Primary_And_Secondary_Analysis

Data Cleaning

1. Data might contain constituencies spelling mismatches and some constituencies may be listed with identical names. Proper validation is required.
2. In 2014, Andhra Pradesh underwent bifurcation. For simplicity, all constituencies from that year should be attributed to Telangana state. This includes constituencies such as Adilabad, Hyderabad, Warangal, etc., which should be considered part of Telangana rather than Andhra Pradesh for the year 2014.

Data Transformation

1. Establish suitable dimension tables and primary keys to link the CSV files effectively.
2. Construct aggregate tables using append and groupby functions as needed to format the data appropriately for answering the queries.

Questions from the available data (Primary)

1. List top 5 / bottom 5 constituencies of 2014 and 2019 in terms of voter turnout ratio?
2. List top 5 / bottom 5 states of 2014 and 2019 in terms of voter turnout ratio?
3. Which constituencies have elected the same party for two consecutive elections, rank them by % of votes to that winning party in 2019
4. Which constituencies have voted for different parties in two elections (list top 10 based on difference (2019-2014) in winner vote percentage in two elections)
5. Top 5 candidates based on margin difference with runners in 2014 and 2019.
6. % Split of votes of parties between 2014 vs 2019 at national level
7. % Split of votes of parties between 2014 vs 2019 at state level.
8. List top 5 constituencies for two major national parties where they have gained vote share in 2019 as compared to 2014.
9. List top 5 constituencies for two major national parties where they have lost vote share in 2019 as compared to 2014.
10. Which constituency has voted the most for NOTA?

11. Which constituencies have elected candidates whose party has less than 10% vote share at state level in 2019?

———

Further analysis & recommendations: (This will require additional research use secondary data)

1. Is there a correlation between postal votes % and voter turnout %?
2. Is there any correlation between GDP of a state and voter turnout %??
3. Is there any correlation between literacy % of a state and voter turnout %?
4. Provide 3 recommendations on what the election commission /government can do to increase the voter turnout %?

———

Note: The above questions are intended to jump-start your analysis. You are encouraged to dig deeper and broaden the scope of your analysis to discover even more valuable findings.