You need to submit a 1-page max document on what techniques you used, your findings, what your data proved or disproved, what else it might prove or disprove with additional data, and what is beyond the scope of your data. Explain what additional resources would be necessary for any follow-up steps.

<div align="right">

James Clark
Generation Data

</div>

Diving into this Data, I appreciated that Generation Data had instructed us on new technological tools such as Tableau for EDA, but found that I was most comfortable with Python. I therefore read a csv into pandas and created pie charts, time series, and bar graphs of Data. Connecting to Civis turned out to be one of the most challenging aspects of this pipeline, and I easily lost 8 hours trying to figure out how to connect to civis API and Postgres respectively. Thankfully with the help of Ryan, I discovered a way to directly load in Data from any Sql server into Pandas, which I then used to connect to my own Schema in Civis. Unfortunately, I might not have enough RAM to pull in the data efficiently this way, because trying to pull the whole table would throttle and my Internet and also semi-crash my notebook. I however love this way of querying sql to Pandas, and I am very happy Ryan helped me with this part of the pipeline.

I found some interesting things while conducting the EDA of the Data. When I featured engineered to find out how long voters held onto ballots I found that they held them for an average of 11 days before returning them. I did a time series find when ballot returns peaked and found that they peaked well into August, so it appears that there alot of people that may hold on to their absentee ballots. I did a graph chart to find what day of the week ballots were issued, and I found overwhelmingly Mondays. Absentee ballots were most often returned during the five day work week, with relatively very few being returned on the weekend.. Perhaps this has something to do with the postal service? Lastly a pie chart I created of Ballots according to Party showed that Democrats were a slight plurality of absentee ballots.

As far as the data is concerned, I believe that there are more secrets to be teased out here. Both in Pandas and in Tableau, I know that if I had gotten Civis working sooner I would have conducted more EDA and probably presented my Data visualizations in a flashier way, but that wasn't in the cards. As far as resources I'd like to have: I would love to have a mentor like Ryan or Michael. Ryan was opening new dimensions for me when I worked with him, and honestly it would be amazing to learn from someone that advanced, and see if they could perhaps pull off some ml on the data. I tried to one hot-encode some categorical variables and do linear regression, but I couldn't get it done so I moved on.

As for making my Data Pipeline robust and Dynamic:

I could theoretically use cron job in conjunction with python/ and or bash script to update my github/code/server. I however did not have time to make these scripts, but I do have much interest in trying to figure out how to do this on my own. I am very thankful that I was able to connect to civis postgres server. Not to mention pull down a table from my own schema, that's a win for me. All in good time.