

## Informe Descriptivo Tarea 1 Data Mining: Beer Clustering.

En este trabajo se utilizara un dataset de reviews cervceras de beeradvocate, un popular portal de evaluacion de cervezas de todo el mundo. El dataset cuenta con 138488 entradas, de ls cuales 4605 tienen un dato faltante en una columna, quedando finalmente en 133883 entradas. Las variables que conforman el dataset son las siguientes:

brewery\_id: Variable cuantitativa discreta, define el id de la cervceria a la que pertenece la cerveza. NO utilizada en el calculo de clusters.

brewery\_name: Variable categórica nominal, define el nombre de la cervceria a la que pertenece la cerveza. NO utilizada en el calculo de clusters.

review\_time: Variable cuantitativa discreta, especifica el tiempo que tomo el review en cuestion. NO utilizada en el cálculo de clusters.

review\_profilename: Variable categorica nominal, define el nombre del usuario que hizo el review. NO utilizada en el calculo de clusters.

review\_overall: Variable cuantitativa continua con valores de 0 a 5. Evaluacion de la impresion general que causo la cerveza, considerando la expertiencia y placer sentidos. Pondera el 20 por ciento en el calculo final de puntaje.

review\_taste: Variable cuantitativa continua con valores de 0 a 5. Evaluacion del sabor, gusto posterior y otras características del gusto. Pondera el 40 por ciento del calculo final de puntaje.

review\_appearance: Variable cuantitativa continua con valores de 0 a 5. Evaluacion correspondiente al color, claridad, y retención del sabor de la cerveza. Pondera el 6 por ciento en el calculo final de puntaje.

review\_aroma: Variable cuantitativa continua con valores de 0 a 5. Evalua el olor de la cerveza, en terminos de alcohol y aroma de escencias. Pondera el 24 por ciento en el calculo final de puntaje.

review\_palate: Variable cuantitativa continua con valores de 0 a 5. Evalua la sensacion en la boca (paladar), describe el cuerpo de la cerveza, suavidad, etc. Pondera el 10 por ciento del puntaje final.

review\_beerstyle: Variable categorica nominal, describe el estilo de la cerveza. NO es utilizada en el calculo de clustering.

beer\_name: Variable categorica nominal, describe el nombre de la cerveza. NO es utilizada en el calculo de clustering.

beer\_beerid: Variable cuantitativa discreta, define el id de la cerveza en cuestion. NO es utilizada en el calculo de clustering.

Cabe destacar, que para efectos de clustering solo se utilizaran las columnas con variables cuantitativas, dado que la reducción de dimensionalidad requiere ser calculada numericamente.

## Conclusión

Se considera más importante el nombre de cervecería pues brewery\_id fue uno de los atributos usados como índice en el uso del dataset, y con este se obtuvieron los clusters esperados en la mayoría de los algoritmos.

**\*\*Se adjunta imagen de clustering dbscan, pues jupyter no arrojó el gráfico obtenido realmente.**