

# Analyzing 54 Years of Pakistan's Democratic Journey

Data-Driven Insights from General  
Elections (1970-2024)

Syed Hunain Abbas Zaidi  
Hammad Malik  
Mian Abdul Wasay  
Hayyan Khan

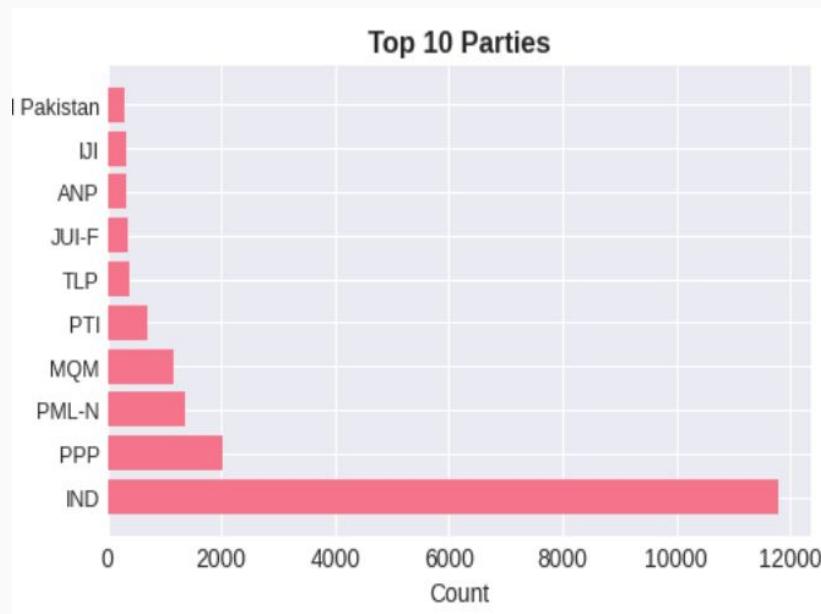


# Project Overview & Objectives

- The core of our study is the Gallup Pakistan election dataset, which gives us **24,585 constituency-level records** spanning **11 national elections**. This rich dataset allows us to study political trends over five decades.
- Our main objective is to establish predictive power in this complex, multi-party system and understand **what factors drive dominance** at the constituency level.
- Research questions which we have completed so far :
  1. Party Evolution: How have regional party strongholds shifted over 54 years, especially PTI's rise in urban areas? (Done in Previous Presentation)
  - 2.

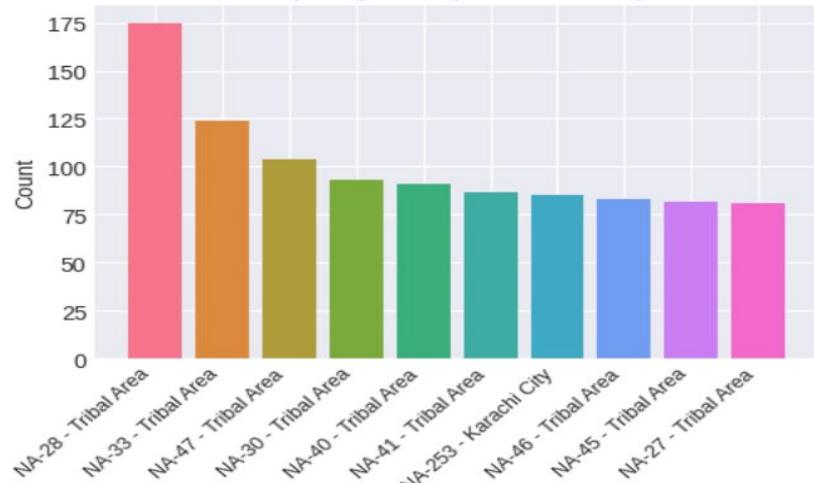


# Data Exploration (EDA)

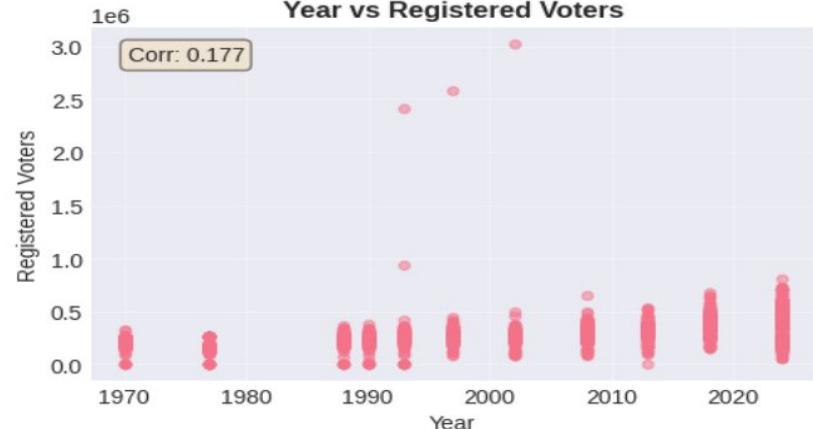


- The raw data had 17 columns, but we focused on **10 key features**, using the geographic and temporal data as our core predictors.
- We chose these because the literature suggests **regional strongholds** are crucial in Pakistani politics.
- Our EDA quickly confirmed the complexity of the data: we have **475 unique parties**, with **independents dominating the raw count**.
- This confirmed our need for the **Top Parties vs. Others** classification approach.

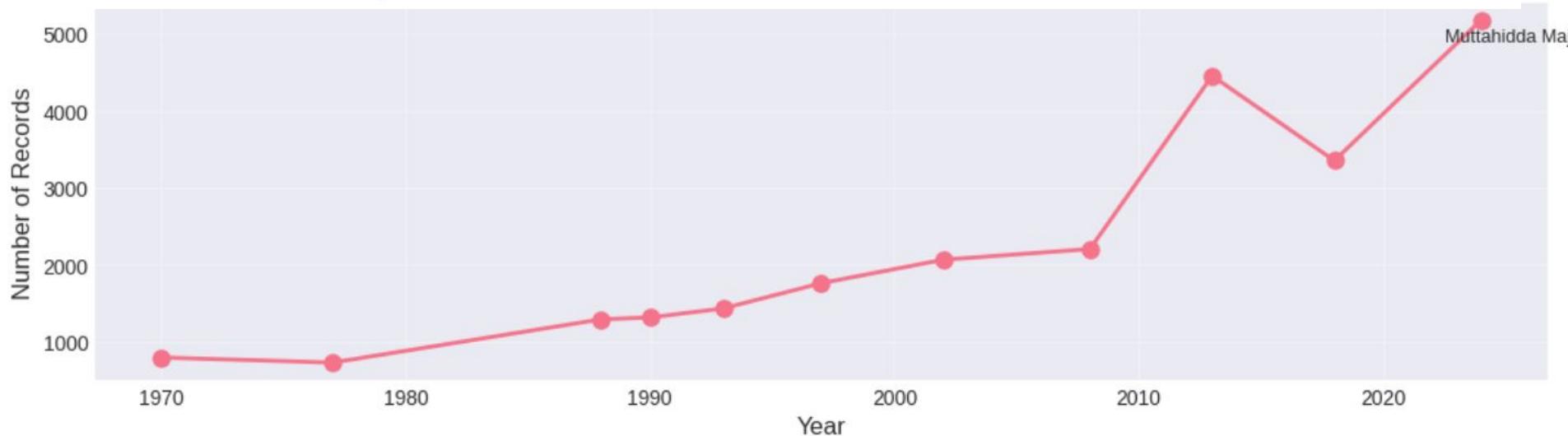
### Top Regions by Constituency



### Year vs Registered Voters

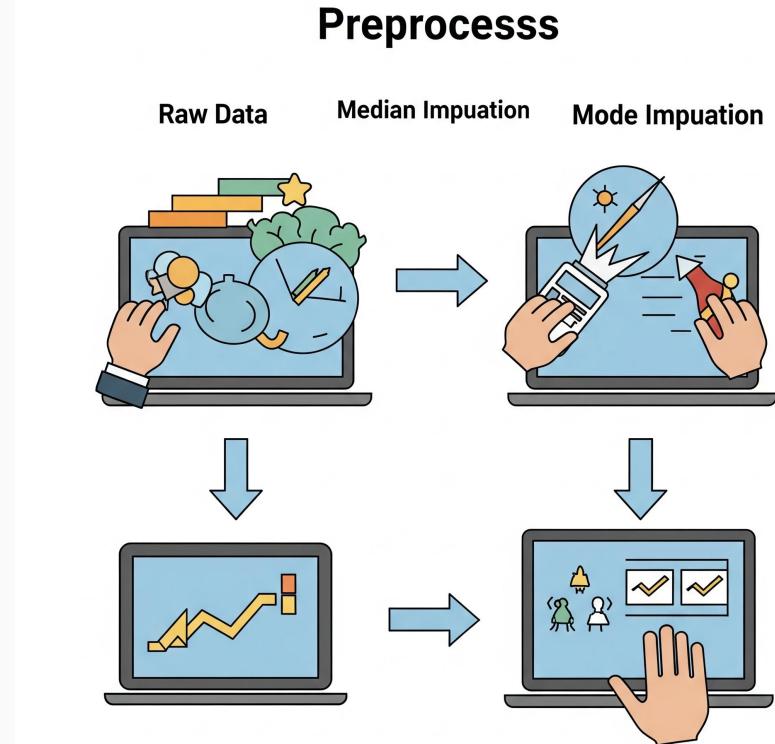


### Number of Records



# Data Preprocessing Steps

- We faced significant missing data, especially in the voter/turnout columns in early years. We used **median imputation for numerical data** to avoid skewing the distributions and **mode imputation for categorical data** to maintain the most common category.
- We transformed 7 categorical features using Label Encoding & applied **StandardScaler** to normalize numerical data like Year and Votes.



# PTI's Urban Revolution

## The Data Story



### PTI's Unique Profile:

- 47.2% urban composition - highest among all major parties
- From 1 seat (1990) to 242 seats in 2018 - exponential growth
- 31.8% vote share in 2018 - historic peak
- 703 total seats won with 25M votes across all elections

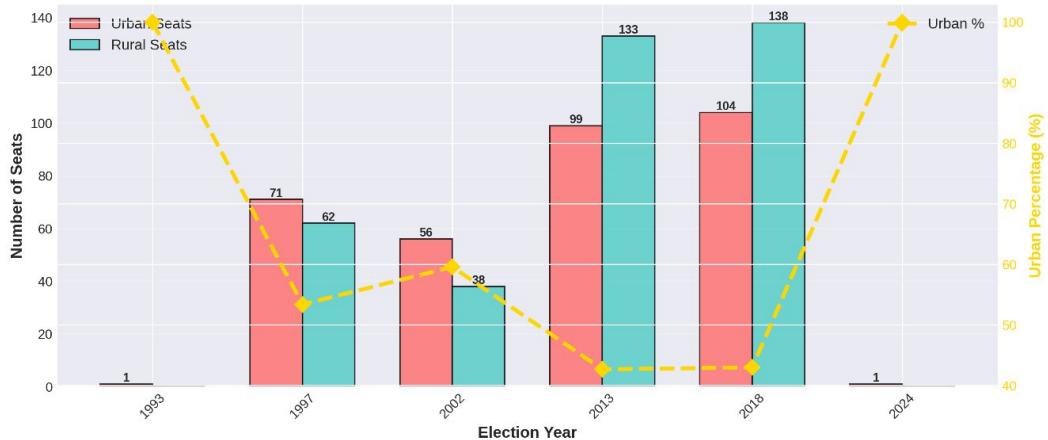
### Comparative Context:

- IND: 40.4% urban (11,790 seats total)
- PPP: 45.5% urban (traditional strength)
- PML-N: 43.6% urban (Punjab-focused)

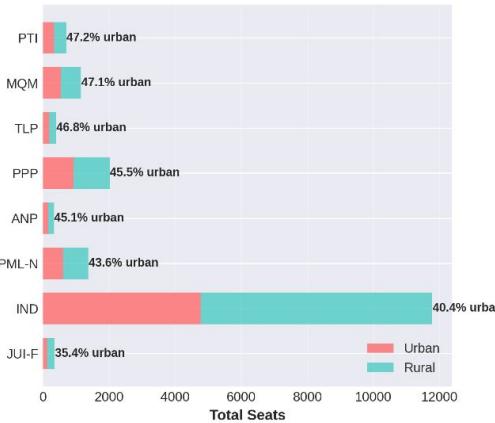
PTI transformed from a fringe party to Pakistan's leading urban political force in just 28 years, fundamentally reshaping the electoral landscape.

# PTI's Urban Revolution: Rise of a New Political Force

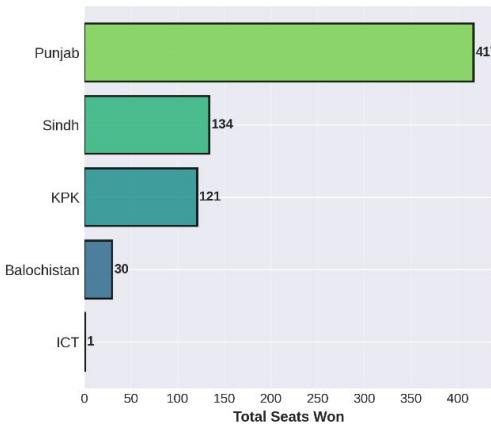
## PTI's Urban Dominance: Seats and Urban Percentage Over Time



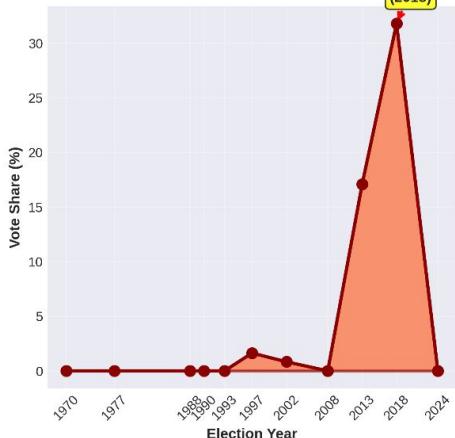
## Urban vs Rural: All Major Parties



## PTI: Provincial Strongholds



## PTI Vote Share Evolution



## Performance Summary: Key Metrics

Party	Seats	Urban %	Total Vote	Avg/Seat
IND	11,790	40.4%	33.9M	2,878
PPP	2,016	45.5%	76.2M	37,802
PML-N	1,371	43.6%	68.9M	50,269
MQM	1,149	47.1%	10.6M	9,268
PTI	703	47.2%	25.0M	35,609
TLP	395	46.8%	5.0M	12,767

# Provincial Strongholds Analysis

Punjab (the battleground):

- PTI: 417 seats | IND: 2,492 seats | PML-N: 891 seats
- Most competitive province with mixed dominance

Sindh (PPP fortress):

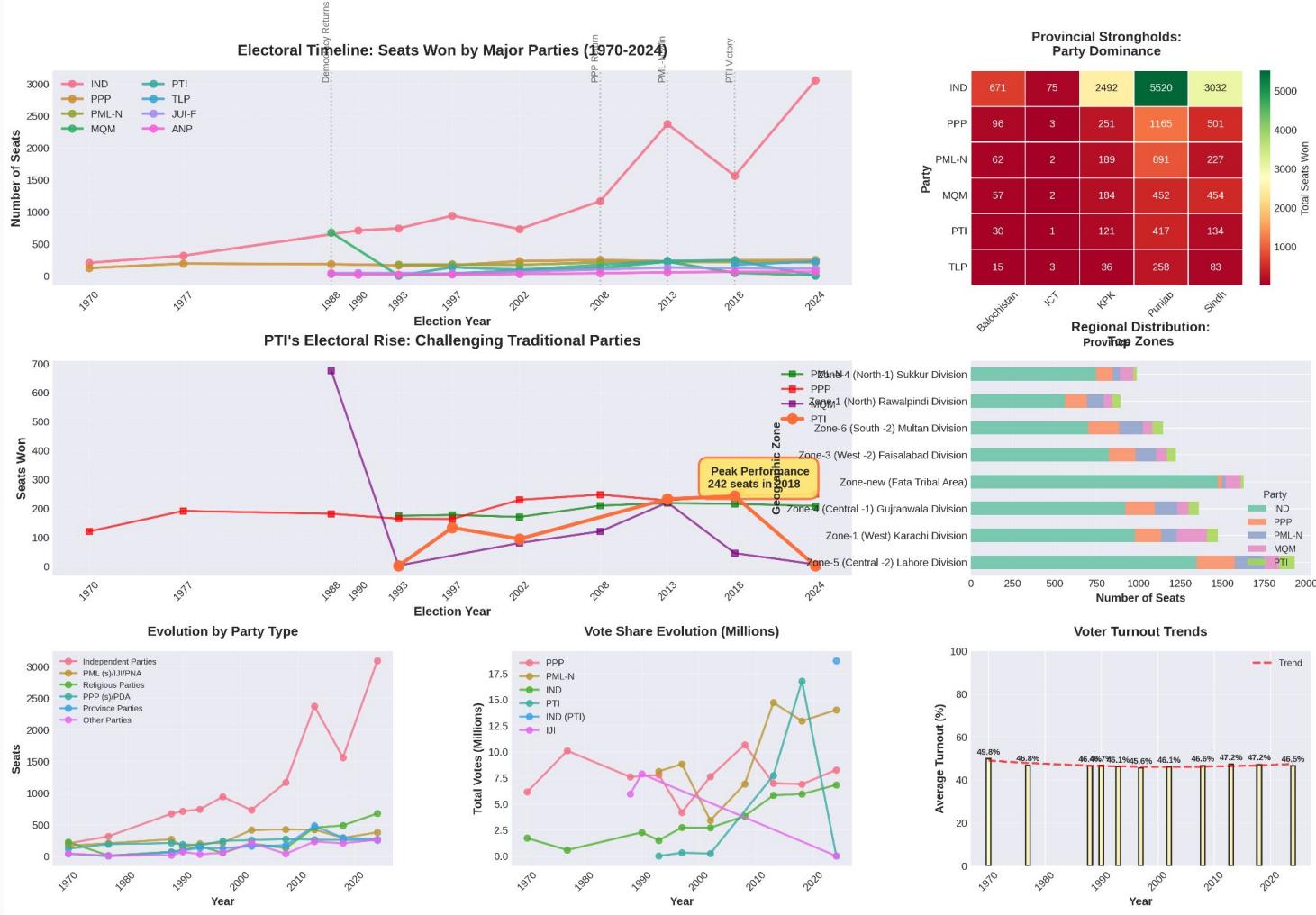
- PPP: 1,165 seats | MQM: 452 seats (urban Karachi)
- Clear urban-rural divide

KPK & Balochistan:

- PTI: 121 seats in KPK (growing influence)
- ANP historically strong in KPK
- Balochistan remains fragmented (30 seats for PTI)

Regional identity remains the strongest predictor of electoral success - no single party dominates nationally.

# Pakistan Elections: 54 Years of Regional Strongholds & Political Evolution (1970-2024)



# Key Insights & Findings

Content: 1. PTI's Urban Revolution is Real

- 47.2% urban base vs. 40-45% for traditional parties
- Explosive growth: 1 seat (1990) → 242 seats (2018)

2. Regional Strongholds Persist

- Punjab: Competitive battleground (PTI, PML-N, IND)
- Sindh: PPP dominance with MQM in urban pockets
- KPK: PTI gaining ground from traditional parties

3. Independent Candidates Remain Powerful

- 11,790 independent candidacies across 54 years
- Reflects local clan/biradari politics over party ideology

4. Electoral Expansion

- Registered voters grew from ~20M (1970) to 128M (2024)
- Candidate pool exploded post-2008 democratic restoration

# Challenges & Limitations

Content: Data Challenges:

- Missing voter data in early elections (pre-1990)
- Inconsistent constituency boundaries across decades
- Party mergers/splits complicate temporal analysis (e.g., PML factions)

Analytical Complexities:

- 475 parties make traditional classification difficult
- Regional heterogeneity requires province-specific models
- Urban/rural binary may oversimplify constituency demographics

Methodological Decisions:

- Label encoding loses ordinal relationships but enables modeling
- Top 6 vs. Others classification sacrifices granularity for predictive power
- Median imputation assumes missing-at-random (may not hold for early data)

Next Steps: Advanced models (Random Forest, XGBoost) to improve prediction and feature importance analysis.

Thank you