Contents

[1. Removing Candidates with Blank Voting Information](#_Toc171931663)

[2. Verification of NaN/Null/None Values](#_Toc171931664)

[3. Inclusion of NOTA as a Valid Candidate](#_Toc171931665)

[4. Removing Duplicates](#_Toc171931666)

[5. Retaining Only Relevant Columns](#_Toc171931667)

[6. Removing Prefix Numbers from Candidate Names](#_Toc171931668)

[7. Calculating the Length of Candidate Names](#_Toc171931669)

[Conclusion](#_Toc171931670)

## 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.

## 5. Retaining Only Relevant Columns

* **Action**: Kept only the columns for candidates and votes and deleted the remaining columns.

## 6. Removing Prefix Numbers from Candidate Names

**Steps Followed**:

1. **Created a New Column**: Added a new column next to the votes column to capture the candidate’s name without the prefix number.
2. **Applied Formula**: In cell B2, used the formula =MID(A2, FIND(" ", A2) + 1, LEN(A2)) to remove the prefix number and dragged it down to apply to all rows.

## 7. Calculating the Length of Candidate Names

**Steps Followed**:

1. **Created a New Column**: Added a new column next to the votes column to capture the length of the candidate’s name.
2. **Applied Formula**: In cell C2, used the formula =LEN(B2) to calculate the name length and dragged it down to apply to all rows.

## Conclusion

By following these data cleanup steps, the dataset is now prepared for analysis to determine the correlation between the length of a candidate's name and the votes they received. This ensures the data is accurate, consistent, and ready for further statistical analysis.