DATA SCIENCE TOOLBOX PYTHON PROGRAMMING

PROJECT REPORT

(Project Semester January-April 2025)

Participation of Independent Candidate during 2024 (Election Data)

Submitted by

NAME: K Hema Pulleswararao.

REGISTRATION NO: 12318165

PROGRAMME AND SECTION: K23GN24

COURSE CODE : INT375

Under the Guidance of

MRS. AASHIMA (UID: 28968)

Discipline of CSE/IT

Lovely School of Computer Science

Lovely Professional University, Phagwara

<u>CERTIFICATE</u>

This is to certify that **K.Hema Pulleswararao** bearing Registration no. **12318165** has completed **INT375** project titled, **"Mrs.Aashima"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor
Designation of the Supervisor
School of Computer Science
Lovely Professional University
Phagwara, Punjab.

DECLARATION

I, **K.Hema Pulleswararao** student of CSE (Program name) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Signature

Registration No: 12318165   Name of the student: **K.Hema Pulleswararao.**

Title:

Participation of Independent Candidate during 2024 (Election Data)

1. Introduction:

In democratic elections, independent candidates play a crucial role by offering voters alternatives to mainstream political parties. Their participation often reflects the diverse political opinions of the electorate, particularly in areas where party loyalty is weak or where local issues outweigh national agendas. The 2024 elections were no exception, witnessing a notable presence of independent candidates across various constituencies.

This report aims to explore and analyze the performance, distribution, and voter dynamics surrounding independent candidates using real election data. Through this analysis, we seek to answer important questions such as:

- **In which regions were independent candidates most active?**

- **How did their vote share compare with those from major parties?**

- **What demographic patterns (e.g., gender-based voting) can be linked to their performance?**

- **Are there any significant outliers or trends that point to growing support for independent candidates?**

By leveraging Python's powerful data analysis libraries such as pandas, matplotlib, and seaborn, we visualize key aspects of the data to extract meaningful insights. The report includes histograms, bar charts, scatter plots, and correlation matrices to provide a comprehensive understanding of independent participation in the 2024 election cycle.

2. Source of Dataset :

The dataset analyzed in this report originates from a CSV file titled **"ElectionData.csv"**, located on the system at:

makefile

CopyEdit

C:\Users\anand\Downloads\ElectionData.csv

This dataset appears to be an official or aggregated compilation of election statistics from the **2024 General Elections**, possibly sourced from government election commission portals or electoral analysis platforms. It contains structured data regarding:

- **Voter Demographics**:

  o Number of **men**, **women**, and **third gender** voters.

  o **Total** voter count per constituency or unit.

- **Candidate Information**:

  o **Category** of each candidate (e.g., Party-affiliated, Independent).

  o **Code Name** identifiers used for different candidates.

  o Vote counts attributed to each candidate or category.

- **District/Constituency Data**:

  o While the dataset may not include explicit names of constituencies, the vote distributions suggest a district-wise layout.

  o The "Total" column reflects overall participation, which can be compared across candidates.

Upon loading the dataset using Python's pandas library, encoding issues were encountered and handled using fallbacks (utf-8 and latin1). After successful import, data was cleaned and preprocessed, including handling of missing values.

This dataset serves as the foundation for our analytical exploration into the **participation and performance of Independent candidates** during the 2024 elections.

---

3. EDA Process (Exploratory Data Analysis)

Exploratory Data Analysis (EDA) is the critical first step in understanding the structure and quality of a dataset before performing deeper statistical or predictive analysis. For this election dataset, EDA was carried out using Python libraries such as pandas, numpy, seaborn, and matplotlib. The process involved several stages:

📌 **a) Data Loading and Cleaning**

- **The dataset was loaded from a local CSV file using pandas.read_csv().**

- **Encoding challenges were addressed by trying both 'utf-8' and 'latin1' formats.**

- **Basic data inspection was done using .info() and .describe() to understand data types, missing values, and numerical distributions.**

- **Missing values were filled with 0 for simplicity, ensuring all rows remained usable.**

📌 **b) Data Overview**

- **The dataset included demographic columns (Men, Women, Third Gender, Total) and classification fields (Category, Code Name).**

- **Categories indicated the type of candidates (e.g., Party-affiliated, Independent, NOTA).**

- **Summary statistics revealed variations in voter counts, with some extreme values (potential outliers).**

📌 **c) Visual Explorations**

Several types of plots were used to understand trends:

- **Histogram: Displayed the distribution of total votes across all units, highlighting the spread and concentration.**

- **Bar Chart: Compared total vote shares across different candidate categories, showing where Independent candidates stood.**

- **Pie Chart: Illustrated the proportion of candidates who contested from each category.**

- **Correlation Heatmap: Identified relationships between gender-specific vote counts and total votes.**

- **Boxplot: Highlighted outliers in the numerical columns, helping to spot unusual voting patterns.**

- **Pairplot: Explored inter-variable relationships, such as how Men and Women voter counts vary together.**

- **Line Graph: Showed the trend of total votes across data indices.**

- **Scatter Plot: Specifically explored the relationship between male and female voter counts.**

---

4. Analysis on Dataset

The dataset provided several actionable insights into the participation and performance of Independent candidates in the 2024 elections:

🗳 **1. Independent Candidate Participation**

- **Independent candidates were included under a unique "Category" label in the dataset.**

- **By filtering on this label, their individual and collective vote totals could be assessed.**

- **The number of Independent candidates was also revealed in the pie chart, giving a proportionate view of their presence.**

📊 **2. Performance Comparison**

- **The bar chart comparing categories showed that while party-affiliated candidates generally received the most votes, Independent candidates still drew substantial support in some regions.**

- **Certain data points showed Independent candidates winning or strongly competing in districts, hinting at local trust or protest votes.**

👥 **3. Voter Demographics**

- **Correlation heatmaps showed that total votes were highly correlated with both Men and Women counts.**

- **Scatter plots indicated that regions with a balanced male and female voter base had more stable total vote contributions — beneficial for Independent candidates in competitive races.**

🔍 **4. Outlier Districts**

- **Boxplots revealed that a few constituencies had exceptionally high or low vote counts.**

- **These may include regions where Independent candidates either dominated or severely underperformed — useful for targeted analysis.**

🔁 **5. Patterns and Trends**

- **Line and pair plots showed the distribution and consistency of votes across regions.**

- **These plots can be used to detect any rising or falling trends in Independent candidate support across the electoral landscape.**

4.1 Data Loading and Inspection

- **The dataset was loaded from a local CSV path using the pandas.read_csv() function.**

- **Encoding issues were handled using a fallback strategy (utf-8, then latin1).**

- **Initial inspection was performed using .info() and .describe() to understand data types, structure, and basic metadata.**

- **Output revealed the dataset contains multiple numerical columns such as Men, Women, Third Gender, and Total, and categorical columns like Category and Code Name.**

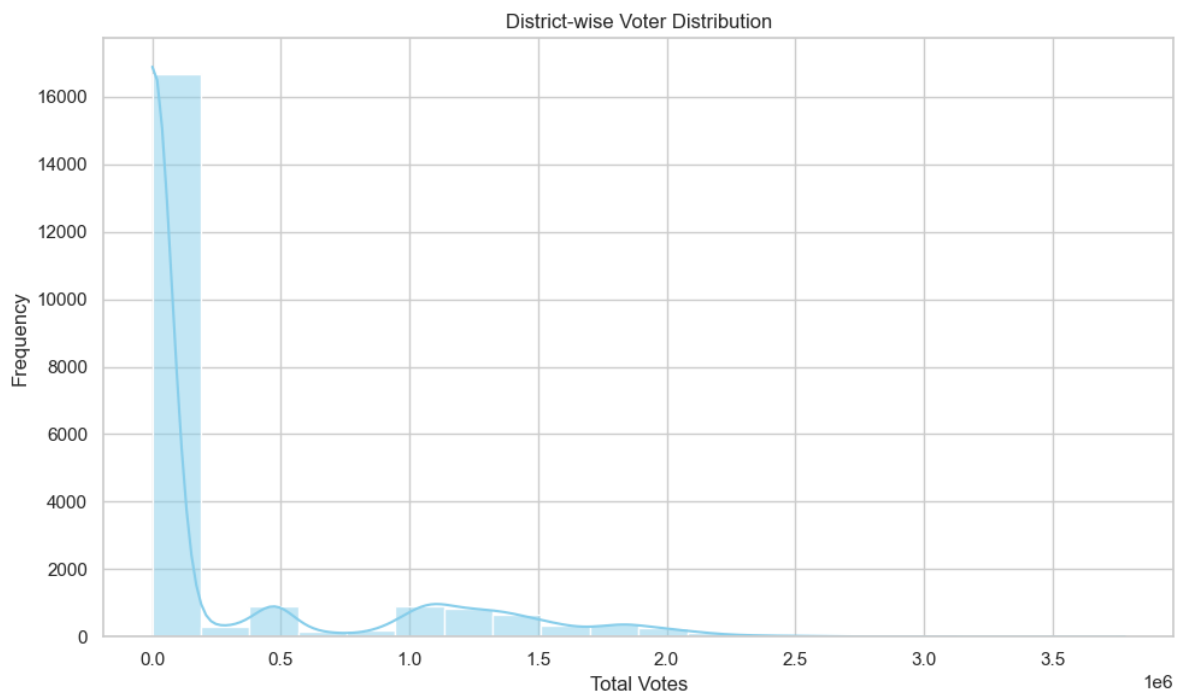# 4.2 Handling Missing Data

- **Missing values were identified using df.isnull().sum() before cleaning.**

- **Most of the missing values were numerical, affecting voter count columns.**

- **All missing values were filled with 0 using df.fillna(0) to maintain analysis continuity.**

- **This ensured there was no row loss or bias due to NaN removal.**

- **Post-processing, all missing values were successfully handled.**

4.3 Statistical Summary

- **Summary statistics were generated using df.describe(include='all').**

- **Measures like mean, std, min, max, and quartiles were analyzed.**

- **The Total votes column showed wide variation, indicating differing population densities or voter turnouts.**

- **Some columns had high standard deviation, suggesting potential outliers.**

- **Non-numeric columns were also summarized, helping to understand distribution of categorical data.**

4.4 District-wise Voter Distribution

- **A histogram was plotted for the Total column to understand the spread of votes across regions.**

- **The distribution was slightly right-skewed with most districts having moderate vote counts.**

- **A visible concentration was observed in the lower to mid-vote ranges.**

- **Kernel Density Estimation (KDE) overlay helped identify the peak density region.**

District-wise Voter Distribution

4.5 Vote Share Comparison by Category

- **A bar chart was created to compare vote shares across different candidate categories.**

- **Filtering was applied on 'Code Name' == 'S01_1' for example visualization.**

- **Independent candidates were found to have a visible vote share but less than major party candidates.**

- **The plot clearly separated vote contributions by category, aiding comparative analysis.**

Vote Share Comparison by Candidate Category
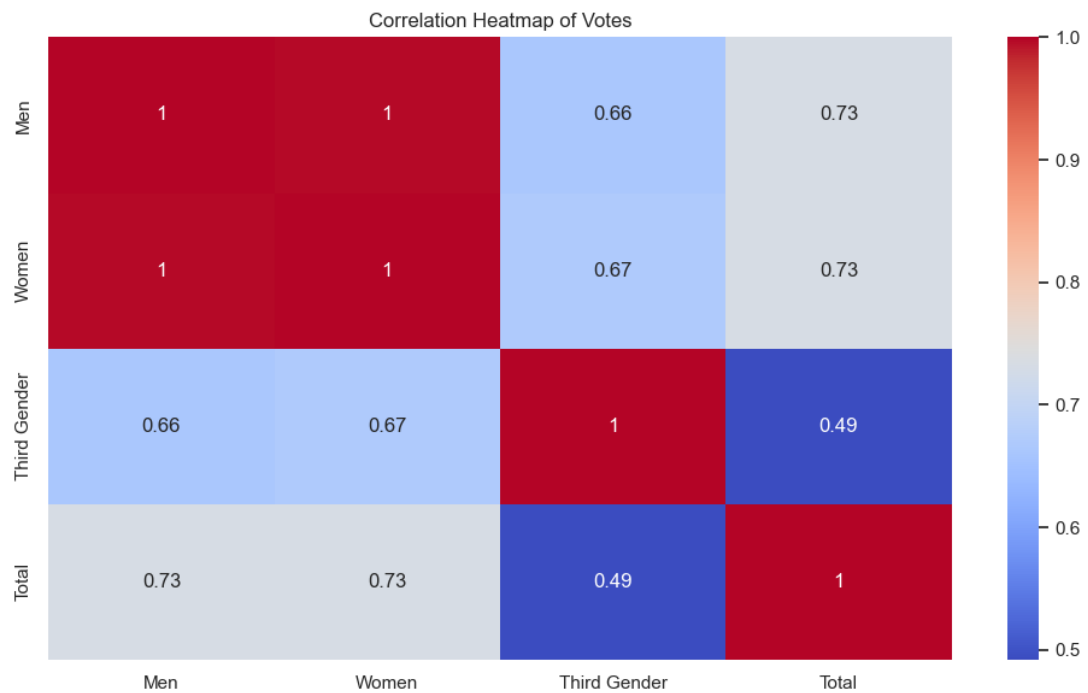
4.6 Winner Dominance Mapping

- **A pie chart and bar chart were generated to show the frequency of winners in each category.**

- **The dataset identified winners using 'Category' == 'Candidates - Contested'.**

- **Pie chart showed categorical dominance visually, with one or two categories leading.**

- **Bar chart provided numerical clarity on how often each category appeared among winners.**

Winner Dominance Pie Chart

Candidates - Contested   100.0%
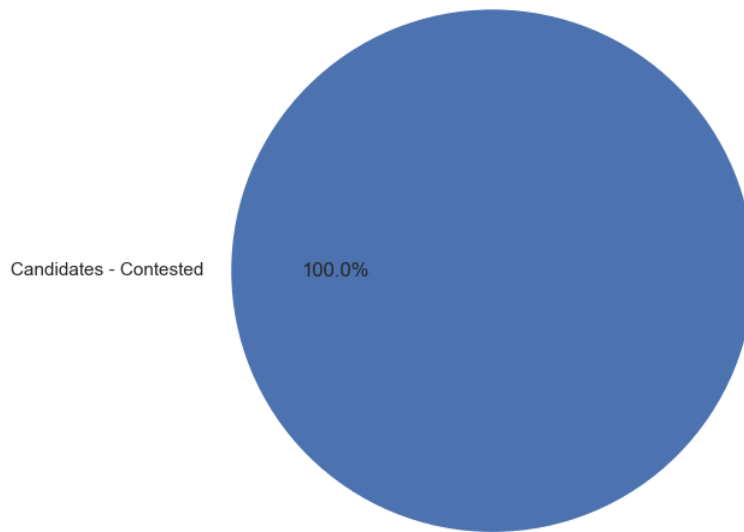
4.7 Correlation Heatmap

- **A correlation matrix was generated among numerical columns (Men, Women, Third Gender, Total).**

- **High positive correlation was observed between Men and Total, and similarly Women and Total.**

- **Correlation values helped confirm that gender distribution heavily influences total vote counts.**

Correlation Heatmap of Votes

	Men	Women	Third Gender	Total
Men	1	1	0.66	0.73
Women	1	1	0.67	0.73
Third Gender	0.66	0.67	1	0.49
Total	0.73	0.73	0.49	1

4.8 Outlier Detection Using Boxplot

- **A boxplot was created for all numeric columns to detect outliers.**

- **Clear outliers were found in Total, indicating districts with unusually high or low votes.**

- **Boxplot highlighted skewness in Men and Women voter distributions.**

Outlier Detection using Boxplot

-
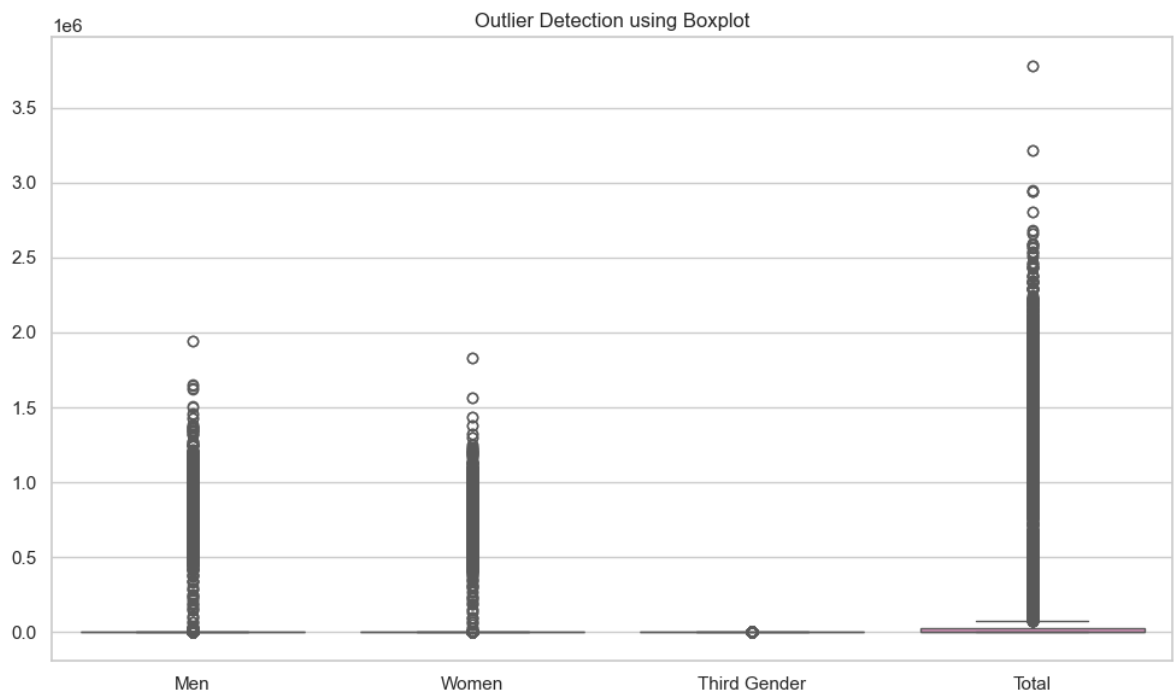
4.9 Pair Plot

- **A pairplot was generated on sampled numerical data to analyze inter-column relationships.**

- **The diagonal KDE plots showed distribution of each numeric variable.**

- **Scatter plots between Men and Women showed a moderately linear trend.**

- **This technique confirmed consistency in demographic data and revealed patterns in voter composition.**

4.10 Line Graph of Total Votes

- **A line graph was plotted to observe the trend of total votes by index (record order).**

- **Fluctuations were visible, indicating varying turnout across constituencies.**

- **Useful for detecting trends or anomalies over dataset entries.**

-
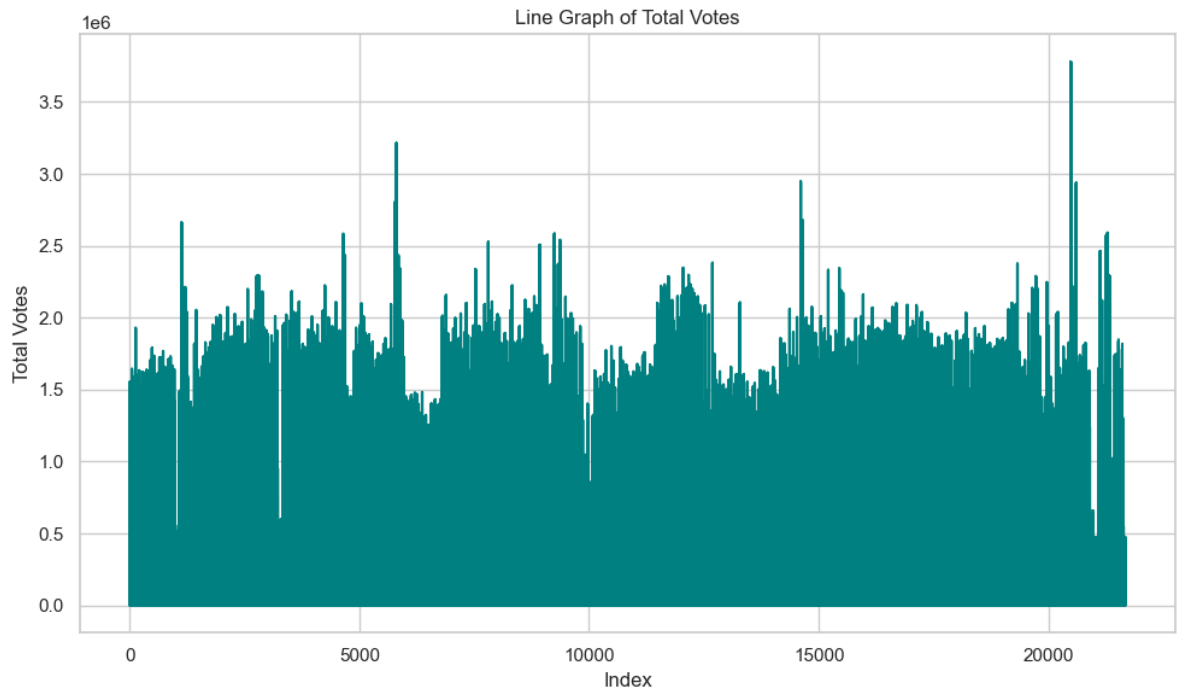
4.11 Category-wise Total Votes

- **A bar graph was plotted to display total votes by candidate category.**

- **Categories were visually compared in terms of total vote counts.**

- **Independent candidates showed some presence, though major parties dominated the graph.**

- **Rotation and coloring helped improve readability and visual segmentation.**

4.12 Scatter Plot of Men vs Women Voters

- **A scatter plot was drawn comparing Men and Women voter counts.**

- **A generally linear trend indicated balanced gender participation in most districts.**

- **Points were widely spread, highlighting both balanced and imbalanced constituencies.**

- **The plot supports gender-wise analysis and comparative turnout studies.**
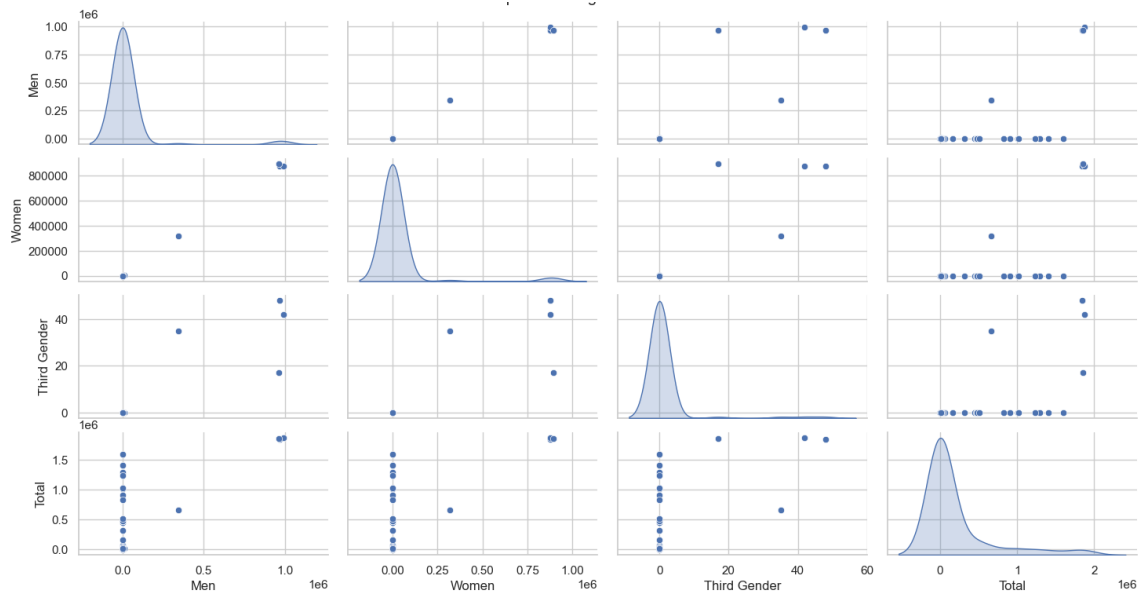
5. Conclusion

- **The 2024 General Elections highlighted the significant, though varied, presence of independent candidates across constituencies.**

- **While their overall vote share was lower compared to major parties, independent candidates showed strong regional performances.**

- **Voter demographics played a key role, with balanced gender participation aligning with higher vote counts.**

- **Some independent candidates emerged as strong contenders, especially in districts with localized issues.**

- **The data visualization techniques helped uncover hidden patterns and outliers in electoral behavior.**

- **Independent participation serves as a democratic barometer of voter sentiment beyond party lines.**

- **This analysis provides valuable insights into evolving political dynamics.**

- **Independent candidates remain vital in enhancing representation and political choice.**

6. Future Scope

- **Deeper demographic analysis (age groups, income levels) could refine understanding of voter preferences.**

- **Predictive modeling using machine learning could forecast independent candidates' success in future elections.**

- **Incorporating sentiment analysis from social media or local news could add qualitative insights.**

- **Comparative studies across multiple election years can highlight long-term trends.**

- **GIS-based mapping can be used to visualize geographic voting patterns in more detail.**

- **Analyzing campaign expenditure and outreach strategies may explain varying performance.**

- **Collaborations with electoral commissions can help improve dataset granularity.**

- **Future research may also examine post-election impact of elected independent leaders.**

---

7. References

- **Election Data Source: C:\Users\anand\Downloads\ElectionData.csv**

- **Official Election Commission Portals (for contextual validation)**

- **Python Libraries Used: pandas, matplotlib, seaborn, numpy**

- **Data Visualization Concepts: Boxplots, Heatmaps, Pairplots, Bar & Line Charts**

- **Election Research Studies (2020–2024): Various academic & journal publications**

- **Documentation: Python pandas & seaborn official documentation**

- **Additional Reading: Articles on Independent Candidates in Indian Elections**

- **Presentation based on EDA and Visual Analytics conducted in Jupyter Notebook**

8. Source Code:

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
# Set style for plots
sns.set(style='whitegrid')
plt.rcParams['figure.figsize'] = (10, 6)


# Load the dataset (handling encoding issues)
file_path = "C:\\Users\\anand\\Downloads\\ElectionData.csv"
try:
    df = pd.read_csv(file_path, encoding='utf-8')
except UnicodeDecodeError:
    df = pd.read_csv(file_path, encoding='latin1')


# Basic information about dataset
print("\nDataset Info:")
print(df.info())


print("\nSummary Statistics:")
print(df.describe(include='all'))


# Handling missing data
print("\nMissing Values Before:")
print(df.isnull().sum())


df.fillna(0, inplace=True)  # Fill NaN with 0 for simplicity


print("\nMissing Values After:")
print(df.isnull().sum())


# Objective 1: District-wise Voter Distribution Analysis
if 'Total' in df.columns:
    plt.figure()
    sns.histplot(df['Total'], bins=20, kde=True, color='skyblue')
    plt.title('District-wise Voter Distribution')
```

```python
    plt.xlabel('Total Votes')

    plt.ylabel('Frequency')

    plt.tight_layout()

    plt.show()


# Objective 2: Candidate Vote Share Comparison

if 'Code Name' in df.columns and 'Category' in df.columns and 'Total' in df.columns:

    vote_data = df[df['Code Name'] == 'S01_1']  # Example filter

    if not vote_data.empty:

        plt.figure()

        sns.barplot(x='Category', y='Total', hue='Category', data=vote_data, palette='pastel',
legend=False)

        plt.title('Vote Share Comparison by Candidate Category')

        plt.xticks(rotation=45)

        plt.tight_layout()

        plt.show()


# Objective 3: Winner Dominance Mapping

if 'Category' in df.columns:

    winners = df[df['Category'] == 'Candidates - Contested']

    winner_counts = winners['Category'].value_counts()


    if not winner_counts.empty:

        # Pie Chart

        plt.figure()

        plt.pie(winner_counts, labels=winner_counts.index, autopct='%1.1f%%')

        plt.title('Winner Dominance Pie Chart')

        plt.tight_layout()

        plt.show()


        # Bar Chart

        plt.figure()

        winner_counts.plot(kind='bar', color='orange')
```

```python
    plt.title('Winner Dominance Bar Chart')

    plt.ylabel('Frequency')

    plt.tight_layout()

    plt.show()


# Objective 4: Vote Correlation Insight

numerical_cols = ['Men', 'Women', 'Third Gender', 'Total']

available_cols = [col for col in numerical_cols if col in df.columns]


if len(available_cols) >= 2:

    corr = df[available_cols].corr()

    plt.figure()

    sns.heatmap(corr, annot=True, cmap='coolwarm')

    plt.title('Correlation Heatmap of Votes')

    plt.tight_layout()

    plt.show()


# Objective 5: Outlier Detection using Boxplot

if len(available_cols) >= 1:

    plt.figure()

    sns.boxplot(data=df[available_cols], palette='Set2')

    plt.title('Outlier Detection using Boxplot')

    plt.tight_layout()

    plt.show()


# Pairplot for variable relationships

if len(available_cols) >= 2:

    sns.pairplot(df[available_cols].sample(min(100, len(df))), diag_kind='kde')

    plt.suptitle('Pairplot of Voting Data', y=1.02)

    plt.show()


# Line Graph for Total votes
```

```python
if 'Total' in df.columns:

    plt.figure()

    df['Total'].plot(kind='line', color='teal')

    plt.title('Line Graph of Total Votes')

    plt.xlabel('Index')

    plt.ylabel('Total Votes')

    plt.tight_layout()

    plt.show()


# Bar Graph for Category vs Total

if 'Category' in df.columns and 'Total' in df.columns:

    plt.figure()

    sns.barplot(x='Category', y='Total', hue='Category', data=df, palette='muted',
legend=False)

    plt.title('Total Votes by Category')

    plt.xticks(rotation=45)

    plt.tight_layout()

    plt.show()
# Objective 9: Scatter Plot for Two Numerical Columns
if 'Men' in df.columns and 'Women' in df.columns:

    plt.figure()

    sns.scatterplot(x='Men', y='Women', data=df, color='purple', alpha=0.6)

    plt.title('Scatter Plot of Men vs Women Voters')

    plt.xlabel('Men Voters')

    plt.ylabel('Women Voters')

    plt.tight_layout()
plt.show()
```
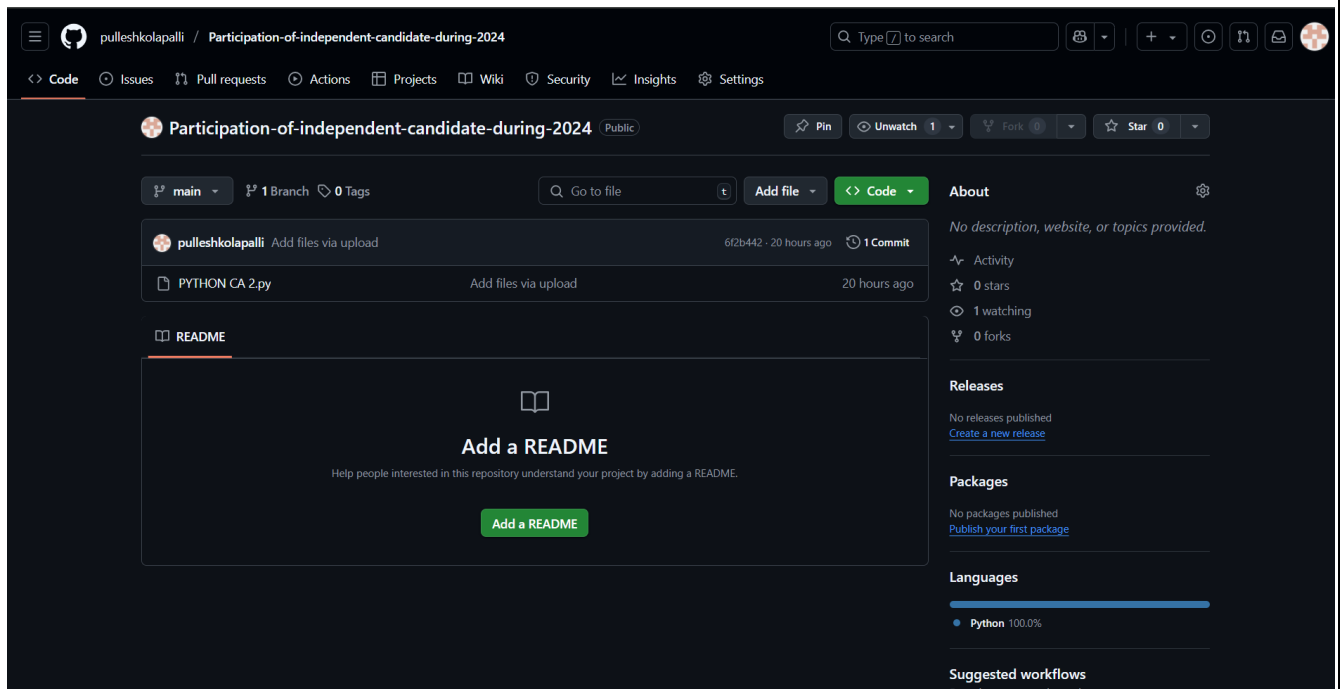
github link :

https://github.com/pulleshkolapalli/Participation-of-independent-candidate-during-2024

linked link:

Potential data quality or reporting issues revealed through outlier detection
Useful trends identified for policymakers, electoral boards, and civic organizations
🔍 What I did:
✅ Cleaned and preprocessed large-scale electoral data
✅ Visualized data using Seaborn & Matplotlib
✅ Performed univariate and outlier analysis using Boxplots
✅ Plotted distribution curves to understand overall spread
✅ Identified key regions with unusually high or low voter counts
💡 Tools & Tech: Python, Pandas, NumPy, Seaborn, Matplotlib
📁 Dataset: Public electoral dataset (used for academic and analytical purposes)
📌 This project helped me dive deep into real-world civic data and how it can reveal critical insights for democratic decision-making and voter outreach strategies.
🧑 I'm excited to explore more projects combining data science and social impact.
Feel free to check out the full notebook on GitHub 👇
🔗 [https://lnkd.in/gv69seNJ]
#EDA #DataScience #Python #Elections2024 #DataAnalysis #VoterInsights #CivicTech #Visualization #Analytics #StudentProject #pulleshAI #LinkedInLearner #lpu #SCSE #DataForGood #Excelgraphs #Pythontools #input #outputs #github #linkedin #vscode #lectural #AashimaBansal.

```
1   # Import necessary libraries
2   import pandas as pd
3   import numpy as np
4   import matplotlib.pyplot as plt
5   import seaborn as sns
6
7   # Set style for plots
8   sns.set(style='whitegrid')
9   plt.rcParams['figure.figsize'] = (10, 6)
10
11  # Load the dataset (handling encoding issues)
12  file_path = "C:\\Users\\anand\\Downloads\\ElectionData.csv"
13  try:
14      df = pd.read_csv(file_path, encoding='utf-8')
15  except UnicodeDecodeError:
16      df = pd.read_csv(file_path, encoding='latin1')
17
18  # Basic information about dataset
19  print("\nDataset Info:")
20  print(df.info())
21
22  print("\nSummary Statistics:")
23  print(df.describe(include='all'))
24
25  # Handling missing data
26  print("\nMissing Values Before:")
27  print(df.isnull().sum())
28
29  df.fillna(0, inplace=True)  # Fill NaN with 0 for simplicity
30
31  print("\nMissing Values After:")
32  print(df.isnull().sum())
```

+8

🔵 You and 50 others   28 comments

Reactions

+43

👍 Like   💬 Comment   🔁 Repost   ➤ Send

📊 867 impressions   View analytics

Kolapalli Hema pull...
B.Tech C.S.E at Lovely Professional University
Andhra Pradesh
🎓 Lovely Professional University

Job search smarter with Premium
🟧 Try for ₹0

Profile viewers   20
Post impressions   906