

VIRTUAL INTERNSHIP 6.0



“Electviz: Election Data Visualization for Media”

A PROJECT SUBMITTED TO -

INFOSYS SPRINGBOARD 6.0

BY

Raut Mayur Shivaji

Master of Science – Statistics

2025-26

INDEX

Sr. No.	Title	Page Number
1)	Abstract	3
2)	Objectives	4
3)	Problem Statement	5
4)	Key Features	6
4)	Data Flow Diagram (DFD)	9-10
5)	Technology stack	11
6)	Module Descriptions	12-13
6)	Project Diagrams	14-19
7)	Pros and Cons Section	20-21
8)	Future Enhancements	22
9)	Conclusions	23
10)	Learnings & Skills Gained	24
11)	References	25

ABSTRACT

This project focuses on building an interactive Election Data Visualization Dashboard using Power BI, designed to analyze the performance of political parties and candidates across different states of India. The dashboard provides real-time insights into electoral outcomes by combining information such as vote share, candidate demographics, education level, financial status, and criminal background.

Through the integration of these multiple data dimensions, the project offers a holistic understanding of election dynamics, enabling users to identify key patterns such as the relationship between candidate wealth, education, and success rate. The visualization dashboards are useful for journalists, researchers, and policymakers to explore and interpret election trends quickly and intuitively.

OBJECTIVES

- To visualize and compare the performance of political parties across multiple states.
- To analyze candidate demographics (age, gender, and education) and influence on election results.
- To study the financial background of candidates, including declared assets and liabilities.
- To summarize election outcomes using KPIs and comparative charts for quick insights.

PROBLEM STATEMENT

Election data is often presented in long tables or text-based reports, which make it difficult to extract insights quickly.

Media professionals and political analysts require clear, visual summaries that highlight key information such as leading parties, vote margins, and regional participation.

Traditional reporting methods lack interactivity, making it hard to explore data in detail or compare different regions.

This project aims to overcome these challenges by using Power BI to design dashboards that offer instant insights, trend visualization, and comparative analytics in a single interactive view.

It simplifies complex datasets and provides a tool that supports faster and more accurate election analysis.

KEY FEATURES

1. **Interactive Visualization:** The dashboard allows users to filter results by state, party, year, and candidate to view customized insights.
2. **Candidate Demographics:** Displays gender ratio, average age, and education levels of candidates and winners.
3. **Financial Overview:** Presents candidates' assets and liabilities, highlighting differences between winners and non-winners.
4. **Criminal Case Analysis:** Visualizes how many candidates have declared criminal cases and explores any relation between criminal record and election results.
5. **State and Party Performance:** Shows comparative analysis of vote shares, number of candidates, and winning percentages.
6. **Dynamic KPIs:** Total Candidates, Total Winners, Average Age, Average Assets, and Total Criminal Cases are displayed as key summary indicators.
7. **Correlation Insights:** Uses scatter plots and bar charts to identify patterns such as “Higher assets → higher chance of winning.”
8. **User-Friendly Design:** Includes interactive slicers, filters, and dropdowns for better exploration of data.

➤ Sample Dataset:

STATE	CONSTITUENCY	NAME	WINNER	PARTY	SYMBOL	GENDER	CRIMINAL
Telangana	ADILABAD	SOYAM BAPU RAO	1	BJP	Lotus	MALE	52
Telangana	ADILABAD	RATHOD RAMESH	0	INC	Hand	MALE	3
Telangana	ADILABAD	Godam Nagesh	0	TRS	Car	MALE	0
Telangana	ADILABAD	NOTA	0	NOTA			
Uttar Pradesh	AGRA	Satyapal Singh Baghel	1	BJP	Lotus	MALE	5
Uttar Pradesh	AGRA	Manoj Kumar Soni	0	BSP	Elephant	MALE	0
Uttar Pradesh	AGRA	Preeta Harit	0	INC	Hand	FEMALE	0
Maharashtra	AHMADNAGAR	SANGRAM ARUNKAKA JAGTAP	0	NCP	Clock	MALE	1
Maharashtra	AHMADNAGAR	Dr. SUJAY RADHAKRISHNA VIKHEPATIL	1	BJP	Lotus	MALE	0
Maharashtra	AHMADNAGAR	SUDHAKAR LAXMAN AVHAD	0	VBA	Cup & Saucer	MALE	0

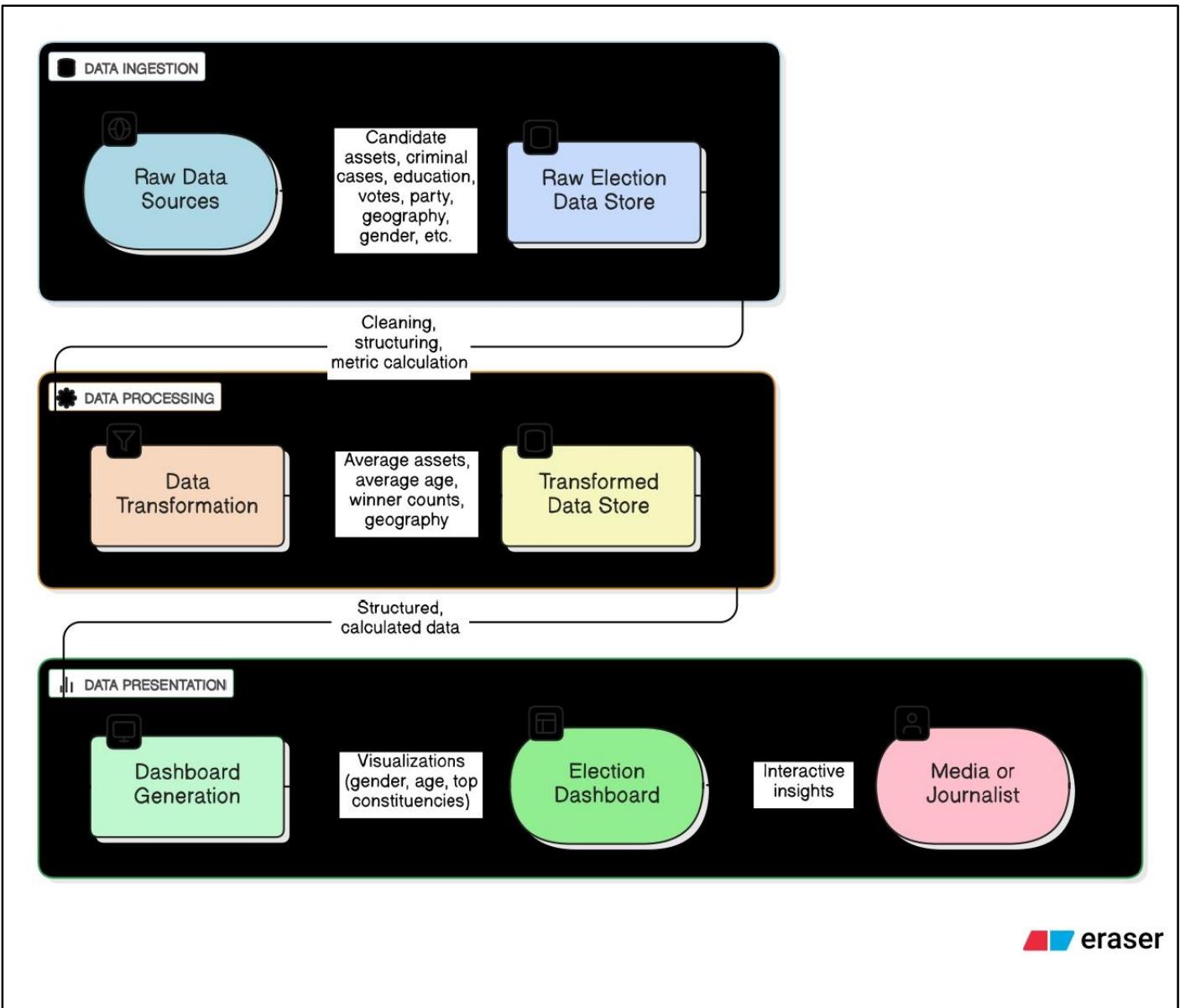
AGE	CATEGORY	EDUCATION	ASSETS	LIABILITIES	GENERAL	POSTAL	TOTAL
52	ST	12th Pass	30,99,414	2,31,450	376892	482	377374
52	ST	12th Pass	3,64,91,000	1,53,00,000	314057	181	314238
54	ST	Post Graduate	1,84,77,888	8,47,000	318665	149	318814
					13030	6	13036
58	SC	Doctorate	7,42,74,036	86,06,522	644459	2416	646875
47	SC	Post Graduate	13,37,84,385	2,22,51,891	434199	1130	435329
54	SC	Post Graduate	5,50,75,753	0	44877	272	45149
34	GENERAL	Graduate	9,44,88,381	3,25,49,132	419364	3822	423186
37	GENERAL	Doctorate	16,86,64,576	26,23,964	696961	7699	704660
62	GENERAL	Graduate	1,39,49,000	14,00,000	31644	163	31807

OVER TOTAL ELECTORS	OVER TOTAL VOTES	TOTAL ELECTORS
25.33068419	35.4682479	1489790
21.09277146	29.53428505	1489790
21.39992885	29.96436953	1489790
0.875022654	1.225214455	1489790
33.38382301	56.46461464	1937690
22.46639039	37.99912537	1937690
2.330042473	3.940979148	1937690
22.73487211	35.08743129	1861396
37.85653348	58.42515898	1861396
1.708771266	2.637199546	1861396

- State: The name of the state where the election took place.
- Constituency: The name of the constituency (area) contested in the election.
- NAME: The full name of the candidate.
- WINNER: Indicates if the candidate won (1 = Winner, 0 = Lost).

- **SYMBOL:** The official election symbol of the political party.
- **GENDER:** Gender of the candidate (Male/Female/Other).
- **CRIMINAL CASES:** The number of criminal cases filed against the candidate.
- **AGE:** The candidate's age.
- **CATEGORY:** The social or reservation category (e.g., SC, ST, OBC, General).
- **EDUCATION:** The highest educational qualification of the candidate.
- **ASSETS:** The total declared assets of the candidate (in ₹).
- **LIABILITIES:** The total declared liabilities or debts (in ₹).
- **GENERAL VOTES:** Number of votes received through general (ballot/EVM) voting.
- **POSTAL VOTES:** Number of votes received through postal ballots.
- **TOTAL VOTES:** Combined total of general and postal votes.
- **OVER TOTAL ELECTORS IN CONSTITUENCY:** Percentage of total electors in the constituency who voted for this candidate.
- **OVER TOTAL VOTES POLLED IN CONSTITUENCY:** Percentage of total votes polled that went to this candidate.
- **TOTAL ELECTORS:** Total number of registered voters (electors) in the constituency.

DATA FLOW DIAGRAM (DFD)



This diagram shows how the election data moves step by step — from collecting raw information to creating a Power BI dashboard that shows clear insights for journalists or analysts.

It has three main parts:

1. Data Ingestion
2. Data Processing
3. Data Presentation

1. Data Ingestion: (Collecting Data)

- In this first step, all the raw election data is collected from different sources such as CSV files or government reports.
- The data includes details like candidate name, age, gender, education, party, assets, liabilities, votes, and constituency.
- All this data is stored together in one place called the Raw Election Data Store.
- This step helps to keep all information organized before cleaning and analysis.

2. Data Processing: (Cleaning & Preparing Data)

- In this stage, the raw data is cleaned and corrected — any missing or repeated information is fixed.
- The data is then structured and transformed to calculate important values like:
 - Average age and assets of candidates
 - Total winners by party or state
 - Gender ratio of contestants
 - Vote share and win percentage
- The cleaned and calculated data is saved in the Transformed Data Store, which is ready for creating visuals.

3. Data Presentation: (Creating Dashboards)

- In this final stage, the processed data is used in Power BI to create interactive visuals and charts.
- These dashboards show different insights like:
 - Which party won the most seats
 - Gender and education of candidates
 - States with the highest votes or richest candidates
- The final output is the Election Dashboard, which is used by media and journalists to view results and share insights.

TECHNOLOGY STACK

The project utilizes a combination of data analysis, visualization, and predictive modeling tools to collect, process, visualize election data.

1. Power BI:

- Used for **interactive data visualization** and **dashboard creation**.
- Used as the main visualization and reporting tool to create interactive dashboards, charts, maps, and filters.
- It provides real-time analytics and visual summaries for users.
- Enables **report publishing** for use by media teams and analysts.

2. Python:

- Used for data cleaning, transformation, and preliminary analysis before importing into Power BI. Python helped ensure accuracy and consistency in datasets.
- Libraries used include:
 - Pandas → for data manipulation and preparation
 - Matplotlib / Seaborn → for additional data visualizations
 - Helps identify trends, party performance patterns, and voter behavior insights.

3. Supporting tools:

Microsoft Excel / CSV Files → Used as the base data storage format. Excel sheets were prepared with well-organized columns such as year, state, constituency, total votes, and party name.

Power Query → Used for transforming raw data (removing duplicates, correcting data types, merging tables) directly inside Power BI

MODULE DESCRIPTION

1. Data Collection Module:

Purpose: To gather and compile raw election data from reliable sources.

Description:

- Collects raw election data from reliable sources such as government or official election reports and stores it in CSV or Excel format.

2. Data Preprocessing & Cleaning Module:

Purpose: To prepare and transform raw data into a usable format.

Description:

- Cleans and prepares the data using Python and Power Query — handling missing values, standardizing names, and ensuring uniform structure across datasets.

3. Data Visualization Module:

Purpose: To represent analytical data in an easy-to-understand visual format.

Description:

- Uses Power BI visuals to represent key insights like total votes, party performance, voter turnout, and top-performing constituencies.

4. Comparative Analysis Module:

Purpose: To help users compare election data across different dimensions such as years, parties, and regions.

Descriptions:

- Compares data across years or regions to highlight changes in party performance, voter turnout, and regional political behavior.

5. Insight & Reporting Module:

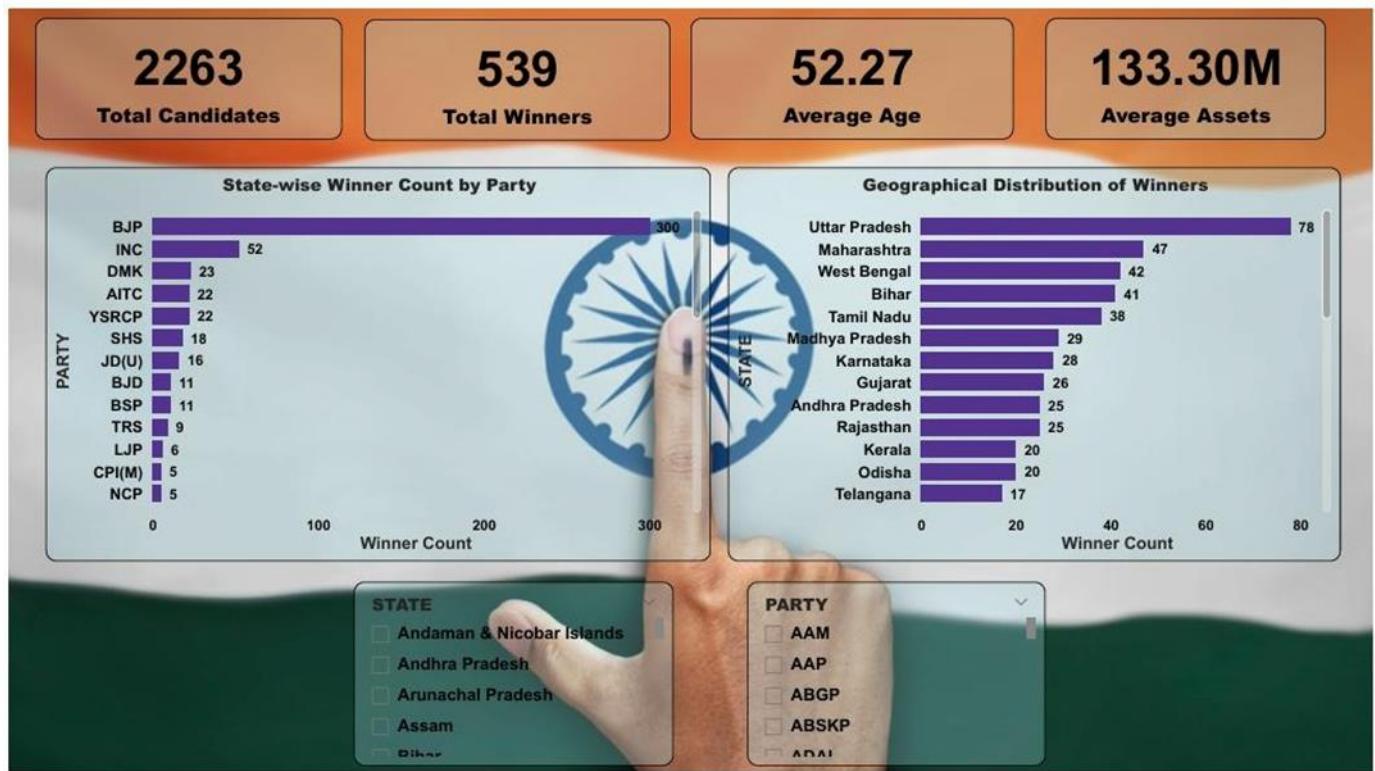
Purpose: To present insights and predictions in a shareable and interactive way.

Description:

- Summarizes all key information into a single dashboard view.
- Allows interactive exploration using filters and slicers for users to analyze results from different angles.
- Produces exportable visuals and reports for presentations and media coverage.

PROJECT DIAGRAMS

1) Slide 1

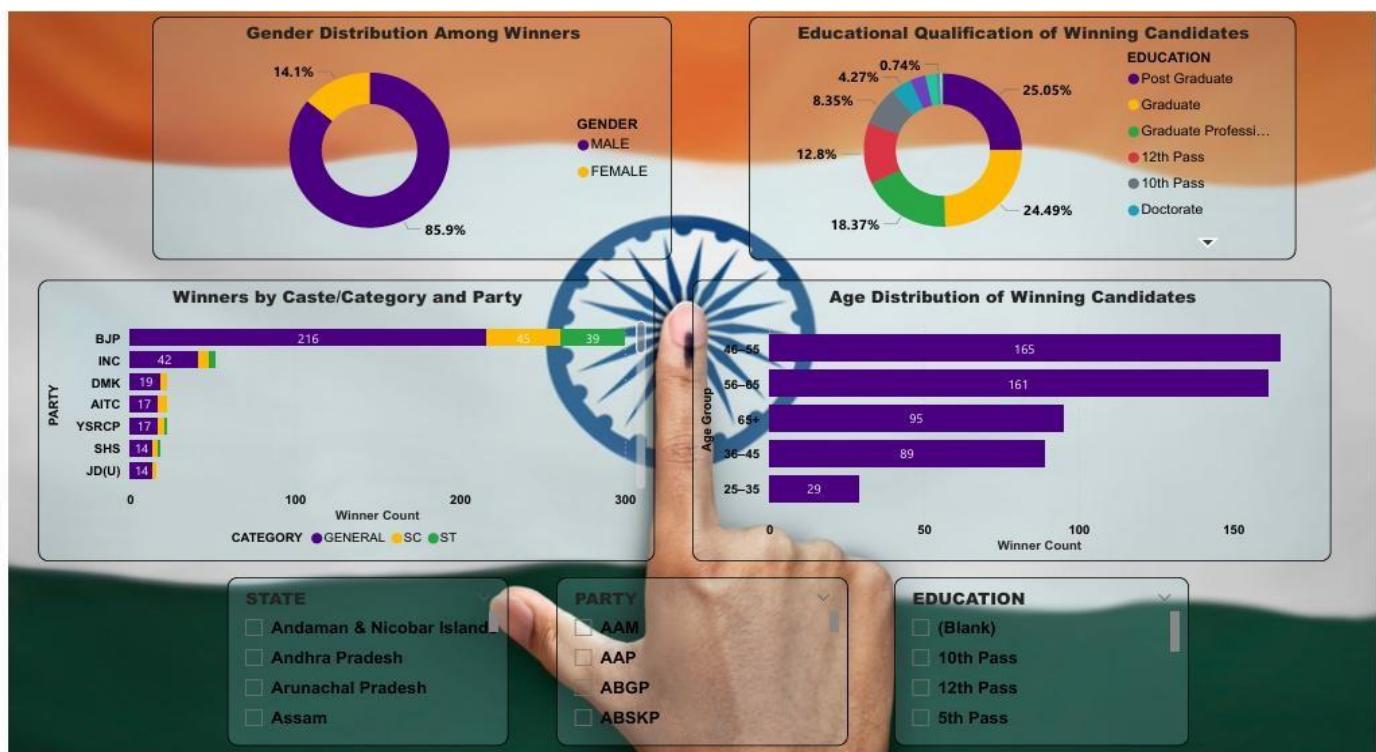


- Total Candidates:** 2263 people contested.
- Total Winners:** 539 candidates secured seats.
- Average Age:** 52.27 years, showing most leaders are middle-aged.
- Average Assets:** ₹133.3 million per candidate, indicating generally high net worth.

➤ Major Insights:

- BJP fielded the largest number of candidates and also won the most seats.
- The geographical map makes it easy to see how victories are distributed across India.
- The visuals show State-wise winner counts and Party-wise performance, highlighting leading states such as Uttar Pradesh, Maharashtra, and West Bengal.

2) Slide 2

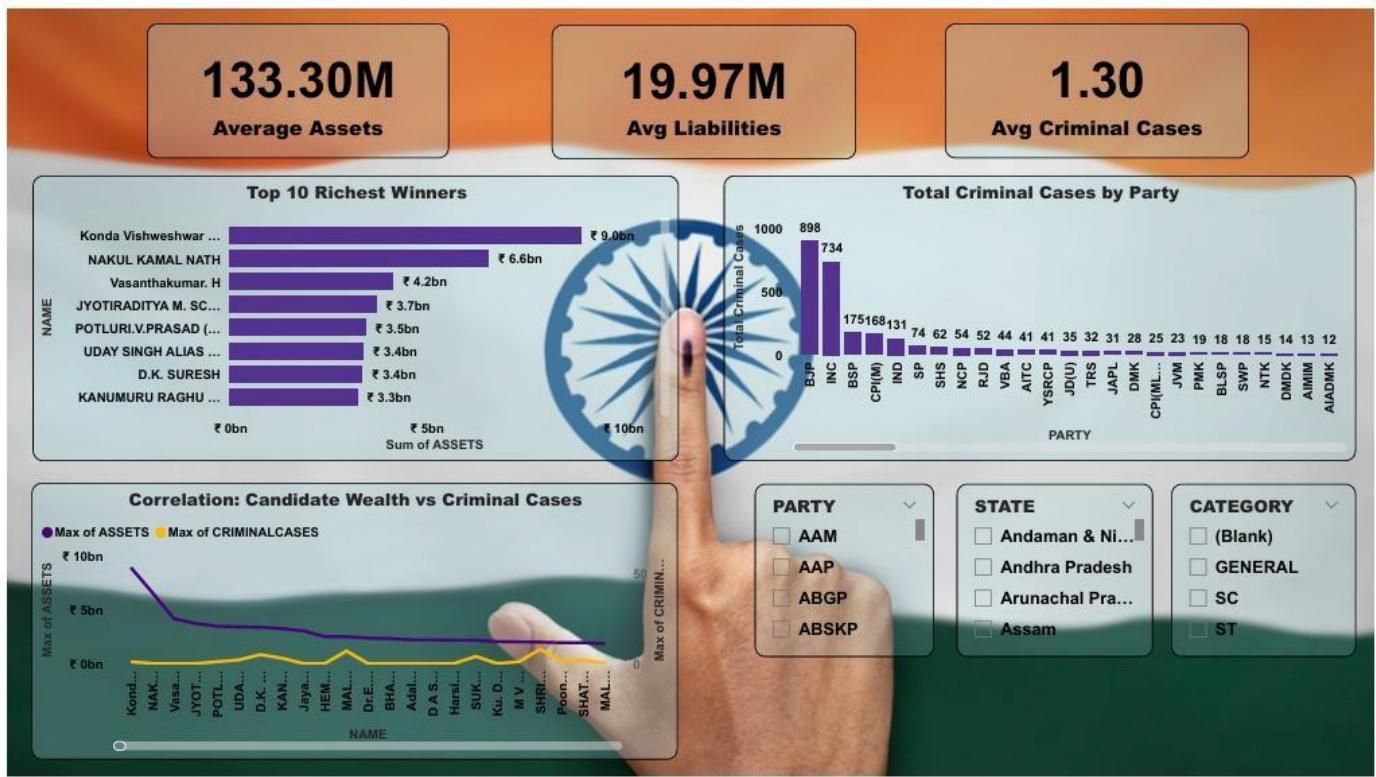


- Gender Distribution:** 86 % male vs 14 % female, revealing a large gender gap.
- Education:** Post-graduates (25 %) and graduates (24 %) form about half of all winners, showing higher education correlates with political success.
- Caste/Category Split:** General-category candidates dominate, followed by SC and ST groups.
- Age Groups:** Most winners are between 46 and 65 years, confirming that experienced candidates perform better electorally.

➤ Overall Purpose of this slide:

To understand social and educational diversity among elected representatives.

3) Slide 3



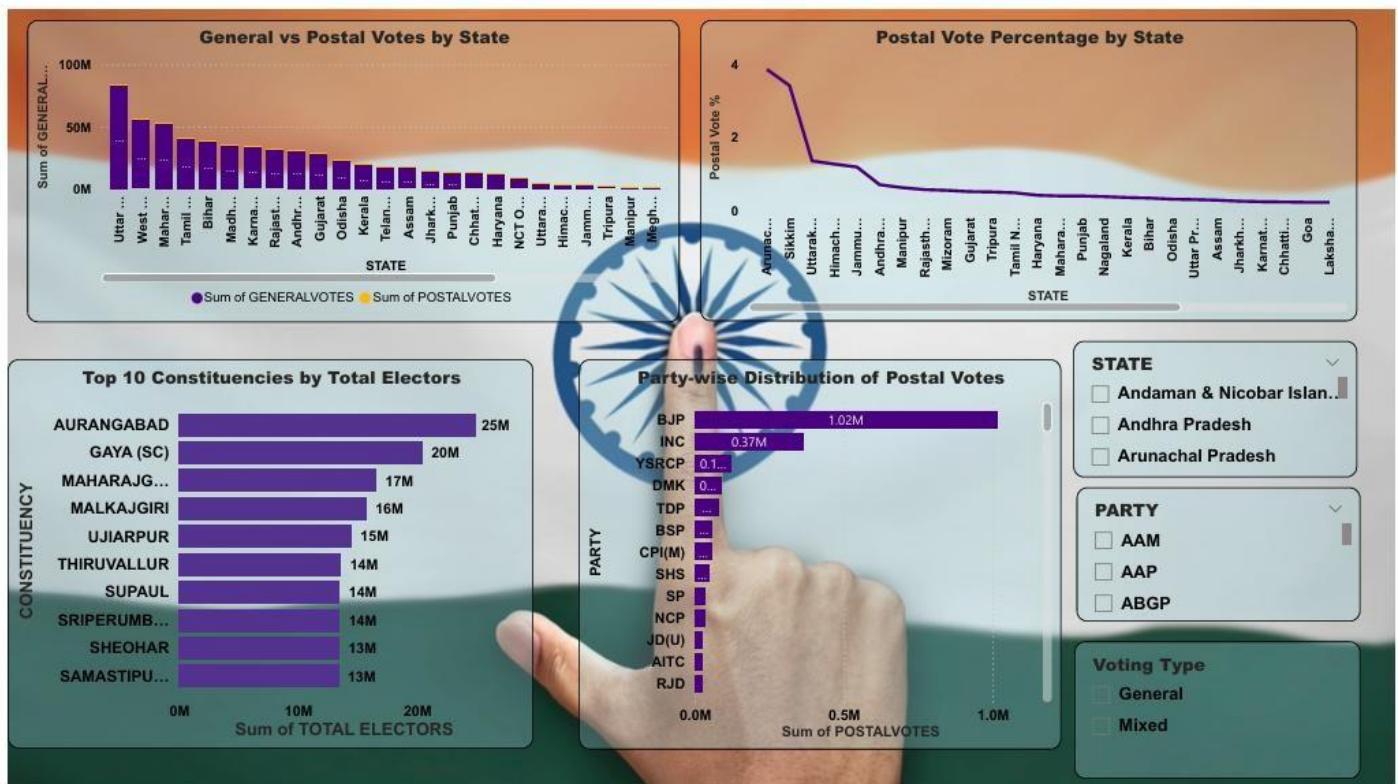
This slide analyses the economic and legal background of candidates.

- Average Assets:** ₹133.3 million
- Average Liabilities:** ₹19.97 million
- Average Criminal Cases:** 1.3 per candidate
- Top 10 Richest Winners:** Mostly from Tamil Nadu and Maharashtra, with assets ranging from ₹3 billion to ₹9 billion.
- Total Criminal Cases by Party:** BJP and INC show the highest totals due to larger candidate counts.
- Correlation Chart:** Demonstrates a mild positive trend – richer candidates often have more cases filed, possibly reflecting greater political visibility.

➤ **Overall Purpose of this Slide:**

To highlight transparency and socio-economic backgrounds of leaders.

4) Slide 4



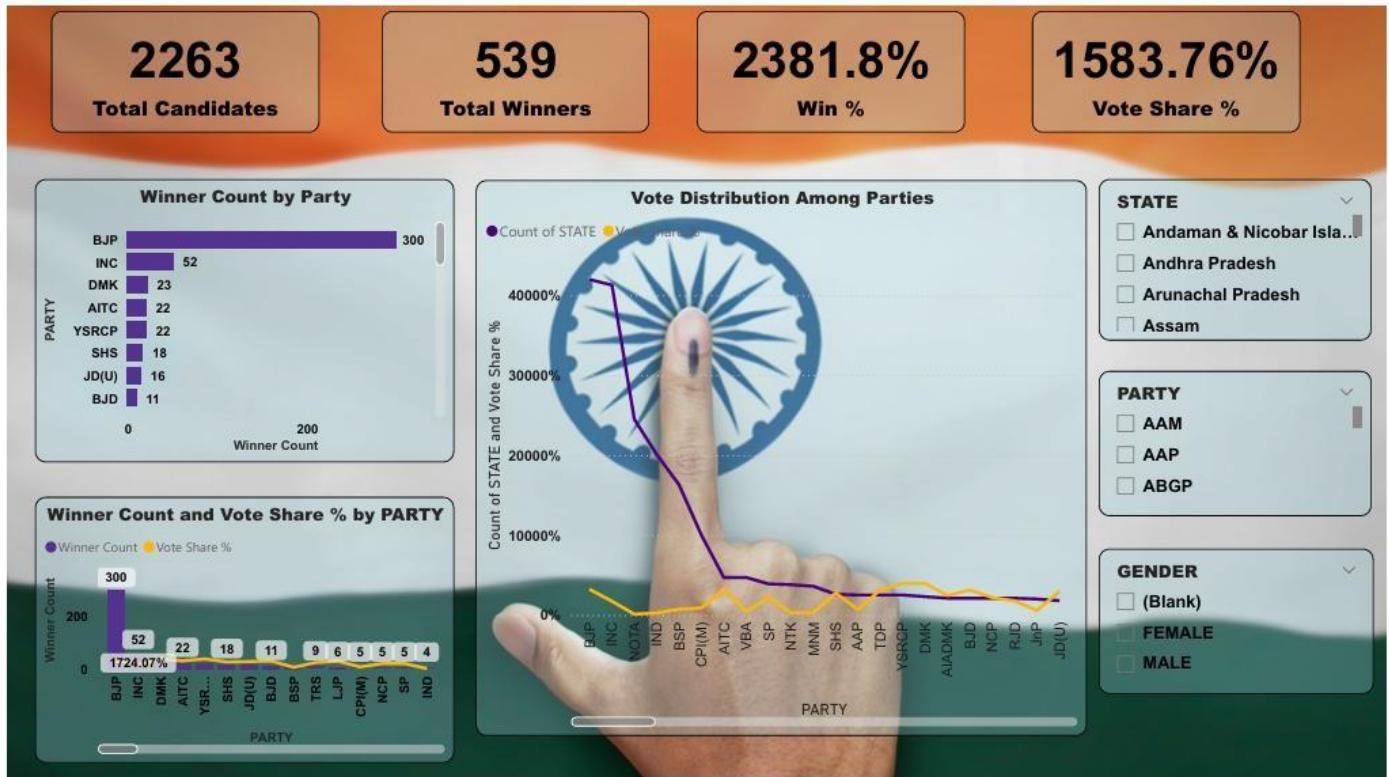
This slide explores how voting occurred across India.

- General vs Postal Votes:** General votes overwhelmingly dominate, while postal votes are under 5 % in most states.
- States with Highest Postal Voting %:** Arunachal Pradesh, Sikkim, Uttarakhand, and Himachal Pradesh.
- Top 10 Constituencies by Electors:** Include Aurangabad, Gaya (SC), and Malkajgiri — each exceeding 10 million voters.
- Party-wise Postal Votes:** BJP leads, followed by INC and YSRCP.

➤ **Overall Purpose of this Slide:**

To understand voter engagement levels and logistical variations across regions.

5) Slide 5



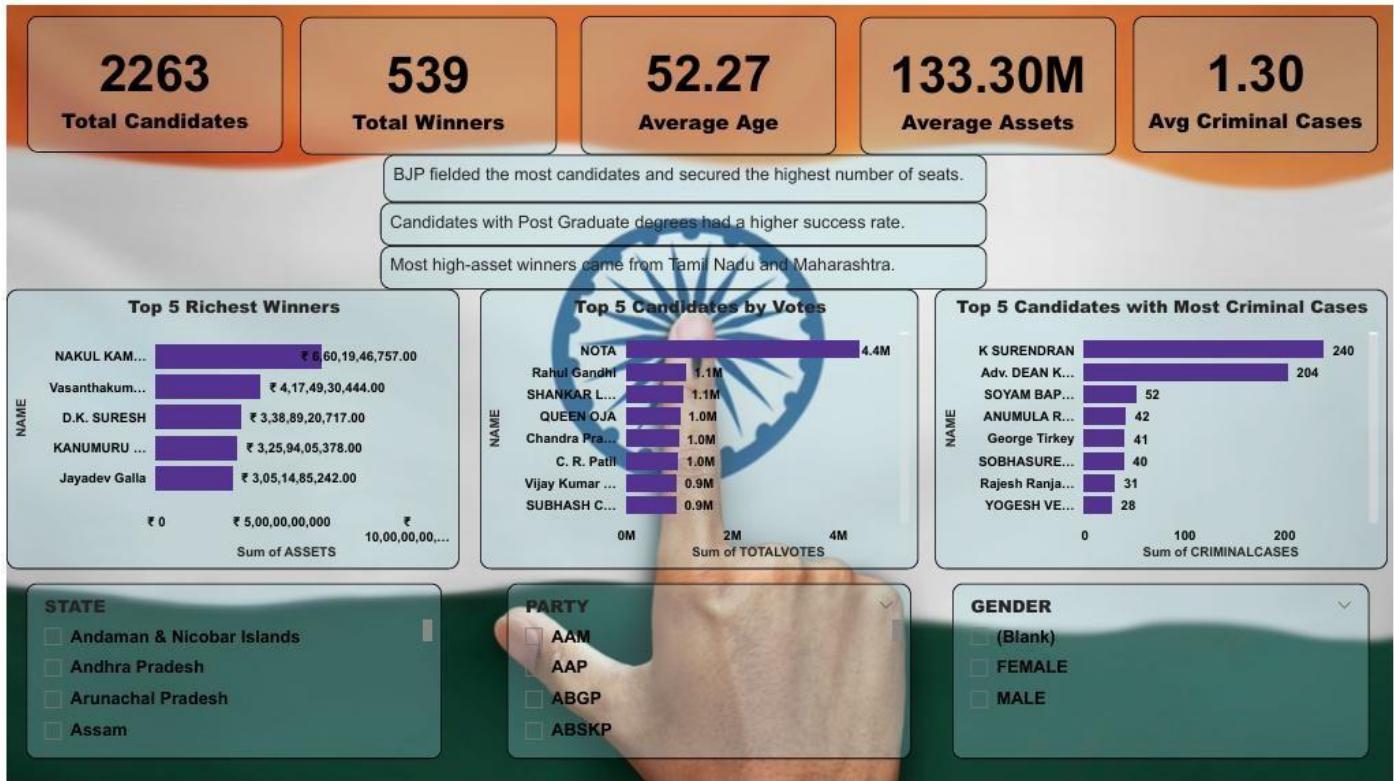
This slide compares party strength nationwide.

1. **Win Rate:** $\approx 23.8\%$ (539 winners out of 2263 candidates).
2. **Vote Share:** $\approx 15.8\%$.
3. **BJP:** Dominates both in number of seats and vote share.
4. **Other Strong Parties:** INC, DMK, AITC, YSRCP, and SHS.

➤ **Charts Included:**

- Winner Count by Party
- Vote Share % by Party
- Distribution of Votes Across States

6) Slide 6:



This slide presents a concise summary of major findings.

1. **BJP** has the highest success rate.
 2. **Higher education** (especially postgraduate degrees) is associated with better winning probability.
 3. **Wealth Concentration:** Most high-asset winners come from Tamil Nadu and Maharashtra.
 4. **Top 5 Richest Winners:** Include Nakul Kamal Nath and Vasanthakumar H., each exceeding ₹3 billion in declared assets.
 5. **Top 5 by Votes:** NOTA and Rahul Gandhi appear among the highest vote tallies.
 6. **Top Criminal Cases:** K Surendran and Adv. Dean Kuriakose lead with 200+ cases each.
- **Overall Purpose of this Slide:**

To summarize the dashboard's overall insights and help policymakers or journalists quickly grasp key trends.

PROS AND CONS SECTION

❖ **PROS (ADVANTAGES):**

1. Instant visual understanding

Provides instant visual understanding of election results and trends.

2. Interpret data quickly:

Helps media and analysts interpret data quickly for live reporting.

3. Combines multiple insights:

Combines multiple insights in one interactive dashboard.

4. Professional design principles:

Uses professional design principles for clarity and presentation.

5. Easy comparisons:

Enables easy comparisons between years, parties, and constituencies.

❖ **CONS (LIMITATIONS):**

1. Static datasets:

The dashboard currently uses static datasets (no live data feed).

2. Accuracy of the source dataset.:

Dependent on the accuracy of the source dataset.

3. Large datasets:

Large datasets may cause slower dashboard loading times.

4. Limited to Power BI's visualization capabilities (custom coding is minimal)

5. Requires manual update for each new election.

FUTURE ENHANCEMENTS

- Integrate real-time data connections to update dashboards automatically as election results are announced.
- Add historical trend analysis showing performance changes of parties over multiple election years.
- Introduce AI-powered insights for anomaly detection and pattern recognition.
- Develop a mobile-friendly version of the dashboard for public access.
- Use cloud-based Power BI services to share dashboards across multiple users securely.
- Include regional demographic filters like gender, age group, or rural/urban segmentation (if data is available).
- Add export-to-report features for generating automatic summary PDFs.

CONCLUSIONS

The project “Electviz: Election Data Visualization for Media Report” successfully demonstrates how data visualization tools like Power BI can transform complex election data into simple, interactive, and meaningful insights.

The developed dashboard highlights total votes, turnout rates, party performance, and leading candidates — enabling media professionals and analysts to interpret results instantly.

By replacing traditional data tables with interactive visuals, the project makes election reporting faster, clearer, and more engaging.

Overall, this project proves the effectiveness of Business Intelligence (BI) tools in improving data presentation, understanding public trends, and promoting data-driven transparency in elections.

Learnings & Skills Gained

- Learned how to collect, clean, and structure large datasets for analysis.
- Developed the ability to design interactive dashboards using Power BI visuals.
- Gained experience in using DAX expressions for calculated metrics like total votes and vote percentages.
- Improved understanding of data storytelling and visualization principles.
- Learned how to connect Python, Excel, and Power BI for data preprocessing and reporting workflows.
- Strengthened knowledge of filtering, relationships, and Power Query transformations.
- Enhanced report presentation and data interpretation skills for professional-level analytics.
- Understood how to create dashboards that communicate insights clearly to both technical and non-technical users.

REFERENCES

1. **Election Commission of India (ECI)** – Official election result datasets and reports.
Source: <https://eci.gov.in>
2. **Microsoft Power BI Documentation** – Used for building data visualizations and dashboards.
Source: <https://learn.microsoft.com/en-us/power-bi/>
3. **Python Official Documentation** – For programming and implementation of data analysis and prediction models.
Source: <https://docs.python.org/3/>
4. **Jupyter Notebook** – For running Python code, data cleaning, and visualization experiments.
Source: <https://jupyter.org/>

