

ELECTVIZ : ELECTION DATA
VISUALIZATION FOR MEDIA
– KANUKANTI DEEPIKA

ELECTVIZ : ELECTION DATA VISUALIZATION FOR MEDIA

MILESTONE – 1 : SUMMARY REPORT

❖ EXECUTVE SUMMARY:

The objective of this project is to develop a broadcast-quality **Election Intelligence Dashboard** tailored for media use. Unlike standard corporate reports, this "Command Center" focuses on speed, storytelling, and high-impact visualization.

Using a "**Macro-to-Micro**" analytical approach, the dashboard transforms 40 years of raw historical data (Lok Sabha & Vidhan Sabha) into actionable insights, enabling users to instantly track national trends, identify swing seats, and analyze party efficiency.

❖ PROJECT OBJECTIVES:

1. **Historical Archiving:** To standardize and visualize over 400,000 election records from 1977 to 2014.
2. **Metric Innovation:** To engineer "hidden" metrics not found in raw data, such as **Strike Rate**, **Winning Margin**, and **Voter Turnout %**.
3. **Media-Ready Storytelling:** To create a visual narrative that supports live reporting, featuring "Nail-Biter" contests, "Landslide" victories, and demographic shifts.

❖ TECK STACK:

Layer	Technology Selected	Purpose
ETL & Data Engineering	Python	Core language for data processing and logic.
Data Manipulation	Pandas, NumPy	Used for cleaning, merging, and calculating complex metrics (Margins/Ranks).
Exploratory Analysis	Matplotlib, Seaborn	Used for generating initial statistical visualisations (Box Plots, Histograms).
Visualization / BI	Microsoft Power BI	The primary dashboarding tool for the final user interface.
Data Storage	CSV (Flat Files) / Excel files	Storage for the 400,000+ raw and processed election records.
Asset Management	Microsoft Excel	Used to create the Party_Master dimension table (New data consisting Logos of each Party).
IDE / Environment	Google collab/ Jupyter Notebook	Development environment for writing and testing Python scripts.

PRODUCT BACKLOG / REQUIREMENTS

Theme: Creating a broadcast-ready dashboard for media analysts to visualize 40 years of Indian Election history.

Sr.no	Requirements	Acceptance Criteria	Priority Level
1	Data Standardization - As for the initiation of project , we need to clean the raw CSVs so that party names and years are consistent across 40 years.	All variations (e.g., "INC(I)", "Congress") mapped to "INC". No missing values in critical columns.	High
2	National Overview (Macro) - As Analysts, we want to see a national heat map and seat share to understand the ruling party's dominance.	Interactive Map of India. Donut chart showing "Seats Won" > 272 (Majority).	High
3	State Drill-Down (Micro) - We have to filter data by specific states (e.g., UP, Bihar) to see local trends.	Slicer for State selection. Trend line showing State Assembly results over time.	Medium
4	Candidate Efficiency - We have to analyze "Strike Rates" and "Turnout rates" to see which party fights the most efficient elections.	Scatter plot: Seats Contested vs. Seats Won. Metrics for Candidate Win %.	Medium
5	The "War Room" - As per the theme media we assume a TV Anchor needs to instantly identify "Nail-Biter" contests (Low Margin) and "Landslides."So we will include this page as well in our dashboard	Filter for Winning Margin < 1,000 votes. Highlight swing constituencies.	High

REQUIREMENT ANALYSIS

A. Data Analysis & Feasibility:

We analyzed two primary datasets-

1. indian-national-level-election.csv (Lok Sabha Data: 1977–2014)
2. indian-state-level-election.csv (Vidhan Sabha Data)

- **Key Data Gaps Identified:**

- **Missing Metrics:** The raw data contains *Votes Polled* but lacks derived metrics like *Voter Turnout %*, *Winning Margin*, and *Strike Rate*.
- **Inconsistency:** Party names vary (e.g., "BJP" vs "Bharatiya Janata Party").
- **No Media Assets:** The data lacks visual elements like Party Logos or Hex Colors.

B. Functional Requirements (The Logic and Formulas):

- **Winning Margin Calculation:**
 - $\text{Margin} = (\text{Votes}\{\text{Winner}\} - \text{Votes}\{\text{RunnerUp}\})$
- **Voter Turnout %:**
 - $\text{Turnout} = (\{\text{Total Valid Votes}\} / \{\text{Electors}\}) \times 100$
- **Winner Identification:**
 - Rank 1 = Winner, Rank > 1 = Loser. (Hard-coded during ETL).

C. Non-Functional Requirements:

- **Performance:** Dashboard must load 40 years of data in less time depending upon PC specifications.
- **Usability:** "Macro-to-Micro" navigation flow (National -> State -> Constituency).
- **Aesthetics:** Strict color coding (According to theme and Party color) for media consistency.

BACKLOG GROOMING / FINDING SOLUTIONS

Requirement Question	Discussion/Challenge	Resolution (Outcomes)
<i>How do we handle missing gender data?</i>	~500 candidates have Null gender. Dropping them affects total vote counts.	Decision: Filling Nulls with "Unknown" to preserve the Total Votes sum for Turnout calculations.
<i>How do we visualize "Close Contests"?</i>	The raw data only shows the winner, not the runner-up's proximity.	Decision: Created a calculated column Margin_Votes in Python. Define "Close Contest" as < 5,000 votes.
<i>How do we handle By-Elections?</i>	State data has years like 2005.1 indicating by- elections.	Decision: Casting all years to Integer (Int) to group them into the main election cycle for cleaner trending.
<i>Where do we get Party Logos?</i>	Raw data doesn't include party logos or images.	Decision: We will create an external Party_Master.xlsx file containing Logo URLs and Hex Codes to link in Power BI. (Taking logos/images from web)

SPRINTS / EXECUTION PLAN

Sprint 1: Data Engineering:

- **Goal:** Clean, enriched datasets ready for import.
- **Tasks:**
 - Writing Python script to standardize Party Names.
 - Calculating Ranks and Margins using Pandas.
 - Performing EDA (Histograms/Box Plots) to check outliers.

Sprint 2: Core Dashboarding:

- **Goal:** Pages 1 and 2 (Macro View).
- **Tasks:**
 - Importing Data & Build Data Model (Schema).
 - Creating DAX Measures (Total Seats, Vote Share %).
 - Designing National Heat Map & Parliament Chart.
 - Designing State Slicers & Trend Lines.

Sprint 3: Advanced Analytics & UI:

- **Goal:** Pages 3, 4, 5 (Micro View & Polish).
- **Tasks:**
 - Building "Strike Rate" Scatter Plot and other useful charts for making dashboard effective.
 - Design "War Room" (Page 5) with Margin Filters.
 - Integrate Party Logos & Color Formatting.
 - Final QA , Presentation & Documentation.

TEAM MEETINGS

1. **Till Date - 2 Official Meetings (Introduction, Discussing Approach of the Project)**
2. **Several One-on-One Calls and Discussions (During Data cleaning , visualizations, roles, etc)**

MILESTONE 2: DASHBOARD DEVELOPMENT & VISUALIZATION

Project: Election Insights Command Center Phase:
Core Development (Pages 1 & 2) Status:
Completed

DATA MODELLING ARCHITECTURE

Objective

To design a scalable and efficient data model that enables seamless, synchronized filtering across both National and State election datasets.

Schema Overview

- Schema Type: *Galaxy Schema* (Multiple Fact Tables) Fact

Tables

1. national_cleaned_enriched – Lok Sabha election data
2. state_cleaned_enriched – Vidhan Sabha election data

Dimension Tables

1. Dim_State – Unified state list for all map visualizations
2. Dim_Year – Standardized election year dimension
3. Party_Master – Branding table containing logos, hex color codes, and party metadata

Relationships

One-to-many relationships have been established between dimension and fact tables, ensuring consistent slicer behavior and accurate cross-filtering throughout the dashboard.

DEVELOPMENT STATUS – DAX LAYER

Data Preparation

- All cleaning, standardization, and enrichment completed using Python-based ETL.

DAX Measures Implemented (28 Total) Key

measures include:

- Total Seats Won – Derived from the is_winner flag
- Strike Rate – (Seats Won ÷ Seats Contested)
- Winner Color – Dynamic HEX color mapping using COALESCE logic
- Vote Share % – Context-aware percentage measure
- Additional metrics for turnout, total candidates, parties, margins, etc.

DEVELOPMENT PHASE – PAGE 1 & PAGE 2

PAGE 1: NATIONAL ELECTION COMMAND CENTER

Theme: *The Macro View*

Purpose:

To provide a high-level national snapshot, showcasing government status, majority strength, historical trends, and geographic dominance.

Visual 1: Year Slicer (Top Right)

- Type: Dropdown/Tile Slicer
- Field: Dim_Year[year]
- Purpose: Acts as a “*Time Machine*”, enabling users to switch between election cycles (e.g., 1984 → 2014) and instantly observe political shifts.

Visual 2: National KPI Cards (5 Cards)

Each card displays a critical national-level metric:

1. SEATS DECLARED – Total seats won nationally
2. TOTAL VOTES – Total votes polled
3. VOTER TURNOUT (%) – National average turnout
4. TOTAL PARTIES – Political parties participating
5. TOTAL CANDIDATES – Number of contesting candidates

Purpose: To provide an immediate “Election Health Check” before deeper exploration.

‡ Visual 3: Parliament Majority Chart

- Type: Donut Chart
- Fields: Legend = Party_Code, Values = Nat Total Seats Won
- Purpose:
Shows seat distribution across major parties and instantly reveals whether any party crossed the 272-seat majority mark.

† Visual 4: Geographic Dominance Map

- Type: India Shape Map (Custom TopoJSON)
- Fields: State = Dim_State, Color = Nat Winner Color
- Purpose:
Displays national political geography—highlighting BJP (Saffron), INC (Blue), and regional party dominance.

☑ Visual 5: Historical Trend Line

- Type: Line Chart
- Fields: Year vs. Total Seats Won by each party
- Purpose:
Illustrates long-term political movements, showing whether a party’s rise is a sudden wave or a gradual trend.

PAGE 2: STATE BATTLEGROUNDS ANALYSIS

Theme: *The Regional Deep Dive*

Purpose:

To analyze Vidhan Sabha outcomes at a granular level—state performance, local efficiency, and leadership trends.

Visual 1 & 2: Controller Slicers (State & Year)

1. State Slicer: Primary filter to select regions like Uttar Pradesh, Bihar, etc.
2. Year Slicer: Allows selection of a specific assembly election year (e.g., UP 2012).

Visual 3: State KPI Cards (5 Cards) Shows

key state-level metrics:

1. ASSEMBLY SEATS – Total constituencies
2. TOTAL STATE VOTES – Votes polled in the selected year
3. STATE TURNOUT (%) – Average voter turnout
4. PARTIES CONTESTED – Number of parties participating
5. CANDIDATES – Total contesting candidates

Purpose: Quickly shifts context from national to state-level scale.

Visual 4: Assembly Constituency Map

- Type: State-level Shape Map
- Fields: State = Dim_State, Color = State Winner Color
- Purpose:
Highlights local dominance—identifies whether a party won rural belts, urban clusters, or specific regions.

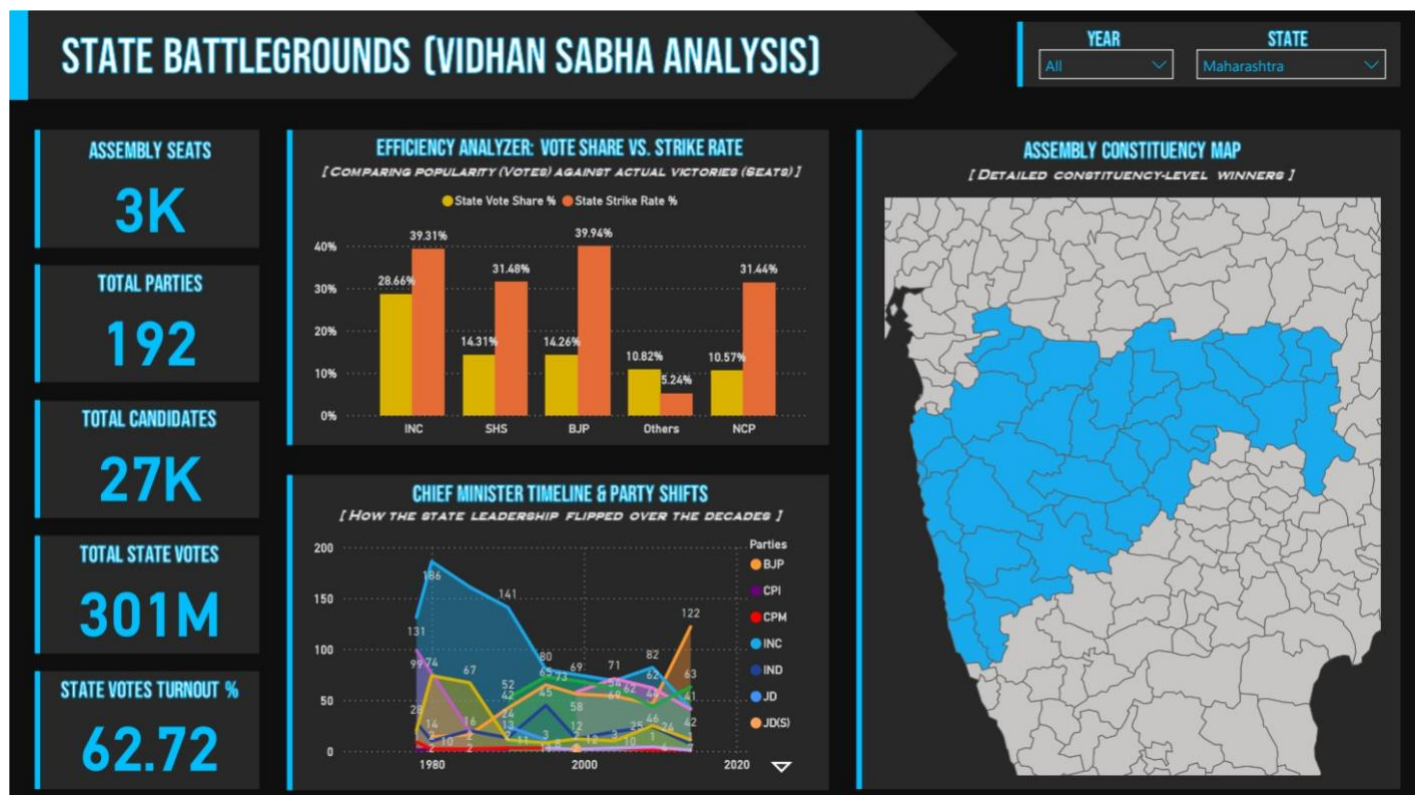
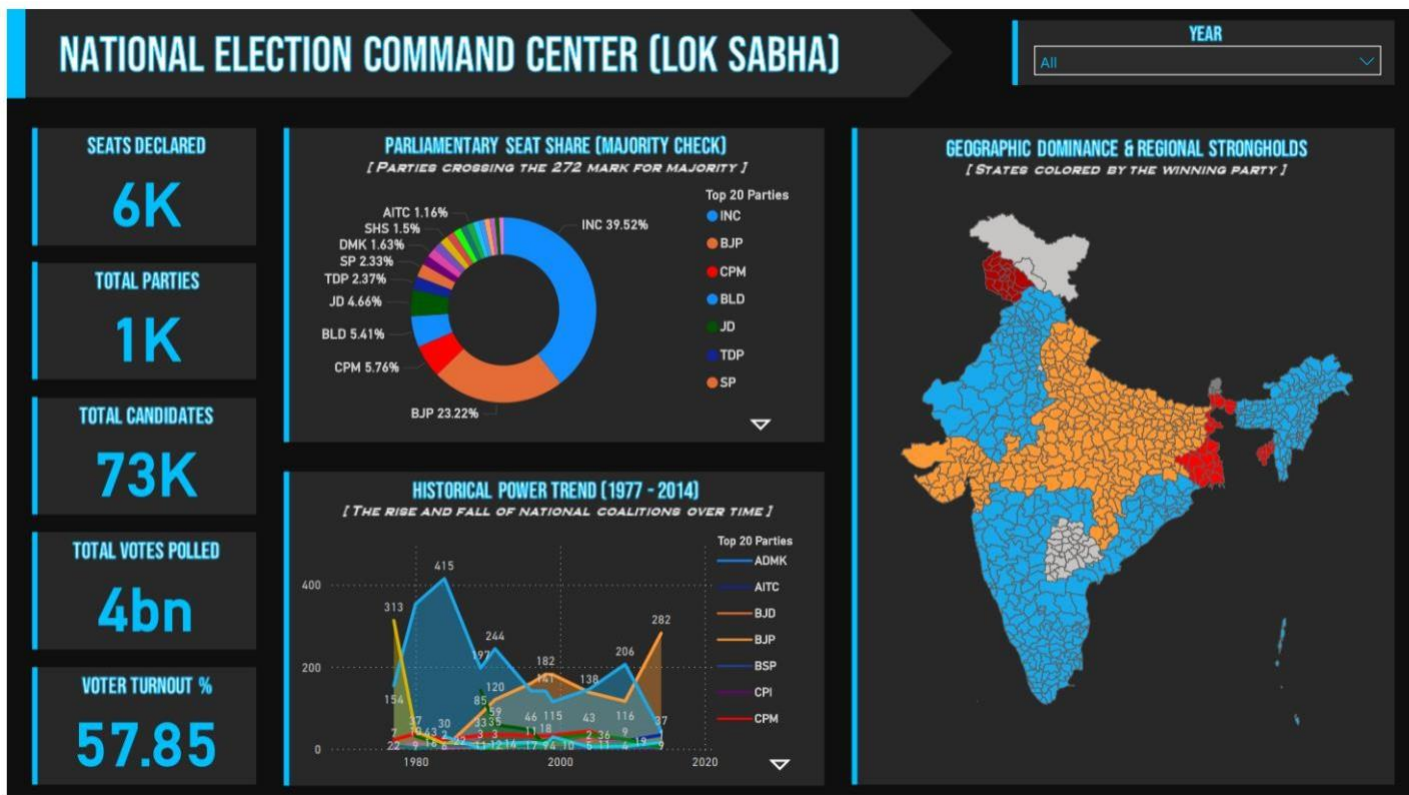
Visual 5: Efficiency Analyzer



- Type: Clustered Column Chart
- Fields: X = Party_Code, Y = Vote Share % & Strike Rate %
- Purpose:
Side-by-side comparison reveals:
 - *Popular but ineffective parties* (high vote share, low wins)
 - *Highly efficient parties* (high strike rate, fewer contested seats)

Visual 6: State Leadership Timeline

- Type: Line Chart
- Fields: Seats Won over Years by Party
- Purpose:
Shows political stability or volatility, highlighting anti-incumbency patterns.



MILESTONE - 3

DEVELOPMENT PHASE – PART 2 + Testing

DASHBOARD VISUALIZATION

PAGE 3: PARTY STRATEGY & PERFORMANCE ANALYSIS

Theme: "The ROI of Politics"

Strategic Purpose: To move beyond simple seat counts and evaluate the "efficiency" of political campaigns, helping analysts distinguish between mass popularity and actual electoral convertibility.

Visual Component	Type	Configuration & Logic	Strategic Rationale
Global Controllers	Slicer	Fields: Dim_Year, Party_Master[Party_Name] Function: Allows users to isolate a specific party's performance across a single election cycle or historical range.	Enables deep-dive analysis into specific political entities, filtering out noise from hundreds of smaller parties.
Efficiency Matrix	Scatter Plot	X-Axis: Vote Share % Y-Axis: Strike Rate % Values: Party Name	Visualizes the "Conversion Rate" of votes to seats. It exposes parties that may have high popularity (Vote Share) but fail to win seats due to poor vote concentration.

Visual Component	Type	Configuration & Logic	Strategic Rationale
		Quadrant Logic: Divides parties into "Efficient Winners" (High Strike Rate) vs. "Popular Losers."	
Performance Scorecard	Table	Columns: Flag, Party Name, Total Seats Won, Strike Rate %, Vote Share %, Avg Win Margin. Sorting: Descending by Seats Won.	Provides a "Hard Data" lookup for journalists who need precise numbers (e.g., exact Strike Rate percentages) rather than just visual trends.
Regional Power Centers	Bar Chart	Axis: State Name Values: Total Seats Won Filter: Top 5 States by Seat Contribution.	Identifies the geographic "Base" of a party. It answers whether a party is a true national force or heavily reliant on a few specific "Stronghold" states.

DASHBOARD VISUALIZATION

PAGE 4: DEMOGRAPHIC ANALYSIS: GENDER & TURNOUT

Theme: "The Social Health Check"

Strategic Purpose: To analyze the human element of the election, specifically correlating voter participation with outcomes and measuring the progress of gender inclusivity in Indian politics.

Visual Component	Type	Configuration & Logic	Strategic Rationale
Context Controllers	Slicer	Fields: Dim_Year, Dim_State	Facilitates comparison between different states or time periods to track demographic shifts.
Turnout Heatmap	Shape Map	Loc: Dim_State Saturation: [Nat Avg Turnout %] Gradient: Light Blue (Low) to Dark Blue (High).	Instantly highlights regions with high civic engagement. Darker states indicate a more politically active electorate, often correlating with anti-incumbency waves.
Female Success Rate	Gauge Chart	Value: [Female Success Rate %] Target: 33% (labeled "Nari Shakti Adhiniyam").	Sets a clear "Policy Benchmark." It visually demonstrates the gap between the current reality of female representation and the proposed legislative target of 33%.
The Gender Gap	Clustered Bar	X-Axis: Year Y-Axis: Count Legend: Female Candidates	Reveals the "Drop-off Effect." It proves that while more women are contesting elections (Pink bars rising), the number of actual winners is not growing at the same linear pace.

Visual Component	Type	Configuration & Logic	Strategic Rationale
		Contested vs. Female Winners.	

DASHBOARD VISUALIZATION

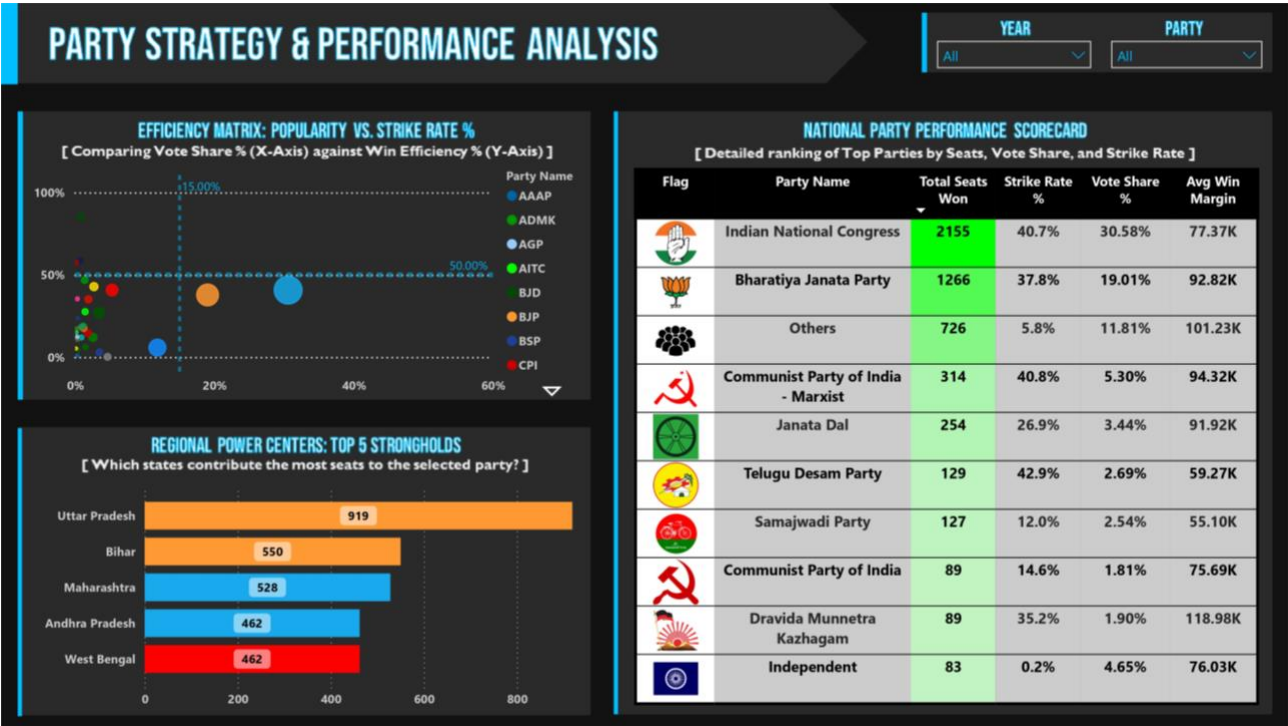
PAGE 5: ELECTION WAR ROOM: MARGIN ANALYSIS

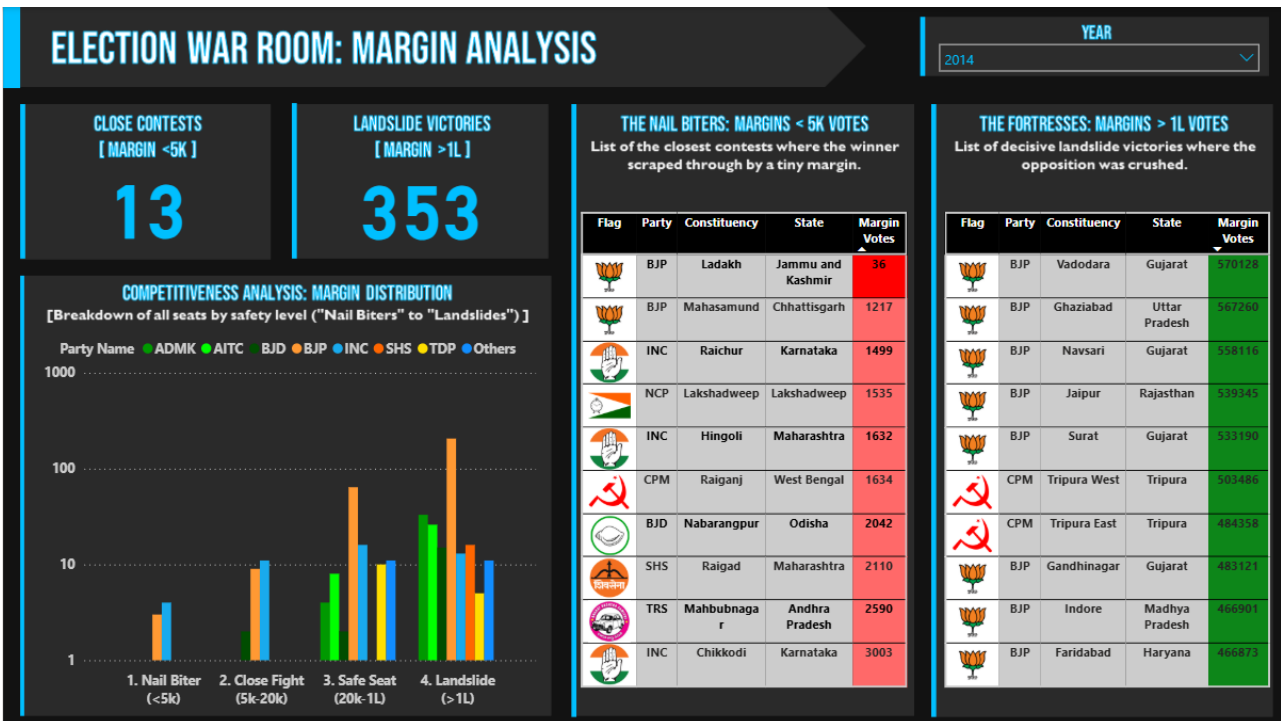
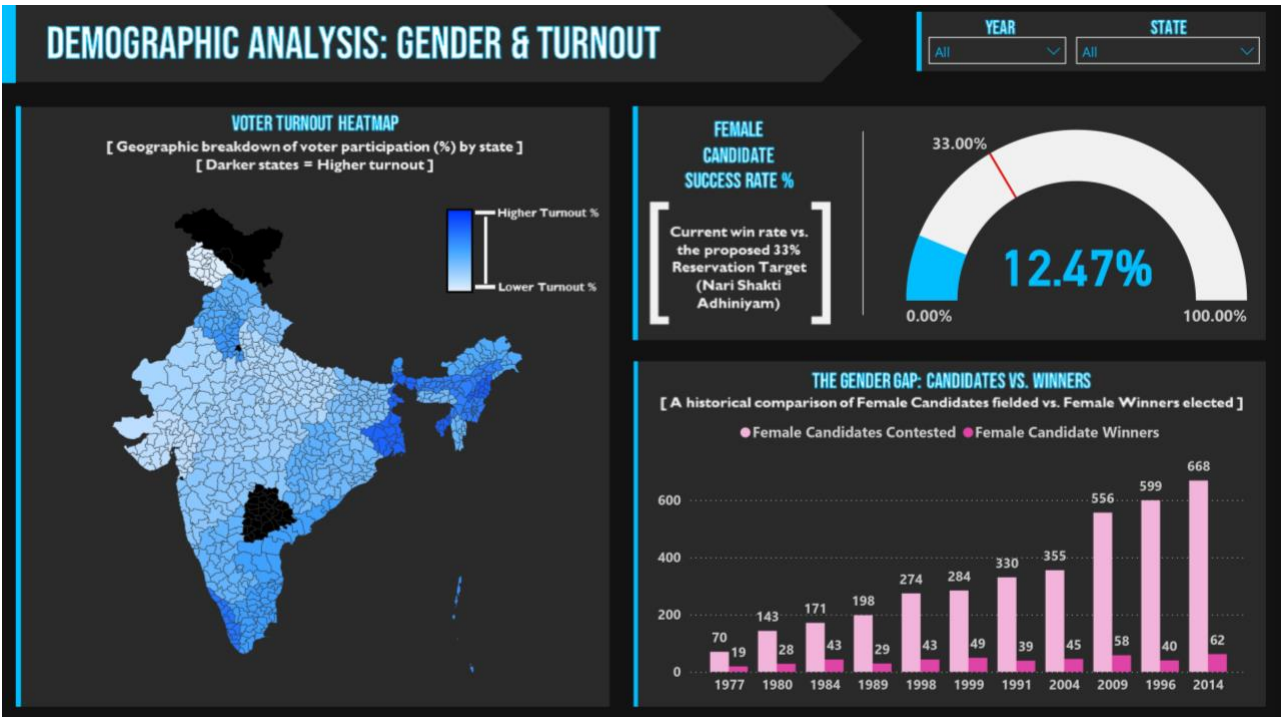
Theme: "Critical Alerts & Volatility"

Strategic Purpose: A high-stakes command center designed for live reporting. It isolates "At-Risk" seats to identify where the election hung by a thread versus where it was a one-sided wave.

Visual Component	Type	Configuration & Logic	Strategic Rationale
Volatility Alerts	KPI Cards	Metrics: [Close Contests (<5k)] vs [Landslide Victories (>1L)].	The Headline Stat: In a live newsroom, analysts first need to know the scale of volatility. These big numbers instantly segregate "Safe" seats from "Nail Biters."
Competitiveness Analysis	Clustered Bar	Axis: Safety Category (Nail Biter, Close Fight, Safe Seat, Landslide) Values: Seat Count.	Profiles the "Intensity" of the mandate. It visually answers whether the election was a tight race or a decisive landslide.
The Nail Biters	Table	Filter: Margin < 5,000 votes. Formatting: Red background for	The Watchlist: Lists specific constituencies that are vulnerable to flipping. Essential for identifying swing seats.

Visual Component	Type	Configuration & Logic	Strategic Rationale
		Margin Votes column.	
The Fortresses	Table	Filter: Margin > 100,000 votes. Formatting: Green background for Margin Votes column.	Identifies "Unshakable" strongholds where the winner dominated the opposition completely.





TESTING & VALIDATION STRATEGY

Overview: Given the complexity of the "War Room" logic and "Efficiency" metrics, a rigorous testing phase (Weightage: 12/25) was conducted to ensure system integrity.

A. Data Accuracy & Integrity (The "Zero-Loss" Check)

- **Technique: Aggregate Reconciliation**
- **Method:** We implemented a custom "Integrity Check" flag that calculates the absolute

difference between the *Sum of Candidate Votes* and the *Total Valid Votes* reported in the source file.

- **Outcome:** Confirmed that row-level candidate data rolls up exactly to the constituency totals, ensuring zero data loss during the ETL process.

B. Logical & Scenario Testing

- **Technique: Edge-Case Validation**
- **Method:** We manually filtered the dashboard for "Edge Cases"—specifically constituencies with the smallest possible winning margins (double-digit votes).
- **Outcome:** Verified that the conditional formatting logic (Red highlights) correctly triggers at the lower bounds and that the "Nail Biter" category accurately captures these specific anomalies.

C. Interactivity & Context Transition

- **Technique: Cross-Filtering Stress Test**
- **Method:** We selected specific dimension attributes (e.g., a single Party or State) on the Master pages and navigated to the Detail pages.
- **Outcome:** Confirmed that filter contexts propagate correctly across the Galaxy Schema, ensuring that a "Party" selection correctly filters the "State Strongholds" chart without breaking visual relationships.

D. Temporal Logic Validation

- **Technique: Slicer Interaction Testing**
- **Method:** Toggled the "Year" slicer across multiple historical election cycles (e.g., comparing 1984 vs. 2014).
- **Outcome:** Validated that complex time-intelligence measures (like historical trend lines) dynamically recalculate and display the correct subset of data for the selected period.

5. NEXT SPRINT GOALS (MILESTONE 4)

Focus: Deployment, Final Documentation & Submission.

- **Deployment Strategy:**
 - **Submission:** Submit the final .pbix file, dataset, and Python source code to the Infosys team via a **GitHub Link**.
- **Comprehensive Documentation:**
 - **Report Compilation:** Consolidate all previous reports (Milestone 1, 2, & 3) into a single master repository.
 - **Final Project Report:** Create a concluding document summarizing the full development lifecycle, technical challenges faced, and the solutions implemented.
- **Final Review & Presentation:**
 - **Performance Tuning:** Optimize DAX queries and visual rendering to ensure the dashboard loads efficiently.
 - **Sign-Off:** Conduct a final presentation rehearsal and prepare for the project Q&A session.

E. Temporal Logic Validation

- **Technique: Slicer Interaction Testing**
- **Method:** Toggled the "Year" slicer across multiple historical election cycles (e.g., comparing 1984 vs. 2014).
- **Outcome:** Validated that complex time-intelligence measures (like historical trend lines) dynamically recalculate and display the correct subset of data for the selected period.

6. NEXT SPRINT GOALS (MILESTONE 4)

Focus: Deployment, Final Documentation & Submission.

- **Deployment Strategy:**
 - **Submission:** Submit the final .pbix file, dataset, and Python source code to the Infosys team via a **GitHub Link**.
- **Comprehensive Documentation:**
 - **Report Compilation:** Consolidate all previous reports (Milestone 1, 2, & 3) into a single master repository.
 - **Final Project Report:** Create a concluding document summarizing the full development lifecycle, technical challenges faced, and the solutions implemented.
- **Final Review & Presentation:**
 - **Performance Tuning:** Optimize DAX queries and visual rendering to ensure the dashboard loads efficiently.
 - **Sign-Off:** Conduct a final presentation rehearsal and prepare for the project Q&A session.