Technical Exercise for Data Scientist Candidates

For this exercise, you will be working with beer data which can be downloaded from here https://drive.google.com/open?id=1e-kyoB97a5tnE7X4T4Es4FHi4g6Trefq

Unzip the file and you should see a CSV file, called "BeerDataScienceProject.csv"

The columns are
beer_ABV beer_beerId beer_brewerId beer_name beer_style
review_appearance review_palette review_overall review_taste review_profileName
review_aroma review_text review_time

You may use any one or more among the following open source programming language(s) (Python, R, C++, Java, Scala, Julia, etc.) if you like to do so (no "SQL" Based Code). You are welcome to use any other open source framework, packages and technologies. Please provide your answers below the questions.

When you submit your answers, please provide a link to your code in public repository like github with instructions on how to run it to reproduce your results.

Include few slides to summarize your findings and to convince the audience.

Answer <u>as many questions as possible.</u>

Questions:

1. Rank top 3 Breweries which produce the strongest beers?
2. Which year did beers enjoy the highest ratings?
3. Based on the user's ratings which factors are important among taste, aroma, appearance, and palette?
4. If you were to recommend 3 beers to your friends based on this data which ones will you recommend?
5. Which Beer style seems to be the favorite based on reviews written by users?
6. How does written review compare to overall review score for the beer styles?
7. How do find similar beer drinkers by using written reviews only?

*Please include all plots you created to complete the project and to explain your results.