

Voter Turnout for Primary Elections

Capstone Project

Miranda Burbridge - Data Analyst

Agenda

- Bio
- Project Context
- Define
- Design
- Deliver
- Summary, conclusions and next steps
- Appendix: list of supporting documents





Bio

Bio

- Education
 - Bachelor of Music from The University of Akron
 - Web Development Bootcamp Graduate from Eleven Fifty Academy
 - Soon to be graduate of Institute of Data Data Science and AI Bootcamp
- Professional experience
 - Data Analyst at DoorDash
- Relevance to the project
 - Currently running for Democratic State Central Committee in Louisiana



Project Context

Project Context

- Industry or domain
 - Politics
- Problem area
 - When running a campaign, especially one with a limited amount of time and budget, it is helpful to know which voters are most likely to turnout and vote.
- Why is this area interesting?
 - Using machine learning to predict the likelihood of voter turnout is interesting, especially in elections with historically low turnout.
- Previous work in this area
 - I am currently using this data for my own campaign.





Define

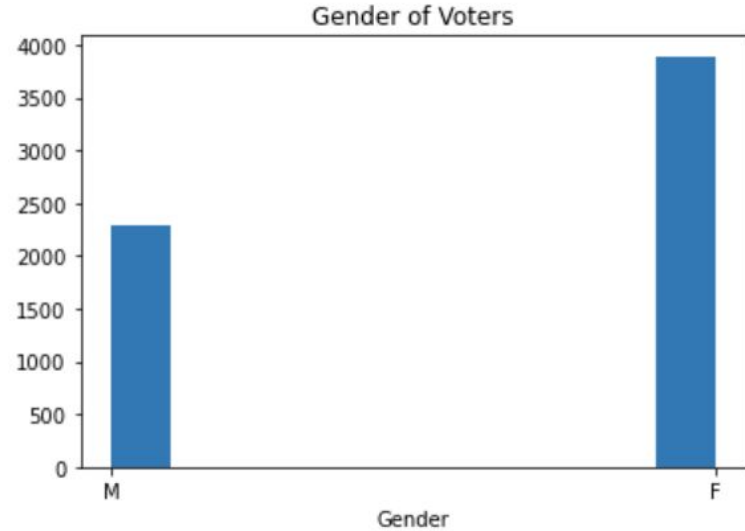
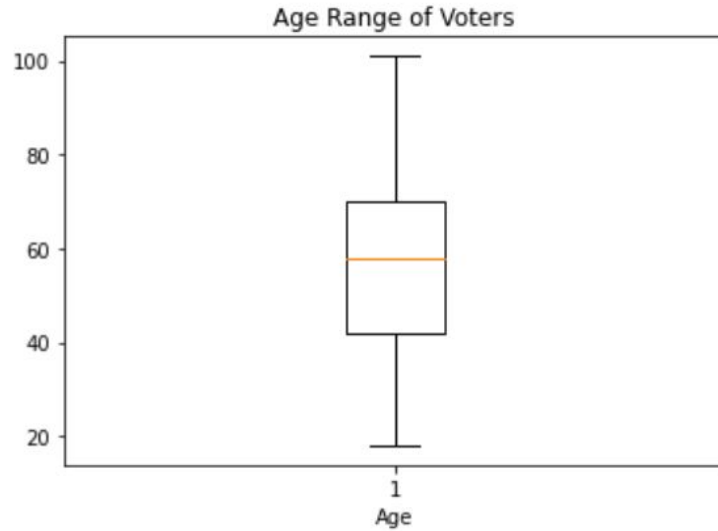
Define

- **Business aspects**
 - Stakeholders - politicians, campaign managers
 - Business question - Which voters should be targeted specifically during voter outreach?
 - Business value - Unfortunately not every registered voter will vote, so knowing which voters to target can help to save time and money during a campaign.
- **Data science aspects**
 - Data question - How can we predict which voters will vote in the presidential primary elections?
 - Data required - Voter List
 - Data sourced
 - Data source, description, volume and quality
 - Data provided by Blue Reboot, containing 6184 registered Democrats in Louisiana House District 90
 - How it can be sourced in the future? Voter lists can be purchased through the Louisiana Secretary of State

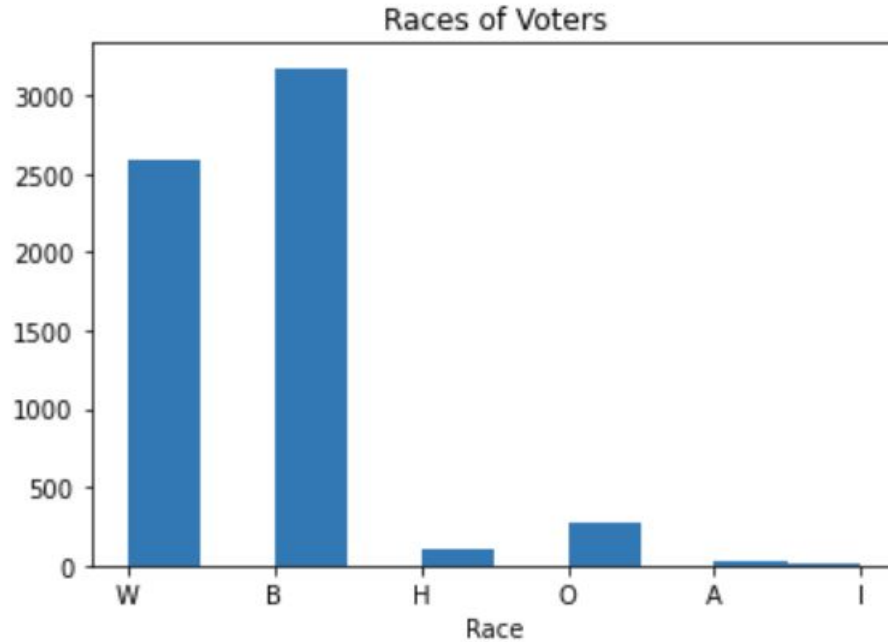


Design

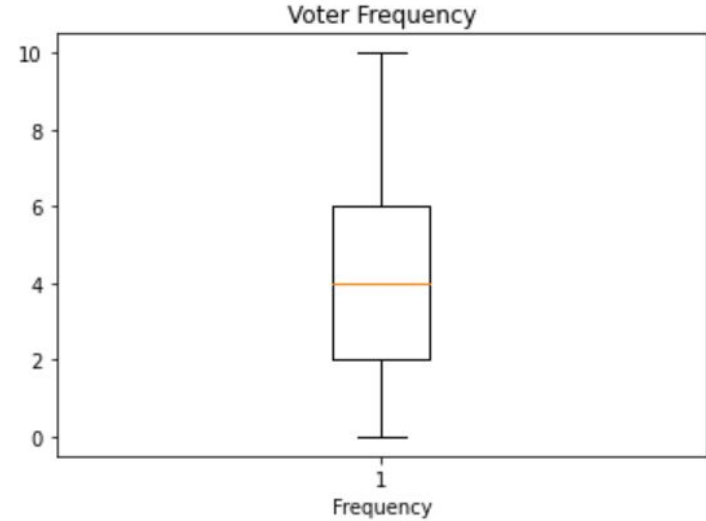
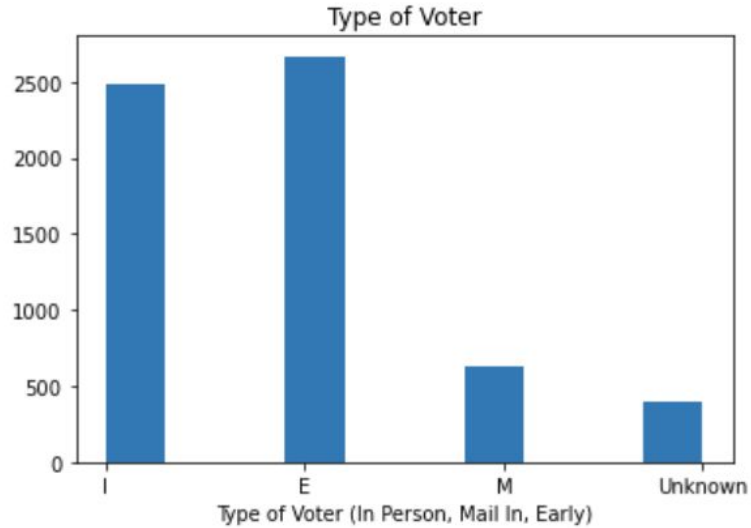
Data exploration, analysis and visualization



Data exploration, analysis and visualization



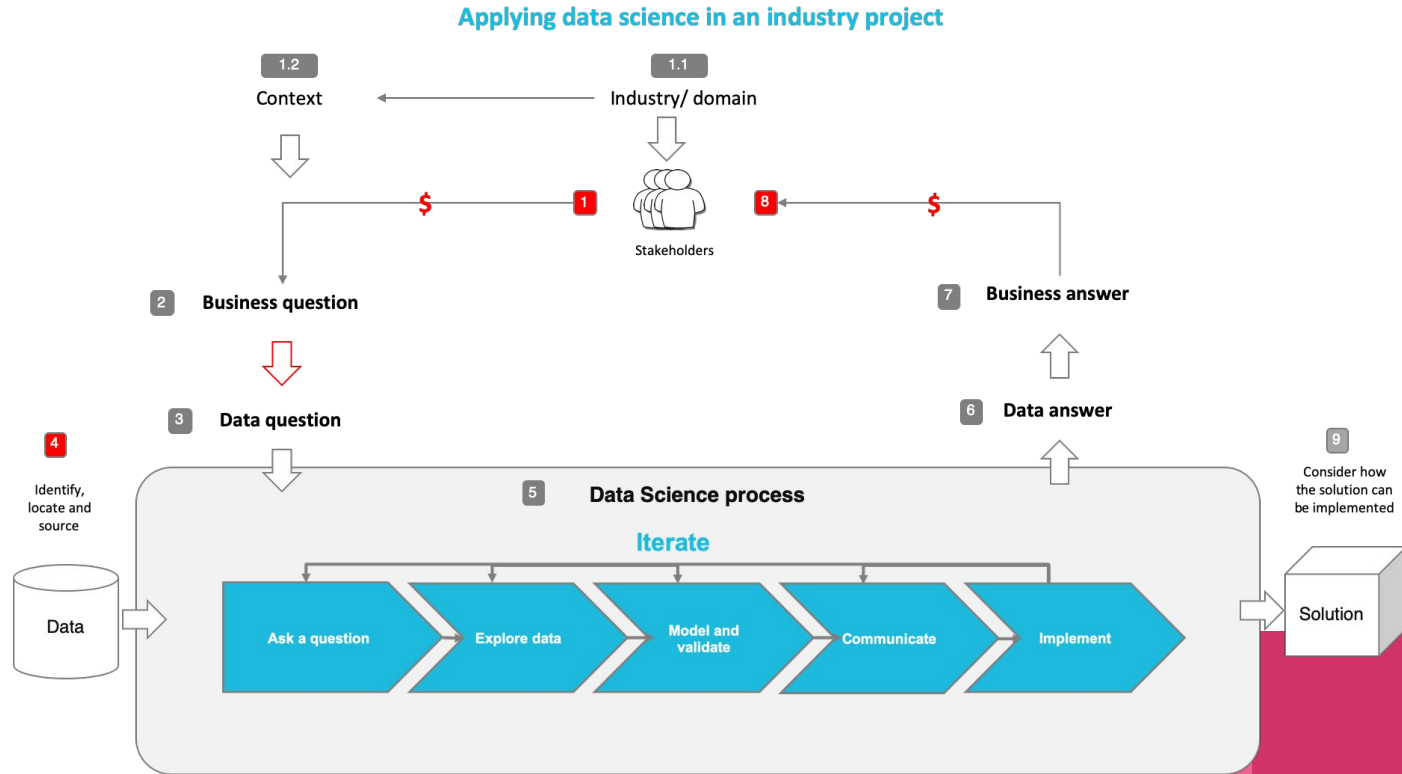
Data exploration, analysis and visualization



Data exploration, analysis and visualization

	HOU,C,3	PARISH,N,2,0	ZIP,C,5	AGE,C,3	GENDER,C,1	RACE,C,1	VOTE_FREQ,C,1	PPP,C,1 (P for presidential primary)	NEW,C,1	How they vote (in person, mail, early)
HOU,C,3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PARISH,N,2,0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
ZIP,C,5	NaN	NaN	1.000000	-0.028081	-0.023329	-0.110789	-0.004891	0.052369	0.000184	0.025024
AGE,C,3	NaN	NaN	-0.028081	1.000000	0.037230	0.149859	0.438710	0.235319	-0.233340	-0.050219
GENDER,C,1	NaN	NaN	-0.023329	0.037230	1.000000	0.012473	0.030711	-0.010321	-0.013969	-0.005250
RACE,C,1	NaN	NaN	-0.110789	0.149859	0.012473	1.000000	0.061420	-0.045722	-0.061562	-0.087369
VOTE_FREQ,C,1	NaN	NaN	-0.004891	0.438710	0.030711	0.061420	1.000000	0.558969	-0.156790	0.071426
PPP,C,1 (P for presidential primary)	NaN	NaN	0.052369	0.235319	-0.010321	-0.045722	0.558969	1.000000	-0.155596	0.078441
NEW,C,1	NaN	NaN	0.000184	-0.233340	-0.013969	-0.061562	-0.156790	-0.155596	1.000000	-0.204180
How they vote (in person, mail, early)	NaN	NaN	0.025024	-0.050219	-0.005250	-0.087369	0.071426	0.078441	-0.204180	1.000000

Process





Deliver

Deliver

- Feature engineering

- When looking at correlations, the feature with the strongest correlation to presidential primary voters was voter frequency at .56 followed by age at .24.
- Zip code, age, voter frequency, gender, race, voting method, and if they are a new voter were all used as features.

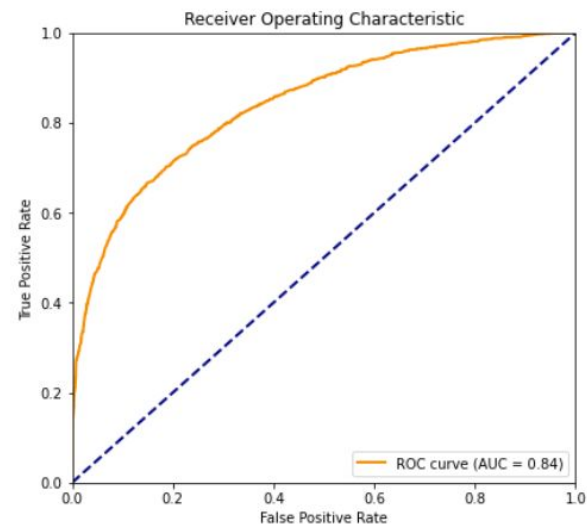
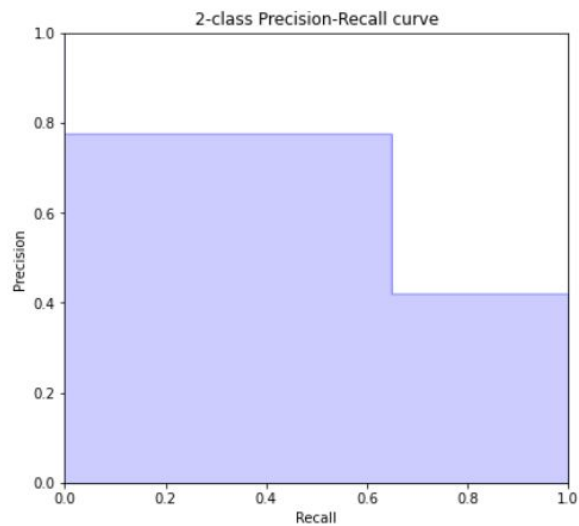
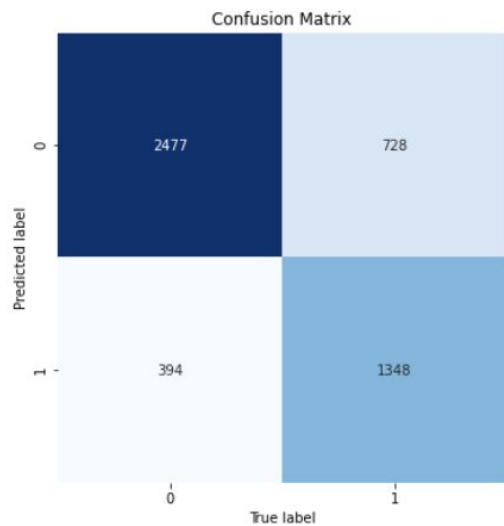


Deliver

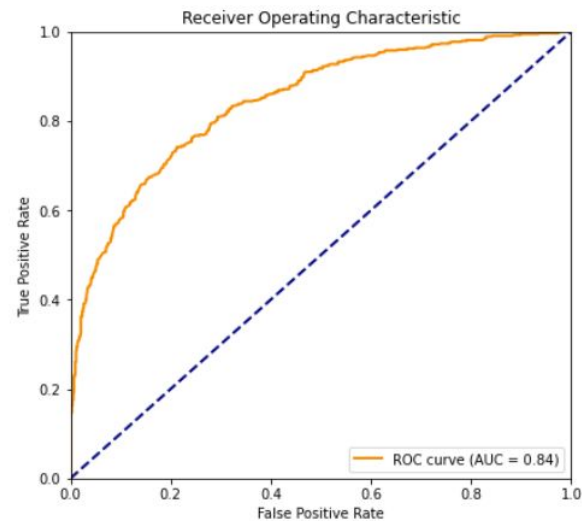
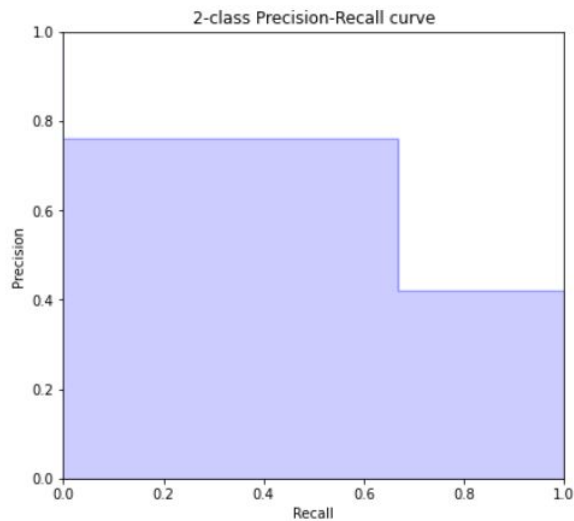
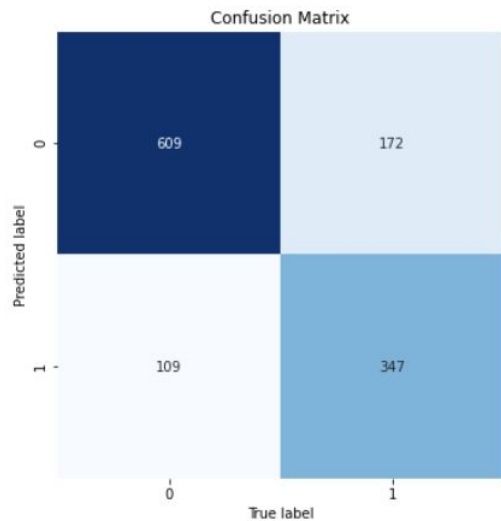
- Machine models used and their evaluation metrics
 - K Nearest Neighbor, RandomForest, GaussianNB, Logistic Regression, Stacking, Boosting, and Bagging were all trained and tested.
 - The highest performing model was Boosting
 - Training Scores
 - Accuracy : 0.7732
 - Precision: 0.7738
 - Recall : 0.6493
 - ROC AUC : 0.8400
 - Testing Scores
 - Accuracy : 0.7728
 - Precision: 0.7610
 - Recall : 0.6686
 - ROC AUC : 0.8433



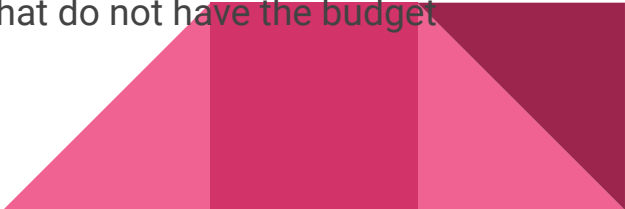
Deliver - Training Scores




Deliver - Testing Scores



Deliver

- How does the model fit in the overall solution
 - Who will use it?
 - Candidates, campaign managers, and other political volunteers can use this model to determine which voters are most likely to participate in a presidential primary election.
 - When is it used?
 - Presidential primaries tend to not get as high of a voter turnout as presidential elections, and “down ballot” races get less attention, so being able to target the voters most likely to participate can make winning one of these elections easier and more cost and time effective for candidates.
 - How does it benefit the business?
 - This model is especially helpful for smaller campaigns that do not have the budget and/or staff to contact every voter in their district.
- 



Summary, Conclusion, and Next Steps

Summary, Conclusion, and Next Steps

- Summary

- By using the Boosting machine learning model, we can predict which voters are most likely to participate in a presidential primary election.

- Conclusions

- Anyone running a campaign that shares the presidential primary election can use this information to better target their voter outreach.

- Next steps

- Expanding the training and testing data to include more districts, political parties, and elections could make this model useful for more campaigns than just those during presidential primaries.





Questions?

Appendices

Appendices

Louisiana Voter Lists:

<https://www.sos.la.gov/ElectionsAndVoting/BecomeACandidate/PurchaseVoterLists/Pages/default.aspx>

Data:

https://drive.google.com/file/d/1GyY1C-H6pKMvtJPv1HRLicpZLVS9rJ2f/view?usp=drive_link

