

**Por favor lea
y entienda el
enunciado
antes de
comenzar su
ejecución.**

Universidad Pontificia Bolivariana
Facultad Ingeniería en Energía, Computación y TIC
Programas de Ingeniería en Sistemas e Informática – Ingeniería en Ciencia de Datos
Curso de Inteligencia Artificial
NRC: 30188, 31393 – Periodo 202520

Simulacro Examen No. 2
Análisis exploratorio de datos – Limpieza y tratamiento de datos
Precipitaciones de lluvia semestrales en el aeropuerto José María Córdova

Tiempo estimado de ejecución: 90 minutos

Tipo de Evaluación:

- **Individual:** Cada estudiante es responsable de su entrega y por lo tanto su calificación refleja su aporte.
- El examen **aplica solo** para los grupos que están cursando con el docente Juan Darío Rodas.

Nota importante:

Por favor responda a la confianza que se le brinda con este examen, trabajando honestamente. Si tiene dificultades para cumplir con el proyecto, por favor realice una valoración de la situación para que trabajemos un plan de mejoramiento específico a sus necesidades. **¡No haga trampa!**

Consideraciones para la entrega de la solución

Método de entrega

- La entrega de la solución del examen se realizará a través de repositorio **privado** creado en la plataforma de GitHub. El repositorio debe llamarse, sin excepciones, **iaupb_simulacro_02**.
- Los estudiantes deben incluir como colaborador del repositorio al docente (usuario **JuanDRodasM**).
- El estudiante debe notificar la entrega vía correo electrónico al correo **juand.rodasm@upb.edu.co**. Este email funciona como evidencia de la participación en la actividad evaluativa.

Contenido de la entrega

Los siguientes documentos deben incluirse en la entrega registrada en el repositorio:

- Documento con extensión “.ipynb”, correspondiente a la solución implementada en lenguaje Python en un Jupyter Notebook. El archivo debe llamarse **“iaupb_solucion_simulacro_02.ipynb”**.
- Archivo “.gitignore” correspondiente al tipo de contenido que se almacenará en el repositorio. En este caso, se requiere usar la plantilla de Python.
- Archivo “README.md” resumiendo el contenido del repositorio y la identificación de los estudiantes que hacen parte de la entrega.

Utilización de herramientas de Inteligencia Artificial

No hay restricciones para la utilización de herramientas de inteligencia artificial. **Por favor úsela a su discreción para resolver los retos del problema.**

Descripción del problema

Eres un científico de datos trabajando para el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). Te han asignado analizar los datos de precipitaciones del Aeropuerto José María Córdova en Medellín durante el primer semestre de 2025. Los datos provienen de sensores automáticos que registran mediciones cada vez que detectan precipitación, excluyendo valores de 0 mm.

Los datos que te entregan están en el archivo “precipitacion_2025sem01_JoseMariaCordova.csv”

En una inspección rápida encuentras:

- **Registros:** 1,600 mediciones
- **Período:** enero - junio 2025
- **Variables:** FechaObservacion, ValorObservado (mm)
- **Características:** Sin valores cero, posibles duplicados, formato mixto de fechas
- **Frecuencia de medición:** cada dos minutos

Adicionalmente:

- Encuentras gaps o brechas en las mediciones, las cuales corresponden a los momentos en los cuales no se registraron lluvias (valorObservado = 0 mm)
- Encuentras outliers muy elevados pero que **no corresponden a errores de medición**. Estos datos deben considerarse como mediciones reales en momentos donde las lluvias fueron intensas.

Acciones a realizar

Etapla 1: Lectura y procesamiento inicial- 10%

1. Cargar el archivo de datos en un *dataframe*.
2. Resolver el problema de formatos mixtos de las fechas.
3. Eliminar duplicados temporales (misma fecha/hora y valor)

Resultados esperados:

- Total registros resultantes luego del procesamiento inicial
- Total registros duplicados temporales eliminados
- Mediciones mínima, máxima y más frecuente (moda)

Etapla 2: Identificar máxima temporada seca. 20%

Identifique cual es el gap más largo sin mediciones. Este valor corresponderá **a la temporada seca más larga durante el semestre.**

Etapla 3: Identificar máxima duración de un evento de lluvia. 20%

Identifique cuál es la duración máxima de un **evento de lluvia** según el conjunto de datos suministrado. Un evento de lluvia es aquel que desde el comienzo hasta el fin, hubo mediciones consecutivas cada dos minutos.

Por ejemplo, el 4 de febrero de 2025 entre las 17:40 y las 18:00 hubo mediciones consecutivas cada dos minutos. Ese evento de lluvia tuvo una duración de 20 minutos.

Etapla 4: Totalización por hora y día para hallar las máximas precipitaciones. 30%

- Totalice todas las mediciones de una fecha y hora para **hallar el día y la hora en que más llovió**. Por ejemplo, para el 4 de febrero de 2025 a las 5:00 pm, sume todas las mediciones que van entre las 5:00 pm y las 5:59 pm.
- Totalice todas las mediciones de una fecha para **hallar el día en que más llovió**. Por ejemplo, para el 4 de febrero de 2025, sume todas las mediciones que van entre las 0:00 am y las 11:59 pm.

Etapla 5: Visualizaciones- 20%

Utilizando matplotlib, realice las siguientes visualizaciones como gráficos independientes, es decir, sin subdividir la zona de graficación en secciones. **Todos los gráficos deben tener títulos descriptivos y ejes etiquetados.**

1. Gráfico que permita entender cuáles son los valores más frecuentes de las mediciones.
2. Gráfico que permita entender cuales son las duraciones más frecuentes de los eventos de lluvia.

**Este examen
SI se hace
en una hora.**

Si no logras realizar ninguna de las actividades propuestas, [Hablemos por favor.](#)