**Group 6 - Assignment 3**

**Names of group members:**

Faiza Quadri, Rebecca Lee, Yichen Pan, Zixi Lei, Briana Ramirez, Stephen Rhee

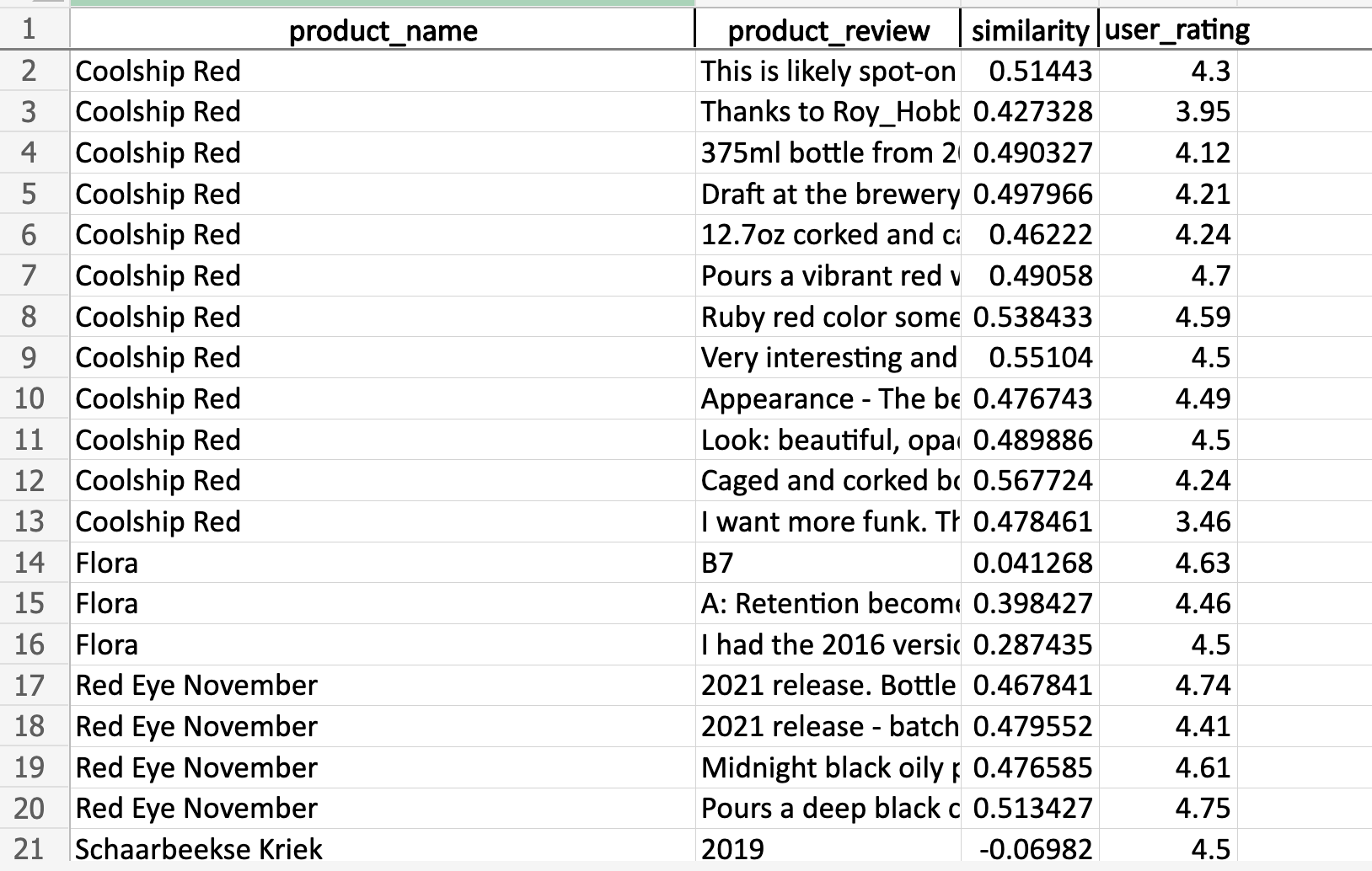
1. Use Web Scraper to extract about 3k reviews of beers from Beeradvocate.com. Scrape 25 reviews and well as the ratings provided by users for each beer. The CSV output file should be processed to retain 3 columns: product\_name, product\_review, and user\_rating. Save it as a Excel file: [beeradvocate\_reviews.xlsx](https://utexas-my.sharepoint.com/:x:/g/personal/rebecca_lee_austin_utexas_edu/EdagonJYYptIlsqofHvSc30BDpxFfisplwJ0jYcGgHZSUg?e=cvPUiB).
2. Five of the most important attributes of beer with word frequency analysis: [beer\_word\_freq.csv](https://utexas-my.sharepoint.com/:x:/g/personal/rebecca_lee_austin_utexas_edu/EYuMR0zIpJpLtZ84cpeFVbcB7gwwM-PKLxSATqFt7fcWgg?e=PxGDDY)

The five most important attributes of beer with word frequency analysis are:

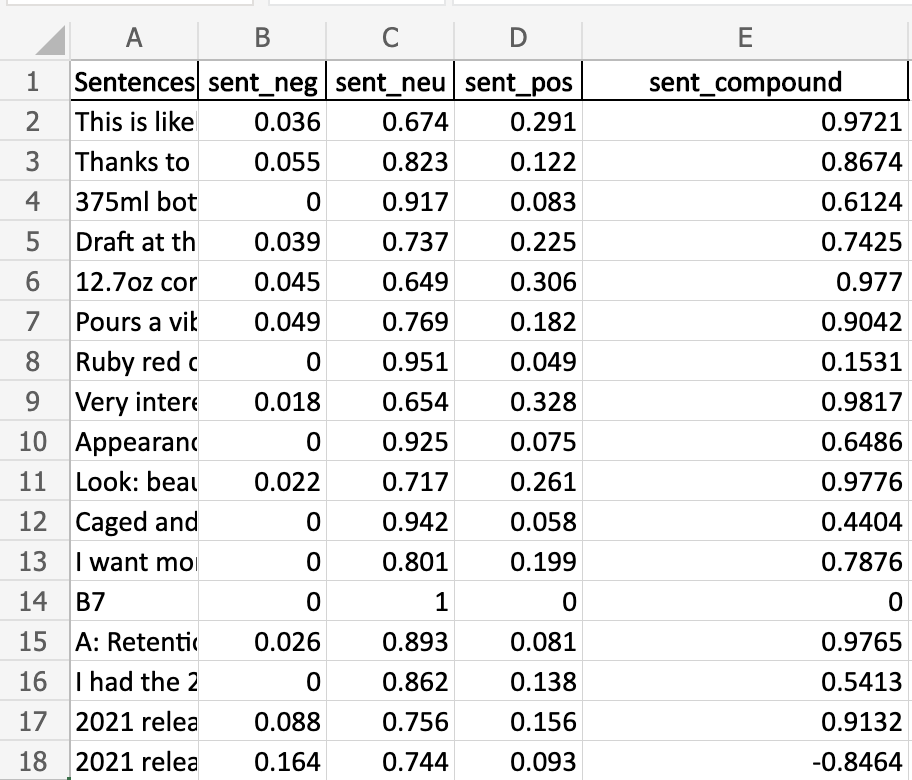
|  |  |
| --- | --- |
| head | 1067 |
| taste | 947 |
| chocolate | 795 |
| dark | 746 |
| aroma | 526 |

We chose head, taste, and aroma for the remaining analysis.

1. **Similarity** analysis with the 3-attribute set and the reviews: We used the above three attribute set: head, taste, and aroma to conduct the similarity analysis and got a similarity score for each review. Below shows what the similarity output sheet looks like.



1. Sentiment analysis on the reviews: We performed sentiment analysis on the reviews and got an overall sentiment on each row. Below shows what the sentiment output sheet looks like.



1. Sort the evaluation scores from high to low, and **recommend 3 beers** (or guitars) to the customer
   1. Combined sentiment scores with similarity scores **(by multiplication)** and combined all ratings into an average per beer: [beer\_attribute\_averages.xlsx](https://utexas-my.sharepoint.com/:x:/g/personal/rebecca_lee_austin_utexas_edu/ERL0dwXKtk9DkPQj931kZjQB6m5MQGzD-xqTX82l88xmVw?e=t24EHq)
   2. The table below shows top 5 beers based on the evaluation scores:

|  |  |  |  |
| --- | --- | --- | --- |
| **Beer** | **Average of overall\_score** | **Similarity** | **Sentiment Compound** |
| **Vanilla Bean Assassin** | 0.48897 | 0.507594 | 0.9637 |
| **Double Dry Hopped Double Mosaic Daydream** | 0.47173 | 0.505031 | 0.934925 |
| **Flora Plum** | 0.45999 | 0.489842 | 0.93772 |
| Ann | 0.45981 | 0.482937 | 0.9523 |
| The Peach | 0.45547 | 0.52634 | 0.8585 |

* 1. The recommended beers based on the evaluation scores are Vanilla Bean Assassin, Double Dry Hopped Double Mosaic Daydream, and Flora Plum.

1. How would your recommendations differ if you ignored the similarity and sentiment scores, and simply chose the 3 highest rated beers from your entire dataset?
   1. Overall highest rated beers: [best\_overall\_averages.xlsx](https://utexas-my.sharepoint.com/:x:/g/personal/rebecca_lee_austin_utexas_edu/EWVBMIIWYxNLlnpFmt4ChrsBXqVnCQsX7zNUml9gmqDAVA?e=pnNYKK)
   2. The table below shows the 5 highest rated beers with their average user ratings:

|  |  |  |  |
| --- | --- | --- | --- |
| **Beer** | **Average of user\_rating** | **Similarity** | **Sentiment**  **Compound** |
| **The Adjunct Trail - Bourbon Barrel-Aged** | 4.93 | 0.477261062 | 0.8126 |
| **Twice the Daily Serving: Raspberry** | 4.83 | 0.48840783 | 0.72876 |
| **Art** | 4.81 | 0.334627507 | 0.32525 |
| Barrel Aged Imperial German Chocolate Cupcake Stout | 4.80 | 0.483257754 | 0.87889 |
| Truth - Vanilla Bean | 4.79 | 0.370209379 | 0.70166 |

* 1. If we simply chose the 3 highest rated beers based only on user ratings, our recommendations would be The Adjunct Trail - Bourbon Barrel-Aged, Twice the Daily Serving: Raspberry, and Art. These are different from the recommended beer list based on similarity and sentiment scores.
  2. Would these three beers meet the requirements of the user looking for recommendations? Why or why not? Justify your answer. Use the user\_rating data and similarity scores to answer this question.

The table above shows the ratings, similarity scores, and sentiments for these top-rated beers. Although our recommended top 3 beers in part F have high ratings, we found that they might have a less desirable similarity and sentiment scores. Therefore, those recommended beers in part F are only based on ratings and may not meet the requirements of the user. Since the similarity scores are lower than the beers we recommended in part E, it shows that the reviews do not positively mention the attributes the user is looking for.