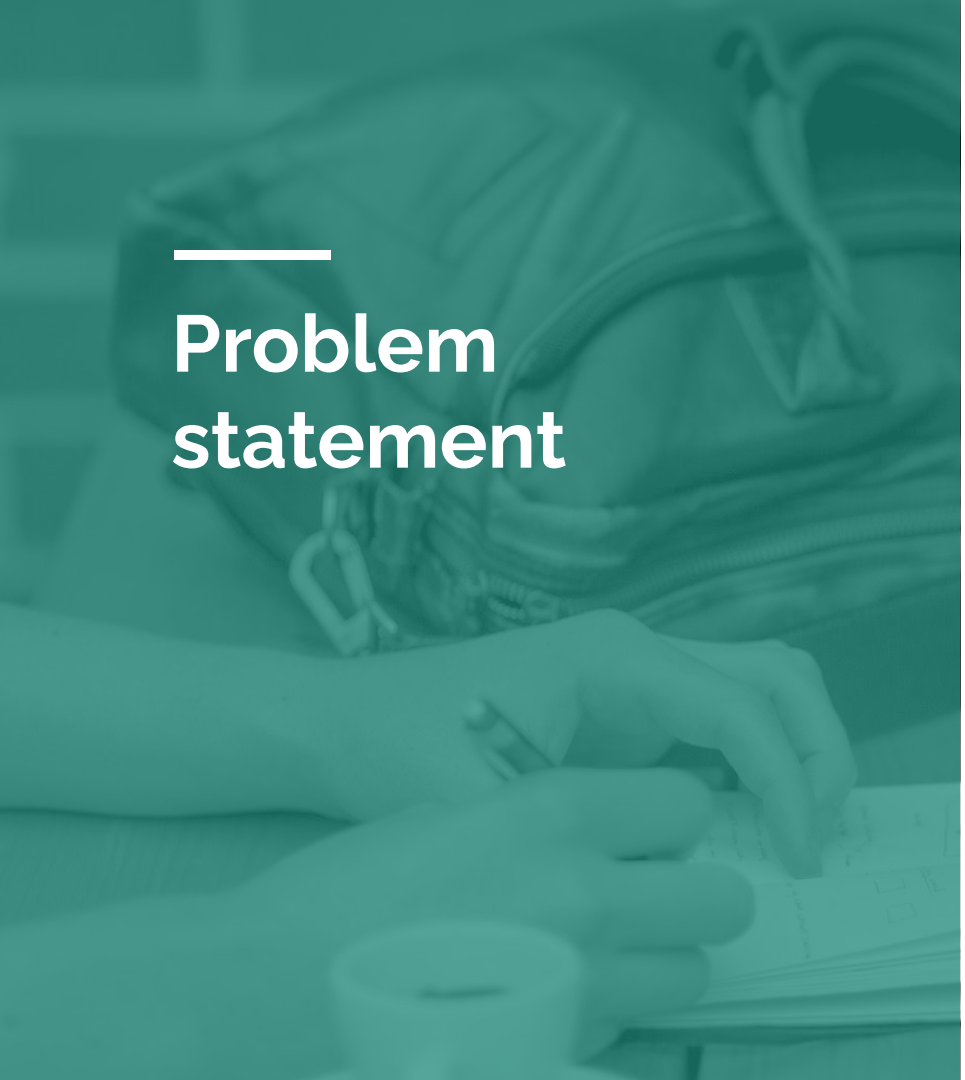
Beer Data Science Project

Name: Abhinandan Pise
Email: abhipise5321@gmail.com



Problem statement

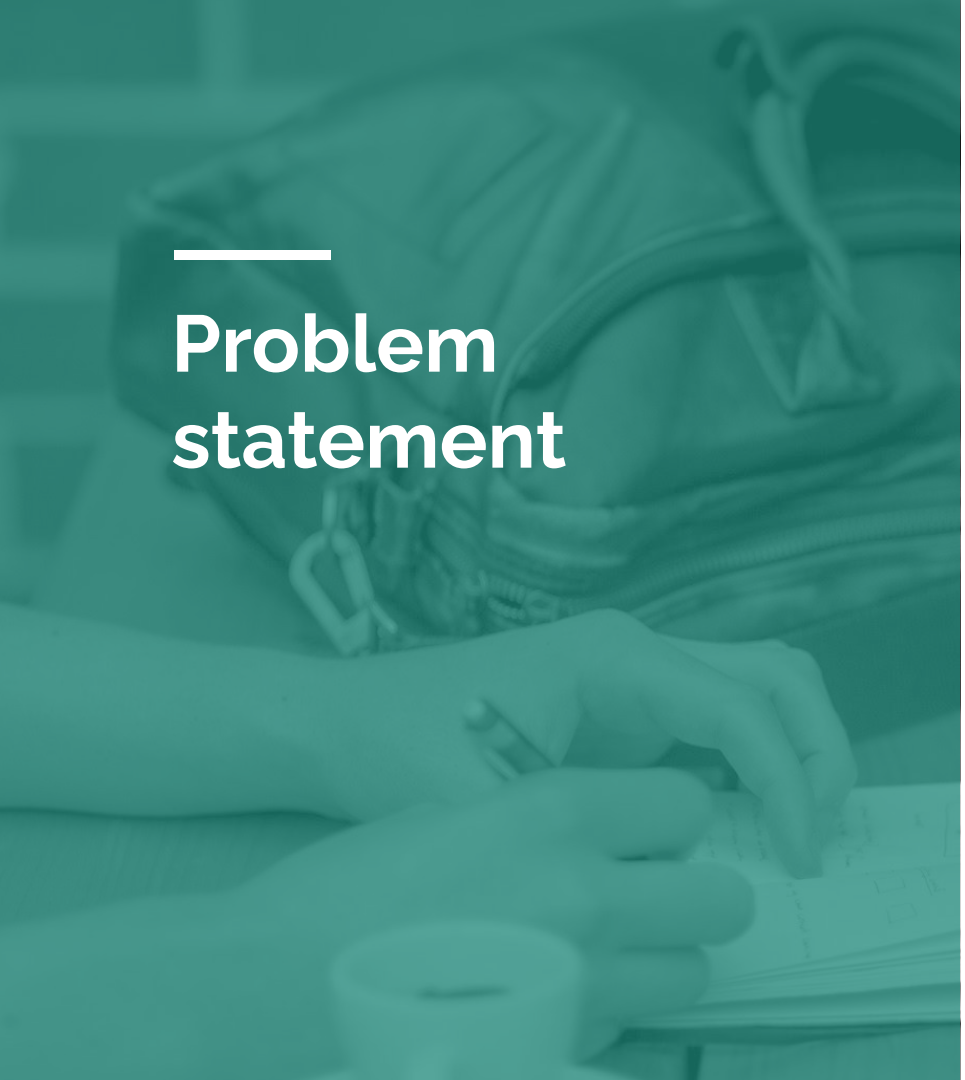
Assessment Questions.

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?



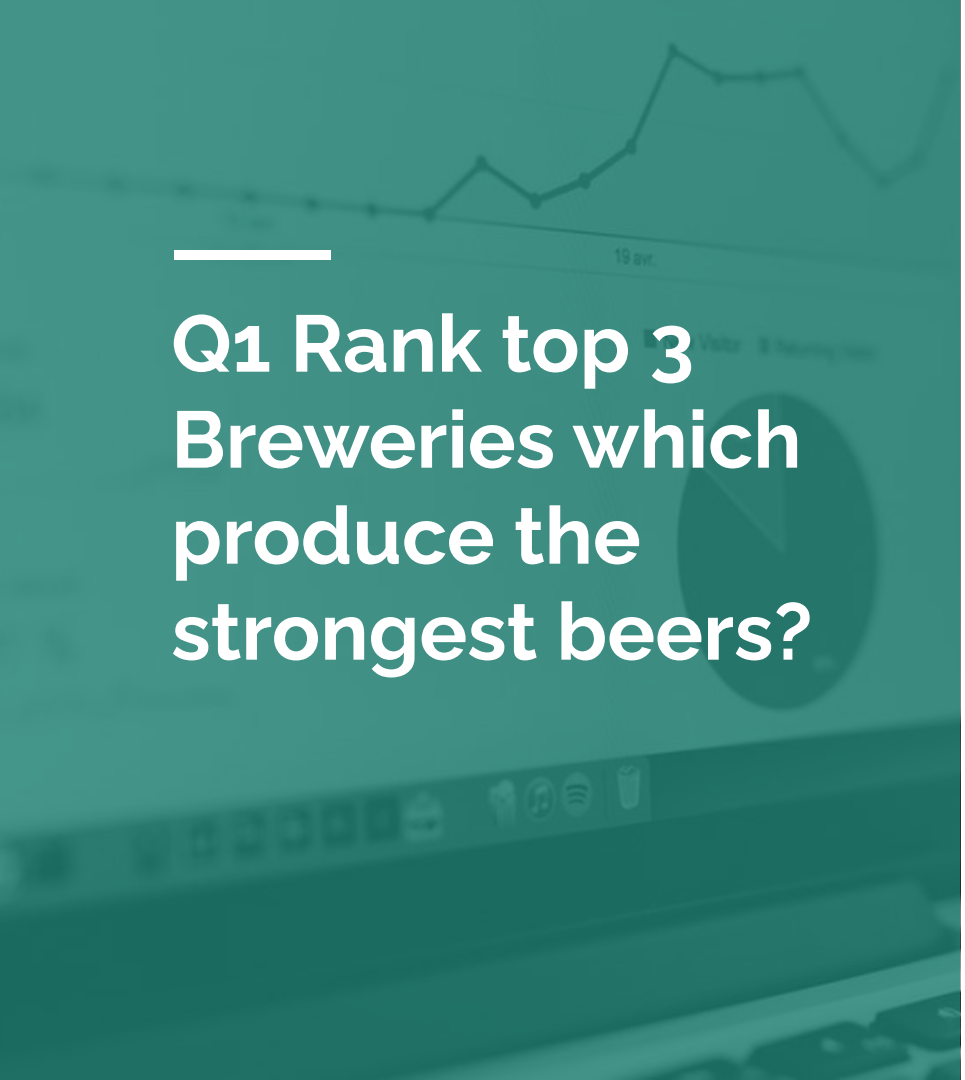
Problem statement

Assessment Questions.

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?



Q1 Rank top 3 Breweries which produce the strongest beers?

Approach

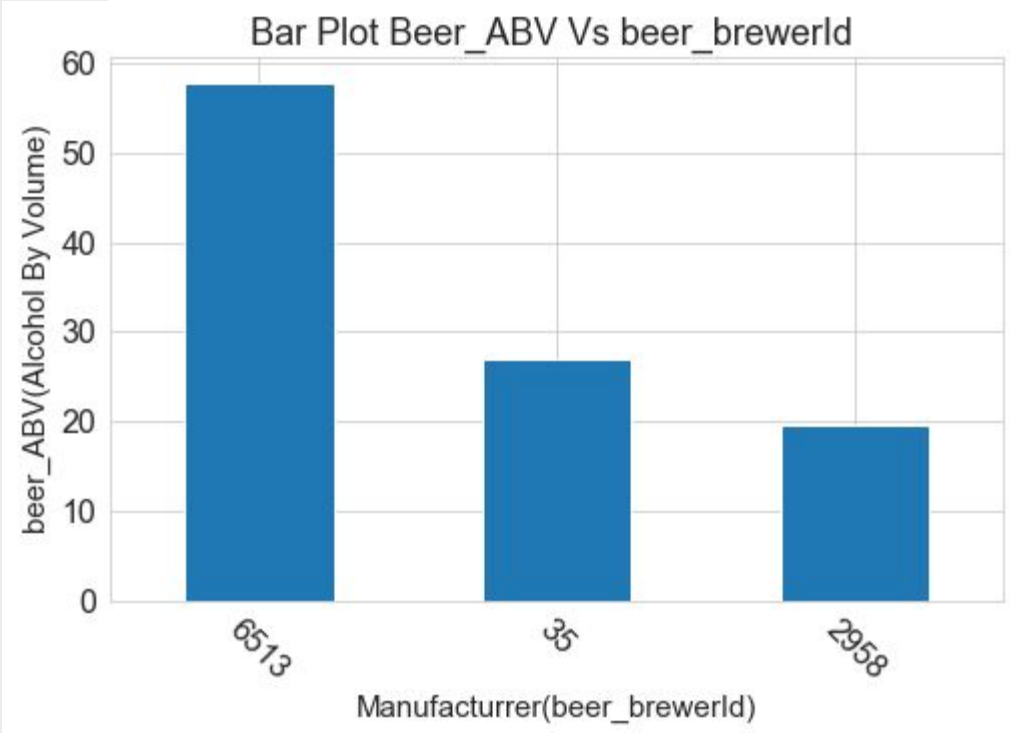
Pandas **GroupBy** command along with Aggregation function **max** is used to tackle the problem.

Supporting information

Q1

Top 3 Breweries which produce the strongest beers

1. beer_brewerId:**6513**
2. beer_brewerId:**35**
3. beer_brewerId:**2958**



Q2 Which year did beers enjoy the highest ratings?

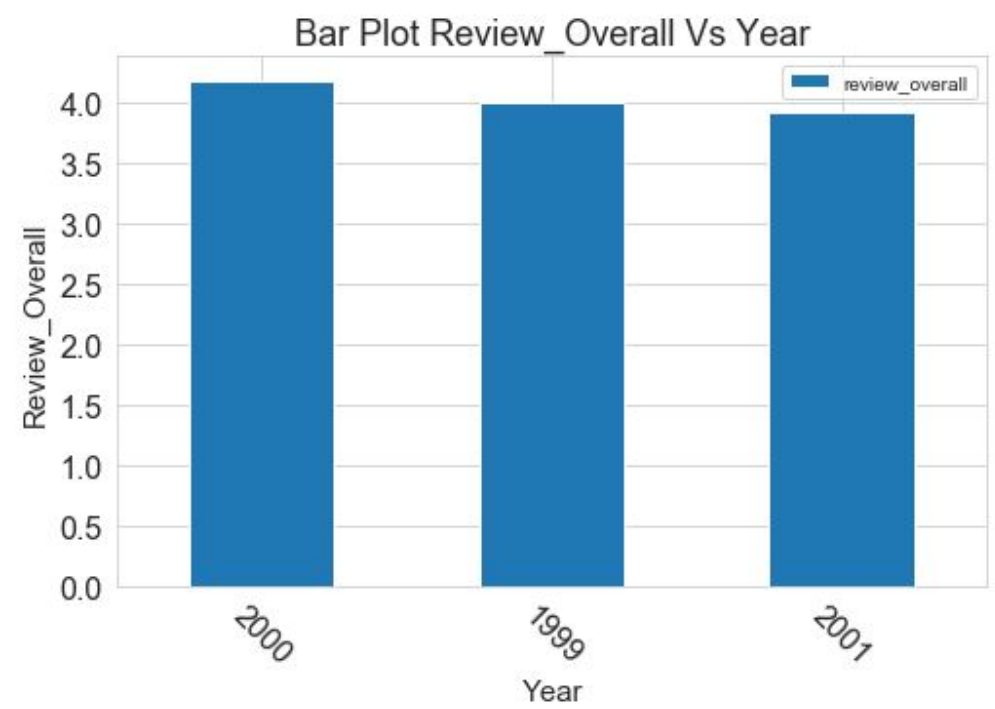
Approach

Pandas **GroupBy** command along with Aggregation function **Mean** is used to tackle the problem.

Supporting information

Q2

Year 2000 enjoy the highest overall ratings with mean of **4.18** ratings.



Q3 Which year did beers enjoy the highest ratings?

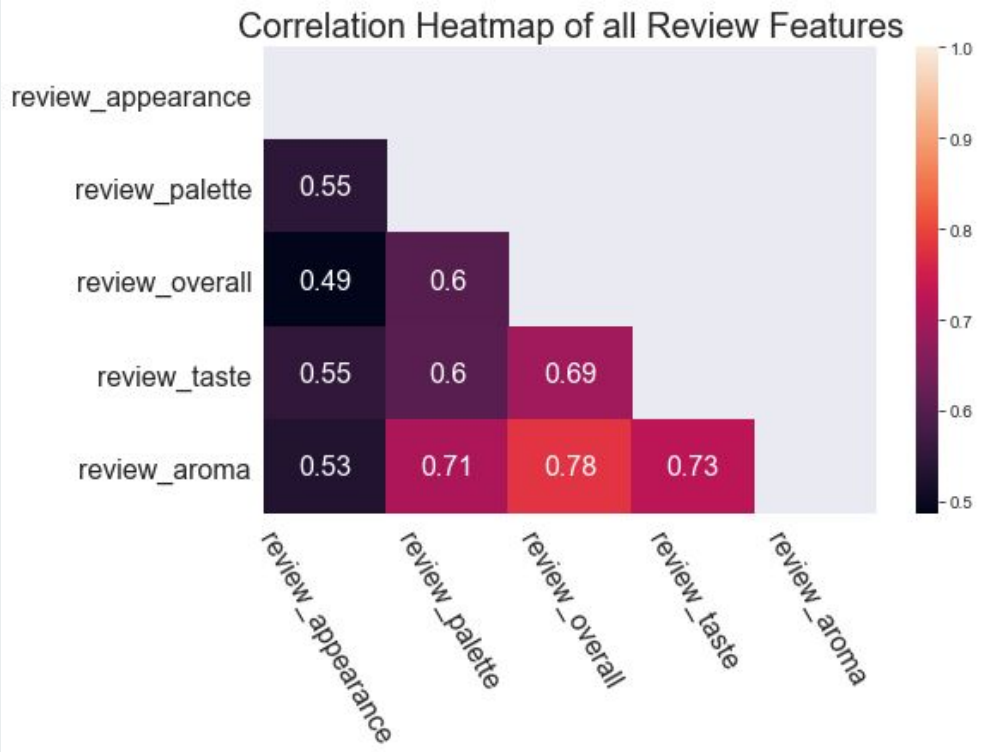
Approach

Pandas **Correlation** command along with Seaborn **Heatmap** is used to tackle the problem.

Supporting information

Q3

Review_overall & **Review_Aroma** are most correlated features with correlation of **0.78**



Q4 If you were to recommend 3 beers to your friends based on this data which ones will you recommend?

Approach

Pandas **GroupBy** command along with aggregation function **mean** is used to tackle the problem by considering all the reviews & BeerABV(Alcohol by Volume) Features

Supporting information

Q4

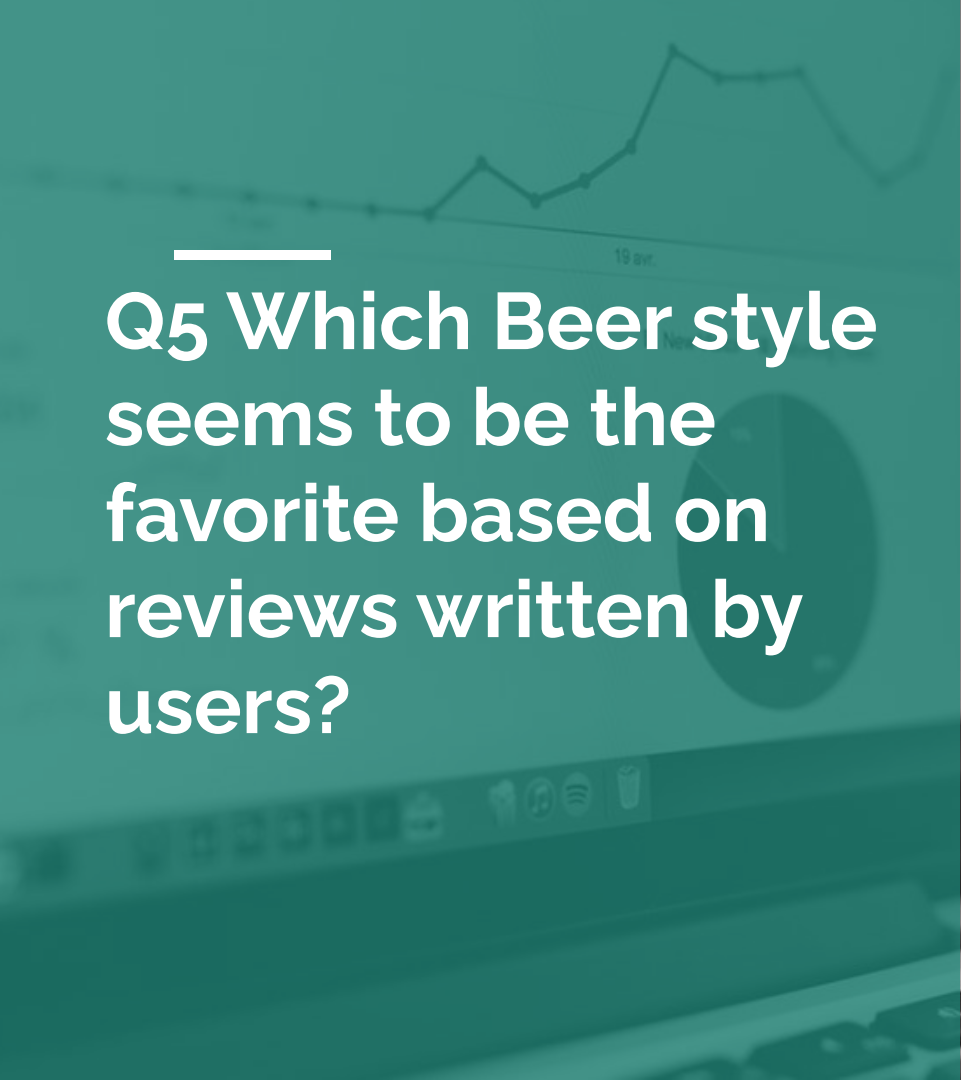
beer_name	review_overall	review_aroma	review_taste	review_appearance	beer_ABV
Edsten Triple-Wit	5.0	5.0	5.0	5.0	10.0
Old Gander Barley Wine	5.0	5.0	5.0	5.0	9.5
Rogue Black Brutal	5.0	5.0	5.0	5.0	9.0

Based on the all Reviews & Alcohol by Volume(Beer_ABV) will recommend the beer to friends as follows:

a. Edsten Triple-Wit

b. Old Gander Barley Wine

c. Rogue Black Brutal



Q5 Which Beer style seems to be the favorite based on reviews written by users?

Approach

- ➔ Based on review_text after doing preprocessing on text data, Vader Sentiment Analyzer is used.
- ➔ Then Pandas **GroupBy** command along with aggregation function **mean** is used to find out the **Best Beer Style** based on **Sentiment of Review Text**.

Supporting information

Q5

For the **Braggot** beer_style, the average sentiment score is max i.e. 0.863941

beer_style	text_sentiment_score
Braggot	0.863941
Quadrupel (Quad)	0.863022

Q6 How does written review compare to overall review score for the beer styles?

Approach

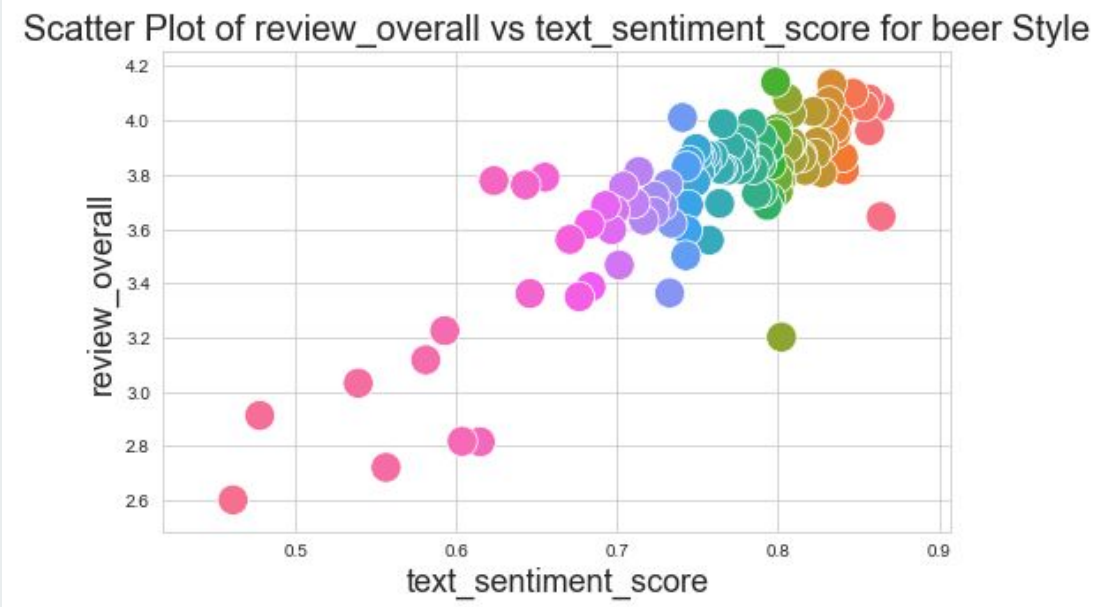
- ➔ Then Pandas **GroupBy** command along with aggregation function **mean** is used, features used **Beer Style**, **Sentiment of Review Text, Review Overall.**
- ➔ Then **Correlation** is found out on feature review_overall & text_sentiment_score

Supporting information

Q6

Correlation of **Review Overall** between Review_Text(**Sentiment_Score**) based on the Beer_Style:**0.82657**

Linear Relationship can be observed between two features **Review Overall** & Review_Text(**Sentiment_Score**)



Q7 How do find similar beer drinkers by using written reviews only?

Approach

- ➔ Cleaned Text is Vectorized using BoW Technique
- ➔ The Cosine Similarity is calculated between Query point & top 5 similar Text vectors

Note:We can use Semantic Based Text Vectorizer for Review text like W2V or Sentence Transformers & to reduce the time Complexity FAISS can be used.

Supporting information

Q7

For **beerId** of Query Datapoint : **33624** most **similar** datapoint is **beer_beerId** : **47647** & beer_name: **YuleSmith (Winter)** with Cosine Score of : **0.6245**