



# Proyecto metodología de ciencias de datos

## 1. IDENTIFICACIÓN DE ASPECTOS BÁSICOS

**Denominación del Programa de Formación:** Análisis y Desarrollo de Software

**Instructor:** Edicson Quiroz

**Fase:** Evaluación

**Actividad de Proyecto:** Desarrollar las tareas de configuración y puesta en marcha del software.

**Competencia:** Desarrollar la solución de software de acuerdo con el diseño y metodologías de desarrollo.

**Resultados de Aprendizaje a Alcanzar:** Codificar la solución que cumpla con el diseño establecido.

**Actividad Evaluativa:** Proyecto metodología de ciencias de datos

Utilizando Python y el editor de Visual Studio Code a través de la línea de comandos, u otro editor que permita realizar la ejecución y utilizando el archivo operations descargado del siguiente link (<https://www.kaggle.com/datasets/usaf/world-war-ii?resource=download>) y desarrolla:

1. Realiza inicialmente actividades de extracción, limpieza y transformación de datos
2. Ver los detalles estadísticos del dataset
3. Generar gráfico estadístico
4. Calcula la media, mediana y moda para la columna Total Weight
5. Genera un reporte en HTML
6. Dividir nuestros datos en “atributos” y “etiquetas”
7. Divide los datos en conjunto de entrenamiento y pruebas
8. Calcula la regresión lineal
9. Evalúa el modelo
10. Exporta el archivo en Excel

**NOTA:** Genera un archivo .txt (block de notas) o .docx (Word) con los scripts. Puedes enviar los archivos anteriores en un archivo comprimido en un zip (no .rar).

*“Cree en ti mismo y en lo que eres”*

