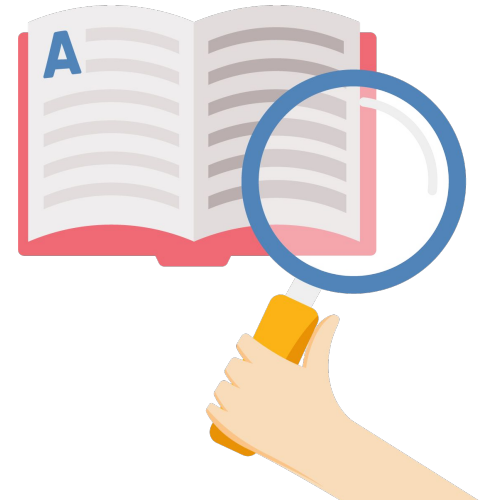




Glosario

Análisis de Datos con Lenguaje R

- **Lenguaje R:** es un lenguaje para el procesamiento y análisis estadístico de los datos. Al ser un software gratuito y de código abierto, se va ampliando por aportes de universidades, investigadores/as, usuarios/as y empresas privadas, organizados en librerías (o paquetes).
- **R Studio:** entorno para utilizar el lenguaje R.
- **Directorio o carpeta de trabajo:** lugar en nuestra computadora donde se encuentran los archivos con los que estamos trabajando. Aquí, el lenguaje R va a buscar los archivos para poder importarlos o exportarlos.
- **Operadores lógicos:** nos ayudarán a ver si un valor o conjunto de valores cumplen con determinadas condiciones.



- **Data Frame:** es la estructura fundamental para manipular un conjunto de datos en R. Se utiliza para guardar tablas de datos y se lo puede considerar como una lista de vectores de igual longitud que no tiene por qué ser del mismo tipo. Por ejemplo, puede haber datos numéricos y caracteres dentro del mismo Data Frame.
- **Data Wrangling:** proceso mediante el cual reestructuramos data perteneciente a distintos formatos para obtener un valor que resulte más conveniente para nuestros propósitos.
- **Data Management:** son prácticas que permiten organizar los datos de manera que podamos sacar un mejor provecho de ellos. Al estar ordenados de forma adecuada, podremos usarlos con mayor practicidad y facilidad.



- **Valores faltantes:** casilleros o registros que no contienen valores. Entre las causas más usuales tenemos: errores de programación, negativas a responder y respuestas no aplicables.
- **Tidyverse:** librería de paquetes disponibles en Lenguaje R orientados a manipular, importar, explorar y visualizar los datos que se utilizan en ciencia de datos.
- **Operador Pipe (Pipeline):** relaciona dos entidades, se usa en R con la simbología `%>%`.
- **Visualización de datos:** presentación de datos en formato gráfico. Nos permite detectar patrones, tendencias o correlaciones que de otro modo podrían pasar desapercibidos en tablas u hojas de cálculo.
- **Ggplot2:** librería utilizada para la visualización de datos.



- **Histogramas:** gráfico para ver las distribuciones univariadas de las variables. Se usa para aquellas variables que son numéricas continuas y nos sirven para detectar valores que se alejan del resto.
- **Scatter Plot:** o diagrama de dispersión. Sirve para graficar los valores de dos variables para un conjunto de datos.
- **Datos georreferenciados:** toda unidad que tiene asociada una referencia geográfica y es relevante o informativa, de tal modo que se puede representar la ubicación, el tamaño y la forma de un objeto en el planeta.
- **Librería Leaflet:** nos permite combinar múltiples capas, exportar mapas en formato .html y agregar tooltips o labels al interactuar con el mapa.



¡Ya estás listo!