

**SEMESTER ONE 2024/2025 ACADEMIC YEAR**

**SCHOOL COMPUTING AND IMFORMATICS TECHNOLOGY**

**DEPARTMENT OF COMPUTER SCIENCE**

**POST GRADUATE DIPLOMA IN COMPUTER SCIENCE**

**MCS 7103: MACHINE LEARNING**

**Assignment 1: Exploratory Data Analysis Process**

**KAMUGISHA ISAAC**

**2400721953**

**2024/HD05/21953U**

# 1. Introduction

## 1.1. Initial Question

How do gender and age distributions differ in parliaments across different countries?

## 1.2. Dataset Overview

The Dataset contains the following features with 922 datapoints.

|  |  |  |
| --- | --- | --- |
| Variable Name | Description | Data Type |
| Country | Name of the country | String |
| Date | Date of data collection | Date |
| Year | Year of data collection (usually an election year) | Integer |
| Number of MPs | Total number of MPs in the Lower House | Integer |
| Percent female MPs | Percentage of female MPs in the Lower House | Float |
| Mean age | Average age of all MPs | Float |
| Median age | Middle age of all MPs | Float |
| Percent MPs aged 30 or under | Percentage of MPs younger than 31 | Float |
| Percent MPs aged 35 or under | Percentage of MPs younger than 36 | Float |
| Percent MPs aged 40 or under | Percentage of MPs younger than 41 | Float |
| Percent MPs aged 41 to 60 | Percentage of MPs aged 41 to 60 | Float |
| Percent MPs aged 61 or over | Percentage of MPs aged 61 or older | Float |
| Percent female MPs aged 30 or under | Percentage of female MPs younger than 31 | Float |
| Percent female MPs aged 35 or under | Percentage of female MPs younger than 36 | Float |
| Percent female MPs aged 40 or under | Percentage of female MPs younger than 41 | Float |
| Percent female MPs aged 41 to 60 | Percentage of female MPs aged 41 to 60 | Float |
| Percent female MPs aged 61 or over | Percentage of female MPs aged 61 or older | Float |
| Age Representation Index 30 or under (ARI 30) | Percentage of MPs under 31 relative to the population | Float |
| Age Representation Index 35 or under (ARI 35) | Percentage of MPs under 36 relative to the population | Float |
| Age Representation Index 40 or under (ARI 40) | Percentage of MPs under 41 relative to the population | Float |
| Age Representation Index 41 to 60 (ARI 41 to 60) | Percentage of MPs aged 41 to 60 relative to the population | Float |
| Age Representation Index 61 or over (ARI 61 plus) | Percentage of MPs 61 or older relative to the population | Float |
| Coverage MPs | Percentage of MPs for whom data was collected | Float |
| Coverage Age | Percentage of MPs for whom age data was collected | Float |
| Coverage Gender | Percentage of MPs for whom gender data was collected | Float |

# 2. Data Wrangling

## 2.1 Handling Missing Data

Most columns in the dataset are complete with all 922 datapoints except the columns indicated in the table below with their corresponding number of missing values

| **Country Code** | **Age Representation Index 30 or under (ARI 30)** | **Age Representation Index 35 or under (ARI 35)** | **Age Representation Index 40 or under (ARI 40)** | **Age Representation Index 41 to 60 (ARI 41 to 60)** | **Age Representation Index 61 or over (ARI 61 plus)** | **Coverage MPs** | **Coverage Age** | **Coverage Gender** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 29 | 54 | 54 | 54 | 54 | 54 | 5 | 5 | 5 |

### 2.1.1 Handling missing data in Country Code

The countries with missing codes were 'Kosovo', 'Northern Ireland', 'Scotland', 'Wales', 'Nagorno-Karabakh', 'South Ossetia', after doing research, I found out that these countries don’t have specific codes according to the ISO 3166-1 encoding that was used in the dataset. I did a manual search and found corresponding codes from official sites like the UK Government’s website.

Using the pandas.DataFrame.apply() method, I replaced corresponding missing codes using the following data;

| **Kosovo** | **Northern Ireland** | **Scotland** | **Wales** | **Nagorno-Karabakh** | **South Ossetia** |
| --- | --- | --- | --- | --- | --- |
| KSV | GB-NIR | GB-SCT | GB-WLS | NGK | SOS |

### 2.1.2 Handling missing values in the Age Representation Index columns

 - Age Representation Index 30 or under (ARI 30)

 - Age Representation Index 35 or under (ARI 35)

 - Age Representation Index 40 or under (ARI 40)

 - Age Representation Index 41 to 60 (ARI 41 to 60)

 - Age Representation Index 61 or over (ARI 61 plus)

I created a simple function to impute using mean of the corresponding columns and replace NaNs using this mean.

### 2.1.3 Handling missing data in Coverage columns

 - Coverage Age

 - Coverage MPs

 - Coverage Gender

Since it’s hard to imagine Coverage data, I decided to create a function that would drop datapoints in columns with missing datapoints less that a given percentage threshold. For these columns I used a threshold of 2.5%.

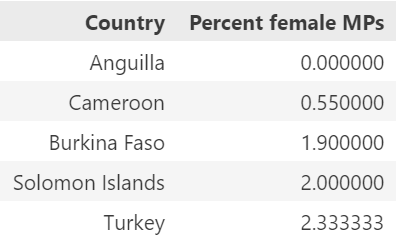
# 3. Exploratory Data Analysis (EDA)

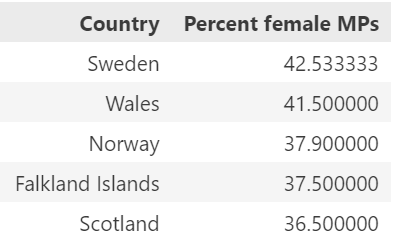
## 3.1 How did the percentage of female MPs vary across countries in the 1990s

### 3.1.1 Analysis

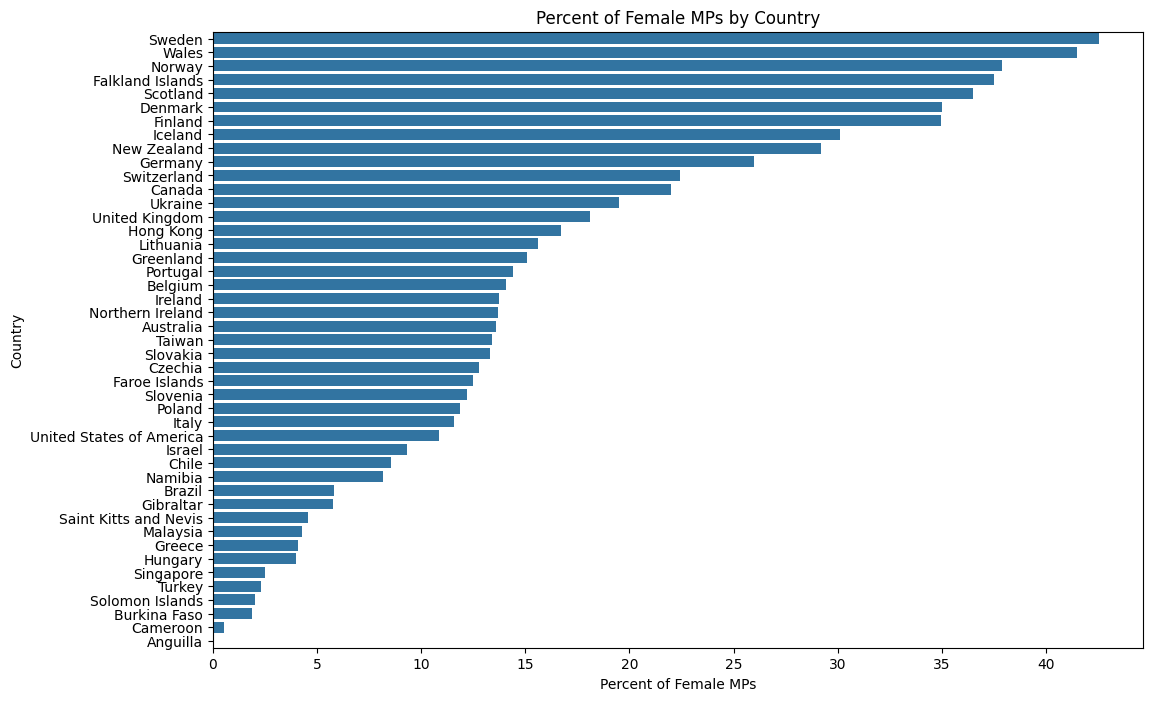
The following summary statistics of the percentage of female MPs across countries in the 1990s were obtained to analyze the distribution;

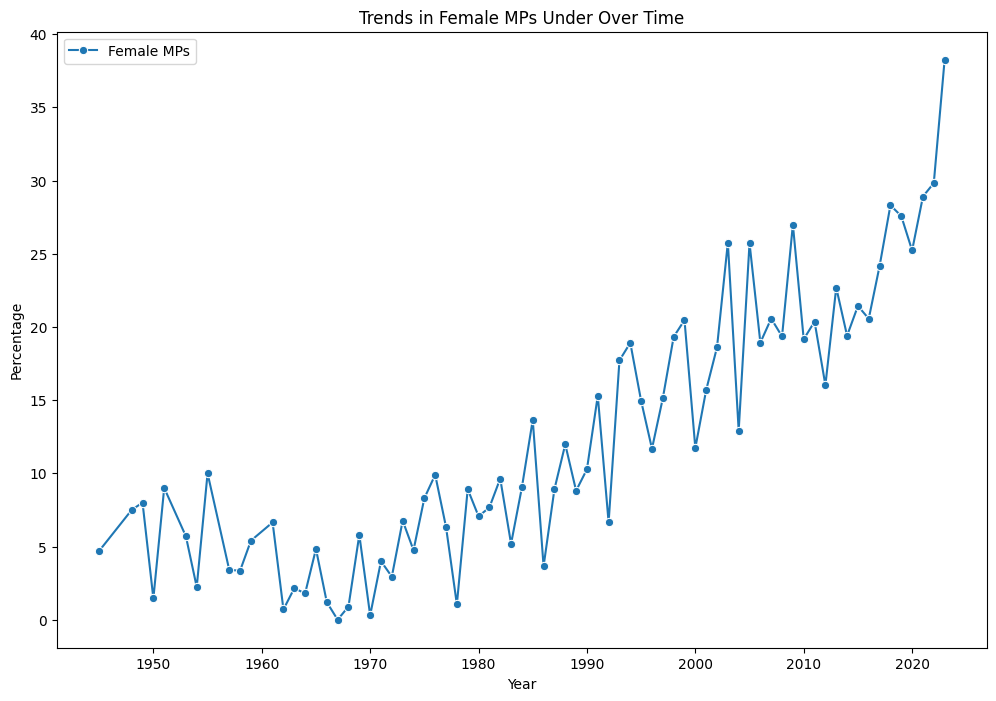
|  |  |
| --- | --- |
| **Summary Statistic** | **Value** |
| Mean | 19.224454 |
| Standard Deviation | 13.807399 |
| Minimum | 0.000000 |
| 25th Percentile | 8.000000 |
| 50th Percentile | 17.400000 |
| 75th Percentile | 29.225000 |
| Maximum | 66.700000 |

Countries with the highest and lowest percentage of female MPs



### 3.1.2 Visualization[[1]](#endnote-1)





### 3.2.3 Findings

In the 1990s, no recorded country had a parliament with up to 50% of females in their Lower House, with countries like Anguilla, Cameroon, Burkina Faso, Solomon Islands and Turkey having less than 3% female representation. However, some countries like Sweden and Wales had female representation going up to 40%.

It is also evident that this up to 2023, the representation of female MPs still lags below 50% showing that males have over time been more involved in politics due to factors like cultural, political or socioeconomic reasons.

# 4. Conclusion

This analysis shows that in the 1990s there were significant disparities of female representation across countries. While some countries made progress towards gender balance in their political systems, others lagged behind with some going as bad as having no female representative in their parliament’s lower house.

Countries like Sweden, Wales, Norway, Falkland Islands and Scotland demonstrated relatively high percentages of female MPs representing progressive movements toward gender equality during the decade. In contrast, countries like Anguilla, Cameroon, Burkina Faso, Solomon Islands and Turkey had a significantly low female representation indicating potential cultural, political, or institutional barriers to women’s participation in politics.

1. Some countries are not shown on the chat because their data was not recorded in the 1990s [↑](#endnote-ref-1)