Infosys Springboard Virtual Internship 6.0
Batch-11 Group 1

Internship Completion Report

Team Details

Team Name: Team A
Team Members: Swayam Bana, Vaibhav Pawar, Nandhini B, Sharmika Shri R, Diksha Palliwal
Mentor Name: Mrs. Nithyasri S J
Internship Duration: 8 Weeks
Year : 2026

---

1. Project Title

ElectViz – Election Data Visualization for Media

---

2. Project Objective

The primary objective of this project is to analyze Indian State Assembly Election data spanning multiple election years and transform it into meaningful, actionable insights through a series of interactive and visually rich dashboards built in Power BI. The project was designed with the following specific objectives in mind:

- To move beyond raw vote counts and understand the deeper dynamics of election outcomes, including how votes are distributed across constituencies, how margins are formed, and what factors truly determine which candidate or party wins a seat
- To perform a comprehensive analysis of party performance across different states, years, and regions, identifying which parties demonstrated consistent dominance, which showed declining trends, and which emerged as growing forces in the Indian political landscape
- To examine the role of electability factors such as winning margins, close contest rates, and seat conversion efficiency, providing a data-

backed understanding of how competitive or decisive individual constituencies and elections have been over the years

- To investigate the influence of candidate demographics including age group, gender, and repeat participation history on electoral outcomes, revealing whether these factors play a statistically significant role in determining winners
- To study long-term electoral trends from 2009 to 2021 and identify patterns in voter participation, party competition, and geographic voting behavior that have evolved over the years
- To identify the key winning factors that contribute to seat conversion and electoral success, going beyond vote share alone to analyze strategic advantages that allow certain parties to win more seats relative to their vote totals
- To enable data-driven storytelling that makes complex electoral data accessible and understandable to a wide range of audiences including political analysts, academic researchers, media professionals, policy makers, and the general public
- To build a scalable and reusable analytical framework that can be adapted for future election datasets, making it a long-term tool for ongoing political analysis rather than a one-time project
- To demonstrate the practical application of Power BI capabilities including Power Query for data transformation, DAX for dynamic measure creation, and interactive dashboard design for real-world data analytics use cases

---

3. Project Description

3.1 Project Approach & Dashboard Modules

Indian elections are among the most complex democratic exercises in the world, involving hundreds of millions of voters, thousands of candidates, multiple political parties, and a vast and geographically diverse set of constituencies spread across different states and union territories. The outcomes of these elections are shaped by a wide range of interacting factors including party ideology, candidate reputation, regional sentiment, voter demographics, campaign strategies, and historical voting patterns. Raw election datasets, while rich in information, often fail to communicate these complexities in a clear and meaningful way to analysts, researchers, and the general public.

- The challenge with raw election data is that it exists in the form of large tabular records containing numerical values such as votes polled, positions secured, and constituency codes, which by themselves do not reveal the patterns, trends, and stories hidden within the data
- Without proper analytical processing and visualization, it becomes extremely difficult to answer important questions such as which party performed best in a particular state, whether elections are becoming more competitive over time, what demographic profile of candidates tends to win, or how efficiently a party converts its votes into seats
- These limitations in raw data interpretation create a clear need for a structured, visual, and interactive analytical solution that can process the data, apply meaningful calculations, and present the findings in a way that is both accurate and easy to understand

ElectViz is a Power BI-based analytical solution developed specifically to address these challenges and transform raw Indian State Assembly Election data into a comprehensive, interactive, and insight-driven reporting platform.

- The name ElectViz reflects the core purpose of the project — combining electoral data analysis with powerful visualization techniques to create a tool that makes election insights visible, accessible, and actionable
- The solution was built entirely within Microsoft Power BI, leveraging the full range of its capabilities including Power Query Editor for data ingestion and transformation, the Power BI data modeling environment for relational schema design, DAX for dynamic measure and KPI creation, and the Power BI report canvas for interactive dashboard development
- The project follows a structured end-to-end data analytics lifecycle, beginning from raw data collection and moving systematically through data cleaning, transformation, modeling, measure creation, dashboard design, and insight generation — ensuring that every stage of the process was handled with rigor and attention to quality
- The dataset used in the project covers Indian State Assembly Election results from 2009 to 2021, spanning over a decade of electoral history across multiple states, providing a rich and comprehensive foundation for both historical analysis and trend identification
- The solution is organized into seven distinct dashboard modules, each designed to focus on a specific analytical dimension of the election data, including overall election overview, party performance, gender distribution, candidate demographics, electability, trend analysis, and winning factor analysis
- Each dashboard module is built to be fully interactive, allowing users to apply slicers and filters for dimensions such as Year, State, Party, Constituency, and Gender, enabling them to explore the data at

whatever level of detail is most relevant to their specific analytical questions

- The project places strong emphasis on clarity and accuracy, ensuring that all calculations are mathematically correct, all visualizations are appropriately chosen for the type of data being displayed, and all insights presented are supported by the underlying data rather than assumptions
- A key design principle of ElectViz is insight generation through data storytelling — each dashboard is not just a collection of charts and numbers, but a carefully structured narrative that guides the viewer from broad patterns to specific findings, making it easy to draw meaningful conclusions without requiring deep technical expertise in data analysis
- The project also demonstrates how advanced analytical concepts such as seat conversion efficiency, close contest rate, repeat candidate impact, and zone-wise performance comparisons can be implemented in Power BI using DAX measures and applied to real-world political datasets
- ElectViz is designed to serve multiple types of users — political analysts can use it to evaluate party strategies, academic researchers can use it to study long-term electoral trends, media professionals can use it to quickly identify newsworthy patterns, policy makers can use it to understand voter engagement and representation issues, and students can use it as a practical example of applied data analytics on a real and socially relevant dataset
- The overall vision of the project is to demonstrate that data analytics and visualization can play a powerful role in making democratic information more transparent, understandable, and useful for everyone who engages with it, from domain experts to the general public

3.2 Project Approach & Dashboard Modules

1. Data Collection

- The project utilizes the State Level Assembly Constituencies Dataset as its primary source of raw data for analysis
- The dataset was sourced from Kaggle, a globally recognized and publicly available data platform widely used for data science and analytics projects
- The data comprehensively covers Indian State Assembly Election results spanning the period from 2009 to 2021, providing over a decade of electoral information
- It includes detailed information on multiple dimensions such as candidate names, constituency names, state names, party affiliations, votes received, positions secured, and final election outcomes
- The dataset also contains demographic details of candidates including age, gender, and category, which enabled deeper analysis beyond just vote counts
- Each record in the dataset represents an individual candidate's performance in a specific constituency during a particular election year, making it a highly granular and detailed source of electoral data
- Before beginning the analysis, the dataset was carefully reviewed to understand its structure, column definitions, data types, and overall quality so that an appropriate cleaning and transformation plan could be designed

---

2. Data Cleaning and Transformation

- The raw dataset was loaded into Power Query Editor in Power BI, which served as the primary environment for all data cleaning and transformation activities
- Duplicate records were identified and removed using the Remove Duplicates function to ensure that each candidate entry per constituency per year appeared only once in the dataset, maintaining the integrity of all aggregated calculations
- Irrelevant columns such as poll_no and delimit_id were dropped from the dataset as they carried no analytical value and only added unnecessary noise to the data model
- Missing values in critical fields such as candidate names, gender, age, and party information were handled systematically by replacing null or blank entries with standardized placeholder labels such as "Not Applicable" or "Not Available" to maintain completeness of records
- Inconsistent text values were corrected, such as replacing "Nota" with "NOTA" in the candidate name column and replacing "None" with "Not

Applicable" in the gender column, to ensure uniform representation across the entire dataset

- Data types of all columns were explicitly assigned and corrected — numerical columns such as _id, id, year, month, votes, and position were set to whole number format, while text-based columns such as state_name, candidate_name, party, and gender were standardized to text format
- Derived columns were created to enrich the dataset with additional analytical dimensions. These included:
  - canditate_type — to classify candidates as Independent or Party-affiliated
  - result_status — to label each candidate as Winner or Non-Winner based on their finishing position
  - candidate_age_group — to group candidates into age categories for demographic analysis
  - state_zone — to geographically classify each state into its respective zone such as North, South, East, West, or Central India
- Column reordering was performed to arrange all fields in a logical sequence, making the dataset easier to read, validate, and work with during the modeling and visualization phases
- The sex column was renamed to gender to align with professional reporting standards and improve clarity in dashboards and visuals
- A final row-level filter was applied to ensure only clean, valid, and relevant records were passed through to the Power BI data model, completing the transformation pipeline

---

3. Data Modeling

- A structured relational data model was designed within Power BI Desktop to organize the cleaned data into a format that supports efficient querying and accurate visual reporting
- The data model followed a Star Schema approach, where the central fact table containing election results was connected to surrounding dimension tables containing reference information about states, parties, candidates, and constituencies
- Relationships were established between key tables based on common fields such as state codes, constituency numbers, party names, and election years to enable seamless cross-filtering and drill-down capabilities
- Cardinality of relationships was carefully defined — one-to-many relationships were used wherever a single dimension record corresponded to multiple fact records, ensuring that filters flowed correctly from dimension to fact tables

- Filter direction was configured appropriately across all relationships to ensure that slicers and visual-level filters applied on one table correctly impacted all related tables and visuals on the dashboard
- The model was validated to confirm that all relationships were active and functioning correctly, with no ambiguous or circular dependency paths that could lead to incorrect results
- The finalized data model served as the backbone of the entire reporting layer, ensuring that all DAX measures, KPIs, and dashboard visuals produced accurate and consistent outputs regardless of the filter context applied by the user

---

4. Measure Creation using DAX

- All analytical calculations in the project were implemented as DAX (Data Analysis Expressions) measures rather than calculated columns, ensuring that computations were dynamic and responsive to user-applied filters and slicers
- The following key measures were developed and used throughout the dashboards:
- Total Votes — Calculated the sum of all votes received across the selected filter context, enabling state-wise, party-wise, and year-wise vote comparisons
- Seats Won — Counted the total number of constituencies won by each party or candidate group based on the result_status column, providing a direct measure of electoral success
- Vote Share Percentage — Computed the proportion of total votes received by a candidate or party relative to the total votes cast in the same constituency or region, expressed as a percentage
- Winning Margin Percentage — Measured the difference in votes between the winner and the runner-up as a percentage of total votes, indicating the competitiveness of each contest
- Close Contest Rate — Calculated the percentage of constituencies where the winning margin fell below a defined threshold, identifying how many seats were won by narrow margins
- Repeat Candidate Rate — Measured the proportion of candidates who had contested in previous elections, helping analyze the influence of experience and incumbent advantage on electoral outcomes
- Trend-based and Comparative KPIs — Developed year-over-year comparison measures to track changes in party performance, vote share, and seat counts across multiple election cycles
- All measures were rigorously tested against multiple filter combinations to verify that they responded correctly to slicers for Year, State, Party, Constituency, and Gender, producing accurate outputs in every scenario

---

5. Dashboard Design and Insight Generation

- A total of seven interactive dashboards were designed in Power BI, each focused on a distinct analytical theme to provide comprehensive coverage of the election dataset
- All dashboards were built using a combination of visual elements including bar charts, line charts, maps, scatter plots, donut charts, KPI cards, tables, and slicers to present data in an engaging and informative manner
- Slicers for Year, State, Party, and Constituency were implemented across all dashboards to allow users to dynamically filter the data and focus on specific subsets of interest
- Visual formatting, color coding, and layout design were applied consistently across all dashboards to ensure a professional and cohesive appearance throughout the report
- Each dashboard was designed to tell a clear data story, guiding the viewer from high-level summaries down to granular insights through logical arrangement of visuals

---

Dashboard Modules:

1. Election Overview Dashboard

- Provides a comprehensive high-level summary of the entire election dataset covering all years, states, and parties
- Displays key metrics through KPI cards, including total votes polled, total number of candidates, total constituencies covered, and total seats won by major parties
- Includes a party-wise seat distribution chart to show how seats were divided among different political parties across election years
- Features a state-wise performance map or bar chart to highlight which states recorded the highest voter participation and seat wins
- Acts as the landing page of the report, giving any viewer a quick and complete overview of the election landscape before diving into deeper analysis modules
- Enables users to immediately identify dominant parties, high-participation states, and overall election scale at a glance

---

2. Party Performance Analysis

- Provides an in-depth examination of how different political parties performed across various years and states throughout the 2009 to 2021 election period
- Includes year-wise trend charts to track how each party's seat count and vote share changed across multiple election cycles, revealing growth or decline patterns
- Analyzes seat conversion efficiency by comparing the percentage of votes received to the percentage of seats won, highlighting parties that converted votes to seats more effectively than others
- Identifies regional strongholds of major parties by showing which states consistently delivered the highest seat counts for each party
- Helps political analysts and researchers understand which parties maintained consistent dominance and which showed significant fluctuations over the years

---

3. Gender Distribution Analysis

- Provides a focused analysis of gender representation and participation in Indian state assembly elections across the study period
- Compares the total number of male versus female candidates who contested elections, revealing the significant gender gap in electoral participation
- Analyzes the success rate of male and female candidates by comparing the percentage of winners from each gender group, providing insight into whether gender influenced electoral outcomes
- Examines voter turnout patterns across gender categories to understand if constituencies with higher female candidate participation also recorded higher overall voter engagement
- Highlights the persistent representation imbalance in Indian elections and tracks whether the gender gap in candidacy and winning rates showed any improvement over the years

4. Candidate Demographic Analysis

- Studies detailed candidate-level attributes including age, gender, repeat participation history, and constituency-level presence to understand what demographic factors influence electoral success
- Identifies frequently winning candidates who have secured victories across multiple election cycles, revealing patterns of incumbent advantage and voter loyalty
- Analyzes age-wise distribution of candidates and winners to determine whether younger or more experienced candidates tend to perform better in different states and constituencies

- Examines regional voting behavior by comparing candidate demographics across different zones and states, revealing how local factors influence electoral preferences
- Tracks repeat candidates — those who contested in more than one election — and measures their success rate compared to first-time contestants
- Helps understand how demographic characteristics of candidates such as age group and gender contribute to their chances of winning in different political and geographic contexts

---

5. Electability Analysis

- Focuses on measuring and understanding the competitiveness of individual constituencies across different election years and states
- Calculates and visualizes winning margins for each constituency, showing whether seats were won by large comfortable margins or by very narrow differences
- Uses the Close Contest Rate measure to identify constituencies where election outcomes were highly competitive and could have gone either way
- Highlights consistently competitive constituencies where winning margins have historically been low, indicating regions where no single party or candidate holds a dominant advantage
- Analyzes the relationship between total votes polled and winning margin, exploring whether higher voter turnout tends to result in more competitive or more decisive outcomes
- Provides political strategists and analysts with a data-backed understanding of which constituencies require the most campaign effort and resources due to their highly competitive nature

---

6. Trend Analysis

- Examines long-term electoral trends across the full 2009 to 2021 period, providing a historical perspective on how Indian state assembly elections have evolved over time
- Tracks year-over-year changes in party dominance, showing which parties gained or lost influence across different election cycles and geographic regions
- Analyzes shifts in voter participation and total votes polled over the years to identify whether overall electoral engagement increased or decreased across states
- Measures changes in the close contest rate over time to determine whether elections became more or less competitive as the years progressed

- Explores the correlation between winning margins and party vote share trends, helping understand whether parties with growing vote shares also showed improvements in their winning margins
- Provides a valuable longitudinal view of the Indian electoral landscape, enabling comparisons between different political eras and helping forecast potential future trends based on historical patterns

---

7. Election Outcome and Winning Factor Analysis

- Identifies and examines the key factors that contributed to election victories across constituencies, states, parties, and years
- Analyzes the impact of repeat candidacy on winning probability, measuring whether candidates who contested multiple times had a significantly higher chance of winning compared to first-time contestants
- Evaluates winning margin patterns across parties, identifying which parties consistently secured victories by large margins and which parties frequently won by narrow differences
- Compares parties on strategic seat conversion efficiency, showing how effectively each party translated its overall vote share into actual seat wins — a critical metric that goes beyond simple vote counts
- Explains the phenomenon where some parties win more seats despite having a similar or lower total vote share compared to other parties, by analyzing vote distribution and constituency-level targeting patterns
- Combines multiple analytical dimensions — including demographics, geography, party strategy, and contest competitiveness — to provide a holistic explanation of what truly drives election outcomes in the Indian assembly election context

---

3.3 Technologies Used

- **Power BI Desktop** – Data visualization and dashboard development
- **Power Query** – Data cleaning and transformation
- **DAX (Data Analysis Expressions)** – Measure and KPI creation
- **Microsoft Excel / CSV** – Data storage and preprocessing
- **GitHub** – Version control, code collaboration, and project repository management
- **GitLab** – Source code management, CI/CD pipelines, and team collaboration
- **Microsoft Word** – Project documentation, report writing, and formatting
- **Microsoft PowerPoint** – Presentation creation and project demonstration
- **Microsoft Teams** – Conducting online sessions, mentor discussions, project reviews, and collaborative meetings

3.4 Key Insights from the Dashboard

- Vote share does not directly translate into seat wins
- Geographical distribution and constituency-level margins play a critical role
- Elections are becoming increasingly competitive over time
- Repeat candidates have a higher probability of winning
- Middle-aged candidates dominate seat victories
- Gender representation remains significantly skewed
- Parties with efficient vote distribution win more seats despite lower overall vote share
- The average winning margin has declined across successive election cycles
- Incumbency advantage is visible but weakens in highly competitive constituencies
- Female candidates who contest show a comparatively competitive win rate
- Regional parties demonstrate stronger vote-to-seat conversion within their dominant states

- Younger candidates below 35 years show an improving trend in recent election cycles
- Independent candidates collectively receive notable votes but win very few seats
- Voter turnout varies significantly across states and election cycles

---

3.5 Real-World Impact for Media and Public Communication

- Helps media houses present fact-based election analysis using verified data rather than assumptions
- Improves transparency in reporting by clearly showing vote share, margins, and regional trends
- Assists news channels in creating visual stories and interactive graphics for better audience engagement
- Enables comparison of current election results with historical trends to highlight shifts in voter behavior
- Supports early identification of swing states and closely contested constituencies
- Reduces misinformation by validating results through structured data analysis
- Helps journalists explain complex electoral systems (like vote share vs seat share differences) in a simplified way
- Strengthens credibility of media organizations through data-backed reporting
- Assists editorial teams in preparing post-election analytical reports and policy discussions
- Encourages responsible journalism by shifting focus from sensationalism to statistical evidence

---

4. Timeline Overview

Week	Activities Done	Activities Summary
Week 1	Dataset collection and understanding	Collected relevant election datasets, reviewed data sources, understood data structure, variables, and overall scope of analysis.
Week 2	Data cleaning and transformation	Cleaned raw data by handling missing values, correcting inconsistencies, standardizing formats, and transforming data using Power Query.
Week 3	Data modeling and DAX measure creation	Designed data relationships, built an optimized data model, and created DAX measures for KPIs, voter metrics, and analytical insights.
Week 4	Dashboard development and optimization	Developed interactive dashboards with filters and visuals, optimized performance, and improved usability and layout for better insights.
Week 5	Insight validation and documentation	Validated analytical results, cross-checked insights, prepared documentation, and finalized explanations for dashboard interpretation.
Week 6	Advanced analytics and performance enhancement	Implemented advanced DAX calculations, improved dashboard performance, refined measures, and enhanced data accuracy.
Week 7	Testing, feedback incorporation, and refinement	Conducted dashboard testing, incorporated mentor feedback, improved visual clarity, and enhanced overall user experience.
Week 8	Final review and project submission	Performed final validation, prepared project report and presentation, ensured documentation completeness, and completed project submission.

5. Project Execution

5.1 Key Milestones

Name	Description	Date Achieved
Milestone 1	Data understanding, dataset exploration, data cleaning, and initial data modeling in Power BI.	9/1/2026
Milestone 2	Development of core dashboards including Election Overview, Party Performance, and Gender Distribution analysis.	16/1/2026
Milestone 3	Creation of advanced dashboards covering Candidate Demographics and Electability analysis with custom DAX measures.	23/1/2026
Milestone 4	Implementation of Trend Analysis and Election Outcome & Winning Factor analysis, final insights, and project documentation.	30/1/2026

5.2 Project Execution Details

Step 1: Data Cleaning and Transformation using Power Query

1. Source (Importing the Dataset)

- The raw election dataset was imported into Power Query Editor as the primary data source
- The data was loaded from the original file and made available for all subsequent transformation steps
- This step established the foundation for all cleaning and preprocessing operations performed throughout the query

2. Promoted Headers

- The first row of the imported dataset was promoted to serve as the column headers of the table
- This ensured that all column names such as _id, id, year, month, state_name, state_code, and others were correctly recognized by Power Query
- This step was essential to give the dataset a proper and readable structure before any transformations were applied

3. Changed Type

- The data types of multiple columns were explicitly defined to match their intended formats
- Columns such as _id, id, year, and month were converted to the Int64 (Whole Number) data type to ensure numerical accuracy
- Assigning correct data types at this early stage helped prevent errors and inconsistencies in later transformation and calculation steps

4. Removed Columns

- Two columns — poll_no and delimit_id — were identified as irrelevant to the analysis and were permanently removed from the dataset
- Removing unnecessary columns reduced the size of the dataset and kept the table clean and focused only on meaningful fields
- This step also helped improve query performance and avoided confusion during visualization and reporting

5. Removed Duplicates

- The Table.Distinct function was applied to the dataset to eliminate all duplicate rows
- Every record in the table was made unique, ensuring that no candidate or election entry appeared more than once
- This step was critical to maintain the accuracy and integrity of the data, as duplicate entries could have led to inflated totals and incorrect analysis results

6. Replaced Value (sex column — "None" to "Not Applicable")

- In the sex column, entries carrying the value "None" were identified as incomplete or undefined records
- These values were replaced with the standardized text "Not Applicable" using the Replacer.ReplaceText function
- This ensured that all missing or unspecified gender entries were handled consistently and professionally across the entire dataset

7. Replaced Value1 (sex column — additional standardization)

- A second replacement operation was performed on the sex column to address any remaining inconsistencies or alternate forms of missing values
- This further ensured that the gender field was fully cleaned and contained only valid, standardized entries
- Having a fully consistent sex/gender column was important for accurate filtering and grouping in the final dashboard

8. Replaced Value2 (candidate_name column — "Nota" to "NOTA")

- In the candidate_name column, the value "Nota" was found to be inconsistently formatted
- It was replaced with the correct uppercase version "NOTA" using the Replacer.ReplaceText function to align with the official electoral terminology
- This correction ensured that all NOTA entries were recognized as a single consistent category during analysis and visualization

---

Step 2: Creation of Calculated Columns

9. Added Custom Column — candidate_type

- A new custom column named candidate_type was created to classify each candidate based on their party affiliation
- A conditional logic was applied: if the party column value, when converted to uppercase, equaled "IND", the candidate was labeled as "Independent"; otherwise a different category was assigned
- This classification allowed for easier analysis and filtering of independent candidates versus party-affiliated candidates throughout the dashboard

10. Changed Type1 — candidate_type column

- After creating the candidate_type column, its data type was explicitly set to Text
- This ensured the column was properly recognized as a categorical text field rather than any other data type
- Correct type assignment at each step helped prevent compatibility issues when using the columns in DAX calculations and visuals

11. Added Custom Column — result_status

- A new column called result_status was introduced to indicate the electoral outcome for each candidate
- The conditional logic used was: if the position column value equaled 1, the candidate was assigned the label "Winner"; all other candidates received a different designation
- This column made it straightforward to identify and analyze winning candidates across different constituencies, states, and years

12. Changed Type2 — result_status column

- The data type of the newly added result_status column was changed to Text to standardize it as a categorical field
- This ensured consistency in how the column was stored and used across all subsequent analysis steps and DAX measures

13. Added Custom Column — candidate_age_group

- A new column named candidate_age_group was created to handle missing age data in the dataset
- The logic applied was: if the age column value was null, the record was assigned the label "Not Available"; otherwise age-based grouping categories were applied
- This ensured that no records were left blank in the age group field, maintaining completeness of data for demographic analysis

14. Changed Type3 — candidate_age_group column

- The data type of the candidate_age_group column was changed to Text to ensure it was treated as a categorical classification field
- This step completed the setup of the age group column and made it ready for use in slicers, filters, and visuals within the dashboard

15. Added Custom Column — state_zone

- A new column called state_zone was created to geographically classify each state into its corresponding zone such as North, South, East, West, or Central
- The List.Contains function was used to check the state name against predefined lists of states belonging to each zone and assign the appropriate zone label
- This column provided a higher-level geographical dimension to the data, enabling zone-wise comparative analysis in the final dashboard

---

Step 3: Column Management and Standardization

16. Changed Type4 — state_zone column

- The data type of the state_zone column was set to Text to finalize it as a categorical text field
- This ensured the zone classifications were properly formatted and consistent before proceeding to the final structural steps of the query

17. Reordered Columns

- All columns in the dataset were rearranged into a logical and meaningful sequence for easier readability and navigation
- The reordered sequence started with _id, id, year, month, state_name, state_code, assembly_no, ac_name, followed by candidate-related and derived columns
- A well-organized column order improved the clarity of the dataset and made it easier for team members to review and validate the data

18. Renamed Columns

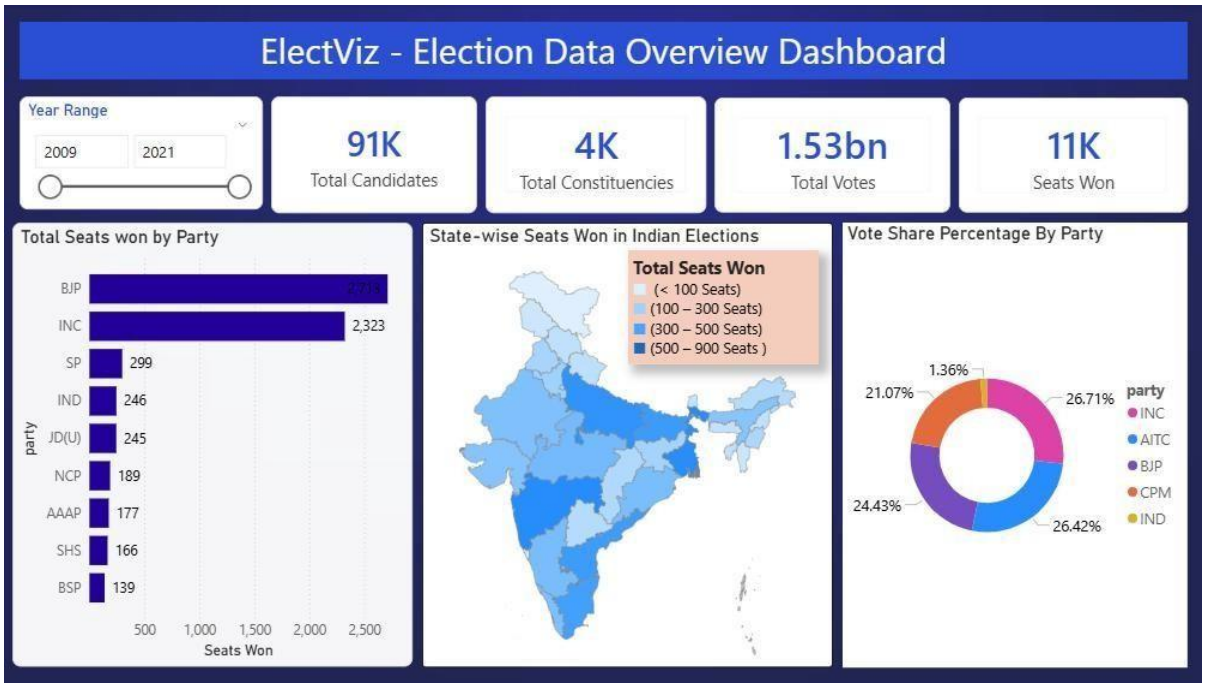
- The column originally named sex was renamed to gender to use more appropriate and professional terminology
- This renaming aligned the column label with standard reporting and data presentation conventions
- Clear and meaningful column names improved the overall readability of the dataset and the dashboard visuals built on top of it

19. Filtered Rows

- A final row-level filter was applied using the Table.SelectRows function to retain all valid and relevant records in the dataset
- This step ensured that only the appropriate rows were passed into the Power BI data model after all cleaning and transformation operations were complete
- The filtered dataset formed the final clean and structured version of the election data, ready for use in building KPIs, measures, and interactive dashboards

6. Snapshots / Screenshots

6.1 Election Overview Dashboard



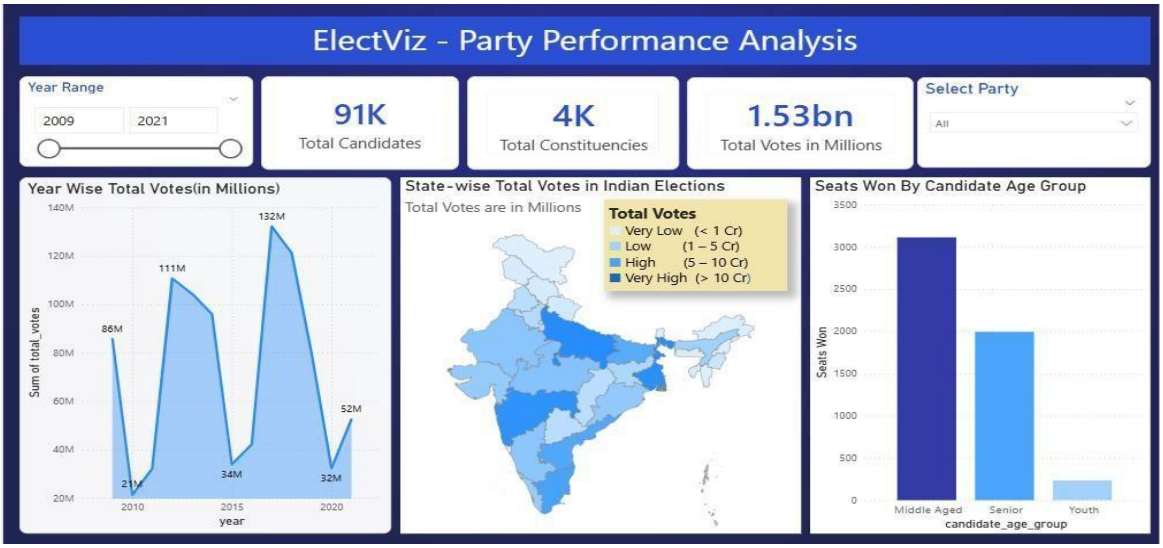
Purpose:
To provide a high-level summary of election participation and overall voting patterns across states and years.

Visuals Used:
KPI cards, shape/filled map, bar charts, slicers (Year, State).

Key Insights:

- Voter turnout varies significantly across states.
- Certain regions consistently show higher electoral participation.
- Overall participation trends change across election years.

6.2 Party Performance Analysis



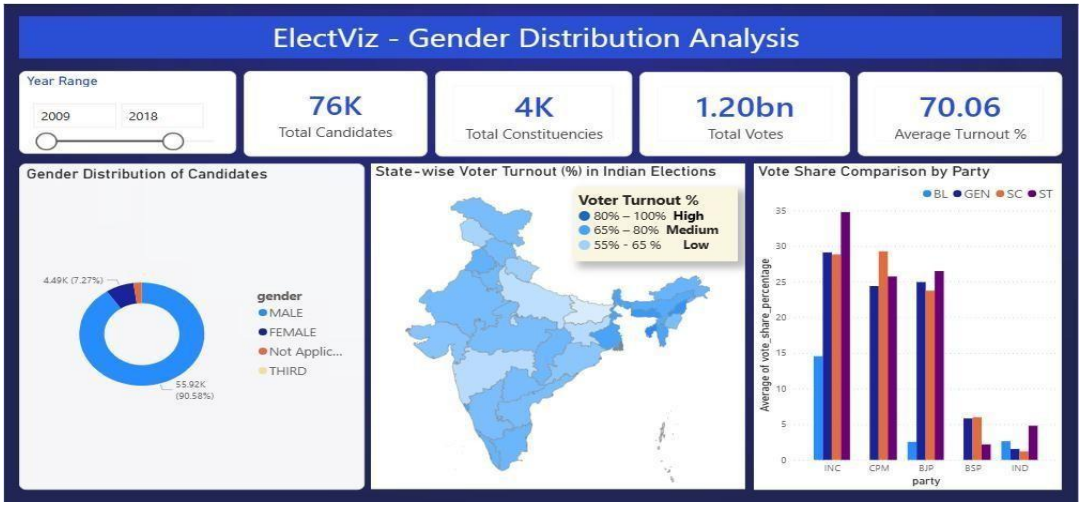
Purpose:
To provide a high-level summary of election participation and overall voting patterns across states and years.

Visuals Used:
KPI cards, shape/filled map, bar charts, slicers (Year, State).

Key Insights:

- Voter turnout varies significantly across states.
- Certain regions consistently show higher electoral participation.
- Overall participation trends change across election years.

6.3 Gender Distribution Analysis



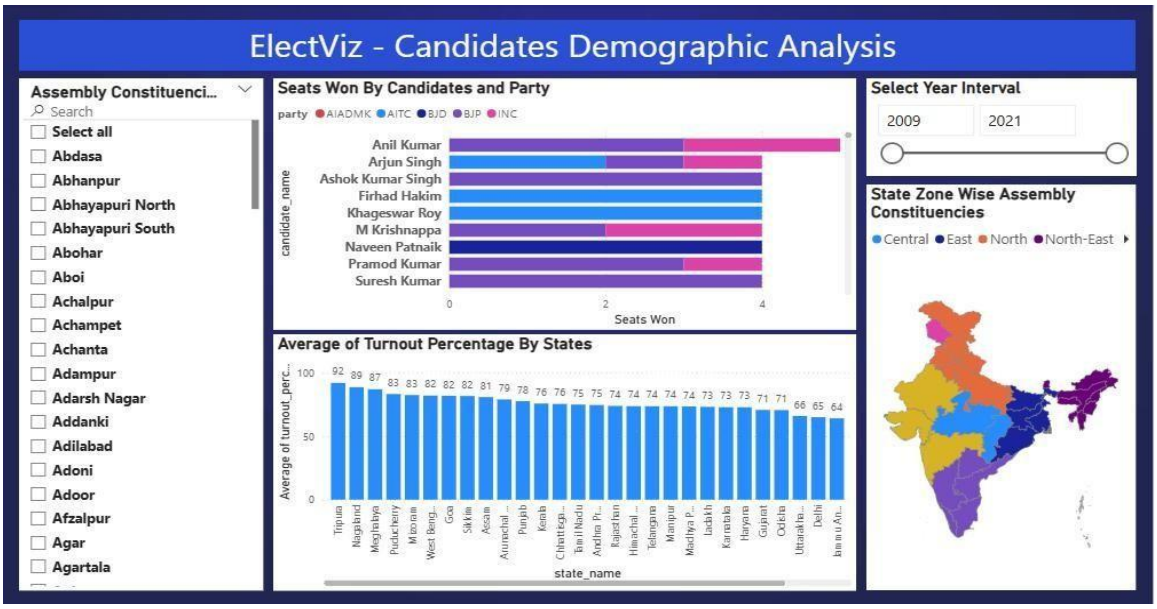
Purpose:
To evaluate party-wise performance in terms of seats won and vote share across elections.

Visuals Used:
Bar charts, stacked bar charts, KPI cards, party slicer.

Key Insights:

- Major parties dominate seat share across multiple years.
- Vote share does not always directly translate into seats won.
- Party performance varies across regions and election cycles.

6.4 Candidates Demographic Analysis



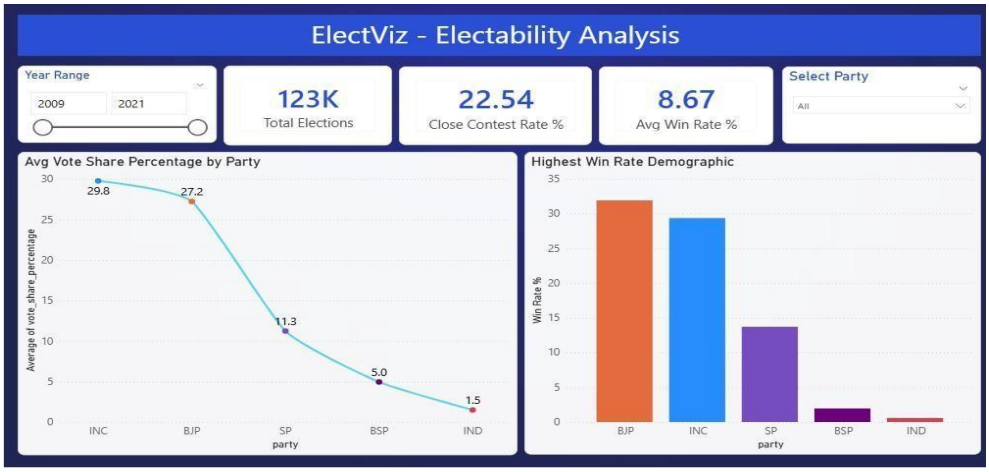
Purpose:
To analyze gender representation among candidates and elected representatives.

Visuals Used:
Bar charts, percentage KPIs, filters by year and state.

Key Insights:

- Male candidates significantly outnumber female candidates.
- Female representation remains limited across most states.
- Gradual improvement is visible in recent election years.

6.5 Electability Analysis



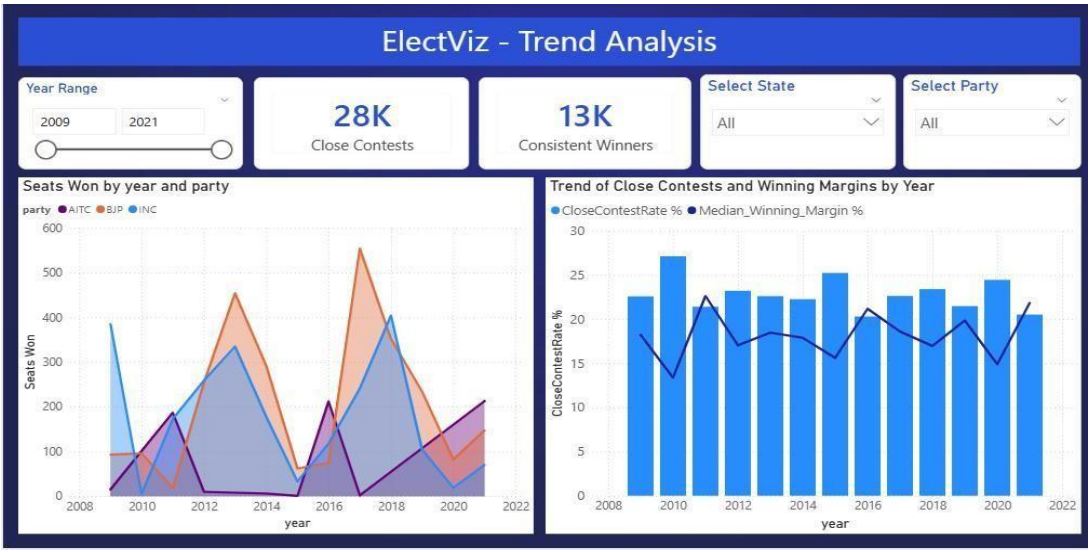
Purpose:
To assess factors influencing a candidate's probability of winning elections.

Visuals Used:
Scatter plots, KPI cards, bar charts, party and state slicers.

Key Insights:

- Higher winning margins often indicate strong candidate dominance.
- Parties may win more seats with lower average margins.
- Close contests reduce overall winning certainty.

6.6 Trend Analysis



Purpose:
To identify long-term trends in voter turnout, winning margins, and close contests over time.

Visuals Used:
Line charts, combined bar-line charts, median-based KPIs.

Key Insights:

- Increase in close contests corresponds to declining winning margins.
- Electoral competition has intensified over the years.
- Margins are becoming narrower in recent elections.

6.7 Election Outcome and Winning Factor Analysis

Purpose:
To analyze election outcomes and identify key factors contributing to consistent wins.

Visuals Used:
Scatter plots, bar charts, reference lines, candidate-level analysis.

Key Insights:

- Some candidates show consistent winning patterns across elections.
- Parties may win fewer seats but with stronger margins.
- Winning outcomes depend on both party strength and candidate performance.

7. Challenges Faced

During the development of the ElectViz project, several technical and analytical challenges were encountered at different stages of the data analytics lifecycle. Each challenge was systematically addressed using appropriate tools, techniques, and problem-solving approaches. The following table outlines the key challenges faced and the resolutions applied:

Challenge	Resolution
Handling large and inconsistent election datasets	• The raw dataset contained a very large number of records spanning multiple election years and states with inconsistent formatting across different fields • Power Query Editor was used to explicitly assign correct data types to every column in the dataset • Irrelevant columns such as poll_no and delimit_id were removed to reduce dataset size and improve processing speed • Column names were standardized to follow a consistent naming convention throughout the dataset • These steps ensured the dataset was clean, lightweight, and optimally structured for smooth processing within the Power BI data model
Cleaning missing and duplicate records	• Data profiling techniques were applied within Power Query to identify the extent and location of null values and duplicate entries across all columns • Duplicate rows were permanently removed using the Remove Duplicates function to ensure every record appeared only once • Missing values in text fields such as gender and age were replaced with standardized labels such as "Not Applicable" and "Not Available" to maintain record completeness • Inconsistent text entries such as "Nota" and "None" were corrected to their proper standardized forms • These steps ensured the final dataset was accurate, complete, and free from any redundant or misleading records
Creating accurate DAX measures for margins and rates	• Each DAX measure was carefully designed step by step before being implemented, with the full formula logic mapped out in advance • Measures for complex metrics such as Winning Margin Percentage, Close Contest Rate, and Vote Share Percentage were broken down into smaller logical components and built incrementally • Each completed measure was tested against known data subsets to verify that it produced the expected output • Results were cross-verified by manually calculating expected values for specific constituencies and comparing

Challenge	Resolution
Interpreting complex electoral relationships correctly	• Comparative analysis was performed across multiple dimensions simultaneously rather than examining any single metric in isolation • Seat conversion efficiency, vote share distribution, and margin analysis were cross-examined together to build a holistic understanding of the data • All insights were validated using statistical consistency checks before being included in the final dashboards • Visual outputs were reviewed multiple times to ensure that the charts and KPIs were correctly representing the underlying data relationships • This careful approach ensured that all conclusions drawn from the analysis were accurate, well-supported, and free from misinterpretation
Designing dashboards that were both informative and easy to understand	• Each dashboard page was designed around a single clear analytical theme to avoid information overload • The most important KPIs were placed prominently at the top of each page using card visuals, with supporting charts and detailed breakdowns arranged below • Visual types were carefully selected to match the nature of the data being displayed, such as bar charts for comparisons, line charts for trends, and maps for geographic analysis • Color schemes, font sizes, and visual spacing were standardized across all pages to ensure a professional and consistent appearance • Multiple rounds of layout review were conducted to ensure every dashboard was clean, readable, and visually balanced
Managing and organizing multiple DAX measures efficiently	• As the number of DAX measures grew, all measures were grouped into clearly named folders within the data model based on their analytical category • Categories such as Vote Metrics, Performance Metrics, and Contest Metrics were created to logically separate different types of measures • A consistent and descriptive naming convention was applied to every measure so that its purpose was immediately clear from its name alone • Measures that were no longer needed were removed promptly to keep the model clean and easy to navigate • This organized approach reduced confusion during development and made it significantly easier to locate and update specific measures when required

8. Learnings & Skills Acquired

8.1 Technical Skills

- **Power BI Dashboard Development** — Gained comprehensive hands-on experience in building end-to-end interactive dashboards within Microsoft Power BI, covering the entire workflow from data import and transformation to final visual design and publishing. Learned how to select the most appropriate chart types for different kinds of data, design clean and professional report layouts, and implement interactive features such as slicers, drill-throughs, and cross-filtering to enhance user experience
- **Power Query and Data Transformation** — Developed strong proficiency in using Power Query Editor to clean, reshape, and enrich raw datasets. This included skills in removing duplicates, handling null values, correcting data types, creating custom conditional columns, renaming and reordering columns, and applying row-level filters — all of which are essential capabilities for any real-world data analytics project
- **Advanced DAX Calculations** — Built significant expertise in writing and optimizing DAX formulas for a wide variety of use cases including aggregations, percentage calculations, conditional logic, row context versus filter context management, and time-based comparisons. Developed a deep understanding of how DAX measures behave differently from calculated columns and how to leverage this distinction for maximum analytical flexibility
- **Data Modeling and Optimization** — Learned how to design efficient relational data models in Power BI using Star Schema principles, define correct cardinality and filter direction for all relationships, and optimize model performance by removing unnecessary fields, avoiding redundant calculations, and structuring the model to support fast and accurate cross-filtering across all dashboard pages
- **KPI Design and Metric Development** — Gained practical experience in translating business and domain questions into measurable KPIs. Learned how to design metrics that are not only mathematically correct but also meaningful, interpretable, and aligned with the analytical objectives of the project

8.2 Analytical and Problem-Solving Skills

- **Electoral Data Interpretation** — Developed a strong ability to read, interpret, and derive insights from complex electoral datasets. Learned how to look beyond simple totals and percentages to uncover deeper patterns such as seat conversion inefficiencies, geographic voting concentration, and the impact of candidate experience on winning probability
- **Pattern Recognition and Trend Analysis** — Built skills in identifying recurring patterns and long-term trends within multi-year datasets. Learned how to use visual and statistical techniques to distinguish genuine trends from random variations, and how to communicate these trends clearly through appropriate chart selections and narrative framing

- **KPI-Driven Insight Generation** — Developed the ability to design and use KPIs not just as reporting metrics but as analytical tools that drive insight discovery. Learned how combining multiple KPIs — such as vote share, seat share, and winning margin — together can reveal strategic patterns that would be invisible when looking at any single metric in isolation
- **Critical Thinking and Validation** — Strengthened the habit of questioning and verifying every analytical output before drawing conclusions. Learned the importance of cross-checking results from different angles and using logical reasoning to identify potential errors in calculations or misinterpretations of data relationships
- **Data Storytelling** — Developed the skill of organizing analytical findings into a coherent and compelling narrative that guides the audience from broad observations to specific conclusions in a logical and engaging way, making complex data insights accessible to non-technical viewers

8.3 Soft Skills and Team Collaboration

- **Time Management and Planning** — Managing a multi-stage analytics project that included data cleaning, modeling, measure development, and dashboard design across seven modules required careful planning and disciplined time allocation. Learned how to break down a large project into manageable phases, set realistic milestones, and prioritize tasks to ensure steady progress and timely completion
- **Documentation and Presentation Skills** — Developed the ability to document every step of the analytical process clearly and comprehensively, from data transformation steps in Power Query to DAX formula logic and dashboard design decisions. This documentation practice not only supported project organization but also made it easier to explain the work to mentors, peers, and evaluators in a clear and professional manner
- **Analytical Storytelling** — Learned how to present data findings in a way that resonates with the audience by combining visual evidence with clear explanations, logical sequencing, and contextual interpretation. This skill is essential for any data analyst who needs to communicate insights to stakeholders who may not have a technical background
- **Attention to Detail** — Consistent work with large datasets and complex DAX calculations reinforced the importance of precision and thoroughness in every aspect of the project, from ensuring data types are correctly assigned to verifying that every visual on the dashboard reflects the intended analytical message

8.4 Domain Knowledge and Application

- **Indian Electoral System Understanding** — Gained a thorough understanding of how Indian State Assembly Elections work, including the structure of constituencies, the role of different political parties, the significance of vote share versus seat share, and the importance of regional political dynamics in shaping national electoral trends

- **Political Data Analytics** — Developed specialized knowledge in applying data analytics techniques to political datasets, including an understanding of what metrics matter most in electoral analysis, how to handle the unique characteristics of election data such as multi-party competition and constituency-level variation, and how to generate insights that are politically meaningful and analytically sound
- **Real-World Application of Business Intelligence** — The project demonstrated how Business Intelligence tools like Power BI can be applied to domains beyond traditional business settings such as sales and finance. Working with election data showed that BI tools are equally powerful and relevant for public policy analysis, social research, and civic data projects, broadening the understanding of where and how data analytics skills can create real value
- **Understanding of Data-Driven Decision Making** — Through the process of building ElectViz, developed a deeper appreciation for how data-driven approaches can improve the quality of decisions and interpretations in any domain. Learned that the true value of a data analytics project lies not just in the technical execution but in the quality and relevance of the insights it produces for its intended audience.

9. Testimonials from Team Members

Swayam Bana

"Contributed to the overall project design and development, focusing on data modeling, DAX measure creation, and dashboard development. Played a key role in building analytical dashboards, extracting insights, and coordinating documentation and final presentation."

Nandhini B

"Actively contributed to data cleaning, preprocessing, and validation to ensure data accuracy. Assisted in exploratory analysis and supported the creation of initial visualizations and insights."

Sharmika Shri R

"Worked on dashboard layout planning and visualization design. Contributed to KPI identification, chart selection, and improving dashboard usability and clarity for end users."

Vaibhav Pawar

"Supported analysis related to election trends and outcome patterns. Assisted in validating results, reviewing insights, and contributing to milestone documentation."

Diksha Palliwal

"Contributed to research, dataset understanding, and documentation. Assisted in interpreting insights and aligning the analysis with real-world election scenarios."

10] Conclusion

The ElectViz project successfully demonstrates how data analytics and visualization techniques can transform complex and large-scale election datasets into meaningful, easy-to-understand insights. By applying systematic data cleaning, transformation, and advanced DAX calculations, the project converts raw electoral data into structured information that highlights voting patterns, margin distributions, regional competitiveness, and overall electoral performance. The interactive dashboards allow users to explore election results dynamically through slicers and filters, making the analysis both flexible and user-friendly. This approach not only improves clarity but also ensures that insights are derived through logical and statistically validated methods.

Furthermore, the project adds significant real-world value by enabling data-driven interpretation of election outcomes rather than opinion-based narratives. The dashboards help media professionals, political analysts, and the general public better understand electoral behavior, shifts in vote share, competitive constituencies, and winning strategies. By presenting complex relationships such as seat share versus vote share in a simplified visual format, ElectViz strengthens transparency and promotes informed discussions. Overall, the project highlights the power of business intelligence tools in supporting accurate reporting, analytical decision-making, and responsible public discourse in the context of elections.

11] Acknowledgements

We sincerely thank **Infosys Springboard**, our mentor, and the internship coordinators for providing continuous guidance, valuable resources, and an excellent learning platform that enabled the successful completion of this project. The structured internship program and supportive environment helped us gain practical exposure and apply theoretical knowledge effectively.

We extend our heartfelt gratitude to the **Infosys Springboard Team** for organizing this valuable internship opportunity and to our mentor, **Nityashree Ma'am**, for her constant encouragement, constructive feedback, and technical guidance throughout the project. Her insights played a crucial role in shaping the project outcomes and improving our understanding of real-world data analytics practices.

We also express our sincere appreciation to our **teammates** for their collaboration, shared ideas, and mutual support during the project. This internship has been a significant milestone in our professional growth, strengthening both our technical skills and teamwork abilities, and has contributed positively to our overall learning journey.