A Brief Analysis on some Economic Factors that Influence Canada's Gender Wage Gap

Table of Contents

1. Objectives

Gender equality is something that our society has been striving towards; since the beginning of the 20th century the rights of women have become an ongoing issue. With the beginning of the second world war there was a lack of men available to work, for the first time married women in addition to single women worked enmasse. By 1951 women accounted for 22% of the labour force in Canada. However, it was not until 1986 until 6 provinces instituted legislation regarding the gap between the pay of men and women. It was not until 2021 that the Canadian Federal government passed the Pay Equity Act to address the wage gap between genders, disability, indigious peoples and visible minorities; the purpose of which is to force companies to account for differences in pay. Across the world there has been a variety of stances on the disparity in the pay between genders.

The Organization for Economic Cooperation and Development (OECD) is an international organization that works with government, policy makers and citizens to build evidence-based international standards to find solutions to social, economic and environmental challenges. OECD provides a large forum of worldwide data and analysis.

This investigation examined the gender wage gap across Canada. We focused on how the gender wage gap has changed in the last 24 years as well as some factors that have potentially influenced the gender wage gap. The gender wage gap is determined by the difference between annual median earnings of women and men relative to the annual median earnings of men.

2. Data Preparation

2.1 Data source

The data set was generated by combining numerous data sets from the OECD. These data sets combine the year, gender wage gap, average wages, wage level (high or low), GDP and the growth rate of the population of Canada.

2.2 Data Quality

This study will measure the data quality from the following 6 aspects:

Accuracy: The dataset comes from the OECD which is a trusted data source that can reflect the real average income of the countries all around the world.

Completeness: The dataset contains 6 columns covering the year, gender wage gap, average wages, wage level (high or low), GDP and population of Canada. Therefore, the dataset effectively covers the aspects that the analysis needed.

Consistency: This dataset has a high consistency.

Validity: The OECD collects data from different countries based on international parameters that ensure the validity of this dataset.

Uniqueness: The dataset is unique and this study will do the data cleaning process to ensure the uniqueness.

Timeliness: The dataset is about the year 1951 to 2020, which is useful for analyzing the current and recent gender pay gap.

2.3 Data Structure

In the raw dataset, there are 6 columns, including: year, gender wage gap, average wage, wage level (high or low), GDP and population of Canada.Detailed as the following chart shows:

```
Float64Index: 71 entries, 1951.0 to 2021.0
Data columns (total 7 columns):
 #  Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0  gender_wage_gap   24 non-null     float64
 1  avg_wages         31 non-null     float64
 2  employment_rate   46 non-null     float64
 3  wage_levels_LPAY  23 non-null     float64
 4  wage_levels_HPAY  23 non-null     float64
 5  gdp               52 non-null     float64
 6  pop               70 non-null     float64
dtypes: float64(7)
```

2.4 Data Cleaning

Dataframe.reset_index, Dataframe.rename(columns) were used to filter out data based on the year for Canada.
Reduce and concat were used to reduce down the content of the data frames and merge them together.
Dataframe.dropna was used to drop null values.

2.5 Challenges

The main challenge with this data was combining the multiple different data sets and then narrowing them down until they only contained useful data to analyze. It was important to have all the data frames combined using time as the index. Then to remove any data that corresponded to null gender wage gap values.

3. Analysis

Initially, when investigating the wage gap we compared Canada to six other countries to provide a frame of reference to how the Canadian wage gap compares on the world stage.

To then examine the gender wage gap in Canada we first were interested in the basic trend of how the wage gap has changed over-time in Canada. We had data from 1997 until 2020.
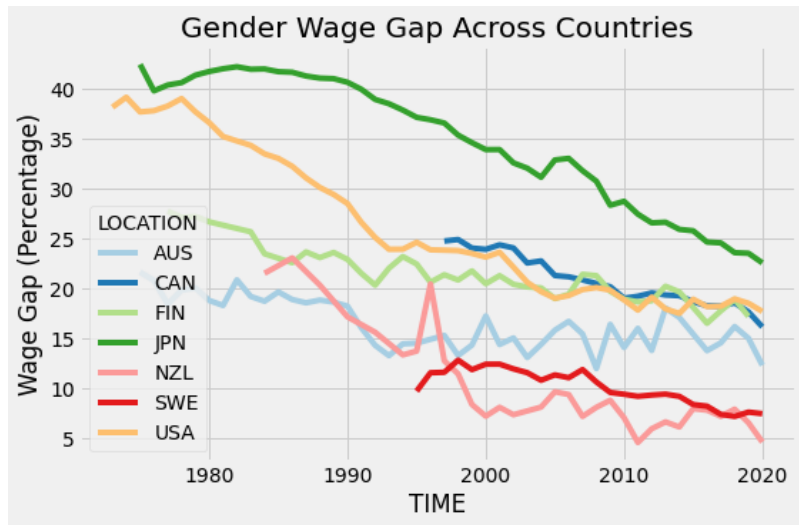
As shown in the graph, the wage gap has dropped almost 9% from 24.86% to 16.11% over the last 24 years. In order to examine the factors that correlate to this decrease we used a scatter matrix to compare the correlation of the gender wage gap to 6 other economic factors. Through viewing this scatter matrix we were able to visualize these relationships and to focus on any with a linear relationship. To ensure that no correlation was missed we performed an ordinary least squares model for each factor. The population growth rate, employment rate, and low paying wage level have little correlation with the gender wage gap, all with an adjusted R squared below 0.3 and F-statistic below 10 (See Appendix 1).
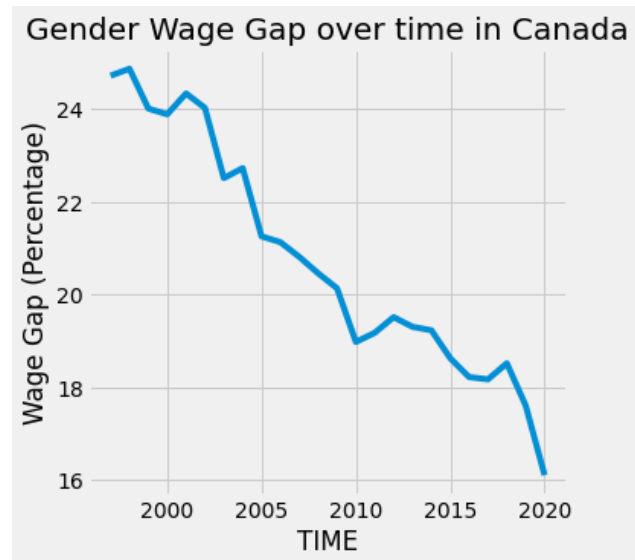
3.1 Average Wage and the Gender Wage Gap Correlation

The average wage was calculated by OECD by "dividing the national-accounts-based total wage bill by the average number of employees in the total economy, which is then multiplied by the ratio of the average weekly hours per full-time employee to the average usually weekly hours for all employees." The average wage is measured in USD. Using the scatter matrix, the effect of the average wage on the gender wage gap appeared linear. It was plotted individually, with the line of best fit to examine its linearity, additionally its residuals were plotted to determine the correct

model was used. There is variability across the residuals, and the residuals are independent.

Effect of the Average Wage on the Gender Wage Gap

Residuals of the effect of the Average Wage on Gender Wage Gap
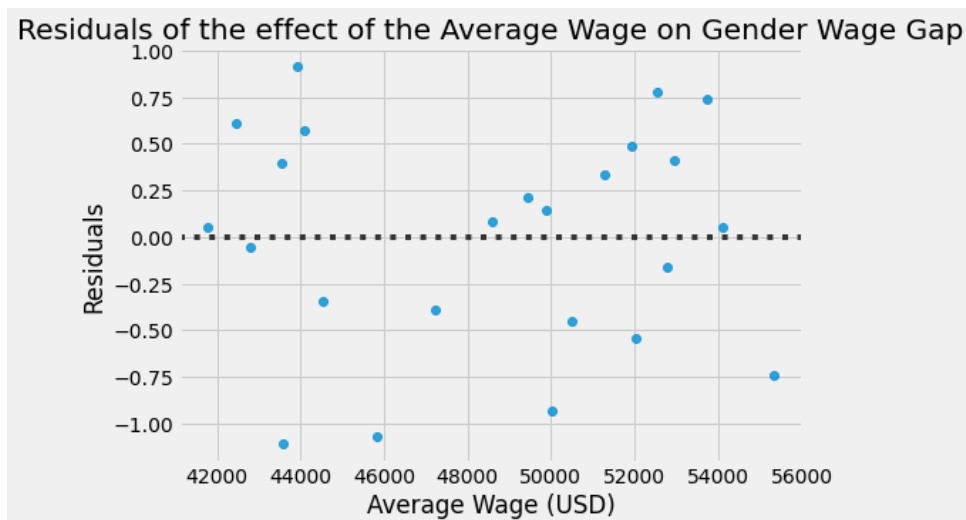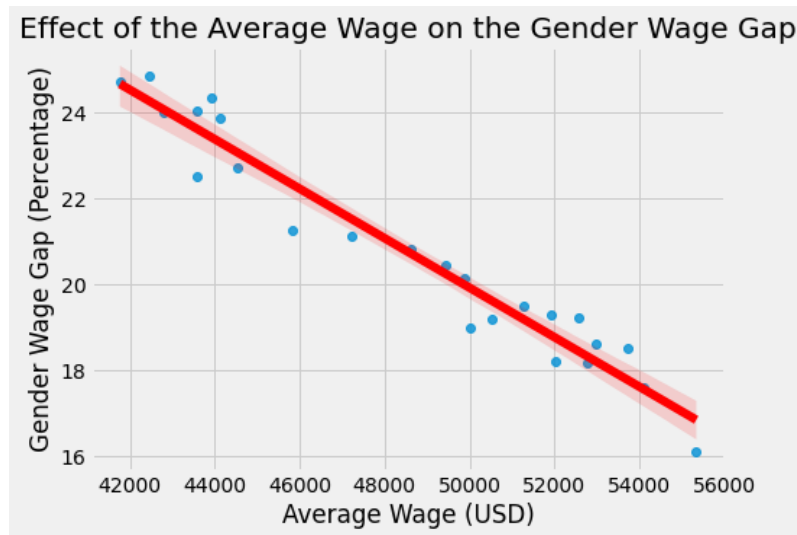
The OLS method summary allows us to confirm that the gender wage gap and the average wage are correlated. The adjusted R squared value is 0.944 suggesting that the model provides a clear fit for the observed values. The high F-statistic and low p-value confirm the model is correct and that the intercept and slope are not equal to zero. Additionally the p-value of the slope is 0, and the confidence interval does not include zero suggesting that the slope is not 0. This allows us to suggest that as the average wage increases the gender wage gap decreases. The OLS model is shown below.

OLS Regression Results

Dep. Variable:	gender_wage_gap	R-squared:	0.946
Model:	OLS	Adj. R-squared:	0.944
Method:	Least Squares	F-statistic:	389.1
Date:	Thu, 14 Apr 2022	Prob (F-statistic):	1.78e-15
Time:	23:36:29	Log-Likelihood:	-20.874
No. Observations:	24	AIC:	45.75
Df Residuals:	22	BIC:	48.10
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	48.6454	1.419	34.280	0.000	45.703	51.588
avg_wages	-0.0006	2.91e-05	-19.726	0.000	-0.001	-0.001

Omnibus:	1.659	Durbin-Watson:	1.347
Prob(Omnibus):	0.436	Jarque-Bera (JB):	1.318
Skew:	-0.389	Prob(JB):	0.517
Kurtosis:	2.155	Cond. No.	5.62e+05

3.2 GDP and Gender Wage Gap Correlation:

Gross domestic product (GDP) which is a measure of the value added through the production of goods and services in a country in a given year. We wanted to determine if there was a correlation between increasing GDP as the gender pay gap decreases. Through the scatter plot matrix, it appears as though the relationship between GDP and the Gender wage gap is linear. This relationship was then plotted individually to better examine this trend, additionally a residuals plot was generated to confirm that the correct model was used.

Residuals of the effect of GDP on the Gender Wage Gap
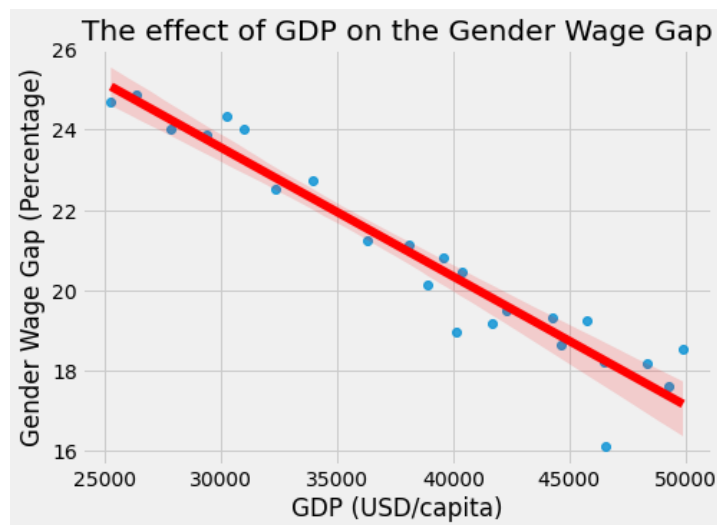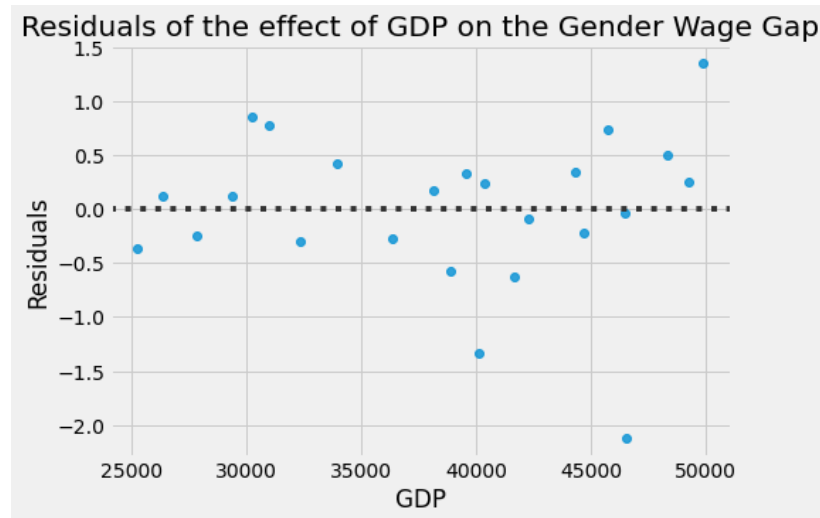
Both the scatter plot and the residuals plot suggest that the linear model is best. There is variability across the residuals, and the residuals are independent, however, there appear to be a few residuals farther from 0. Therefore an influence plot was generated.
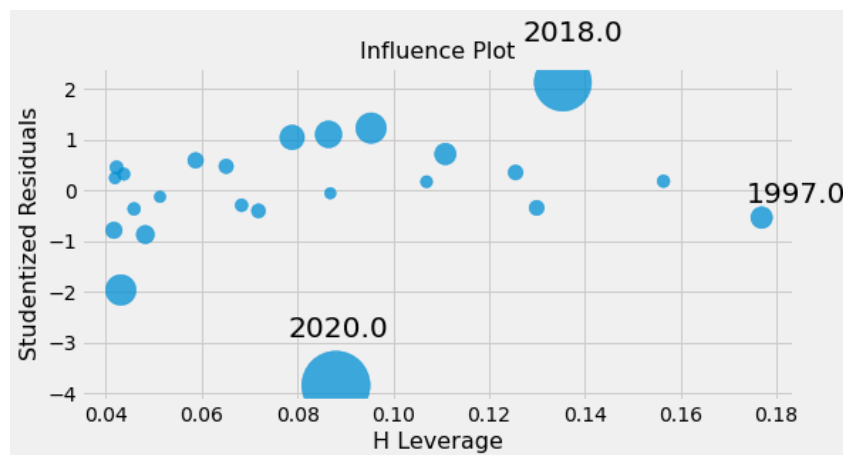
The influence plot suggests that the 2020 and 2018 data points have more of an effect on the model generated. The OLS model was then performed. The OLS method summary allows us to confirm that the gender wage gap and the GDP are related. The adjusted R squared value is 0.917 suggesting that the model provides a clear fit for the observed values. The high F-statistic and low p-value confirm the model is correct and that the intercept and slope are not equal to zero. Additionally the p-value of the slope is 0, and the confidence interval does not include zero suggesting that the slope is not 0. Omnibus and Jarque-Bera are both relatively high, suggesting a normal distribution and limited skewness. This suggests that as Canada's GDP increases the gender wage gap decreases. The OLS summary is below.

```
                    OLS Regression Results
    Dep. Variable:    gender_wage_gap      R-squared:         0.920
          Model:     OLS                Adj. R-squared:     0.917
         Method:     Least Squares         F-statistic:      253.9
           Date:     Thu, 14 Apr 2022   Prob (F-statistic): 1.45e-13
           Time:     23:36:33           Log-Likelihood:    -25.658
  No. Observations:  24                        AIC:          55.32
     Df Residuals:   22                         BIC:         57.67
        Df Model:    1
  Covariance Type:   nonrobust
                coef     std err      t      P>|t|  [0.025  0.975]
  Intercept  33.1803   0.794      41.799   0.000  31.534  34.827
       gdp   -0.0003   2.01e-05  -15.936   0.000  -0.000  -0.000
      Omnibus:      8.871    Durbin-Watson:    1.220
  Prob(Omnibus):   0.012    Jarque-Bera (JB): 7.002
         Skew:     -0.979         Prob(JB):    0.0302
      Kurtosis:     4.779         Cond. No.    2.08e+05
```

3.3 Correlation of the Percentage of High Pay Workers with the Gender Wage Gap

The High Wage level refers to the share of workers earning more than one-and-a-half times median earnings. We wanted to determine if there was a correlation between increasing the percentage of high wage earners and a decrease in the gender pay gap. Through the scatter plot matrix, it appears as though the relationship between wage level (High) and the Gender wage gap is linear. This relationship was then plotted individually to better examine this trend, additionally a residuals plot was generated to confirm that the correct model was used. There is variability across the residuals, and the residuals are independent.

Residuals of the effect of Wage Level on the Gender Wage Gap
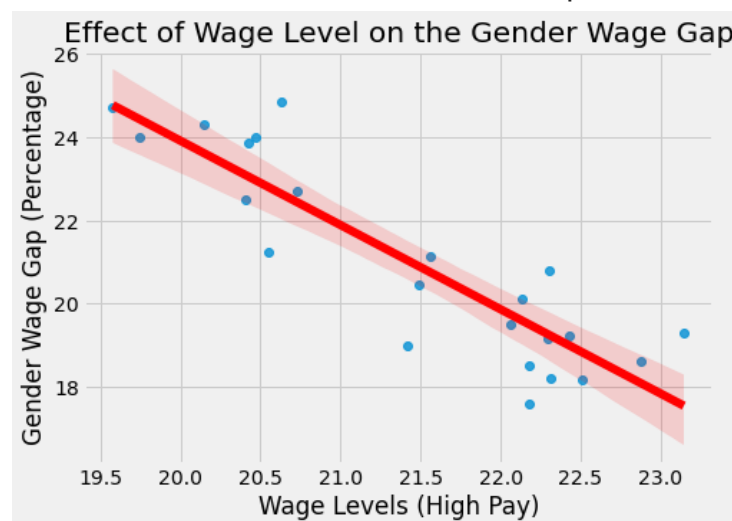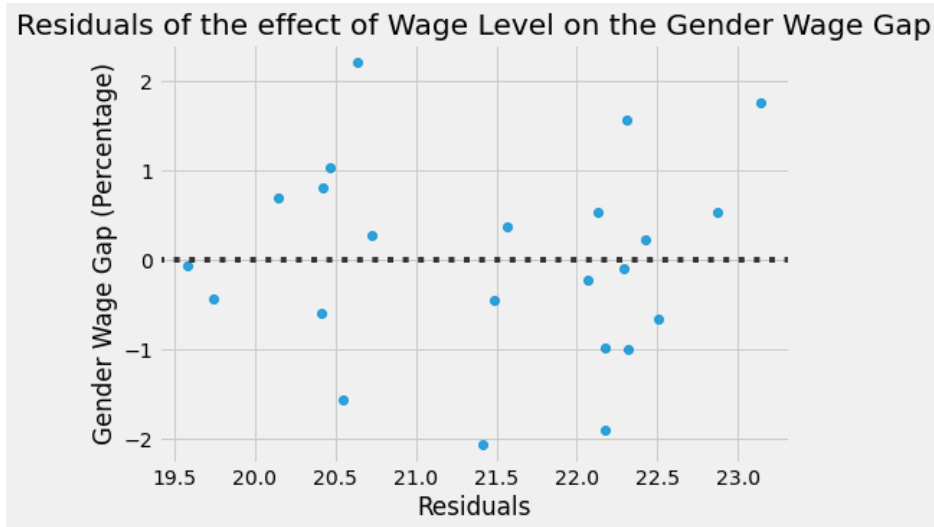
The OLS method summary allows us to confirm that the gender wage gap and the percentage of high wage earners are correlated. While the adjusted R squared value is not as significant as the previous factors, it is 0.777. The relatively high F-statistic and low p-value confirm the model is a good fit and that the intercept and slope are not equal to zero. Additionally the p-value of the slope is 0, and the confidence interval does not include zero suggesting that the slope is not 0. This suggests that as the percentage of high wage earners increases the gender wage gap decreases. The OLS method summary is shown below:

OLS Regression Results

Dep. Variable:	gender_wage_gap	R-squared:	0.787
Model:	OLS	Adj. R-squared:	0.777
Method:	Least Squares	F-statistic:	77.44
Date:	Thu, 14 Apr 2022	Prob (F-statistic):	1.73e-08
Time:	23:36:32	Log-Likelihood:	-34.518
No. Observations:	23	AIC:	73.04
Df Residuals:	21	BIC:	75.31
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	64.4685	4.950	13.025	0.000	54.175	74.762
wage_levels_HPAY	-2.0275	0.230	-8.800	0.000	-2.507	-1.548

Omnibus:	0.007	Durbin-Watson:	1.578
Prob(Omnibus):	0.997	Jarque-Bera (JB):	0.192
Skew:	0.013	Prob(JB):	0.909
Kurtosis:	2.554	Cond. No.	450.

3.4 GDP, Average Wage, High Wage level and Gender Wage Gap Correlation

As mentioned previously, the average wage, the percentage of high wage earners and GDP are linearly related to the gender wage gap in Canada. Herein, we discuss how these factors coupled together affect the gender wage gap. To ensure linearity between variables GDP, wage level and average wage were tested together as sets of twos. Finally, multiple linear regression was used to find the correlation of the gender wage gap when you combine GDP, average wage and wage level. Through the comparison of these multiple linear regression summary tables (see in Appendix 2), the wage level was excluded because the p-value of its coefficient was consistently above the 10% significance level, and the confidence interval always included zero. Therefore, the best model includes only the GDP and the average wage to predict the gender wage gap. The model generated is:

Gender wage gap = 45.1792 - 0.0004(average wage) - 0.00007751(GDP)

4. Conclusions

In conclusion we can see that the gender wage gap is slowly closing, it has decreased almost 9% over the last 24 years. Additionally we can conclude that closing the gender wage gap is influenced by a couple economic factors,primarily the increase in Canada's GDP, and an increase in the average wage for Canadian workers. Using linear models we were able to see a correlation between decreasing the gender wage gap and increasing the average wage and increasing Canada's GDP. We were able to generate two useful models based on our data. One to calculate the increase in the average wage of a Canadian worker as the gender wage gap decreases, and a second model to calculate the increase in the GDP as the gender wage gap decreases. Furthermore, when combined into multiple linear regression, GDP and average wage used in conjunction provided the best model. The employment rate, wage level and the population growth of Canada were all investigated but determined that they did not provide an adequate model. Further data analysis will need to be generated to ensure that this increase in gender wage gap was not caused by unseen confounding variables.

Appendix 1

Graph and OLS Summary of the effect of the Employment Rate on the Gender Wage Gap:

Residuals of the effect of the Employment Rate on Gender Wage Gap
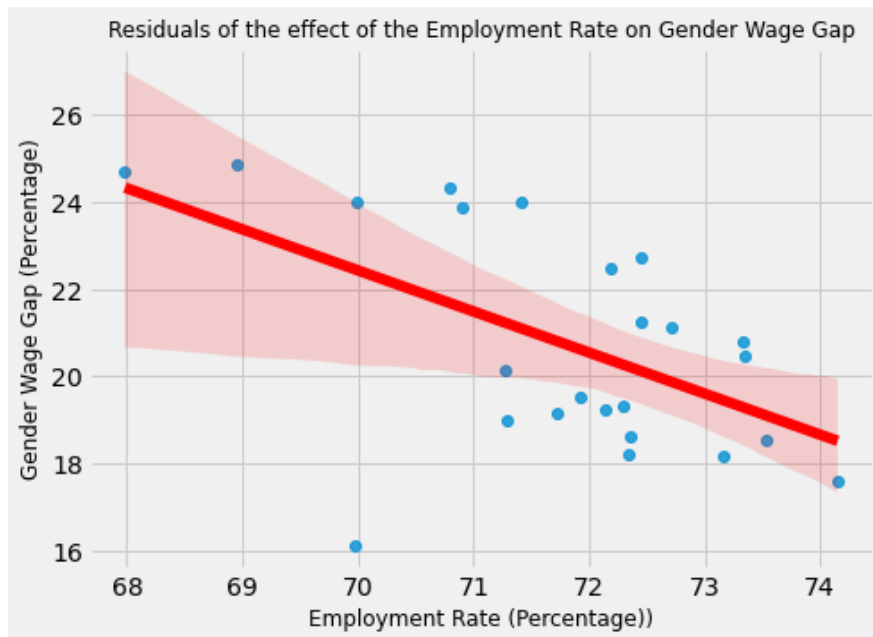
OLS Regression Results

Dep. Variable:	gender_wage_gap	**R-squared:**	0.296
Model:	OLS	**Adj. R-squared:**	0.264
Method:	Least Squares	**F-statistic:**	9.253
Date:	Mon, 18 Apr 2022	**Prob (F-statistic):**	0.00598
Time:	21:26:12	**Log-Likelihood:**	-51.795
No. Observations:	24	**AIC:**	107.6
Df Residuals:	22	**BIC:**	109.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	88.4786	22.267	3.973	0.001	42.299	134.658
employment_rate	-0.9435	0.310	-3.042	0.006	-1.587	-0.300

Omnibus:	6.693	**Durbin-Watson:**	0.412
Prob(Omnibus):	0.035	**Jarque-Bera (JB):**	4.520
Skew:	-0.885	**Prob(JB):**	0.104
Kurtosis:	4.178	**Cond. No.**	3.58e+03

Graph and OLS Summary of the effect of the Percentage of Low Wage Earners on the Gender Wage Gap:

Effect of the Low Wage Level on the Gender Wage Gap
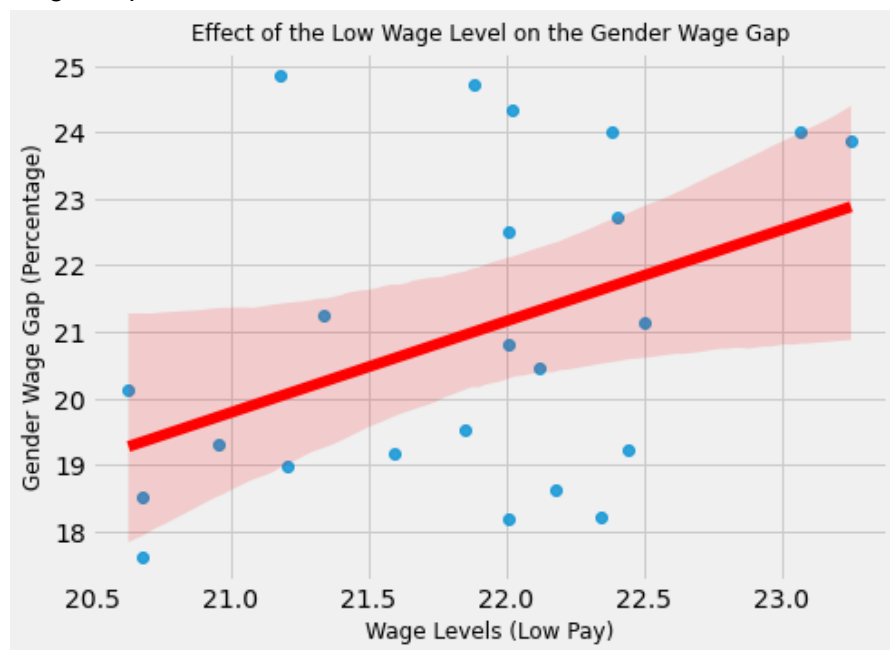
OLS Regression Results

Dep. Variable:	gender_wage_gap	R-squared:	0.172
Model:	OLS	Adj. R-squared:	0.133
Method:	Least Squares	F-statistic:	4.369
Date:	Mon, 18 Apr 2022	Prob (F-statistic):	0.0489
Time:	21:33:43	Log-Likelihood:	-50.112
No. Observations:	23	AIC:	104.2
Df Residuals:	21	BIC:	106.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-9.1318	14.405	-0.634	0.533	-39.088	20.824
wage_levels_LPAY	1.3767	0.659	2.090	0.049	0.007	2.747

Omnibus:	1.037	Durbin-Watson:	0.396
Prob(Omnibus):	0.595	Jarque-Bera (JB):	0.951
Skew:	0.435	Prob(JB):	0.622
Kurtosis:	2.515	Cond. No.	677.

Graph and OLS Summary of the effect of the Growth Rate of the Population of Canada on the Gender Wage Gap:

The effect of the Population Growth Rate on the Gender Wage Gap
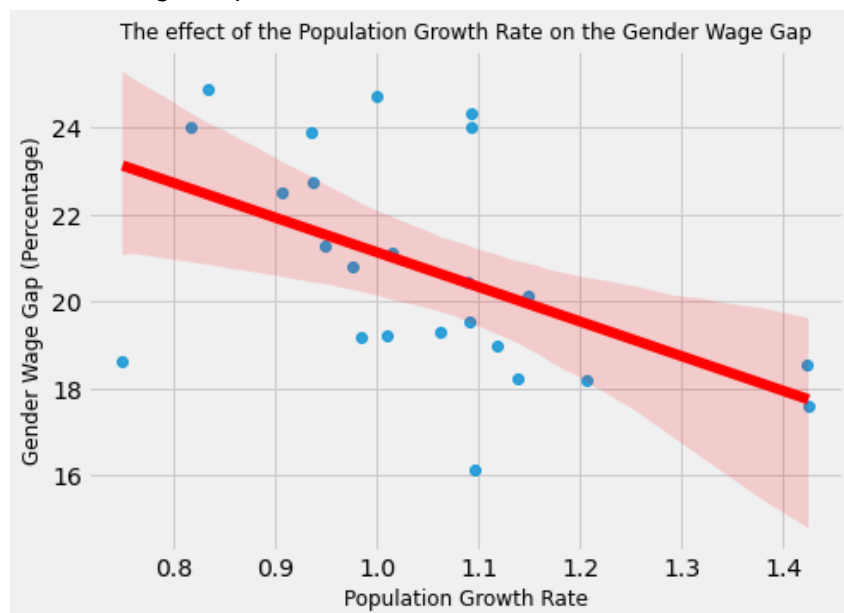
OLS Regression Results

Dep. Variable:	gender_wage_gap	R-squared:	0.254
Model:	OLS	Adj. R-squared:	0.220
Method:	Least Squares	F-statistic:	7.477
Date:	Mon, 18 Apr 2022	Prob (F-statistic):	0.0121
Time:	21:36:14	Log-Likelihood:	-52.497
No. Observations:	24	AIC:	109.0
Df Residuals:	22	BIC:	111.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	29.0772	3.077	9.451	0.000	22.697	35.458
pop	-7.9556	2.909	-2.734	0.012	-13.989	-1.922

Omnibus:	0.054	Durbin-Watson:	0.531
Prob(Omnibus):	0.973	Jarque-Bera (JB):	0.073
Skew:	-0.043	Prob(JB):	0.964
Kurtosis:	2.743	Cond. No.	13.3

Appendix 2

OLS Summary of the effect of Canada's GDP and Average Wage on the Gender Wage Gap:

```
                           OLS Regression Results
==============================================================================
Dep. Variable:        gender_wage_gap   R-squared:                       0.949
Model:                            OLS   Adj. R-squared:                  0.944
Method:                 Least Squares   F-statistic:                     196.5
Date:                Sun, 17 Apr 2022   Prob (F-statistic):           2.54e-14
Time:                        22:31:34   Log-Likelihood:                -20.232
No. Observations:                  24   AIC:                             46.46
Df Residuals:                      21   BIC:                             50.00
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      45.1792      3.523     12.823      0.000      37.852      52.506
gdp         -7.751e-05   7.22e-05     -1.074      0.295      -0.000    7.26e-05
avg_wages      -0.0004      0.000     -3.465      0.002      -0.001      -0.000
==============================================================================
Omnibus:                        1.010   Durbin-Watson:                   1.400
Prob(Omnibus):                  0.603   Jarque-Bera (JB):                0.981
Skew:                          -0.363   Prob(JB):                        0.612
Kurtosis:                       2.327   Cond. No.                     1.80e+06
==============================================================================
```

OLS Summary of the effect of Canada's GDP and the High Wage level on the Gender Wage Gap:

```
                           OLS Regression Results
==============================================================================
Dep. Variable:        gender_wage_gap   R-squared:                       0.949
Model:                            OLS   Adj. R-squared:                  0.944
Method:                 Least Squares   F-statistic:                     185.1
Date:                Sun, 17 Apr 2022   Prob (F-statistic):           1.25e-13
Time:                        22:33:50   Log-Likelihood:                -18.121
No. Observations:                  23   AIC:                             42.24
Df Residuals:                      20   BIC:                             45.65
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                    coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept         37.7645      4.179      9.038      0.000      29.048      46.481
wage_levels_HPAY   -0.2991      0.246     -1.214      0.239      -0.813       0.215
gdp               -0.0003    3.4e-05     -7.952      0.000      -0.000      -0.000
==============================================================================
Omnibus:                        4.531   Durbin-Watson:                   1.442
Prob(Omnibus):                  0.104   Jarque-Bera (JB):                2.587
Skew:                          -0.708   Prob(JB):                        0.274
Kurtosis:                       3.833   Cond. No.                     1.37e+06
==============================================================================
```

OLS Summary of the effect of the Average Wage and the High Wage level on the Gender Wage Gap:

```
                        OLS Regression Results
==============================================================================
Dep. Variable:          gender_wage_gap   R-squared:                       0.942
Model:                              OLS   Adj. R-squared:                  0.936
Method:                   Least Squares   F-statistic:                     163.0
Date:                Sun, 17 Apr 2022    Prob (F-statistic):           4.17e-13
Time:                        22:33:10    Log-Likelihood:                -19.505
No. Observations:                  23    AIC:                             45.01
Df Residuals:                      20    BIC:                             48.42
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
Intercept          47.3367      3.525     13.428      0.000      39.983      54.690
avg_wages          -0.0006   7.85e-05     -7.334      0.000      -0.001      -0.000
wage_levels_HPAY    0.0658      0.311      0.212      0.834      -0.582       0.714
==============================================================================
Omnibus:                        1.235   Durbin-Watson:                   1.351
Prob(Omnibus):                  0.539   Jarque-Bera (JB):                1.116
Skew:                          -0.392   Prob(JB):                        0.572
Kurtosis:                       2.258   Cond. No.                     1.36e+06
==============================================================================
```

OLS Summary of the effect of Canada's GDP, Average Wage and the High Wage level on the Gender Wage gap:

```
                        OLS Regression Results
==============================================================================
Dep. Variable:          gender_wage_gap   R-squared:                       0.954
Model:                              OLS   Adj. R-squared:                  0.946
Method:                   Least Squares   F-statistic:                     130.1
Date:                Sun, 17 Apr 2022    Prob (F-statistic):           7.63e-13
Time:                        22:35:19    Log-Likelihood:                -16.981
No. Observations:                  23    AIC:                             41.96
Df Residuals:                      19    BIC:                             46.50
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
Intercept          40.5186      4.525      8.955      0.000      31.048      49.989
avg_wages          -0.0002      0.000     -1.407      0.175      -0.001       0.000
wage_levels_HPAY   -0.0656      0.292     -0.225      0.825      -0.677       0.546
gdp                -0.0002   7.88e-05     -2.159      0.044      -0.000   -5.23e-06
==============================================================================
Omnibus:                        1.163   Durbin-Watson:                   1.465
Prob(Omnibus):                  0.559   Jarque-Bera (JB):                0.532
Skew:                          -0.371   Prob(JB):                        0.767
Kurtosis:                       3.058   Cond. No.                     2.42e+06
==============================================================================
```

References

Data Reference:

OECD (2022), Wage levels (indicator). doi: 10.1787/0a1c27bc-en (Accessed on 17 April 2022).

OECD (2022), Average wages (indicator). doi: 10.1787/cc3e1387-en (Accessed on 17 April 2022).

OECD (2022), Gender wage gap (indicator). doi: 10.1787/7cee77aa-en (Accessed on 17 April 2022).

OECD (2022), Employment rate (indicator). doi: 10.1787/1de68a9b-en (Accessed on 17 April 2022).

OECD (2022), Gross domestic product (GDP) (indicator). doi: 10.1787/dc2f7aec-en (Accessed on 17 April 2022).

OECD (2022), Population (indicator). doi: 10.1787/d434f82b-en (Accessed on 17 April 2022).