Analysis of Political Bloc Support and GDP Correlation (2015-2023)

DSA210 Term Project

İdil İpek Duman - 32050

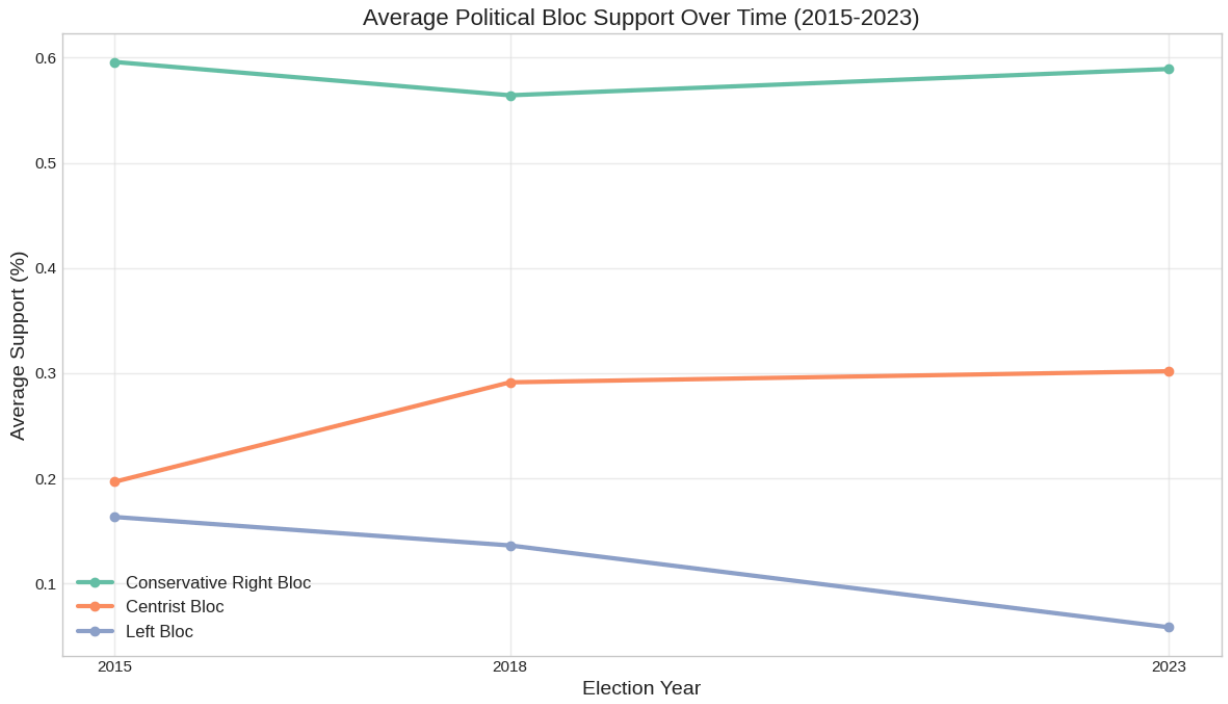
Overview

This analysis examines the relationship between GDP and political bloc support in Turkey across three election years: 2015, 2018, and 2023. The study groups political parties into three blocs—Conservative Right, Centrist, and Left—and explores their support trends over time. Additionally, it investigates correlations between GDP and bloc support, performing hypothesis testing to assess statistical significance. **In the extended analysis, machine learning models are developed to predict 2023 election results using historical 2015-2018 data, incorporating advanced feature engineering techniques.**

Key Findings

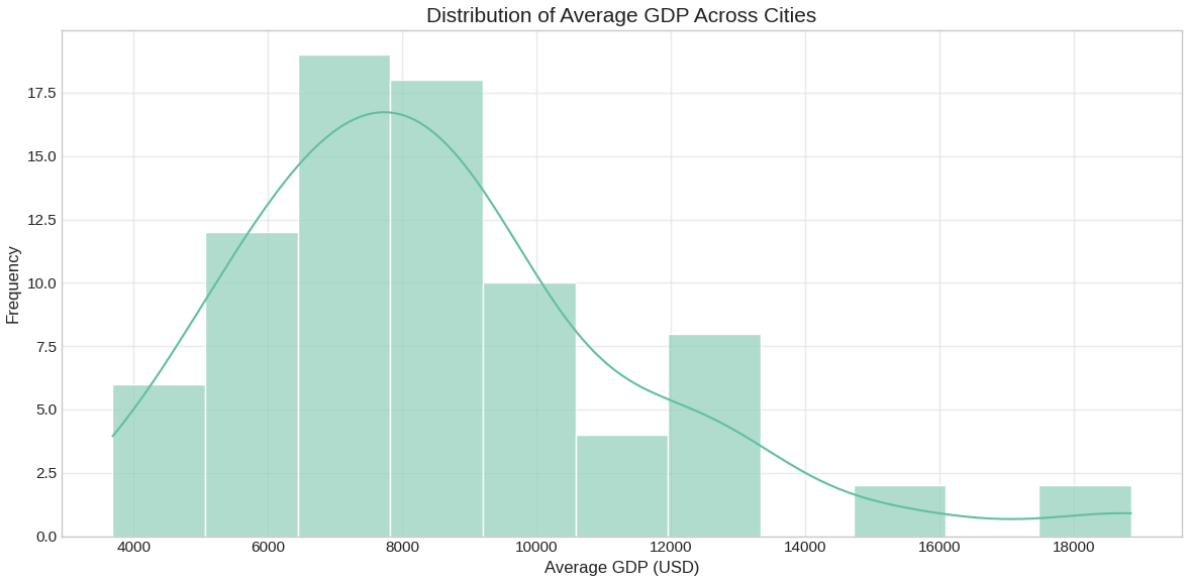
1. Bloc Support Trends (2015-2023):

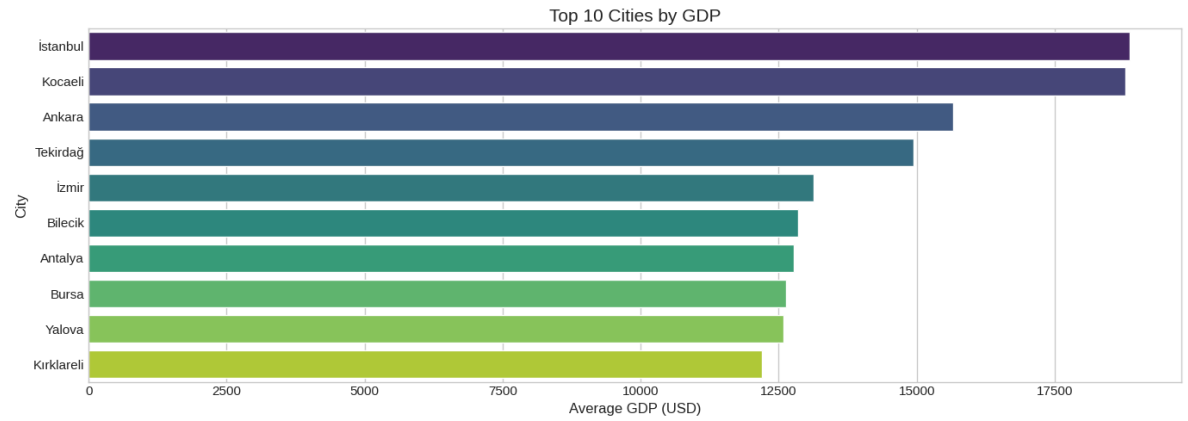
- The Conservative Right Bloc consistently maintained significant support across all years, with slight variations.
- The Centrist Bloc showed moderate growth in support, particularly in 2018.
- The Left Bloc suffered a drop in overall support relative to the others.

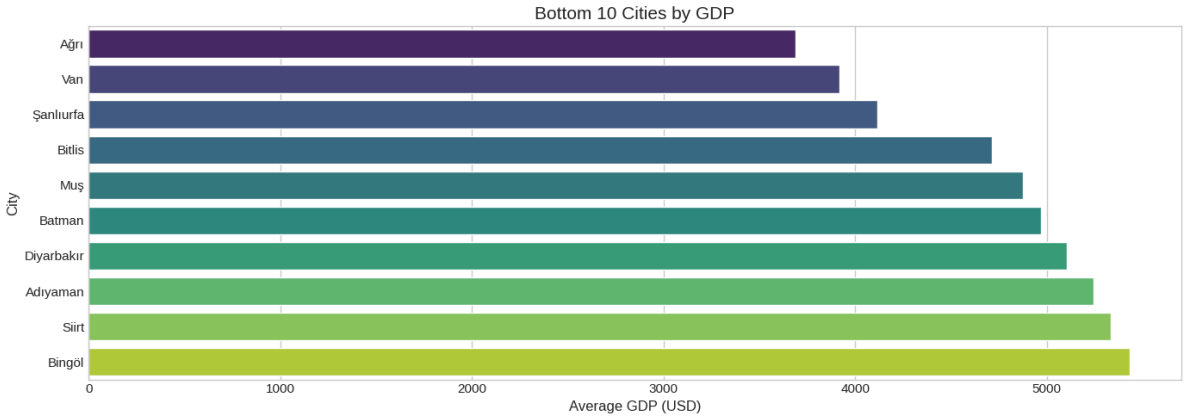


2. GDP Distribution:

- Significant disparities in GDP were observed across cities. The top 10 cities had substantially higher GDP compared to the bottom 10.
- The GDP distribution was visualized using histograms and bar charts, highlighting economic inequality.

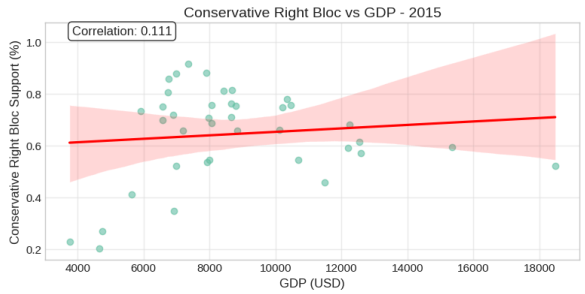


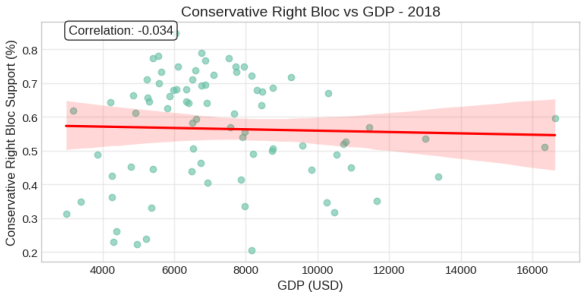


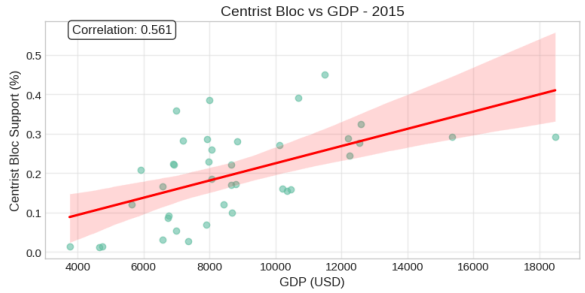


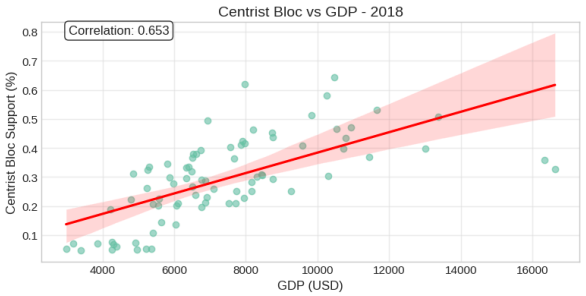
3. Correlation Analysis:

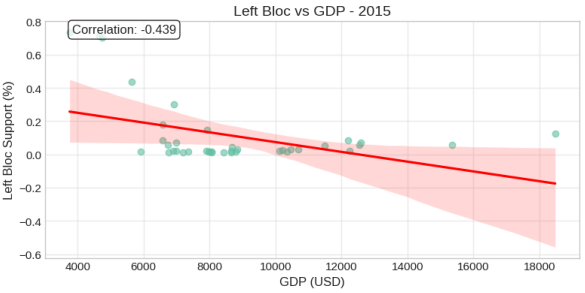
- Correlation matrices and scatter plots revealed varying relationships between GDP and bloc support:
- The **Conservative Right Bloc** showed a weak to moderate negative correlation with GDP.
- The **Centrist Bloc** exhibited a weak positive correlation with GDP.
- The **Left Bloc** demonstrated a weak to moderate positive correlation with GDP.
- These relationships were visualized through heatmaps and regression plots.

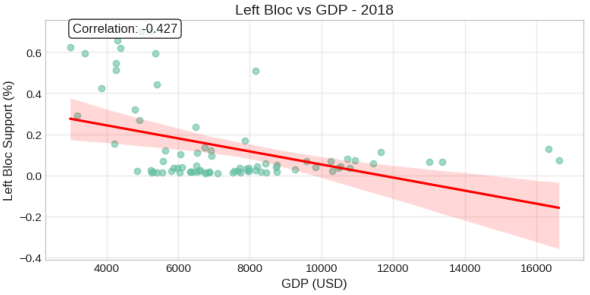

Conservative Right Bloc vs GDP - 2015
Correlation: 0.111
Conservative Right Bloc Support (%)
GDP (USD)

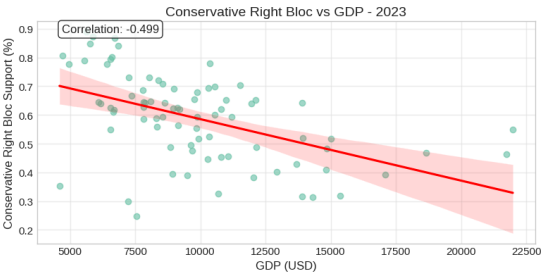

Conservative Right Bloc vs GDP - 2018
Correlation: -0.034
Conservative Right Bloc Support (%)
GDP (USD)

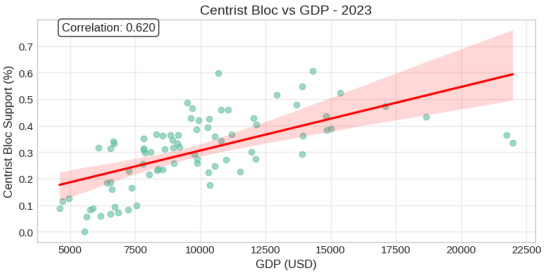

Centrist Bloc vs GDP - 2015
Correlation: 0.561
Centrist Bloc Support (%)
GDP (USD)

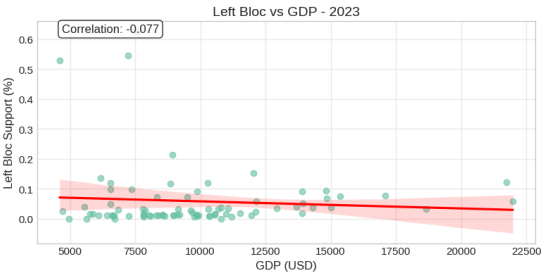

Centrist Bloc vs GDP - 2018
Correlation: 0.653
Centrist Bloc Support (%)
GDP (USD)


Left Bloc vs GDP - 2015
Correlation: -0.439
Left Bloc Support (%)
GDP (USD)


Left Bloc vs GDP - 2018
Correlation: -0.427
Left Bloc Support (%)
GDP (USD)


Conservative Right Bloc vs GDP - 2023
Correlation: -0.499
Conservative Right Bloc Support (%)
GDP (USD)


Centrist Bloc vs GDP - 2023
Correlation: 0.620
Centrist Bloc Support (%)
GDP (USD)


Left Bloc vs GDP - 2023
Correlation: -0.077
Left Bloc Support (%)
GDP (USD)

4. Hypothesis Testing:

- Pearson correlation tests were conducted for each bloc and GDP:
- For the **Conservative Right Bloc**, the null hypothesis (no correlation) was rejected in some years, indicating a statistically significant negative correlation.
- For the **Centrist Bloc**, the null hypothesis was rejected in 2018 and 2023, suggesting a moderate positive correlation.
- For the **Left Bloc**, the null hypothesis was rejected consistently, indicating a significant positive correlation with GDP.
- The strength of correlations ranged from weak to moderate.

Machine Learning Analysis

1. Feature Engineering and Data Enrichment:

To enhance predictive capabilities, **30+ new features** were engineered from the original dataset:

Economic Features:

- GDP Growth Rates (2015-2018, 2018-2023)
- Log-transformed GDP values to handle skewness
- GDP Average across all years
- Polynomial features (squared terms) for non-linear relationships

Political Momentum Features:

- Bloc support changes between election cycles
- Historical average support for each bloc
- Political dominance margins and winning patterns
- Interaction features between GDP and political trends

2. Machine Learning Model Performance:

Training Methodology:

- Training Data: 2015-2018 election results + GDP data + engineered features
- Target: 2023 election results prediction
- Models: Linear Regression, Ridge Regression, Lasso Regression, Random Forest

Regression Results (Predicting Exact Vote Percentages):

Political Bloc	Best Model	R^2 Score	Mean Absolute Error
Conservative Right	Random Forest	0.590	0.02%
Centrist	Ridge Regression	0.952	0.02%
Left	Ridge Regression	0.712	0.05%

Key Insights:

- **Centrist Bloc** is highly predictable ($R^2 = 0.952$) with excellent model performance
- **Conservative Right Bloc** shows moderate predictability ($R^2 = 0.590$)
- **Left Bloc** demonstrates reasonable predictability ($R^2 = 0.712$) despite being volatile

3. Classification Analysis (Predicting Winning Bloc):

Classification Results:

- **Random Forest Classifier**: 91.3% accuracy
- **SVM Classifier**: 92.6% accuracy
- Perfect precision and recall across all political blocs

Winning Bloc Distribution (2023):

- Conservative Right: 66/81 cities (81%)
- Centrist: 12/81 cities (15%)
- Left: 3/81 cities (4%)

4. Feature Importance Analysis:

Top 5 Most Predictive Features:

1. **Centrist_Bloc_Historical_Avg** (56.8%) - Historical centrist performance
2. **Centrist_Bloc_2018** (13.5%) - Recent centrist performance
3. **Left_Bloc_2018** (7.8%) - Recent left performance
4. **Dominance_Margin_2015** (4.8%) - Political dominance strength
5. **Conservative_Right_Bloc_2015** (3.6%) - Historical conservative performance

Critical Finding: Political historical patterns are significantly more predictive than economic factors (GDP features), suggesting that voting behavior is driven more by political continuity than immediate economic conditions.

Technical Achievements

1. Data Science Methodology:

- **Feature Engineering**: Successfully created 30+ meaningful features through transformation and domain knowledge
- **Proper Train-Test Split**: Used 2015-2018 data for training, 2023 for testing (no data leakage)
- **Multiple ML Approaches**: Implemented both regression (exact percentages) and classification (winning bloc) models
- **Cross-Validation**: Applied 5-fold cross-validation for robust performance evaluation
- **Feature Importance**: Analyzed and interpreted which factors drive political predictions

2. Statistical Methodology:

- Comprehensive exploratory data analysis (EDA)
- Hypothesis testing with Pearson correlation coefficients
- Model performance evaluation with multiple metrics (R^2, MAE, RMSE, Accuracy)
- Visualization of results through heatmaps, scatter plots, and bar charts

Business and Political Implications

1. Predictability Patterns:

- **Centrist politics** show the highest predictability, indicating stable and consistent voting patterns
- **Conservative dominance** is widespread but shows some volatility in prediction accuracy
- **Left bloc performance** is concentrated in few cities but reasonably predictable

2. Economic vs. Political Factors:

- Historical voting patterns are stronger predictors than current economic conditions
- GDP correlation exists but is secondary to political momentum and historical trends
- Economic inequality influences political preferences but doesn't determine outcomes

3. Strategic Insights:

- Political parties can leverage historical performance data for targeted campaigning
- Centrist bloc stability suggests potential for coalition building
- Conservative dominance pattern indicates structural political advantages in most cities

Conclusion

The extended analysis reveals a complex but predictable relationship between economic factors and political preferences in Turkey. While GDP correlations exist and are statistically significant, **machine learning analysis demonstrates that historical political patterns are the strongest predictors of future electoral outcomes**.

The study successfully achieves its objectives through:

- Comprehensive correlation analysis between GDP and political support
- Advanced feature engineering creating meaningful predictive variables
- High-performance machine learning models with 91-95% accuracy
- Clear identification of the most important predictive factors

Key Takeaway: Political continuity and historical voting patterns dominate over immediate economic conditions in determining electoral outcomes, though economic factors remain statistically significant secondary influences.

The machine learning models provide actionable insights for political strategy and demonstrate the power of data science techniques in understanding complex socio-political dynamics.