Beer\_Profile\_Kaggle

<https://www.kaggle.com/datasets/ruthgn/beer-profile-and-ratings-data-set?resource=download>

# Beer Profile and Ratings Data Set

Beer Dataset with Tasting Profiles and Consumer Ratings

## Beer Profile and Ratings Data Set

arrow\_drop\_up49

[Data Card](https://www.kaggle.com/datasets/ruthgn/beer-profile-and-ratings-data-set/data)[Code (7)](https://www.kaggle.com/datasets/ruthgn/beer-profile-and-ratings-data-set/code)[Discussion (0)](https://www.kaggle.com/datasets/ruthgn/beer-profile-and-ratings-data-set/discussion)[Suggestions (0)](https://www.kaggle.com/datasets/ruthgn/beer-profile-and-ratings-data-set/suggestions)

## About Dataset

### Data Set Information

This data set contains tasting profiles and consumer reviews for 3197 unique beers from 934 different breweries. It was created by integrating information from two existing data sets on Kaggle:

* [Beer Tasting Profiles Dataset](https://www.kaggle.com/stephenpolozoff/top-beer-information)
* [1.5 Million Beer Reviews](https://www.kaggle.com/rdoume/beerreviews)

The purpose of the data integration is to create **a new data set that contains comprehensive consumer review (appearance, aroma, palate, taste and overall review scores) for different brews, combined with their detailed tasting profiles**—this is that data set.

In the future it might be possible to go through and scrape more detailed information for each brew, such as the brewery's country of origin or brew recipe information such as its original gravity vs. final gravity.

### Contents

The main data set (beer\_profile\_and\_ratings.csv) contains the following columns:

1 - **Name**: Beer name (label)  
2 - **Style**: Beer Style  
3 - **Brewery**: Brewery name  
4 - **Beer Name (Full)**: Complete beer name (Brewery + Brew Name) -- unique identifier for each beer  
5 - **Description**: Notes on the beer if available  
6 - **ABV**: Alcohol content of beer (% by volume)  
7 - **Min IBU**: The minimum IBU value each beer can possess. IBU was not a value available for each beer, but the IBU range for each style was  
8 - **Max IBU**: The maximum IBU value each beer can possess. IBU was not a value available for each beer, but the IBU range for each style was

The next eleven columns represent the tasting profile features of the beer, and are defined by word counts found in up to 25 reviews of each beer. The assumption is that people writing reviews are more than likely describing what they do experience rather than what they do not. (Refer to the file Beer Descriptors Simplified to see the list of words that are used to calculate the values contained in each of the feature columns below)

(Mouthfeel)  
9 - **Astringency**  
10 - **Body**  
11 - **Alcohol**

(Taste)  
12 - **Bitter**  
13 - **Sweet**  
14 - **Sour**  
15 - **Salty**

(Flavor And Aroma)  
16 - **Fruits**  
17 - **Hoppy**  
18 - **Spices**  
19 - **Malty**

The last six columns contain information from beer reviews--they include the number (count) of consumer/user reviews, the average overall rating score, and the average rating scores for the aroma, appearance, palate, and taste of each individual beer.

20 - **review \_ aroma**  
21 - **review \_ appearance**  
22 - **review \_ palate**  
23 - **review \_ taste**  
24 - **review \_ overall**  
25 - **number \_ of \_ reviews**

The next two files (Brewery Name Fuzzy Match List.csv and Beer Name Fuzzy Match List.csv) only contain lists of breweries and beers that are found in both source datasets--and consequently, included in this dataset. To see the data integration process in more details, check out [this notebook](https://www.kaggle.com/ruthgn/data-cleaning-integration-pandas-fuzzywuzzy).

The last file (Beer Descriptors Simplified) contains list of words that are used to calculate the values contained in the tasting profile feature columns.

### Acknowledgements

Source: BeerAdvocate  
Credits:

* [Beer Tasting Profiles Dataset](https://www.kaggle.com/stephenpolozoff/top-beer-information) by [sp1222](https://www.kaggle.com/stephenpolozoff).
* [1.5 Million Beer Reviews](https://www.kaggle.com/rdoume/beerreviews) by Tanya Cashorali (uploaded by [Datadoume](https://www.kaggle.com/rdoume" \t "_blank)).

### Context

As previously mentioned, this data set was created to put existing consumer ratings data for different brews and their detailed tasting profiles in one place. Possible uses for the dataset include:

* Analyzing the properties that make a highly-rated beer
* Clustering and building a beer recommendation system based on similarities
* Classifying different beer styles based on tasting profile information
* Predicting a brew's alcohol content (ABV) using known characteristics
* A lot more!

This one is for all the beer lovers out there—cheers! 🍻