## ****Trabajo práctico de investigación y aplicación****

**Tema:** Visualización de datos con Matplotlib y Seaborn

### ****Parte 1 – Investigación****

Elabora un informe que responda las siguientes preguntas:

1. **Definición de Matplotlib y Seaborn**
   * ¿Qué son estas librerías?

Son librerrias del lenguaje deprogramacion Python que sirven para plotear gráficos de ámbito numérico.

La librería Matplotlib es de bajo nivel comparada a la de Seaborn, esta ultima está construida sobre Matplotlib y da acceso a una interfaz fácil de usar para los gráficos mas usados en estadística y ciencia de datos

* + ¿En qué se diferencian y complementan?

Seaborn esta basado en Matplotlib  
Seaborn ofrece una API con métodos y clases amigables para obtener visualizaciones con pocas líneas de código.

En Matplotlib se tiene que hacer de forma “manual”, por ejemplo uno puede usar la interfaz OOP de Matplotlib para obtener varias subplots en una figura. En los Axes uno tiene que, explícitamente pasar los ejes(x,y) como secuencias, establecer las etiquetas y posicionamiento.

Texto

El contenido generado por IA puede ser incorrecto.

En cambio en Seaborn se podría usar directamente la clase FacetGrid que recibe como argumento el dataset de pandas directamente, y en otro argumento el nombre de las columnas usar, que también serán las etiquetas de los ejes.

Interfaz de usuario gráfica, Texto

El contenido generado por IA puede ser incorrecto.

Si bien se ve que Seaborn es mas práctico, Matplotlib es muy útil cuando se quiere personalizar a detalle los gráficos requeridos.

* + Ejemplos de uso en análisis de datos.

Del tutorial de Seaborn:

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

Se ve una visualización del dataset de tips(propinas).  
Se han hecho sub graficas divididas en sexo.  
Aunque es una grafica 2D se pueden ver 2 variables cuantitativas: tip, total\_bill; y una categórica: smoker.

1. **Histogramas y su importancia**
   * ¿Qué es un histograma?

Es una representación grafica en forma de barras donde el eje horizontal es la variable a analizar y el eje vertical es la frecuencia. El eje horizontal esta segmentado.

* + ¿En qué situaciones es útil?

Sirve para obtener una primera vista de la distribución de una muestra de una variable cuantitativa

* + Ejemplos reales de aplicación en ciencia de datos.

En análisis de datos exploratorio es una de las primeras visualizaciones que se hace para ver la distribución de una variable y verificar si tiene “cola” que pueda sesgar futuros análisis.

1. **Importancia de la visualización de datos**
   * ¿Por qué es fundamental visualizar datos antes de analizarlos?

Porque dependiendo de la distribución de la variable a analizar se pueden usar diferentes métodos estadísticos por ejemplo: las pruebas de hipótesis paramétricas requieren que la muestra tenga una distribución normal, en cambio si la distribución tiene fuertes sesgos, es mejor usar pruebas no paramétricas.

* + Beneficios y posibles errores al no realizar una correcta visualización.

Cuando uno tiene un dataset este es una muestra, no es todo el universo. Al hacer un análisis inferencial se busca generalizar. Si la muestra contiene sesgos o valores atípicos esta generalización y/o predicción del comportamiento de la variable no será adecuada.

1. **Tipos de gráficos estadísticos**
   * Enumera y explica brevemente los gráficos más comunes (barras, líneas, dispersión, cajas, violín, etc.)
   * ¿En qué casos se recomienda cada uno?
2. **Exportar y guardar imágenes en Python**
   * Métodos para guardar gráficos creados con Matplotlib y Seaborn.

En Matplotlib se puede pyplot.savefig() si se usa la API procedural. Si en cambio se prefiere la orientada objetos se usa Figure.savefig()

* + Diferencias entre formatos (PNG, JPG, SVG, PDF).

PNG: Es tipo mapa de bits sin perdidas de calidad, también soporta transparencias

JPG: Es un formato comprimido, tiene perdidas de calidad y no soporta transparencias.

SVG: Tipo vectorial, escalabilidad infinita. Se recomienda graficas y geometrías simples

El trabajo debe incluir **imágenes de ejemplo** y referencias de las fuentes utilizadas.

### ****Parte 2 – Práctica****

Usando Python y las librerías **Matplotlib** y **Seaborn**, realiza los siguientes pasos:

1. **Carga de datos**
   * Utiliza el dataset: dataset\_ventas.csv.
2. **Visualización inicial**
   * Crea un **histograma** de una variable numérica relevante.
   * Explica qué información observas.
3. **Gráficos estadísticos variados**
   * Un gráfico de **líneas** para mostrar la evolución de un dato en el tiempo (si aplica).
   * Un gráfico de **barras** comparando categorías.
   * Un gráfico de **dispersión** para ver la relación entre dos variables.
   * Un gráfico de **caja** o **violín** para analizar distribución.
4. **Personalización de gráficos**
   * Cambia colores, títulos, etiquetas y leyendas.
5. **Exportación**
   * Guarda al menos dos gráficos en formato PNG y uno en PDF.
6. **Informe breve**
   * Explica en un párrafo qué conclusiones obtuviste de las visualizaciones.

**Criterios de evaluación:**

* Claridad y profundidad en la investigación.
* Correcto uso de Matplotlib y Seaborn.
* Creatividad en las visualizaciones.
* Calidad y formato del informe final.