Contents

[1. Removing Candidates with Blank Voting Information](#_Toc171935630)

[2. Verification of NaN/Null/None Values](#_Toc171935631)

[3. Inclusion of NOTA as a Valid Candidate](#_Toc171935632)

[4. Removing Duplicates](#_Toc171935633)

[Conclusion](#_Toc171935634)

## 1. Removing Candidates with Blank Voting Information

**Steps Followed**:

1. **Selected Filter Option**: Applied a filter to the dataset.
2. **Applied Filter on VOTES Column**: Set the filter criteria to identify rows where the VOTES column is blank.
3. **Selected the Filtered Records**: Highlighted the records that met the blank criteria.
4. **Deleted**: Removed the 17 candidates with blank voting information from the dataset.

## 2. Verification of NaN/Null/None Values

* **Observation**: No candidates had NaN, null, or None values in the VOTES column. This was verified before proceeding with further data cleanup.

## 3. Inclusion of NOTA as a Valid Candidate

* **Consideration**: NOTA (None of the Above) votes are treated as valid entries in the dataset, ensuring that these votes are considered in the analysis.

## 4. Removing Duplicates

**Steps Followed**:

1. **Selected Columns for Duplicate Check**: Focused on the following columns, excluding 'Unnamed: 0' due to its unique values:
   1. ST\_NAME (State Name)
   2. YEAR (Election Year)
   3. AC (Assembly Constituency)
   4. CANDIDATE (Candidate Name)
   5. SEX (Gender)
   6. AGE (Age)
   7. CATEGORY (Category)
   8. PARTY (Political Party)
   9. VOTES (Vote Count)
2. **Selected Data Tab**: Navigated to the Data tab in the data processing tool.
3. **Selected 'Remove Duplicates' Ribbon**: Used the 'Remove Duplicates' function to eliminate 146 duplicate records from the dataset.

## Conclusion

The data cleanup strategy ensures that the dataset is accurate, complete, and free of redundancies. By removing records with blank voting information, verifying the absence of NaN/null/None values, including NOTA as a valid candidate, and eliminating duplicate records, the dataset is prepared for robust analysis and reliable insights into election voting patterns.