

# Data Preparation

Date	19 October 2023
Team ID	NM2023TMID07655
Project Name	Quantitative Analysis of Candidates in 2019 Lok Sabha Elections

## Prepare The Data for Visualization: -

Analyzing Lok Sabha election data involves collecting and preparing various types of data for quantitative analysis. Here is a step-by-step guide on how you can prepare the data for visualization:

### 1. Data Collection:

- **Geographical Data:** Obtain geographical data, such as latitude and longitude of constituencies, for mapping purposes.
- **Party Data:** Gather information about political parties, their ideologies, leaders, and alliances.
- **Voter Turnout:** Collect data on voter turnout in each constituency.

### 2. Data Cleaning:

- **Missing Data:** Check for missing or incomplete data in any of the datasets and decide how to handle them (remove, impute, etc.).
- **Data Integrity:** Ensure consistency in the data formats. For example, ensure that numbers are consistently formatted as integers or floats.
- **Duplicate Data:** Remove any duplicate entries in the datasets to maintain data integrity.

### 3. Data Integration:

- **Merge Datasets:** Combine the election results data with demographic and geographical data using a common identifier like constituency ID.
- **Data Transformation:** If needed, transform the data. For example, you can calculate percentages, change absolute numbers to per capita figures, etc.

### 4. Data Exploration:

- **Descriptive Statistics:** Calculate basic statistics such as mean, median, standard deviation, etc., to understand the central tendency and dispersion of the data.

- **Correlation Analysis:** Explore relationships between variables, for instance, between literacy rates and voter turnout.
- **Outlier Detection:** Identify and handle outliers that might affect the analysis and visualizations.

## 5. Data Visualization:

- **Bar Charts:** Use bar charts to represent the number of votes obtained by each party or candidate in different constituencies.
- **Pie Charts:** Represent the percentage of votes secured by each party in the overall election using pie charts.
- **Maps:** Utilize geographical data to create maps showing the election results by constituency.
- **Heatmaps:** Show voter turnout or party-wise votes using a heatmap to identify patterns.
- **Scatter Plots:** Use scatter plots to explore relationships between variables like literacy rates and votes secured.
- **Time Series Plots:** If you have data over time, create time series plots to visualize trends in voter turnout or party popularity.

## 6. Interpretation and Insights:

- **Analyze Patterns:** Look for patterns and trends in the visualizations. For example, identify regions where a particular party performed exceptionally well or areas with low voter turnout.
- **Compare Data:** Compare the election results with demographic data to draw meaningful insights. For instance, explore if there is a correlation between literacy rates and the votes received by a certain party.
- **Communicate Findings:** Clearly communicate your findings through charts, graphs, and concise summaries. Ensure that your audience understands the key takeaways from the analysis.

Remember, the choice of visualization depends on the type of data and the insights you want to convey. Always provide clear labels, legends, and context to your visualizations to make them understandable to your audience.

## Preparing A Data Module: -

preparing a data module for quantitative analysis involves organizing the data in a structured manner, creating functions to load and manipulate the data, and documenting the process for future reference. Below is a Python-based example

of how you can create a data module for the quantitative analysis of Lok Sabha Elections 2019.

## **1. Data Collection and Storage:**

First, store your data in a CSV file, a common format for tabular data. Let us assume you have two CSV files: `election_results.csv` containing election results and `constituency_data.csv` containing demographic and geographical data for each constituency.

## **2. Creating the Data Module:**

Create a Python file, for example, `election_data.py`, and define functions to load and manipulate the data.

## **3. Documentation:**

Add comments and docstrings to your functions explaining what each function does, the parameters it accepts, and the format of the returned data. This documentation is crucial for anyone using your data module in the future.

## **4. Testing:**

Test your functions with sample data to ensure they work as expected.

By organizing your data and code in this manner, you create a reusable and understandable data module for quantitative analysis of Lok Sabha Elections 2019. Remember to adapt the functions and processing steps according to the specific requirements of your analysis.