### **HISTORIAPARAGANDULES**

### **Descripción del Proyecto:**

El proyecto **HistoriaParaGandules** busca transformar la forma en que consumimos contenido histórico en Instagram. Actualmente, para encontrar un video específico en un perfil, es necesario desplazarse manualmente por el feed, lo que puede resultar tedioso y poco eficiente.

**Solución propuesta:**

Un mapa interactivo que permita localizar los videos en función de las ubicaciones a las que pertenecen, ya sea donde fueron grabados o sobre los lugares que tratan. Esta herramienta no solo facilita el acceso a contenido específico, sino que también mejora la experiencia del usuario al combinar geografía e historia en una interfaz visual.

### **Paso 1: Scraping con Instaloader**

Para comenzar, se realizó un scraping del contenido del perfil de Instagram utilizando la librería **Instaloader**. Este proceso permitió extraer información valiosa de cada reel, incluyendo:

* **Texto del reel**: La descripción o título que acompaña al video.
* **Fecha de publicación**: Cuándo se subió el video.
* **Número de likes y comentarios**: Para analizar el impacto de cada publicación.
* **Visualizaciones**: Un indicador clave de popularidad.
* **Otros metadatos relevantes**: Como hashtags, etiquetas o ubicación si están disponibles.

#### **Razones para usar Instaloader:**

* Es una herramienta robusta y sencilla para descargar datos directamente de perfiles públicos.
* Permite obtener información estructurada que luego puede procesarse fácilmente con Python.

#### **Resultados del scraping:**

* Los datos extraídos se guardaron en un archivo CSV para su posterior análisis y procesamiento. (Más adelanto lo cambio a excel por comodidad)
* Este archivo servirá como base para el análisis exploratorio y la generación del mapa interactivo.

### **Paso 2: Asignación de Ubicaciones**

Después de recopilar los datos iniciales de los reels con el scraping, el siguiente paso fue identificar manualmente las ubicaciones relacionadas con cada video utilizando **Google Maps**.

#### **Proceso realizado:**

1. **Consulta del contenido del reel:**
   1. Se analizó el texto y el contenido de cada reel para identificar referencias geográficas o lugares mencionados.
2. **Búsqueda en Google Maps:**
   1. Se utilizó **Google Maps** para determinar las coordenadas exactas (latitud y longitud) de cada ubicación.
3. **Registro de datos:**
   1. Las ubicaciones encontradas se añadieron manualmente al archivo de datos (excel\_info\_1.xlsx), creando nuevas columnas para la latitud, longitud y el nombre del lugar.

#### **Resultado:**

Este paso permitió enlazar cada reel con una ubicación específica, lo cual es clave para generar el mapa interactivo que se presentará en pasos posteriores.

### **Paso 3: Creación del Mapa Interactivo**

Con los datos recopilados y las ubicaciones asignadas, se generó un mapa interactivo utilizando la librería **Folium**. Este mapa permite a los usuarios visualizar los videos en sus respectivas ubicaciones.

#### **Proceso:**

1. **Preparación de datos:**
   1. Se utilizó el archivo Excel actualizado (excel\_info\_1.xlsx), que contiene las coordenadas geográficas (latitud y longitud) junto con los datos de cada reel.
2. **Uso de Folium:**
   1. Se creó un mapa base.
   2. Para cada reel, se añadió un marcador en la ubicación correspondiente.
   3. Cada marcador incluye un pop-up con información relevante del video:
      1. Texto del reel.
      2. Enlace al video.
3. **Exportación del mapa:**
   1. El mapa generado se guardó como un archivo HTML (mapita.html), listo para ser compartido o integrado en una web.

#### **Resultados:**

El mapa interactivo no solo facilita la localización de los videos, sino que también enriquece la experiencia del usuario al combinar contenido geográfico e histórico.

### **Paso 4: Limpieza, Procesamiento y Enriquecimiento de Datos**

En este paso, se procesaron y enriquecieron los datos recolectados para prepararlos para el análisis exploratorio y el mapa interactivo. Este trabajo se realizó en el archivo **testeando.ipynb**.

#### **Tareas principales:**

1. **Eliminación de emojis:**
   1. Se desarrolló la función eliminar\_emojis, que utiliza expresiones regulares para limpiar los textos del reel, eliminando todos los emojis.
2. **Extracción de hashtags:**
   1. Con la función extraer\_hashtags, se identificaron y extrajeron hashtags del texto de los reels, almacenándolos en una nueva columna.
3. **Clasificación temática de los textos:**
   1. Se implementó la función clasificartexto, que utiliza la API de OpenAI para clasificar automáticamente los reels en una de las siguientes categorías:
      1. Acontecimientos Históricos
      2. Toponimia de Lugares
      3. Biografías de Personajes Históricos
      4. Arquitectura
      5. Curiosidades Históricas
   2. Esto permitió añadir una nueva columna al dataset llamada Categoría, que facilita el análisis y la organización de los contenidos según su temática.
4. **Generación de títulos llamativos para redes sociales:**
   1. La función crear\_titulo\_redes\_sociales utiliza la API de OpenAI para generar títulos dinámicos, breves (máximo 10 palabras) y optimizados para redes sociales. Estos títulos no incluyen hashtags y están diseñados para captar la atención de los usuarios.
5. **Combinación de texto limpio y hashtags:**
   1. Se creó una nueva columna que combina el texto del reel (limpio de emojis) con los hashtags extraídos, proporcionando un campo enriquecido para análisis.
6. **Exportación de datos procesados:**
   1. Todos los datos enriquecidos y estructurados se guardaron en un archivo Excel denominado excel21deenero.xlsx.

#### **Resultados obtenidos:**

* Los datos están organizados y categorizados temáticamente.
* Los textos de los reels son más claros y fáciles de analizar.
* Los títulos generados automáticamente ofrecen una manera rápida de optimizar el contenido para redes sociales.
* Los datos están listos para el análisis exploratorio y la integración en el mapa interactivo.

### **Paso 5: Análisis Exploratorio de Datos (EDA)**

En este paso, se realizó el **Análisis Exploratorio de Datos (EDA)** para obtener una comprensión más profunda de las características y relaciones de los datos. Este proceso fue realizado en el archivo **EDA2.py**, y consistió en analizar las métricas clave de los reels, identificar patrones en los datos y visualizarlos de manera efectiva. A continuación, se detallan las tareas realizadas:

#### **1. Estadísticas Descriptivas**

Se generaron estadísticas descriptivas para las columnas numéricas del dataset: **Likes**, **Comentarios**, **Visualizaciones** y **Duración del video (s)**. Esto permitió obtener una visión general de la distribución de estos valores, como la media, mediana, desviación estándar, mínimo y máximo. Estas estadísticas ayudan a identificar si los datos contienen valores extremos (outliers) o si hay alguna anomalía en el comportamiento de las métricas clave.

#### **2. Top 5 Videos con Más Likes**

Se identificaron los 5 videos más populares en cuanto a **Likes**. Este análisis ayuda a comprender qué tipo de contenido genera mayor interacción con los usuarios. Al observar los textos asociados a estos videos, se pueden identificar patrones o temas que resuenan más con la audiencia.

#### **3. Visualización de Datos**

Se crearon varios gráficos para visualizar las métricas clave y detectar tendencias o patrones en los datos. Los gráficos generados fueron los siguientes:

* **Gráfico de Barras Apiladas (Engagement por Categoría):** Este gráfico muestra el **compromiso (engagement)** total por categoría, dividiendo los **Likes**, **Comentarios** y **Visualizaciones** para cada categoría de contenido. Esto permite observar qué categorías generan más interacción en términos de estas tres métricas, y se puede identificar qué tipo de contenido atrae más a la audiencia en general.
* **Gráfico de Barras (Comentarios por Categoría):** Este gráfico visualiza la cantidad total de **comentarios** generados por cada categoría. Al observar esta visualización, se puede evaluar si alguna categoría está generando más discusión entre los usuarios, lo que podría indicar que ciertos temas son más propensos a generar interacción.
* **Mapa de Calor de Correlaciones:** Utilizando un **mapa de calor**, se visualizó la correlación entre las variables numéricas (Likes, Comentarios, Visualizaciones y Duración del video). Este análisis permitió identificar si alguna de estas métricas está estrechamente relacionada entre sí, lo que puede proporcionar información sobre cómo ciertos factores (por ejemplo, duración del video) influyen en el engagement.

#### **4. Gráfico Interactivo de Relación entre Likes y Comentarios**

Se utilizó **Plotly** para crear un gráfico interactivo que muestra la relación entre **Likes** y **Comentarios**, donde el tamaño de cada punto está determinado por el número de **Visualizaciones**. Este gráfico proporciona una manera visualmente atractiva de analizar cómo interactúan estas métricas entre sí, y puede ofrecer insights sobre qué videos tienen una mayor tasa de comentarios en relación con sus Likes y visualizaciones.

#### **5. Conclusiones Generadas del EDA**

Al finalizar el análisis exploratorio, se extrajeron varias conclusiones clave:

* **Categorías más populares:** Se pudo identificar qué categorías generan más engagement en términos de **Likes**, **Comentarios** y **Visualizaciones**.
* **Impacto de la duración:** A través del análisis de la correlación, se observó cómo la **Duración del video** influye en las demás métricas. Por ejemplo, si los videos más largos tienden a generar MENOS interacciones.
* **Contenido con mayor interacción:** Se identificaron los videos con más **Likes** y **Comentarios**, lo que puede ayudar a detectar qué tipo de contenido atrae más a la audiencia y a ajustar futuras publicaciones en función de estos insights.

### **Posibles Mejoras Técnicas para el Proyecto HistoriaParaGandules**

#### **Automatización de Ubicaciones**

* **Geolocalización Automática**: Implementar un sistema que asigne coordenadas automáticamente, reduciendo la necesidad de realizar esta tarea de forma manual. Esto podría lograrse integrando APIs como Google Maps o OpenStreetMap.
* **Reconocimiento Geográfico Avanzado**: Desarrollar algoritmos de inteligencia artificial que analicen los textos y los videos de los posts para identificar referencias geográficas con mayor precisión, facilitando la asignación de ubicaciones.

#### **Enriquecimiento de Datos**

* **Ampliación Temática**: Mejorar la clasificación de contenido mediante la incorporación de nuevas subcategorías históricas, permitiendo una organización más detallada de los datos.
* **Etiquetado Semántico Detallado**: Crear un sistema avanzado de etiquetado semántico que ofrezca búsquedas más precisas y relevantes. Esto podría incluir la identificación de temáticas específicas dentro del contenido histórico.

#### **Mejoras en la Experiencia de Usuario**

* **Filtros Dinámicos**: Introducir filtros que permitan a los usuarios personalizar su experiencia, como:
  + Buscar contenido por época histórica.
  + Filtrar publicaciones por categorías específicas.
  + Ordenar resultados según su popularidad o interacción.
* **Interactividad y Redes Sociales**: Incorporar funciones que permitan compartir ubicaciones directamente desde el mapa interactivo, así como la posibilidad de que los usuarios contribuyan con nuevas ubicaciones históricas de manera colaborativa.

#### **Optimización del Contenido**

* **Análisis de Engagement**: Profundizar en el análisis de las métricas de interacción para entender qué categorías de contenido generan mayor interés. Por ejemplo, explorar por qué temas como la toponimia de lugares tienen tanto éxito.
* **Estrategias Basadas en Datos**: Ajustar la creación de contenido en función de los datos obtenidos sobre likes, comentarios y visualizaciones, maximizando así el impacto de cada publicación.

La clave está en continuar innovando, aprovechando herramientas tecnológicas emergentes y perfeccionando el análisis de datos para ofrecer una experiencia enriquecedora, interactiva y altamente atractiva para los usuarios.