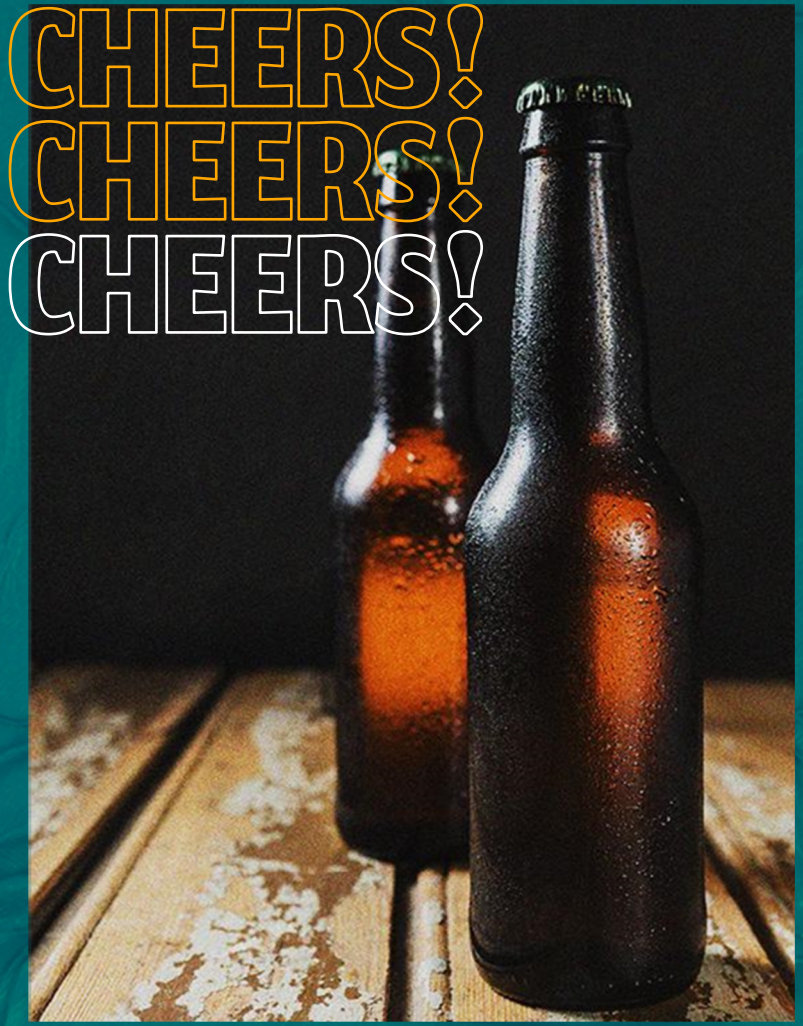
# BREWED AWAKENING



Which one do you think would like?

What do you think influences your choice?

# Table of Contents


01 Introduction

- Data Set
- Questions of Interest

02 Data Processing

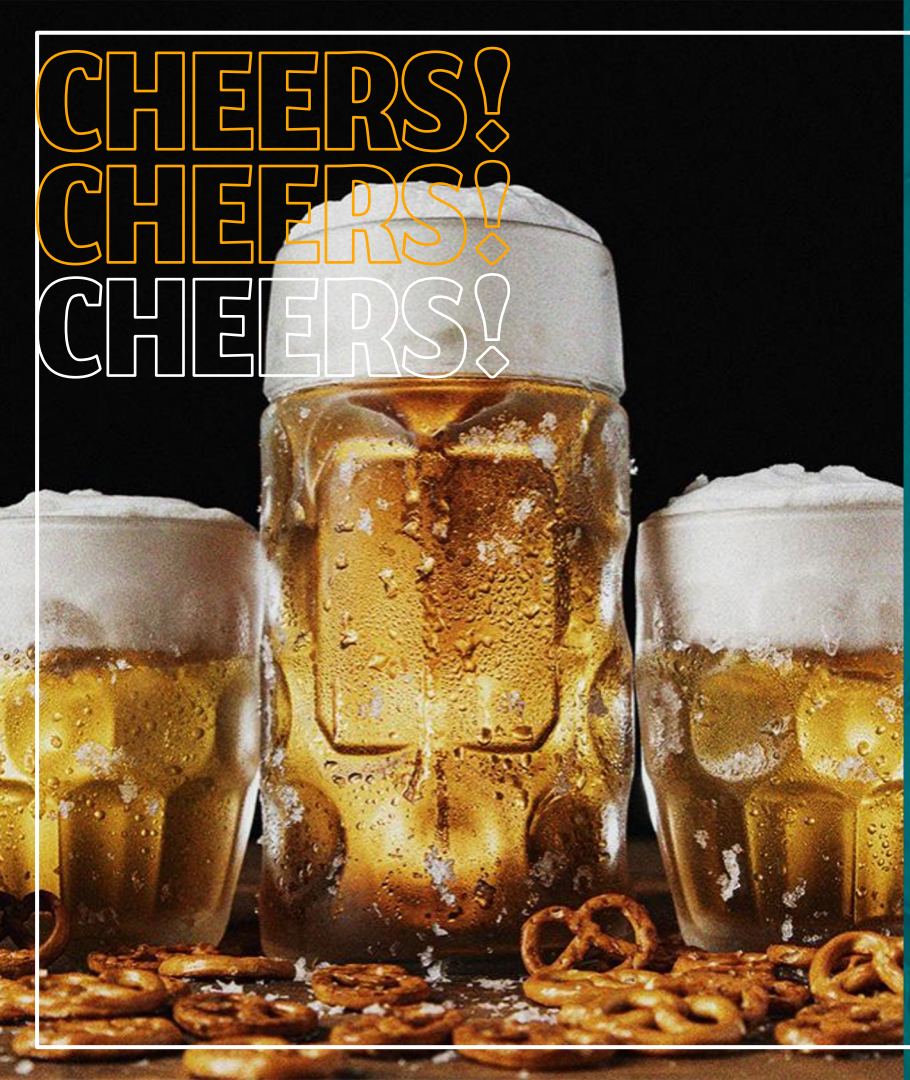
- Data Cleaning
- Data Clustering

03 Data Analysis

- Correlation Heatmap
- Distribution of Review Scores
- Distribution of Beer Style

04 Conclusion

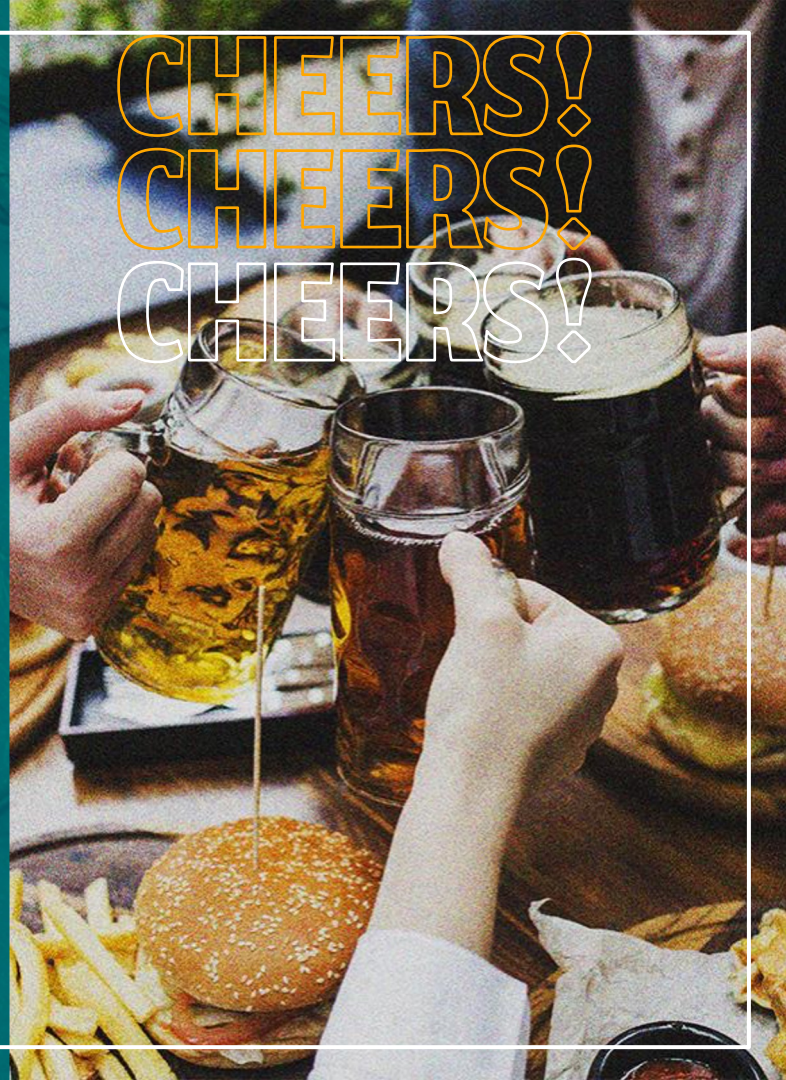
- Summary
- Significance
- Future Scope




# 01

# Introduction

Imagine this: the annual consumption of **beer** by the average person is an astonishing 28.2 gallons. Yes, you read that right - **28.2** gallons! Now, picture a world where 500 different kinds of beer exist, each with its distinct flavour, style and brew.



# Data Set

## 'Beer Reviews from Beer Advocate'(Kaggle)

- 1.5 million observations of 13 different features obtained from 1996 to 2012
- Beer Advocate, a blog site, is the source of the data that allows users to rate and review beers.
- The aroma, appearance, palate, taste, and overall impressions of the beer are rated on a scale of 0 to 5 based on the user’s preferences.
- User’s feedback on approximately 6,000 distinct breweries.

# Questions of Interest

1. Does alcohol content influence the overall beer rating? If yes, how?

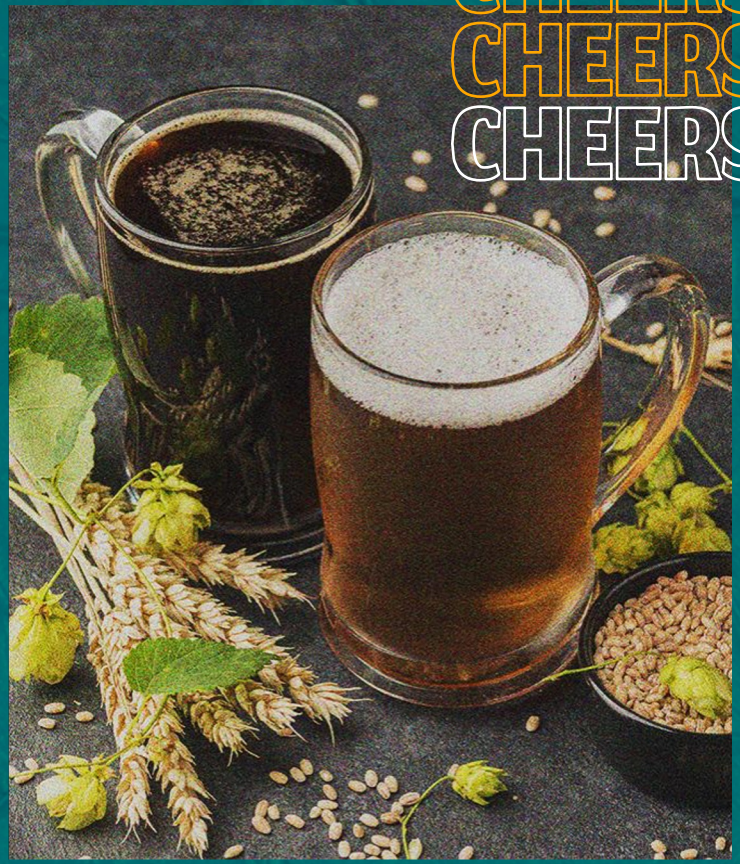
1. Which beer styles consistently resonate with consumers?

1. What features significantly contribute to a beer's success as a brand?

1. Can we pinpoint factors leading to poor reception of a beer brand?

# 02

# Data Processing




Data Cleaning
Data Clustering

# 2.1 Data Cleaning

CHEERS!
CHEERS!
CHEERS!

- ❑ Check for null values
- ❑ Treat missing values
  - Replacing missing 'brewery_name' and 'review_profilename' values with 'unknown'
  - Imputing for null values in 'beer_abv'
- ❑ Correct the data type and format of 'review_time' column
- ❑ Handle duplicate reviews
  - Take the latest review made by that person which would reflect their most recent take on the beer

# 2.2 Processing for Clustering

1

Label encoding the categorical variables

2

Scaling numerical features

3

Determine number of clusters

# Classifications

Quality of a beer:

- Poor: Ratings below 3
- Decent: Ratings between 3 and 4
- Good: Ratings between 4 and 4.5
- Best: Ratings of 4.5 and above

Alcohol Strength:

- Low: Alcohol by volume (ABV) ranging from 0% to 5%
- Mild: ABV between 5% and 15%
- Strong: ABV between 15% and 30%
- Extreme: ABV of 30% and above

# Profiling of Clusters

## Cluster 0

3rd in overall, appearance, palate, taste scores

3rd in alcohol by volume %

Mostly consists of beers that can be considered decent

## Cluster 1

2nd in overall, appearance, palate, taste scores

2nd in alcohol by volume %

Well received by the consumers on average. High proportion of decent beers, brings average score down

## Cluster 2

4th in overall, appearance, palate, taste scores

4th in alcohol by volume %

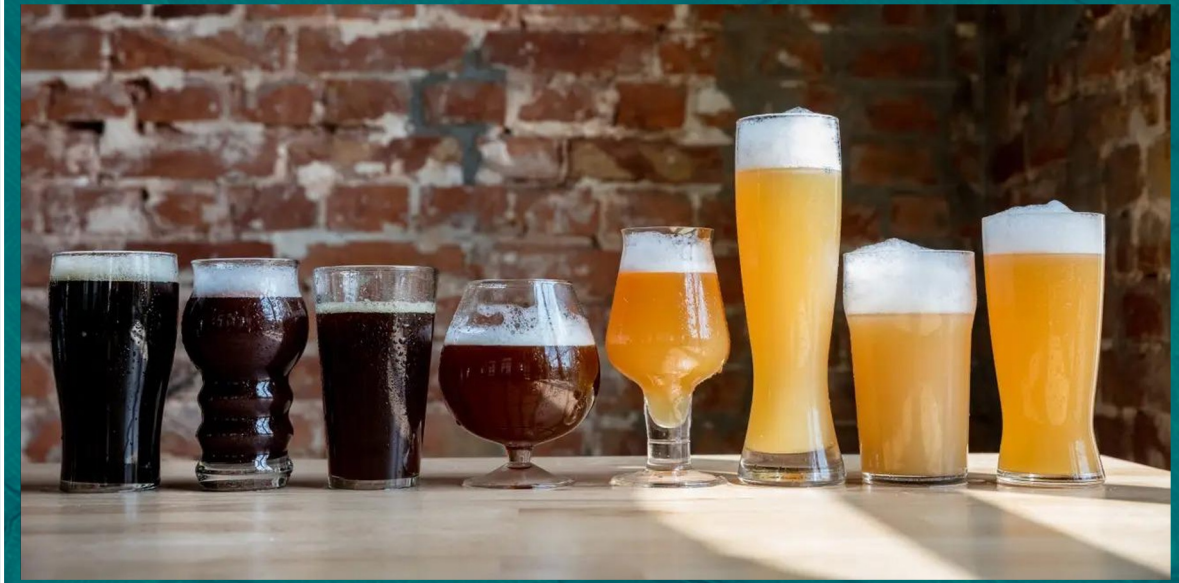
Mostly consists of beers that were considered poor

## Cluster 3

1st in overall, appearance, palate, taste scores

1st in alcohol by volume %

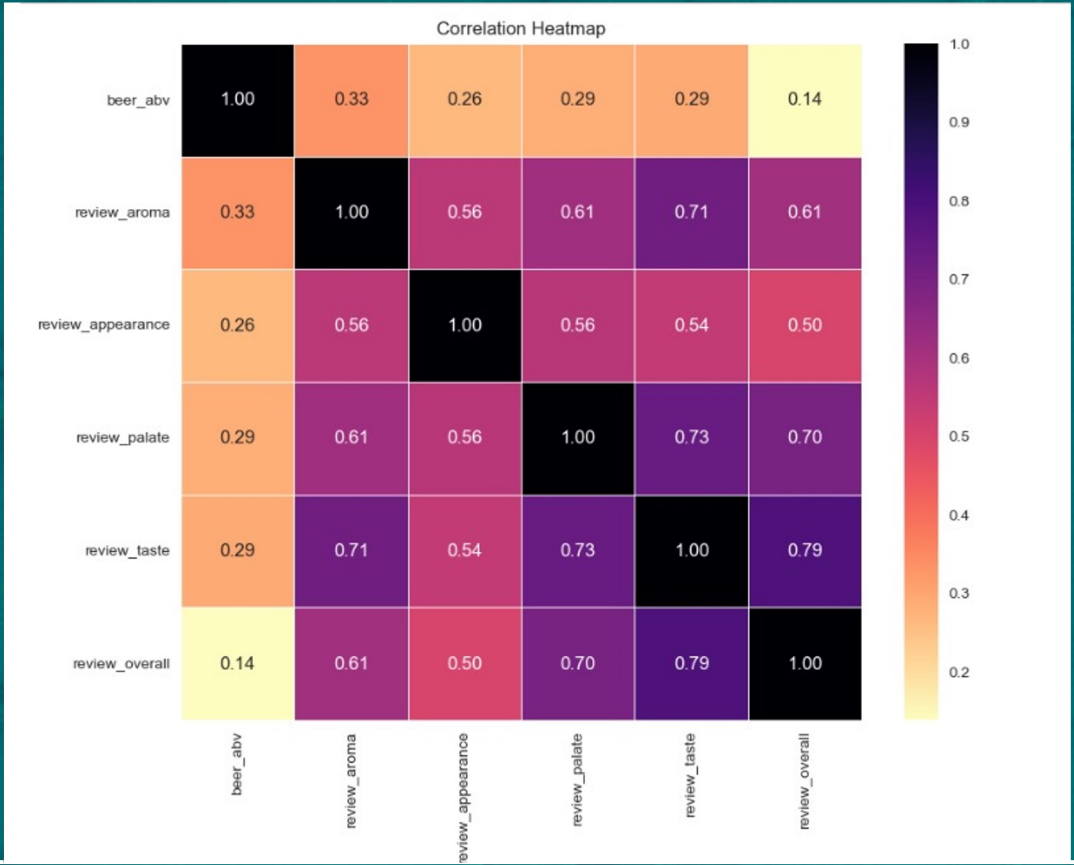
These are beers that were most well received by the consumers on average



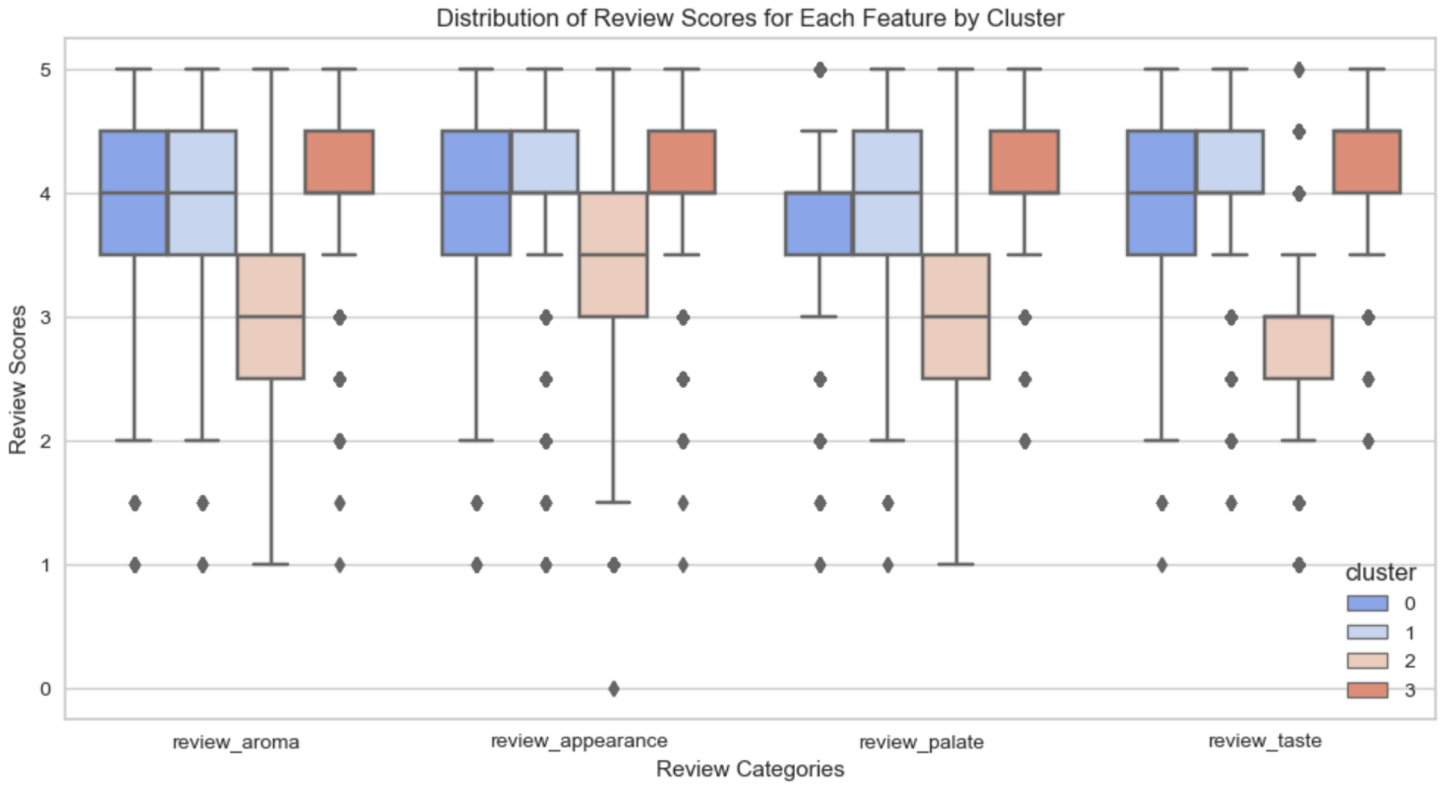
# 03

# Data Analysis

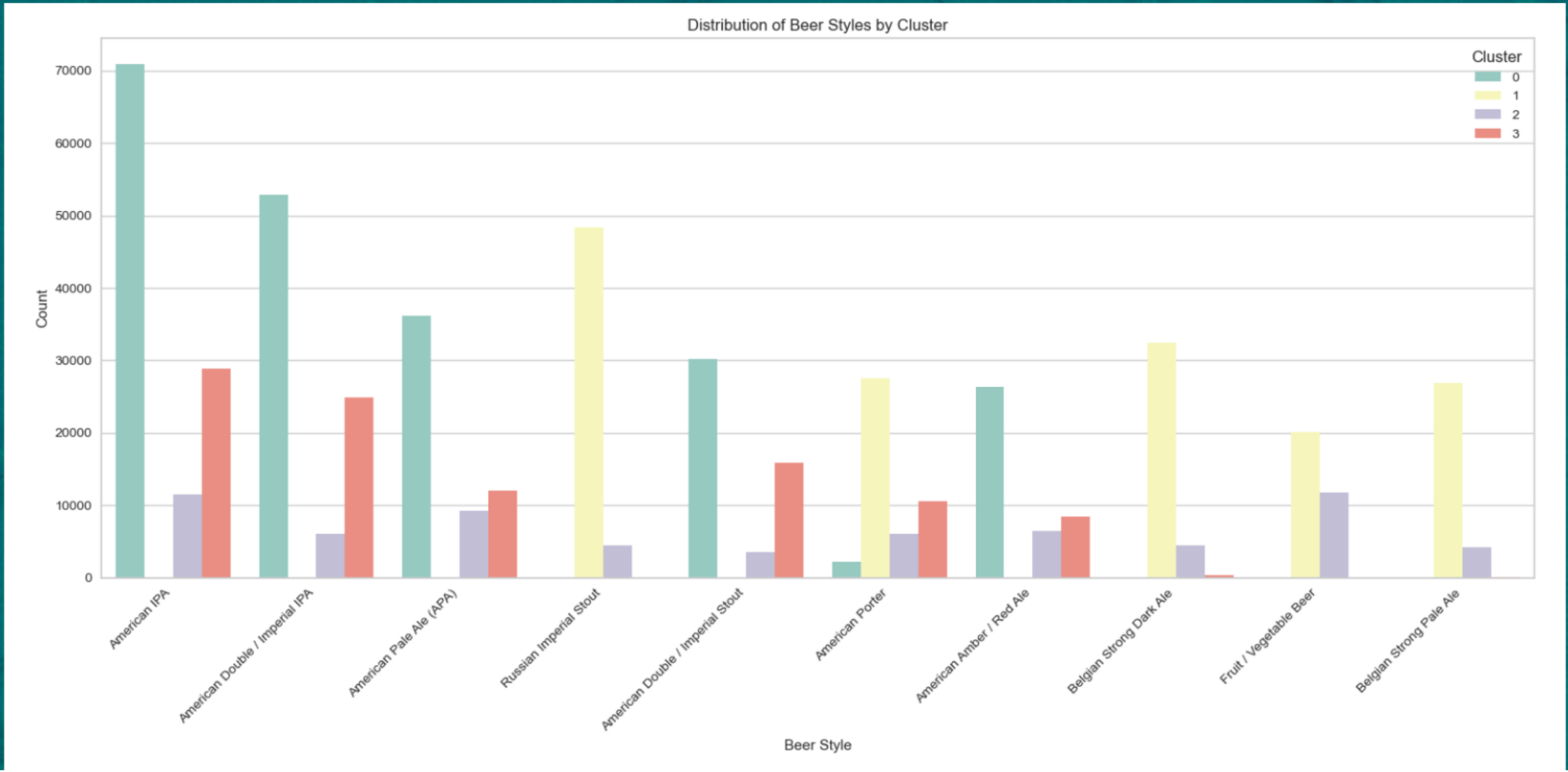
# Correlation Heatmap

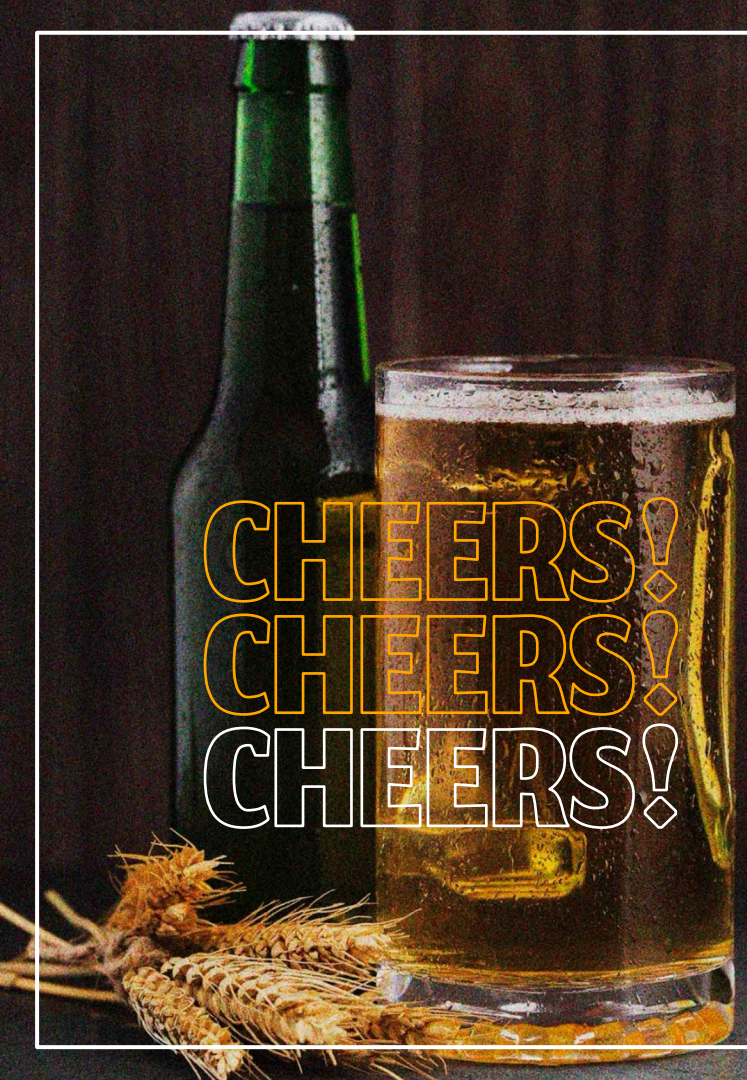


# Distribution of Review Scores for Each Feature by Cluster



# Distribution of Beer Styles by Cluster






# 04

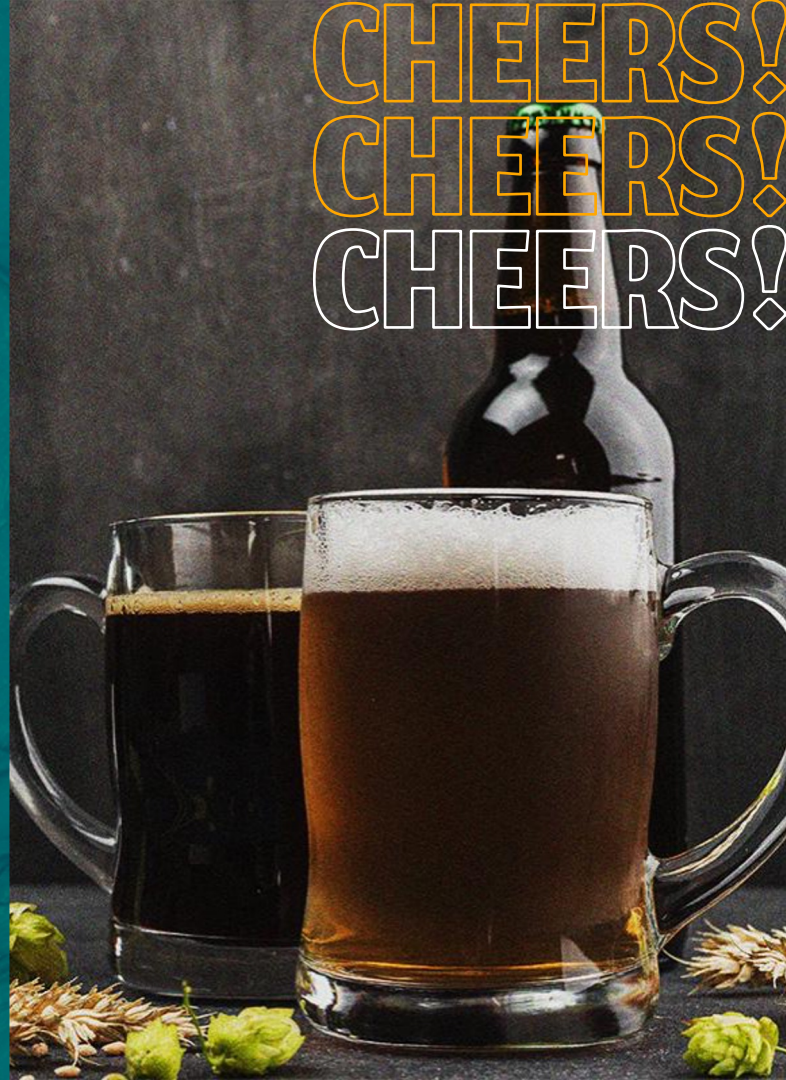
# Conclusion

# Summary

- Identify the extent to which each attribute of a beer affects its reception by consumers

- Identify the beer styles that resonated consistently with them and how the characteristics varied across segments of beers

- Provide insights regarding their success and failure in the market

# THANKS!

QUESTIONS?