

Hoja de Trabajo 1. Análisis Exploratorio

INSTRUCCIONES:

Esta hoja de trabajo debe realizarse en parejas. Para esto es necesario que cada miembro de la pareja se inscriba en uno de los grupos de Canvas que se han conformado para esta actividad. Si no, se inscriben, no tendrán nota.

Para este laboratorio utilizarán el Jupyter Notebook titulado “03-Vida_facil.ipynb” que se encuentra en el Canvas. Este notebook consta de dos partes:

1. La primera parte consta de un tutorial con diferentes instrucciones que les serán útiles cuando tengan que hacer EDA y trabajos de limpieza de datos. En el tutorial se incluye el uso del paquete `Pandas_profiler`, que simplifica mucho los procesos mencionados.
2. La segunda parte consta de múltiples secciones que les solicitarán realizar diferentes acciones necesarias para explorar los datos y hacer limpieza de los mismos. En cada una de estas secciones ustedes deben proveer el código en Python para realizar las acciones solicitadas. La nota obtenida se basa en esta segunda parte.

DESCRIPCIÓN DEL CONJUNTO DE DATOS

El conjunto de datos que estarán utilizando es uno clásico que se utiliza para estudiantes iniciándose en la Ciencia/Minería de Datos. Son datos reales obtenidos de la tragedia del hundimiento del Titanic.

El conjunto de datos se encuentra en el archivo “titanic3.csv” y contiene datos de 1309 pasajeros, y 14 características (variables) que los describen.

Características:

Las descripciones de las características están dadas en el archivo “Variables_titanic3.pdf”, incluido en el Canvas

EJERCICIOS

1. **(10 puntos c/sección, Total 90 puntos)** Escriban el código necesario para realizar cada acción. Si es necesario, deben justificar el o los métodos utilizados.
2. **(10 puntos)** Generen otras seis preguntas que le parezcan interesantes porque les permitirán realizar otras exploraciones y respóndalas. No puede repetir ninguna de las instrucciones anteriores.

MATERIAL A ENTREGAR

- Vínculo del archivo de Google docs (se utilizará el historial para calificar) con el informe de análisis exploratorio que debería tener:
 - Enunciado de la pregunta que se está respondiendo.
 - Respuesta con su respectiva explicación
 - Gráfico, si aplica de acuerdo con la pregunta.
- Jupyter Notebook que utilizaron para responder las preguntas con el código utilizado
- NOTA. En vez del archivo de Google Docs, pueden escribir sus observaciones, respuestas en el mismo Jupyter Notebook, utilizando la opción de "Markdown" en celdas que usarán para texto.