

Primera entrega

Santiago Aillón, Juan José Caballero y Laura Ortiz

28 de octubre de 2022

1. Contenido de la base de datos

Para el proyecto utilizaremos la siguiente base de datos <https://www.kaggle.com/datasets/rtatman/chocolate-bar-ratings>, la cuál contiene los siguientes datos:

1. **Company (Maker-if known)**: Nombre de la empresa que fabrica el chocolate.
2. **Specific Bean Originor Bar Name**: La región geográfica específica, del origen de la barra.
3. **REF**: Un valor relacionado con el momento en que se ingresó la reseña en la base de datos. El más alto equivale al más reciente.
4. **ReviewDate**: Fecha de la publicación de la reseña.
5. **CocoaPercent**: Que es el porcentaje de cacao que tiene la barra de chocolate.
6. **CompanyLocation**: País base del fabricante del chocolate.
7. **Rating**: Calificación por un experto del chocolate.
8. **BeanType**: La variedad (tipo) de grano utilizado, si se proporciona.
9. **Broad BeanOrigin**: La región geográfica del origen del grano.

2. Análisis que se quieren realizar

Con la base de datos mencionada anteriormente, queremos lograr los siguientes objetivos:

- Analizar en que países se produce la mejor barra de chocolate (*teniendo en cuenta la clasificación*).
- Ver en donde se cultivan los mejores granos de cacao (*teniendo en cuenta la clasificación*).
- Realizar una comparación entre la relación entre el porcentaje de cacao y la calificación de la barra.
- Ver que tipo de grano suelen utilizar las compañías de los diferentes países.
- Comparar que países manufacturan mas barras sin generar tantos granos y que países generan más granos sin manufacturar tantas barras. Hallar la relación entre estas 2 variables para los países.
- Analizar la relación entre los países y sus respectivos niveles de Cacao en las barras (promedio).
- Analizar cuál tiene más influencia para una barra de calidad entre el país manufacturador de barras y el país productor de granos. (incluso se podría tener en cuenta la compañía que las crea).
- Encontrar que países importan más granos de que otros países (se puede seleccionar alguno en particular como Colombia, o bajo algún parámetro como el más productor y el menos productor).

De los anteriores análisis existe la posibilidad de que cambien, se quiten algunos, o se agreguen otros diferentes.