

# Creación de la visualización y entrega del proyecto (Práctica)

## 1. Presentación de la visualización:

Para la realización de la práctica, se utilizará el conjunto de datos “**Adult dataset**” del repositorio de datos <https://archive.ics.uci.edu/ml/datasets/Adult>.

Con estos datos crearemos diferentes visualizaciones basándonos en los ingresos de diferentes individuos utilizando el conjunto de atributos que forman este conjunto de datos.

Realizaremos una transformación sobre el conjunto de datos, para generar un nuevo juego de datos que nos permita generar una correcta visualización de datos para dar respuesta a las preguntas planteadas en la PRA1.

**Título:** Análisis y visualización del censo de ingresos de un individuo adulto en Estados Unidos.

**URL de la visualización:**

[https://public.tableau.com/app/profile/waziri.ajibola.lawal.mohammed/viz/PRA2\\_Income\\_Dataset/Story1](https://public.tableau.com/app/profile/waziri.ajibola.lawal.mohammed/viz/PRA2_Income_Dataset/Story1)

## 2. Explicación de la visualización

Las visualizaciones que se generarán con el conjunto de datos escogido permitirán a los usuarios en general y a los ciudadanos estadounidenses conocer o deducir qué factores o atributos hacen que el ingreso anual de un individuo adulto exceda o no los \$50.000 anuales. Cabe recordar que se considera un individuo adulto si es mayor de 16 años y menor de 100 años ((**edad**>16) && (**edad**>100)).

Para las visualizaciones que crearemos utilizaremos el gráfico de barras apiladas y estas visualizaciones, nos ayudarán a dar respuesta a las siguientes preguntas:

- Nivel de ingresos respecto la edad de los individuos.
- Nivel de ingreso respecto el sector en la que trabaja los individuos.

- Nivel de ingresos respecto el tipo de trabajo realizado por los individuos.
- Nivel de ingresos respecto el número de años de estudio de los individuos.
- Nivel de ingresos respecto el estado civil de los individuos.
- Nivel de ingreso respecto la raza de los individuos.
- Nivel de ingresos basándonos en el género de los individuos.

### 3. Descripción técnica del proyecto: lenguajes, librerías, licencias, código y datos.

Para la realización de la práctica se ha seguido varias de las etapas del ciclo de vida de los datos.

- **Captura:** como se ha indicado anteriormente se ha capturado datos ya existentes mediante la adquisición de los datos por medio de una descarga en un repositorio digital publicado en abierto.
- **Almacenamiento:** posteriormente se ha almacenado los datos un en un formato o representación adecuada que nos permitirá su utilización posterior de la manera más simple posible.
- **Preprocesado:** en esta fase se ha realizado una serie de técnicas para la preparación de los datos que nos permitirán realizar un análisis posterior:
  - o **Filtrado** de los datos para seleccionar los datos de interés
  - o **Conversión y transformación** de los datos original a un formato más plano y entendible (ej: **normalización, discretización y reducción de la dimensionalidad**).
  - o **Limpieza** de datos mediante la eliminación de datos inconsistentes (ej: **valores nulos**).
  - o **Análisis** de los datos mediante modelos que nos permitan explicar cómo son los datos y cuáles son sus características principales.
  - o **Visualización o representación** gráfica que nos permitirá la detección de patrones, tendencias y extracción un conocimiento mediante herramientas que permitan interactuar con los resultados de la fase de análisis.

Para las fases de filtrado, conversión, limpieza y análisis, se ha utilizado el lenguaje de programación R utilizando el entorno de desarrollo “RStudio”. Se ha escogido este lenguaje de programación

porque me resulta más sencillo para la transformación de los datos, ya que se ha utilizado anteriormente para este tipo de tareas.

Y para una visualización gráfica de los resultados, utilizaremos la herramienta “**Tableau**”. La razón por la que se ha escogido esta herramienta es porque nos ofrece todo lo necesario para realizar las visualizaciones que queremos.

#### 4. Visualización de datos

Como se ha indicado anteriormente, las visualizaciones generadas se pueden encontrar en el siguiente enlace:  
[https://public.tableau.com/app/profile/waziri.ajibola.lawal.mohammed/viz/PRA2\\_Income\\_Dataset/Story1](https://public.tableau.com/app/profile/waziri.ajibola.lawal.mohammed/viz/PRA2_Income_Dataset/Story1)

Se ha creado una narración con las visualizaciones generadas para responder a las preguntas planteadas anteriormente.