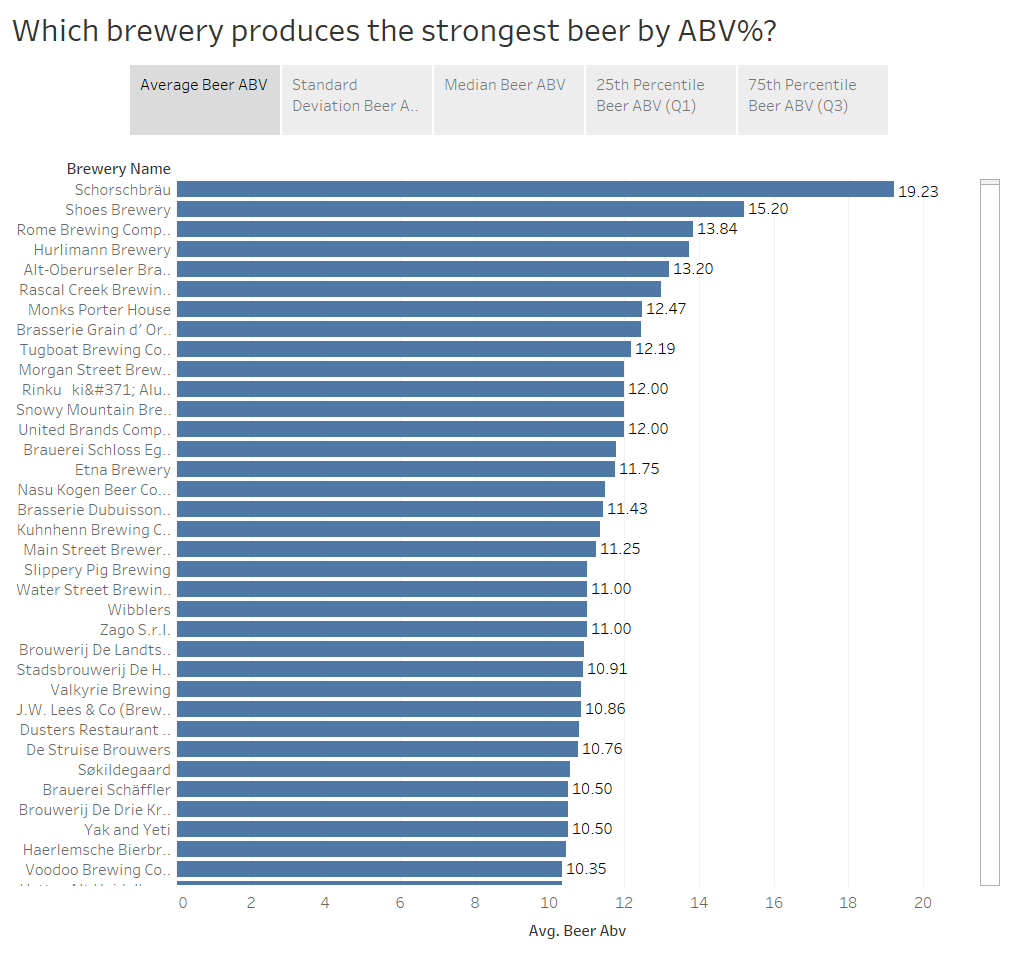
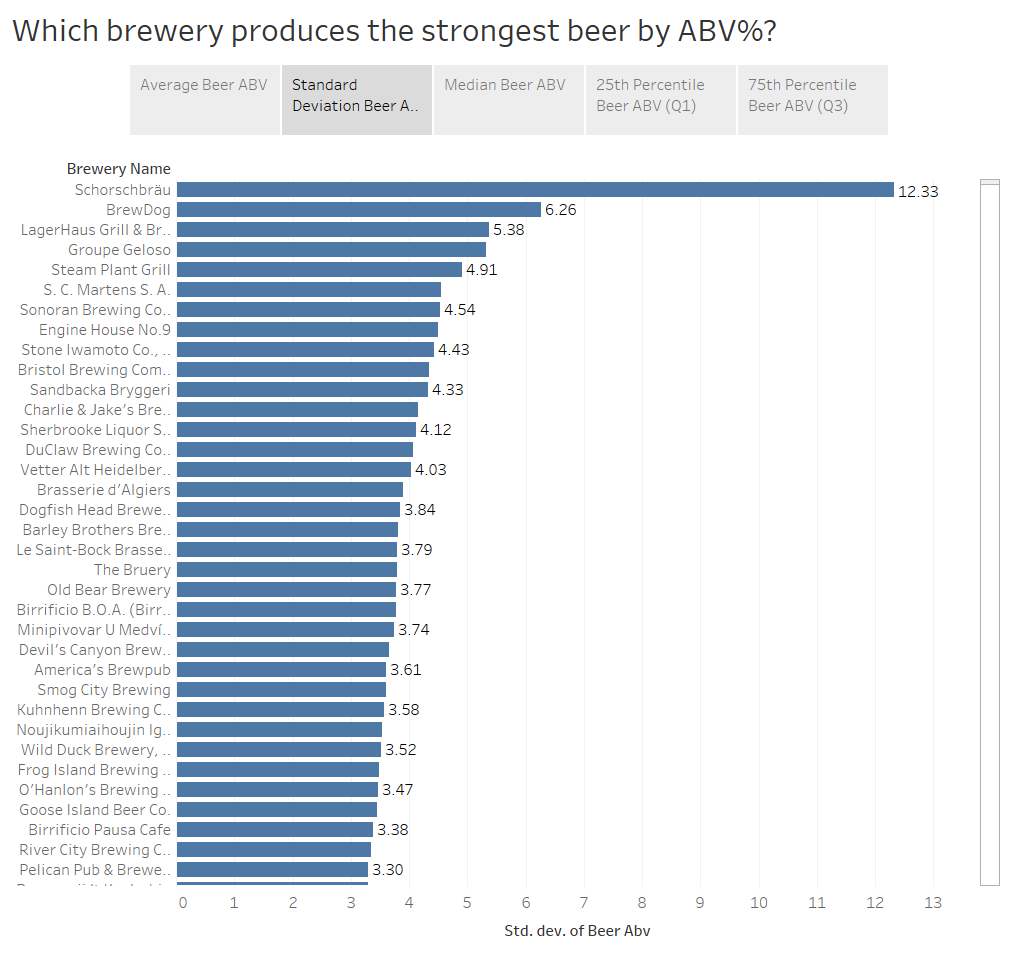
**Data Scientist Exercise**

1. Which brewery produces the strongest beers by ABV%?

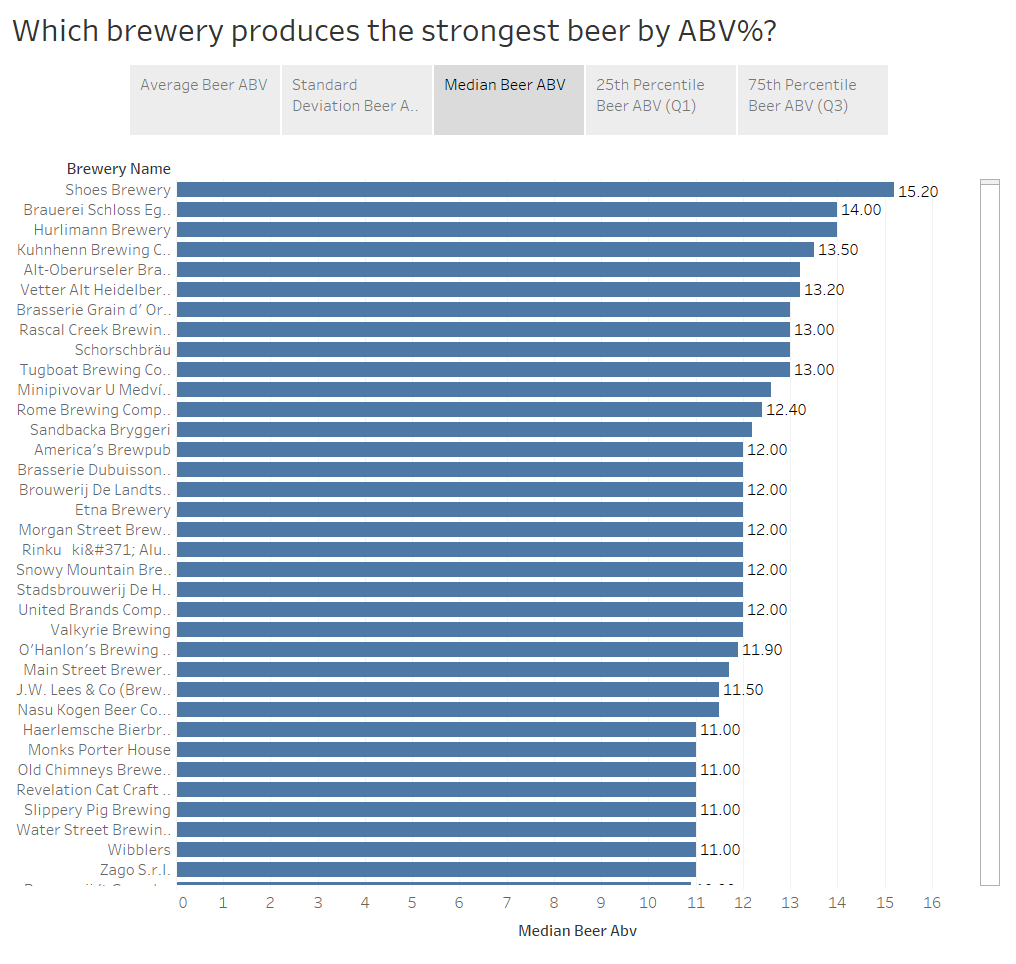
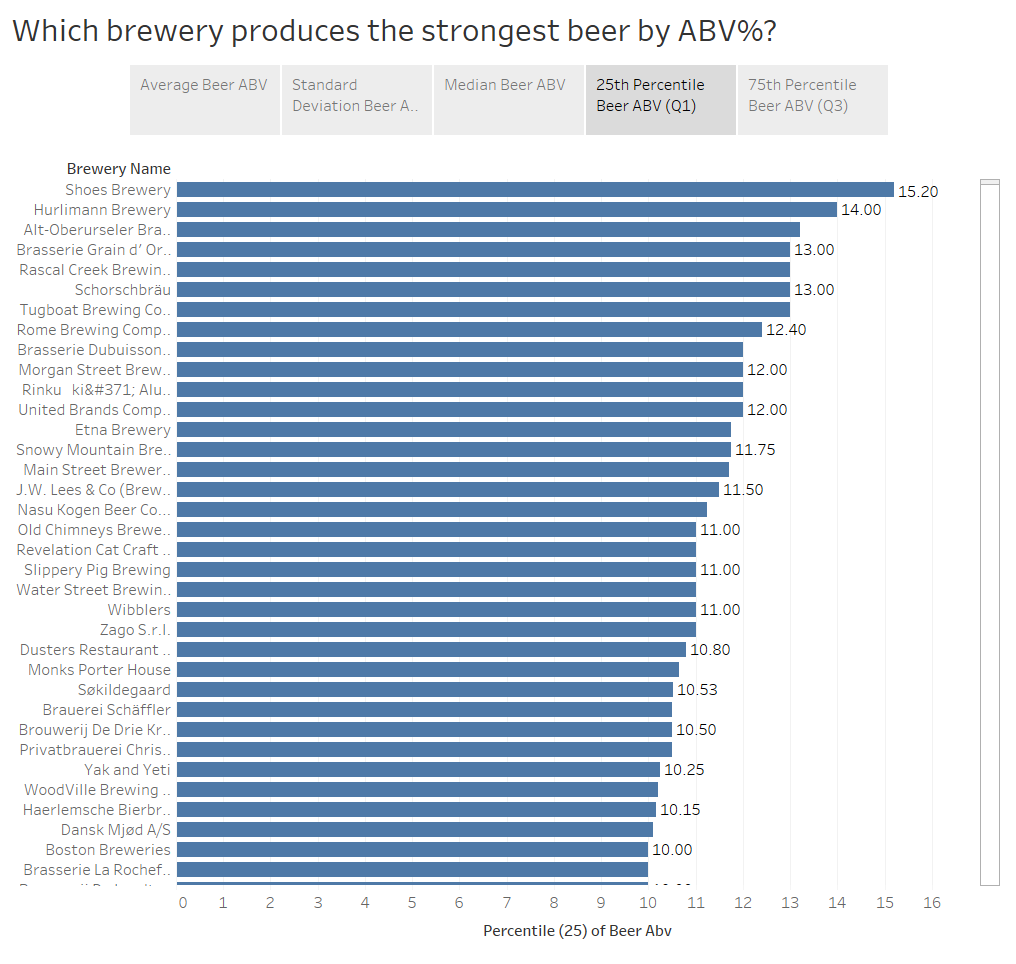
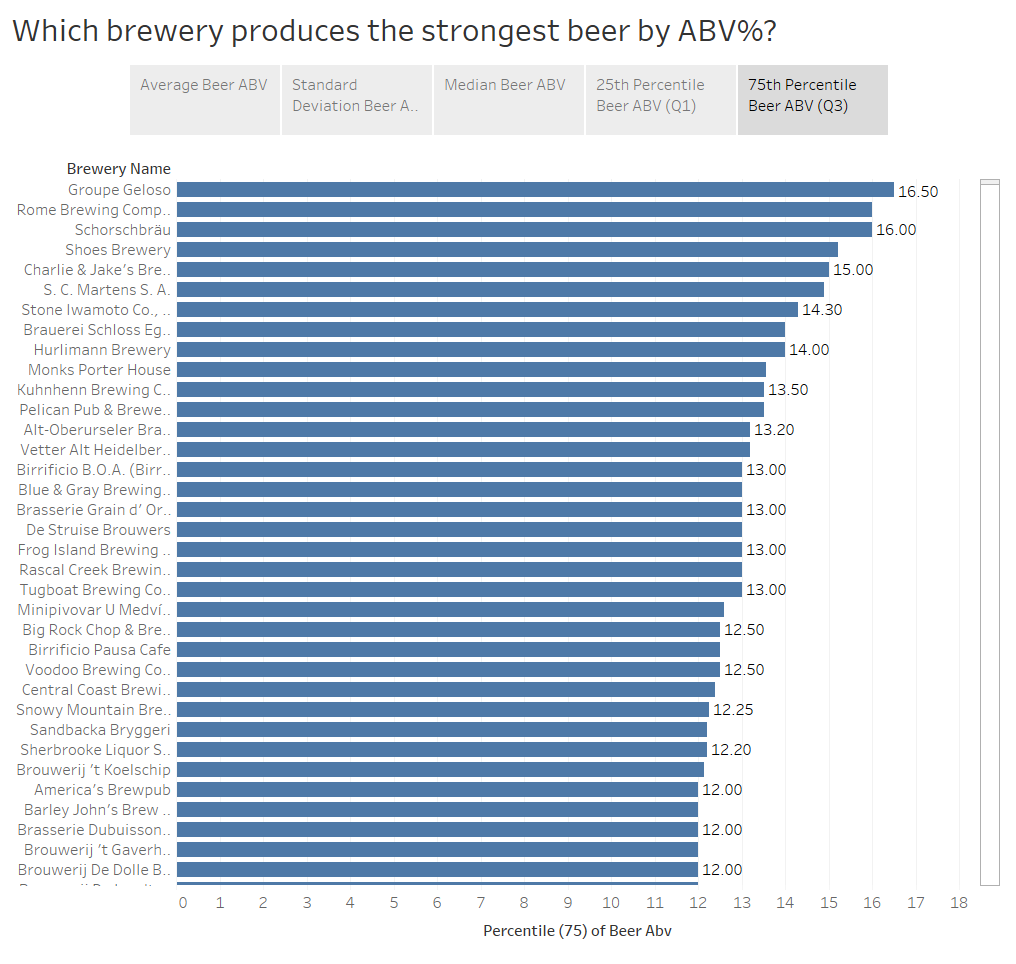
This appears to be a relatively straightforward question. We have the option to simply find the average ABV% for each brewery and sort from largest to smallest. However, the average ABV% can easily be affected by outliers, especially considering the high variance in the number of beers each brewery produces. To get a better understanding, I looked at the mean ABV% for each brewery alongside the standard deviation, median, and Q1 and Q3 values.



We can see above that the brewery Schorschbräu appears to have the highest average (mean) ABV% by a significant margin followed by Shoes Brewery. However, Schorschbräu also has the highest standard deviation between their beers by an even larger margin (see below). In other words, while the brewery does produce beers with extremely high ABV%, those values are not *consistent* among all of their products. If you were to select a random beer from Schorschbräu, you might get their most alcoholic beverage with a staggering 57.7% ABV or you could get a beer containing 4.9% ABV.



Next, I examined the median ABV% for each brewery along with the 25th and 75th percentiles. Those visuals can be seen below (I hope you’re not tired of looking at blue lines).

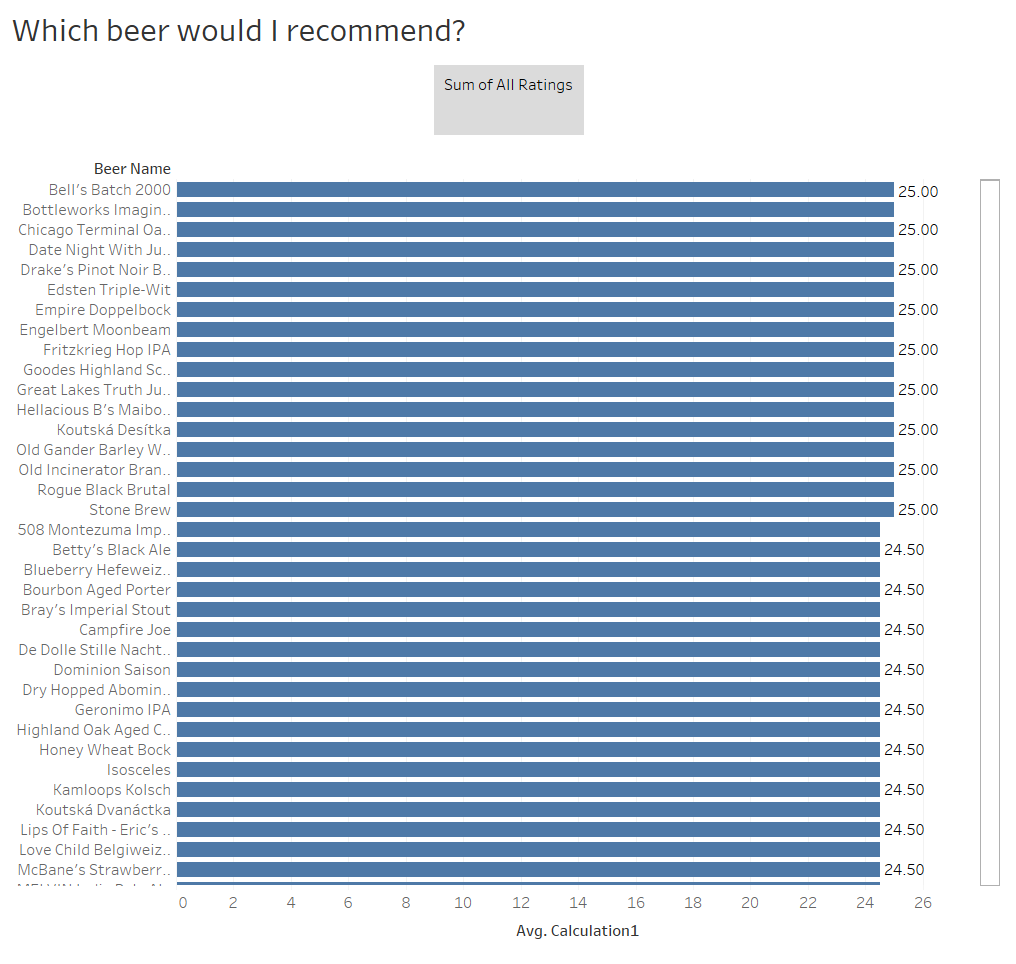
Here we can observe that Shoes Brewery (in addition to having the second highest mean ABV%) has the highest median ABV%, the highest ABV% in the 25th percentile, and has one of the highest ABV% in the 75th percentile. It is interesting to note that all of these values (mean, median, Q1, Q3) are the same at 15.2. This leads me to conclude that all of the beers produced by Shoes Brewery have a 15.2 ABV%.

So to finally answer the original question, I would claim that Shoes Brewery produces the strongest beers by ABV%. A single randomly selected beer from Shoes Brewery will likely have a higher alcohol content than a single randomly selected beer from any other brewery.

1. If you had to pick 3 beers to recommend using only this data, which would you pick?

Before we even look at the data, common sense would tell us to recommend the three beers with the highest overall average ratings. Unfortunately, there are hundreds of beers that all have the maximum 5-star review so it won’t be quite that easy.

It is possible (and common) for a beer to have a 5-star overall rating while falling short in the contributing categories. For example, there may be a beer with a 5-star overall rating but only a 4-star rating in taste or appearance. To find the beers with the best reviews in all categories, I found the sum of the ratings in all five areas (overall, appearance, aroma, palate, and taste). With five categories and a maximum of 5 points in each category, the absolute perfect score would be a 25. I found that there were only 17 beers that received this perfect score.



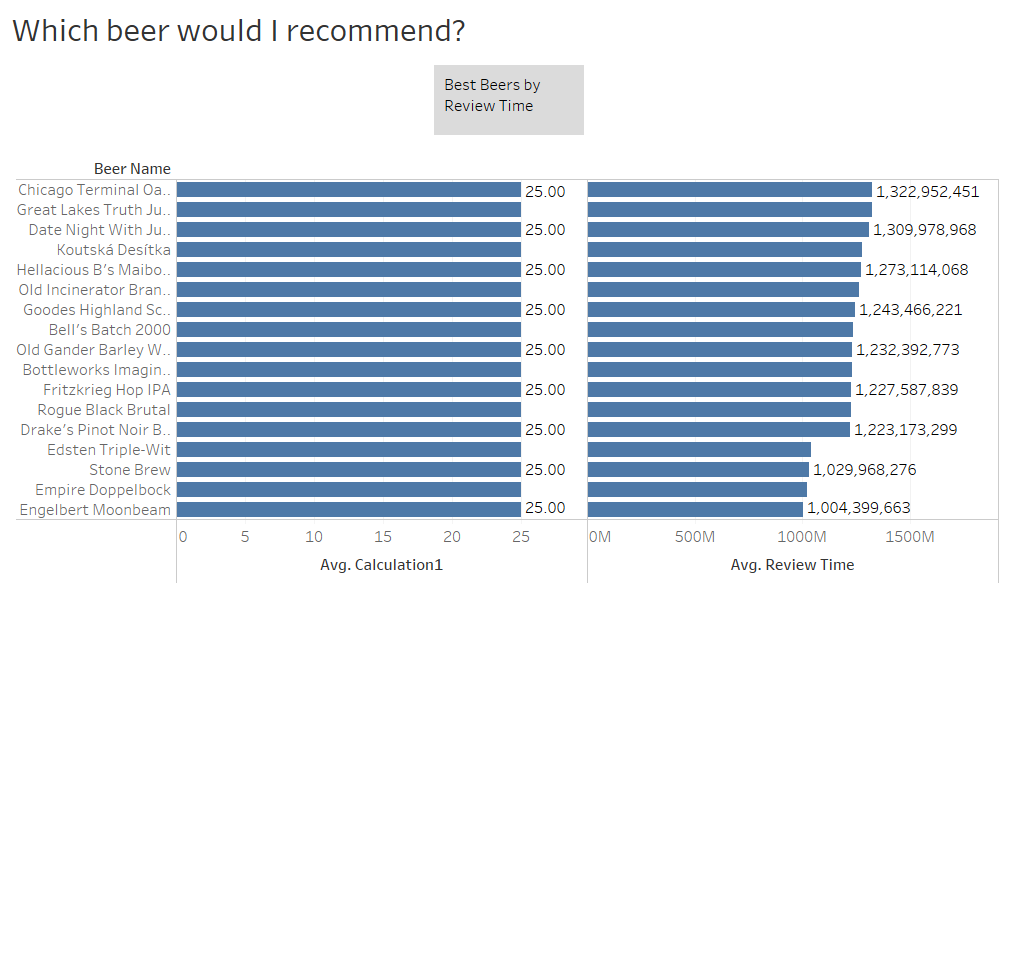
While each of these seventeen of these beers received 5-out-of-5 stars in all categories, they were not all reviewed for an equal amount of time. If we take the assumption that a greater review time would lead to greater accuracy, then the three beers I would recommend would be:

***Chicago Terminal Oatmeal Bourbon Stout***

***Great Lakes Truth Justice and The American Ale***

***Date Night with Jumbo Love***

It should also be noted that the number of reviewers was taken into consideration. However, only one of these 17 beers had been reviewed by more than one person. That beer was ‘Date Night with Jumbo Love’ which is already on my recommended list.



1. Which of the factors (aroma, taste, appearance, palate) are most important in determining the overall quality of a beer?

There are a few different ways we can determine the relationship (or lack thereof) between the overall quality of a beer and its aroma/taste/appearance/palate. I chose to find the Pearson correlation, which can measure the strength and direction of the linear relationship between two variables. In this case, we are more interested in the strength of the relationship as it’s unlikely for any of the variables to have a negative relationship. For example, it would not make much sense for the overall ratings of a beer to go *down* as the ratings of a beer’s taste went up.

A value of 1.00 indicates a perfect positive relationship while a value of 0 indicates no relationship between the variables. To answer this question, we are not necessarily looking to meet or exceed a certain number. Rather, I want to find which of the factors is most closely related to the increase and decrease of the overall ratings (which of the correlation coefficients is closest to 1).

The results are as follows:

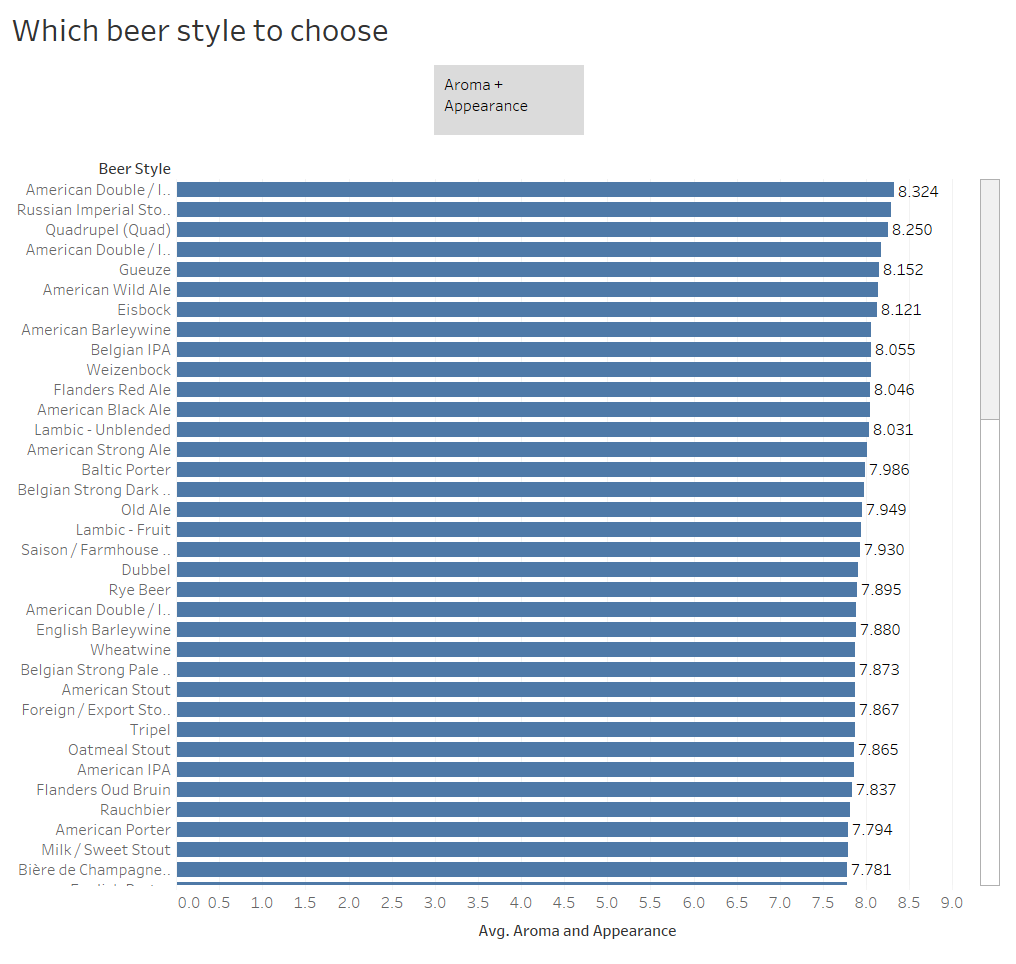
|  |  |
| --- | --- |
| **Pearson Correlation Coefficient** | |
| *Overall Ratings and Taste* | *R = 0.790* |
| *Overall Ratings and Palate* | *R = 0.702* |
| *Overall Ratings and Aroma* | *R = 0.616* |
| *Overall Ratings and Appearance* | *R = 0.501* |

We can see that taste is the most important factor in determining the overall quality of a beer, followed closely by palate. Aroma and appearance are less important in determining the quality of a beer, but they still play a part.

1. If I typically enjoy a beer due to its aroma and appearance, which beer style should I try?

First and foremost, I would like to inform you that aroma and appearance are actually the least important attributes to consider when determining the quality of a beer. Everyone knows that taste and palate are more closely correlated with overall quality!

Second, while answering this question, I was working under the assumption that both aroma and appearance were of equal importance to you. With that in mind, I wanted to find the beer styles with the highest average *combined* scores for aroma and appearance.



It appears that the ***American Double/Imperial Stout*** has the highest combined average scores for aroma and appearance, so that is the beer style I would recommend to you.