ElectViz : Election Data Visualization for Media – Team C

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:**
  - Margin = (Votes{Winner} - Votes{RunnerUp})
- **Voter Turnout %:**
  - Turnout = ({Total Valid Votes}/ {Electors}) x 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)
How do we handle missing gender data?	~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.
How do we visualize "Close Contests"?	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.
How do we handle By-Elections?	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.
Where do we get Party Logos?	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)**