Leap Skills - Data Scientist Hiring Challenge

Beer Reviews Analysis

OVERVIEW

1. You must be able to read the data into R or Python and be able to summarize and explore the data.
2. Aggregate and manipulate the data accordingly (simple means, thresholds, grouping and subsetting).
3. Visualize and communicate results (extra points for presenting the findings and code in a well-documented RMarkdown or iPython Notebook).

Answer the following questions using the dataset provided.

1. Which brewery produces the strongest beers by ABV%?
2. If you had to pick 3 beers to recommend using only this data, which would you pick?
3. Which of the factors (aroma, taste, appearance, palette) are most important in determining the overall quality of a beer?
4. Lastly, if I typically enjoy a beer due to its aroma and appearance, which beer style should I try?

Data set

Link: https://s3.ap-south-1.amazonaws.com/leapskills-cdn/beer-review-data-set.csv

Data fields

- index - an identifier for the review
- beer/ABV - the alcohol by volume of the beer
- beer/beerId - a unique ID indicating the beer reviewed
- beer/brewerId - a unique ID indicating the brewery
- beer/name - name of the beer
- beer/style
- review/appearance - rating of the beer's appearance (1.0 to 5.0)
- review/aroma - rating of the beer's aroma (1.0 to 5.0)
- review/overall - rating of the beer overall (1.0 to 5.0)
- review/palate - rating of the beer's palate (1.0 to 5.0)

- review/taste - rating of the beer's taste (1.0 to 5.0)
- review/text - the text of the review
- review/timeStruct - a dict specifying when the review was submitted
- review/timeUnix
- user/ageInSeconds - age of the user in seconds
- user/birthdayRaw
- user/birthdayUnix
- user/gender - gender of the user (if specified)
- user/profileName - profile name of the user