

Objetivo

El presente trabajo tiene como finalidad la integración de las técnicas estudiadas durante la materia aplicadas en el análisis de datos.

Condiciones

1. El trabajo puede ser realizado de forma individual o grupal (hasta 3 personas).
2. Se debe respetar el formato de los entregables detallados a continuación.

Entregables

El trabajo final consta de dos entregables con sus formatos detallados correspondientes.

1. Un informe analítico-estadístico en formato Word o PDF con las siguientes características.
 - a. Carátula con nombre, apellido y DNI de los estudiantes.
 - b. Hojas de informe del análisis realizado.
 - c. *Características del documento:* Letra Arial 12, interlineado 1.5.
 2. Un notebook python en formato .ipynb o link del notebook de Colab que contenga el código que respalde las conclusiones del informe. La notebook debe tener tanto el código como así la explicación y comentario de lo que se realiza en esa celda.
-

Consigna

El trabajo consta de un informe analítico-estadístico sobre un conjunto de datos analizado mediante las técnicas y herramientas trabajadas a lo largo de la cursada. El informe debe ser un escrito articulado que describa estructural y estadísticamente la información contenida en el conjunto de datos.

Datos

La recomendación es trabajar con los datos de intervenciones de los centros de salud de la Provincia de Buenos Aires. El conjunto de datos es una unión de tablas a partir de las publicaciones del gobierno provincial en su [portal de datos abiertos](#). El archivo en sí pueden descargarlo [aquí](#).

Como alternativa, los estudiantes podrán seleccionar un conjunto de datos que contenga al menos 100 registros, 2 variables cuantitativas y 2 cualitativas para analizar.

Selección de variables para el análisis

Para trabajar con este dataset deberán formular una serie de preguntas que guíen el trabajo. Las preguntas deberán contemplar el uso de al menos cuatro variables (una o más cuantitativas y una o más cualitativas) que permitan realizar un análisis relevante de los datos brindados.

Guía de preguntas orientadoras

La siguiente guía presenta en forma de pregunta las cuestiones que deberán ser consideradas en el trabajo y que representan las etapas del análisis de datos tal como vimos en la cursada. **No deben responder una por una**, sino que pretenden servir de orientación de los pasos a seguir en el análisis de datos. Las respuestas a muchas de ellas deberán estar contenidas en el desarrollo tanto del notebook como del informe.

1. *Extracción: Sobre la estructura general de la información.*

- a. *¿Qué tipo de archivo es? ¿Cómo pudieron leer los datos?*
- b. *¿Cómo está compuesta la tabla? ¿Cuántas columnas tiene y filas tiene?*
- c. *¿Cuál es la unidad de análisis de la tabla?*

2. *Análisis exploratorio y limpieza: Sobre la estructura específica de la información.*

- a. *¿Qué cantidad de valores nulos tengo por columna?*
- b. *¿Qué porcentaje de valores nulos tiene cada columna? ¿Hay columnas con más del 10% de valores nulos?*
- c. *¿Qué tipo de dato tiene cada columna? ¿Tienen los tipos correctos según el tratamiento volcado?*
- d. *¿Cuáles son las columnas que se podría formatear el tipo de dato?*

3. *Análisis de datos: Sobre las variables numéricas del conjunto de datos (trabajar con las variables seleccionadas).*

- a. *¿Cuáles son las variables numéricas en el conjunto de datos?*
- b. *¿Cuál es la distribución de cada variable numérica? ¿Son simétricas o asimétricas?*

- c. ¿Existen valores atípicos o outliers en las variables numéricas?
- d. ¿Cuál es la media, mediana de cada variable numérica?
- e. ¿Hay alguna relación o correlación entre las variables numéricas? ¿Cómo se puede visualizar esta relación?

4. Análisis de datos: Sobre las variables categóricas del conjunto de datos (trabajar con las variables seleccionadas).

- a. ¿Cuáles son las variables categóricas en el conjunto de datos?
- b. ¿Cuántas categorías únicas hay en cada variable categórica?
- c. ¿Cuál es la frecuencia de cada categoría en las variables categóricas?
- d. ¿Hay alguna relación o dependencia entre las variables categóricas? ¿Cómo se puede explorar esta relación?

5. Análisis de datos: Sobre la visualización de datos (trabajar con las variables seleccionadas).

- a. ¿Cuáles son las variables más relevantes o interesantes para visualizar?
- b. ¿Qué tipo de gráficos o visualizaciones serían adecuadas para cada variable o combinación de variables?
- c. ¿Se pueden identificar patrones, tendencias o insights a través de la visualización de los datos?

6. Análisis de datos: Sobre el análisis comparativo.

- a. ¿Se pueden hacer comparaciones o análisis entre distintas subpoblaciones del conjunto de datos? (por ejemplo, comparar resultados entre diferentes grupos o categorías)
- b. ¿Hay diferencias significativas en alguna variable entre las subpoblaciones?