Exit Poll Analysis

CAPSTONE PROJECT PYTHON

Submitted by :

1. Mayank Kanojia

2. Sushant Dutta Patil

3. Dheeraj Singh

4. Eshwar

Submitted to:

Sir Madhav Dubey

Exit Poll Analysis - Capstone Project Report

ABSTRACT

This Exit Poll Analysis project leverages Python for analyzing and visualizing voter patterns, turnout rates, and party performance. By using tools such as pandas for data manipulation and matplotlib for visualizations, the project aims to uncover trends in voter behavior, political participation, and election outcomes. The focus is on generating actionable insights from exit poll data, making it a valuable resource for election analysts and political strategists.

INTRODUCTION

Exit polls provide a snapshot of voter behavior and election trends by collecting data directly from voters after they have cast their ballots. This project aims to analyze such data to reveal key insights into party performance, voter demographics, and overall turnout. By employing Python as the analytical tool, the project demonstrates the power of data-driven decision-making in understanding electoral trends. The analysis covers various metrics, including votes, electors, party performance, and regional dynamics, and presents the findings through impactful visualizations.

OBJECTIVE :

The primary objectives of this project are:

1. To clean and preprocess the exit poll dataset to ensure data accuracy and usability.

2. To analyze the relationships between key variables such as electors, votes, and turnout rates.

3. To create visualizations that effectively communicate trends and insights, including:

   - Party-wise seat distribution.
   - Top cities by total votes.
   - Correlation between electors and votes cast.

DESCRIPTION

The dataset used for this project contains exit poll information, including:

- State: The state where the elections were conducted.
- City Name: Cities or constituencies within the states.
- Party: Political parties contesting in the elections.
- Electors: The number of registered voters.
- Votes: The number of votes polled in each constituency.
- Caste Type: Demographic distribution of voters categorized into ST, SC, GEN, etc.

Steps in the Project:

1.Data Collection: We Collected Data from https://www.indiavotes.com website.

2. Data Cleaning: Removal of unnecessary characters (e.g., commas) from numerical data and conversion of string columns to appropriate numeric or categorical types.

3. Exploratory Data Analysis (EDA): Aggregation of data to analyze votes and turnout across cities, parties, and states.

4. Visualization: Creation of scatter plots, bar charts, and horizontal bar charts to present the findings.

SOURCE CODE

The project was implemented using Python, with the following libraries and functions:

1. Libraries:
- pandas: For reading, cleaning, and aggregating data.
Functions used: read_excel, groupby, reset_index, astype, sort_values, and general DataFrame operations.
- matplotlib.pyplot: For visualizing trends in the dataset.
Functions used: scatter, barh, title, xlabel, ylabel, tight_layout, show.

2. Key Code Snippets:

```python
import pandas as pd

# Load the data from Sheet1
data = pd.read_excel("C:\\Pyhton_Capstone\\Final_Dataset.xlsx")
(data.head())
```

```python
# Standardize column names
data.columns = data.columns.str.lower().str.replace(' ', '_')

# Convert 'electors' and 'votes' columns to numeric, handling non-numeric
entries
for col in ['electors', 'votes']:
    data[col] = pd.to_numeric(data[col].replace(',', '', regex=True),
errors='coerce')

# Replace missing or invalid values in numeric columns
data['electors'] = data['electors'].fillna(data['electors'].mean())
data['votes'] = data['votes'].fillna(data['votes'].median())

# Replace missing values in 'turnout' with a numeric placeholder or drop
them
data['turnout'] = pd.to_numeric(data['turnout'], errors='coerce')
data['turnout'] = data['turnout'].fillna(data['turnout'].median())

# Replace missing values in categorical columns
categorical_columns = data.select_dtypes(include=['object']).columns
data[categorical_columns] = data[categorical_columns].fillna('Unknown')

# Ensure 'no._of_seats' is integer
data['no._of_seats'] = pd.to_numeric(data['no._of_seats'],
errors='coerce').fillna(0).astype(int)

# Drop duplicates if any
data = data.drop_duplicates()

# Save the cleaned data (optional)
data.to_excel('Cleaned_Dataset.xlsx', index=False)

print("Data preprocessing completed")
```

INPUT/OUTPUT

Input: An Excel file with structured exit poll data, including columns like State, City Name, Party, Electors, Votes, and Caste Type.

Output:

Basic Stats

```
Total number of seats: 9870
Total number of candidates: 543
Unique parties: 42
           turnout        electors          votes
count   543.000000   5.430000e+02   5.430000e+02
mean      0.668880   1.800474e+06   1.190679e+06
std       0.099748   3.601687e+05   2.444487e+05
min       0.386000   5.778400e+04   4.920000e+04
25%       0.587000   1.633828e+06   1.059064e+06
50%       0.659000   1.820854e+06   1.174982e+06
75%       0.755000   1.980864e+06   1.346307e+06
max       0.922000   3.779596e+06   2.453608e+06
```

```
Seats won by party:
 party
National Unity Party                                           3746
Samajwadi Party                                                1644
People's Democratic Alliance                                   1295
Green Future Party                                              708
Dravida Munnetra Kazhagam                                      351
Nationalist Congress Party – Sharadchandra Pawar               241
Shiv Sena (Uddhav Balasaheb Thackrey)                          234
Telugu Desam                                                   217
Alliance for Equality (United)                                 183
Shiv Sena                                                      181
Progressive Reform Congress                                    137
Lok Janshakti Party(Ram Vilas)                                 125
Apna Dal (Soneylal)                                             79
Independent                                                     71
Yuvajana Sramika Rythu Congress Party                           69
Communist Party Of India (Marxist)                              68
Communist Party Of India (Marxist-Leninist) (Liberation)        67
Alliance for Equality (Secular)                                 48
Indian Union Muslim League                                      48
Communist Party Of India                                        47
Viduthalai Chiruthaigal Katchi                                  40
Hindustani Awam Morcha (Secular)                                38
Nationalist Congress Party                                      32
```

```
...
Voice Of The People Party                               1
Sikkim Krantikari Morcha                                1
Zoram People's Movement                                 1
```
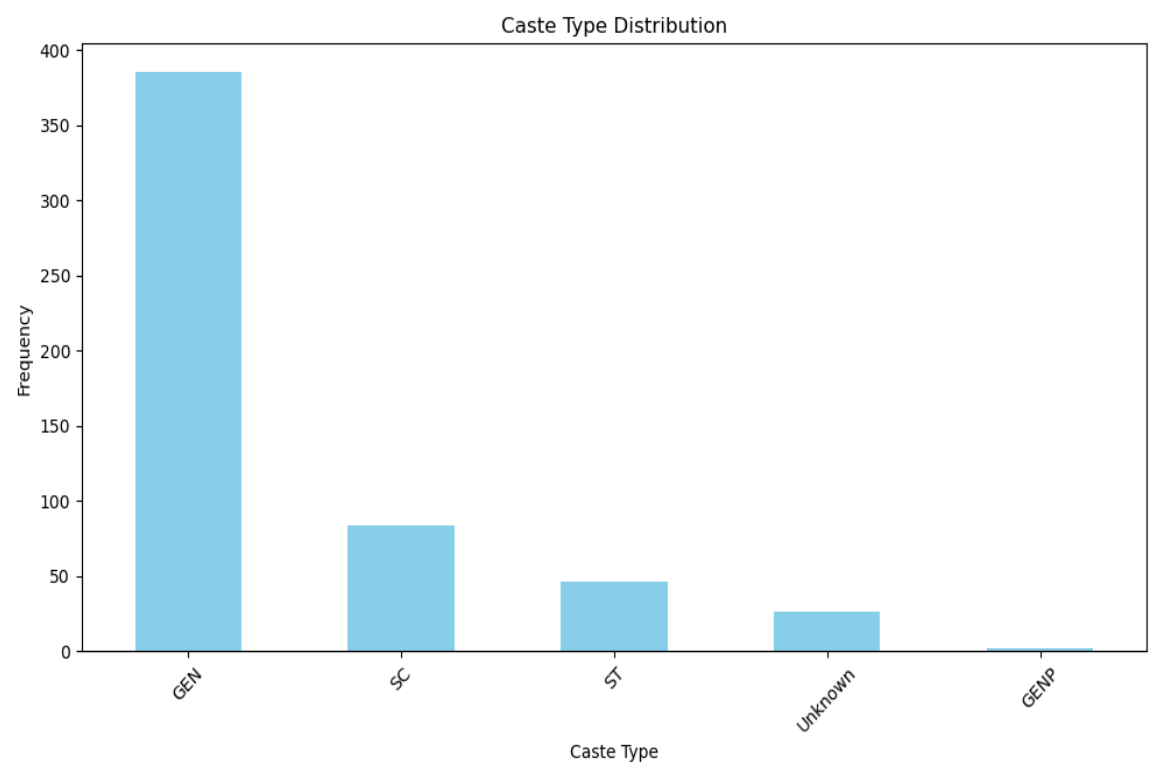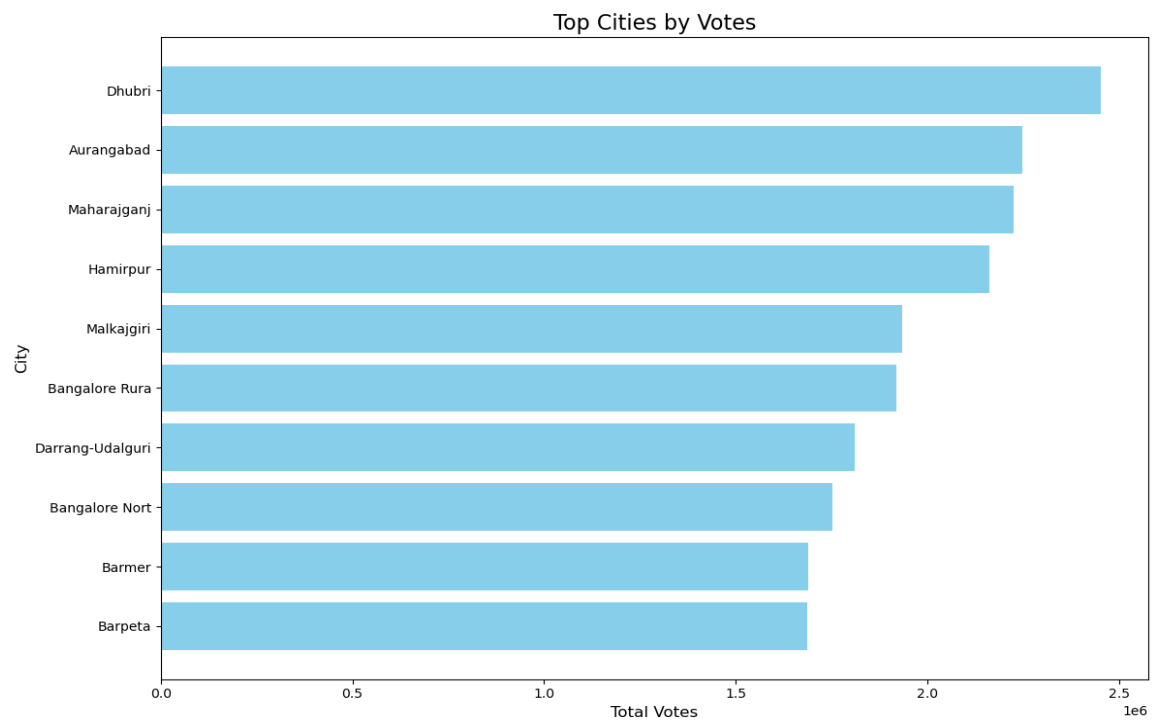
Turout Percentage by State :
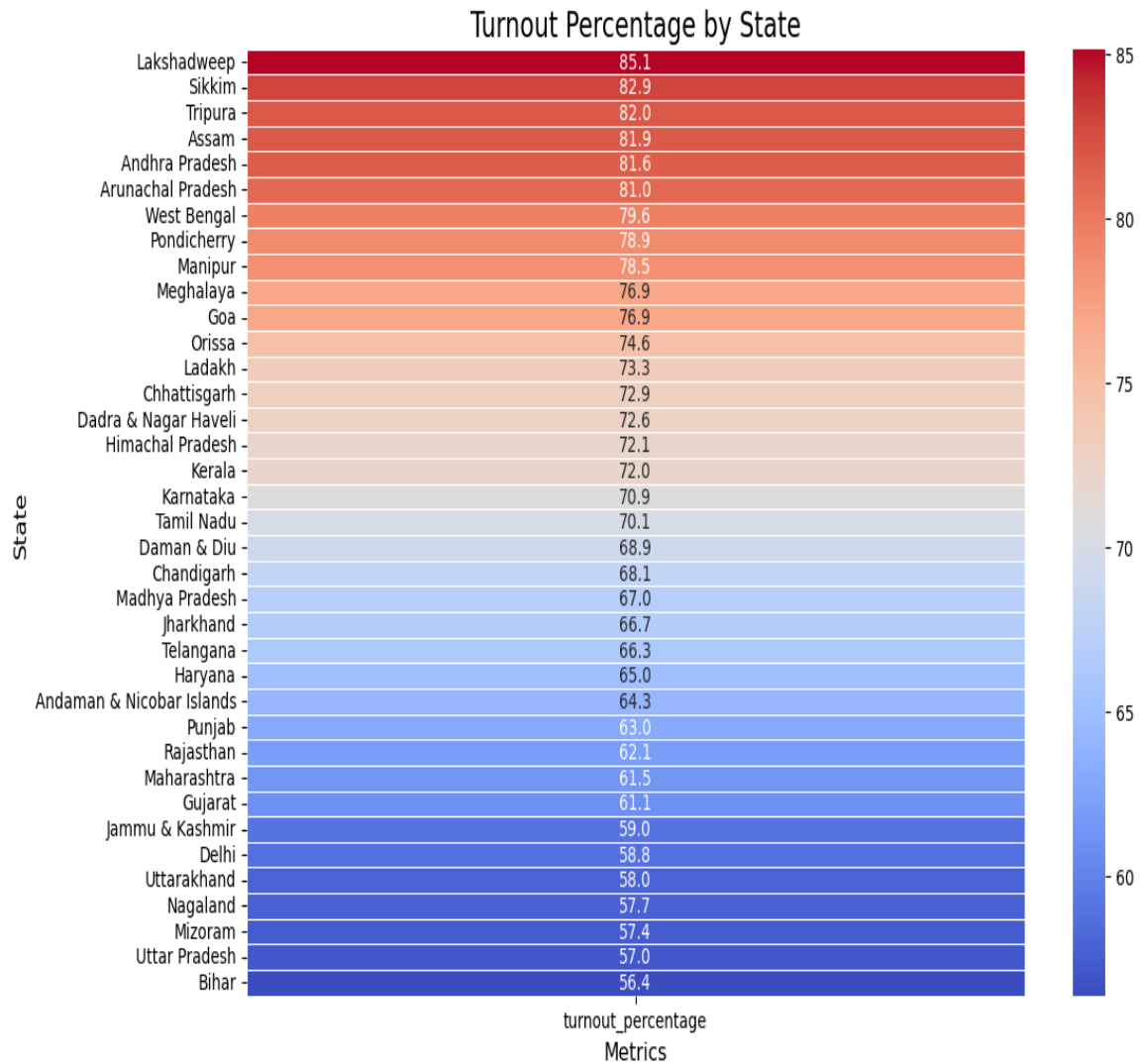
- Insights:
  - Top-performing cities based on voter turnout.
  - Party-wise seat analysis.
  - Correlation between registered voters and votes cast.
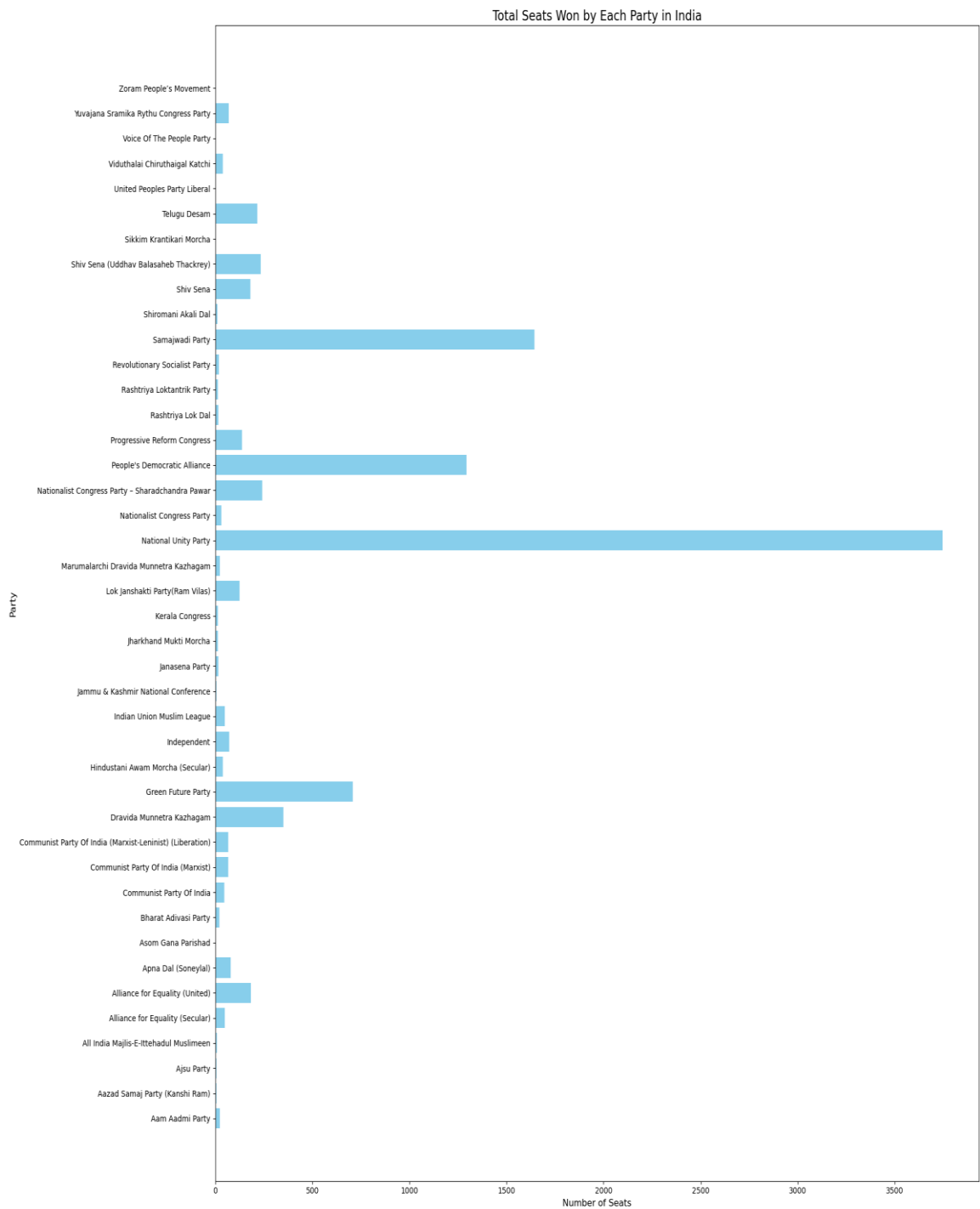
Caste Type Distribution :

Top Cities by Voters :

Overall Party Perfomance :

Total Seats Won by Each Party in India

FUTURE SCOPE

The project has potential for expansion in several directions:
1. Predictive Analytics: Use machine learning models to predict election outcomes based on historical and exit poll data.
2. Sentiment Analysis: Analyze voter opinions through social media or survey data to gain deeper insights into voter preferences.
3. Real-Time Data Integration: Extend the project to include real-time data collection and visualization during ongoing elections for dynamic analysis.

DEVELOPMENT

Tools and Libraries:
- Python Libraries:
  - pandas: Used extensively for data cleaning, grouping, and aggregation.
  - matplotlib: Used for all visualizations, such as scatter plots and bar charts.

Steps in Development:
- Dataset Preprocessing: Ensured numerical consistency and proper data types for analysis.
- Exploratory Data Analysis (EDA): Generated summaries and identified key trends.
- Visualization: Created charts to effectively communicate insights.

Environment:

- Jupyter Notebook: Used for coding and iterative analysis.
- Microsoft Excel: Used as the input dataset format.

CONCLUSION

The Exit Poll Analysis project effectively demonstrates the use of Python for analyzing and visualizing voter trends. By focusing on metrics such as votes, electors, and party performance, the project highlights actionable insights into voter behavior and election outcomes. With its current capabilities, this project serves as a foundation for deeper analysis and predictive modeling in future studies.