

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

4 de diciembre de 2024

M.Sc. Leydi Cruz

---

## Limpieza de datos Climáticos con ruido

Se cuenta con 1 dataset que contiene datos climáticos recopilados durante las últimas dos gestiones. Durante la adquisición de datos se presentaron algunos errores, generando que la base de datos se llene con datos que no corresponden a lecturas válidas. Use este dataset para:

1. (a) Limpiar, filtrar y corregir bajo algún criterio los datos ruidosos. **(45 Puntos)**
  - (b) Una vez garantizada la limpieza y confiabilidad de la información del dataset generar estadísticas diarias 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 por día? Cuánto fue la precipitación pluvial acumulada diaria? Cuánto fue la presión atmosférica media diaria? Cuál 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)**
  - (c) Una vez garantizada la limpieza y confiabilidad de la información del dataset principal selecciona el mes de mayor cantidad de precipitaciones pluviales, realiza el siguiente análisis: que relación existe entre la precipitación pluvial con las demás variables climáticas? **(35 Puntos)**

## 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.