

FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS
PROGRAMA ACADÉMICO DE INGENIERÍA DE DATOS E INTELIGENCIA
ARTIFICIAL
ASIGNATURA: PROGRAMACIÓN (G02)

PROYECTO-3: ANALÍTICA DE DATOS CON PYTHON

Objetivos

- Familiarizar al estudiante con algunas de las fases fundamentales de un proyecto de analítica de datos: recolección de datos, formulación de preguntas de interés, análisis exploratorio básico y limpieza.
- Aplicar Python y algunas de sus librerías (Pandas, Matplotlib, Seaborn, Numpy, etc.) en el desarrollo de un proyecto básico de analítica de datos.

Metodología

- El proyecto debe ser desarrollado en grupos de hasta cuatro personas (mínimo dos personas).
- Cada grupo debe seleccionar una problemática de interés y comunicar su elección enviando un correo electrónico. Recuerde incluir también los nombres de los integrantes del grupo.
- Con base en la problemática seleccionada, el grupo debe realizar las siguientes actividades:
 1. Recolección de datos: Buscar, seleccionar y descargar un conjunto de datos adecuado para la problemática seleccionada.
 2. Contextualización: Comprender el contexto del conjunto de datos y documentar el significado de cada una de las columnas.
 3. Análisis exploratorio de datos (EDA) básico:
 - a. Realizar una descripción general del conjunto de datos.
 - b. Ejecutar análisis univariado sobre las variables más relevantes.
 - c. Realizar análisis bivariado si se considera pertinente para la problemática abordada.
 - d. Detectar posibles problemas en los datos (nulos, inconsistencias, etc.) y proponer soluciones. Implementar una estrategia de limpieza en las columnas que se usarán.
 4. Preguntas de análisis: Plantear **mínimo 8 y máximo 10** preguntas de interés que puedan ser contestadas a partir de los datos. Para dar solución a estas preguntas se deben usar visualizaciones y al menos en alguna de ellas se debe utilizar groupby().
 5. Notebook: Todo el desarrollo del proyecto debe realizarse en un solo notebook debidamente comentado. También pueden usar los campos de texto para incluir descripciones más detalladas, el análisis que realicen de las gráficas generadas o lo que ustedes consideren necesario.

Entregables

- 1 (40%). Carpeta que contenga: el notebook con código en Python debidamente comentado (archivo .ipynb) y el conjunto de datos.
- 2 (60%). Sustentación: *jueves 29 de mayo en el horario de clase.

Problemáticas

1. Salud Pública

- Descripción: Analizar estadísticas de enfermedades, factores de riesgo, datos de vacunación, o hábitos de actividad física, etc.
- Ideas: Estudiar tendencias de enfermedades, impacto de campañas de vacunación, análisis de calidad de vida por región, etc.

2. Educación

- Descripción: Explorar datos sobre el rendimiento académico, factores socioeconómicos, encuestas de satisfacción, deserción escolar, etc.
- Ideas: Predicción del rendimiento académico, análisis de factores que impactan la deserción escolar, estudio de la efectividad de programas educativos, etc.

3. E-commerce y Retail

- Descripción: Análisis de ventas, patrones de compra de clientes, análisis de carrito de compras, satisfacción del cliente, etc.
- Ideas: Análisis de ventas estacionales, segmentación de clientes, identificación de productos populares y tendencias de compra, etc.

4. Medio Ambiente

- Descripción: Análisis de datos meteorológicos, calidad del aire, deforestación, niveles de contaminación, etc.
- Ideas: Estudio de patrones de temperatura y clima, análisis de calidad del aire en ciudades, predicción de áreas de deforestación, etc.

5. Transporte y Movilidad

- Descripción: Datos de tráfico, uso de transporte público, accidentes de tráfico, tiempos de viaje, etc.

- Ideas: Análisis de congestión en horarios pico, identificación de rutas de alto riesgo, análisis de uso de transporte público, etc.

6. Deporte y Fitness

- Descripción: Datos de rendimiento deportivo, actividad física, seguimiento de rutinas de ejercicio, etc.
- Ideas: Análisis de progreso en rutinas, predicción de rendimiento en deportes, estudio de tendencias de actividad física en distintos grupos de edad, etc.

7. Finanzas Personales

- Descripción: Datos de gastos personales, ahorros, inversión, patrones de consumo, etc.
- Ideas: Análisis de gastos mensuales, patrones de ahorro, predicción de gastos futuros y estrategias de ahorro, etc.

8. Mercado Laboral

- Descripción: Análisis de tendencias de empleo, salarios, competencias, satisfacción laboral, etc.
- Ideas: Predicción de demanda en profesiones específicas, análisis de salario promedio por sector, estudio de habilidades más demandadas, etc.

9. Redes Sociales

- Descripción: Análisis de interacciones, hashtags, contenido popular y tendencias de usuario, etc.
- Ideas: Identificación de temas populares, análisis de sentimientos en publicaciones, estudio de patrones de interacción en horarios específicos, etc.

10. Bienes Raíces

- Descripción: Datos sobre precios de inmuebles, ubicaciones, tamaño de propiedades, características de vecindarios, etc.
- Ideas: Predicción de precios inmobiliarios, análisis de zonas de alto crecimiento, estudios de factores que influyen en el valor de una propiedad, etc.

11. Agricultura y Alimentación

- Descripción: Datos de producción agrícola, consumo de alimentos, precios, distribución, etc.

- Ideas: Análisis de producción por región, estudio de precios de alimentos, predicción de demanda de productos agrícolas, etc.

12.Turismo y Viajes

- Descripción: Análisis de tendencias de destinos turísticos, ocupación hotelera, satisfacción del cliente, etc.
- Ideas: Predicción de temporadas altas, análisis de satisfacción de visitantes, identificación de destinos en crecimiento, etc.

13.Ciencia y Tecnología

- Descripción: Datos sobre patentes, publicaciones científicas, avances tecnológicos y participación en investigación, etc.
- Ideas: Análisis de tendencias de investigación, mapeo de tecnologías emergentes, estudio de colaboraciones entre instituciones, etc.

14.Energía

- Descripción: Datos sobre consumo de energía, generación de energía y demanda de energía, así como diferentes tipos de fuentes de energía utilizadas, , etc.
- Ideas: análisis de consumo de energía, predicción de la demanda, análisis de uso de combustibles fósiles y fuentes renovables, etc.

15.Otro tema

- Pueden escoger otro tema que les resulte de interés.