HERE IS THE LINK FOR Q2 (place cursor on file icon) : [Q2.TDS](https://drive.google.com/drive/folders/1NdHx6uT7mx0SyqZiK-JNK0Zo7gxclmEU?usp=sharing)

**Data Cleanup Strategy**

**1. Handling Missing Data:**

- Replace empty strings with NaN values for consistency

- Remove rows where 'VOTES' is "None" to ensure data integrity

**2. Assembly Constituency (AC) Name Cleanup:**

- Strip leading numbers and spaces (e.g., "01 Constituency" becomes "Constituency")

- Remove non-alphabetic characters, except spaces and periods, to standardize names

**3. Candidate Name Cleanup:**

- Similar to AC names, remove leading numbers and spaces

- Eliminate non-alphabetic characters (excluding spaces and periods) for consistency

**4. Whitespace Management:**

- Trim excess whitespace from all string columns to improve data quality

**5. Duplicate Entry Removal:**

- Perform initial removal of all duplicate rows

- Conduct targeted de-duplication based on key fields (ST\_NAME, YEAR, AC, CANDIDATE, SEX, AGE, CATEGORY, PARTY, VOTES) to maintain unique election entries

**6. Data Type Standardization:**

- Treat all data as strings during the cleaning process for consistent handling

- Convert specific columns to appropriate data types (e.g., YEAR to integer, VOTES to float) for analysis

**7. Data Saving Approach:**

- Save complete state data to a CSV file

- Filter and save data for specific Assembly Constituencies separately for focused analysis

**8. Error Management:**

- Implement checks to verify state data existence before proceeding with scraping and cleaning

- Provide appropriate feedback if data is unavailable for a given state

*This strategy aims to clean and standardize election data, ensuring consistency and reliability for subsequent analysis. By addressing issues such as missing data, inconsistent naming conventions, and duplicates, we create a robust dataset suitable for in-depth electoral studies.*

*The approach balances thoroughness with efficiency, tackling common data quality issues while preserving the integrity of the original information. This cleaned dataset will serve as a solid* foundation for exploring voting patterns, candidate performance, and broader electoral trends in Assam.