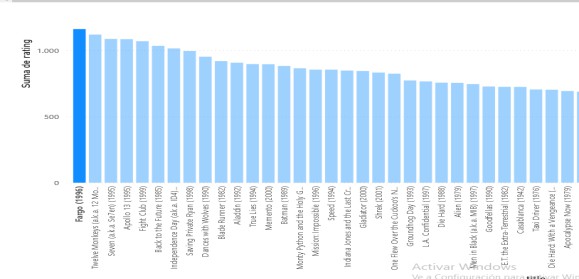
Los datos fueron extraídos de fuentes confiables y estructurados en tres ejes principales:

1. **Historial de Conciertos y eventos públicos:** Archivo JSON que contiene información detallada sobre registros desde 1986 hasta 2024 en algunos casos, permitiendo identificar patrones en la frecuencia y concurrencia.
2. **Ciudades o lugares:** Información sobre los lugares donde se ha presentado con mayor frecuencia.
3. **Análisis y Visualización de Información**
   1. **Canciones más escuchadas con sus localizaciones**

****Se generó un gráfico de barras que representa el número de conciertos por año desde 1986 hasta 2024. Se observa que hubo picos en ciertos periodos donde Metallica realizó giras extensas, mientras que en otros años hubo menos presentaciones.

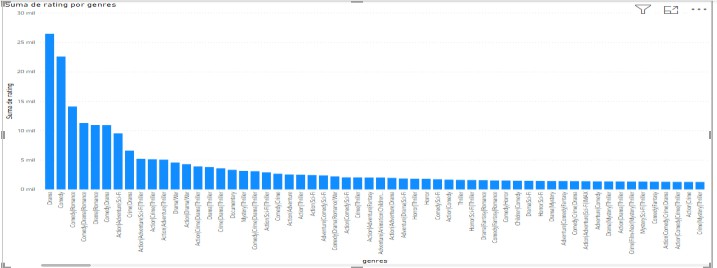


* 1. **Ratings de Películas por titulo**

Se generó un gráfico de barras representando la distribución de ratings en cada titulo. Se observó que algunos títulos presentaron un mayor número de películas con calificaciones altas, lo que sugiere una tendencia en la mejora de la calidad cinematográfica percibida.

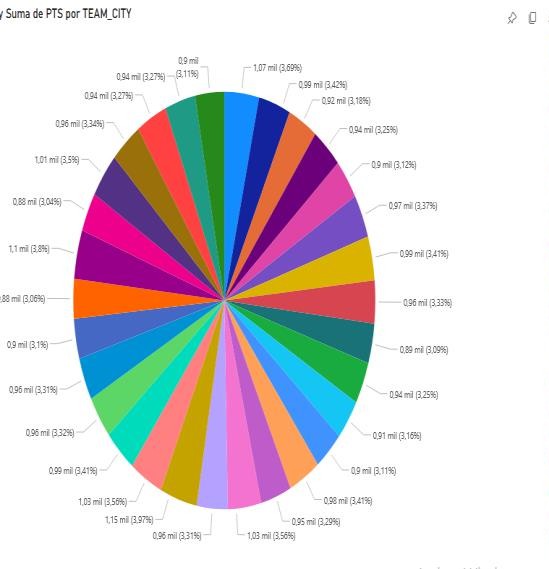
* 1. **Géneros Más Populares**

Se visualizó la frecuencia con la que cada género de película fue

estrenado y evaluado. Géneros como acción, drama y comedia fueron consistentemente populares a lo largo de los años.

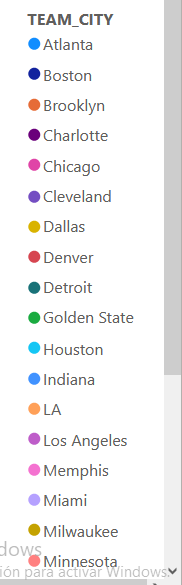
* 1. **Eventos de hockey por Ciudad**

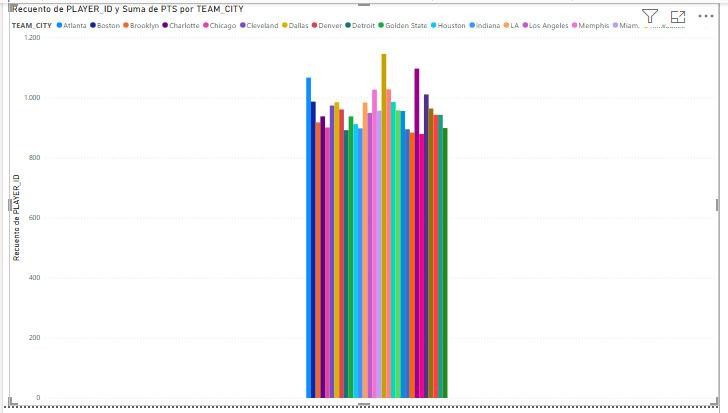
Se generó diagrama que representa la distribución de eventos por ciudad, identificando aquellas con mayor actividad en hockey.



* 1. **Sumatoria de Puntos Anotados**

Se creó un gráfico de barras para mostrar la cantidad total de puntos anotados en cada ciudad, permitiendo identificar tendencias en la ofensiva de los equipos.





1. **Interpretaciones de los datos obtenidos**

* **Frecuencia de Conciertos:** La banda ha mantenido una actividad constante en giras, con algunos años de mayor intensidad debido al lanzamiento de nuevos álbumes.
* **Recintos Más Frecuentes:** Metallica ha priorizado ciertos lugares icónicos para sus conciertos, reflejando una estrategia de consolidación en mercados clave.
* **Géneros Más Frecuentes:** Géneros como acción, comedia y drama dominaron el mercado, reflejando las preferencias del público.
* **Evolución de la Calidad Percibida:** A lo largo de los años, se evidencia una tendencia por títulos que se realizaron entre el 96 hasta el 2005 en la mejora de la calificación promedio de las películas, especialmente en géneros/títulos específicos.

**Frecuencia de Eventos:** Se identificaron ciudades con una alta cantidad de eventos de hockey, reflejando la popularidad del deporte en dichas regiones.

* **Puntuación por Ciudad:** Algunas ciudades registraron una mayor sumatoria de puntos, lo que podría indicar estilos de juego más ofensivos o una alta competitividad entre equipos.

1. **Conclusión**

El análisis de los conciertos de Metallica entre 1986 y 2024 muestra una evolución en sus giras, con una tendencia a realizar presentaciones en recintos masivos y festivales internacionales. La distribución geográfica de sus conciertos indica una fuerte presencia en América del Norte y Europa, aunque también han explorado mercados emergentes.

En el análisis de eventos de cinematografía permite comprender la evolución de los eventos de cine y sus calificaciones entre 1996 y 2016, destacando tendencias en la percepción de la calidad cinematográfica y la popularidad de los distintos géneros.

El análisis de eventos de hockey por ciudad y la sumatoria de puntos anotados proporciona una visión clara sobre la distribución y el impacto del deporte en distintas regiones donde asisten a eventos públicos de hockey. Estos datos permiten comprender mejor la dinámica de los partidos y apoyar estrategias tanto para equipos como para organizadores de eventos.

El uso de herramientas de visualización como Power BI ha permitido resaltar patrones clave en la frecuencia de conciertos, los lugares más visitados y la evolución en la asistencia del público, proporcionando información valiosa para entender la trayectoria de Metallica en el escenario mundial.

1. **Link del grafico en Línea**

* [**https://app.powerbi.com/groups/me/repor**](https://app.powerbi.com/groups/me/reports/2c03fd5b-0f72-4185-90dd-2d07a27333a9/8e90131108a2ca50b298?experience=power-bi)[**ts/2c03fd5b-0f72-4185-90dd-**](https://app.powerbi.com/groups/me/reports/2c03fd5b-0f72-4185-90dd-2d07a27333a9/8e90131108a2ca50b298?experience=power-bi)[**2d07a27333a9/8e90131108a2ca50b298?ex**](https://app.powerbi.com/groups/me/reports/2c03fd5b-0f72-4185-90dd-2d07a27333a9/8e90131108a2ca50b298?experience=power-bi)[**perience=power-bi**](https://app.powerbi.com/groups/me/reports/2c03fd5b-0f72-4185-90dd-2d07a27333a9/8e90131108a2ca50b298?experience=power-bi)

**Actividades y Hobbies**

**Nivel de estrés y estilo de vida de estudiantes, como comentarios al momento de utilizar Chatgpt , tiempo de utilización de redes sociales**

**1.- Definición del Caso de Estudio**

El presente estudio analiza la relación entre las actividades diarias, el nivel de estrés y el estilo de vida de los estudiantes, enfocándose en el uso de herramientas tecnológicas como ChatGPT y las redes sociales. En un contexto donde la carga académica y las responsabilidades personales pueden generar altos niveles de estrés, el uso de estas plataformas digitales se ha vuelto parte fundamental de la rutina estudiantil, ya sea como apoyo académico, entretenimiento o socialización. Sin embargo, el tiempo dedicado a estas herramientas puede influir tanto positiva como negativamente en la salud mental, la gestión del tiempo y el rendimiento académico de los estudiantes. Este estudio busca comprender cómo la interacción con ChatGPT y el uso de redes sociales impactan en su estilo de vida, analizando hábitos, motivaciones y percepciones para determinar si funcionan como mecanismos de afrontamiento o si, por el contrario, contribuyen al aumento de la ansiedad y la distracción.

1. **Enfoque**

El enfoque de este estudio es de tipo descriptivo y correlacional, basado en el análisis de datasets existentes que contienen información sobre actividades estudiantiles, niveles de estrés, estilo de vida y uso de herramientas tecnológicas como ChatGPT y redes sociales. Dado que los datos analizados provienen de fuentes diversas y no necesariamente están directamente relacionados, el estudio busca identificar patrones y posibles asociaciones entre estas variables, sin establecer causalidad directa. A través del procesamiento y la interpretación de estos datos, se pretende obtener una visión general sobre cómo el uso de tecnologías digitales puede influir en la rutina, el bienestar y el rendimiento académico de los estudiantes.

1. **Actividades Realizadas**
   * **Recopilación de Datos:** Se obtuvieron datasets existentes que contienen información sobre las actividades de los estudiantes, su nivel de estrés, su estilo de vida y el uso de herramientas tecnológicas como ChatGPT y redes sociales.
   * **Depuración de Datos:** Se limpiaron los datos eliminando valores atípicos, datos inconsistentes y registros incompletos para asegurar la calidad del análisis.
   * **Exploración y Análisis Preliminar:** Se revisaron las características principales de los datos mediante estadísticas descriptivas y visualizaciones básicas para comprender su distribución y detectar posibles patrones.
   * **Visualización de Datos:** Se generaron gráficos y representaciones visuales para identificar tendencias, correlaciones y relaciones entre las variables analizadas.
   * **Interpretación de Resultados:** Se analizaron los hallazgos obtenidos a partir de las visualizaciones, enfocándose en cómo el uso de ChatGPT y redes sociales puede estar relacionado con el nivel de estrés y el estilo de vida de los estudiantes.
2. **Recursos y Herramientas Utilizadas**
   * **Kaggle:** Se utilizó como fuente principal para la recopilación de datasets relacionados con actividades estudiantiles, nivel de estrés, estilo de vida y uso de herramientas tecnológicas.

**Jupyter Notebook:** Se empleó para la depuración y análisis preliminar de los datos, permitiendo realizar limpieza, transformación y exploración mediante librerías como Pandas, NumPy y Matplotlib.

* + **Power BI:** Se usó para la creación de visualizaciones interactivas, facilitando la identificación de patrones, correlaciones y tendencias en los datos de manera clara y comprensible.

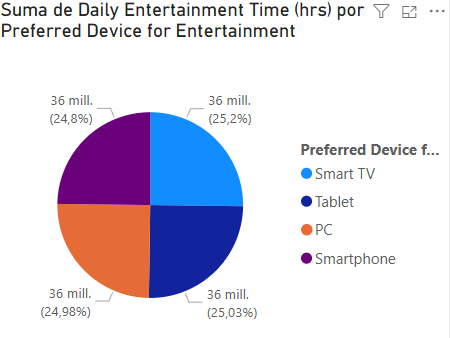
1. **Extracción de Datos**

Para el análisis, se seleccionaron columnas clave de los datasets disponibles en Kaggle, enfocándose en variables que permiten identificar patrones en el estilo de vida, nivel de estrés y desempeño académico de los estudiantes en relación con el uso de tecnología.

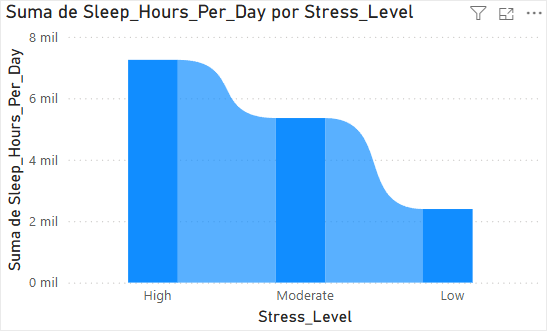
1. **ChatGPT:** Se analizaron los comentarios realizados por los usuarios, identificando las palabras más repetidas para detectar tendencias en su uso y posibles emociones asociadas.
2. **Redes Sociales:** Se extrajeron datos sobre el tipo de dispositivo utilizado, las horas de pantalla y las horas de sueño para evaluar su impacto en la rutina diaria de los estudiantes.
3. **Estudiantes - Nivel de Estrés:** Se consideraron tres niveles de estrés reportados en el dataset, así como las horas de sueño y las horas de estudio para determinar posibles correlaciones entre estos factores.
4. **Estudiantes - Rendimiento Académico:** Se analizaron los resultados de exámenes en función del nivel educativo de los padres, las horas de sueño y estudio, además del tipo de escuela (pública o privada) para evaluar su influencia en el desempeño escolar.
5. **Análisis y Visualización de Información**

**Análisis de comentarios en ChatGPT:** Se realizó un análisis de frecuencia de palabras en los comentarios de los usuarios utilizando un **mosaico de palabras**. Esta visualización permitió identificar los términos más recurrentes en los comentarios, revelando patrones de uso y posibles emociones asociadas al empleo de la herramienta. A través del mosaico, se destacaron las palabras clave más frecuentes, lo que facilitó la comprensión de cómo los estudiantes interactúan con ChatGPT y qué aspectos son más relevantes para ellos.

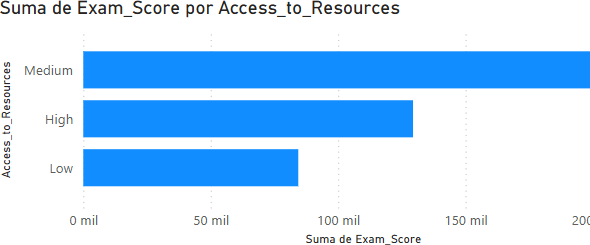
**Redes Sociales y Hábitos de Uso:** Se utilizó un **gráfico de pastel** para representar la suma de horas de entretenimiento entre diferentes dispositivos (como teléfonos, computadoras y tabletas), lo que permitió visualizar cómo se distribuye el tiempo dedicado al entretenimiento digital. Esto facilitó la evaluación de cómo las horas de pantalla varían según el tipo de dispositivo, lo que permitió identificar posibles impactos en la productividad y en las horas de sueño de los estudiantes.



**Niveles de Estrés en Estudiantes:** Se utilizaron **gráficos de cintas** para representar la suma de horas de sueño por día, lo que facilitó la visualización de la relación entre las horas de sueño y los niveles de estrés. Estos gráficos ayudaron a identificar tendencias y patrones en los estudiantes, mostrando cómo la cantidad de sueño influye en los niveles de estrés y bienestar general.



**Desempeño Académico:** Se emplearon **barras horizontales** para analizar la relación entre los resultados de exámenes y variables como los recursos económicos de los estudiantes (por ejemplo, nivel educativo de los padres) y la suma de los resultados del examen. Este gráfico permitió observar cómo los recursos económicos influyen en el desempeño académico, comparando la suma de los resultados del examen entre los diferentes grupos de recursos económicos y cómo se relaciona con el tipo de escuela (pública o privada).



1. **Interpretaciones de los datos obtenidos**

**Análisis de comentarios en ChatGPT:** Al analizar los comentarios de los usuarios mediante el mosaico de palabras, se observó que las palabras más repetidas fueron **“good”**, **“nice”** y **“excellent”**, lo que indica que, en general, los estudiantes tienen una percepción positiva sobre el uso de ChatGPT. Estas palabras reflejan satisfacción con la herramienta, sugiriendo que los usuarios encuentran útil y efectiva la asistencia proporcionada. El predominio de términos positivos también podría indicar que los estudiantes recurren a ChatGPT principalmente para obtener respuestas satisfactorias o asistencia, lo que puede contribuir a una experiencia más eficiente en sus actividades académicas. Sin embargo, aunque los comentarios son mayormente positivos, es importante destacar que no todos los usuarios comparten la misma opinión, y sería útil profundizar más para identificar posibles áreas de mejora en la herramienta.

**Redes Sociales y Hábitos de Uso:** A través del gráfico de pastel que muestra la distribución del tiempo de entretenimiento entre dispositivos, se encontró que los estudiantes distribuyen su tiempo de uso de tecnología de manera bastante equilibrada entre diferentes dispositivos. Los porcentajes son los siguientes:

* + **Smart TV**: 25.2%
  + **Tablet**: 25.03%
  + **PC**: 24.98%
  + **Smartphone**: 24.8%

Esto indica que no hay un dispositivo dominante para el consumo de redes sociales y entretenimiento digital, lo que sugiere que los estudiantes utilizan diversos dispositivos para acceder a contenido. Esta distribución equilibrada también puede reflejar la flexibilidad en el uso de tecnología, dependiendo de la situación o el tipo de contenido. Sin embargo, el hecho de que casi el 25% de los estudiantes utilicen smartphones, que son más propensos a ser distracciones, sugiere que el uso de estos dispositivos podría estar afectando la productividad y las horas de sueño, lo que a su vez impacta su bienestar y rendimiento académico.

**Niveles de Estrés en Estudiantes:** Al analizar las horas de sueño en función del nivel de estrés, se encontró lo siguiente:

* + **Nivel de Estrés Alto**: Los estudiantes con un nivel alto de estrés sumaron un total de **7,25 mil horas** de sueño. Esto podría indicar que, aunque estos estudiantes duermen más, el estrés se acumula debido a otros factores, como la carga académica o el tiempo limitado para actividades relajantes.
  + **Nivel de Estrés Medio**: Los estudiantes con un nivel de estrés medio sumaron **5,36 mil horas** de sueño. Este grupo muestra un equilibrio en las horas de descanso, pero podría beneficiarse de ajustes en sus hábitos de sueño o en la gestión de sus tareas para reducir el estrés.
  + **Nivel de Estrés Bajo**: Los estudiantes con un nivel de estrés bajo sumaron **2,40 mil horas** de sueño. Este bajo total de horas de sueño puede estar vinculado a un menor nivel de estrés, ya que aquellos que duermen más tienden a sentir menos presión, aunque es importante destacar que este grupo también podría estar menos concentrado debido a la falta de descanso.

Estos resultados reflejan una correlación interesante entre las horas de sueño y los niveles de estrés, lo que sugiere que una mayor cantidad de descanso puede estar asociada a menores niveles de ansiedad y mayor bienestar general.

**Desempeño Académico:** En el análisis del desempeño académico, se utilizaron **gráficos de barras** para comparar los resultados de los exámenes en función del nivel económico de los estudiantes. Los resultados sumados de los exámenes para cada nivel económico fueron los siguientes:

* + **Nivel Económico Alto**: Los estudiantes en este grupo sumaron **129 mil** puntos en los exámenes. Esto sugiere que los estudiantes con mayor poder adquisitivo tienden a tener un mejor desempeño académico, lo que podría estar relacionado con el acceso a mejores recursos educativos, materiales de estudio y apoyo extracurricular.
  + **Nivel Económico Medio**: Los estudiantes en este grupo sumaron **215 mil** puntos en los exámenes, lo que sugiere que, aunque su nivel económico no es tan alto, tienen un desempeño académico relativamente sólido. Este resultado podría reflejar que, a pesar de las limitaciones económicas, tienen acceso a recursos educativos suficientes para un rendimiento académico adecuado.
  + **Nivel Económico Bajo**: Los estudiantes con un nivel económico bajo sumaron **84 mil** puntos en los exámenes. Este grupo muestra un desempeño académico más bajo en comparación con los otros dos grupos, lo que podría indicar que las limitaciones económicas afectan el acceso a recursos educativos de calidad, lo que influye negativamente en su rendimiento académico.

Estos resultados reflejan una relación entre los recursos económicos y el desempeño académico, destacando la importancia de mejorar el acceso a recursos educativos para estudiantes de niveles económicos más bajos, lo que podría ayudar a reducir las disparidades en los resultados académicos.

**8. Conclusión**

El análisis de los datos recopilados ha permitido obtener una visión integral sobre diversos aspectos que influyen en la vida de los estudiantes, incluyendo el uso de herramientas como ChatGPT, el consumo de redes sociales y la relación entre el estrés y el desempeño académico.

1. **Uso de ChatGPT:** Los comentarios de los usuarios reflejan una percepción mayormente positiva sobre la herramienta, destacándose palabras como "good", "nice" y "excellent". Esto sugiere que los estudiantes encuentran útil la asistencia de ChatGPT, aunque también es importante considerar que podrían existir áreas de mejora, que podrían ser identificadas mediante un análisis más profundo.
2. **Redes Sociales y Hábitos de Uso:** La distribución equilibrada del tiempo de uso entre diferentes dispositivos (Smart TV, Tablet, PC y Smartphone) indica una preferencia por utilizar múltiples plataformas para el consumo de entretenimiento. Sin embargo, el uso intensivo de dispositivos móviles, como los smartphones, podría estar contribuyendo a la disminución de las horas de sueño y aumentando la distracción, lo que afecta la productividad y el bienestar general de los estudiantes.
3. **Niveles de Estrés en Estudiantes:** El análisis de los niveles de estrés revela una clara relación entre las horas de sueño y el estrés de los estudiantes. Aquellos con menos horas de sueño tienden a tener niveles más altos de estrés, lo que resalta la importancia de un descanso adecuado para mejorar la salud mental y el rendimiento académico. Los estudiantes con niveles de estrés bajos suelen dormir menos, lo que podría estar relacionado con su capacidad para manejar mejor las presiones académicas.

**Desempeño Académico:** Los resultados muestran una fuerte correlación entre el nivel económico de los estudiantes y su desempeño académico. Los estudiantes con mayor nivel económico obtuvieron mejores resultados en los exámenes, mientras que aquellos con menos recursos económicos mostraron un desempeño académico más bajo.

Esto pone de manifiesto la importancia de garantizar una educación accesible y equitativa, independientemente del contexto económico de los estudiantes, para reducir las disparidades en el rendimiento académico.

1. **Link de los gráficos en Línea**

[https://app.powerbi.com/view?r=eyJrIjoiNT](https://app.powerbi.com/view?r=eyJrIjoiNTdkMTdiN2UtMWI3ZC00ZTU5LWJkYzMtOTQ5ODZlZGU0NTQ1IiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9) [dkMTdiN2UtMWI3ZC00ZTU5LWJkYzMt](https://app.powerbi.com/view?r=eyJrIjoiNTdkMTdiN2UtMWI3ZC00ZTU5LWJkYzMtOTQ5ODZlZGU0NTQ1IiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9) [OTQ5ODZlZGU0NTQ1IiwidCI6IjY4MmE](https://app.powerbi.com/view?r=eyJrIjoiNTdkMTdiN2UtMWI3ZC00ZTU5LWJkYzMtOTQ5ODZlZGU0NTQ1IiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9) [0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMj](https://app.powerbi.com/view?r=eyJrIjoiNTdkMTdiN2UtMWI3ZC00ZTU5LWJkYzMtOTQ5ODZlZGU0NTQ1IiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9) [Y2ZDE4YWEzNyIsImMiOjR9](https://app.powerbi.com/view?r=eyJrIjoiNTdkMTdiN2UtMWI3ZC00ZTU5LWJkYzMtOTQ5ODZlZGU0NTQ1IiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9)

**Noticias y Eventos Mundiales**

1. **Definición del Caso de Estudio**

El caso de estudio se centra en el análisis global de tres aspectos clave: la temperatura mundial a lo largo de los años, representada en un archivo XLSM; la cantidad de radiación solar registrada en diferentes países por año, obtenida de un archivo XML; y el crecimiento poblacional a nivel global, también registrado por año en un XLMS. La finalidad es visualizar estos datos de manera clara e interactiva en Power BI, permitiendo identificar casos de estudio relevantes, como el país con mayor crecimiento poblacional, las regiones con mayor exposición solar a lo largo del tiempo o las tendencias en el aumento o disminución de la temperatura global. Estas visualizaciones facilitarán una comprensión más profunda de los datos y permitirán realizar comparaciones entre países y periodos específicos.

1. **Enfoque**

El enfoque para esta temática es la visualización y análisis de los datos relacionados con la temperatura mundial, la cantidad de radiación solar y el crecimiento poblacional para identificar patrones y tendencias significativas que permitan comprender mejor estos fenómenos a nivel global. Para ello, se recolectaron datos relevantes desde diferentes fuentes fiables, incluyendo archivos XLSM, XML y bases de datos públicas, y se transformaron a formatos compatibles como JSON mediante Jupyter Notebook con sus bibliotecas para que sean exportadas a MongoDB y finalmente terminen en un archivo tipo CSV. Posteriormente, se utilizaron plataformas como Power BI para crear gráficos y mapas interactivos que facilitan la visualización de información clave, como los países con mayor crecimiento poblacional, las regiones con más radiación solar, o las variaciones de temperatura a lo largo de los años. Finalmente, se interpretaron los resultados obtenidos, proporcionando hallazgos significativos que destacan relaciones y patrones entre los distintos conjuntos de datos analizados.

1. **Actividades Realizadas**
   * *Extracción de Datos:* Patricio Ponce se encargó de la extracción de datos relevantes sobre los temas mencionados en plataformas como UnData y bases de datos abiertas.
   * Transformación de Datos: Se desarrollaron scripts en Jupyter Notebook para transformar y limpiar los datos obtenidos, asegurando su coherencia y formato adecuado.
   * Visualización: Se integraron los datos en Power BI, creando gráficos y mapas interactivos que representan la evolución de la temperatura global a lo largo de los años, la distribución de la radiación solar en diferentes países y el crecimiento poblacional en distintas regiones.

Análisis de Resultados: A partir de la interpretación de los datos visualizados, se identificaron patrones relevantes, como los países con mayores incrementos de población, las regiones con una mayor cantidad de horas de sol y las áreas más afectadas por variaciones en la temperatura. Basándose en estos resultados, se generaron recomendaciones enfocadas en aprovechar los datos recopilados para investigaciones futuras, tales como analizar posibles relaciones entre el cambio climático y el crecimiento poblacional, o estudiar cómo la exposición solar podría influir en la calidad de vida de ciertas regiones.

1. **Recursos y Herramientas Utilizadas**
   * Fuentes de Datos: UnData, Bases de Datos Abiertas.
   * Herramientas de Transformación y Limpieza: Scripts personalizados en Jupyter Notebook para conversión de formatos y limpieza de datos.
   * Herramientas de Visualización: Power BI para la creación de gráficos y análisis visuales.
   * Bases de Datos SQL: Todos los datos se exportaron a MongoDb para que este los transforme en un formato compatible para su producto final, el cual es un Dashboard en MySQL Sever.
2. **Extracción de Datos**

Los datos fueron extraídos de UnData y Bases de datos abiertas, centrados en 3 casos de estudio clave:

* **Temperatura Mundial:** Se trata de un archivo XLSM que contiene información sobre la temperatura mundial a lo largo de los años, permitiendo identificar tendencias climáticas y cambios en diferentes periodos históricos.
* **Radiación Solar:** Un archivo XML que proporciona datos detallados sobre la cantidad de horas de sol registradas en diferentes países por año, lo que facilita analizar patrones

de exposición solar en distintas regiones del mundo.

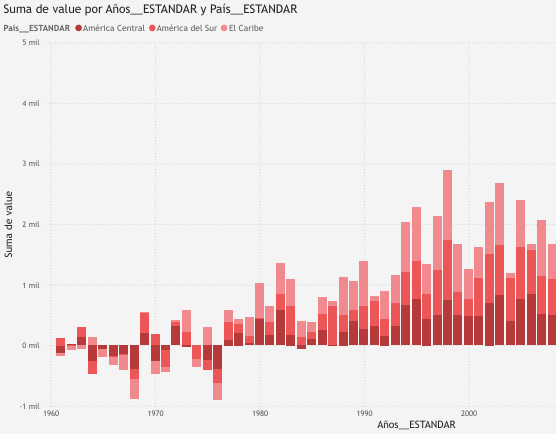
* **Crecimiento Poblacional**: Un archivo XML de datos con información sobre el crecimiento poblacional global por año, permitiendo observar las variaciones en la densidad poblacional y su distribución entre países.

1. **Análisis y Visualización de Información**

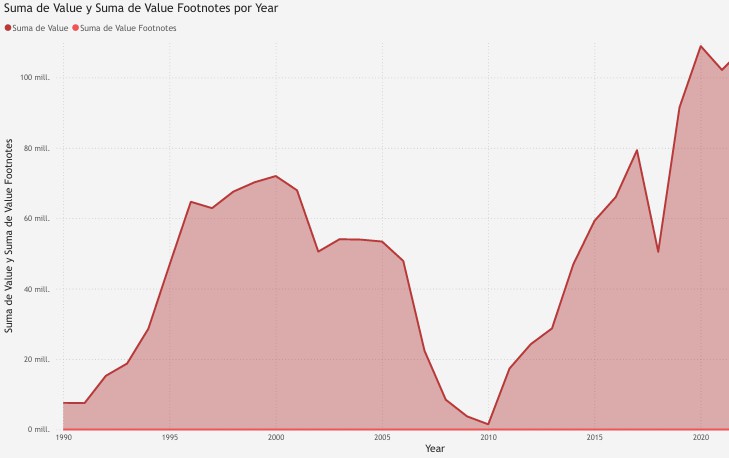
**Crecimiento Poblacional:** Se creó un gráfico de mapa de burbujas geográficas el cual representa el crecimiento poblacional en cada uno de los países el cual cuenta con registros del año 1960 incluso hasta con registros del año 2022, en este las burbujas más grandes señalan los países con mayor crecimiento poblacional, mientras que las más pequeñas indican un crecimiento menor.

* + 

**Temperatura Mundial***:* Se generó un gráfico de barras apiladas que representa la temperatura global en América Central, América del Sur y el Caribe a lo largo de varios años. Cada barra apilada muestra los valores de temperatura para las regiones mencionadas, y el apilamiento permite observar la contribución individual de cada región al total.



**Radiación Solar***:* Se generó un gráfico de áreas que representa la radiación solar mundial a lo largo de varios años. El área sombreada debajo de la línea muestra la magnitud acumulativa de los valores, lo que permite identificar patrones de aumento o disminución de la radiación solar en diferentes periodos.



1. **Interpretaciones de los datos obtenidos**

**Interpretación Gráfico 1 (Crecimiento poblacional en varios países):**

Los resultados obtenidos evidencian un crecimiento poblacional sostenido a nivel global, con variaciones significativas entre regiones y países. China e India presentan un incremento notable en comparación con

otras zonas, reflejado por los tamaños de las burbujas en el mapa.

Estas diferencias destacan la necesidad de estrategias demográficas específicas para afrontar los retos asociados con el aumento poblacional, como el acceso a recursos y la planificación urbana.

**Interpretación Gráfico 2 (Temperatura global en América Central, América del Sur y El Caribe):**

Este gráfico muestra un aumento progresivo de las temperaturas globales en las regiones de América Central, América del Sur y El Caribe a lo largo de las décadas, en las que se puede apreciar que El Caribe tiene la mayor cantidad de temperatura. Además, América del Sur también cuenta con una temperatura bastante alta y finalmente América Central es la región que se encuentra más abajo en esta lista de zonas en temperatura mundial. Cabe recalcar que datos resaltan tendencias alarmantes hacia el calentamiento global, especialmente después del año 2000.

Este comportamiento subraya la urgencia de implementar políticas ambientales que mitiguen el impacto del cambio climático en estas regiones vulnerables.

**Interpretación Gráfico 3 (Radiación solar mundial):**

Los datos reflejan fluctuaciones en la radiación solar mundial a lo largo de los años, con picos pronunciados en los períodos cercanos al año 2000 y 2020. No obstante, el período comprendido entre 2006 y 2010 destaca por una disminución abrupta en la radiación solar, interrumpiendo una tendencia ascendente general iniciada a partir de 2011. Esta anomalía temporal resulta particularmente interesante y amerita un análisis más detallado.

Estas variaciones podrían estar relacionadas con cambios climáticos o actividades humanas que impactan la atmósfera. Este análisis sugiere la importancia de profundizar en estudios que conecten estas tendencias con fenómenos como el calentamiento global y su influencia en la capacidad de los ecosistemas para absorber radiación solar.

1. **Conclusión**

A partir de las interpretaciones de los gráficos presentados, se ha evidenciado la necesidad de realizar estudios más detallados que aborden las interrelaciones entre el crecimiento poblacional, el cambio climático y la radiación solar. En un futuro, este análisis puede ser de gran ayuda para generar políticas de urbanización sostenible y la gestión de recursos naturales pueden mitigar los efectos adversos del crecimiento poblacional y el calentamiento global.

Además, sería valioso investigar cómo las variaciones en la radiación solar influyen en los ecosistemas locales y globales, y cómo estas interacciones pueden ser gestionadas para promover un desarrollo ambientalmente responsable. De esta manera, se permitirá desarrollar estrategias más efectivas para enfrentar los desafíos ambientales y demográficos del siglo XXI.

1. **Link de los gráficos en Línea**

[https://app.powerbi.com/view?r=eyJrIjoiOT](https://app.powerbi.com/view?r=eyJrIjoiOTVjMjU1ZDQtZGIzNS00ZjhhLThhNzItZWZlODI4MDYzZDMwIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9) [VjMjU1ZDQtZGIzNS00ZjhhLThhNzItZW](https://app.powerbi.com/view?r=eyJrIjoiOTVjMjU1ZDQtZGIzNS00ZjhhLThhNzItZWZlODI4MDYzZDMwIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9) [ZlODI4MDYzZDMwIiwidCI6IjY4MmE0Z](https://app.powerbi.com/view?r=eyJrIjoiOTVjMjU1ZDQtZGIzNS00ZjhhLThhNzItZWZlODI4MDYzZDMwIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9) [TZhLWE3N2YtNDk1OC1hM2FjLTllMjY2](https://app.powerbi.com/view?r=eyJrIjoiOTVjMjU1ZDQtZGIzNS00ZjhhLThhNzItZWZlODI4MDYzZDMwIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9) [ZDE4YWEzNyIsImMiOjR9](https://app.powerbi.com/view?r=eyJrIjoiOTVjMjU1ZDQtZGIzNS00ZjhhLThhNzItZWZlODI4MDYzZDMwIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9)

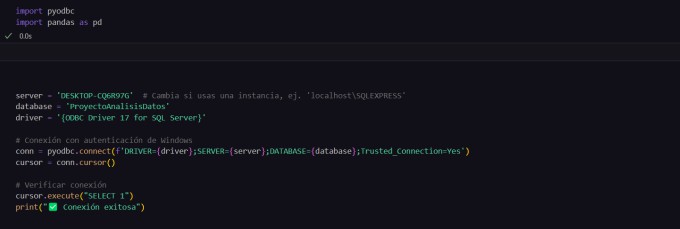
**Migración y Conteo de Datos en SQL Server**

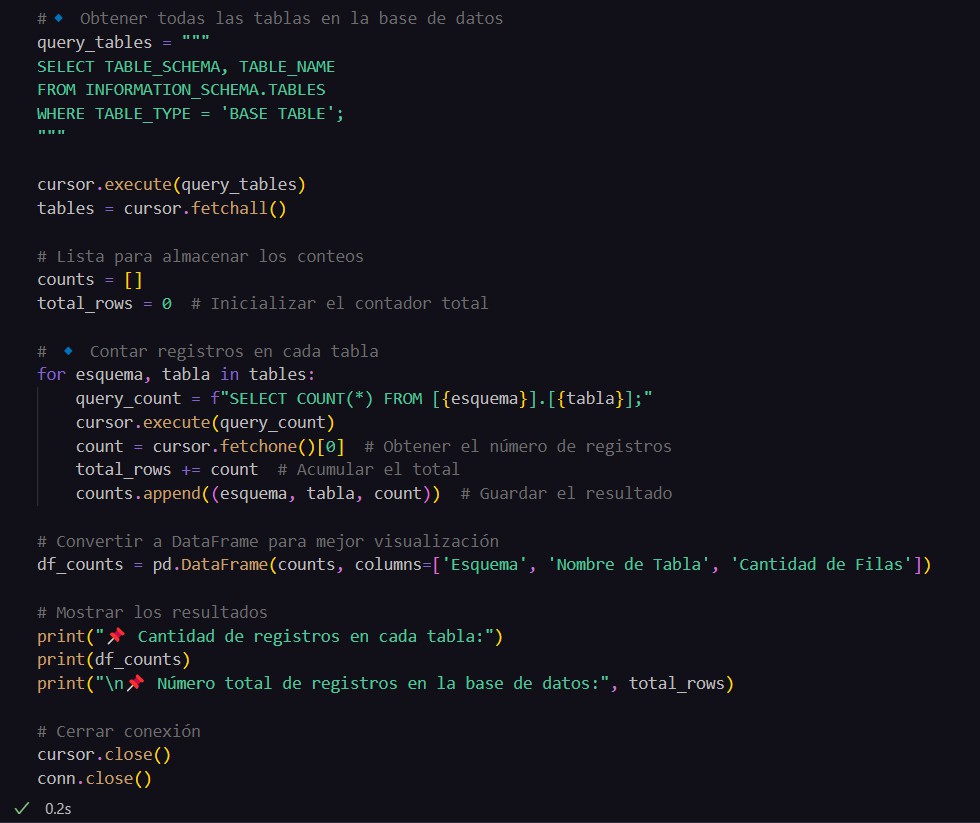
1. **Conteo de Datos en SQL Server**

**Objetivo**

Contar cuántos registros hay en cada tabla de la base de datos.

**Código en Python para contar registros**





Resultado esperado



1. **Migración de SQL Server a CSV usando Task Flat File**

**Objetivo**

Exportar las tablas de SQL Server a archivos CSV para su migración a otro sistema.

**Pasos en SQL Server Management Studio (SSMS)**

* 1. Abrir SQL Server Management Studio (SSMS).
  2. Seleccionar la base de datos ProyectoAnalisisDatos.

1. Clic derecho en la base de datos → Tasks → Export Data.
2. En la ventana "Choose a Data Source":
   * 1. Origen de datos: "SQL Server Native Client"
     2. Servidor: localhost
     3. Base de datos: ProyectoAnalisisDatos
3. En "Choose a Destination", seleccionar:
   * 1. Flat File Destination → Para

exportar a CSV.

* + 1. Ruta del archivo:

C:\Datos\mi\_tabla.csv

* + 1. Codificación: UTF-8.

1. Seleccionar la tabla que deseas exportar y hacer clic en "Next".
2. Finalizar la exportación.

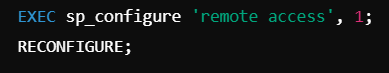
El archivo CSV estará guardado en C:\Datos\mi\_tabla.csv y listo para ser importado en otro sistema.

1. **Recomendaciones para el Hosting y Acceso Remoto**

Dado que no contamos con la posibilidad de utilizar un hosting en este momento, existen algunas formas en las que se podría alojar o compartir la base de datos SQL Server de manera remota.

**Opción 1: Configurar Acceso Remoto usando TCP/IP**

SQL Server permite que otros dispositivos se conecten de manera remota utilizando el protocolo TCP/IP. Para habilitar esto, se deben seguir los siguientes pasos:

* 1. **Habilitar conexiones remotas en SQL Server**
     1. En SQL Server Management Studio (SSMS)
  2. También se debe habilitar el protocolo TCP/IP en el **SQL Server**

**Configuration Manager**.

**Abrir el puerto 1433 en el Firewall**

SQL Server utiliza el puerto **1433**

para conexiones TCP/IP.

Es necesario configurar el firewall para permitir conexiones entrantes a ese puerto.

**Obtener la dirección IP pública**

Para compartir el acceso, se debe conocer la dirección IP pública del servidor SQL Server.

**Conectarse desde otro dispositivo**

En SSMS, en el otro equipo, se debe ingresar la dirección IP pública del servidor como:

* TU\_IP\_PUBLICA,1433

**Nota de Seguridad:**

A pesar de que este método permite el acceso remoto, **no se recomienda exponer SQL Server directamente en Internet sin medidas de seguridad adecuadas**. Dejar una base de datos abierta al público puede generar vulnerabilidades como accesos no autorizados y ataques de fuerza bruta.

Se adjunta un link con el back up de la base de datos si acaso se la quiera usar por medio remoto:

https://epnecuador- my.sharepoint.com/:u:/g/personal/mateo\_par edes\_epn\_edu\_ec/ESNLagtarUJNgppgX522 UF4BFSJ2c9ckcwWQwOQYdgDMtQ?e=L pJuWR

**Opción 2: Alternativas para Alojar SQL Server en un Servidor**

* Si en el futuro se requiere alojar la base de datos en la nube, se pueden utilizar opciones como: **Google Cloud SQL**: Servicio en la nube para bases de datos SQL Server.
* **Azure SQL Database**: Alternativa de Microsoft con integración directa con SSMS.
* **Oracle Cloud VPS**: Proporciona máquinas virtuales gratuitas donde se puede instalar SQL Server.

Por ahora, como no se tiene acceso a un hosting, se recomienda utilizar el **método de exportación a CSV** para compartir los datos cuando sea necesario.

**Cronograma:**

* **Recopilación de información (datasets)**
  1. Identificación de fuentes de datos
  2. Extracción y almacenamiento de datos
  3. Organización inicial
* **Depuración de datos**
  1. Eliminación de valores nulos y duplicados
  2. Corrección de errores y estandarización
  3. Transformación de datos según necesidad
  4. ​
* **Manejo de la arquitectura**
  1. Definición del entorno de trabajo (local, nube, bases de datos)
  2. Estructura de almacenamiento y optimización de datos
  3. Configuración de herramientas y pipelines de datos
* **Creación de las visualizaciones**
  1. Exploración de datos con gráficos
  2. Generación de dashboards y reportes interactivos
  3. Comparación de métricas y tendencias
* **Interpretación y conclusiones**
  1. Análisis de patrones y hallazgos
  2. Resumen de insights clave
  3. Recomendaciones para la toma de decisiones



**Recomendaciones**

Para realizar un proyecto de análisis de datos sólido, es fundamental utilizar fuentes confiables como Kaggle, Google Dataset Search, UCI Machine Learning Repository o datos oficiales de organismos como el Banco Mundial y el INEGI, evitando fuentes sin respaldo o sin metadatos claros. Una vez obtenidos los datos, se debe realizar una adecuada limpieza y preprocesamiento mediante la detección de valores nulos, eliminación de duplicados, estandarización de variables y manejo de valores atípicos utilizando herramientas como Pandas en Python y visualizaciones con boxplots. La selección de métodos de análisis debe estar alineada con los objetivos del estudio, empleando herramientas como Python para análisis estadístico y Power BI o Tableau para crear dashboards interactivos, asegurando el uso adecuado de gráficos como heatmaps para correlaciones, histogramas para distribuciones y gráficos de líneas o barras para tendencias, evitando errores como gráficos mal etiquetados o visualizaciones confusas. La interpretación de los resultados debe contextualizarse con estudios previos, evitando correlaciones engañosas y empleando storytelling con datos para una mejor comprensión, destacando no solo los hallazgos sino también su impacto y posibles aplicaciones. Finalmente, una documentación clara y estructurada es clave para la reproducibilidad del análisis, incluyendo la descripción del dataset, metodología utilizada, código fuente en plataformas como GitHub y conclusiones bien fundamentadas que respalden las recomendaciones derivadas de los datos, asegurando así un proyecto riguroso, claro y aplicable.

**Desafíos y problemas encontrados**

Durante el desarrollo del proyecto se presentaron diversos desafíos, comenzando con la selección y calidad de los datasets, ya que algunos contenían datos incompletos, desbalanceados o con información poco relevante, lo que requirió una depuración exhaustiva para garantizar análisis precisos. En la fase de limpieza y preprocesamiento, se identificaron valores nulos y atípicos que afectaban la consistencia de los resultados, por lo que se aplicaron técnicas de imputación y eliminación de datos irrelevantes. Otro desafío fue la integración de diferentes fuentes de datos, ya que cada dataset tenía estructuras y unidades de medida distintas, lo que complicó la comparación y relación de variables. En el análisis de comentarios en ChatGPT, el procesamiento del lenguaje natural (NLP) presentó dificultades debido a la necesidad de filtrar palabras sin significado relevante (stopwords) y determinar el sentimiento real de los comentarios sin sesgos. En cuanto al análisis de redes sociales y hábitos de uso, se encontró que los datos sobre tiempo de pantalla eran autodeclarados, lo que pudo introducir imprecisiones debido a sesgos en la percepción del tiempo por parte de los usuarios. En la parte de niveles de estrés y desempeño académico, uno de los principales problemas fue determinar relaciones causales, ya que la correlación entre variables como horas de sueño y nivel de estrés no implica necesariamente causalidad, requiriendo un análisis más profundo para evitar interpretaciones erróneas. En términos de visualización de datos, hubo desafíos en la selección de gráficos adecuados que permitieran representar la información de manera clara y sin ambigüedades, evitando el

uso de visualizaciones que pudieran inducir a confusión. Además, surgió una limitación técnica relacionada con el almacenamiento y compartición del proyecto, ya que GitHub no permitió subir los archivos debido a su gran tamaño, lo que requirió buscar soluciones alternativas, como el uso de almacenamiento en la nube o la reducción del tamaño de los archivos. Finalmente, la interpretación y comunicación de resultados supuso un reto al sintetizar grandes volúmenes de información en conclusiones claras y aplicables, asegurando que los hallazgos fueran comprensibles y relevantes para la toma de decisiones.

**Link de GitHub del Proyecto**

* El repositorio del proyecto, que incluye todo el código, datasets, y documentaciones, está disponible

**Link de los Videos**

Youtube

<https://youtu.be/dWVoasb7ggA>

<https://youtu.be/tEkiUDjBjhY>

**Referencias**

1. OpenWebinars, "¿Qué es MySQL?" OpenWebinars, [En línea]. Disponible en: [https://openwebinars.net/blog/que-](https://openwebinars.net/blog/que-es-mysql/) [es-mysql/](https://openwebinars.net/blog/que-es-mysql/). [Accedido: 2-agosto-2025].
2. MongoDB, "¿Qué es MongoDB?" MongoDB, [En línea]. Disponible en: [https://www.mongodb.com/es/compa](https://www.mongodb.com/es/company/what-is-mongodb)ny/what-is-mongodb. Accedido: 02- agosto-2025
3. IONOS. (2020). CouchDB

[Online]. Available: [https://www.ionos.es/digitalguide/ho sting/cuestiones- tecnicas/presentacion-de-couchdb/](https://www.ionos.es/digitalguide/ho%20sting/cuestiones-%20tecnicas/presentacion-de-couchdb/)

1. Elastic. (2021). Elasticsearch [Online]. Available: https://www.elastic.co/es/what- is/elasticsearch
2. Davinci. (2020). ¿Qué es Logstash? [Online]. Available: https://www.davincigroup.es/que-es- logstash-ejemplo-practico-de-uso/
3. Tableau. (2021). ¿Qué es Tableau? [Online]. Available: https://www.tableau.com/es-es/why- tableau/what-is-tableau