



Beer Analysis

E.K. Itoku
5/20/2015

Objective

- **DS objectives**
 - Classify beer types
 - Identify review topics
 - Extract key words for each topic
- **Other objectives**
 - Improve beer skills
 - Have a good time!

Data Source

- RatBeer.com

- The most popular beer review websites
- Founded in May 2000

The screenshot shows the product page for Westvleteren 12 (XII) on RateBeer.com. At the top, there's a banner for 'Kibana 4 - Newly Released' with a message about reimaged data visualization demos. Below the banner, the main navigation menu includes Local, Ratings, People, Shop, Events, Places, and Forums. The main content area features the beer's name, 'Westvleteren 12 (XII)', in large blue text. To the left is an image of a dark brown bottle of the beer. To the right, there's a summary section with a large 'overall 100' rating, 'style 100', and a note that it's 'Brewed by Westvleteren Abdij St. Sixtus'. It also mentions the style is 'Abt/Quadrupel' and it's from 'Westvleteren, Belgium'. Below this, there are links for 'send corrections', 'shelftag', 'edit barcodes', and 'update pic'. On the far right, there are buttons for 'Rate This', 'Statistics', and 'Locate Sellers'. A 'tags' section at the bottom lists '#trappist', '#sugar', '#bottle conditioned', and '#cellar favorite'. The footer contains social media links and a note about the brewery's small output.

Data Transformation

RateBeer



MongoDB



PostgreSQL

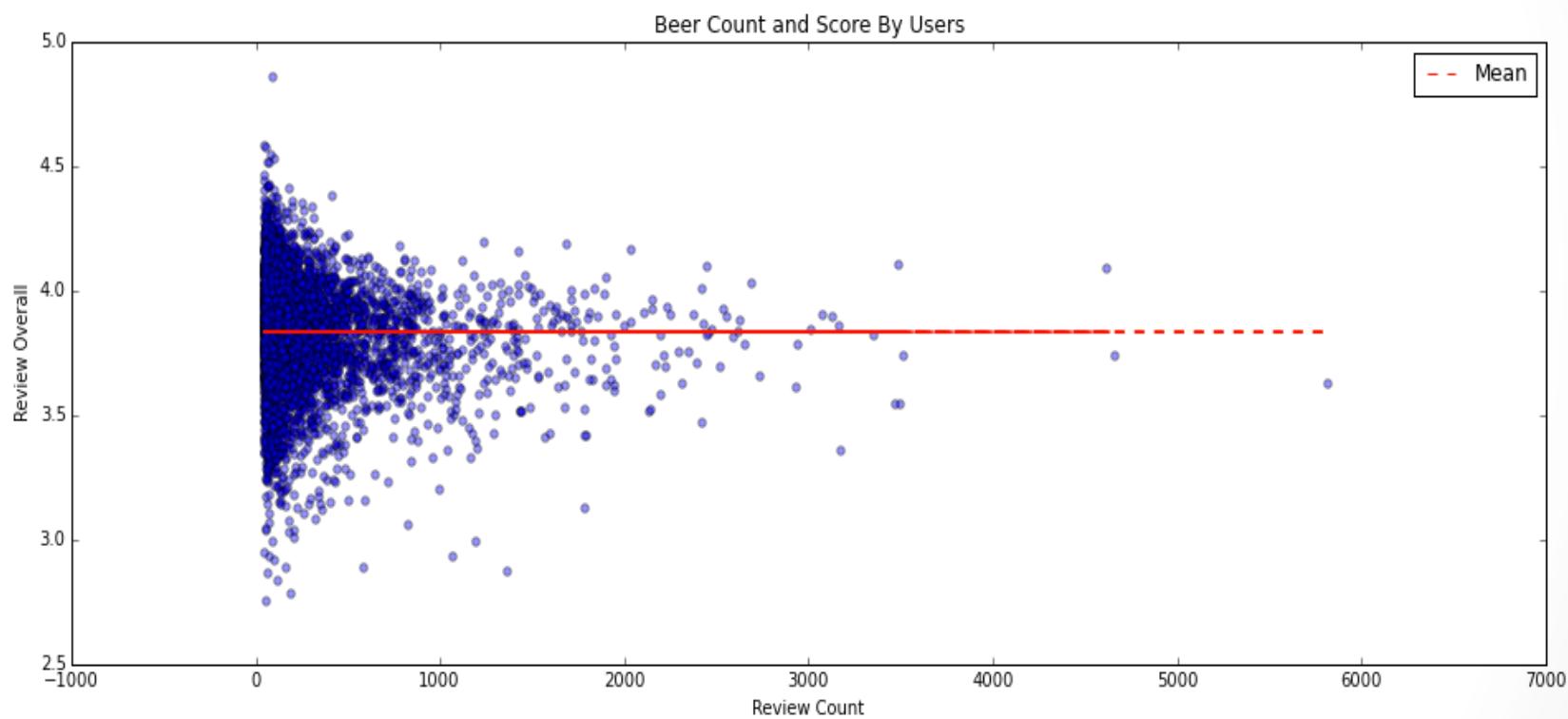


Key Stats

	Brewer	Beer	Review	User
Count	18,680	291,913	6,898,186	60,481

- 15 beers / brewer
- 114 reviews / user

Ratings by User

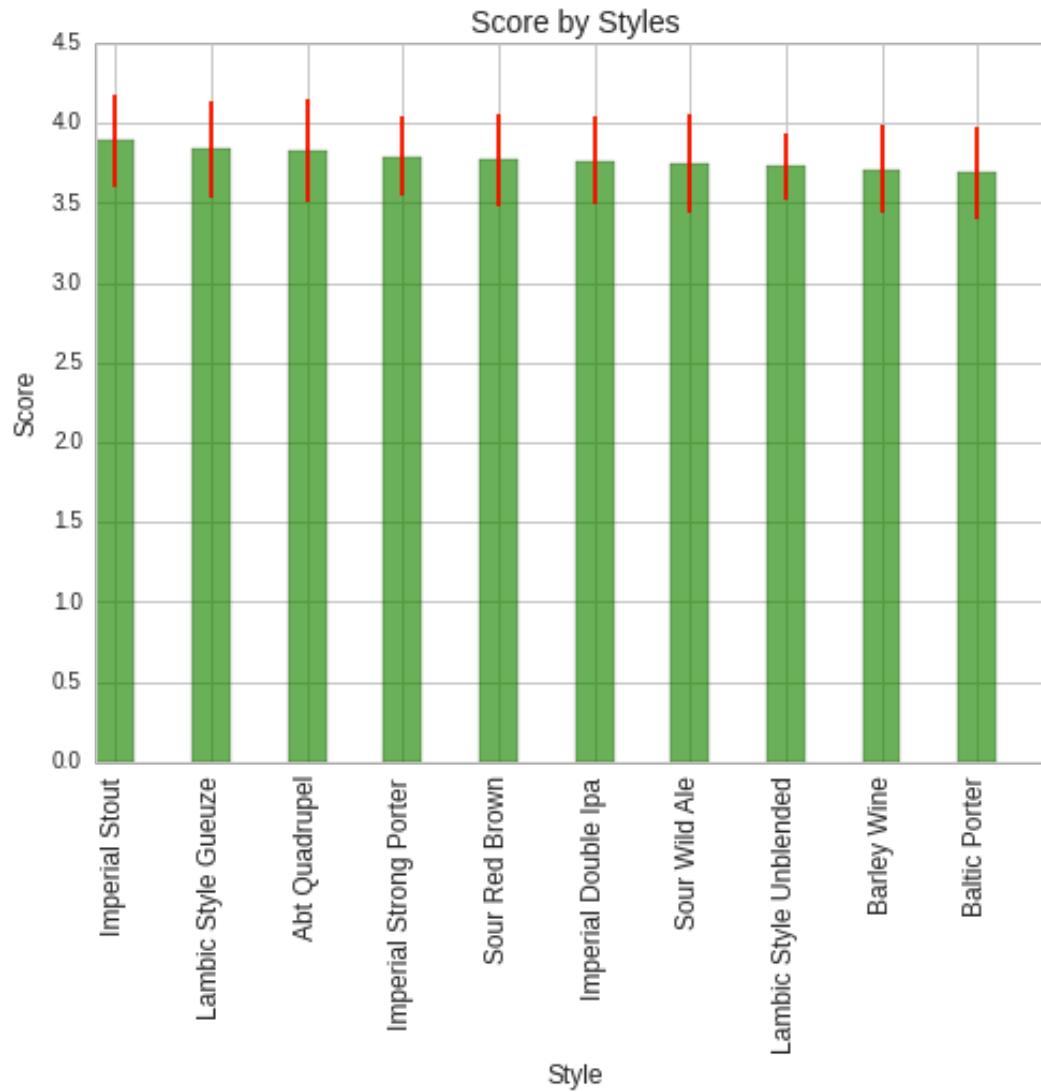


Ratings

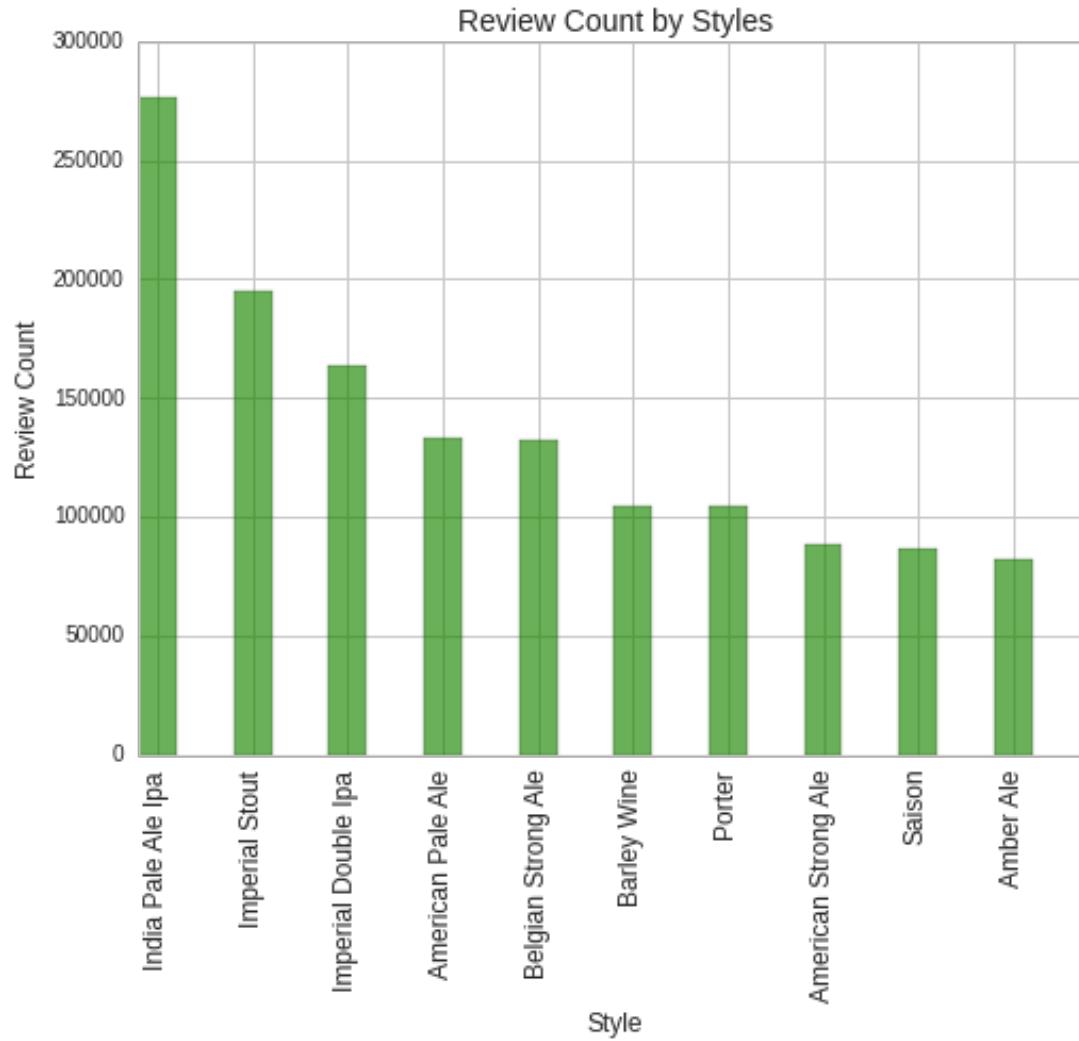
	Overall	Taste	Aroma	Palate	Appearance
Average	13.23	6.50	6.41	3.28	3.44
Max	20	10	10	5	5
Normalized	3.30	3.25	3.20	3.28	3.44

- Final score = $(\text{overall} + \text{taste} + \text{aroma} + \text{palate} + \text{appearance}) / 10$
- RateBeer does weight adjustment

Ratings by Style

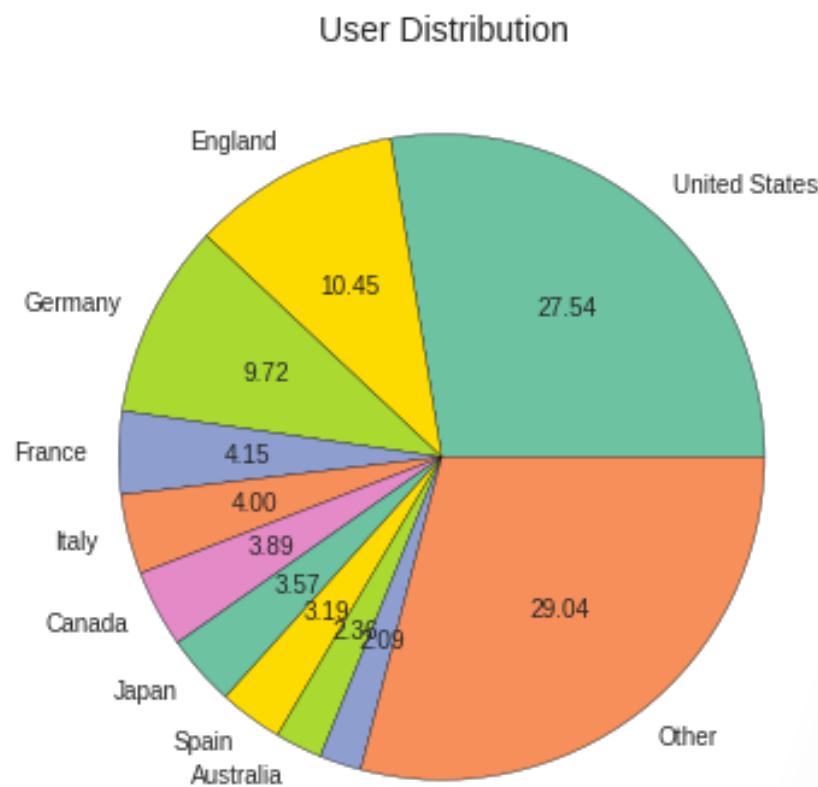


Ratings by Style



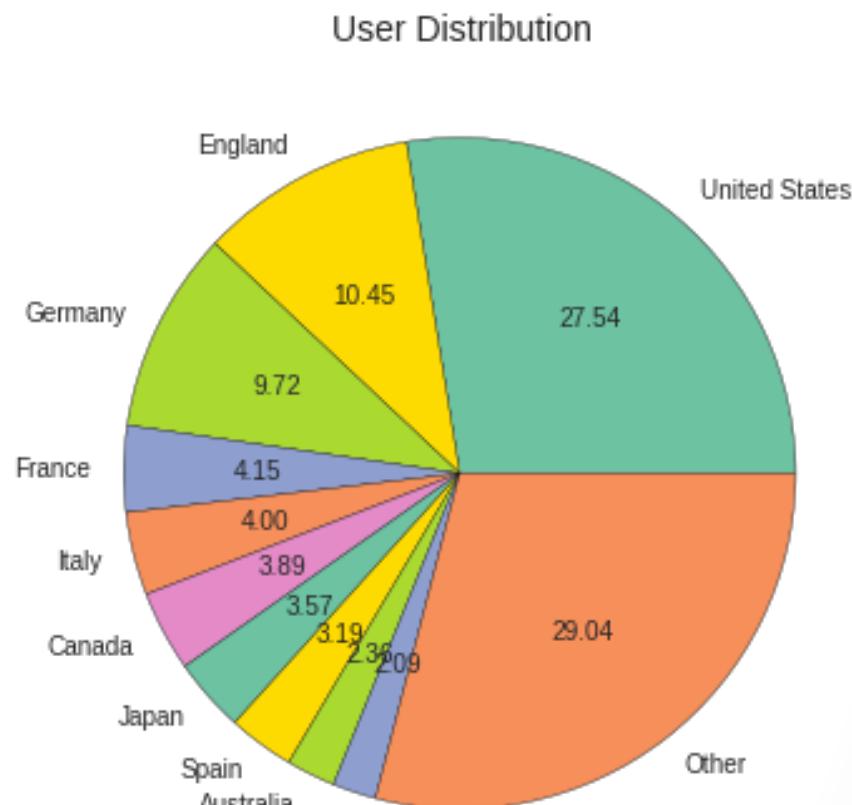
Brewer by Country

Country	Count
United States	5,145
England	1,952
Germany	1,816
France	775
Italy	748
Canada	726
Japan	667
Spain	595
Australia	441
Switzerland	391
Other	5424



User by Country

Country	Count
United States	25,798
Canada	2,793
England	1,957
Australia	1,175
Netherlands	851
Denmark	864
Belgium	817
Sweden	816
Italy	689
Poland	632
Other	6356



Beer Styles

Level 1



Level 2: type



Level 3: group



Level 4: style

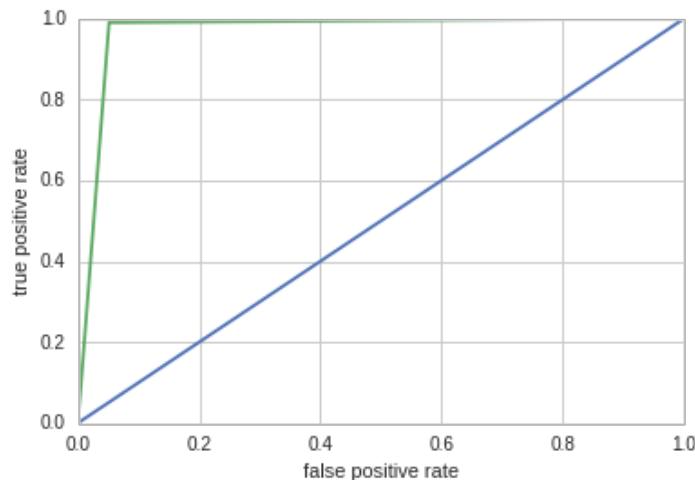


Level2: Ale or Lager?

- Algo: Random Forest Classifier
 - n_estimators=1,000
 - max_features='sqrt'
- Vectorizer: Tf-idf
 - Only alphabets
 - Remove stop words
 - Remove additional words
 - At least two characters
 - Stem words

Level2: Ale or Lager?

- Accuracy 96.88%
- Precision 94.78%



feature	importance
Dimpled mug	0.174
weighted_avg	0.068
English pint	0.060
est_calories	0.038
Shaker	0.037
alcohol_by_volume	0.036
Snifter	0.024
Commercial Brewery	0.019
Flute	0.018
Trappist glass	0.017

Level2: Ale or Lager?



Ale



Lager

Level3: 6 Beer Groups

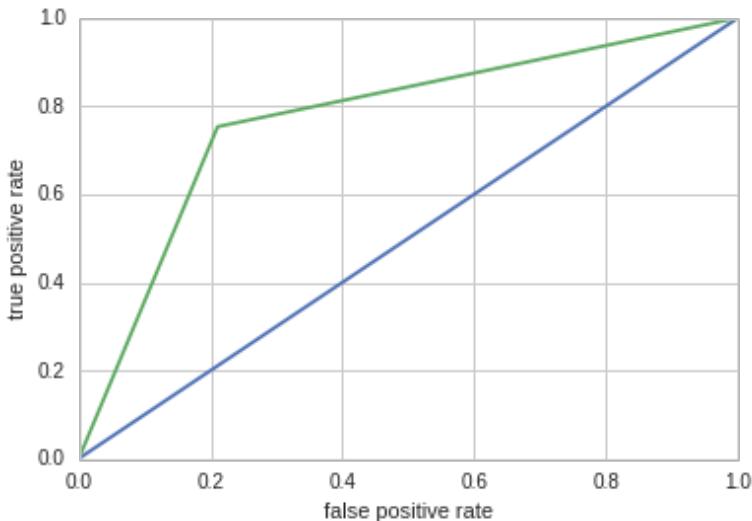
- Accuracy 92.96%
- Precision 92.97%
- Confusion matrix

```
[247  0  45  0   6  1]  
[ 2  89  0   0   0  0]  
[ 5  0  626  0   0  0]  
[ 4  0   3  31   2  1]  
[ 12  0   0   1 129  0]  
[ 2   1   3   0   0  40]
```

feature	importance
Dimpled mug	0.094
Trappist glass	0.047
English pint	0.047
weighted_avg	0.041
Shaker	0.034
alcohol_by_volume	0.031
est_calories	0.029
Snifter	0.024
black	0.016
chocol	0.014

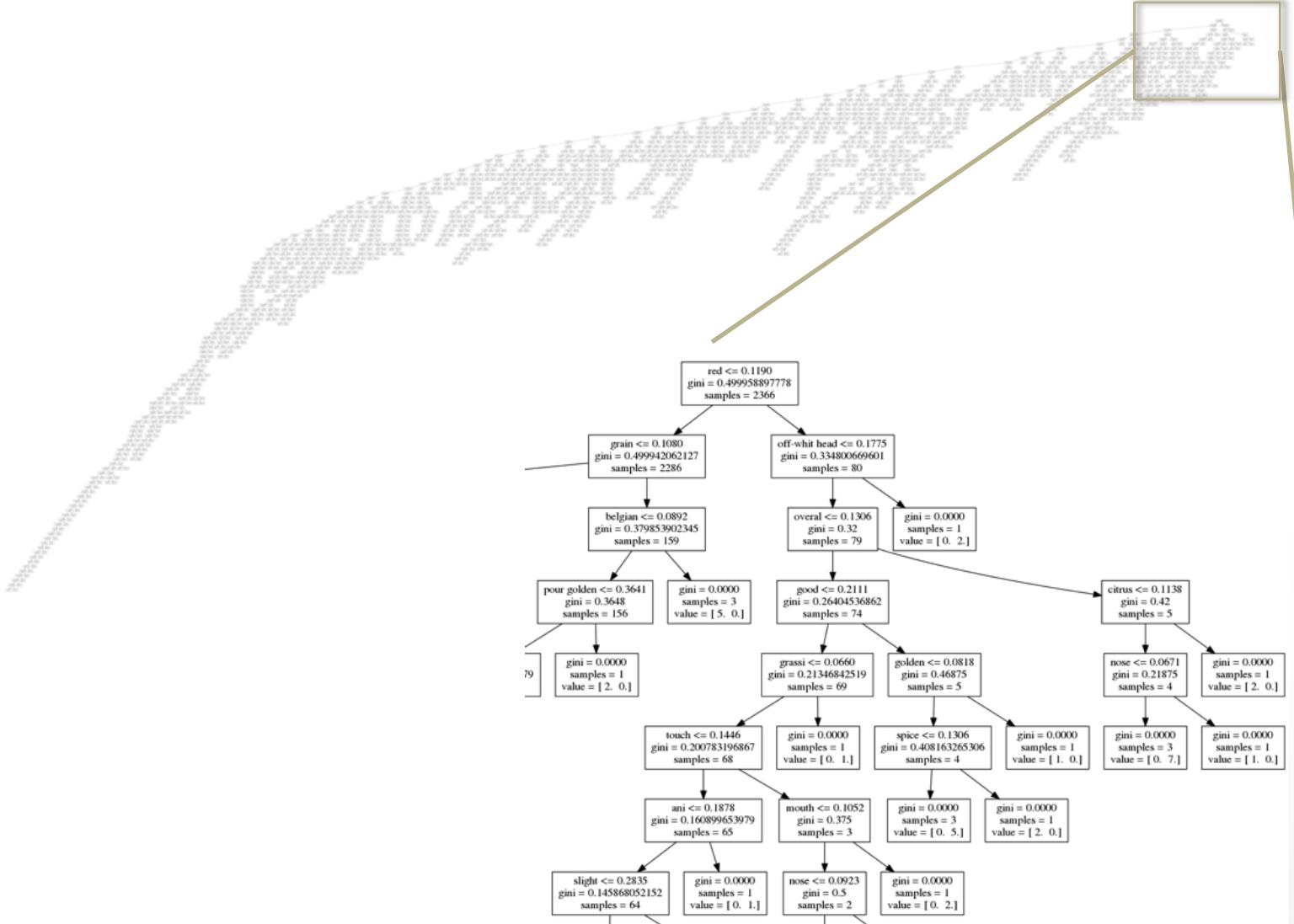
Level2: Ale or Lager, Using Just Reviews

- Accuracy 77.12%
- Precision 78.31%



feature	importance
citrus	0.0140
pour	0.0114
head	0.0107
orang	0.0107
aroma	0.0097
light	0.0096
malt	0.0095
tast	0.0094
sweet	0.0087
flavor	0.0083

Level2: Ale or Lager, Using Just Reviews

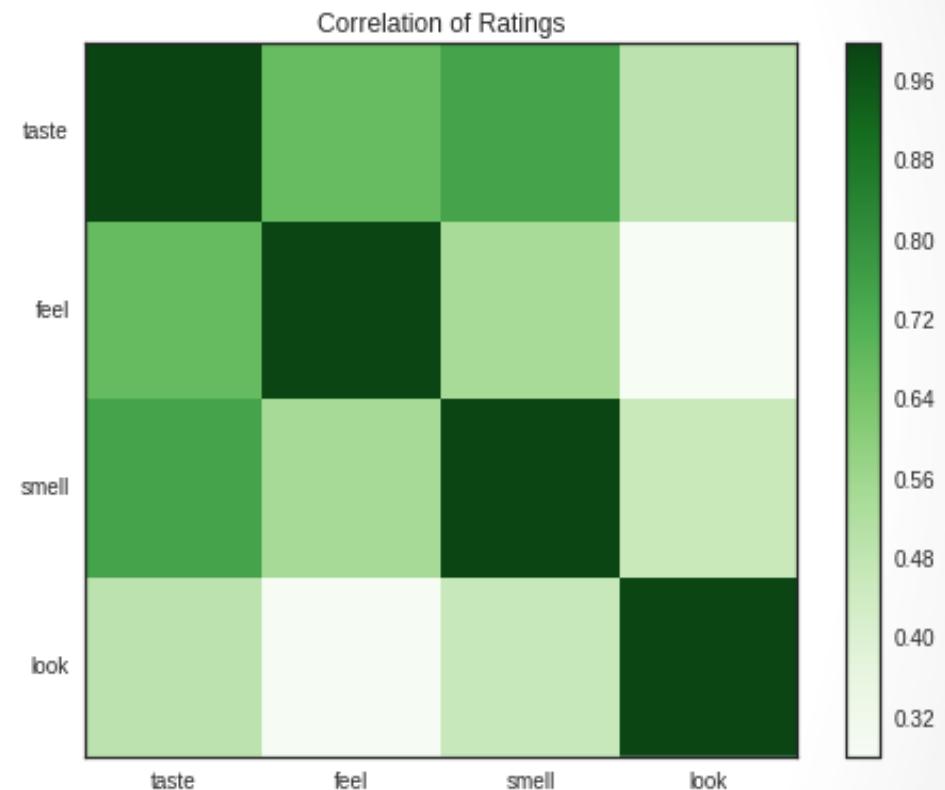


Review Topic Detection

- Use 1.0 as cut off
- Compare one rating field against others
- For example...
 - {taste: 5.0, palate3.5, aroma2.7, appearance:3.0}
 - = > taste
 - {taste: 2.0, palate2.5, aroma2.7, appearance:5.0}
 - = > appearance
 - {taste: 4.5, palate4.2, aroma3.7, appearance:4.0}
 - = > not able to identify

Review Topic Detection

Topic	Count
appearance	306,540
palate	106,726
aroma	25,550
taste	14,190

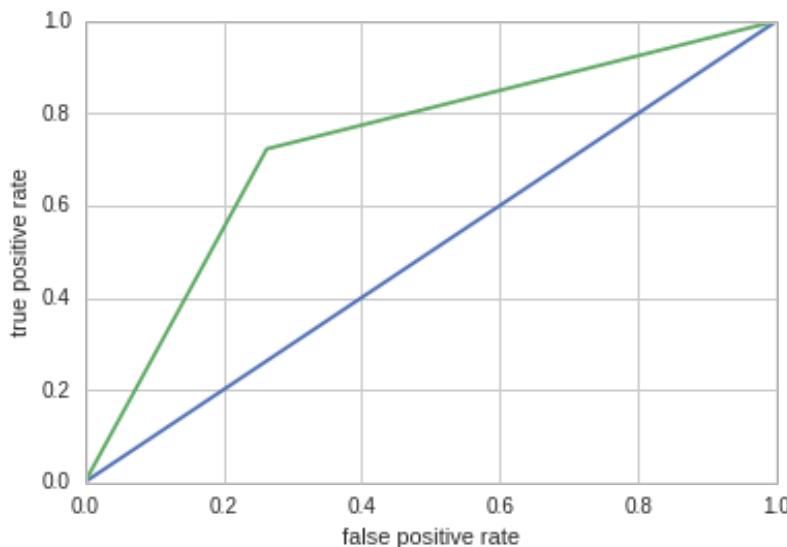


Review Topic Detection

- Algo: Random Forest Classifier
 - n_estimators=10,000
 - max_features='sqrt'
- Vectorizer: Tf-idf
 - Only alphabets
 - Remove stop words
 - Remove additional words
 - At least two characters
 - Stem words

Review Topic Detection

- Accuracy 72.98%
- Precision 73.29%



Feature	Importance
weighted_avg	0.0129
num_ratings	0.0119
alcohol_by_volume	0.0109
est_calories	0.0108
head	0.0107
aroma	0.0080
flavor	0.0077
pour	0.0076
tast	0.0075
veri	0.0065

Extract Topic Key Words

- **Clean up1:** Sum of tf-idf

Look

Feature	tf-idf
flavor	790.9990
light	790.7167
head	783.1871
aroma	774.2346
malt	766.2756
sweet	755.3147
tast	703.7829
hop	694.7205
pour	693.4842
veri	683.0662

Taste

Feature	tf-idf
flavor	800.6873
veri	791.3475
tast	762.3294
head	731.2342
aroma	730.4333
light	720.7826
hop	716.3707
sweet	692.9077
nice	661.6315
malt	655.9216

Extract Topic Key Words

- **Clean up2:** Part-of-Speech
- Only select these below tags

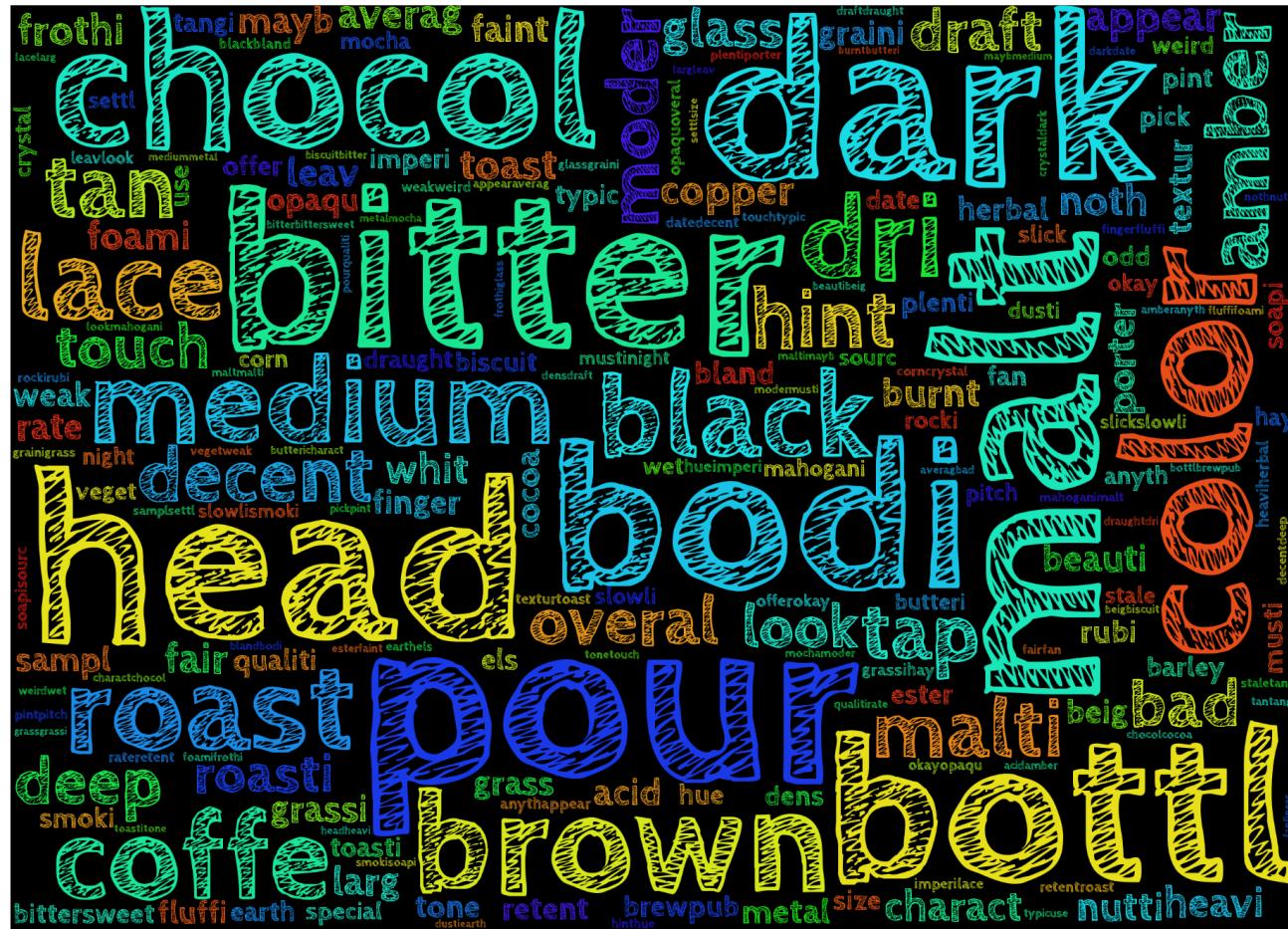
Tag	Description	Example
NN	Noun, singular or mass	sweet
JJ	Adjective	good
JJR	Adjective, comparative	better
RB	Adverb	long
PRP\$	Possessive	sour
NNS	Noun,	banana

Extract Topic Key Words

- **Top 10 Result**

Rank	Look	Feel	Smell	Taste
1	head	light	aroma	flavor
2	malt	hop	tast	veri
3	sweet	finish	bitter	nice
4	pour	smooth	littl	good
5	bottl	carbon	bit	great
6	dark	realli	smell	drink
7	bodi	golden	nose	hoppi
8	color	creami	citrus	balanc
9	chocol	mild	orang	ale
10	brown	yellow	fruit	ipa

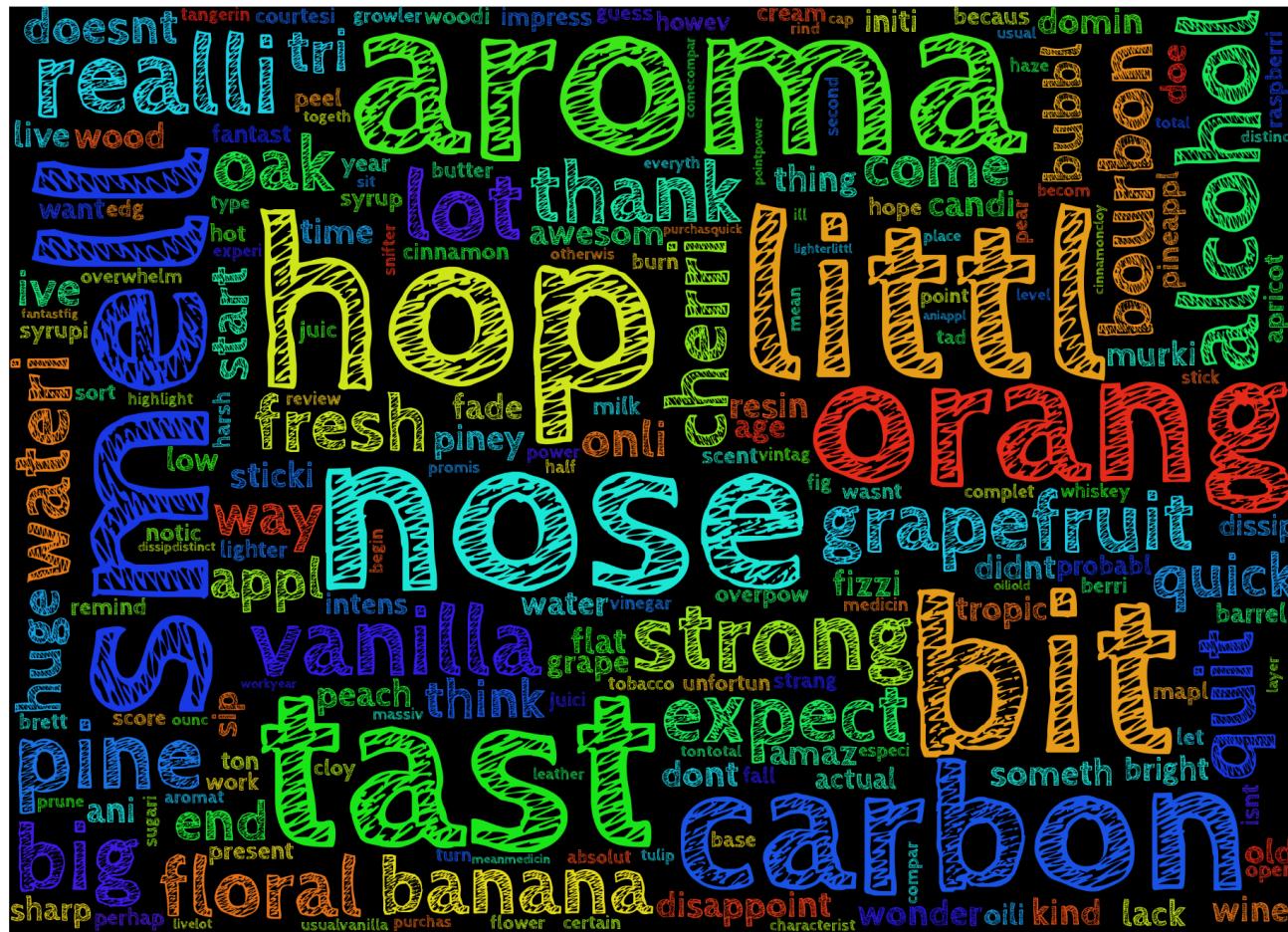
Look



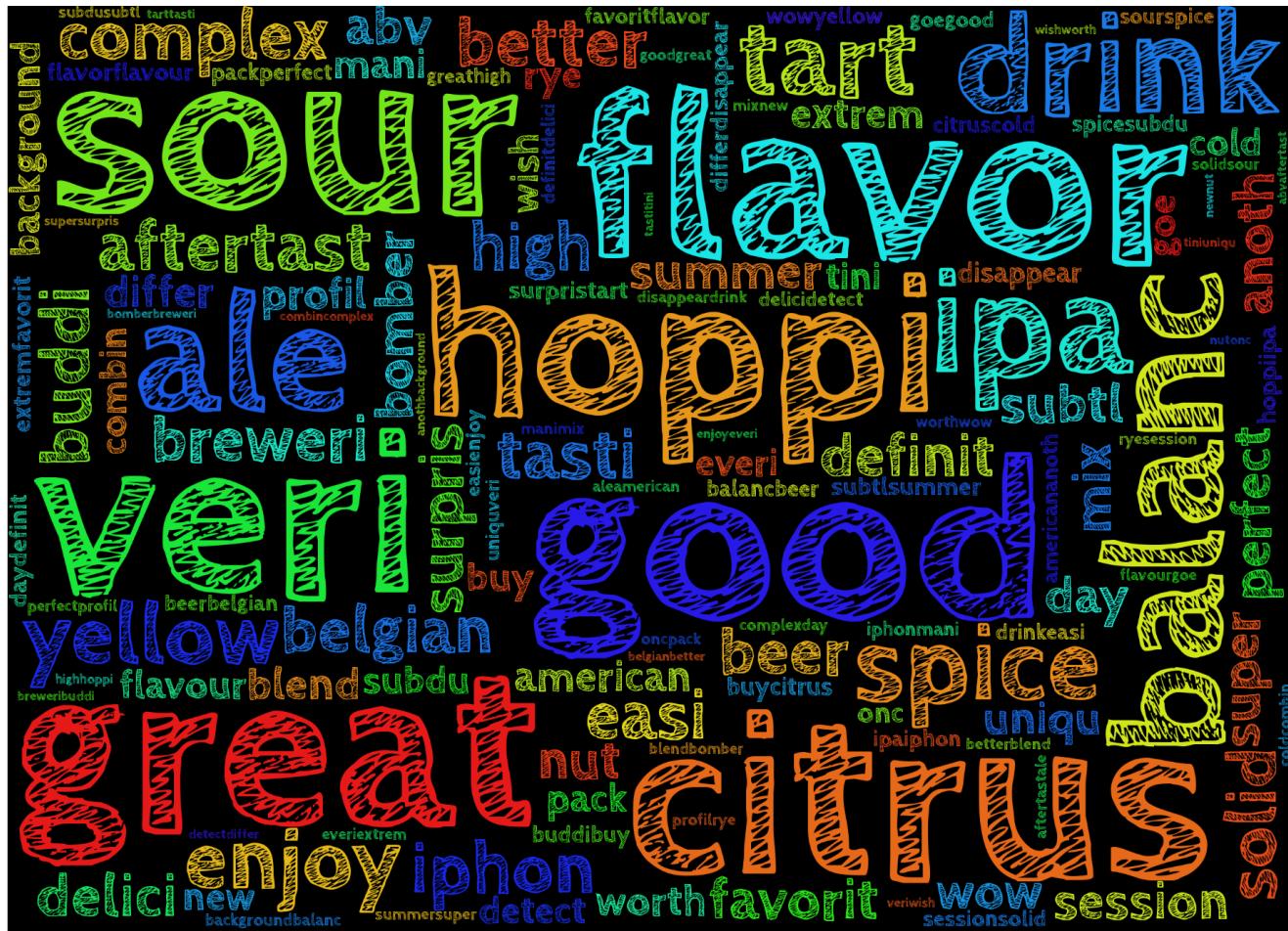
Feel



Smell



Taste



Tank you!