Demographic effects on Kuwait's elections results

by

Khaled Nasser Alsalmi

B.S., Gulf University for Science and Technology, 2008

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Computer Science
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Joshua L. Weese

Copyright

Abstract

Kuwaiti political analysts and pundits rely solely on their personal experience, intuition, and subjective perspectives to analyze election results and make predictions. However, this approach neglects the potential benefits and insights that data analysis can provide in uncovering influential factors behind various outcomes. In this report, a data science analysis is carried out on the Kuwaiti parliamentary elections that took place in September 2022, considering voters' age, gender, tribe, religion, occupation, and district. The analysis model was developed by comparing various classification algorithms depending on their suitability to the nature of the data. Most algorithms exhibited a similar trend; However, the RandomForest algorithm consistently achieved the highest accuracy. The interaction between all features excluding tribe and all features excluding religion exhibited a statistically significant result indicating their suitability as predictors for the election outcome of the 2022 elections, making it a potential predictor of Kuwaiti election results.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude and appreciation to several individuals who have made significant contributions to the completion of my master's dissertation. Firstly, I want to acknowledge my dear friends Yousef AlEnizi, Bader AlOtibi, Fahad Almutairi, Saad Alshammari, Rakan Alajmi, Loay AL Bader, Hammad AlNomisy, Ghazi AlEnizi, Mustafa AlDakheel, Mohammed AlAdwani, and Majid AlThufiry for their invaluable assistance in identifying the affiliations of voters. Their unwavering efforts and dedication have been essential in making this dissertation a reality. I would also like to extend my heartfelt thanks to Mr. Salah Aljasim, the founder of Aljasim Systems, for his invaluable assistance in the data collection processes. His expertise and guidance have played a crucial role in ensuring the accuracy and reliability of the data used in this research. Lastly, I am sincerely grateful to my academic advisor, who has provided me with constant guidance, support, and feedback throughout the entire dissertation process. His invaluable insights and recommendations have helped shape this work and refine its content. I would like to express my appreciation to everyone who has contributed to this dissertation, whether directly or indirectly. Your unwavering efforts and support have been instrumental in my success, and I am deeply grateful for your contributions.

Introduction

Kuwait has a unicameral parliament, the National Assembly, with 50 elected members serving four-year terms. The voting system in Kuwait is based on a single vote, where voters cast a vote for one candidate in their district, and the candidate with the most votes wins the seat. The country is divided into five electoral districts, with 10 seats allocated to each district. The system allows for multiple candidates to run from the same political party, which has led to a fragmented parliament with a large number of independent members. While many political analysts and experts argue that demographic features are the primary factors that affect election outcomes, there is a lack of scientific analysis to support their opinions [1]. It is crucial for political campaign analysts to incorporate data science into their analysis to ensure the accuracy of the information provided and minimize any biases.

The aim of this research is to create a data science analysis that can validate political experts' parliamentary election's results claims. This project used data science framework that used five stages to generate the output which are: Strategy, Data Gathering, Data Cleaning, Data Discovery, Machine Learning and Testing. The progress was not sequential as we had to go through cycles until we developed the final analysis.

The data gathering phase of the study of the September 2022 elections in Kuwait involved collecting three primary datasets. The first dataset (dataset-1) contained digital information on eligible voters obtained from an electoral candidate committee. The second dataset (dataset-2) included voting results from a Kuwaiti Newspaper website. The third dataset (dataset-3) contained information on the religious and tribal affiliations of voters and candidates, which had to be obtained from societal knowledge due to the lack of formally documented resources.

During the data cleaning phase six key challenges were faced which includes the inability of the database to recognizing Arabic characters, spaces in between the names, different values with the same meaning, last names not in a distinct column, and inaccurate datatypes.

In data preparation and discovery stages some features were forked, and new columns were added. From voter dataset (dataset-1) the name was divided into first and last name. then the last name was used to create the Tribe and Religion columns from it as a connector to determine the appropriate values from dataset-3. The "tribe" feature has 31 groups. The religious consists of "Sunni," "Shi'a," and "others." From the DOB, age was calculated then used to create eleven age groups each representing a ten-year interval. Also, the "Occupation" feature initially had 563 jobs that were clustered into 5 categories: Employee, Student, Home Maker, Retired, and Unknown. From the winner dataset (dataset-2) the candidates' gained votes were used as indicator to create two new columns: number of the losers and winners for each district. In addition, the number of the winners attribute was used to create a Boolean 'has-winner' column. Lastly, a gender/occupation column was added that indicates the count of voters within similar gender- Occupation group. The Data Description phase describes the distribution of all features.

The Machine Learning and Tools phase focused on identifying the necessary tools for the study. The paper utilized various Data Science techniques, including descriptive statistics, SQL, cluster, and classification using tools such as MS Access, Excel, R studio, Weka and Power Bi and an online P-value calculator. The findings revealed that all of the classification algorithms had a similar trend when using the dataset, but RandomForest algorithm consistently achieved the highest accuracy among all the algorithms. Based on the assessment, the interaction between all features excluding tribe and all features excluding religion exhibit statistically significant results indicating their suitability as predictors for the election outcome.

Background

In this section a description of the history of the political system, the demography of Kuwait's population, and the methodology used for this research. In addition to the information of the related work.

Kuwait's Election System

The political system in Kuwait has evolved over time since the early 18th century, and it currently consists of three major sectors: the appointed government, the self-appointed judiciary, and the elected parliament. At the head of these three sectors is the Amir [2] , who serves as the President of Kuwait. The Kuwaiti parliament, also known as the National Assembly, is a key component of the country's political system. It is comprised of 50 elected members who are elected from different electoral districts- groups of small cities-. The legislative laws governing the parliament have changed four times since the first elected parliament in 1963, and 19 elections have been held since then. The last legislative change happened in 2012, reducing the number of votes from 4 to 1 vote, along with another change in district size by including/excluding cities [3]. Correspondingly the results witnessed dramatic turnout where several new faces won the election [4]. The current election law divides Kuwait into five electoral districts. Ten members represent every district. The parliament gets re-elected after four years or after a ”dissolvement act” by the Amir or the high constitutional court [5]. The voter must be a Kuwaiti citizen who has turned 21 before the registration period. The process of voter registration in Kuwait is based on residence location, with eligible voters being registered according to their place of residence in one of the five electoral districts. Once registered, each voter has the right to cast a single vote for one of the candidates in their respective electoral district. The ten candidates who receive the majority of votes become the parliament members representing their district. Table 1. shows the districts and the voting power a voter has , since the 1st election held in Kuwait . This report followed the regulation under the 2012-2022 elections laws.

Table 1. Number of District/Voting Power Through the Years [6]

Years	District	Voting Power	No of election(s)	period
1961-1976	10	5	4	15 years
1980 -2006	25	2	9	26 years
2008 – 2012	5	4	1	4 years
2012 – 2022	5	1	5	10 years

Kuwaiti Districts

Districts in Kuwait are comprised of small cities, and there are 5 districts that are formed by grouping 105 cities for election purposes [7, 8]. These districts are labeled by numbers (District 1-5). Table 2. shows the population and number of cities in each district [9]. The District data is used as a feature for the analysis of this project.

Table 2. Population Distribution in all Districts [9]

District Name	Population	No of cities
District 1	14.9%	18
District 2	11.4%	17
District 3	17.8%	15
District 4	26.45%	28
District 5	29.2%	27

Kuwaiti Tribes

Kuwait's tribal population originates from neighboring countries : Saudi Arabia, Iraq, and Iran. The tribes are classified into northern and southern tribes based on their historical lands. Some Iranian descent tribes are referred to as "Arab Faris" or Arabs of Persia, and they no longer identify themselves as a tribe[10]. However, some groups of Arab Faris evolved into a tribe based on their geographical descent from Persia or their specialization in certain crafts such as ship crafting or freshwater distribution, which made them early settlers in Kuwait [11]. On the other hand, some descendants from all tribal ancestries do not identify themselves as a tribe but rather as citizens of a national state, and they are called "Hadar" or urban dwellers [12]. The selected tribes fall under three categories: a) Tribes by blood lineage b) Tribes by historical geographic existence c) Groups of people who do not recognize tribal heritage as an identity in a modern state [13].

Kuwaiti Religious Groups

Kuwait's population religious background is primarily composed of Arab Muslims, with an estimated 1.5 million citizens. The Muslim community is divided into two main sects, Sunni and Shi'a, with the former constituting approximately 75 percent of the citizen population, while the latter accounts for the remaining 30 percent. Sunni Muslims follows the teachings of the Sunni school of thought, while Shi'a Muslims follows the Shi'a school of thought. Both communities have their unique beliefs and practices, which play a significant role in shaping Kuwaiti society [11]. Individuals with Christian [14] [15]or unknown religious affiliations will be categorized as "others" for the purpose of this research. The absence of Jewish citizens in Kuwait has been confirmed by the Public Authority for Civil Information (PACI) [15].Understanding the religious composition and dynamics of Kuwait's population is crucial for comprehending the country's complex interplay of religion and politics.

Data Science Methodology

Element61, a consulting firm specializing in Business Analytics, Performance Management, and data science developed a framework for Data Science. This framework is utilized to suit organizations of varying sizes and industries. The methodology comprises five distinct phases: Strategy, Data gathering, Data Discovery, Machine Learning, and Fine Tuning &

A/B Testing. The first phase involves identifying use-cases and extracting insights from them. The second phase includes collecting relevant data to be used in the analysis. In the third phase, data is explored and analyzed to gain further insights. The Machine Learning phase involves developing the appropriate ML model and optimizing it through feature engineering, balancing data, running the model, and evaluating its performance. Finally, in the Fine Tuning & A/B Testing phase, the model is tested in real-life scenarios to provide recommendations for the business and gather feedback for further improvement[16] .A modified version of this methodology is employed in this report.

Related work

To our knowledge, no analysis investigated the Kuwaiti elections using the data science neither before or after 2012 (last regulation update). However, an adequate number of published articles and papers used machine learning to predict and analyze voting around the world.

Silvia Kim and Jan Zilinsky investigated the five demographic characteristics (income, age, gender, education, and ethnicity)- which are similar to this report features-, researching what meaning they carry over the accuracy of prediction on ideological sorting and polarization. They used three data sets of public opinion surveys to test the hypothesis using tree-based methods compared with random guesses. They split into training and testing datasets with an 80:20 ratio using R, in addition to using random forests, logistic regression, and a CART model. Their results, in contrast to this report show that demographic characteristics are not vital to the prediction, where they find no support for vote-based demographic sorting [17].

A group of four researchers from different universities, explored the feasibility of using a Deep Neural Network to predict bipartisan elections in the US and Hong Kong. They investigated distinct geographic parameters: age, and sex. They suggested that those factors can result in a more precise voter count, and therefore affect the prediction, while this report shows that age and gender are not good predictors. They recommended using methods such as disaggregation, multi-level regression, and poststratification [18].

Szkatuła et la. applied forecasting machine learning technique to forecast the voting behavior of individual members of parliament. A special issue of parliamentary debates was chosen based on 60 votes concerning the ideological left-right voting criterion. They investigate the number of parliament members, the parliamentary group membership, the number of votes, types of votes (“for”, “against”, “abstain”), the matter of the voting (e.g., “pro-government”, “anti-government”, “left”, “right”), etc. Results obtained indicate that the method under consideration can be used to forecast. with a slight modification it can also determine the most representative member of a given parliamentary group [19].

Fairstein et la. proposed a model for predicting people's voting behavior using poll information. The model is based on a Bayesian framework and considers various factors such as the demographics of the voters, their political affiliations, and the poll results. The authors test the model using data from the 2012 US presidential election and find that it performs better than other models in predicting the voting behavior of the electorate. They suggests that the proposed model can be used to improve the accuracy of election forecasts and help political campaigns to target their messages more effectively [20].

Methodology

In this study, a modified version of the Element61's Methodology [16] was utilized, as presented in Table 3. The original phases were adapted to provide better clarity and to suit the project's requirements. The methodology comprises five main phases that are explained below. The objective of this research is to examine the impact of demographic information of voters and candidates on the election results in Kuwait. Although this is a commonly held belief among Kuwaitis, it lacks empirical evidence.

Table 3. Data Science Methodology Phases

Strategy	Reflect and understand the context and objectives of the desire problem. What problem are we trying to solve, what has been tried in the past, what do we expect from the solution
Data Gathering	Translate the problem into data. Identify the right data sources and fields
Data Cleaning/ Data Preparation	Detecting and correcting corrupt or inaccurate records and identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting
Data Description/ Discovery	Get to know the data, spend sufficient time on describing it until basic discoveries can be made and create required columns that can be used to build the final data set
Machine Learning	do feature definitions based on the data discovery analysis. run and compare different ML algorithm, evaluate their performance

Strategy

While many political analysts and experts argue that demographic and religion backgrounds are the primary factors that affect election outcomes, there is a lack of scientific analysis to support their opinions [1]. It is crucial for political campaign analysts to incorporate data science into their analysis to ensure the accuracy of the information provided and minimize any biases.

Data Gathering

In the Data Gathering phase, three primary datasets were collected for this study. The first dataset was collected digitally in Arabic language, containing information for 795,910 eligible voters for the September 2022 elections. This data was obtained from an electoral candidate committee, which was initially provided by the governmental election committee to every candidate. The data includes details such as the current voter registration status, occupation, date of birth, registration date, voting form number, the first letter of the voting form, voting table (male/Female), voting area, voters table number, election district, and the full name of the voter (first, second, third, and last name).

The second dataset includes 298 rows of voting results from the September 2022 elections, collected from the Kuwaiti Newspaper website "Al-Anbaa"[21]. The data includes the full names of all the candidates, their electoral districts, and the number of votes they gained. It is worth noting that the data in the article was not in tabular format.

Finally, the third dataset collected during the Data Gathering phase consisted of information on the religious and tribal affiliations of the voters and candidates. Obtaining this data posed a challenge due to the absence of formally documented resources, and it had to be obtained from societal knowledge rather than relying on formal references.

Data Cleaning

The process of cleaning and merging the data to create the required data sets was time consuming and involved using Microsoft Access and Microsoft Excel. One of the primary challenges encountered during the process was the unavailability of information on tribal and religious affiliations. While it was possible to infer most of this information from the last names of individuals, a significant number of individuals required personal correspondence to obtain accurate data. Another major issue was creating the last name attribute since it was not initially distinct in the dataset, and creating it required addressing various issues related to the Arabic language. The data cleaning process addressed six key challenges, which are summarized in Table 4.

Table 4. Key Challenges of the Data Cleaning Process

The issue	Example	Action Taken
Database can’t recognize Arabic characters	Lastname: [????????????]	Encode the data into a recognizable format.
Spaces between a compound last name	Abdu allah and Abdullah "عبد الله"and "عبدالله"	Removed the Spaces.
Last names with Different writing in the Arabic language but similar pronunciation.	مطيري And مطيرى	Standardize the writing.
Different Values with the same meaning	employee , an employee , employed , works in the government, etc.	Set to a single value
Last names are not in a distinct column	Names were in a full name column [1st , 2nd , 3rd ,…, last name]	Extract the last name into a distinct column.
Unavailable documented data	Tribal and Religious affiliation	Personal societal investigation
Inaccurate datatypes	Date of birth was entered into a text field.	Converted the data to a Date Type

Data Preparation

The data preparation section was conducted twice in this paper. The first round involved transforming the raw data into a valid format for data descriptions. This included creating and adding features such as "tribe," " age," " occupation," and "religion" to data set 1. The "tribe" and "religion" features were derived from the last name, while the "age" feature was calculated by subtracting the date of birth from the registration date column and then categorized into eleven age bins, each representing a ten-year interval starting from the legal voting age of twenty. The bin size was determined after testing multiple formulas, including Sturge's Rule [22], Scott's Rule [23], and Rice's Rule [24], without significant deviation. The "occupation" feature was also cleaned and modified to include five occupation choices, as there were 536 repetitive job entries with typos that required cleaning. In the second round, a new data set was created by connecting data sets one and two along with the newly added features to form the final data set.

Data Discovery

The data discovery phase created the final dataset which was used for the final analysis. It was created by combining previous phases' outcomes. *District* and *Gender* columns were obtained from the voters' dataset (dataset-1). In addition to the newly generated columns: *Tribe*, *Religion*, *Age groups*, *occupation*, and *Gender-Occupation* column - count of the voters with similar Gender and Occupation features within the same tribe. Using the candidates' dataset (dataset-2) new columns were generated. The first one is *Tribe-Loser* which consists of the number of losers in a tribe and the 2nd column *has-winner* is a Boolean that represents whether the tribe has a winner or not.

Data Description

Tribe

The "tribe" feature in Figure 1 was divided into 29 groups, including the "Unknown" category. These groups were Murrah, Suhool, Sulaylat, Dufhier, Fudhool, Bani Tameem, Bani Khaled, Bani Ghanim, Bani Hajar, Harb, Kandari, Duwasir, Reshaydah, Zuoob, Subaie, Shararat, Shammar, Oteibah, Ejman, Adaween, Enizah, Awazim, Qahtan, Mutair, Hirshan, Failakawi, Qena'at, Hadar, Small Tribes, and Unknown. The "Unknown" category represents last names that are uncommon in society, usually resulting from marriages outside the society or from different countries that have led to citizenship. The "Small Tribes" category includes families with fewer than 60 members, and these two categories were excluded from the prediction analysis. The "Hadar Sunna and Shiaa" tribe is the dominant tribe, comprising 36.1%

of the total number of votes across all districts and having a significant presence in districts 1, 2, and 3 at the district level. The next largest category is the "Unknown" tribes, which are four times smaller than the Hadar tribe, accounting for 8.5% of the votes. The remaining tribes vary in size, with Shararat having only 58 voters (.01%) and Mutair having the highest number of voters at 62,584 (7.9%). After excluding the Hadar Sunna and Shiaa tribe, the Awazm tribe dominates districts 1 and 5 with 7.3% and 17%, respectively. In District 2, the Enizah tribe is the dominant group, while in District 3, the unknown tribes account for 18.2% of the votes. In District 4, the Mutair tribe holds the majority, with 20.1% of the votes as represented in Figure 2. and Table 5.

Figure 1. Tribal Distribution of the Voters in all Districts

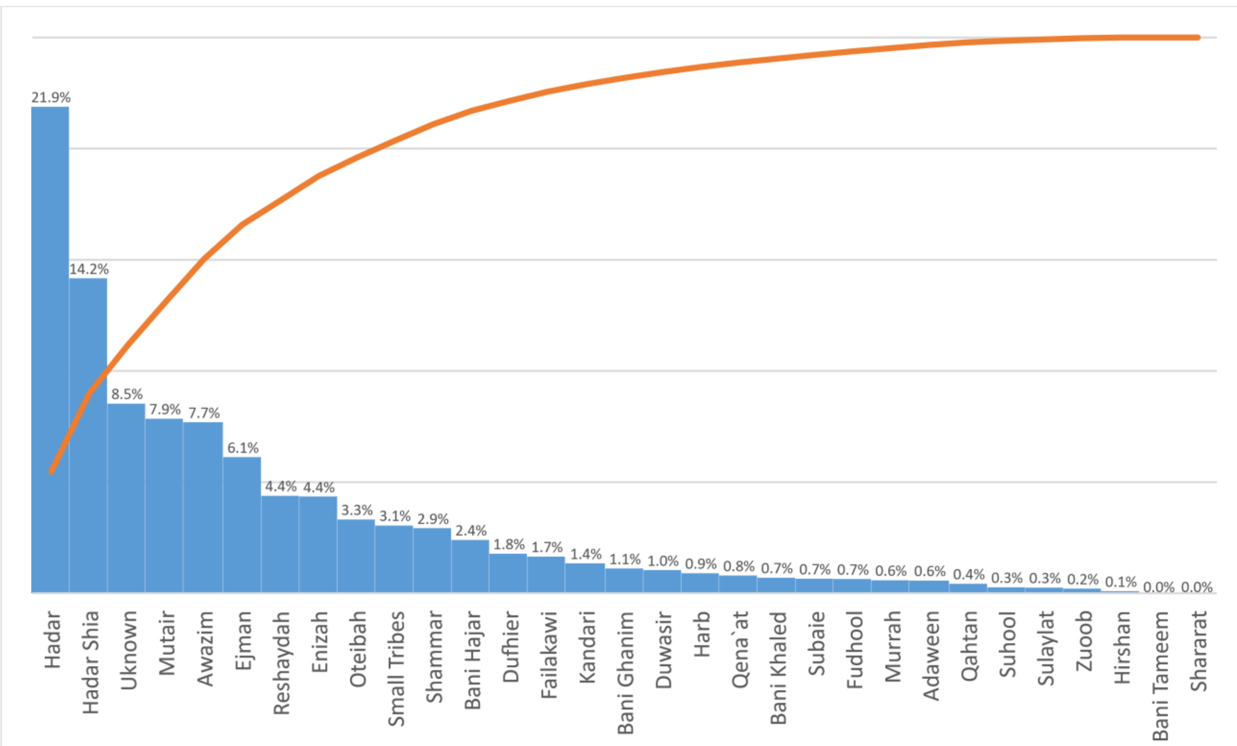


Figure 2. Tribal Distribution of Voters in Different Districts

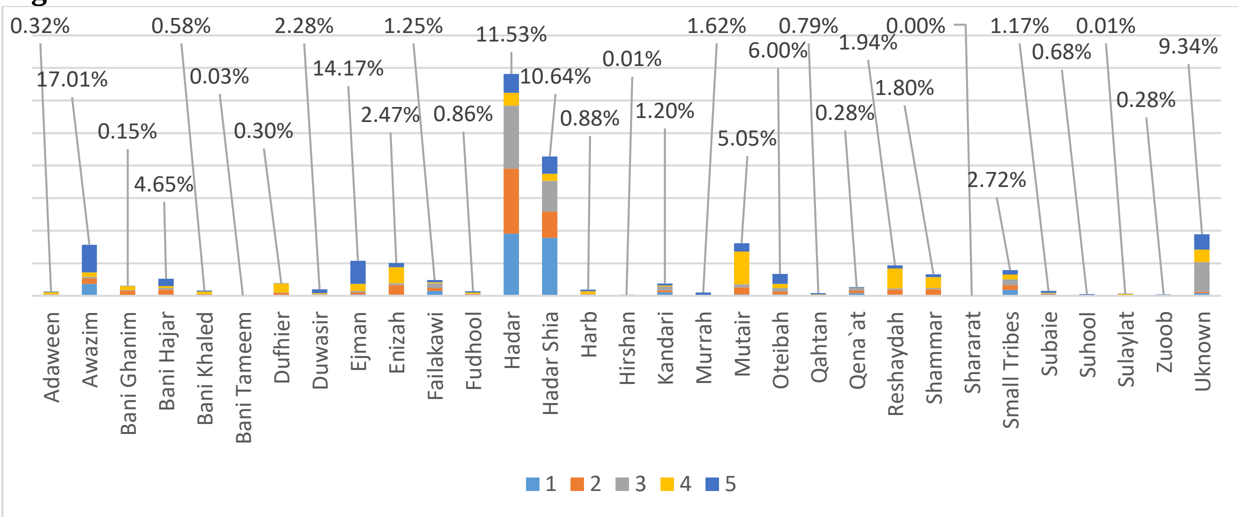


Table 5. Tribal Percentage Distribution in all Districts

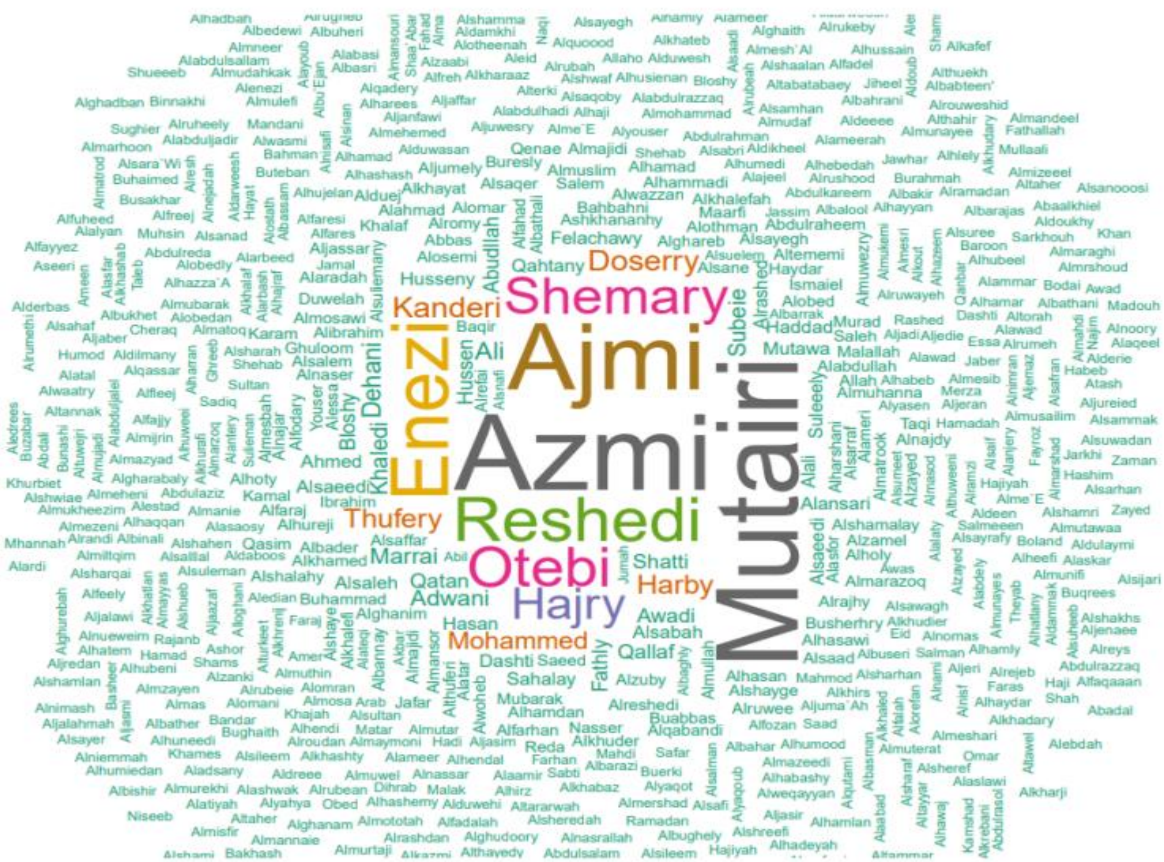
District	Adaween	Awazim	Bani Ghanim	Bani Hajar	Bani Khaled	Bani Tameem	Dufhier	Duwasir	Ejman	Enizah
1	0.08%	7.33%	0.22%	0.58%	0.19%	0.10%	0.23%	0.21%	0.79%	0.65%
2	0.43%	3.42%	2.94%	3.30%	0.64%	0.03%	1.51%	0.37%	1.31%	5.87%
3	0.30%	0.91%	0.50%	0.95%	0.28%	0.08%	0.35%	0.51%	0.92%	1.31%
4	1.41%	2.74%	2.39%	1.09%	1.43%	0.02%	5.43%	0.58%	4.30%	9.84%
5	0.32%	17.01%	0.15%	4.65%	0.58%	0.03%	0.30%	2.28%	14.17%	2.47%

District	Failakawi	Fudhool	Harb	Hirshan	Kandari	Murrah	Mutair	Oteibah	Qahtan	Qena`at
1	3.00%	0.18%	0.15%	0.00%	2.14%	0.03%	0.79%	0.93%	0.15%	1.62%
2	1.74%	0.77%	0.57%	0.31%	1.29%	0.12%	4.32%	1.68%	0.22%	1.56%
3	3.06%	0.26%	0.23%	0.01%	2.32%	0.07%	2.01%	2.13%	0.18%	1.75%
4	0.57%	0.82%	1.91%	0.21%	0.54%	0.18%	20.13%	2.68%	0.39%	0.10%
5	1.25%	0.86%	0.88%	0.01%	1.20%	1.62%	5.05%	6.00%	0.79%	0.28%

District	Reshaydah	Shammar	Shararat	Small Tribes	Subaie	Suhool	Sulaylat	Zuoob	Hadar Shia	Hadar Shia	Uknown
1	0.59%	0.48%	0.00%	3.68%	0.43%	0.03%	0.01%	0.03%	38.21%	35.64%	1.55%
2	3.17%	3.51%	0.01%	2.87%	0.51%	0.13%	0.38%	0.08%	39.97%	16.07%	0.88%
3	0.87%	0.90%	0.00%	3.55%	0.50%	0.11%	0.01%	0.05%	38.62%	19.00%	18.25%
4	12.11%	6.62%	0.02%	2.94%	0.33%	0.07%	0.79%	0.39%	7.96%	4.22%	7.78%
5	1.94%	1.80%	0.00%	2.72%	1.17%	0.68%	0.01%	0.28%	11.53%	10.64%	9.34%

Observing the count of the last names see Figure 3 , lead to finding 8003 unique last names (family names) and observing that 87% of the family sizes are less than 100 individuals. In addition, the distribution of the voter's last names in Kuwait shows that there are eight families dominating the county by their number of individuals; those eight families identify themselves as a tribe. At the same time, 44% of winners in the election belong to those tribes.

Figure 3. Word Cloud of the Last Names



Religion

The religious background feature can be "Sunni", "Shi'a", or "others" for Christian & unknown religious affiliations. Figure 4. shows the majority of the voters identified as Muslims, with 74.9% Sunah and 14.2% Shiaa. The remaining religions and unidentified families were a minority, with 11.1%. As shown in Table 6. District 1 has the highest percentage of Shia in

Kuwait, with 35.6% of the total of voters. For "others", the highest percentage of "other" can be found in District 3, with 21.4% of the district's total number of voters.

Figure 4. Religious Background Distribution of the Voters in all Districts

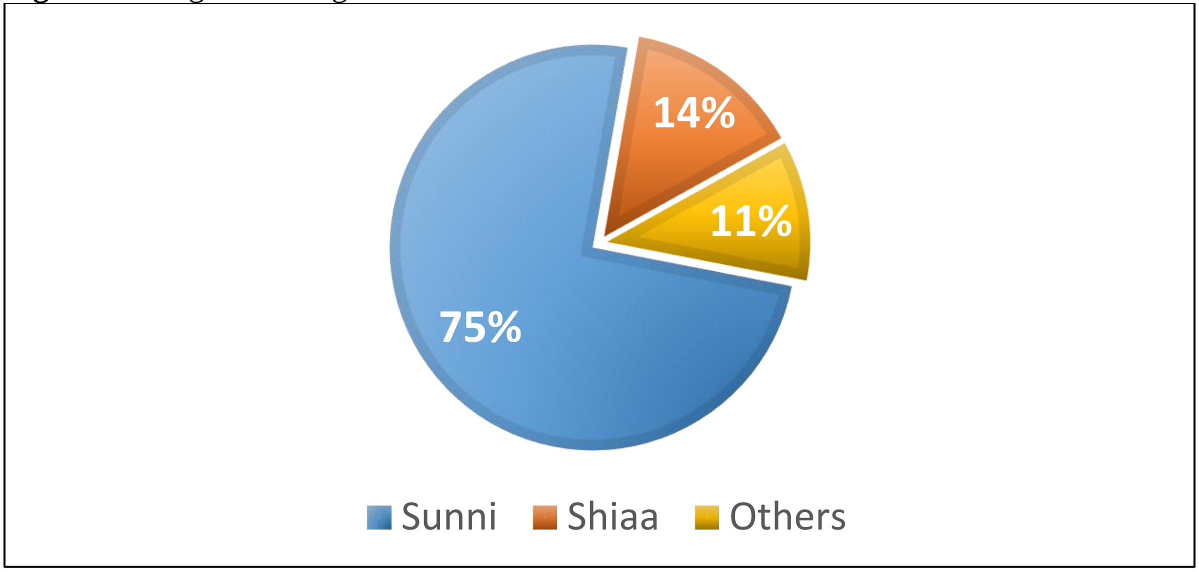


Table 6. Religious Affiliation of the Voters in all Districts

District	M	F	Sunni	Shia`a	Other
1	48.5%	51.5%	60.7%	35.6%	3.7%
2	49.0%	51.0%	81.1%	16.1%	2.9%
3	48.0%	52.0%	59.6%	19.0%	21.4%
4	48.3%	51.7%	84.7%	4.2%	11.0%
5	49.5%	50.5%	77.9%	10.6%	11.5%
total	48.8%	51.2%	74.7%	14.2%	11.1%

Age

The Age feature comprises eleven age bins, each representing a ten-year interval as shown in Figure 5. A significant portion of voters, 54%, were under 40, while 32% fell within the age range of 40-60. The remaining age bins accounted for only 14% of the voters, indicating that most of the voting bodies consisted of early adults. As voters age, their numbers decrease, with only 2% of the total being made up of those aged 90 and above. Figure 6. reveals that this trend of decreasing voter numbers with aging is consistent across genders, with males and females showing similar patterns in each district, differing by only 1%+-, except in the case of the 60+ age group in districts 1 and 4.

Figure 5. Age Percentage Distribution in Different Bins

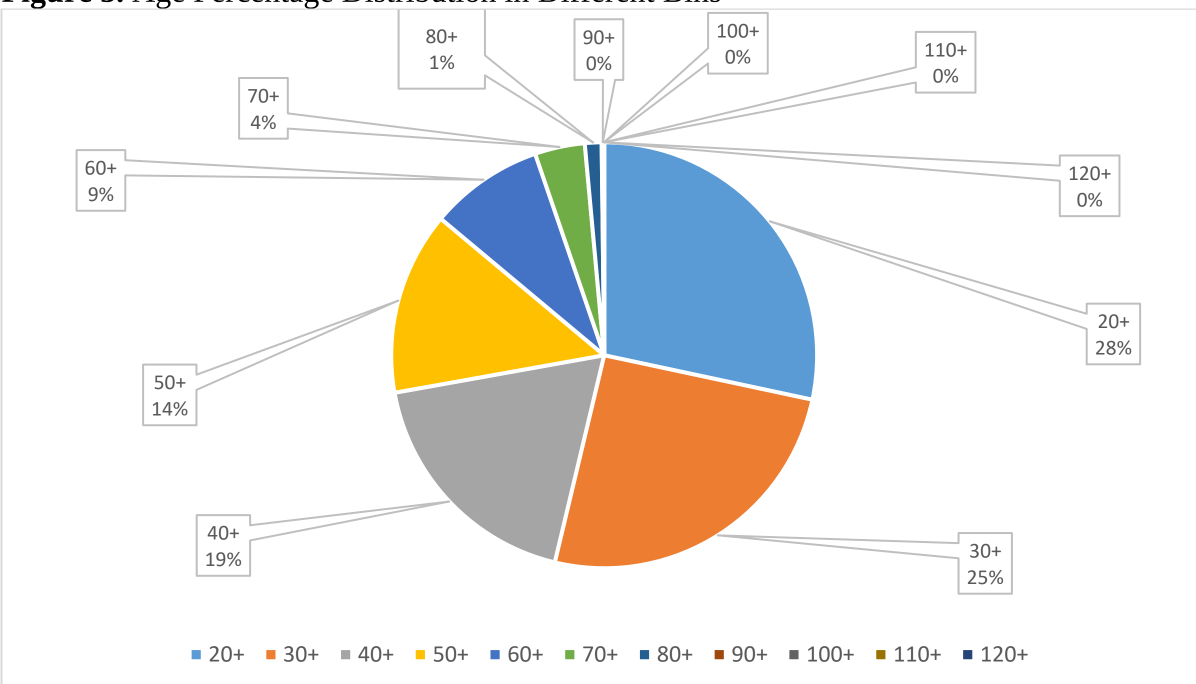
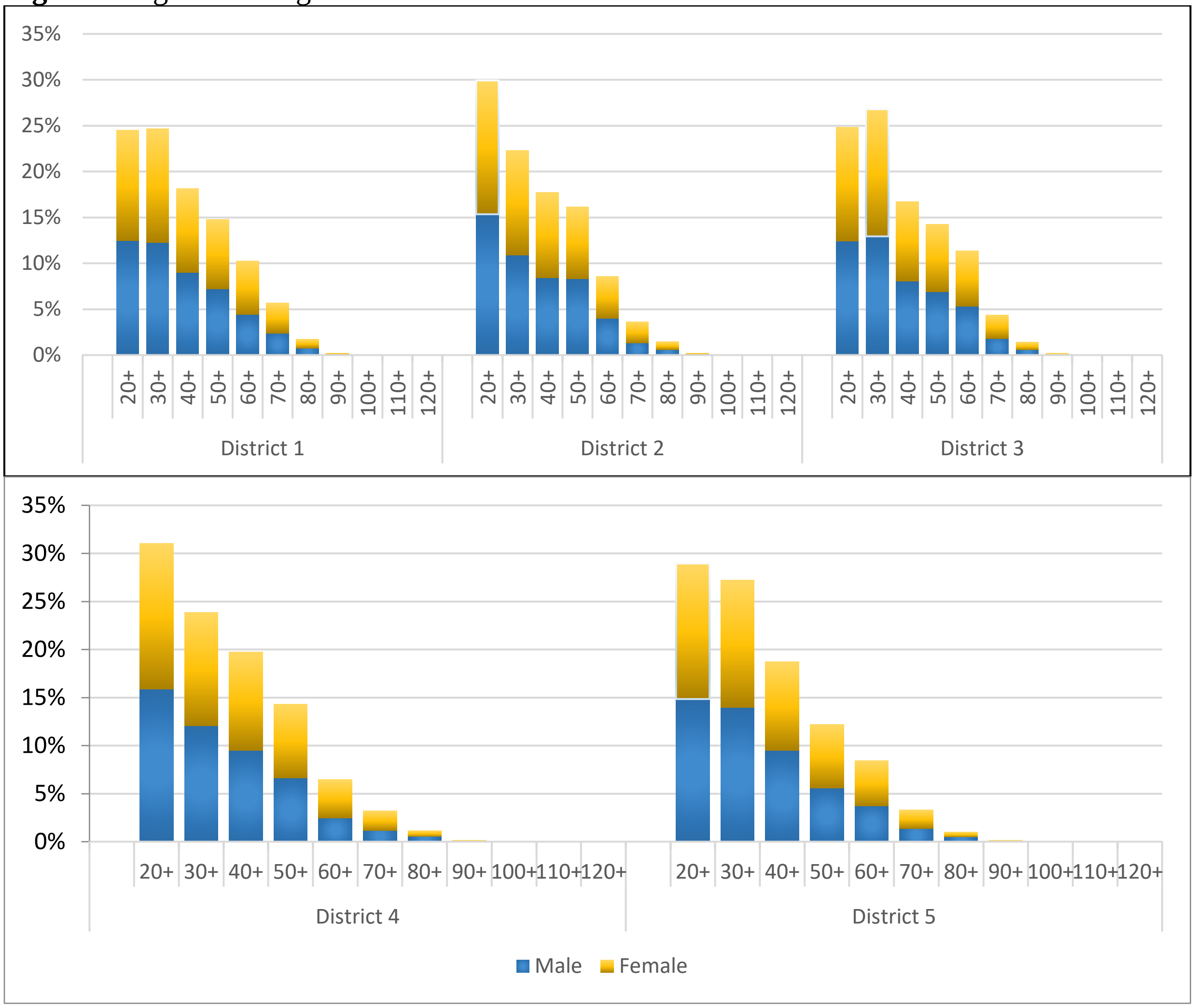


Figure 6. Age Percentage Distribution in Different Bins divided into District and Gender



Occupation

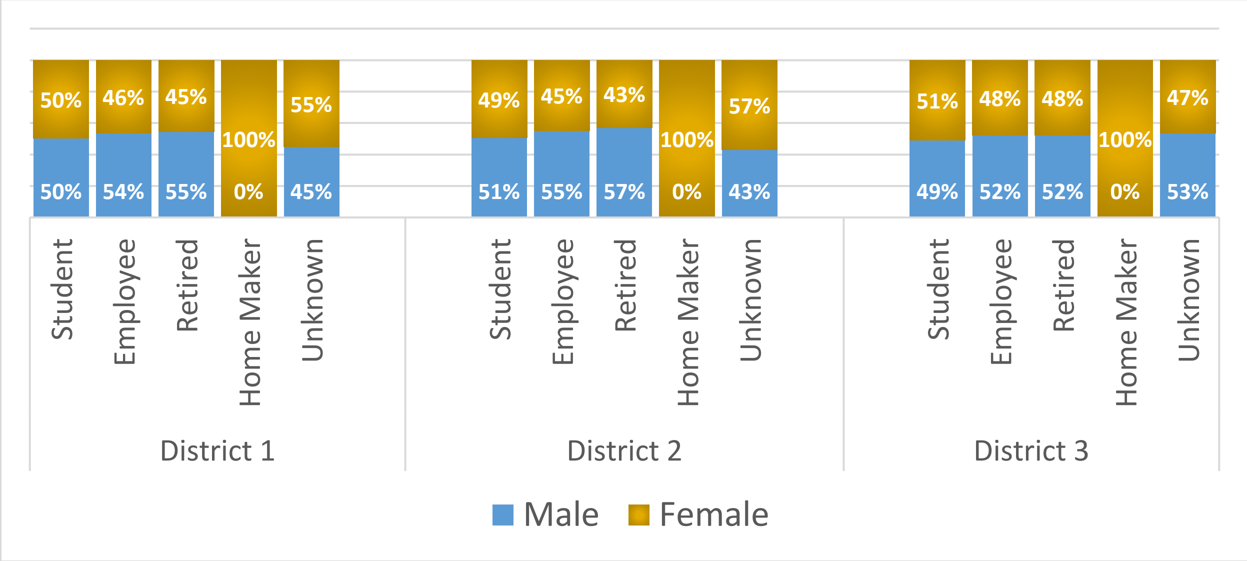
Figure 8. presents "Occupation" feature which include Employee, Student, Home Maker, Retired, and Unknown. These jobs are distributed in descending order, with 54%, 27.5%, 11%, 7.5%, and 0.2% respectively, over the total number of voters. Students and Employees consistently have the highest number of voters in all districts, for both males and females (see Figure 7Figure 8). Table **7** shows that home maker occupation is predominantly for females, with only two males identified as homemakers in district 5. Districts 4 and 5 show a similar trend in terms of occupations, with homemakers comprising the highest percentage of voters at 13.3%

and 12.9% respectively. In districts 1, 2, and 3, home makers rank as the third highest occupation, while in the remaining districts, it falls to the fourth or fifth position. The percentage of unknown occupation is minimal across all districts, always remaining below 1%.

Table 7. Percentage Voters' Occupation Distribution in All Districts by Gender.

D	G	Student	Employee	Retired	Homemaker	Unknown
1	M	11.00%	31.71%	5.77%	0.00%	0.04%
2	M	13.68%	31.02%	4.27%	0.00%	0.03%
3	M	10.69%	32.25%	5.07%	0.00%	0.03%
4	M	17.28%	27.11%	3.83%	0.00%	0.08%
5	M	13.60%	30.56%	5.28%	0.00%	0.09%
D	G	Student	Employee	Retired	Home Maker	Unknown
1	F	10.90%	27.40%	4.79%	8.34%	0.05%
2	F	13.40%	25.30%	3.21%	9.04%	0.04%
3	F	11.07%	29.55%	4.66%	6.64%	0.03%
4	F	17.03%	20.19%	1.05%	13.31%	0.12%
5	F	13.77%	21.59%	2.03%	12.94%	0.14%

Figure 7. Age Percentage Distribution in Different Bins Divided by Job and Gender



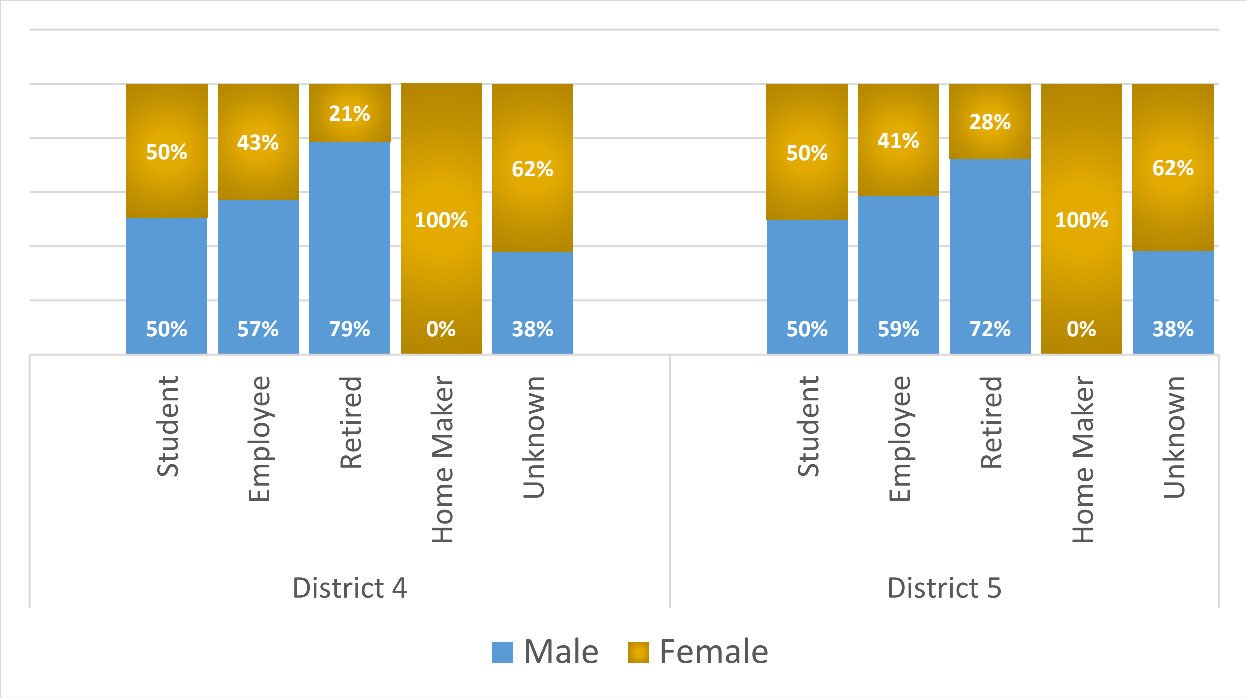


Figure 8. Occupations Distribution Over All Districts.

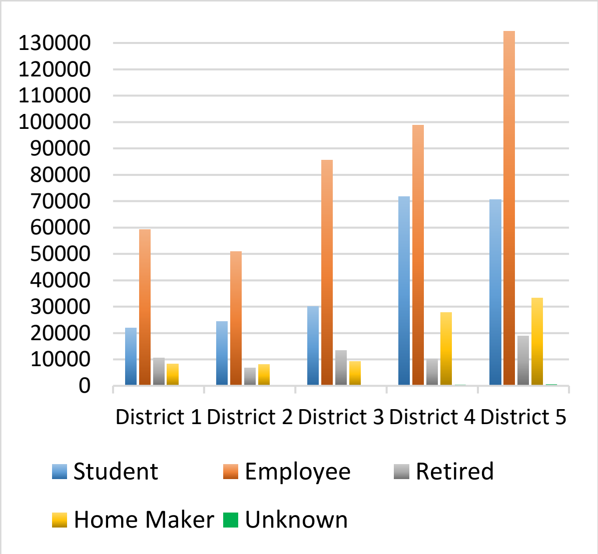
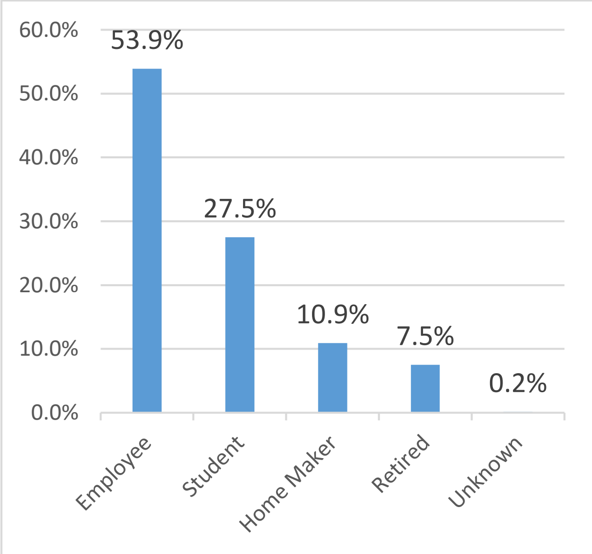


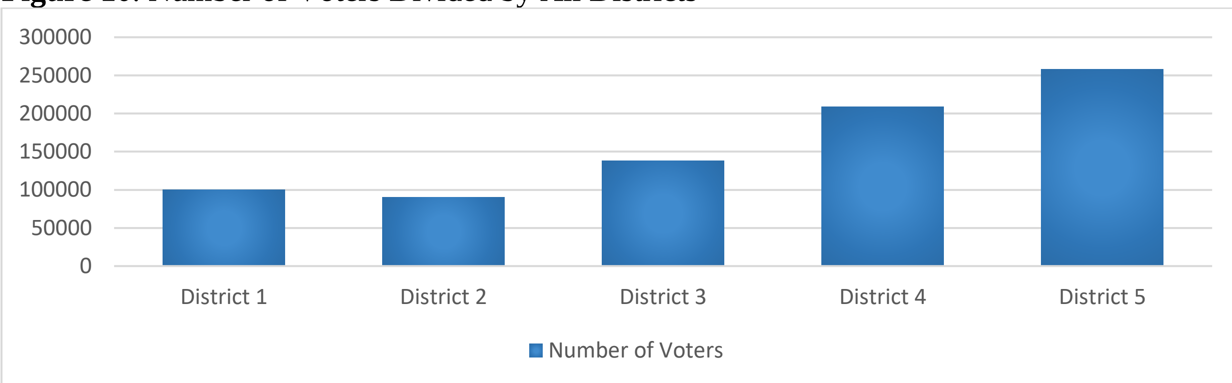
Figure 9. Percentage Distribution of Occupations



District

The district feature comprises 5 districts, with district 5 having the highest percentage of voters at 32%. District 4 follows as the second highest with 26% of the voters, while district 3 accounts for 17% of the voters. District 1 has 12% of the voters, and district 2 has 11% of the voters. Figure 10. shows the number of voters divided by all districts.

Figure 10. Number of Voters Divided by All Districts

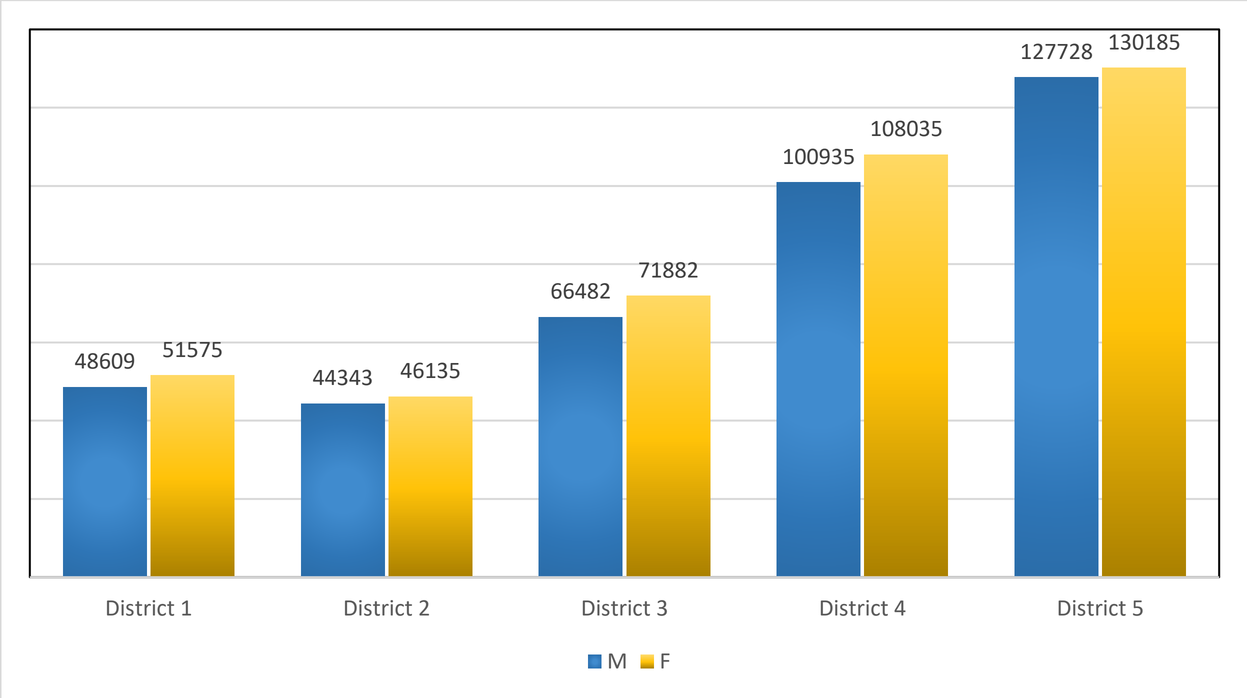


Gender

The biological sex is the gender recognized in Kuwait and only includes males and females. The number of female voters in Kuwait is slightly higher, with 1+% than the number of males, which is 51.2%. The same trend can be seen in all districts, with 52% to District 3, 51% to District 1, 2, and 4, and 50% to District 5.

Figure 11. shows the number of voters by gender divided according to the district.

Figure 11. Male/Female Voters Divided by All Districts



Election Results

Table 8. shows the number of winners and losers in each district, with a total of 492,419 votes. District 5 had the highest percentage of votes with 32.6%, followed by district 4 with 28.3%, district 3 with 16.9%, district 2 with 10.9% , and district 1 with 11.3%. Female candidates represented a range of 2.4% to 16% of the total candidates from all districts, with only one female winner from district 2 and 3, resulting in a total of 2 out of the 50 total seats.

Table 8. The percentage of winners and losers in Each District

Category	District 1	District 2	District 3	District 4	District 5
Female losers	6.70%	8.7%	15.9%	6.10%	2.40%
female winners	0.00%	10%	10%	0.00%	0.00%
Male Losers	93.3%	91.3%	84.1%	93.9%	97.6%
Male winners	100.%	90%	90%	100%	100%
Total Losers	77.8%	78.3%	77.3%	87.8%	87.8%
Total # votes	56111	53999	83638	140185	161374
% Total votes	11.3%	10.9%	16.9%	28.3%	32.6%

Appendix A Table A.1. represents tribes' number of candidates, the number of winners, and losers over the 5 districts. The table shows that 10% of the winners came from the Ejman tribe. The combination of the five tribes of (Mutair, Awazim, Reshayda, and Kanadra) won third of the seats with 34%. Otaibah, Bani Hajar, Bani Ghanim have same power were each control the 4% of the votes. And Dufhier and Qenaat also had the same power with 2% each.

Features Interaction

Appendix A Table A.1 presents the number of Kuwaiti voters in various age categories according to the voters' district 'D', religious affiliation 'R', and gender 'G'. The rows represent the religion and gender distributed in the five districts. The columns-rows intersection is the number of voters in each combination. The combination of gender and the religion with Sunnah and others shows different trends, as shown in Table 9. It is not 50~% distribution anymore with Sunnis in districts 3,4, and 5. District 5 males Sunni has 10% more than females Sunni from the number of voters in the same district, with 43.5% for Sunni males and 34.4% for Sunni females. Same trend with District 4, with 45.5% for Sunni males and 39.2% for Sunni females. For the others, districts 5 and 4 have female percentages that are ten times more than males. District 3 has a number of females 18 times more than the number of males.

Table 9. Percentage Distribution in All Districts by Religious Affiliation

Male	1	2	3	4	5
Sunnah	29.8%	40.1%	37.5%	45.5%	43.5%
Shiaa	17.4%	7.9%	9.2%	1.9%	5.2%
Others	1.3%	1.0%	1.3%	0.9%	0.9%

Female	1	2	3	4	5
Sunnah	30.8%	41.0%	22.1%	39.2%	34.4%
Shiaa	18.3%	8.1%	9.8%	2.3%	5.5%
Others	2.4%	1.9%	20.0%	10.2%	10.6%

Machine Learning and Tools

The Data Science techniques used in this report were descriptive statistics, SQL, clustering, and classification. The descriptive statistics were used to get the count, percentage, max, min and total of the features using MS Access, Excel, and Power Bi. The clustering algorithm was used to have a better understanding for the distribution of the last name feature in the dataset. The implemented algorithm is the Euclidean distance exists in the word cloud package of R using R studio. The classification was used to evaluate the" features validity" to be used for classifying using Weka and online P-value calculator.

Results

In the results chapter, the findings of the study are presented in detail, providing a thorough analysis of the data . The chapter discusses the main findings, highlighting the most significant and interesting results that emerged from the research.

Classification

Correlation

Table 10. Shows the correlation coefficient values between features in the dataset. According to the results, there is a significant positive correlation between the total number of individuals belonging to a particular tribe and the election outcome. Additionally, there is a strong positive correlation between the tribe feature and the religious background feature. On the other hand, there is a strong negative correlation between the district feature and the number of losers in the tribe feature.

Table 10. Correlation Matrix of the Features in the Dataset

Features	District	Tribe	Religious B/G	Gender	Age Group	Occupation	Total of Gen/Occu.	Total of tribe members	Losers in tribe	Has a winner
District	**1.000**	-0.013	0.098	0.016	0.006	0.072	0.066	0.183	**-0.197**	0.095
Tribe	-0.013	**1.000**	**0.393**	0.038	0.056	0.041	0.078	0.197	-0.049	-0.055
Religious BG	0.098	**0.393**	**1.000**	0.165	-0.010	0.021	0.040	0.180	-0.052	-0.053
Gender	0.016	0.038	0.165	**1.000**	-0.053	0.208	-0.037	0.025	-0.008	-0.011
Age Group	0.006	0.056	-0.010	-0.053	**1.000**	0.359	-0.165	0.062	0.035	0.050
Occupation	0.072	0.041	0.021	0.208	0.359	**1.000**	-0.153	0.117	0.026	0.081
Total of Gen/Occu.	0.066	0.078	0.040	-0.037	-0.165	-0.153	**1.000**	0.360	0.102	0.227
Total of tribe	0.183	0.197	0.180	0.025	0.062	0.117	0.360	**1.000**	0.242	**0.607**
Losers in tribe	**-0.197**	-0.049	-0.052	-0.008	0.035	0.026	0.102	0.242	**1.000**	0.217
Has a winner	0.095	-0.055	-0.053	-0.011	0.050	0.081	0.227	**0.607**	0.217	**1.000**

Classification Algorithm Performance

When evaluating the performance of different classification algorithms, the accuracy of the predictions is a crucial metric. The percentage of instances that were classified correctly can determine which algorithm is performing better on the given task. Figure 12. compares seven classification algorithms based on the percentage of instances that were correctly classified. The RandomForest Algorithm Registered the highest percentage of correctly classified instances with a 98.54% while the NaiveBayes algorithm registered the lowest correctly classified instances with 79.25%. Table 11. shows the comparison between 7 classification algorithms based on the percentage of correctly classified instances. Table 12. shows the confusion matrix for all features by classification algorithm. And Table 13. shows the list of rates computed for the above confusion matrix.

Figure 12. Percentage of Correctly Classified Instances

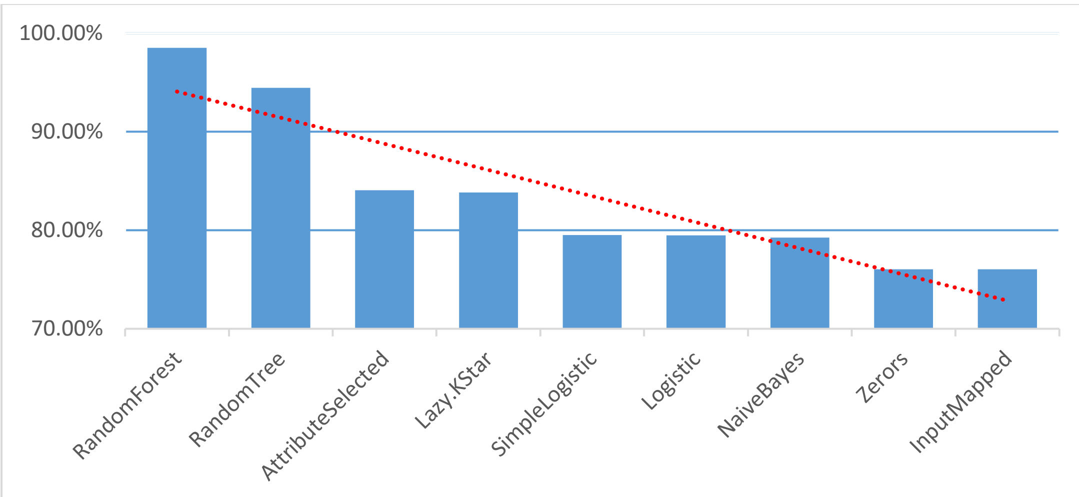


Table 11. Percentage of Correctly Classified Instances

Algorithm	Random Forest	Random Tree	Attribute Selected	Lazy KStar	Simple Logistic	Logistic	Naïve Bayes
Correctly Classified Instances	98.54%	94.46%	84.06%	83.84%	79.51%	79.49%	79.25%

Table 12. Confusion Matrix of All Features by Classification Algorithm

RandomForest			RandomTree			AttributeSelected		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4327	48	a = winner	4227	148	a = winner	4121	254	a = winner
36	1343	b = not winner	171	1208	b = not winner	663	716	b = not winner
Lazy.KStar			SimpleLogistic			Logistic		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4257	118	a = winner	4331	44	a = winner	4328	47	a = winner
812	567	b = not winner	1135	244	b = not winner	1133	246	b = not winner
NaiveBayes								
a	b	<-- classified as						
4310	65	a = winner						
1129	250	b = not winner						

Table 13. Computations of Confusion Matrix Rates for all Features

	Random Forest	Random Tree	Attribute Selected	Lazy. KStar	Simple Logistic	Logistic	Naive Bayes
Total N	5755	5755	5755	5755	5755	5755	5755
Accuracy	0.9852	0.9444	0.8405	0.8382	0.7950	0.7948	0.7924
Misclassification rate	0.0146	0.0554	0.1593	0.1616	0.2049	0.2050	0.2075
True positive rate	0.9890	0.9662	0.9419	0.9730	0.9899	0.9893	0.9851
False positive rate	0.0261	0.1240	0.4808	0.5888	0.8231	0.8216	0.8187
True negative rate	0.9739	0.8760	0.5192	0.4112	0.1769	0.1784	0.1813
Precision	0.9917	0.9611	0.8614	0.8398	0.7924	0.7925	0.7924
Prevalence	0.7602	0.7602	0.7602	0.7602	0.7602	0.7602	0.7602

Predictive Features

To determine which feature can be used to predict the outcome of the elections, several tests were made by removing a selected feature and comparing the correctly classified instances with and without that feature.

Tribe Feature

Figure 13. shows the comparison between the percentage of correctly classified instances with and without the tribe feature. The results show that the RandomForest algorithm showed the highest percentage of correctly classified instances after excluding the tribe feature , and NaïveBayes algorithm showed the lowest classified instances after excluding the tribe feature. Table 14. shows the percentages of correctly classified instances of all the features vs all the features excluding the tribe. Where a significant change is observed in the comparison.

In the confusion matrix tests, similar trends are observed .Table 15. shows the confusion matrix for all attributes excluding tribe by classification algorithm.

Figure 13. Correctly Classified Instances Comparison of the Tribe

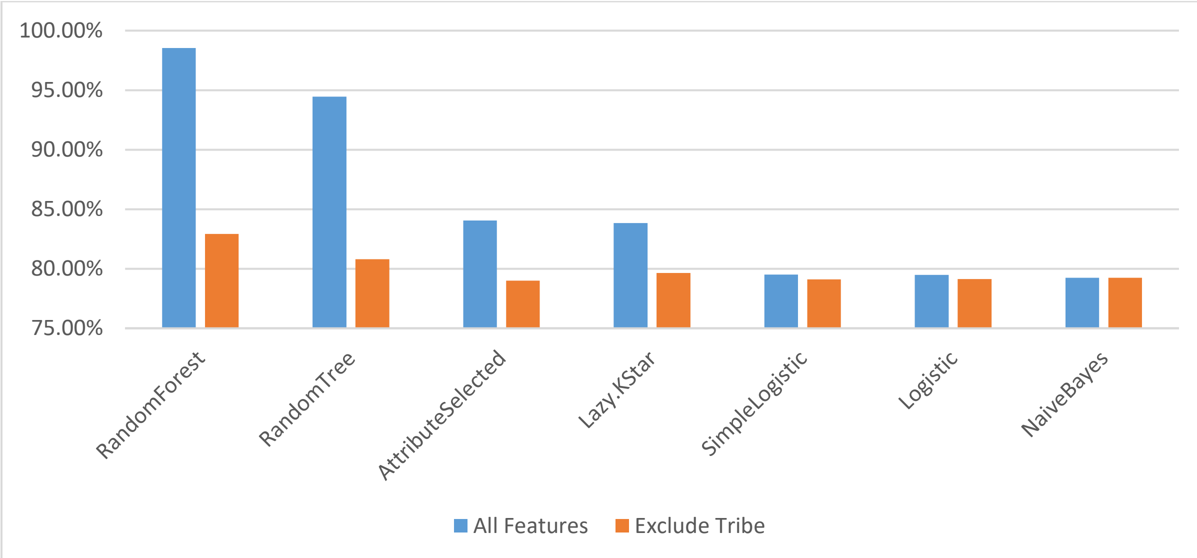


Table 14. Instances Comparison: All vs. Excluding Tribe

Algorithm	RandomForest	RandomTree	AttributeSelected	Lazy.KStar	SimpleLogistic	Logistic	NaiveBayes
All Features	98.54%	94.46%	84.06%	83.84%	79.51%	79.49%	79.25%
Exclude Tribe	82.93%	80.80%	78.99%	79.65%	79.11%	79.14%	79.25%

Table 15. Confusion Matrix of All Features Excluding the Tribe by Classification Algorithm

RandomForest			RandomTree			AttributeSelected		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4000	375	a = winner	3919	462	a = winner	4306	69	a = winner
607	772	b = not winner	643	736	b = not winner	1140	239	b = not winner
Lazy.KStar			**SimpleLogistic**			**Logistic**		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4298	77	a = winner	4316	59	a = winner	4316	59	a = winner
1094	285	b = not winner	1143	236	b = not winner	1141	238	b = not winner
NaiveBayes								
a	b	<-- classified as						

4310	65	a = winner	
1129	250	b = not winner	

Gender Feature

Figure 1Figure 14. shows the comparison between the percentage of correctly classified instances with and without the Gender feature. Table 16. shows the percentages of correctly classified instances off all the features vs all the features excluding the Gender. The results show that the RandomForest algorithm showed the highest percentage of correctly classified instances after excluding the Gender feature , and NaïveBayes algorithm showed the lowest correctly classified instances after excluding the Gender feature. The data shows minimal differences in the comparison of the tests. In the confusion matrix tests, similar trends are observed.

Table 17.Table 16 shows the confusion matrix for all attributes excluding the gender feature by classification algorithm.

Figure 14. Correctly Classified Instances Comparison of the Age Instances Comparison of the Gender

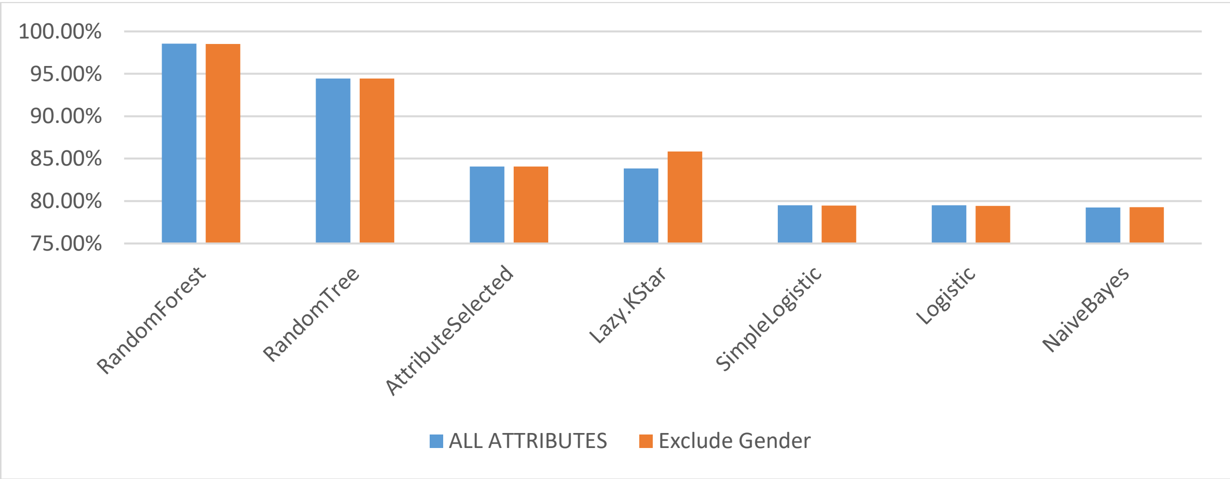


Table 16. Correctly Classified Instances Comparison: All vs. Excluding Gender

Algorithm	RandomForest	RandomTree	AttributeSelected	Lazy.KStar	SimpleLogistic	Logistic	NaiveBayes
All Features	98.54%	94.46%	84.06%	83.84%	79.51%	79.49%	79.25%
Exclude Gender	98.52%	94.43%	84.06%	85.86%	79.46%	79.42%	79.27%

Table 17. Confusion Matrix of All Features Excluding the Gender by Classification Algorithm

RandomForest			RandomTree			AttributeSelected		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4313	62	a = winner	4219	156	a = winner	4121	254	a = winner
50	1329	b = not winner	190	1189	b = not winner	663	716	b = not winner
Lazy.KStar			**SimpleLogistic**			**Logistic**		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4273	102	a = winner	4329	46	a = winner	4325	50	a = winner
713	666	b = not winner	1136	243	b = not winner	1134	245	b = not winner
NaiveBayes								
a	b	<-- classified as						
4310	65	a = winner						
1128	251	b = not winner						

Occupation Feature

Figure 15. shows the comparison between the percentage of correctly classified instances with and without the Occupation feature. Table 18. shows the percentages of correctly classified instances of all the features vs all the features excluding the Occupation. The results show that the RandomForest algorithm showed the highest percentage of correctly classified instances after excluding the Occupation feature , and NaïveBayes algorithm showed the lowest classified instances after excluding the Occupation feature. The data shows minimal differences in the

comparison of the tests. In the confusion matrix tests, similar trends are observed. Table 19. Shows the confusion matrix for all attributes excluding the occupation feature by classification algorithm.

Figure 15. Correctly Classified Instances Comparison of the Occupation

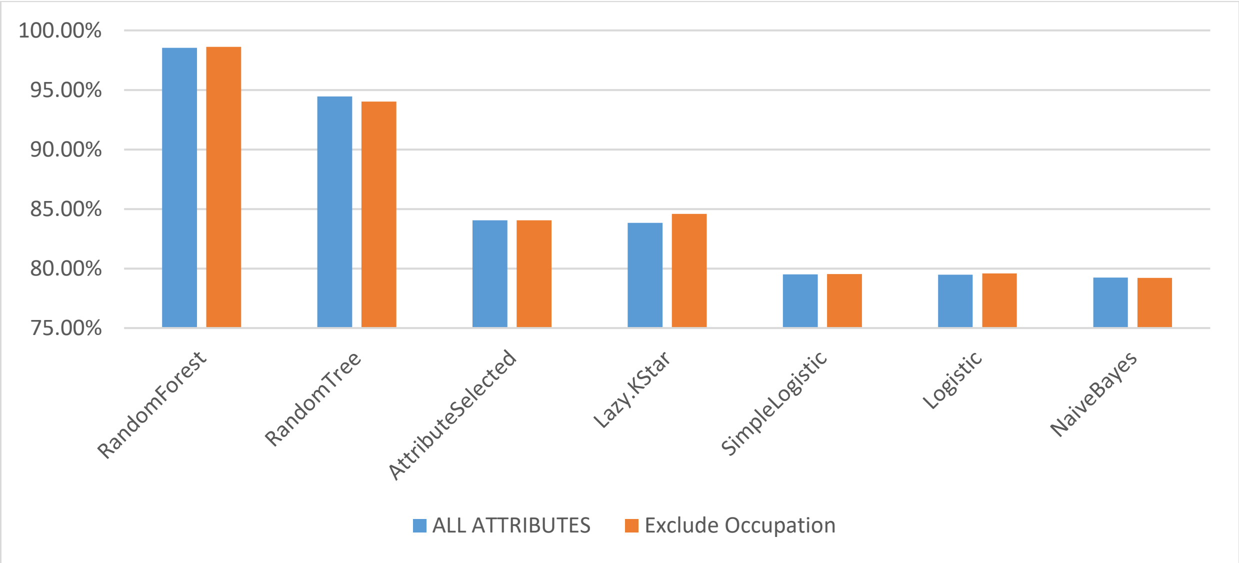


Table 18. Instances Comparison: All vs. Excluding Occupation

Algorithm	RandomForest	RandomTree	AttributeSelected	Lazy.KStar	SimpleLogistic	Logistic	NaiveBayes
All Features	98.54%	94.46%	84.06%	83.84%	79.51%	79.49%	79.25%
Exclude Occupation	98.61%	94.02%	84.06%	84.58%	79.54%	79.60%	79.21%

Table 19. The Confusion Matrix for all Attributes Excluding the Occupation Feature by Classification Algorithm

RandomForest			RandomTree			AttributeSelected		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4339	36	a = winner	4192	183	a = winner	4121	254	a = winner
44	1355	b = not winner	161	1218	b = not winner	663	716	b = not winner
Lazy.KStar			SimpleLogistic			Logistic		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4294	81	a = winner	4332	43	a = winner	4330	45	a = winner
806	573	b = not winner	1134	245	b = not winner	1129	250	b = not winner
NaiveBayes								
a	b	<-- classified as						
4308	67	a = winner						
1129	250	b = not winner						

Religious Feature

Figure 16. shows the comparison between correlation coefficients of all features vs all features without the Religious Background feature. Table 20. shows the percentages of correctly classified instances off all the features vs all the features excluding the religious background. The results show that the RandomForest algorithm presented the highest percentage of correctly classified instances after excluding the religious background feature , and NaïveBayes algorithm showed the lowest classified instances after excluding the religious background feature. Where a significant change is observed in the comparison .In the confusion matrix tests similar trends are observed. Table 21. shows the confusion matrix for all attributes excluding the religious background feature by classification algorithm.

Figure 16. Correctly Classified Instances Comparison of the Religious Background

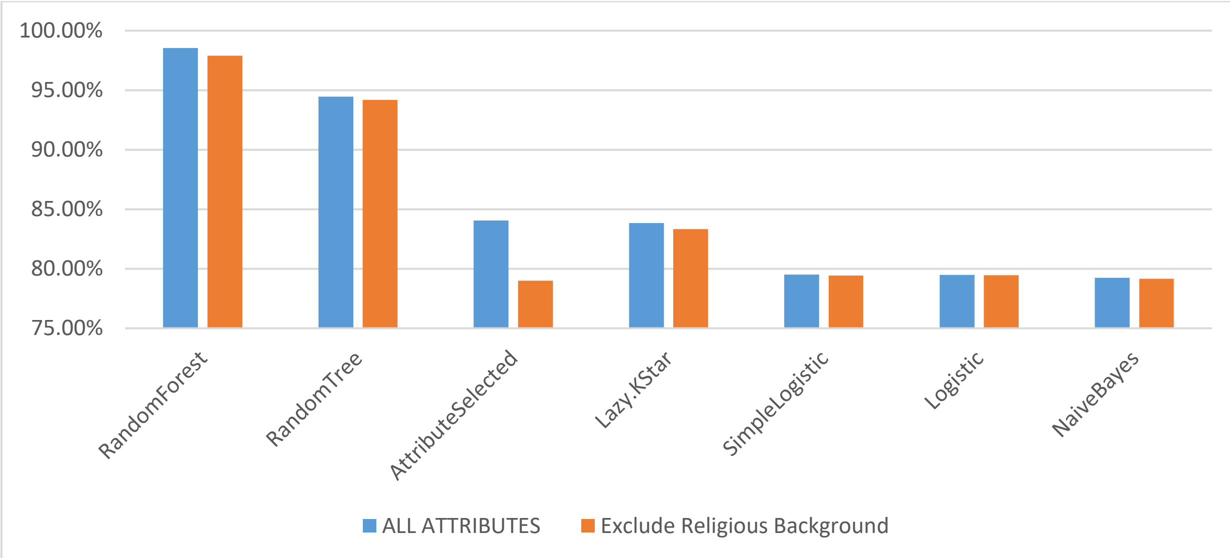


Table 20. Correctly Classified Instances Comparison: All vs. Excluding Religious Background

Algorithm	RandomForest	RandomTree	AttributeSelected	Lazy.KStar	SimpleLogistic	Logistic	NaiveBayes
All Features	98.54%	94.46%	84.06%	83.84%	79.51%	79.49%	79.25%
Exclude Religious Background	97.90%	94.18%	78.99%	83.32%	79.42%	79.46%	79.16%

Table 21. The Confusion Matrix for all Attributes Excluding the Religious Background Feature by Classification Algorithm

RandomForest			RandomTree			AttributeSelected		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4318	57	a = winner	4214	161	a = winner	4121	254	a = winner
64	1315	b = not winner	174	1205	b = not winner	663	716	b = not winner
Lazy.KStar			SimpleLogistic			Logistic		
a	b	<-- classified as	a	b	<-- classified as	a	b	<-- classified as
4268	107	a = winner	4328	47	a = winner	4328	47	a = winner
853	526	b = not winner	1137	242	b = not winner	1135	244	b = not winner
NaiveBayes								
a	b	<-- classified as						
4307	68	a = winner						
1131	248	b = not winner						

Age Feature

Figure 17. shows the comparison between correlation coefficients of all features vs all features without the Age feature. Table 22. shows the percentages of correctly classified instances of all the features vs all the features excluding the age feature. The results show that the RandomForest algorithm showed the highest percentage of correctly classified instances after excluding the Age feature , and NaïveBayes algorithm showed the lowest classified instances after excluding the age feature. The data shows minimal differences in the comparison of the tests. In the confusion matrix tests, similar trends are observed.

Table 23. shows the confusion matrix for all attributes excluding the ag feature by classification algorithm.

Figure 17. Correctly Classified Instances Comparison of the Age

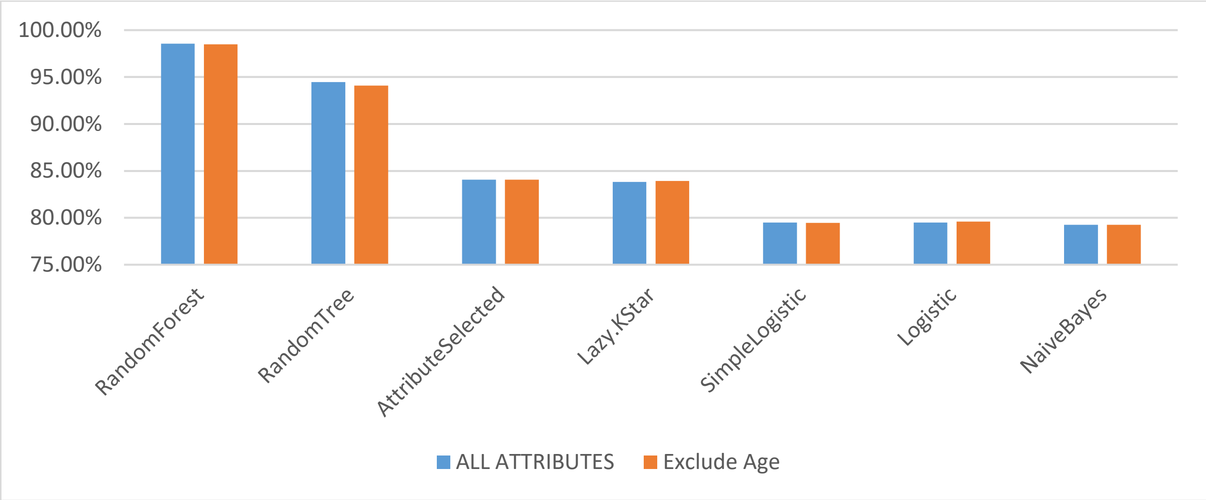


Table 22. Instances Comparison: All vs. Excluding Age Feature

Algorithm	RandomForest	RandomTree	AttributeSelected	Lazy.KStar	SimpleLogistic	Logistic	NaiveBayes
All Features	98.54%	94.46%	84.06%	83.84%	79.51%	79.49%	79.25%
Exclude Religious Background	98.48%	94.07%	84.06%	83.93%	79.48%	79.58%	79.25%

Table 23. The Confusion Matrix for all Attributes Excluding the Age Feature by Classification Algorithm

RandomForest			RandomTree			AttributeSelected		
a	b	<-- classified as	a	b	<-- classified as	A	b	<-- classified as
4326	49	a = winner	4196	179	a = winner	4121	254	a = winner
66	1313	b = not winner	162	1217	b = not winner	663	716	b = not winner
Lazy.KStar			SimpleLogistic			Logistic		
a	b	<-- classified as	a	b	<-- classified as	A	b	<-- classified as
4268	107	a = winner	4328	47	a = winner	4328	47	a = winner
853	526	b = not winner	1134	245	b = not winner	1128	251	b = not winner
NaiveBayes								
a	b	<-- classified as						
4309	66	a = winner						
1128	251	b = not winner						

Discussion

Classification

Correlation Matrix

The correlation matrix shown in Table 10. exhibits the correlation coefficients among various pairs of variables in a dataset. These coefficients assess the strength and direction of the linear association between two variables, with values ranging from -1 to 1. In the context of correlation coefficients, a value of 1 represents a complete positive correlation, -1 denotes a complete negative correlation, and 0 indicates no correlation between the variables.

The correlation between the total number of individuals belonging to a specific tribe and the election outcome registered a significant positive correlation of 0.607. This finding is not surprising, as belonging to a tribe with dominance in a district increases the likelihood of winning. Additionally, the tribe feature and the religious background feature have a positive correlation of 0.393, which is to be expected since most tribal members share a common religious background as shown in Table 6. and Figure 4.

In contrast, there is a negative correlation of -0.197 between the district feature and the number of losers in the tribe feature. This finding is also anticipated, but within a different understanding as the district feature represents a categorical feature . Based on the data presented in Table 2, candidates expect to have a better chance of winning in smaller population districts. This is because the data reveals that as the district number increases, so does the population, which implies that smaller districts are more likely to attract tribal-based anticipation for an easy win. However, the actual turnout results demonstrate the opposite outcome.

Classification Algorithm Performance

Various algorithms may perform differently on different datasets. Therefore, selecting the most effective classification algorithm for a given task is crucial. Therefore, assessing the performance of classification algorithms is a critical step in machine learning and data analysis, and it can aid in choosing the most suitable algorithm for a specific task.

In this report, several classification algorithms were tested to gain a better understanding of their ability to identify instances correctly. The findings revealed that all classification algorithms had a similar trend when using the dataset, but the top three algorithms produced the same general outcome as shown in Table 11. Moreover, the RandomForest algorithm consistently achieved the highest accuracy among all the algorithms in all of the test cases as shown in Table 14, Table 16, Table 18, and Table 22. To confirm the above findings, a confusion matrix analysis was performed, as shown in Table 12 and Table 13, which shows the two possible predicted classes (a or b) where (a) presents a winner, and (b) presents no winner. The classifier made a total of 5755 predictions. Out of those predictions, the RandomForest algorithm registered the highest accuracy rate with .9852 and the highest precision rate with .9997; hence it will be used as a benchmark for the rest of this report.

The RandomForest algorithm registered accuracy rate of .9852 which means how often the algorithm is correct in predicting the output. In contrast, the misclassification rate of (0.0146) shows how often it makes a wrong prediction. The true positive rate of (.9890) means how many times the actual value was a win, and the algorithm predicted a win. The false positive rate of (0.261) means how often the algorithm predicted a win when the actual value was a no-win. The true negative rate of (.9739) means how many times the actual value was a no-win, and the algorithm predicted a no-win also. The precision rate of (.9739) means how often the algorithm predicted a win and it was a correct prediction. Finally, the prevalence rate of (.07602) means how much an actual win is present in the dataset.

Prediction

An assessment was carried out using the "significance of the difference between two correlations" method to evaluate the predictive features [25]. This approach compared the correlation coefficients of all the features with those obtained by excluding a selected feature. If the resulting probability value is less than 0.05, it implies that the two correlation coefficients are significantly different from each other. Based on the assessment, as shown in Table 24. the interaction between all features excluding tribe and all interaction excluding religion exhibit a significant difference between them, with a probability value close to 0, indicating their suitability as predictors for the election outcome. Conversely, the interaction between all features of gender, occupation, and age features has probability values of 0.3694, 0.6873, and 0.343, respectively, higher than 0.05. As a result, these features cannot be utilized as predictors in this dataset.

Table 24. The Probability Values of Correlated Coefficients

RandomForest	All Features	Exclude Tribe	Exclude Religious Background	Exclude Gender	Exclude Occupation	Exclude Age
Correlation coefficient	0.9456	0.5648	0.9263	0.9438	0.9448	0.9437
Probability	-	0.0000	0.0000	0.3694	0.68733	0.3437

.

Conclusion

This report delves into the question of which demographic attributes contribute to the outcome of the Kuwaiti elections. By utilizing data from the September 2022 Kuwaiti parliamentary elections, five attributes were chosen from both voters' and the election winner's information, namely the Tribe, Religious Background, Gender, Age and Occupation.

The results of this report provide valuable insights into the factors that influence the election's outcome. Notably, the analysis revealed statistically significant results by the interaction of Religious Background, Gender, Age and Occupation. The interaction of Tribe and Gender indicated a positive correlation between these demographic features and the election's outcome. However, the remaining features did not contribute significantly to the election's outcome based on the available dataset.

Overall, the findings of this data analysis align with the experiences of Kuwaiti political analysts who suggest that certain demographic features, like tribe and gender, can influence the results of Kuwait's elections. This report offers valuable insights into the significant features that can contribute to the outcome of an election, which could inform future political strategies and decision-making processes.

Limitations

The research encountered several limitations, particularly in collecting information related to tribal and religious backgrounds. These sensitive issues in Kuwait posed a challenge in finding a reliable data source, resulting in data being collected through personal communication. This process was time-consuming and led to discrepancies, requiring additional efforts to improve the data quality. The researcher assigned some religious affiliations based on their general societal knowledge, which may have introduced bias into the results.

Future Work

The study exhibits promising prospects for further development to advance the precision of forecasting upcoming elections and offer a better understanding of the socio-political dynamics in Kuwait for individuals interested in this field. It is crucial to examine several aspects, such as economic and political groups and smaller geographic divisions within districts, to determine their role in the election outcome. In addition, the number of influencers involved in political campaigns and campaign budgets might influence election results. Although the research employed five classification algorithms, exploring other classification methods could boost the accuracy of the predictions.

References

[1] A. Hagagy. "Kuwaiti opposition wins big in election, standoff with government to endure." Reuters. (accessed 2023).

[2] S. Ghabra, *Kuwait: at the crossroads of change or political stagnation*. Middle East Institute., 2014.

[3] O. B. John. "Will Kuwait's new parliament resolve its political impasse?" *he Middle East Institute* . (accessed 2022).

[4] A. Alqaed. "New Faces, New Elections, Unchanging System in Kuwait." gulf internation forum. (accessed 2023).

[5] P. Salem, "Kuwait: Politics in a participatory emirate," 2007.

[6] A. K. UPADHYAY, "Democratic Politics in Kuwait."

[7] S. Ghabra and T. Alterkait, "Kuwait: The First Arab Gulf Constitutionalism," *Al-Abhath,* vol. 70, no. 1-2, pp. 53-82, 2022.

[8] D. L. Tavana, "The Evolution of the Kuwaiti 'Opposition': Electoral Politics after the Arab Spring," *Issue Brief,* vol. 8, 2018.

[9] A. J. Saadoun, "A geographical analysis of the distribution of the population of the State of Kuwait, according to the estimates of the year 2020," *journal of sustainable studies,* vol. 4, no. 1, 2022.

[10] T. Fibiger, "Invisible and Visible Shi'a: Ashura, State and Society in Kuwait," *Anthropology of the Middle East,* vol. 17, no. 1, pp. 63-78, 2022.

[11] M. E. Alhabib, "The Shia migration from southwestern Iran to Kuwait: Push-pull factors during the late nineteenth and early twentieth Centuries," 2010.

[12] F. Al-Nakib, "Revisiting Ḥaḍar and Badū in Kuwait: citizenship, housing, and the construction of a dichotomy," *International Journal of Middle East Studies,* vol. 46, no. 1, pp. 5-30, 2014.

[13] F. Al-Nakib, "Towards an urban alternative for Kuwait: Protests and public participation," *Built Environment,* vol. 40, no. 1, pp. 101-117, 2014.

[14] F. M. Abu Sulaib, "Christians in Kuwait: A Challenge for Political Tolerance," *Middle East Policy,* 2022.

[15] U. S. D. o. State. "2021 Report on International Religious Freedom: Kuwait." U.S Department of State. (accessed 2023).

[16] element61. "Data Science Methodology." element61. (accessed 2023).

[17] S.-y. S. Kim and J. Zilinsky, "The Divided (But Not More Predictable) Electorate: A Machine Learning Analysis of Voting in American Presidential Elections," 2021.

[18] J. A. Garcia, N. C. Tao, F. Betancourt, and K. Wong, "Deep Learning and Visualization of Election Data."

[19] G. Szkatuła, J. Hołubiec, and D. Wagner, "Forecasting voting behaviour using machine learning–Poland in transition," *Annals of Operations Research,* vol. 97, no. 1-4, pp. 31-41, 2000.

[20] R. Fairstein, A. Lauz, K. Gal, and R. Meir, "Modeling peoples voting behavior with poll information," *arXiv preprint arXiv:1909.10492,* 2019.

[21] A.-a. a. N. Paper. "Al-anba`a News Paper." https://www.alanba.com.kw/elections/ (accessed 2023).

[22] D. W. Scott, "Sturges' rule," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 1, no. 3, pp. 303-306, 2009.

[23] D. W. Scott, "Scott's rule," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 2, no. 4, pp. 497-502, 2010.
[24] StatisticsHowTo.com. "Choose Bin Sizes for Histograms in Easy Steps + Sturge’s Rule." (accessed may,2023, 2023).
[25] D. D. Soper. "Significance of the Difference between Two Correlations Calculator." The Free Statistics Calculators index. (accessed May,2023, 2023).

Appendix A - Reference Tables

Appendix A Table A.1 Percentage Distribution of Voters by Religious Affiliation and Gender

D	R	G	20+ %	30+ %	40+ %	50+ %	60+ %	70+ %	80+ %	90+ %	100+ %	110+ %	120+ %	Total G %
1	Sunni	M	0.08	0.08	0.06	0.04	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.30
		F	0.08	0.08	0.06	0.04	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.31
	Shia`a	M	0.04	0.05	0.03	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.17
		F	0.04	0.05	0.03	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.18
	Uknown	M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
		F	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02
	Total Age Range		0.25	0.25	0.18	0.15	0.10	0.06	0.02	0.00	0.00	0.00	0.00	1.00
2	Sunni	M	0.13	0.09	0.07	0.07	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.40
		F	0.12	0.09	0.08	0.06	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.41
	Shia`a	M	0.03	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.08
		F	0.02	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.08
	Uknown	M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
		F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
	Total Age Range		0.30	0.22	0.18	0.16	0.09	0.04	0.02	0.00	0.00	0.00	0.00	1.00
3	Sunni	M	0.10	0.10	0.06	0.05	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.38
		F	0.06	0.06	0.04	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.22
	Shia`a	M	0.02	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.09
		F	0.02	0.03	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.10
	Uknown	M	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
		F	0.05	0.05	0.03	0.03	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.20
	Total Age Range		0.25	0.27	0.17	0.14	0.11	0.04	0.01	0.00	0.00	0.00	0.00	1.00
4	Sunni	M	0.15	0.11	0.09	0.06	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.46
		F	0.12	0.10	0.08	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.39
	Shia`a	M	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
		F	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
	Uknown	M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
		F	0.02	0.02	0.02	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.10
	Total Age Range		0.31	0.24	0.20	0.14	0.07	0.03	0.01	0.00	0.00	0.00	0.00	1.00
5	Sunni	M	0.13	0.12	0.09	0.05	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.44
		F	0.11	0.09	0.07	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.34
	Shia`a	M	0.01	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.05
		F	0.01	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.06
		M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
		F	0.02	0.03	0.02	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.11
	Total Age Range		0.29	0.27	0.19	0.12	0.08	0.03	0.01	0.00	0.00	0.00	0.00	1.00
	Total Over All Age Range		0.28	0.25	0.19	0.14	0.09	0.04	0.01	0.00	0.00	0.00	0.00	1.00

Appendix A Table A.2. Descriptive Analysis of Candidate's Tribes Distribution over Districts

Districts	Tribes	Murrah	Suhool	Sulaylat	Duthier	Fudhool	Bani Tameem	Bani Khaled	Bani Ghanim	Bani Hajar	Harb	Kandari	Duwasir	Reshaydah	Zuoob	Subaie	Shararat
1	Votes	0	0	0	0	0	0	0	0	0	0	4638	0	0	0	0	0
	Candidate	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
	Losers	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	winner	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	8.30%	0.00%	0.00%	0.00%	0.00%	0.00%
2	Votes	0	0	0	511	295	0	0	3374	2923	0	395	0	1433	0	0	0
	Candidate	0	0	0	1	1	0	0	1	2	0	2	0	2	0	0	0
	Losers	0	0	0	1	1	0	0	0	1	0	2	0	2	0	0	0
	winner	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
	%	0.00%	0.00%	0.00%	0.90%	0.50%	0.00%	0.00%	6.20%	5.40%	0.00%	0.70%	0.00%	2.70%	0.00%	0.00%	0.00%
3	Votes	0	0	0	0	0	0	0	0	0	0	8911	0	634	0	0	0
	Candidate	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	0
	Losers	0	0	0	0	0	0	0	0	0	0	1	0	3	0	0	0
	winner	0	0	0	0	0	0		0	0	0	1	0	0	0	0	0
	%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.80%	0.00%	0.80%	0.00%	0.00%	0.00%
4	Votes	0	0	0	7287	0	0	0	4735	7	3041	0	0	31358	0	0	0
	Candidate	0	0	0	3	0	0	0	1	1	1	0	0	14	0	0	0
	Losers	0	0	0	2	0	0	0	0	1	1	0	0	11	0	0	0
	winner	0	0	0	1	0	0	0	1	0	0	0	0	3	0	0	0
	%	0.00%	0.00%	0.00%	5.20%	0.00%	0.00%	0.00%	3.40%	0.00%	2.20%	0.00%	0.00%	22.60%	0.00%	0.00%	0.00%
5	Votes	2079	2450	1278	21	396	0	0	0	18990	0	6251	6726	2479	0	3022	0
	Candidate	1	1	1	1	2	0	0	0	6	0	3	5	2	0	1	0
	Losers	1	1	1	1	2	0	0	0	5	0	2	5	2	0	1	0
	winner	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
	%	1.30%	1.50%	0.80%	0.00%	0.20%	0.00%	0.00%	0.00%	11.80%	0.00%	3.90%	4.20%	1.50%	0.00%	1.90%	0.00%
	Total /	2079	2450	1278	7819	691	0	0	8109	21920	3041	20195	6726	35904	0	3022	0
	% votes for each tribe	0%	0%	0%	2%	0%	0%	0%	4%	4%	0%	6%	0%	6%	0%	0%	0%

Districts	Tribes	Shammar	Oteibah	Ejman	Adaween	Enizah	Awazim	Qahtan	Mutair	Hirshan	Failakawi	Qena`at	Hadar	Small Tribes	Uknown	Total / % votes in each district
1	Votes	0	0	0	0	0	9738	0	0	0	0	55	41680	0	0	56111
	Candidate	0	0	0	0	0	9	0	0	0	0	1	33	0	0	11%
	Losers	0	0	0	0	0	8	0	0	0	0	1	25	0	0	
	winner	0	0	0	0	0	1	0	0	0	0	0	8	0	0	
	%	0.00%	0.00%	0.00%	0.00%	0.00%	17.40%	0.00%	0.00%	0.00%	0.00%	0.10%	74.30%	0.00%	0.00%	
2	Votes	634	1336	0	0	3522	2022	0	2489	1537	0	2056	31472	0	0	53999
	Candidate	3	1	0	0	4	3	0	3	1	0	1	21	0	0	11%
	Losers	3	1	0	0	4	3	0	3	1	0	0	14	0	0	
	winner	0	0	0	0	0	0	0	0	0	0	1	7	0	0	
	%	1.20%	2.50%	0.00%	0.00%	6.50%	3.70%	0.00%	4.60%	2.80%	0.00%	3.80%	58.30%	0.00%	0.00%	
3	Votes	0	6055	3784	0	46	0	0	2794	0	0	0	60569	0	0	82793
	Candidate	0	4	1	0	1	0	0	1	0	0	0	33	0	0	17%
	Losers	0	3	0	0	1	0	0	1	0	0	0	26	0	0	
	winner	0	1	1	0	0	0	0	0	0	0	0	7	0	0	
	%	0.00%	7.30%	4.60%	0.00%	0.10%	0.00%	0.00%	3.40%	0.00%	0.00%	0.00%	73.20%	0.00%	0.00%	
4	Votes	10508	385	7638	2488	14418	3858	0	47029	1901	0	0	4280	0	0	138933
	Candidate	4	2	2	1	9	2	0	24	2	0	0	14	0	0	28%
	Losers	3	2	1	1	8	2	0	22	2	0	0	14	0	0	
	winner	1	0	1	0	1	0	0	2	0	0	0	0	0	0	
	%	7.60%	0.30%	5.50%	1.80%	10.40%	2.80%	0.00%	33.90%	1.40%	0.00%	0.00%	3.10%	0.00%	0.00%	
5	Votes	0	14899	32725	0	186	43585	231	10890	0	93	0	14281	0	0	160582
	Candidate	0	7	13	0	2	18	1	5	0	1	0	12	0	0	33%
	Losers	0	6	10	0	2	16	1	4	0	1	0	11	0	0	
	winner	0	1	3	0	0	2	0	1	0	0	0	1	0	0	
	%	0.00%	9.30%	20.40%	0.00%	0.10%	27.10%	0.10%	6.80%	0.00%	0.10%	0.00%	8.90%	0.00%	0.00%	
	Total /	11142	22675	44147	2488	18172	59203	231	63202	3438	93	2111	152282	0	0	492419
	% votes for each tribe	2%	4%	10%	0%	2%	6%	0%	6%	0%	0%	2%	46%	0%	0%	