

## Desafío - Calculando probabilidades

- Para realizar este desafío debes haber estudiado previamente todo el material disponibilizado correspondiente a la unidad.
- Una vez terminado el desafío, comprime la carpeta que contiene el desarrollo de los requerimientos solicitados y sube el `.zip` en el LMS.
- Desarrollo desafío:
  - El desafío se debe desarrollar de manera Individual.
  - Para la realización del desafío necesitarás apoyarte del archivo *Apoyo Desafío - Calculando probabilidades*.

### Desafío 1: Lectura de archivos

- Importe `pandas` y `numpy` siguiendo las convenciones.
- Lea la base de datos `worldcup2014.csv` y asígnela a un objeto `df`.
- Solicite las primeras 5 observaciones con `head`.

### Desafío 2: Estime las frecuencias de `continent`

Utilizando `value_counts`. Responda lo siguiente:

- ¿Cuál es el continente con una mayor presencia en la muestra?
- ¿Cuál es la probabilidad de elegir un equipo asiático al azar?
- ¿Cuál es la probabilidad de elegir un equipo africano al azar?

### Desafío 3: Generación de Subconjunto de datos

Por cada continente, genere un nuevo objeto que almacene sólo las observaciones del continente.

**Tip:** Para ello puede utilizar la siguiente sintaxis: `df[df['variable'] == condicion]`.

### Desafío 4: Calcule la probabilidad de clasificación a la siguiente ronda

- Calcule la probabilidad de que un país pase a la siguiente ronda (clasifique) por continente.
- **Consejos:** El problema puede ser resuelto de varias formas:
  - La manera más intuitiva corresponde a calcular la probabilidad en cada subconjunto creado en el problema 3.
  - Una forma alternativa es copiar todos los subsets creados en el problema 3 en una lista e iterar sobre dicha lista.
  - Una forma más elegante consiste en utilizar una operación de agrupación por continente, y luego calcular la probabilidad.
- ¿Cuál es la probabilidad de que un país asiático pase a la siguiente ronda?
- ¿Cuáles son los dos continentes con la mayor probabilidad de clasificar?
- ¿Cuál es la probabilidad de que un país europeo no clasifique?

## Desafío 5: Probabilidades de Victoria y Derrota

- Se pide utilizar `value_counts` para calcular la probabilidad de que un continente contenga países con o sin victorias.
- Para ello cree una columna adicional llamada `'ha_ganado'` que contenga un 0 si es que un país no tiene juegos ganados y un 1 si es que ha ganado al menos un juego.
- **Sugerencias:** Para crear dicha columna puede revisar:
  - `np.where`, de Numpy
  - `pd.where`, de Pandas
  - Además puede utilizar los mismos procedimientos sugeridos en el Desafío 4.
- En base al resultado del procedimiento recién descrito, responda lo siguiente:
  - ¿Qué continente tiene la mayor probabilidad de tener países con al menos 1 victoria?
  - ¿Qué continente tiene un nivel similar entre países que tienen o no tienen victorias?
- Reformule el código para juegos perdidos y analice los continentes con mayor probabilidad de tener países sin juegos perdidos.