

Election Analysis

Erma, Hiep, Jason, Mike,
Blake

Project Summary: How To Win

For our final group project, we are examining the relationship between how much money is spent by United States Senate campaigns, how the campaigns spend their money, and the success of these campaigns.

We are interested in how often the Senate candidate that spends the most wins the election, but also *how* the campaigns spend their funds. Some examples of how campaigns spend their money include staff salaries, advertising expenses, travel expenses, and campaign materials.

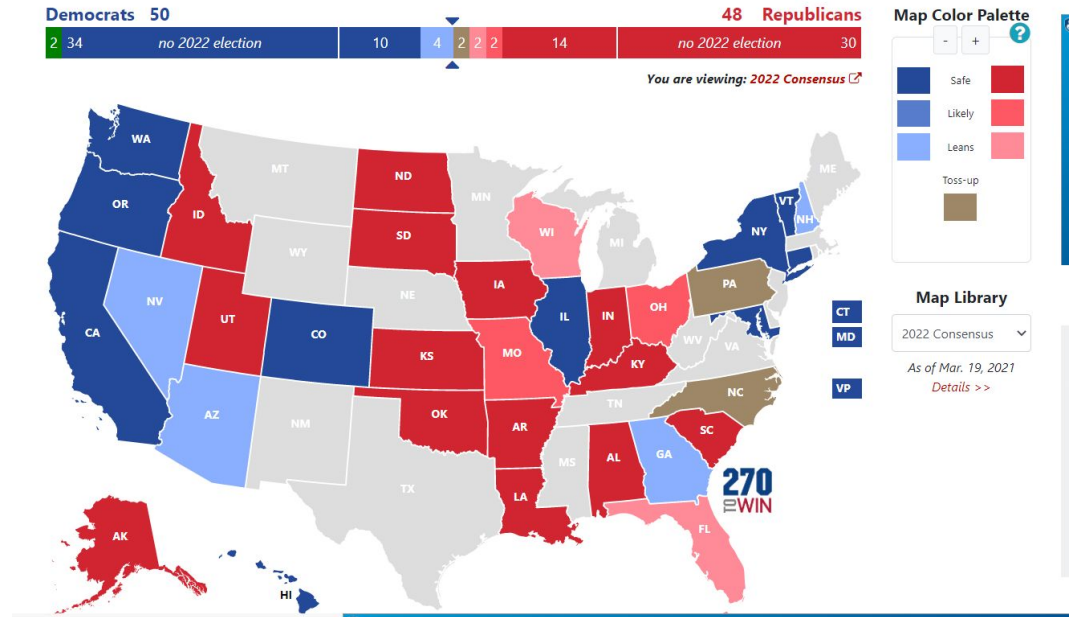
We will measure success in multiple ways including whether or not the Senate candidates win their campaigns, but also how well they do relative to expectations. Regardless of victory, gaining a couple of percentage points, in a state that has historically been partisan, demonstrates not only that campaigns can successfully reach people, but also that the USA is forever demographically and ideologically changing.

We have chosen to look at a handful of competitive purple states that dictate what will happen on Capitol Hill. The Senate is home to some of the world's most powerful people, and understanding how this power is acquired is not only good for candidates hoping to gain access, but also for the health of American Democracy. The states we have chosen to examine will dictate where power lies come 2022.

Competitive States

- Georgia
- Pennsylvania
- Ohio
- Arizona
- Colorado
- New Hampshire
- South Carolina

We chose these states because they have a wide range of demographics, histories, and cultural differences. We hope to balance the states that lean blue with those that lean red and states from similar regions of the country with one another.



Machine Learning - Hiep

1. Base on the [president_dataset.csv](#), we found that the more population the more total votes. So we use **Linear regression** to **predict** the value of a target variable based on given **predictor** variable. Refer to [Test Results](#)
2. Use RandomForestRegressor to predict the total votes and who will win next election for Senate_dataset due to this is supervised learning where we have input variables (x) and an output variable (Y). Refer to [Test Results](#)
3. Apply deep-learning neural networks model to analyze from [senate_dataset.csv](#). The outcome of this dataset is **totalvotes**, so we calculated the r2_score due to R2 is suitable for predicting continuous variable. A higher r-squared indicates a better fit for the model. Refer to [Test Results](#)
4. Apply deep-learning neural networks model to analyze from [independent_expenditures.csv](#). The outcome of this dataset is **expenditure_amount**, so we also calculated the r2_score due to R2 is suitable for predicting continuous variable. A higher r-squared indicates a better fit for the model. Refer to [Test Results](#)

Refer to [Elections Results](#) and [Web App](#) for details.

Database

- The database has several different interacting datasets
 - The information ranges from 1976-2020
 - Most of the data overlaps from 2000-2020
- We plan to use our database to better understand how spending influences the success of political candidates
- We have data that helps us not only understand spending, but also turnout in the several states we intend to use.
- Demographics and turnout will also be considered in relation to spending and victory

Demographic Information

Data Output

	index bigint	Year bigint	Gender text	Hispanic text	Race text	Population text	State text
1	0	2010	Male	Non Hispanic	White	1,832,998	Arizona
2	1	2010	Male	Non Hispanic	Black o...	127,328	Arizona
3	2	2010	Male	Non Hispanic	Americ...	126,085	Arizona
4	3	2010	Male	Non Hispanic	Asian	80,886	Arizona
5	4	2010	Male	Non Hispanic	Native ...	6,227	Arizona
6	5	2010	Male	Non Hispanic	Two or ...	54,380	Arizona
7	6	2010	Male	Hispanic	White	863,391	Arizona
8	7	2010	Male	Hispanic	Black o...	19,772	Arizona
9	8	2010	Male	Hispanic	Americ...	40,230	Arizona
10	9	2010	Male	Hispanic	Asian	7,985	Arizona
11	10	2010	Male	Hispanic	Native ...	2,509	Arizona
12	11	2010	Male	Hispanic	Two or ...	22,131	Arizona
13	12	2010	Female	Non Hispanic	White	1,874,603	Arizona
14	13	2010	Female	Non Hispanic	Black o...	114,756	Arizona
15	14	2010	Female	Non Hispanic	Americ...	132,838	Arizona
16	15	2010	Female	Non Hispanic	Asian	93,087	Arizona
17	16	2010	Female	Non Hispanic	Native ...	4,972	Arizona

- This data is important because it not only tells us how many people turned out to vote, but who actually did vote.
- Winning elections is becoming more and more of a science, and understanding demographics and how to reach them is essential to victory.

Senate Data

Data Output

	index bigint	year bigint	state text	state_po text	state_fips bigint	state_cen bigint	state_ic bigint	office text	district text	stage text	special boolean	candidate text	party_detailed text	writen boolean	mode text	candidatevote bigint
1	0	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	SAM STEIGER	REPUBLICAN	false	total	
2	1	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	WM. MATHE...	INDEPENDENT	false	total	
3	2	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	DENNIS DECO...	DEMOCRAT	false	total	
4	3	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	ALLAN NORW...	LIBERTARIAN	false	total	
5	4	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	BOB FIELD	INDEPENDENT	false	total	
6	5	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	JACK MCCOY	AMERICAN INDEP...	false	total	
7	6	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	S. I. (SAM) HA...	REPUBLICAN	false	total	
8	7	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	JOHN V. TUN...	DEMOCRAT	false	total	
9	8	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	OMARI MUSA	INDEPENDENT	false	total	
10	9	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	DAVID WALD	PEACE AND FREED...	false	total	
11	10	1976	CONNEC...	CT	9	16	1	US SENA...	statewide	gen	false	LOWELL P. WE...	REPUBLICAN	false	total	
12	11	1976	CONNEC...	CT	9	16	1	US SENA...	statewide	gen	false	SCATTER	[null]	false	total	
13	12	1976	CONNEC...	CT	9	16	1	US SENA...	statewide	gen	false	ROBERT BAR...	AMERICAN INDEP...	false	total	
14	13	1976	CONNEC...	CT	9	16	1	US SENA...	statewide	gen	false	GLORIA SCHA...	DEMOCRAT	false	total	
15	14	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	THOMAS C. M...	DEMOCRAT	false	total	
16	15	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	WILLIAM V. R...	REPUBLICAN	false	total	
17	16	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	DONALD G. GI...	AMERICAN	false	total	
18	17	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	JOHN A. MAS...	PROHIBITION	false	total	
19	18	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	JOSEPH F. MC...	NONE	false	total	
20	19	1976	FLORIDA	FL	12	59	43	US SENA...	statewide	gen	false	LAWTON CHIL...	DEMOCRAT	false	total	
21	20	1976	FLORIDA	FL	12	59	43	US SENA...	statewide	gen	false	SCATTER	[null]	false	total	
22	21	1976	FLORIDA	FL	12	59	43	US SENA...	statewide	gen	false	JOHN GRADY	REPUBLICAN	false	total	

With this dataset alone, tracking victory since 1976, can be used to predict winners based on trends. By combining this dataset with others that help us understand where and why the nation is changing, we can build an accurate model to predict how and why different candidates can win in purple states.

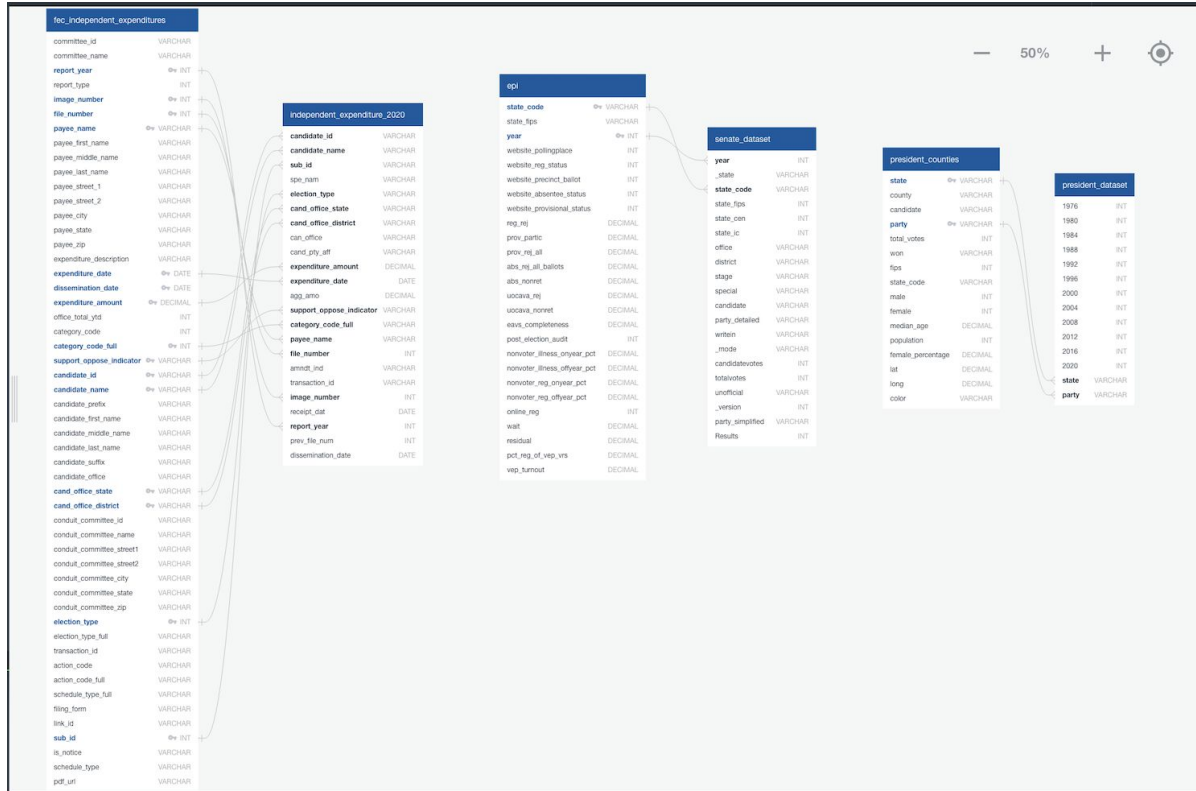
FEC Data

Data Output

	index bigint	Unnamed: 0 bigint	report_year bigint	image_number double precision	file_number double precision	payee_name text	expenditure_date text	dissemination_date text	expenditure_amount double precision	category_code_full text	sup text
1	0	0	2020	2.0201e+17	1444283	Good Works Matt...	5-Oct-20	[null]	35000	Phone bankers for feder...	S
2	1	1	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	85262.15	Digital ads	S
3	2	2	2020	2.0201e+17	1466866	BERLIN ROSEN, L...	15-Oct-20	25-Oct-20	80000	Projected phone bank co...	S
4	3	3	2020	2.0201e+17	1466866	CENTURY DIREC...	19-Oct-20	25-Oct-20	3742	Mailer	S
5	4	4	2020	2.02011e+17	1470736	AB PARTNERS PBC	31-Oct-20	31-Oct-20	48093.24	Digital ads	S
6	5	5	2020	2.0201e+17	1467573	AB PARTNERS PBC	23-Oct-20	27-Oct-20	290000	Projected digital ad buy	S
7	6	6	2020	2.0201e+17	1445694	CallHub	9-Oct-20	7-Oct-20	1000	Phone bank platform	S
8	7	7	2020	2.0201e+17	1445694	CallHub	8-Oct-20	7-Oct-20	1000	Phone bank platform	S
9	8	8	2020	2.0201e+17	1445694	CallHub	7-Oct-20	7-Oct-20	500	Phone bank platform	S
10	9	9	2020	2.0201e+17	1445694	1199SEIU United ...	3-Nov-20	7-Oct-20	0	Projected staff compens...	S
11	10	10	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	120451.5	Digital ads	S
12	11	11	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	120451.5	Digital ads	S
13	12	12	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	120451.5	Digital ads	S
14	13	13	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	120451.5	Digital ads	S
15	14	14	2020	2.0201e+17	1466866	AB PARTNERS PBC	23-Oct-20	25-Oct-20	200000	Digital ads	S
16	15	15	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	85262.15	Digital ads	S
17	16	16	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	85262.15	Digital ads	S
18	17	17	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	85262.15	Digital ads	S
19	18	18	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	2700	Digital ad setup	S
20	19	19	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	168.75	Digital ad setup	S
21	20	20	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	8979	Digital ads	S
22	21	21	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	6398.25	Digital ads	S

Understanding not only how much but also on what campaigns spend their money is essential to deconstructing how the senate works, and how to better run campaigns.

Database - Relationships



As mentioned above, the data we have found interacts with itself in many useful and interesting ways.

Tableau - Jason