

The Ultimate Halloween Candy Power Ranking

1. Enlace:

El dataset original usado en este curso puede ser encontrado en:

<https://www.kaggle.com/fivethirtyeight/the-ultimate-halloween-candy-power-ranking>

2. Descripción:

Este dataset publicado en 2017 contiene información de la composición de **86** diferentes golosinas y de la preferencia de varias personas que fueron encuestadas para descubrir cuál es la favorita entre los consumidores. En total se recolectó información de un total de 8.371 IPs diferentes y 269.000 votos.

Para calcular la información de la columna 'winpercent' que recoge el porcentaje de veces en las que una determinada golosina fue seleccionada como favorita, se crearon rondas donde competían dos golosinas elegidas al azar. Al final se cuenta el número de veces en que una golosina salió victoriosa en las rondas donde compitió y se obtiene el porcentaje.

3. Columnas

candy.csv tiene atributos categóricos y numéricos para cada una de las 86 golosinas. Para estas variables categóricas 1 significa presencia y 0 ausencia.

1. **chocolate**: ¿Contiene chocolate?
2. **fruity**: ¿Tiene algún saborizante frutal?
3. **caramel**: ¿Hay presencia de caramelo en la golosina?
4. **peanutalmondy**: ¿Contiene maní, mantequilla de maní o almendras?
5. **nougat**: ¿Tiene turrón?
6. **crispedricewafer**: ¿Contiene arroz inflado, wafers o galletas en sus componentes?
7. **hard**: ¿Es de textura dura?
8. **bar**: ¿Viene en forma de barra?
9. **pluribus**: ¿Es una golosina individual o son muchas en una bolsa o caja?
10. **sugarpercent**: El percentil de azúcar en el que recae dentro del mismo dataset.
11. **pricepercent**: El percentil de precio por unidad dentro del que se encuentra respecto al dataset.
12. **winpercent**: Porcentaje de victorias de acuerdo a 269.000 emparejamientos al azar.