

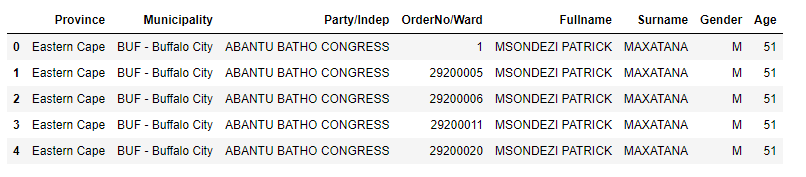
**TASK**

**Exploratory Data Analysis on the Electoral candidates for the 2021 Municipal elections-South Africa Data Set**

[](http://www.hyperiondev.com/portal/)

**Introduction**

The data set that is being analysed contains 95 427 records and 9 variables attributed to the electoral candidates that stood for election in the municipal election of 2021. The data includes information like political party the candidate is affiliated to, the age and gender of candidate, the province etc. The table below gives a sample of 5 records extracted from the data set.



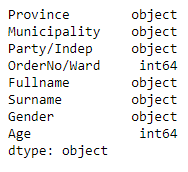
In this project, we seek to determine how far South Africa has come in terms of leadership by females and the youth. As a country where leadership roles have predominantly been filled by older men, both in politics and the corporate environment, it is of interest to assess where we are in this regard. Are there political parties or provinces that have a better gender or age group representation than others? These are the questions we seek to answer.

**DATA CLEANING**

Data cleaning is a very crucial stage prior data analysis. In this stage, we assess the data to ensure that the data is suitable for reliable data analysis.

**HANDLING INCORRECT DATA TYPES**

It is important that we ensure that all data types are suitable for computations. Data types were determined and all the data was found to be of the correct type. The data types are given in the table below.

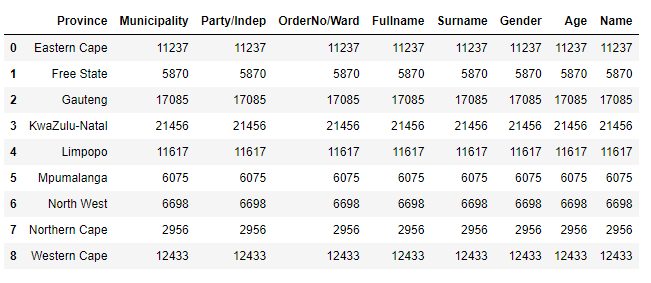
****

**OPTIMISING DATA**

Upon inspection of the data, it was found that the ‘Fullname’ and Surname’ columns would at times contain the same name or surname, but for different individuals. These two columns were combined into a column called ‘Name’ that contains both ‘Fullname’ and ‘Surname’, thereby avoiding possible inaccurate analysis.

MISSING DATA

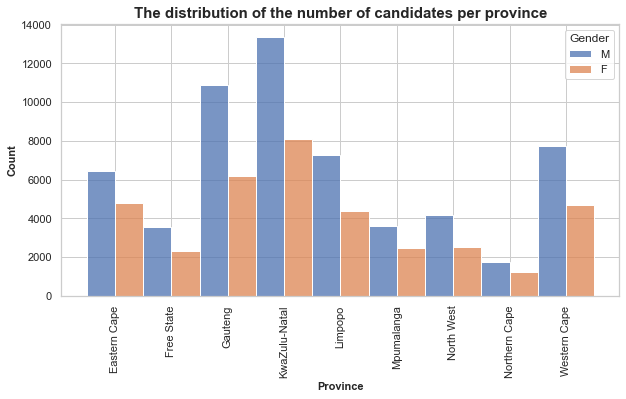
The data set was inspected for any missing data. It was found that there were no missing data. The number of data points per province is tabulated below.



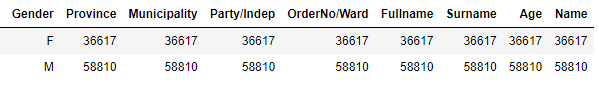
DATA STORIES AND VISUALIZATIONS

**DETERMINING THE NUMBER OF MALE AND FEMALE CANDIDATES PER PROVINCE**

The number of candidates per province were determined. These were further sub-divided according to gender. The graph below gives a representation of the numbers.



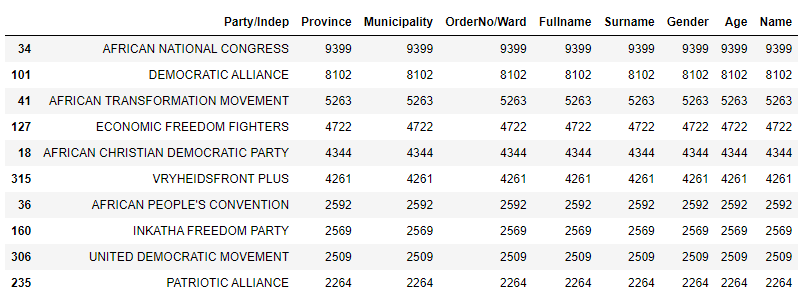
The bar chart above has the blue bars representing the number of male candidates while the red bars represent the number of female candidates per province. The most candidates are in the KZN province, followed by Gauteng and then the Western Cape. There is generally a higher count of males compared to females in all provinces. The 3 provinces with significantly higher male representation are Gauteng, KwaZulu Natal and the Western Cape. The Northern Cape has the smallest difference between the genders.  The table below shows that 62% of the candidates were male.



These proportions are not representative of the total gender ratios in South Africa.

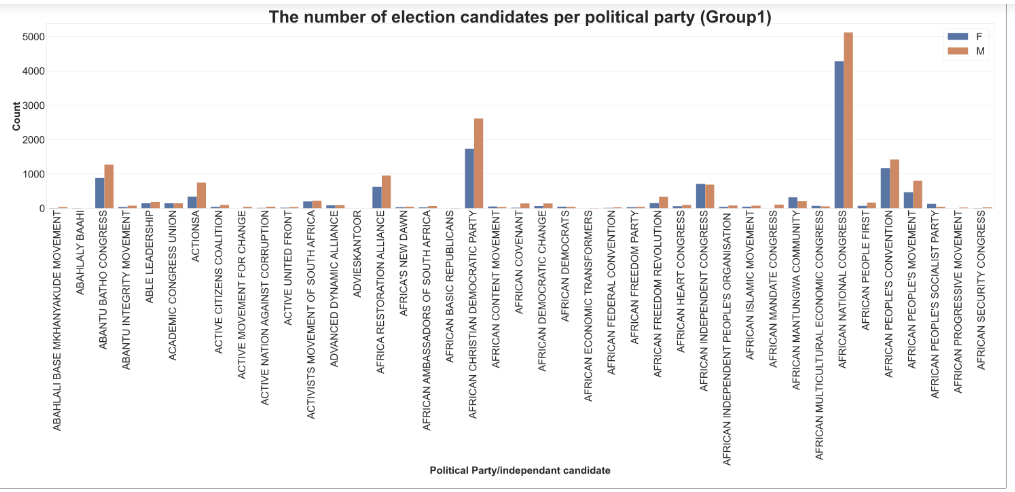
**DETERMINING THE NUMBER OF MALE AND FEMALE CANDIDATES PER POLITICAL PARTY OF PRIVATE CANDIDACY**

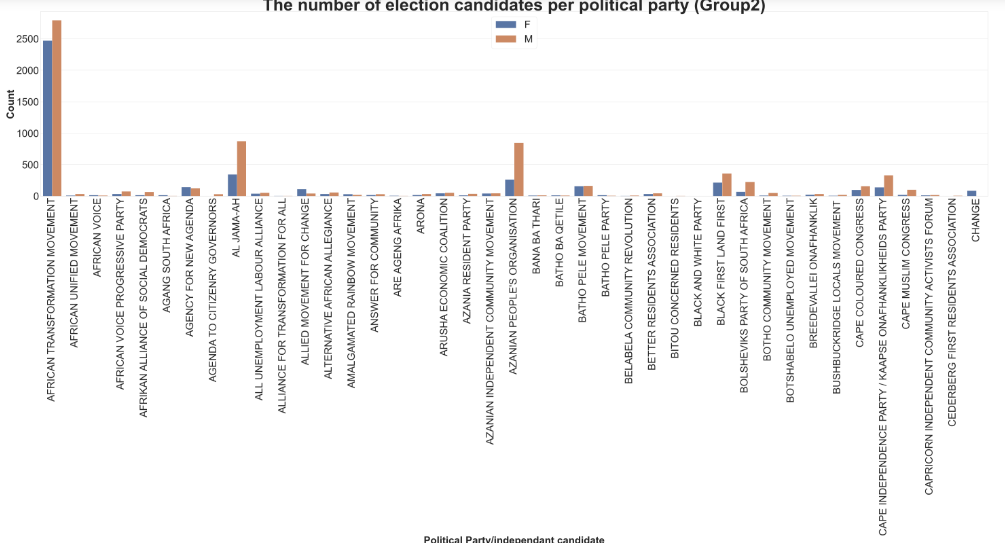
The total number of candidates from each political party were determined. The following table presents the 10 political parties with the most candidates.

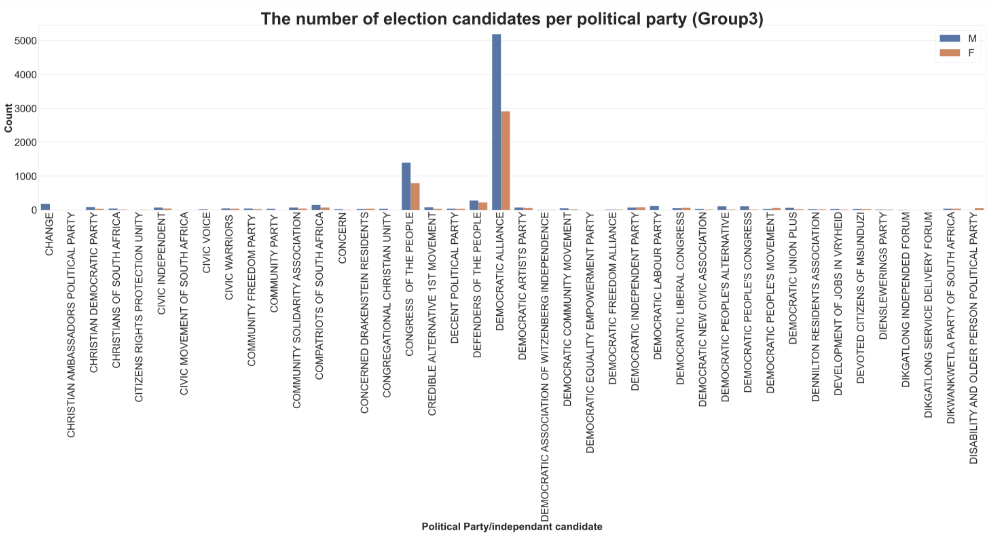


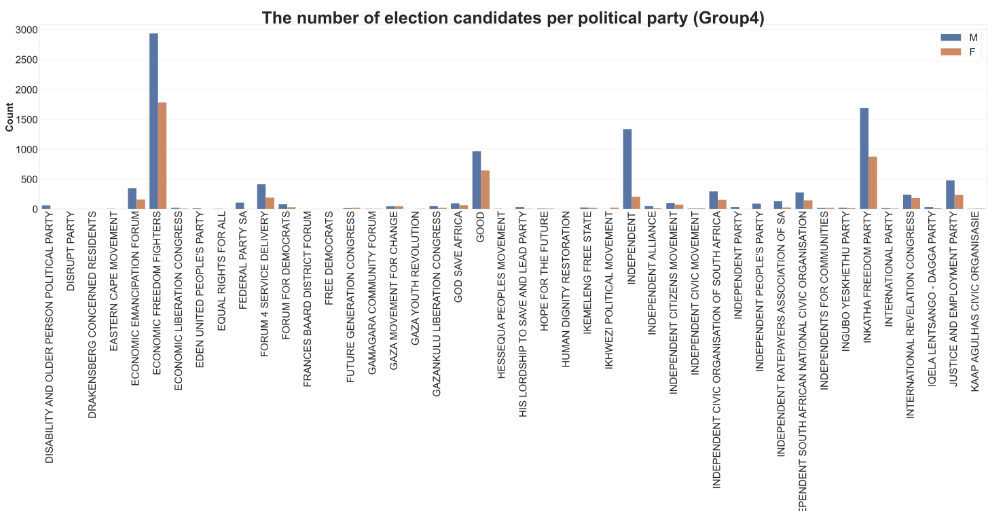
The ANC had the most candidates (9399) out of all the participating political parties, followed by the official opposition party, DA. Because political parties are expected to pay a fee for every candidate who runs for election, decisions on the number of candidates entered and which municipalities are contested, are made strategically. These decisions will be ruled by the financial position of the political party as well as the popularity in a particular municipality. The less popular parties will therefore tend to have fewer candidates because of these reasons.

We are also interested in the distribution of this candidacy according to gender to see which parties have evolved in addressing the countries’ gender inequalities. The following graphs presents the gender distribution in the parties’ candidates. The data was split into 4 groups because there was a total of 324 political parties and independent candidates.







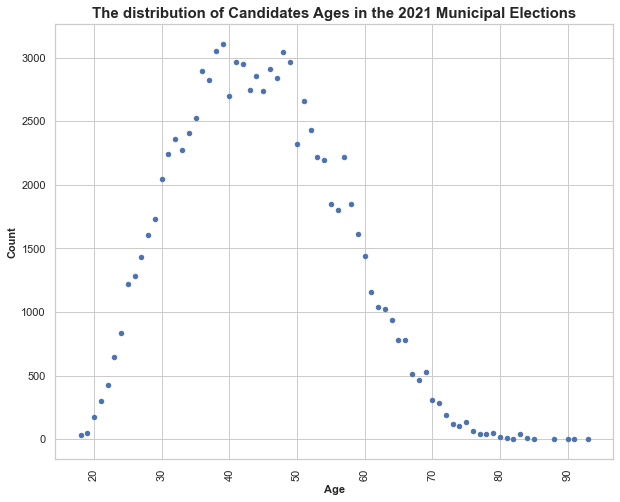


The graphs above show that there is generally a higher male candidate count in the political parties. A few exceptions exist in some small parties, namely;

* Agency for New Change
* Allied Movement for Change
* Batho Pele Movement

**DISTRIBUTION OF CANDIDATES’ AGE**

A healthy distribution of age in leadership is one that allows the older and more experienced leadership working with the younger leaders who are less experienced. This ensures swift continuity in cases of retirement or any loss. The graph below shows the distribution of age in the 2021 candidates.



The ages of the candidates are the focus of the analysis in the above cell. We can see that the youngest candidate is 18 years old and the oldest 93 year old. The average age of the candidates is 44 years. Age in political leadership is another topic of great importance as representation of age groups is seen as important for skills transfer from more experienced (older) to the younger who may be less experiences. The age distribution of this candidate list can be viewed as a healthy one that would ensure continuity. The question to be asked is whether the election outcome was also as representative in terms of age

**CONCLUSIONS**

* There is room for improvement in female representation in political leadership
* The age representation in the candidacy is wide and is considered good for continuity and skills transfer

