Documentation

Infosys Springboard Virtual Internship 6.0 – Batch 11

Presented By: Arya Mishra

(Group 2 – Team B)
Domain: Data Visualization
Project Title: *ElectViz – Election Data Visualization*

Milestone – 1

Problem Statement

Analyzing and interpreting election-related data is a complex task due to the massive size, diversity, and multi-dimensional nature of electoral datasets. Election data typically includes multiple attributes such as states, constituencies, candidates, political parties, vote counts, voter turnout, margins of victory, and reservation categories. When such extensive data is presented in raw tables or spreadsheets, it becomes difficult to extract actionable insights, thereby reducing its effectiveness for meaningful political analysis and reporting.

Media agencies, researchers, and analysts require clear, interactive, and visually rich representations of election data to communicate outcomes accurately and efficiently to the public. However, the lack of a comprehensive visual analytics system often restricts deeper exploration of voting trends, party dominance, and regional dynamics.

The ElectViz project addresses this challenge by developing an interactive election analysis platform using Power BI. The objective is to convert raw and complex election datasets into intuitive dashboards that preserve analytical depth while improving clarity and accessibility. The solution involves systematic data cleaning, preprocessing, and metric identification to ensure accuracy. By leveraging Power BI's advanced visualization and interactivity features, the project enables transparent, data-driven insights that support informed political reporting and analysis.

About the Dataset

The dataset used in this project contains candidate-level election data from India's parliamentary and state legislative elections, with a primary focus on Lok Sabha elections. It spans election years from 1977 to 2015, where each row

represents an individual candidate contesting from a specific constituency in a particular year.

This dataset supports in-depth analysis of electoral trends, constituency-wise performance, political party dominance, and voting behavior across states and election cycles. The data was sourced from Kaggle and consists of **73,082 rows and 11 columns**.

Dataset Source:
https://www.kaggle.com/datasets/awadhi123/indian-election-dataset

---

Dataset Columns and Description

Column Name	Description
st_name	Name of the state
Year	Year of the general election
pc_no	Parliamentary constituency number
pc_name	Name of the parliamentary constituency
pc_type	Reservation category of the constituency
cand_name	Name of the contesting candidate
cand_sex	Gender of the candidate
partyname	Political party name
partyabbre	Party abbreviation
totvotpoll	Total votes received by the candidate
electors	Total registered voters in the constituency

Data Cleaning and Preprocessing

To ensure consistency, accuracy, and meaningful visual outputs, initial data cleaning and preprocessing were carried out before dashboard development. The following steps were performed:

1. Verified data types for all columns; Power BI automatically detected correct data formats.
2. Identified missing or blank values; approximately 1% null entries were found in the `pc_type` column and were removed using column-level filters.
3. Checked for duplicate records by selecting all columns and applying the "Remove Duplicates" option in Power Query Editor.
4. Standardized inconsistent state names in the `st_name` column (e.g., *Chattisgarh* corrected to *Chhattisgarh*), resolving six such inconsistencies.
5. Unified inconsistent party names in the `partyname` column (e.g., *BJP* vs *Bhartiya Janta Party*), reducing distinct party entries from 45 to 36.
6. Removed the `pc_no` column as it served only as an internal identifier and was not required for analysis.
7. Created a calculated column named `voters_turnout_%` by computing the ratio of total votes polled to total electors.
8. Converted the new column's data type to percentage for better readability and analysis.

Milestone – 2

1. Election Overview Page

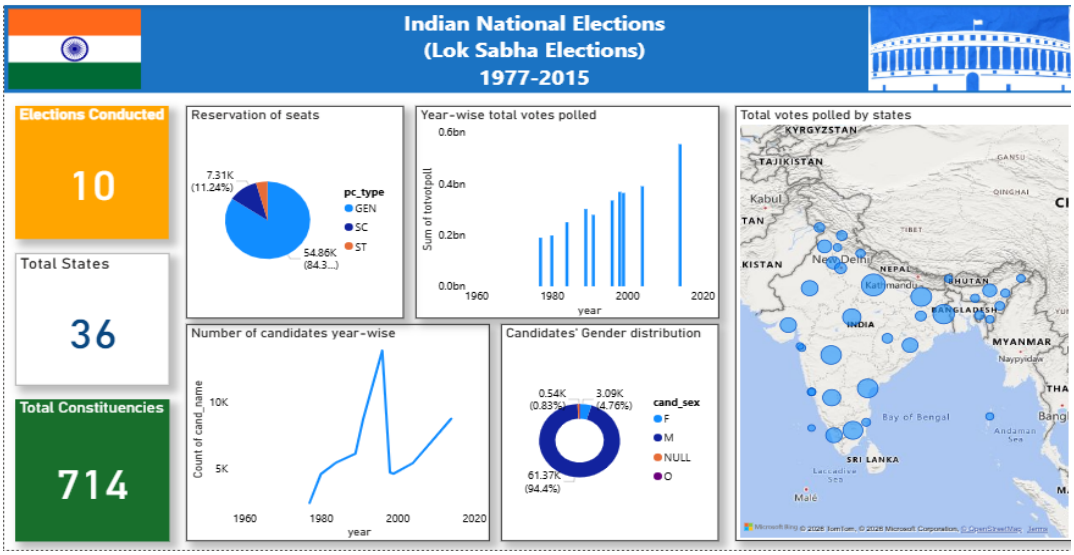
1.1 Visuals Used and Their Purpose

- **Card Visual – Total Years**
  Displays the number of election years covered in the dataset, giving a quick overview of the time span.
- **Card Visual – Number of States**
  Shows how many states are included, helping users understand the geographical coverage.

- **Card Visual – Number of Constituencies**
  Highlights the total constituencies represented, indicating the dataset's scale.
- **Pie Chart – Constituency Type Distribution (General, SC, ST)**
  Visualizes the proportional distribution of constituency reservation categories.
- **Column Chart – Total Votes Polled by Year**
  Compares total votes across election years to identify participation trends.
- **Line Chart – Candidate Count by Year**
  Displays how the number of candidates has evolved over time.
- **Donut Chart – Gender-wise Candidate Distribution**
  Shows the proportion of male, female, and other gender candidates.

1.2 Key Insights

- The dataset spans multiple decades, offering a strong historical perspective.
- A large number of states and constituencies are represented.
- General constituencies dominate, while SC and ST categories hold notable shares.
- Voter participation fluctuates significantly across election years.
- Candidate participation shows varying trends over time.
- Male candidates dominate elections, though female representation is gradually increasing.



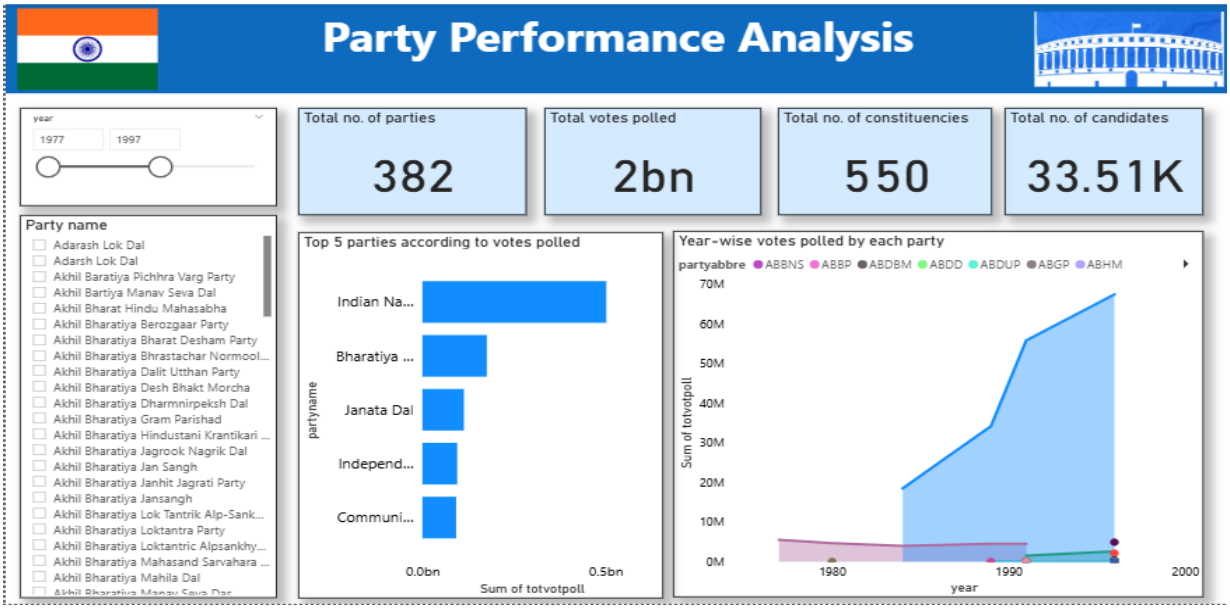
2. Party Performance Analysis Page

2.1 Visuals and Objectives

- **Year Slicer:** Enables year-wise filtering for comparative analysis.
- **Party Name Slicer:** Allows selection of individual or multiple parties.
- **Card – Number of Parties:** Indicates political diversity for the selected context.
- **Card – Total Votes Polled:** Reflects party popularity and performance.
- **Card – Constituencies Contested:** Shows party reach across regions.
- **Card – Candidate Count:** Measures participation intensity.
- **Bar Chart – Top 5 Parties by Votes:** Highlights dominant political parties.
- **Line Chart – Votes Trend Over Years:** Tracks performance evolution.

2.2 Insights

- A few major parties account for a large share of total votes.
- Party dominance varies across election cycles.
- Some parties demonstrate stable growth, while others decline.
- Larger parties contest more constituencies and field more candidates.
- Interactive filters support detailed comparative political analysis.



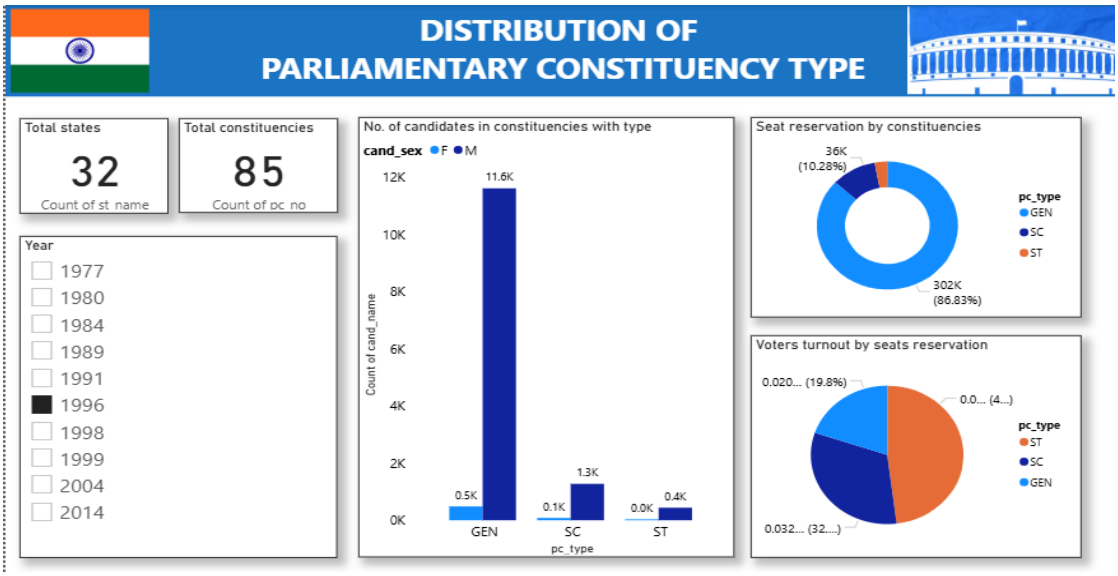
3. Parliamentary Constituency Type Distribution

3.1 Visuals and Purpose

- **Year Slicer:** Filters analysis by election year.
- **Card – Total States:** Displays number of states represented.
- **Card – Total Constituencies:** Shows total parliamentary constituencies.
- **Donut Chart – Constituency Type Share:** Highlights General, SC, and ST proportions.
- **Clustered Column Chart – Gender vs Constituency Type:** Compares candidate participation by gender.
- **Pie Chart – Average Voter Turnout:** Shows turnout variation by constituency type.

3.2 Insights

- SC and ST constituencies show higher average voter turnout than General.
- Turnout percentages increase steadily over the years.
- General constituencies dominate numerically but show year-wise growth.
- Female participation is increasing despite male dominance.
- Political instability is evident between 1996 and 1998.



Milestone – 3

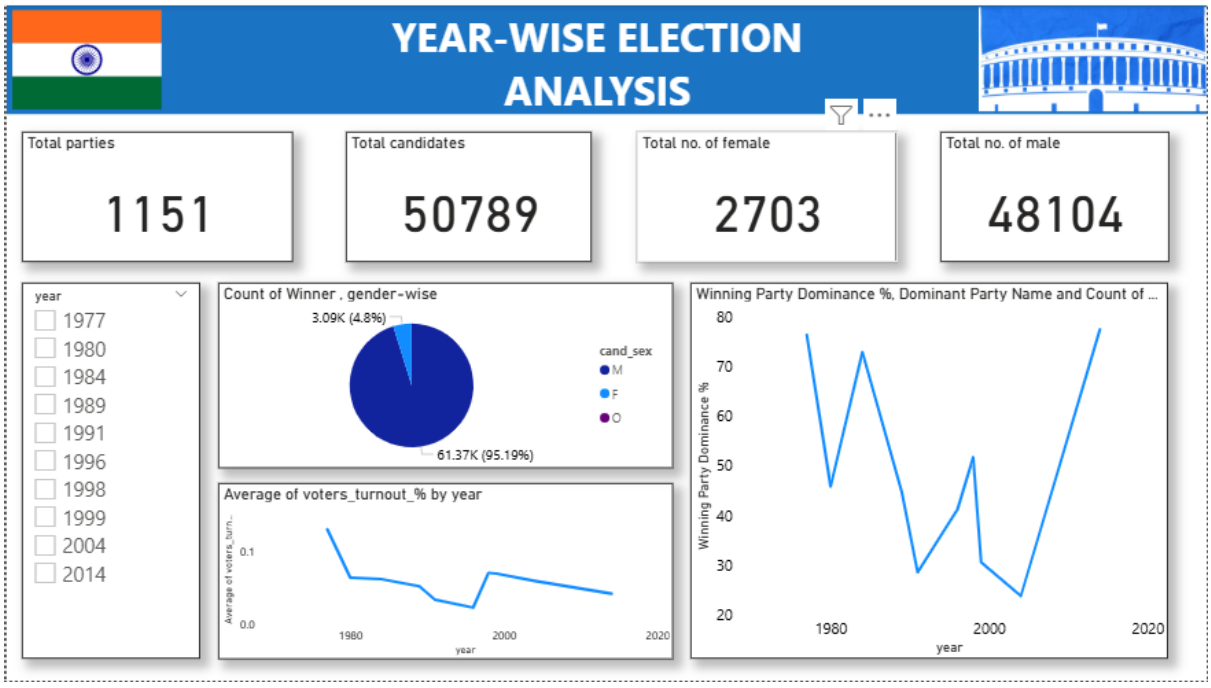
4. Year-wise Analysis of Parties and Candidates

4.1 Visuals Used

- **Year Slicer:** Enables year-specific filtering.
- **Cards – Total Parties & Candidates:** Show election scale changes.
- **Cards – Male & Female Candidates:** Highlight gender distribution.
- **Pie Chart – Winners by Gender:** Shows winning gender proportions.
- **Line Chart – Average Turnout:** Displays voter engagement trends.
- **Line Chart – Winning Party Dominance %:** Shows dominance vs coalition trends.

4.2 Insights

- Higher dominance values indicate single-party dominance.
- Lower dominance reflects coalition politics.
- Turnout peaked in 1977, dropped in 1996, and stabilized post-1998.
- Female winners are increasing, though fluctuations exist.



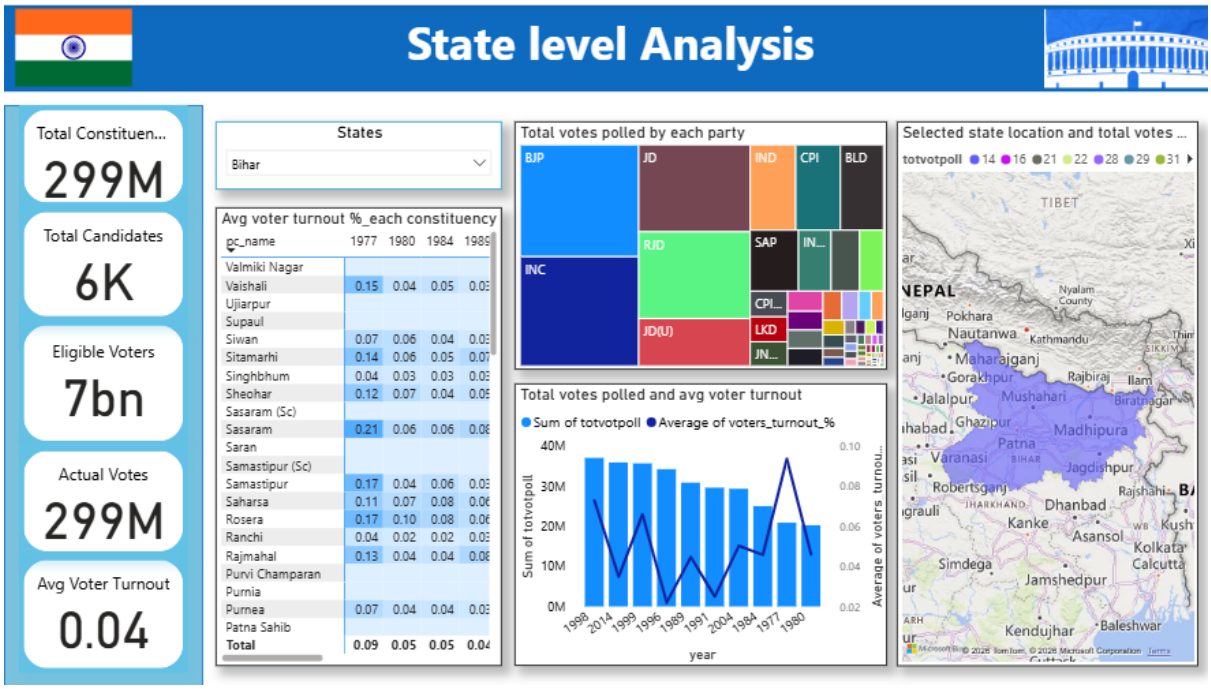
5. State-Level Analysis Page

5.1 Visuals and Purpose

- **KPI Cards:** Summarize constituencies, voters, turnout, candidates, and votes.
- **Shape Map:** Enables state-wise filtering and geographic context.
- **Treemap – Votes by Party:** Shows party dominance visually.
- **Combo Chart:** Compares votes and turnout trends together.
- **Matrix Table:** Displays constituency-wise turnout over years.

5.2 Insights

- Covers a wide electoral landscape with a massive voter base.
- Turnout does not always correlate with total votes.
- Two to three parties usually dominate vote share.
- Constituency turnout varies in stability.
- Dashboard supports scalable, nation-wide analysis.



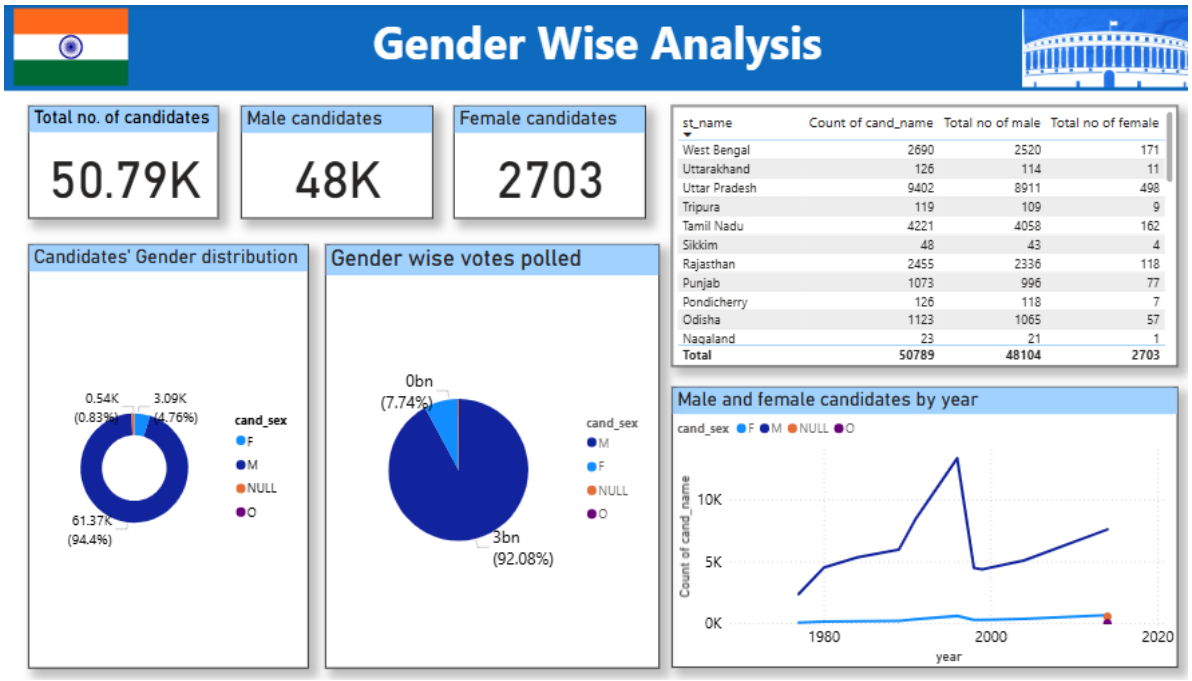
6. Gender Wise Analysis Page

6.1 Visuals and Purpose

- **KPI Cards:** Show total, male, and female candidates for quick gender overview.
- **Donut Chart:** Displays percentage distribution of candidates by gender.
- **Pie Chart:** Shows gender-wise share of total votes polled.
- **Table:** Compares state-wise male and female candidate participation.
- **Line Chart:** Tracks male vs female candidate trends over years.

6.2 Insights

- Male candidates dominate across all election years.
- Female participation is lower but shows gradual growth.
- Vote distribution largely mirrors candidate participation.
- Female representation varies significantly by state.
- Year-wise trends indicate slow improvement in gender inclusivity.



---

7. Candidate Level Analysis Page

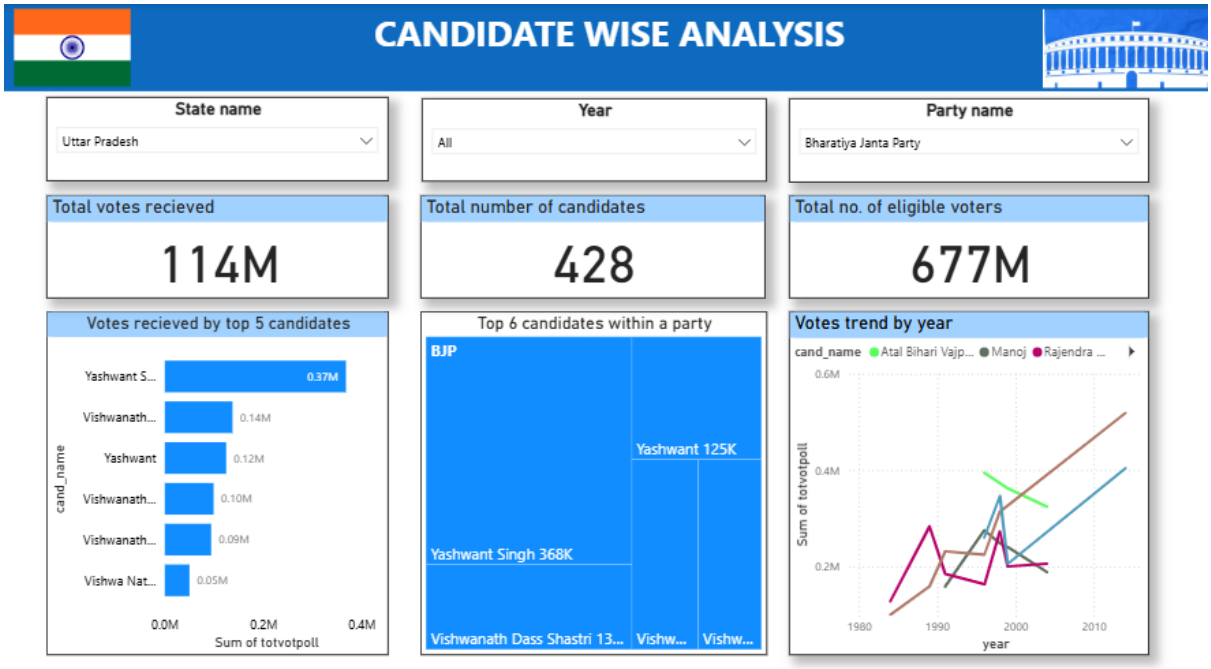
7.1 Visuals and Purpose

• **Slicers:** Filter data by state, year, and party.
• **Cards:** Show total votes, candidates, and voter participation %.
• **Bar Chart:** Compares votes received by individual candidates.

• **Treemap:** Shows candidate contribution to party vote share.
• **Scatter Plot:** Analyzes candidate vote performance across years.

7.2 Insights

• Vote share is often concentrated among a few top candidates.
• Voter participation varies widely across parties.
• Significant performance gaps exist between candidates of the same party.



---

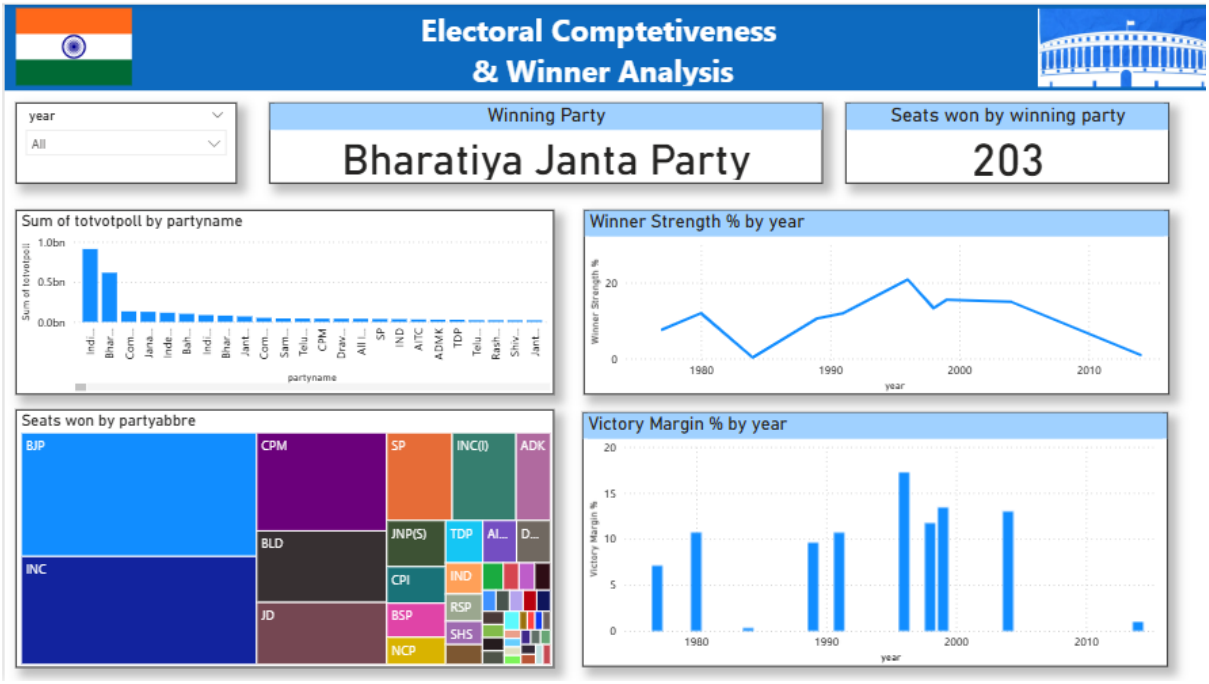
8. Electoral Competitiveness Page

8.1 Visuals and Purpose

• **Year Slicer:** Enables time-based analysis.
• **Column Chart (Victory Margin %):** Shows closeness of elections.
• **Line Chart (Winner Strength %):** Measures dominance over runner-up.
• **Treemap:** Displays seats won by each party.
• **Column Chart:** Shows total votes received by parties.

8.2 Insights

• Most elections are closely contested with low victory margins.
• Winner dominance varies significantly across years.
• National parties receive higher overall votes.
• Few major parties dominate seats, while regional parties remain strong locally.



Conclusion

This project successfully demonstrates the use of data analytics and visualization to analyze Indian national-level election data in a structured and interactive manner. By transforming complex, candidate-level datasets into meaningful dashboards, the project enables clear interpretation of voting patterns, party performance, electoral competitiveness, and representation trends. The use of Power BI, along with optimized DAX measures, ensures dynamic analysis across multiple dimensions such as year, state, constituency, party, gender, and candidate. Overall, the project highlights how data-driven visual storytelling can support transparent election analysis and provide valuable insights for informed decision-making.
