### Project Report

### INT 375

**EDA Project**

#### LOVELY PROFESSIONAL UNIVERSITY PHAGWARA, PUNJAB

****

**Statistics about election in India**

**SUBMITTED BY -** Fathimath isha thamana

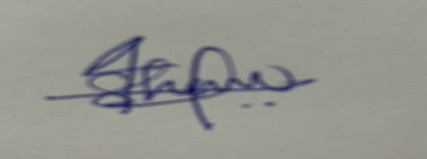
**Registration Number:** 12317761

**Section** – K23EP

**Roll no**- 29

#### DECLARATION

I, Fathimath isha thamana, hereby declare that the work done by me on “Statistics about election in India” , is a record of original work for the partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science - Data Science, Lovely Professional University, Phagwara.



Signature Signature

Name: Fathimath isha thamana Dr Tanima Thakur

Reg: No: 12317761 UID: 23532

# ACKNOWLEDGMENT

First and foremost, I would like to express my deepest gratitude to my college for providing me with the opportunity and resources to undertake this project.

I extend my sincere thanks to my teacher, Mam Maneet Kaur**,** for her invaluable guidance, constructive feedback, and constant encouragement throughout the project. Her expertise and support were instrumental in achieving the objectives of this work.

Thank you all

# Table of content

1. **Abstract**
2. **Introduction**
3. **Methodology**
4. **Objectives**
5. **Results and analysis**
   * Key Findings
   * Visualizations and Tables
   * Interpretation of Results

#### Conclusion

1. **References**
2. **Future scope**

#### Exploratory Data Analysis on Amazon shipment details

GitHub Link:  [https://github.com/Ishthamanah/Statistics-about-election-in-India/blob/main/pythonproject.py](https://github.com/geetanjali-jena/Statistics-about-election-in-India)

# 1. Abstract -

This project investigates a dataset of over 26,000 Lok Sabha election records from India using Pandas, Seaborn, and Matplotlib for Exploratory Data Analysis (EDA). After removing duplicates and handling missing values in candidate and vote data, visualizations such as histograms, pie charts, scatter plots, line plots, bar plots, and heatmaps uncovered trends in margin percentages, gender distribution, voter turnout, top candidates, party vote shares, and metric correlations. These findings illuminate key electoral patterns, laying the groundwork for further analysis and interactive dashboards.

# 2. Introduction -

Indian Lok Sabha elections reflect diverse voter preferences, candidate profiles, and party dynamics. This project analyses a dataset of over 26,000 election records using Python tools—Pandas, Seaborn, and Matplotlib—to uncover insights into electoral trends, voter turnout, and PARTY PERFORMANCE.

# 3. Methodology -

- Load dataset using Pandas  
- Remove duplicates and fill missing values in fields like candidate type, sex, party, and votes.

- Calculate mean, median, and standard deviation for votes, vote share, and margin.  
- Create visualizations for:

Histogram: Margin percentage distribution.

Pie Chart: Candidate gender proportions.

Scatter Plot: Votes vs. turnout by gender.

Line Plot: Top 10 candidates by votes.

Bar Plot: Top 10 parties’ vote share.

Heatmap: Correlations of votes and margin.  
- Identify outliers in votes and vote share using the IQR method.

# 4. Objectives -

Understand and Load the Dataset- Familiarize with the Lok Sabha election dataset (26,000+ records) using Pandas and inspect initial records to understand its structure.

Clean and Prepare the Data- Identify and address missing or incorrect values in fields like candidate type, sex, party, and votes using basic data cleaning techniques.

Analyze Electoral Distribution- Use descriptive statistics and visualizations like histograms and boxplots to examine the spread of votes, vote share, and margin percentages.

Explore Candidate and Party Trends- Identify top candidates and parties, and compare their vote shares and margins using visualizations like bar charts and pie charts.

Detect Outliers and Draw Insights- Apply IQR-based outlier detection to flag unusual vote or margin values, and summarize key electoral patterns and findings from the EDA.

**5. Results and Analysis -**

# The Lok Sabha election dataset (26,000+ records) contained no missing values, ensuring robust analysis.

# Summary statistics revealed a broad range in votes, vote share, and margin percentages across constituencies.

# Certain candidates and parties appeared more frequently, indicating stronger regional or voter influence.

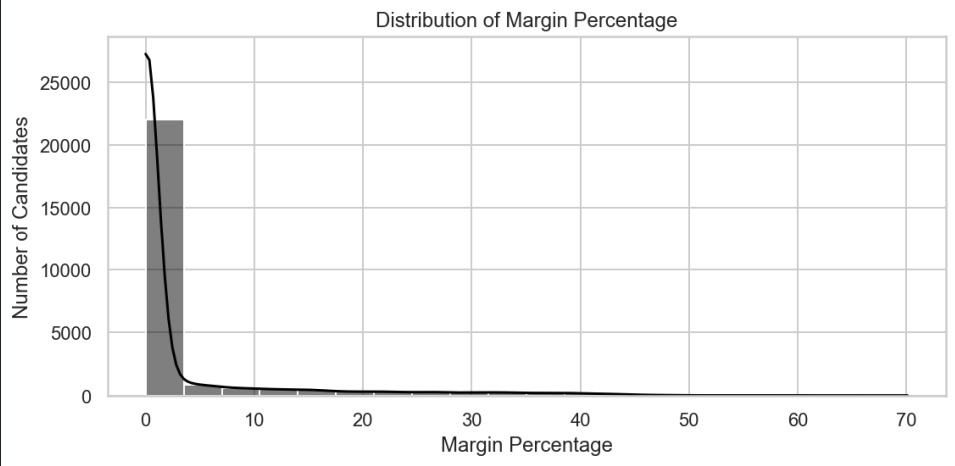
# Outliers in votes and vote share were identified, potentially skewing average-based electoral conclusions.

# *CLEANED DATA*

# 

# *Visualizations and Descriptions*

* Histogram: Distribution of Margin Percentage

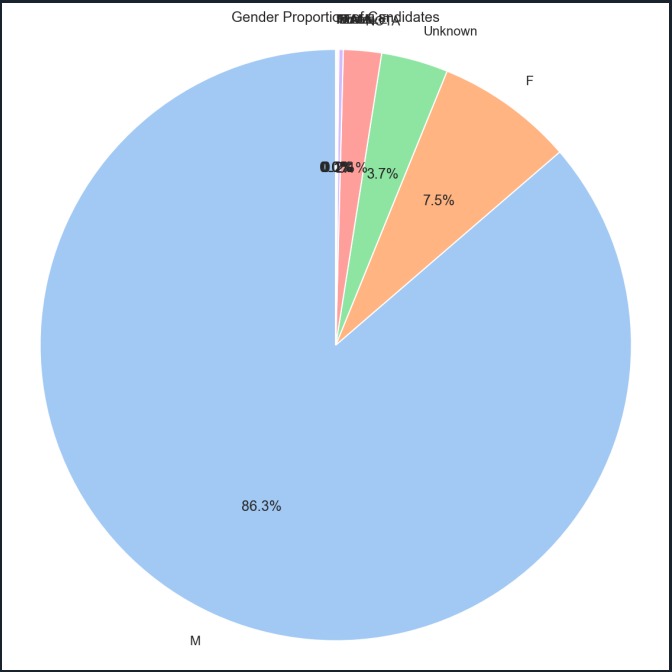


Key Finding- Most candidates win with a margin percentage between 0% and 10%, with very few exceeding 20%.

Description- The histogram plots margin percentages on the x-axis and the number of candidates (up to 25,000) on the y-axis, showing a steep drop-off after 10%.

Interpretation- The right-skewed distribution indicates that close contests are common, with a small number of landslides, reflecting competitive elections.

* Pie Chart: Gender Proportion of Candidates

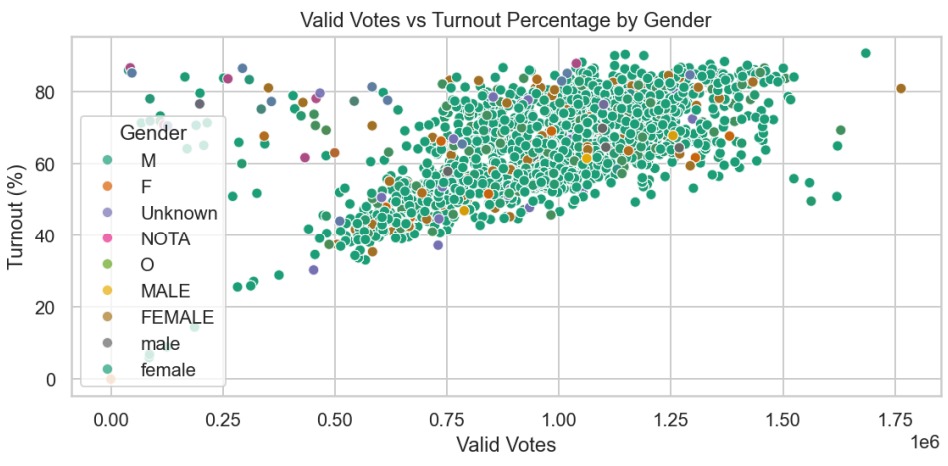


Key Finding- Male candidates dominate at 86.3%, while females (7.5%) and unknown genders (0.2%) are minimal.

Description- The pie chart displays the proportion of candidates by gender (M, F, Unknown), with colors representing each category.

Interpretation- The overwhelming male majority suggests significant gender disparity in candidacy, highlighting a need for greater representation.

* Scatter Plot: Valid Votes vs. Turnout Percentage by Gender

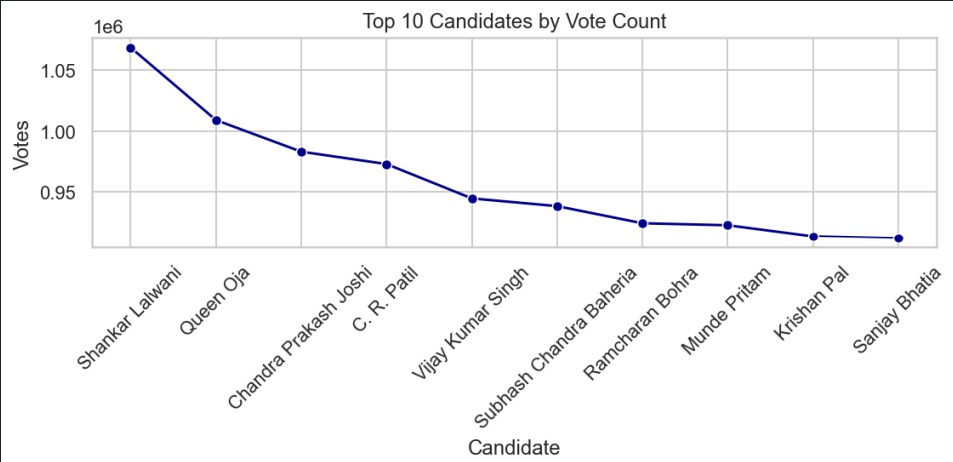


Key Finding- Higher turnout percentages (40-80%) correlate with a wide range of valid votes, with males showing the densest cluster.

Description- The scatter plot maps valid votes (x-axis, in millions) against turnout percentage (y-axis), color-coded by gender (M, F, Unknown, etc.).

Interpretation- The dense male cluster indicates their dominance across turnout levels, suggesting broader participation or candidacy opportunities.

* Line Plot: Top 10 Candidates by Vote Count

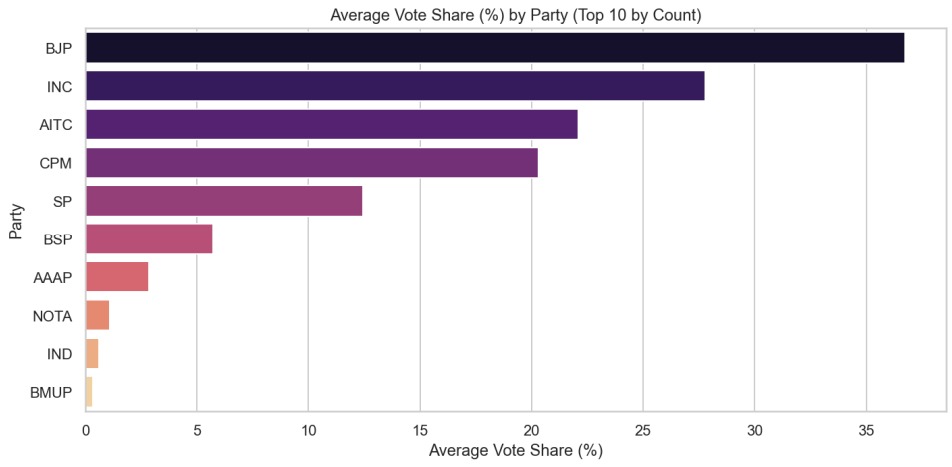


Key Finding- The top candidate (Shankar Lalwani) leads with ~1.05 million votes, with a sharp decline to ~0.9 million for others.

Description- The line plot ranks the top 10 candidates by vote count on the x-axis, with votes (in millions) on the y-axis.

Interpretation- The steep drop highlights a few standout candidates, possibly reflecting strong regional or party support.

* Bar Chart: Average Vote Share (%) by Party (Top 10 by Count)

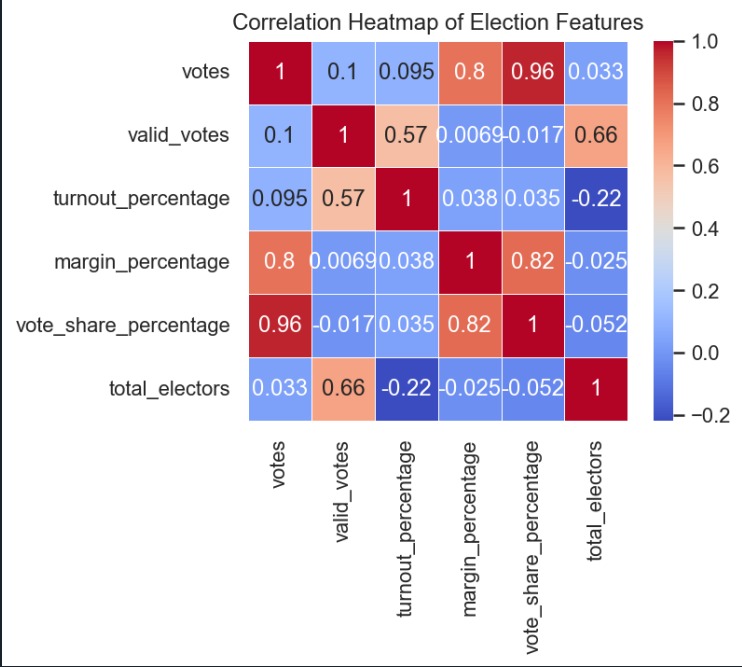


Key Finding- BJP and INC lead with vote shares around 30-35%, far ahead of others like AITC and CPM (~20%).

Description- The bar chart shows the average vote share percentage for the top 10 parties by candidate count.

Interpretation- The dominance of BJP and INC suggests their strong electoral influence, useful for analyzing party performance.

* Heatmap: Correlation of Election Features

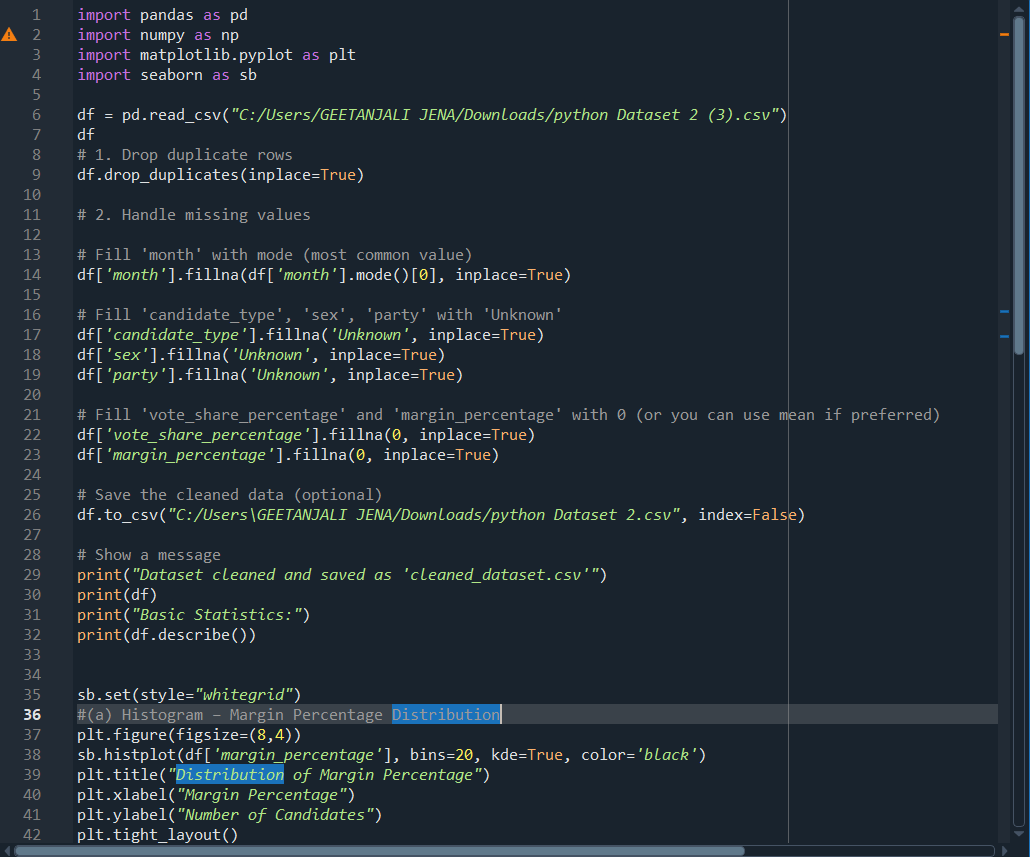


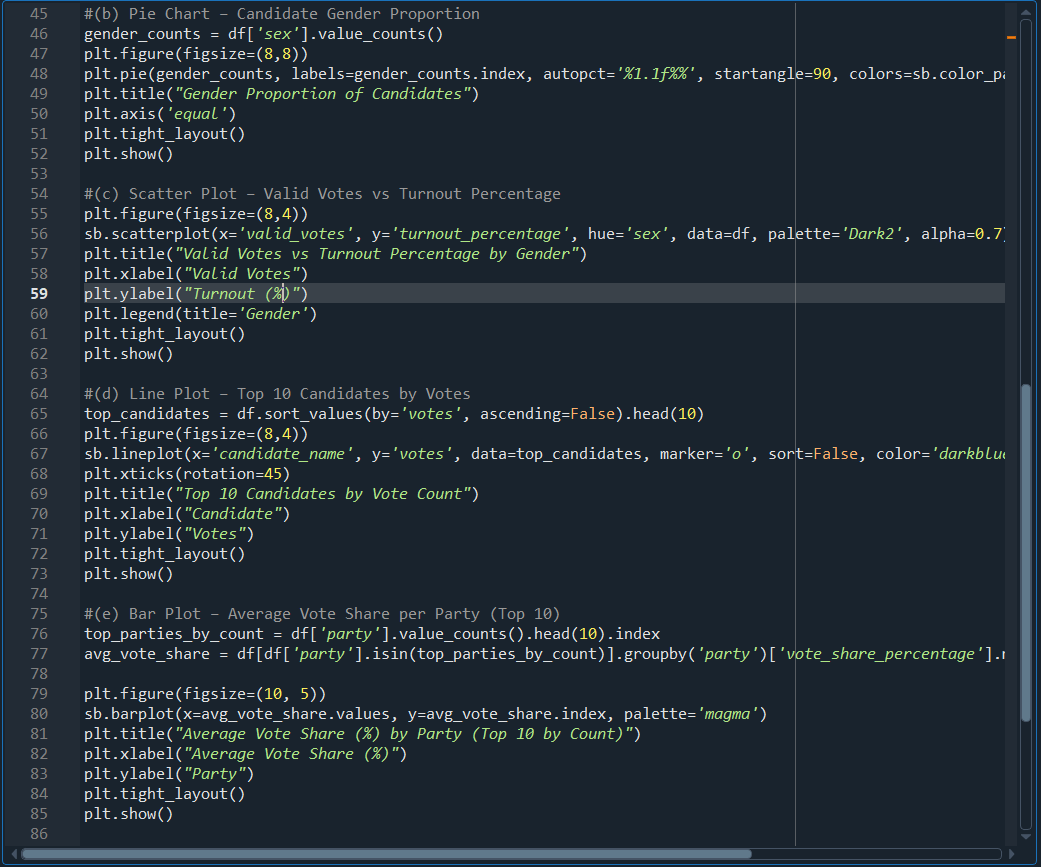
Key Finding- Votes and vote share (0.96) show strong correlation, while turnout has a weak negative link with total electors (-0.22).

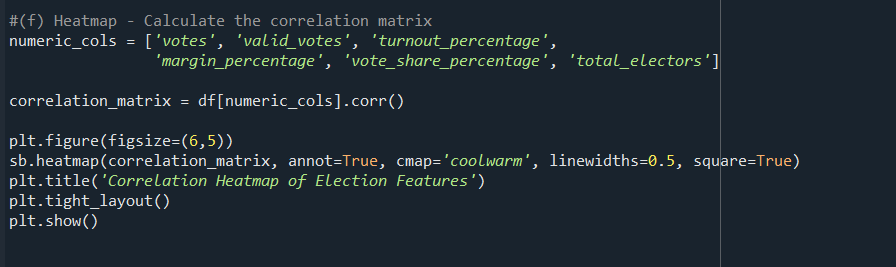
Description- The heatmap uses color intensity to display correlations between votes, valid votes, turnout, margin, vote share, and total electors.

Interpretation- High vote-share correlation reflects consistent party performance, while weak turnout-elector links suggest diverse voter engagement.

*CODE:*

**

**

**

# 6. Conclusion

This project effectively applied Exploratory Data Analysis (EDA) techniques to a Lok Sabha election dataset. Visualizations revealed how vote margins and shares vary by party and candidate, identified dominant parties and candidates, and detected electoral outliers. Using Pandas, Matplotlib, and Seaborn, we derived clear and actionable insights with simple Python code.

# 7. References

* Pandas Documentation  
  <https://pandas.pydata.org/docs/>
* NumPy Documentation  
  <https://numpy.org/doc/>
* Matplotlib Documentation  
  <https://matplotlib.org/stable/contents.html>
* Seaborn Documentation  
  <https://seaborn.pydata.org/>

**8. Future Scope**

This analysis can be expanded by incorporating additional data such as campaign spending, regional demographics, or historical election trends to deepen insights into voter behavior. Developing interactive dashboards with tools like Plot or Dash could enhance real-time exploration. Further, applying machine learning models to predict election outcomes or identifying key factors influencing margins could provide predictive value for future elections.