CHARLES UNIVERSITY

Faculty of Social Sciences

Institute of Economic Studies

Bachelor thesis

2021 **Eliška Velková**

CHARLES UNIVERSITY
Faculty of Social Sciences
Institute of Economic Studies

Eliška Velková

Gender Index in the Czech Public Firms

Bachelor thesis

Prague 2021

Author: Eliška Velková

Supervisor: doc. Ing. Tomáš Cahlík, CSc.

Academic Year: 2020/2021

Bibliographic note

VELKOVÁ, Eliška. *Gender Index in the Czech Public Firms.* Prague 2021. 47 pages. Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor doc. Ing. Tomáš Cahlík, CSc.

Abstract

This bachelor thesis examines the extent of gender inequality in the Czech labour market. More specifically, it explores the under-representation of women on board positions using an analysis prepared by Open Society which contains data of more than 500 Czech public firms. The data analysis is made using the method of Kohonen self-organizing maps. SOMs represent a type of artificial neural network which allows to uncover possible patterns in a dataset and also visualize the multidimensional input data as a two-dimensional mapping while preserving topological properties of the input. To date, there is no academic paper examining gender inequality on decision-making positions in the Czech labour market using the method of Kohonen maps. The used dataset includes 77 Czech regions and 14 variables. A choice of appropriate factors that may influence the participation of women in the labour market is essential. The results are presented in 5 clusters of regions which differ in level of gender gap. In conclusion, our results prove that self-organizing maps are a useful data mining tool which can simply interpret high-dimensional data sets.

Abstrakt

Tato bakalářská práce se zabývá rozsahem genderových nerovností na trhu práce České republiky. Konkrétněji zkoumá nedostatečné zastoupení žen ve statutárních orgánech firem, a to za použití analýzy zhotovené Otevřenou společností, která obsahuje data více než 500 českých veřejných firem. Data jsou analyzována za použití metody Kohonenových samoorganizujících se map. Tyto mapy představují druh umělých neuronových sítí, který umožňuje odhalit skryté vzorce v souborech dat a také vizualizovat mnohadimenzionální vstupní data jako dvoudimenzionální mapu, která zároveň zachovává topologické vlastnosti na vstupu. Doposud se žádná odborná práce nezabývala genderovou nerovností ve vedoucích pozicích na českém trhu práce za použití Kohonenových map. Použitý dataset zahrnuje 77 českých okresů a 14 proměnných. Volba vhodných parametrů, které ovlivňují zastoupení žen na trhu práce, je důležitá. Výsledky jsou představeny jako 5 skupin okresů, které se mezi sebou liší v úrovni genderové nerovnosti. V závěru obdržené výsledky dokazují, že samoorganizující se mapy jsou užitečným nástrojem analýzy dat, který dokáže jednoduše interpretovat mnohadimenzionální soubory dat.

Keywords
gender, inequality, self-organizing map, Kohonen map, the Czech Republic, public firms

Klíčová slova

pohlaví, nerovnost, samoorganizující se mapa, Kohonenova mapa, Česká republika, veřejné firmy

Název práce

Genderový index v českých veřejných firmách

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, January 5, 2021

Signature

Acknowledgment

I would like to express my sincere gratitude to doc. Ing. Tomáš Cahlík, CSc. for his guidance, valuable insights and professional advice in the process of writing this Bachelor thesis, and to my family for their endless support during my studies.

Bachelor Thesis Proposal

Research question and motivation

The fundamental aim of this thesis is to explore gender inequality in Czech public firms. I would like to determine which of the potential factors may have the most substantial impact on the gender gap and to make a comparison across different regions of the Czech Republic. The issue of gender inequality on the labour market has been discussed for many years. Not only do women get paid less than men for a similar kind of work but also it is more difficult for women to achieve board or leading positions. The society has confronted this issue, and gender inequality became a highly important topic over the last few years. Despite this fact, according to the Global Gender Gap Report of the World Economic Forum, the Czech Republic is steadily falling behind the international comparison of gender equality.

Contribution

As already mentioned above, existing researches of the gender inequality on the labour market mostly deal with a problem of a pay gap trying to analyse defining factors or consequences of this phenomenon (Kunze, 2000). In this thesis, I would like to determine the essential factors that influence a gender index in the Czech Republic (Myslíková and Večerník, 2007). On top of that, I am keen on identifying a correlation between gender indices in different Czech regions and the importance of particular determinants. By using knowledge of labour economics (Ehrenberg Smith, 2012), I am going to analyse theories of gender discrimination in the labour market (Becker, 1957). The thesis should result in indicating the most essential factors of gender inequality across the Czech Republic.

Methodology

The thesis focus on Czech public firms since the state, as the main regulator, should represent the values it enforces. Therefore, I am going to use an analysis of the Czech public firms prepared with the contribution of the Economics Institute of the Czech Academy of Sciences. This analysis was also visualised by

Open Society creating a special website which clearly indicates gender indices of 547 Czech public firms. Furthermore, the data are updated every month from the public register of justice. The model adopted in the thesis will be based on a regression of relevant effects of gender inequality across different Czech regions which will be first discussed theoretically and then included into the model. I will consider various sociological and economic influences.

Outline

1. Abstract
2. Introduction
    - Overview of already existing knowledge
    - Organization of the thesis
3. Literature review and hypotheses
    - Literature on gender index, gender board diversity
    - Evolution of gender inequalities on the labour market
    - Factors influencing a gender gap
4. Methodology
    - Description of data
    - Set up of a regression model, examination of each variable
    - Comparison based on region differences
5. Conclusion
    - Interpretation of results
    - Summary of major findings
6. Bibliography

List of academic literature

- Global Gender Gap Report 2017 - Reports - World Economic Forum. Reports - World Economic Forum [online]. Copyright © 2019 World Economic Forum [cit. 10.05.2019]. Available at: http://reports.weforum.org/global-gender-gap-report-2017/measuring-the-global-gender-gap/
- Gendrová mapa. Gendrová mapa [online]. Available at: http://www.genderovamapa.cz/
- Kunze, Astrid, The Determination of Wages and the Gender Wage Gap: A Survey (August 2000). IZA Discussion Paper No.193. Available at SSRN: https://ssrn.com/abstract=251995
- Srovnání krajů v České republice - 2018 | ČSÚ. Český statistický úřad | ČSÚ [online]. Available at: https://www.czso.cz/csu/czso/srovnani-kraju-v-ceske-republice-2018
- Jaumotte, Florence, Female Labour Force Participation: Past Trends and Main Determinants in OECD Countries (December 12, 2003). OECD Working Paper No. 376. Available at SSRN: https://ssrn.com/abstract=2344556 or http://dx.doi.org/10.2139/ssrn.2344556
- Mysíková, M. (2007), "Gender Wage Gap and Its Determinants," in: Večerník, J., ed., Czech Labour Market: Changing Structures and Work Orientations. Sociological Studies 2007:4. Prague: Institute of Sociology of the Academy of Sciences.
- Becker, Gary S. The economics of discrimination. Chicago: University of Chicago Press, c1957. ISBN 0-226-04116-6.
- Ehrenberg, Ronald G. a Robert Stewart Smith. Modern labor economics: theory and public policy. Eleventh ed. Upper Saddle River, N.J.: Prentice Hall, c2012. ISBN 0132540649

Contents

Introduction

Even today women and girls are discriminated against in access to education and healthcare, in political representation, as well as in the labour market. Gender equality, besides representing a fundamental human right, refers to a concept where women's and men's rights, responsibilities and opportunities will not depend on whether they are born male or female. This attribution of presupposed social roles to each sex, while disregarding skills of an individual, has a big impact on the status of women in our society. In consequence, all this creates gender stereotypes that remain a serious problem in countries and societies all around the world (Global Gender Gap Report 2020, World Economic Forum) .

However, gender inequality is much more than just a moral issue. Since women represent half of the world's population, empowering them is essential to achieve sustainable development and economic growth as well (former International Monetary Fund Counsellor M. Obstfeld, IMF Annual Report 2017). Therefore, the matter of gender inequality has become an issue of high political importance in countries all over the world. Major international organizations have provided incentives to reduce gender imbalances. For instance, the United Nations acknowledged gender issue as one of the top priorities among their Sustainable Development Goals (the UN, 2030 Agenda for Sustainable Development, 2019), the European Union set up Gender Equality Strategy 2020-2025 (European Commission, 2019). Moreover, the European Union has also established a special agency EIGE, the European Institute for Gender Equality, to fight gender-based discrimination, and we could find many non-governmental organizations as well. Finally, several countries have introduced gender quotas in order to increase female representation on company boards.

Large number of studies were conducted to inspect the problematic of gender inequality. In case of developing countries, researchers find the main obstacles in their education and healthcare system. On the contrary, developed world finds the gender-related issues primarily in disparities in the labour market. In addition, academic papers analyse gender inequality from different perspectives, for example discrimination on ground of income, representation in decision-making positions, or other employment opportunities. Nowadays, the crucial problem is represented by gender segregation. While its horizontal dimension refers to the tendency of women and men to work in different sectors of the labour market, the vertical one highlights the concept of women maintaining mostly occupations at the lower levels of the organizational pyramid.

One of the methods of data analysis which has proven to be a very useful tool is the self-organizing map (SOM). SOMs, also known as Kohonen maps, refer to a type of artificial neural networks which was developed in 1980's by a Finnish researcher Tuevo Kohonen. This method is widely popular as it allows to transform high-dimensional data sets into usually two-dimensional maps while preserving their topological relations. Therefore, Kohonen maps are extensively used as a clustering and visualization instrument in analyses of explanatory data.

The purpose of this thesis is to analyse the differences in gender inequality across regions of the Czech Republic in order to identify the relationship between socio-demographic and economic determinants and the gender gap in representation on board positions. Moreover, the analysis focuses on firms in the public sector since the state, as the main regulator, should also lead by example. Given that the number of studies covering this specific phenomena in the Czech Republic is very low, this thesis aims at offering a new perspective of evaluating gender inequality on boards which may consequently serve as a step towards implementation of any necessary policies. Moreover, the application of self-organizing maps may provide an innovative approach of data analysis which has not yet been used for examination of this problematic in the Czech Republic.

This thesis is structured as follows: the first chapter reviews the theory of gender and gender inequality mostly from the sociological point of view. Moreover, a description of women's position in the labour market is discussed and the key influences of gender inequality are pointed out. In the second chapter, a methodology of Kohonen self-organizing maps is theoretically explained. The third section describes the dataset and practical application of Kohonen maps on these data. Furthermore, the results are presented. The final chapter contains the conclusion.

1. Gender Inequality

1.1 Theory of Gender Inequality

To fully understand the problematic of gender inequality, it is important to define the notion gender since there is a great difference between sex and gender. While sex refers to a set of biological attributes that represent a dichotomy of two main groups - male and female, gender is much more complex than that.

The term gender was introduced to sociology by a British feminist and sociologist Ann Oakley in her book Sex, Gender and Society (1972). Oakley shows that distinction between women and men is not only biological, but there is a great importance of social aspects as well. She extended her theories by cross-cultural differences in perceiving masculinity and femininity. More specifically, Oakley believes that a basic criterion is always the biological sex of an individual, but apart of that, every culture adopts its specific characteristics of gender. Some cultures acknowledge even more than only two gender categories.

In addition, a relationship between sex and gender was described by Dr. Robert J. Stoller in his book Sex and Gender (1974) where he defines sex as a set of physical conditions, such as chromosomes, internal and external genitalia, gonads, hormonal states and other secondary sex characteristics. On the other hand, according to Stoller, gender has rather a psychological or cultural connotation. It represents social roles that society attributes to each sex and that are socially appropriate in the context of a specific culture as well. In other words, gender can be interpreted as a pattern or a social structure of our everyday practises. Being a man or a woman, a girl or a boy, directly influences the way we dress, behave, choose a profession, and the way our environment perceives us in general. Sociologists and psychologists refer to this concept of behavioural expectations based on one's gender status as a gender role. Traditionally, men are expected to be strong, competitive, independent. On the other hand, women are believed to be more sensitive and caring, expressing their emotions more openly, one could even say weaker compared to men (e.g. Eagley Wood, 2016; Greenglass, 2001). Regarding the gender roles, researchers believe that children start to be aware of gender stereotypes early in childhood. During the preschool years, children already have some knowledge of attributes and activities that are socially associated with each sex. With age, children are becoming more aware of gender stereotypes as they start attending school and engaging in sports or other group activities (Kelly, 2008).

Such a division of social roles creates stereotypical views about women and men and expectations regarding their behaviour. As a consequence, men and women often face gender role stress (also gender role discrepancy) which arises when an individual cannot live up to an appropriate gender role prescribed by society (Kazmierczak, 2010). Higgins (1989) describes in his theory how self-discrepancies between the ideal and actual self may have a negative effect on person's well-being. This problematic is closely linked with a concept of gender identity. The term gender identity can be seen as an individual's internal sense of feeling male or female (sometimes even both or neither of them). It refers to a deep internal self-identification, not necessarily matching the biological sex (M Pilar Sanchez-Lopez and Liminana-Gras, 2017).

Finally, the above sociological analysis of gender helps to better understand the problematic of gender inequality itself. The stereotypical distinction between social roles of women and men remains a serious problem in today's society. Even though the world has made a great progress over the years, women are still often placed into disadvantage when compared to men. Generally, gender equality means that individuals, no matter the gender, are not disadvantaged in relation to decision-making and access to economic and social resources, and that needs of both genders are considered and valued equally. It does not mean that women and men become the same, only that their rights and life opportunities will not be constrained by their sex. Furthermore, it is important to mention a concept of gender equity which refers to a process of achieving the gender equality. More specifically, gender equity ensures fairness in treatment of women and men in order to compensate for social and historical disadvantages since women are not usually in the same starting position as men (International Labour Office, 2000).

There are various aspects of discrimination of women that drive the gender gap in countries all around the world. Not only do women have an uneven access to education or medical care, lower decision making power, but they also suffer from discrimination in the workplace. The latter one include several factors that all together make quite an impact on the labour market and consequently, on the economy as a whole.

1.2 Women in the Labour Market

Narrowing gender differences is desirable not only from the sociological point of view, but also from the perspective of economic growth or development. That is why an increasing number of researchers, policymakers, as well as organizations all around the world study the impact of gender inequality on the overall economy and strive for the attainment of an equitable society (United Nations, 2015). Given that women account for approximately a half of the world's population, the world is missing out on a considerable potential when excluding them. In general, the empowerment of women plays a crucial role in achieving gender equality. Allowing women to reach their full potential, increasing their decision-making power, giving them equal opportunities for personal development, education, and access to resources as well - all this is necessary to make a progress towards the desired equality.

Education

There are several aspects that directly influence the level of gender inequality and consequently affect women's position in the labour market. The key area is definitely education. Investment in human capital foster economic growth, this holds especially in case of developing countries where the gender gap in education represents more serious problem than in the developed world.

Furthermore, Barro and Lee (1994), who first estimated a relationship between gender inequality and economic growth in their influential empirical study, claim that whereas the growth is positively related to male schooling, the correlation between female secondary-school attainment and economic growth is negative. These assumptions were challenged by other studies, such as the one of Dollar et al. (1999), which divides a dataset into two separate parts representing developed and developing countries in order to control for country-specific factors. In contrast to Barro and Lee (1994), they find a positive correlation between the female secondary-school attainment and the growth of per capita income, while controlling for male secondary-school attainment. Thus, the conflicting results across literature are mainly affected by the type of chosen data as well as the econometric strategy used in a model. Another paper by authors Knowles, Owen and Lorgelly (2002) offers a further analysis when suggesting that women's education has a statistically significant positive effect on the productivity per worker, on the other hand, implications concerning men's education are less obvious and are linked to other control variables.

To sum up, in emerging countries, poor families with only limited financial means typically tend to prioritize the education for their sons, while girls are being prepared for their future role of a mother. Paradoxically, such attitude prevents those countries from progress since educated girls would probably postpone starting a family, and more likely send their own children to school as well. Moreover, they would have more opportunities to earn money or to participate in politics (Khan, 2018).

Healthcare

Last but not least, health status is definitely worth mentioning too as health conditions directly influence mental and physical well-being of an individual. There are gender-related differences regarding life expectancy, mortality or morbidity risks. An interesting example may be a study by Miguel and Kremer (2004) which discovers a link between education and health. Miguel and Kremer tried to evaluate treatments (such as deworming or nutritional supplements) on children in developing countries and found out that these treatments considerably improved health and also school participation among children and that they also have a stronger impact on the school outcomes of girls.

However, some economic theories suggest that health status is one of several troublesome variables which can have influence on both economic growth and gender inequality at the same time. As a result, when estimating the effect of gender inequality on economic growth, we obtain a biased coefficient of gender inequality since health status is captured in both variables (Bandiera and Natraj, 2013). In addition, there exist other determinants of economic growth that are correlated with gender inequality (including for instance fertility rate or amount of savings) and may cause the same problem during a regression analysis.

Labour Market Segregation

As already mentioned, the relationship between the level of gender inequality and economic growth varies across countries. In developing countries, the gender issues are primarily connected with situation in education or health care. On the contrary, in more advanced economies, where health and education disparities have significantly narrowed over the years, the main hurdle is represented by gaps in economic and political parity.

One might think that gender inequality in the labour market is not a problem of the developed world's society anymore. However, analysis of labour-related indicators in OECD countries would prove the exact opposite (Oecd.org, 2014). These statistics offer an overview of several measures including labour force participation, employment rate by sex, gender wage gap, or share of female managers. Although these indicators offer an important overview of a situation in the labour market, there is a different approach which rather compares the quality and structure of labour participation for both sexes. That includes characteristics such as a level of gender segregation or labour market discrimination.

Gender segregation remains one the crucial sources of inequality between men and women in the labour market. Bettio et al. (2009) identified the root causes of segregation as differential biological advantages, unequal investment in human capital, preferences and prejudices, socialisation and stereotypes, or organisational practices. In addition, the influence of institutional interventions of state (such as anti-discriminatory policies) and various factors of labour force demand and supply need to be highlighted (Valentová, Šmídová and Katrňák, 2007). Robert M. Blackburn in his numerous papers defines a term overall segregation which consists of two components - vertical and horizontal one (Blackburn, Brooks and Jarman, 2001; Blackburn et al., 2002; Blackburn, 2012). Also, it is important to mention that only the vertical segregation includes the element of inequality.

The horizontal segregation is understood as under or over representation of women and men in occupations or sectors, not ordered by any criterion (Bettio et al., 2009). Usually it is linked with different physical or mental capabilities of women and men that are needed for work in specific sectors. In practice, jobs predominated by women are mostly in health care, education or public administration which also tend to be paid less. Furthermore, women intending to have children concentrate in occupations where an interruption of career is not so problematic. They are also more likely to choose the kind of employment, for instance part-time, that allows them to balance their work with family responsibilities

(International Labour Office, 2000).

On the contrary, the vertical segregation is the one entailing inequality. It indicates the under or over representation of women and men workers in occupations or sectors at the top of an ordering based on desirable attributes, such as income, prestige, job stability, etc., independent of the sector of activity (Bettio et al., 2009). In other words, it describes a disproportional involvement of women and men in different levels of work hierarchy, with women being typically excluded from holding board or senior positions.

Moreover, there exists a relationship between vertical and horizontal segregation as well. A research launched in 2016 by Leanin.org and McKinsey & Company discovered that most CEO's tend to be promoted from line or operational roles rather than staff roles which are usually held by women. As a consequence, chances of women to get promoted to senior positions are significantly reduced (LeanIn.Org and McKinsey & Company, 2016). Valentová, Šmídová and Katrňák (2007) also mention the fact that occupational segregation usually pursues both dimensions at the same time. Thus, referring to the occupational segregation as the horizontal one may become inaccurate. Similarly, Coré (1999) adds that female-dominated professions usually possess lower standing in terms of income, career prospects or social recognition.

Finally, labour segregation largely contributes to other gender-related problems. One of them being a gender pay gap which refers to a difference in average gross hourly earnings between sexes. Based on the latest Eurostat data, the average gender pay gap in the EU currently stands at $14,1\%$, meaning that women earn $14,1\%$ less per hour than men (European Commission, 2020). For instance, the gender pay gap in the Czech Republic now equals $20,1\%$, which is far above the EU average. Pay gap is probably one of the most argued topics in the matter of gender inequality. Wage disparity between genders reflects several factors, most of them have been already mentioned. These include for example the under-representation of women in better paid jobs and over-representation in low-paid work, a wage penalty for motherhood due to their household care responsibilities which undermines their career prospects as well (OECD, 2012), concentration of women in lower-paid sectors (Laura D'Andrea Tyson and Parker, 2019). In addition, in some cases women get paid less for doing the same or very similar work as men. That is referred to as an unequal pay since it does not describe an average value, but rather a discrimination between individuals (Smethers, 2020).

Last noteworthy issue directly linked to the vertical segregation is called the glass ceiling. It can be imagined as an artificial discriminatory barrier that prevents minorities, in this case women, from rising to positions of leadership and power, regardless of their qualification or achievements (Kulik and Rae, 2019). Consequently, such a lack of opportunities for women to get promoted impede income equality.

Gender Board Diversity

At the same time, vertical segregation draws attention to another widely discussed issue which is gender diversity in decision-making positions. Even though the situation is gradually improving, women still have not attained equal representation in company boards. A survey, regularly conducted by the European Institute for Gender Equality, shows the evolution of a percentage of women holding executive positions in the EU countries on a sample of the largest publicly listed companies. For instance, women currently hold approximately 20% of executive seats, but only in 7.5% of cases they act as CEO's. However, compared to 2015, it is a 6 percentage points increase in executive positions and a 4 percentage points increase in case of CEO's (EIGE, 2020).

Moreover, such low levels of women representation seem astonishing as various studies suggest that their presence on company boards are associated with positive influence on firm governance as well as its performance (Terjesen, Aguilera and Lorenz, 2014). Adams and Ferreira (2004) found out that firms with more diverse boards provide their directors with more pay-performance incentives and also more board meetings are held. In addition, Carter, Simkins and Simpson (2002) examined a relationship between the fraction of women or minorities on the board and firm value after controlling for different firm's characteristics, and found a significant positive correlation. An estimation by Wright et al. (2013) concludes that the board diversity has a positive effect also on financial performance as measured by Tobin's q. Last, Joy, Wagner and Narayanan (2019) contribute to the research with an analysis of 520 companies based upon the four-year average of financial measures, such as return on equity (ROE), return on sales (ROS), and return on invested capital (ROIC). In all of these cases, firms with higher number of female directors outperform those with only a small share of women board directors. Finally, several studies supporting this point of view also mention the following arguments - a more gender-diversified board leads to a better understanding of markets which are gender-diversified themselves, it increases firm creativity, as well as it adds new perspectives. Women in the top management may also act as role models for women in the lower-level positions.

On the other hand, there are also several studies that do not find any significant relationship between gender diversity of board members and firms' performance (Rose, 2007), or even a negative one. For example, Daunfeldt and Rudholm (2012) offer an analysis of a large sample of small-sized companies in Sweden which results in a conclusion that an increase in gender diversity on the board has actually a negative effect on returns on total assets with the time lag of two years. They highlight possible methodological problems of other studies including unobserved heterogeneity and reverse causality of those models. Mínguez-Vera and Martin (2011) add their analysis of small and medium Spanish enterprises with a result that presence of women on board positions has a negative influence on firm's performance possibly due to less risky strategies. However, such results might be connected with existence of a gender quota system in a given country. To sum up, despite a large amount of studies, the actual effect of gender board diversity is still not completely clear.

In order to accelerate the achievement of gender-balanced representation on corporate boards, the European Commission decided to take a legislative action by establishing gender quotas. In 2012, a proposal for a Directive was introduced setting up an objective of a 40% female directors on corporate boards by 2020 (European Commission, 2019a). Respecting the principle of subsidiarity within the EU, this directive ensured flexibility in means of achieving the goal, whether it would be legally binding or only on a voluntary basis. The discussion regarding this directive remains open. Nevertheless, several European countries have decided to impose female quotas (The Economist, 2018). In 2008, Norway became the first country to oblige listed companies to secure at least 40% of director seats for women with a threat of possible dissolution. Furthermore, France, Italy, or Belgium have introduced binding quotas with similar sanctions. In other 13 Member States, such as the Netherlands, Spain, Sweden, or Finland, they adopted legislative measures with some restrictions or without sanctions. Still, several European countries have not taken any action (European Union, 2019).

Situation in the Czech Republic

In case of the Czech Republic, the proportion of women in statutory boards or supervisory boards is below the EU average. The Czech government has not implemented any gender quotas, binding or recommended, in relation to the Directive of the European Commission. However, there are some incentives to improve the situation in the labour market. For instance, in 2001, the Government Council for

Gender Equality was established as an advisory body. The council has introduced Government Strategy for Equality of Women and Men in the Czech Republic for 2014 – 2020 (The Office of the Government of CR, 2014) which sets several targets, including increasing the degree of representation of women in decision-making positions in public and private sphere to 40%. In order to achieve this goal, an Action Plan for Equal Representation of Women and Men in Decision-making Positions for 2016 to 2018 was adopted (The Office of the Government of the Czech Republic, 2016). The plan included 30 different tasks from which only 6 were accomplished by the end of 2017.

Moreover, a number of existing studies covering the situation in the Czech Republic is very low. An example may be an analysis provided by Deloitte (Deloitte Global Center for Corporate Governance's, 2019) using data of TOP 100 Czech companies. The analysis reports 6.8% of women on boards of directors and 15.2% on supervisory boards. Another noteworthy study was conducted by McKinsey Company (McKinsey Company, 2012). Although this document also find a high under representation of women on company boards, only a small sample of 23 Czech companies was used so it does not allow to draw any general conclusions of the overall situation in the country. The last one and probably the most important study with respect to this thesis is an analysis conducted by Petr Janský, Kryštof Krotil and Tomáš Pavlas (2015). This analysis was conducted within the project that was designed by Business for Society, Open Society, and the Economics Institute of the Czech Academy of Sciences (Otevřená společnost o.p.s.). They used data of Czech publicly listed companies, i.e. companies where the majority shareholder is state, region or municipality, collected by the public register of justice. The main advantage of this analysis is the completeness of data and also inclusion of state-owned companies. Results show that the average representation of women equaled 17% in case of supervisory boards and only 7% in boards of directors. Apart of that, they also included a survey on companies listed on Prague stock exchange and registered in the Czech Republic, which are in fact the largest Czech companies. In addition, the proposed directive of the European Commission from 2012 would concern primarily these firms. Based on the analysis of 13 such companies, they found out that only 5% of women figure in supervisory boards and 13% in boards of directors. To sum up, there is obviously still a long way to go to achieve the gender equality in decision-making in the Czech Republic.

2. Self-Organizing Maps

This chapter aims at introducing self-organizing maps as a special computational technique which has been used widely for data analysis purposes, in information visualisation, and especially in the visualisation of high-dimensional systems (Kohonen et al., 1996). Self-organizing maps represent a type of a larger group referred to as artificial neural networks. Thus, first, it is essential to explain a theoretical background of neural networks in general which will consequently help to understand a concept of self-organizing maps.

2.1 Theory of Artificial Neural Networks

Artificial neural networks (ANNs) are computational networks based on biological studies of the human brain and nervous system. More specifically, they are inspired by a reduced number of concepts from the biological neural systems as they primarily focus on the electrical signals and behaviour of neurons (Walczak and Cerpa, 2003).

The fundamental unit of the biological neural network is a nerve cell called a neuron. Every neuron consists of a cell body and several nerve fibres. There are two types of these fibres - dendrites which receive signals from other neurons, and a single long fibre called the axon whose function is to transmit information to other nerve cells. The connections between individual neurons are referred to as synaptic junctions, or synapses (Figure 1). Yegnanarayana (2006) specifies that a neuron is estimated to receive on average an input from 10^4 synapses. As the total number of neurons in human brain is approximately 10^{11}, it results in a total of 10^{15} synaptic junctions in the human cortex.

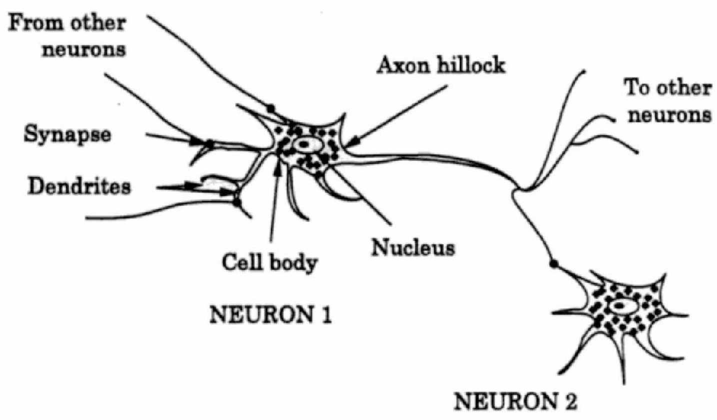

Figure 1: Structure of a Biological Neuron (Yegnanarayana, 2006)

Artificial neural networks are trying to simulate such a process in order to interpret real-world situations. ANNs consist of interconnected units, sometimes also called neurodes or perceptons. These artificial neurons are arranged in three types of layers: input, hidden, and output. The first one is an input layer which represents the data that are supposed to be processed. From the input layer, data go through one or more hidden layers (Figure 2).

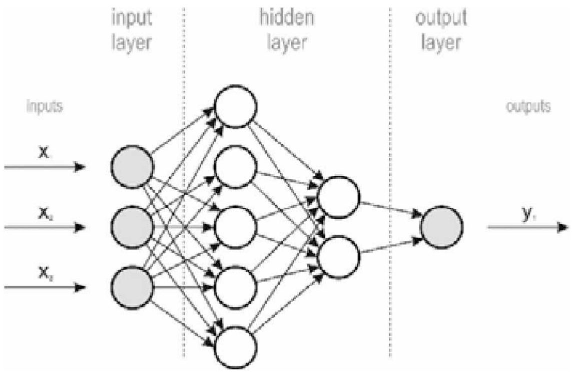

Figure 2: Layers of Artificial Neural Networks (Krenek et al., 2016)

The hidden layers are the most essential of the whole structure, because here the data are transformed into a form that can be used in the output layer afterwards. Each layer may perform different transformations in the input data. Furthermore, an artificial neuron is usually fully connected to neurons in the subsequent layer, which is in fact reproducing the synaptic connections in a human brain. Most importantly, these connections between layers are weighted. The weight reflects the strength of the signal at a connection, meaning that the higher the weight is, the largest impact one unit has on another. Then an aggregated value of the input is calculated, for example by using a summation. There may exist a threshold which refers to a minimum value needed for a signal to pass through. Finally, the aggregated value is transformed by an activation function and the output value for the whole neuron is determined.

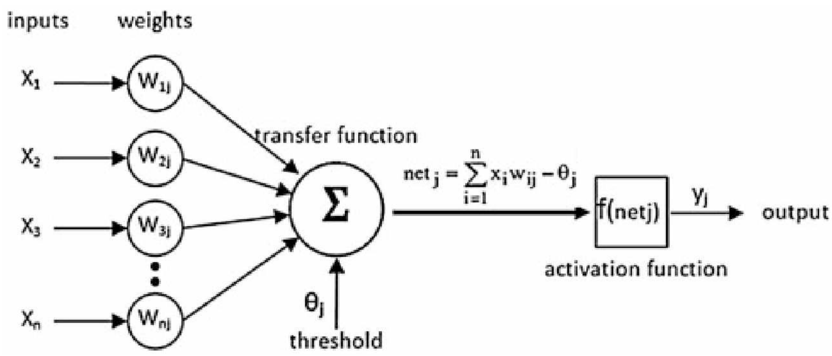

Figure 3: Process of Artificial Neural Networks (Varol, Canakci and Ozsahin, 2013)

Moreover, the artificial neural networks possess the ability to learn by processing training examples. Thus, it can improve the results by adjusting weights, changing links between individual neurons, and consequently reducing the error rate (Schalkoff, 1997). A typical example of such training may be illustrated by an image recognition, where the artificial neural network is supposed to detect an object in a picture. In such case, hidden layers include analyses of picture's features, such as shapes, textures, similar pixels. The process is repeated for a large set of images so that the weights of individual neurons provide accurate results (it resembles to the way people learn from experience) and in the end, the network is able to classify what is in a picture correctly and independently (Basheer and Hajmeer, 2000).

2.2 Kohonen Maps

Self-organizing maps represent a type of artificial neural network that uses an unsupervised learning technique for data mining purposes. In other words, it is keen on finding previously undetected patterns in data sets and consequently grouping similar data points together (Kohonen et al., 1996). The method of self-organizing maps was introduced in the 1980's by Finnish professor Tuevo Kohonen who largely contributed to the research of artificial neural networks and pattern recognition. Therefore, the self-organizing maps are usually referred to as Kohonen maps.

The SOMs algorithm offers a useful method which transforms high-dimensional space of input data onto a lower-dimensional representation scheme, usually a two-dimensional map. Moreover, such a projection preserves the topological relationships of data that results in automatic data clustering. Clustering refers to classifying objects in a data set while the similar objects are stored in proximity and projected into the same unit or into neighboring units. Thus, it can be used to effectively visualize high-dimensional systems (Ong and Sibte Raza Abidi, 1999).

Structure

The structure of the Kohonen maps network differs from those of other neural networks that are described in the previous chapter. In case of SOMs, there are only two layers of neurons. The input layer, representing the source data set, is directly connected to the output layer, while the hidden area is not present at all.

The input layer consists of real n-dimensional vectors $\mathbf{x}$ which form a sequence $\{\mathbf{x}_t\}$, where an integer t signifies the t^{th} step in the repetitive learning process. In case of significant differences among values of an input variable x, it is necessary to normalize the data in order to avoid any dominance of such outliers. That is usually achieved by z-score transformation, i.e. subtracting the mean from the value x and dividing it by its standard deviation:

$$z_j = \frac{x_j - \mu_{x_j}}{\sigma_{x_j}} \tag{1}$$

Furthermore, the output layer, also called a SOM or Kohonen layer, is formed by units called nodes which are arranged in a grid. Typically, the grid has either rectangular or hexagonal shape (Figure 4). The hexagonal topology is usually preferred as it offers a more various neighbourhood compared to the rectangular one. Also, the number of nodes indicates the maximum amount of resulting clusters in the output layer.

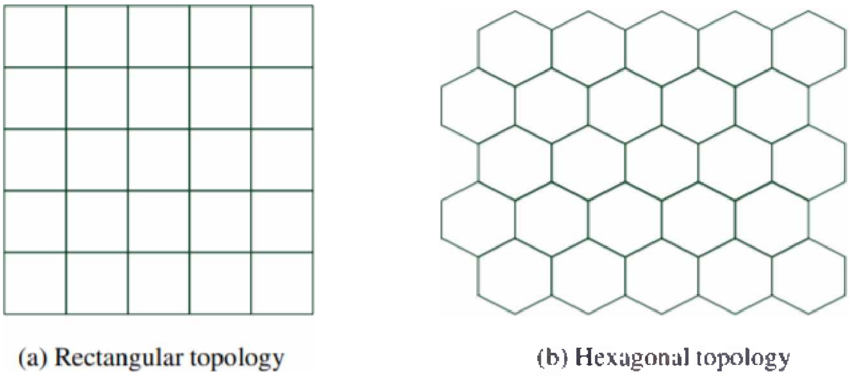

Figure 4: SOM grid topology (Asan and Ercan, 2012)

The input attributes are linked to output nodes by so-called weights. Each node in the SOM array (output layer) is associated with a n-dimensional weight vector $\mathbf{w}_i$, where i refers to the node i. The initial values of weight vectors are typically chosen arbitrarily since they are then updated through the learning process. In his numerous academic papers, Kohonen refers to the weight vectors as models $\mathbf{m}_i$. Then let $\{\mathbf{m}_i(t)\}$ be another sequence of approximation of model $\mathbf{m}_i$ in the iteration step t.

Analysis

After initializing the values of models $\mathbf{m}_i$, the output nodes start to compete between each other to become the best matching unit, BMU. The winner node c is the one whose model $\mathbf{m}$ is the most similar to the input vector x, i.e. the distance between model $\mathbf{m}_c$ and the vector x is minimized:

$$c(t) = argmin_i\{||x(t) - m_i(t)||\} \tag{2}$$

The distance measure is usually calculated using the Euclidan distance.

In the next step, the model of the winning node and its spatial neighbours in the grid are adjusted so that they match better the input. Such a modification is made by using the following formula:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \tag{3}$$

where t is the order number of a current iteration and $h_{ci}(t)$ refers to a neighbourhood function which plays an important role in the whole SOM algorithm. The neighbourhood function defines the rate of change of the neighbourhood around the winning node. There are different types of neighbourhood functions, but the most preferred is the Gaussian function defined as follows:

$$h_{ci}(t) = \alpha(t) exp\left(-\frac{d_{ci}^2}{2\sigma^2(t)}\right) \tag{4}$$

where $\alpha(t)$ is a monotonically decreasing scalar function of t representing a learning rate, d_{ci}^2 stands for the distance between the winning node c and the excited neuron i, and $\sigma(t)$ is another monotonically decreasing function of t referring to the neighbourhood radius at iteration t (Kohonen, 2013). Moreover, the neighbourhood function shrinks with time as the learning rate and the neighbourhood's radius decreases in every iteration step. Finally, the parameters need to be adjusted by setting $t = t + 1$ and so do the neighbourhood size with the learning rate.

This algorithm is repeated for a large number of iterations. In the end, when the the SOM becomes trained, the final network should capture the natural clusters in the input data.

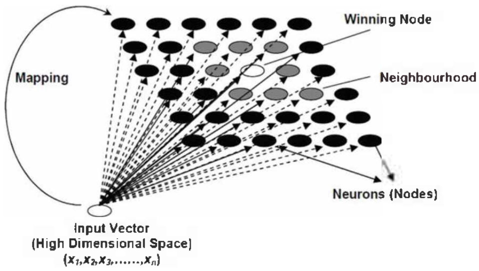

Figure 5: Illustration of Kohonen Self-Organizing Map (Adamala, 2019)

Application of SOMs

The self-organizing map (SOM) principle has become a popular method since its introduction in 1981. Researchers working in diverse fields utilize unsupervised learning method when analysing data. The main sectors include engineering sciences, finance, medicine, or biology (Kohonen, 1998). The practical applications range from process analyses, image and pattern recognition, intelligent systems design, finance applications, retrieval of textual documents, to data mining or biomedical analyses. The most promising areas of application of the Kohonen maps are biomedicine, data mining used for statistical visualisation, and other data analyses in finance or macroeconomics.

During the 8^{th} Workshop on Self-Organizing Maps held in Finland in 2011, Tuevo Kohonen announced that by 2005 approximately 8000 works have been based on the SOM method. Furthermore, this number has increased to 10000 bibliographic items in 2011.

Not so many academic papers dealing with the problematic of gender inequality use neural network methodology as a mean of analysis. However, there are papers using data mining techniques to investigate gender-related data across a specific region and consequently visualize the results in clusters (Rahman and Sultana, 2013). Clustering and projection seem to be a useful tool when classifying labour markets differences across regions. For instance, Mirzabalaeva, Kvachev and Kuksova (2018) used Kohonen maps to group the Russian regions according to the level of labour market depressiveness. In addition, Gaubert and Cottrell (1999) analyse the unemployed on French labour market using Kohonen maps as well. Finally, Pö, Honkela and Kohonen (2009) collected a list of more than 5000 scientific papers that contain the SOM algorithm in their analyses.

3. Application of SOMs on the Data of Czech Public Firms

3.1 Dataset

This bachelor thesis explores the level of gender inequality across regions of the Czech Republic using an analysis of possible determinants that may have influence on gender equality disparities. Most of the dataset includes publicly available data of the Czech Statistical Office, while the statistics related to the gender index was used from the website Genderová mapa (Otevřená společnost o.p.s.).

Officially, the Czech Republic consists of 76 regions. However, regarding this thesis, a slight modification was made and another sample accounting for the capital city Prague was also included in the data set. As a result, the analysis could cover the whole are of the Czech Republic. Other districts remained unchanged. Therefore, there are 77 districts included in the data set. The complete list of all regions and their territorial location in the Czech Republic is also added to the Appendix of this thesis.

Variables

In this section of the thesis, the variables are introduced. They represent important factors that may have influence on the level of gender inequality in regions of the Czech Republic. These factors include various demographic and socio-economic characteristics defining population in all districts. It is also important to mention that gathering data on the regional level in case of the Czech Republic is often a demanding task. Furthermore, selecting appropriate factors correlated with the level of gender inequality is essential. However, in such a small country as the Czech Republic, the available data may not exhibit such distinctive differences among regions as it would be in case of larger countries.

Gender Index This variable depicts the representation of women on boards of Czech public firms included in the analysis available on the Genderova mapa website (Otevřená společnost o.p.s.). The value varies between 0 and 1, and it is calculated as a ratio of women on boards to the total number of the board positions in each region.

Number of board positions To reflect the amount of board positions, this variable was added to the model as it may show the amount of possible opportunities for women to achieve a deciosion-making position. It shows the number of board positions of public firms used in the analysis of Genderova mapa (Otevřená společnost o.p.s.).

Unemployment rate The unemployment rate was calculated as the share of unemployed people aged 15 to 64 years from the total population of the same age category. The number of job seekers are reported by the Ministry of Labour and Social Affairs.

Average age The average age of the population is publicly available on the website of the Czech Statistical Office.

Average monthly pension In order to reflect the wealth of the population in a region, this variable was included. Moreover, data for the level of income on the regional level were not available in case of the Czech Republic. Therefore, average pension was used instead.

Life expectancy As discussed in previous chapters, health accounts for an important determinant of gender inequality is health, measured in years. Thus, the variable of life expectancy was used to account for regional differences in the quality of healthcare institutions or other indirect characteristics that may have influence on population's health.

Population growth This variable is represented as a percentage increase in the number of population in a region. The reason of adding this variable to the model is the assumption that regions with higher population growth are attractive for living as people decide to settle there. Thus, such regions are believed to have a higher level of equality as well.

Population density per km^2 Population density was calculated as the ratio of region's total population and its area measured in km^2. The idea behind it is similar to the one presented in case of population growth. Therefore, a positive correlation between the population density and the level of equality is expected.

Share of population having a university degree Another variable which was presented in previous chapters as one of the determinants of gender inequality is education. Thus, education was included into the model as a variable reflecting a share of inhabitants who obtained a university degree. It is generally believed that more educated population would have a positive effect on gender equality in a region.

Mother's age when having her first child Furthermore, two variables which aim at describing the degree of women's emancipation were included. Ii is expected that women focusing on their careers would also have their first child later in life.

Marriage rate The second variable linked to the career perspectives of women was expressed as a marriage rate. The reasoning behind it is similar to the previous variable. The marriage rate was calculated as a ratio of number of marriages and the population older than 15 years old in a district.

Income inequality Finally, a proxy variable for income inequality was used since it could be linked to the gender inequality (Mikulikova, 2020). It was computed using the number of housing allowances paid to households in a region. Since the amount of benefit depends on the size of a household and it is paid annually, the following formula was used for every region i:

$$incomeinequality = \frac{housingallowances_i \times householdsize}{population_i \times 12} \tag{5}$$

where i refers to the specific region, housing allowances describe the total number of paid housing allowances, and the household size was set to the average number of household's members in the Czech Republic which was equal to 2.34 in 2019. Finally, in denominator the total population of a region was multiplied by 12 in order to obtain a monthly value of income inequality.

The following Figures 6 and 7 show the descriptive statistics for all explanatory variables included in the analysis. However, these values do not tell much about the importance of those variables or relationships among them.

	Gender_index	Number_board_positions	Unemployment	Average_age_M	Average_age_W	Average_pension	Life_exp_M
Minimum	0.000000	5.00000	1.165733	37.800000	39.70000	11200.0000	73.142276
1. Quartile	0.110000	18.00000	2.352289	40.900000	43.50000	11642.0000	75.158611
Median	0.150000	31.00000	2.892648	41.300000	44.00000	11803.0000	75.914997
3. Quartile	0.210000	52.00000	3.450908	41.800000	44.70000	12000.0000	76.566704
Maximum	0.600000	416.00000	6.815066	42.600000	45.50000	12934.0000	78.064986
Stdev	0.100422	50.47669	1.109910	0.798247	1.02034	292.1179	1.096492

Figure 6: Descriptive statistics of the variables - part 1

	Life_exp_W	Population_growth	Population_density	Share_Uni_degree	Age_First_child	Marriage_rate	Income_inequality
Minimum	78.938448	-0.009444	37.02805	0.054070	26.800000	0.003814	0.011755
1. Quartile	81.289370	-0.000308	75.15267	0.083244	28.200000	0.004101	0.028098
Median	81.804840	0.001932	115.91131	0.089606	28.700000	0.004259	0.035230
3. Quartile	82.257078	0.006077	155.90518	0.105537	29.200000	0.004414	0.050031
Maximum	83.207009	0.023394	2668.78338	0.247172	31.300000	0.004791	0.141535
Stdev	0.961533	0.005787	363.98173	0.033289	0.827329	0.000220	0.024787

Figure 7: Descriptive statistics of the variables - part 2

In the Figure 8, the covariance matrix is represented. The more distinctive colour, the stronger relationship between each pair of variables. More specifically, the blue colour represent a positive correlation and red the negative one, as depicted on the scale on the right. Nevertheless, based on this matrix, no strong relationships are detected between the gender index and other explanatory variables.

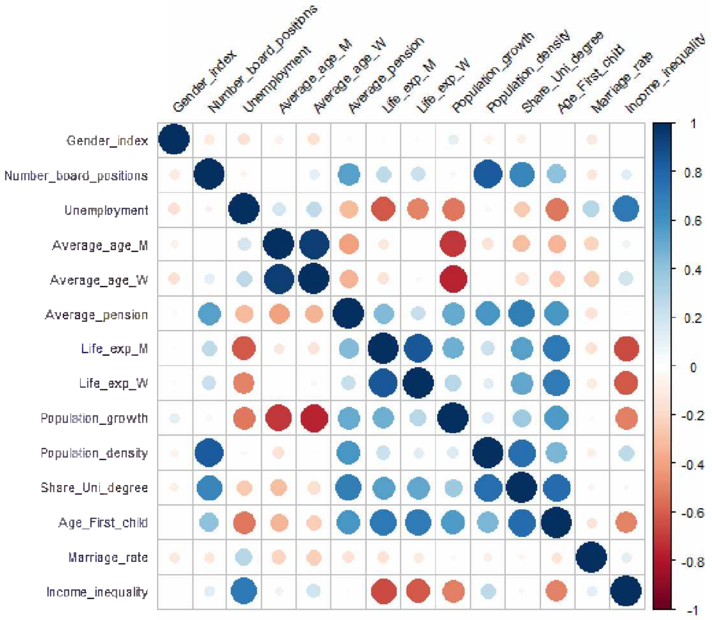

Figure 8: Covariance matrix

3.2 Parameters of the Model

As already mentioned in the previous chapter, a self-organizing map is a very useful tool which is able to produce a two-dimensional representation of multi-dimensional input data set. Furthermore, the algorithm uses an unsupervised learning approach as no prior knowledge of data is assumed.

However, the quality of learning of Kohonen maps highly depends on initialization settings, such as the value of initial weights, the neighbourhood function, the learning rate, or the number of iterations. The best initialization method is influenced by characteristics of the specific dataset. After that, the process of training and the data clustering follow on.

The most popular programming languages, including Python, R, or Matlab, already contain packages for data analysis using SOMs. For instance, the SOM Toolbox in Matlab offers SOMs training with different topologies or learning rates, and also visualization in various ways (Vesanto et al., 1999). In this thesis, the R package **kohonen** was used since it provides a simple-to-use functions with an increased ephasis on visualisation (Wehrens and Buydens, 2007).

Initialization of SOMs

There are several ways how to the weight vectors can be initialized. One of them is random initialization where the weights are assigned with small random values between the minimum and maximum values of given variables. Other methods include sample initialization where the initial weights values are random samples drawn from the input dataset, or the linear initialization which utilizes two principal eigenvectors of the input dataset. In this case, the random initialization was applied. As neural networks are stochastic in nature, we may ensure the reproducibility in the R code by including **set.seed** function generating random numbers.

Furthermore, the size of the grid, referring to the number of neurons, needs to be determined at the beginning as well. A higher number of neurons makes the mapping better, but at the same time, the computational complexity in case of a large grid may become a burden in the training phase. Therefore, numerous empirical papers suggest a heuristic equation defining a recommended number of neurons (Vesanto, 1999; Spanakis and Weiss, 2016):

$$m = 5\sqrt{n} \tag{6}$$

where **m** refers to the number of map units and **n** is the number of samples in a data set. In addition, another measure helps to determine an appropriate grid size, more specifically the ratio of grid side lengths. This ratio should correspond to the ratio between two largest eigenvalues of the covariance matrix of training data (Vesanto, 1999).

Regarding the shape of the map grid, usually a hexagonal lattice is recommended as it offers more advantages than only a better visualization effect. The distance between the center of each and every adjacent cell is the same in case of hexagons. Moreover, two neighbouring cells always share whole edges, not only corners as it happens for rectangular grids, which reduces sampling bias from the edge effects.

Another parameter that is supposed to be defined is the neighbourhood function which determines the correlation between neurons. There are two fundamental types of neighbourhood function - bubble and the gaussian. The bubble function is quite simple since it is constant over the neighbourhood and zero otherwise. On the other hand, the gaussian function which gradually decreases with rising distance from the winning neuron.

Finally, a few indicators are often used to evaluate the quality of a map which also reflects suitability of chosen parameters for the given data set. The learning quality indicator is measured by quantization error (QE), whereas topographical error (TE) estimates the projection quality. More specifically, QE is the average distance between each model vector and its BMU and TE indicates the number of all data vectors for which the first and the second BMUs are not adjacent. The smallest values of these two indicators are, the better is the quality of the map (Anh Tu, 2020). Another useful measure of goodness of a SOM may be the number of empty cells (the lower the number of empty cells, the better), or the distance of each codebook from the others. Many researchers focused in their work on improving the SOM algorithm in order to increase the quality of resulting maps (Kamimura, 2014; Lee and Verleysen, 2002).

Having a dataset which includes data of 77 regions, the approximate number of neurons equals $5\sqrt{77} = 43.87$. Moreover, calculating the ratio of eigenvalues of the covariance matrix, the recommended size of a grid is 8x5. Furthermore, the bubble neighbourhood function was used since it yielded better quality parameters of the model. The quality plot function was added in order to show the mean distance of objects mapped to a neuron to the codebook vector of that neuron. The smaller the distance, the better the representation by the weight vectors. Thus, this quality plot seems to offer a quite good representation. Moreover, the grey node represents a node where no region is located.

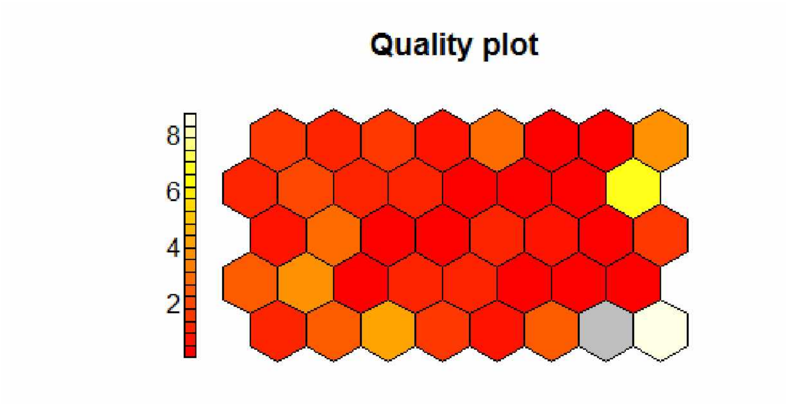

Figure 9: Quality plot

SOMs Training

During the learning phase, the SOM gets trained. It is important to standardize all data first which is achieved using the function **scale** in R. As a consequence, extremely high or low values will not negatively influence the training process.

Two parameters need to be determined with regards to the learning of SOMs - number of iterations and learning rate. As the training of SOM progresses with every iteration step, the mean distance of a node's weights to the samples represented by that node decreases. This can be represented by a plot showing the progress over time. Ideally, at the end of the training, the minimum plateau is reached where no more iterations are required. In this case, based on the below plot, the amount of 200 of iterations was applied.

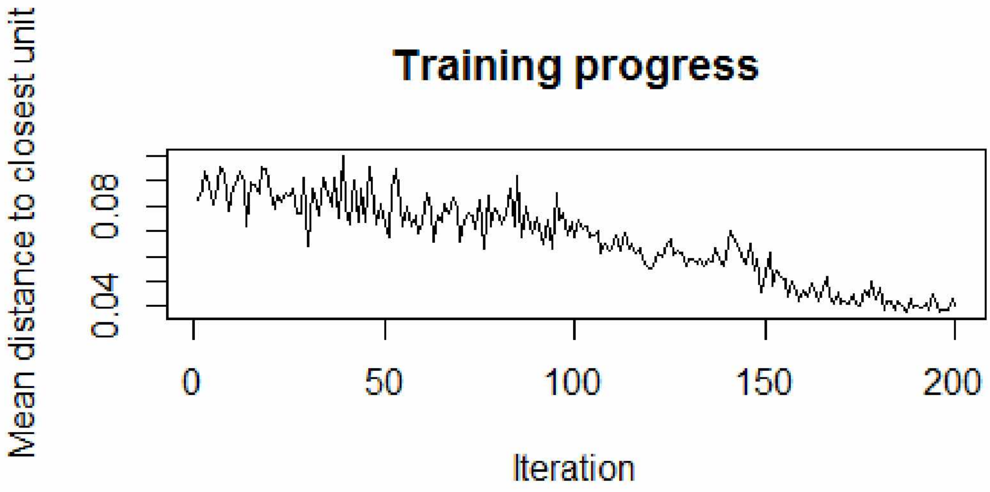

Figure 10: Training progress

Next, the learning rate, referring to the size of adjustments during the training, is to be determined. The initial level of learning rate is set higher and it decreases with time. Moreover, two learning-rate forms are mostly used - a linear function and an inverse-of-time function. The rate is chosen based on the quality measures which should be minimized.

Clustering

In the final step, once the map is trained, the visualization into groupings with similar metrics takes place. First, the optimal number of clusters has to be determined. In this case, the *elbow* method was implemented. This algorithm defines the clusters such that the total within-cluster sum of square is minimized. The sum measures the squared average distance between all points included in a cluster and the cluster centroid. Therefore, the smallest sum possible is preferred. Next, the elbow method looks at a plot of the total within-cluster sum of squares as a function of the number of clusters. The point of a bend in the plot is then considered as an optimal number of clusters that should be used. In this case, the total number of 5 clusters was set and used in the analysis.

Furthermore, having the number of clusters specified, the k-means algorithm partions the data set into the clusters, such that the objects within each cluster are as similar as possible. The algorithm begins with a random selection of k objects from the data set as the initial centroids for the clusters. After that, the remaining objects are assigned to its closest centroid. The whole process is repeated until the cluster assignment stops changing.

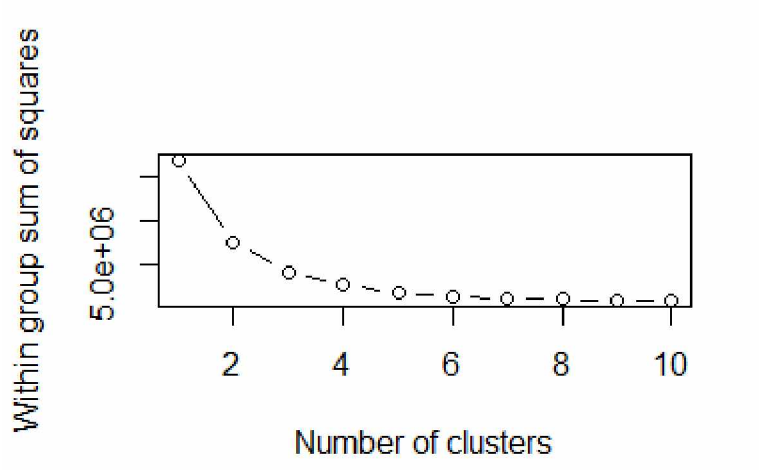

Figure 11: Elbow method

3.3 Results

The Figure 12 describes the location of each region of the Czech Republic in the grid of the Kohonen map. Moreover, the thick line depicts the boundaries between the clusters. From the very definition of self-organizing maps, regions sharing similar characteristics are located close to each other. As a result, the grid of the Kohonen map allows to simply visualize resemblances between Czech regions with respect to the variables used in the analysis.

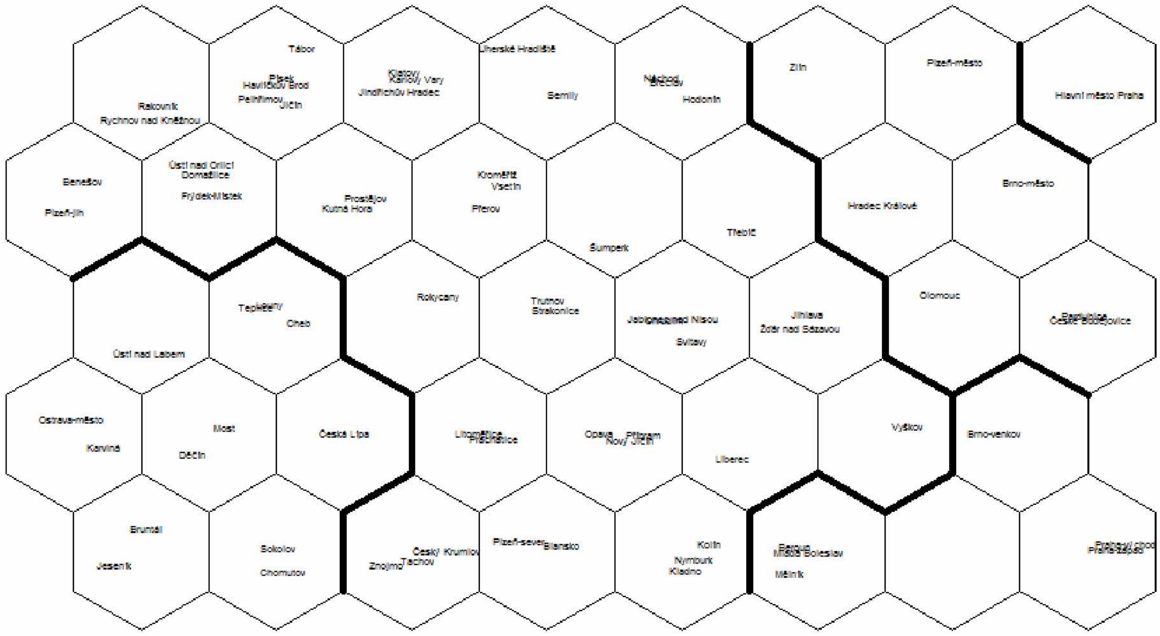

Figure 12: Location of regions in the clusters

In addition, there are other plots visualizing the input data. It is also important to mention that the location of every region remains the same for every plot. For instance, Figure 13 illustrates a structure of single codebook vectors for every artificial neuron in the grid. Such a visualisation of the influence of all explanatory variables is very practical and easy to interpret. The larger the segment referring to a specific variable, the more important that variable is in case of that neuron. However, this plot serves rather as a general overview, than an accurate measure since the segments may become difficult to recognize in case of large number of variables.

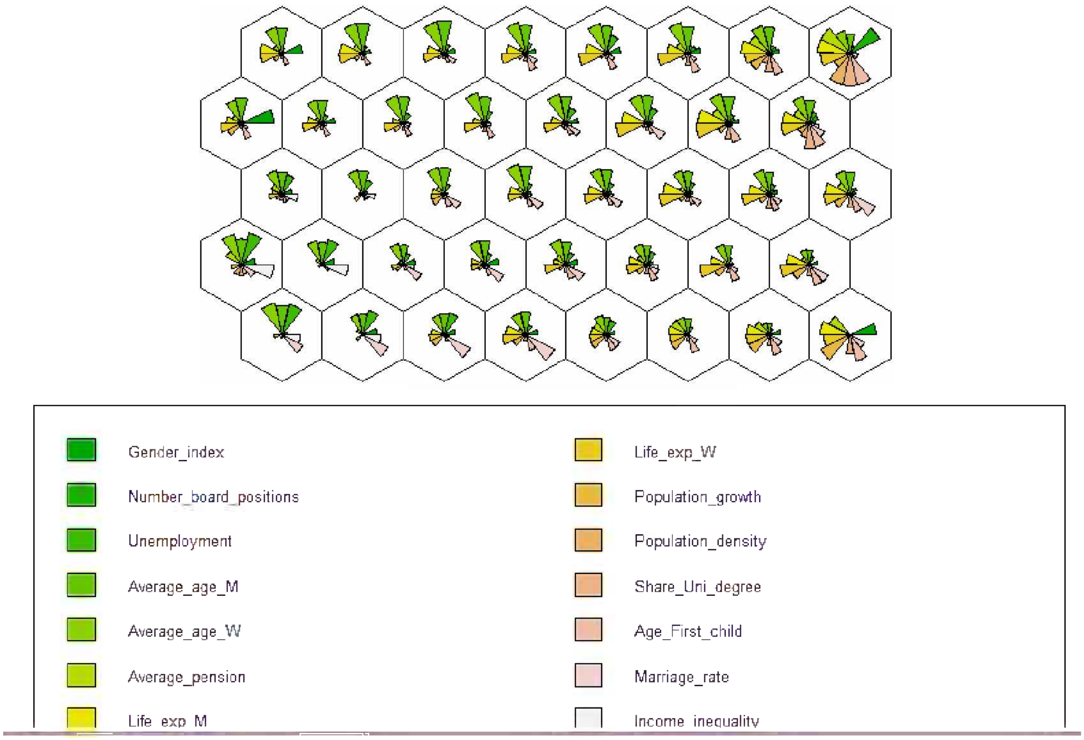

Figure 13: Codebook vectors

Moreover, Figure 14 offers another interactive hierarchical visualization of the input data. The so-called heat maps, also called feature maps, provide another useful tool which normalize the input data and then transform those data to colour scale. As a consequence, each value is assigned a colour and it is possible to easily distinguish between low and high values of a given variable across regions. The Figure 14 shows a very similar pattern for the variables number of board positions, share of university degree and population density, which draws the same conclusion as the covariance matrix in the previous section. Another similar pattern in heat maps is noticeable in case of variables for average pension and population growth. This similarity seems logical since these two factors are to some extent positively correlated.

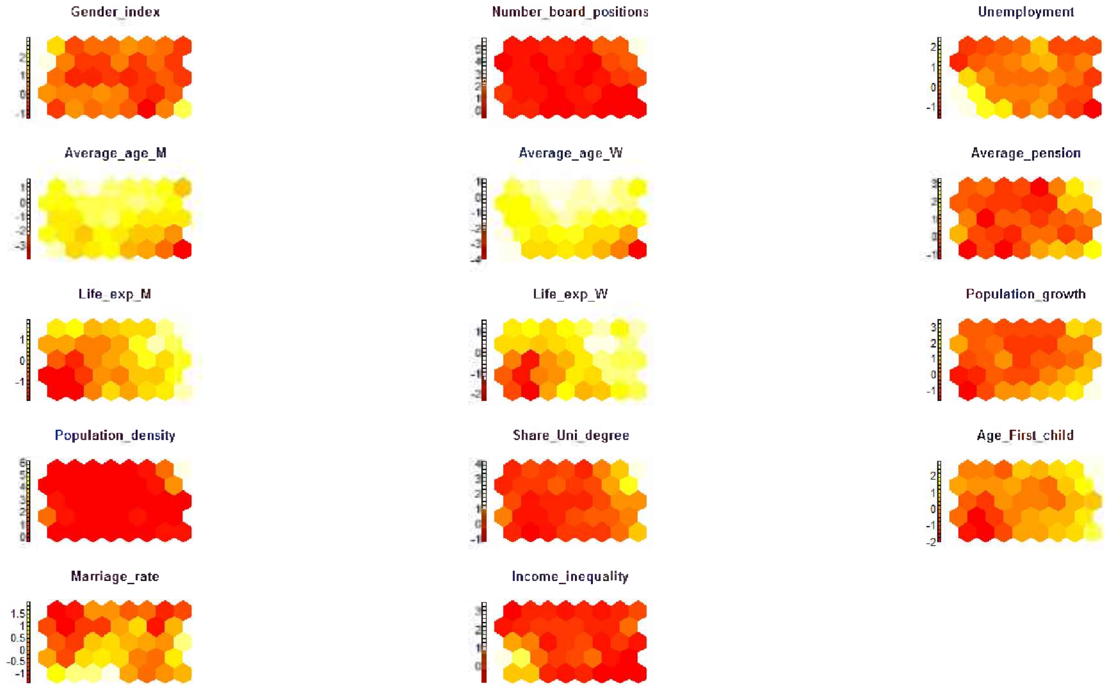

Figure 14: Heatmaps of the variables

Next, U-matrix is depicted in Figure 15. Unified distance matrix transforms the self-organizing map into a greyscale image which reflects the distance between the codebook vectors of neurons. In this case, the darker shade of grey defines a higher level of similarity between neurons. As a result, the most similar regions are those located in the middle part of the grid.

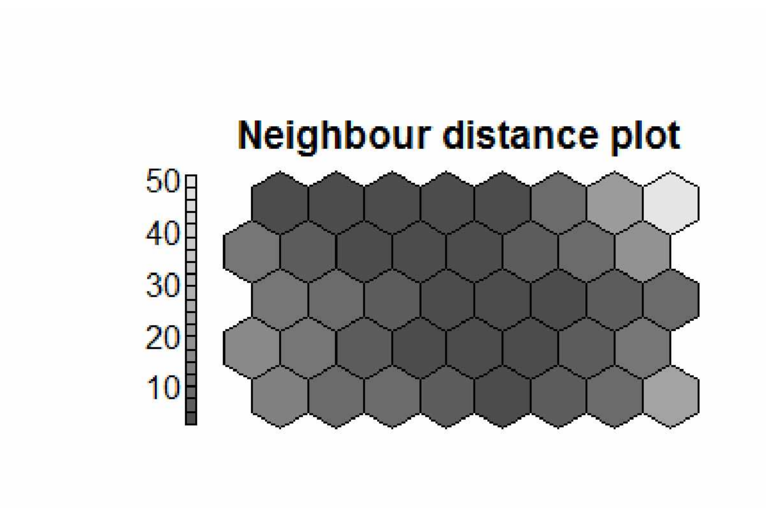

Figure 15: U-matrix

The following plot (Figure 16) defines the final clusters. Each cluster is coloured differently in order to easily distinguish the clusters, the thick line is dividing the clusters.

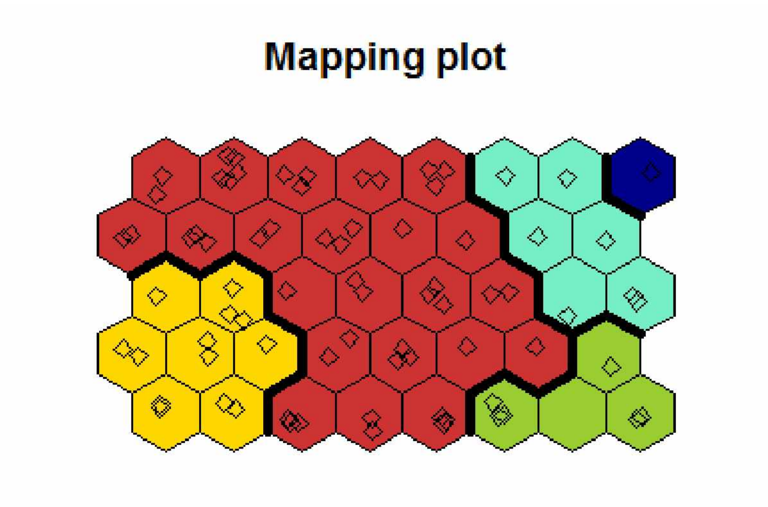

Figure 16: Visualization of clusters

Cluster 1

This cluster includes a single district - the capital city Prague. As expected, this region is defined by a high gender index referring to the higher representation of women on board positions. Furthermore, the capital city exhibits other positive characteristics, such as a higher share of tertiary educated inhabitants, lower income inequality and increased average pension. These results are not surprising as Prague is a modern city representing the business center of the Czech Republic which offers a large scale of work opportunities without gender disparities.

Cluster 2

This cluster includes the following regions: Zlín, Plzeň - město, Brno - město, Hradec Králové, Olomouc, Pardubice, and České Budějovice. These regions are represented by large cities of the Czech Republic that are characterized by low unemployment rate, but also by a low gender index which may be influenced by a lower number of board positions for these regions in the dataset. On the other hand, the income inequality is also quite low.

Cluster 3

Regions included in this cluster are: Brno - venkov, Praha - východ, Praha - západ, Beroun, Mladá Boleslav, and Mělník. The common feature of these districts is their proximity to the two largest Czech cities - Praha and Brno. Therefore, the influence of these large cities is obvious. Moreover, many people commute to work to Prague from these surrounding regions. Also Škoda Mladá Boleslav has a great inlfuence on the labour market in that district while it lowers the unemployment rate and increases the relative wealth.

Cluster 4

This cluster includes Ústí nad Labem, Cheb, Teplice, Louny, Ostrava - město, Karviná, Most, Děčín, Česká Lípa, Bruntál, Jeseník, Sokolov, and Chomutov. All these regions are located in the border areas with a tradition of mining and heavy industry which is a domain of men. As a result, the gender index is also quite low as women are underrepresented in the decision-making positions. Finally, this cluster yields a high level of income inequality and low levels of life expectancy as well.

Cluster 5

The last cluster is also the largest. It includes the rest of Czech regions that are not yet included in the previously mentioned clusters. These regions are characterized by a medium standard of living. The representation of women expressed by the gender index varies across the cluster, probably according to the type of industry represented in each region. Furthermore, the income inequality is quite low in these regions.

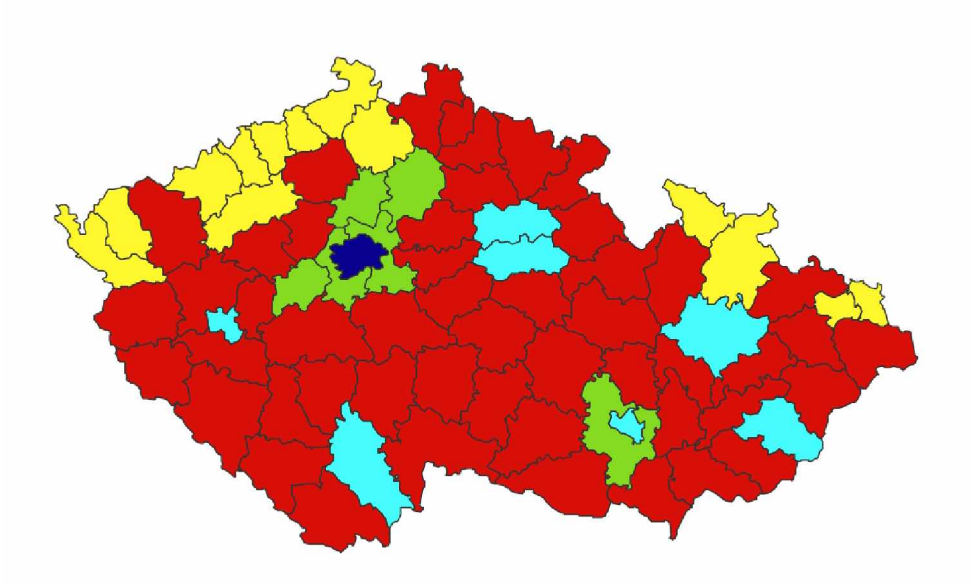

Figure 17: Map of the Czech Republic divided into clusters

To sum up, the clusters are depicted on the map of the Czech Republic (Figure 17) which visualizes the patterns across the country. The clusters are coloured as follows: Cluster 1 - dark blue, Cluster 2 - light blue, Cluster 3 - green, Cluster 4 - yellow, Cluster 5 - red.

Conclusion

Gender inequality represents a widely discussed topic in today's society. Women are still discriminated in various domains of everyday life, one of them being the labour market. It is therefore very important to investigate and define possible factors that may have influence on the level of gender inequality. In order to explore the situation concerning gender disparities in the Czech Republic, an innovative method of self-organizing maps was applied.

The possible factors of gender inequality were represented in 14 explanatory variables included in the data set of 77 regions of the Czech Republic. However, it is important to mention that the algorithm of self-organizing maps is quite sensitive to the choice of variables and also to the parameters used in the algorithm, such as size and shape of the grid, the learning rate, or the resulting number of clusters of the map. Changing even one of those parameters may result in very different outcomes. On the other hand, the visualization of results is very interesting as it offers diverse plots of results.

Moreover, no such analysis was conducted regarding the problematic of gender inequality in decision-making in the Czech Republic. Also, there are no defined or fixed determinants of gender gap on board positions. Thus, the choice of variables, that should be used in the analysis, might be to some extent subjective, or based on the trial and error method in order to estimate the most appropriate ones.

On the other hand, the method of self-organizing maps provides a very useful tool with a great ability of transforming multidimensional input data into a two-dimensional map. In addition, the clustering method simply visualizes the patterns in gender inequality across the regions in the Czech Republic. As a result, the total number of 77 regions was divided into 5 separate clusters while every cluster has its own specifics. Cluster 1 is represented by the capital city Prague has a the highest ratio of women represented on the boards. Cluster 2 does not yield such a high gender index, but the income inequality and unemployment are low. Furthermore, the cluster 3 comprises regions located close to the two largest cities and are highly influenced by them. On the contrary, in the cluster 4 is the gender inequality higher. This cluster includes regions with low life expectancy and higher level of unemployment as well. Finally, the largest cluster includes most of the Czech regions.

To sum up, policy makers could find the results of the analysis useful when looking for optimal policies which would potentially reduce the gender inequality in the Czech Republic.

Bibliography

[1] World Economic Forum. 2020. Global Gender Gap Report 2020. [online] Available at: <https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality> [Accessed 18 October 2020].

[2] Stoller, R.J., 1974. Sex and gender: the development of masculinity and femininity, London: Karnac Books.

[3] Oakley, A., Polackova Milena amp; Polacek Martin, 2000. Pohlavi, gender a spolecnost, Praha: Portal.

[4] Greenglass, E.R. (2001). Gender Role Stress and Health. International Encyclopedia of the Social Behavioral Sciences, pp.6027–6029.

[5] Eagly, A.H. and Wood, W. (2016). Social Role Theory of Sex Differences. The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies, pp.1–3.

[6] Cho, S.-J., Hudley, C., Lee, S., Barry, L. and Kelly, M. (2008). Roles of gender, race, and SES in the college choice process among first-generation and nonfirst-generation students. Journal of Diversity in Higher Education, 1(2), pp.95–107.

[7] Kazmierczak, M. (2010). The feminine and masculine gender role stress — conclusions from Polish studies. Polish Psychological Bulletin, 41(1), pp.20–30.

[8] Higgins, E.T. (1989). Self-Discrepancy Theory: What Patterns of Self-Beliefs Cause People to Suffer? Advances in Experimental Social Psychology, pp.93–136.

[9] M Pilar Sanchez-Lopez and Liminana-Gras, R.M. (2017). The psychology of gender and health: conceptual and applied global concerns. London; San Diego: Academic Press.

[10] International Labour Office (2000). ABC of women workers' rights and gender equality. Geneva: International Labour Office.

[11] Mikkola, A Miles, CA 2007, Development and gender equality: consequences, causes, challenges and cures. Discussion papers / Helsinki Center of Economic Research, no. No 159, Helsinki Center of Economic Research, Helsinki. <http://ethesis.helsinki.fi/julkaisut/eri/hecer/disc/159>

[12] Khan, N. (2018). Critical Review of Past Literature on Gender Inequality in the World. Research on Humanities and Social Sciences, [online] 8(13), p.11. Available at: https://www.iiste.org/Journals/index.php/RHSS/article/view/43559 [Accessed 14 Nov. 2020].

[13] Elborgh-Woytek, K., Newiak, M., Kochhar, K., Fabrizio, S., Kpodar, K., Wingender, P., Clements, B. and Schwartz, G. (2013). Women, Work, and the Economy: Macroeconomic Gains from Gender Equity. Staff Discussion Notes, [online] 13(10), p.1. Available at: https://www.imf.org/external/pubs/ft/sdn/2013/sdn1310.pdf [Accessed 21 Feb. 2019].

[14] Barro, R.J. and Lee, J.-W. (1994). Sources of economic growth. Carnegie-Rochester Conference Series on Public Policy, 40, pp.1–46.

[15] United Nations (2015). United Nations: Gender equality and women's empowerment. [online] United Nations Sustainable Development. Available at: https://www.un.org/sustainabledevelopment/gender-equality/.

[16] Dollar, D., Gatti, R., World Bank. Development Research Group and World Bank. Poverty Reduction And Economic Management (1999). Gender inequality, income, and growth: are good times good for women? Washington, D.C.: World Bank, Development Research Group/ Poverty Reduction And Economic Management Network.

[17] Knowles, S., Owen, D. and Lorgelly, P. (2002). Are educational gender gaps a brake on economic development? Some cross-country empirical evidence. Oxford Economic Papers, 54(1), pp.118–149.

[18] Miguel, E. and Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. Econometrica, 72(1), pp.159–217.

[19] Bandiera, O. and Natraj, A. (2013). Does Gender Inequality Hinder Development and Economic Growth? Evidence and Policy Implications.

The World Bank Research Observer, [online] 28(1), pp.2–21. Available at: https://www.jstor.org/stable/24582370?seq=1 [Accessed 28 Nov. 2020].

[20] Oecd.org. (2014). employment - OECD. [online] Available at: https://www.oecd.org/gender/data/employment/ [Accessed 29 Nov. 2020].

[21] Blackburn, R.M., Brooks, B. and Jarman, J. (2001). The Vertical Dimension of Occupational Segregation. Work, Employment and Society, 15(3), pp.511–538.

[22] Blackburn, R.M., Browne, J., Brooks, B. and Jarman, J. (2002). Explaining gender segregation. British Journal of Sociology, [online] 53(4), pp.513–536. Available at: https://pdfs.semanticscholar.org/68f2/55c81a2cfa09917e20f0551227ed38e072bf.pdf [Accessed 29 Nov. 2020].

[23] Blackburn, R.M. (2012). The Measurement of Occupational Segregation and Its Component Dimensions. International Journal of Social Research Methodology, 15(3), pp.175–198.

[24] Bettio, F., Verashchagina, A., Comissio Europea. Direccio General D'ocupacio, Assumptes Socials I Igualtat D'oportunitats and Comissio Europea. Expert Group On Gender And Employment (2009). Gender segregation in the labour market: root causes, implications and policy responses in the EU. Luxembourg: Publications Office Of The European Union.

[25] Valentová, M., Šmídová, I. and Katrňák, T. (2007). Gender Segregation in the Labour Market Placed in the Context of Educational Segregation: Cross-National Comparison. www.muni.cz, [online] 39(3). Available at: https://www.muni.cz/en/research/publications/719624 [Accessed 5 Dec. 2020].

[26] European Commission - European Commission. (2020). The gender pay gap situation in the EU. [online] Available at: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-euendocuments.

[27] LeanIn.org and McKinsey Company (2016). Women in the Workplace. [online] Available at: http://wit.berkeley.edu/wp-content/uploads/2017/08/Women-in-the-Workplace-2016-1.pdf [Accessed 6 Dec. 2020].

[28] Reskin, B. (1993). Sex Segregation in the Workplace. Annual Review of Sociology, 19(1), pp.241–270.

[29] Coré, F. (1999). The continuing saga of labour market segregation - OECD Observer. [online] oecdobserver.org. Available at: https://oecdobserver.org/news/archivestory.php/aid/88/The_continuing_saga_of_labor_market_segregation.html [Accessed 7 Dec. 2020].

[30] OECD (2012). Lack of support for motherhood hurting women's career prospects, despite gains in education and employment, says OECD. [online] www.oecd.org. Available at: http://www.oecd.org/newsroom/lackofsupportformotherhoodhurtingwomenscareer-prospectsdespitegainsineducationandemploymentsaysoecd.htm.

[31] Laura D'Andrea Tyson and Parker, C. (2019). An Economist Explains Why Women Are Paid Less. [online] World Economic Forum. Available at: https://www.weforum.org/agenda/2019/03/an-economist-explains-why-women-get-paid-less/ [Accessed 7 Dec. 2020].

[32] Smethers, S. (2020). Six Stubborn Myths about Equal Pay between Men and Women. [online] www.ft.com. Available at: https://www.ft.com/content/32791f78-4775-11ea-aee2-9ddbdc86190d [Accessed 7 Dec. 2020].

[33] Kulik, C.T. and Rae, B. (2019). The Glass Ceiling in Organizations. Oxford Research Encyclopedia of Business and Management. [online] Available at: https://oxfordre.com/business/view/10.1093/acrefore/9780190224851.001.0001/acrefore-9780190224851-e-41 [Accessed 7 Dec. 2020].

[34] EIGE (2020). Browse Gender Statistics | Gender Statistics Database. [online] European Institute for Gender Equality. Available at: https://eige.europa.eu/gender-statistics/dgs/indicator/wmidm_bus_bus_wmid_comp_compex/datatable.

[35] Terjesen, S., Aguilera, R.V. and Lorenz, R. (2014). Legislating a Woman's Seat on the Board: Institutional Factors Driving Gender Quotas for Boards of Directors. Journal of Business Ethics, 128(2), pp.233–251.

[36] Carter, D.A., Simkins, B.J. and Simpson, W.G. (2002). Corporate Governance, Board Diversity, and Firm Performance. SSRN Electronic Journal. [online] Available at:

https://www.researchgate.net/publication/228292359_Corporate_Governance_Board_Diversity_and_Firm_Performance [Accessed 8 Dec. 2020].

[37] Wright, M., Siegel, D.S., Keasey, K. and Igor Filatotchev (2013). The Oxford Handbook of Corporate Governance. [online] Oxford: Oxford University Press. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=972763 [Accessed 8 Dec. 2020].

[38] Joy, L., Wagner, H.M. and Narayanan, S. (2019). Report: The Bottom Line: Corporate Performance and Women's Representation on Boards | Catalyst. [online] Catalyst. Available at: https://www.catalyst.org/research/the-bottom-line-corporate-performance-and-womens-representation-on-boards/ [Accessed 8 Dec. 2020].

[39] Rose, C. (2007). Does Female Board Representation Influence Firm performance? the Danish Evidence. Corporate Governance: An International Review, 15(2), pp.404–413.

[40] Daunfeldt, S.-O. and Rudholm, N. (2012). Does Gender Diversity in the Boardroom Improve Firm Performance? [online] CiteSeer. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.935.9933rep=rep1type=pdf [Accessed 8 Dec. 2020].

[41] Mínguez-Vera, A. and Martin, A. (2011). Gender and management on Spanish SMEs: an empirical analysis. The International Journal of Human Resource Management, 22(14), pp.2852–2873.

[42] European Commission (2019). Gender equality strategy. [online] European Commission. Available at: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en.

[43] European Commission (2019a). Gender balance in decision-making positions. [online] European Commission. Available at: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/gender-balance-decision-making-positions_en [Accessed 9 Dec. 2020].

[44] European Union (2019). 2019 Report on Equality between Women and Men in the EU. [online] Luxembourg: Publications Office of the European Union. Available at:

https://ec.europa.eu/info/sites/info/files/aid_development_cooperation _fundamental_rights/annual_report$_{g}e$_2019$_{e}n$_1.$pdf[Accessed9Dec.$2020].

[45] The Economist (2018). Ten years on from Norway's quota for women on corporate boards. [online] The Economist. Available at: https://www.economist.com/business/2018/02/17/ten-years-on-from-norways-quota-for-women-on-corporate-boards [Accessed 10 Dec. 2020].

[46] The Office of the Government of CR (2014). The Office of the Government of the Czech Republic Department of Human Rights and Minorities Government Strategy for Equality of Women and Men in the Czech Republic for. [online] Available at: https://www.vlada.cz/assets/ppov/rovne-prilezitosti-zen-a-muzu/dokumenty/Government_Strategy_for-Gender_Equality_2014_2020.pdf [Accessed 11 Dec. 2020].

[47] The Office of the Government of the Czech Republic (2016). Úřad vlády České republiky Action Plan for Equal Representation of Women and Men in Decision-making Positions for 2016 to 2018. [online] Available at: https://www.vlada.cz/assets/ppov/rovne-prilezitosti-zen-a-muzu/dokumenty/Action_plan_for_Equal_Representation_Decision_Making_Positions.pdf [Accessed 11 Dec. 2020].

[48] McKinsey Company (2012). Unlocking the Full Potential of Women in the Czech Republic. [online] www.mckinsey.com. Available at: https://www.mckinsey.com/featured-insights/europe/unlocking-the-full-potential-of-women-in-the-czech-republic [Accessed 11 Dec. 2020].

[49] Petr Jansky, Krystof Krotil and Tomas Pavlas (2015). Bereme zeny na palubu: analyza zastoupeni zen a muzu ve vedeni verejnych firem. [online] Praha: Otevrena Spolecnost O.P.S. Available at: https://www.otevrenaspolecnost.cz/knihovnaprosazovani-genderove-rovnosti [Accessed 11 Dec. 2020].

[50] Kohonen, T., Oja, E., Simula, O., Visa, A. and Kangas, J. (1996). Engineering applications of the self-organizing map. Proceedings of the IEEE, [online] 84(10), pp.1358–1384. Available at: https://ieeexplore.ieee.org/document/537105 [Accessed 20 Dec. 2020].

[51] Walczak, S. and Cerpa, N. (2003). Artificial Neural Networks. Encyclopedia of Physical Science and Technology, [online] pp.631–645. Available at:

https://www.sciencedirect.com/science/article/pii/B0122274105008371?via%3Dihub [Accessed 20 Dec. 2020].

[52] Krenek, J., Kuca, K., Blazek, P., Krejcar, O. and Jun, D. (2016). Application of Artificial Neural Networks in Condition Based Predictive Maintenance. Recent Developments in Intelligent Information and Database Systems, pp.75–86.

[53] Varol, T., Canakci, A. and Ozsahin, S. (2013). Artificial neural network modeling to effect of reinforcement properties on the physical and mechanical properties of Al2024–B4C composites produced by powder metallurgy. Composites Part B: Engineering, [online] 54, pp.224–233. Available at: https://www.sciencedirect.com/science/article/pii/S1359836813002588 [Accessed 20 Dec. 2020].

[54] Schalkoff, R.J. (1997). Artificial neural networks. New York: Mcgraw-Hill.

[55] Basheer, I.A. and Hajmeer, M. (2000). Artificial Neural networks: fundamentals, computing, design, and Application. Journal of Microbiological Methods, [online] 43(1), pp.3–31. Available at: https://www.sciencedirect.com/science/article/pii/S0167701200002013 [Accessed 21 Dec. 2020].

[56] Ong, J. and Sibte Raza Abidi, S. (1999). Data Mining Using Self-Organizing Kohonen maps: a Technique for Effective Data Clustering Visualisation. [online] Available at: https://www.researchgate.net/publication/220835585_Data_Mining_Using_Self-Organizing_Kohonen_Maps_A_Technique_for_Effective_Data_Clustering_Visualization [Accessed 25 Dec. 2020].

[57] Asan, U. and Ercan, S. (2012). An Introduction to Self-Organizing Maps. Atlantis Computational Intelligence Systems, pp.295–315.

[58] Adamala, S. (2019). Nonlinear Evapotranspiration Modeling Using Artificial Neural Networks. Advanced Evapotranspiration Methods and Applications.

[59] Kohonen, T. (1998). The self-organizing Map. Neurocomputing, 21(1–3), pp.1–6.

[60] Rahman, R.M. and Sultana, N. (2013). Geospatial Data Mining Techniques to Investigate Gender Equality and Empowerment of Women Status in Bangladesh. International Journal of Knowledge Engineering and Soft Data Paradigms, 4(2), p.166.

[61] Mirzabalaeva, F.I., Kvachev, V.G. and Kuksova, O.D. (2018). Clusterization of Regional Labor Markets in Russia Using Kohonen Maps. Revista ESPACIOS, [online] 39(24). Available at: https://www.revistaespacios.com/a18v39n24/18392435.htmliden [Accessed 27 Dec. 2020].

[62] Gaubert, P. and Cottrell, M. (1999). Neural Network and Segmented Labour Market. European Journal of Economic and Social Systems, [online] 14, pp.19–40. Available at: https://hal.archives-ouvertes.fr/hal-00134146/document [Accessed 27 Dec. 2020].

[63] Pö, M., Honkela, T. and Kohonen, T. (2009). Bibliography of Self-Organizing Map (SOM) Papers: 2002-2005 Addendum. [online] Available at: http://www.cis.hut.fi/research/som-bibl/ [Accessed 27 Dec. 2020].

[64] Otevřená společnost o.p.s. Gendrová mapa. [online] www.genderovamapa.cz. Available at: http://www.genderovamapa.cz/about [Accessed 28 Dec. 2020].

[65] Vesanto, J., Himberg, J., Alhoniemi, E. and Parhankangas, J. (1999). Self-organizing Map in Matlab: the SOM Toolbox. [online] Laboratory of Computer and Information Science, Helsinki University of Technology, Finland. Available at: http://lib.tkk.fi/Diss/2002/isbn951226093X/article2.pdf.

[66] Wehrens, R. and Buydens, L.M.C. (2007). Self- and Super-organizing Maps in R: the Kohonen Package. Journal of Statistical Software, 21(5).

[67] Vesanto, J. (1999). SOM-based Data Visualization Methods. Intelligent Data Analysis, 3(2), pp.111–126.

[68] Spanakis, G. and Weiss, G. (2016). AMSOM: Adaptive Moving Self-organizing Map for Clustering and Visualization. Proceedings of the 8th International Conference on Agents and Artificial Intelligence.

[69] Anh Tu, L. (2020). Improving Feature Map Quality of SOM Based on Adjusting the Neighborhood Function. Sustainability in Urban Planning and Design.

[70] Kamimura, R. (2014). Input Information Maximization for Improving self-organizing Maps. Applied Intelligence, 41(2), pp.421–438.

[71] Lee, J.A. and Verleysen, M. (2002). Self-organizing Maps with Recursive Neighborhood Adaptation. Neural Networks, 15(8–9), pp.993–1003.

[72] MIKULÍKOVÁ, Pavla. Criminality Analysis in the Czech Republic using Self-Organizing Maps. Praha, 2020. Bakalářská práce. Univerzita Karlova, Fakulta sociálních věd, Institut ekonomických studií. Vedoucí práce Cahlík, Tomáš.

List of Figures

Appendix

i	Region
1	Hlavní město Praha
2	Praha-východ
3	Praha-západ
4	Plzeň-město
5	Plzeň-sever
6	Plzeň-jih
7	Benešov
8	Sokolov
9	Rakovník
10	Náchod
11	Rychnov nad Kněžnou
12	Prachatice
13	Český Krumlov
14	Děčín
15	Ostrava-město
16	Klatovy
17	Příbram
18	Domažlice
19	Karviná
20	Zlín
21	Frýdek-Místek
22	Břeclav
23	Šumperk
24	Opava
25	Česká Lípa
26	Jihlava
27	Karlovy Vary
28	Havlíčkův Brod
29	Jičín
30	Cheb
31	Hradec Králové
32	Blansko
33	Uherské Hradiště
34	Kladno
35	Semily
36	Litoměřice
37	Pardubice
38	Kroměříž
39	Nový Jičín

i	Region
40	Chrudim
41	Tachov
42	Olomouc
43	Kolín
44	Most
45	Tábor
46	Znojmo
47	Hodonín
48	Chomutov
49	Ústí nad Orlicí
50	Brno-město
51	Brno-venkov
52	Nymburk
53	Bruntál
54	Ústí nad Labem
55	Liberec
56	Trutnov
57	Písek
58	Mladá Boleslav
59	Třebíč
60	Prostějov
61	Vsetín
62	Jeseník
63	Jindřichův Hradec
64	Kutná Hora
65	Vyškov
66	Strakonice
67	Pelhřimov
68	Svitavy
69	Rokycany
70	České Budějovice
71	Přerov
72	Jablonec nad Nisou
73	Teplice
74	Louny
75	Žďár nad Sázavou
76	Beroun
77	Mělník

Figure 18: List of regions