****Análisis de Datos con Herramientas Avanzadas****

****Objetivo:**** Seleccionar uno de los 5 posibles datasets propuestos, responder una serie de preguntas específicas utilizando código, análisis de datos, visualización gráfica y técnicas avanzadas como regresión lineal y K-Means. Puedes usar IA generativa para resolver tus dudas.

****Instrucciones:****

1. ****Selección del Dataset:**** Elige uno de los siguientes datasets:
   * Dataset 1: House Prices: Advanced Regression Techniques - Contiene información detallada sobre casas y sus precios de venta.
   * Dataset 2: Auto MPG Dataset - Incluye datos sobre vehículos y su eficiencia de combustible, adecuado para predecir el consumo de gasolina.
   * Dataset 3: eCommerce Customer Behavior Dataset - Incluye información sobre el comportamiento de compra de clientes en línea.
   * Dataset 4: FIFA 19 Players Dataset - Proporciona información sobre jugadores de fútbol y sus atributos, ideal para estudiar correlaciones entre habilidades y rendimiento.
   * Dataset 5: Global Temperature Dataset - Contiene datos sobre temperaturas globales a lo largo del tiempo, perfecto para analizar la dispersión de los datos.
2. ****Preguntas a Responder:**** En función del dataset elegido, resuelve las siguientes preguntas:
   * ****Dataset 1: House Prices: Advanced Regression Techniques****
     + ¿Cuáles son las características más importantes que determinan el precio de venta de una casa?
     + ¿Cómo afecta el tamaño de la vivienda (superficie habitable) al precio de venta?
     + ¿Cómo se pueden agrupar las casas en diferentes clusters basados en características como tamaño, número de habitaciones, y antigüedad?
     + ¿Qué características comunes tienen las casas en cada cluster?
     + ¿Cuál es la relación entre los clusters formados y los precios de venta de las casas? ¿Hay clusters con precios significativamente más altos o más bajos?
   * ****Dataset 2: Auto MPG Dataset****
     + ¿Cuáles son las características más importantes que afectan la eficiencia de combustible (MPG) de los vehículos?
     + ¿Cómo se pueden agrupar los vehículos en diferentes clusters basados en características como tamaño del motor, peso y eficiencia de combustible?
     + ¿Qué impacto tienen el tamaño del motor (displacement) y la potencia (horsepower) en la eficiencia de combustible?
     + ¿Existen patrones de eficiencia de combustible en los diferentes clusters de vehículos formados?
     + ¿Cuál es la relación entre los clusters formados y la eficiencia de combustible de los vehículos? ¿Hay clusters con mayor o menor eficiencia de combustible?
   * ****Dataset 3: eCommerce Customer Behavior Dataset****
     + ¿Cuáles son las características más importantes que afectan el gasto total de un cliente en la plataforma de comercio electrónico?
     + ¿Cómo influye el número de visitas a la web en el gasto total de un cliente?
     + ¿Cómo se pueden segmentar los clientes en diferentes grupos basados en su comportamiento de compra, como la frecuencia de compras, el gasto total y el tiempo promedio de visita?
     + ¿Qué características comunes tienen los clientes en cada cluster?
     + ¿Cuál es la relación entre los clusters formados y el gasto total de los clientes? ¿Hay clusters con clientes que gastan significativamente más o menos?
   * ****Dataset 4: FIFA 19 Players Dataset****
     + ¿Cuáles son los atributos más importantes que determinan el valor de mercado de un jugador?
     + ¿Cómo influye la edad de un jugador en su valoración general (overall rating)?
     + ¿Cómo se pueden agrupar los jugadores en diferentes clusters basados en sus atributos como velocidad, dribbling, y defensa?
     + ¿Qué características comunes tienen los jugadores en cada cluster?
     + ¿Cuál es la relación entre los clusters formados y la valoración general de los jugadores? ¿Hay clusters con jugadores con valoraciones significativamente más altas o más bajas?
   * ****Dataset 5: Global Temperature Dataset****
     + ¿Cuáles son las tendencias a largo plazo en la temperatura global?
     + ¿Cómo ha cambiado la temperatura global promedio por década?
     + ¿Cómo se pueden agrupar diferentes periodos de tiempo basados en las variaciones de temperatura?
     + ¿Existen patrones o ciclos de temperatura en los diferentes clusters de años formados?
     + ¿Cuál es la relación entre los clusters formados y los eventos climáticos significativos (por ejemplo, El Niño, La Niña)?
3. ****Generación de Código:****
   * Escribe el código necesario para obtener las respuestas.
4. ****Visualización de Datos:****
   * Crea gráficas para respaldar tus respuestas. Esto puede incluir gráficos de líneas, barras, scatter plots, box plots, etc.
5. ****Herramientas Avanzadas:****
   * Realiza un análisis de datos sencillos (estadísticas descriptivas).
   * Implementa una regresión lineal para uno de los análisis.
   * Usa el algoritmo K-means para agrupar los datos en clusters.
6. ****Uso de IA Generativa:****
   * Puedes hacer uso de herramientas de IA generativa para resolver dudas específicas durante el desarrollo de la actividad.

****Reglas para el Uso de IA Generativa:****

1. ****Uso Permitido:****
   * Puedes utilizar herramientas de IA generativa para obtener ayuda con la sintaxis de programación, corrección de errores, y sugerencias para mejorar el código.
   * La IA puede ser usada para obtener explicaciones sobre conceptos teóricos y metodológicos.
   * Puedes pedir a la IA ejemplos de códigos aplicables a problemas específicos para orientar su solución.
2. ****Uso No Permitido:****
   * No debes copiar y pegar soluciones completas obtenidas de herramientas de IA sin comprender y sin hacer ninguna modificación.
   * No puedes utilizar la IA para obtener respuestas directas a las preguntas de análisis de datos sin intentar el análisis por su propia cuenta.
   * La IA no debe ser usada para obtener artículos o reportes completos que sustituyan su propio trabajo.
3. ****Transparencia y Reporte del Uso:****
   * Debes especificar claramente en tu reporte cuándo y cómo utilizaron la IA generativa.
   * Indica qué herramientas de IA utilizaron (por ejemplo, ChatGPT, GitHub Copilot, etc.) y en qué partes del proyecto.
   * Describe cómo la IA les ayudó a resolver problemas específicos o mejorar tu código.
   * Incluye fragmentos de las interacciones con la IA que fueron relevantes para su trabajo, destacando cómo los implementaron en el proyecto.

****Formato para Reportar el Uso de IA Generativa:****

En el reporte final que entregues, incluye una sección dedicada al uso de IA generativa siguiendo este formato:

1. ****Herramientas Utilizadas:****
   * Indica todas las herramientas de IA generativa que emplearon. Ejemplo: ChatGPT, GitHub Copilot.
2. ****Descripción del Uso:****
   * Describe brevemente cómo utlizaste la IA en las diferentes etapas del proyecto.
     + ****Ejemplo:****
       - "Use ChatGPT para obtener una explicación sobre cómo implementar K-means clustering en Python."
       - "GitHub Copilot fue usado para sugerir mejoras en la estructura de mi código."
3. ****Interacciones Relevantes:****
   * Inlcuye fragmentos de las conversaciones/interacciones con la IA que fueron clave para su análisis o código.
     + ****Ejemplo:****
       - Pregunta a ChatGPT: "¿Cómo puedo crear un scatter plot usando Matplotlib?"
       - Respuesta de ChatGPT: "Puedes usar el siguiente código para crear un scatter plot: ..."
       - Implementación en el Proyecto: "Tomé el código sugerido y lo adapté para mis necesidades específicas."
4. ****Impacto en el Proyecto:****
   * Explica cómo el uso de la IA mejoró tu proyecto.
     + ****Ejemplo:****
       - "La IA me ayudó a corregir un error en mi código que estaba afectando los resultados del análisis de regresión lineal."
       - "Gracias a las sugerencias de la IA, logré optimizar el código, lo cual redujo significativamente el tiempo de ejecución."

****Evaluación:****

* Elección correcta del dataset y justificación (10%)
* Precisión y corrección de las respuestas a las preguntas propuestas (25%)
* Calidad y claridad del código generado (20%)
* Efectividad y claridad de las visualizaciones de datos (20%)
* Correcta implementación de regresión lineal y K-means clustering (15%)
* Uso eficaz de la IA generativa para resolver dudas (10%)

****Entrega:****

* Entrega un reporte en formato PDF que incluya las respuestas a las preguntas, el código utilizado, gráficas generadas y conclusiones. Además, sube el Cuaderno de Jupyter con el código generado.

Buena suerte y disfruten del análisis de datos!