

Desafío - Calculando probabilidades

- Para realizar este desafío debes haber estudiado previamente todo el material disponibilizado en la sesión online correspondiente a la unidad.
- Crea una carpeta de trabajo y guarda todos los archivos correspondientes (notebook y csv).
- Una vez terminado el desafío, comprime la carpeta y sube el `.zip`

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`

Utilice `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: Probabilidades por continente

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

- Pase todos los objetos creados en un loop que imprima la probabilidad de pasar a la siguiente ronda por continente.

tip: Genere un array que contenga los objetos a pasar en el loop.

- ¿Cuál fue la probabilidad de que un país asiático pase a la siguiente ronda?
- ¿Cuáles fueron los dos continentes que tuvieron la mayor probabilidad de clasificar?
- ¿Cuál fue la probabilidad de no clasificar un país europeo?

Desafío 5: Refactorización

- A continuación se presenta un loop que cuenta las probabilidades de juegos ganados. Se pide que `value_counts` cuente los juegos no ganados y ganados, ignorando de la cantidad de juegos.

```
for d in [africa_df, europe_df, asia_df, northamerica_df, southamerica_df]:  
    print(d['juegos_ganados'].value_counts('%'))
```

```
1    0.6  
0    0.4  
Name: juegos_ganados, dtype: float64  
1    0.461538  
2    0.230769  
3    0.153846  
0    0.153846  
Name: juegos_ganados, dtype: float64  
0    1.0  
Name: juegos_ganados, dtype: float64  
2    0.50  
1    0.25  
0    0.25  
Name: juegos_ganados, dtype: float64  
2    0.500000  
3    0.333333  
1    0.166667  
Name: juegos_ganados, dtype: float64
```

- Para ello puede utilizar `np.where`, que devuelve un array con valores imputados cuando se satisface una condición. En este caso debemos clasificar como 0 cuando los juegos ganados sean igual a 0, y 1 cuando sean igual o mayor a 1.
- Este array es difícil de trabajar, por lo que pueden transformarlo con `pd.Series` para utilizar `value_counts`.
- En base a la refactorización del código, responda lo siguiente:
 - De acuerdo a la totalidad de juegos, ¿Qué continente tuvo una mayor probabilidad de ganar juegos?
 - ¿Qué continente presentó un nivel similar entre juegos ganados y perdidos?

- Utilize el código para el caso de juegos perdidos y analice el continente con una mayor probabilidad de perder.