

Ejercicio 2

1 Estadísticas de los juegos olímpicos de Rio 2016

Medallero de los juegos olímpicos de 2016

2 Acerca del conjunto de datos

2.1 Contexto

Este conjunto de datos contiene estadísticas de las medallas logradas por los 11 mejores países que participaron en los juegos olímpicos de Río de Janeiro 2016.

index	ranking	country	gold_medal	silver_medal	bronze_medal	total
0	1	Estados Unidos (USA)	46	37	38	121
1	2	Reino Unido (GBR)	27	23	17	67
2	3	China (CHN)	26	18	26	70
3	4	Rusia (RUS)	19	18	19	56
4	5	Alemania (GER)	17	10	15	42
5	6	Japón (JPN)	12	8	21	41
6	7	Francia (FRA)	10	18	14	42
7	8	Corea del Sur (KOR)	9	3	9	21
8	9	Italia (ITA)	8	12	8	28
9	10	Australia (AUS)	8	11	10	29
10	13	Brasil (BRA)	7	6	6	19

Table 1:

3 Contenido

11 filas y 6 columnas.

La descripción de las columnas se enumera a continuación.

Variable	Descripción
ranking	posición
country	país
gold_medal	medalla de oro
silver_medal	medalla de plata
bronze_medal	medalla de bronce
total	total

Table 2:

3.1 Información del DataFrame

De la tabla 3 podemos observar que nuestras variables son *strings*, pero si observamos la tabla 1 a excepción de la columna **country** las demás parecen ser numéricas.

#	Column	Non-Null Count	Dtype
0	ranking	11 non-null	object
1	country	11 non-null	object
2	gold_medal	11 non-null	object
3	silver_medal	11 non-null	object
4	bronze_medal	11 non-null	object
5	total	11 non-null	object

Table 3:

4 Conclusión

Hemos visto la manera de obtener datos a través de WebScraping, pero para hacer estudios como un modelado estos datos no están listos para su estudio, por ejemplo, en los datos de la tabla 1 para hacer el preprocesado se tienen que transformar estas variables, **country** como categórica y el resto como numéricas.