Beer Reviews

Managing Big Data Final Project
A Recommender System by Fab Team 5

MEET THE TEAM

MARY BOKEN

Chicago, Illinois

SAMI FELLER

Coral Springs, Florida

DAVID LI

Shandong, China

WHATS ON THE MENU

TOPICS AND HIGHLIGHTS

Beeradvocate

EXECUTIVE SUMMARY

WHAT WE DID

- Found dataset from Kaggle with 1.5 million rows and 13 columns with information on BeerAdvocate user reviews
- Used SQLContext in Pyspark to run queries on dataset and answer initial questions
- Created a recommender system using the Alternating Least Squares (ALS) algorithm from pyspark.ml to predict the ratings for beers which a user has not yet reviewed

INTRODUCTION

MOTIVATION

- Want to chose best possible beer for taste preferences
- Interested in consumer preferences of beer

OVERVIEW

- Used data from reviews on BeerAdvocate, a global beer review platform
- Created recommender system using ALS algorithm

POTENTIAL VALUE

- Reliability – organic means of spreading information
- Inform new market segments
- Partner with breweries, restaraunts and producers

Beeradvocate®

DATA OVERVIEW

- OVER 1.5 MILLION BEER REVIEWS
- SOURCE HTTPS://WWW.KAGGLE.COM/RDOUME/BEERREVIEWS?SELECT=BEER_REVIEWS.CSV

summary	review_aroma	review_appearance	review_overall	review_palate	review_taste
count	1586266	1586266	1586266	1586266	1586266
mean	3.735685565976955	3.84167125816225	3.8156280220341356	3.7437532544982997	3.7929209224682365
stddev	0.6975674474672164	0.6160649705043265	0.7205948113732451	0.682175283752888	0.7319121675111695
min	1.0	0.0	0.0	1.0	1.0
max	5.0	5.0	5.0	5.0	5.0

VARIABLES WE ARE WORKING WITH

brewery_id
brewery_name
review_time
review_overall
review_aroma
review_appearance
profilename
beer_style
review_palate
review_tast

Beeradvocate®

Data Wrangling Process

LOADING DATA INTO SPARK

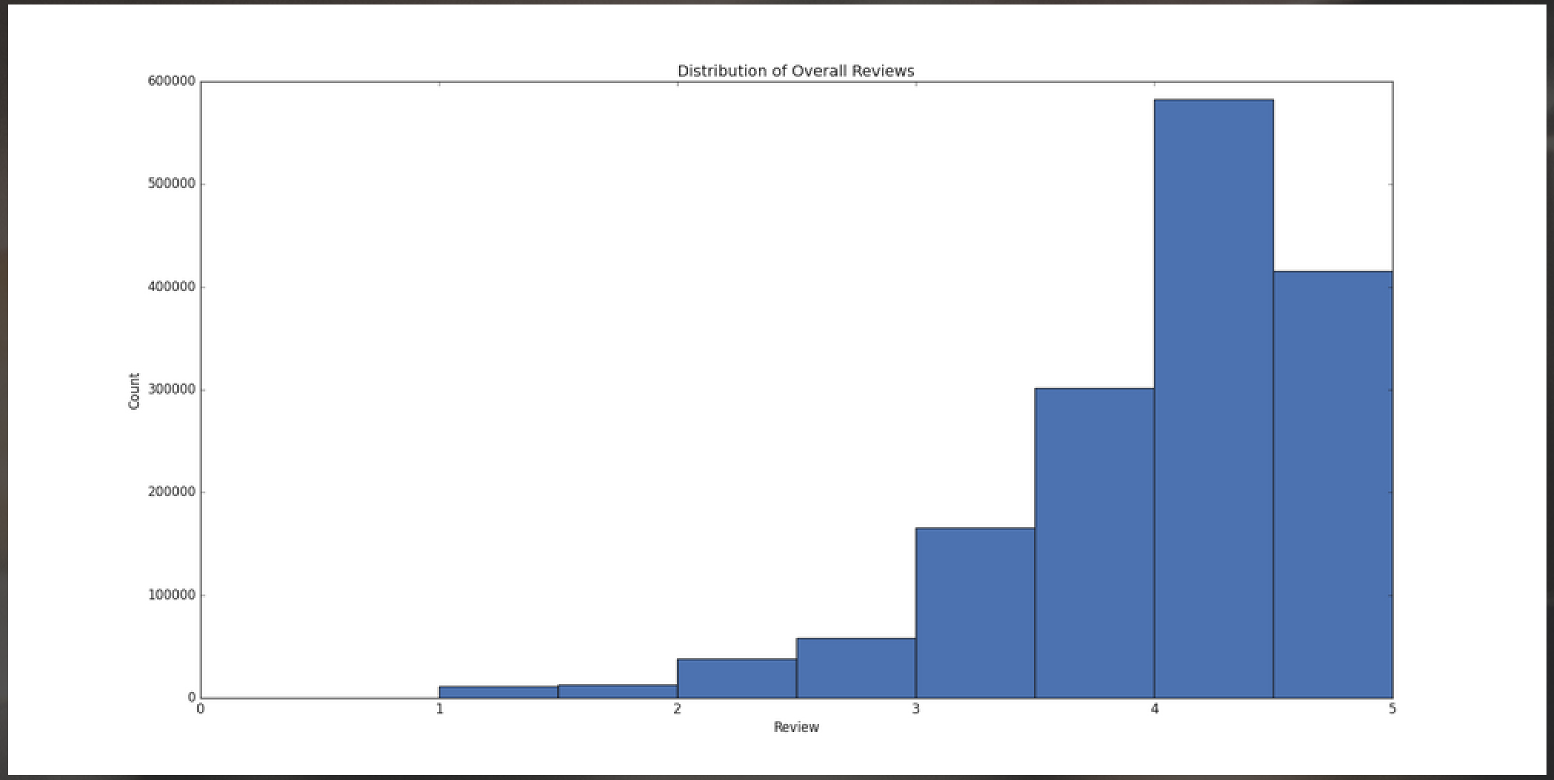
VISUALIZATIONS

QUERIES

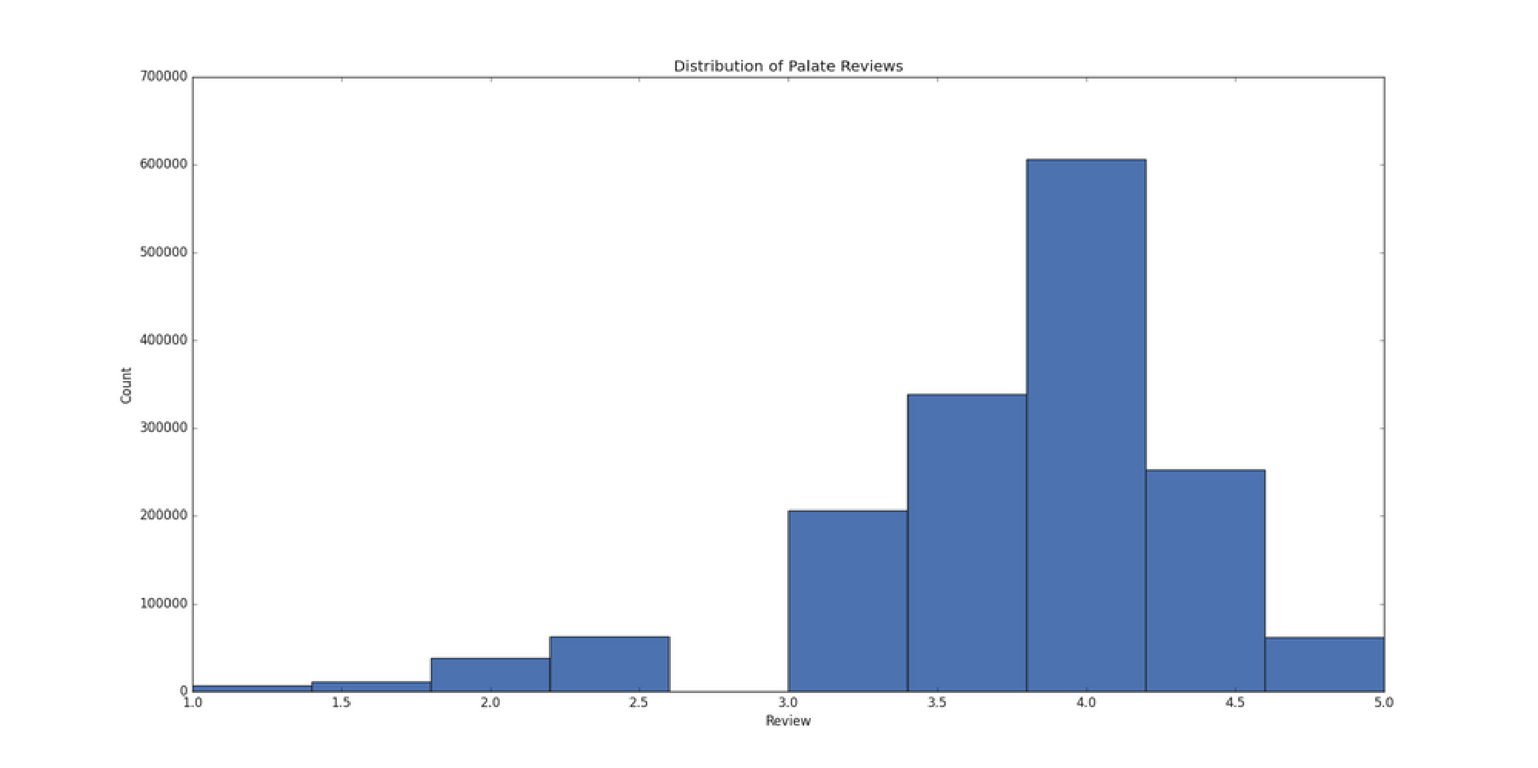
RECOMMENDER SYSTEM

VISUALIZING OUR DATA - HISTOGRAMS OF REVIEW DATA

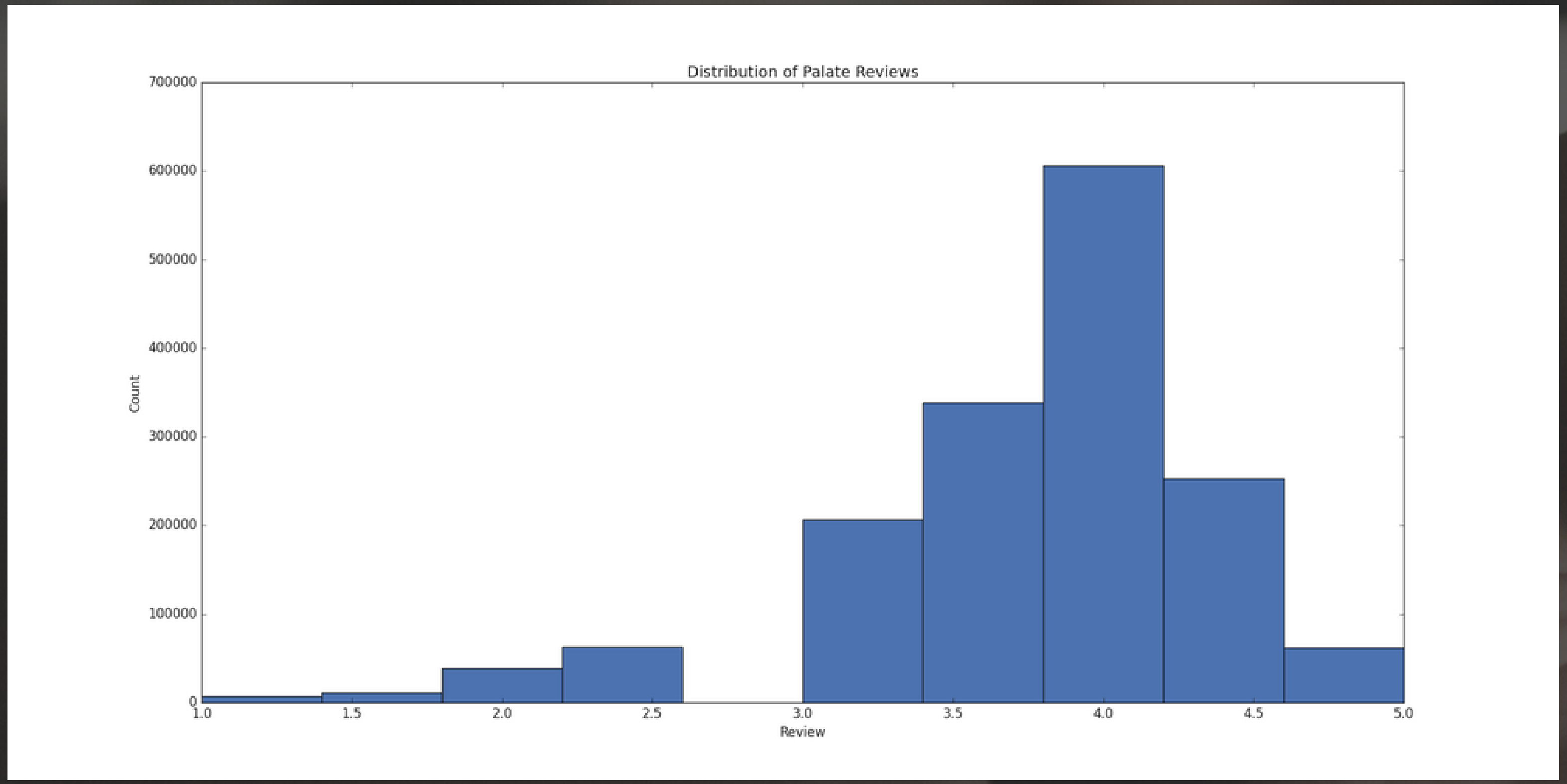
OVERALL



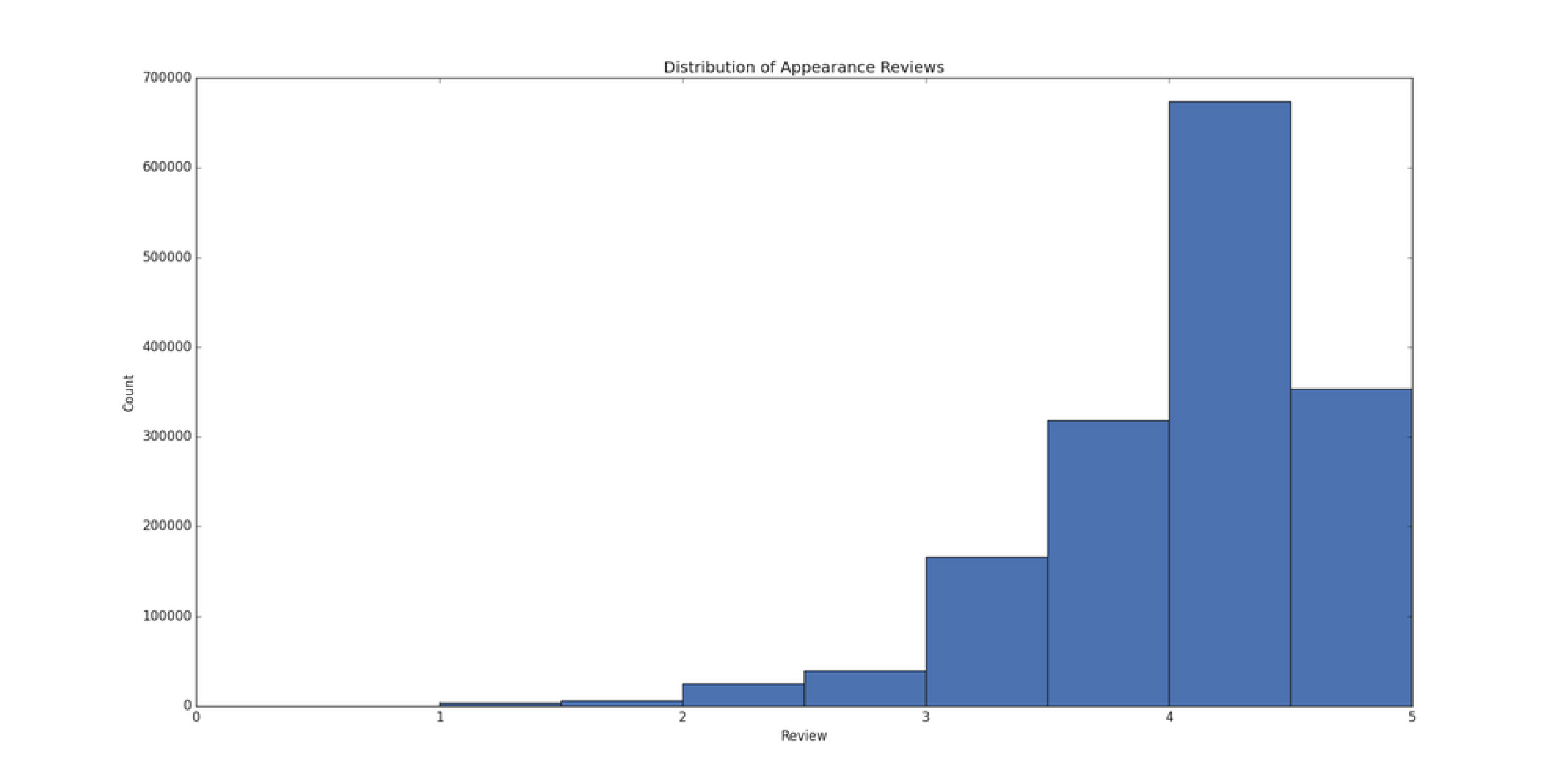
PALATE



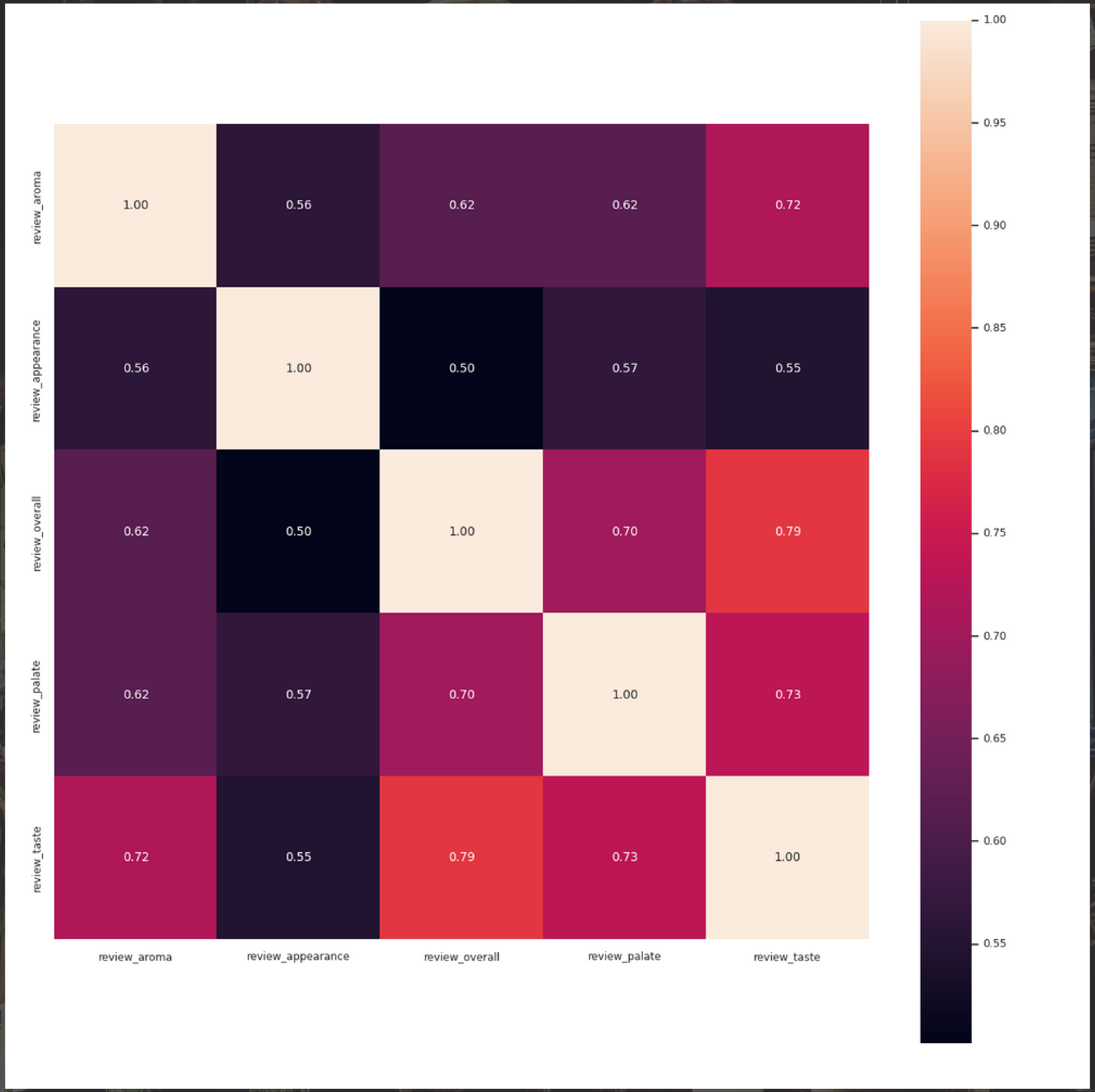
TASTE



APPEARANCE



VISUALIZING OUR DATA -
HEATMART FOR CORRELATION AMONG REVIEW TYPES



INSIGHTS

MODEL-FREE INSIGHTS FROM DATA

Run queries to understand top beers based on volume of rating, top beers based on average score of rating, most popular style of beer, and highest rated style of beer based on average score

MODEL-DRIVEN INSIGHTS

Use collaborative filtering techniques to recommend beers which users have not yet reviewed

Beeradvocate®

RESULTS OF QUERIES

WHICH BEER HAD THE HIGHEST RATING IN DATASET (AT LEAST 1000 REVIEWS)

beer_name	average_rating	num_reviews
Trappist Westvlet...	4.62	1272
Pliny The Elder	4.59	2527
Weihenstephaner H...	4.52	1980
The Abyss	4.45	1412
Sculpin India Pal...	4.44	1352
Supplication	4.43	1053
Founders KBS (Ken...	4.4	1930
Tröegs Nugget Nectar	4.4	1954
Gumballhead	4.4	1234
Bell's Hopslam Ale	4.38	2443

WHICH BREWERY IS MOST POPULAR?

brewery_name	num_reviews
Boston Beer Compa...	39438
Dogfish Head Brewery	33829
Stone Brewing Co.	33053
Sierra Nevada Bre...	28746
Bell's Brewery, Inc.	25189
Rogue Ales	24079
Founders Brewing ...	20000
Victory Brewing C...	19473
Lagunitas Brewing...	16832
Avery Brewing Com...	16105

THE BOSTON BEER COMPANY

WHICH BEER STYLE IS MOST POPULAR?

beer_style	num_reviews
American IPA	117586
American Double /...	85977
American Pale Ale...	63469
Russian Imperial ...	54129
American Double /...	50705
American Porter	50477
American Amber / ...	45751
Belgian Strong Da...	37743
Fruit / Vegetable...	33861
American Strong Ale	31945

How the Recommender System Works

Beeradvocate

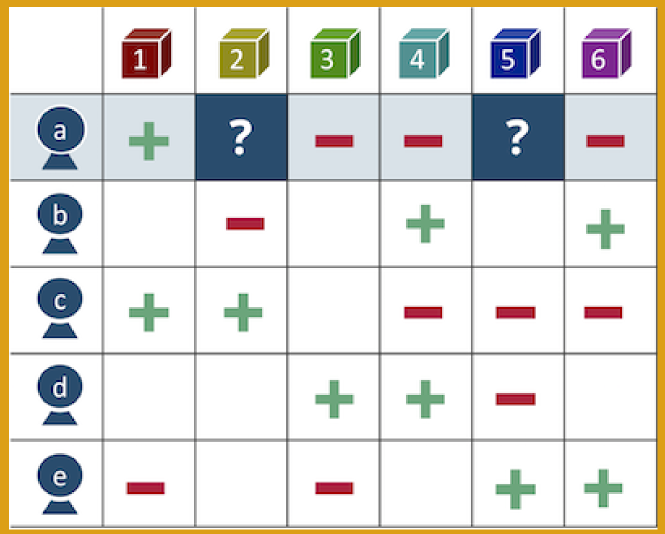
INPUTS

User ID – a list of each user

Item Column – list of beer names

Rating Column – overall review each user assigned to each beer

MATRIX REPRESENTATION



MATRIX FACTORIZATION

Generate latent factors to represent user preferences in a lower dimensionality space

PREDICT RATING AND RECOMMEND

Predicts ratings and recommends beers to users based on the highest predicted ratings

source: https://www.google.com/search?q=collaborative+filtering&client=firefox-b-1-d&sxsrf=ALeKk03XTYy3NbGuqzRhx8fU1ccO_Su99A:1605996405698&source=lnms&tbm=isch&sa=X&ved=2ahUKEwj4qZbq0pTtAhU3QzABHc0JCEoQ_AUoAnoECBIQBA&biw=1920&bih=966#imgrc=DkzRjUGTqwA0UM&imgdii=ot2TYv-958MowM

RECOMMENDER SYSTEM RESULTS

- ROOT-MEAN-SQUARE ERROR = 0.55
- ON AVERAGE, THE DIFFERENCE BETWEEN THE ACTUAL RATING AND THE PREDICTION IS 0.55 (REVIEW SCALE IS 1 - 5)

review_profilename	review_overall	beer_name	prediction
beertunes	3.0	Pecan Harvest Ale	3.2725961
CortexBomb	5.0	Trois Pistoles	3.8026361
CortexBomb	4.0	Éphémère (Apple)	3.3211913
CortexBomb	5.0	Edition 2004	3.6699762
CortexBomb	4.0	Quelque Chose	3.3831587
CortexBomb	4.5	Unibroue 15	3.629299
beertunes	4.0	Maudite	3.7953036
beertunes	4.0	Raftman	3.623886
BeefyMee	5.0	Samuel Adams Crea...	3.740396
BeefyMee	2.5	Samuel Adams Cher...	2.9721463

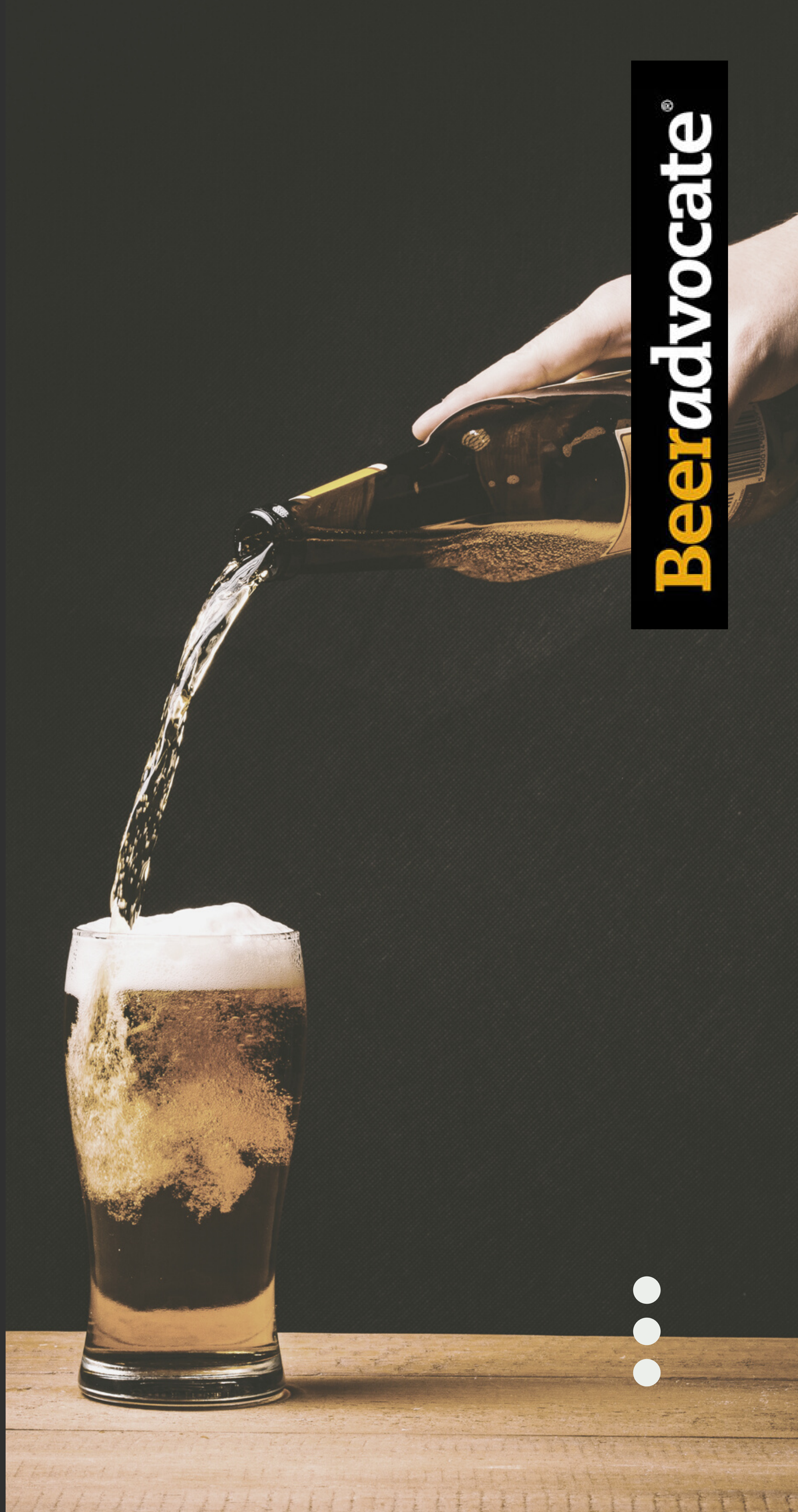
only showing top 10 rows

EXAMPLE RECOMMENDATIONS

RECOMMENDING NEW BEERS FOR USER 1

user_id	rating	brewery_name	beer_name	beer_style
1	5.2295203	Triumph Brewing C...	Snakebite	Fruit / Vegetable...
1	5.1477723	Five Seasons North	Whiskey And Oak B...	Dubbel
1	5.198488	Carter's Brewing	Black Magic Porter	American Porter

TRIUMPH BREWING CO



REASONS FOR SELECTING APPROACH

EXTREMELY LARGE DATASET

Spark is a memory based data processing framework that is much faster than Hadoop Map Reduce

MODULARITY

Spark offers modules for both SQL queries and machine learning, which were both central to our project

SPARK MLLIB – ALS

ALS algorithm – collaborative filtering to build a recommendation system

Beeradvocate

GENERALIZIBILITY OF INSIGHTS

WHAT WE LEARNED

- THIS TECHNOLOGY IS NOT LIMITED TO THIS TOPIC AND CAN BE USED TO CREATE RECOMMENDATIONS FOR OTHER FOOD AND BEVERAGE INDUSTRIES
- LEVERAGE SELF-REPORTED INFORMATION AS AN ORGANIC MEANS OF ADVERTISING AND COLLECT ADDITIONAL DEMOGRAPHIC INFORMATION TO GET MORE ROBUST UNDERSTANDING OF CUSTOMERS
- USE TOOL TO PARTNER WITH BREWERIES, RESTAURANTS, PRODUCERS

SWOT Analysis

STRENGTHS	WEAKNESSES	OPPORTUNITIES	THREATS
Provides dynamic information on preferences to keep up with how consumers react to trends and new products to generate value of customers and businesses	Data last updated in 2011 so not as up to date as possible and recommender system could be limiting as customer tastes are dynamic with many likes and dislikes	Potential partnerships with restaurants, breweries, and beer producers can bring BeerAdvocate additional revenue streams	Competitors, change in preferences, first-movers on an alternative recommender system

Beeradvocate®

CONCLUSION

THERE IS CLEAR POTENTIAL VALUE FOR CUSTOMERS LOOKING TO FIND THE PERFECT PRODUCT FOR THEIR TASTE, BUT OUR PROJECT IS UNIQUE BECAUSE THERE IS NOTABLE VALUE FOR SELLERS AS WELL AS THIS IS A FORM OF ORGANIC ADVERTISING FOR EACH BREWERY AND BEER THAT APPEAR IN OUR RECOMMENDER SYSTEM

WORKS CITED

https://www.kaggle.com/rdoume/beerreviews
https://www.beeradvocate.com/user/location/?status=no-location
https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-2-alternating-least-square-als-matrix-4a76c58714a1https://towardsdatascience.com/building-a-recommendation-engine-to-recommend-books-in-spark-f09334d47d67https://towardsdatascience.com/intro-to-recommender-system-collaborative-filtering-64a238194a26https://towardsdatascience.com/build-recommendation-system-with-pyspark-using-alternating-least-squares-als-matrix-factorisation-ebe1ad2e7679

Beeradvocate®

THANK YOU!
QUESTIONS?