

## Aprendizaje Estadístico.

Máster Interuniversitario en Tecnologías de Análisis de Datos Masivos: Big Data

Curso 2021/2022

Primera actividad de evaluación de prácticas

### Análisis de reviews sobre Amazon

Trustpilot.com es una página web que publica comentarios acerca de empresas. El fichero `amazon.tsv`<sup>1</sup> contiene información de comentarios en inglés sobre la empresa Amazon. Importa los datos en R utilizando la función `read_tsv`, a la que podrás acceder instalando la librería `tidyverse`. Contesta a las siguientes preguntas razonadamente utilizando R:

1. ¿Cuántos comentarios han sido registrados? ¿Cómo es la distribución del número de comentarios a lo largo de los años?
2. ¿Cómo resumirías la valoración de Amazon según los usuarios? ¿Cómo ha evolucionado la valoración media por año?
3. Representa y comenta la distribución del número de caracteres de los comentarios. (Puedes usar la función `nchar`).
4. Representa mediante un diagrama de cajas la distribución del número de caracteres de los comentarios en función de la valoración (De 1 a 5). ¿Qué observas?
5. Completa el análisis descriptivo de los datos con información que consideres relevante.

---

<sup>1</sup>El fichero `amazon.tsv` fue generado ejecutando el código que puedes consultar en el fichero `webscraping.R`. Dicho código es una modificación del código original obtenido en <https://www.datacamp.com>