Gender Pay Gap Analysis

Anshika Rai

2024-06-08

Introduction

The gender pay gap remains a significant issue within the labor market, reflecting deep-rooted inequities that persist across industries, education levels, and job titles. Despite numerous initiatives and ongoing societal discourse aiming to address gender equality, the disparities in compensation remain a stark reality of the workplace. Motivated by a commitment to further understand and contribute to the global conversation on gender equality, this report utilizes the comprehensive Glassdoor Gender Pay Gap dataset to delve into how these disparities manifest in various professional settings and demographic segments.

My objective is to uncover underlying trends that might explain the persistence of the pay gap and to provide actionable insights that can help mitigate these disparities. By analyzing this data, we seek to answer critical questions such as: **How does the gender pay gap vary by industry and job title? Are higher education levels associated with a reduced pay gap? What roles do factors like seniority and performance evaluations play in influencing gender disparities in pay?** The answers to these questions are crucial for developing targeted strategies that organizations and policymakers can implement to promote gender pay equity.

Dataset Overview and Preliminary Data Cleaning

The dataset comprises detailed employee information from Glassdoor, including gender, job title, department, age, performance evaluations, education level, seniority, base pay, and bonuses. This rich dataset allows for an in-depth exploration of pay disparities across different segments of the workforce.

```r
# required libraries
library(ggplot2)
library(dplyr)
library(readr)
library(tidyr)
library(plotly)



# Load the dataset
pay_gap_data <- read_csv("Glassdoor-Gender-Pay-Gap.csv")

# Detecting and handling outliers
pay_gap_data <- pay_gap_data %>%
  mutate(BasePay_Z = (BasePay - mean(BasePay)) / sd(BasePay)) %>%
  filter(abs(BasePay_Z) < 3) # Filtering out extreme outliers
```

The initial cleaning process involvs removing incomplete entries and converting categorical data into factors to facilitate accurate analyses. This step ensures our dataset is robust and reliable for drawing meaningful conclusions about gender-based pay inequalities.

```r
# Data Inspection
glimpse(pay_gap_data)
```

```
## Rows: 998
## Columns: 10
## $ JobTitle  <chr> "Graphic Designer", "Software Engineer", "Warehouse Associat~
## $ Gender    <chr> "Female", "Male", "Female", "Male", "Male", "Female", "Femal~
## $ Age       <dbl> 18, 21, 19, 20, 26, 20, 20, 18, 33, 35, 24, 18, 19, 30, 35, ~
## $ PerfEval  <dbl> 5, 5, 4, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, ~
## $ Education <chr> "College", "College", "PhD", "Masters", "Masters", "PhD", "C~
## $ Dept      <chr> "Operations", "Management", "Administration", "Sales", "Engi~
## $ Seniority <dbl> 2, 5, 5, 4, 5, 4, 4, 5, 5, 5, 5, 3, 3, 5, 4, 3, 5, 5, 5, 5, ~
## $ BasePay   <dbl> 42363, 108476, 90208, 108080, 99464, 70890, 67585, 97523, 11~
## $ Bonus     <dbl> 9938, 11128, 9268, 10154, 9319, 10126, 10541, 10240, 9836, 9~
## $ BasePay_Z <dbl> -2.05662227, 0.55267294, -0.16831393, 0.53704393, 0.19699451~
```

```r
# Check for missing values
summary(pay_gap_data)
```

```
##    JobTitle            Gender               Age           PerfEval    
##  Length:998         Length:998         Min.   :18.00   Min.   :1.000  
##  Class :character   Class :character   1st Qu.:29.00   1st Qu.:2.000  
##  Mode  :character   Mode  :character   Median :41.00   Median :3.000  
##                                        Mean   :41.36   Mean   :3.038  
##                                        3rd Qu.:54.00   3rd Qu.:4.000  
##                                        Max.   :65.00   Max.   :5.000  
##   Education             Dept             Seniority        BasePay      
##  Length:998         Length:998         Min.   :1.000   Min.   : 34208  
##  Class :character   Class :character   1st Qu.:2.000   1st Qu.: 76821  
##  Mode  :character   Mode  :character   Median :3.000   Median : 93313  
##                                        Mean   :2.968   Mean   : 94305  
##                                        3rd Qu.:4.000   3rd Qu.:111388  
##                                        Max.   :5.000   Max.   :165229  
##      Bonus         BasePay_Z       
##  Min.   : 1703   Min.   :-2.378477  
##  1st Qu.: 4852   1st Qu.:-0.696671  
##  Median : 6507   Median :-0.045768  
##  Mean   : 6469   Mean   :-0.006627  
##  3rd Qu.: 8032   3rd Qu.: 0.667592  
##  Max.   :11293   Max.   : 2.792555  
```

```r
# Remove rows with any NA values
pay_gap_data <- pay_gap_data %>%
  filter(!is.na(JobTitle), !is.na(Gender), !is.na(Age), !is.na(PerfEval),
         !is.na(Education), !is.na(Dept), !is.na(Seniority), !is.na(BasePay), !is.na(Bonus))

# Convert necessary columns to factors
pay_gap_data$Gender <- as.factor(pay_gap_data$Gender)
pay_gap_data$Dept <- as.factor(pay_gap_data$Dept)
pay_gap_data$JobTitle <- as.factor(pay_gap_data$JobTitle)
```

Summary Statistics:

In our analysis of the Glassdoor Gender Pay Gap dataset, we examined several key metrics across different demographics and professional criteria. Here's an overview of the primary attributes included in our dataset:

1. **Gender:** Consists of male and female categories.
2. **Education Level:** Ranges from high school to PhD.
3. **Job Title:** Includes a variety of positions from Data Scientist to Warehouse Associate.
4. **Department:** Encompasses Administration, Engineering, Management, Operations, and Sales.
5. **Seniority:** Measured in years, from 1 to 5.
6. **Base Pay and Bonuses:** Quantified in monetary values.

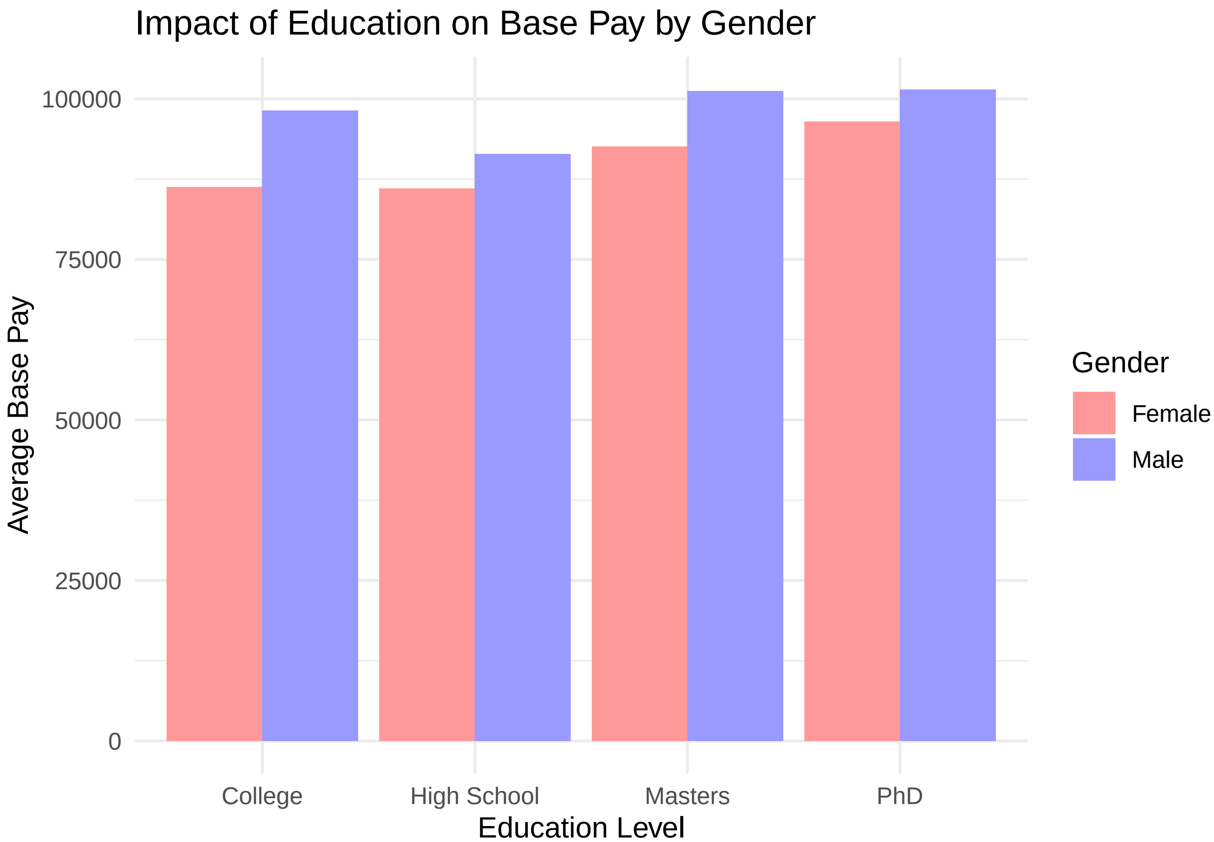
The dataset consists of several thousand entries, ensuring a robust analysis of gender disparities in pay across various sectors.

Data Visualization

- Impact of Education on Base Pay by Gender

Bar chart with education levels on the x-axis and average base pay on the y-axis, distinguished by color for each gender. I used this chart as Bar charts are ideal for comparing the sizes of different groups. This chart clearly shows the pay differences between genders across various education levels, highlighting the increasing disparity as educational attainment rises.

```
# Impact of Education on Base Pay by Gender
ggplot(pay_gap_data, aes(x = Education, y = BasePay, fill = Gender)) +
  geom_bar(stat = "summary", fun = "mean", position = "dodge") +
  labs(title = "Impact of Education on Base Pay by Gender", x = "Education Level", y = "Average Base Pa
  scale_fill_manual(values = c("#FF9999", "#9999FF")) +
  theme_minimal()
```

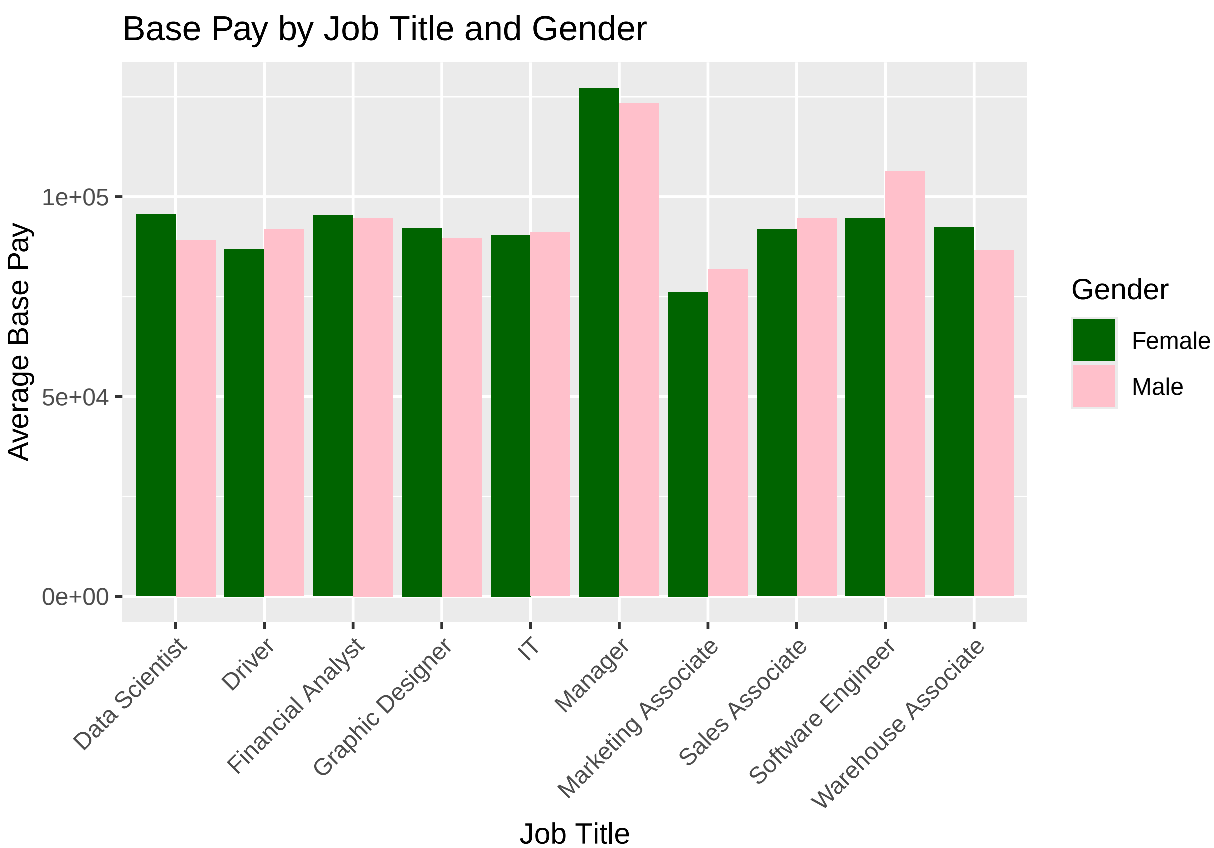


Analysis shows that higher education levels generally correlate with higher base pay across both genders. However, the pay gap widens at higher educational tiers, with men consistently earning more than women, particularly noticeable at the PhD level.

- Base Pay by Job Title and Gender

Bar chart with job titles on the x-axis and average base pay on the y-axis, color-coded by gender. This visualization helps compare the average base pay for males and females across different job titles. It is particularly effective in showing which positions have the most significant gender pay gaps, allowing for easy comparison across categories.

```
# Base Pay by Job Title and Gender
ggplot(pay_gap_data, aes(x = JobTitle, y = BasePay, fill = Gender)) +
  geom_bar(stat = "summary", fun = "mean", position = "dodge") +
  labs(title = "Base Pay by Job Title and Gender", x = "Job Title", y = "Average Base Pay") +
  scale_fill_manual(values = c("darkgreen", "pink")) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate x labels for readability
```



The graph illustrates a positive change where women are earning more than men in prominent positions such as 'Manager' and 'Data Scientist.' This is a notable shift towards addressing the gender pay gap in high-responsibility roles. However, the data also reveals persistent disparities in other job titles, underscoring the ongoing challenges in achieving comprehensive pay equity across all positions.

- Base Pay Distribution by Department and Gender

Stacked area chart showing total base pay distribution across different departments, segmented by gender. Area charts are useful for understanding how different segments stack up over a particular variable—in this case, departments. This chart provides a visual representation of the cumulative base pay, emphasizing disparities in total compensation between genders within departments.

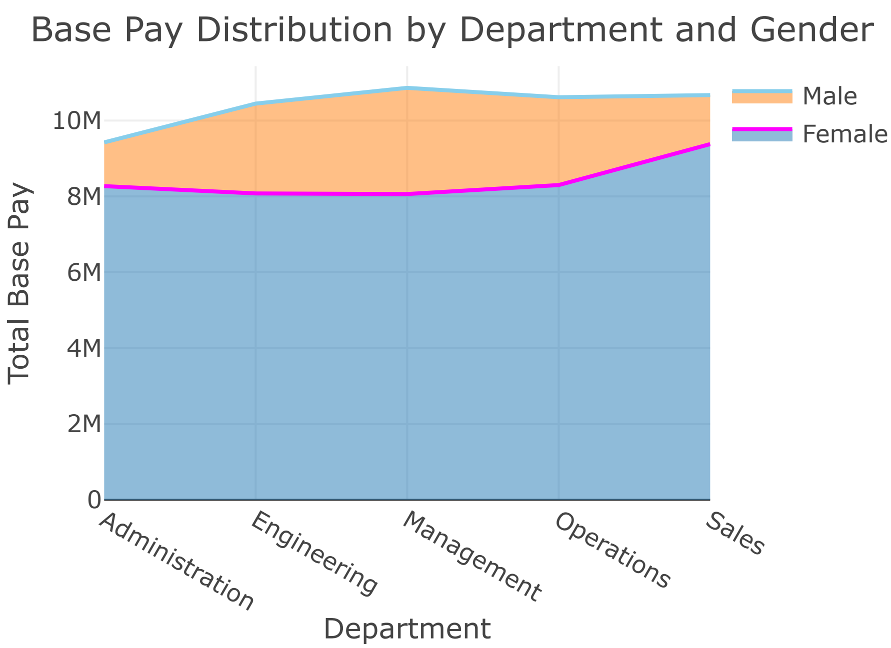
```
# Aggregate base pay by department and gender
pay_summary <- pay_gap_data %>%
  group_by(Dept, Gender) %>%
  summarise(TotalBasePay = sum(BasePay), .groups = 'drop') %>%
```

```r
  arrange(Dept, Gender)

# Create a wider format for Plotly area plot
pay_wide <- pay_summary %>%
  pivot_wider(names_from = Gender, values_from = TotalBasePay)

# Check if any NA and replace them with 0
pay_wide[is.na(pay_wide)] <- 0

# Create the Plotly stacked area plot
plot_ly(pay_wide, x = ~Dept, y = ~`Female`, name = 'Female', type = 'scatter', mode = 'lines', fill = '
  add_trace(y = ~`Male`, name = 'Male', mode = 'lines', fill = 'tonexty', line = list(color = 'skyblue'
  layout(title = 'Base Pay Distribution by Department and Gender',
         xaxis = list(title = 'Department'),
         yaxis = list(title = 'Total Base Pay'),
         hovermode = 'compare')
```

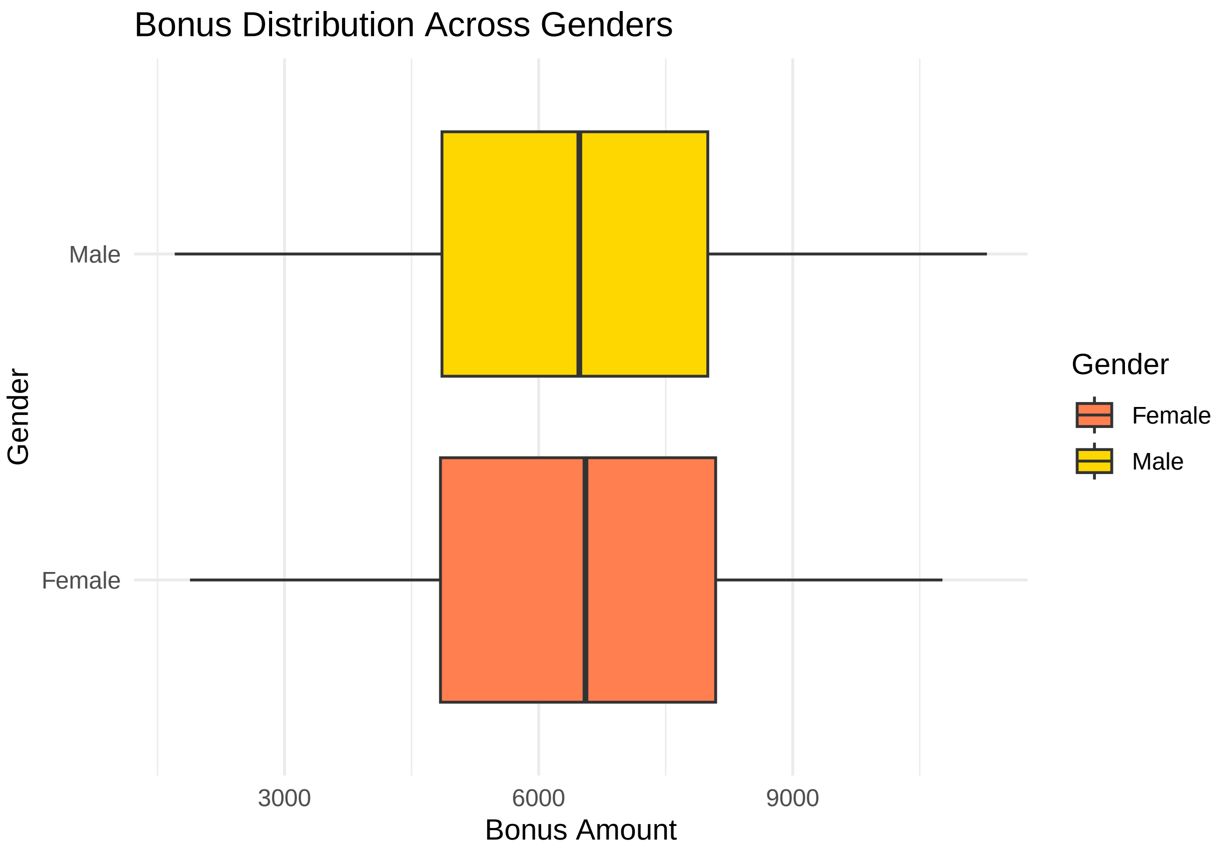


Our findings indicate that the pay gap extends across different departments. The stacked area plot illustrates that while both genders see an increase in total base pay in departments such as Engineering and Management, the proportion of pay received by females is substantially lower compared to their male colleagues.

- Bonus Distribution Across Genders

Horizontal box plot displaying the distribution of bonuses received by each gender. Box plots are excellent for showing the distribution of data based on a five-number summary (minimum, first quartile, median, third quartile, and maximum). This chart was chosen to compare the central tendency and variability of bonuses between genders, highlighting outliers and overall spread.

```r
# Bonus Distribution Across Genders - Horizontal Boxplot
ggplot(pay_gap_data, aes(x = Gender, y = Bonus, fill = Gender)) +
  geom_boxplot() +
  coord_flip() + # This flips the x and y axes making the plot horizontal
  labs(title = "Bonus Distribution Across Genders", x = "Gender", y = "Bonus Amount") +
  scale_fill_manual(values = c("coral", "gold")) +
  theme_minimal() # Adding minimal theme for cleaner look
```

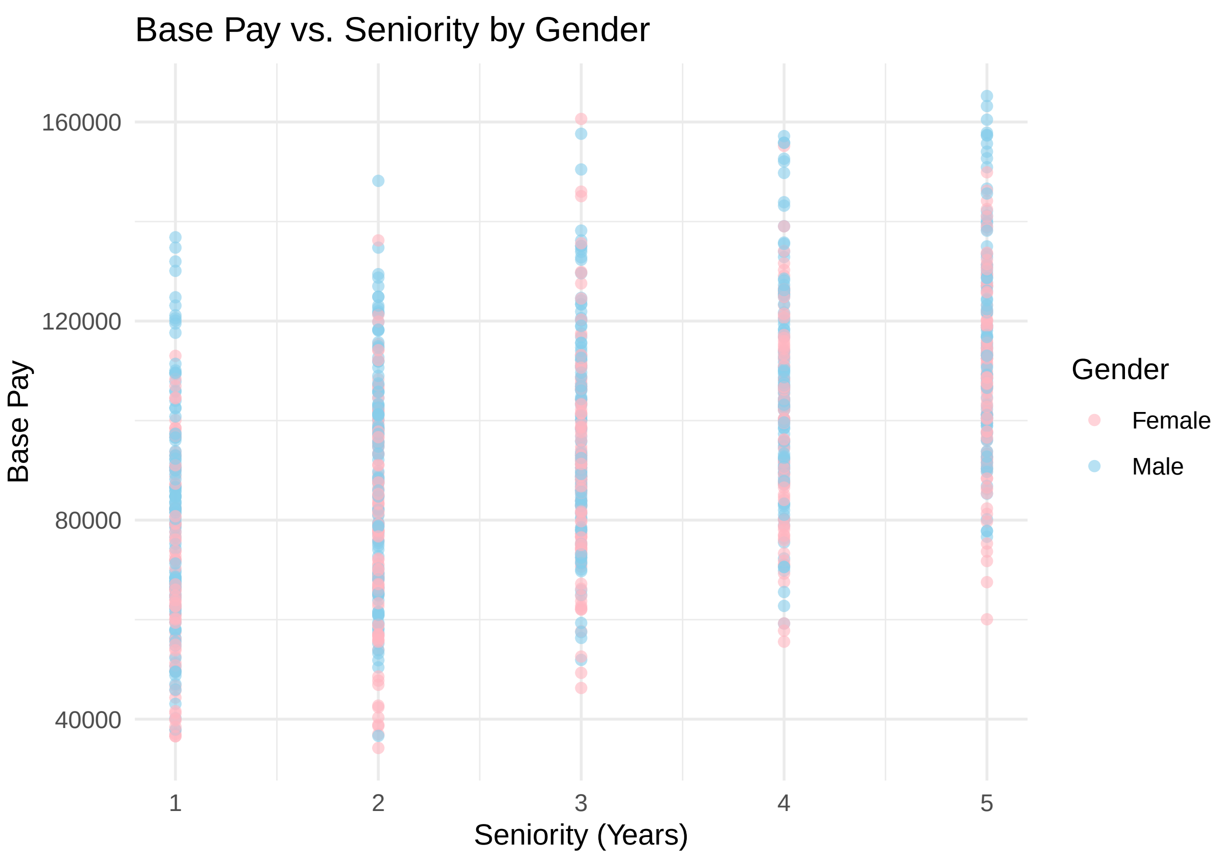


The analysis of bonus payments further underscores the gender pay gap, with males receiving higher bonuses across nearly all categories. This disparity suggests that gender biases might extend into the reward systems within organizations.

- Scatter Plot of Base Pay vs. Seniority by Gender

Scatter plot with seniority on the x-axis and base pay on the y-axis, differentiated by color for each gender. Scatter plots are effective for examining the relationship between two quantitative variables. This plot shows how base pay correlates with years of seniority for each gender, illustrating the trajectory of pay progression over time.

```
# Scatter Plot of Base Pay vs. Seniority by Gender
ggplot(pay_gap_data, aes(x = Seniority, y = BasePay, color = Gender)) +
  geom_point(alpha = 0.6) +
  labs(title = "Base Pay vs. Seniority by Gender", x = "Seniority (Years)", y = "Base Pay") +
  scale_color_manual(values = c("lightpink", "skyblue")) +
  theme_minimal()
```



The scatter plot reveals that although base pay increases with seniority for both genders, males consistently have a higher pay trajectory compared to females with equivalent years of experience.

- Performance vs. Base Pay with Bonus by Gender

Scatter plot with performance evaluation scores on the x-axis and total compensation on the y-axis, color-coded by gender.This visualization is used to analyze the relationship between performance scores and total compensation. It helps identify patterns of compensation equity or disparity, particularly useful for assessing if higher performance correlates with better pay, and whether this trend is consistent across genders.

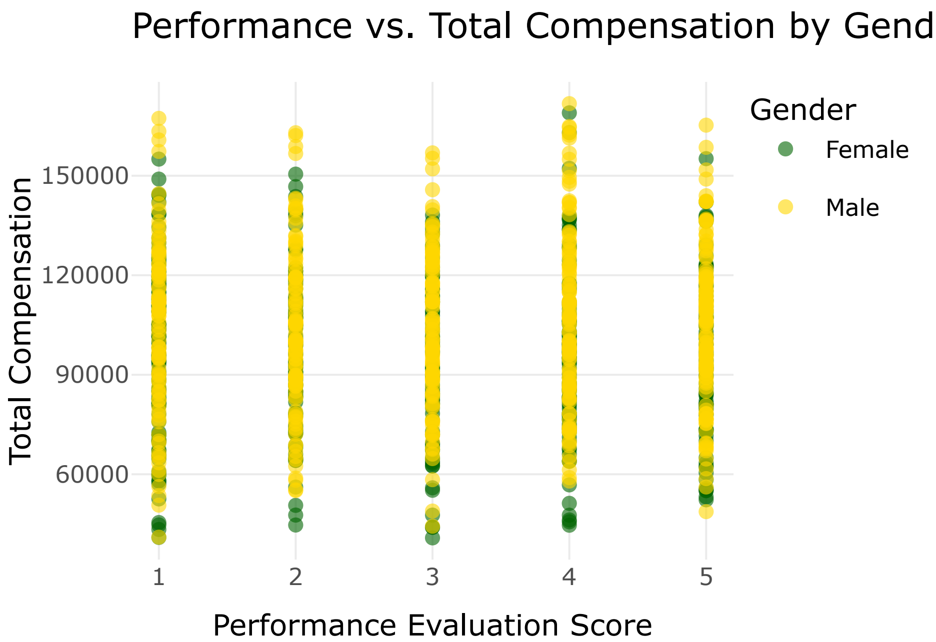
```r
library(plotly)

# Calculate total compensation including bonuses
pay_gap_data$TotalCompensation <- pay_gap_data$BasePay + pay_gap_data$Bonus

# Create a ggplot
p <- ggplot(pay_gap_data, aes(x = PerfEval, y = TotalCompensation, color = Gender)) +
  geom_point(alpha = 0.6) +
  labs(title = "Performance vs. Total Compensation by Gender",
       x = "Performance Evaluation Score", y = "Total Compensation") +
  scale_color_manual(values = c("darkgreen", "gold")) +
  theme_minimal()

# Convert ggplot to plotly
plotly_p <- ggplotly(p)

# Show the plot
plotly_p
```



Interestingly, the relationship between performance evaluations and total compensation shows that despite similar performance scores, females often receive less total compensation than males, highlighting a crucial area for policy review and adjustment.

Conclusion

This comprehensive analysis of the Glassdoor Gender Pay Gap dataset revealed persistent and significant disparities in pay across various factors such as gender, education level, job title, department, and seniority. Key findings include:

- Higher educational attainment generally correlates with higher pay for both genders, but the gender pay gap widens significantly at higher educational levels, especially at the PhD level.
- Male employees tend to receive higher base pay than their female counterparts across similar job roles, with larger gaps noted in high-responsibility and technical positions such as 'Manager' and 'Data Scientist'.
- Departmental analysis showed that while pay increases with seniority and within certain departments like Engineering and Management, the proportion of compensation received by females is consistently lower compared to males.
- Bonus distributions also reflect gender disparity, with males frequently receiving larger bonuses than females across most sectors.
- The analysis of base pay versus seniority and performance versus total compensation highlighted that despite similar levels of experience and performance scores, women generally receive less pay and slower pay progression.

Limitations:

- **Data Scope and Representativeness:** The dataset is limited to employees who have shared their pay details on Glassdoor, which may not comprehensively represent all industry sectors or geographical regions.
- **Causality and Correlations:** While this report identifies correlations between gender and pay disparities, it does not establish causality. Factors such as negotiation skills, maternity leave, and other socio-economic factors were not available in the dataset and could influence the gender pay gap.

- **Temporal Dynamics:** The dataset does not cover changes over time, preventing analysis of trends that could inform whether the gender pay gap is widening or narrowing.

Recommendations

Actionable Steps:

- **Regular Pay Audits:** Organizations should conduct regular pay audits to ensure equitable compensation across all levels and positions. Discrepancies should be corrected, and the criteria for pay scales should be transparent.
- **Revise Performance Evaluations:** To mitigate unconscious bias, companies should standardize performance evaluations and ensure they are tied directly to pay increases and bonuses. This process should be regularly reviewed by a diverse committee.
- **Promote Pay Transparency:** Companies should adopt and promote pay transparency to help reduce the gender pay gap. Open discussions about salary ranges and remuneration policies can empower employees to understand and advocate for fair pay.
- **Career Development Programs:** Implement targeted career development programs for women, particularly in industries where they are underrepresented or in higher-paying technical roles. This includes mentorship programs, sponsorship, and professional training.
- **Policy Advocacy:** Encourage policymakers to examine and enforce equal pay legislation rigorously. Companies can also advocate for policies that support work-life balance, such as flexible working conditions, which can help retain high-performing employees regardless of gender.

Citation

Jauhari, Neelima. "Glassdoor- Analyze Gender Pay Gap." Kaggle, https://www.kaggle.com/datasets/nilimajauhari/glassdoor-analyze-gender-pay-gap.ca