

// Spring 2021

COGS 118B

FINAL PROJECT

Analysis of Beer Reviews as
Predictors of Beer Type

K-MEANS!
PCA!
BEER!

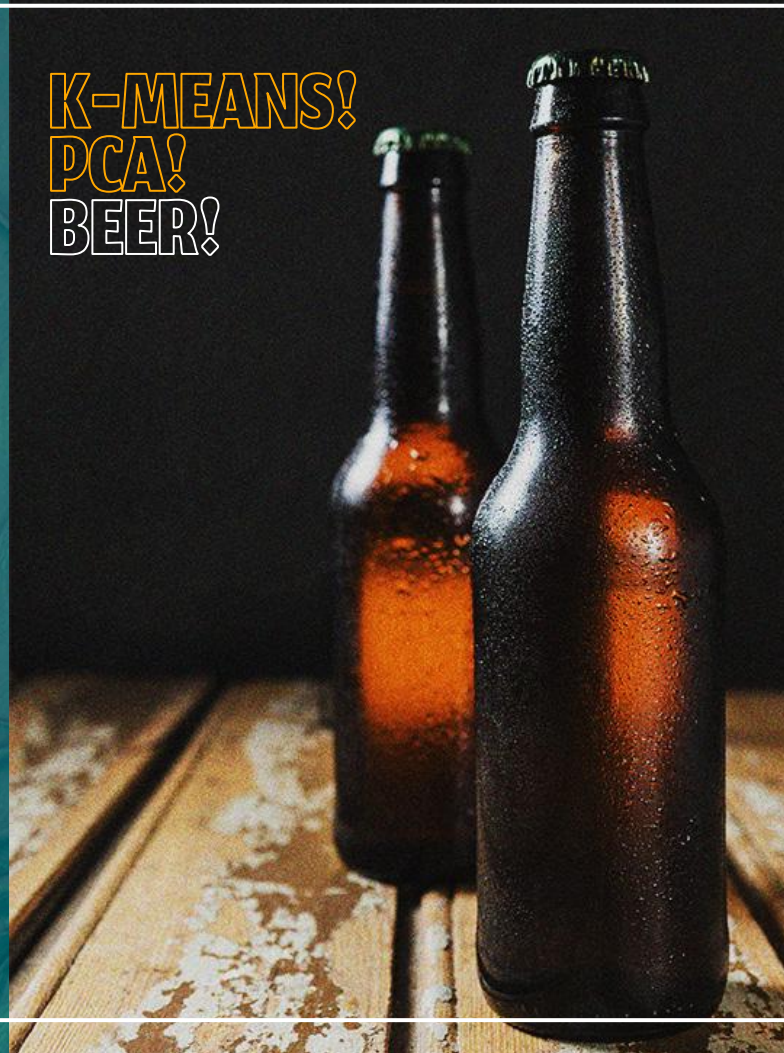


TABLE OF CONTENTS

CHEERS!
CHEERS!

01

INTRODUCTION

Why we chose beer, and
what we wanted to do

02

METHODS

How we analyzed our beer
dataset

03

RESULTS

What we could conclude
from our analysis

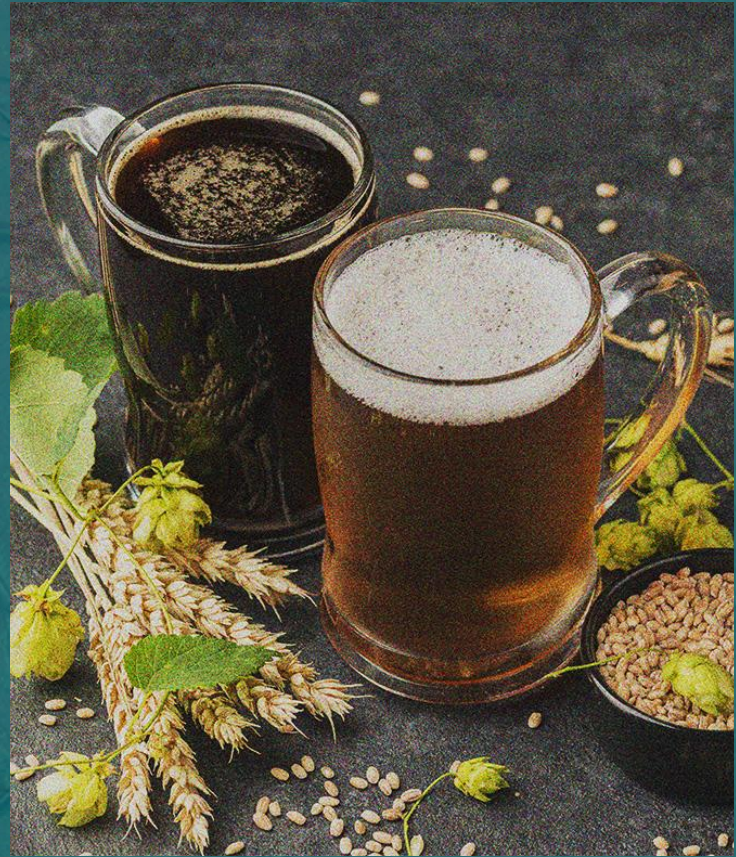
04

DISCUSSION

Limitations with our
analysis and next steps

OUR MOTIVATION

- Thousands of years of history
- Brewing method informs taste, ABV, color, hoppiness and more
- Lots of review data online
- Gives insight into most valued features in types of beer



RELATED WORK



PCA + K-means

Class lectures aided
in application of PCA
and implementation of
K-means grouping



RoboBEER

Beer quality assessed
through ML/CV-aided
color, foamability and
foam stability

CHEERS!
CHEERS!



Category	Percent of Reviews
Dark ale	20.68%
Pale ale	28.85%
Lager	15.01%
Stout	16.10%
Other	19.36%

METHODS: Data Cleaning

First, we split 100+ unique beer styles into 5 major categories:

- Dark Ales (1)
- Pale Ales (2)
- Lagers (3)
- Stouts (4)
- Other (0)

This split the beer by brewing method, separating the data by color & ABV

METHODS: Review Data

TASTE



AROMA

PALATE

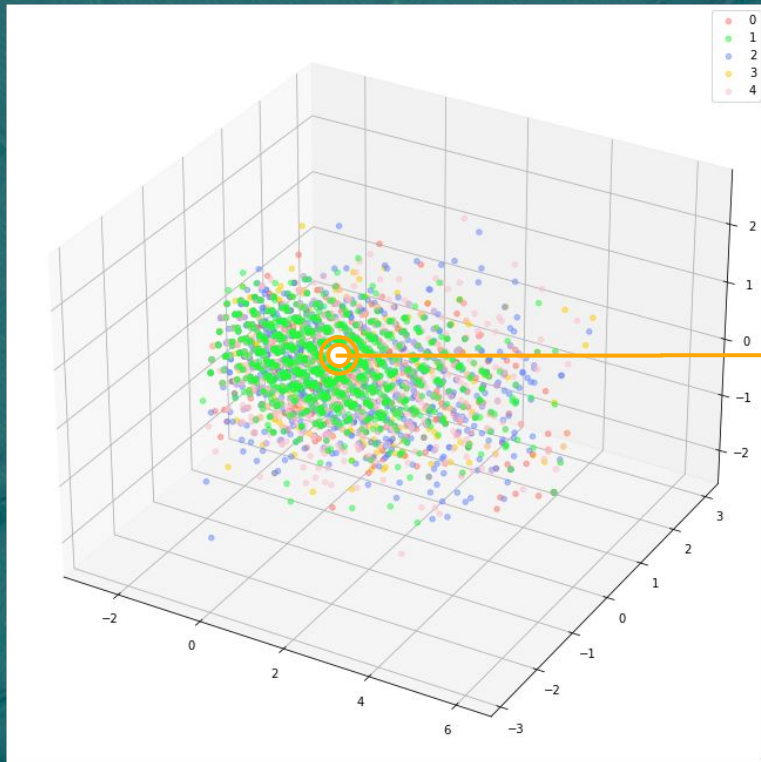


OVERALL SCORE



APPEARANCE

METHODS: Principal Component Analysis



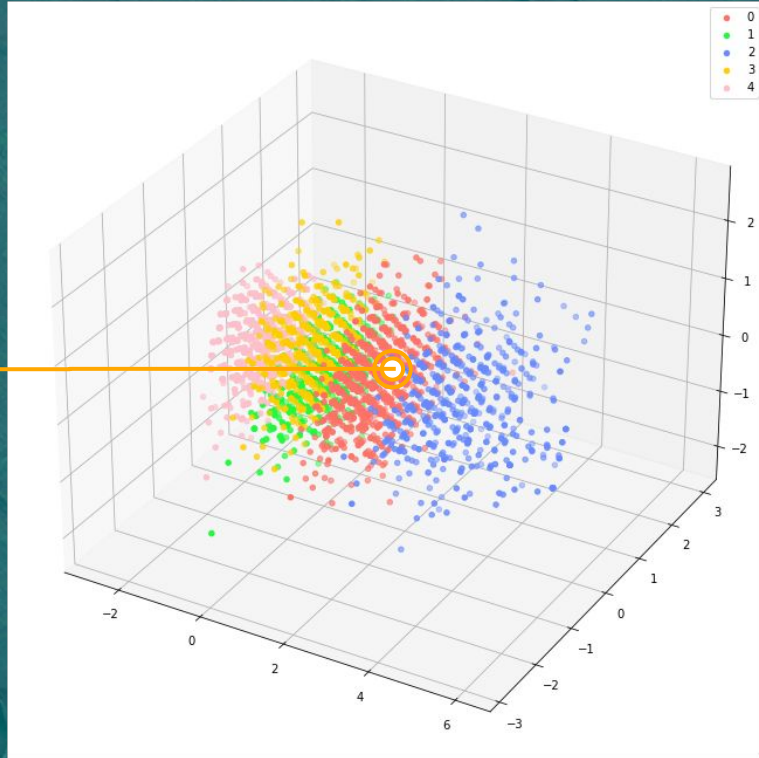
Using the sklearn package, we applied PCA to the cleaned dataset.

Unfortunately, each group has equal spread over the graph, with category 2 showing in the highest layer.

METHODS: K-means

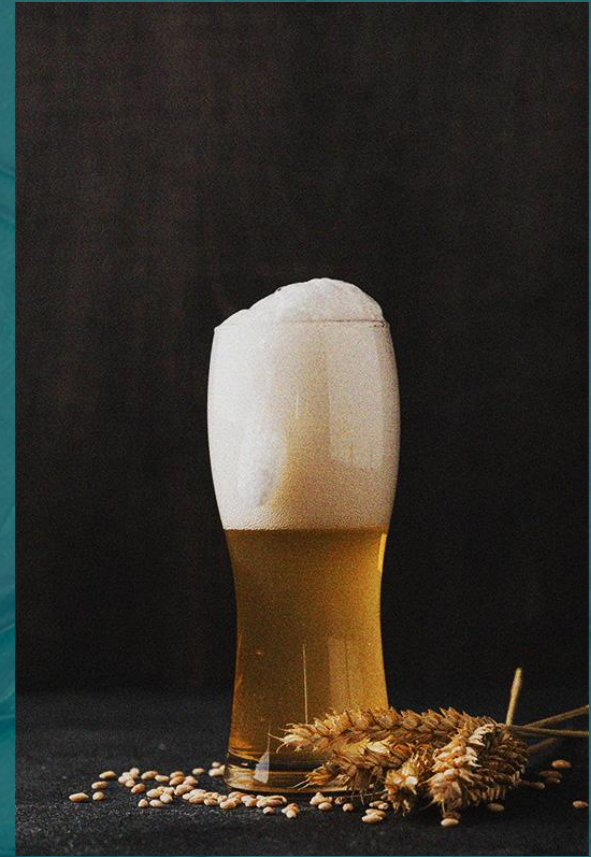
We implemented K-means from scratch, using lecture slides as a reference.

After applying the function to the PCA-processed data, it yielded the following graph.



RESULTS

- K-means clusters only matched with our 5 categories 20% of the time
- This implies random selection, demonstrating no correlation between beer reviews and beer category
- Utilizing PCA & K-means was not enough to uncover a relationship between reviews and beer
- This result is useful for brewers and sellers, allowing them to focus on the craft and marketing without worrying about the brewing method



DISCUSSION: Limitations



CLASSIFICATION

A more rigorous classification system could have improved groups



MORE FEATURES

More numerical features or encoding of extra categorical data could have led to more conclusive results



PLOT CHANGES

3-D plots are complex, and rotation features could help better interpret data

DISCUSSION: **Future Work**

- A recommendation system utilizing unsupervised learning could be ideal
- Users could receive recommendations for new beers they'd like based on reviews
- Brewers could analyze the most popular beers and styles to inform future brewing and marketing choices





CHEERS!
CHEERS!
CHEERS!

THANK YOU!