



BEER CHALLENGE

Marina Sanchez
10/21/2022

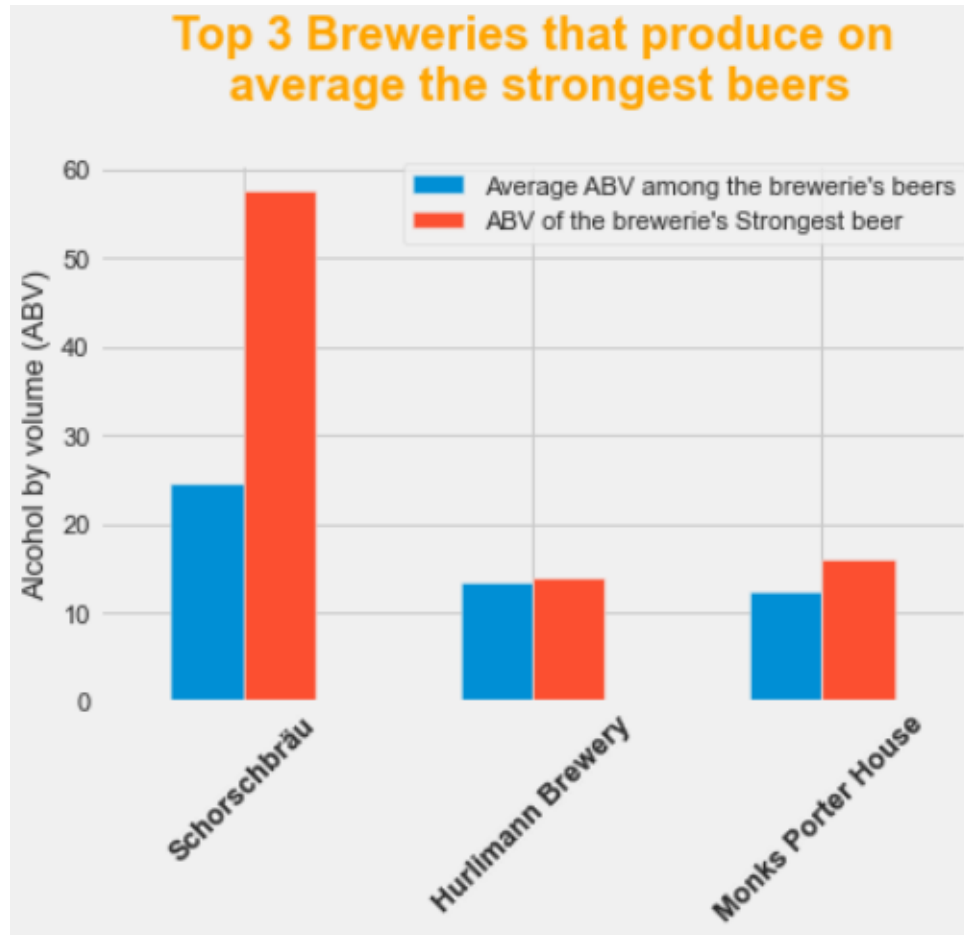


CONTENT

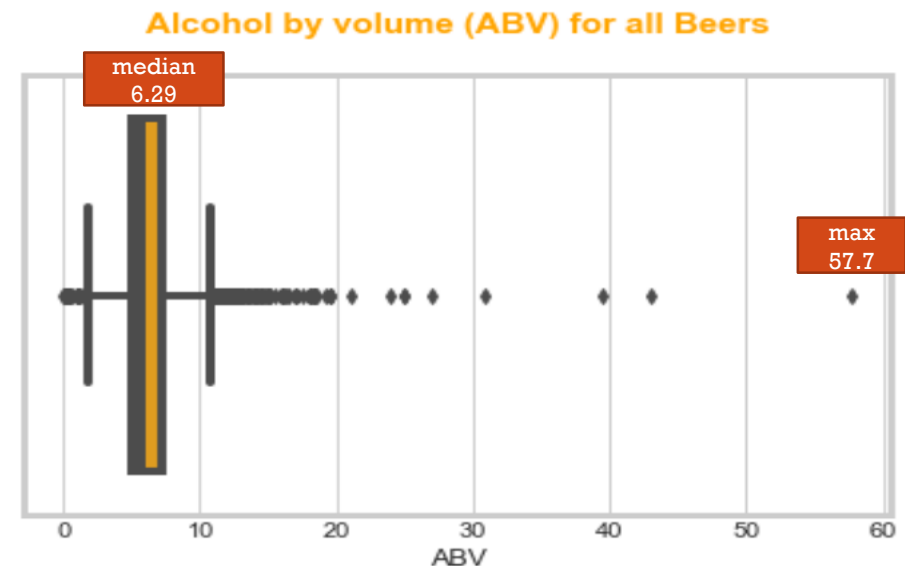
- Top Breweries With The Strongest Beers
- The Best Years For The Beer Lovers
- Fruity Beers I Would Recommend
- People's Favourite Beer Styles
- Important Features According To Users
- Sentiment Vs Overall Score In Beer Styles
- Similar Drinkers
- Links



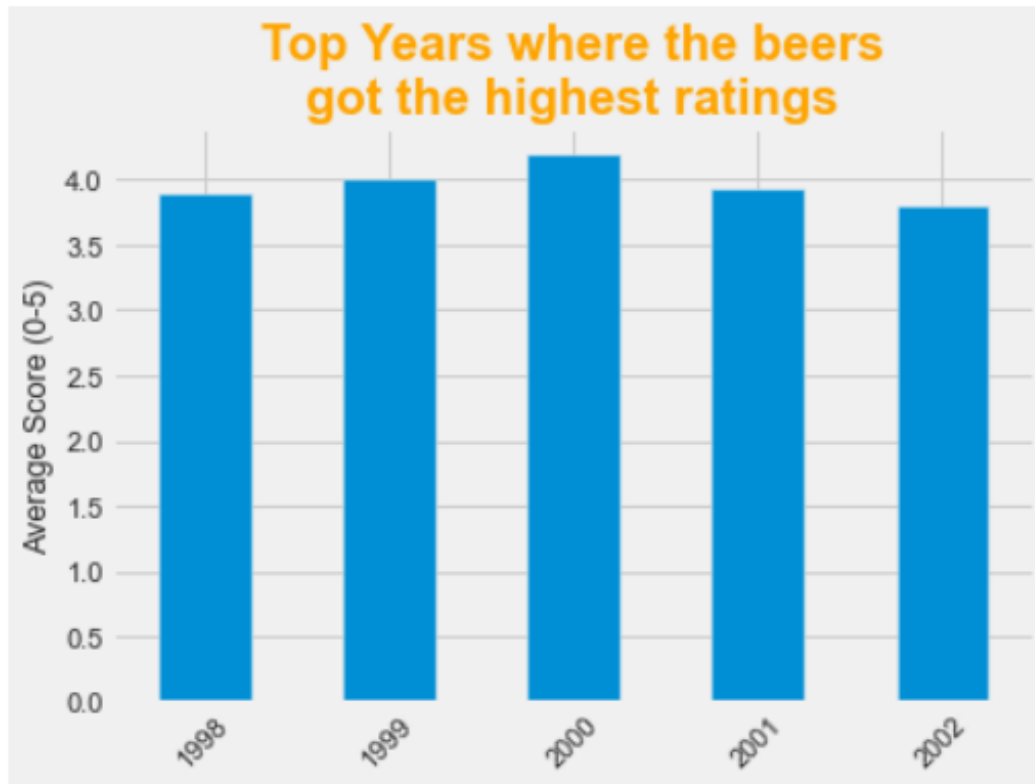
TOP BREWERIES WITH THE STRONGEST BEERS



- The breweries **Schorschbräu**, **Hurlimann Brewery** and **Monks Porter House** produce, on average, the strongest beers
- Schorschbräu produces the strongest beer (ABV 57.7)
- Hurlimann Brewery and Monks Porter **do not** produce any of the strongest beers (ABV above 15).



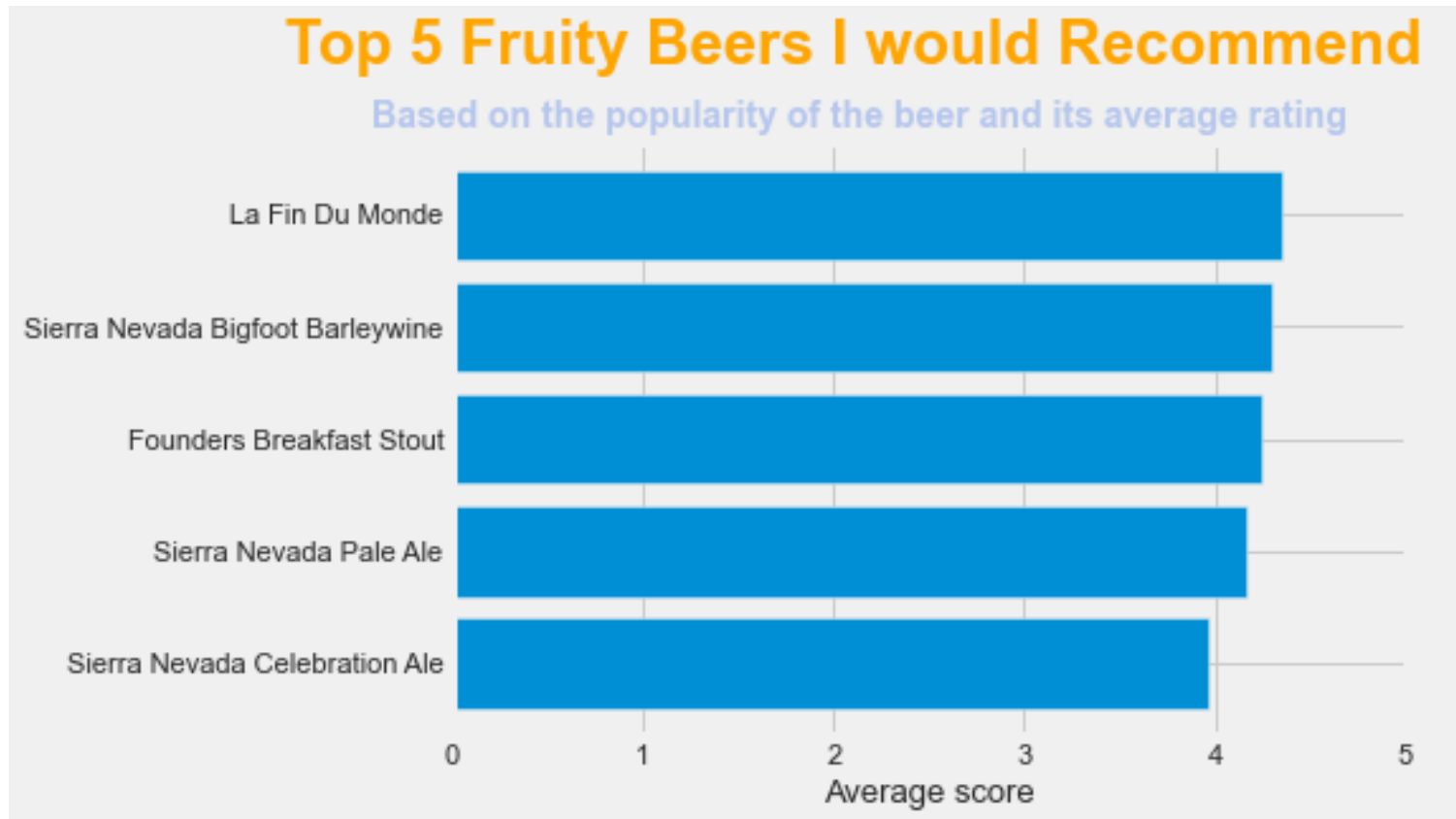
THE BEST YEARS FOR THE BEER LOVERS



- It looks like the period from 1998 to 2002 were good years.
- But year 2000 was the best year for the beer lovers!



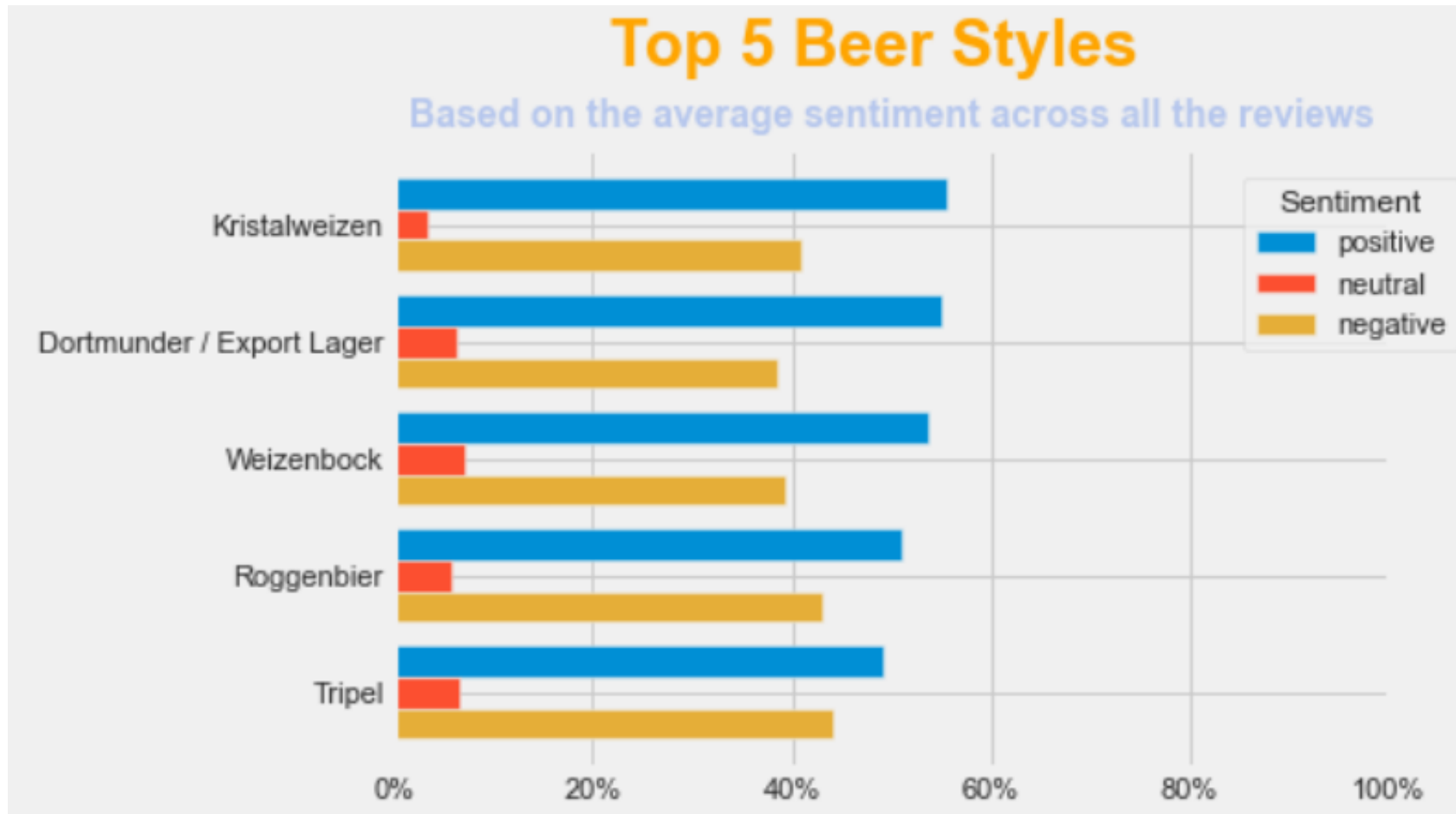
FRUITY BEERS I WOULD RECOMMEND



- I love **fruity** beers...
- The **written reviews** of these popular beers say they are fruity!
- And when I like something I **recommend** it!



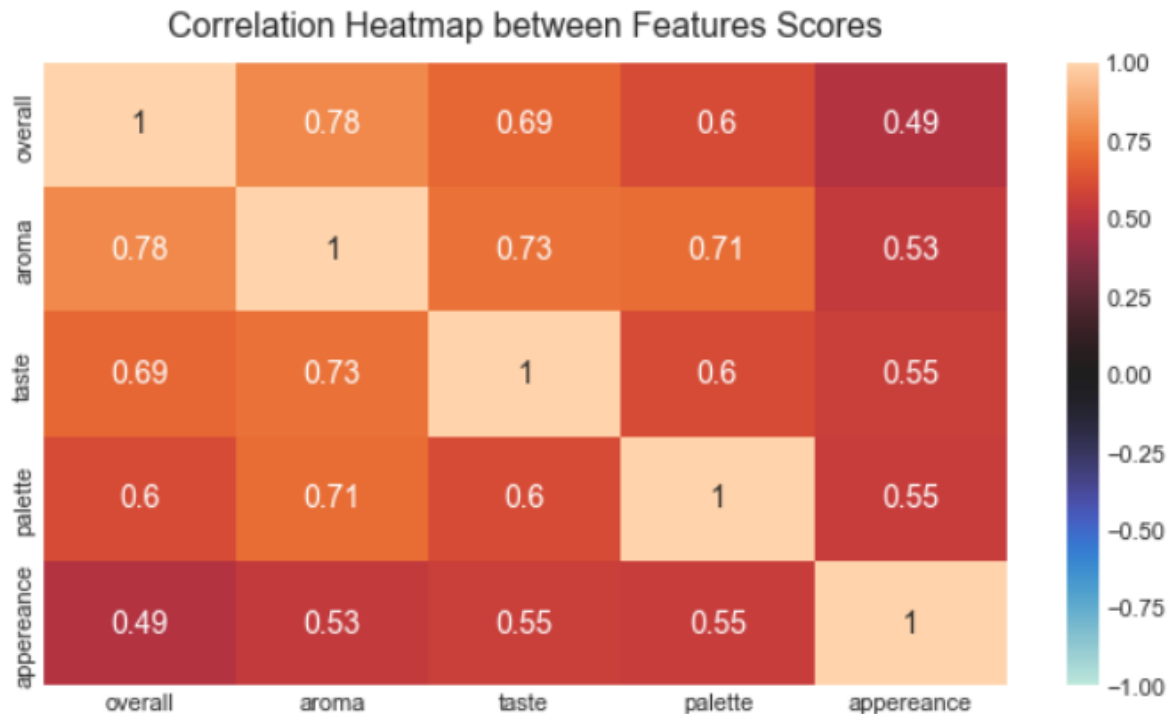
PEOPLE'S FAVORITE BEER STYLE



MORE INSIGHTS



IMPORTANT FEATURES



- We can see a pretty strong **positive correlation** between overall score and:
 - Aroma (0.78)
 - Taste (0.69)
 - Palette (0.60)
 - Appereance (0.49)
- But which one of them has the **most influence** over the overall score? We will find out in the next slide!



IMPORTANT FEATURES

- To find the **most influencing factor** I have fitted four **simple linear regressions** between the overall review and each of the predictors individually

Dependent variable: overall review score | **Independent variables:** Aroma, Taste, Palette, Appereance

Factor	Coefficient	R-squared
Aroma	0.996	0.985
Taste	1.009	0.980
Palette	1.007	0.974
Appereance	0.982	0.975

- All the models **explain** more than **95%** of the variability in overall: **great fits!**
- These four factors are **relevant** to the users when defining the quality of a beer in the following order: **Taste > Palette > Aroma > Appearance**



SENTIMENT VS. OVERALL SCORE IN BEER STYLES

Do the emotions expressed in the reviews match the average numerical score for each beer style?

Spoiler... No!

PROCESS

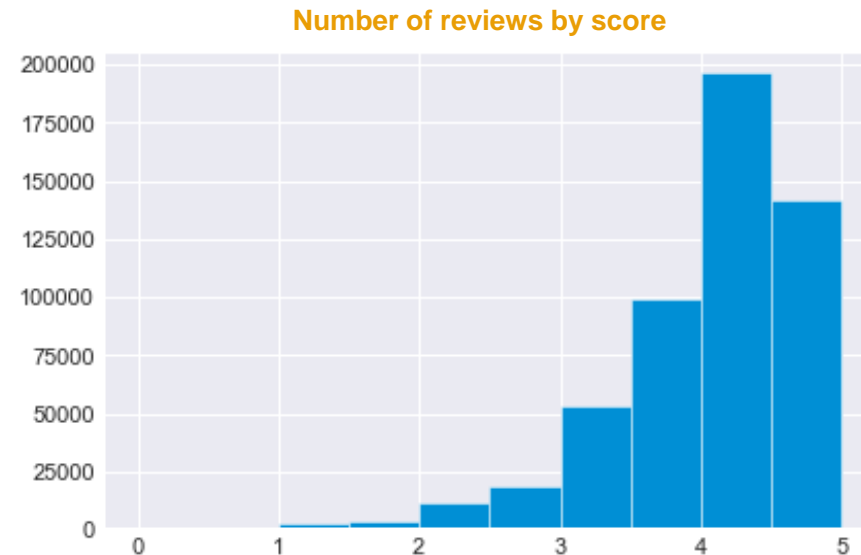
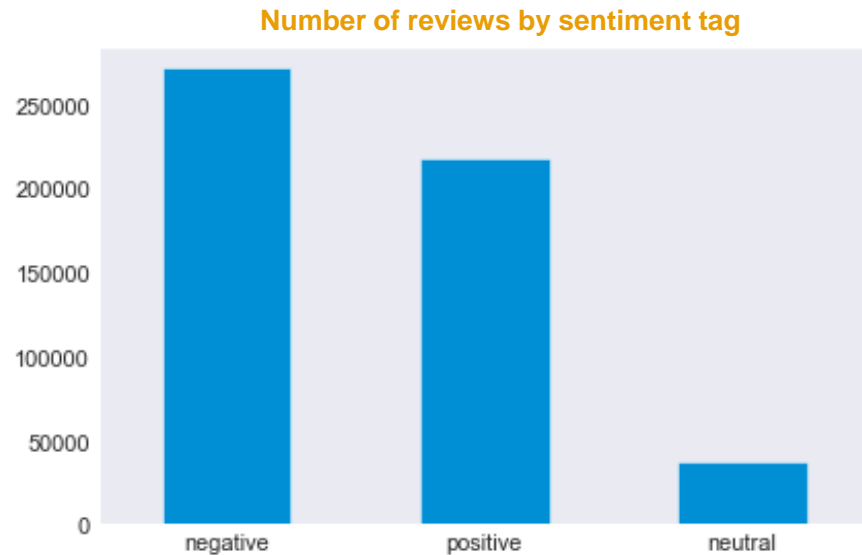
- **Sentiment Analysis** Classification: tag the reviews as **Positive**, **Neutral** or **Negative**
- **Crosscheck** between predominant sentiment and average numerical **scores**

➤ Con't in next slide



SENTIMENT VS OVERALL SCORE IN BEER STYLES

In the sentiment plot predominate **negative emotions** | but in the scores plot predominate **good ratings!**



- Considering Negative (★ or ★★), Neutral (★★★) and Positive(★★★★ and ★★★★★)

We find a **missmatch** in **86%** of the reviews!

- Why?
 - Maybe users tend to express more their **negative thoughts**, even though their overall opinion is positive.
 - The **Sentiment Model** didn't perform very well in detecting the overall sentiment of the reviews.



SIMILAR DRINKERS

- **Recommender System** based on **TFIDF** and **Cosine Similarity** of the written reviews
 - Similar users are paired according to the similarity in the content of their written reviews

review_profileId	0	3	8	28	36	42	53	78	83	85	...
review_profileId											
0	1.000000	0.483075	0.493136	0.415329	0.513497	0.417000	0.528434	0.359105	0.552174	0.399219	...
3	0.483075	1.000000	0.473516	0.490897	0.487994	0.486469	0.520533	0.401002	0.559191	0.436085	...
8	0.493136	0.473516	1.000000	0.481470	0.598935	0.650622	0.595328	0.618701	0.543657	0.388774	...
28	0.415329	0.490897	0.481470	1.000000	0.498915	0.603996	0.488314	0.501433	0.451463	0.362354	...
36	0.513497	0.487994	0.598935	0.498915	1.000000	0.619143	0.547832	0.541339	0.532874	0.358998	...
...
22764	0.589563	0.508729	0.665936	0.474009	0.590224	0.605349	0.657412	0.522858	0.634965	0.447959	...
22765	0.487732	0.537032	0.628344	0.572130	0.598275	0.670492	0.631491	0.641150	0.555457	0.378396	...
22769	0.468455	0.417078	0.528164	0.465376	0.579044	0.614130	0.544376	0.449702	0.504533	0.371830	...
22773	0.463272	0.557181	0.480788	0.657137	0.543521	0.585521	0.511541	0.431591	0.528680	0.445367	...
22785	0.413179	0.435211	0.635383	0.456950	0.497676	0.642659	0.579557	0.604876	0.459556	0.371563	...

Cosine Similarity Matrix



SIMILAR DRINKERS

Results

- Example of the **output** generated for some of the users (see code in GIT for all the details)
- The similar users are chosen according to their cosine similarity in the table of previous slide

```
User with ID 0 is similar to the following users:
```

- User Id 14598 : 67.42 % similarity
- User Id 9532 : 67.32000000000001 % similarity
- User Id 8893 : 67.05 % similarity

```
User with ID 3 is similar to the following users:
```

- User Id 19911 : 62.96000000000001 % similarity
- User Id 14518 : 62.25000000000001 % similarity
- User Id 483 : 61.8 % similarity

```
User with ID 8 is similar to the following users:
```

- User Id 19118 : 74.77000000000001 % similarity
- User Id 10842 : 74.46000000000001 % similarity
- User Id 2735 : 73.83 % similarity



LINKS

- GIT Repository with source code: <https://github.com/msmillan7/BeerChallenge>

