Annabel Keppel-Palmer
Professor Scharff

Spring 2023

CS 668 Capstone Project

Title : Measuring the attitudes of governments towards their female population : A cluster analysis on discriminatory policies

Name : Annabel Keppel-Palmer

Github : https://github.com/annabelkeppelpalmer/capstone-project

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Research Question(s) :

1) By assessing discriminatory regulations, can we state which governments are actively working to lessen the gender gap?

2) Which governments/ countries have the best policies in place to protect and empower their female citizens?

3) Which areas face the biggest discrimination for women?

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Motivation

1) My own personal experience as a woman having lived in multiple countries

2) Constant changing laws in the United States that affect women on a daily basis

3) Going into a male dominated workforce, starting the conversation with those that do not identify as female

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Dataset : The Gender, Institutions and Development Database (GID-DB)

- Covers 180 countries and territories.
- Helps to analyse women's empowerment and understand gender gaps
- Uses the Social Institutions and Gender Index to assign scores for each category

Each category has a different methodology to assign scores.

E.g. Discrimination in the family

1) Laws on child marriage
2) Laws on household responsibilities
3) Laws on Divorce
4) Laws on Inheritance

Variable		Discrimination in the family					
		Child marriage		Household responsibilities			Divorce
		Law	Practice	Law	Attitudes	Practice	Law
Country			i	i	i	i	i
Australia	i	0.5	0.6	0.5	21.1	1.82	0
Austria	i	0.5	2.8	0.5	58.3	1.95	0
Belgium	i	0.5	2.2	0.5	32	1.55	0
Canada	i	0.75	1.7	0.5	27.7	1.51	0
Chile	i	0.5	5.9	0.75	36	2.24	0.25
Colombia	i	0.5	14.5	0	42.4	3.43	0
Costa Rica	i	0.5	8.4	0.5	..	2.53	0.5
Czech Republic	i	0.5	0.2	0.5	32.5	..	0.25
Denmark	i	0	0.1	0.5	22.1	1.39	0
Estonia	i	0.5	0.5	0.5	23.9	1.54	0
Finland	i	0.5	0.3	0.5	21.1	1.55	0
France	i	0.5	2.8	0.5	34.7	1.61	0
Germany	i	0.25	0.4	0.5	31.7	1.51	0
Greece	i	0.5	1.9	0.5	..	2.81	0.5
Hungary	i	0.5	0.7	0.5	50.9	1.86	0
Iceland	i	0.5	0.5	0.5	13	..	0
Ireland	i	0.5	0.3	0.5	26	2.29	0
Israel	i	0.5	2.5	0.5	47.6	..	0.75
Italy	i	0.5	0.2	0.5	..	2.82	0
Japan	i	0.5	0.6	0.25	15.3	4.76	0.25
Korea	i	0.5	0.2	0.5	55.2	4.43	0
Latvia	i	0.5	0.7	0.5	59.3	2.15	0
Lithuania	i	0.5	0	0.5	43.3	1.95	0.25
Luxembourg	i	0.5	1.1	0.5	..	1.98	0

180 rows (Countries) **35**
columns (laws/ practices)

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Literature Review

- Many papers analyze the gender pay gap only
- Combination of econometric and machine learning approaches
- **Clustering algorithms*** heavily used across many research papers
- For predictions, regression and KNN used
- Gender gap is decreasing, however at a slow rate

Authors Referenced

1. *H. Alatrista-Salas, B. Esposito, M. Nunez-del-Prado, M. Valdivieso (2017)*
2. *K. Gupta, R. Rani, A. Sharma, P. Bansal, A. Dev, R. Gandhi (2023)*
3. *Strittmatter A, Wunsch C (2021)*
4. *Michael Danquah, Abdul Malik Iddrisu, Ernest Owusu Boakye, Solomon Owusu (2021)*
5. *Jaanika Meriküll, Merike Kukk,Tairi Rõõm (2020)*
6. *Albanesi S, Şahin A (2018)*
7. *Hugo Alatrista-Salas, Pilar Hidalgo-Leon, Miguel Nunez-del-Prado (2018)*
8. *Andrey Shastri (2014)*

*From reading these papers, I have decided to use clustering techniques for my project as I am looking to group the countries based on the data.

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Preliminary Data Preprocessing and EDA :

```
df_discrimination
```

	Country	CM Law	CM Practice	HR Law	HR Attitudes	HR Practice	Divorce Law	Inheritance Law
0	Australia	0.50	0.6	0.50	21.1	1.82	0.00	0
1	Austria	0.50	2.8	0.50	58.3	1.95	0.00	0
2	Belgium	0.50	2.2	0.50	32	1.55	0.00	0
3	Canada	0.75	1.7	0.50	27.7	1.51	0.00	0
4	Chile	0.50	5.9	0.75	36	2.24	0.25	0

```
df_physical.head()
```

	Country	Violence Law	Violence Attitudes	Violence Practice	FGM Law	FGM Attitudes	FGM Practice	Missing Practice
0	Australia	0.75	3.2	16.9	0.00	..	..	105.5
1	Austria	0.25	3	13	0.00	..	..	106.2
2	Belgium	0.50	2	24	0.00	..	..	104.7
3	Canada	0.25	7.8	1.9	0.25	..	..	105.1
4	Chile	0.75	10.3	6.7	1.00	..	..	103.9

```
#Check number of rows and columns in dataset
df_discrimination.shape
```

(181, 8)

After removing row

```
df_discrimination.shape
```

(180, 8)

Separated the dataset into four tables based on the different categories

1) Discrimination in the family
2) Restricted physical integrity
3) Restricted access to productive and financial resources
4) Restricted civil liberties

Preprocessing undertaken so far includes:

- Checking for duplicates
- Checking for null values
- Dropping rows
- Checking data types

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Preliminary Data Preprocessing and EDA :

	Country	Access_to_land Law	Access_to_land Practice	non_land_assets Law	non_land_assets Practice	access_to_formal Law	ac
0	Australia	0.00	..	0.0	..	0.0	
1	Austria	0.00	34.5	0.0	..	0.0	
2	Belgium	0.00	15.1	0.0	..	0.0	
3	Canada	0.50	27.4	0.0	..	0.0	
4	Chile	0.75	29.9	1.0	..	0.0	

Further Data Preprocessing needs to be taken, as some of my values are inputted as objects ".."

Before I can continue with EDA, I need to replace these with mean values or remove columns so continue with my analysis

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Methodology & Experimentation

Link to the dataset : https://colab.research.google.com/drive/1sOK8F1BNC4HxUWBjld55qTilAwXep2Km?usp=sharing

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Methodology & Experimentation

I knew that I wanted to use a clustering machine learning technique in order to group together my data.

The goal of this is to group together countries that have similar legal systems (based on their scores in the dataset.

From the list of 180 countries, I initially started to chose to separate the data into 5 clusters. This number is simply just a placeholder while I build the model. I later plan to use the elbow method in order to find the optimal number of clusters

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Methodology & Experimentation

Model:

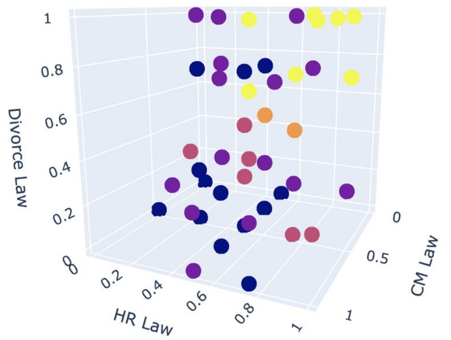
The model I chose to use was the K-Means Clustering algorithm. K means clustering is very useful when you do not have a specific outcome to predict; data scientists use K-Means to find patterns and observe similarities between data points.

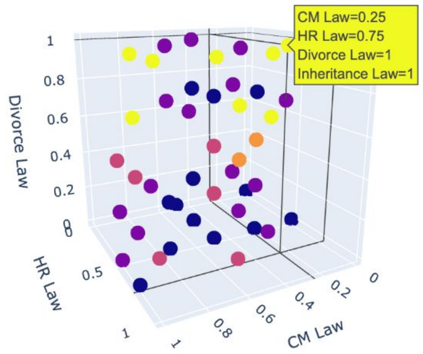
As these clusters were not predetermined, this unsupervised learning technique was one of the most suitable for my choice in data.

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Methodology & Experimentation

A big goal of mine for this project, was to challenge myself by experimenting with new visualization methods.





Above are screenshots of 3d scatterplots showing the correlation between the different variables in the data. Each attribute is represented by a different color.

***The results are just for one table in my dataset - "Discrimination in the family"**

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Methodology & Experimentation

The K Means algorithm was successful in separating the data into 3 different clusters (labelled 0-2).

Using these results, I added an additional column to my dataset containing the cluster of that specific country based on the 4 attributes (‘CM Law’, ‘HR Law’, ‘Divorce Law’, ‘Inheritance Law’).

```
df_discrimination['cluster'] = kmeans.labels_
df_discrimination
```

Country	CM Law	HR Law	Divorce Law	Inheritance Law	cluster
Australia	0.50	0.50	0.00	0.00	1
Austria	0.50	0.50	0.00	0.00	1
Belgium	0.50	0.50	0.00	0.00	1
Canada	0.75	0.50	0.00	0.00	1
Chile	0.50	0.75	0.25	0.00	1
...	...	...	...	...	...
Venezuela	0.50	0.50	0.75	0.00	2
Viet Nam	0.25	0.25	0.25	0.25	1
Yemen	0.75	1.00	1.00	1.00	0
Zambia	0.50	0.50	0.25	0.50	1
Zimbabwe	0.50	0.75	0.25	0.25	1

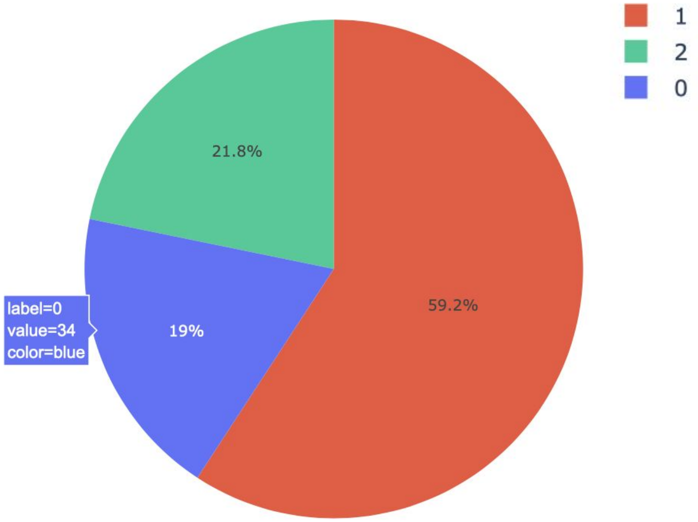
179 rows × 5 columns

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Methodology & Experimentation

Again, to work on my goal of challenging the visualization methods I use in this project, I used the results from the K-Means clustering to represent the data as a pie chart.

This pie chart is interactive, and when the user hovers over the segment, they are able to learn more about that cluster.



In the example to the right, we can see that label 0 (cluster 0) contains 34 countries.

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Results

Once completing this for each table in the dataset, I was able to find the highest and lowest performing clusters for each category. After doing so, it enabled me to see the countries that are the most and least discriminatory for each category (table).

Country	CM Law	HR Law	Divorce Law	Inheritance Law	cluster	Discrimination Mean
Switzerland	0.0	0.00	0.00	0.00	1.0	0.20
Panama	0.0	0.50	0.00	0.00	1.0	0.30

Palestinian Authority or West Bank and Gaza Strip	0.750000	1.000000	1.000000	1.000000	0.0	0.750000
Bahrain	1.000000	1.000000	1.000000	1.000000	0.0	0.800000
Qatar	1.000000	1.000000	1.000000	1.000000	0.0	0.800000

The results above show that for Discrimination in the Family, Switzerland has the lowest mean score meaning the least discriminatory, whereas Bahrain and Qatar have the highest mean score, representing the most discriminatory laws.

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Results

For Restricted Physical Integrity:

Austria	0.250000	0.000000	0.000000	1.0	0.312500
Croatia	0.250000	0.000000	0.000000	1.0	0.312500
Sweden	0.250000	0.000000	0.000000	1.0	0.312500

Haiti	0.75	1.00	1.00	2.0	1.1875
Equatorial Guinea	1.00	1.00	0.75	2.0	1.1875
Congo	0.75	1.00	1.00	2.0	1.1875

Austria, Croatia and Sweden hold the lowest mean scores, whilst Haiti, Equatorial Guinea and Congo have the highest.

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Results

For Restricted Access to Productive and Financial Resources:

Country						
France	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Malta	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Luxembourg	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Portugal	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Sweden	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Netherlands	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Finland	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Norway	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Denmark	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Peru	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Belgium	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Austria	0.000000	0.000000	0.000000	0.000000	1.0	0.200000
Kuwait	1.000000	1.000000	0.500000	1.000000	0.0	0.700000
Guinea-Bissau	1.000000	1.000000	1.000000	1.000000	0.0	0.800000

Twelve different (mostly European) countries hold the lowest mean scores, whilst Guinea-Bissau has the highest.

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Results

For Restricted Civil Liberties:

Argentina	0.0	0.00	0.0	0.00	0.0	0.00
Poland	0.0	0.00	0.0	0.00	0.0	0.00
Slovenia	0.0	0.00	0.0	0.00	0.0	0.00
Spain	0.0	0.00	0.0	0.00	0.0	0.00
Honduras	0.0	0.00	0.0	0.00	0.0	0.00
Viet Nam	0.0	0.00	0.0	0.00	0.0	0.00
Luxembourg	0.0	0.00	0.0	0.00	0.0	0.00
Kyrgyzstan	0.0	0.00	0.0	0.00	0.0	0.00
Moldova	0.0	0.00	0.0	0.00	0.0	0.00
Switzerland	0.0	0.00	0.0	0.00	0.0	0.00
Nicaragua	0.0	0.00	0.0	0.00	0.0	0.00
Italy	0.0	0.00	0.0	0.00	0.0	0.00
Ireland	0.0	0.00	0.0	0.00	0.0	0.00
Iceland	0.0	0.00	0.0	0.00	0.0	0.00
Panama	0.0	0.00	0.0	0.00	0.0	0.00
China (People's Republic of)	0.0	0.00	0.0	0.00	0.0	0.00
Peru	0.0	0.00	0.0	0.00	0.0	0.00
France	0.0	0.00	0.0	0.00	0.0	0.00
Finland	0.0	0.00	0.0	0.00	0.0	0.00
Senegal	0.0	0.00	0.0	0.00	0.0	0.00
Denmark	0.0	0.00	0.0	0.00	0.0	0.00
United Kingdom	0.0	0.00	0.0	0.00	0.0	0.00
Costa Rica	0.0	0.00	0.0	0.00	0.0	0.00
Burkina Faso	0.0	0.00	0.0	0.00	0.0	0.00
Chile	0.0	0.00	0.0	0.00	0.0	0.00
Albania	0.0	0.00	0.0	0.00	0.0	0.00

Belgium	0.0	0.00	0.0	0.00	0.0	0.00
Venezuela	0.0	0.00	0.0	0.00	0.0	0.00
Portugal	0.0	0.00	0.0	0.00	0.0	0.00
Dominican Republic	0.0	0.00	0.0	0.00	0.0	0.00

Saudi Arabia	1.000000	0.250000	1.000000	0.750000	1.0	0.800000
Qatar	1.000000	0.500000	1.000000	0.750000	1.0	0.850000
Oman	1.000000	0.750000	0.750000	0.750000	1.0	0.850000
Iran	1.000000	0.500000	1.000000	0.750000	1.0	0.850000
Yemen	1.000000	0.500000	1.000000	0.750000	1.0	0.850000

For this category, 30 countries had a perfect score of 0.00. The lowest mean scores for restricted civil liberties came from Qatar, Oman, Iran and Yemen.

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Conclusion:

From the results produced, and the study undertaken, what does this show? It shows that we are still very far from truly achieving gender equality in the world today.

It was very surprising to see that some of the most financially powerful and economically developed countries (such as United States, United Kingdom and Germany), whilst were all in the top cluster of each category, they were not the world leaders.

As a woman that lives in one of these countries, it is sad to see just how well some smaller countries such as Switzerland and Finland are doing in comparison. The study hopes to shed light on the realities of living in the United States as a woman, and hopes to bring about positive change.

Measuring the attitudes of governments towards their female population: *A cluster analysis on discriminatory policies*

Future Work:

To continue with this study, it would be extremely useful to assess the "practice" column that was initially dropped during this study. This column shows the effectiveness of how each government actually implement these laws in their countries.

It would be very interesting to compare this column to the initial results we have to see if the clusters and results remain the same, or is there is a drastic change based on the implementation of the law.

This study can also be applied to other social topics in society today, such as racial tension, homophobia and disability laws. With the same methodology, I would love to apply this model to other critical topics that are not spoken about enough.