Predicting Election Results:

2016 Primary Election

Eric Vandament, Jacob Brandis, Eve Dean

About the Dataset

1. Kaggle dataset of 2016 primary election results
    a. Fraction of votes received per candidate per county
        i. primary election: vote fraction split by party
    b. Missing Minnesota
    c. South Dakota had limited data reported
2. County facts dataset
    a. Population demographics (age, race, gender)
    b. Economic conditions (median income, percent below poverty, retail sales)

PREDICTING WINNERS

VOTE!

RACE & ETHNICITY:
Predicting Democratic Primary Outcomes

Can you predict if Hillary Clinton or Bernie Sanders will win in a county based on demographic conditions?

Input:

- ★ % Black (2014)
- ★ % Indian American (2014)
- ★ % Asian (2014)
- ★ % Hawaiian (2014)
- ★ % Biracial (2014)
- ★ % Hispanic (2014)
- ★ % Caucasian (2014)

Output:

- ★ Predict Clinton win

		Predicted	
		Won	**Lost**
Actual	**Won**	80.23%	19.77%
	Lost	21.04%	78.96%

RACE & ETHNICITY:
Predicting Republican Primary Outcomes

Can you predict if Donald Trump or Ted Cruz will win in a county based on demographic conditions?

Input:

- ★ % Black (2014)
- ★ % Indian American (2014)
- ★ % Asian (2014)
- ★ % Hawaiian (2014)
- ★ % Biracial (2014)
- ★ % Hispanic (2014)
- ★ % Caucasian (2014)

Output:

- ★ Predict Clinton win

		Predicted	
		Won	**Lost**
Actual	**Won**	20.92%	79.08%
	Lost	2.83%	97.17%

GENDER:
Predicting Democratic Primary Outcomes

Is Hillary Clinton more likely to win in counties with a higher proportion of women?

Input:

★ % Female (2014)

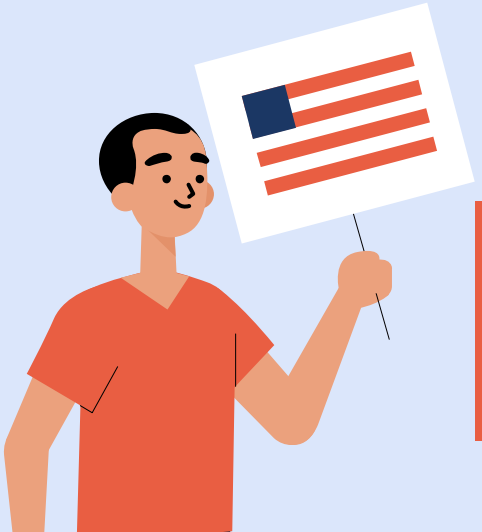
Output:

★ Predict Clinton win

Actual		Predicted: Won	Predicted: Lost
	Won	1.77%	98.23%
	Lost	3.62%	96.38%

PREDICTING NUMBER OF VOTES

DEMOGRAPHIC TRENDS

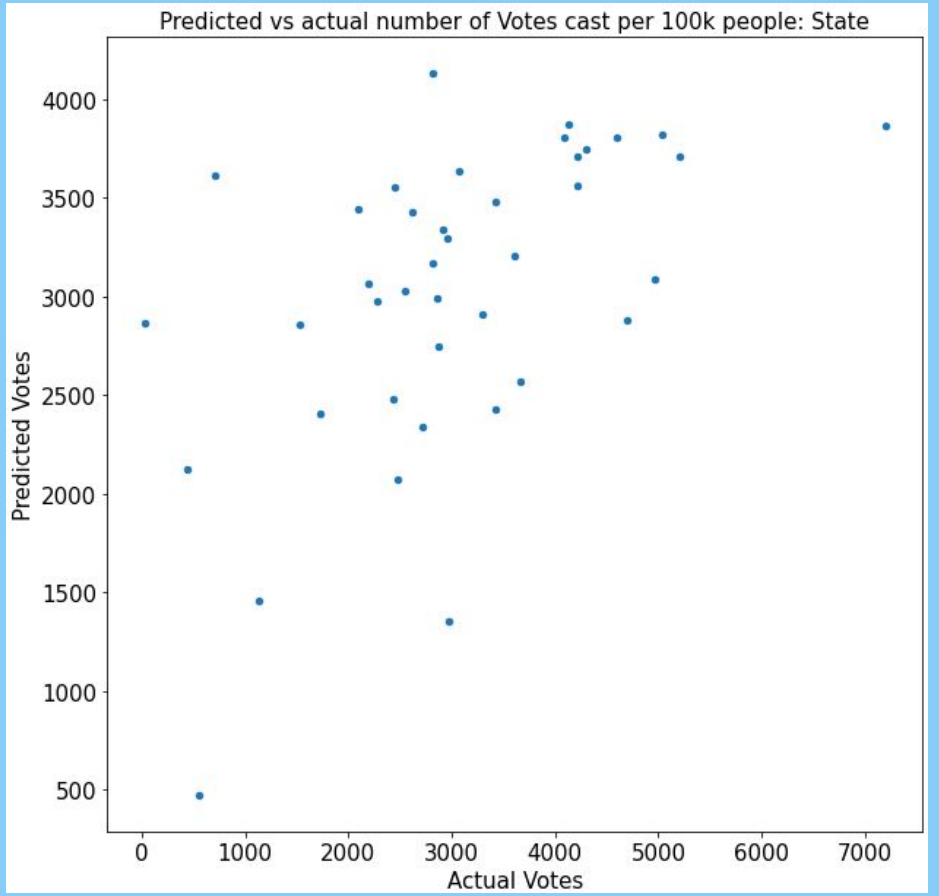


VOTE!

RACE & ETHNICITY:

Predicting number of votes

Can you predict the number of votes cast based on demographic conditions?



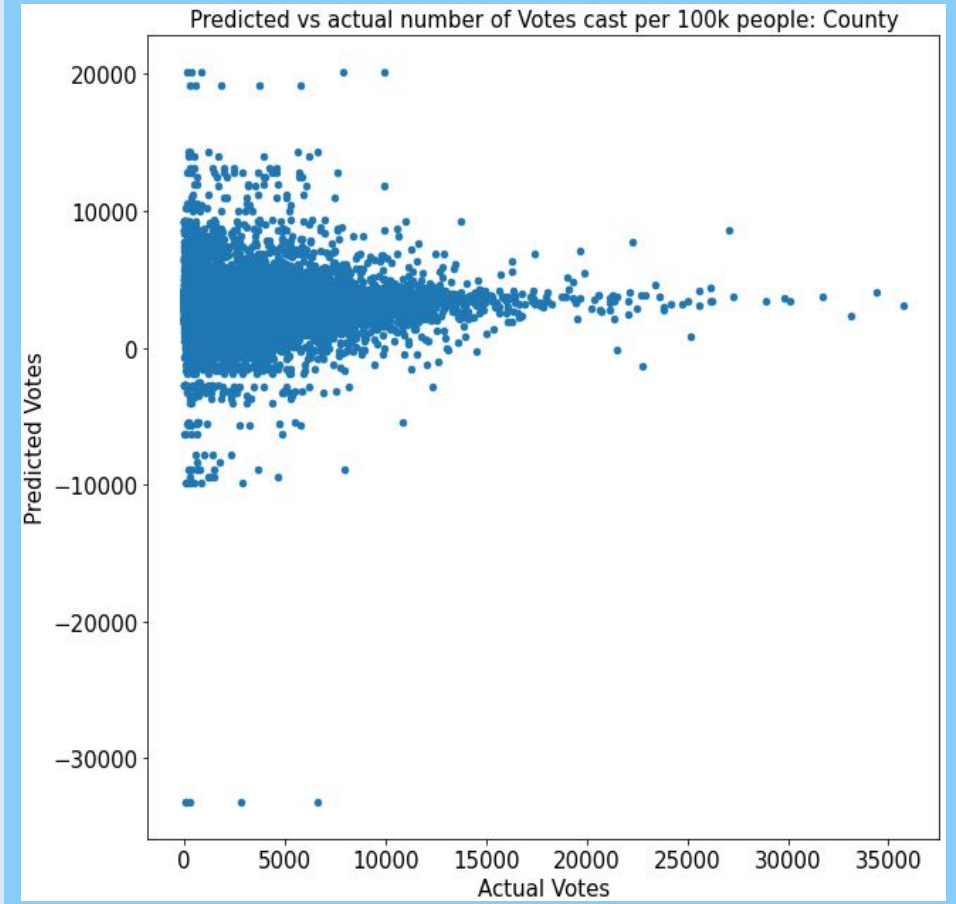
MSE = 1460084

R^2 = 0.2860650364766063

RACE & ETHNICITY:

Predicting number of votes

Can you predict the number of votes cast based on demographic conditions?



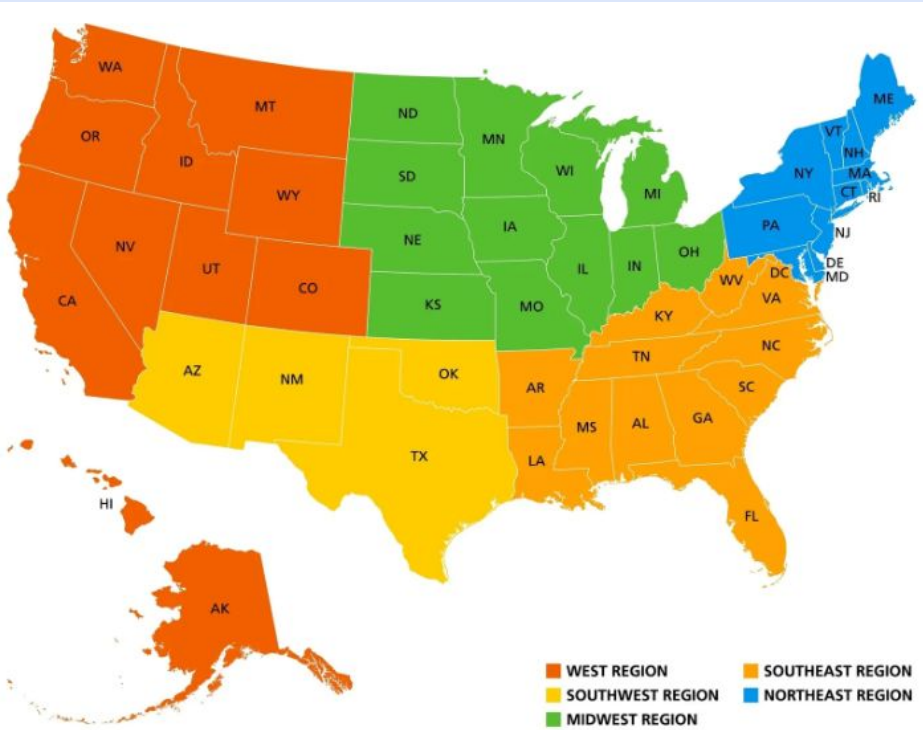
MSE = 12614795

R^2 = -0.24903105036492867

REGIONAL PREDICTIONS

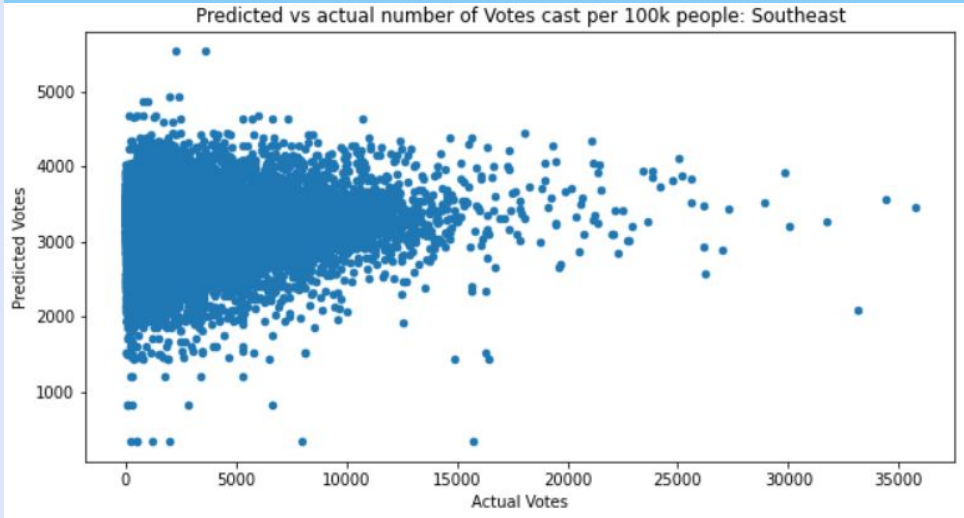
VOTE!

How Do Different Regions Affect Votes?



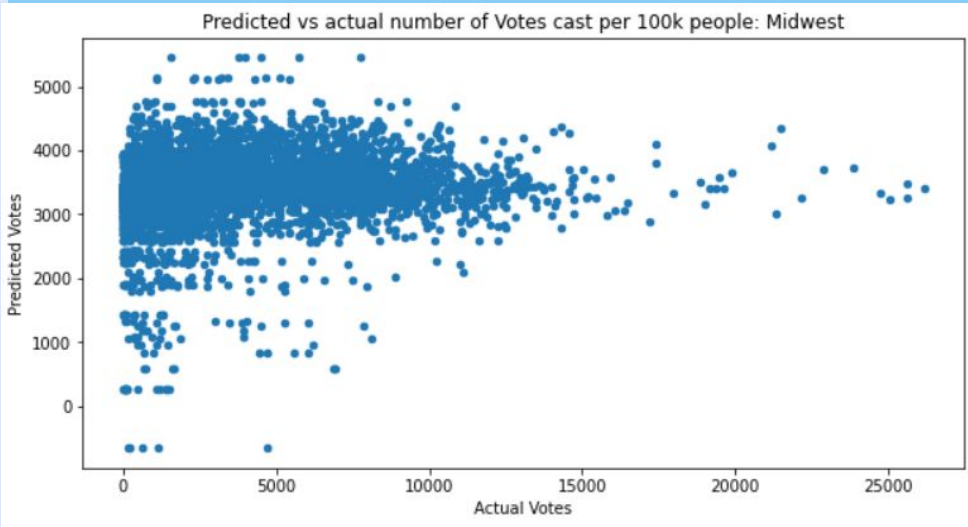
Southeast

Using demographics to predict votes by 100k people (points represent county)



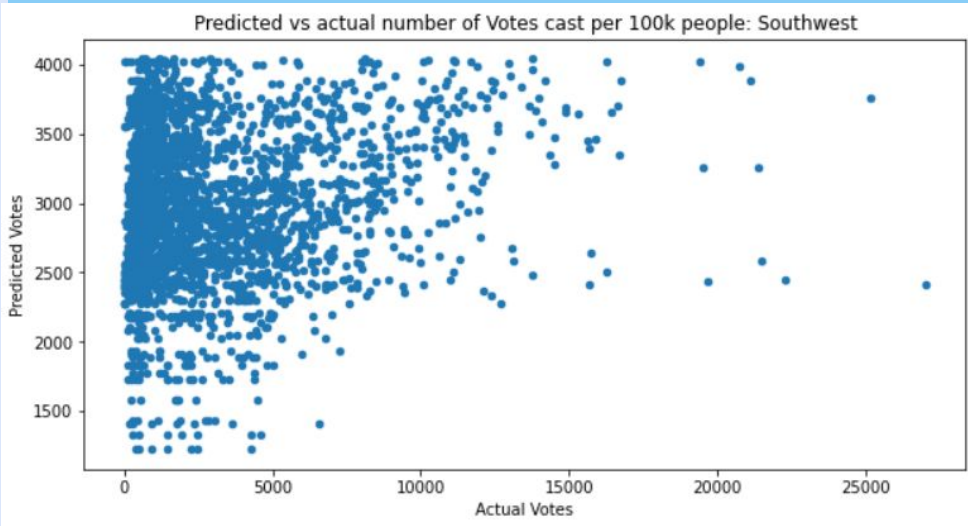
Midwest

Using demographics to predict votes by 100k people (points represent county)



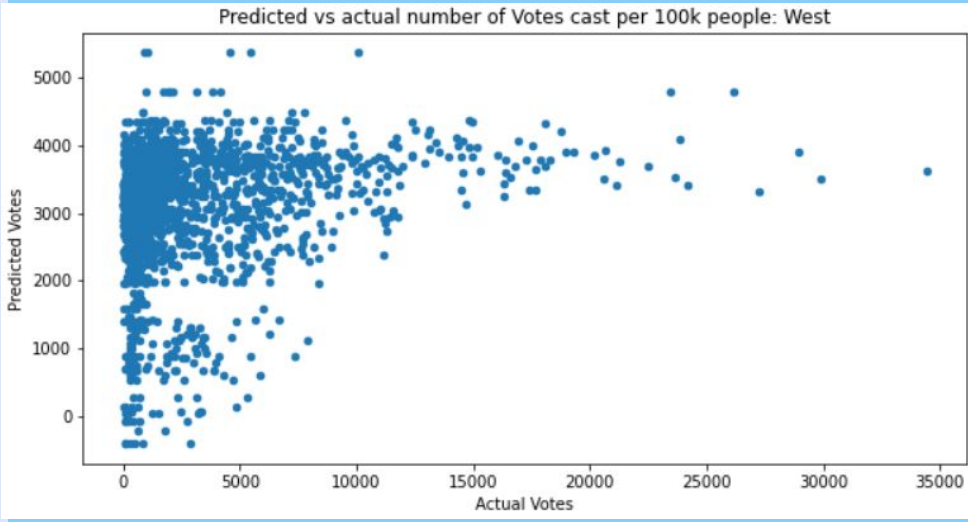
Southwest

Using demographics to predict votes by 100k people (points represent county)



West

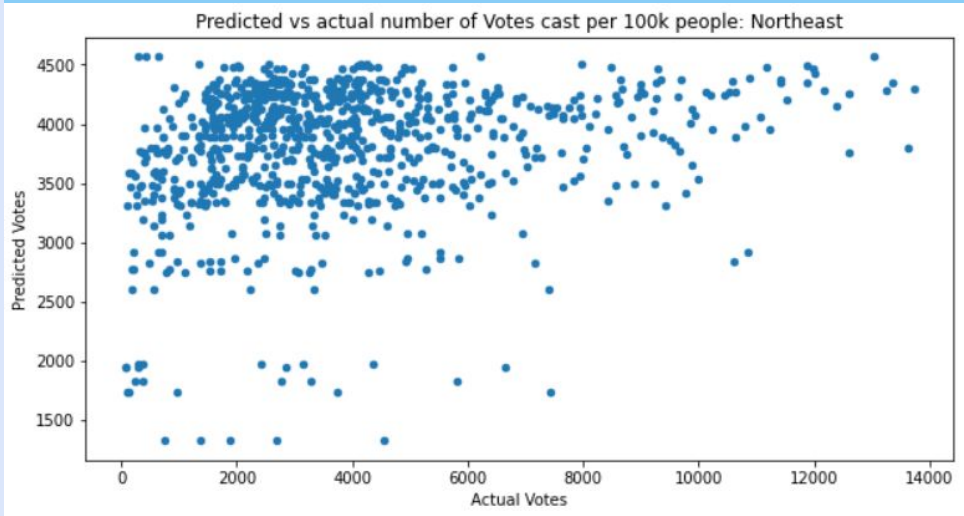
Using demographics to predict votes by 100k people (points represent county)



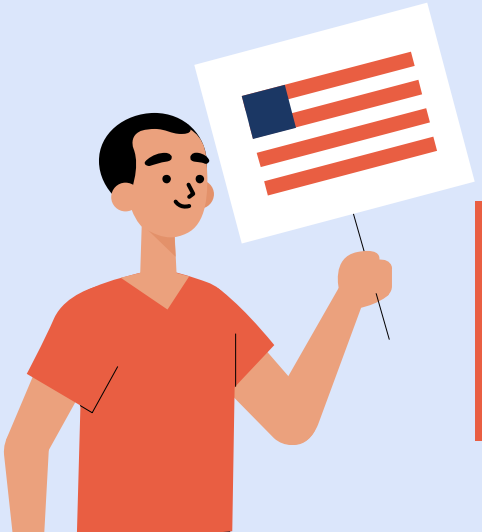
Northeast

- High Percentage of White Population
- **Highest** Median Income
- **Highest** Bachelor's Degree or Higher rate
- **Highest** by Population by Sq. Mile (Insanely High!)
- **Lowest** Poverty Level
- Many of the coefficients are **high and positive**

Using demographics to predict votes by 100k people (points represent county)



ECONOMIC PREDICTIONS



VOTE! ★

Sales:

Predicting number of votes

Columns Used:

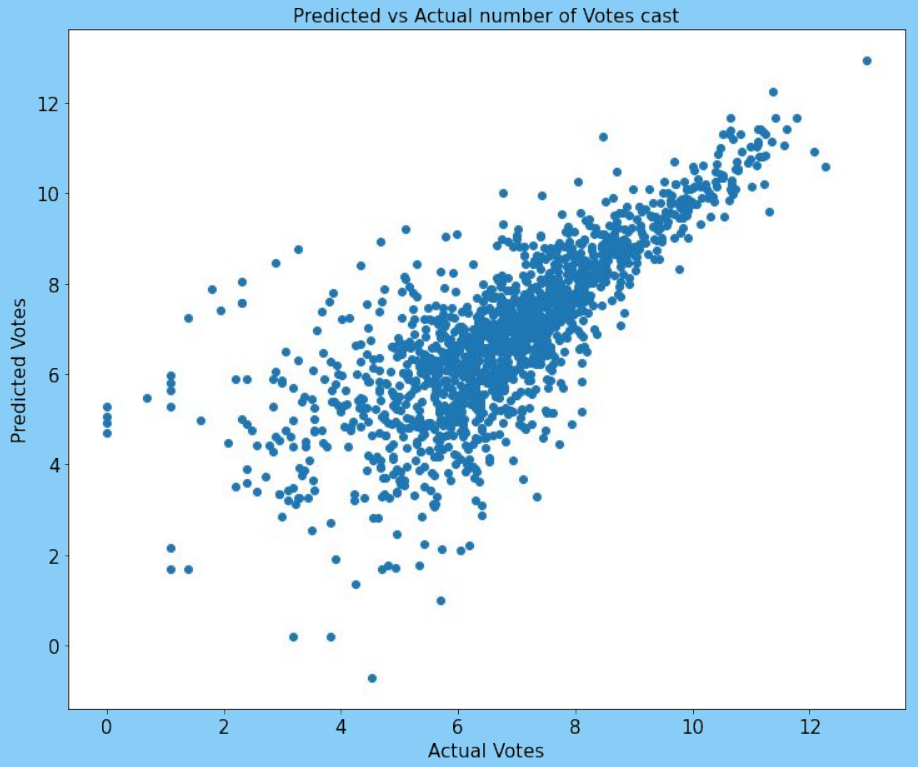
- Manufacturer shipments by $1000
- Merchant wholesaler sales by $1000
- Retail Sales by $1000
- Food Service Sales by $1000

Grouped by Parties

Democrats Vs. Republicans

Sales:
Predicting Number of votes
Democratic

Democratic candidates see a higher level of correlation



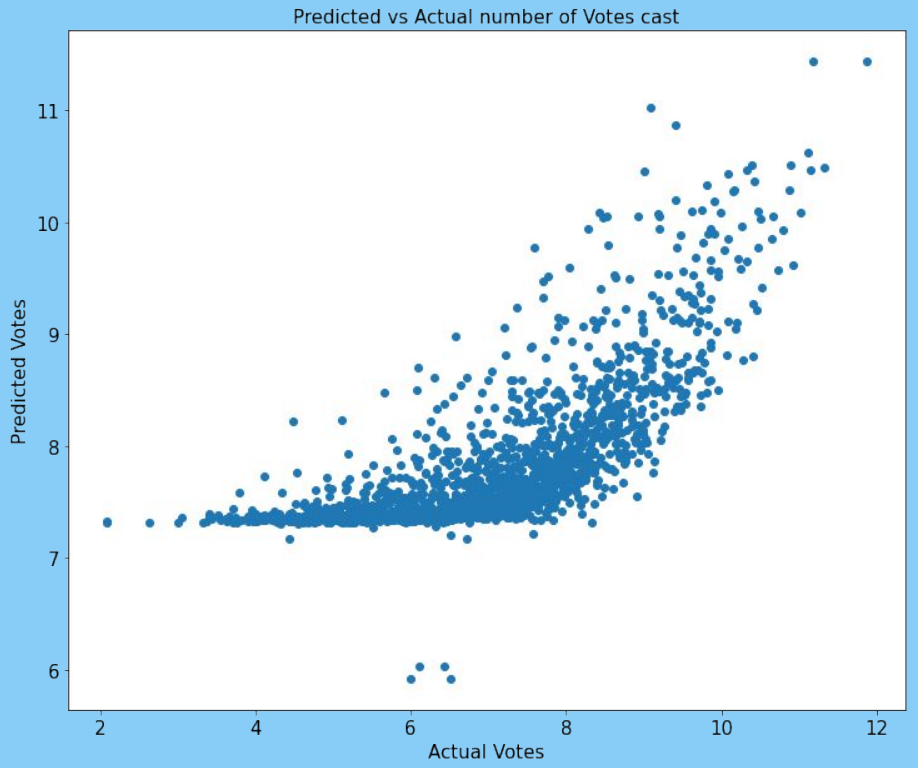
R^2 = 0.771

Sales:

Predicting Number of votes

Republican

Republican candidates see lower level of correlation than Democrats



R^2 = 0.612

Sales:
Predicting Primary Outcomes With Logistic Regression

Columns Used:

- Manufacturer shipments by $1000
- Merchant wholesaler sales by $1000
- Retail Sales by $1000
- Food Service Sales by $1000

Grouped by Parties

Democrats Vs. Republicans

Sales:

Predicting Primary Outcomes With Logistic Regression

Hillary Clinton
Democrats

Percentage of predicted primary elections won for Clinton that are correct

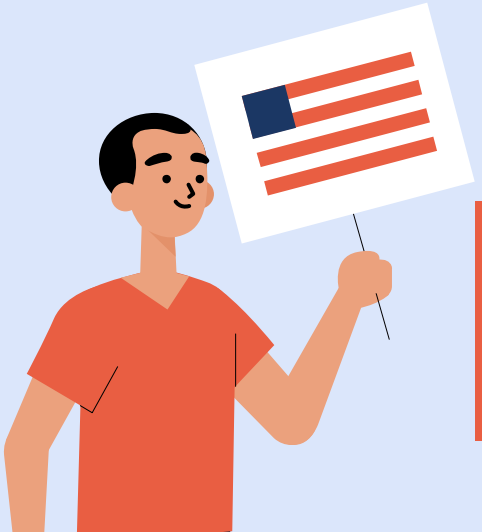
63.02%

Donald Trump
Republicans

Percentage of predicted primary elections won for Trump that are correct

78.01%

USING ALL CONDITIONS

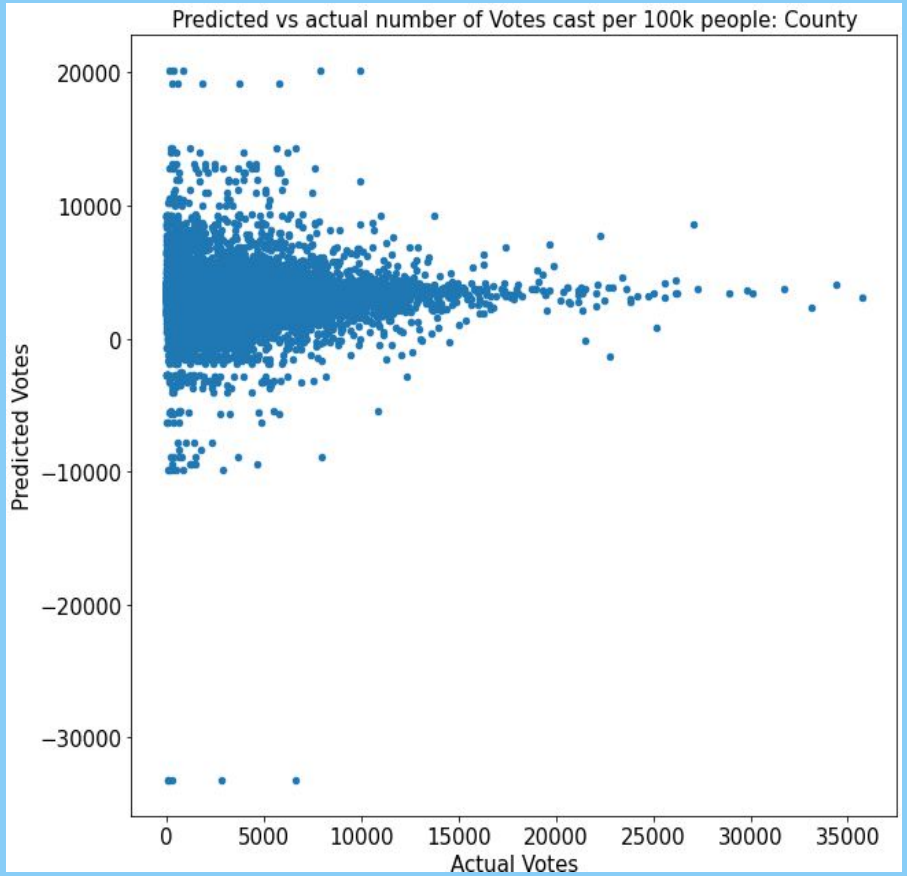


VOTE!

USING ALL CONDITIONS:

Predicting number of votes

Can you predict the number of votes cast based on demographic conditions?



MSE = 12614795

R^2 = -0.24903105036492867

Conclusions

VOTE!

Conclusions

- Voting trends **vary** by state, county and region
- There are a **variety of factors** involved in election outcomes
- Using demographic and economic conditions independently **does not** make valuable predictions
- Using a combination of factors might be more useful, how do you choose **which to use**?
- The Northeast voting patterns when looking at demographic are **quite differen**t than the rest of the other regions of the US