

Desafío: Calculando probabilidades

- Para realizar este desafío debes haber estudiado previamente todo el material disponible 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 del desafío: Individual
- Para la realización del desafío necesitarás apoyarte del archivo Apoyo Desafío –Calculando probabilidades.
- Puntaje total: 10 puntos.

Habilidades a evaluar

- Utilizar conceptos básicos de probabilidad

Descripción

Un amigo, al cual le gusta mucho el fútbol, tiene en mente poder crear una app de apuestas para el próximo mundial. Para ello, solicita tu ayuda como analista de datos de manera de poder trabajar con las probabilidades de victoria o derrota de los equipos, y que así él diseñe el resto de la app. Tú aceptas y ordenas el trabajo que debes realizar a partir de los siguientes requerimientos.

Requerimientos

A continuación, revisaremos los requerimientos y acciones que la empresa a la cual postulas te pide realizar.

1. Lectura de archivos (1 pto):

- 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**

2. Frecuencia de continent (1 pto): utilizando **value_counts** responde 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?

3. **Generación de un subconjunto de datos (1 pto):** 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]`.

4. **Probabilidad de clasificación a la siguiente ronda (1 pto):** Calcule la probabilidad de que un país pase a la siguiente ronda (clasifique) por continente.

- ¿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?



Tip: 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.

5. **Probabilidad de victoria y derrota (1 pto):** 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:
 - a) ¿Qué continente tiene la mayor probabilidad de tener países con al menos 1 victoria?
 - b) ¿Qué continente tiene un nivel similar entre países que tienen o no tienen victorias?
 - c) Reformule el código para juegos perdidos y analice los continentes con mayor probabilidad de tener países sin juegos perdidos.