Spark Mini Project Report

Prepared By: **P.U.G. SASANK**
Course: **Big Data Analytics**
Date: **October 2025**

Dataset Description (State-wise_Voters_Information_2024_10000.csv)
- **Columns Identified:**
• State Name → Name of the state.
• Electors - Male / Female / Total → Total eligible voters (electors).
• Voters - Male / Female / Total → Actual votes cast.
• Voters - Poll % → Turnout percentage.
• NOTA Votes → "None of the Above" option count.
• EVM Rejected Votes → Rejected EVM votes.
• Tendered Votes → Votes re-cast due to identity issues.
• Valid Votes Polled → Count of accepted valid votes.

■ Observations from Executed Cells
1. Data Loading & Cleaning: Handled state-level totals ("Grand Total") by filtering. Columns renamed/aggregated for clarity.
2. State-Wise Analysis: Larger states (Uttar Pradesh, Maharashtra, Bihar) contribute maximum votes; smaller states/UTs like Sikkim, Goa, Mizoram have low numbers.
3. Gender-Wise Participation: Male and female participation fairly balanced. Some states show higher female turnout.
4. Turnout Analysis: Wide variation in turnout %, with highest >80% and lowest <60%.
5. NOTA Votes: Small but significant category, highest in politically competitive states.
6. EVM Rejected Votes: Very small fraction of valid votes. Top 5 states with highest rejection % highlighted.
7. Tendered Votes: Negligible across all states, on average <0.01% of valid votes.
8. Final National-Level Comparison: Valid Votes dominate (>98%), Rejected and Tendered votes negligible, NOTA reflects dissatisfaction.

■ Plots Observed
• State-wise voter totals (bar chart)
• Male vs Female voters (grouped bar chart)
• Turnout % by state (horizontal bar)
• Top 10 states with highest NOTA votes (bar chart)
• EVM rejected vs valid votes (pie + multi-pie)
• Tendered vs valid votes (absolute + %)
• National-level breakdown (final bar chart)

Key Insights
- ■ Valid Votes make up almost all of the electoral participation.
- ■ EVM Rejected votes are extremely rare, showing machine reliability.
- ■ Tendered Votes are negligible.
- ■ NOTA Votes reflect democratic choice and voter dissatisfaction.
- ■ Larger states dominate voter numbers, but trends are consistent across India.

Recommendations
1. Voter Awareness Campaigns → improve turnout in low-performing states.
2. Gender-Focused Outreach → in states with lower female turnout.
3. Electoral Confidence Building → maintain low EVM rejection, reduce tendered votes.
4. Study NOTA Trends → policy research on high-NOTA states.

Conclusion
The State-wise Voters Information 2024 dataset highlights strong participation across states, balanced gender representation, and high validity of votes.
- Valid votes dominate (>98%).

- Rejected/Tendered votes are negligible.
- NOTA votes exist in measurable numbers, indicating democratic maturity.

■ Final Note: The analysis confirms a robust election process with high participation and minimal anomalies.