In this work, were calculated the arithmetic mean and median for the respective data for the subsequent years. The described results of the average indicate that the data analysis points to positive changes in the studied areas over recent years, such as a reduction in the gender pay gap, an increase in the minimum wage, and a decrease in the level of poverty. At the same time, an increase in the disparity in the number of employed individuals was observed, which may be subject to further research and analysis. Based on the median results, it can be observed that the median analysis confirms the trends seen in the average graphs, suggesting that the changes in the studied areas over recent years are consistent with both averages and medians.

The analysis began with examining the relationships between various variables and the "gender pay gap." The results showed a moderate negative correlation between the gender pay gap and the minimum wage. This means that gender pay differences are smaller when the minimum wage is higher, suggesting that a higher minimum wage impacts reducing pay inequalities.

Furthermore, it was shown that there is a correlation between poverty and pay inequalities, as a greater pay gap between genders was correlated with a higher likelihood of poverty. Additionally, the research indicated that a smaller gender division in the labor market could result in smaller pay inequalities, as smaller differences in employment between genders were correlated with smaller pay differences.

Next, linear regression was conducted to assess the impact of independent variables on the gender pay gap. According to the regression model, about 61% of the variability in the gender pay gap can be attributed to independent variables such as employment differences, poverty risk, and the minimum wage.

To assess heteroscedasticity and autocorrelation of the residuals, model diagnostic tests such as the Breusch-Pagan test and the Durbin-Watson test were used. A histogram of the residuals showed that the residuals of the model were close to a normal distribution.

Moreover, groups of observations were examined using cluster analysis (PCA) based on the characteristics of the variables that were examined. Using KMeans clustering analysis, five clusters of observations were identified, differing in employment differences, poverty risk, and minimum wage values. Clusters 2 had the highest poverty risk, while clusters 0 and 4 showed the greatest differences in earnings.

Ultimately, the Q-Q plot analysis of the residuals confirmed that the model residuals were consistent with the normal distribution in most cases, with minor deviations at the extremes.

It can be concluded that gender pay differences will decrease with an increase in the minimum wage. However, larger gender pay differences are associated with a higher risk of poverty. The linear regression model showed that about 61% of the variability in the gender pay gap is caused by the minimum wage, poverty risk, and employment differences. Identifying five clusters facilitated the understanding of economic differences between observation groups. Diagnostic tests confirmed that the linear regression model was appropriate for the data. However, the model cannot explain all the variability in the pay gap. According to the ANOVA analysis, there are statistically significant differences between clusters regarding at least one examined variable, confirming their usefulness.

The conducted analysis provides significant information about the studied variables and their interrelations. It also allows for a better understanding of the dynamics and trends in the studied areas over the analyzed period. A downward trend for the "gender gap" variable can also be observed, suggesting an improvement in women's labor market situation in Europe.

From the conducted study, several conclusions can be drawn. It can be observed that a higher minimum wage is correlated with a reduction in gender pay differences. This suggests that policies increasing the minimum wage can be an effective tool in combating pay inequalities. There is also a strong correlation between larger pay differences and a higher risk of poverty. This indicates that pay differences may be one of the key factors influencing the level of poverty, emphasizing the importance of policies reducing these differences.

A smaller gender division in the labor market (smaller differences in employment between genders) is associated with smaller pay differences. Therefore, increasing gender equality in the labor market may lead to a reduction in pay inequalities.

The linear regression model explains 61% of the variability in gender pay differences. This indicates that the analyzed variables (minimum wage, poverty risk, employment differences) have a significant but not exclusive impact on pay differences, suggesting the existence of other, unexamined factors. Cluster analysis, in turn, showed the existence of five clusters differing in terms of poverty risk, employment differences, and minimum wage. These clusters help better understand economic diversification among different observation groups, which can be useful in designing public policies.

The observed downward trend for the "gender gap" variable suggests that the situation of women in the European labor market is improving. This may indicate the effectiveness of current actions and policies aimed at reducing pay inequalities.

The data analysis confirmed the hypothesis that there are other factors influencing the gender gap than those listed in the first chapter. This suggests the need for further research to identify and analyze additional variables affecting gender pay differences.

In conducting the above analysis, the Python programming language played a crucial role. Using Python in analyzing factors affecting gender pay differences significantly simplified, accelerated, and systematized the entire research process. Python proved to be an extremely useful tool at every stage of the analysis, from data preparation to result visualization.

The Pandas library facilitated data import from various sources and its transformation into the appropriate format. Using methods like read_csv() allowed for quick data loading into DataFrames. Pandas enabled easy data processing and cleaning, including filtering, removing missing data, and creating new columns based on existing data. Using functions like corr() in Pandas allowed for quick calculation of correlation matrices, helping to understand the relationships between variables. Thanks to libraries like Pandas, processing and analyzing large datasets became much more efficient, enabling work on data that would be difficult to manage manually.

Thanks to Seaborn and Matplotlib, it was possible to create clear and intuitive correlation plots, making result interpretation easier. Using these libraries allowed for creating various plots, such as scatter plots, residual histograms, and Q-Q plots, which facilitated result interpretation and conclusion communication.

Scikit Learn, on the other hand, made it easier to conduct linear regression analysis, providing functions for model creation, training, and result evaluation. This library enabled the application of clustering algorithms like KMeans, which allowed for identifying groups with similar characteristics in the data.