

PEC1 Otoño 2025

UOC

En esta actividad no está permitido el uso de herramientas de inteligencia artificial. En el plan docente y en la [web sobre integridad académica y plagio de la UOC](#) encontraréis información sobre qué se considera conducta irregular en la evaluación y las consecuencias que puede tener.

Esta PEC se basará en los datos de las películas de Netflix estrenadas entre 1942 y 2019 (los datos se han obtenido de la web *ExcelDemy*). Hay incluida información de las siguientes variables:

1. *Name* = variable cualitativa que indica el título de la película
2. *Year* = año del estreno
3. *Age_Rating* = variable cualitativa que indica la clasificación de la película por edad
4. *Duration* = duración de la película en minutos
5. *Category* = variable cualitativa que indica la categoría de la película
6. *IMDb_Rating* = puntuación de la película (sobre 10)

Observación: las categorías de la variable *Age_Rating* son

1. *PG (Parental Guidance)*. Se sugiere la supervisión de los padres; algunos materiales podrían no ser aptos para niños pequeños.
2. *PG-13 (Parents Strongly Cautioned)*. Se advierte a los padres que algunos materiales podrían ser inapropiados para menores de 13 años.
3. *R (Restricted)*. Los menores de 17 años necesitan la compañía de un parent o tutor adulto para ver la película.

Para importar los datos podéis usar las siguientes instrucciones:

```
library(readxl)
datos <- read_excel("Netflix-Movies-Sample-Data.xlsx", skip = 5)
```

Os puede ser útil consultar el siguiente material del reto 1:

1. El entorno estadístico R. Estructura, lenguaje y sintaxis
2. Análisis de datos y estadística descriptiva con R
3. Actividades Resueltas del Reto 1 (Estadística Descriptiva)

Hay que entregar la práctica en formato “.pdf” en esta misma tarea.

NOMBRE:

PEC1

Una vez importados los datos...

Pregunta-1 (40%)

1.1 Ordenad la base de datos según el orden decreciente de la variable *IMDb_Rating* y mostrad solo las 3 primeras filas de esta base de datos ordenada. Dad el resumen numérico (mínimo, Q1, mediana, media, Q3 y máximo), la varianza y la desviación estándar de la variable *IMDb_Rating* (20%).

1.2 Dad el resumen numérico de la variable *IMDb_Rating*, pero solo cuando la variable *Age_Rating* vales *R*. Comentad los resultados obtenidos (20%).

Pregunta-2 (10%)

Dad el valor mínimo de la variable *IMDb_Rating* junto con las variables *Name* y *Category* donde se da este valor mínimo.

Pregunta-3 (30%)

Dad la tabla de frecuencias absolutas de la variable *Age_Rating*, y otra tabla con los porcentajes de los diferentes niveles de esta misma variable *Age_Rating* (podéis usar la instrucción *prop.table*). Haced el gráfico adecuado de las frecuencias o de los porcentajes. Comentad los resultados obtenidos.

Pregunta-4 (20%)

Haced los boxplots de la variable *Duration* estratificando por la variable *Age_Rating*. Comentad el resultado obtenido.