



INF- 316 Programación para la Analítica de Gestión

Tarea 1: Definición del problema a abordar

1.- Portada del informe: Debe detallar el nombre de los integrantes del equipo (entre 2 y 3), rut y mail de contacto más Nombre del proyecto a desarrollar.

2.- Leer documento Ciencia de Datos y Estudios Globales para que cada integrante del equipo, se haga una idea de qué se entiende por Ciencia de datos y sus aplicaciones.

3.- En base al documento leído, c/integrante debe escribir un párrafo de qué entiende por Ciencia de datos y cómo lo aplicaría en su vida profesional (debe identificar c/párrafo con el nombre de quién lo realizó. [El párrafo debe tener entre 400 a 500 palabras]

4.- Descripción del contexto en donde se piensa resolver el proyecto, por ejemplo en Salud, Educación, Negocios, Hobbies, ciencia, entre otros. Debe quedar bien en claro el problema o desafío que se quiere enfrentar en este proyecto. (máximo una hoja).

5.- Definición del problema/oportunidad de negocio – Antes de iniciar cualquier proyecto, se deben plantear preguntas estratégicas para definir el problema, entender los datos y establecer las metas del proyecto. Se debiera tener más menos claras las siguientes preguntas

- ✓ ¿Cuál es el objetivo del proyecto?
- ✓ ¿Qué problema de negocio se está tratando de resolver?
- ✓ ¿Qué decisiones se podrían tomar en base a los resultados de este análisis?

6.- Debe comenzar a buscar fuentes de datos (data set) y adjuntarlos con el informe de la tarea 1. Puede buscar en:

- ✓ <https://www.kaggle.com/> - Plataforma líder en ciencia de datos
- ✓ <https://archive.ics.uci.edu/> - Repositorio de datos de la Universidad de California
- ✓ <https://data.gov/> - Portal oficial de datos abiertos de gobierno de los EEUU
- ✓ <https://aws.amazon.com/es/opendata/?wwps-cards.sort-by=item.additionalFields.sortDate&wwps-cards.sort-order=desc> – Datos abiertos de AWS

7.- Una vez determinado el data set con el que van a trabajar deben responder las siguientes consultas.

- ✓ ¿Qué tipo de datos se tiene? ¿Son estructurados (tablas), semiestructurados (Json, XML) o no son estructurados (texto, imágenes)?
- ✓ ¿Cuál es el tamaño y calidad de los datos? ¿Están completos?, ¿Hay valores faltantes o con errores?