**Statistics 101A Group Project Report - Group 11**
Topic: Irish Companies' Gender Pay Gap Distribution
Group 11: Alan Wu, Alton Law, Oscar Su, Sean Gee, Shelby Jackson

## Introduction:

As a result of the Gender Pay Gap Information Act in Ireland, from 2022 Irish companies with more than 250 employees were required to create and publish a report outlining their gender pay gap. This research project aims to develop a predictive model of pay gaps among different Irish companies to show how factors such as report years, industry, and gender composition of each company affect the pay gap within Ireland. Our data set was sourced from PayGap.ie, which contained 1945 observations and 29 variables to analyze. From these variables, we chose to analyze:

- Report.Year: the specified year in which data was collected (there are only three years: 2022, 2023, 2024)
- Mean.Hourly.Gap: The difference between the mean hourly remuneration of employees of the male gender and that of employees of the female gender expressed as a percentage of the mean hourly remuneration of employees of the male gender
- Median.Hourly.Gap: The difference between the median hourly remuneration of employees of the male gender and that of employees of the female gender expressed as a percentage of the median hourly remuneration of employees of the male gender
- Mean.Bonus.Gap: The difference between the mean bonus remuneration of employees of the male gender and that of employees of the female gender expressed as a percentage of the mean bonus remuneration of employees of the male gender
- Median.Bonus.Gap: The difference between the median bonus remuneration of employees of the male gender and that of employees of the female gender expressed as a percentage of the median bonus remuneration of employees of the male gender
- Mean.Hourly.Gap.Part.Temp: The difference between the mean hourly part-time temp remuneration of employees of the male gender and that of employees of the female gender expressed as a percentage of the mean hourly remuneration of employees of the male gender
- Mean.Hourly.Gap.Part.Time: The difference between the mean hourly part-time remuneration of employees of the male gender and that of employees of the female gender expressed as a percentage of the mean hourly remuneration of employees of the male gender
- Percentage.Bonus.Paid.Female: The percentage of all employees of the female gender who were paid a bonus
- Percentage.BIK.Paid.Female: The percentage of all employees of the female gender who received benefits in kind
- Percentage.Employees.Female: The percentage of all employees who are of the female gender
- Group: Binary variable, where 1 denotes a member of the NACE letters: 'K', 'L', 'M', 'N', "G", "J", and 0 denotes a company that is not. These specific industry labels are chosen based on this official Irish report that detailed the industries that tend to have larger pay gaps. We wanted to corroborate the report's findings with our model.

From the perspective of large companies, it is of the utmost importance to minimize pay gaps between individuals to maintain egalitarian principles. In this project, we investigate gender pay gap trends, the influence of industry, and other contributing factors across Irish companies within a three-year period. To achieve this, we will employ multiple linear regression, multicollinearity analysis, variable selection, and Box-Cox transformation. These techniques allow us to gain deeper insights into the factors driving pay disparities and better understand their impact across different industries within Ireland.

## Data Cleaning:

In this project, we faced the challenge of cleaning our data. There were many NA values. When removing them, our observations dropped significantly, making our data less valuable. We made a standardized decision to replace these NA values with the mean of the remaining observations (better known as mean imputation). This made our data appear normally distributed.

**Statistics:**

In the beginning, we had 16 variables. We plotted their correlation pairs to see which ones are highly correlated with each other and picked one from each highly correlated pair (see Appendix). After the process, 10 variables were left. Below are the summary statistics and distributions of these variables:

| Variable | N | Mean | Std. Dev. | Min | Pctl. 25 | Pctl. 75 | Max |
|---|---|---|---|---|---|---|---|
| Report.Year | 1945 | 2023 | 0.822 | 2022 | 2022 | 2024 | 2024 |
| Mean.Hourly.Gap | 1945 | 11.2 | 12.5 | -52.1 | 3.4 | 18 | 73 |
| Mean.Bonus.Gap | 1945 | 25.5 | 58.8 | -644 | 14.8 | 38.8 | 2028 |
| Median.Bonus.Gap | 1945 | 10.3 | 49.8 | -893 | 4.69 | 24.3 | 264 |
| Mean.Hourly.Gap.Part.Time | 1945 | -2.74 | 29.7 | -267 | -5.56 | 5.09 | 189 |
| Mean.Hourly.Gap.Part.Temp | 1945 | 3.09 | 18.2 | -165 | 0 | 5.1 | 100 |
| Percentage.Bonus.Paid.Female | 1945 | 46.5 | 40.3 | 0 | 0 | 87.3 | 100 |
| Percentage.BIK.Paid.Female | 1945 | 34.5 | 40.4 | 0 | 0 | 83.7 | 100 |
| Percentage.Employees.Female | 1945 | 46.1 | 18.5 | 0 | 34 | 56.2 | 98.2 |
| group | 1945 | 0.37 | 0.483 | 0 | 0 | 1 | 1 |

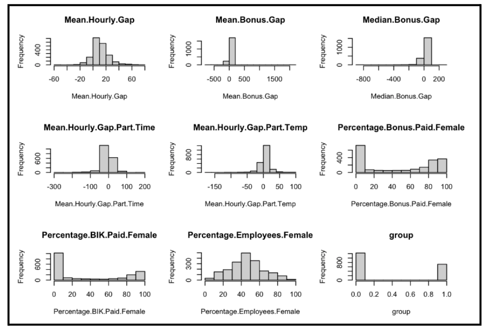



***Table 1: Summary Statistics for Initial Variables / Graph 1: Distribution of Initial Variables***

Judging from the summary and distribution outputs, our response variable (Mean.Hourly.Gap) and some other predictors (Mean.Hourly.Gap.Part.Time, Mean.Hourly.Gap.Part.Temp, and Percentage.Employees.Female) are normally distributed. The other predictors are either bimodal or skewed to one side. Regardless, it is still reasonable to fit a linear regression model to investigate the relationship between the Mean Hourly gap and the predictor variables.

**Model Fitting:**

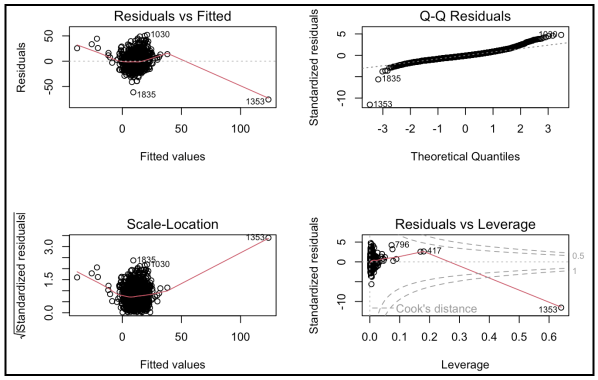
After selecting our variables, our initial full model comes out to become:

$$\widehat{\mathbf{Mean.Hourly.Gap}} = 1,308 - 0.6465(Report.year) + 5.5455 * 10^{-2}(Mean.Bonus.Gap) + 2.071 * 10^{-2}(Median.Bonus.Gap) + 5.434 * 10^{-2}(Mean.Hourly.Gap.Part.Time) + 8.436 * 10^{-2} + 8.426 * 10^{-2}(Mean.Hourly.Gap.Part.Temp) + 6.811 * 10^{-2}(Percentage.Bonus.Paid.Female) + 3.482 * 10^{-2}(Percentage.BIK.Paid.Female) + 8.622 * 10^{-2}(Percentage.Employees.Female) + 1.433(group)$$

```
## Call:
## lm(formula = Mean.Hourly.Gap ~ ., data = new_table)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -75.674  -6.315  -0.960   5.500  52.058
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  1.308e+03  6.121e+02   2.137 0.032687 *
## Report.Year                 -6.465e-01  3.026e-01  -2.136 0.032770 *
## Mean.Bonus.Gap               5.455e-02  4.601e-03  11.856  < 2e-16 ***
## Median.Bonus.Gap             2.071e-02  5.444e-03   3.803 0.000147 ***
## Mean.Hourly.Gap.Part.Time    5.434e-02  8.434e-03   6.443 1.47e-10 ***
## Mean.Hourly.Gap.Part.Temp    8.426e-02  1.366e-02   6.168 8.37e-10 ***
## Percentage.Bonus.Paid.Female 6.811e-02  8.457e-03   8.054 1.39e-15 ***
## Percentage.BIK.Paid.Female   3.482e-02  8.254e-03   4.219 2.57e-05 ***
## Percentage.Employees.Female  8.622e-02  1.405e-02   6.136 1.02e-09 ***
## group                        1.433e+00  5.781e-01   2.479 0.013259 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.93 on 1935 degrees of freedom
## Multiple R-squared:  0.2398, Adjusted R-squared:  0.2363
## F-statistic: 67.83 on 9 and 1935 DF,  p-value: < 2.2e-16
```

***Figure 1: Full Model Summary***

The full model shows every variable to be significant. The summary shows an $R^2$ value of 0.2398, indicating 24% of the variation in Mean.Hourly.Gap can be explained by the model. Our F-statistic is 67.83 with a p-value less than 0.05, meaning at least one of our predictors is significant.




***Graph 2: Full Model Diagnostic Plot***

From the Residuals vs Fitted, the residuals are centered around 0 with an obvious outlier for case 1353, and the standardized residuals showed constant variance. The Q-Q Residuals plot has a bit of heavy tailing deviating from normality, suggesting the need for a Box-Cox transformation later on. Finally, there is one bad leverage point in the case of company 1353, which we analyzed further below.

**Transformation:**

To alleviate the normality issue exposed by the diagnosis plots, we wanted to use Box-Cox transformation on both our response and predictor variables. However, since the Box-Cox transformation requires all values to be strictly positive, we had to add the minimum value of each variable plus one to each value so that all values are strictly positive.

The subsequent Box-Cox transformation suggests transforming the original variables by raising them to different powers. Here, we encountered a problem where negative values within a vector will produce NaNs when raised to a non-integer power. As a result, we rounded the suggested powers to their nearest integers and raised the variables accordingly (ex: 0.72 → 1, 0.29 → log transformation, etc.). We did not transform the 'group' predictor as it is a binary categorical variable, and transforming such a variable would make no sense.

```
## bcPower Transformations to Multinormality
##                              Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Mean.Hourly.Gap                 0.7315        0.73       0.6343       0.8287
## Report.Year                     3.0009        3.00          NaN          NaN
## Mean.Bonus.Gap                  0.7194        0.72       0.6869       0.7518
## Median.Bonus.Gap                6.4964        6.50       6.1399       6.8529
## Mean.Hourly.Gap.Part.Temp       1.6799        1.68       1.5515       1.8083
## Mean.Hourly.Gap.Part.Time       2.2307        2.23       2.0707       2.3906
## Percentage.Bonus.Paid.Female    0.2902        0.29       0.2573       0.3231
## Percentage.BIK.Paid.Female     -0.0745       -0.07      -0.1057      -0.0432
## Percentage.Employees.Female     0.8677        0.87       0.7885       0.9470
## group                          -2.3547       -2.35      -2.5915      -2.1179
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                                           LRT df       pval
## LR test, lambda = (0 0 0 0 0 0 0 0 0 0) 16135.31 10 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                                           LRT df       pval
## LR test, lambda = (1 1 1 1 1 1 1 1 1 1) 8747.803 10 < 2.22e-16
```

***Figure 2: Box-Cox Transformation Suggestions***

We fit another linear model with the transformed variables and obtained the following summary (Figure 3). Compared to the original full model, the transformed model is notably worse. Not only were there fewer significant predictors, but the $R^2$ was also lower than the original model. Its diagnostic plots are also not promising (Graph 3), as both the Q-Q plot and Scale-Location plot look identical to that of the original model. We suspect that the compromise necessary to make the transformation possible eliminated the correction effect, hence no improvement in the goodness of fit. Since the transformation did not yield a better-fitted model, we decided to move forward with the original full model.

```
tMHG <- Mean.Hourly.Gap^1
tRY <- Report.Year^3
tMBG <- Mean.Bonus.Gap^1
tMeBG <- Median.Bonus.Gap^7
tMHGPTe <- `Mean.Hourly.Gap.Part.Temp`^2
tMHGPTi <- `Mean.Hourly.Gap.Part.Time`^2
tPBPF <- log(`Percentage.Bonus.Paid.Female`+0.00001)
tPBIKPF <- log(`Percentage.BIK.Paid.Female`+0.00001)
tPEF <- `Percentage.Employees.Female`
tGroup <- group
```

***Figure 3: Transformed Variable Values***

```
##
## Call:
## lm(formula = tMHG ~ tRY + tMBG + tMeBG + tMHGPTe + tMHGPTi +
##     tPBPF + tPBIKPF + tPEF)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -89.654  -6.504  -1.230   5.884  52.843
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.746e+02  2.090e+02   2.271   0.0233 *
## tRY         -5.669e-08  2.524e-08  -2.246   0.0248 *
## tMBG         6.354e-02  4.401e-03  14.438  < 2e-16 ***
## tMeBG       -8.316e-21  1.826e-20  -0.455   0.6489
## tMHGPTe      3.959e-04  1.844e-04   2.147   0.0319 *
## tMHGPTi      5.308e-05  7.775e-05   0.683   0.4949
## tPBPF        4.197e-01  5.222e-02   8.038 1.58e-15 ***
## tPBIKPF      2.069e-01  5.125e-02   4.036 5.64e-05 ***
## tPEF         1.048e-01  1.427e-02   7.342 3.08e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.2 on 1936 degrees of freedom
## Multiple R-squared:  0.2024, Adjusted R-squared:  0.1991
## F-statistic: 61.41 on 8 and 1936 DF,  p-value: < 2.2e-16
```

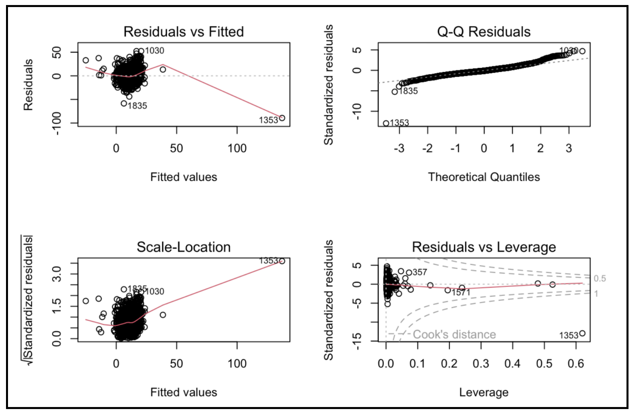



***Figure 4 & Graph 3: Model Summary and Diagnostic Plots of Transformed Model***

**<u>Variable Selection:</u>**

While our initial variables do not display a high correlation with each other, we still wanted to see if any predictor variables should be removed. All subsets, and forward and backward selections suggest that the full model is the best (see Table 2, Figure 5, and Table 3), and this is not surprising as the VIF analysis also shows that none of the predictors have a VIF > 5, suggesting that the full model does not suffer from multicollinearity (see Figure 6).

| ## | Rad | AIC | AICc | BIC |
|---|---|---|---|---|
| ## M1 | 0.08923467 | 9648.482 | 9648.495 | 9659.628 |
| ## M2 | 0.16856970 | 9472.218 | 9472.239 | 9488.937 |
| ## M3 | 0.18935329 | 9423.978 | 9424.009 | 9446.270 |
| ## M4 | 0.20460492 | 9388.034 | 9388.077 | 9415.899 |
| ## M5 | 0.21811170 | 9355.719 | 9355.777 | 9389.157 |
| ## M6 | 0.23280873 | 9320.804 | 9320.878 | 9365.388 |
| ## M7 | 0.23280873 | 9320.804 | 9320.897 | 9365.388 |
| ## M8 | 0.23489560 | 9316.502 | 9316.615 | 9366.659 |
| ## M9 | 0.23630163 | 9313.919 | 9314.055 | 9369.649 |

***Table 2: Selection Criteria of All Possible Subsets***

```
## Step:  AIC=9316.5
## Mean.Hourly.Gap ~ Mean.Bonus.Gap + Percentage.Bonus.Paid.Female +
##     Mean.Hourly.Gap.Part.Time + Percentage.Employees.Female +
##     Mean.Hourly.Gap.Part.Temp + Percentage.BIK.Paid.Female +
##     Median.Bonus.Gap + group
##
##                Df Sum of Sq    RSS    AIC
## + Report.Year  1    545.55 231279 9313.9
## <none>                     231825 9316.5
##
## Step:  AIC=9313.92
## Mean.Hourly.Gap ~ Mean.Bonus.Gap + Percentage.Bonus.Paid.Female +
##     Mean.Hourly.Gap.Part.Time + Percentage.Employees.Female +
##     Mean.Hourly.Gap.Part.Temp + Percentage.BIK.Paid.Female +
##     Median.Bonus.Gap + group + Report.Year
```

***Figure 5: Portion of Forward Selection Output***

```
## Start:  AIC=9313.92
## Mean.Hourly.Gap ~ Report.Year + Mean.Bonus.Gap + Median.Bonus.Gap +
##     Mean.Hourly.Gap.Part.Time + Mean.Hourly.Gap.Part.Temp + Percentage.Bonus.Paid.Female +
##     Percentage.BIK.Paid.Female + Percentage.Employees.Female +
##     group
##
##                                Df Sum of Sq    RSS    AIC
## <none>                                      231279 9313.9
## - Report.Year                   1     545.5 231825 9316.5
## - group                         1     734.5 232014 9318.1
## - Median.Bonus.Gap              1    1729.0 233008 9326.4
## - Percentage.BIK.Paid.Female    1    2127.5 233407 9329.7
## - Percentage.Employees.Female   1    4500.5 235780 9349.4
## - Mean.Hourly.Gap.Part.Temp     1    4547.9 235827 9349.8
## - Mean.Hourly.Gap.Part.Time     1    4962.2 236242 9353.2
## - Percentage.Bonus.Paid.Female  1    7752.6 239032 9376.0
## - Mean.Bonus.Gap                1   16801.6 248081 9448.3
```

***Table 3: Backward Selection Output***

```
##                Report.Year               Mean.Bonus.Gap
##                   1.005280                     1.191411
##           Median.Bonus.Gap    Mean.Hourly.Gap.Part.Time
##                   1.197570                     1.023238
##  Mean.Hourly.Gap.Part.Temp Percentage.Bonus.Paid.Female
##                   1.007753                     1.888046
## Percentage.BIK.Paid.Female  Percentage.Employees.Female
##                   1.805576                     1.095043
##                      group
##                   1.267227
```

***Figure 6: VIF of the Full Model***

**Discussion:**

Based on the summary output of the full model (Figure 1), we can conclude that several variables significantly impact the Mean Hourly Gap. The Mean Bonus Gap, Median Bonus Gap, Mean Hourly Gap for Part-Time and Temporary Employees, and the Proportion of Female Employees Receiving Bonuses are all highly significant ($p < 0.001$) and positively associated with the Mean Hourly Gap, indicating that workplaces with higher bonus gaps and female participation in pay structures tend to have a larger wage gap. The Report Year variable is slightly significant ($p = 0.032770$) with a negative coefficient (-0.6465), suggesting that the Mean Hourly Gap has slightly decreased over time. The group variable ($p = 0.013259$), which represents companies in specific industries (K, L, M, N, G, J), is positively associated with the wage gap, implying that these sectors tend to have higher disparities in pay, which is consistent with the official report.

The model has an adjusted $R^2$ of 0.2363, meaning it explains approximately 23.63% of the variance in the Mean Hourly Gap, indicating that while the included predictors are meaningful, other unaccounted factors contribute to wage differences. This is due to starting with predictors that already had low individual $R^2$

values. When we removed highly correlated variables to reduce multicollinearity, the model lost shared variance, further lowering $R^2$.

From the diagnostic plots (Figure 2), the Residuals vs Fitted has a mean centered around 0, and the standardized residuals show constant variance. From the Q-Q plot, there is heavy tailing so our fitted values deviate from normality, especially the lower end, but we determined that it was because of a few specific observations, such as case 1353. If we exclude that point and consider that our sample number is large (n = 1945), the normality issue is not a big concern. There is also a bad leverage point in case 1353, which can be seen in every diagnostic plot.

We observed that the company associated with observation 1353 is a Mining and Quarrying company by the name of Breedon. According to their reported data, Breedon has exhibited notably high gender pay gaps over recent years, which matches our dataset. A key factor contributing to this extreme gap is the significant gender imbalance, with only 13.7% of employees being female compared to 86.55% male. As a result, Breedon emerges as a strong statistical outlier, inflating overall variance and skewing the dataset, which helps explain why removing highly correlated variables led to an even lower $R^2$. Their case underscores structural gender inequalities in traditionally male-dominated industries and highlights the impact of outliers on statistical modeling.

**Final Model and Conclusion:**

In the end, we finalized our decision to keep the full model and yielded the equation:

$$\mathbf{Mean.\widehat{Hourly}.Gap} = 1,308 - 0.6465(Report.Year) + 5.5455 * 10^{-2}(Mean.Bonus.Gap) + 2.071 * 10^{-2}(Median.Bonus.Gap) + 5.434 * 10^{-2}(Mean.Hourly.Gap.Part.Time) + 8.436 * 10^{-2} + 8.426 * 10^{-2}(Mean.Hourly.Gap.Part.Temp) + 6.811 * 10^{-2}(Percentage.Bonus.Paid.Female) + 3.482 * 10^{-2}(Percentage.BIK.Paid.Female) + 8.622 * 10^{-2}(Percentage.Employees.Female) + 1.433(group)$$

Our final model indicates that besides Report.Year, all other variables result in slight increases towards Mean.Hourly.Gap, which collectively adds up to increase Mean.Hourly.Gap. Each variable influences the Mean.Hourly.Gap positively by a small amount per unit increase, and only Report.Year lowers the Mean.Hourly.Gap by 0.05434 per unit.

The main limitations regarding our project model analysis was the prevalence of NA variables. Due to incomplete information from the data set, we had to impute such values with the column's mean, which may fail to represent a given company's actual values accurately due to the variety of industry types included in the model. In addition to the influence of estimating NA values, the $R^2$ value of 0.2398 proves to be underwhelming when it comes to explanation regarding Mean.Hourly.Gap by the model.

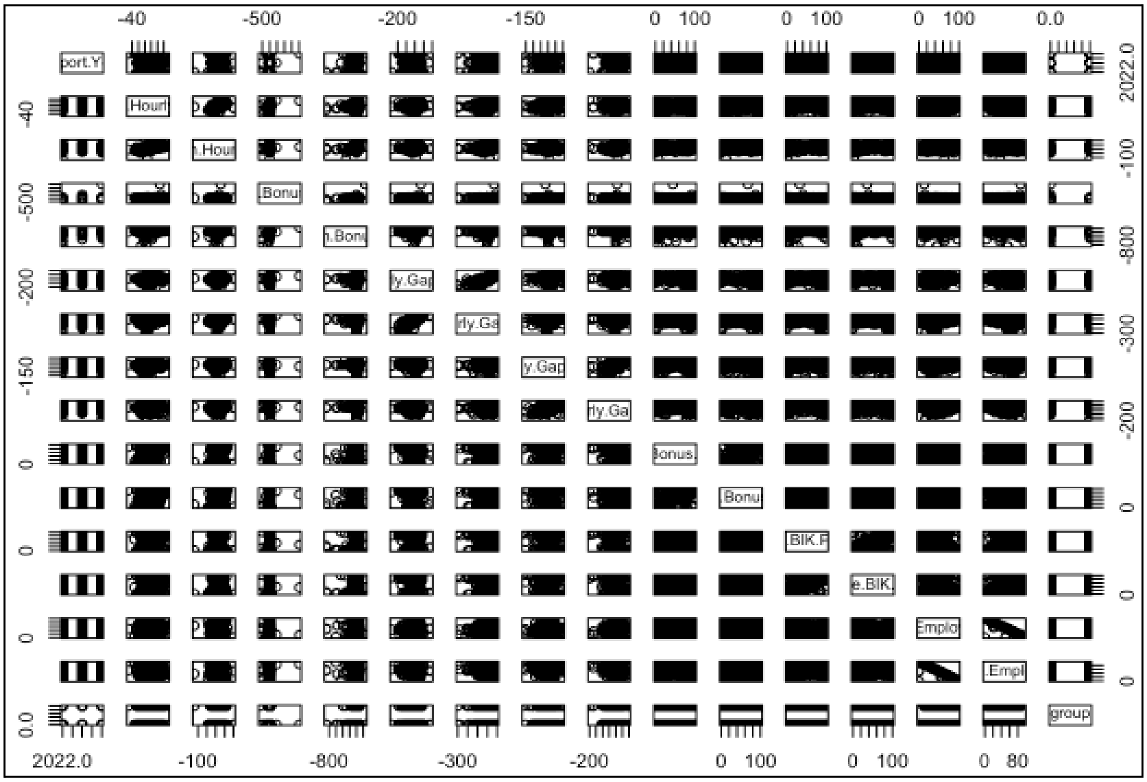
These findings, however, still provide good insight into the distribution of hourly gender wage gaps among Ireland's workers and help in the process of minimizing pay gaps and promoting equality. By utilizing the model, companies are able to strive to uphold equality standards and to promote equal growth with every one of their employees. Policy makers can also extract factors from this model that are affecting the pay gap and develop corresponding policies to mitigate the impact of these factors, further diminishing the discrepancy between the wages of Irish female and male workers.

## Appendix:

- Correlation Pairs Table for Original Data

```
new_table <- data %>% dplyr::select(-c(`Company.Name`, `NACE.Section`, `NACE.Letter`, `NACE.Division`, `Q1.Female`, `Q2.Female`, `Q3.Female`, `Q4.Female`, `Q1.Men`, `Q2.Male`, `Q3.Male`, `Q4.Male`))
```

```
pairs(new_table)
```




From the pairs table, we decided to remove the highly correlated variables `Median.Hourly.Gap`, `Median.Hourly.Gap.Part.Time`, `Median.Hourly.Gap.Part.Temp`, `Percentage.Bonus.Paid.Male`, `Percentage.BIK.Paid.Male`, and `Percentage.Employees.Male`.

- Additional Plot Outputs
  - Variable Selection (all subsets):

```
X <- cbind(Report.Year, Mean.Bonus.Gap, Median.Bonus.Gap, Mean.Hourly.Gap.Part.Time, Mean.Hourly.Gap.Part.Temp, Percentage.BIK.Paid.Female, Percentage.Bonus.Paid.Female, Percentage.Employees.Female, group)
b <- regsubsets(as.matrix(X), Mean.Hourly.Gap, nvmax = 9)
summary(b)
```

```
## Subset selection object
## 9 Variables  (and intercept)
##                              Forced in Forced out
## Report.Year                      FALSE      FALSE
## Mean.Bonus.Gap                   FALSE      FALSE
## Median.Bonus.Gap                 FALSE      FALSE
## Mean.Hourly.Gap.Part.Time        FALSE      FALSE
## Mean.Hourly.Gap.Part.Temp        FALSE      FALSE
## Percentage.BIK.Paid.Female       FALSE      FALSE
## Percentage.Bonus.Paid.Female     FALSE      FALSE
## Percentage.Employees.Female      FALSE      FALSE
## group                            FALSE      FALSE
## 1 subsets of each size up to 9
## Selection Algorithm: exhaustive
##          Report.Year Mean.Bonus.Gap Median.Bonus.Gap Mean.Hourly.Gap.Part.Time
## 1  ( 1 ) " "         "*"            " "              " "
## 2  ( 1 ) " "         "*"            " "              " "
## 3  ( 1 ) " "         "*"            " "              "*"
## 4  ( 1 ) " "         "*"            " "              "*"
## 5  ( 1 ) " "         "*"            " "              "*"
## 6  ( 1 ) " "         "*"            " "              "*"
## 7  ( 1 ) " "         "*"            "*"              "*"
## 8  ( 1 ) " "         "*"            "*"              "*"
## 9  ( 1 ) "*"         "*"            "*"              "*"
##          Mean.Hourly.Gap.Part.Temp Percentage.BIK.Paid.Female
## 1  ( 1 ) " "                       " "
## 2  ( 1 ) " "                       " "
## 3  ( 1 ) " "                       " "
## 4  ( 1 ) " "                       " "
## 5  ( 1 ) "*"                       " "
## 6  ( 1 ) "*"                       "*"
## 7  ( 1 ) "*"                       "*"
## 8  ( 1 ) "*"                       "*"
## 9  ( 1 ) "*"                       "*"
##          Percentage.Bonus.Paid.Female Percentage.Employees.Female group
## 1  ( 1 ) " "                          " "                         " "
## 2  ( 1 ) "*"                          " "                         " "
## 3  ( 1 ) "*"                          " "                         " "
## 4  ( 1 ) "*"                          "*"                         " "
## 5  ( 1 ) "*"                          "*"                         " "
## 6  ( 1 ) "*"                          "*"                         " "
## 7  ( 1 ) "*"                          "*"                         " "
## 8  ( 1 ) "*"                          "*"                         "*"
## 9  ( 1 ) "*"                          "*"                         "*"
```

- <u>Shifted Tables</u>

```
shifted_table <- new_table %>%
  mutate(`Mean.Hourly.Gap` = `Mean.Hourly.Gap` + 1 - min(`Mean.Hourly.Gap`)) %>%
  mutate(`Mean.Bonus.Gap` = `Mean.Bonus.Gap` + 1 - min(`Mean.Bonus.Gap`)) %>%
  mutate(`Median.Bonus.Gap` = `Median.Bonus.Gap` + 1 - min(`Median.Bonus.Gap`)) %>%
  mutate(`Mean.Hourly.Gap.Part.Time` = `Mean.Hourly.Gap.Part.Time` + 1 - min(`Mean.Hourly.Gap.Part.Time`)) %
>%
  mutate(`Mean.Hourly.Gap.Part.Temp` = `Mean.Hourly.Gap.Part.Temp` + 1 - min(`Mean.Hourly.Gap.Part.Temp`)) %
>%
  mutate(`Percentage.Bonus.Paid.Female` = `Percentage.Bonus.Paid.Female` + 1) %>%
  mutate(`Percentage.BIK.Paid.Female` = `Percentage.BIK.Paid.Female` + 1) %>%
  mutate(`Percentage.Employees.Female` = `Percentage.Employees.Female` + 1) %>%
  mutate(group = group + 1)
```

- <u>Full Model VIF</u>

```
##                     Report.Year               Mean.Bonus.Gap
##                        1.005280                     1.191411
##                Median.Bonus.Gap     Mean.Hourly.Gap.Part.Time
##                        1.197570                     1.023238
##       Mean.Hourly.Gap.Part.Temp Percentage.Bonus.Paid.Female
##                        1.007753                     1.888046
##      Percentage.BIK.Paid.Female  Percentage.Employees.Female
##                        1.805576                     1.095043
##                           group
##                        1.267227
```

- References & Citations:

  - PayGap.ie Dataset Download Link: https://paygap.ie/downloads
  - Breedon Group PLC Pay Gap Data: https://paygap.ie/company/breedon
  - Gender Pay Gap Structure Analysis of Earnings Survey 2022: https://www.cso.ie/en/releasesandpublications/ep/p-ses/structureofearningssurvey2022/genderpaygap/