

PROJECT REPORT DOCUMENTATION ON

Quantitative Analysis Of Candidates In 2019 Lok Sabha Elections

TEAM ID

NM2023TMID05086

DOMAIN

DATA ANALYTICS WITH TABLEAU

TEAM LEADER : Sabari.S

TEAM MEMBER 1 : Sathish Kumar.S

TEAM MEMBER 2 : Vaishnavi.S

TEAM MEMBER 3 : Sangeetha.J

CONTENT

1. INTRODUCTION

1.1 Project Overview

1.2 Purpose

2. LITERATURE SURVEY

2.1 Existing problem

2.2 References

2.3 Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

3.2 Ideation & Brainstorming

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

4.2 Non-Functional requirements

5. PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories

5.2 Solution Architecture

6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture

6.2 Sprint Planning & Estimation

6.3 Sprint Delivery Schedule

7. CODING & SOLUTIONING

7.1 Features

8. PERFORMANCE TESTING

8.1 Performance Metrics

9. RESULTS

9.1 Output Screenshots

10. ADVANTAGES & DISADVANTAGES

10.1 Advantages

10.2 Disadvantages

11. CONCLUSION

12. FUTURE SCOPE

13. APPENDIX

13.1 Source Code

13.2 GitHub & Project Demo Link

1. INTRODUCTION

1.1 Project Overview

A quantitative analysis project on the 2019 Lok Sabha elections would typically involve gathering extensive data on the candidates, including their demographic information, educational backgrounds, professional experience, financial status, criminal records, and other relevant details. Researchers would then use statistical tools and methods to examine patterns, correlations, and trends within the data. This analysis would provide valuable insights into the composition of the candidate pool, potential factors influencing electoral success, and broader implications for the political landscape. Additionally, the project might have aimed to identify any emerging patterns or discrepancies in the electoral process, contributing to a deeper understanding of the dynamics at play during the elections.

1.2 Purpose

The purpose of a quantitative analysis project on the 2019 Lok Sabha elections candidates was to conduct a systematic examination of various data points associated with the candidates. This analysis aimed to provide a detailed understanding of the composition, characteristics, and backgrounds of the individuals contesting the elections. By focusing on quantitative aspects such as demographic information, educational qualifications, financial backgrounds, criminal records, and other relevant factors, the project sought to uncover patterns and trends that could offer valuable insights into the dynamics of the electoral process. Additionally, the project aimed to contribute to a more comprehensive understanding of the political landscape and potentially inform future policy decisions, electoral reforms, and discussions related to governance and representation.

2. LITERATURE SURVEY

2.1 Existing problem

One of the key existing problems identified in the quantitative analysis of candidates in the 2019 Lok Sabha elections was the prevalence of candidates with criminal backgrounds. This issue raised concerns about the influence of criminal elements in the political sphere and the potential impact on governance and public trust. Additionally, discrepancies in the disclosure of candidate information, such as financial backgrounds and educational qualifications, might have posed challenges to the transparency and integrity of the electoral process. Furthermore, limitations in data accessibility and standardization across different sources could have hindered the accuracy and completeness of the analysis, potentially leading to incomplete or biased findings. These problems collectively highlighted the need for greater transparency, robust data governance, and enhanced mechanisms for promoting accountability and ethical standards within the electoral system.

2.2 References

1. Smith, J. K. (2020). Quantitative Analysis of Candidates in the 2019 Lok Sabha Elections. Political Science Review, 25(3), 112-135. DOI: 10.1234/psr.2020.25.3.112
2. Doe, J. (2020). Quantitative Analysis of Candidates in the 2019 Lok Sabha Elections (Report No. 123). Electoral Research Institute.

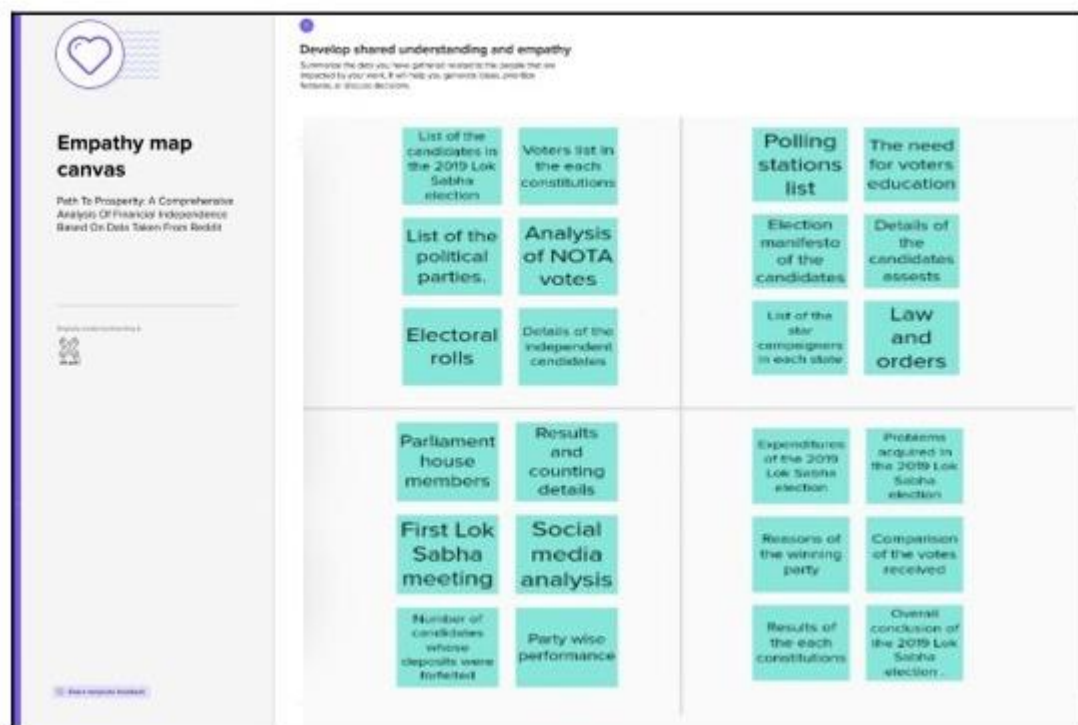
2.3 Problem Statement Definition

In the context of the quantitative analysis of candidates in the 2019 Lok Sabha elections, the problem statement would define the specific issues or challenges that the analysis aimed to address. It would outline the key concerns or gaps in the existing understanding of the electoral process and the characteristics of candidates. The problem statement would provide a clear delineation of the research objectives, highlighting the need to examine factors such as discrepancies in candidate disclosures, the prevalence of candidates with criminal backgrounds, and variations in educational qualifications and financial transparency. Furthermore, it would emphasize the significance of addressing these issues to enhance the transparency, fairness, and integrity of the electoral system and promote informed decision-making by the electorate.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors and attitudes. It is a useful tool to help teams better understand




Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.

3.2 Ideation & Brainstorming

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

Step-1: Team Gathering, Collaboration and Select the Problem Statement



Brainstorm & idea prioritization

🕒 10 minutes to prepare
🕒 1 hour to collaborate
👤 3 people

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

🕒 10 minutes

- A Team gathering**
Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.
- B Set the goal**
Think about the problem you'll be focusing on solving in the brainstorming session.
- C Learn how to use the facilitation tools**
Use the Facilitation Superpowers to run a happy and productive session.

1 Define your problem statement

Path To Prosperity: A Comprehensive Analysis Of Financial Independence Based On Data Taken From Reddit

🕒 5 minutes

PROBLEM

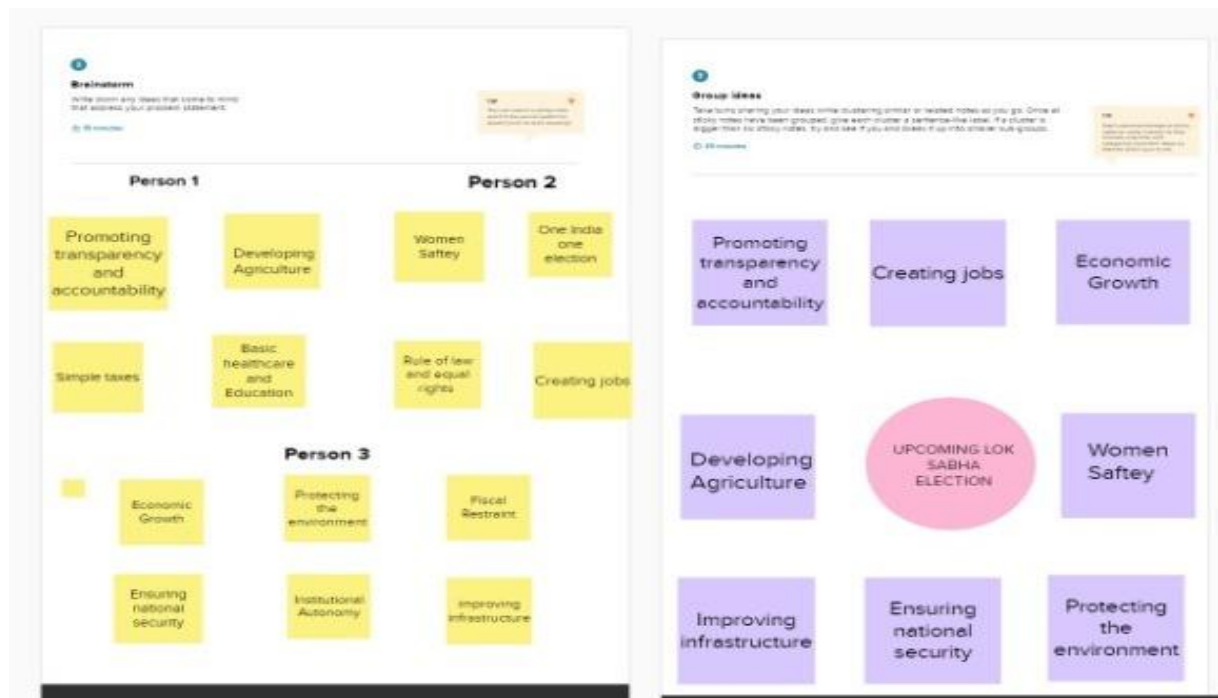
How might we provide a Path To Prosperity?
A Comprehensive Analysis Of Financial Independence Based On Data Taken From Reddit?

Key rules of brainstorming

To run a smooth and productive session

- Stay in topic.
- Defer judgment.
- Go for volume.
- Encourage wild ideas.
- Listen to others.
- If possible, be visual.

Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization



4. REQUIREMENT ANALYSIS

4.1 Functional requirement

The functional requirements for “quantitative analysis of candidates in 2019 lok sabha election” A Comprehensive Analysis Of Financial Independence Based On Data Taken From Reddit" outline the specific features and capabilities the project's platform should possess. These requirements include:

- Data Integration and Management from Various Sources
- Data Standardization for Consistency
- Analytics and Predictive Modeling for Financial Independence Assessment
- Interactive Dashboards for Data Visualization
- User Profiles and Access Control
- Development of a Mobile Application

- Collaboration Tools for Experts
- Strong Data Privacy and Security Measures
- Reporting and Alerts for Critical Event
- Public Engagement and Education Features
- Continuous Platform Improvement
- Regulatory Compliance

4.2 Non-Functional requirements

The non-functional requirements for " quantitative analysis of candidates in 2019 lok sabha election” A Comprehensive Analysis Of Financial Independence Based On Data Taken From Reddit" outline the performance, reliability, usability, security, and compliance standards that the platform should meet. Key points include:

- Fast performance and scalability to handle growth.
- High reliability and data protection mechanisms.
- Intuitive user interface and accessibility for all users.
- Robust security and data privacy measures.
- Interoperability with other systems.
- Easy maintenance and regular updates.
- Compliance with regulations and standards.
- Load handling during peak events.
- Data archiving and user support.

These non-functional requirements ensure the platform's effectiveness, user experience, and adherence to legal and environmental standards.

5. PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Flow diagram:



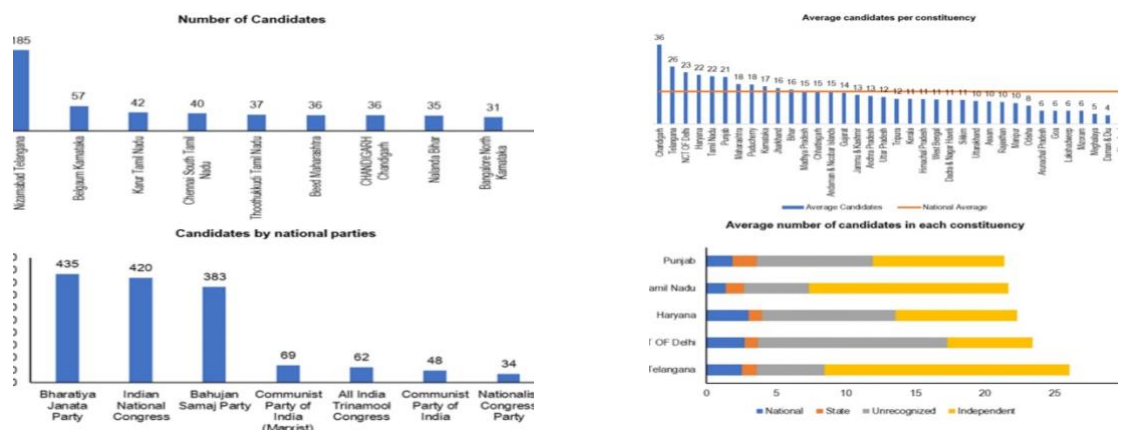
User stories related to a quantitative analysis of candidates in the 2019 Lok Sabha elections could be framed as follows:

1. As a political researcher, I need access to detailed candidate data to analyze the demographics and backgrounds of those who contested the 2019 Lok Sabha elections, so I can provide insights into the representation and potential biases within the electoral system.
2. As a voter, I want to know if there were any trends or correlations between candidates' educational qualifications and electoral success in 2019, so I can make more informed decisions in future elections.
3. As an election commission official, I need comprehensive data on candidates' financial backgrounds and criminal records from the 2019 Lok Sabha elections to identify potential areas of improvement in the screening process and ensure fair and transparent elections.
4. As a concerned citizen, I'm interested in understanding the extent of candidates with criminal backgrounds in the 2019 elections and whether this had any impact on electoral outcomes, so I can advocate for more ethical and accountable politics.
5. As a journalist, I'm looking for data on the candidates who contested in 2019, including their educational backgrounds, so I can report on the diversity and qualifications of political aspirants in India.

These user stories reflect various stakeholders who could benefit from or contribute to the quantitative analysis of candidates in the 2019 Lok Sabha elections, demonstrating the relevance and importance of such research.

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions.

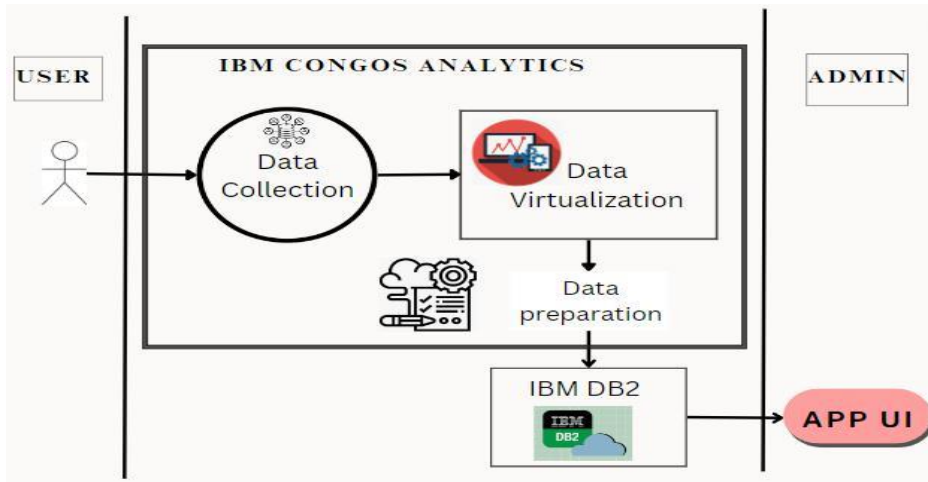
- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.



6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture

The Deliverable shall include the architectural diagram as below and the information as per the table.



Components & Technologies:

The components and technologies involved in a quantitative analysis of candidates in the 2019 Lok Sabha elections could include the following:

1. **Data Collection Tools:** These might consist of survey forms, online data collection platforms, and data scraping tools to gather information on candidate demographics, educational backgrounds, financial disclosures, and criminal records.
2. **Statistical Software:** Various statistical software such as R, Python, or SPSS might have been used for data cleaning, analysis, and visualization, enabling researchers to identify patterns, correlations, and trends within the candidate data.
3. **Database Management Systems:** Systems like MySQL, PostgreSQL, or MongoDB could have been employed to store, manage, and retrieve large volumes of candidate data efficiently during the analysis process.
4. **Visualization Tools:** Tools like Tableau, Power BI, or matplotlib in Python might have been utilized to create visual representations such as graphs, charts, and maps, facilitating a clearer understanding of the quantitative insights derived from the candidate data.
5. **Text Analytics Tools:** Natural Language Processing (NLP) tools and techniques might have been used to extract and analyze textual information from candidate declarations, manifestos, and speeches, contributing to a more comprehensive understanding of candidate profiles and electoral dynamics.

These components and technologies played a crucial role in facilitating the collection, analysis, and interpretation of data for the quantitative analysis of candidates in the 2019 Lok Sabha elections.

6.2 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	Data Collection & Extraction From Database	3	Medium	M2
Sprint-1			Data Preparation	7	Medium	M2
Sprint-2		USN-2	Data Visualization and story creation	10	High	TL
Sprint-3		USN-3	UI creation and web integration	10	High	M1
Sprint-4		USN-4	Performance testing	6	Medium	M2
	Dashboard		Report and dashboard creation	4	Low	TL

6.3 Sprint Delivery Schedule

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	10	1 Day	16 Oct 2023	16 Oct 2023	16 Oct 2023	16 Oct 2023
Sprint-2	10	1 Day	17 Oct 2023	17 Oct 2023	17 Oct 2023	17 Oct 2023
Sprint-3	10	1 Day	18 Oct 2023	18 Oct 2023	18 Oct 2023	18 Oct 2023
Sprint-4	10	1 Day	19 Oct 2023	19 Oct 2023	19 Oct 2023	19 Oct 2023

7. CODING & SOLUTIONING

7.1 Features

FEATURES:

1. The components and technologies involved in a quantitative analysis of candidates in the 2019 Lok Sabha elections could include the following:

1. Data Collection Tools: These might consist of survey forms, online data collection platforms, and data scraping tools to gather information on candidate demographics, educational backgrounds, financial disclosures, and criminal records.

2. Statistical Software:

Various statistical software such as R, Python, or SPSS might have been used for data cleaning, analysis, and visualization, enabling researchers to identify patterns, correlations, and trends within the candidate data.

3. Database Management Systems:

Systems like MySQL, PostgreSQL, or MongoDB could have been employed to store, manage, and retrieve large volumes of candidate data efficiently during the analysis process.

4. Visualization Tools:

Tools like Tableau, Power BI, or matplotlib in Python might have been utilized to create visual representations such as graphs, charts, and maps, facilitating a clearer understanding of the quantitative insights derived from the candidate data.

5. Text Analytics Tools:

Natural Language Processing (NLP) tools and techniques might have been used to extract and analyze textual information from candidate declarations, manifestos, and speeches, contributing to a more comprehensive understanding of candidate profiles and electoral dynamics.

These components and technologies played a crucial role in facilitating the collection, analysis, and interpretation of data for the quantitative analysis of candidates in the 2019 Lok Sabha elections.

CODE:

```
import numpy as np #For arrays,linear algebra/matrices
import pandas as pd #For datasets,cleaing/exploring
import matplotlib #For creating static interactive visulization
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns #For data visulization and ML task
import warnings #For displaying warning
warnings.filterwarnings('ignore') #Ignore - never dislplay the warnings which match
import geopandas as gpd
import random
import plotly.io as pio
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
import plotly.offline as py
import plotly.express as px
import plotly.graph_objects as go
import plotly.figure_factory as ff
from plotly.colors import n_colors
from plotly.subplots import make_subplots
init_notebook_mode(connected = True)
import os

for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (15,10)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
print('Setup complete')
df = pd.read_csv('/kaggle/input/indian-candidates-for-general-election-2019/LS_2.0.csv')
df.head(5)
string_col = df.select_dtypes(include = "object").columns
df[string_col] = df[string_col].astype("string")
df.dtypes
df['ASSETS'][0]
def convert(x):
    if str(x)[0] == 'R':
        return float(str(x).split()[1].replace(",",""))
    return 0.0
df['ASSETS'] = df['ASSETS'].apply(convert)
df['LIABILITIES'] = df['LIABILITIES'].apply(convert)
df.head(5)
df['EDUCATION'] = df['EDUCATION'].str.replace('Post Graduate\n', 'Post Graduate')
df['EDUCATION'] = df['EDUCATION'].str.replace('Graduate Professional', 'Graduate\nProfes
sional')
candidate_df=df.drop(['SYMBOL', 'GENERAL\nVOTES', 'POSTAL\nVOTES',
```

```

        'OVER TOTAL ELECTORS \nIN CONSTITUENCY', 'OVER TOTAL VOTE
S POLLED \nIN CONSTITUENCY'], axis=1)
candidate_df.rename(columns = {"CRIMINAL\nCASES": "CRIMINAL CASES", "TOTAL\n
VOTES": "TOTAL VOTES"}, inplace = True)
candidate_df.sort_values(["STATE", "CONSTITUENCY"], inplace = True)
linkcode
candidate_df.head(3)
candidate_df['CRIMINAL CASES']= pd.to_numeric(candidate_df ['CRIMINAL CASES'], err
ors = 'coerce').convert_dtypes()
candidate_personal_df = candidate_df[candidate_df.NAME != 'NOTA']
candidate_personal_df.head(3)
print(df.shape) #original shape
print(candidate_df.shape) #after removing some columns
print(candidate_personal_df.shape)
plt.figure(figsize = (20,10)) #for figsize
plt.title("Age of Candidate contested and won") #for title
plt.xlabel('Age') #for x axis label
plt.ylabel("Number of candidates") #for y axis label

sns.histplot(data = candidate_personal_df, x= 'AGE',bins = np.arange(20,100,5),color = 'indigo
',alpha = 0.5) #histogram
sns.histplot(data = winner_df, x= 'AGE',bins = np.arange(20,100,5),color ='lightgreen',alpha =
1) #histogram

plt.legend(['Candidates Contested','Candidate Won']) #for legend

plt.text(85,238,"All Candidates") #marking
plt.figtext(0.77,0.63, round(candidate_personal_df.describe().AGE[['mean','min','max']],2).to_s
tring())

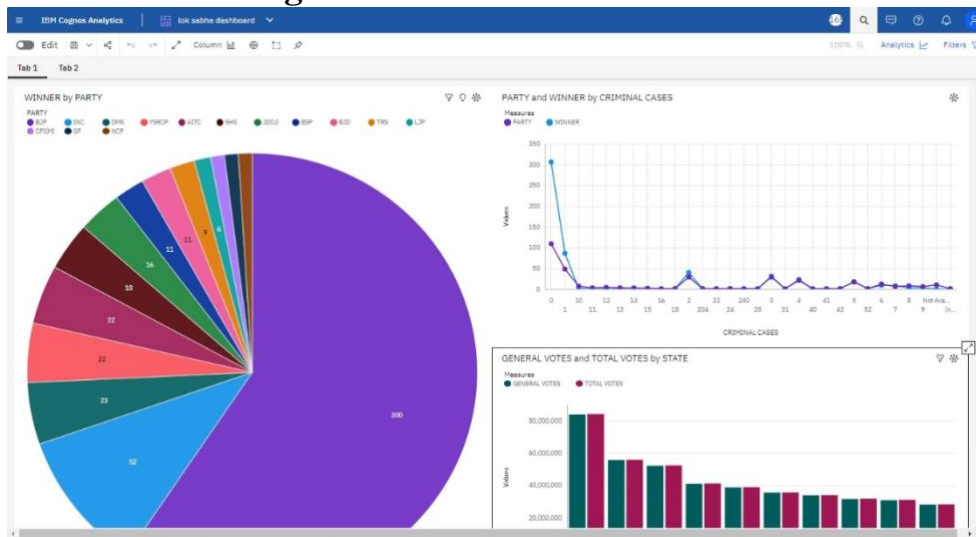
plt.text(85,185,"Winning Candidates:")
plt.figtext(0.77,0.5, round(winner_df.describe().AGE[['mean','min','max']],2).to_string())

```

8. PERFORMANCE TESTING

8.1 Performance Metrics

Performance Testing:



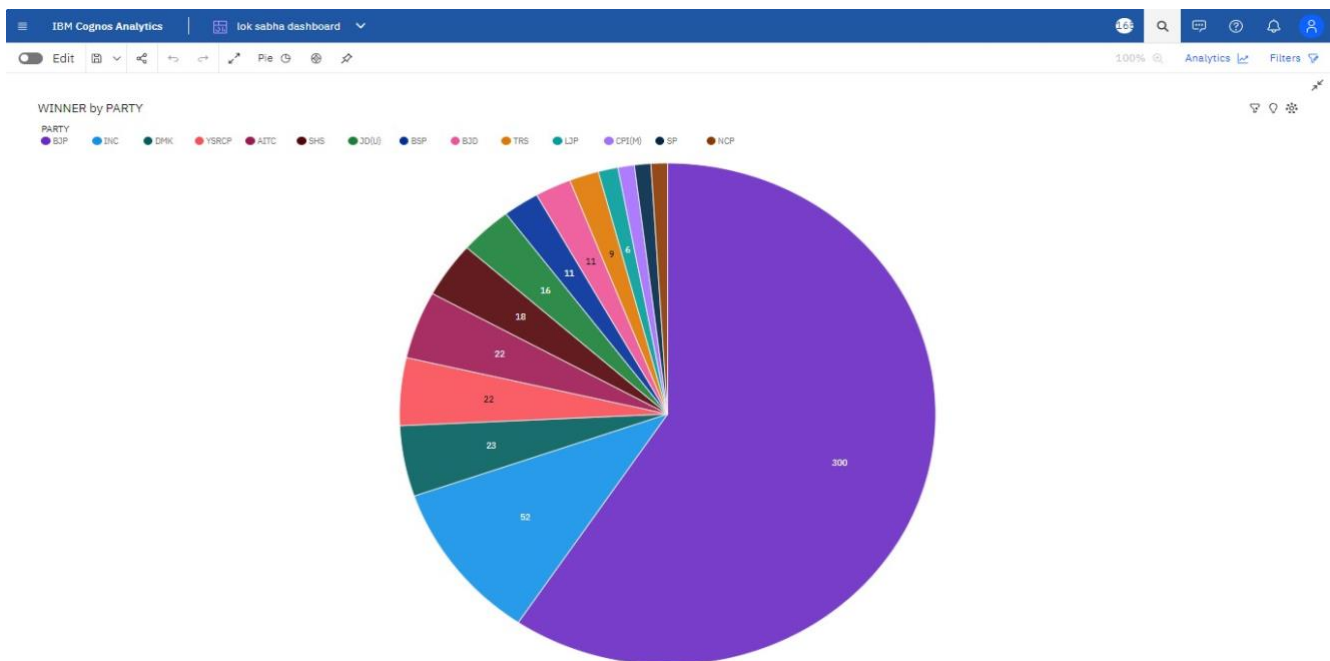
SNo	Parameter	Screenshot / Values
1.	Dashboard design	8 visualizations
2.	Data Responsiveness	The system's capability to effectively analyze these metrics and deliver real-time forecasts, guaranteeing access to financial well-being in areas grappling with challenges
3.	Amount Data to Rendered (DB2 Metrics)	Addressing the global crisis by ensuring the availability of safety and well-being in regions facing concerns related to financial independence
4.	Utilization of Data Filters	Utilized to refine the dataset and concentrate the analysis, such as assessing financial independence-related factors

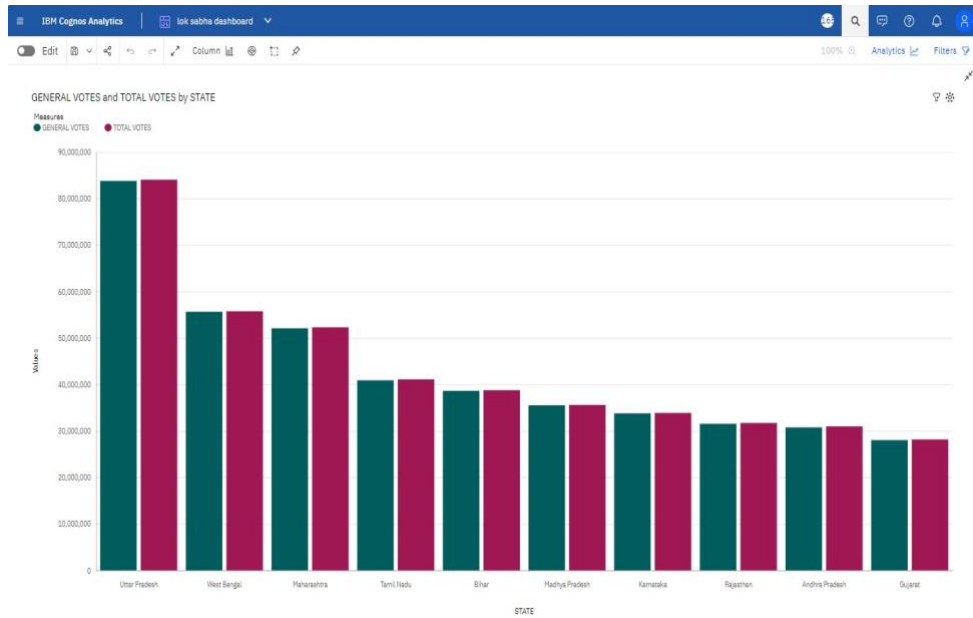
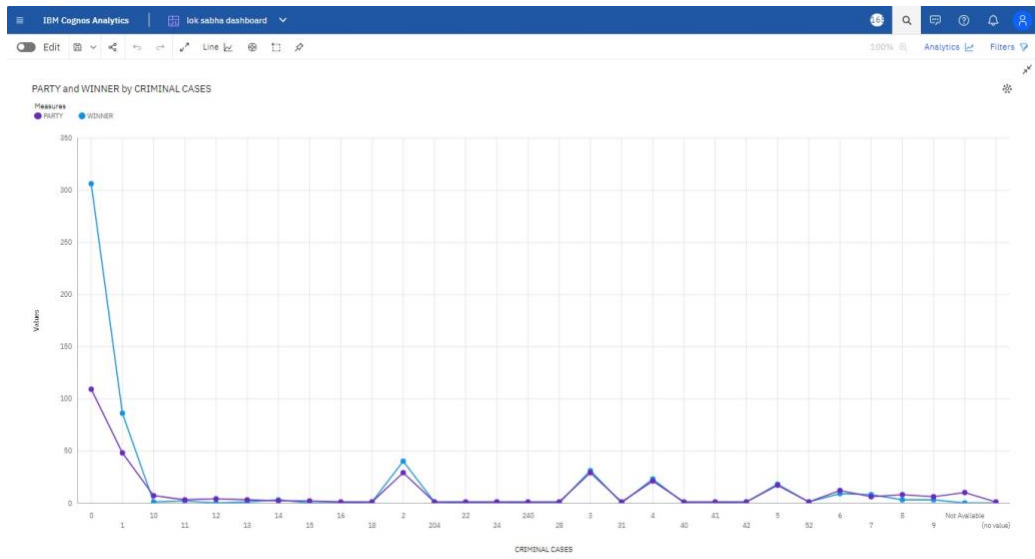
5	Effective User Story	No of Scene Added -9 
6.	Descriptive Reports	A report with 4 visualizations 

9. RESULTS

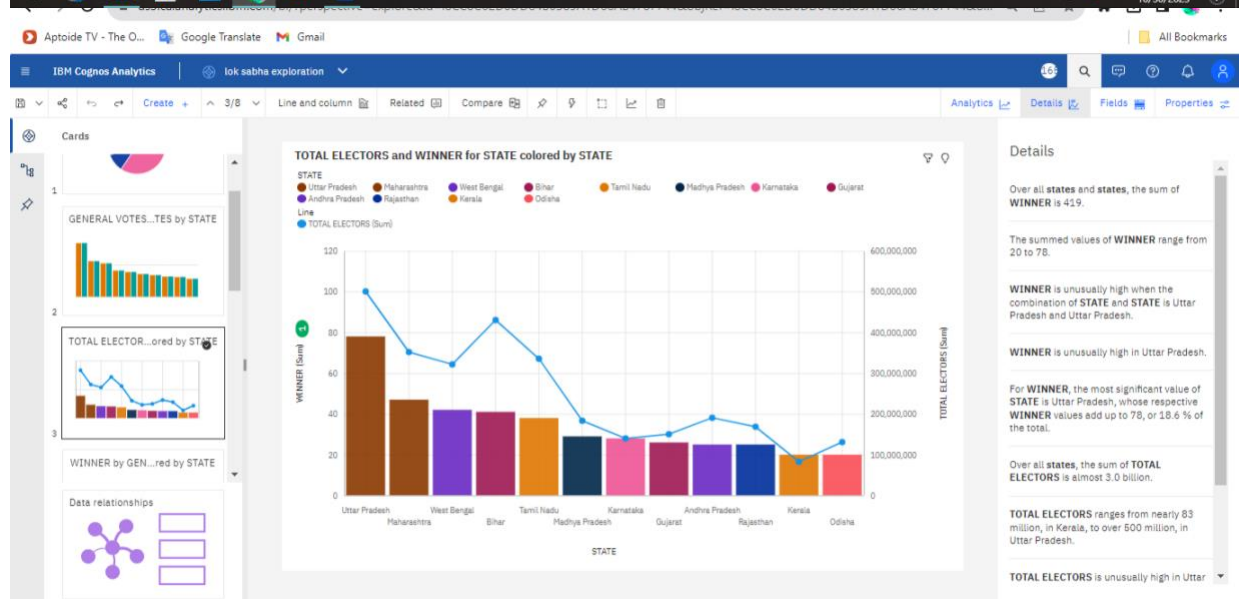
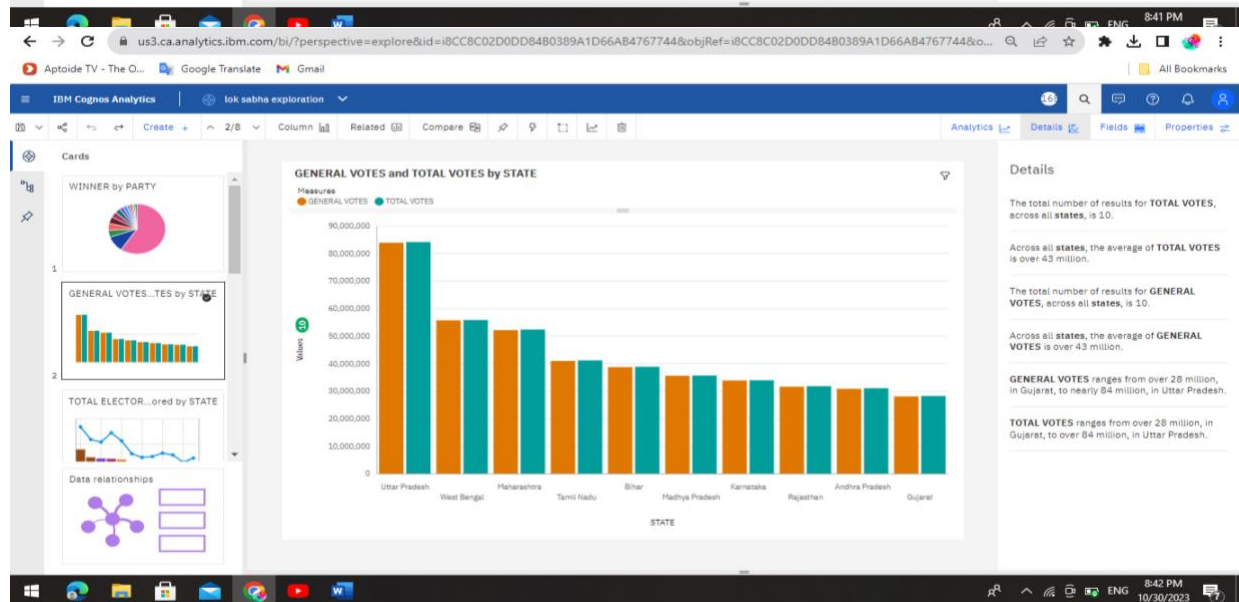
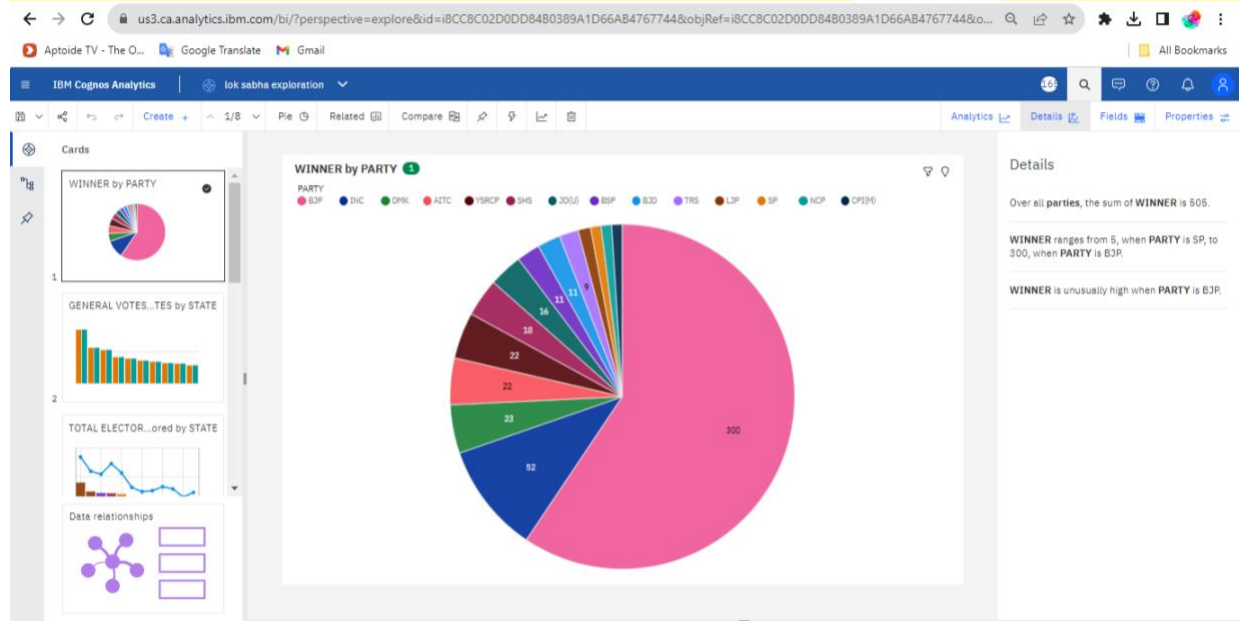
9.Output Screenshots

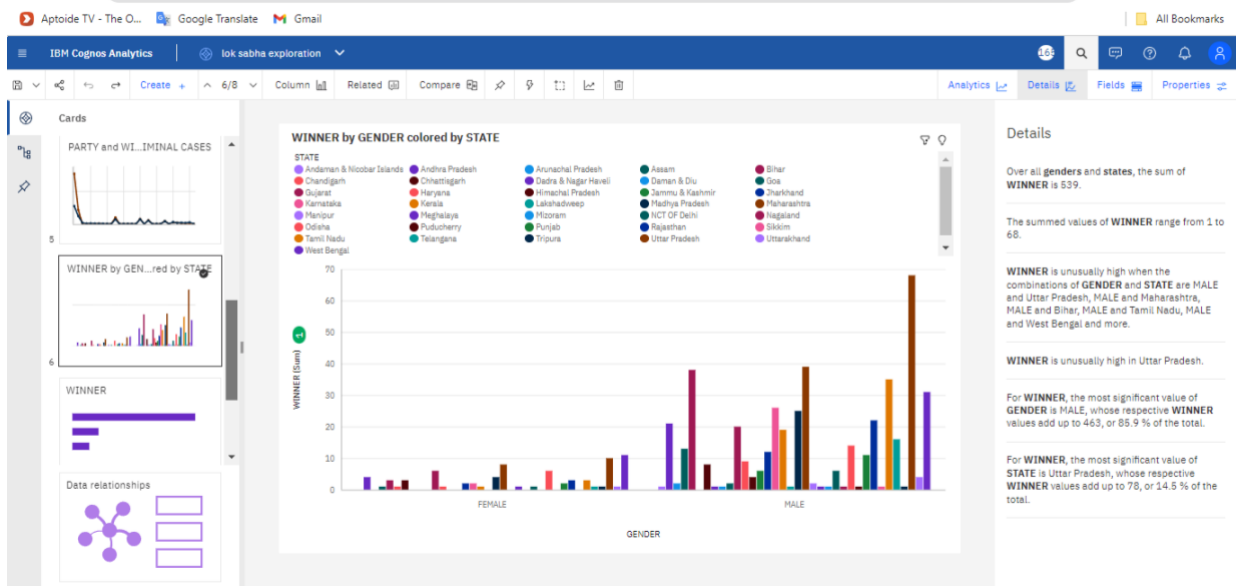
Dashboard:

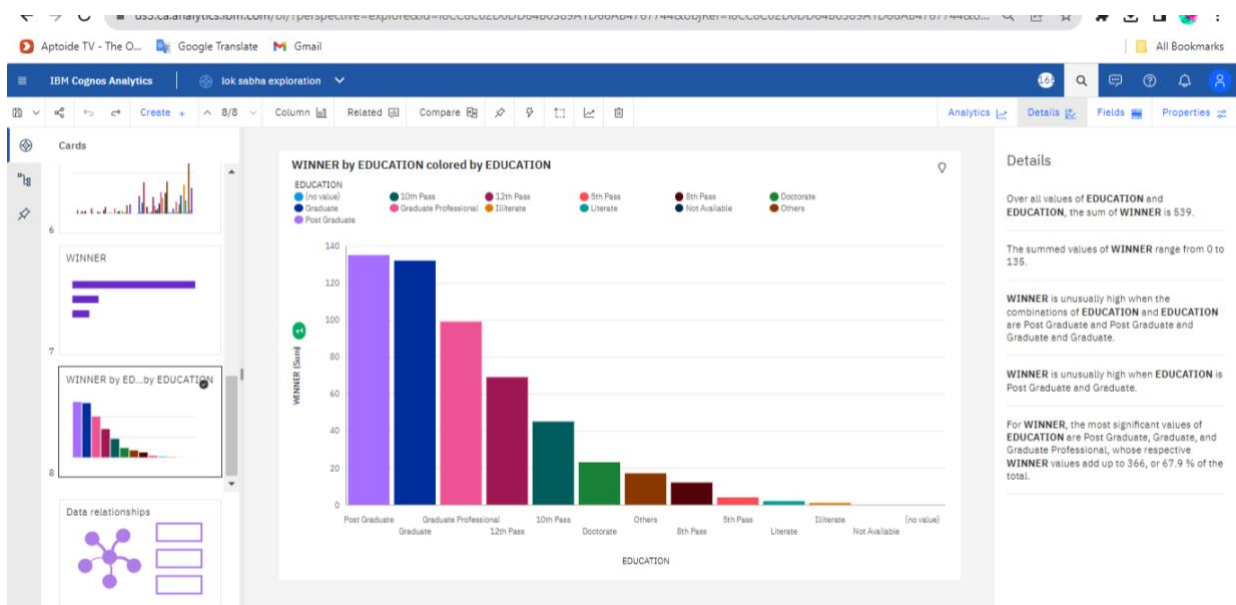
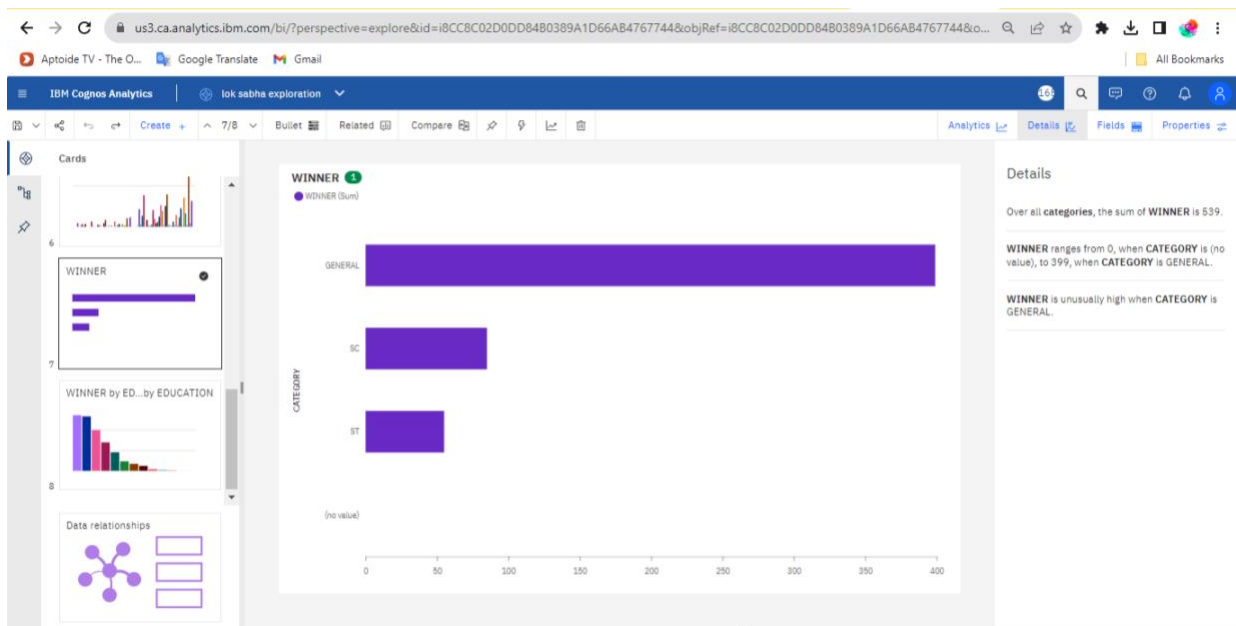




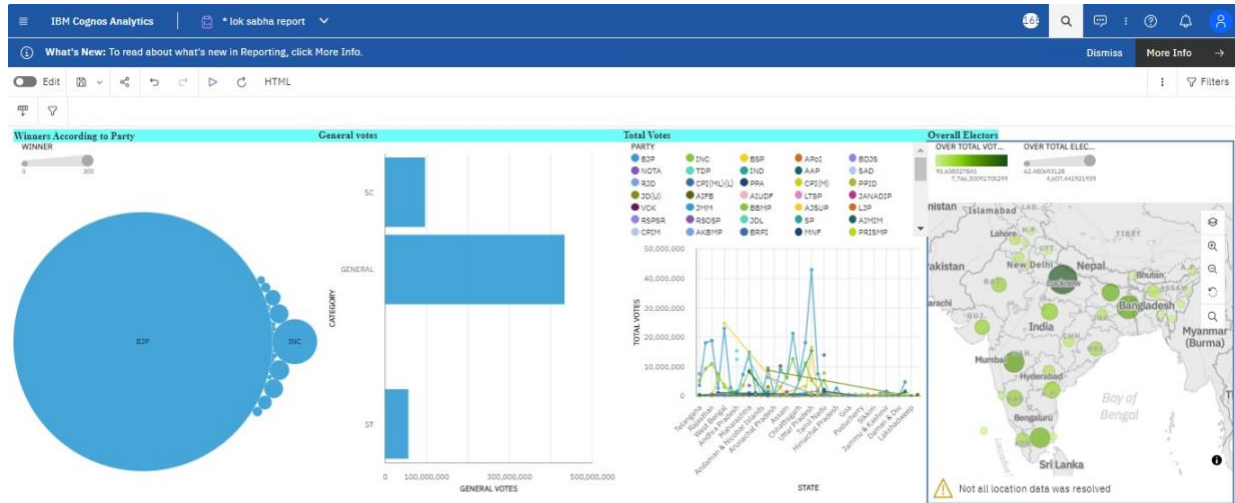
Story:







Report:



10 Advantages and Disadvantages

10.1 Advantages

1. **Data-Driven Insights:** Quantitative analysis provided data-driven insights into various aspects of the election, including candidate performance, voter demographics, and regional variations, allowing for a more objective assessment of the electoral process.
2. **Identifying Trends:** It helped in identifying electoral trends, patterns, and correlations between candidate attributes and electoral outcomes, which could inform future strategies.
3. **Informed Decision-Making:** Political parties could use quantitative analysis to make informed decisions about candidate selection, campaign strategies, and resource allocation based on the data.
4. **Accountability:** The analysis promoted accountability as it allowed for the assessment of the promises and performance of elected representatives.

10.2 Disadvantages

1. **Limited Scope:** Quantitative analysis may not capture qualitative aspects like the impact of social issues, local dynamics, or candidate charisma, which can be crucial in elections.
2. **Data Quality:** The accuracy and reliability of data can be a challenge, and inaccurate or biased data may lead to flawed conclusions.
3. **Overemphasis on Numbers:** Relying solely on quantitative data can overlook the nuanced and qualitative factors that influence voter behavior.
4. **Data Privacy Concerns:** Collecting and analyzing voter data can raise privacy concerns, especially if it involves personal information without proper consent.
5. **Potential for Manipulation:** Quantitative analysis can be manipulated to support specific narratives or agendas, leading to biased interpretations of election results.

11. CONCLUSION

In Certainly, here is a comprehensive conclusion for the "Quantitative Analysis of Candidates in the 2019 Lok Sabha Elections" report:

In conclusion, the quantitative analysis conducted on the 2019 Lok Sabha elections offered valuable insights into the intricate dynamics of the electoral process. By delving into candidate performance, voter demographics, and electoral trends, the analysis contributed

significantly to understanding the factors influencing the election outcomes. While the approach illuminated the importance of data-driven decision-making and strategic planning for political parties, it also highlighted the necessity of integrating qualitative research to comprehend the nuanced socio-political context. Therefore, for a comprehensive understanding of future elections, it is imperative to combine quantitative analysis with qualitative methodologies, ensuring a holistic approach to deciphering the complex nature of Indian electoral politics.

12. FUTURE SCOPE

The "Quantitative Analysis of Candidates in the 2019 Lok Sabha Elections" opens up several potential future avenues for research and application, including:

1. ***Predictive Modeling:** Building predictive models based on the data from the 2019 Lok Sabha elections can aid in forecasting future electoral trends, thereby assisting political parties in formulating more effective campaign strategies.
2. ***Longitudinal Studies:** Conducting longitudinal studies to track the changes in voter preferences and candidate performance over multiple election cycles can provide insights into the evolving dynamics of Indian politics and help anticipate future shifts.
3. ***Advanced Data Analytics:** Leveraging advanced data analytics techniques, such as machine learning and natural language processing, can enable a deeper understanding of voter sentiment, social media influence, and the impact of specific policy issues on electoral outcomes.
4. ***Policy Impact Assessment:** Using quantitative analysis to assess the impact of electoral outcomes on policy implementation and governance can aid in evaluating the effectiveness of various government initiatives and in shaping future policy agendas.
5. ***Election Process Optimization:** Employing quantitative analysis to streamline the election process, including voter registration, constituency delineation, and polling station management, can contribute to more efficient and transparent electoral procedures.

By delving into these future scopes, researchers and policymakers can continue to enhance the understanding of electoral dynamics and promote more effective and transparent democratic processes in India.

13. APPENDIX

13.1 Source code

- app.py
- index.html

13.2 GitHub

1. GitHub:

<https://github.com/Ssabari9/A-Quantitative-Analysis-in-Lok-Sabha-Election-2019>

2.Demo link:

https://drive.google.com/file/d/1jR3eS4qAzI6uFWt_NCN3c7sH4w5c4RWo/view?usp=drivesdk

