Predicting Anna Eshoo's Lead in 2024 Election

Overview

In this analysis I am using election results from the 2020 election between Anna Eshoo and Rishi Kumar, and 2020 and 2021 census data of some census tracts from districts 16 and 18. The goal of this analysis is to predict by how much Anna Eshoo will win the 2024 election. I am considering the census variables population size, white population, black population, Asian population, male to female ratio, mean income, and percent of population over 25 with bachelor's degrees.

Table of training data

| Precincts	Census_tract	pop_size	white_%	black_%	asian_%	med_age	male:female	mean_income	bach_degree+_%_(25 yrs +)	Eshoo_lead
2729, 2734, 3783	5070.03	2793	73.47	0.43	15.83	50.7	1.08	219,688	61.58	872
2751, 3783	5068.01	6454	62	0.84	23.75	43.4	0.972	97,717	44.16	54
3434	5098.01	6196	51.53	1.07	30.57	42.2	0.866	86,681	51.05	868
2317, 2330	5100.01	6295	53.12	0.49	36.36	50.1	1.013	149742	59.17	1773
4688, 4687, 3783	5075	5984	27.47	0.17	66.41	48.7	1.06	132736	61.93	-19
3804	5065.05	4788	24.69	5.4	21.78	31.6	1.218	51875	26.55	276
8049, 3783, 5567	5021.03	4449	30.28	5.37	15.73	32.7	1.02	36717	24.9	183
2542	5130	14272	53.6	5.63	25.8	23	1.212	37548	23.12	518
3783	5119.15	3562	32.65	4.13	18.75	36.1	0.891	49274	29.17	56
2726, 3783	5067.03	3937	54.1	3.1	21.54	39.6	0.912	81048	32.84	227

Graphs of independent variables vs. Eshoo lead

In Tableau, I created a calculated field in order to adjust Eshoo's lead according to population size. Next, I compared this adjusted lead to each independent variable. The graphs of each are below:
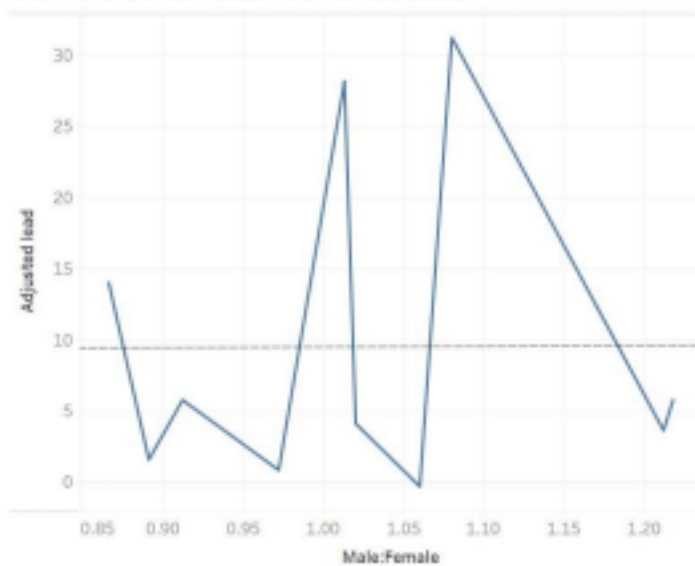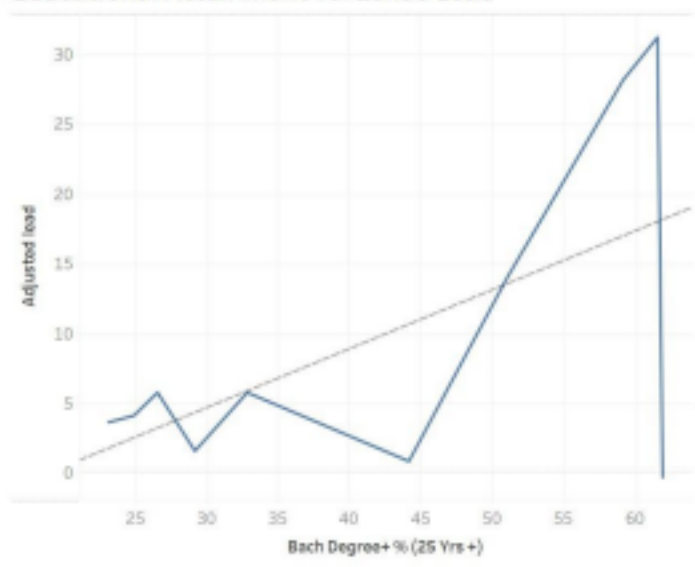
Race Proportion vs. Eshoo Lead
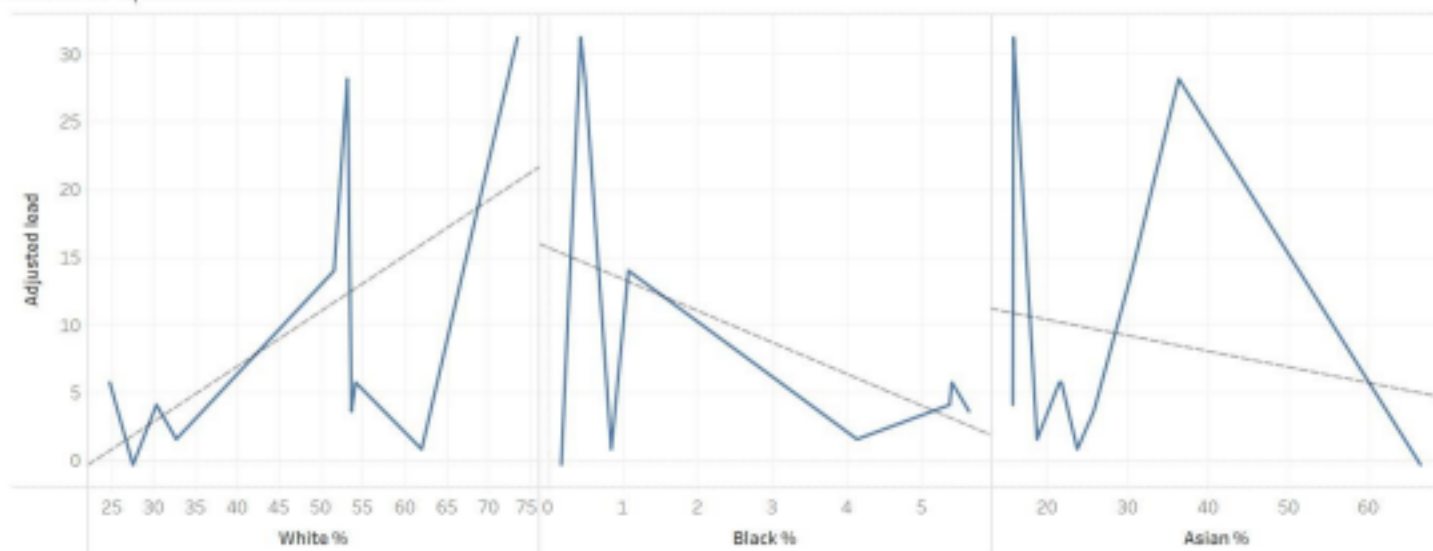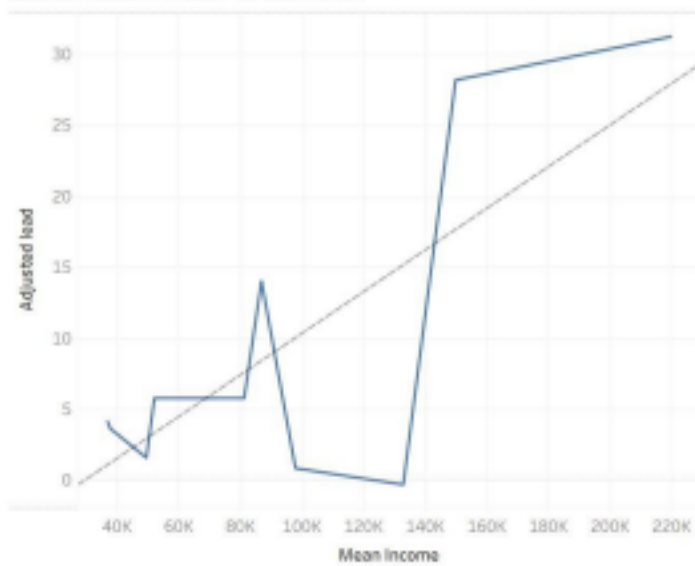

Median Age vs. Eshoo Lead
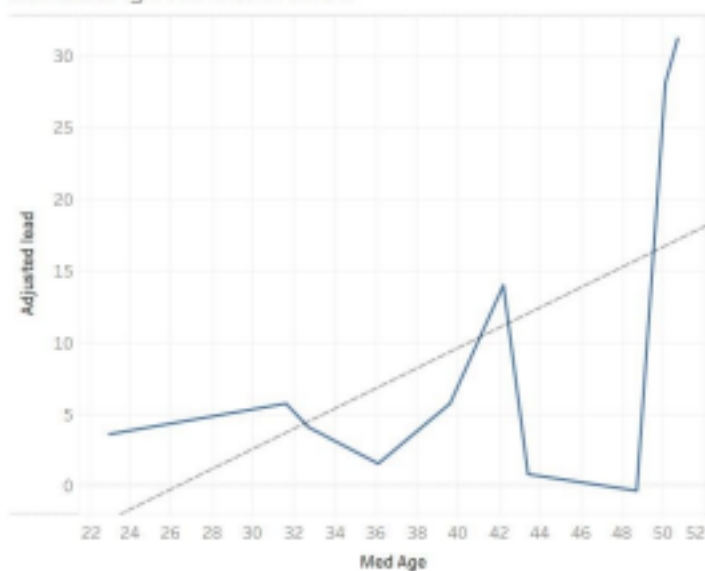
Male to Female Ratio vs. Eshoo Lead

Educational Attainment vs. Eshoo Lead

Mean Income vs. Eshoo Lead

	White %	Black %	Asian %	Med age	Male/Female ratio	Income	Bach degrees or higher
p-value	0.07	0.17	0.67	0.09	0.99	0.01	0.07
r-squared	0.35	0.22	0.02	0.31	0.00002	0.57	0.35
Slope	Positive	Negative	Slight negative	Positive	Flat	Positive	Positive

Conclusions from graphs

Due to the small sample size, these results are not reliable. However, the most reliable indicators of Eshoo's lead (adjusted for population size of census tract) were educational attainment, income, whites as percent of population, and median age. I found that male to female ratio was an irrelevant data point. The slope of each graph tells us about independent variable's support for Eshoo. The income graph suggests that higher income and college educated people are more likely to vote for Eshoo. The graph of black percent of population suggests that black people may be less likely to vote for Eshoo.

Making training set

I used the Sklearn library in Python to make a linear regression model with this my training data

Eshoo_lead	Prediction	Difference
872	939.535959	67.535959
54		2 2
868	795.655455	72.344545
	1773 16	

		1
-19		7 47
276		9 51
183		3 3
518	519.484367	1.484367
56		9 9
227	131.252968	95.747032

This table shows the actual Eshoo lead, versus what my linear regression model predicted, as well as the difference between them.

Actual lead vs. predicted lead
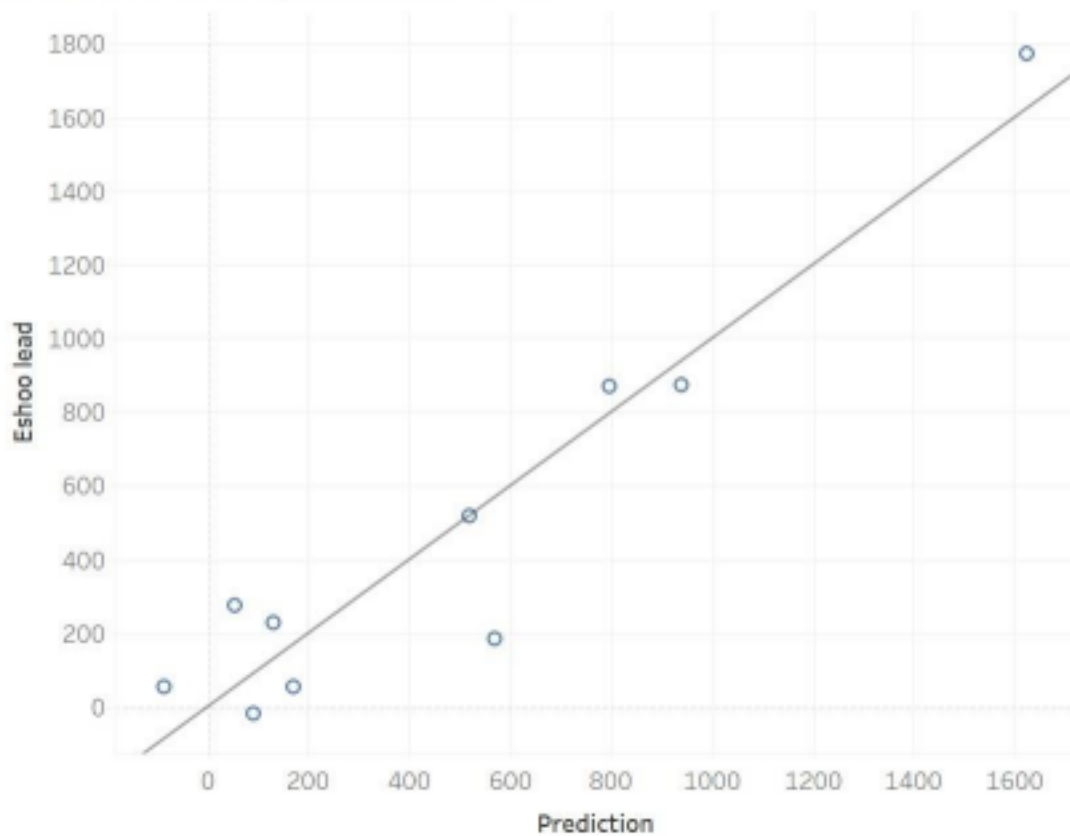
I graphed predicted lead versus actual lead to test the accuracy of my model. The p-value is less that 0.000001 and the r squared value is 0.9.

Predicting 2024 Election

Next, I'm going to use census data from 2021 (the most recent year available) to predict the outcome of the 2024 election in district 16, using census tracts common to the old district 18 and the new district 16. It is highly likely that this seat will go to a Democrat, since they have held this seat for decades and *The Cook Report* rates this district as 26 percentage points more democratic than the national average. Therefore, I will be predicting by how much Eshoo wins the seat.

Results

The data table for my predictions table is too large to paste here, so here is a link:
file:///C:/Users/franc/OneDrive/RANDOM/Predictions%20-%20Sheet1.pdf

The predictions from my linear regression model are highly inaccurate. For many of the predictions, Eshoo's predicted lead is much larger than the population size of the area. One of my predictions is above 1 million when the population is only about 4000. I looked at the coefficients of my linear regression equation and the p-values and r-squared values of my independent variables to see where my model went wrong. The coefficients, p-values, and r-squared values are in the table below:

	Pop. size	White %	Black %	Asian %	Age	Sex ratio	Income	Education
Coefficients	0.126	46.706	1293.831	-17.204	255.323	1224.473	-0.03	155.99
p-value	0.52	0.07	0.17	0.67	0.09	0.99	0.01	0.07
r-squared	0.05	0.35	0.22	0.02	0.31	0.00002	0.57	0.35

Black percentage and sex ratio are having a large effect on the model, despite having large p-values and low r-squared values. Also, population size is having a surprisingly low effect on the model. I will remake my linear regression model, this time excluding variables with p-values larger than 0.1, which means I will exclude population size, black percentage, Asian percentage, and sex ratio.

Predicting 2024 Election, Attempt #2

Link to new table: file:///C:/Users/franc/Downloads/Predictions%20-%20Sheet1.pdf

The new predictions are a big improvement over the old predictions. Now, there are no prediction values that are larger than the population size.

Final results and advice

I wanted to predict by how much Eshoo will win the 2024 election. The prediction table accounts for 134,519 people. The total population of district 16 is 751,432. From my updated prediction table, Eshoo had a total surplus of votes of 26,811. Therefore, my model predicts that Eshoo will win the 2024 election by 149,768.

I found that the variables that had the largest impact on Eshoo's win were education, income, age, and percent of population that is white. All of these variables had a positive correlation with Eshoo's lead. If Eshoo wants to have an even more commanding lead, she should focus on campaigning for people that are not college educated, have a lower income, are younger, or are non-whites.

Limitations and concluding thoughts

There were many limitations for this analysis. The most important limitation on the accuracy of my results was the sample size. If I had more time, I would've collected more data, using all of the census tracts from district 16 for my training data, rather than my small sample size of 10., Secondly, many important variables were left out. For example,

sex is generally considered an important variable when it comes to voting, but the only applicable variable from the census data was male to female ratio, which I found to be irrelevant to the results. My data did not analyze how different groups voted, but rather how different places voted when certain groups were a larger percent of the population. If I had more time, I would have also liked to analyze how the variables had an impact on one another. For example, what would the effect of each race population be if we controlled for education and income?