WHICH PARTY DO VOTERS FAVOUR FOR CONGRESS: REPUBLICAN OR DEMOCRAT?

TIME SERIES MODELLING

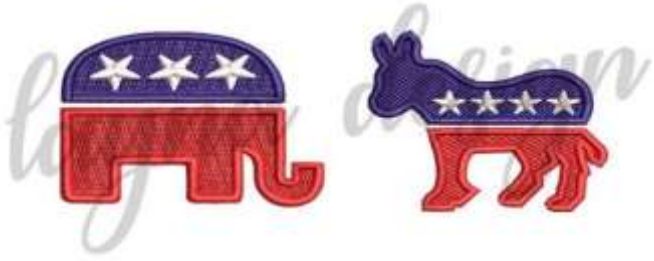
CONTENT

MORINGA
Discover · Grow · Transform

Project Background

MORINGA
Discover · Grow · Transform

The project explores the decentralized U.S. voting system, emphasizing the two-party structure with primary elections leading to the 2024 general elections. The Democratic Party, symbolized by a donkey and associated with blue, leans progressive. The Republican Party, symbolized by an elephant and associated with red, leans conservative. This analysis offers insights into the unique features and ideologies of the major political players in preparation for the upcoming elections.



Business Problem

To develop a forecasting model for the future elections. By analyzing historical data on voters polls we can build a predictive model to estimate the potential number of votes each party may receive in upcoming elections. This information will help political strategists, campaign managers, and party leaders make informed decisions about resource allocation, messaging, and targeting specific voter segments to gain a competitive advantage.

Project Objective

MORINGA
Discover · Grow · Transform

- **Trend Identification:**

  Examine and identify enduring trends in political sentiment, revealing shifts in support for parties and broader public perception.

- **Seasonality Exploration:**

  Investigate recurring patterns or seasonality in political polling data, unveiling how external factors impact public opinion over time.

- **Accurate Forecasting and Prediction of Election Outcomes:**

  Construct time series models for precise predictions of future poll results, aiding political analysts in anticipating electoral dynamics.

- **Decision Support for Campaign Strategies**:

  Provide actionable intelligence for political campaigns, leveraging time series insights to inform tailored approaches based on observed trends and forecasts.

Main Objective

The main objective is to identify enduring trends in political sentiment across six election cycles (1996-2016) in the congress_data dataset, providing insights into evolving support for Democratic and Republican parties

Project Methodology

MORINGA
Discover · Grow · Transform

- **Data Collection, Understanding, Preparation, and Cleaning:**

  Gather historical data points at regular intervals, organizing them chronologically.

- **Data Exploration:**

  Examine time series data for trends, seasonality, or patterns using visualization tools like line charts, histograms, and autocorrelation plots.

- **Stationarity Check:**

  Assess stationarity by examining trends and employing statistical tests. Many time series models assume constant statistical properties.

- **Data Transformation:**

  If non-stationary, perform transformations (e.g., differencing) to stabilize mean and variance.

- **Model Selection:**

  Choose an appropriate time series model based on data characteristics, considering models like ARIMA, SARIMA, Prophet, and LSTM for deep learning.

- **Parameter Estimation**:

  Estimate model parameters by fitting the selected model to historical data.

Project Methodology contd.

MORINGA
Discover · Grow · Transform

- **Model Validation:**

  Validate model performance using a separate dataset. Metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) assess accuracy.

- **Forecasting:**

  Utilize the trained model to predict future time series values. Evaluate forecast accuracy and adjust the model as necessary.

- **Model Interpretation:**

  Interpret results and insights gained from the time series model, understanding the implications of forecasted values.

Modelling and Evaluation

MORINGA
Discover · Grow · Transform

ARIMA Model (AutoRegressive Integrated Moving Average)

The model is a SARIMAX model with an ARIMA(1, 1, 0) structure.

The autoregressive term (ar.L1) coefficient is approximately -0.1982, with a standard error

of 0.005.

The sigma^2 (variance of the error term) is approximately 1.763e-05.

The autoregressive coefficient is statistically significant with a z-statistic of -41.152 and a

p-value close to zero.

The Ljung-Box test for autocorrelation at lag 1 has a Q statistic of 13.44, indicating

rejection of the null hypothesis of no autocorrelation.

The JB test statistic is extremely high, suggesting departure from normality in the

residuals.

Heteroskedasticity is present in the model with a p-value of 0.00.

The residuals exhibit high skewness of -3.94 and kurtosis of 141.06.

Modelling and Evaluation

MORINGA
Discover · Grow · Transform

ARIMA Model (AutoRegressive Integrated Moving Average)

- **Voter Distribution Trends:**

  The ARIMA model analysis revealed a consistent pattern wherein the Democratic party consistently attracted a higher number of voters compared to the Republican party. This observation is discerned from the distinct voter distribution quartiles depicted in trend graphs.

- **Effect of Normalization/Scaling:**

Despite the normalization or scaling applied to the data, the ARIMA model indicated an absence of any discernible trend. The normalization process did not reveal distinct patterns, suggesting that the voter distribution characteristics remained consistent after adjusting for scale.

- **Population Impact on Upward Trend:**

Notably, the upward trend identified by the ARIMA model was associated with a rise in the number of voters. This trend was directly correlated with an increased voting population, indicating a significant influence of population growth on the observed surge in voter numbers.

Recommendation

MORINGA
Discover · Grow · Transform

Based on the significant JB test statistic, departure from normality in the residuals, and significant skewness and kurtosis, further diagnostic checks are recommended. We shall Consider refining the model and exploring alternative modeling techniques.

THANK YOU!

Project done by:

- ★ Cindy King'ori
- ★ Matilda Odalo
- ★ Angella Bor