IS507 PROJECT FINAL REPORT

Gender Pay Gap Analysis Using Statistical and Predictive Modeling

Suruthe Jayachandran

1. Introduction

Gender-based compensation differences are a persistent and widely discussed concern in organizational policy and public discourse. This project examines whether men and women receive equal pay for comparable roles and qualifications by applying statistical analysis and predictive modeling to assess the influence of factors such as job role, education, performance, and experience. The findings will be valuable for HR teams, executive leadership, DEI committees, compliance groups, and employees by helping identify where disparities exist and guiding fair salary and bonus review practices.

2. Research Questions

We have four major research questions to have a comprehensive understanding of the inequality of the job market.

RQ1: Is there evidence of the gender pay gap in general and across job titles, departments, and education levels?
We will analyze whether average salaries differ significantly between male and female employees across various job titles, functional departments, and education levels. This will identify whether pay disparities exist, and if so, assess the magnitude of inequality.

RQ2: Which organizational and employee factors best explain variations in the gender pay gap?
If salary or bonus disparities are observed, we will examine which factors contribute most to these differences. We will evaluate how employee characteristics (i.e. year of experience) and organizational factors (i.e. department) relate to pay inequality. This aims to find which groups or companies are more likely to exhibit significant gender pay gaps.

RQ3: Does higher education or performance evaluation mitigate the gender pay gap?
Our motivation is to identify the most effective ways to reduce the gender pay gap. Specifically, we will examine whether higher education levels and stronger performance evaluation scores are associated with smaller differences in salary and bonus between male and female employees. This investigates whether education and performance moderate pay inequality.

RQ4: Are women less likely to be among the top earners?
One possible confounding variable making inequality of salary is the less representation of women in higher-level positions which may influence access to top-paying roles. We aim to study whether female employees have lower odds of being in the top percentile of earners compared to their male counterparts.

3. Data Description

The dataset used in this project was collected from Glassdoor and contains detailed information related to compensation across various job roles. It includes nine key attributes: Job Title, Gender, Age, Performance Evaluation score, Education level, Department, Seniority, Base Pay, and Bonus. These features collectively offer a structured view of employee profiles and their annual compensation. Because the data spans multiple job categories, such as Software Engineer, Marketing Associate, Data Scientist, Graphic Designer, and more, it provides a broad representation of the workforce across different departments and educational backgrounds.

The primary motivation behind using this dataset is to investigate the presence and extent of gender-based pay gaps for similar roles and qualifications. Prior research highlights that women are often paid less than men for comparable work, making this dataset valuable for validating and quantifying such trends. With detailed salary components (Base Pay and Bonus) and relevant demographic and performance features, the dataset will enable us to perform statistical and predictive analysis of income inequality.

4. Methodology

Across different research questions, we employed different methodological approaches to conduct the analyses. For RQ1, we used *t*-tests and ANOVA to compare mean differences between female and male employees. As these are parametric statistical tests, we applied the Shapiro–Wilk test and Levene's test to assess normality and homogeneity of variance, respectively. When running the ANOVA while controlling for department and job title, we found that the normality assumption of the residuals was violated. Consequently, we employed the Scheirer–Ray–Hare test, a non-parametric alternative to ANOVA, to conduct the analysis. We used the Dunn test for post hoc comparisons.

For RQ2 and RQ3, we used multiple regression models for the analysis. We examined residuals versus fitted value plots for each model to assess linearity and homoscedasticity, and we used Q–Q plots to evaluate the normality of residuals. We also computed correlation matrices to assess potential collinearity among predictors.

For RQ4, we used logistic regression as our modeling approach. We constructed an additional predictive variable, *top earner*, which indicates whether an individual falls within the top 10th percentile of earners across the dataset.

5. Results

RQ1: Is there evidence of the gender pay gap in general and across job titles, departments, and education levels?

The initial objective of this analysis was to determine whether a gender-based salary discrepancy exists within the organization. To address this question, we first conducted an independent sample t-test comparing annual salaries of male and female employees. The results indicated a statistically significant difference in mean salary, with men earning more on average ($M = 104{,}375.24$) than women ($M = 96{,}261.48$), $t(993) = 5.249$, $p < .001$. The effect size was moderate (*Cohen's* $d = 0.33$, 95% CI [0.21, 0.46]), and the 95% confidence interval for the mean difference ranged from *5,080* to *11,147* USD.

At the high level, these findings suggested the presence of a gender pay gap. However, we need to be cautious of the validity of this raw comparison as salaries are strongly influenced by confounding variables such as job title, department, and educational level. To isolate the unique contribution of gender and avoid the Simpson's paradox, we conducted a series of two-way ANOVAs with salary as the dependent variable.

Analysis Controlling for Job Title
A two-way ANOVA including Gender and Job Title revealed that Gender was no longer a significant predictor of salary when job title was controlled, $F(1, 975) = 0.203, p = .652$. In contrast, Job Title showed a robust main effect, $F(9, 975) = 168.158, p < .001$, demonstrating substantial salary variation across roles. The Gender × Job Title interaction was non-significant *($p = .447$)*, meaning that the relationship between job title and compensation did not differ for men and women. The post-hoc comparisons indicated that managerial and technical positions (e.g., Manager, Software Engineer, Data Scientist) consistently earned significantly higher salaries than administrative, operational, or entry-level roles *(Result 8)*. These findings suggest that job roles are the primary driver of salary differences.

Analysis Controlling for Department
A similar pattern emerged when analyzing salary by department. While departments differed in average salary levels, these differences were unrelated to gender as well. Post hoc comparisons showed that certain departments (e.g., Engineering, Operations) offered higher salaries than others (*Result 12*), but there were no gender-based differences within any department.

Analysis Controlling for Education Level
A two-way ANOVA including Gender and Education Level found a significant main effect of Education on salary, $F(3, 987) = 9.896, p < .001$. Employees with higher degrees; particularly PhDs earned substantially more than those with lower levels of education (*Result 16*). However, Gender remained non-significant *($p = .388$)*, and the Gender × Education interaction was also non-significant.

RQ2: Which organizational and employee factors best explain variations in the gender pay gap?

The next question we aim to address is whether the pay gap between males and females is driven by industry-related factors or by differences among employees themselves. To answer this, we used linear regression to examine how organizational factors (department and job type) and employee factors (seniority, performance evaluation, and age) explain pay differences across genders.

We fitted three models to assess how these factors account for variance in salary:

Base_model:	Salary ~ Gender
Organization_model:	Salary ~ Gender + Department + Job Type
Employee_model:	Salary ~ Gender + Seniority + Performance + Age

We found that the Employee Model ($\Delta R^2 = 0.591$) explains substantially more variance than the Organization Model ($\Delta R^2 = 0.201$) . In the Organization Model, job type accounts for the most variance, while gender accounts for the least. In the Employee Model, age and seniority explain similar amounts of variance, and performance evaluation explains the least. (Result 18, 19)

In terms of effect sizes, when considering only employee-related factors, being male and having more years of experience have the strongest positive effects on salary, while performance evaluation has the weakest effect. In the Organization Model, being a manager or working in the sales department is associated with higher pay. Employees who are marketing associates earn the least compared to other job categories.

Employee Model

Coefficient	Effect Size
Being Male	9736.56***
Seniority	9759***
Performance	799.89*
Age	956.25***

Corporational Model

Coefficient	Effect Size
Software Engineer	12083.3***
Manager	30383.2***
Marketing Associate	-16974.4***
Sales Department	5130***

Table 1*: Presents the estimated effect sizes from two regression models examining determinants of employee compensation. The Employee Model (Top) includes individual-level predictors: gender, seniority, performance, and age. While the Corporational Model (Bottom) incorporates job-role and departmental indicators. Effect sizes are reported as coefficient estimates, with significance levels denoted by *** p < 0.01, ** p < 0.05, and * p < 0.1.*

RQ3: Does higher education or performance evaluation mitigate the gender pay gap?

To examine whether education or performance evaluation reduce gender-based salary differences, we fitted a full regression model that included both organizational factors (department, job title) and employee-level factors (seniority, age, education, performance evaluation), along with two interaction terms: Gender × Education and Gender × Performance.

We fitted the full model with all the factors from both organizational and employee-related factors with two interaction terms: *Gender x performance* and *Gender x education*

Full_mode: Salary ~ Gender + All Factors + Gender x performance + Gender x education

---

The results show that none of the Gender × Education interaction terms were statistically significant for high school (*$\beta = -1952.3$, $p = .284$*), master's degree (*$\beta = 490.8$, $p = .790$*), or PhD level (*$\beta = -3184.5$, $p = .088$) (Result 20).* Likewise, the Gender × Performance interaction was non-significant (*$\beta = -402.8$, $p = .376$*). This indicates that the salary returns to education and performance evaluation do not differ between men and women. In other words, we find no statistical evidence that obtaining higher education qualifications or achieving stronger performance evaluation scores reduces the gender pay gap. One thing is worth pointing out is that the marginal interaction effect for gender × PhD (*$p = .088$*) suggests a possible trend where the highest education level might slightly narrow the gap.

RQ4: Are women less likely to be among the top earners?

Salaries in most organizations follow a right-skewed distribution, where the majority of employees earn around the middle of the pay range, and only a small proportion earn very high salaries. Because of this skewed structure, analyzing average salary differences (as in

RQ1–RQ3) may miss an important dimension of inequality: who reaches the top of the salary distribution?

We dichotomize the salary into either top 10 percentile earners or not and use a logistic regression to study how the model predicts. From the logistic regression model, we found that male have significantly lower odds (65.3% lower) of being among the top earners than females, even after controlling for seniority, age, job title, department, education, and performance evaluation (*Result 20*).

As expected, seniority ($\beta = 1.85$, $p < .001$, OR = 6.33) and age ($\beta = 0.21$, $p < .001$, OR = 1.23) strongly increased the odds of being in the top earner group. Certain job titles such as Manager (OR = 468.93) and Software Engineer (OR = 19.24), and departments such as Engineering (OR = 7.30) and Sales (OR = 13.30), also substantially increased the likelihood of being a top earner. Higher educational qualifications (Master's OR = 3.57; PhD OR = 4.58) and higher performance evaluations (OR = 1.47) were additionally associated with improved odds.

6. Discussion

During RQ1, when we control for job titles and functional departments, the main effect of gender on pay disappears. This effect was significant in the baseline model without additional controls. This pattern suggests that differences in occupational roles may help explain the observed gender pay gap, rather than gender itself being the direct driver of pay differences. We will elaborate on this interpretation in the addressing-comments section.

Interestingly, when we instead control for educational level, the main effect of gender becomes significant again. This aligns with the workplace intuition: higher educational attainment is associated with higher pay, and certain job titles or departments also receive higher compensation. However, the insignificant interaction terms indicate that there is no evidence of a gender pay gap across titles or departments. By contrast, differences may appear among employees with the same educational level.

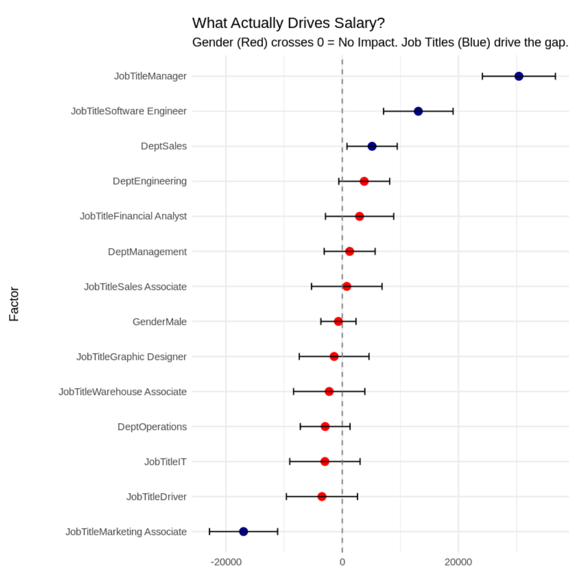

Figure 1*: Impact on Salary across different factors. Gender effects hover around zero, while job titles have the strongest impact.*

In RQ2, we compared the explanatory power of employee characteristics versus job-related factors in accounting for pay variance. We found that characteristics such as age, years of experience, and gender account for a substantially larger portion of the variance than job-related factors like job title or functional department. These findings provide additional evidence of a gender pay gap, as pay appears to be driven largely by education and seniority. Higher education and greater experience predict higher salaries, which is consistent with firms' willingness to pay more for specialized or experienced talent. One notable finding is that performance evaluations do not meaningfully explain pay variance. In other words, measured performance does not appear to translate into higher compensation. However, this may not reflect actual career dynamics, as employees often receive promotions or transfer to different departments; which processes that are not captured in this dataset.

RQ3 examines how the gender effect could be mitigated and whether education or performance moderates gender differences. Interaction terms in the full regression model show that gender is not dependent on either educational level or performance ratings; changes in these factors do not significantly alter the gender effect. The reduced model (excluding organizational variables) yields similar results. Even so, there is a slight trend toward decreasing p-values for the gender term as educational level increases, suggesting a possible moderating effect. Because this trend is not statistically robust, we treat this as a tentative observation that requires validation with a larger and more generalizable dataset.

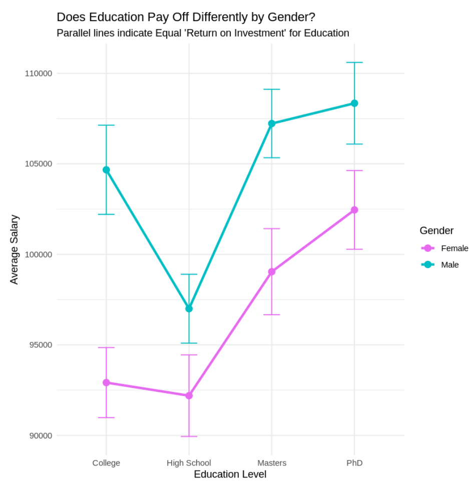

Figure 2*: Mean salaries across education levels. The nearly parallel lines show that returns to education are similar for both male and females. Thus, education does not widen the gender gap nor narrow it.*

For RQ4, we hypothesized that a skewed distribution of opportunities might contribute to lower pay for women. However, our logistic regression results show the opposite: after full controls, men have significantly lower odds of being top earners. This unexpected result likely reflects dataset limitations, which are discussed in a later section. Unsurprisingly, seniority and age strongly predict top-earner status. Organizational role and department also exert extremely large influences, underscoring the importance of structural and positional factors in determining high-income outcomes.

Figure 3*: Strong predictors for top 10% earners. Job role, seniority, and education are the overwhelming drivers, with the residual odds at this level favoring female employees (OR_Male= 0.35)*

7. Limitation

There are several limitations to our dataset. First, the number of job types represented is relatively small compared with real-world organizational structures, which may limit the generalizability of our findings. Second, the distribution of salaries appears unusually close to a normal distribution. In most real organizations, salary distributions are right-skewed due to the presence of a small number of high earners and executives. The more symmetric pattern in our dataset raises concerns about its comprehensiveness (Figure 4). One possible explanation is the data source: Glassdoor submissions often reflect salaries from mid-level or lower-level positions, as higher-paid employees are less likely to self-report. As a result, the dataset may be biased toward the mid-range of the salary distribution and may underrepresent top earners or executive-level roles.

8. Conclusion

In summary, across RQ1–RQ3, the apparent gender pay gap diminishes once controlling for job titles, departments, and other organizational factors, though education and experience related characteristics still explain substantial pay variance. There is no statistical significant evidence showing that gender pay gap does not significantly get minimized by education or performance. For RQ4, logistic regression unexpectedly shows that men have lower odds of being top earners after full controls, suggesting reinforcing that structural factors such as seniority, age, and organizational role primarily determine high-income outcomes.

9. Future Work

- Expand the Dataset
    1) Incorporate additional variables such as promotions, leadership roles, firm size, industry and geographic location to better capture structural drivers of pay differences
    2) Extend the dataset across multiple years to analyse salary growth and changes in gender pay gap over time.
- Examine Promotions and Career Progression
    1) Analyse whether performance evaluations translate into promotions at equal rates for men and women
    2) Move beyond salary levels to study promotions rates, time-to-promotion, and role transitions by gender.
- Improve model robustness

1) Explore causal inference methods (such as matching or instrumental variables) to better isolate gender-related effects.
2) Conduct sensitivity analyses and robustness check to validate findings.

10. Appendix

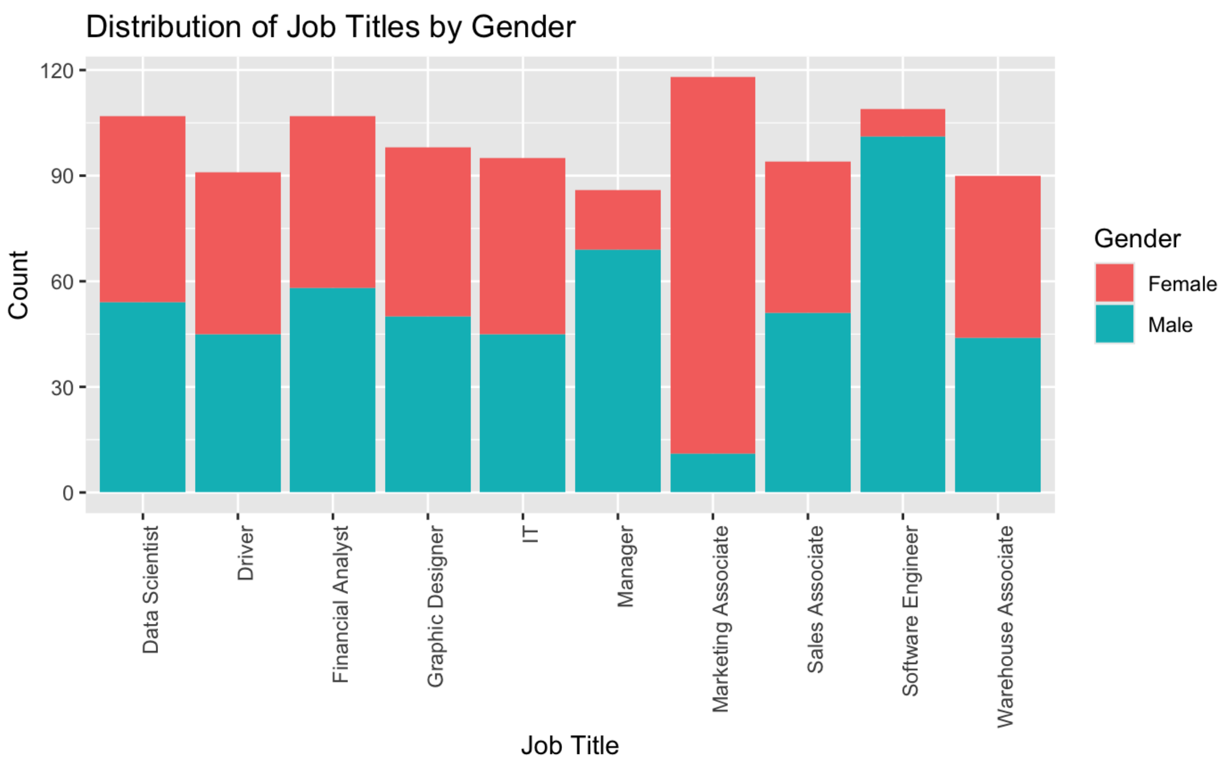

Figure 4*: The Distribution of each Job Title in the Dataset*

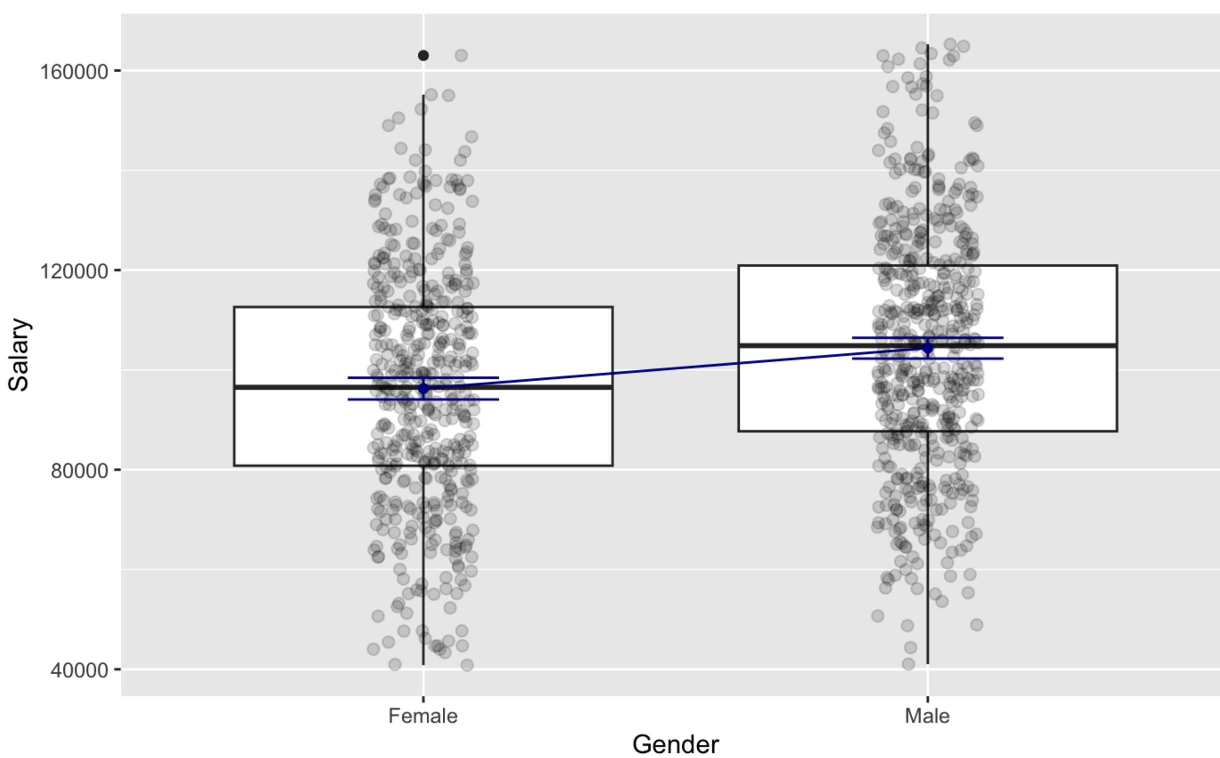

Figure 5*: The distribution of salary for female and male*

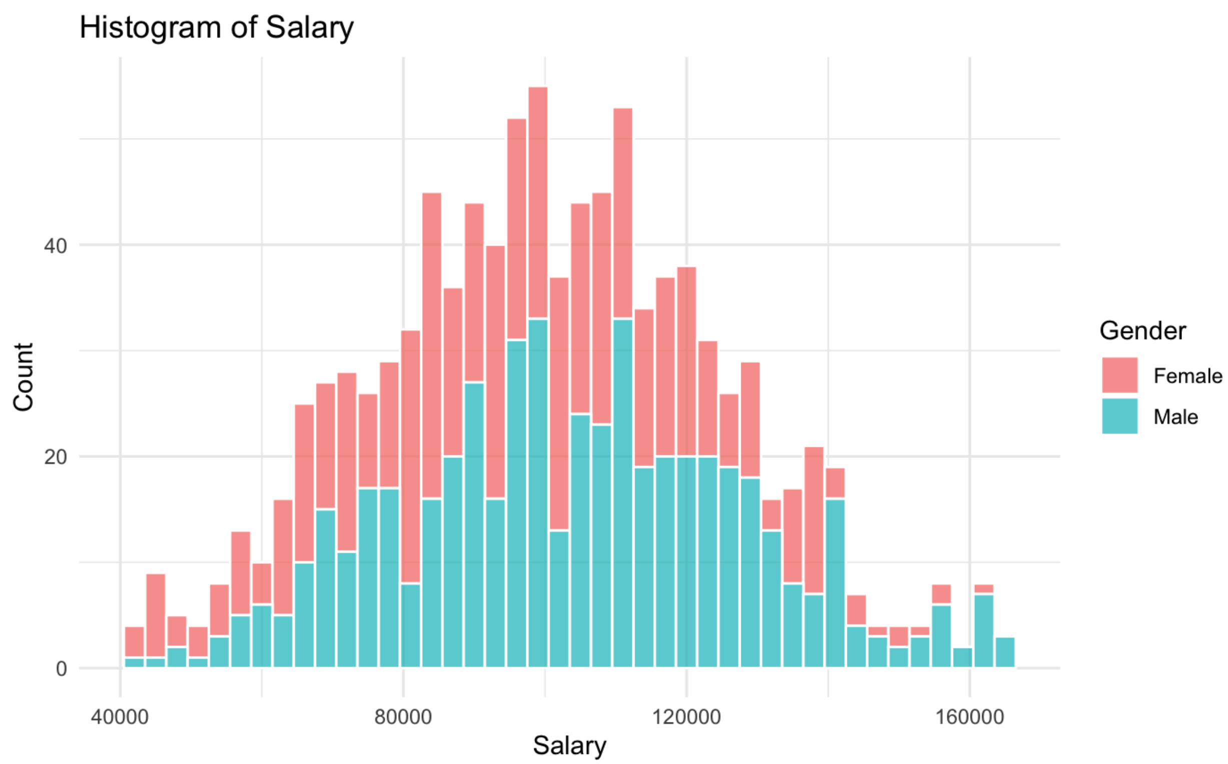

***Figure 6**: Distribution of salaries*

Test Results for RQ1:

```
Shapiro-Wilk normality test


data:  female$Salary
W = 0.99528, p-value = 0.1678
```

***Result 1**: Shapiro-Wilk Normality Test for female salary*

```
Shapiro-Wilk normality test

data:  male$Salary
W = 0.99459, p-value = 0.0592
```

***Result 2**: Shapiro-Wilk Normality Test for male salary*

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   1  0.3088 0.5785
      993
```

***Result 3**: Levene's Test for Salary ~ Gender*

```
Two Sample t-test

data:  male$Salary and female$Salary
t = 5.249, df = 993, p-value = 1.87e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  5080.382 11147.148
sample estimates:
mean of x mean of y
104375.24  96261.48

Cohen's d |       95% CI
------------------------
0.33      | [0.21, 0.46]
```

***Result 4**: Levene's Test for Salary ~ Gender*

```
Shapiro-Wilk normality test

data:  res
W = 0.99555, p-value = 0.005521
```

***Result 5**: Shapiro-Wilk Normality Test for residual of Salary ~ Gender * JobTitle*

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group  19   0.987 0.4738
      975
```

***Result 6**: Levene's Test of Salary ~ Gender * JobTitle*

```
DV:  Salary
Observations:  995
D:  0.9999999
MS total:  82585

                 Df   Sum Sq       H p.value
Gender            1    16765   0.203  0.6523
JobTitle          9 13887336 168.158  0.0000
Gender:JobTitle   9   734615   8.895  0.4470
Residuals       975 65523243
```

***Result 7**: Scheirer–Ray–Hare Test for Salary ~ Gender * JobTitle*

```
"Job Title 1","Job Title 2","Z statistic","Adjusted p-value"
"Data Scientist","Driver",-0.936513364784581,1
"Data Scientist","Financial Analyst",0.794847852948062,1
"Data Scientist","Graphic Designer",-0.195832461830762,1
"Data Scientist","IT",-0.636263615253325,1
"Data Scientist","Manager",8.0366096610294,3.60197344385632e-14
"Data Scientist","Marketing Associate",-4.77169595147957,6.21115852703611e-05
"Data Scientist","Sales Associate",0.206253854470138,1
"Data Scientist","Software Engineer",3.61094303300292,0.00762714186226366
"Data Scientist","Warehouse Associate",-0.643401117659081,1
"Driver","Financial Analyst",1.69856974641505,1
"Driver","Graphic Designer",0.729266389258829,1
"Driver","IT",0.29897419426266,1
"Driver","Manager",8.62716665005492,2.76718228812614e-16
"Driver","Marketing Associate",-3.60858569844385,0.00762714186226366
"Driver","Sales Associate",1.10636168728576,1
"Driver","Software Engineer",4.40116633867358,0.000333778678012601
"Driver","Warehouse Associate",0.279311888589846,1
"Financial Analyst","Graphic Designer",-0.973036611267622,1
"Financial Analyst","IT",-1.40714067385255,1
"Financial Analyst","Manager",7.28624983437275,1.17919642694347e-11
"Financial Analyst","Marketing Associate",-5.58574157892951,8.37738196248658e-07
"Financial Analyst","Sales Associate",-0.562460341574124,1
"Financial Analyst","Software Engineer",2.81242380784857,0.108173275943561
"Financial Analyst","Warehouse Associate",-1.40317987145744,1
"Graphic Designer","IT",-0.432778222026067,1
"Graphic Designer","Manager",8.06236776327749,2.99322700538912e-14
"Graphic Designer","Marketing Associate",-4.46042922674684,0.000261746136160909
"Graphic Designer","Sales Associate",0.391625637172878,1
"Graphic Designer","Software Engineer",3.726768423047,0.00504271312388315
"Graphic Designer","Warehouse Associate",-0.44276644798575,1
"IT","Manager",8.42215946490411,1.55235157026918e-15
"IT","Marketing Associate",-3.97039995009953,0.00200905830646697
"IT","Sales Associate",0.81694867051906,1
"IT","Software Engineer",4.14010925065465,0.00104142136318959
"IT","Warehouse Associate",-0.01584694690981,1
"Manager","Marketing Associate",-12.7015994178924,2.60700582340368e-35
"Manager","Sales Associate",-7.60447063098234,1.08707520795965e-12
"Manager","Software Engineer",-4.66254555753255,0.000103066232982546
"Manager","Warehouse Associate",-8.32861309240285,3.35359133032522e-15
"Marketing Associate","Sales Associate",4.81843183918591,5.06418282994723e-05
"Marketing Associate","Software Engineer",8.49381621762535,8.59829789127658e-16
```

```
"Marketing Associate","Warehouse Associate",3.89401974473756,0.00266210871607447
"Sales Associate","Software Engineer",3.28402977961998,0.0235368449657144
"Sales Associate","Warehouse Associate",-0.821696926341006,1
"Software Engineer","Warehouse Associate",-4.0963631022647,0.00121710530061763
```

***Result 8**: Dunn Test for Salary ~ Gender * JobTitle*

```
Shapiro-Wilk normality test

data:  res
W = 0.99606, p-value = 0.01246
```

***Result 9**: Shapiro-Wilk Normality Test for residual of Salary ~ Gender * Dept*

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   9  0.6711 0.7357
      985
```

***Result 10**: Levene's Test of Salary ~ Gender * Dept*

```
DV:  Salary
Observations:  995
D:  0.9999999
MS total:  82585

             Df   Sum Sq       H p.value
Gender        1  1947685 23.5840 0.00000
Dept          4   971410 11.7625 0.01921
Gender:Dept   4    59449  0.7199 0.94886
Residuals   985 79114336
```

***Result 11**: Scheirer–Ray–Hare Test for Salary ~ Gender * Dept*

```
"Job Title 1","Job Title 2","Z statistic","Adjusted p-value"
"Administration","Engineering",1.97727449673391,0.336074275708078
"Administration","Management",1.52203359191175,0.640003319601275
"Administration","Operations",-0.614193447901988,1
"Administration","Sales",1.92790438841258,0.336074275708078
"Engineering","Management",-0.462305280290929,1
"Engineering","Operations",-2.63179912582636,0.0849340633691772
"Engineering","Sales",-0.0814591169030587,1
"Management","Operations",-2.16981474837059,0.240167061109869
"Management","Sales",0.388088559114545,1
"Operations","Sales",2.59318087120903,0.0855834711601272
```

***Result 12**   Dunn Test for Salary ~ Gender * Dept*

```
Shapiro-Wilk normality test

data:  res
W = 0.99757, p-value = 0.1486
```

***Result 13**   Shapiro-Wilk Normality Test for residual of Salary ~ Gender * Education*

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value  Pr(>F)
group   7   1.834 0.07746 .
      987
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

***Result 14**   Levene's Test of Salary ~ Gender * Education*

```
                 Df    Sum Sq   Mean Sq F value   Pr(>F)
Gender             1 1.631e+10 1.631e+10  28.293 1.29e-07 ***
Education          3 1.712e+10 5.706e+09   9.896 1.97e-06 ***
Gender:Education   3 1.746e+09 5.821e+08   1.009    0.388
Residuals        987 5.691e+11 5.766e+08
```

Result 15 *ANOVA of Salary ~ Gender * Education*

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Salary ~ Gender * Education, data = data)

$Education
                          diff        lwr       upr     p adj
High School-College -4159.888 -9665.3433  1345.568 0.2102193
Masters-College      4383.608 -1168.0191  9935.235 0.1769245
PhD-College          6527.614   862.3636 12192.864 0.0163450
Masters-High School  8543.496  3117.6218 13969.370 0.0003189
PhD-High School     10687.501  5145.4256 16229.577 0.0000049
PhD-Masters          2144.006 -3443.9390  7731.950 0.7566900
```

Result 16 *Tukey Test of Salary ~ Gender * Education*

Test Results for RQ2:

```
Call:
lm(formula = Salary ~ Gender, data = data)


Residuals:
   Min     1Q Median     3Q    Max
-63345 -16194    375  16402  66750


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    96262       1126  85.487  < 2e-16 ***
GenderMale      8114       1546   5.249 1.87e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 24330 on 993 degrees of freedom
Multiple R-squared:  0.027,   Adjusted R-squared:  0.02602
F-statistic: 27.55 on 1 and 993 DF,  p-value: 1.87e-07
```

Result 17 *Linear Regression for Salary ~ Gender*

```
Call:
lm(formula = Salary ~ Gender + Dept + JobTitle, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-55615 -15448   -221  14682  52817

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                  98002.1     2618.1  37.432  < 2e-16 ***
GenderMale                    -655.9     1539.0  -0.426   0.6701
DeptEngineering               3782.1     2230.4   1.696   0.0903 .
DeptManagement                1274.5     2231.5   0.571   0.5680
DeptOperations               -2934.2     2179.9  -1.346   0.1786
DeptSales                     5130.2     2198.9   2.333   0.0198 *
JobTitleDriver               -3492.5     3114.2  -1.121   0.2624
JobTitleFinancial Analyst     2978.9     2988.2   0.997   0.3191
JobTitleGraphic Designer     -1395.1     3055.4  -0.457   0.6481
JobTitleIT                   -2981.5     3079.6  -0.968   0.3332
JobTitleManager              30383.2     3200.6   9.493  < 2e-16 ***
JobTitleMarketing Associate -16974.4     2986.2  -5.684 1.73e-08 ***
```

```
JobTitleSales Associate        780.6     3090.6   0.253   0.8007
JobTitleSoftware Engineer    13083.3     3046.0   4.295 1.92e-05 ***
JobTitleWarehouse Associate  -2246.1     3123.6  -0.719   0.4723
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21820 on 980 degrees of freedom
Multiple R-squared:  0.2278,  Adjusted R-squared:  0.2168
F-statistic: 20.65 on 14 and 980 DF,  p-value: < 2.2e-16
```

Result 18 *Linear Regression for Salary ~ Gender + Dept + JobTitle*

```
Call:
lm(formula = Salary ~ Gender + Seniority + PerfEval + Age, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-42262 -10438  -1157   9408  45259

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 24550.11    2217.90  11.069   <2e-16 ***
GenderMale   9736.56     973.81   9.998   <2e-16 ***
Seniority    9759.39     348.13  28.034   <2e-16 ***
PerfEval      799.89     341.81   2.340   0.0195 *
Age           956.24      34.04  28.090   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15280 on 990 degrees of freedom
Multiple R-squared:  0.6176,  Adjusted R-squared:  0.616
F-statistic: 399.7 on 4 and 990 DF,  p-value: < 2.2e-16
```

Result 19 *Linear Regression for Salary ~ Gender + Seniority + PerfEval + Age*

Test Results for RQ3:

```
Call:
lm(formula = Salary ~ Gender * Education + Gender * PerfEval +
    Seniority + Age + JobTitle + Dept, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-33719  -6961    409   7078  28078

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      21441.1     2178.5   9.842  < 2e-16 ***
GenderMale                        2712.8     1949.2   1.392  0.16433
EducationHigh School              -506.0     1272.3  -0.398  0.69094
EducationMasters                  4231.7     1334.5   3.171  0.00157 **
EducationPhD                      7549.1     1344.0   5.617 2.54e-08 ***
PerfEval                          1385.0      329.0   4.210 2.79e-05 ***
Seniority                         9843.9      231.4  42.536  < 2e-16 ***
Age                                942.1       22.5  41.862  < 2e-16 ***
JobTitleDriver                   -3853.7     1439.3  -2.677  0.00754 **
JobTitleFinancial Analyst         3544.9     1383.6   2.562  0.01055 *
JobTitleGraphic Designer         -2380.8     1417.5  -1.680  0.09335 .
JobTitleIT                       -2429.6     1423.1  -1.707  0.08809 .
JobTitleManager                  30948.4     1485.0  20.841  < 2e-16 ***
JobTitleMarketing Associate     -16982.1     1379.6 -12.309  < 2e-16 ***
JobTitleSales Associate            401.3     1427.0   0.281  0.77862
JobTitleSoftware Engineer        13468.6     1404.8   9.587  < 2e-16 ***
JobTitleWarehouse Associate       -509.8     1443.8  -0.353  0.72412
DeptEngineering                   3201.6     1031.8   3.103  0.00197 **
DeptManagement                    2711.0     1030.3   2.631  0.00864 **
```

```
DeptOperations                    -639.7     1010.5  -0.633  0.52689    
DeptSales                         5947.1     1017.5   5.845 6.93e-09 ***
GenderMale:EducationHigh School  -1952.3     1822.0  -1.072  0.28418    
GenderMale:EducationMasters        490.8     1845.4   0.266  0.79032    
GenderMale:EducationPhD          -3184.5     1862.1  -1.710  0.08755 .  
GenderMale:PerfEval               -402.8      454.9  -0.885  0.37615    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10050 on 970 degrees of freedom
Multiple R-squared:  0.8378,  Adjusted R-squared:  0.8338 
F-statistic: 208.8 on 24 and 970 DF,  p-value: < 2.2e-16
```

Result 20 *Linear Regression for Salary ~ Gender * Education + Gender * PerfEval + Seniority + Age + JobTitle + Dept*

Test Results for RQ4:

```
Call:
glm(formula = TopEarner ~ Gender + Seniority + Age + JobTitle + 
    Dept + Education + PerfEval, family = "binomial", data = data)

Coefficients:
                             Estimate Std. Error z value Pr(>|z|)    
(Intercept)                 -23.61733    2.49015  -9.484  < 2e-16 ***
GenderMale                   -1.05857    0.40827  -2.593 0.009520 ** 
Seniority                     1.84565    0.21204   8.704  < 2e-16 ***
Age                           0.20982    0.02427   8.645  < 2e-16 ***
JobTitleDriver               -0.51464    0.85145  -0.604 0.545559    
JobTitleFinancial Analyst     0.58785    0.80553   0.730 0.465533    
JobTitleGraphic Designer     -0.26881    0.83951  -0.320 0.748817    
JobTitleIT                    0.45766    0.79182   0.578 0.563274    
JobTitleManager               6.15045    0.89877   6.843 7.74e-12 ***
JobTitleMarketing Associate  -4.69026    1.47599  -3.178 0.001484 ** 
JobTitleSales Associate       1.04576    0.78413   1.334 0.182315    
JobTitleSoftware Engineer     2.95718    0.79740   3.709 0.000208 ***
JobTitleWarehouse Associate  -0.16067    0.88703  -0.181 0.856263    
DeptEngineering               1.98807    0.65513   3.035 0.002408 ** 
DeptManagement                1.28916    0.66414   1.941 0.052244 .  
DeptOperations                0.46617    0.75201   0.620 0.535330    
DeptSales                     2.58741    0.67777   3.818 0.000135 ***
EducationHigh School         -0.50607    0.54728  -0.925 0.355120    
EducationMasters              1.27290    0.51223   2.485 0.012955 *  
EducationPhD                  1.52265    0.50842   2.995 0.002746 ** 
PerfEval                      0.38201    0.12250   3.119 0.001817 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Result 20 *Logistic Regression for TopEarner ~ Gender + Seniority + Age + JobTitle + Dept + Education + PerfEval*

Extra Test Results for RQ1:

```
Call:
multinom(formula = JobTitle ~ Gender + Seniority + Age, data = data)

Coefficients:
                     (Intercept)  GenderMale    Seniority           Age
Driver              -0.294643787 -0.03711696 -0.003492877  0.0038656186
Financial Analyst    0.047156689  0.14682075 -0.004168948 -0.0027185714
Graphic Designer    -0.213146735  0.02465605  0.081268916 -0.0032579137
IT                  -0.008780129 -0.12539513 -0.052663778  0.0024966769
Manager             -1.157446661  1.38258482 -0.041976102  0.0034115107
Marketing Associate  0.638828048 -2.29183323  0.030321455 -0.0006668065
Sales Associate     -0.463295095  0.15768136  0.010591199  0.0052459885
Software Engineer   -1.881779407  2.51721105  0.002076780 -0.0003748319
```

```
Warehouse Associate  0.173233095 -0.07206054  0.032169693 -0.0100732684

Std. Errors:
                    (Intercept) GenderMale  Seniority         Age
Driver                0.5675383  0.2855099 0.10259903 0.010037641
Financial Analyst     0.5404311  0.2742273 0.09832673 0.009608132
Graphic Designer      0.5565394  0.2801042 0.10058484 0.009843961
IT                    0.5569115  0.2824885 0.10164704 0.009926496
Manager               0.6047010  0.3330078 0.10536419 0.010182025
Marketing Associate   0.5391696  0.3711886 0.09831093 0.009751918
Sales Associate       0.5660334  0.2835995 0.10170043 0.009941041
Software Engineer     0.6345149  0.4153149 0.10007366 0.009657463
Warehouse Associate   0.5613016  0.2866503 0.10282214 0.010106759
```

Result 21 *Multi nomial Regression for JobTitle ~ Gender + Seniority + Age*

```
Response: JobTitle
          LR Chisq Df Pr(>Chisq)
Gender     215.593  9     <2e-16 ***
Seniority    2.343  9     0.9848
Age          3.315  9     0.9505
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Result 22 *ANOVA JobTitle ~ Gender + Seniority + Age*