

PAC2 (Otoño 2025)

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.

Introducción

Consideramos la tabla de datos `jobs_in_data.csv` que proporciona información sobre salarios y trabajos en el ámbito de la Ciencia de Datos.

Contiene, entre otros, los siguientes campos:

work_year: el año en que se registraron los datos. Este campo indica el contexto temporal de los datos, importante para comprender las tendencias salariales a lo largo del tiempo.

job_title: el título específico del puesto de trabajo, como “Científico de datos”, “Ingeniero de datos” o “Analista de datos”, entre otros. Esta columna es crucial para comprender la distribución salarial entre diversos roles especializados dentro del campo de datos.

job_category: clasificación del puesto de trabajo en categorías más amplias para facilitar el análisis. Esto podría incluir áreas como “Análisis de datos”, “Aprendizaje automático”, “Ingeniería de datos”, etc.

salary: el salario bruto anual convertido a dólares estadounidenses (USD). Esta conversión uniforme ayuda en las comparaciones y análisis salariales globales.

Employee_residence: el país de residencia del empleado. Este dato se puede utilizar para explorar las diferencias salariales geográficas y las variaciones del coste de vida.

experience_level: clasifica el nivel de experiencia profesional del empleado. Las categorías comunes pueden incluir “Nivel de entrada”, “Nivel medio”, “Senior” y “Ejecutivo”, lo que proporciona información sobre cómo la experiencia influye en el salario en roles relacionados con datos.

employment_type: especifica el tipo de empleo, como “tiempo completo”, “tiempo parcial”, “contrato”, etc. Esto ayuda a analizar cómo los diferentes acuerdos laborales afectan las estructuras salariales.

work_setting: el entorno de trabajo, como “Remoto”, “Presencial” o “Híbrido”. Esta columna refleja el impacto de los entornos laborales en los niveles salariales en la industria de datos.

`company_location`: el país donde está ubicada la empresa. Ayuda a analizar cómo la ubicación de la empresa afecta las estructuras salariales.

`company_size`: el tamaño de la empresa empleadora, a menudo categorizada en tamaño pequeño (S), mediano (M) y grande (L). Esto permite analizar cómo el tamaño de la empresa influye en el salario.

Para importar los datos podemos usar la siguiente instrucción:

```
dades <- read.csv("jobs_in_data.csv")
n_total <- nrow(dades) # Número total de registros
```

Es necesario entregar la práctica en forma de archivo PDF (exportando el resultado final a PDF, por ejemplo) **únicamente** en esta misma tarea.

Indicar las fórmulas usadas del tipo $P(A|B)$, $P(A \cap B)$, etc.

Os puede ser útil la función `table` para tabular los datos.

Consultad las actividades resueltas de probabilidad del reto 2.

NOMBRE:

PAC2

Una vez importados los datos:

Problema 1 (30 puntos)

Trabajaremos con el entorno de trabajo (`work_setting`) y el nivel de experiencia (`experience_level`).

- Calculad la probabilidad de que un trabajador **trabaje en modalidad híbrida** (`work_setting == "Hybrid"`). *(15 puntos)*
- Calculad la probabilidad de que **un trabajador en modalidad híbrida** tenga un nivel de experiencia **Medio** (`experience_level == "Mid-level"`). *(15 puntos)*

Problema 2 (70 puntos)

Ahora consideraremos la relación entre **año de registro** (`work_year`) y **tipo de empleo** (`employment_type`).

- Construid la tabla de contingencia **número de trabajadores por año** (`work_year`) y **tipo de empleo** (`employment_type`). *(15 puntos)*
- Calculad la probabilidad de que **un trabajador sea del año 2023 y de tipo 'Full-time'**. *(10 puntos)*

- c) Calculad la probabilidad de que **un trabajador Freelance** esté registrado en el año **2023**. *(15 puntos)*
- d) Calculad la probabilidad de que **un trabajador registrado en 2022** tenga tipo de empleo **Contract**. *(15 puntos)*
- e) ¿Son **independientes** los eventos A : “ser Full-time” y B : “estar registrado en el año 2023”?
(15 puntos)