**Gujarat-Demographic Voting pattern and the help of AAP to the BJP in ensuring the defeat of the Congress Party**

PM Narinder Modi was the Chief Minister of Gujarat for more than 12 years before becoming the Prime Minister of India. This makes Gujarat one of the most critical political States in India. Elections to the Legislative Assembly of Gujarat were held in the first week of December 2022. BJP created history by securing 52.50% of votes and winning 156 seats out of 182. Never in the history of Gujarat has any party won these many seats in an election. The opposition, Indian National Congress, could secure just 28.28% of votes and win only 17 seats- the lowest in history.

To begin analysing these elections, it becomes vital to compare them with the last elections in 2017. In 2017 the BJP secured 49.1% votes and could win just 99 seats. Indian National Congress and its alliance and supported independents won 81 seats with 41.4% votes.

From 99 to 156 seats and 49.1% to 52.50% vote share, the BJP, with just 3.45% additional votes, has won 57 extra seats, which is 31.32% of the total seats.

This project is an attempt to understand this anomaly and to have a look at voters’ preferences based on demography. The election data has been taken from the publicly available data at TCPD in Ashoka University, and the demography data of Gujarat has been taken from the website chanakyya.com. Both the data are in CSV files in the project folder.

NumPy, pandas, matplotlib, and sklearn are the libraries used. The code consists of four functions other than the main function. Using the input method, the initiation of code prompts the user first to enter the year of elections between 2017 and 2022. Then it asks the user to select one of the four demographic features, namely “URBAN”, “RURAL”, “SC”, and “ST”. SC stands for scheduled caste, and ST stands for the scheduled tribe. One of these features would be the “basis” of regression analysis.

The function year\_party\_basis (year, basis) will first merge the two CSV files on constituency number using pandas, then select the column to be used based on the feature set by the user input. It would also filter the rows based on the ‘year’ prompted by the user. It would return the data frame PARTY\_BASIS\_YEAR, with three columns, party, vote share percentage and one of the four demographic features selected by the user.

The three other functions are based on the three parties. They would filter for their respective parties, train, test, and conduct linear regression analysis using the ordinary least square OLS method through sklearn. Also, the Pearson coefficient r and p-value shall be returned to the main function under these three functions. 80% of the data is used for training and 20% for testing.

The main function shall print the summary statistics and scatter chart with linear regression for all three parties for comparative analysis. Also provided in the chart are details, such as the total average votes polled by each political party, the values of the r coefficient and the p-values.

These charts indicate a correlation between the demographic feature selected and the political parties. For example, in both elections, the charts clearly depict the decrease in BJP vote share as the rural votes in an assembly constituency increase and vice-versa. Similarly, SC votes increase also favours the Congress Party.

The anomaly in terms of a disproportionate jump in the number of seats compared to the increase in the percentage of Votes can be attributed to the presence of the AAP party in the 2022 elections. In rural areas, if we compare 2022 with 2017, we can see the impact of the AAP party in favour of the BJP.