MODELING US VOTING
USING CHAIN STORES

by Aaron Lee

OUR PROJECT

Goal: To model and predict party voting using only the   presence of chain stores within those boundaries.
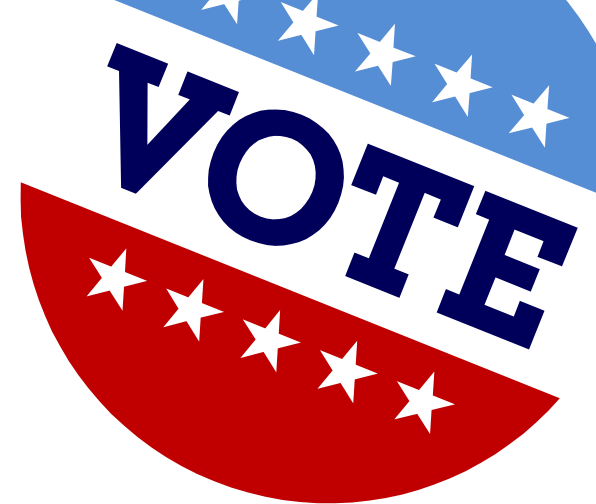
Data:

- Every location of 20 national chain stores (+45,000 stores in total)
- Political boundary maps for counties and US congressional districts
- Results of 2018 midterm election
- Population and area of every county and congressional district

Data Sources: Kaggle, POI Factory, MIT Election Lab, data.gov and web scraping.

BIG QUESTIONS

- How accurately can we predict party voting?
- Which stores have the greatest influence on our model?
- Which stores trend Democratic or Republican?
- How do we apply this model toward future elections?

POLITICAL LEANINGS

- <u>NBC</u>, <u>Time</u>, <u>Washington Post</u>, and <u>The New York Times</u> have recent articles and conducted polls of political leanings of certain brands.
- Grayed items not used in any of our models

- H&M
- Starbucks
- Target
- Whole Foods
- Trader Joe's
- REI
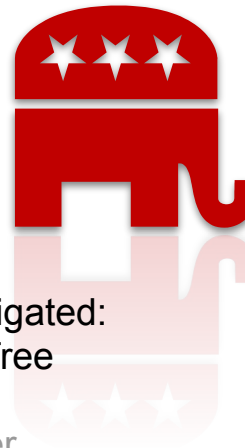
- Tractor Supply Company
- Walmart
- Chick Fil A
- Hobby Lobby
- Cracker Barrel
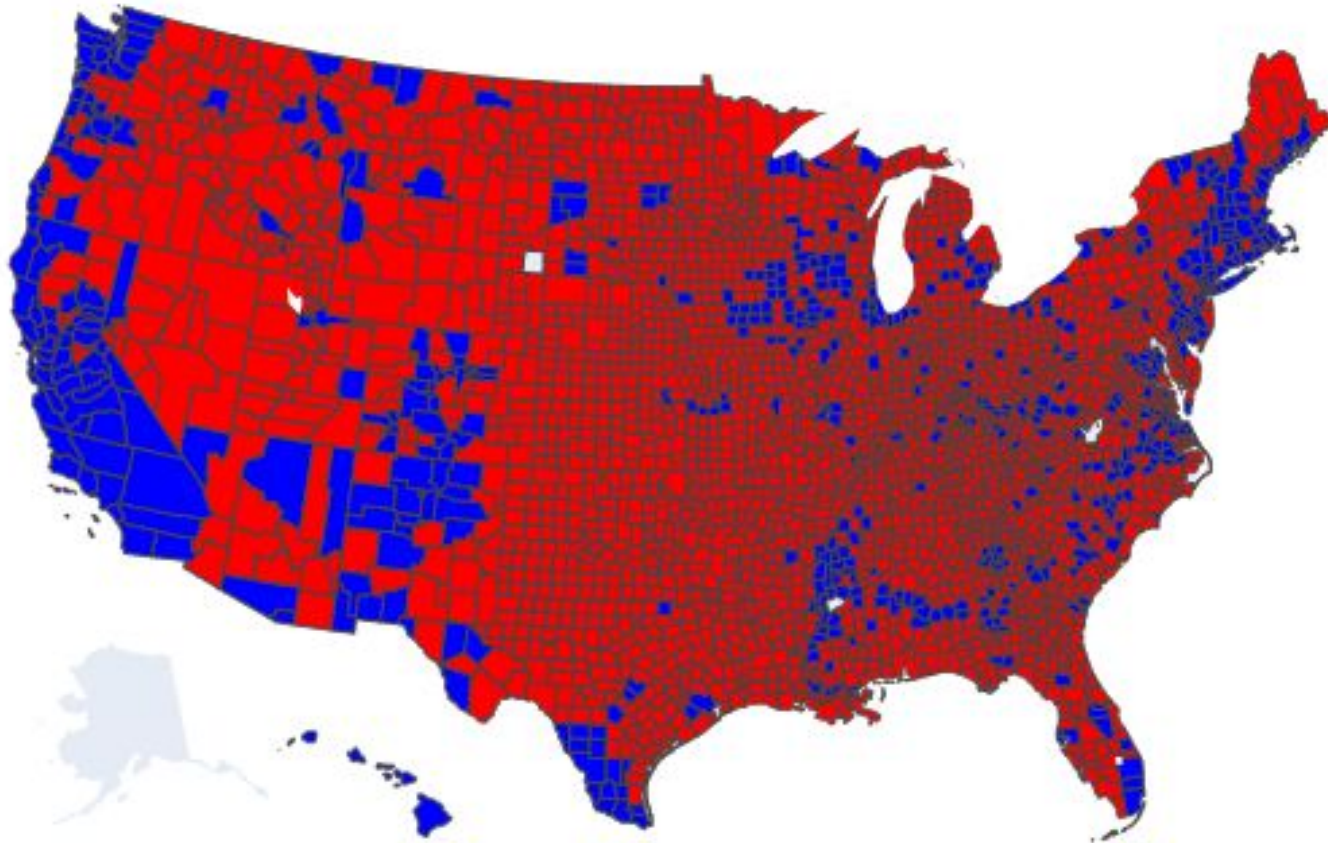- Bass Pro Shops
- Harley Davidson

Others investigated:
- Dollar Tree
- Costco
- Grainger
- Dick's
- Academy Sports
- LL Bean
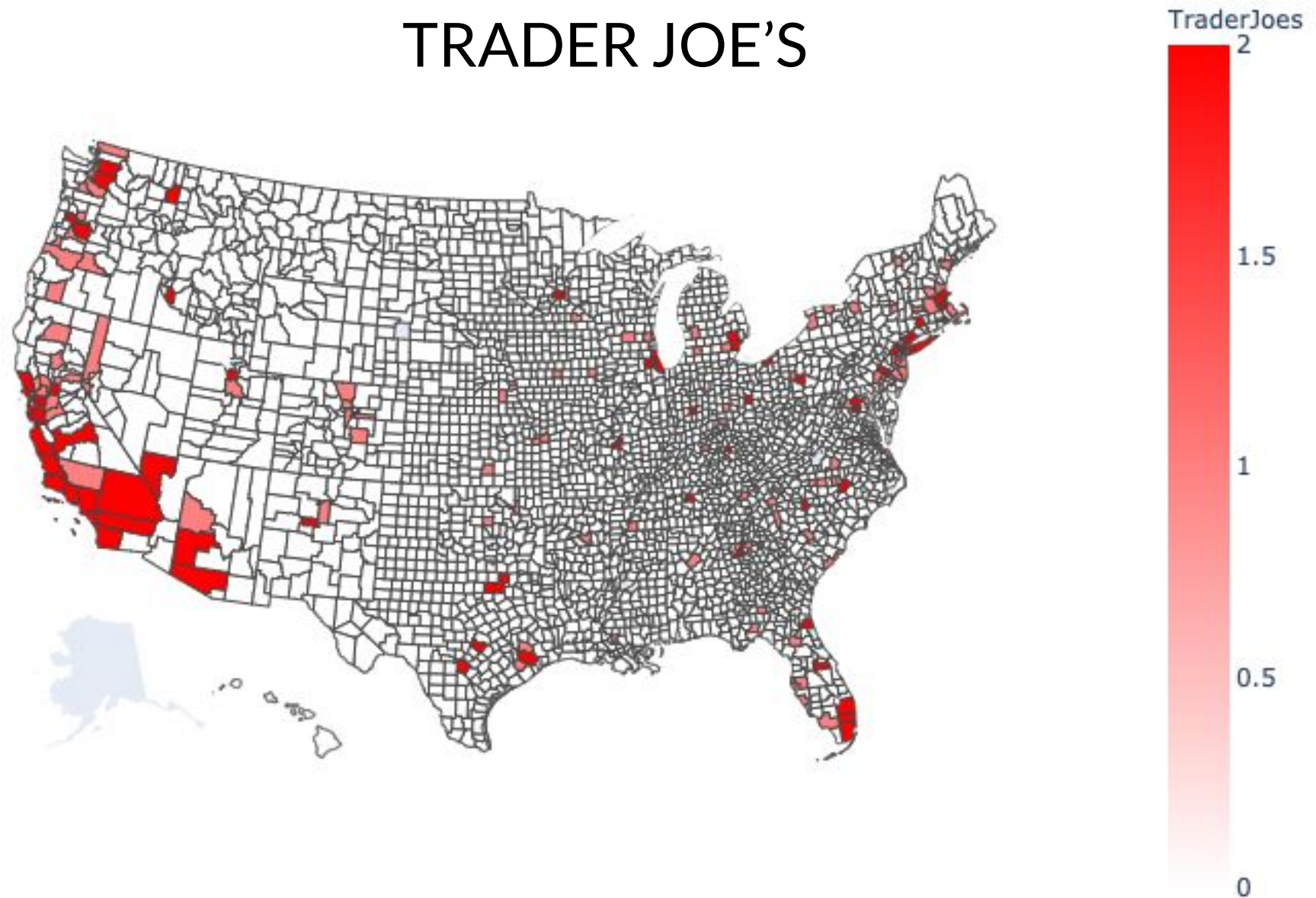- Pottery Barn

UNDERSTANDING OUR DATA

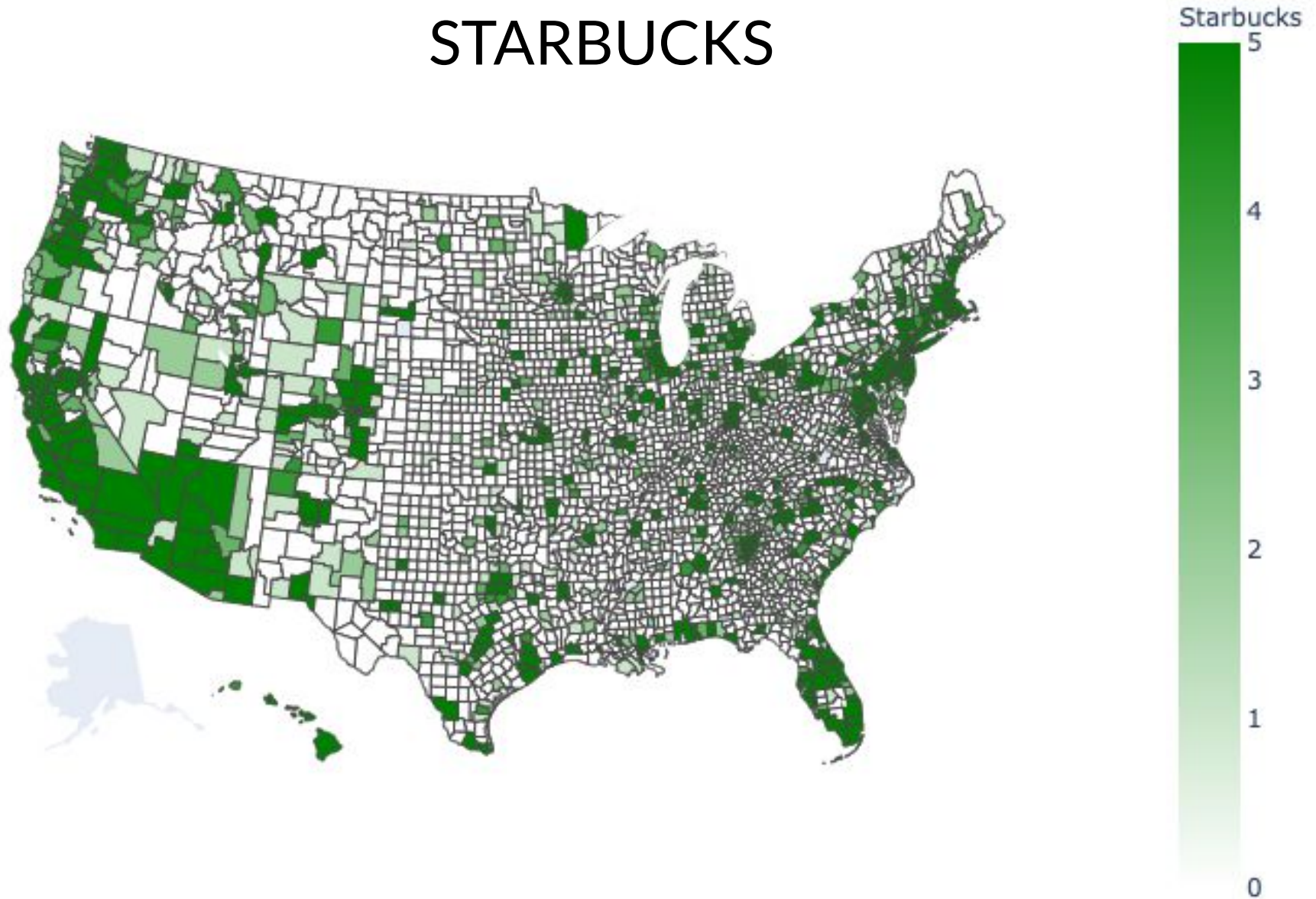|	Democratic	Republican
Number of Counties	19.5%	80.5%
Population of Counties	58.1%	41.9%
Total Stores in Counties	59.4%	40.6%

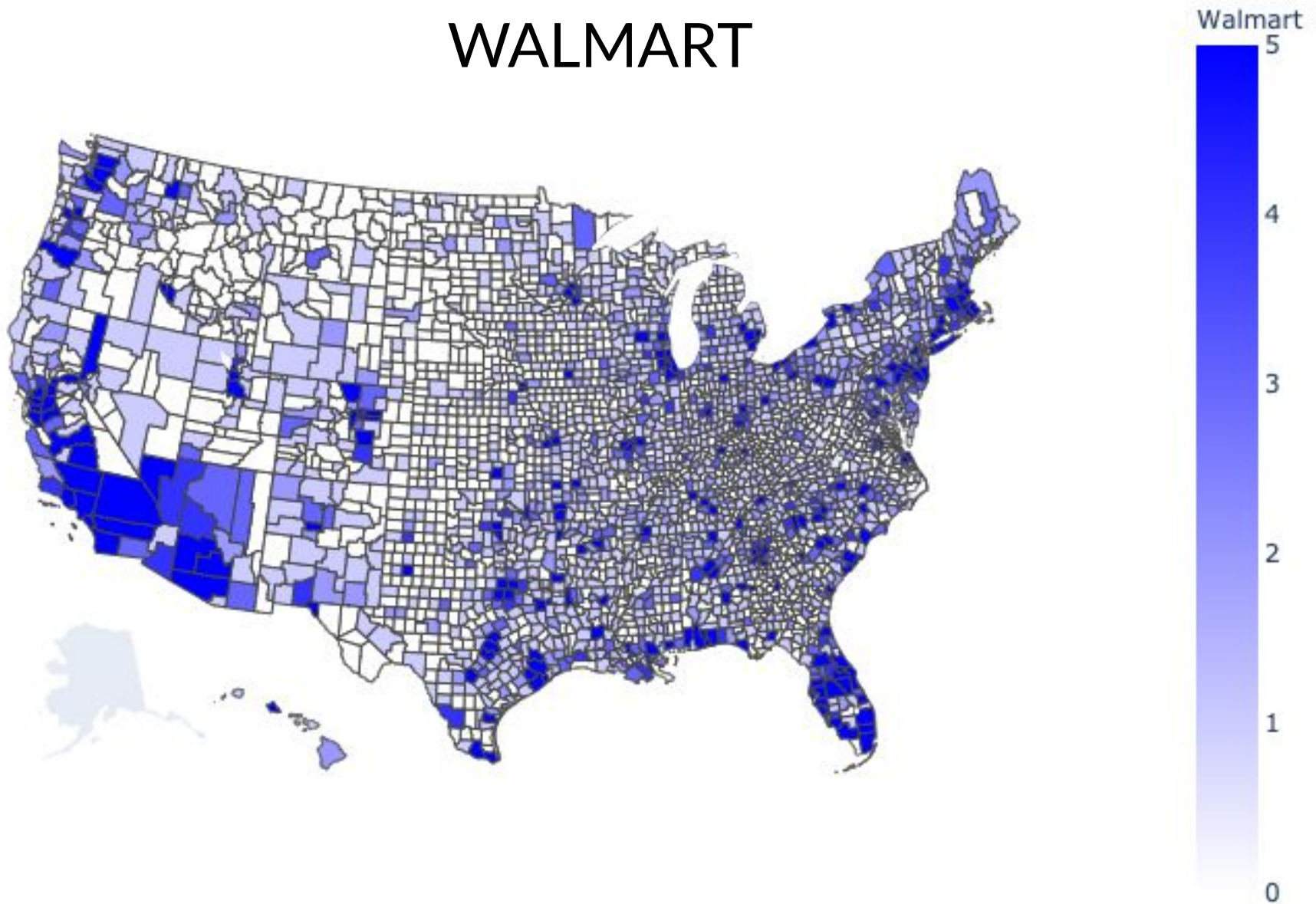Stores in Dem/GOP Counties
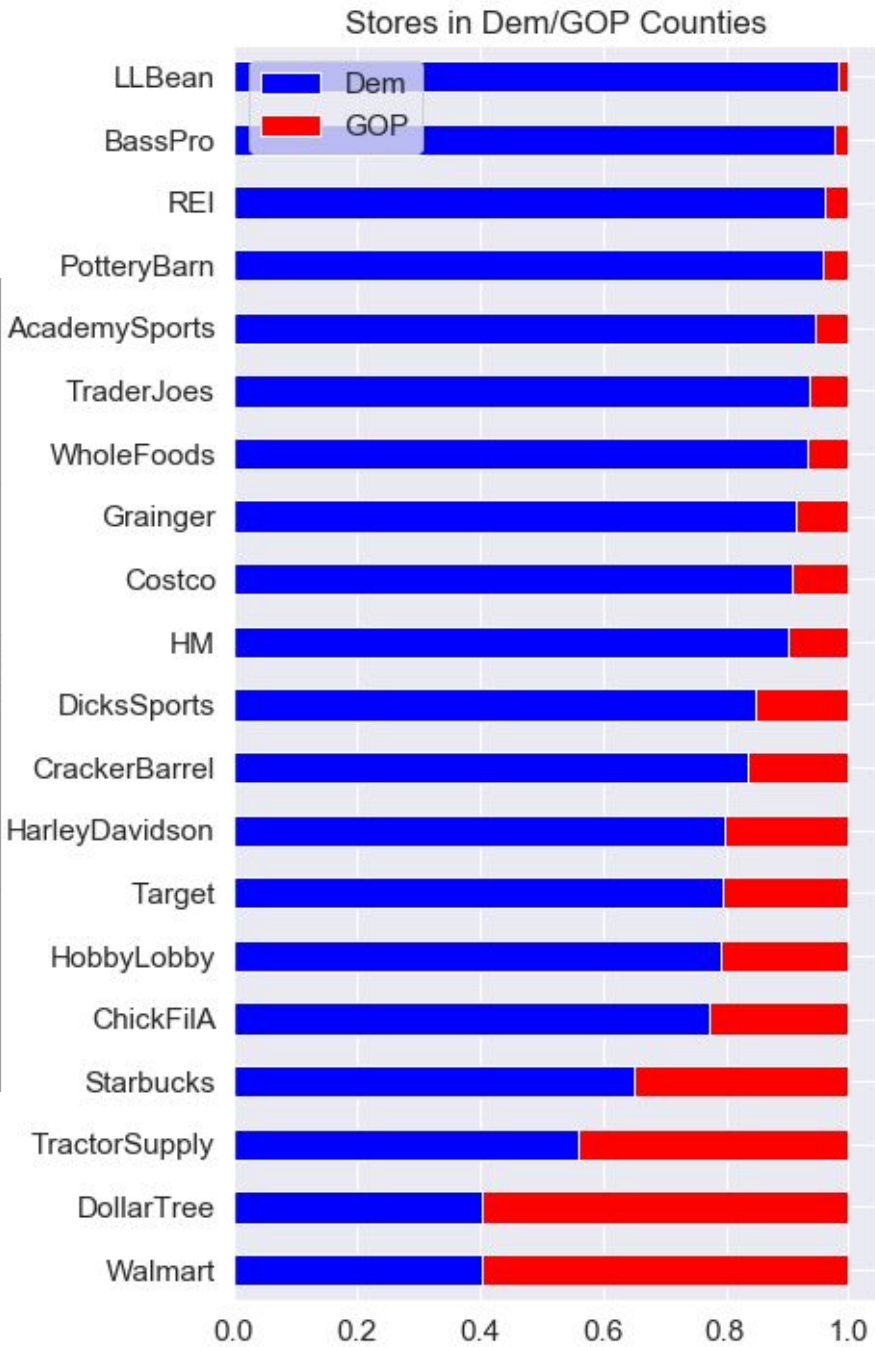
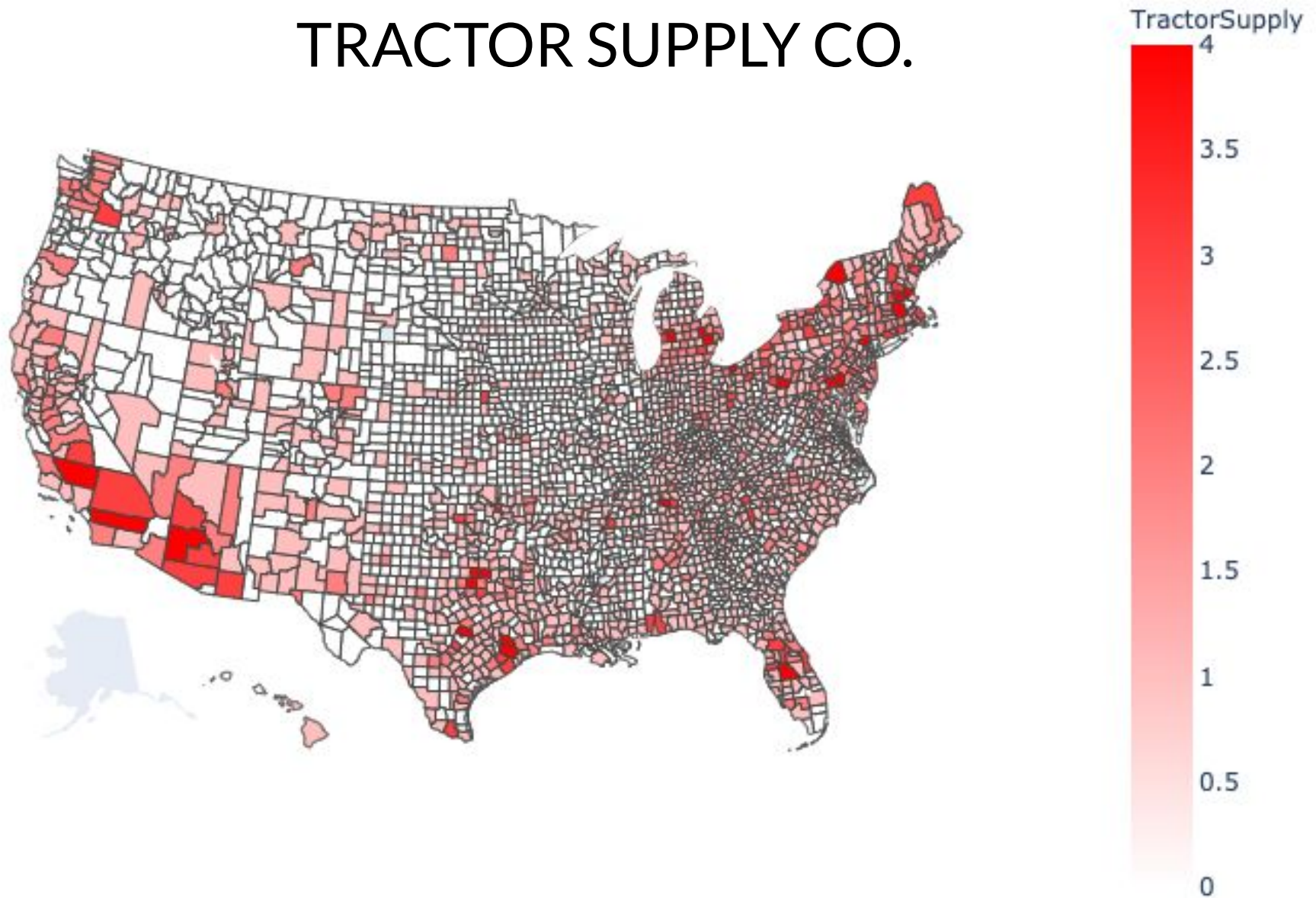# ELECTORATE BY COUNTY

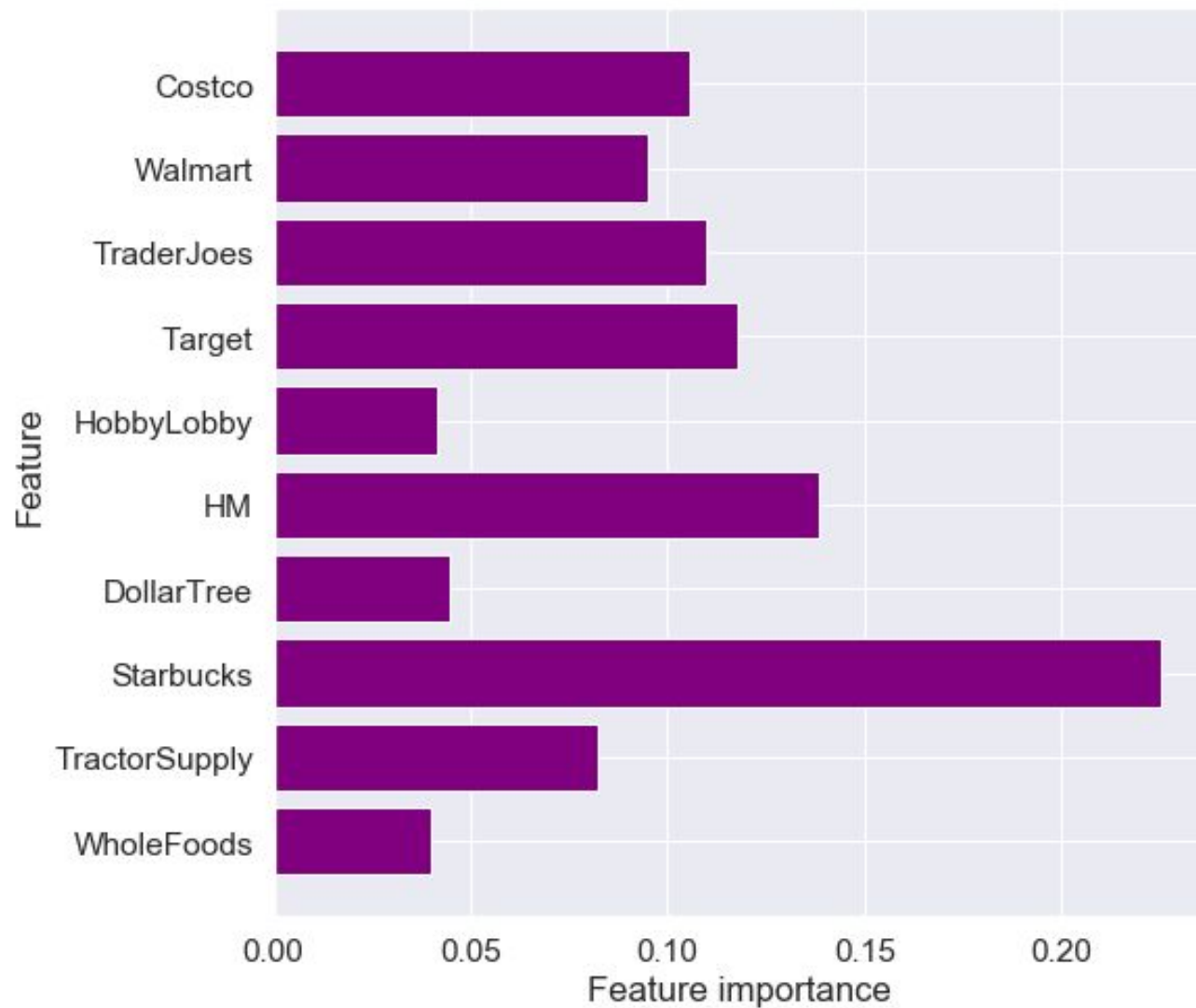# TRADER JOE'S
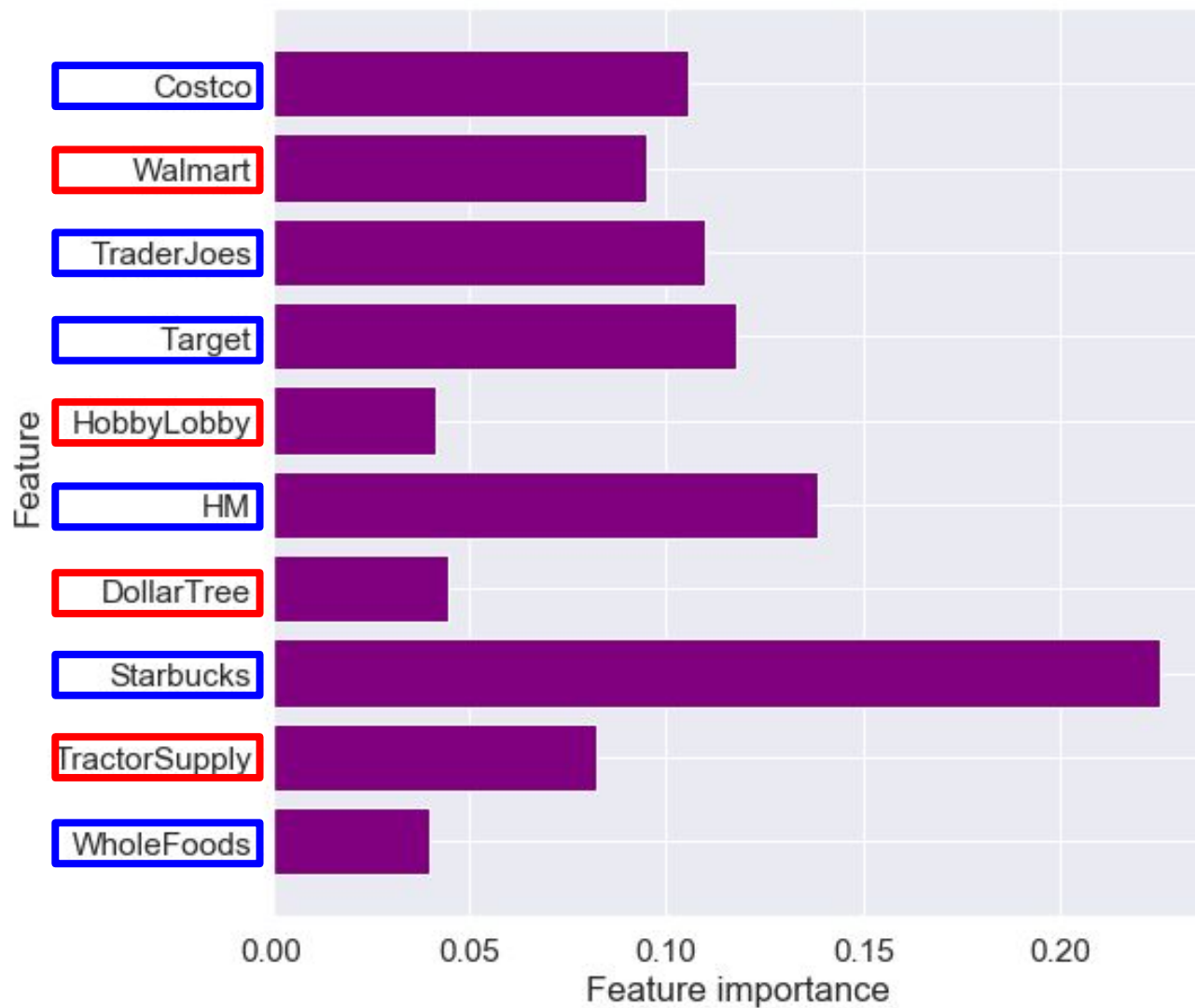
STARBUCKS

WALMART

TRACTOR SUPPLY CO.

OUR MODEL

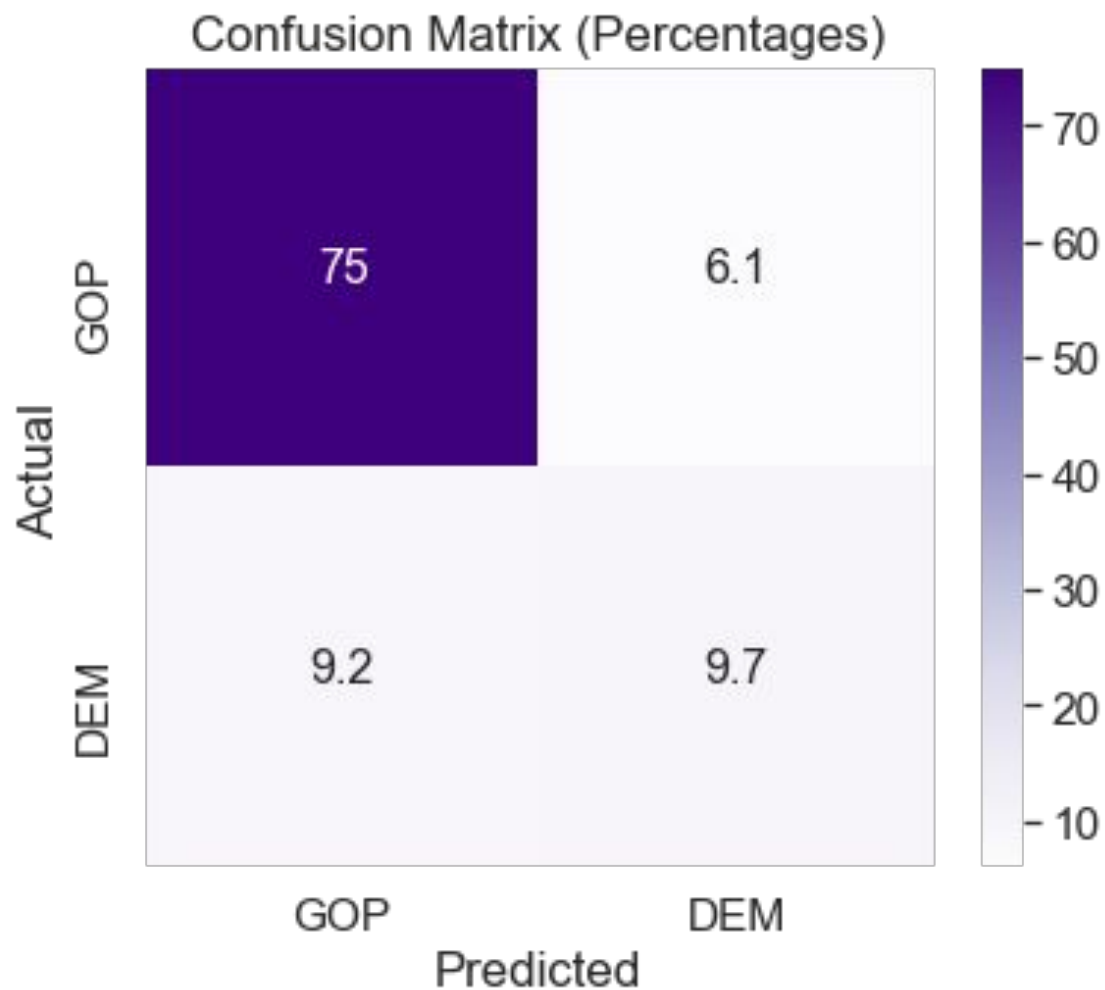Overall Accuracy

85.3%
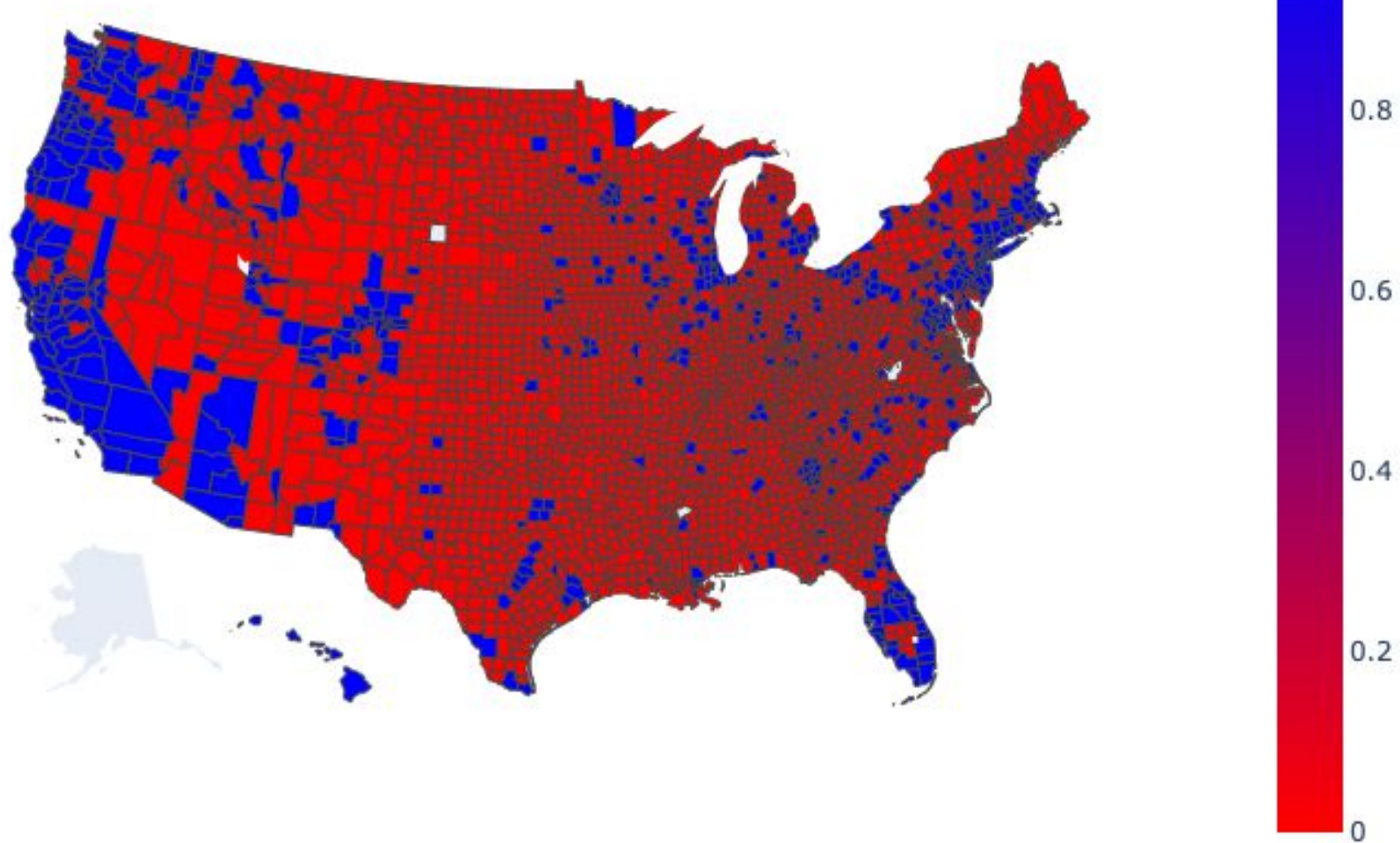
OUR MODEL

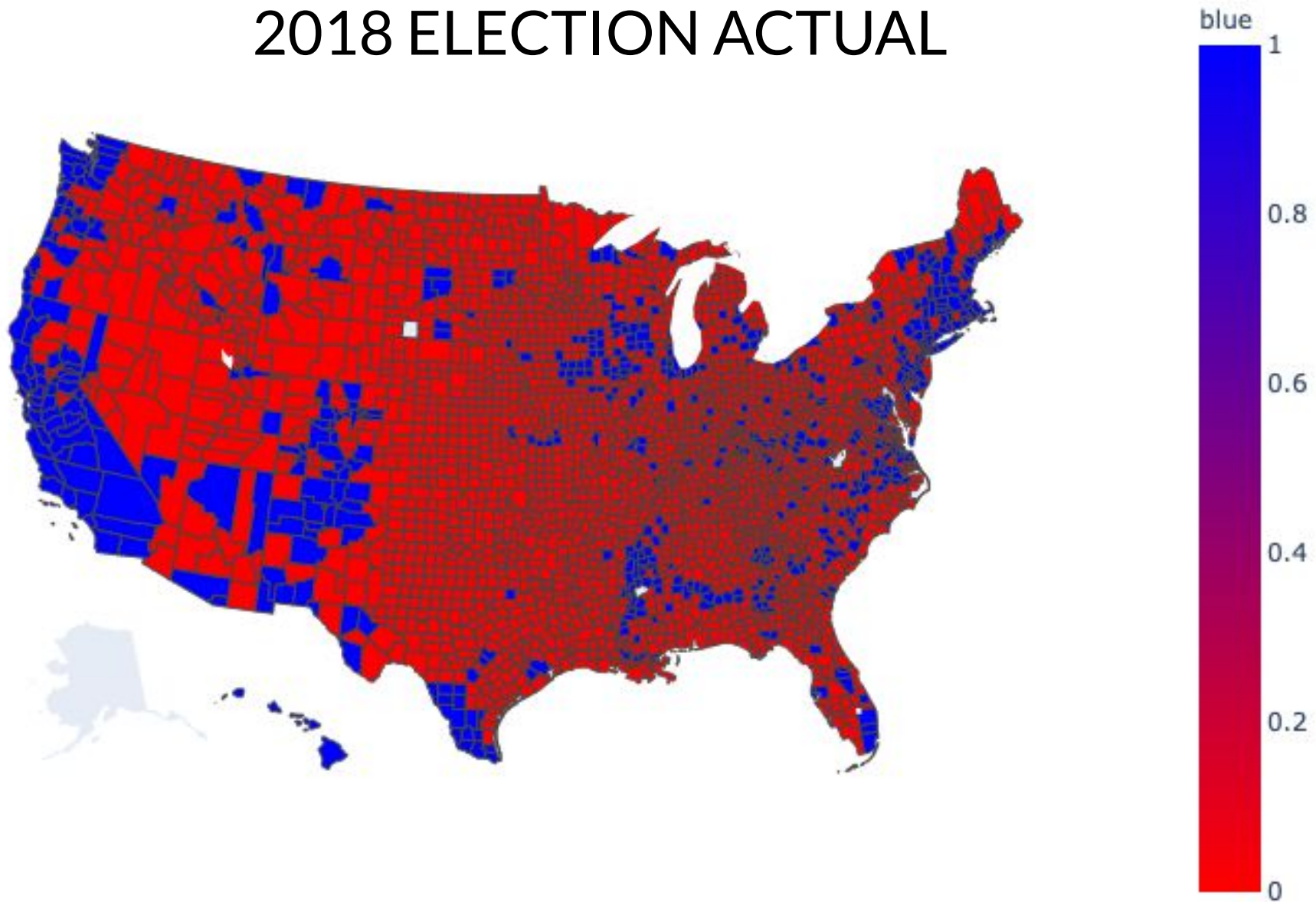Overall Accuracy

85.3%

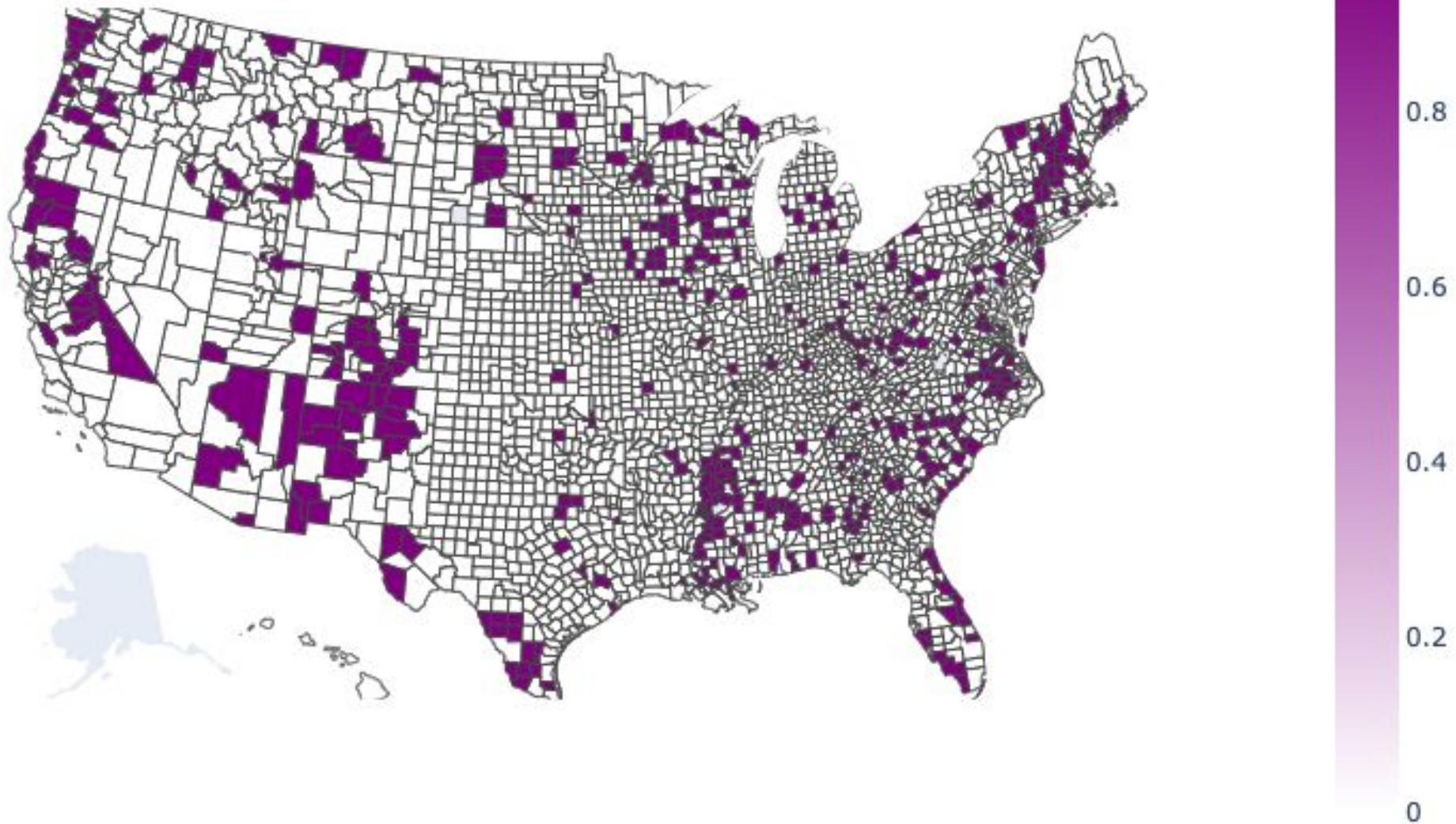# OUR MODEL

Overall Accuracy

85.3%
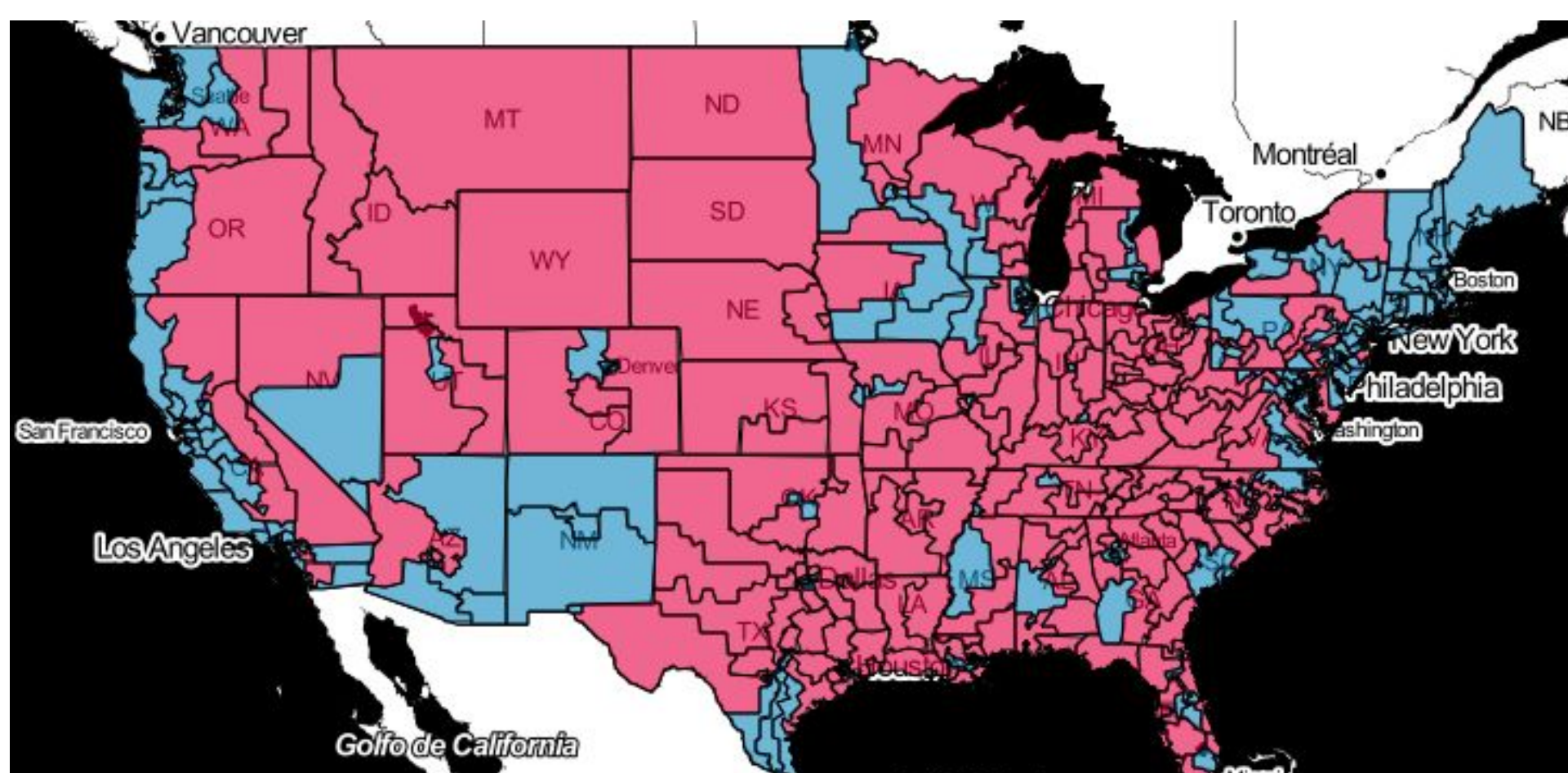
OUR 2018 ELECTION MODEL
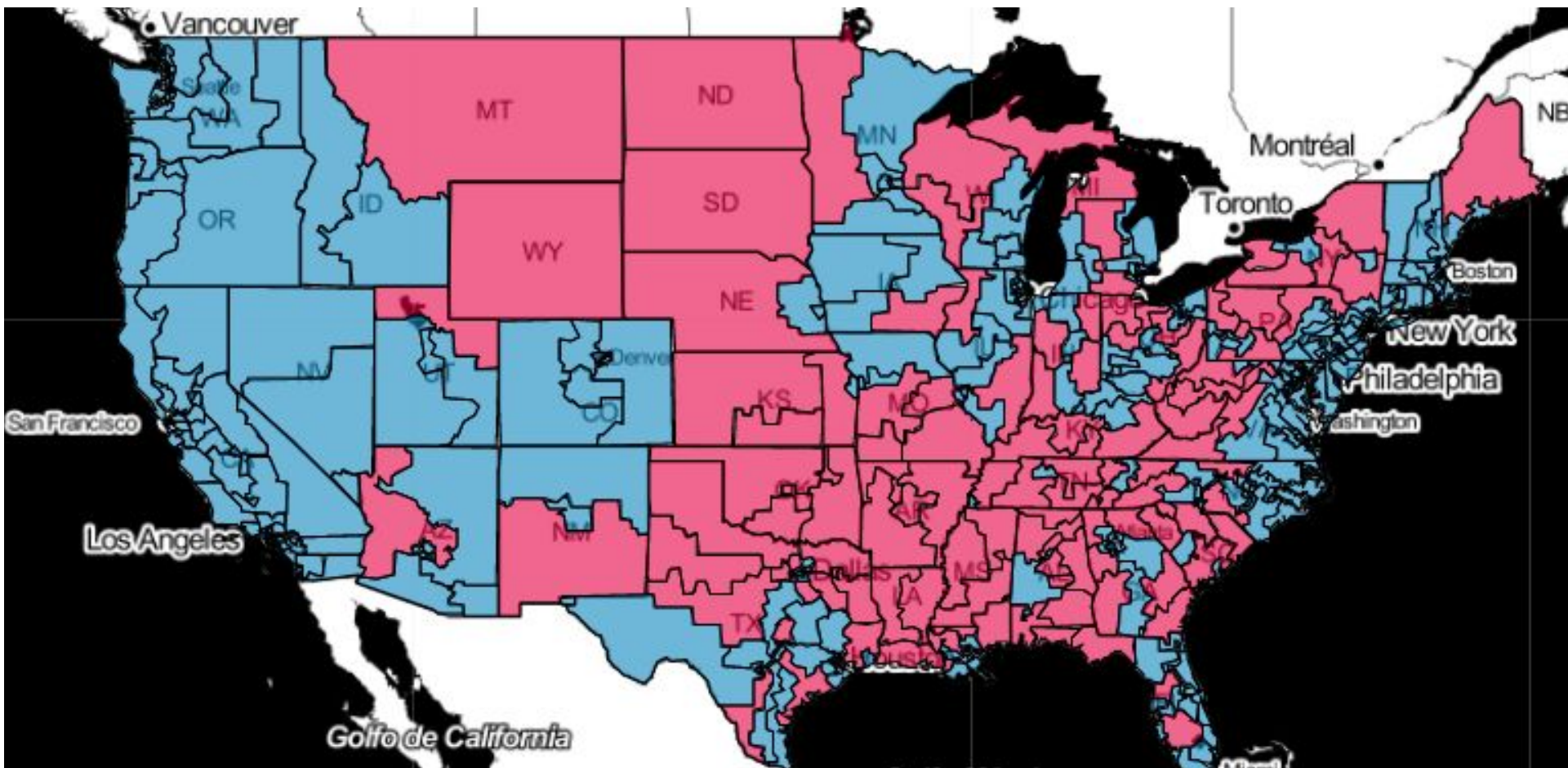
2018 ELECTION ACTUAL

MISSED COUNTIES

The county model was then applied directly to US
Congressional district maps following the same method.

Actual Results 2018

The county model was then applied directly to US Congressional district maps following the same method.

Model Prediction (72.6% Accuracy)

CONCLUSION

Models focus on Democratic 'upmarket' retail (Whole Foods, Trader Joe's, Starbucks)

Model performs poorly in regions where other factors might be more important (demographics, unique candidates, regional leanings)

Model Advantages

- Independent of census or polling
- Readily available data (updated more easily than polling or census)
- Responds to demographic changes

Future Use

- Election predictions
- Identify under/over performing districts
- Spot demographic changes to the electorate

NEXT STEPS

- Use these predictions to inform other more traditional election models
- Test model
  - Identify under/over performing districts
  - Evaluate with 2020 election data
- Improve model
  - Update model with new store data
  - Change or add features
- Create other model versions:
  - Regional (homogenous)
  - Senate model
  - Swing districts

Questions
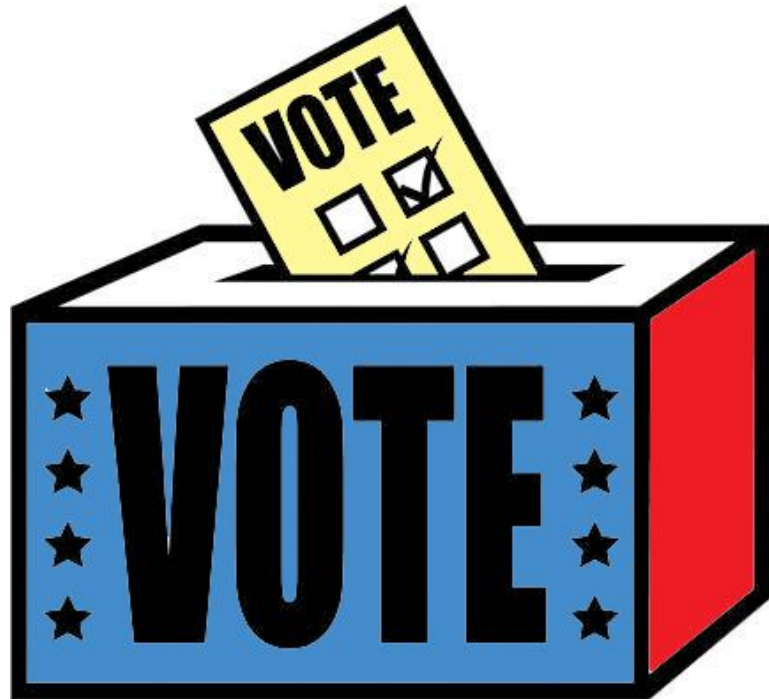
Regional Politics of US

There are clear areas on our map the model missed significant numbers of counties. Some areas stand out:

- In the New England area, particularly Vermont and New Hampshire, Democrats significantly outperformed the model. Rural areas in these states have a Democratic lean which goes against the model which assumes rural areas to be Republican.
- New Mexico is an outlier. It is one of three states where White voters are the minority. Hispanic voters in New Mexico have shown enthusiastic support for Democrats in recent years.
- Western Mississippi has one of the greatest concentrations of African American voters in the US, and they overwhelmingly (90%) support the Democratic party. Similar demographics exist in parts of Alabama which likely account for those misses.
- The model also misses counties in the Pacific Northwest and the Deep South where conservatism and demographics play outsized roles in the voting habits.
- Other big misses occurred in swing states of Florida and Wisconsin where political leanings are split and in flux.

See this article for more detail:

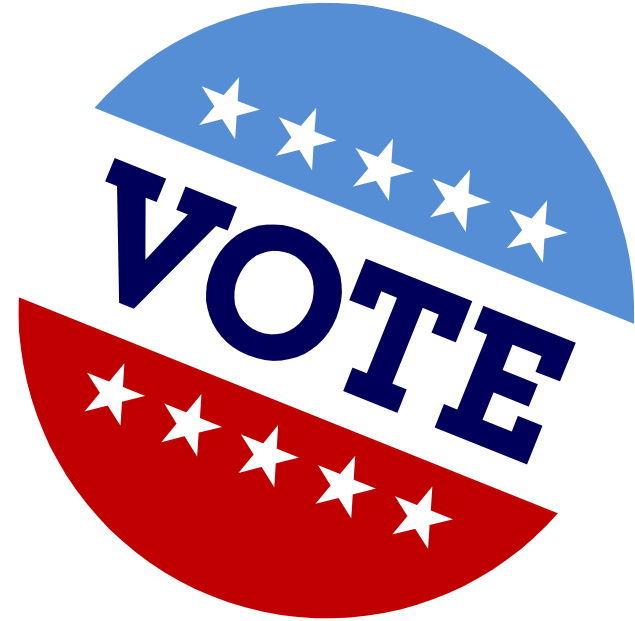https://towardsdatascience.com/are-you-a-trader-joes-democrat-or-a-walmart-republican-a7b156131435
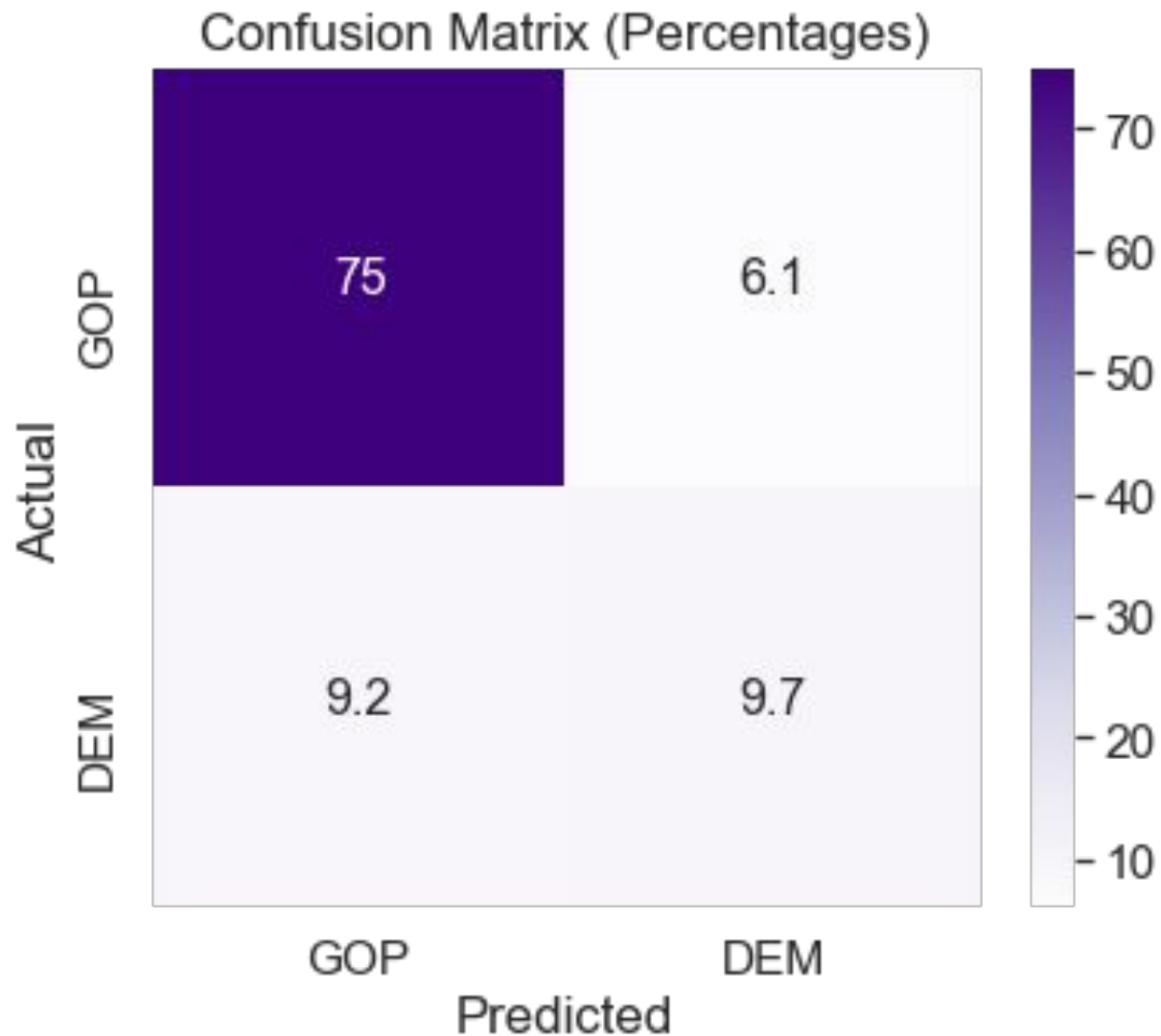
Data Resources

Kaggle

POI Factory

MIT Election Lab

data.gov

OTHER MODELS (XGBoost)

Overall Accuracy

82.5%

Confusion Matrix (Percentages)

OTHER MODELS (XGBoost)

Overall Accuracy

82.5%

Feature importance