



Evidencias Challenge 1 | Ramón S.

Aa Name	📅 Date Created	☰ RETO_1	🕒 Status	☰ Text
🔗 "Pandas documentation" CICLO 1 INVESTIGA	@March 8, 2023		Done 🟢	Se realizó la consulta acerca de la biblioteca pandas y se encontró valioso aprendizaje acerca de como utilizar la herramienta para la manipulación de datos, estructuras de datos flexibles, funciones de lectura y escritura, visualización de datos, integración con otras bibliotecas como Matplotlib y Seaborn, funciones estadísticas, funciones de series de tiempo, indexación y selección de datos, etc.
🔗 "Get started visualizing data in Python using Matplotlib, Pandas and Seaborn", etc."	@March 10, 2023		Done 🟢	Se realizó la consulta acerca de los métodos de análisis exploratorio de datos y se encontró valiosa información acerca de como utilizar las distintas técnicas de análisis y bibliotecas de python para analizar conjuntos de datos.
🔗 "Python MySQL Database Connection using MySQL Connector" CICLO 1 INVESTIGA	@March 12, 2023		Done 🟢	se investigó y consultó la documentación acerca de como gestionar la conexión y la utilización de bases de datos, específicamente de MySQL, con lo cual se instaló en WorkBench y también se averiguó la manera de visualizar las tablas de una forma amigable y fácil, por lo cual se llegó hasta la manera de ver tablas a través de la conexión asistida con XAMPP y su visualizador o administrados de BD llamado phpmyadmin lo cual facilita enormemente la visualización de la data que va ingresando a la misma en el momento de creación, edición y carga de la tabla de datos, también para ir verificando que los cambios se estén realizando de manera satisfactoria, también, se encontró la manera de conectar la base de datos local para ser utilizada desde Google Colab con un servicio llamado NGROK, el cual hace de forwarding y asegura la conexión puente entre los servicios de terceros como colab y la máquina local.
🔗 "Python MySQL Database Connection using MySQL Connector" CICLO 1 INVESTIGA	@March 14, 2023		Done 🟢	Se investigó las distintas maneras de hacer la conexión entre python y MySQL, además de como crear y borrar tablas, editar tablas, agregar y eliminar registros, cambiar la posición de las filas y columnas e incluso hacer correcciones a los datos internamente en la tabla, y finalmente de como obtener o recuperar los datos de la tabla con distintos criterios o queries, y todo esto manejado con instrucciones desde el notebook de Python.
🔗 "Consulta de Plantillas e instructivos" CICLO 1 INVESTIGA	@March 14, 2023		Done 🟢	Se consultó las plantillas donde vamos a trabajar tanto en miro como en Notion para ir documentando el progreso de las actividades, además se consultó los instructivos para elaborar mapas mentales y el Road Map del Challenge.

Aa Name	📅 Date Created	☰ RETO_1	🕒 Status	☰ Text
👤 "Sesión ID 1" CICLO 1 DESARROLLA	@March 14, 2023		Done 🟢	Se utilizó el código Python en GitHub para: <ul style="list-style-type: none"> Extraer tweets: la biblioteca para manejar la API (Tweepy) no conto con datos de acceso para ser utilizada, por lo tanto se cambió para investigar y hacer uso de las distintas bibliotecas para hacer scrapping de información en twitter. Se eligió "snsrape" Verificar tipos de datos: en este punto, para verificar los tipos de datos, se utilizó el método 'dtypes' de Pandas, que devuelve una lista de los tipos de datos de cada columna en el conjunto de datos, los tipos de datos son . Verificar estructura de datos: para verificar la estructura de los datos, se utilizó el método 'info()', que proporciona información sobre el número de filas y columnas, los nombres de las columnas y los tipos de datos de cada columna. Por último para el ciclo 1 de desarrollo se comenzó con el desarrollo de las distintas fases del challenge , comenzando con la sección de instalacion de dependencias o bibliotecas necesarias para realizar las diferentes secciones de codigo subsiguientes, en esta etapa se definió la lógica de scrapping de tweets con los criterios determinados por el challenge, como lo son , region (AR), cantidad de tweets (50000), principales trending topics, idioma (es-ar), etc, para finalmente extraer los tweets de interés que van a ser utilizados en el análisis.
🔧 "MySQL Installation, python MySQL conexión, Querys: comandos de consulta para BD tablas and manipulación de data, relaciones entre tablas y datos, primary and foreign keys" CICLO 2 INVESTIGA	@March 16, 2023		Done 🟢	En éste ciclo se investigó sobre la instalación de MySQL Workbench, el conector de python con MySQL, y los principales query para manipular tablas en MySQL, tales como crear y borrar tablas , modificar tablas en cuanto a indices, filas, columnas para su organización .
👤 "Sesión ID 2" CICLO 2 DESARROLLA	@March 18, 2023		Done 🟢	Como evidencia y novedad se creo una rutina para cargar data a la BD con un bucle para la carga por baches y una rutina para retomar la carga en donde quedó en caso de caída de la conexión, también se logró hacer la conexión directa de los recursos de computo de colab con los recursos locales a través de un servicio llamado ngrok , que, entre otras cosas, permite usar y conectar la BD local y cargarla con resultados obtenidos en colab con el cómputo de google que es más rapido y poderoso que los recursos locales del ordenador. Paralelamente se desarrollaba las rutinas de limpieza y traducción de datos, y la rutina de aplicación de la biblioteca textBlob a los datos resultantes de la limpieza y traducción al idioma inglés.

Aa Name	📅 Date Created	☰ RETO_1	🕒 Status	☰ Text
🧠 "Análisis exploratorio de datos, técnicas de ETL, visualización y presentación de resultados" CICLO ID 3 INVESTIGA	@March 22, 2023		Done 🤖	En este ciclo se investigó sobre las técnicas EDA (Exploratory Data Analysis) de análisis de datos y los procesos ETL (Extract, Transform, Load), los cuales son fundamentales en el campo de la ciencia de datos. La EDA es un proceso que implica analizar y visualizar datos con el fin de resumir sus principales características y encontrar patrones, tendencias y anomalías en ellos. Esta técnica es especialmente útil cuando se trabaja con grandes conjuntos de datos, ya que permite identificar rápidamente problemas o anomalías que podrían afectar la calidad de los datos. Por otro lado, el proceso ETL se refiere a la extracción, transformación y carga de datos desde múltiples fuentes de datos a un sistema de destino. Es un proceso fundamental en la integración de datos y se utiliza para combinar datos de diferentes fuentes en un solo sistema para su análisis. Durante el ciclo también se investigó sobre el proceso de visualización de datos, el cual es fundamental para comunicar los resultados del análisis de datos a los interesados y tomar decisiones informadas. La visualización de datos permite resumir grandes cantidades de datos en gráficos o tablas fáciles de interpretar, lo que permite una mejor comprensión de los resultados y una toma de decisiones más eficiente.
👤 "Sesión ID 3" CICLO 3 DESARROLLA	@March 24, 2023		Done 🤖	Como evidencia en esta etapa se realizó la logica de análisis exploratorio de datos para procesar graficar y analizar en total 50330 tweets extraidos en el rango de fechas entre el 01-01-2023 hasta el 19-03-2023, esta cantidad de tweets se deprenden de alrededor de 20 Trending topics, el analisis exploratorio de datos se realizó tomando en consideracion el top 10 de trending topics a los cuales se les realizó un analisis univariado y bivariado, se realizaron gráficas de barras , de torta, gráficos de línea, diagrama de dispersión , mapa de calor , diagrama de cajas y bigotes , diagrama hexadecimal, y mapa conceptual, con los cuales se analizan todas las aristas que se pueden abordar para poder desentrañar las claves y significado de los datos en resultados concretos. Finalmente Análisis y visualización de datos: • Utilizar bibliotecas de procesamiento de lenguaje natural para realizar análisis de sentimiento en el texto de los tweets. • Asignar una polaridad positiva, negativa o neutral al texto del tweet. • Utilizar herramientas de visualización de datos para presentar los resultados del análisis de sentimiento, como gráficos de barras, diagramas de dispersión y mapas de calor, etc. Además, se utilizó Python y bibliotecas como NLTK, TextBlob, VADER y otras bibliotecas de análisis de datos para realizar este análisis de sentimientos y el EDA. Por último se utilizó una base de datos relacional como MySQL para almacenar y gestionar los datos recopilados y los resultados del análisis.
👤 "Sesión ID 3" Link a Repositorio CICLO 3 DESARROLLA	@March 27, 2023		Done 🤖	https://github.com/ramonserrano76/Digital_Nao_Python_Dev_Retos