

Documentation

Infosys Springboard Virtual Internship 6.0-Batch 11

Milestone-1

Presented By: Group 2- Team B

Team Members' Name: Arya Mishra,Kritika Balpande,Radha sri Yakkala

Domain: Data Visualization

Project Name : ElectViz - Election Data Visualization

Problem Statement:

The analysis and interpretation of election data pose significant challenges due to the large volume, complexity, and multidimensional structure of electoral datasets. Election data consists of numerous attributes such as constituencies, candidates, political parties, vote counts, victory margins, and regional classifications. When this data is presented in raw or tabular formats, it becomes difficult to derive meaningful insights, limiting its usefulness for accurate and data-driven political reporting. Media organizations and analysts often require timely, clear, and visually intuitive representations of such data to communicate election outcomes effectively to the public.

The problem addressed by the ElectViz project is the absence of an interactive and visual analytical system that can simplify complex election datasets while maintaining accuracy and depth of analysis. Using Power BI, the project aims to transform raw election data into dynamic dashboards that highlight key indicators such as voting trends, party performance, and regional election dynamics. The solution involves data cleaning, preprocessing, and the identification of relevant metrics to ensure reliable analysis. By leveraging Power BI's visualization and interactivity features, the project enables users to explore election data efficiently, supporting transparent, fact-based interpretation and enhanced political reporting.

About the Dataset:

This dataset contains detailed **candidate-level election data** for India's national and state legislative elections, specifically the **Lok Sabha** elections. It covers election years from **1977 to 2015**, with each record representing an individual

candidate contesting in a specific constituency during a given election year. The dataset enables comprehensive analysis of voting patterns, party performance, and constituency-level dynamics across different states and years. This dataset was collected from kaggle containing 73082 rows and 11 columns.

Link: <https://www.kaggle.com/datasets/awadhi123/indian-election-dataset>

Dataset Columns and Description

Column Name	Description
st_name	Name of the state
Year	General election year
pc_no	Parliamentary constituency number
pc_name	Name of the parliamentary constituency
pc_type	Reservation status of the constituency
cand_name	Name of the contesting candidate
cand_sex	Gender of the candidate
partyname	Name of the political party
partyabbre	Abbreviation of the political party
totvotpoll	Total votes received by the candidate
electors	Total number of registered voters in constituency

Data Cleaning and Preprocessing:

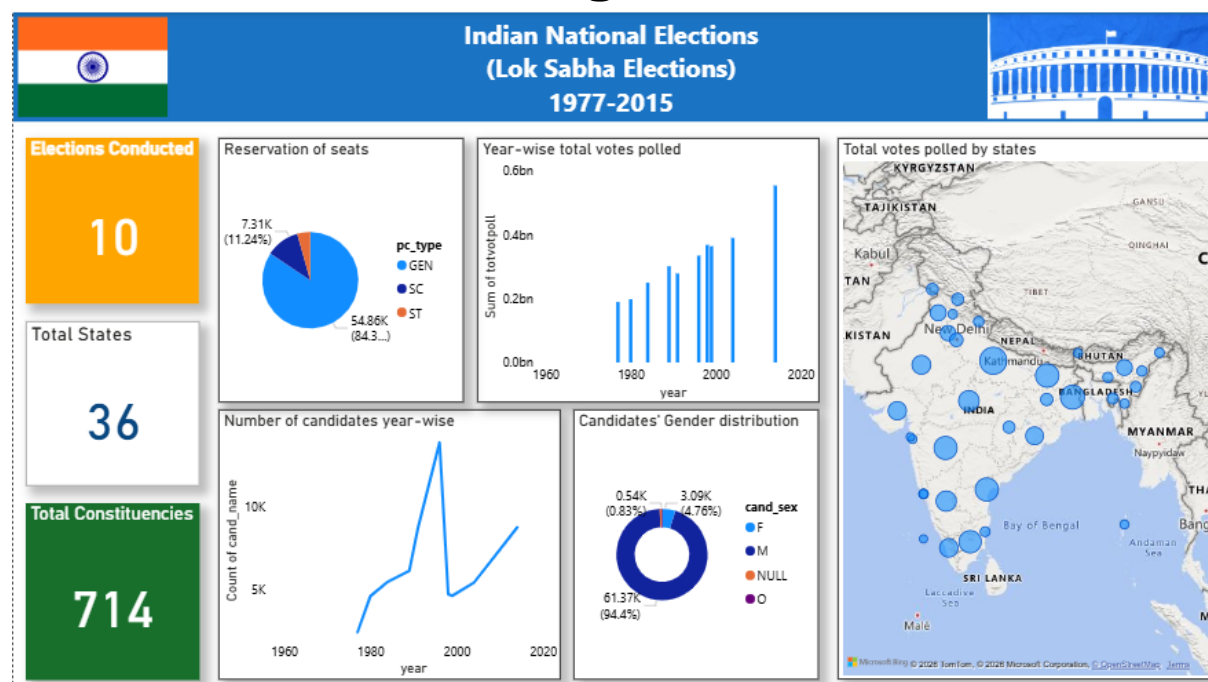
The first step is to clean the dataset and perform early preprocessing so the data is uniform and dashboard made and correct and insightful. Steps performed are:

1. Checking the data types of each column; all data types were correctly detected by PowerBI data loader
2. Checking for null /empty values ; 1% empty values were found in 'pc_type' column and those entries were removed by using the dropdown filter in the particular column
3. Checking for duplicates; selected all columns and entries and then used remove duplicate rows feature in the power query editor

4. Replaced mismatching or different spelling values in 'st_name' column like Chhattisgarh and Chattisgarh and similarly 6 values
5. Replaced mismatching or different spelling values in 'party_name' column like BJP and Bhartiya Janta Party and similarly 9 values. Got 45 distinct values reduced to 36 distincts.
6. Removed column 'pc_no' as it was just an identification to constituencies
7. Added a custom column 'voters_turnout_%' to calculate percentage of totvotpoll over electors
8. Changed the type of this column to percentage type

Milestone-2

1. Election Overview Page



1.1 Visuals Used and Their Purpose

- **Card Visual – Count of Years**
 - Displays the total number of years across which the election data varies.
 - Used to give users an immediate understanding of the temporal span of the dataset.
 - Cards are ideal for showing single, high-level KPIs clearly and prominently.
- **Card Visual – Count of States**
 - Represents the total number of states included in the dataset.

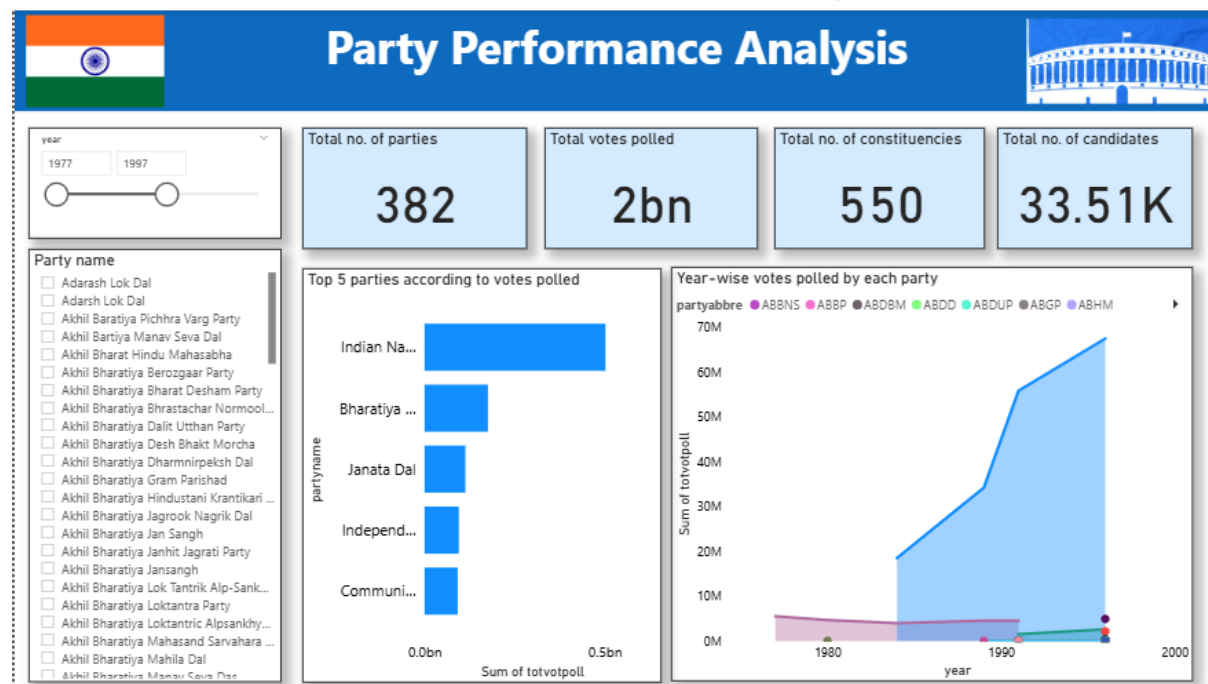
- Helps users quickly understand the geographical coverage of the election data.
- Cards are used here for simplicity and instant readability.
- **Card Visual – Count of Constituencies**
 - Shows the total number of constituencies present in the dataset.
 - Provides insight into the scale and granularity of the election data.
 - Cards are effective for highlighting key summary statistics.
- **Pie Chart – Distribution of Election Types (General, SC, ST)**
 - Illustrates the proportion of different election types.
 - Pie charts are suitable for showing percentage-based distribution of categories.
 - Helps users visually compare the share of General, SC, and ST constituencies.
- **Column Chart – Total Votes Polled by Year**
 - Displays the sum of total votes polled for each election year.
 - Column charts are effective for comparing numerical values across discrete time periods.
 - Helps identify years with higher or lower voter participation.
- **Line Chart – Count of Candidates per Year**
 - Shows how the number of candidates has changed over time.
 - Line charts are ideal for trend analysis across years.
 - Helps observe growth or decline in political participation.
- **Donut Chart – Gender Distribution of Candidates**
 - Represents the ratio of male, female, and other candidates.
 - Donut charts provide a clearer center-focused view of category proportions.
 - Highlights gender diversity in elections in an intuitive manner.

1.2 Insights Generated from the Election Overview Page

- The dataset spans multiple election years, providing a comprehensive historical perspective.
- A wide number of states and constituencies are covered, indicating a large-scale dataset.
- General elections form the majority, while SC and ST elections represent smaller but significant portions.

- Total votes polled vary across years, reflecting changes in voter turnout and political engagement.
- The number of candidates fluctuates over time, showing shifts in electoral competition.
- Gender distribution indicates a strong dominance of male candidates, with relatively lower female and other gender representation.

2. Party Performance Analysis Page



2.1 Visuals Used and Their Purpose

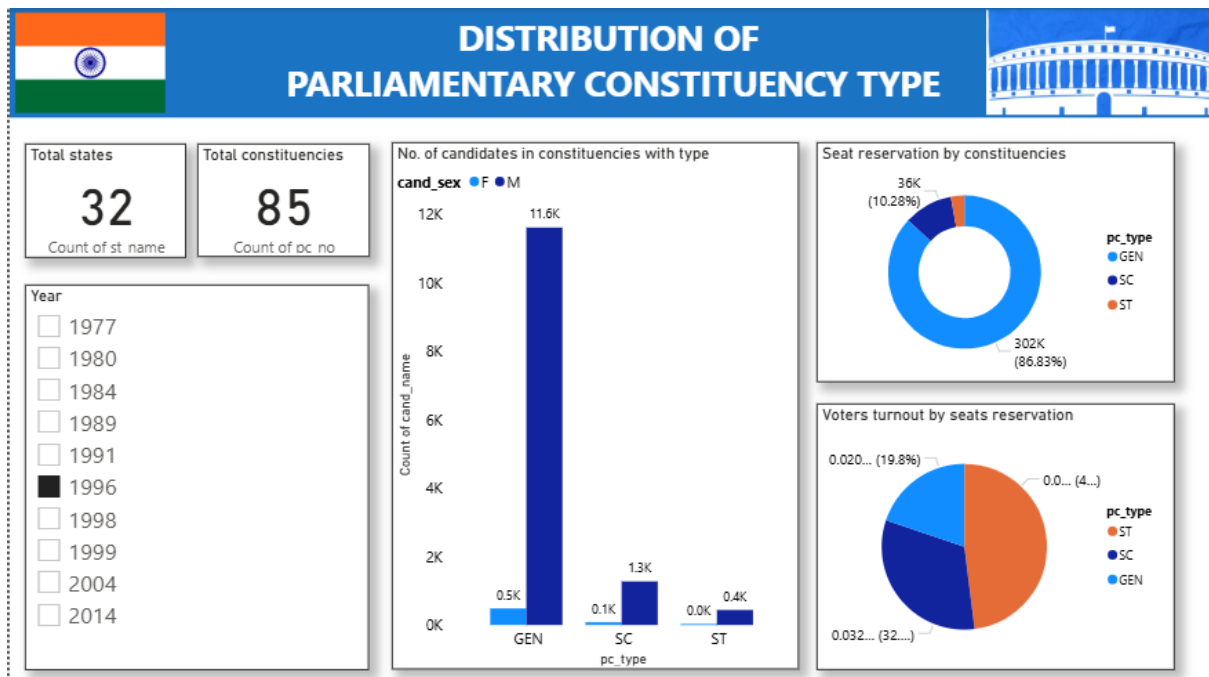
- **Year Slicer**
 - Enables users to filter party performance for a specific election year.
 - Helps analyze temporal changes in party influence.
- **Party Name Slicer**
 - Allows selection of one or multiple political parties.
 - Supports detailed party-wise analysis.
- **Card Visual – Count of Parties**
 - Displays the number of parties participating in the selected context.
 - Helps assess political diversity in elections.

- **Card Visual – Total Votes Polled**
 - Shows total votes secured by the selected party or parties.
 - Acts as a key indicator of party popularity.
- **Card Visual – Constituencies Covered**
 - Displays the number of constituencies contested by the party.
 - Helps evaluate the geographical reach of a party.
- **Card Visual – Number of Candidates**
 - Represents the total candidates fielded by the party.
 - Indicates the party's level of participation.
- **Bar Chart – Top 5 Parties by Total Votes Polled**
 - Highlights the five most dominant parties based on vote share.
 - Bar charts are ideal for ranking and comparison.
 - Helps quickly identify leading political parties.
- **Line Chart – Trend of Total Votes Polled by Year**
 - Shows how a party's vote count has changed over time.
 - Useful for analyzing growth, decline, or stability in party performance.
 - Supports long-term performance evaluation.

2.2 Insights Generated from the Party Performance Analysis Page

- A small number of parties dominate the total votes polled.
 - Party performance varies significantly across different years.
 - Some parties show consistent growth, while others experience decline.
 - Larger parties contest more constituencies and field more candidates.
 - Year-wise and party-wise filtering allows detailed comparative analysis of political trends.
-

3. Distribution of Parliamentary Constituency Type



3.1 Visuals Used and Their Purpose

- **Year slicer:**
 - Allows users to filter the entire dashboard by election year.
- **Card Visual - Total no of state**
 - Represent the total no of states present in the dataset.
- **Card Visual - Total no of parliamentary constituency number**
 - Represent the total no of parliamentary constituency numbers present in the dataset.
- **Donut Chart – Distribution of Parliamentary Constituency Types**
 - This chart displays the distribution of constituencies by reservation type (GEN, SC, ST).
 - It highlights that General constituencies dominate, while SC and ST constituencies form a smaller but constitutionally significant share.
- **Clustered Column Chart – Candidate Participation by Constituency Type & Gender**
 - This chart compares the number of candidates by gender across different constituency types.
 - It shows strong male dominance in candidate participation across all constituency types, with relatively lower female representation, especially in SC and ST constituencies.
- **Pie Chart – Average Voter Turnout by Constituency Type**
 - This chart presents the average voter turnout percentage across constituency types.

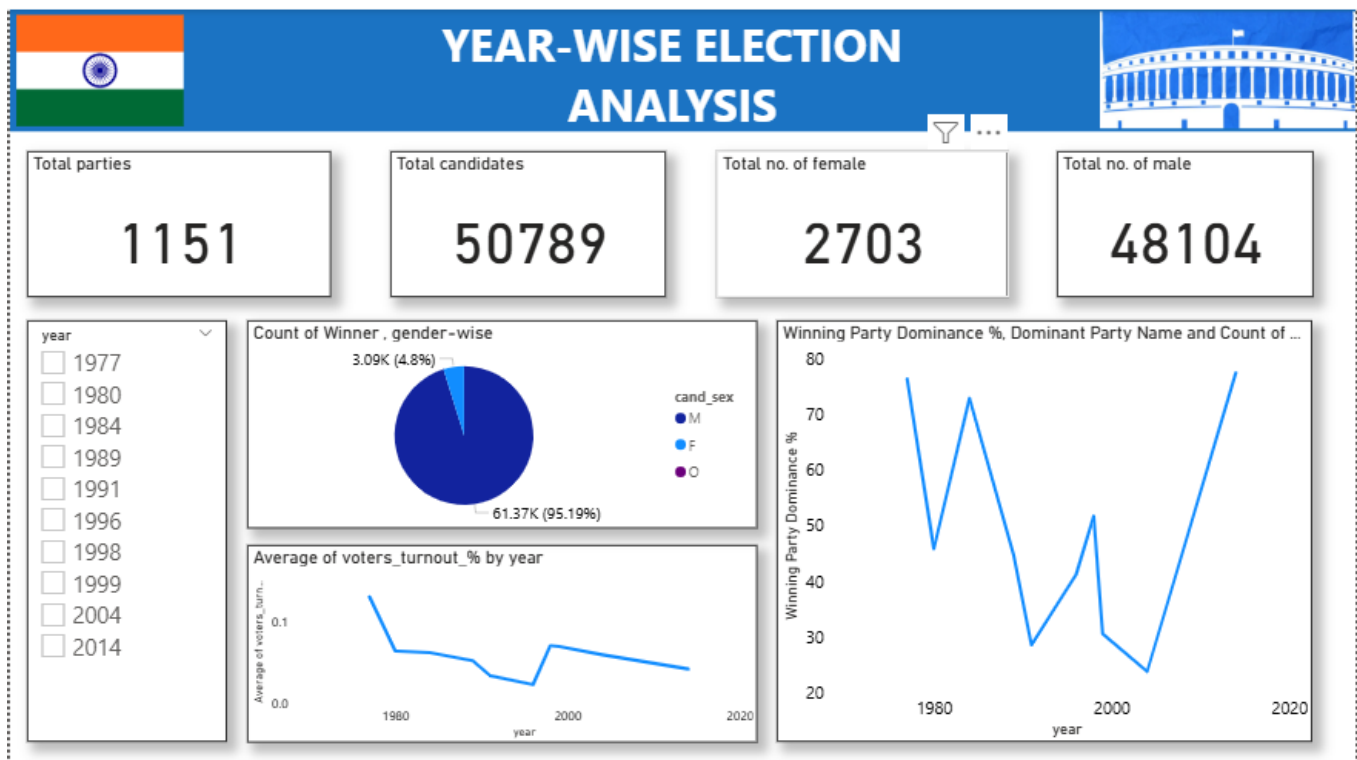
- It indicates that ST constituencies generally record higher turnout, followed by SC and General constituencies, reflecting varying voter engagement levels.

3.2 Insights generated from distribution of parliamentary constituency type page

- There is a higher share of average turnout% in ST and SC compared to General .
- The share of average turnout % keeps on increasing for all Parliamentary constituency types year by year.
- There is high dominance of General compared to SC , ST but shows significant increase year wise.
- Though there is male dominance in every participation type, female participation keeps on increasing year by year steadily.
- There is a decrease in no of candidates and average turnout % from 1996 to 1998.
- The period from 1996 to 1998 clearly shows instability in Indian politics .

Milestone-3

4. Year wise analysis of Total no of parties and Total no of candidate



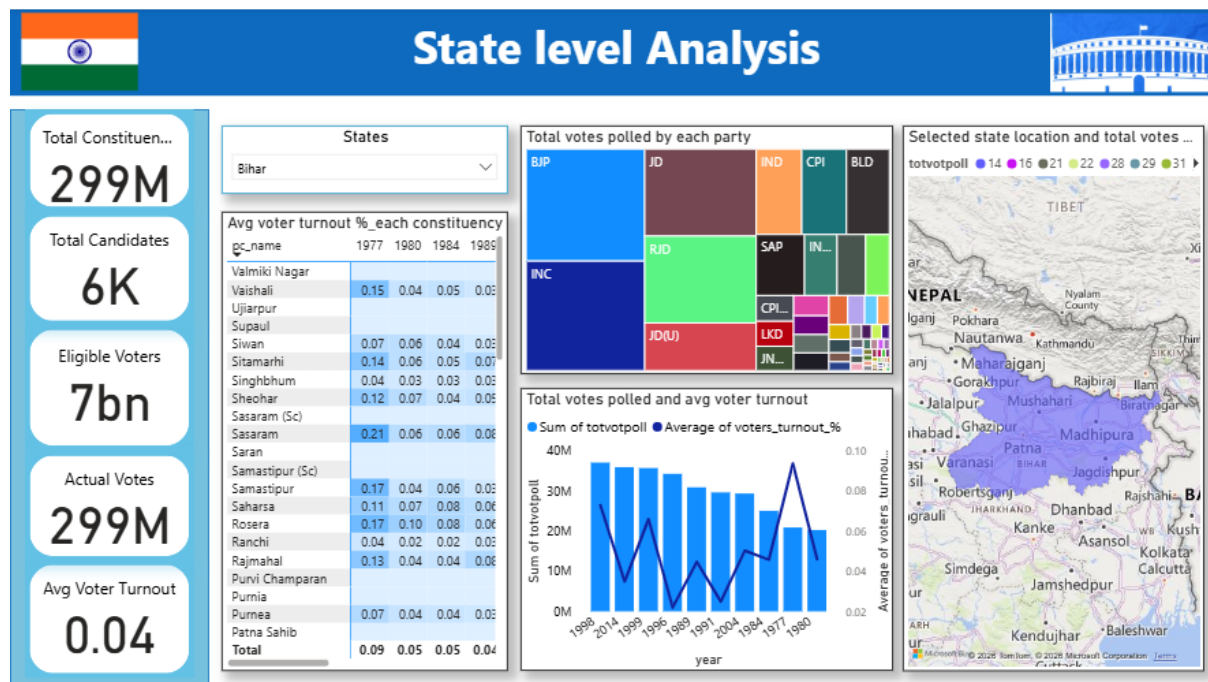
4.1 Visuals Used and Their Purpose

- **Year slicer:**
 - Allows users to filter the entire dashboard by election year
- **Card Visual Total no of parties**
 - Represent Total no of parties present in the dataset and varies year by year
- **Card Visual Total no of candidate**
 - Represent Total no of candidates present in the dataset and varies year by year
- **Card Visual Total no of female candidate**
 - Represent Total no of female candidate out of total present in the dataset and varies year by year
- **Card Visual Total no of male candidate**
 - Represent Total no of male candidate out of total present in the dataset and varies year by year
- **Gender Distribution of Winner Candidate (Pie Chart)**
 - Displays the proportion of male and female winning candidates.
 - Highlights gender imbalance in election outcomes.
- **Average Turnout Percentage by Year (Line Chart)**
 - Shows trends in voter participation across election years.
 - Helps analyze changes in public engagement over time.
- **Winning Party Dominance% (Line Chart):**
 - This chart shows the Winning Party Dominance Index, which measures the percentage of constituencies won by the top party in each election year.
 - The tooltip/card highlights which party held that dominance, allowing us to clearly distinguish single-party dominance from coalition eras.

4.2 Insights generated from year wise analysis page

- Higher values indicate a dominant single-party system, while lower values reflect coalition-based politics .
 - When observed alongside the number of participating parties, it clearly highlights the transition from dominance to coalition eras.
 - There was the highest turnout % in 1977 then it kept on decreasing and lowest in 1996 then it increased from 1998 and then became stable in 2014.
 - There are more male candidate winners but female candidate winners also keep on increasing but showing decrement in 1996-1998.
-

5. State Level Analysis page



5.1 Visuals Used and Their Purpose

- Card Visuals – Key Performance Indicators (KPIs)**
 - Displays high-level metrics: Total Constituencies (55), Eligible Voters (6bn), Avg Voter Turnout (5.87%), Total Candidates (4K), and Actual Votes (290M).
 - Used to provide an immediate summary of the state's electoral scale at a glance.
- Shape Map – Geographical State Selection**
 - An interactive map of India allowing users to filter the entire dashboard by selecting a specific state (e.g., Andhra Pradesh).
 - Helps users visualize the geographical context of the data being analyzed.
 - Selected regions are highlighted to show the physical distribution of constituencies.
- Treemap – Sum of Votes by Party**
 - Illustrates the distribution of the total vote poll across different political parties.

Total Votes by Party =
 $SUM('indian-national-level-election'[totvotpoll])$

 - The size of each rectangle represents the party's vote share, making it easy to identify dominant parties like INC and TDP.
 - Useful for comparing the relative strength of multiple parties simultaneously.
- Combination Chart (Line and Stacked Column) – Trends by Year**

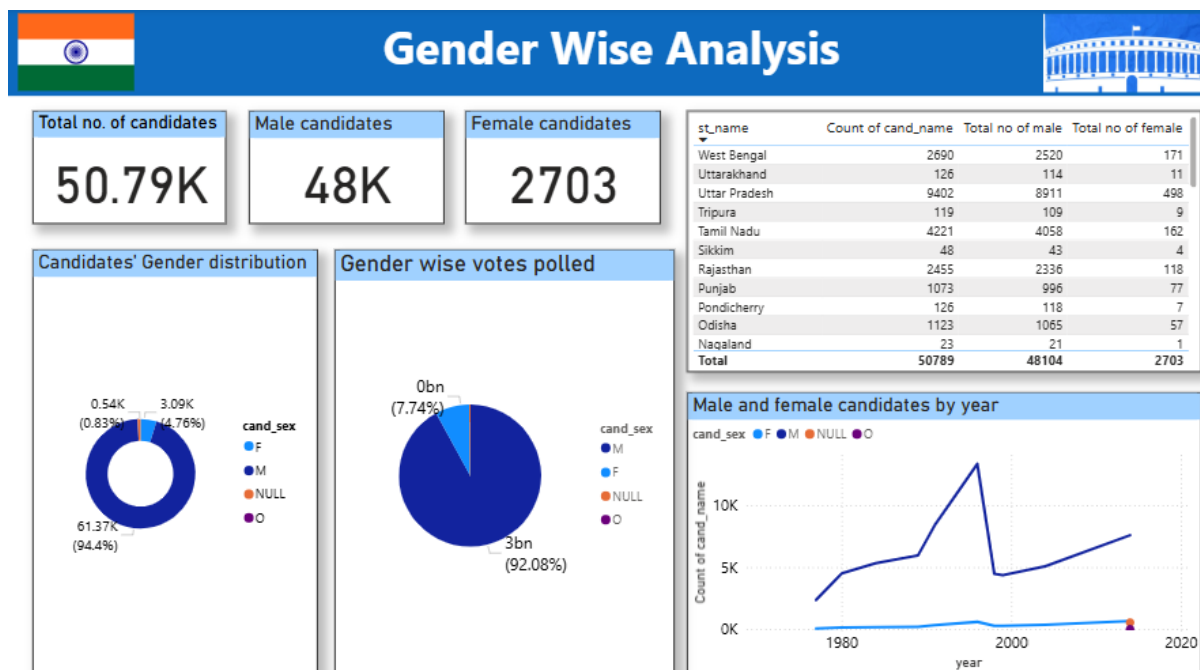
- Displays the Sum of totvotpoll as bars and the Average of voters_turnout_% as a line over different election years.
- Ideal for identifying correlations between the volume of votes and the percentage of voter participation over time.
- Dual axes allow for the comparison of two different scales (millions of votes vs. percentage) in one visual.
- **Matrix Table – Constituency Wise Voter Turnout**
 - Provides a detailed breakdown of voter turnout percentages for each constituency (e.g., Adilabad, Amalapuram) across various years.
 - Allows for a deep dive into localized performance trends and historical comparisons.
 - The "Total" row at the bottom provides a state-wide average for each election cycle.

5.2 Insights Generated from the State Level Analysis Page

- The dashboard covers a significant electoral span, showing data for 55 constituencies with a massive voter base of 6 billion eligible individuals.
 - Historical trends show that voter turnout percentage is not always directly proportional to the total volume of votes, indicating varying levels of mobilization.
 - The Treemap reveals a competitive landscape where two or three major parties typically capture the vast majority of the total votes polled.
 - Specific constituencies show consistent turnout patterns, while others exhibit high volatility across different election years.
 - The interactive map and slicers allow for seamless navigation between different states, proving that the platform is scalable for national-level analysis.
-

Milestone-4

6. Gender Wise Analysis Page



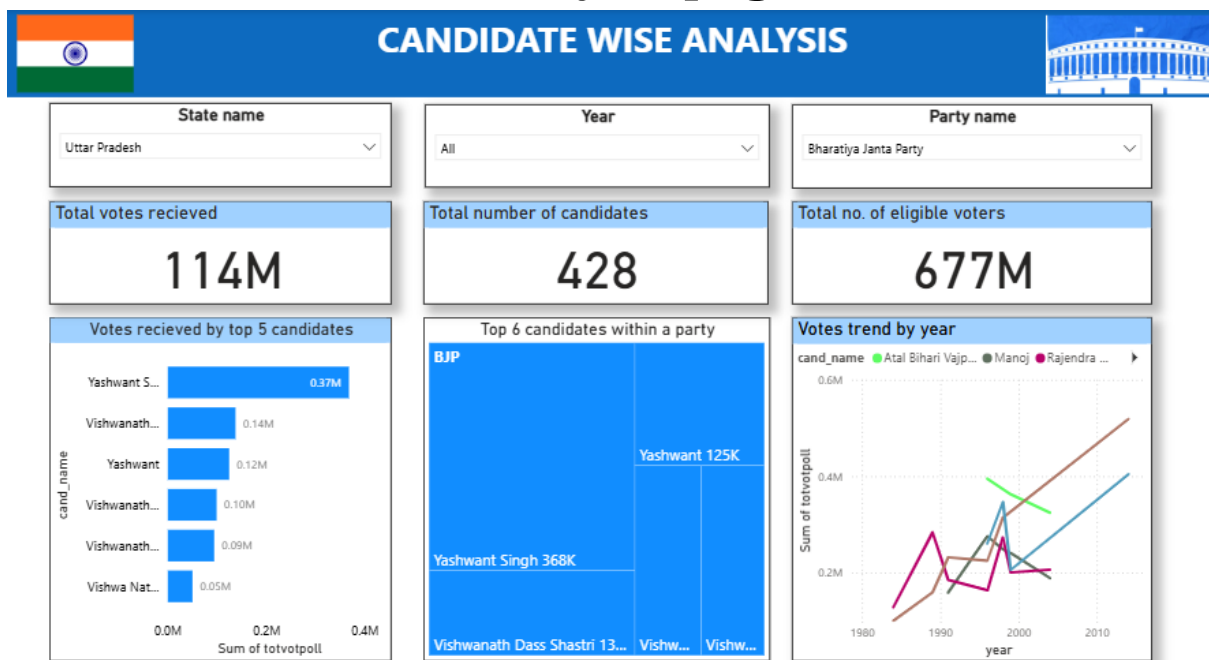
6.1 Visuals Used and Their Purpose

- **KPI Cards – Total, Male, and Female Candidates**
Display the total number of candidates along with a breakdown of male and female candidates, providing a quick overview of gender representation in elections.
- **Donut Chart – Candidates' Gender Distribution**
Represents the percentage distribution of candidates based on gender, highlighting the proportion of male and female participation.
- **Pie Chart – Gender-wise Votes Polled**
Represents the share of total votes polled by male and female candidates, reflecting voter support distribution across genders.
- **Table – State-wise Gender Representation**
Displays state-wise statistics including total candidates, male candidates, and female candidates, enabling comparison of gender participation across different states.
- **Line Chart – Male and Female Candidates by Year**
Represents the trend of male and female candidate participation over different election years, showing changes in gender representation over time.

6.2 Insights Generated from the Gender Wise Analysis Page

- Male candidates form a dominant majority across all election years, indicating significant gender imbalance in electoral participation.
- Female representation remains comparatively low, though a gradual increase in female candidate participation can be observed over time.
- The distribution of votes polled closely follows candidate participation patterns, with male candidates receiving a larger share of total votes.
- State-wise analysis reveals variation in female participation, with certain states contributing more female candidates than others.
- Year-wise trends suggest slow but positive progress toward increased gender inclusivity in the electoral process.

7. Candidate Level Analysis page



7.1 Visuals Used and Their Purpose

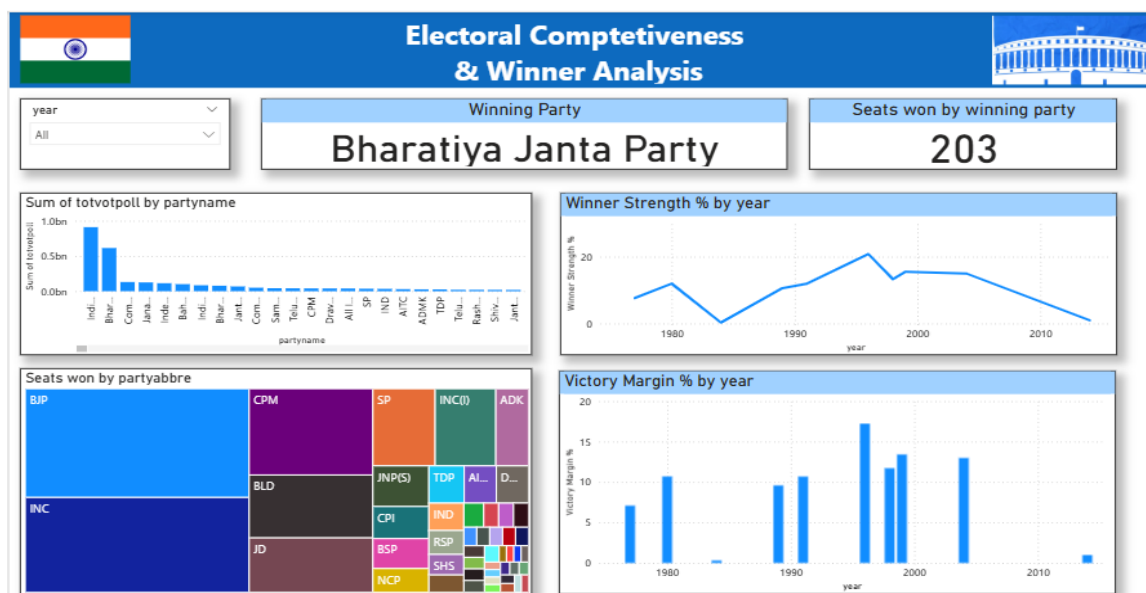
- **Slicer Visual – State, Year, and Party Name**
 - Allows users to drill down into specific election cycles and political organizations.
 - Essential for narrowing the dataset from millions of records to a specific, comparable group.

- **Card Visual – Total Votes Polled, Total Candidates, and Actual Voters (%)**
 - Provides an immediate snapshot of the scale of the filtered data.
 - The "Actual Voters" card (25.20%) offers a quick look at engagement for that specific segment.
- **Horizontal Bar Chart – Votes for Each Candidate**
 - Compares individual performance side-by-side using the **sum of votes received**
 - Effective for identifying top performers and the "vote gap" between rivals.
- **Treemap – Total Votes for Each Party and Candidate**
 - Visualizes the proportional "weight" of each candidate's contribution to the total party vote.
 - Helps in understanding the distribution of influence within a single party.
- **Scatter Plot – Total Votes for Each Candidate According to Year**
 - Maps candidate performance over time to identify outliers or performance clusters.
 - Useful for observing if vote shares are consolidating or fragmenting.

7.2 Insights Generated from the Candidate Level Analysis Page

- Specific years and parties show high vote concentration among a few top candidates, while others show a more even spread.
- Voter participation percentages vary significantly when filtered by party, suggesting different levels of base mobilization.
- The bar chart highlights that even within the same party, there is a substantial variance in individual candidate pull.

8. Electoral Competitiveness page



8.1 Visuals Used and their purpose

- **Slicer Visual – Year**
 - Allows users to filter the entire dashboard using slicer.
- **Card - Winner Party**
 - Represent winner party across each year.
- **Column Chart – Victory Margin %**
 - Represent How big was the win compared to all voters who participated?
 - Small % → highly competitive election
 - Large % → one-sided contest
- **Line Chart – Winner Strength %**
 - Represent by what fraction did the winner outperform the runner-up?
 - Small % → narrow win
 - Large % → dominant win
- **Treemap – Seats Won By Party**
 - The chart represent Total number of parliamentary seats won by each party.
 - It shows rise and fall of political parties.
- **Column Chart - Total Constituency Votes received by party**
 - This chart represent Cumulative votes received by each political party across all constituencies and years.

8.2 Insights Generated from the Electoral competitive page

- Low victory margins across years indicate that most elections are closely contested rather than one-sided.
 - Winner strength varies significantly by year, showing changing levels of dominance and electoral competitiveness.
 - National parties receive higher cumulative votes, reflecting wider voter reach compared to regional parties.
 - A few major parties dominate seat share, while regional parties maintain influence in specific constituencies.
-

Measures/Formulas Used in overall project

1.Voter Turnout %

Voter Turnout % =
DIVIDE(
 SUM('indian-national-level-election'[totvotpoll]),
 SUM('indian-national-level-election'[electors])
*) * 100*

2.Male Candidates

Male Candidates =
CALCULATE(
 DISTINCTCOUNT('indian-national-level-election'[cand_name]),
 'indian-national-level-election'[cand_sex] = "M"
)

3.Female Candidates

Female Candidates =
CALCULATE(
 DISTINCTCOUNT('indian-national-level-election'[cand_name]),
 'indian-national-level-election'[cand_sex] = "F"
)

4.Winner Votes

Winner Votes =
VAR TopCandidate =
 TOPN(
 1,
 'indian-national-level-election',
 'indian-national-level-election'[totvotpoll],
 DESC
)
RETURN
 MAXX(TopCandidate, 'indian-national-level-election'[totvotpoll])

5.Runner –Up Votes

RunnerUp Votes =


```

VAR TopTwo =
  TOPN(
    2,
    'indian-national-level-election',
    'indian-national-level-election'[totvotpoll],
    DESC
  )
RETURN
  MINX(TopTwo, 'indian-national-level-election'[totvotpoll])

```

6.Victory Margin

Victory Margin Votes =
 $[Winner\ Votes] - [RunnerUp\ Votes]$

7.Victory Margin %

Victory Margin % =
 $DIVIDE([Victory\ Margin\ Votes], [Winner\ Votes]) * 100$

8.Winner Strength %

Winner Strength % =
 $DIVIDE([Victory\ Margin\ Votes], [RunnerUp\ Votes]) * 100$

9.Seats Won

```

Seats Won =
COUNTROWS(
  FILTER(
    VALUES('indian-national-level-election'[pc_name]),
    VAR MaxVotesInPC =
      CALCULATE(
        MAX('indian-national-level-election'[totvotpoll])
      )
    RETURN
      CALCULATE(
        MAX('indian-national-level-election'[totvotpoll])
      ) = MaxVotesInPC
  )
)

```

10. Winning Party Name

Winning Party Name =

VAR TopParty =

```
    TOPN(  
        1,  
        VALUES('indian-national-level-election'[partyname]),  
        [Seats Won],  
        DESC  
    )
```

RETURN

```
    MAXX(TopParty, 'indian-national-level-election'[partyname])
```

11. Winning Party Seats

Winning Party Seats =

VAR TopParty =

```
    TOPN(  
        1,  
        VALUES('indian-national-level-election'[partyname]),  
        [Seats Won],  
        DESC  
    )
```

RETURN

```
    MAXX(TopParty, [Seats Won])
```

Conclusion

ElectViz successfully bridges the gap between raw electoral data and actionable public information by providing a centralized, automated visualization framework. By transforming complex datasets into intuitive, interactive dashboards featuring choropleth maps and swing-analysis charts, the platform overcomes the high latency and manual errors inherent in traditional reporting. The implementation of a robust data pipeline ensures that real-time insights enables newsrooms to identify critical "voter-shift" and "anti-incumbency" patterns instantly. Ultimately, this project offers a scalable, low-latency solution that enhances democratic transparency and provides broadcast-ready graphical insights for the digital age.