Predicting brewery locations

with Machine Learning

Sarah Robinson

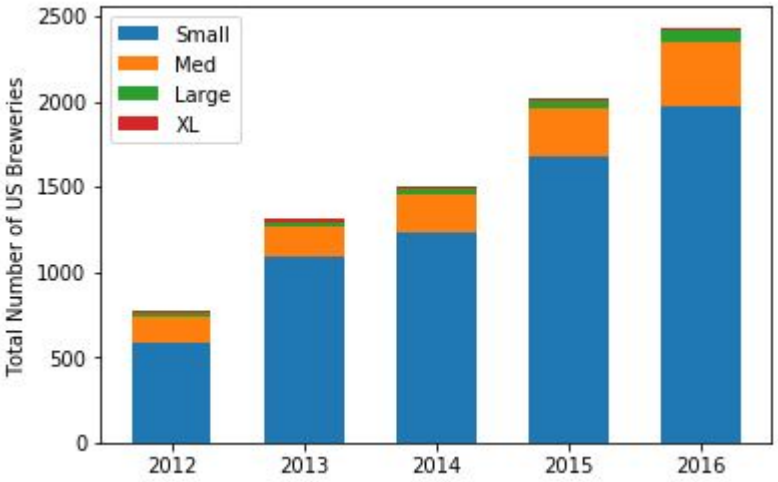
Breweries on the Rise

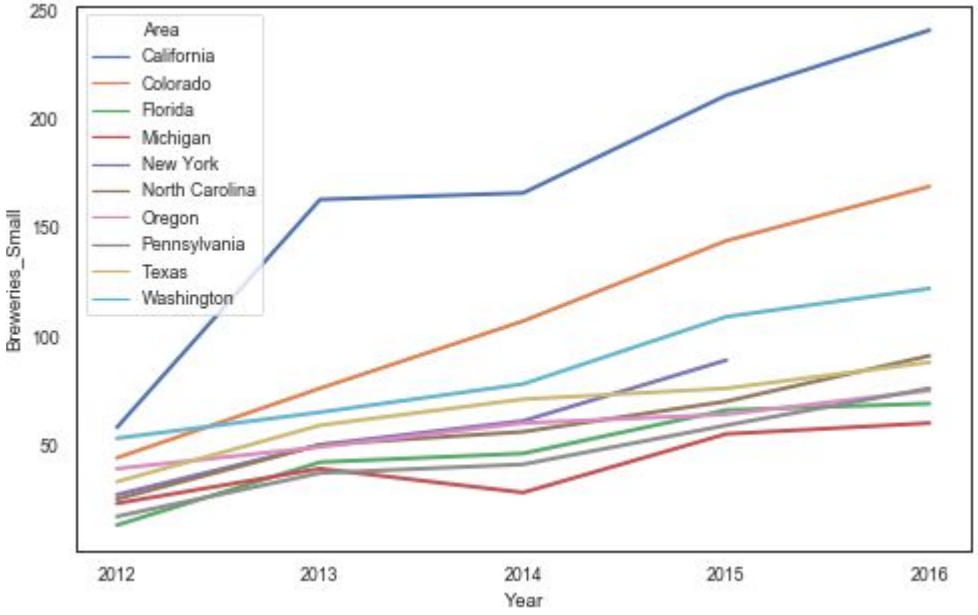
The number of US craft breweries and brewpubs:

2013	2014	2015	2016	2017
2,898	3,739	4,544	5,424	6,266

[According to the BreweriesAssociation.org]

Breweries on the Rise





The question:

Can we make accurate predictions about where breweries will thrive?

This would allow brewery, or potential brewery, owners to make informed business decisions about location.

U.S. Department of Commerce

United States Census Bureau

AMERICAN FactFinder

The Data

FRED®

ECONOMIC DATA | ST. LOUIS FED

Brewery Sizes

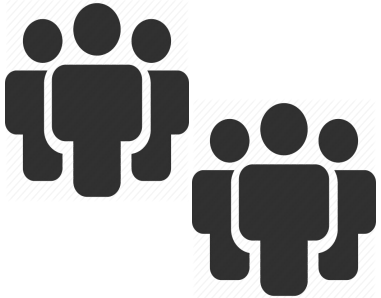
(and other business sizes)

SMALL:

1-19 employees

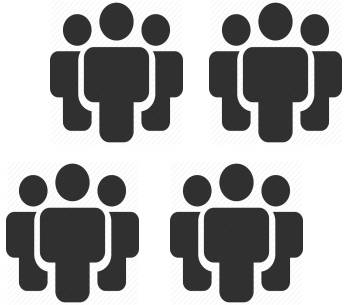
MEDIUM:

20-99 employees



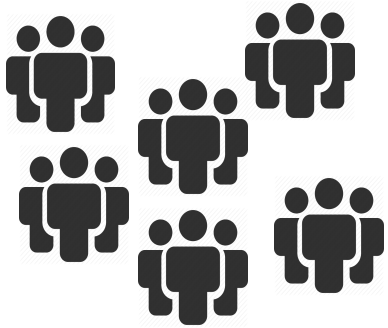
LARGE:

100-499 employees

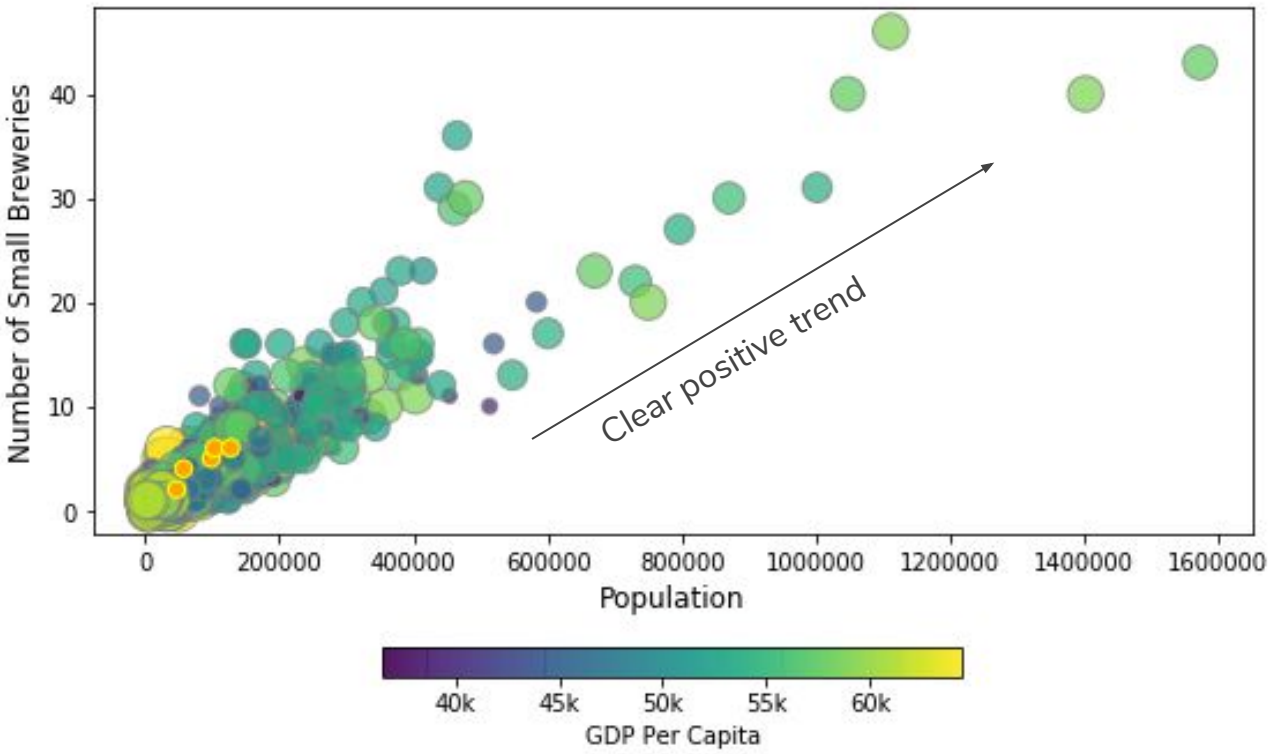


X-LARGE :

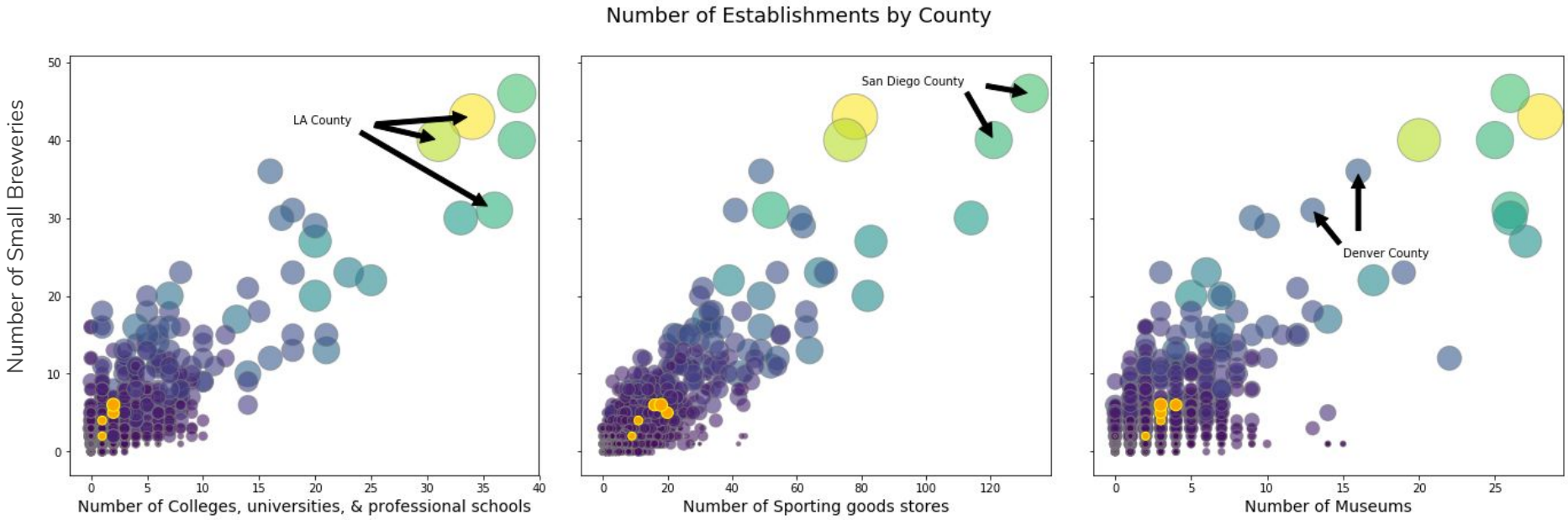
500 + employees



Population

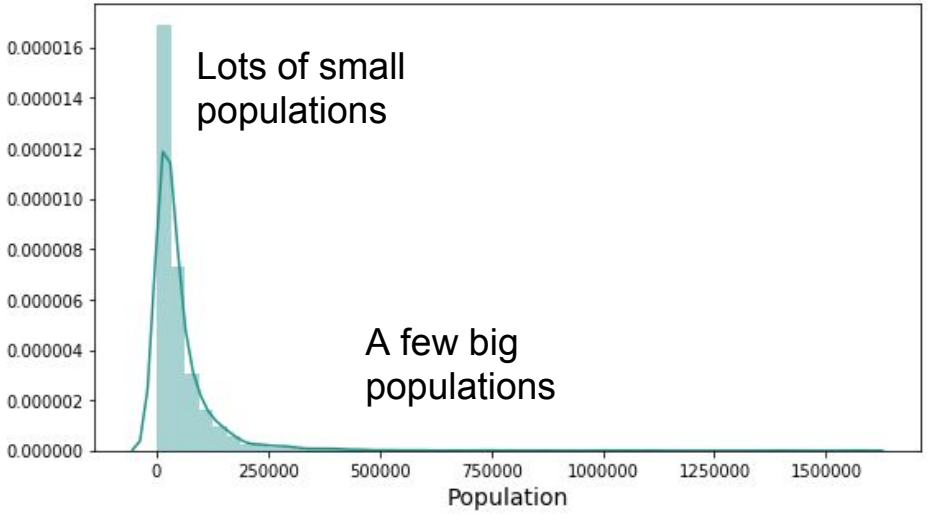

Number of Small Breweries
0
10
20
30
40
0
200000
400000
600000
800000
1000000
1200000
1400000
1600000
Population
Clear positive trend
40k
45k
50k
55k
60k
GDP Per Capita

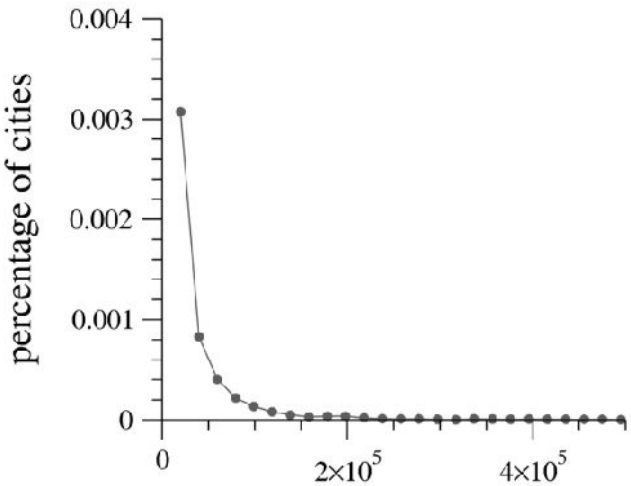
Other businesses as they relate to number of breweries and population (bubble size and color)

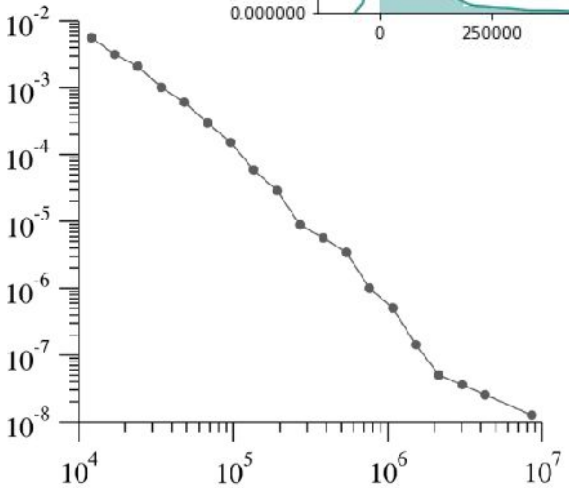


Orange bubbles indicate Santa Cruz county

Zipf's Law
(power law)





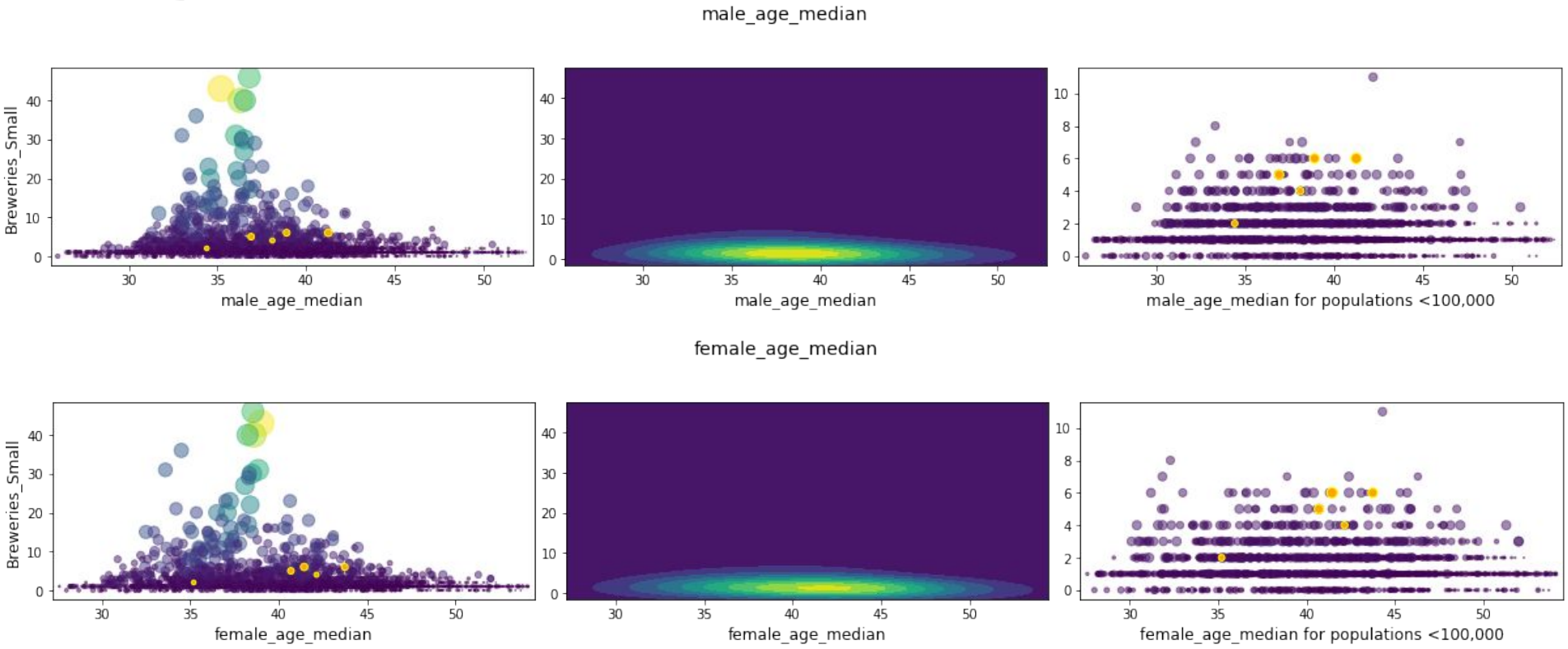


population of city

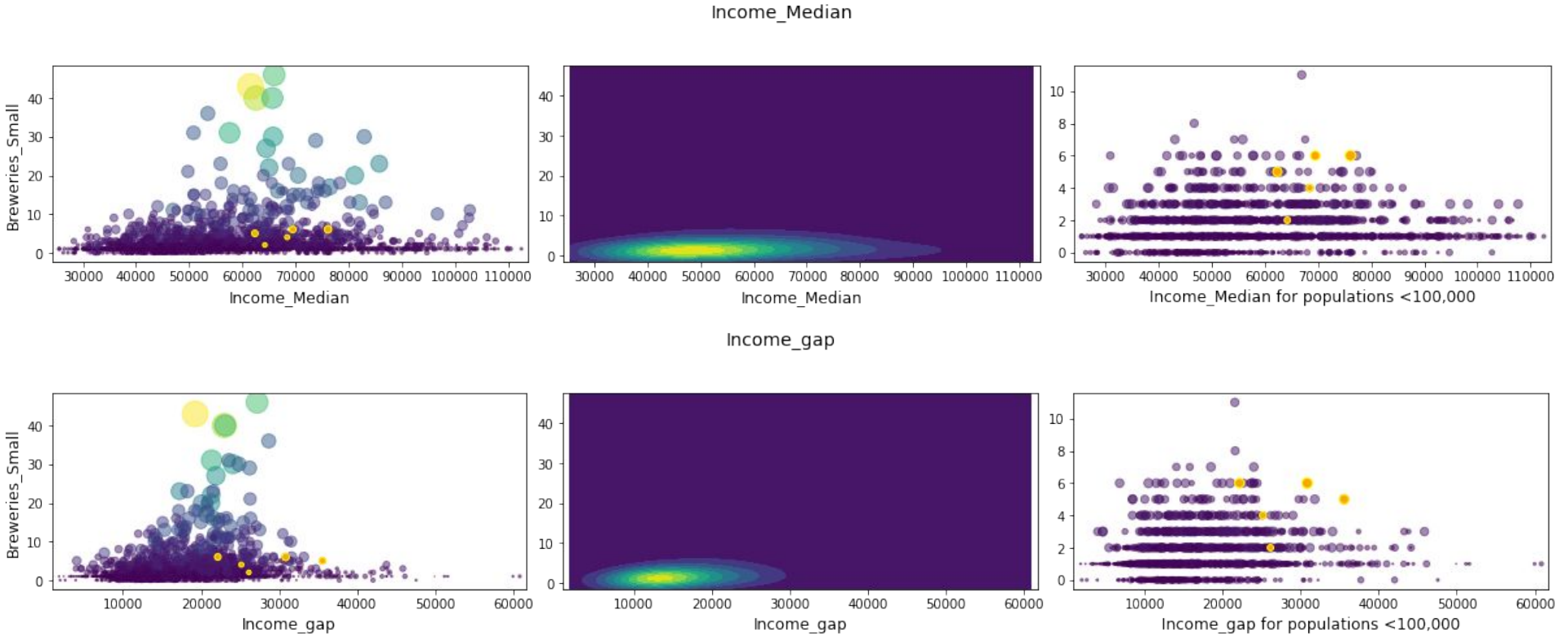
Why is this important?

Small populations are more likely to be noisy and affected by outliers.

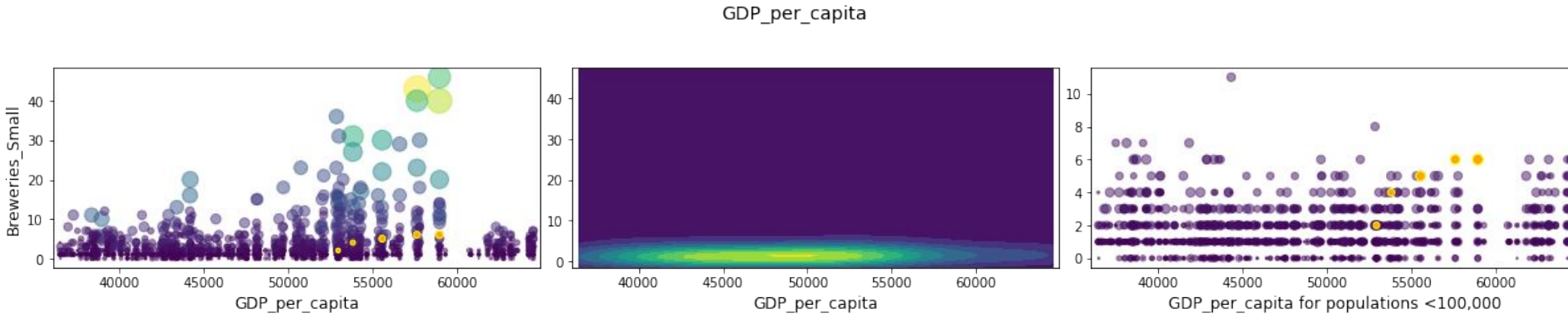
AGE


male_age_median
Breweries_Small
male_age_median
male_age_median
male_age_median for populations <100,000
female_age_median
Breweries_Small
female_age_median
female_age_median
female_age_median for populations <100,000

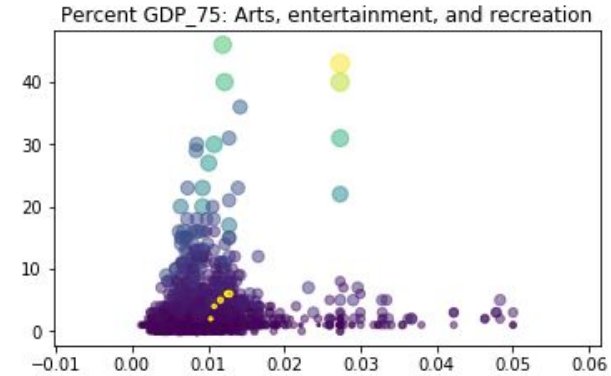
Income

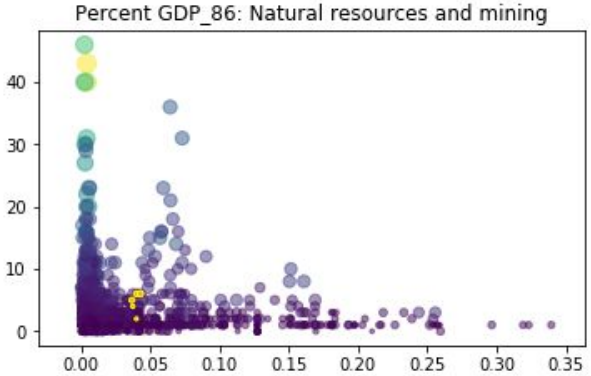

Income_Median
Breweries_Small
Income_Median
Income_Median
Income_Median for populations <100,000
Income_gap
Breweries_Small
Income_gap
Income_gap
Income_gap for populations <100,000

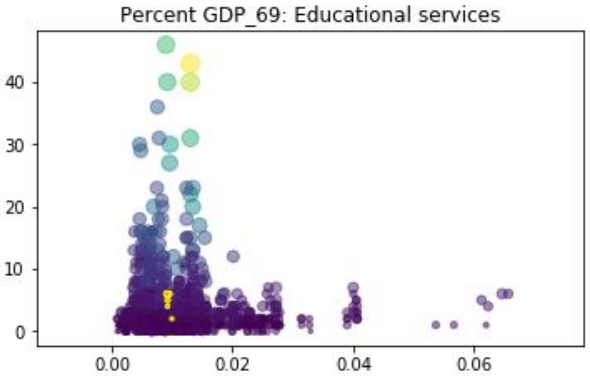
GDP per Capita

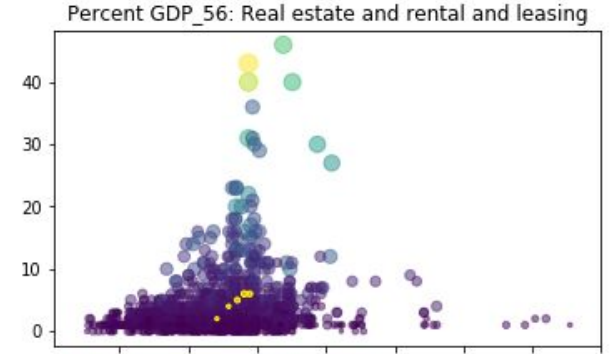


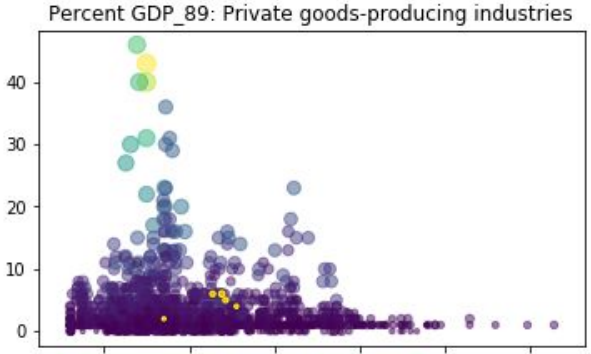
Industries and GDP

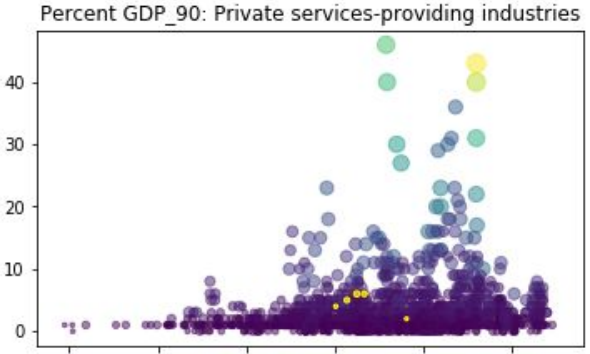








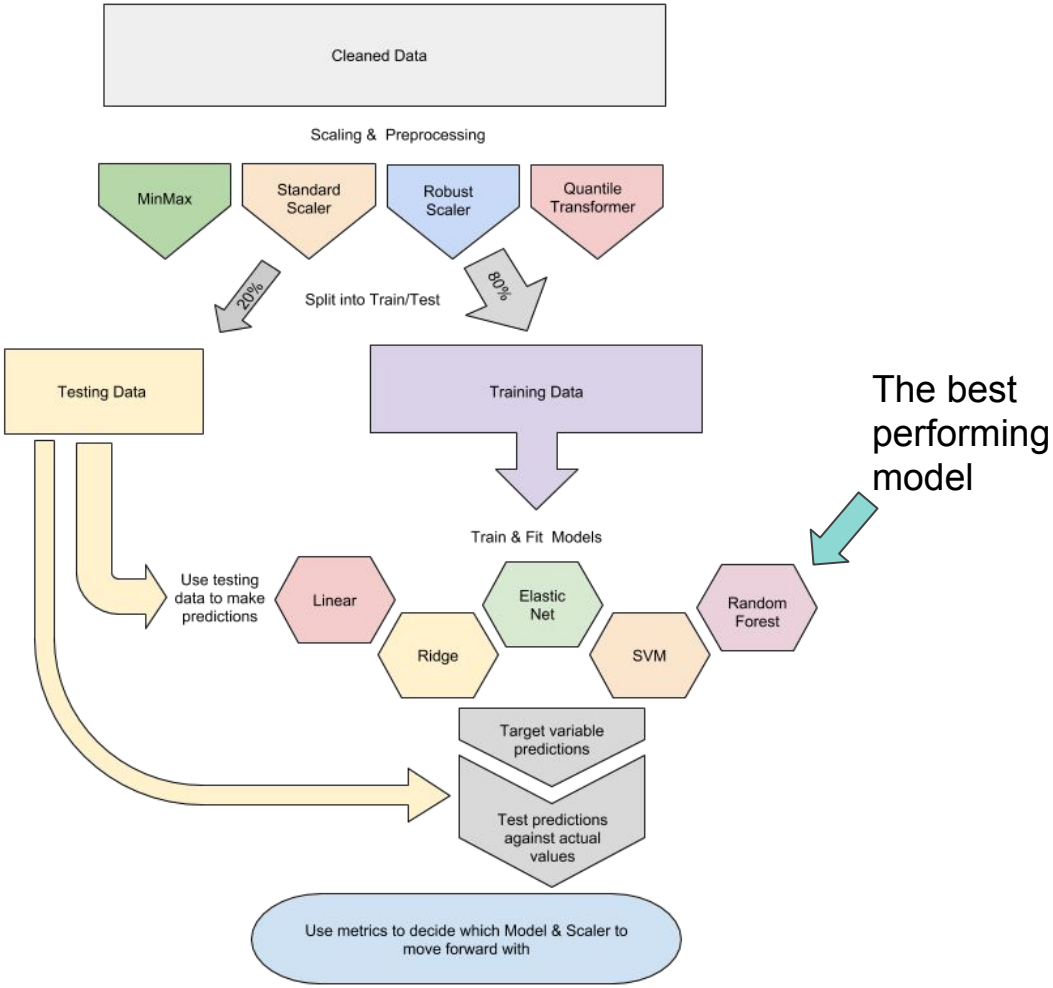




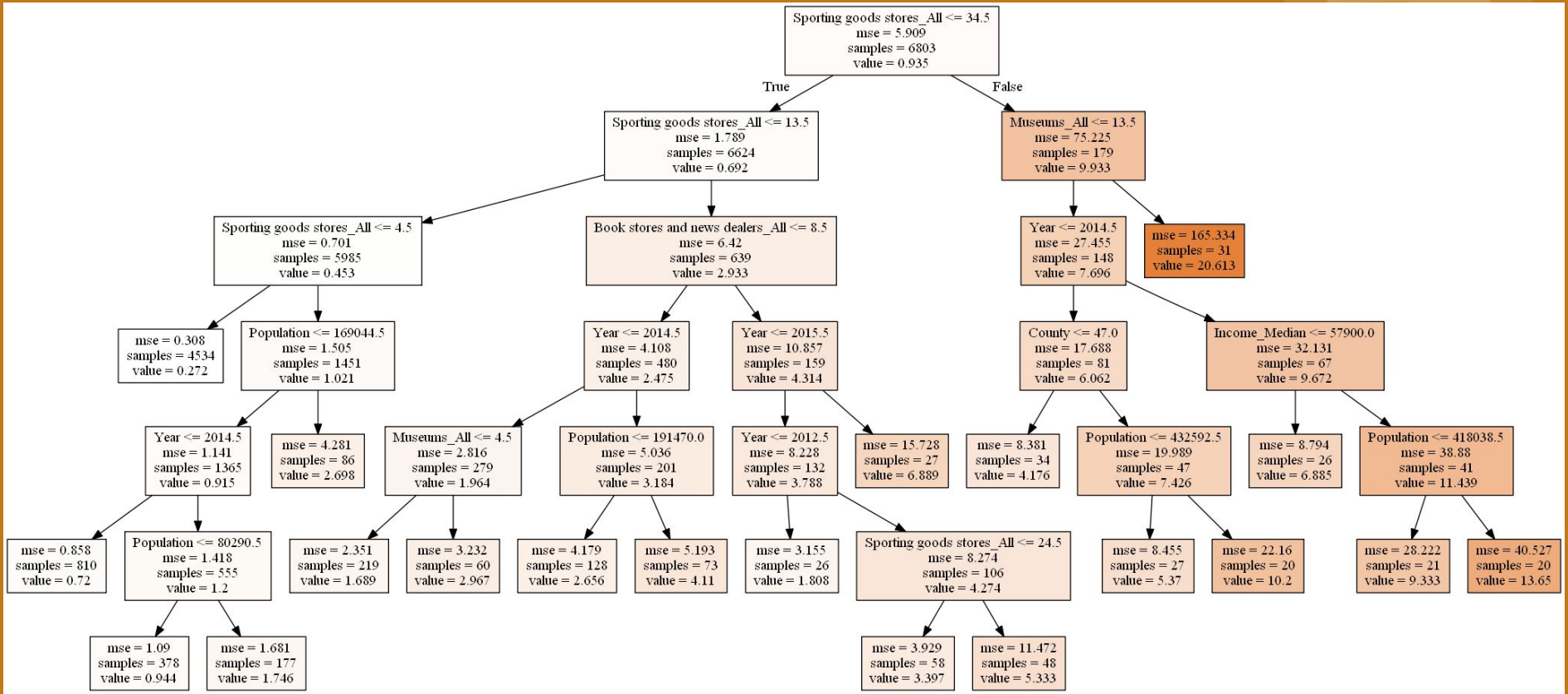
Making Predictions with Machine Learning

My Process

1. Try different methods of scaling combined with different model types
2. Select the combination with the lowest error
3. Tune parameters to make this model the best it can be



Decision tree (simplified)



Predictions for Santa Cruz

	Year	all_age_median	population	Income_Median	Breweries_Small	predictions
466	2012	40.00	266794.0	72250.5	2.0	1.961667
467	2013	43.70	269395.0	83507.0	4.0	3.898333
468	2014	41.60	271529.0	69872.0	5.0	3.723333
469	2015	41.65	274299.0	69476.0	6.0	3.045000
470	2016	42.50	275196.0	77677.5	6.0	5.648333

Next time...

- Use rank data instead of actual numbers (because of Zipfs Law)
- Try different data sources (like the brewers association or google places API)
- Try Gradient boosting