MILESTONE – 1
ElectViz: Election Data Visualization for Media

Problem statement:

ElectViz is an interactive data visualization project designed to analyze, interpret, and present election data in a clear and engaging way. Developed using Power BI, this project transforms large-scale election datasets into intuitive dashboards that help media organizations, analysts, and the public gain meaningful insights into voting patterns, party performance, and regional election dynamics.

The project demonstrates how data analytics and visualization can enhance political reporting by offering fact-based, data-driven stories. By simplifying complex datasets into visual narratives, ElectViz empowers users to make informed interpretations of election outcomes.

Dataset:

Indian Election Dataset

Resource:

Kaggle :- [Indian Election Dataset](#)

About dataset :-

This database contains detailed candidate-level data for elections to the lower houses of India's national and state legislatures, i.e., the Lok Sabha and Vidhan Sabhas. The data span 1977-2015, with each row representing a candidate that ran for office in that state-year.

The variables in the national election data file are:

Variable Name	Storage Type	Data Type	Description/Contents
st_name	str35	String	The name of the State or Union Territory.
Year	Int	Integer	The year the General Election (Lok Sabha) was held.
pc_no	Byte	Integer	Unique numerical identifier for the Parliamentary Constituency.
pc_name	str25	String	The official name of the Parliamentary Constituency.
pc_type	str3	String	Reservation status: **GEN** (General), **SC**, or **ST**.
cand_name	str70	String	The full name of the contesting candidate.
cand_sex	str1	String	Candidate's gender (M, F, or O).
partyname	str57	String	The full official name of the political party.
partyabbre	str10	String	Short-form abbreviation of the party (e.g., BJP, INC).
totvotpoll	Long	Numeric	Total number of valid votes the candidate received.
electors	Long	Numeric	Total number of registered voters in that constituency.

The variables in the state election data file are:

Variable Name	Storage Type	Data Type	Description/Contents
st_name	str35	String	The name of the State where the election was held.
Year	float	Numeric	The year of the Assembly Election (Vidhan Sabha).
ac_no	int	Integer	Unique identifier for the Assembly Constituency within the state.
ac_name	str28	String	The official name of the Assembly Constituency.
ac_type	str5	String	Reservation status: **GEN**, **SC**, or **ST**.
cand_name	str98	String	The full name of the candidate contesting the seat.
cand_sex	str1	String	Candidate's gender (M, F, or O).
partyname	str60	String	The full official name of the political party.
partyabbre	str11	String	Short-form abbreviation of the party.
totvotpoll	long	Numeric	Total valid votes received by the candidate.
electors	long	Numeric	Total number of registered voters in that AC.

Data Preprocessing & Cleaning Report: ElectViz

1. Data Standardization (Type Casting):

To ensure mathematical consistency across National and State datasets, all numeric fields were unified.

- Converted Year, totvotpoll, and electors to **Whole Number** (Int64).

- The State dataset initially held Year as a float. Standardizing to Integer is required for joining datasets and performing time-series analysis.

2. Text Normalization & Sanitization:

Election data often contains "dirty" text due to manual entry at the booth level.

- Applied Trim and Clean functions to cand_name, partyname, and st_name.

- Applied Capitalize Each Word to state and constituency names.

- Removes leading/trailing spaces and non-printable characters that cause "Duplicate" entries in visualizations (e.g., treating "BJP " and "BJP" as different entities).

3. Handling Missing Values (Imputation)

- Replaced null values in pc_type and ac_type with the string "GEN".

- In ECI data, General (non-reserved) seats are often left blank. Replacing them ensures filters and slicers in Power BI function correctly without an "Empty" category.

4. Feature Engineering: Unique Constituency ID (UID)

- Created a composite column Constituency_UID by merging [st_name] and [ac_name/pc_name].

- Prevents data collision. For example, it distinguishes between "Aurangabad" in Bihar and "Aurangabad" in Maharashtra, ensuring geo-spatial accuracy.

5. Party Categorization & Binning

- Implemented conditional logic to group hundreds of minor parties into an "Others" category while preserving major National and Regional parties.

- Reduces "visual noise" in charts. A pie chart with 50 parties is unreadable; grouping ensures the focus remains on key political shifts.

6. Electoral Logic: Winner Identification

- Grouped data by Year and Constituency, then applied a ranking transformation based on totvotpoll.

- Since the raw data is candidate-level, we needed a boolean "Is_Winner" flag to calculate seat counts and color-code the maps on the dashboard.

Milestone-2

1. Election Overview Page

1.1 Visuals Used and Their Purpose

- **Card Visual – Count of Years**

  - Displays the total number of years across which the election data varies.

  - Used to give users an immediate understanding of the temporal span of the dataset.

  - Cards are ideal for showing single, high-level KPIs clearly and prominently.

- **Card Visual – Count of States**

  - Represents the total number of states included in the dataset.

  - Helps users quickly understand the geographical coverage of the election data.

  - Cards are used here for simplicity and instant readability.

- **Card Visual – Count of Constituencies**

  - Shows the total number of constituencies present in the dataset.

  - Provides insight into the scale and granularity of the election data.

  - Cards are effective for highlighting key summary statistics.

- **Pie Chart – Distribution of Election Types (General, SC, ST)**

  - Illustrates the proportion of different election types.

  - Pie charts are suitable for showing percentage-based distribution of categories.

  - Helps users visually compare the share of General, SC, and ST constituencies.

- **Column Chart – Total Votes Polled by Year**
  - Displays the sum of total votes polled for each election year.
  - Column charts are effective for comparing numerical values across discrete time periods.
  - Helps identify years with higher or lower voter participation.
- **Line Chart – Count of Candidates per Year**
  - Shows how the number of candidates has changed over time.
  - Line charts are ideal for trend analysis across years.
  - Helps observe growth or decline in political participation.
- **Donut Chart – Gender Distribution of Candidates**
  - Represents the ratio of male, female, and other candidates.
  - Donut charts provide a clearer center-focused view of category proportions.
  - Highlights gender diversity in elections in an intuitive manner.

1.2 Insights Generated from the Election Overview Page

- The dataset spans multiple election years, providing a comprehensive historical perspective.
- A wide number of states and constituencies are covered, indicating a large-scale dataset.
- General elections form the majority, while SC and ST elections represent smaller but significant portions.
- Total votes polled vary across years, reflecting changes in voter turnout and political engagement.
- The number of candidates fluctuates over time, showing shifts in electoral competition.
- Gender distribution indicates a strong dominance of male candidates, with relatively lower female and other gender representation.

2. Party Performance Analysis Page

2.1 Visuals Used and Their Purpose

- **Year Slicer**

  - Enables users to filter party performance for a specific election year.

  - Helps analyze temporal changes in party influence.

- **Party Name Slicer**

  - Allows selection of one or multiple political parties.

  - Supports detailed party-wise analysis.

- **Card Visual – Count of Parties**

  - Displays the number of parties participating in the selected context.

  - Helps assess political diversity in elections.

- **Card Visual – Total Votes Polled**

  - Shows total votes secured by the selected party or parties.

  - Acts as a key indicator of party popularity.

- **Card Visual – Constituencies Covered**

  - Displays the number of constituencies contested by the party.

  - Helps evaluate the geographical reach of a party.

- **Card Visual – Number of Candidates**

  - Represents the total candidates fielded by the party.

  - Indicates the party's level of participation.

- **Bar Chart – Top 5 Parties by Total Votes Polled**

  - Highlights the five most dominant parties based on vote share.

  - Bar charts are ideal for ranking and comparison.

- Helps quickly identify leading political parties.
- **Line Chart – Trend of Total Votes Polled by Year**
  - Shows how a party's vote count has changed over time.
  - Useful for analyzing growth, decline, or stability in party performance.
  - Supports long-term performance evaluation.

2.2 Insights Generated from the Party Performance Analysis Page

- A small number of parties dominate the total votes polled.
- Party performance varies significantly across different years.
- Some parties show consistent growth, while others experience decline.
- Larger parties contest more constituencies and field more candidates.
- Year-wise and party-wise filtering allows detailed comparative analysis of political trends.

---

3. Distribution of Parliamentary Constituency Type

3.1 Visuals Used and Their Purpose

- **Year slicer:**
  - Allows users to filter the entire dashboard by election year.
- **Card Visual - Total no of state**
  - Represent the total no of states present in the dataset.
- **Card Visual - Total no of parliamentary constituency number**
  - Represent the total no of parliamentary constituency numbers present in the dataset.

- **Donut Chart – Distribution of Parliamentary Constituency Types**
  - This chart displays the distribution of constituencies by reservation type (GEN, SC, ST).
  - It highlights that General constituencies dominate, while SC and ST constituencies form a smaller but constitutionally significant share.

- **Clustered Column Chart – Candidate Participation by Constituency Type & Gender**
  - This chart compares the number of candidates by gender across different constituency types.
  - It shows strong male dominance in candidate participation across all constituency types, with relatively lower female representation, especially in SC and ST constituencies.

- **Pie Chart – Average Voter Turnout by Constituency Type**
  - This chart presents the average voter turnout percentage across constituency types.
  - It indicates that ST constituencies generally record higher turnout, followed by SC and General constituencies, reflecting varying voter engagement levels.

3.2 Insights generated from distribution of parliamentary constituency type page

- There is a higher share of average turnout% in ST and SC compared to General .

- The share of average turnout % keeps on increasing for all Parliamentary constituency types year by year.

- There is high dominance of General compared to SC , ST but shows significant increase year wise.

- Though there is male dominance in every participation type, female participation keeps on increasing year by year steadily.

- There is a decrease in no of candidates and average turnout % from 1996 to 1998.

- The period from 1996 to 1998 clearly shows instability in Indian politics .

MILESTONE – 3
POWER BI DASHBOARD

State level analysis :-

1. Data Preparation:-

Before building the visuals, ensure your dataset (likely a CSV or SQL database) contains the following columns:

- Geography: State Name, Constituency Name (pc_name).
- Electoral Data: Total Voters, Actual Votes Polled (totvotpoll), Number of Candidates.
- Time: Year of the election.
- Categorical: Party Abbreviations (partyabbre).
- Calculated Measures: Voter Turnout % (usually $\frac{\text{Actual Votes}}{\text{Eligible Voters}} \times 100$).

2. Page Layout & Theme:-

Background: Set a light blue background for the left sidebar and a very light gray or white for the main canvas.

Sidebar: Place a text box on the left with the title "State Level Analysis".

Styling: Use a consistent color palette of blues, grays, and whites. Apply a slight "Drop Shadow" to the white cards to give them a 3D effect.

3. Top Ribbon: KPI Cards

Create five Card Visuals and place them in a horizontal row at the top. Each should have a white background and a blue border.

1. Total Constituencies: Use a Distinct Count of the constituency name field. (Result: 55)
2. Eligible Voters: Use Sum of the total voters field. (Result: 6bn)
3. Avg Voter Turnout: Use the Average of your turnout percentage measure. (Result: 5.87%)
4. Total Candidates: Use a Count or Distinct Count of the candidate names. (Result: 4K)
5. Actual Votes: Use Sum of the total votes polled. (Result: 290M)

4. Left Sidebar: Interactive Filters

- Slicer: Add a Slicer visual. Drag the "State" field into it. Format it as a Dropdown and select "Andhra Pradesh."
- Map Visual: Add a Shape Map or Filled Map.
  - Location: State Name.
  - Color Saturation: You can use "Total Votes" to shade the map. Ensure the map is filtered to show the India sub-regions.

5. Main Content Visuals

A. Treemap (Sum of totvotpoll by party)

- Visual Type: Treemap.
- Group: partyabbre (Party Abbreviation).
- Values: Sum of totvotpoll.

- Formatting: Manually assign colors to major parties (e.g., Dark Blue for INC, Blue for TDP, Yellow for IND).

B. Combination Chart (Trends by Year)

- Visual Type: Line and Stacked Column Chart.
- X-Axis: Year.
- Column Values (Bars): Sum of totvotpoll.
- Line Values (Line): Average of voters_turnout_%.
- Secondary Axis: Ensure the Line value is plotted on the secondary Y-axis (right side) to show the percentage scale.

C. Matrix Table (Constituency Breakdown)

- Visual Type: Matrix.
- Rows: pc_name (Constituency Name).
- Columns: Year (1977, 1980, 1984, etc.).
- Values: Average of voters_turnout_%.
- Formatting: Enable "Stepped layout" off to see the years clearly in columns and add a "Total" row at the bottom.

MILESTONE – 4
POWER BI DASHBOARD

Candidate performance analysis :-

1. Page Configuration & Theme

- Canvas Settings: Use a light gray background to make the white visual containers pop.
- Top Header: Add a Text Box spanning the top with the title "CANDIDATE PERFORMANCE ANALYSIS" in bold, uppercase font.
- Visual Styling: Apply a dark gray border (approx. 2-3 pt) to every container on the page to maintain the structured, professional look seen in the screenshot.

2. Top Section: Global Slicers

Create three Slicer visuals at the top to allow for granular filtering. Wrap each in a container with a title header.

- State name: Filter using the State field (e.g., set to "Andhra Pradesh").
- Year: Filter using the Year field (e.g., set to "2004").
- Party name: Filter using the partyabbre or Party Name field (e.g., set to "BJP").

3. Middle Section: Key Performance Indicators (KPIs)

Use three Card Visuals placed horizontally. Each card should have a dark gray title bar.

- Total Votes Polled: Sum of the votes for the selected filters. (Result: 3M)
- Total candidates: Distinct Count of the cand_name field. (Result: 9)

- Actual voters: This appears to be a percentage calculation (likely $\frac{\text{Total Votes Polled}}{\text{Total Electors}}$). (Result: 25.20%)

4. Bottom Section: Comparative Analysis Visuals

A. Horizontal Bar Chart (Individual Performance)

- Title: "Votes for each candidate"
- Y-Axis: cand_name (Candidate Name).
- X-Axis: Sum of totvotpoll.
- Formatting: Enable Data Labels so the vote counts (e.g., 0.42M) are visible on the bars. Sort the chart by Sum of totvotpoll in descending order.

B. Treemap (Party & Candidate Distribution)

- Title: "Total votes for each party and candidate"
- Group/Category: partyabbre (Party) and then cand_name (Candidate).
- Values: Sum of totvotpoll.
- Formatting: In this specific view, since a single party is selected, it shows the relative "weight" of each candidate's contribution to that party's total.

C. Scatter Plot (Trend Analysis)

- Title: "Total votes for each candidate according to yr"
- X-Axis: Year.
- Y-Axis: Sum of totvotpoll.
- Legend: cand_name.
- Formatting: This visual shows the distribution of votes across different candidates for a specific point in time. It helps identify "clusters" or outliers in performance.

Electoral Competitiveness & Winner Analysis :-

This page focuses on the performance of political parties and the strength of their mandates across various election years.

Key Visuals & Logic

- Slicer (Year): A dropdown filter that allows users to select specific election years to update the entire dashboard.
- KPI Cards:
  - Winning Party: Displays the name of the party with the most seats. Logic: partyname filtered where Winner = 1.
  - Seats Won: Shows the total seat count for the winner. Logic: Count of Winner filtered by 1.
- Sum of Votes by Party (Bar Chart): Ranks parties based on the total volume of votes polled (totvotpoll).
- Winner Strength % (Line Chart): Analyzes the percentage of electors who voted for the winning candidate over time.
- Seats Won by Abbreviation (Treemap): A proportional view of seat distribution across different party abbreviations (partyabbre).
- Victory Margin % (Column Chart): Tracks how closely contested elections were by year.

Gender Wise Analysis:-

This page provides a demographic breakdown of candidates and their performance at the polls.

Key Visuals & Logic

- Candidate Summary Cards:
  - Total Candidates: A distinct count of all cand_name entries.
  - Male & Female Candidates: Filtered counts of candidates based on the cand_sex column ("M" or "F").
- Gender Distribution (Donut Chart): Visualizes the ratio of Male to Female candidates in the total dataset.
- Votes Polled by Gender (Pie Chart): Compares the total sum of totvotpoll categorized by gender.
- Historical Participation (Line Chart): Tracks the growth or decline of male and female candidates over time. Logic: year on the X-axis and count of candidates on the Y-axis with cand_sex as the Legend.

- State-wise Analysis (Table): Provides a granular look at candidate counts and gender splits for every state (st_name).

Technical Specifications

- Dataset: Indian national-level election data (1977–2014).
- Visual Styling: Consistent use of visual borders, shadows, and a blue-themed header with the Indian national flag and parliament icons.
- Primary Measures:
  - $Total\ Candidates = \text{Count}(\text{cand\_name})$
  - $Seat\ Count = \text{Count}(\text{Winner})\ \text{where}\ \text{Winner} = 1$

Constituency level insights:-

1. Project Overview & Design Philosophy

The primary objective of this page is to provide a granular, constituency-level view of electoral outcomes, identifying individual winners and tracking historical voter engagement.

- Aesthetic Theme: The design follows a minimalist "grid" layout using a white canvas with light gray accents.
- Visual Consistency: Every visual element is contained within a white box with a thin black border to ensure high readability and a professional structure.
- Interactive Hierarchy: The report is structured to lead the user from high-level "Winner" results down to detailed candidate lists and historical trends.

2. Step-by-Step Implementation

Step 1: Defining Core Metrics (KPIs)

To provide an immediate answer to "who won?", two high-prominence cards were created at the top of the dashboard:

- Winner Identification: A dynamic measure was implemented to display the name of the candidate with the highest vote count based on the selected filters.
- Winning Party: A corresponding field shows the party abbreviation associated with the leading candidate.

Step 2: Implementing Interactive Controls (Slicers)

Two dropdown slicers were added to allow for multi-level data exploration:

- State Level Slicer: Uses the State field to narrow results to a specific region.
- Constituency Level Slicer: Uses the pc_name field.
  - *Technical Configuration:* This slicer is set to "Show only relevant values," ensuring that once a state is chosen, the list only displays valid constituencies for that region.

Step 3: Visualizing Data Insights

Three distinct visuals were configured to break down the electoral data:

1. Party Vote Distribution (Donut Chart):
   - Data: Uses partyabbre for the legend and Sum of totvotpoll for the values.
   - Formatting: Configured with detail labels to show both the absolute vote volume and the percentage share per party.
2. Historical Engagement (Area Chart):
   - Data: Tracks voters_turnout_% across the year axis.
   - Formatting: A light-blue shaded area was applied to visualize the rise and fall of voter participation over time.
3. Candidate Detailed List (Table):
   - Data: Includes cand_name, partyabbre, and Sum of totvotpoll.
   - Formatting: Sorted by vote count in descending order to prioritize the most relevant results.

3. Data Transformation & DAX Logic

As part of the development process in Power BI, specific attention was given to data integrity:

- Data Cleaning: Power Query was used to ensure totvotpoll and year were correctly typed as numeric values for accurate summation and axis sorting.
- Voter Turnout Logic: The turnout was calculated to ensure percentages correctly reflect the ratio of actual votes to eligible electors for each election cycle.