COVID-19: Coming soon to a county near you?

Project 2

Julian Cheng

Table of Contents

The question:

Is it possible to predict COVID-19 cases in United States counties using county-wide statistics?

My answer:

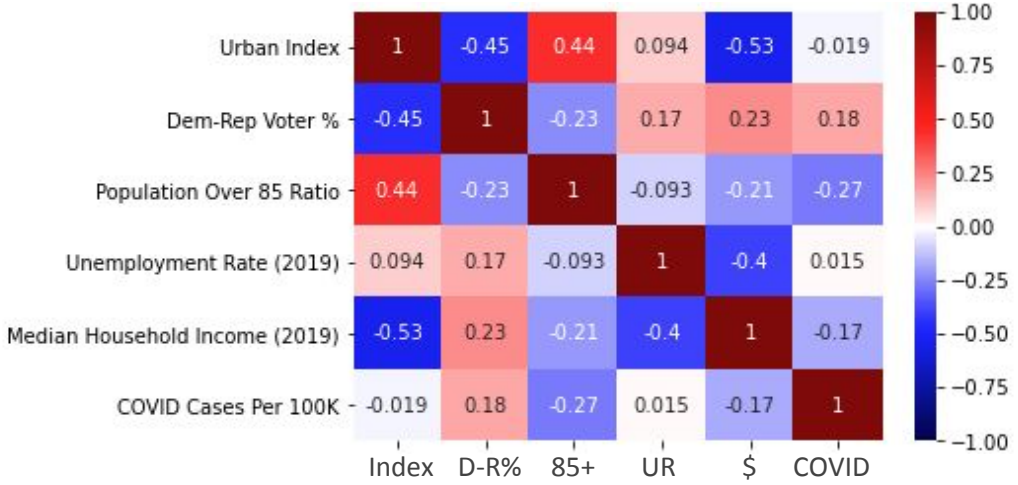
I developed a model that can predict how many COVID-19 cases per 100,000 individuals occur in each county

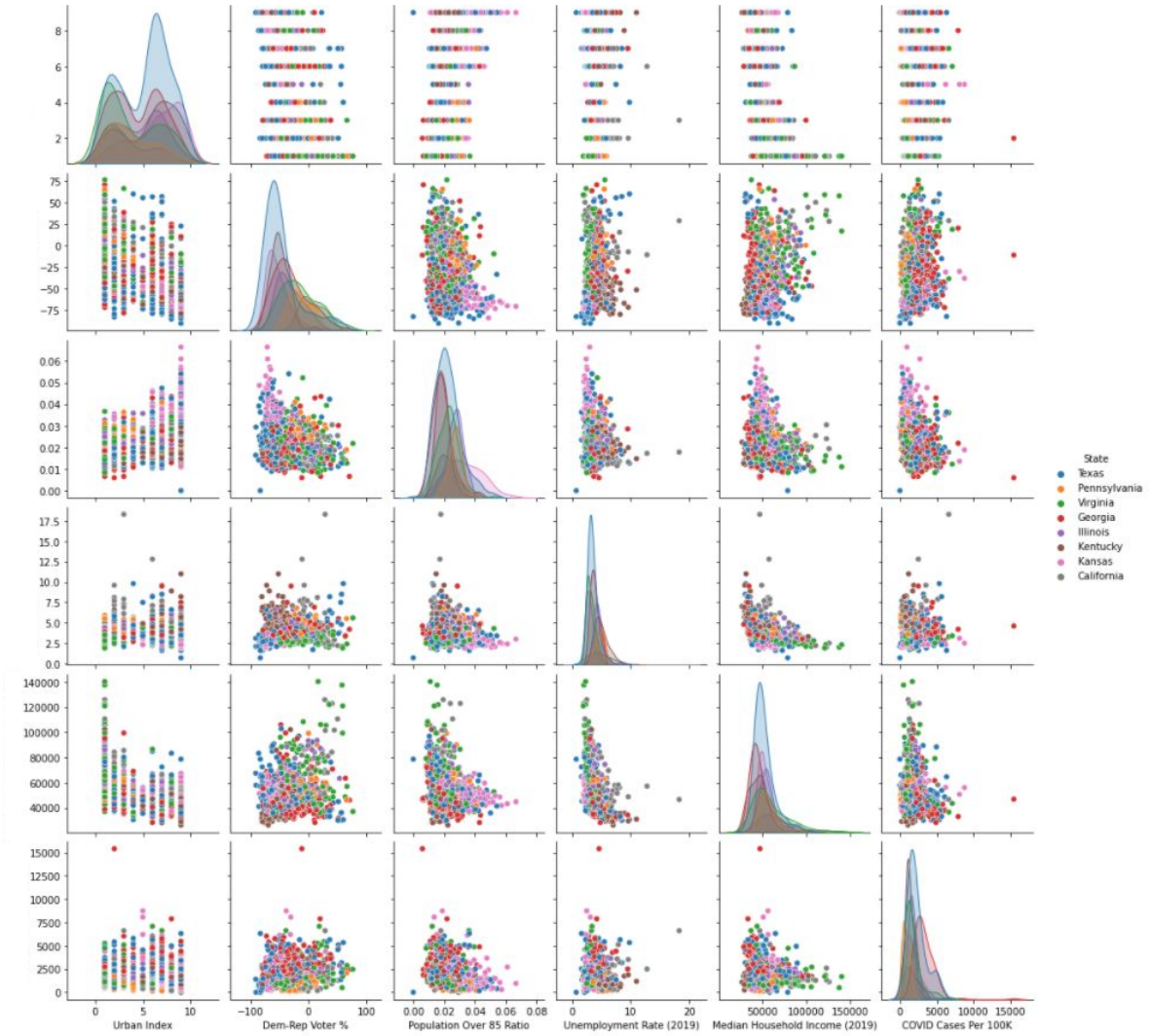
Methodology - Features

- State

Methodology - Collinearity

- Urban Index
- Vulnerable Populations
- Mean Household Income



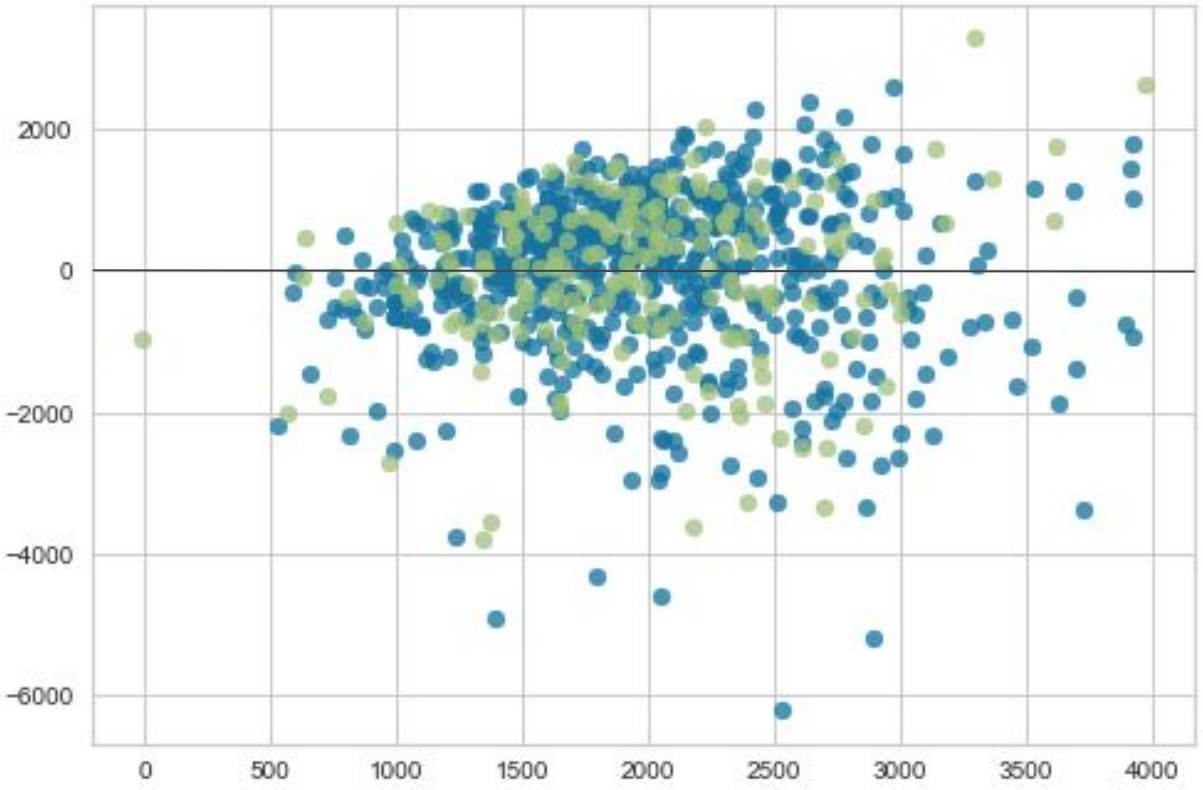


Methodology - Outliers

- Positive Cases
- Unemployment
- 85+ Population

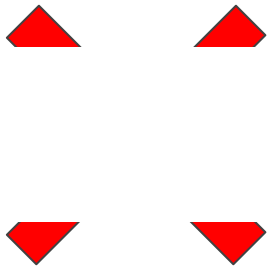
The Process

Results - Residual Plot



MAE = 811.120

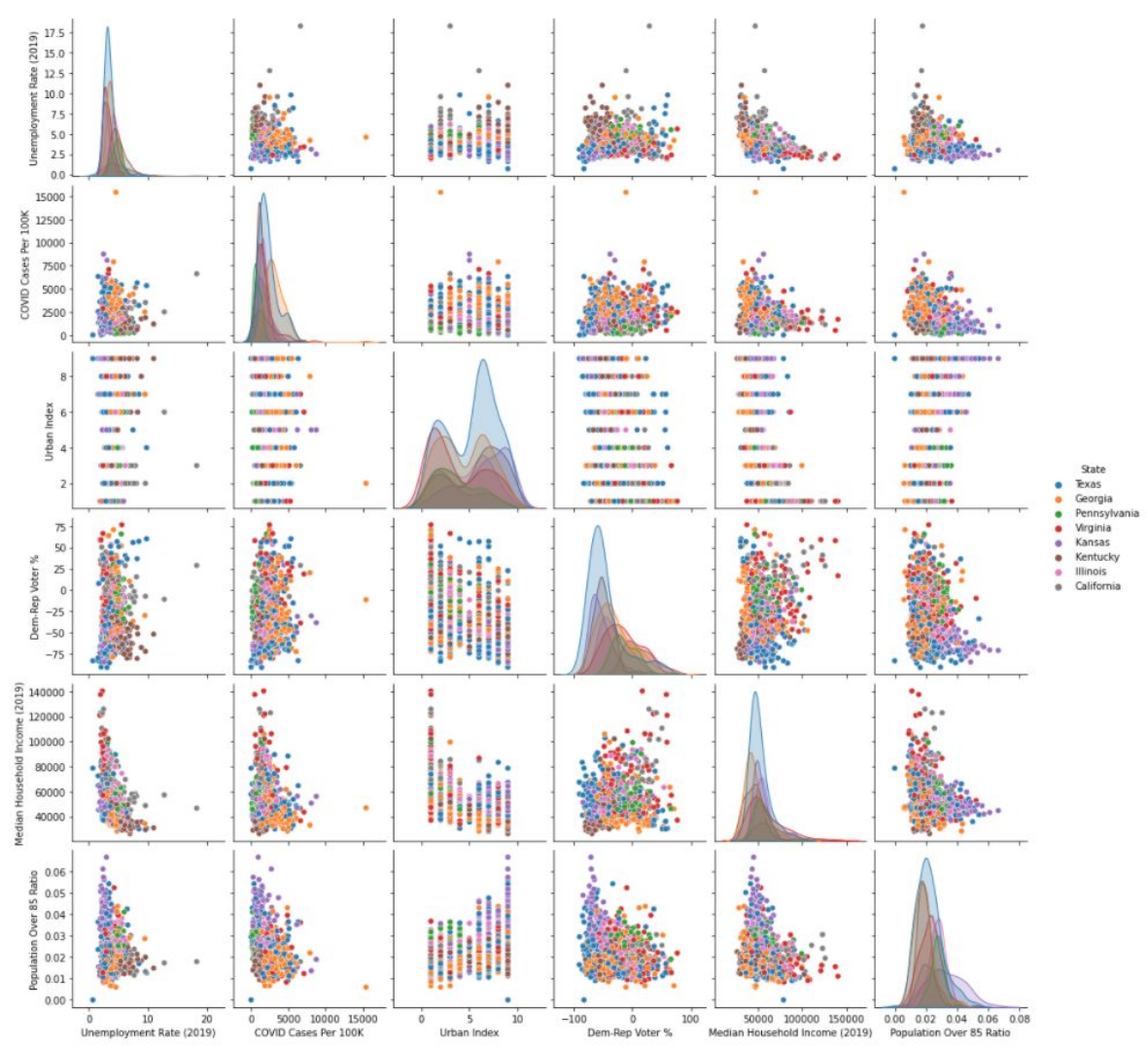
Next Steps

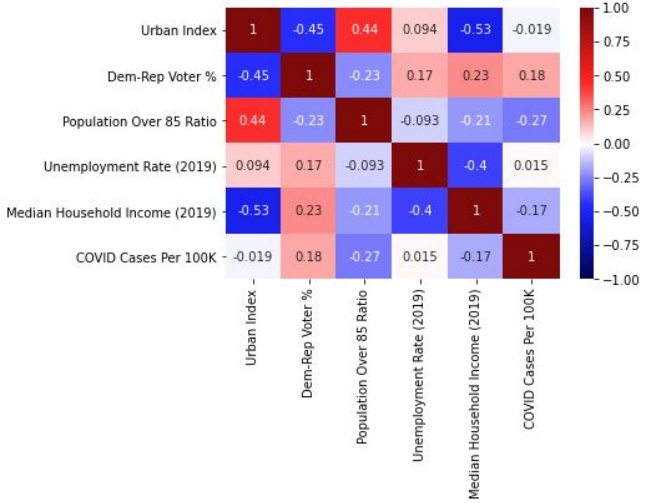


Appendix - Sources:

Density - USA.com
Urban Index - Department of Agriculture
Unemployment % - Bureau of Labor Statistics
Household Income - Bureau of Labor Statistics
Political Distribution - Wikipedia
Vulnerable Populations - US Census Bureau
Racial Demographics - US Census Bureau
COVID-19 Cases - USAfacts.org

Appendix - Tables and Graphs



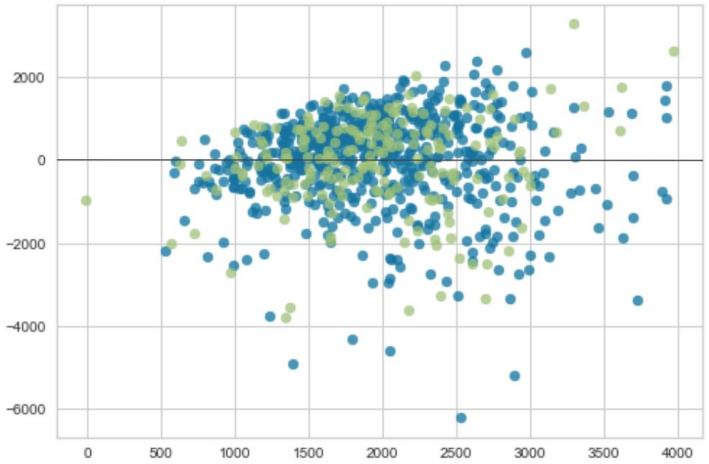


Appendix - Tables and Graphs

Dep. Variable:	Y	**R-squared:**	0.208
Model:	OLS	**Adj. R-squared:**	0.204
Method:	Least Squares	**F-statistic:**	52.02
Date:	Thu, 08 Oct 2020	**Prob (F-statistic):**	5.80e-48
Time:	22:22:28	**Log-Likelihood:**	-8401.8
No. Observations:	994	**AIC:**	1.682e+04
Df Residuals:	988	**BIC:**	1.685e+04
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	6050.0575	319.706	18.924	0.000	5422.677	6677.438
X1	45.7721	18.166	2.520	0.012	10.124	81.420
X2	10.6385	1.296	8.207	0.000	8.095	13.182
X3	-5.118e+04	4736.824	-10.804	0.000	-6.05e+04	-4.19e+04
X4	-282.0718	34.525	-8.170	0.000	-349.823	-214.321
X5	-0.0311	0.003	-9.822	0.000	-0.037	-0.025

Omnibus:	218.033	**Durbin-Watson:**	1.890
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	497.045
Skew:	1.191	**Prob(JB):**	1.17e-108
Kurtosis:	5.515	**Cond. No.**	7.22e+06




```
Linear Regression test R^2: 0.204
Ridge Regression test R^2: 0.123
Lasso Regression test R^2: 0.193
Degree 2 polynomial regression test R^2: 0.242

Linear Regression val R^2: 0.220
Ridge Regression val R^2: 0.220
Lasso Regression val R^2: 0.219
Degree 2 polynomial regression val R^2: 0.278
```