Project - HotMess Brewery
Team Members: Arsel Vargas, David Hutchinson, Christine Hemphill, Maninder Bawa

Overall Project: As a new Brewery in the US, we want to make our own beer, but what type will our first beer be, an ale or a lager?

Hypothesis: For beer drinkers in the US, lagers are more preferred than ales.

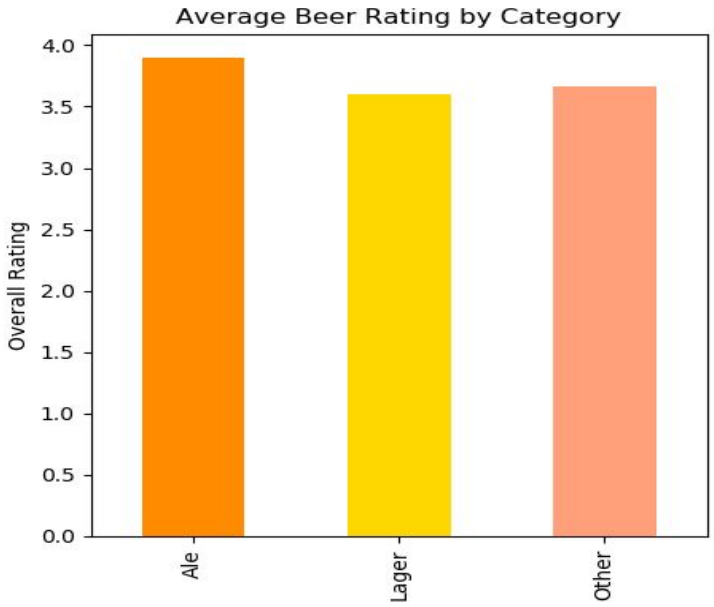
Alternative: For beer drinkers in the US, ales are more preferred than lagers.

Null: There is no significant difference in preference of beer between lagers and ales.

Data Source:
https://www.kaggle.com/rdoume/beerreviews

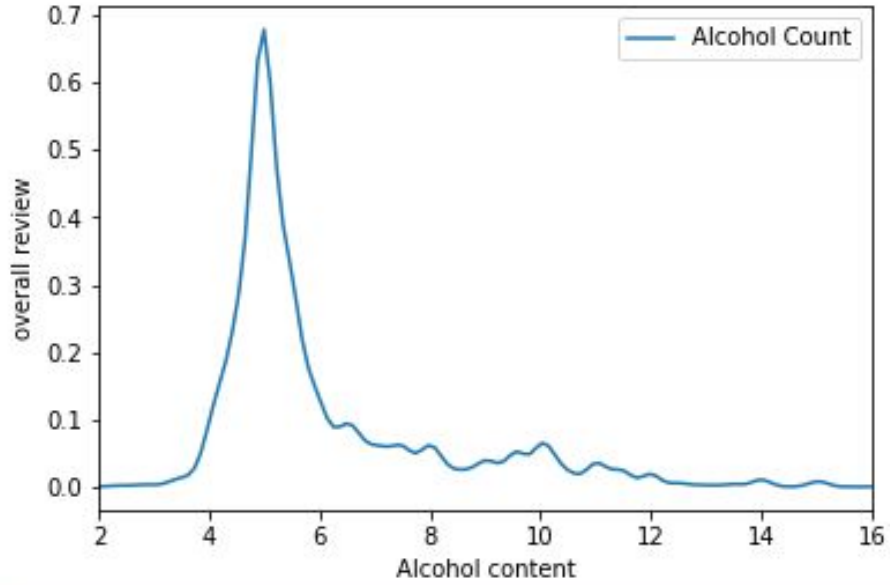
Questions:
What is the most popular type of beer among reviewers?

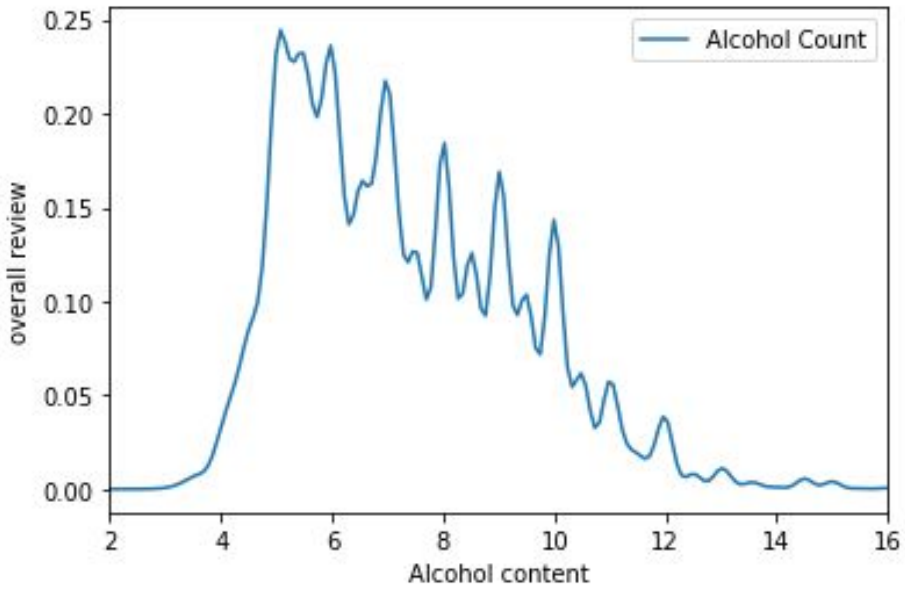
Though the average mean between ales (3.9) and lagers (3.6), ales were the most popular beer among reviewers. Another possible indication of popularity could be the number of reviews for the types of beers. There were over one million ale reviews while there were only three hundred thousand lagers. The same could be said also of the number of beer styles we allocated into the three categories. There were a total of one hundred and four beer styles that were divided into 54 ales, 26 lagers, and 24 that were other.



How does alcohol content contribute to the ratings received?

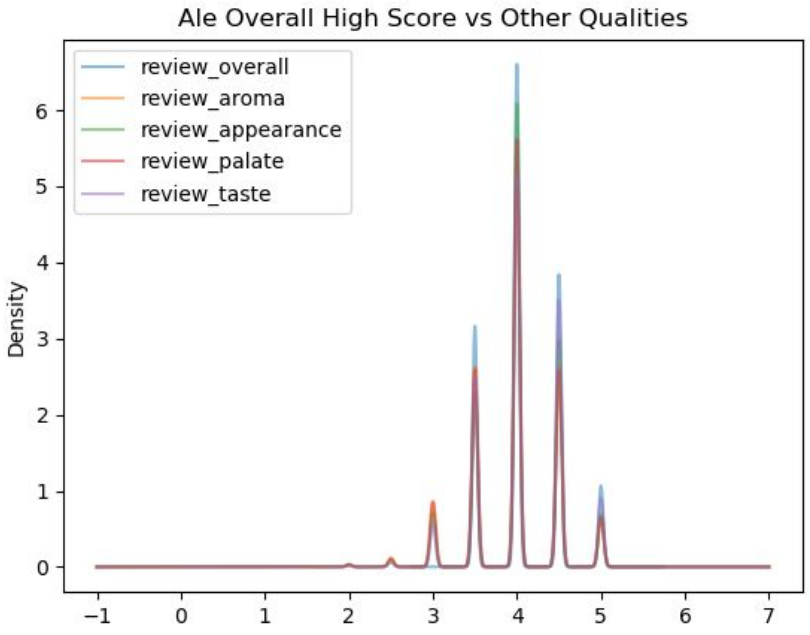
After filtering data between ales and lagers we compared the overall rating given by the reviewer with the alcohol content of the beer. Reviewing the plot it showed people like Lager with ABV~5% and as the ABV goes up review went down very drastically. On the other hand people like Ale's with ABV between 5% - 7% and reviews decline steadily till 10% ABV. As a result we can say people prefer ale's with higher alcohol content when Compared to lagers

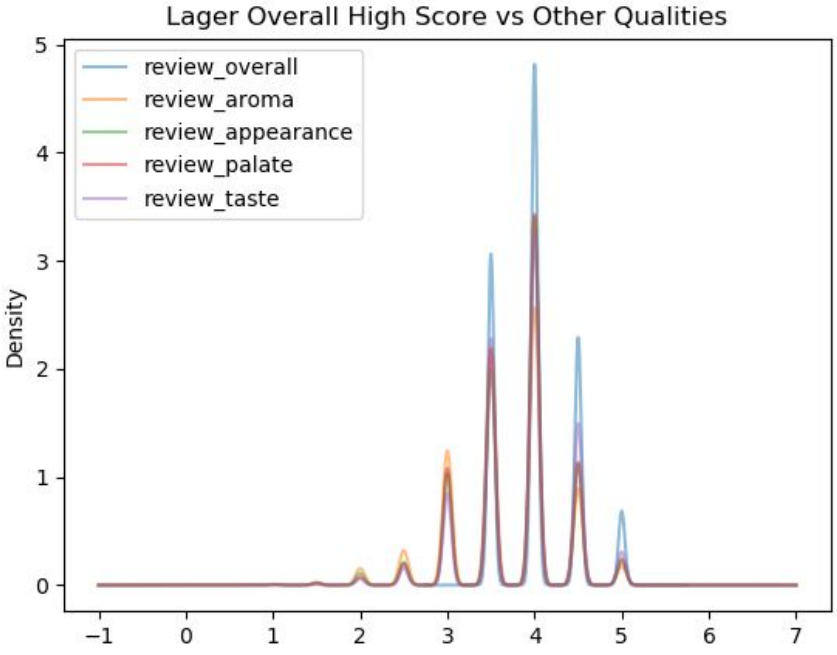


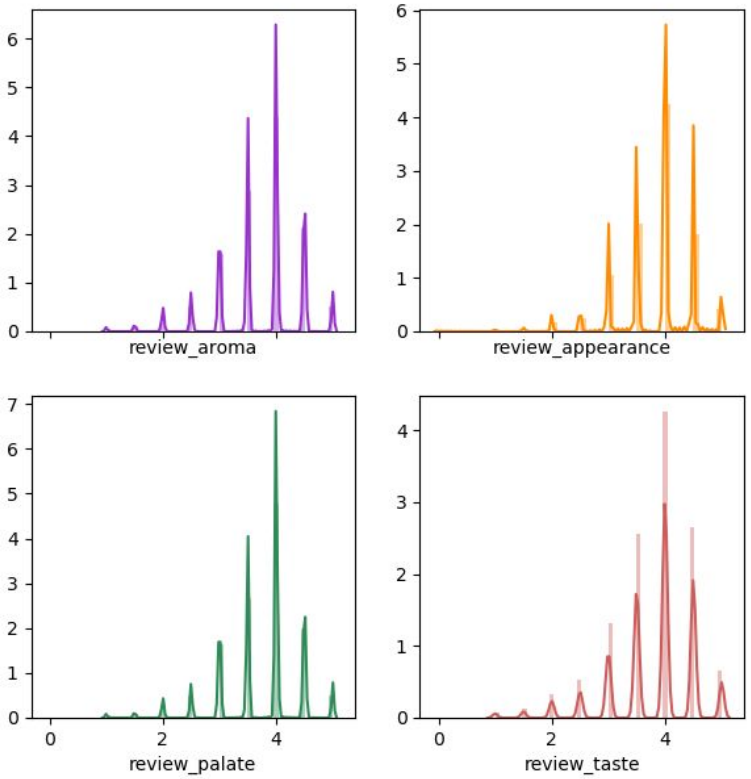


How does the overall review correlate to the individual reviews of appearance, aroma, palate, and taste?

When looking at the data we observed that the overall score was not an average of the other scores. That led to trying to find a correlation between the overall score of a beer style and the scores of aroma, appearance, palate, and taste. For both ales and lagers there were more scored as 4 than any other. Looking at the density plot, ales that scored higher than 4 overall also had a higher score for taste. For lagers, the palate score was more closely tied to the 4 score while the aroma and palate were tied to a score of 5. This was important because looking at the scores of an individual beer you could see the overall score was not a mathematical equation of the other scores. Basic histogram plots were made for the lager and ale overall to other qualities review but were not included in presentation or write up because density plots were better at distinguishing the different scores.

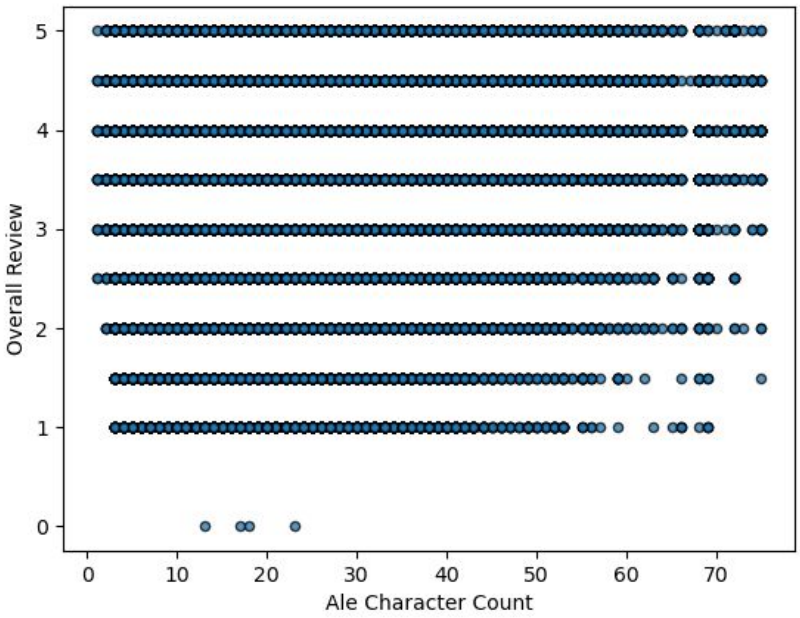


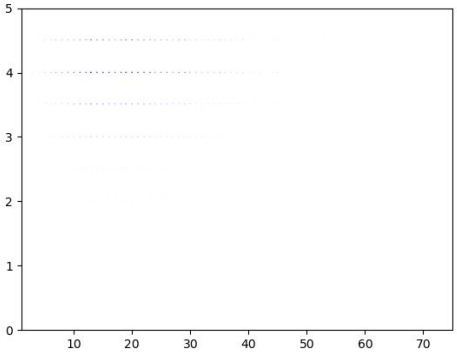

Lager Overall High Score vs Other Qualities
review_overall
review_aroma
review_appearance
review_palate
review_taste
Density

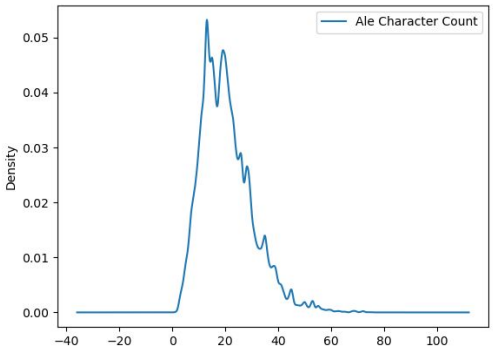

review_aroma
review_appearance
review_palate
review_taste

Can the complexity of the product name affect the overall rating?

Upon further review of the character count of the product name, the initial scatter plot provided little insight, due to the sheer amount of data points.







Using a density chart to focus on our most active section of reviews, we were able to narrow our focus to a more informative range. We have a dramatic loss in review quality as the number of characters increases. Assuming this may be due to difficulty in memory retention, especially among our target audience, we suggest a beer name between 7 to 15 characters as the optimal range.

Does the public awareness of the brewery affect the overall score?

After reorganizing the data to represent the top 3 most reviewed breweries, we compared the average overall review scores of each. Even though Sam Adams had the highest quantity of reviews, Stone had the highest quality of reviews. In conclusion, we have determined that "popularity" and/or more widespread distribution didn't necessarily make a better beer. The average scores may only be nominally different, but when the sheer number of reviews are taken into consideration, it is clear that having a wide range of different beers spread out over a larger region doesn't make one brewery better than another. Additional demographic data would have been instrumental in demonstrating this.