I worked on this project alone using jupyter notebooks. I used the FEC Financial Campaign Data from 2005-2006 as that is the most recent year they recorded wins and losses for the various elections. The data includes many details about campaign financing during the election cycle, including the total money made, where the funds came from, whether or not they won the election, and more. The data was given in a .txt document and came without column headers, so I had to convert the .txt file to a .csv file and created column names based on the documentation from the FEC website. I made sure any strings or characters I was going to be working with were uppercase because when I was initially working with the data there was both "DEM" and "Dem" when referring to a candidate who was a democrat. So to avoid any issues that I wouldn't be able to see, I made them everything I was working with uppercase.

I then wanted to explore the different political parties and the funds that they got. I created a bar plot with seaborn - pretty much every plot or chart is done with seaborn or matplotlib - to show the total funding/receipts that each party got. After that I took the top six parties that had the most total receipts. These parties are the Democratic Party(DEM), the Republican Party(REP), Independents(IND), the Democratic the Minnesota Democratic–Farmer–Labor Party(DFL), the Libertarian Party(LIB), and the Constitution Party(CST). I showed a simple bar graph of the total receipts again just to show a clearer picture of the parties I would be working with. I wanted to show where the funding for each party came from, so I created a stacked bar chart to show this information. I found that individual contributions are the driving force for most of the parties, with the Libertarian and Democratic-Farmer parties being the exceptions. Libertarians seem to have a lot of candidate loans and the DFL gets a lot from other political committee contributions. This is likely because the DFL is a part of the democratic party. I then decided to show what the distribution of funds

was in general to show overall where the money comes from. I used a pie chart to display this information clearly. When working with the specific parties, I split the data and created a new dataset just with those parties.

I also split the data when I showed a logistic model trying to determine the impact total receipts has on winning the general election. This was important because a lot of the data points had NaN values because they didn't participate in the general election. To do the model, I imported sklearn modules that had functions to do the logistic regression, standardize the data, test and train the data, and more. After splitting the data into candidates who had participated in the general election, I gathered the variables I wanted to test - total receipts and win/loss - I split the data into testing and training data, then I standardized the data using the StandardScaler() from sklearn, and finally I fit the model. When fitting the model, I initially didn't standardize the data which led to weird and incorrect results. This part was the most challenging and time consuming, just remembering how to train the data, figuring out how to standardize the data, and remembering how to understand and fit logistic models. The results ended up coming out pretty well. The model was generally a good predictor of whether a candidate would win or lose based on the total receipts, but it wasn't amazing. The results showed that, yes, total receipts does seem to indicate a candidate will have a better chance of winning or losing, but there are also many other factors that go into it as well. I go more into detail about the results in the notebook. I also plotted an AUC-ROC Curve with an AUC score to show that the logistic model was a good fit for the data.

If I were to continue this project, I would definitely take a look at the differences between the data we see now and the data from back then. There has been a pretty huge shift in the political landscape since 2006 and seeing how that change affected the money being received.