A Study on the Gender Pay Gap in Egypt's Tech Market

Supervised by:
   Dr. Mohamed Taher Alrefaie

Submitted by:
   Mohamed Elmosalamy 23-101283
   Omar Mohammad 23-101288
   Omar Hesham 23-101302
   Mohamed Kotb 24-101222

2025-05-24

Contents

1 Abstract

The gender pay gap remains a contentious and highly debated issue worldwide. This analysis investigates whether women earn less than their peers in **Egypt's tech industry**.

The study aims to quantify salary differences between genders using publicly available data, offering insights into whether a significant gender-based wage gap exists.

2 Introduction

The gender pay gap has been a global issue across labor markets. While considerable research has been done in wester countries, limited data-driven analysis exists in the Middle East. In Egypt This remains an unstudied topic.

To reason about this, we had to define 3 methods to attack this problem.

1. Assessing the pay gap between genders

The first hypotehsis testing aims to address the gender pay gap in Egypt's tech industry, by conducting **Welch's independant t-test** on salaries for both genders. This analysis aims to answer the question **"Is there a pay gap between men and women in Egypt's tech industry?"**

TODO OMAR -> *Describe 2st hypothesis here*

1. Hypothesis 2nd

2. Estimating **The Cost of Being a Woman**

we define *The Cost of Being a Woman* as the amount of disparity between salaries that a women loses (or potentially earns) compared to men per month. It's the amount a man with her skills, title, and years of experience would be expected to earn. The analysis aims to provide a confidence interval with $\alpha = 95\%$ to find this cost.

3 Research Questions

1. Is there a statistically significant difference in mean salaries between genders?
2. After controlling for years of experience, and other contributing factors, does the pay gap persist/vanish?

3. What is the cost of being a woman?, in other words, how much does a woman lose/gain in a month in comparison to a man with the same qualifications?

4 Hypothesis

4.1 Difference in mean salaries

- Null hypothesis (H_o): There is no significant pay gap between men and women.
- Alternative hypothesis (H_α): There is a significant pay gap between men and women.
- Alpha value (α): 0.05

4.2 Controlled difference in mean salaries

- Null hypothesis (H_o): After controlling for all contributing factors such as years of experience, job title, job level, etc., there is no significant pay gap between men and women.
- Alternative hypothesis (H_α): After controlling for all contributing factors such as years of experience, job title, job level, etc., there is a significant pay gap between men and women.
- Alpha value (α): 0.05

5 Population of Interest

All professionals working in the tech field in Egypt.

6 Dataset

The dataset used in this study comes from the popular Egyptian Tech Market Survey API conducted in 2024 on the technology sector in Egypt.

Below is a brief description of the relevant columns included in the dataset:
- **Gender:** gender of the subject (Male, Female).
- **Degree:** whether or not the subject has a bachelor degree (Yes, No)
- **Title:** professional title of the subject. (Data Analyst, Scrum Master, etc.)
- **Level:** professional level of the subject (Junior, Senior, Team lead, etc.)
- **YearsOfExperience:** number of years the subject has been working in the field.
- **Salary:** amount of money a subject is paid monthly in EGP currency.
- **IsEgp:** currency of the subject's salary. (EGP, Another Currency [e.g. USD], Hybrid)
- **ProgrammingLanguages:** programming languages that can be written by subject (Python, C++, etc.)
- **BusinessMarket:** market scope of the subject's company (Regional, Global, Local)

- **BusinessSize:** size of the subject's company (Start-up, Small and Medium, Large)
- **WorkSetting:** working condition of the subject. (Office, Remote, etc.)
- **CompanyLocation:** state/city where the subject's company operates.

Number of samples used: **2649**.

7 Analysis

TODO MOHA -> insert visualizations and descriptions from the python notebook

8 Hypothesis Testing Steps

8.1 Difference in mean salaries

We chose **Welch's Independent T-Test** for this hypothesis.

• Step 1: First, we formulated the null and alternative hypotheses (mentioned in section 4.1).

• Step 2: Then, we assumed our significance level (α) to be the standard 0.05.

• Step 3: Next, we selected and cleaned our dataset to ensure no missing or erroneous values existed within the comparison groups.

• Step 4: We used the `scipy.stats.ttest_ind()` function with `equal_var=False` to perform Welch's t-test, which is ideal for comparing the means of two groups when their variances are unequal.

• Step 5: Finally, we compared the p-value produced by the test with our alpha value to decide whether to reject or fail to reject the null hypothesis.

8.2 Controlled difference in mean salaries

For this complicated hypothesis test, we used **Blinder–Oaxaca decomposition**.

• Step 1: First, we formulated the null and alternative hypotheses (mentioned in section 4.2).

• Step 2: We set the significance level (α) to 0.05, as is standard in most social science research.

• Step 3: We selected and cleaned our dataset, and then processed the data to define the outcome variable and explanatory variables clearly.

• Step 4: We ran separate regressions for the two groups and applied the Blinder–Oaxaca decomposition using the `statsmodels`. This allowed us to break down the mean outcome gap into an explained component and an unexplained component.

• Step 5: We interpreted the decomposition output by examining the coefficients, p-values, and contribution of each variable. We assessed whether the overall gap, explained portion, or unexplained portion were statistically significant and derived conclusions about which factors contributed to inequality and how much remained unexplained.

8.3 Cost of Being a woman

We can define the cost of begin a woman as following

$$\text{Cost} := \text{Expected salary based on objective factors} - \text{Current Salary}$$

To estimate the cost of being a woman, we need to find the objective factors that should dictate one's salary and use them to predict their salaries regardless of gender.

The objective factors used in this study are:

1. Years of Experience
2. Title
3. Level

We used a multiple regression model to predict each woman's expected salary. This model captures how salary is determined for male employees based on their qualifications, creating a baseline for equitable pay.

The idea behind this is to have the explanatory variables all the columns in the dataset except for `gender` and train this model solely on men data. Therefore, this model treats the entire workforce as men. The trained model is applied to the female workforce to find their expected salary based on objective factors.

Now, we use the current salary given in the dataset, and use this distribution to create a confidence interval.

9 Conclusion

9.1 Difference in mean salaries

P-value: 0.2668 The result is far greater the significance level 0.05, so we fail to reject the null hypothesis: There is no significant evidence of a pay gap between men and women.

9.2 Controlled difference in mean salaries

After controlling for years of experience, the resultant data was used to generate the following graph
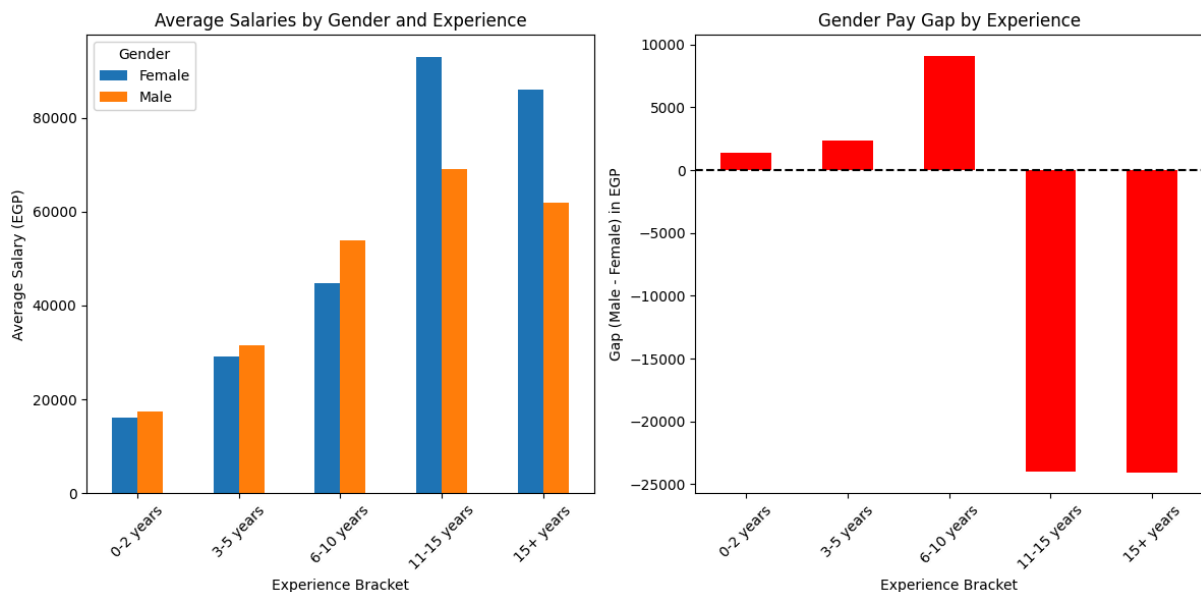

Key observations:

1. The average salary for both genders increases with experience
2. In the first two years, the average salary is around 18000 for both genders
3. In the first 10 years, men appear to have higher average salaries, small at the first two years, and increases as time goes on
4. After the first 10 years, women have far higher salaries them men in the tech industry. After 15+ years, women have 12000 EGP more than men on average.

9.3 The Cost of Being a Woman

The resultant confidence interval is the following:

Distribution of Women Residuals with 95% Confidence Interval
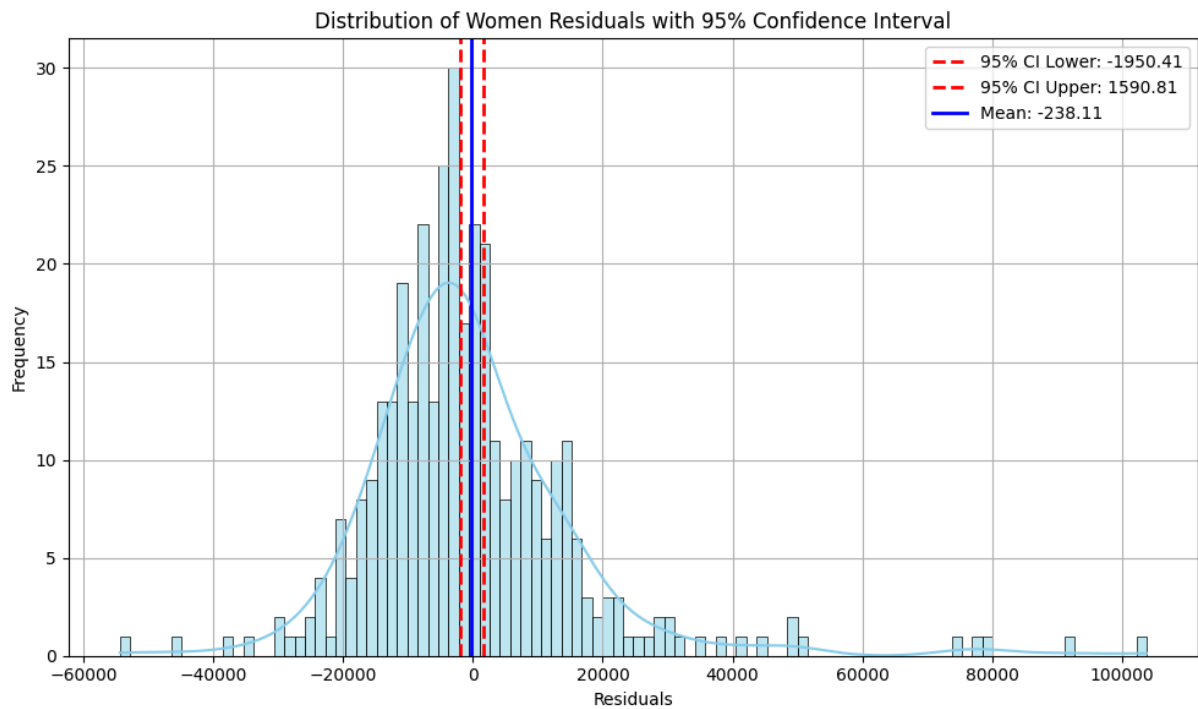
We are 95% confident that the *cost of being a woman* lies between −1950.41 and 1590.81.

we can reach an important conclusion: the value **0** is included in the 95% interval, showing that that 0 is a possible estimate for the cost of being a woman, indicating that the cost might be zero.

10 Any potential issues

TODO MOHA -> discuss outliers and the scarcity of female records in the dataset