



Data Incubator Finalist Presentation

Yue JIN

16 May 2018

About Me



Yue JIN

ISTJ Personality

Introverted | **O**bservant | **T**hinking | **J**udging

2013-2016

- Worked as Business Analyst / Management Consultant, mostly serving clients from healthcare industry
- Gradually fell in love with **Data**

2017

Completed Udacity Data Analyst NanoDegree

Present

Master Student in Biostatistics
University of Michigan

2013

Bachelor of Clinical Medicine
SJTU, China

2016

- Started to learn programming for data visualization and machine learning
- Obtained Stanford Machine Learning Online Certificate

Project Introduction

BeerRadar

A Beer Recommendation APP for Craft Beer Lovers



User Side

1

- Rate beers you have tried before
- Set additional preference (manufacturer region, beer style, bitterness, alcohol content and etc.)

3

Provide feedback on recommended beers after you tried them



App Side

2

Recommend new beers you may like

4

Improve future recommendations according to user feedback

Project Motivation

A Hobby for Me:

- Personally, I love craft beer
- From 2016 to Present, I tasted **152** different beers

Great Market Potential:

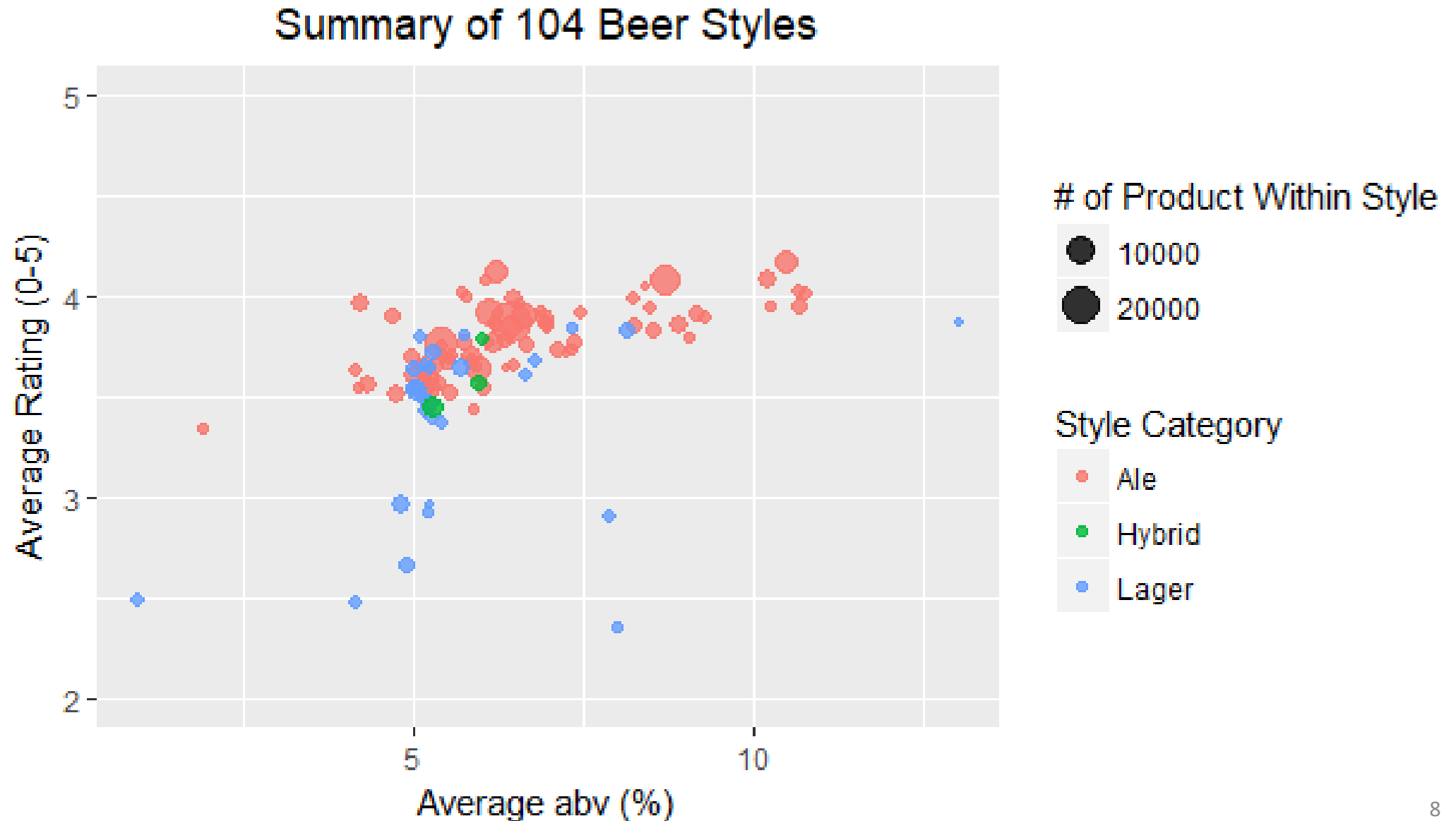
- **Huge user base:** **4.71M** monthly users of *BeerAdvocate.com*, **2.99M** monthly users of *Ratebeer.com*, **3.07M** monthly users of *Untappd.com**
- **Business model succeeded in similar field:** Recommendation algorithm + Product merchandise for wine - [BrightCellar](#)

Data Source

Data was scraped from [BeerAdvocate.com](https://beeradvocate.com):

- **104** Beer Styles
- **13K** Breweries from around the world
- **200K** Beer Items
- **7 Million+** Ratings from **300K+** Users
- **3 GB** Total Data Size

For Beer Styles... Stronger Ale is always better!



Algorithm Test Run: Mean error of 0.3/5.0

- **SVD matrix factorization** implemented on **2.2 Million** ratings from **9000 Users** on **1300 Beers***
- **Cross validation result:**
 - RMSE (Root Mean Square Error): **0.37/5.00**
 - MAE (Mean Absolute Error): **0.27/5.00**

Evaluating RMSE, MAE of algorithm SVD on 5 split(s).

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std
RMSE (testset)	0.3713	0.3713	0.3715	0.3713	0.3710	0.3713	0.0001
MAE (testset)	0.2658	0.2664	0.2668	0.2665	0.2665	0.2664	0.0003

Performance will be further improved:

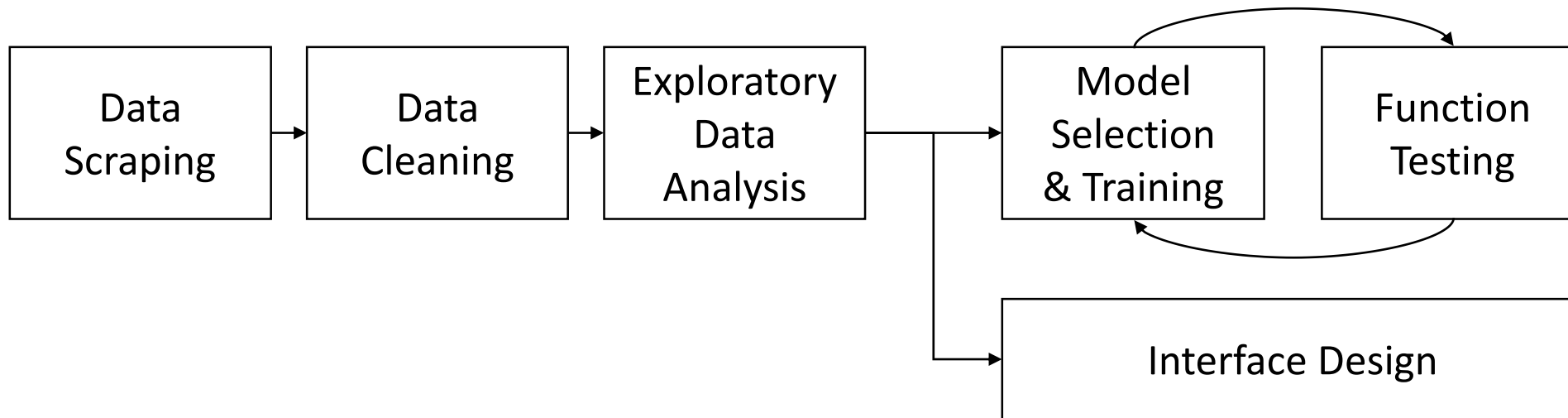
- Algorithm parameter tuning
- Leverage text features extracted from textual comments
- Incorporate popularity based/content based algorithms

* Subset of the full dataset, which includes ratings from users who submitted >100 ratings on beers with >1000 ratings

Project Timeline

Phase One: 19th April ~ 31st May

Phase Two: 1st June ~ 10th August



...And Cheers!