



Cátedra 1: Informe de Análisis de Datos

La cátedra 1 consistirá en un informe de análisis de datos. Se medirán los siguientes resultados de aprendizaje:

1. Implementar códigos en lenguaje de programación R para la resolución de problemas de análisis de datos.
2. Aplicar los principios y paquetes de *Tidyverse* en la resolución de problemas del área de la ciencia de datos.

El informe debe contener los siguientes elementos:

0. Definición del problema/objetivo de investigación.

1. **Introducción:** Descripción del conjunto de datos seleccionado y los objetivos del análisis.
2. **Preprocesamiento de Datos:** Descripción detallada de los pasos tomados para limpiar y preparar los datos para el análisis utilizando Tidyverse.
3. **Análisis Exploratorio de Datos:** Implementación de códigos en R para explorar los datos. Esto puede incluir visualizaciones de datos, medidas de tendencia central, etc.
4. **Modelado de Datos:** Aplicación de técnicas de modelado de datos y algoritmos de aprendizaje automático, si corresponde.
5. **Interpretación de Resultados:** Discusión de los resultados del análisis y su relevancia para los objetivos establecidos en la introducción.
6. **Código:** Incluir todo el código R utilizado en el análisis como un apéndice o integrado en el informe. El código debe ser legible y propiamente comentado.

Pauta de Evaluación

Criterio	Puntaje
----------	---------

1. Selecciona a lo menos 2 variables del tipo cuantitativa y cualitativa. Redacta el problema/objetivo en términos de una pregunta de investigación o una hipótesis que contemple la(s) variables(s) y un parámetro a estimar.	0-5
2. Claridad y coherencia de la introducción. Dimensión de los datos, descripción y clasificación de las variables, mención de datos perdidos o irregularidades en los datos si fuese el caso.	0-5
3. Profundidad y relevancia del análisis exploratorio de datos. Exploración realizada según cada tipo de variable.	0-15
4. Calidad y eficacia del preprocesamiento de datos. Describe de forma esquemática el proceso completo del análisis.	0-20
5. Adecuación y eficacia del modelado de datos. Describe en detalle el modelo estadístico o algoritmo empleado. Argumenta los motivos de la sección de estos últimos.	0-15
6. Calidad de la interpretación de los resultados. Interpreta los resultados en términos del modelo. Realiza una interpretación aplicada.	0-5
7. Calidad, eficiencia y organización del código R. El código empleado es estructurado en bloques. Cada bloque está bien programado. El código presenta funciones bien definidas. El código es subido a un repositorio de GitHub.	0-5

El puntaje total se calculará sumando los puntajes de cada criterio. El puntaje máximo posible es 70.

Fecha de entrega: sábado 22 de junio del 2024.

Formato: pdf o GitHub.

