**PC1: Consolidación de conocimientos de Pandas, SQL y NoSQL**

**Objetivo**

Este proyecto tiene como objetivo consolidar conocimientos sobre la manipulación de datos con Pandas, el manejo de bases de datos relacionales (SQLite) y la preparación de los datos para una posible inserción en bases de datos no relacionales (MongoDB). El tiempo estimado de trabajo es de 2-3 horas por alumno.

Además, este proyecto integra elementos de cómo contribuir a bases de código ya existentes el GIT, por lo que lee cuidadosamente las instrucciones de entrega al final del README.

**Instrucciones detalladas**

**1. Descarga de datos**

* Descarga los datasets para vinos tintos y blancos desde la siguiente URL: [Wine Quality Dataset](http://archive.ics.uci.edu/dataset/186/wine+quality).
* Asegúrate de que ambos datasets se encuentren disponibles en tu entorno de trabajo como archivos CSV.
* Idealmente, intenta llevarlo a cabo de manera programática para no tener que hacer operaciones manualmente.

**2. Combinar los datos**

* Usa Pandas para cargar ambos datasets en dataframes en memoria.
* Combina los datos en un único dataframe añadiendo una columna adicional que indique el tipo de vino (red o white).
* ¿Cuántos registros tenemos? ¿Cuántas variables y de qué tipo?

**3. Filtrar atípicos y manejar datos ausentes**

* Realiza un análisis estadístico o inspección visual de cada columna numérica para identificar valores atípicos.
* Usa Pandas para filtrar y eliminar los datos atípicos y los valores ausentes. Explica en tu entrega qué criterios utilizaste para identificar los atípicos.

**4. Almacenar los datos limpios en SQLite**

* Usa SQLite para almacenar el dataframe limpio en una base de datos persistente.
* Sigue la documentación oficial de SQLite: [SQLite Python Documentation](https://docs.python.org/3/library/sqlite3.html).

**5. Realizar 3 consultas en SQLite**

Basándote en los datos y las columnas del dataset, realiza las siguientes consultas:

1. **Consulta 1**: ¿Cuál es el promedio de calidad (quality) por tipo de vino (type)?
2. **Consulta 2**: ¿Cuántos vinos tienen un nivel de alcohol superior a 10.5, agrupados por tipo?
3. **Consulta 3**: Obtén el conteo de vinos por nivel de acidez (fixed acidity) agrupados en rangos (por ejemplo, de 0-5, 5-10, 10-15).

**6. Exportar datos a JSONLines**

De cara a una potencial insercion en una base de datos noSQL como mongoDB, podemos servirnos de pandas para preparar los datos.

* ¿Qué estructura de datos de python es la más similar a un documento noSQL?
* Usa Pandas para transformar los datos de una de las consultas en un archivo JSONLines.
* Usa la librería jsonlines para guardar el archivo.
* ¿Qué problemas podrían surgir al transformar un dataframe en jsonlines?
* Añade una columna que sea originalmente un np.array,¿qué sucede al transformarlo en jsonlines?
* Añade una columna que sea originalmente un pd.datetime,¿qué sucede al transformarlo en jsonlines?

**7. Análisis de calidad de los vinos**

* Inspecciona qué caracteriza a los vinos tintos y blancos con mayor calidad (quality).
* Usa análisis estadístico, gráficos o cualquier técnica que consideres relevante para identificar patrones.

**Instrucciones para la entrega**

La entrega de este proyecto de consolidación simula la contribución a un proyecto ya existente. En estos casos es vital evitar que impactéis el trabajo de otros compañeros.

1. No clonéis el [repositorio](https://github.com/MBIT-School-Academy/MIA-MDS-MDE-PC1.git) sino que en su lugar, haced un fork del mismo para tener un repositorio de trabajo propio para vuestro grupo.
2. Cread una rama dentro de vuestro repositorio siguiendo la convención feature/MAR25\_GRUPO\_X.
3. Logead el esfuerzo en diferentes commits, asegurándoos de realizar al menos un commit por persona.
4. Realizad un commit final con el mensaje: **ENTREGA FINAL**.
5. Cuando esté listo, uno de los miembros del grupo creará una Pull Request en el [repositorio](https://github.com/MBIT-School-Academy/MIA-MDS-MDE-PC1.git) original. En este [enlace](https://docs.github.com/es/pull-requests/collaborating-with-pull-requests/proposing-changes-to-your-work-with-pull-requests/about-pull-requests) tenéis información sobre las Pull Requests.

**Notas adicionales para la entrega**

* Asegúrate de documentar cada paso en tu entrega y justificar tus decisiones (por ejemplo, cómo identificaste los atípicos).
* El entregable principal será un Jupyter notebook con todos los pasos que habéis seguido y las explicaciones que consideréis pertinentes.
* Incluye en un fichero requirements.txt todas las librerías que estés utilizando.
* La entrega se hará tanto a través de Campus como en el repositorio de Git asignado para el proyecto.

¡Buena suerte con el proyecto! 😊