Project Title: Analysis of Lok Sabha Elections (India)

Overview / Objective

This project focuses on analysing historical data from **Lok Sabha elections**, the lower house of India's bicameral Parliament. With the rise in data-driven decision-making, this analysis helps to understand voting trends, party performance, demographic impact, and regional variations over different election years.

The primary objective is to draw meaningful insights from past election data using Power BI, and to present them visually in an interactive dashboard format. This analysis can assist political analysts, researchers, students, and even political parties in understanding historical performance and voter behaviour across Indian states.

Problem Statement

Indian general elections are a complex and large-scale democratic exercise. With thousands of constituencies and hundreds of political parties, analysing this data manually is tedious and inefficient.

Hence, I framed my problem statement as:

"How can past Lok Sabha election data be visualized and analysed to understand voting patterns, party dominance, and regional performance?"

This project seeks to use Power BI to simplify and explore this large dataset in an interactive and insightful way.

Dataset Description

The data was provided by the online edtech for analytics called “Codebasics”, it was a data challenge that they conduct every month. It contains constituency-wise results of multiple Lok Sabha elections (mainly from the years 2009, 2014, and 2019). The dataset includes over **5000+ rows** and relevant fields that help perform detailed analysis.

Key columns used in the analysis include:

- **Election Year**: The year in which the general election took place.
- **State**: Name of the state or union territory.
- **Constituency**: Name of the parliamentary constituency.
- **Candidate Name**: Name of the candidate who contested the election.
- **Party**: Political party to which the candidate belongs.
- **Votes Received**: Number of votes received by the candidate.
- **Position**: Position based on vote count (1 = winner).
- **Total Electors**: Number of registered voters in the constituency.
- **Turnout**: Total number of voters who cast their votes.
- **Margin**: Difference between the winner and the runner-up.
- **Gender**: Gender of the candidate (Male/Female/Others).
- **Age**: Age of the candidate (if available).

Data Cleaning & Preprocessing

The raw data was initially unstructured and required extensive cleaning. The following steps were taken in **Power Query Editor (Power BI)**:

- Removed null values and duplicates.
- Standardized party names (e.g., "BJP" vs "Bhartiya Janata Party").
- Created calculated columns for **vote share**, **voter turnout percentage**, and **winning margin percentage**.
- Converted election years and state names to proper formats for filtering.
- Grouped and aggregated data where necessary (e.g., party-wise seat count per year).

Once cleaned, the data was loaded into Power BI for visualization and analysis.

Data Visualization & Dashboard

An interactive dashboard was developed using **Power BI**, featuring a wide range of visuals including:

- **Bar Charts**: Showing top-performing parties by seats and vote share.
- **Slicers**: For filtering by year, party, state, gender, and more.
- **Tables and Cards**: Highlighting total seats won, vote margin, voter turnout, etc.

👉 *Refer to the Power BI dashboard screenshot in the project folder for a visual overview.*

Key Findings & Insights

Here are some key insights derived from the analysis:

- The **BJP and INC** have been the two dominant parties across the last three elections, though their performance varies widely by state.
- **Uttar Pradesh**, **Maharashtra**, and **West Bengal** contribute significantly to the overall seat count and are key battleground states.
- **Voter turnout** has steadily increased over the years, indicating stronger democratic participation.
- Many **regional parties** (like TMC, DMK, AAP, and BJD) have strongholds in their respective states and significantly influence the overall results.
- The **winning margin** tends to be narrow in urban and metro constituencies, whereas rural areas often see wider margins.
- **Female participation**, both in terms of candidacy and voter turnout, is on the rise, though still significantly lower than male counterparts.

Tools & Technologies Used

- **Power BI Desktop**:
  - Power Query for data transformation.
  - DAX for calculated columns and measures.
  - Custom visualizations for interactive exploration.
- **Excel** (for preliminary data checking and formatting).

Conclusion & Recommendations

Conclusion:

Analysing the Lok Sabha election data through Power BI enabled an interactive and scalable view of Indian electoral trends. It allows users to visualize which parties dominate where, how voter behaviour has evolved, and where regional shifts have occurred.

- **High voter turnout regions** should be studied to understand successful voter engagement strategies.
- **Emerging regional parties** should analyse their vote conversion rates to optimize candidate selection.
- **Women candidates** need greater representation—data shows an upward trend but still room for growth.
- The **BJP and INC** have remained the two dominant national parties across the last three general elections, although their performance significantly varies by state and election year.
- States like **Uttar Pradesh**, **Maharashtra**, and **West Bengal** play a decisive role in determining the majority, given their high number of parliamentary constituencies.
- **Voter turnout ratio** (Turnout ÷ Total Electors) has shown a **gradual increase** over the years:
  - In **2019**, the **average turnout across the country was around 67%**, the highest in recent decades.
  - **North-eastern states** like Nagaland and Mizoram consistently record high turnout ratios (above 75%), reflecting strong electoral participation.

    - In contrast, **urban canters** such as Mumbai, Delhi, and Bengaluru show **lower voter turnout ratios**, often falling below 55%, despite high literacy rates.
- **Regional parties** like **TMC (West Bengal)**, **DMK (Tamil Nadu)**, **BJD (Odisha)**, and **AAP (Delhi)** have emerged as strong contenders in their respective states, limiting the influence of national parties in those regions.
- **Winning margins** are typically tighter in **urban constituencies**, indicating more competitive elections, whereas **rural constituencies** often display wider margins.
- **Female participation**—both as candidates and voters—has been increasing steadily, although still underrepresented compared to male counterparts.
- The **vote share vs. seats won** analysis shows that in several cases, parties with a relatively small vote share have managed to secure more seats due to concentrated regional support.