

# Answers from analysing beer review data

## 1. Rank top 3 Breweries which produce the strongest beers?

According to Wikipedia - The strength of beer is measured by its alcohol content by volume expressed as a percentage, that is to say, the number of millilitres of absolute alcohol (ethanol) in 100 mL of beer.

This parameter is called 'Alcohol by Volume' (ABV)

Strongest Beers by ABV are -

- Schorschbräu Schorschbock 57% (with ABV 57.70)
- Schorschbräu Schorschbock 43% (with ABV 43.00)
- Schorschbräu Schorschbock 40% (with ABV 39.44)

## 2. Which year did beers enjoy the highest ratings?

For this we need to extract 'year' information from 'review time' column.

After that we can group on 'beer name' and 'review year', then aggregate them on 'overall rating' with maximum rating

Some of highest rated beers with the year they rated highest are -

- Hop, Drop n Roll (2011)
- Cantillon Kriek 100% Lambic (2008)
- Slap & Tickle (2010)
- Hell For Certain (2004)
- Cantillon La Dernière Cuvée Du 899 (2008)

## 3. Based on the user's ratings which factors are important among taste, aroma, appearance, and palette?

Here we need to calculate correlation of 'review\_aroma', 'review\_appearance', 'review\_palette' and 'review\_taste' with 'review\_overall'

There is high correlation between Aroma and Overall review, after that Taste has high correlation.

## 4. If you were to recommend 3 beers to your friends based on this data which ones will you recommend?

To recommend 3 beers to my friends as I do not have experience and preference, I would first find most popular beers from data and then find highest rated beers among them.

This will increase chance that my recommendations will work for my friends. Also there is always a favour for famous things among public.

Top 3 recommendations from my side will be (with average rating) -

- Founders KBS (Kentucky Breakfast Stout) (4.4)
- Tröegs Nugget Nectar (4.39)
- Founders Breakfast Stout (4.3)

## 5. Which Beer style seems to be the favorite based on reviews written by users?

To find most favourite styles I collected all beer styles and maximum sentiment score from user rating and then selected beer styles with maximum sentiment score

Most Favourite Beer Styles with positive review sentiment only -

- Dubbel,
- Irish Red Ale,
- Belgian Strong Dark Ale,
- American Double / Imperial Stout,
- American Stout, Scotch Ale / Wee Heavy,
- Belgian Strong Pale Ale,
- Winter Warmer,
- American Amber / Red Lager

To find most favourite styles I collected all beer styles and maximum sentiment score from user rating and overall rating, then selected beer styles with maximum of both scores

Most Favourite Beer Styles from positive sentiment and overall rating -

- American Amber / Red Lager,
- American Double / Imperial Stout,
- American Stout,
- Belgian Strong Dark Ale,
- Belgian Strong Pale Ale,
- Dubbel,
- Irish Red Ale,
- Scotch Ale / Wee Heavy,
- Winter Warmer

## 6. How does written review compare to overall review score for the beer styles?

From analysis of reviews written we can say that Positive reviews tend to be lengthy, and negative reviews are shorter in number of sentences as well as words

## 7. How do find similar beer drinkers by using written reviews only?

- To find similar beer drinkers by only their reviews we can use -
  - Content based filtering

- Collaborative filtering
- Content based filters are used to make recommendations based on similar products/services according to their attributes
- Collaborative filtering uses algorithms to filter data from user reviews to make personalized recommendations for users with similar preferences. With this we can find out similar beer drinkers.
- Three types of collaborative filtering commonly used in recommendation systems are
  - neighbor-based
  - item-to-item and
  - classification- based.
  - In neighbor-based filtering, users are selected for their similarity to the active user. This similarity is determined by matching users who have posted similar reviews. Based on the previous similarity, it is presumed that future likes and dislikes will also be similar. From the average rating of the group, recommendations are made for the active user.
  - An item-to-item filtering process uses a matrix to determine the likeness of pairs of items. Item-to-item processes then compare the current user's preference to the items in the matrix for similarities upon which to base recommendations.
  - A classification-based collaborative filtering system recommends things based on how similar users liked that classification or genre. It is assumed that users that enjoy or dislike similar experiences within a classification will also enjoy others within that classification.
- I tried out content based filtering on sample of data to find out if we can get similar users if given a user
  - Steps involved are
    - Find TFIDF vectors of review text
    - then calculate cosine similarity between vectors
    - from this try to find out similar beer drinkers from their reviews
  - The function take input as profile name and returns similar users based on their cosine similarity
- This may not be a perfect method to find ot similar users using text data.
  - We can try out Collaboratiove filtering for doing so as described in articles like this - <https://medium.com/@medfordxie/what-to-drink-next-a-simple-beer-recommendation-system-using-collaborative-filtering-b65dd32b600d>