

Módulo 1: Introducción a la Ciencia de los Datos (2 horas)

- ¿Qué es la Ciencia de los Datos?
- Aplicaciones y tendencias actuales.
- Introducción a R Console
- R Studio y R Studio Cloud
- Instalación de paquetes y configuración del entorno.

Módulo 2: Fundamentos de R (4 horas)

- Estructuras de datos en R: vectores, listas, matrices, data frames y factores.
- Importación y exportación de datos.
- Operaciones básicas y manipulación de datos.
- Introducción a la programación en R: funciones, control de flujo y bucles.

Módulo 3: Limpieza y Transformación de Datos (5 horas)

- Principios de datos ordenados.
- Uso de paquetes tidyverse: dplyr, tidyr y stringr.
- Manejo de datos faltantes.
- Creación de nuevas variables y filtrado de datos.

Módulo 4: Análisis Exploratorio de Datos (8 horas)

- Estadísticas descriptivas.
- Visualización de datos con ggplot2:
 - Gráficos de dispersión, barras, histogramas y boxplots.
- Identificación de patrones, tendencias y valores atípicos.

Módulo 5: Modelado Predictivo (8 horas)

- Introducción al aprendizaje supervisado:
 - Modelos de regresión
 - Regresión lineal simple
 - Regresión polinómica

- Regresión múltiple
- Árboles de regresión
- Otro ...
- Evaluación de modelos
- Modelos de clasificación
 - Modelos logística.
 - Árboles de clasificación
 - Modelos de MSV
- Validación de modelos: partición de datos, métricas de evaluación.
- Introducción al aprendizaje no supervisado:
 - Clústeres (K-means).

Módulo 6: Proyecto Final y Aplicaciones Avanzadas (3 horas)

- Publicación de *dashboard* con rpubs y/o github
- Integración de lo aprendido: desarrollo de un proyecto práctico.
- Aplicaciones en la toma de decisiones basada en datos.
- Perspectivas futuras y recursos adicionales para profundizar en R.

Recursos:

Computadora y conexión a internet

Correo electrónico (gmail de preferencia)

Conocimientos previos

- Conocimiento de estadística elemental
- Conocimiento de programación básica
- Uso de ofimática

