



# Fundamentos de Big Data

## Lección 4: Introducción a Data Mining

# Introducción a Data Mining

## ACTIVIDAD LECCIÓN 4

### Objetivos

- | En el presente tema hemos hablado de algunos conceptos básicos del Data Mining, hemos hecho muchos gráficos, hemos creado DataFrames teniendo en cuenta que podamos seleccionar para algunas columnas parámetros concretos.
- | De modo que el objetivo de esta tarea, en este tema 4, será aprender lo mejor posible los aspectos del propio manual. Se pedirán algunas cosas más como actividad, lo cual será explicado a continuación, pero, en este caso, la clave es que se trate de analizar muy bien el manual.
- | También se tratará de fomentar la proactividad y el aprendizaje autodidacta, no obstante, dado que el tema es bastante denso no se pedirán muchas cosas más.

### Contenido correspondiente a Lección 4:

1. Introducción a Data Mining
  - 1.1. Competición Titanic
  - 1.2. Selección de información en DataFrames
  - 1.3. Obtención de información de los gráficos

## Actividad relacionada con la Lección 4:

### Primera parte de la actividad 4 (hasta 6 puntos)

El/la alumno/a deberá entregar el propio .ipynb del manual.

La idea es que se trate de entender lo mejor posible, se podrían preguntar las dudas respecto a esta primera parte.

*A continuación deja alguna celda en blanco en el propio .ipynb*

*Y continúa con la segunda parte de la actividad..*

*Responde a cada pregunta a continuación de la misma, y todas ellas en celdas del .ipynb diferentes (debes entregar todo en el mismo archivo .ipynb)*

*Pregunta 1*

*Respuesta 1*

*Pregunta 2*

*Respuesta 2*

*..*

### Segunda parte de la actividad 4: (hasta 4 puntos)

Pregunta -1-

Si estás trabajando con un set de datos, y necesitas ver rápidamente la información de una columna con un gráfico.

¿Qué librería usarías? ¿Por qué?

Si quieres puedes añadir un ejemplo. No es necesario, no obstante.

Pregunta -2-

¿Te parece difícil la forma en que creamos “nuevos DataFrames”? (seleccionando solo ciertas filas o columnas, bajo ciertas condiciones o premisas).

Nota: Sobre esto hemos hablado en el punto 5 del Manual

(Esta parte será puntuada solamente por el mero hecho de responder a la misma, independientemente de lo que se diga).

Se trata de yo saber lo que habéis avanzado.

Esta materia, o algo muy similar, probablemente se explique nuevamente en asignaturas posteriores. (En Machine Learning, quizá).

Pregunta -3-

Busca algo de información acerca de “Heatmap” y explica con tus palabras lo que consideres oportuno

Puedes añadir ejemplos o lo que quieras.

Pregunta -4-

Hemos usado diferentes tipos de gráficos.

¿Se entiende bien cuando usar un “violinplot” o un “factorplot” por ejemplo?

¿Serías capaz de emplearlo en otro set de datos?

(Esta parte de la actividad será puntuada solamente por el mero hecho de ser respondida, independientemente de lo que se diga).

Recuerda que existen más tipos de Gráficos, no obstante, y lo más importante es la “extracción de información” de los mismos.