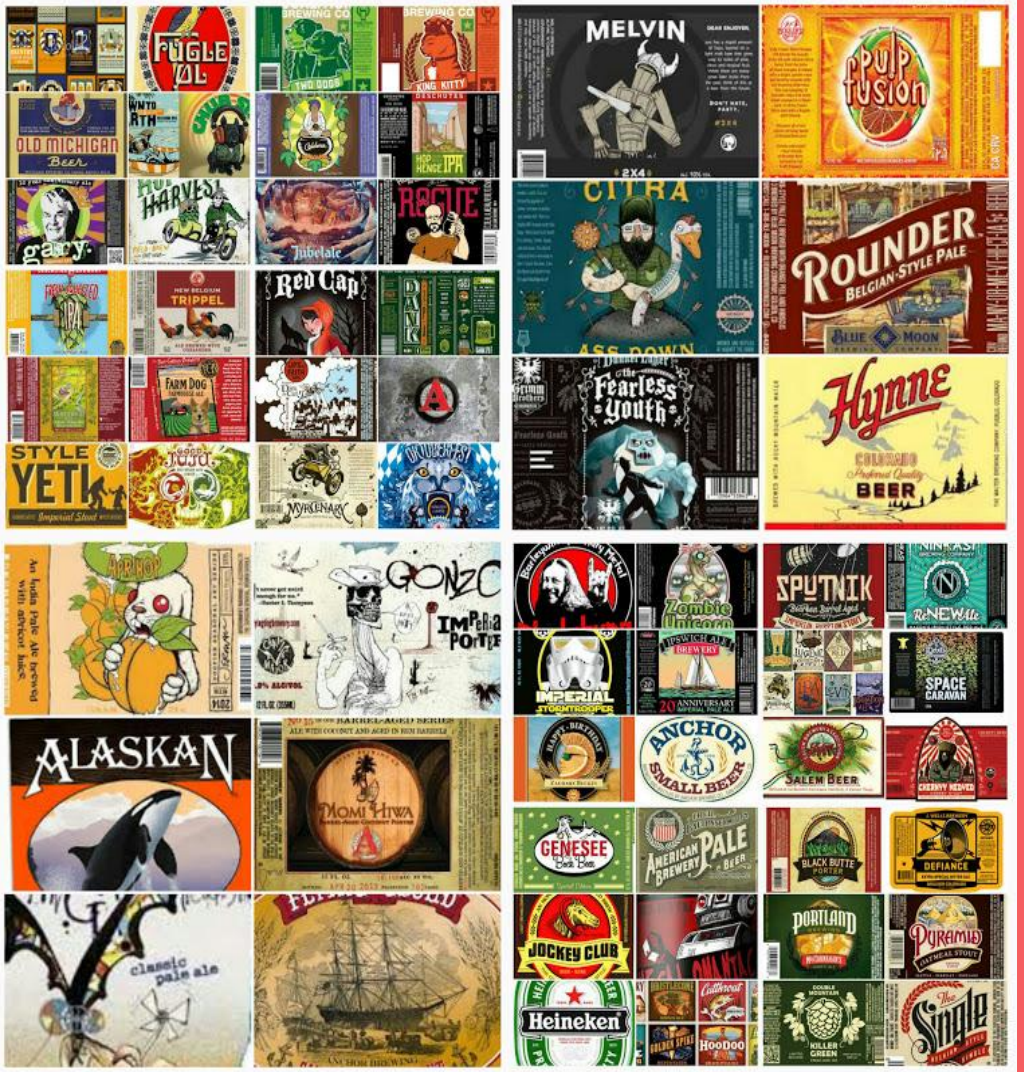


The Beer Genome Project

An extensive analysis of beers for the sake of exploration and understanding

Anthony Gesek

General Assembly

Some people try to recreate dishes from restaurants they like. I like to recreate beer I've tried.

As a beer lover and amateur homebrewer, I know that beer is more than the sum of its parts. There is a lot of nuance to beer recipes.

To recreate beer, I'd need a LOT of information.

What information about beer is publically available?

What is proprietary?

Could I find those missing pieces from what already is published?

Can I recreate the beers that I love at home, using public information?

Data Gathering

API called BreweryDB

beeradvocate.com

ychhops.com

id	0
name	0
abv	11010
ibu	38270
originalGravity	54366
ogMin	7278
isOrganic	1
categoryId	2684
category_name	2684
styleId	2684
style_name	2684
style_description	2868
servingTemperatureDisplay	55576
foodPairings	55945
description	23632
labels	37299
glasswareId	43274
glass_name	43274
srmId	53147
srm_hex	53147
srmMin	8215
srmMax	8260
brewery_id	29
brewery_name	29
brewery_description	10961
brewery_established	13687
brewery_locality	1752
loc_country_code	81

Explore the Data

American-Style India Pale Ale	5453
American-Style Pale Ale	3466
Imperial or Double India Pale Ale	2195
French & Belgian-Style Saison	2000
American-Style Amber/Red Ale	1894
Golden or Blonde Ale	1430
American-Style Brown Ale	1371
Brown Porter	1359
American-Style Imperial Stout	1328
American-Style Stout	1291
South German-Style Hefeweizen / Hefeweissbier	1046
Specialty Beer	959
Belgian-Style White (or Wit) / Belgian-Style Wheat	910
Fruit Beer	855
Herb and Spice Beer	840

Explore the Data

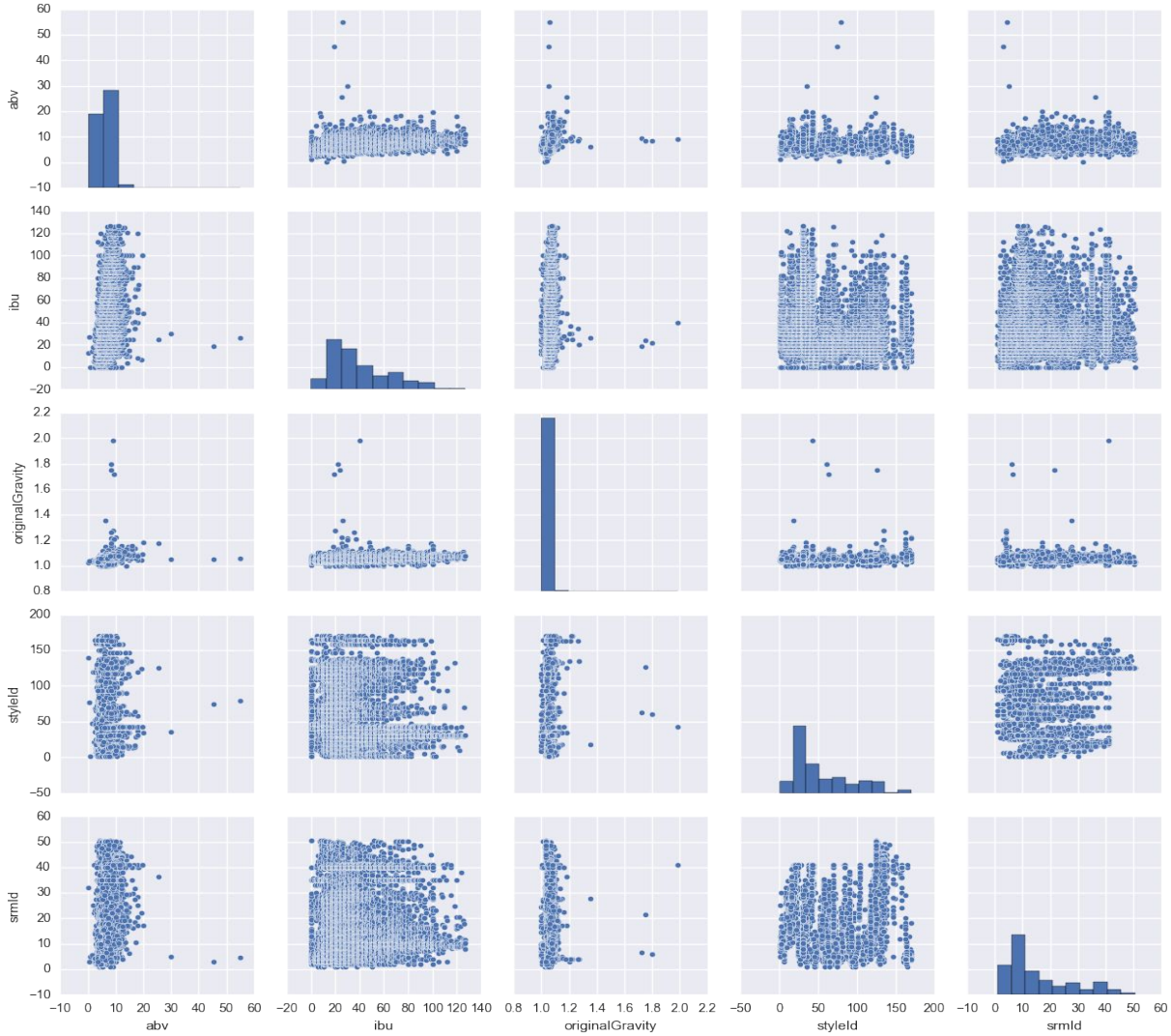
American-Style India Pale Ale	5453
American-Style Pale Ale	3466
Imperial or Double India Pale Ale	2195
French & Belgian-Style Saison	2000
American-Style Amber/Red Ale	1894
Golden or Blonde Ale	1430
American-Style Brown Ale	1371
Brown Porter	1359
American-Style Imperial Stout	1328
American-Style Stout	1291
South German-Style Hefeweizen / Hefeweissbier	1046
Specialty Beer	959
Belgian-Style White (or Wit) / Belgian-Style Wheat	910
Fruit Beer	855
Herb and Spice Beer	840

UNITED STATES	47813
GERMANY	3865
UNITED KINGDOM	1792
CANADA	1404

Explore the Data

Gross! (By definition)

This data set is very spread out without much discernible shape

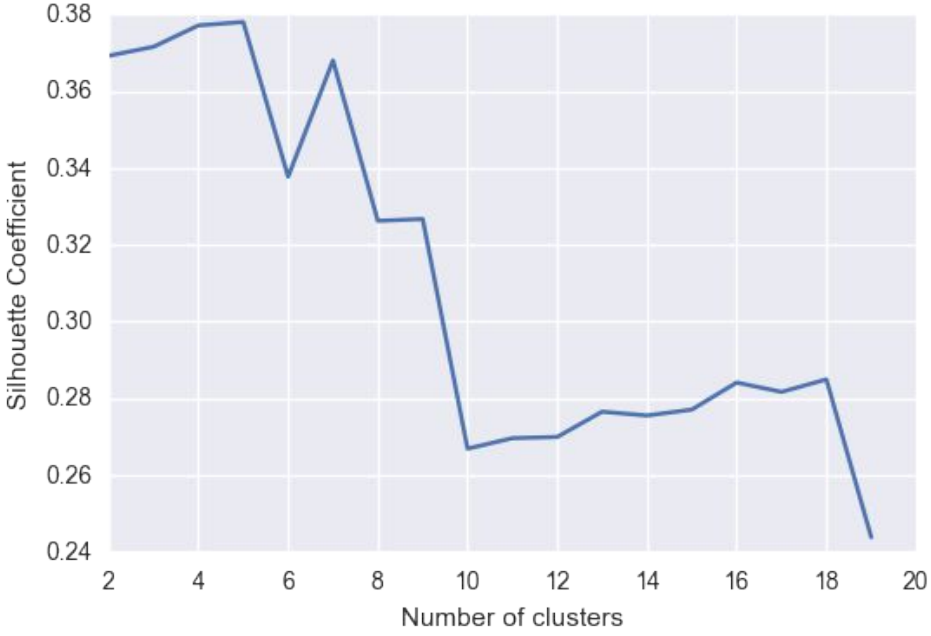


Prediction Time!

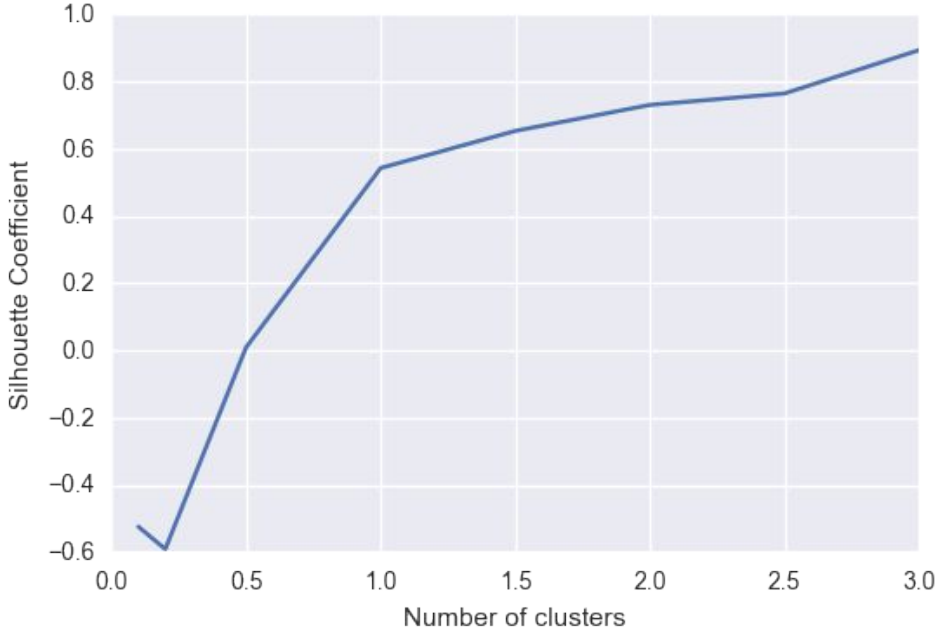
Target	Didn't Work	Kinda Worked	Worked!
Original Gravity (1.000 - 1.099)	KNN(duh)	Linear Regression: RMSE 0.015 BagReg(DecisionTreeReg): RMSE 0.014	RandomForestRegressor: RMSE 0.0135
SRM(color) (1-50)			BagReg(DecisionTreeReg) : RMSE 5.009
ABV (1-25)		DecisionTreeReg: RMSE 1.475	RandomForestRegressor: RMSE 1.185

Cluster Attempts

K-Means



DB Scan



What About Other Information?

Ingredients:

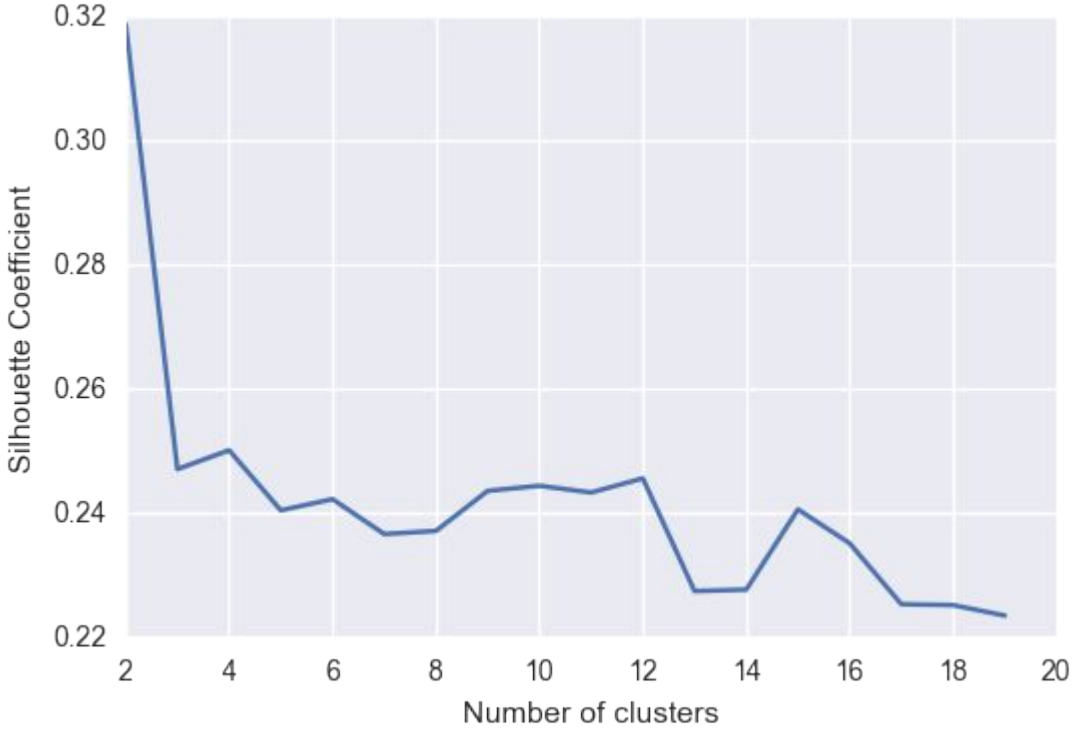
-Hops

-Malted Barley,Oats, etc.

id	0
name_x	0
abv_x	0
ibu_x	0
originalGravity_x	0
isOrganic_x	0
styleId_x	0
style_name_x	0
style_description_x	0
description	0
srmId_x	0
loc_country_code	0
hop_names	0
full_beer_name	0
hop_alphaacid	0
hop_betaacid	0
hop_caryophyllene	0
hop_myrcene	0
hop_humulene	0
hop_cohumulone	0
hop_geraniol	0
hop_totaloil	0

Cluster with Ingredients

K - Means for k = 2 - 20



0	abv_x	ibu_x	originalGravity_x	srmId_x
count	841.000000	841.000000	841.000000	841.000000
mean	6.299962	56.940166	1.054741	10.024095
std	0.953722	18.055757	0.010368	4.568949
min	3.400000	1.090000	1.000000	2.000000
25%	5.700000	45.000000	1.047174	8.938776
50%	6.400000	58.000000	1.059126	9.914286
75%	7.000000	70.000000	1.060494	10.165576
max	10.000000	108.000000	1.084682	41.000000

1	abv_x	ibu_x	originalGravity_x	srmId_x
count	473.000000	473.000000	473.000000	473.000000
mean	7.204217	44.610782	1.056896	34.771501
std	1.771679	21.415869	0.015339	7.481522
min	3.200000	3.500000	1.008000	17.039184
25%	6.000000	27.000000	1.045290	28.588605
50%	6.900000	42.000000	1.055809	35.024490
75%	8.000000	61.000000	1.066000	40.000000
max	14.500000	109.000000	1.101000	50.267347

American-Style India Pale Ale	424
American-Style Pale Ale	142
American-Style Amber/Red Ale	34
French & Belgian-Style Saison	24
Session India Pale Ale	24
Imperial or Double India Pale Ale	18
Session Beer	16

American-Style Black Ale	61
American-Style Stout	49
American-Style Imperial Stout	41
Brown Porter	32
Specialty Beer	32
Robust Porter	27
German-Style Doppelbock	15

2	abv_x	ibu_x	originalGravity_x	srmId_x
count	677.000000	677.000000	677.000000	677.000000
mean	8.866544	84.408227	1.074658	11.675216
std	1.943298	25.372273	0.010137	6.056075
min	5.300000	10.000000	1.044436	3.000000
25%	8.000000	71.000000	1.069000	9.000000
50%	8.800000	85.000000	1.075037	9.829640
75%	9.500000	99.000000	1.079260	12.500000
max	45.250000	280.000000	1.117000	41.000000

3	abv_x	ibu_x	originalGravity_x	srmId_x
count	1560.000000	1560.000000	1560.000000	1560.000000
mean	5.414060	30.647401	1.046753	9.544875
std	0.842588	13.667792	0.007301	5.600200
min	1.059000	1.068000	1.000000	1.000000
25%	4.900000	20.000000	1.043977	5.000000
50%	5.300000	28.000000	1.045516	9.000000
75%	5.900000	40.000000	1.050000	13.095153
max	9.000000	112.000000	1.082018	30.912566

Imperial or Double India Pale Ale	341
American-Style India Pale Ale	146
American-Style Barley Wine Ale	34
Imperial Red Ale	33
Belgian-Style Tripel	20
American-Style Imperial Stout	11
British-Style Barley Wine Ale	9

American-Style Pale Ale	238
American-Style Amber/Red Ale	116
Golden or Blonde Ale	98
French & Belgian-Style Saison	60
American-Style Brown Ale	59
American-Style India Pale Ale	56
German-Style Kölsch / Köln-Style Kölsch	45

4	abv_x	ibu_x	originalGravity_x	srmId_x
count	788.000000	788.000000	788.000000	788.000000
mean	5.409712	24.196475	1.047265	9.627120
std	0.977771	13.208652	0.008496	6.590022
min	2.700000	1.000000	1.000000	1.000000
25%	4.837500	15.000000	1.043316	5.000000
50%	5.200000	22.000000	1.046965	7.768385
75%	5.800000	30.000000	1.051612	12.873003
max	9.700000	85.000000	1.074000	41.428571

5	abv_x	ibu_x	originalGravity_x	srmId_x
count	374.000000	374.000000	374.000000	374.000000
mean	6.095268	36.606016	1.052155	14.545560
std	1.285706	18.733931	0.011145	9.946376
min	2.300000	1.054000	1.009000	2.000000
25%	5.100000	23.000000	1.044515	7.236735
50%	5.880000	32.000000	1.049460	11.000000
75%	6.800000	45.000000	1.059798	20.000000
max	11.100000	107.000000	1.106000	49.864286

South German-Style Hefeweizen / Hefeweissbier 69
French & Belgian-Style Saison 63
Belgian-Style White / Belgian-Style Wheat 44
German-Style Kölsch / Köln-Style Kölsch 29
Berliner-Style Weisse (Wheat) 27
Irish-Style Red Ale 23
American-Style Pale Ale 22

American-Style India Pale Ale 50
American-Style Amber/Red Ale 31
American-Style Pale Ale 31
American-Style Brown Ale 19
Light American Wheat Ale or Lager with Yeast 15
French & Belgian-Style Saison 14
Brown Porter 13

6	abv_x	ibu_x	originalGravity_x	srmId_x
count	987.000000	987.000000	987.000000	987.000000
mean	5.601369	32.000051	1.048666	10.273072
std	0.963214	15.330979	0.008246	5.905643
min	3.200000	2.000000	1.006000	1.000000
25%	5.000000	21.000000	1.044165	5.000000
50%	5.500000	29.000000	1.047644	9.738440
75%	6.050000	40.000000	1.053714	14.000000
max	9.200000	100.000000	1.077226	33.000000

7	abv_x	ibu_x	originalGravity_x	srmId_x
count	1047.000000	1047.000000	1047.000000	1047.000000
mean	5.749408	41.039035	1.049606	10.394108
std	0.880279	17.912916	0.009337	5.006523
min	1.600000	1.050000	1.000000	2.000000
25%	5.100000	28.150000	1.044380	7.195374
50%	5.600000	39.000000	1.047972	9.935107
75%	6.300000	52.500000	1.057000	12.000000
max	9.000000	110.000000	1.083191	41.000000

American-Style India Pale Ale	78
American-Style Pale Ale	77
American-Style Amber/Red Ale	53
German-Style Pilsener	39
French & Belgian-Style Saison	37
Extra Special Bitter	35
Golden or Blonde Ale	33

American-Style Pale Ale	285
American-Style India Pale Ale	197
American-Style Amber/Red Ale	85
Golden or Blonde Ale	58
French & Belgian-Style Saison	37
Light American Wheat Ale or Lager with Yeast	23
American-Style Brown Ale	22

8	abv_x	ibu_x	originalGravity_x	srmId_x
count	421.000000	421.000000	421.000000	421.000000
mean	7.663014	35.342280	1.063131	29.549107
std	2.068023	18.050302	0.016425	11.268019
min	3.100000	5.000000	1.015000	4.000000
25%	6.000000	24.000000	1.050763	21.058603
50%	7.500000	30.300000	1.062000	29.987755
75%	9.000000	42.000000	1.074000	40.000000
max	16.200000	118.000000	1.111000	50.500000

9	abv_x	ibu_x	originalGravity_x	srmId_x
count	976.000000	976.000000	976.000000	976.000000
mean	7.925840	75.854406	1.068214	10.917842
std	1.448337	23.079573	0.011655	4.295099
min	4.100000	4.500000	1.039700	2.000000
25%	6.854312	65.000000	1.060031	9.000000
50%	7.700000	75.000000	1.064059	9.989047
75%	8.900000	90.000000	1.075289	11.532162
max	15.000000	200.000000	1.153656	40.000000

American-Style Imperial Stout	23
American-Style Stout	21
Brown Porter	19
Robust Porter	19
Belgian-Style Dubbel	18
Specialty Beer	17
Classic Irish-Style Dry Stout	17

American-Style India Pale Ale	450
Imperial or Double India Pale Ale	243
American-Style Barley Wine Ale	43
Imperial Red Ale	28
Belgian-Style Tripel	22
American-Style Amber/Red Ale	17
American-Style Pale Ale	11

10	abv_x	ibu_x	originalGravity_x	srmId_x
count	399.000000	399.000000	399.000000	399.000000
mean	6.866756	42.130674	1.056896	36.505549
std	1.558232	19.707396	0.014319	6.433368
min	3.300000	1.059000	1.014000	19.561224
25%	5.700000	28.000000	1.046153	32.000000
50%	6.700000	39.000000	1.055850	40.000000
75%	7.700000	53.000000	1.067923	41.000000
max	14.000000	111.000000	1.103000	50.500000

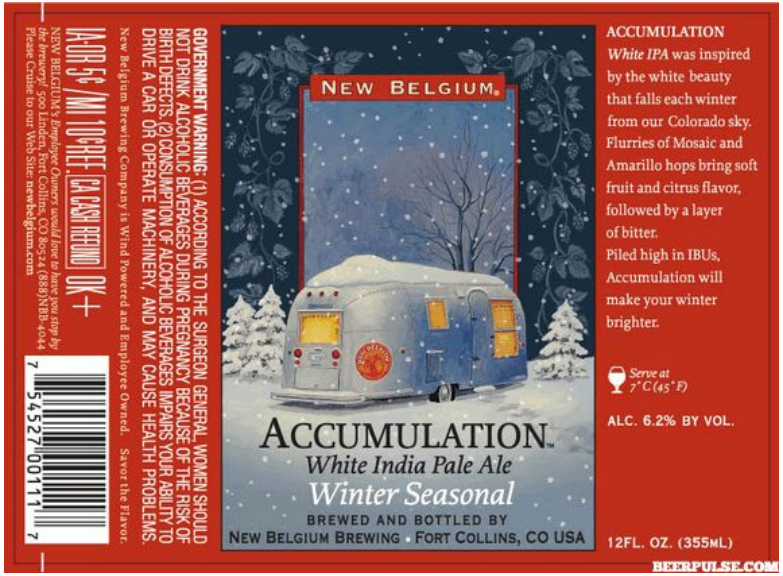
American-Style Black Ale	46
American-Style Stout	41
Brown Porter	35
American-Style Imperial Stout	33
Specialty Beer	29
Robust Porter	26
Sweet or Cream Stout	20

11	abv_x	ibu_x	originalGravity_x	srmId_x
count	172.000000	172.000000	172.000000	172.000000
mean	8.764754	52.895349	1.069598	20.024846
std	2.621238	26.850569	0.023060	13.634852
min	4.200000	5.000000	1.011000	3.000000
25%	6.875000	30.000000	1.055244	9.025017
50%	8.650000	49.500000	1.071623	15.321939
75%	10.000000	73.500000	1.080186	33.147213
max	18.000000	120.000000	1.214114	50.035714

Imperial or Double India Pale Ale	21
American-Style Imperial Stout	18
American-Style India Pale Ale	17
American-Style Barley Wine Ale	9
French & Belgian-Style Saison	7
British-Style Barley Wine Ale	7
American-Style Pale Ale	5

Examples!

Accumulation is an example of a beer that fits into Cluster 0 vey well. It's an IPA that is light in color(10 srm) with 6.2% ABV and 70 IBU



This Mississippi Red beer is a bit of an outlier in Cluster 0 as it is low in IBU(29) but is close in color(12.5 srm) and ABV(5.8%)

What really brings these two together similarities in Caryophyllene, Geraniol, and Cohumulone oils.(Aroma)

Brew Time!

All that work brings us back to the original question:

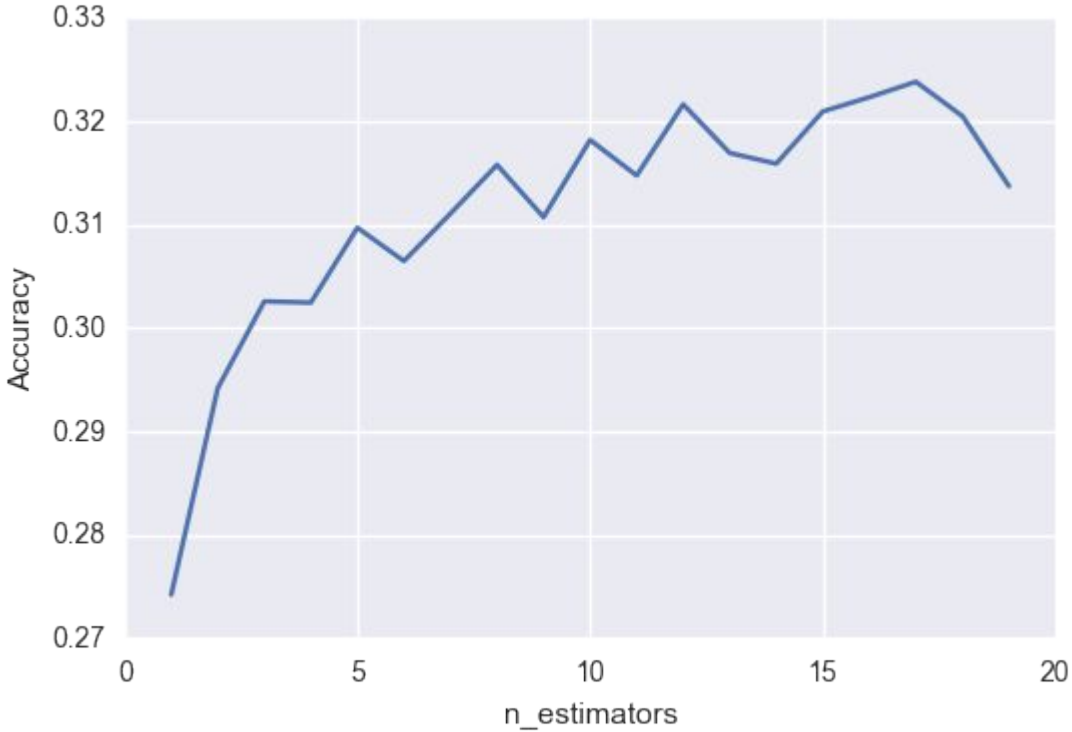
Can I Brew These at Home Now?

Can machine learning help with this?

Clustering created a classification system where I didn't have one before.

RandomForests

Grid Search a RandomForestClassifier to see if it is possible to predict which cluster a beer would fit in given a desired ABV, IBU, SRM and Style



Results!

Using the RandomForest Model I now have a list of hops that I can use to create a beer with my desired characteristics.

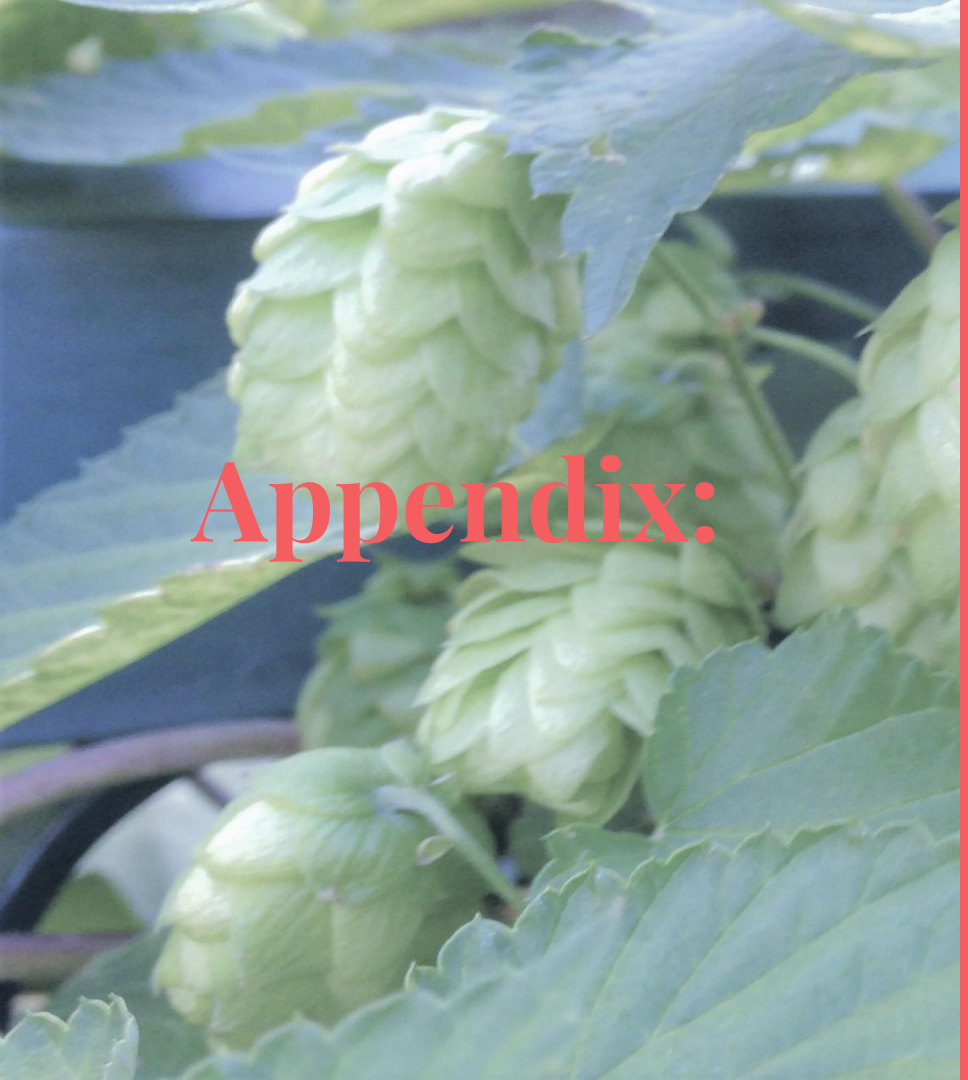
```
new_test = [5,75,30,50]
randc = RandomForestClassifier(n_estimators = 17)
randc.fit(X,y)
```

```
probabs = randc.predict_proba(np.array(new_test))
cluster = np.where(probabs == probabs.max())[1]
print cluster
print probabs
```

```
[10]
[[ 0.          0.35294118  0.11764706  0.          0.          0.          0.
   0.          0.05882353  0.          0.47058824  0.        ]]
```

```
print beers_US_hops.groupby(beers_US_hops['cluster'] == int(cluster)).hop_names.value_counts()[1].head()
```

```
hop_names
Hops                396
Columbus             10
Centennial Hops       9
Chinook Hops          7
Cluster Hops          5
Name: hop_names, dtype: int64
```



Appendix:

Issues and To Do's:

-Still a lot of missing data(original list of 59,000 was cut down to 8,700 beers to fit a model to)

-Relies heavily upon machine learning models for some of the missing data

-Fine tune the accuracy of prediction models

-Add Malt information