

# **Enunciado Entrega 1 Laboratorio TIC I**

A continuación, se explica lo solicitado para la entrega 1 del proyecto de laboratorio de la asignatura Técnicas de Inteligencia Computacional I. El proyecto consiste en emular el desarrollo de un problema clásico de Ciencia de Datos.

Primero los estudiantes deben conformar grupos de dos personas (tres personas solo en casos excepcionales y con mayor exigencia en sus evaluaciones), quienes deben elegir una base de datos clásica de Ciencia de Datos para trabajar. Dicha base de datos debe ser de algún problema de clasificación. Una vez elegida la base de datos, deben realizar tres presentaciones durante el semestre. En dichas presentaciones expondrán de manera incremental la solución del problema de Ciencia de Datos elegido. Todas las presentaciones deben ser los códigos y resultados obtenidos en Google Colab.

En la primera presentación deben exponer el Google Colab del Análisis Descriptivo e Inferencial de su problema. Aquí deben utilizar a criterio los códigos vistos en clases, pero también pueden ejecutar códigos obtenidos desde Internet. De todas formas, el profesor le puede pedir explicar cada línea de código utilizada, por lo que es su responsabilidad entender a cabalidad los códigos utilizados.

Para el análisis descriptivo se recomienda seguir los pasos de un Análisis Exploratorio de Datos (EDA por siglas en inglés). Se dejan 6 links que pueden ayudar a guiar este trabajo, los links están ordenados según la utilidad que la coordinación considera estos tienen, sin embargo, se recomienda leer todos y utilizar lo mejor de cada uno. Otra recomendación es utilizar como referencia los EDAs disponibles en la red del mismo problema que trabajen.

1. <https://growingdata.com.au/a-guided-introduction-to-exploratory-data-analysis-eda-using-python/>
2. <https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/>
3. [https://miykael.github.io/blog/2022/advanced\\_eda/](https://miykael.github.io/blog/2022/advanced_eda/)
4. <https://towardsdatascience.com/13-key-code-blocks-for-eda-classification-task-94890622be57>
5. <https://medium.com/geekculture/10-automated-eda-libraries-at-one-place-ea5d4c162bbb>

La evaluación será 30% de la forma y 70% del fondo de la presentación. En el archivo Excel está el detalle de dicha evaluación. Los contenidos de fondo a evaluar en la primera presentación son los siguientes:

1. **Carga de datos, preprocesamiento, medidas estadísticas y visualizaciones:**
  - a. Introducción.
  - b. Cargar correctamente el conjunto de datos.
  - c. Explicar el problema, explicar las variables e identificar los tipos de datos.
  - d. Identificar valores atípicos, datos nulos u observaciones faltantes.
  - e. Realiza el preprocesamiento de los valores atípicos, nulos o faltantes.
  - f. Construye gráficos de caja, violín, barra, tortas, histogramas y de correlación según corresponda.
  - g. Obtiene medidas estadísticas según corresponda.
  - h. Obtiene conclusiones no triviales en base al análisis de los gráficos y los estadísticos generados.
2. **Análisis estadístico inferencial:**
  - a. Genera distintas hipótesis en base a la estadística descriptiva.
  - b. Comprueba las hipótesis planteadas utilizando estadística inferencial.