**Implementation**

EDA

Data Cleaning

Data cleaning is a critically important step in any machine learning project. Data cleaning refers to all kinds of tasks and activities to detect and repair errors in the data. Comparing Tweets and survey datasets, we found tweets from the messiest dataset since we restricted our survey questioners with required fields. In contrast, Twitter web scrap pulls all the tweets without any necessity. Further, we identified and removed the columns that contain a single value in survey data and duplicated tweets from the Twitter dataset as the measure of the data clearing process.

Handling missing values-

Removing outliers- Understanding the core domain of our project, we identified the uncertain situation and handled it as an outliers.

[Outilers Identifications and Deletions, Tableau Dashboard](https://public.tableau.com/views/Case2_16194646770520/Dashboard1?:language=en&:display_count=y&:origin=viz_share_link)

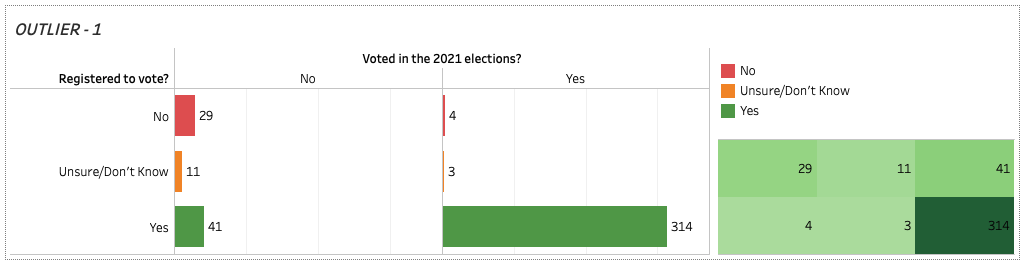


Figure 1

Narrowing the questioners to ‘Registered to vote?’ and ‘Voted in the 2021 elections?’ we spotted out an outlier. We have undergone a situation where a citizen has not registered to vote but has voted in the 2021 election, which is not possible. Hence we recognize four records as outliers and handled them.

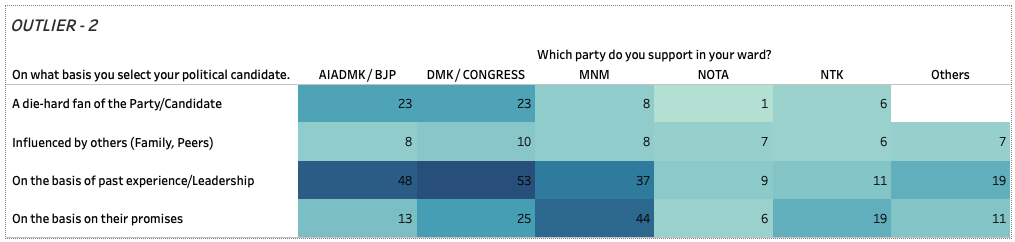


Figure 2

*Here, we can see that more people have voted based on their political candidate's past experience and leadership skills. Their support level has cast them a high number of votes for DMK of 53 votes, and AIADMK of 48 votes have been recorded based on past experience and leadership. On the line, MNM has 44 votes based on their promises.*

*The data also shows that AIADMK and DMK received more support and votes from their die-hard fans. The Outlier that can be identified here is The die-hard fans who have voted for NOTA. NOTA is not a party, and it indicates that the voter has not chosen to vote for any of the parties. Therefore, we cannot proceed with this DATA as it is not logical that a die-hard fan has voted for NOTA.*

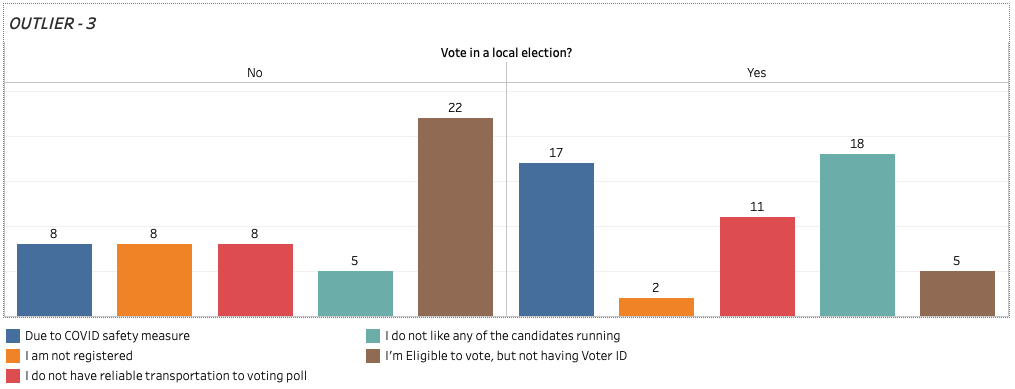


Figure 3

In this Data, we could see those who have voted and not voted in the local election. We can see that people have not voted due to factors like no reliable transportation, COVID safety measures, and others by analysing the data.

On the other side, we can see the data of people who have voted. The outlier spotted here is ‘2’ people have said that they have voted, but they are not registered to vote, and ‘5’have voted, but they do not have Voter Id to vote. Without registration and Voter ID, people cannot cast their votes. Therefore, the Data is not logical and unclear. So, we reject this Data before proceeding to the next step.

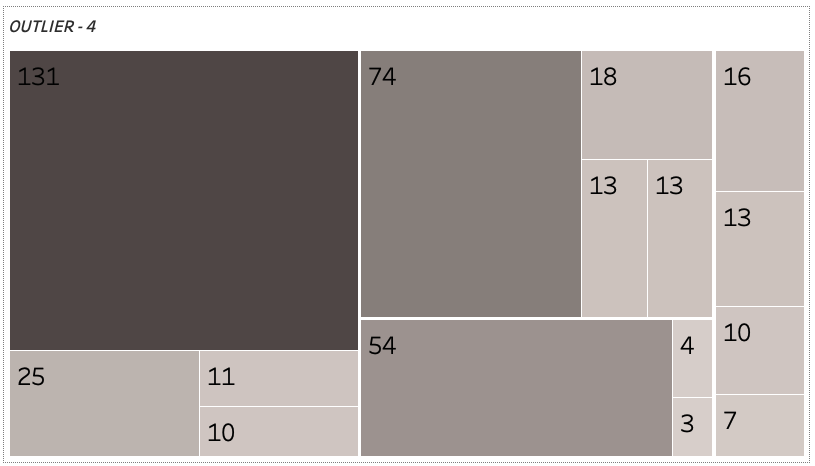


Figure 4

*Here we analyze that whether the Government is going in the right direction. People answering this should also have to give ratings for Edappadi Palanisamy. ‘111’ People who have said that it’s not going in the right direction have given a neutral rating, and ‘44’ people have rated it to be bad, and ‘44’ have said that it is very bad. The Outlier spotted here is ‘3’ People who have said that the Government is going in the right direction have also given bad and very bad ratings to Edappadi. This is because that they like the Government but personally do not like Edappadi. There is a bias in the dataset. So, we remove this data before proceeding to the next step.*

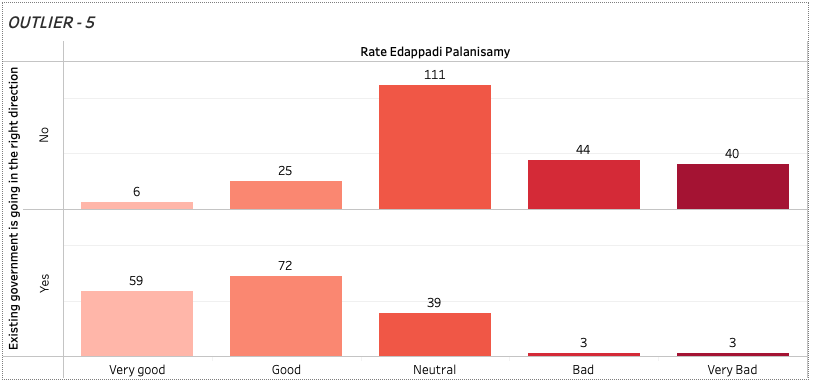


Figure 5

*This data shows how people have cast votes based on specific criteria. A high number of '131' people have voted based on past experience of the political candidate, and '74' people have voted based on the candidate's promises. '54'die-hard fans voted to the specific party they like and also were sure about whom they were voting for. The Outlier that can be identified here is '3' die-hard fan of the particular party said that they would vote for the party they like, but they also said they don't care about who wins. Being a die-hard fan for a particular party but doesn't care who wins is not logical. So, we reject this data before proceeding to the next step.*