cct College Dublin
Computing • IT • Business

# Assessment Cover Page

| | |
|---|---|
| *Student Full Name* | Yumiko Maria Bejarano Azogue |
| *Student Number* | 2024144 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA 1 – Capstone Project Proposal |
| *Assessment Due Date* | 29th March 2024 23:59 |
| *Date of Submission* | 29.03.2024 |

## Declaration

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

# Contents

# Women in Technology Industry

## Introduction

The present report focuses on analysing women's participation in the technology industry, highlighting the challenges and opportunities we constantly face. The low representation of women in technical and leadership roles in the technology industry is a problem that affects gender equity and limits the potential for innovation and growth in that sector.

Women face persistent gender barriers in the technology industry, the lack of role models and stereotypes that perpetuate gender inequality in this field, to quote the civil rights activist Marian Wright Edelman, “You can’t be what you can’t see”.

## Scope

This project will focus on investigating the gender gap in the technology industry, exploring women's participation in the technical, leadership, and entrepreneurial roles. Gender distribution will be analyzed across different contexts, including representation by country and age group, and will suggest potential areas for improvement to promote gender equity in technology will be identified.

The project scope will include analyzing gender distribution by country and age group, as well as exploring the most common job titles among surveyed women and their educational levels. Advanced data analysis methods such as machine learning and text mining will be used to identify patterns and trends in women's participation in technology. Any analysis not directly related to the gender gap in the technology industry will be excluded.

## Problem definition

The gender gap in different roles and leadership positions in the technology industry limits professional advancement opportunities and decision-making for women. Gender stereotypes deeply ingrained in society perpetuate the perception that women are not as competent in technological fields as men, affecting their confidence and recognition at work.

The male-dominated work culture in technology companies creates an environment that is not inclusive for women, making it difficult for them to integrate and progress within the organization. Furthermore, gender bias in evaluation and promotion processes result in lower salaries, limited promotion opportunities, and less recognition for women.

# Objectives

## General Objective:

To investigate and understand the gender gap in the technology industry, focusing on women's participation in technical, leadership, and entrepreneurial roles.

## Specific Objectives:

Analyze gender distribution in different contexts, including by country and age group, using data from the Kaggle ML & DS Survey.

Identify the most common job titles among surveyed women and their educational levels to better understand trends in women's participation in technology.

Explore potential areas for improvement to promote gender equity in the technology industry by analyzing the challenges and barriers faced by women in technical and leadership roles.

# Data Sources (Technologies used)

## Libraries

Different libraries have been used to perform different tasks and modelling of algorithms. These may include: Pandas, Numpy, Seaborn, Matplotlib, etc.

## Accomplishment Data

The Gender Statistics database is a comprehensive source for the latest sex-disaggregated data and gender statistics covering demography, education, health, access to economic opportunities, public life and decision-making, and agency.

# Attributes

### Multiple_choice_responses.csv

This file contains responses to single-choice questions in separate columns. For questions with multiple responses, each option was split into its own column. Text responses were encoded to safeguard user privacy, and countries with fewer than 50 respondents were grouped as the “other.”

- The dataset comprises 19,718 rows and 246 columns.
- Each column represents a different question or provides additional metadata related to the survey.

### Other_text_responses.csv

If "Other" was selected, respondents had the option to provide a text response. These responses were separated and shuffled to protect privacy.

- This DataFrame includes responses to open-ended survey questions.
- It consists of 19,718 rows and 28 columns.
- Each column represents an open-ended question, with responses stored as text.

### Questions_only.csv

This file lists the questions from the 2019 Kaggle Data Science and Machine Learning Survey.

- All columns are of the 'object' data type.
- The DataFrame has dimensions of 1 row and 35 columns.

### Survey_schema.csv

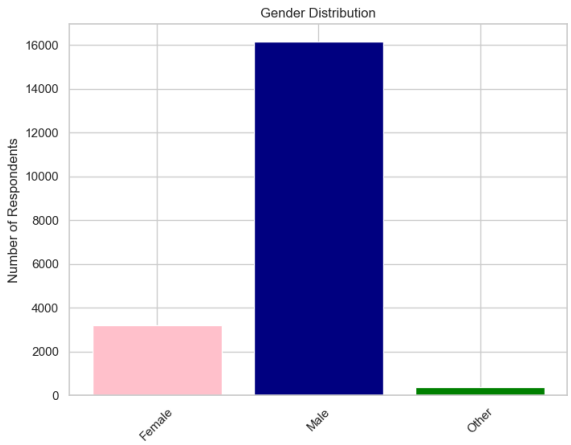
This dataset describes which questions were presented to which respondents in the survey. Generally, respondents with more experience were asked more questions.

- It contains 10 rows and 35 columns.
- All columns have the data type 'object'.
- The survey scheme dataset serves as a key reference for understanding the structure and content of the data.

# Data Analysis and Insights

## Gender Distribution Among Participants

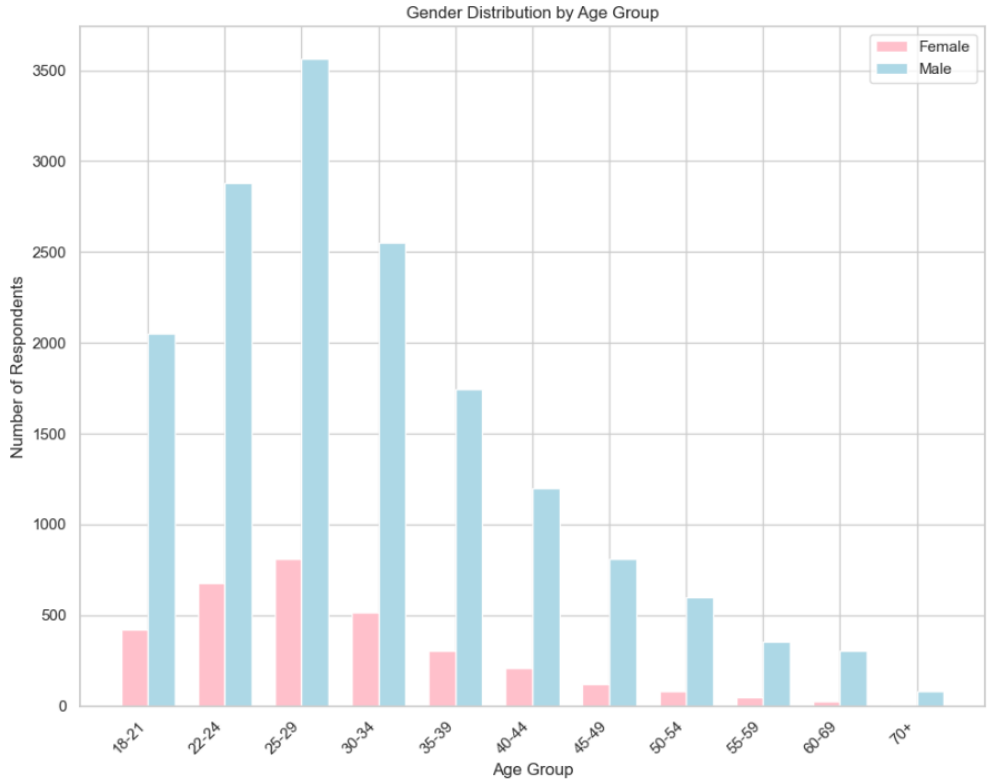
The data reveals a significant gender disparity among participants. Out of the total respondents, 16,138 identified as male, while only 3,212 identified as female. Additionally, there were 367 respondents who identified as 'Other' gender. This highlights a notable difference in gender representation within the dataset.



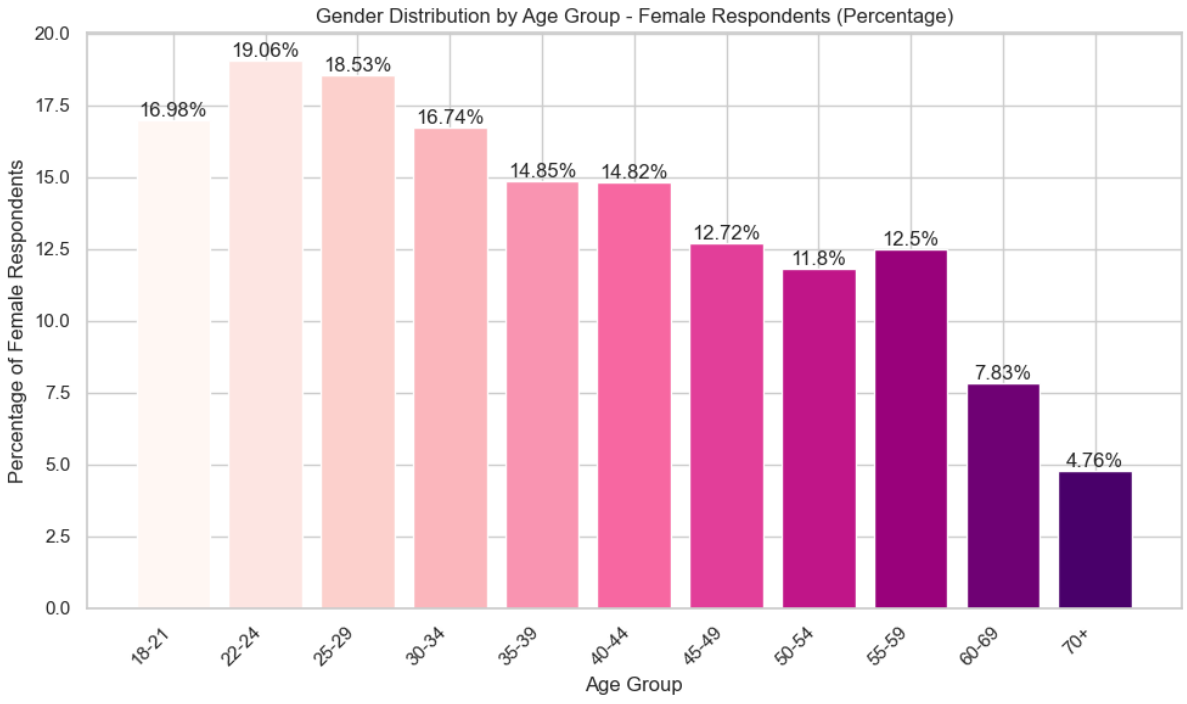

| Gender | Respondents | Percentage |
|---|---|---|
| Female | 3212 | 16.29 |
| Male | 16138 | 81.85 |
| Other | 367 | 1.86 |

## Age Distribution

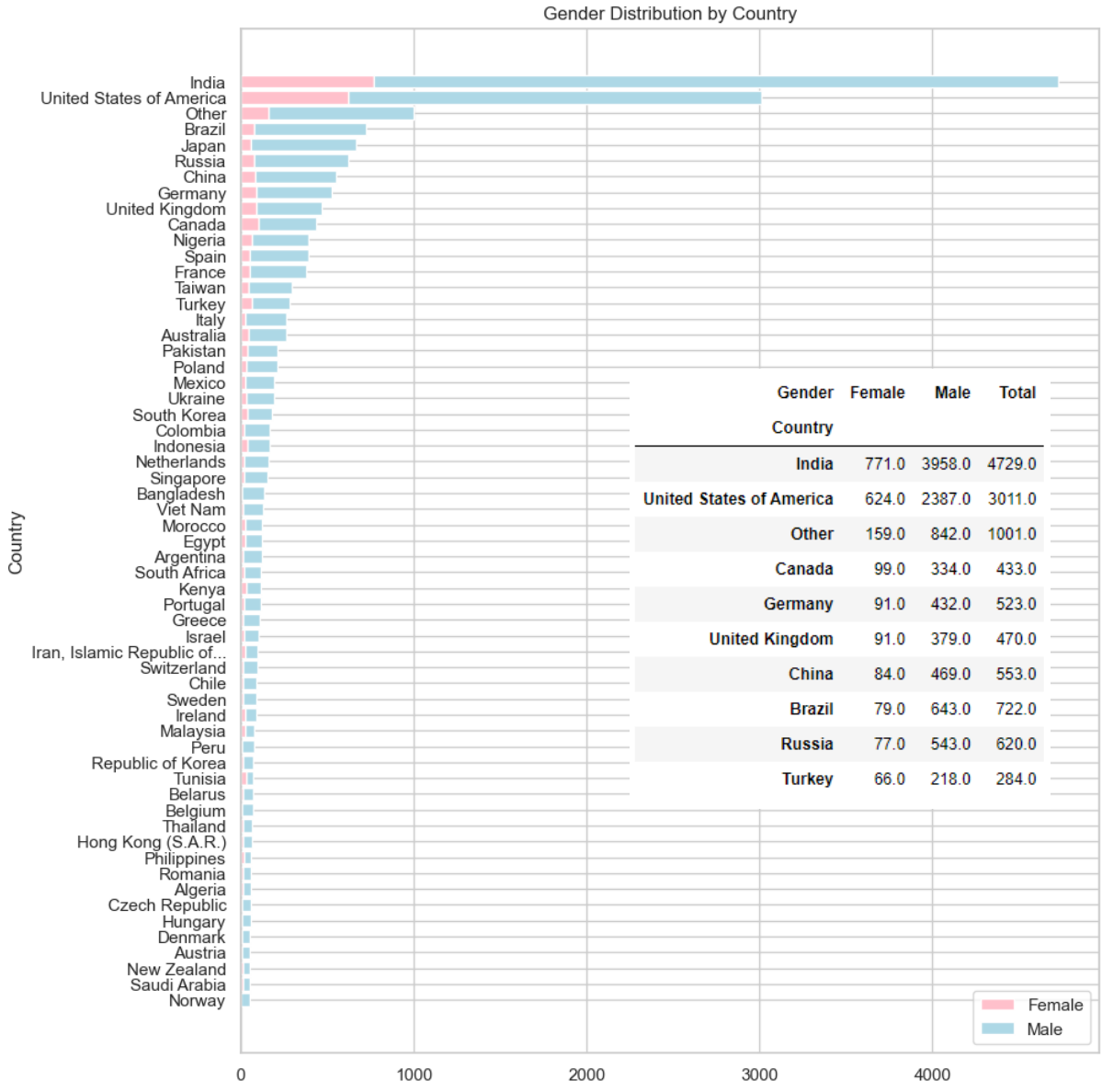
When examining the age distribution among male and female respondents, we notice a chart. There's a moderate bias towards respondents under 40 years old, which could be attributed to the relatively low median age of employees in the technology field.




The percentage of surveyed women within each of the eleven age groups follows a similar pattern with some interesting differences. Overall, we can see that there are fewer women in older age groups, indicating a lower representation of older women in technology. However, the highest percentages are not found among young adults aged 18 to 25, but among respondents in their twenties.



An initial analysis of the participant count by country reveals that India and the United States have the highest numbers, followed by Canada, Germany, and the United Kingdom. This is expected since the densely populated countries with a strong technological presence tend to attract higher participation rates overall.

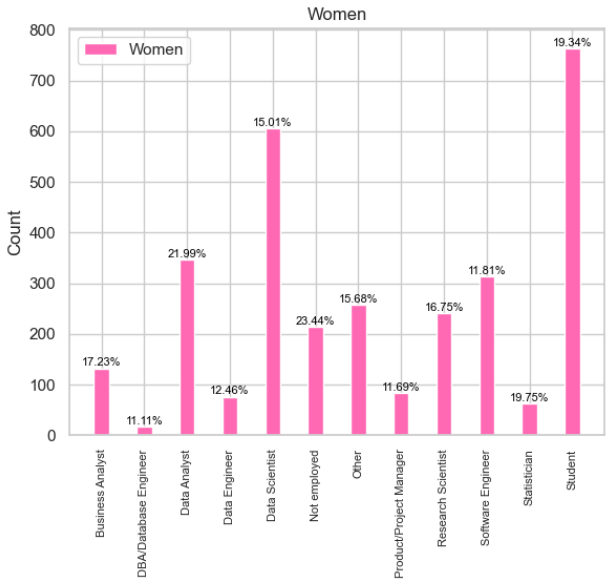


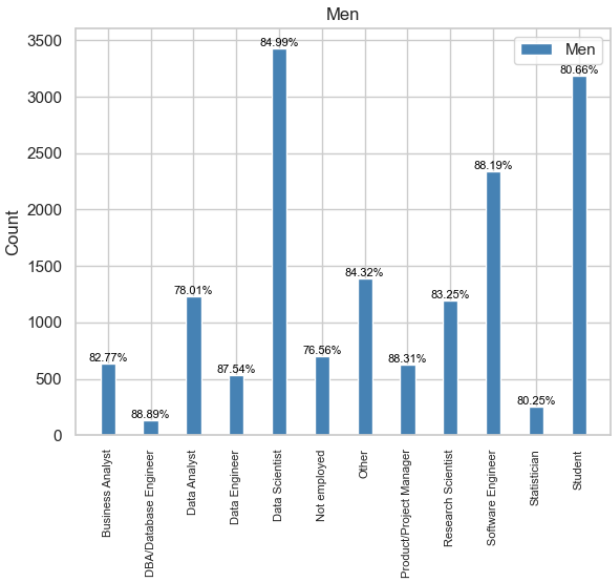

| Gender | Female | Male | Total |
|---|---|---|---|
| Country | | | |
| India | 771.0 | 3958.0 | 4729.0 |
| United States of America | 624.0 | 2387.0 | 3011.0 |
| Other | 159.0 | 842.0 | 1001.0 |
| Canada | 99.0 | 334.0 | 433.0 |
| Germany | 91.0 | 432.0 | 523.0 |
| United Kingdom | 91.0 | 379.0 | 470.0 |
| China | 84.0 | 469.0 | 553.0 |
| Brazil | 79.0 | 643.0 | 722.0 |
| Russia | 77.0 | 543.0 | 620.0 |
| Turkey | 66.0 | 218.0 | 284.0 |

# General Data Analysis

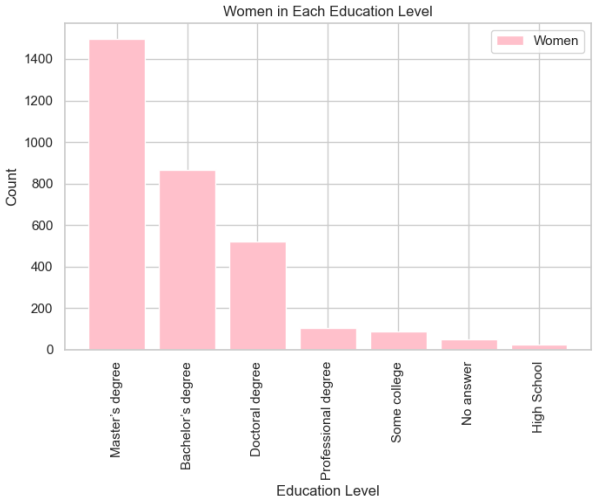
## Career and Education

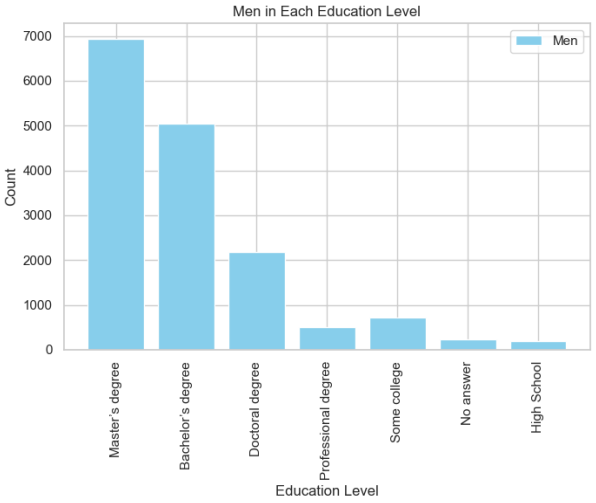
A detailed analysis of job role distribution reveals significant differences between men and women. While students make up the largest category, women represent a considerably higher percentage than men in this area (approximately 32% of women are students, while only 24% of are men). In specific technical roles like Data Scientist, Data Analyst, and Software Engineer, women are underrepresented compared to men. For instance, only around 15% of data scientists are women, contrasting with approximately 20% of men in this field.




In terms of education, women show a higher proportion in obtaining master's degrees compared to men, representing approximately 18% of surveyed women compared to only 13% of men. However, the gender gap is less pronounced in other educational levels, such as bachelor's and doctoral degrees, where the percentage differences are less significant. For example, approximately 14% of women hold a doctoral degree, compared to 12% of men.

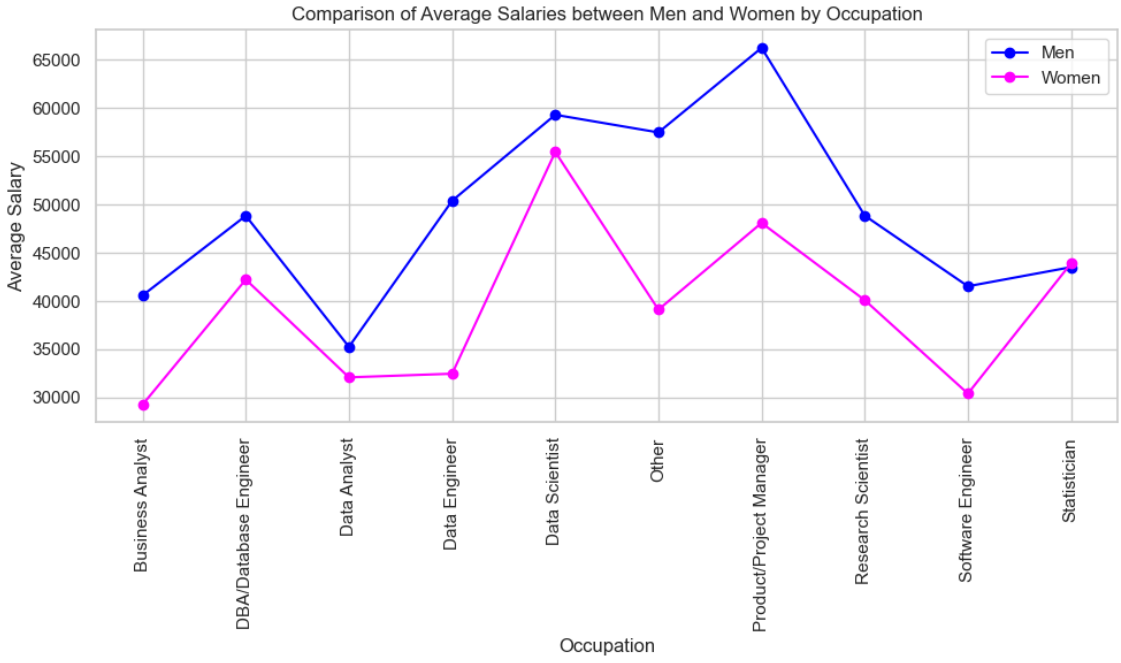



## Salary

The information we have depicts that women earn less than men in various tech jobs. For example, in roles like Business Analyst or Product Manager, women earn much less, sometimes over $18,000 less on average. Even in jobs where women earn a bit more, like in the role of a Statistician, the difference is meagre. This elaborates on the need to create more equitable accommodations in the required areas in order to facilitate more gender equality and ensure equal pay for all regardless of their gender.

| Occupation | Range_Start_women | Range_Start_men |
|---|---|---|
| Business Analyst | 29288.29 | 40594.05 |
| DBA/Database Engineer | 42230.27 | 48832.84 |
| Data Analyst | 32104.29 | 35264.37 |
| Data Engineer | 32480.84 | 50386.65 |
| Data Scientist | 55440.04 | 59289.54 |
| Other | 39116.88 | 57462.01 |
| Product/Project Manager | 48093.66 | 66221.14 |
| Research Scientist | 40094.98 | 48845.81 |
| Software Engineer | 30422.05 | 41534.98 |
| Statistician | 43910.97 | 43515.05 |




# Conclusion

The data analysis shows a significant wage gap between men and women in the tech industry. This gap exists across many roles, from Business Analysts to Product Managers, where women consistently earn less than their male counterparts. This situation raises concerns about gender fairness in the tech sector and highlights the need for actions to address this wage discrimination and promote equal pay between men and women in the tech industry.

As Cathy Engelbert, CEO of Deloitte, stated, "Women continue to be the most underutilized natural resource" which emphasizes a binarized world where talent is underutilised because of existing gender stereotypes. Additionally, it's essential to recognize and tackle gender biases ingrained in hiring, evaluation, and promotion processes, which perpetuate the wage gap and limit professional advancement opportunities for women in technology. Implementing inclusive policies and gender equity awareness programs are crucial steps toward building a more diverse, fair, and inclusive tech industry for all its participants.

# References

Mooney, P. (2019). 2019 Kaggle Machine Learning & Data Science Survey. Kaggle. URL: https://www.kaggle.com/competitions/kaggle-survey-2019.

Engelbert, C. 2016, 'Starting A New Conversation for Women: Deloitte CEO Catherine Engelbert on Gender Equality', TIME.

OpenAI. (2024). ChatGPT [Artificial Intelligence model]. Conversation with ChatGPT on March 29, 2024, at 10:00 a.m. https://chat.openai.com/

Douglas M. Branson (2018) The Future of Tech Is Female : How to Achieve Gender Diversity. New York: NYU Press. Available at: https://research.ebsco.com/linkprocessor/plink?id=e3b09355-de83-37e4-b1bf-678f33012b68 (Accessed: 27 March 2024).