



Día 1: Introducción y Preparación de los Datos

Objetivo del día:

- Conocer las bases de datos que van a analizar.
- Aprender a cargar, limpiar y explorar los datos con pandas.
- Formulación de preguntas y una hipótesis clara para el análisis.

Tareas del día:

1. Presentación del taller:

- a. Explicar los objetivos del taller y cómo se evaluará.
- b. Explicar cómo usar Git y cómo crear un repositorio para almacenar su trabajo.
- c. Asegurarse de que todas las integrantes del equipo tengan acceso a las bases de datos y el código necesario.

2. Carga y limpieza de los datos:

- a. Las integrantes del equipo deben cargar las bases de datos proporcionadas.
- b. Usar pandas para inspeccionar las primeras filas con `df.head()` y entender la estructura de los datos.

c. Limpiar los datos eliminando duplicados (`df.drop_duplicates()`) y manejando los valores nulos (`df.dropna()`).

d. Documentar los pasos en un informe en PDF.

3. Exploración de los datos:

a. Visualizar distribuciones básicas con **matplotlib** y **seaborn** (ejemplo: histogramas, gráficos de cajas).

b. Crear una lista de preguntas basadas en los datos, por ejemplo:

- i. ¿Cuál es el diagnóstico más frecuente?
- ii. ¿Cuál es la distribución de edades en diferentes unidades funcionales?
- iii. ¿Existen diferencias significativas entre los géneros?

4. Formulación de hipótesis:

a. Las integrantes del equipo deben formular una o dos hipótesis basadas en sus preguntas. Por ejemplo, si están analizando enfermedades, podrían formular la hipótesis: *"La fiebre tifoidea es más común en un grupo de edad específico."*

5. Entregables del día:

- a. Subir los datos limpiados a Git.
- b. Incluir el código para cargar, limpiar y explorar los datos en el repositorio.
- c. Escribir el primer borrador del pdf con título, objetivos, preguntas y las primeras observaciones.



Objetivo del día:

- Realizar un análisis más profundo usando pandas, numpy, y crear visualizaciones avanzadas.
- Comenzar a escribir conclusiones basadas en el análisis.

Tareas del día:

1. Análisis de los datos:

- a. Usar **pandas** para realizar cálculos específicos como la suma y el producto de ciertos campos de interés.
- b. Crear filtros para responder las preguntas planteadas. Por ejemplo, filtrar por diagnóstico o por unidad funcional.

2. Visualización de los datos:

- a. Crear gráficos de barras, gráficos de líneas, gráficos de dispersión y gráficos de cajas usando **matplotlib** y **seaborn**.

3. Hypothesis testing:

- a. Realizar un análisis que permita confirmar o refutar las hipótesis planteadas el primer día.
- b. Por ejemplo, si la hipótesis es que "La fiebre tifoidea es más común en un grupo de edad específico", pueden usar **pandas** para calcular la distribución de edades para este diagnóstico.

4. SQL y NoSQL:

- a. Si están trabajando con bases de datos SQL o NoSQL, deben escribir los scripts o dumps para consultas importantes que hayan utilizado en su análisis.

5. Documentación y Reporte:

- a. Escribir las primeras conclusiones del análisis en el informe en PDF.
- b. Agregar capturas de pantalla del código y gráficos generados.
- c. Subir el archivo .pbix de power BI

6. Entregables del día:

- a. Subir el análisis y las visualizaciones al repositorio de Git.
- b. Incluir el código de análisis y visualización con comentarios explicativos.
- c. Escribir las primeras conclusiones en el informe.

Día 3: Refinamiento, Conclusiones y Entrega

Objetivo del día:

- Refinar el análisis y concluir el proyecto.
- Generar los entregables finales y asegurar que todo esté bien documentado.

Tareas del día:

1. Refinamiento del análisis:

- a. Revisar el análisis realizado y hacer ajustes si es necesario.

- b. Realizar análisis más detallados o pruebas adicionales si las preguntas o hipótesis no se resolvieron por completo.

2. Finalización del reporte:

- a. Redactar la conclusión final del análisis, resumiendo los hallazgos más importantes.
- b. Asegurarse de que todo esté claramente documentado en el informe en PDF, incluyendo:
 - i. Título y objetivo del análisis.
 - ii. Preguntas y hipótesis.
 - iii. Pasos realizados.
 - iv. Resultados y hallazgos.
 - v. Conclusiones finales.

3. Revisión de los entregables:

- a. Comprobar que el repositorio de Git contiene el código bien organizado.
- b. Verificar que todos los pasos estén documentados y que se haya usado numpy, pandas, matplotlib y seaborn en el análisis.
- c. Revisar que los scripts SQL o dumps NoSQL estén correctos y bien documentados.

4. Entrega final:

- a. Subir todos los archivos finales a GitHub, asegurándose de que todo esté organizado (código, análisis, gráficos, informes).
- b. Comprimir y entregar el repositorio junto con el informe en PDF

5. Entregables del día:

- a. Subir el repositorio final a Git.

- b. Entregar el archivo .pbix (si se utiliza Power BI).
- c. Entregar el informe en PDF con todos los detalles

Requisitos del proyecto:

- **Repositorio de Git** con el código completo.
- **Código de Python** con el uso de pandas, numpy, matplotlib, seaborn.
- **Scripts SQL o dumps NoSQL.**
- **Archivo .pbix** si se está usando Power BI.
- **Capturas de pantalla:** Estas son complementarias y deben ir en el informe PDF.

Liderado por:



Fomentado por:



Con el Apoyo de:

