



Análisis de Datos con Python : Trabajo

Se puede elegir entre 2 tipos de trabajo:

Trabajo 1: Hacer un programa de sistematización para análisis de datos. El programa debe permitir, mediante la terminal, ingresar datos, guardarlos en una base de datos en excel, y permitir mediante opciones que el usuario pueda hacer estas acciones :

- Ingresar datos a la base de datos y que se actualicen.
- Filtrar datos por medio de características.
- Solicitar gráficos de análisis entre algunas variables.
- Solicitar datos estadísticos de los datos.
- Generar nuevas características para los datos. (Nuevas columnas)

Se pueden usar las siguientes librerías para su ejecución: Numpy, Pandas, Matplotlib, Seaborn y Plotly.

Características del informe

La estructura del informe es libre, pero es recomendable tener :

- Carátula
- Introducción
- Análisis del sistema
- Ejemplos al correr el programa
- Conclusiones.

Trabajo 2: Se debe presentar un informe estadístico y un modelamiento de datos que se pueda hacer respecto a un dataset escogido de Kaggle (<https://www.kaggle.com/datasets>).

Recomendable elegir datasets acerca de regresión y análisis de datos. No es recomendable elegir datasets que tengan que ver con series de tiempo (Stock prediction, o NLP). Se solicita:

- Análisis estadístico con seaborn, pandas, matplotlib o plotly.
- Gráficas estadísticas y analíticas con matplotlib o Plotly.
- Modelo de regresión lineal analizando los conjuntos de entrenamiento y prueba y el sobreajuste.

Características del informe

La estructura del informe es libre, pero es recomendable tener :

- Carátula.
- Introducción.
- Análisis del problema.
- Análisis de datos.
- Conclusiones.

Se debe presentar para la revisión:

Link de un repositorio en github de un participante del grupo que contenga:

- Informe descriptivo del proyecto
- Script del programa o scripts en el caso de ser varios para el caso del trabajo tipo 1.
- Blog de notas de google Colab para el caso del Trabajo tipo 2.