

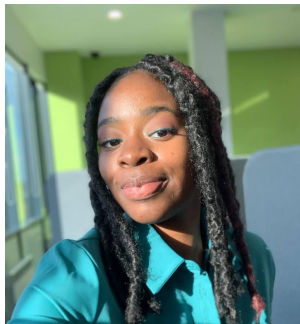
CRAFT BEER REVIEWS

Recommendation System



Our Team

Beeradvocate



**Ozioma
Okonicha**

Data Engineer

1+ year of
experience



**Abdulmueez
Emiola**

ML Engineer

2+ years of
experience

Business Problem

Beeradvocate

Determine the elements that influence
consumer satisfaction in the craft beer sector



Business Objectives

Beeradvocate



Find the most important attributes that consumers look for in a craft beer, such as taste, aroma, appearance, or abv



Which particular styles of craft beer are more popular among consumers



Use consumer reviews and ratings to predict which craft beers are most likely to be successful in the market



Process

Beeradvocate



Data Collection & Ingestion

Build a relational database in PostgreSQL and

Import the database into HDFS via Sqoop



Data Storage / Preparation & EDA

Create Hive tables and

Perform EDA

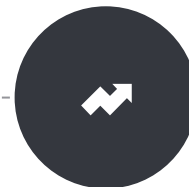


Predictive Data Analysis

Build Spark ML model,

Train and evaluate the model and

Hyperparameter tuning and Grid search



Presentation

Build a web dashboard in Streamlit

Descriptive Data Analysis

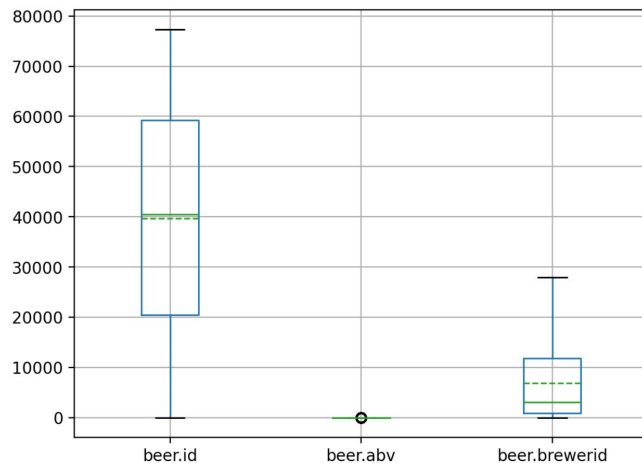
Data Characteristics

	Tables	Features	Instances
0	Beers	66,055	5
1	Brewers	5,840	2
2	Reviews	1,586,614	9
3	Persons	33,388	2

Descriptive Data Analysis

beers table

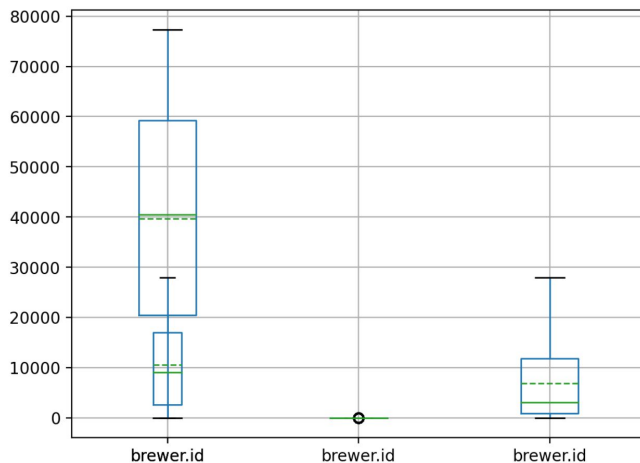
	beer.id	beer.abv	beer.brewerid
count	66,055	49,012	66,055
mean	39,718.1351	6.2719	6,855.2653
std	22,416.1778	2.0509	7,482.9543
min	3	0.01	1
25%	20,503.5	5	879
50%	40,506	5.7	3,073
75%	59,206.5	7.2	11,814
max	77,317	57.7	28,003



Descriptive Data Analysis

brewers table

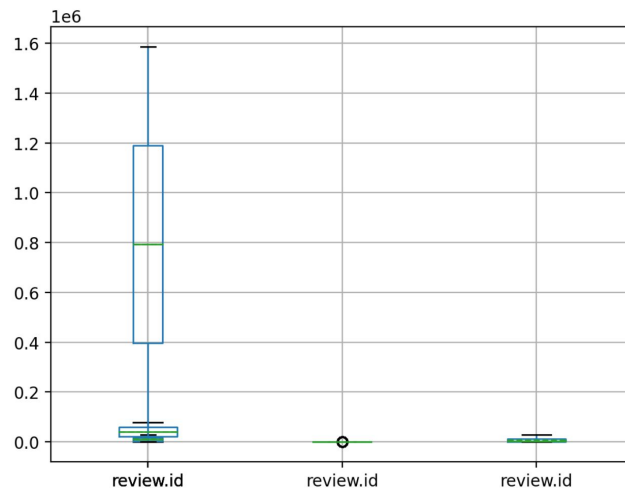
	brewer.id
count	5,840
mean	10,541.9015
std	8,496.0369
min	1
25%	2,641.75
50%	9,138.5
75%	17,032.25
max	28,003



Descriptive Data Analysis

reviews table

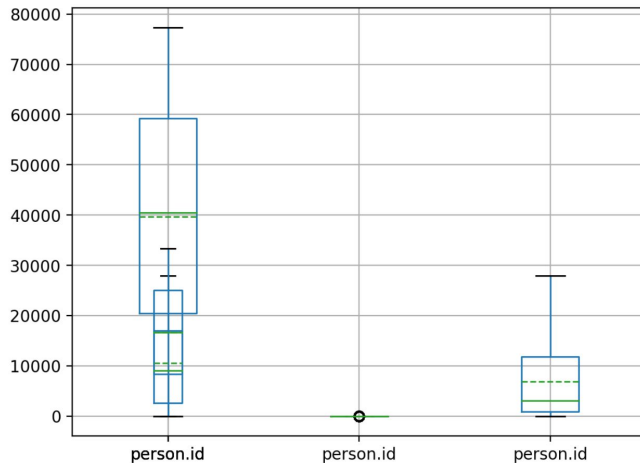
	review.id	↑ review.appearance	review.aroma	review.palate	review.taste	review.time
min	1	0	1	1	1	840,672,001,000
std	458,016.1543	0.6161	0.6976	0.6822	0.732	76,544,274,539.2376
25%	396,654.25	3.5	3.5	3.5	3.5	1,173,224,188,250
mean	793,307.5	3.8416	3.7356	3.7437	3.7929	1,224,089,280,012.2107
50%	793,307.5	4	4	4	4	1,239,202,881,500
75%	1,189,960.75	4	4	4	4.5	1,288,568,405,250
max	1,586,614	5	5	5	5	1,326,285,348,000
count	1,586,614	1,586,614	1,586,614	1,586,614	1,586,614	1,586,614



Descriptive Data Analysis

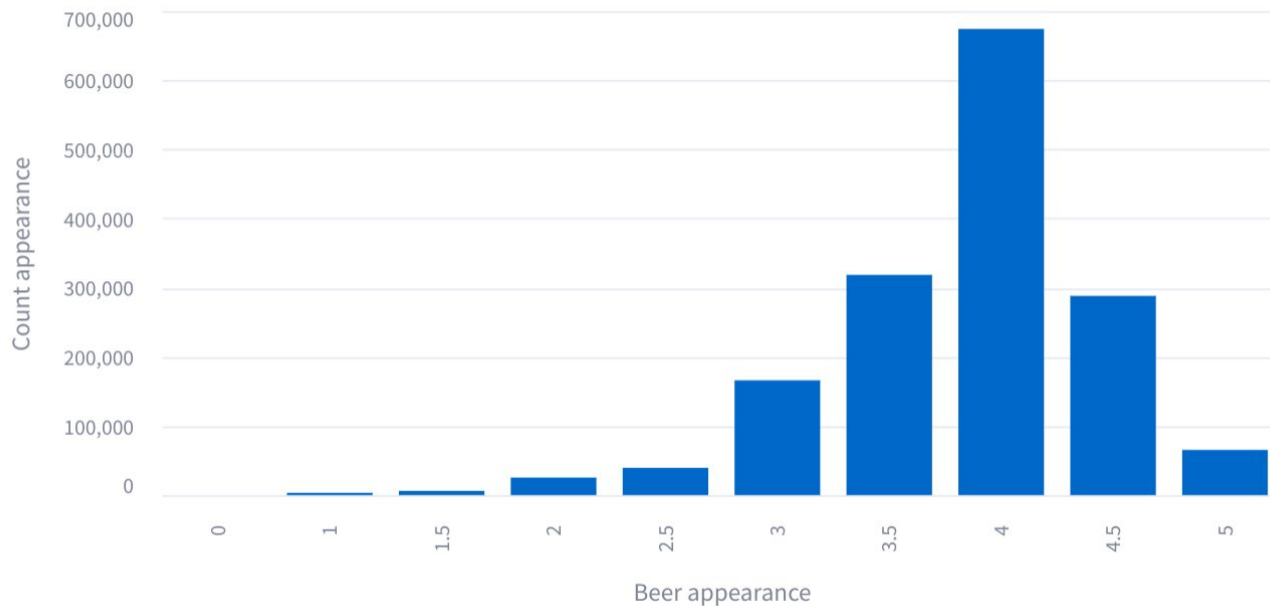
persons table

	person.id
count	33,388
mean	16,694.5
std	9,638.4297
min	1
25%	8,347.75
50%	16,694.5
75%	25,041.25
max	33,388



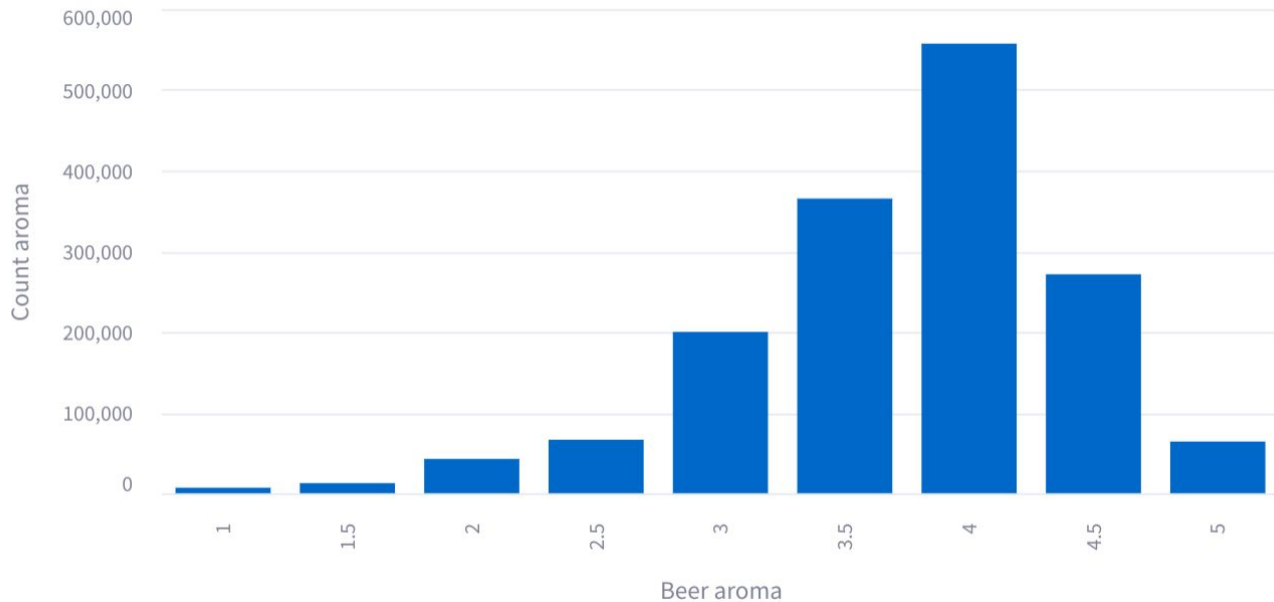
Exploratory Data Analysis

The number of beers for each appearance score



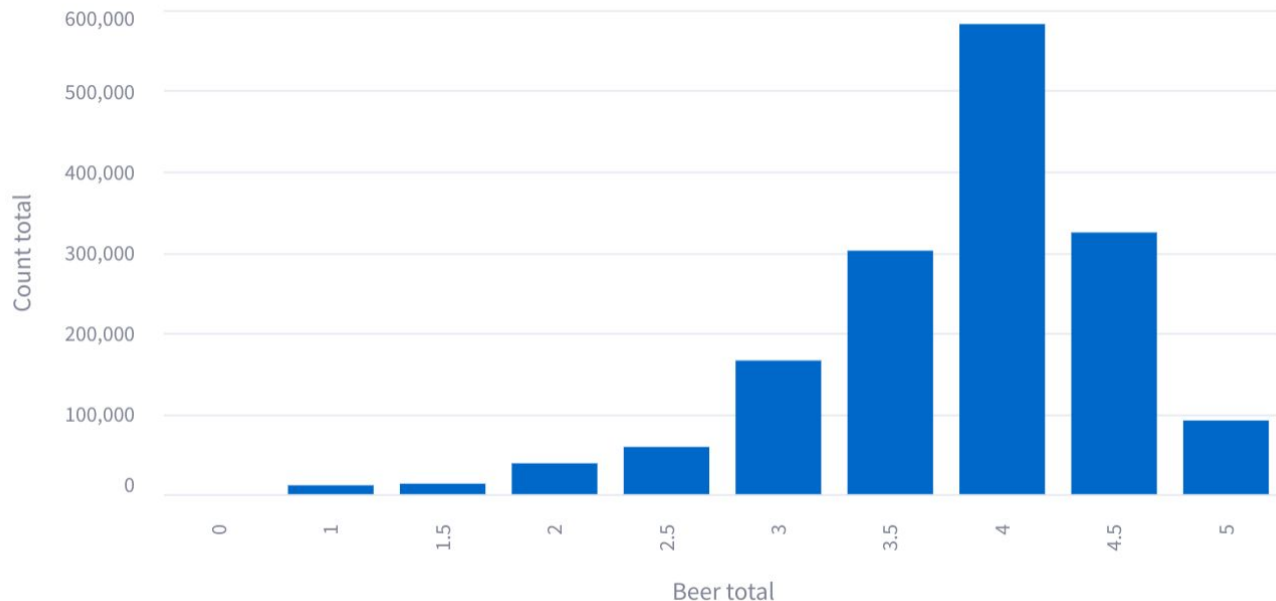
Exploratory Data Analysis

The number of beers for each aroma score



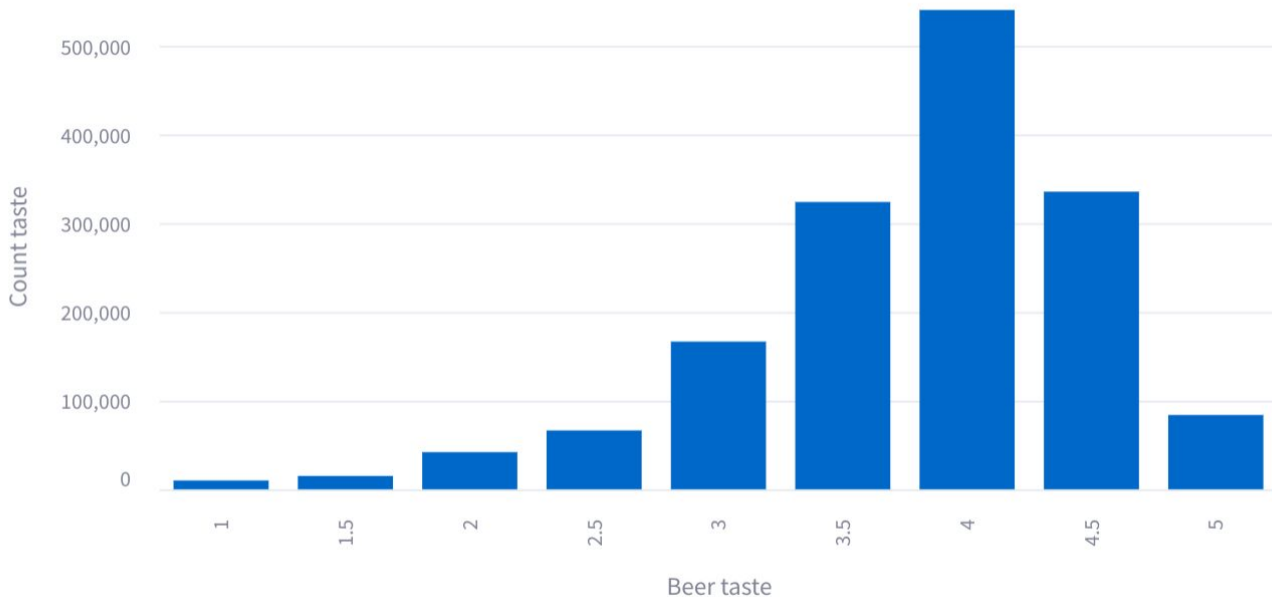
Exploratory Data Analysis

The number of beers for each total score



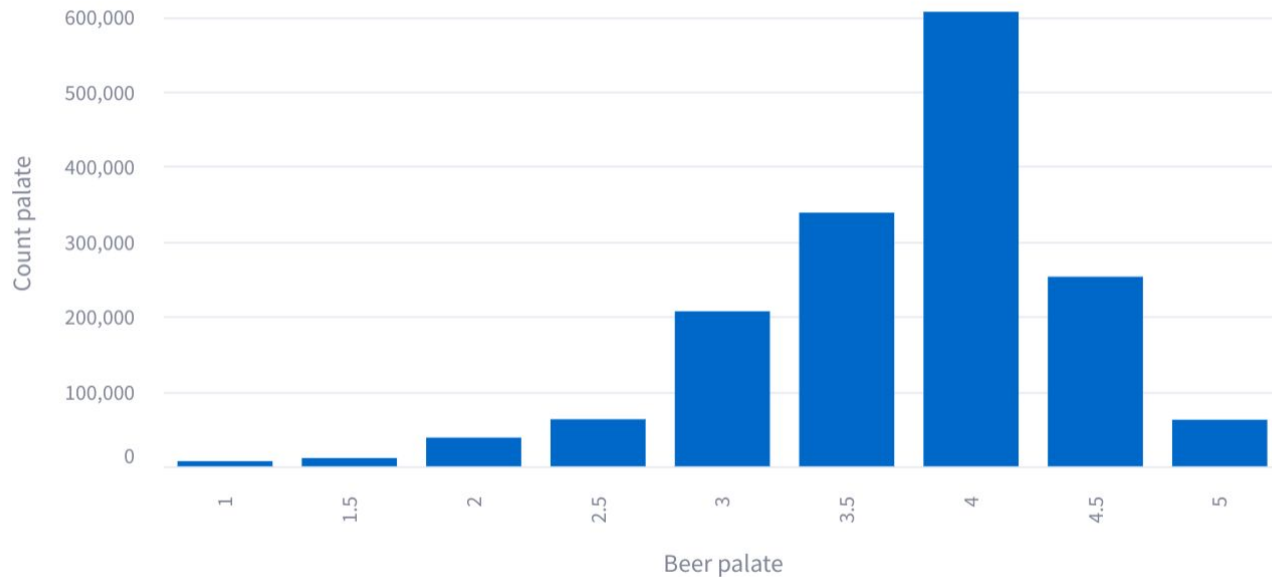
Exploratory Data Analysis

The number of beers for each taste score



Exploratory Data Analysis

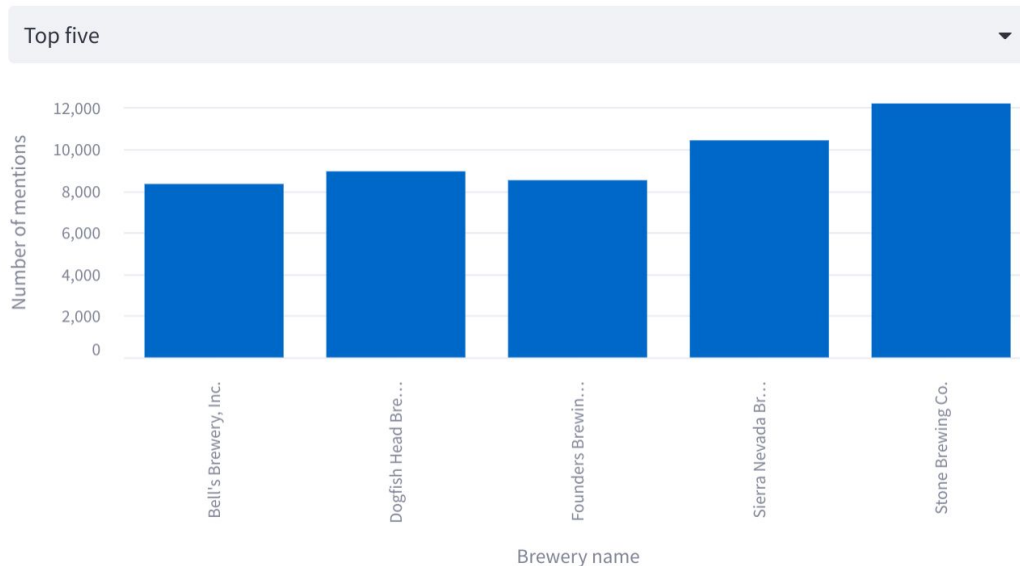
The number of beers for each palate score



Exploratory Data Analysis

The breweries which are mentioned most often in positive reviews (aka >4)

Choose the style of comparison



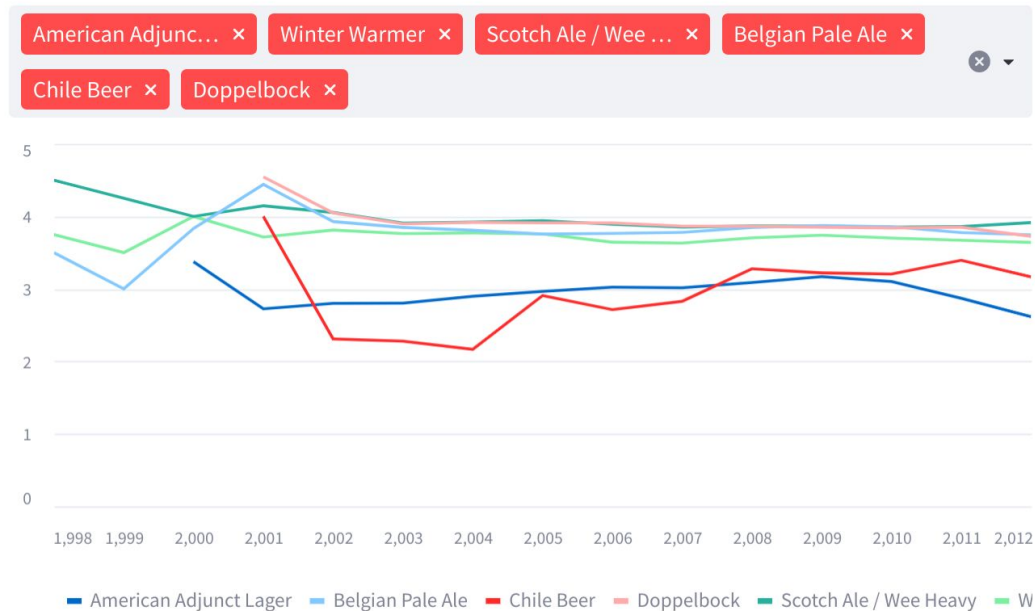
Hyperparameters & models performance

	LR
RMSE	0.656
maxIter	10
ElasticNetParam	0
regParam	0.01

Exploratory Data Analysis

How the popularity of different beer styles changed over years

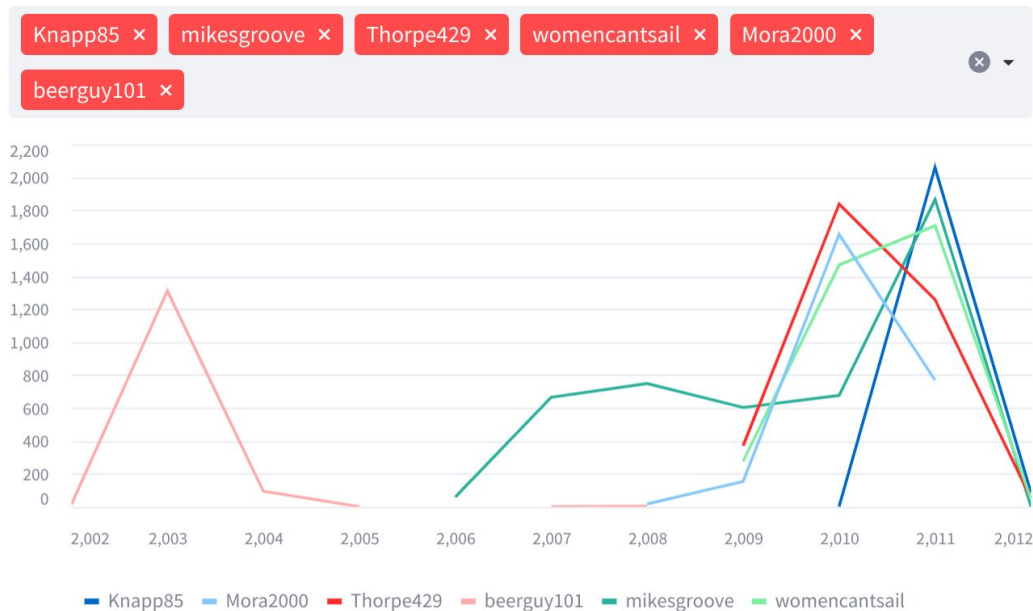
Select Styles



Exploratory Data Analysis

The number of reviews per year of active users

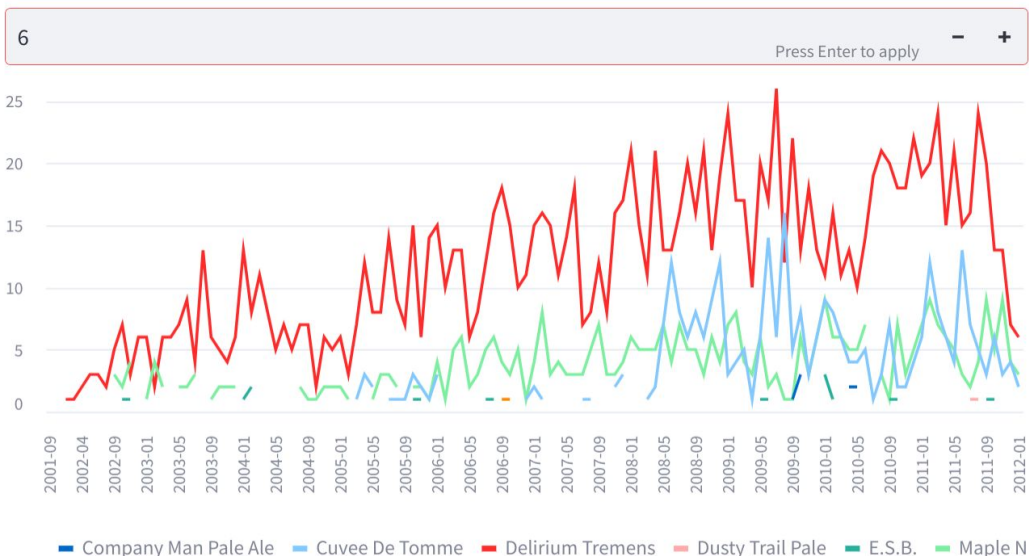
Select Users



Exploratory Data Analysis

The beer reviews over time for a particular beer

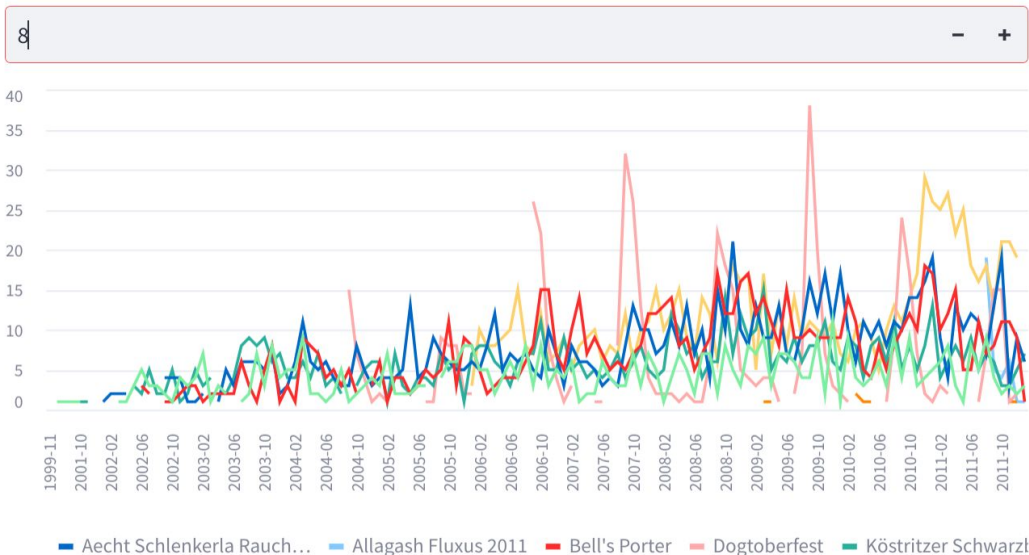
Enter the number of beers you want to display



Exploratory Data Analysis

The beer reviews over time for a particular beer

Enter the number of beers you want to display



Hyperparameters & models performance

	ALS		KNN
NDCG	0.985		
maxIter	10	bucket length	2
rank	4	number of hash tables	3
regParam	0.01		

Results

×

Exploratory data analysis

Here, we list different exploratory data analysis actions

What action will you like to perform?

Checkout the descriptive d... ▾

Show

Recommendation System

Different Recommendation tasks

What action will you like to perform?

Predict Beers a user will like ▾

Which model will you like to use?

Best Model ▾

What users will you like to make predictions for

User 1 × ▾

How many beers will you like to be recommended

3 - +

Submit

Big Data Project 2023



CRAFT BEER

Beer Reviews

Descriptive Data Analysis

Challenges



File format compatibility with PostgreSQL



Limited resources for partitioning and complex models



Low software version and hence less friendly interface



Future Prospects



Obtain more information such as location of reviewers, the review text content and missing abv values



Experiment with more sophisticated models to see how performance can improve



Get access to more resources like hardware with better specs and upgraded software version



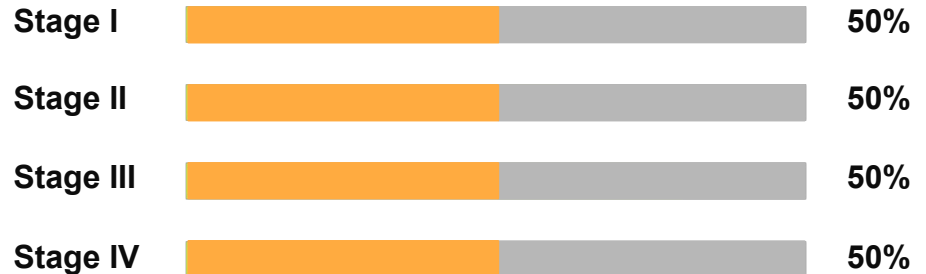
Participation



Beeradvocate

Teamwork

For all stage of the project, the workload was equally shared among us





Beeradvocate

