

MCD407  
Ingeniería de datos  
**Proyecto de fin de curso**

5 de diciembre de 2024

M.Sc. Leydi Cruz

---

## **Datos Climáticos con ruido y Sensor Descalibrado**

Se cuenta con 2 datasets que contienen datos climáticos recopilados durante las últimas dos gestiones. En el dataset principal el sensor de humedad ambiente presenta lecturas con error a partir de cierta fecha; se dispone de un segundo dataset el cual contiene datos de temperatura y humedad de la misma zona, use este dataset como referencia para:

1. (a) Descubrir a partir de que fecha el sensor de humedad del dataset principal empezó a fallar en su lectura. **(5 Puntos)**
- (b) Corriga los datos de la humedad ambiente del dataset principal con los datos del dataset secundario. En caso de no tener información del dataset secundario, utilice algún criterio para corregir los valores leídos. **(20 Puntos)**
- (c) Limpiar el dataset principal, filtrar los datos ruidosos. **(30 Puntos)**
- (d) Una vez garantizada la limpieza y confiabilidad de la información del dataset principal generar estadísticas diaras de un mes (Seleccione un mes para trabajar). Genere un archivo excel que tenga información por día de: cuál fue la temperatura máxima, mínima y media día? Cuánto fue la precipitación pluvial acumulada diaria? Cuanto fue la presión atmosférica media diaria? Cual fue la velocidad máxima del viento día? Cual fue la velocidad media del viento día?. **(15 Puntos)**  
Por último cuál fue la temperatura máxima, mínima y media de ese mes? Cuánto fue la precipitación pluvial acumulada mensual? Cuánto fue la presión atmosférica media mensual?Cuál fue la velocidad máxima del viento mensual? Cual fue la velocidad media del viento mensual?. **(5 Puntos)**
- (e) Una vez garantizada la limpieza y confiabilidad de la información del dataset principal selecciona el mes de mayor cantidad de precipitaciones pluviales, rea-

liza el siguiente análisis: que relacion existe entre la precipitación pluvial con las demás variables climáticas?.. (25 puntos)

## Rúbrica de Evaluación para el Proyecto de Fin de Curso

## Rúbrica de Evaluación para el Proyecto de Fin de Curso

En el desarrollo del proyecto se espera que se cubran los siguientes puntos:

Criterio	Descripción
<b>Definición del Problema</b>	Claridad en la formulación del problema, objetivos bien definidos y descripción adecuada de los datos proporcionados.
<b>Preprocesamiento</b>	Limpieza de los datos: manejo de valores faltantes, detección/corrección de outliers y validación del dataset.
<b>Metodología</b>	Aplicación adecuada de técnicas analíticas y uso de herramientas relevantes (NumPy, Pandas, etc.). Estructura clara en el flujo de trabajo.
<b>Análisis Estadístico</b>	Cálculo y presentación de estadísticas descriptivas, correlaciones y otros análisis relevantes al problema planteado.
<b>Visualización</b>	Uso de gráficos claros y efectivos (Matplotlib, Seaborn, etc.) para respaldar el análisis y comunicar hallazgos.
<b>Conclusiones e Insights</b>	Resumen de los hallazgos, interpretación de resultados y conexión con el problema original. Incluye respuesta clara a las preguntas planteadas.
<b>Presentación</b>	Claridad y organización del reporte o presentación final. Código bien comentado y estructurado.