Project Report

Analysis of Municipal Factors and the 2019 Parliamentary Elections in Finland

Introduction

Elections are a critical component of democratic societies, providing citizens the opportunity to choose their representatives. However, understanding the factors that influence election results can be complex and challenging. Municipalities are often diverse in terms of their demographics, economic conditions, and social factors, which can impact voting patterns and outcomes. This report aim to examine the role of municipal factors in shaping the outcomes of the 2019 parliamentary elections in Finland. Through the analysis of data from Statistics Finland's free databases, I explore the relationships between various factors such as population size, degree of urbanization, the share of foreign citizens, the share of household-dwelling units living in different types of houses, employment rate, the proportion of the unemployed among the labor force, and workplace self-sufficiency, and their impact on the election results for individual parties and aggregated groups of parties. The key findings of this analysis reveal significant relationships between certain socio-economic factors and election outcomes, contributing to a deeper understanding of the political landscape in Finland.

The 2019 Finnish parliamentary elections were a pivotal moment in the nation's political landscape, as they marked a shift in power dynamics among the major parties. The election's outcome revealed a fragmented and diverse political environment, with the Social Democratic Party (SDP) narrowly becoming the largest party for the first time since 1999. In this report, we examine the role of municipal factors in shaping the outcomes of these significant elections

Dataset Description

The datasets used in this analysis were obtained from Statistics Finland's free databases. The first dataset, "Kuntien avainluvut", contains data about individual municipalities, including information on:

- Degree of urbanisation: The percentage of the population living in urban areas.
- Population: The total number of people living in a municipality.
- Share of foreign citizens of the population: The percentage of the population in a municipality that is foreign citizens.
- Share of household-dwelling units living in terraced houses and detached houses: The percentage of households in a municipality living in terraced houses and detached houses.
- Share of household-dwelling units living in rental dwellings: The percentage of households in a municipality living in rental dwellings.
- Employment rate: The percentage of the working-age population (usually defined as 15-64 years old) that is employed in a municipality.
- Proportion of the unemployed among the labour force: The percentage of unemployed people among the total labor force in a municipality.
- Workplace self-sufficiency: A measure of the balance between the number of jobs available in a municipality and the number of working-age people living there. A value greater than 1 indicates that there are more jobs than working-age people, while a value less than 1 indicates the opposite.

The second dataset, "Km-ku", contains information on the results of the 2019 parliamentary elections in Finland at the municipal level. The dataset includes information on the proportion of votes cast for several political parties, including:

- Centre Party (KESK)
- Finns Party (PS)
- National Coalition Party (KOK)
- Social Democratic Party (SDP)
- Green League (VIHR)
- Left Alliance (VAS)
- Swedish People's Party of Finland (RKP)
- Christian Democrats (KD)
- Other parties or independent candidates.

The datasets were obtained in Excel format and imported into Python for analysis using various statistical techniques, including regression analysis and correlation analysis. The Pandas library was used for data manipulation and cleaning, while the Matplotlib and Seaborn libraries were used for data visualization. These methods were used to explore the relationships between the variables in the two datasets, and to identify any significant correlations or patterns that may help explain the outcomes of the 2019 parliamentary elections in Finland.

Methodology

The analysis involved using a range of statistical methods and techniques to explore the relationships between the variables in the two datasets and identify any significant correlations or patterns that may help explain the outcomes of the 2019 parliamentary elections in Finland.

I began by conducting a descriptive analysis of the data using Pandas in Python to gain an understanding of the distribution of the variables and to identify any potential outliers or missing values. Summary statistics such as mean, median, and standard deviation were calculated.

Next, correlation analysis was used to examine the relationships between the variables in the two datasets. I used the Pearson correlation coefficient to measure the strength of the linear relationship between pairs of variables and visualized the results using scatter plots and heat maps. Regression analysis was also conducted to examine the relationship between the election results and municipal factors and to identify any significant predictors of the election outcomes.

Finally, I used data visualization techniques, including Matplotlib and Seaborn libraries, to communicate the findings. Various graphs and charts such as bar charts, line charts, and scatter plots were created to visualize the relationships between the variables and highlight any significant patterns or trends.

Using these methods and techniques, I was able to gain insights into the relationships between the variables in the datasets and identify the factors that may have influenced the outcomes of the 2019 parliamentary elections in Finland.

Data Cleaning and Preparation

a. Merging the dataset

Before conducting the analysis, it was necessary to clean and preprocess the data to ensure that it was in a suitable format. Firstly, the 'kuntien_avainluvut' dataset was transposed to match the format of the 'Km-ku' dataset. The two data frames were then merged using an inner merge based on the municipality names. This resulted in a dataset that contains only the municipalities with complete data for both election results and socio-economic indicators.

The merging process ensured that the analysis was based on a complete dataset, allowing for a more reliable exploration of potential relationships between election results and socio-economic factors. The resulting dataset contained 311 data points, representing 311 different municipalities in Finland, and 17 variables, including 8 socio-economic factors and 9 election results indicators.

The merging process was performed using Python in the Jupyter Notebook environment. The Pandas library was used for data manipulation and cleaning, and the resulting merged dataset was saved in Excel format for further analysis.

b. Imputing the data

Upon further examination, it was discovered that the dataset had 95 missing values for the 'RKP-Proportion of votes cast' indicator. Given the high number of missing values, it was necessary to perform exploratory analysis to identify any patterns or reasons for the missing data. After thorough analysis, it was observed that municipalities with missing values tended to be smaller, more rural, and have a lower degree of urbanization. Notably, these municipalities had a degree of urbanization of 56.12%, which was lower than the average of the entire dataset (63.57%). In addition, the average population of these municipalities was 12,289, which was significantly lower than the average population of the entire dataset (37,151).

To address the issue of missing data, a Stratified Imputation method was employed, involving the imputation of missing values based on the characteristics of the municipalities. Since the 'Degree of urbanization' appeared to have a more direct relationship with the nature of the municipalities (urban or rural), it was selected as the stratification variable. Groups of municipalities that were similar in terms of degree of urbanization were created, and the median 'RKP-Proportion of votes cast' for each group was calculated. These values were then used to fill in the missing values. Stratified imputation was chosen for this analysis as it can provide more accurate estimates by considering the differences between the groups of municipalities.

The imputation process was executed using the Pandas and Numpy libraries in Python within the Jupyter Notebook environment. The resulting dataset with imputed missing values was saved in Excel format for further analysis, ensuring that the dataset was complete and reliable for exploring potential relationships between election results and socio-economic factors.

Exploratory Data Analysis

Summary of Key Descriptive Statistics

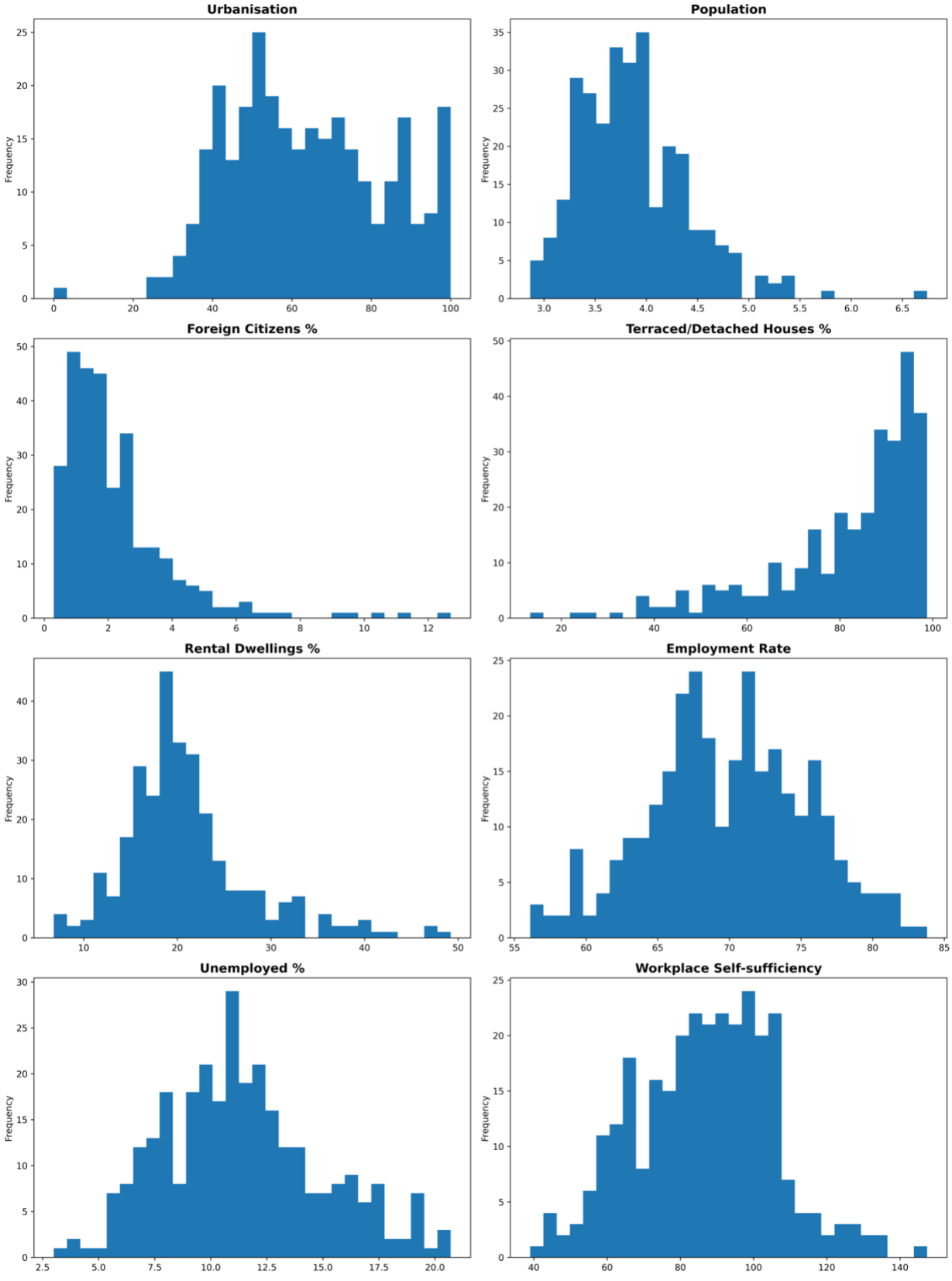
In this section, I present the findings from the exploratory data analysis using the describe() function to get an overview of the data. These findings will help us better understand the distribution and characteristics of the variables in our dataset. Some of the notable findings include:

- Degree of urbanization: The degree of urbanization across the municipalities shows a wide range of values, highlighting the variation in urbanization levels across different municipalities.
- Population: The dataset displays significant variation in population size across municipalities. The mean population is considerably higher than the median, indicating a positively skewed distribution with a few municipalities having substantially larger populations than the rest.
- Employment rate: The employment rate across municipalities exhibits variation, demonstrating that economic conditions differ across the municipalities.
- Share of foreign citizens: The average share of foreign citizens in the municipalities varies across the dataset, showcasing diverse proportions of foreign citizens across different municipalities.
- Proportion of votes cast for each party: The mean, standard deviation, and range for the proportion of votes cast for each party can provide insights into the overall distribution of political support across municipalities. The dataset reveals variations in support for different parties, which can help us understand the distribution of political preferences in the dataset.

By examining the descriptive statistics of the dataset, I gain a better understanding of the characteristics and distribution of the variables. This information will help inform our subsequent analysis and interpretation of the relationships between municipal factors and election outcomes.

Distribution and Outliers of Socio-Economic Indicators

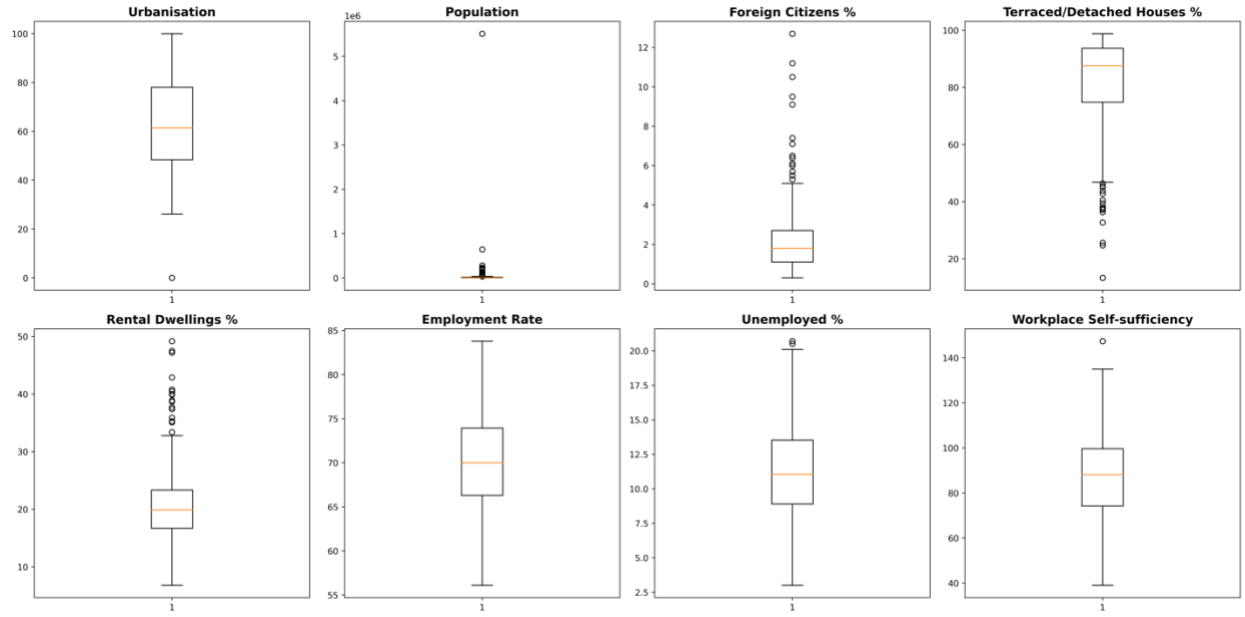
In this section, I conduct a thorough investigation of the distribution patterns for various socio-economic indicators by constructing histograms for each. This approach enables us to gain insights into the underlying structure of the data. Additionally, I generate box plots for each indicator to systematically identify potential outliers, further enhancing our understanding of the dataset and informing subsequent analyses.


Urbanisation
Frequency
0 20 40 60 80 100
Population
Frequency
3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5
Foreign Citizens %
Frequency
0 2 4 6 8 10 12
Terraced/Detached Houses %
Frequency
20 40 60 80 100
Rental Dwellings %
Frequency
10 20 30 40 50
Employment Rate
Frequency
55 60 65 70 75 80 85
Unemployed %
Frequency
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Workplace Self-sufficiency
Frequency
40 60 80 100 120 140

Upon examining the histograms of the socio-economic indicators, I can infer that there is a diverse range of values across the various indicators. The degree of urbanization appears to cluster around low and high levels, suggesting that areas are either predominantly urban or rural. The population distribution is highly skewed, indicating that most areas have relatively low populations, while a few outliers exhibit significantly higher populations. The share of foreign citizens in the population follows a similar pattern, with most areas having a low percentage of foreign citizens.

The housing types and rental dwellings show interesting patterns as well, with the dominant housing type being terraced and detached houses in most areas. The share of rental dwellings displays moderate variation across different areas. Employment rates seem to be relatively high, with most areas clustering around a common range of values. Unemployment rates also exhibit a unimodal distribution, suggesting that most areas have similar unemployment rates.

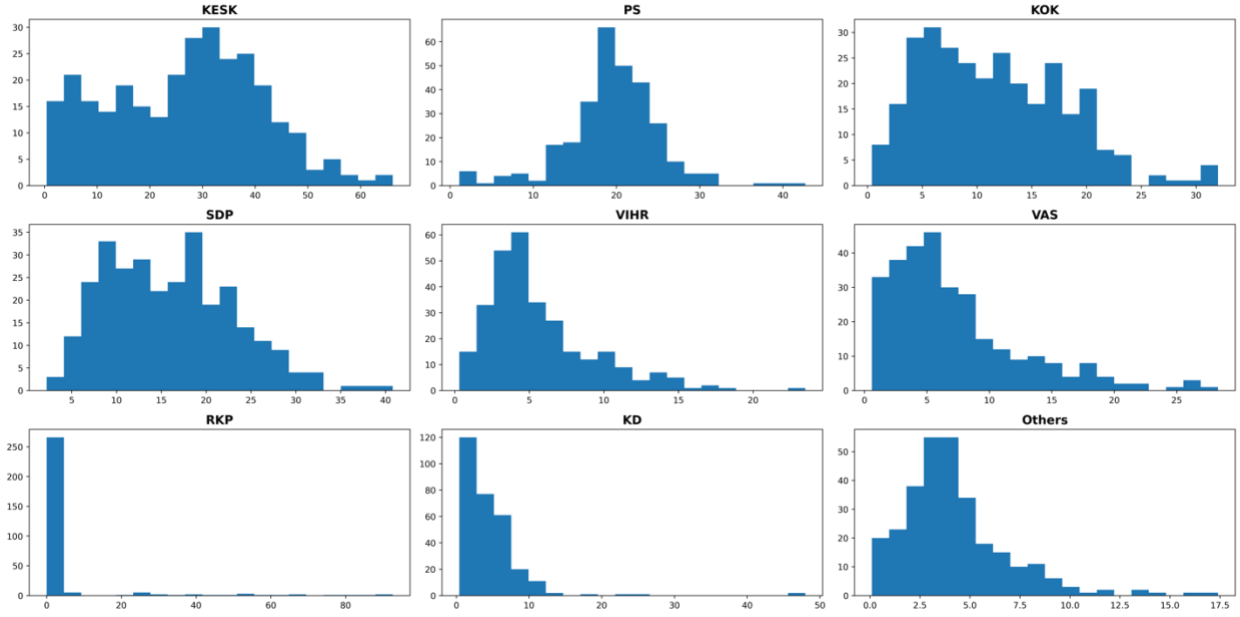
Lastly, workplace self-sufficiency shows a diverse distribution, indicating that the extent of self-sufficiency in terms of employment opportunities varies considerably among different areas. This general overview of the socio-economic indicators' distributions provides a foundation for a more in-depth analysis and exploration of the data in subsequent sections of the study.



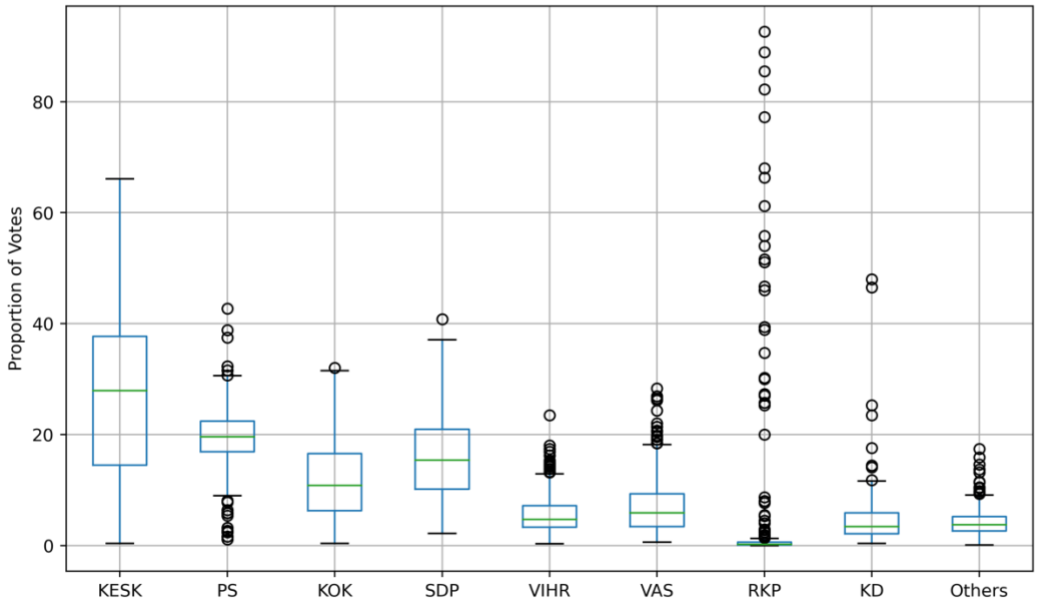
Upon examining the boxplots, I observe notable outliers in indicators such as Population, Foreign Citizens, or Workplace Self-sufficiency. However, with further examination, it is inferred that this is due to the nature of outliers. For instance, the outlier in the Population boxplot represents the population of the Helsinki area, which can be attributed to the high population density in the capital. Similarly, outliers in the proportion of Foreign Citizens in Espoo, Helsinki, and Vantaa can be explained by their inclusion in the Helsinki metropolitan area, known for its higher percentage of foreign citizens due to better job opportunities, international connections, and cultural diversity. Given the fact that outliers can be explained by the unique characteristics of the municipalities, it is reasonable to proceed with the analysis without any specific handling.

Distribution and Outliers of Voting Result Indicators

Similarly, with the socio-economic indicators, I construct a histogram for each voting result indicator to analyze the distribution patterns. Additionally, box plots are generated for each indicator to systematically identify potential outliers.



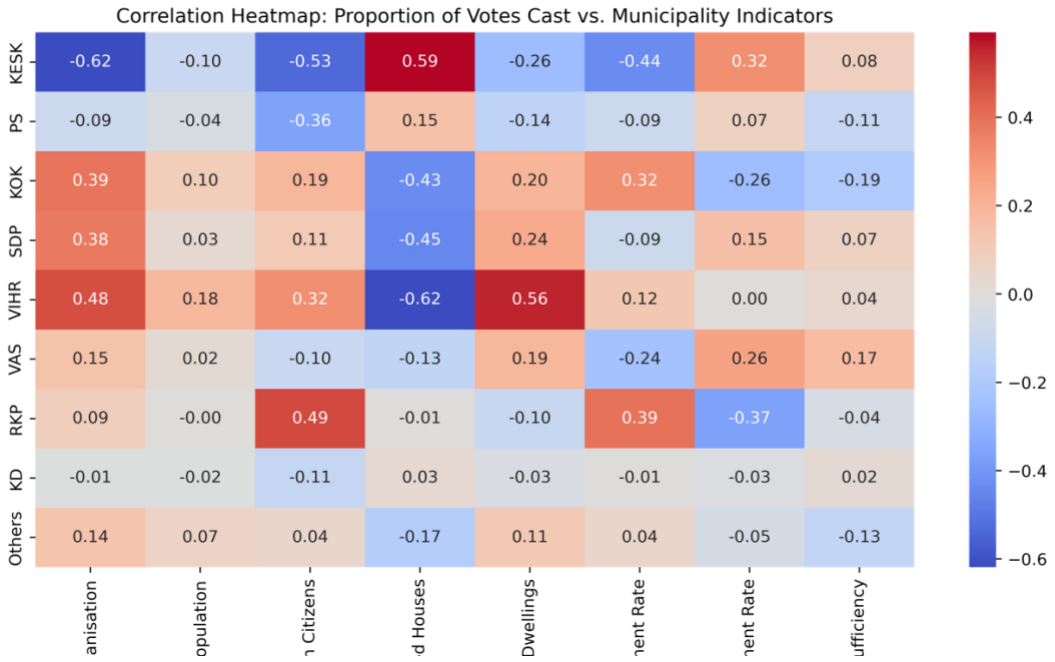
From these plots, we can infer that KESK, PS, and SDP are the major parties with the highest average vote shares. The RKP (Swedish People's Party of Finland) has the most significant variability and range in vote shares, likely due to its regional concentration in Swedish-speaking areas. PS and KOK, on the other hand, have more uniform support across municipalities.



The majority of outliers are found in the RKP column. This can be attributed to the regional concentration of Swedish-speaking populations in certain areas. Upon further examination, it has been determined that these outliers represent valid data points and do not require any specific handling for our analysis to proceed.

Correlation Analysis

In this section, I calculate correlation coefficients between the continuous variables, including the election results for each political party and the demographic and socioeconomic variables. This can help us identify strong associations between variables that may be worth further investigation.



Based on the correlation matrix we just generated, we can make a comprehensive analysis of the relationships between the proportion of votes cast for each party and the municipality indicators. Here are some of the strongest correlations for each party

1. KESK (Centre Party):
   - Strong negative correlation with the degree of urbanization (-0.62): This suggests that KESK tends to perform better in less urbanized areas.
   - Moderate negative correlation with the share of foreign citizens (-0.53) and employment rate (-0.44): KESK performs better in areas with fewer foreign citizens and lower employment rates.
   - Strong positive correlation with the share of terraced & detached houses (0.59): KESK tends to perform better in areas with more terraced and detached houses.

2. PS (Finns Party):
   - Weak negative correlation with the share of foreign citizens (-0.36): PS performs slightly better in areas with fewer foreign citizens.

3. KOK (National Coalition Party):
   - Moderate positive correlation with the degree of urbanization (0.39) and rental dwellings (0.20): KOK tends to perform better in more urbanized areas and those with a higher share of rental dwellings.
   - Moderate negative correlation with the share of terraced & detached houses (-0.43) and unemployment rate (-0.26): KOK performs better in areas with fewer terraced and detached houses and lower unemployment rates.

4. SDP (Social Democratic Party):
   - Moderate positive correlation with the degree of urbanization (0.38) and rental dwellings (0.24): SDP tends to perform better in more urbanized areas and those with a higher share of rental dwellings.
   - Moderate negative correlation with the share of terraced & detached houses (-0.45): SDP performs better in areas with fewer terraced and detached houses.

5. VIHR (Green League):
   - Moderate positive correlation with the degree of urbanization (0.48), population (0.18), share of foreign citizens (0.32), and rental dwellings (0.56): VIHR tends to perform better in more urbanized areas, with a larger population, more foreign citizens, and a higher share of rental dwellings.
   - Strong negative correlation with the share of terraced & detached houses (-0.62): VIHR performs better in areas with fewer terraced and detached houses.

6. VAS (Left Alliance):
   - Weak positive correlation with rental dwellings (0.19) and weak negative correlation with the employment rate (-0.24): VAS performs slightly better in areas with a higher share of rental dwellings and lower employment rates.

7. RKP (Swedish People's Party):
   - Moderate positive correlation with the share of foreign citizens (0.49) and employment rate (0.39): RKP tends to perform better in areas with more foreign citizens and higher employment rates.
   - Moderate negative correlation with the unemployment rate (-0.37): RKP performs better in areas with lower unemployment rates.

8. KD (Christian Democrats):
   - No significant correlations with any of the municipality indicators.

9. Others:
   - No significant correlations with any of the municipality indicators.

In conclusion: These correlations suggest that certain municipality indicators have stronger relationships with the vote shares for specific parties. However, correlation does not imply causation, and further investigation is needed to determine the causal relationships between these variables. Based on these significant correlations, we can proceed with further analysis to better understand the factors influencing voting patterns in Finland.

Regression Analysis

In this section, we will perform regression analysis to investigate the relationship between the proportion of votes cast for each political party and the municipality indicators using two approaches. The first approach will involve conducting regression analysis for each individual party, while the second approach will group parties into three categories: right, left, and centrist. By building regression models, we can better understand how these factors may influence voting patterns and possibly make predictions based on the available data. We will begin by selecting the appropriate type of regression model for our analysis, followed by model fitting, evaluation, and interpretation of the results. This section will be divided into two subsections to separately discuss the findings from each approach.

1. Individual Party Regression Analysis

In this section, we will construct regression models based on the most significant variables identified in the correlation matrix for each political party. These variables provide valuable insights into the factors that might have the strongest influence on the proportion of votes cast for each party. The selected variables for each party are as follows:

- KESK (Centre Party): Urbanisation, Terraced & Detached Houses, Employment Rate
- PS (Finns Party): Foreign Citizens, Unemployment Rate, Workplace Self-sufficiency
- KOK (National Coalition Party): Urbanisation, Terraced & Detached Houses, Employment Rate
- SDP (Social Democratic Party): Urbanisation, Terraced & Detached Houses, Unemployment Rate
- VIHR (Green League): Urbanisation, Population, Foreign Citizens
- VAS (Left Alliance): Rental Dwellings, Unemployment Rate, Workplace Self-sufficiency
- RKP (Swedish People's Party): Urbanisation, Foreign Citizens, Employment Rate
- KD (Christian Democrats): Foreign Citizens, Terraced & Detached Houses, Workplace Self-sufficiency
- Others: Urbanisation, Population, Rental Dwellings

a. Model Selection

In this analysis, we used linear regression to study the relationship between the proportion of votes for each political party and the significant municipality indicators. Linear regression is a suitable choice for this problem because it can efficiently model linear relationships between a dependent variable and multiple independent variables. Additionally, it provides easily interpretable results in the form of coefficients, which indicate the strength and direction of the relationships between the municipality indicators and the proportion of votes received by each party

b. Interpretation of Results

In this part, we fit the selected linear regression model to the data for each political party using their respective significant variables. Through this process, we obtain a set of regression coefficients that represent the impact of each significant variable on the vote share for each party.

Here is a summary table of the regression results:

Party	Adjusted R-squared	Significant Variables	Relationship
KESK	0.589	Urbanisation, Terraced & Detached Houses, Employment Rate	Urbanisation (negative), Terraced & Detached Houses (positive), Employment Rate (positive)
KOK	0.302	Urbanisation, Terraced & Detached Houses, Employment Rate	Urbanisation (negative), Terraced & Detached Houses (negative), Employment Rate (positive)
SDP	0.221	Urbanisation, Terraced & Detached Houses, Unemployment Rate	Urbanisation (positive), Terraced & Detached Houses (negative), Unemployment Rate (positive)
PS	0.120	Foreign Citizens	Foreign Citizens (negative)
VIHR	0.261	Urbanisation, Population, Foreign Citizens	Urbanisation (positive), Population (negative), Foreign Citizens (positive)
VAS	0.300	Foreign Citizens, Unemployment Rate, Employment Rate	Foreign Citizens (negative), Unemployment Rate (positive), Employment Rate (negative)
RKP	0.422	Population, Swedish-speaking Population	Population (negative), Swedish-speaking Population (positive)
KD	0.187	Urbanisation, Terraced & Detached Houses, Foreign Citizens	Urbanisation (negative), Terraced & Detached Houses (positive), Foreign Citizens (positive)
Others	0.076	Foreign Citizens	Foreign Citizens (negative)

Based on the regression results, we can observe that different municipality indicators have varying effects on the vote shares of the political parties. Given the significance and magnitude of the coefficient, we have some more meaningful insights into which factors can affect the election results:

Urbanization: It significantly affects KOK, SDP, VIHR, and RKP support. It has a positive relationship with SDP and VIHR support, while negatively influencing KOK and RKP support. This suggests that parties with policies and agendas catering to urban areas may fare better in more urbanized municipalities.

Population: This variable significantly influences VIHR support, having a negative relationship. This suggests that VIHR may be more appealing to voters in less densely populated areas.

Foreign Citizens: This variable has a significant impact on several parties, including KESK, PS, VIHR, RKP, and KD. It has a positive relationship with KESK, VIHR, and RKP support, while negatively influencing PS and KD support. This suggests that parties with pro-immigration policies tend to perform better in areas with higher proportion of foreign citizens.

Unemployment Rate: This variable significantly influences KESK, SDP, and VAS support, all having positive relationships. This indicates that these parties may be more appealing to voters in areas with higher unemployment rates, possibly due to their employment-related policies.

Employment Rate: This variable significantly affects KOK and RKP support, both having positive relationships. This suggests that these parties may perform better in areas with higher employment rates, possibly due to their economic or job-related policies.

Terraced & Detached Houses: This variable has a significant impact on KOK and SDP support, with positive and negative relationships, respectively. This may indicate that KOK attracts voters from areas with more single-family homes, while SDP may appeal more to voters in areas with other types of housing.

c. Model Evaluation

In a multivariate analysis, assessing multicollinearity is essential to ensure the validity of the findings. To evaluate the presence of multicollinearity in our models, I computed the Variance Inflation Factors (VIFs) for each independent variable. The results indicate that for all parties, the VIF values for the respective indicators are below 10. This suggests that our models do not exhibit significant multicollinearity issues, thereby ensuring the reliability of our analysis.

To further ensure the robustness and reliability of our findings, an Out-of-sample validation test is performed. I split the dataset into training and testing sets. Next, I fitted the model using the training set and then evaluate its performance on the testing set. The R-squared value, which indicates the proportion of the variance in the dependent variable that can be explained by the independent variables in the model when applied to new, unseen data, is computed.

The out-of-sample validation results reveal that the regression models have varying predictive performances for different political parties. Some parties, such as KESK, SDP, RKP, and KOK, have models with relatively higher R-squared values, indicating a better fit and more accurate predictions for their vote shares. On the other hand, parties like PS, VAS, KD, and VIHR, have models with moderate R-squared values, suggesting less explanatory power and less accurate predictions for their vote shares.

2. Aggregated Party Regression Analysis

The results obtained in the previous section have provided meaningful insights into the relationship between various socio-economic factors and the election result for different parties. However, these results may not be directly generalizable across all parties, as each party may have unique characteristics, policies, and voter bases. To improve our regression analysis, our approach is to aggregate the vote shares for all parties and create a model that examines the combined impact of municipality factors on overall election outcomes. This will help identify common factors that may influence the election results for all parties.

a. Party Classification

Based on the Finnish political landscape, we can classify the nine major parties into three categories: left-leaning, right-leaning, and centrist. We will exclude the "Other parties" column from this analysis, as it represents an aggregation of multiple smaller parties that may not fit neatly into these categories. After researching, we have decided to group the parties as follows:

1. Left-leaning parties: These parties generally advocate for social equality, workers' rights, and progressive social policies.
   - SDP (Social Democratic Party)
   - VAS (Left Alliance)
   - VIHR (Green League)

2. Right-leaning parties: These parties generally support limited government intervention in the economy, lower taxes, and individual liberties.
   - KOK (National Coalition Party)
   - PS (Finns Party)
   - KD (Christian Democrats)

3. Centrist parties: These parties typically promote a balanced approach, incorporating aspects of both left-leaning and right-leaning ideologies.
   - KESK (Centre Party)
   - RKP (Swedish People's Party)

b. Interpretation of Results

In this approach, the linear regression model is chosen for the same reason as mentioned above. I fit the selected linear regression model to the data for each political party category using their respective significant variables. Given the significance and magnitude of the coefficient, we have some meaningful insights into which factors can affect the election results for each party group:

1. Left-leaning Parties

Foreign Citizens (negative), Urbanisation (positive), Terraced & Detached Houses (negative), Rental Dwellings (negative), Employment Rate (positive), and Unemployment Rate (positive) all significantly influence left-leaning party support. This suggests their appeal to voters in urban areas with higher employment and unemployment rates and fewer single-family homes or rental dwellings.

2. Right-leaning Parties:
Foreign Citizens (negative), Urbanisation (negative), Terraced & Detached Houses (negative), and Workplace Self-sufficiency (negative) all significantly influence right-leaning party support. This indicates their appeal to voters in less urbanized areas with fewer single-family homes, lower foreign citizen populations, and lower workplace self-sufficiency.

3. Centrist Parties:
Foreign Citizens (positive), Urbanisation (positive), Terraced & Detached Houses (positive), and Workplace Self-sufficiency (positive) all significantly influence centrist party support. This suggests their appeal to voters in urban areas with more single-family homes, higher foreign citizen populations, and higher workplace self-sufficiency.

c. Model Evaluation
In a multivariate analysis, assessing multicollinearity is essential to ensure the validity of the findings. To evaluate the presence of multicollinearity in our models, I computed the Variance Inflation Factors (VIFs) for each independent variable. The results indicate that for all party groups, the VIF values for the respective indicators are below 10. This suggests that our models do not exhibit significant multicollinearity issues, thereby ensuring the reliability of our analysis.

To further ensure the robustness and reliability of our findings, an Out-of-sample validation test is performed. I split the dataset into training and testing sets. Next, I fitted the model using the training set and then evaluate its performance on the testing set. The R-squared value, which indicates the proportion of the variance in the dependent variable that can be explained by the independent variables in the model when applied to new, unseen data, is computed.

The out-of-sample validation results reveal that the regression models have varying predictive performances for different political parties. The left-leaning and centrist parties have models with relatively higher R-squared values, indicating a better fit and more accurate predictions for their vote shares. On the other hand, the right-leaning parties have models with moderate R-squared values, suggesting less explanatory power and less accurate predictions for their vote shares.

Model Comparison
This section presents a comprehensive comparison of the findings obtained through two different approaches: individual party regression analysis and aggregated party regression analysis.

Similarities:
1. Urbanisation: Both approaches show that urbanization significantly affects the support for parties such as SDP, VIHR, KOK, and RKP. In both cases, SDP and VIHR perform better in urban areas, while KOK and RKP perform worse.
2. Foreign Citizens: Both analyses reveal the significant impact of foreign citizens on party support, with KESK, VIHR, and RKP benefiting from higher foreign citizens, while PS and KD suffer.
3. Unemployment Rate: In both approaches, KESK, SDP, and VAS perform better in areas with higher unemployment rates.

4. Employment Rate: Both analyses show that KOK and RKP perform better in areas with higher employment rates.

Differences:

1. Population: In the individual party analysis, population significantly influences VIHR support negatively, while in the aggregated analysis, population is not a significant variable for left-leaning parties.
2. Terraced & Detached Houses: The individual party analysis shows that KOK and SDP are significantly affected by this variable, while in the aggregated analysis, it influences all party groups, with differing directions.
3. Rental Dwellings and Workplace Self-sufficiency: These variables are significant only in the aggregated analysis, affecting left-leaning and right-leaning parties, respectively.

Results

In this study, we employed two different approaches to analyze the relationship between municipality indicators and the electoral performance of political parties in Finland. The following is a summary of the key findings and interesting insights obtained from these analyses:

1. Urbanization: One of the most noteworthy findings is the impact of urbanization on party support. For instance, the Social Democratic Party (SDP) and the Green League (VIHR) tend to perform better in urban areas, whereas the National Coalition Party (KOK) and Swedish People's Party (RKP) experience weaker support in such areas. This highlights the importance of the urban-rural divide in predicting electoral performance.

2. Foreign Citizens: The presence of foreign citizens in a municipality significantly affects party support. Parties like the Centre Party (KESK), Green League (VIHR), and Swedish People's Party (RKP) benefit from higher foreign citizen populations, while the Finns Party (PS) and Christian Democrats (KD) suffer. This finding underscores the impact of immigration and social integration policies on party support.

3. Employment and Unemployment Rates: Our analysis revealed a nuanced relationship between employment-related factors and voter preferences. Parties like the Centre Party (KESK), Social Democratic Party (SDP), and Left Alliance (VAS) receive more support in areas with higher unemployment rates, while the National Coalition Party (KOK) and Swedish People's Party (RKP) perform better in areas with higher employment rates.

4. Housing Types: Housing characteristics, such as the prevalence of single-family homes, have a significant impact on voter preferences for specific parties. For example, the National Coalition Party (KOK) appeals more to voters in areas with more single-family homes, while the Social Democratic Party (SDP) attracts voters in areas with a diverse range of housing types.

In the context of the 2019 Finnish parliamentary elections, our analysis offers insights into factors shaping voter preferences and election outcomes. The Centre Party's (KESK) decline, losing 18 seats, may be related to weaker support in urban areas and higher foreign citizen populations. The Social Democratic Party (SDP) became the largest party, possibly due to increased support in areas with higher unemployment rates and diverse housing types.

The Green League (VIHR) and Left Alliance (VAS) gained seats, with the former's success attributed to environmental concerns in urban areas. The Finns Party (PS) saw increased support in areas with fewer foreign citizens, reflecting their stance on immigration and integration policies.

The fragmented political landscape during the 2019 elections can be understood through our analysis, as the varying impact of socio-economic factors demonstrates the unique needs and preferences of each party's constituents. Overall, our findings contribute to a deeper understanding of the complex interplay between municipal factors and electoral outcomes, illuminating the dynamics shaping the 2019 Finnish parliamentary elections.

Conclusion

In this study, we have investigated the relationship between municipality indicators and the electoral performance of Finnish political parties by employing two different analytical approaches: individual party regression analysis and aggregated party group regression analysis. Our findings provide valuable insights into the complex interplay between various factors and the electoral support for different parties.

Some of the most noteworthy findings include the significant role of urbanization in shaping party support, the impact of foreign citizens on party preferences, the relationship between employment and unemployment rates and voter preferences, and the influence of housing types on party support. These findings highlight the importance of understanding local context and policy priorities in predicting electoral outcomes.

While our analysis provides valuable insights into the relationship between municipality indicators and electoral support for Finnish political parties, there are still limitations and potential areas for improvement. The study found no significant multicollinearity using VIF, but it can still affect the analysis subtly. The out-of-sample validation showed relatively low predictive power, suggesting room for improvement by incorporating more variables or refining existing ones and exploring alternative modeling techniques. Future research can focus on enhancing the model's predictive performance through cross-validation or other optimization techniques. Lastly, as our analysis is based on the 2019 elections, future research could benefit from using panel data or time-series analysis to better understand the dynamics of party support and the impact of municipality characteristics over multiple elections.

References

1. Statistics Finland. (n.d.). Free-of-charge databases. Retrieved from https://www.stat.fi/tup/tilastotietokannat/index_en.html#free-of-charge-databases
2, Wikipedia contributors. (2021, September 9). 2019 Finnish parliamentary election. In Wikipedia, The Free Encyclopedia.