Predicting States' Political Alignment w/ Consumerism trends

Dillon Clark, Max Gannett, Jay Ghosh, Tanner Richard

Project Motivation:
How do we compare to polls?

- Traditional Polling methods are highly inaccurate and inconsistent
  - Rely on calling people's landline
  - Require people to be honest
  - Dependent on sample size
  - Take time to produce
- Unpredictable election results can result in instability in democracies.

We intend to use machine learning methods to produce more accurate results and gain insight into how consumerism impacts politics

Google Trends

```
pytrend.interest_by_region(resolution = "country", inc_low_vol = True, inc_geo_code = False)
```

Google Trends data is queryable for any search string, at any time frame back to 2004:

Each request includes:

Search query (brand name)
Timeframe
Location specificity (country, metro region, global)

	Unnamed: 0	3M Company	AT&T	Adidas	Airbnb	Aldi	Allstate	Amazon	American Airlines	American Express	...	Walgreens	Walmart	Wayfair	Wegmans	Wells Fargo	Wendy's	Wish	Yum!	Zoom Video	eBay
0	Alaska	0	63	28	63	1	65	82	29	32	...	31	37	21	0	84	39	48	0	66	50
1	Alabama	34	74	60	61	44	93	78	34	51	...	58	82	59	1	61	50	57	0	32	65
2	Arkansas	28	82	49	56	21	47	73	45	35	...	65	100	51	0	11	46	60	0	40	65
3	Arizona	28	28	59	80	7	56	85	100	58	...	74	56	53	1	86	43	58	0	48	52
4	California	31	48	85	88	17	42	82	42	58	...	34	37	53	1	61	28	55	0	100	52
5	Colorado	27	30	51	78	3	56	80	47	45	...	62	45	55	1	65	57	49	0	60	49

Particularly partisan companies:

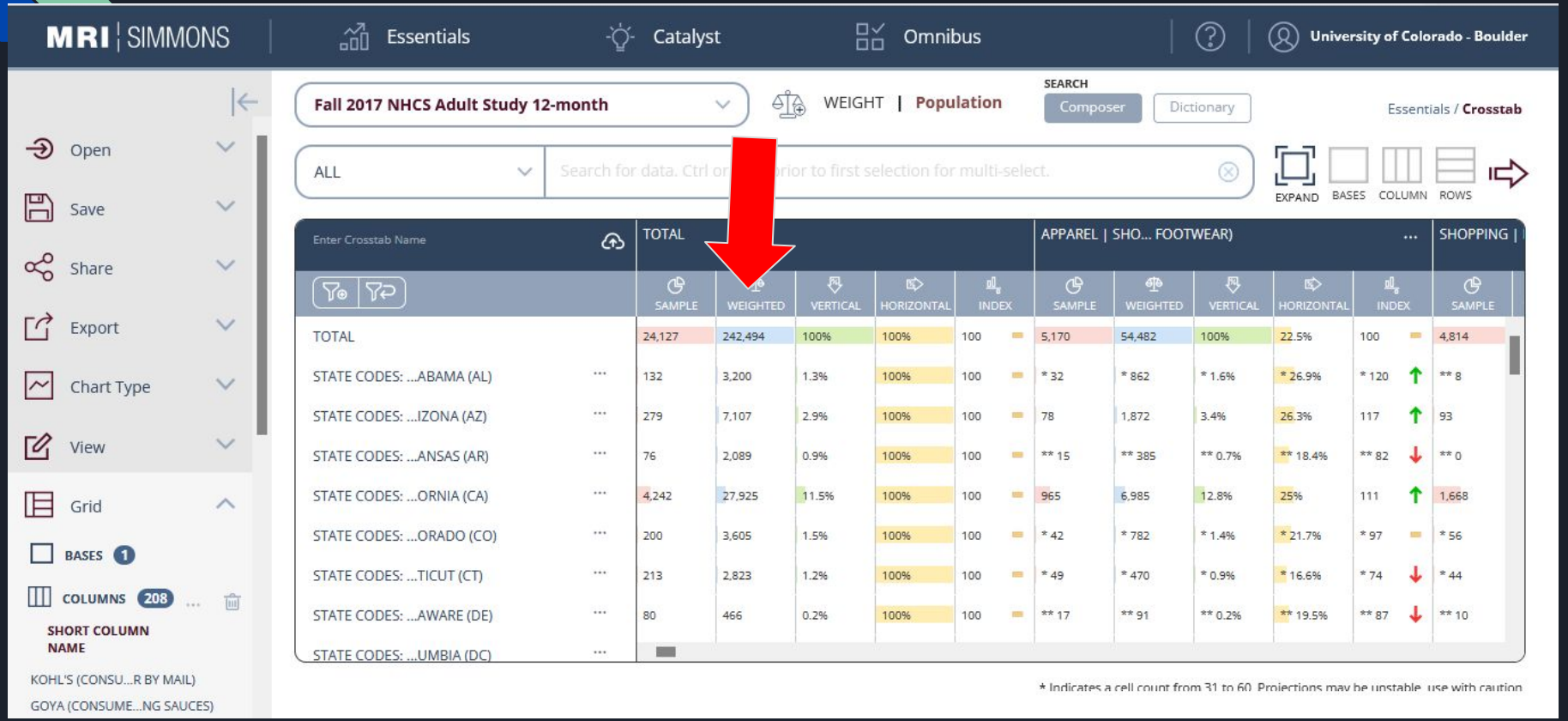
	Alaska	Alabama	Arkansas	Arizona	California	Colorado	Connecticut	Delaware	Florida	Georgia	...	Tennessee	Texas	Utah	Virginia	Vermont	Washington	Wisconsin	West Virginia	Wyoming	District of Columbia
Apple	69	65	64	70	88	68	84	75	72	72	...	70	71	77	83	77	75	74	69	55	82
Papa John's	30	59	32	28	11	24	4	28	45	65	...	69	30	24	46	1	16	17	58	23	24
Chick-fil-A	4	59	41	36	32	40	22	52	65	100	...	59	66	39	66	8	23	24	47	16	28

Data collection

How we aggregated labels:

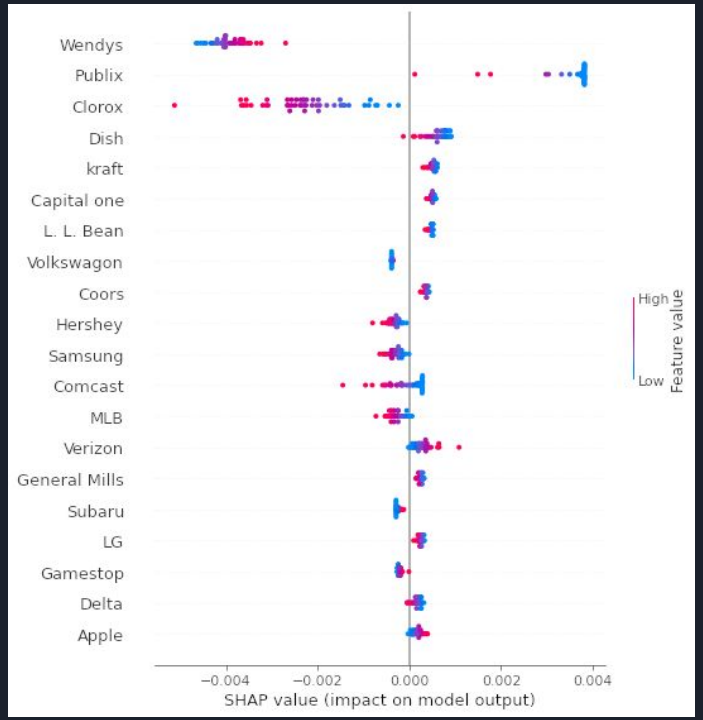
- Sum of all Congressional votes by dem and GOP divided by total
- For all 50 states (excluded D.C.)
- Better outcome for total populations of electorate
- Elections every two and presidential since 2004-2020

Simmons Consumption Data:
Found to be nonoptimal vs Google Trends

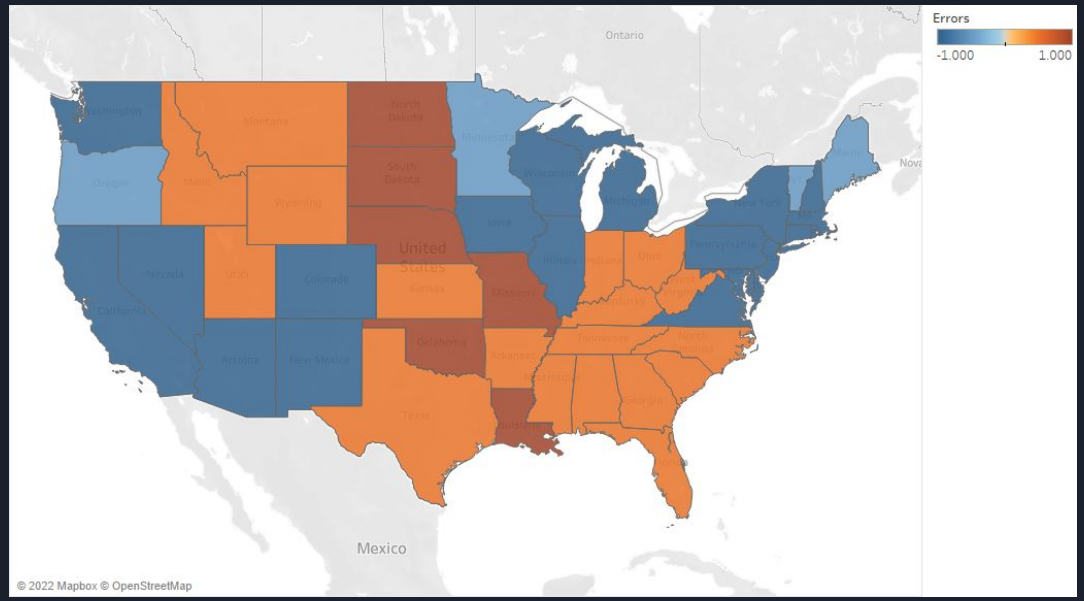


Model results:
Exclusively Consumption Data (Simmons)

Trained on 2016 data: 0.833 accuracy | Tested on 2018 data: 0.54166 accuracy

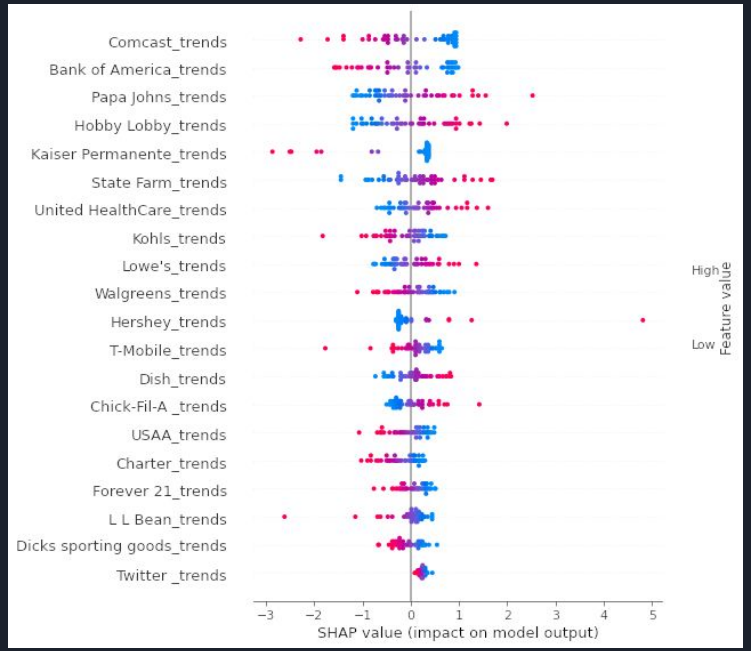


2018

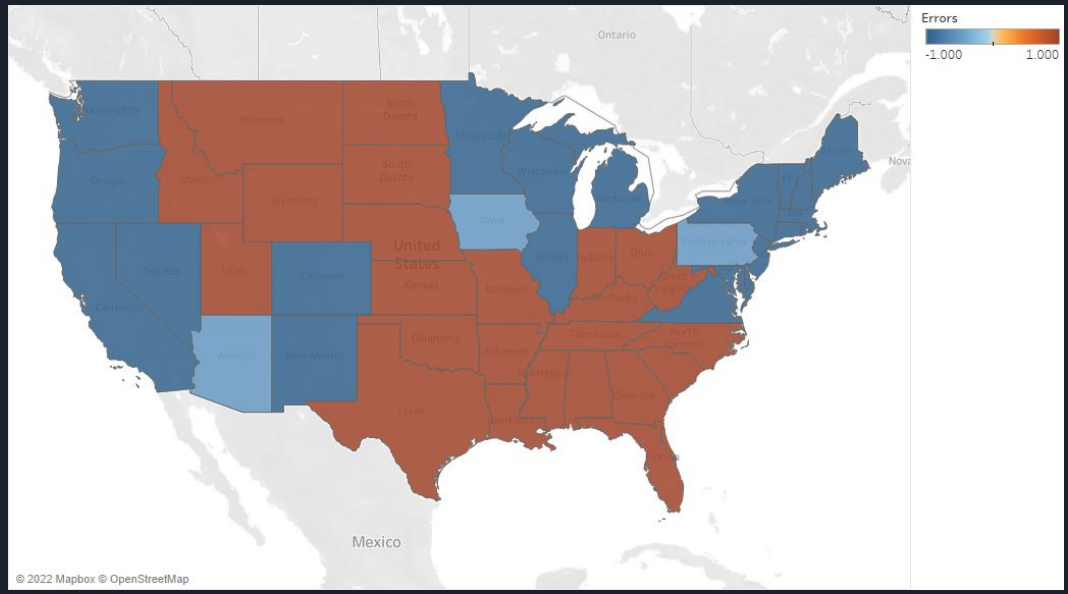


Model results:
Exclusively Search volume data (Google Trends)

Trained on 2016: 1.00 accuracy; Tested on 2018: 0.9375 accuracy

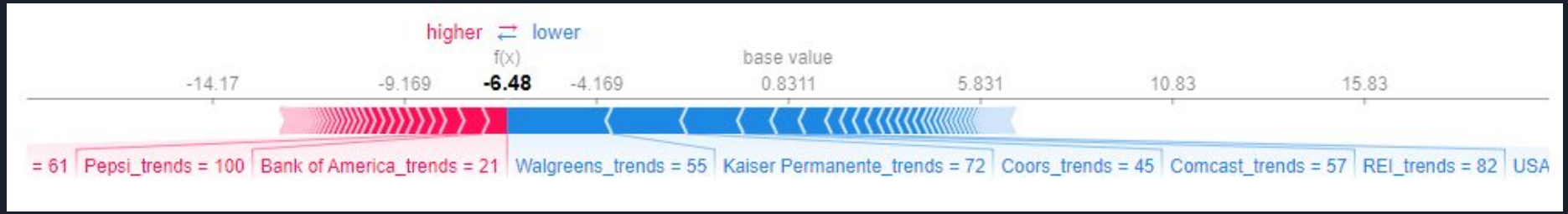


2018



CO:

How the Model Predicts Colorado 2018

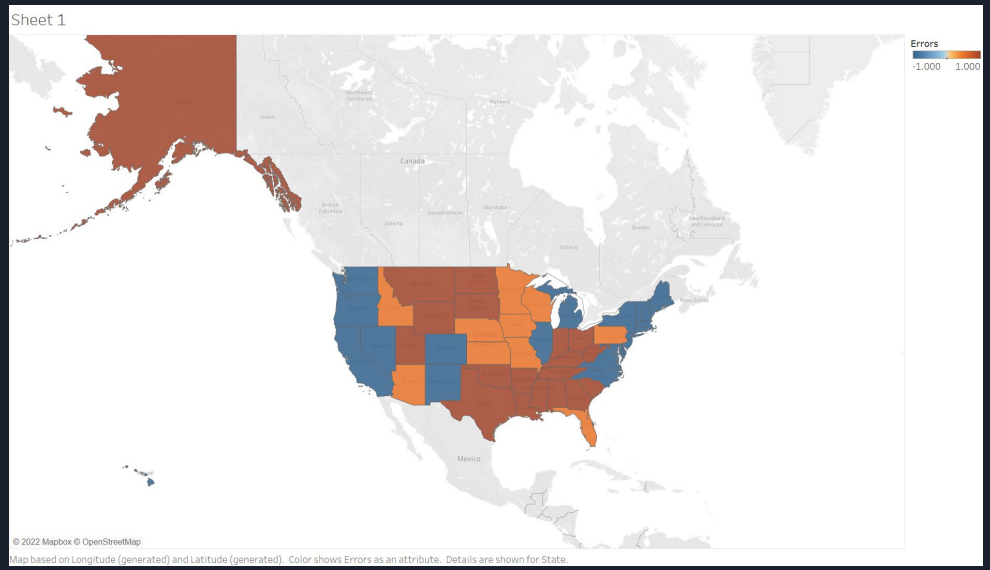


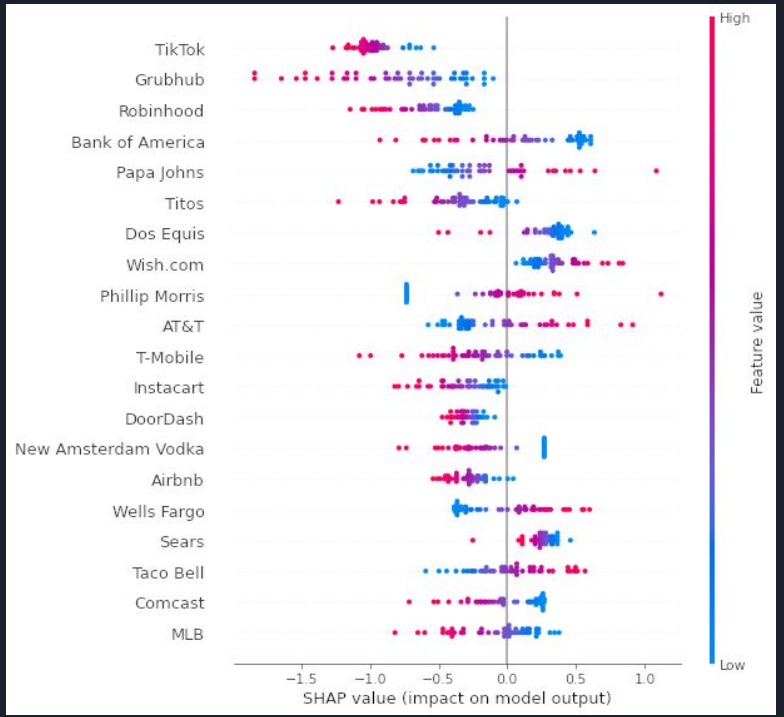
New Results

LogReg on just Google Trends (2010-20)

Train Accuracy (2010-18): 0.988

Test Accuracy (2020): 0.8

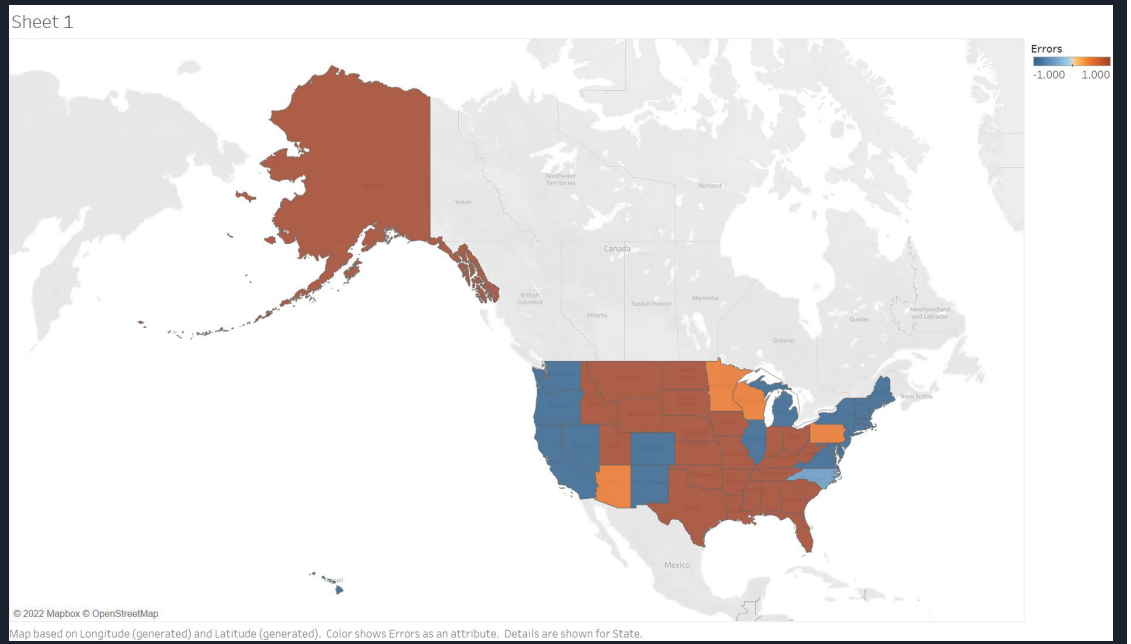


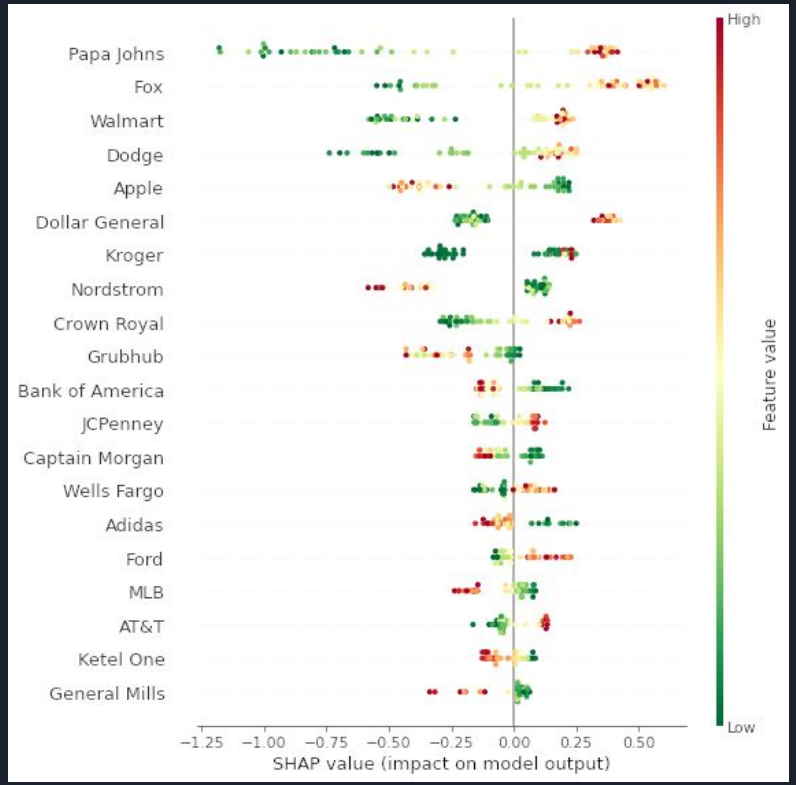


XGBoost

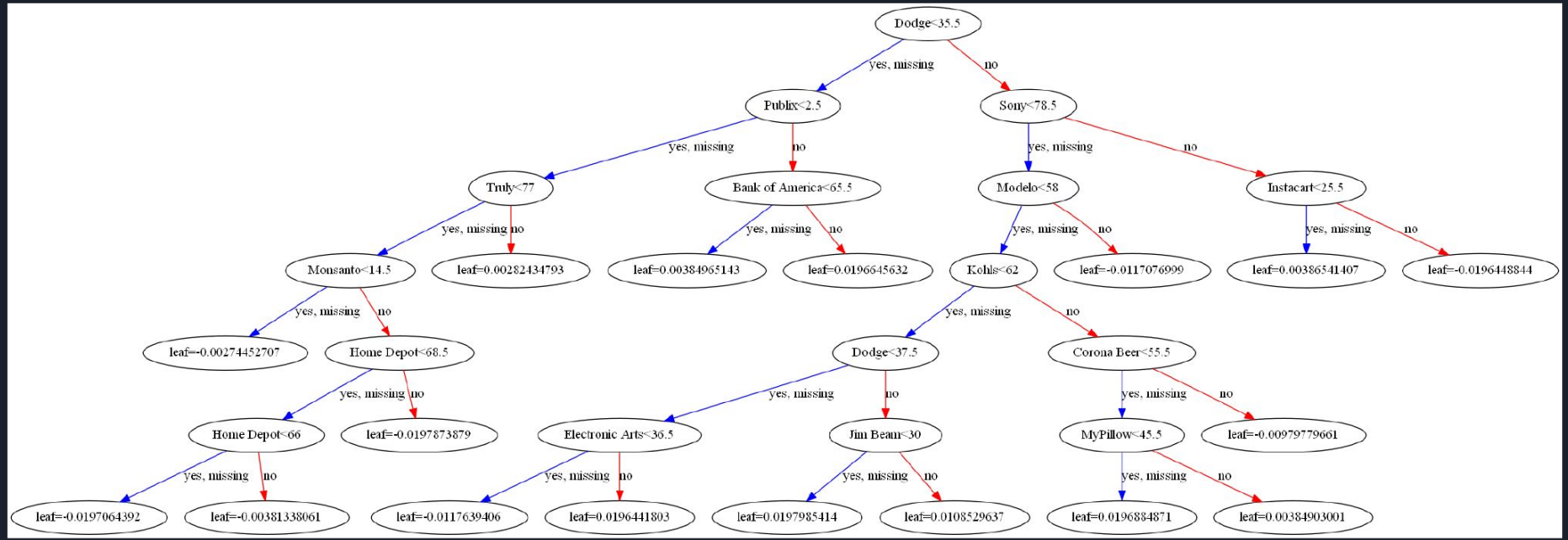
Training Accuracy (2010-18): 1.00

Testing Accuracy (2020): 0.90





500 Trees Total



Regression

XGB Regressor - Just Google Trends

<u>Trained on 2010-18, Tested on 2020</u>

Training Mean Absolute Error: 0.00147 degrees.

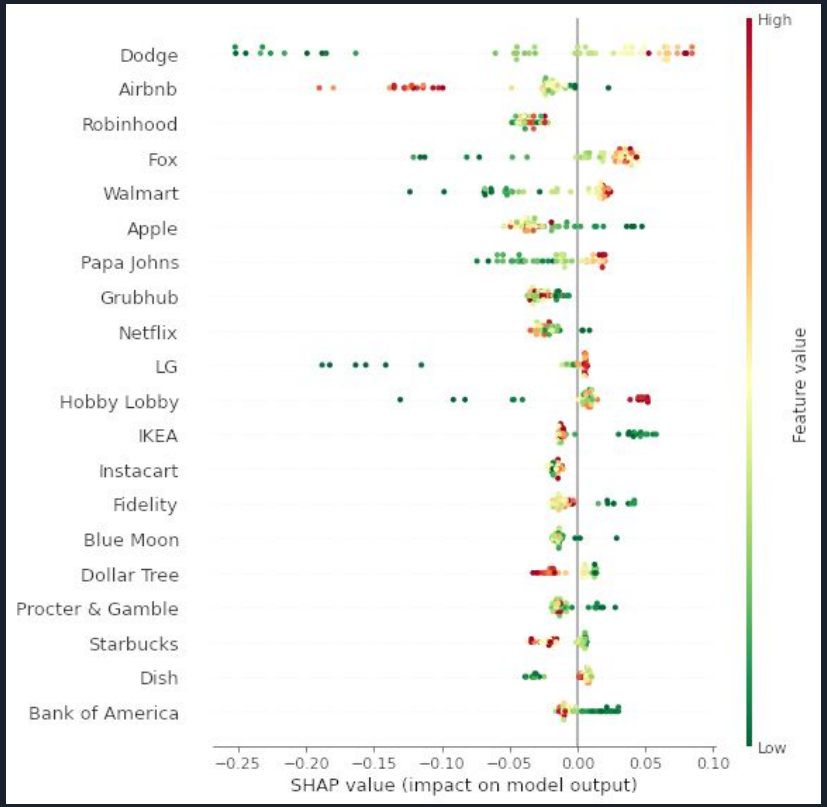
Testing Mean Absolute Error: 0.0599 degrees

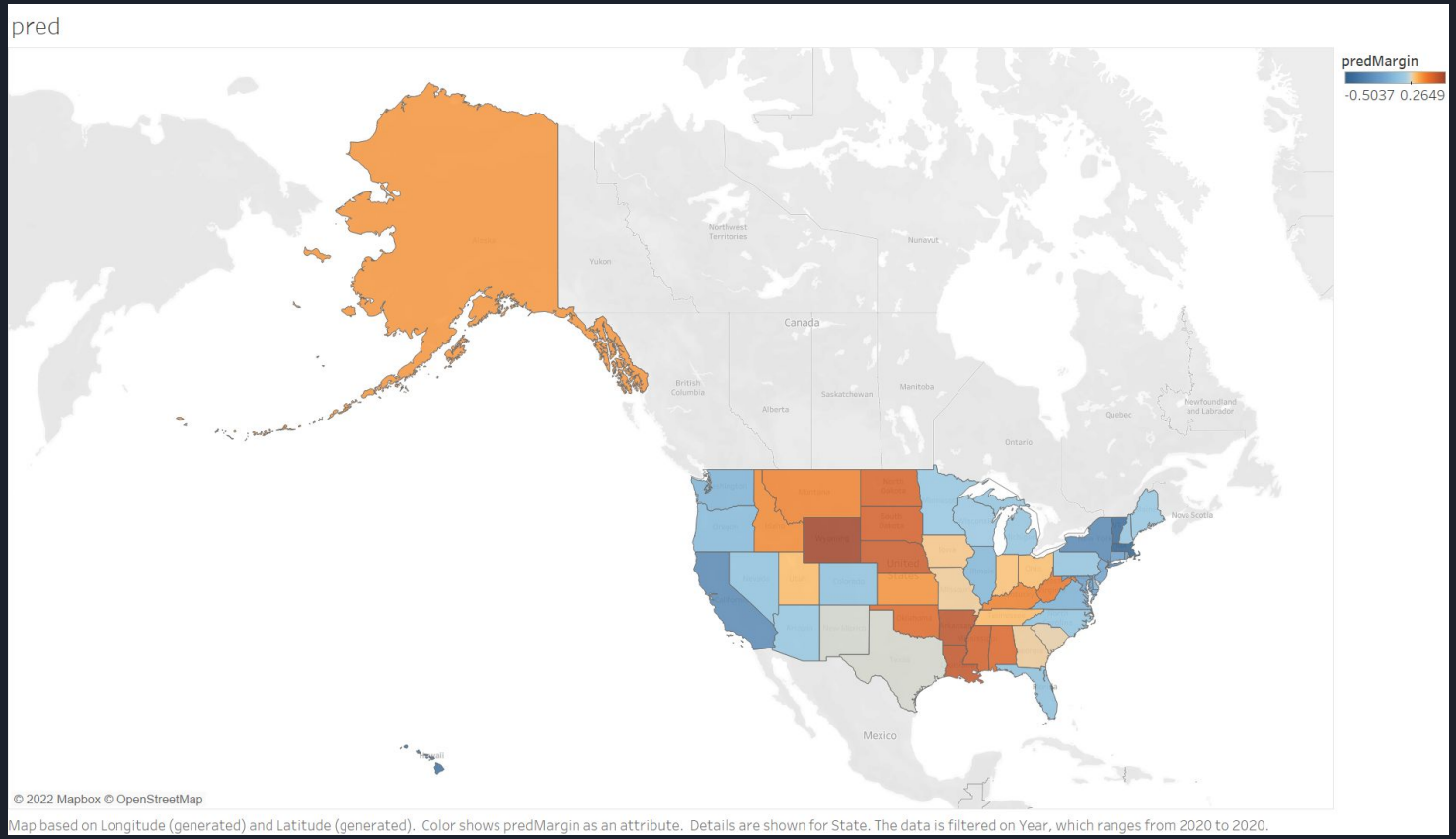
Training Squared Error: 3.607957750636028e-06 degrees

Testing Mean Squared Error: 0.00777 degrees

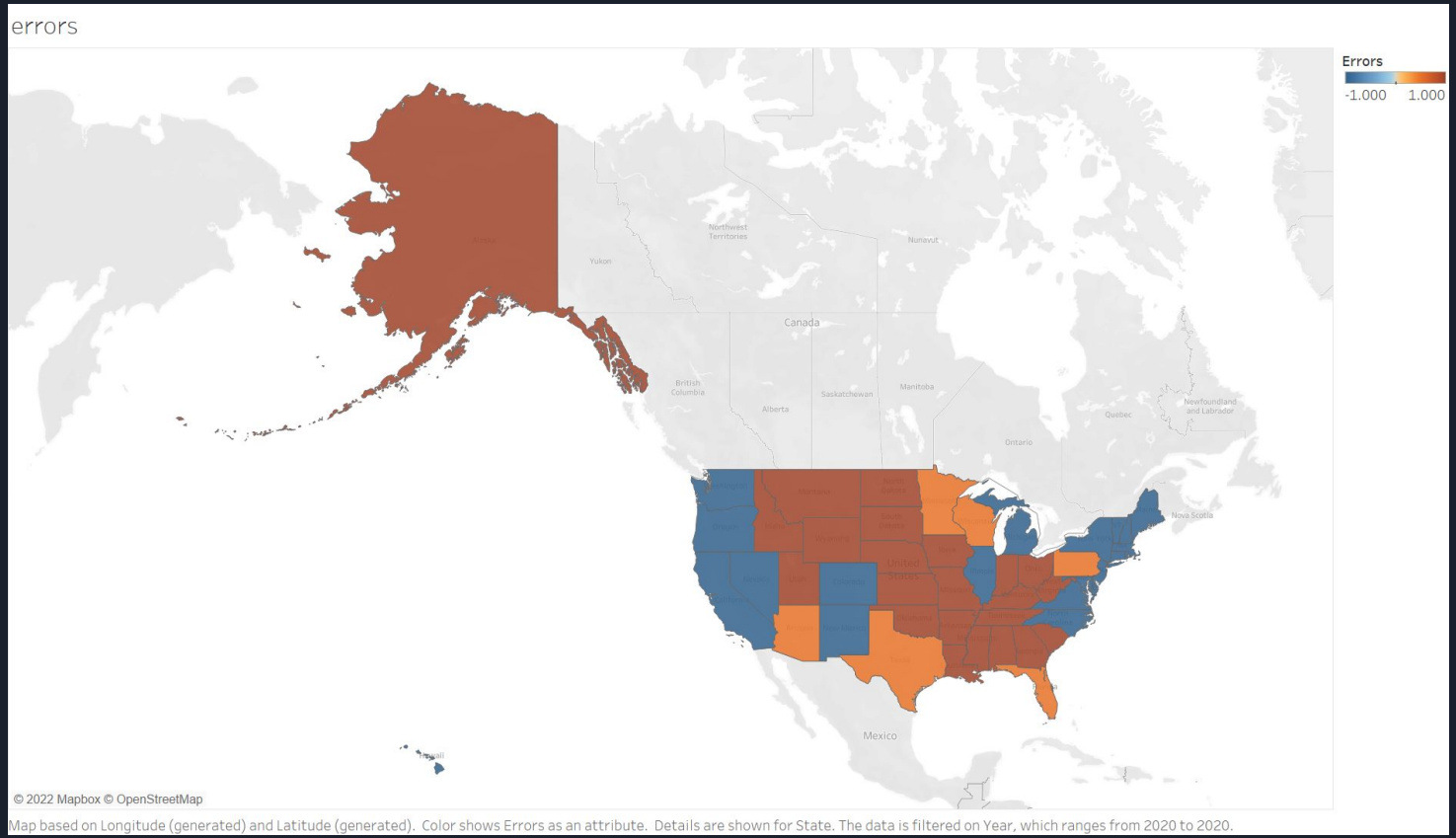
Training Classification Accuracy: 0.848

Testing Classification Accuracy: 0.88





XGB Regressor Predictions



XGB Regressor Classification Errors

Checking just Classical Economic Variables

Training Mean Absolute Error: 0.079227049 degrees.

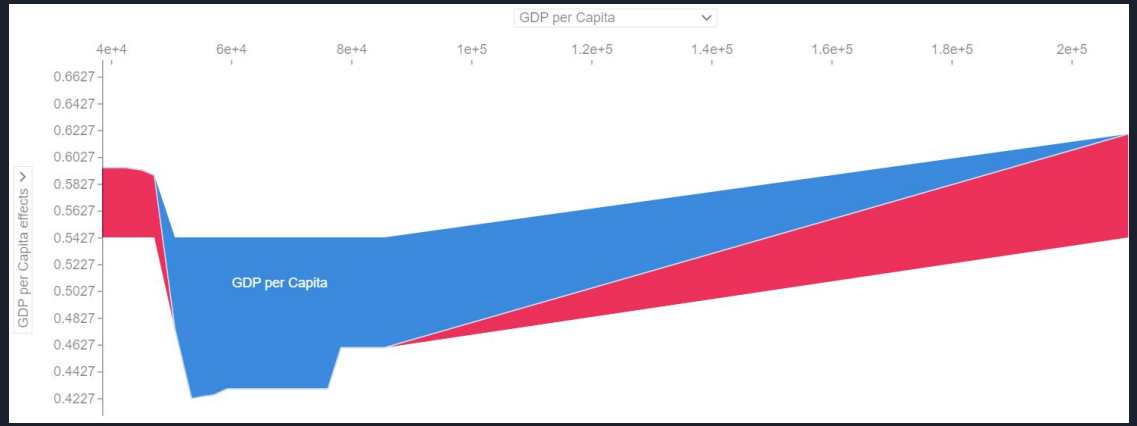
Testing Mean Absolute Error: 0.116001006 degrees

Training Squared Error: 01111136 degrees

Testing Mean Squared Error: 0.024310883 degrees

Training Classification Accuracy: 0.744

Testing Classification Accuracy: 07059



Checking Both Economic and Google Trends

Training Mean Absolute Error: 0.002257748 degrees.

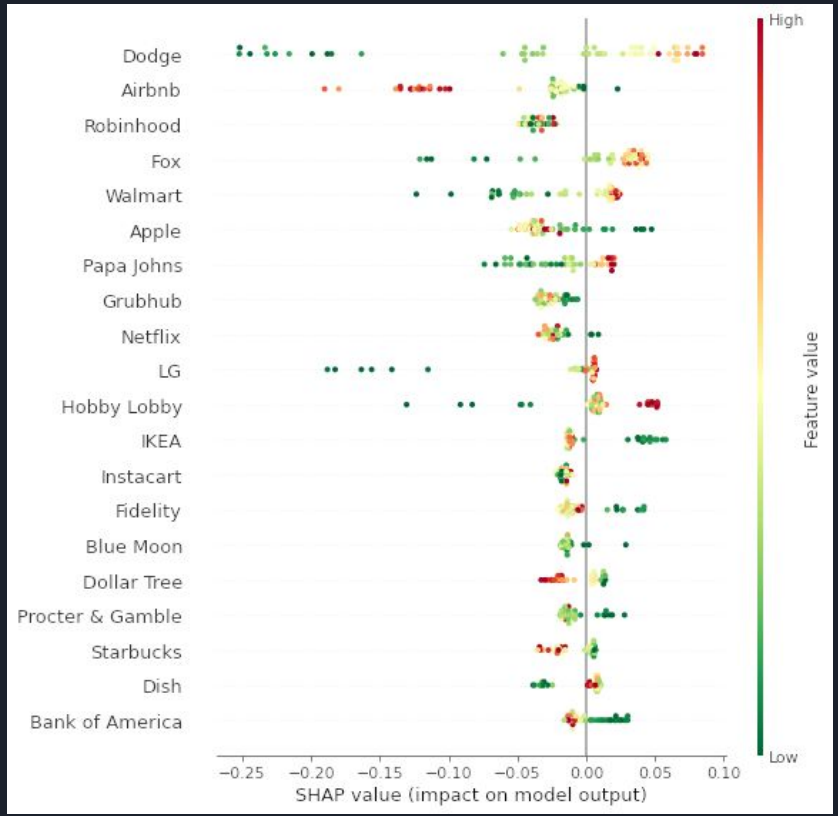
Testing Mean Absolute Error: 0.061140151 degrees

Training Squared Error: 1.65E-05 degrees

Testing Mean Squared Error: 0.007881735 degrees

Training Classification Accuracy: 1.00

Testing Classification Accuracy: 0.8824



Ablation table

	Mean Absolute Error		Mean Squared Error		Classification Accuracy *(Regressor Classification)*	
Model	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Economic	0.08	0.11	0.01	0.02	0.74	0.71
Google Trends	0.00	0.06	0.00	0.01	0.85	0.88
Both	0.00	0.06	0.00	0.01	1.00	0.88

Ablation is the removal of individual component features of a machine learning training set. An ablation study investigates the performance of a system by removing certain components to understand the contribution of each component to the overall system.