A logo for college computing

Description automatically generated

**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* | 2\*.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

[Introduction 1](#_Toc162602522)

[Scope 1](#_Toc162602523)

[Problem definition 1](#_Toc162602524)

[Objectives 2](#_Toc162602525)

[General Objective: 2](#_Toc162602526)

[Specific Objectives: 2](#_Toc162602527)

[Data Sourses (Technologies used) 2](#_Toc162602528)

[Libraries 2](#_Toc162602529)

[Accomplishment Data 2](#_Toc162602530)

[Source 2](#_Toc162602531)

[Attributes 2](#_Toc162602532)

[Descriptive statistics and Data 2](#_Toc162602533)

[multiple\_choice\_responses.csv 3](#_Toc162602534)

[other\_text\_responses.csv 3](#_Toc162602535)

[questions\_only.csv 3](#_Toc162602536)

[survey\_schema.csv 3](#_Toc162602537)

[Data Analysis and Insights 3](#_Toc162602538)

[Gender Distribution Among Participants 3](#_Toc162602539)

[Age Distribution 5](#_Toc162602540)

[General Data Analysis [proponer un nombre alternativo a este capítulo] 7](#_Toc162602541)

[Career and Education 7](#_Toc162602542)

[Salary 8](#_Toc162602543)

[Conclusion 9](#_Toc162602544)

[References 10](#_Toc162602545)

**Women in Technology Industry**

# Introduction

The present report focuses on analyzing 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, as highlighted in Tracy Chou's article "Women in Tech: The Facts," which underscores the lack of role models and stereotypes that perpetuate gender inequality in this field.

# 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 (Source: "Women in Tech: Breaking Barriers and Overcoming Challenges" by Tech Insights). 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 (Source: "Gender Diversity in Tech: The Key to Innovation" by Tech Solutions).

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 (Source: "Closing the Gender Gap: Strategies for Promoting Women in Tech" by Tech Trends).

The lack of female role models and specific support networks also hinder professional development and access to mentoring opportunities for women in technology (Source: "Empowering Women in Tech: Building Support Networks" by Tech Insights).

# 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, scipy, missingno, 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.

## Source

The data has been taken from an online source that is Kaggle. Kaggle links needs to be provided and referenced!

Attributes are the variables in the machine learning model that may be used as a predictor (Khanal et al., 2018). In this paper, the main attributes include symptoms of patients, vaccination name, and days spent in hospital after contacting covid19.

## Descriptive statistics and Data

In the descriptive statistics, we have gone through the overview of our dataset using head or simple description codes. The following results showed the statistics of numerical features.

## 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.

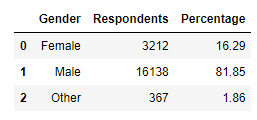
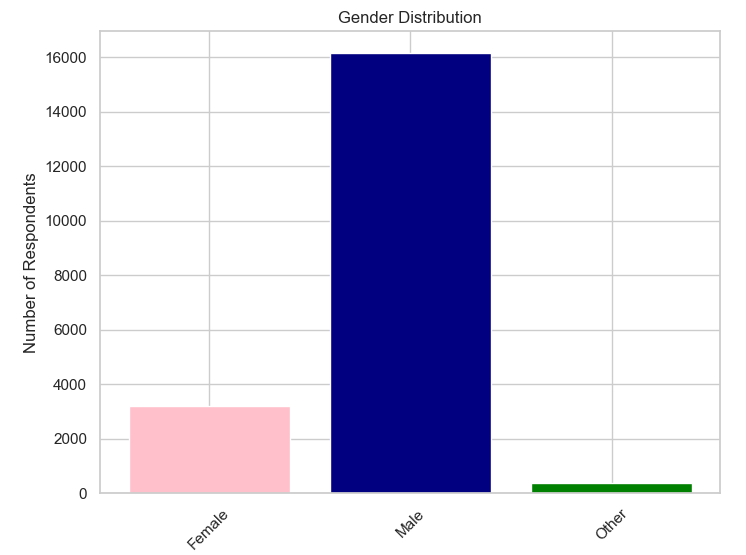
All columns have the data type 'object'.

The survey scheme dataset serves as a key reference for understanding the structure and content of the survey data.

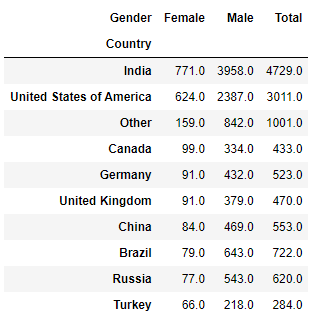
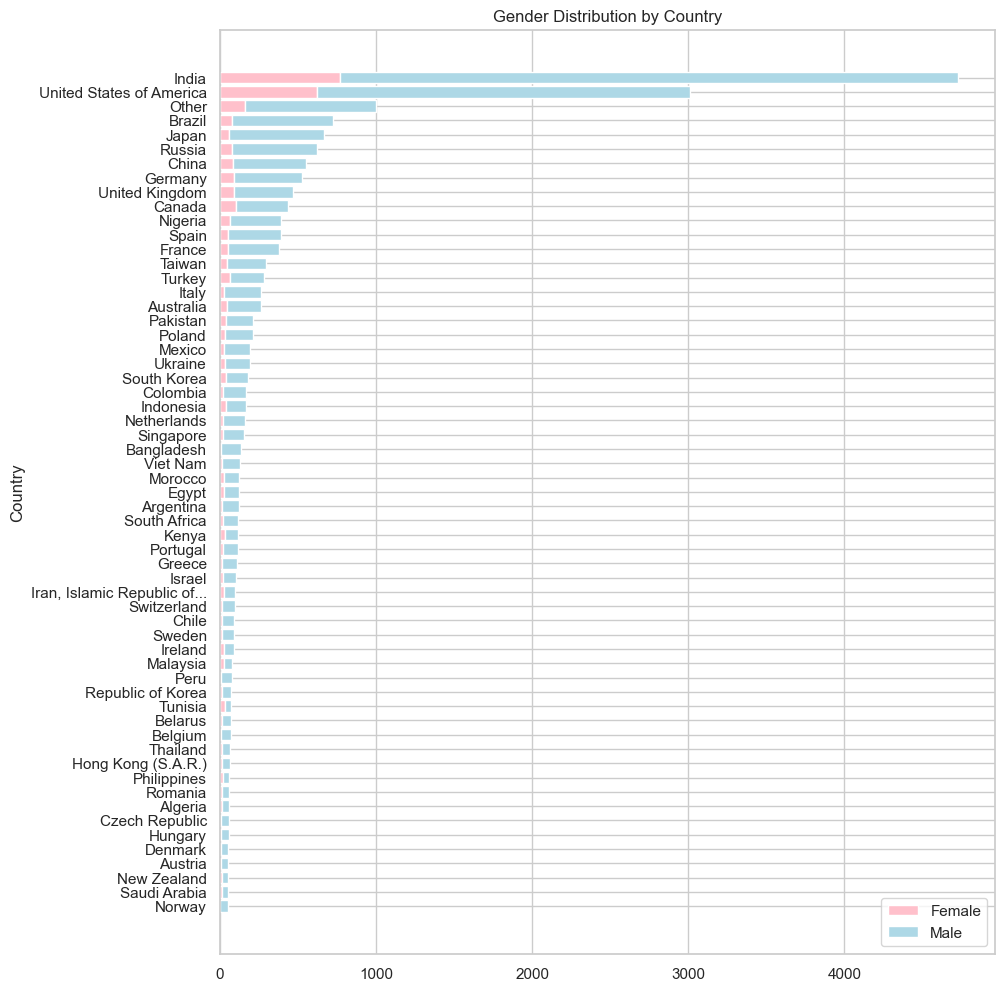
# Data Analysis and Insights

## 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.

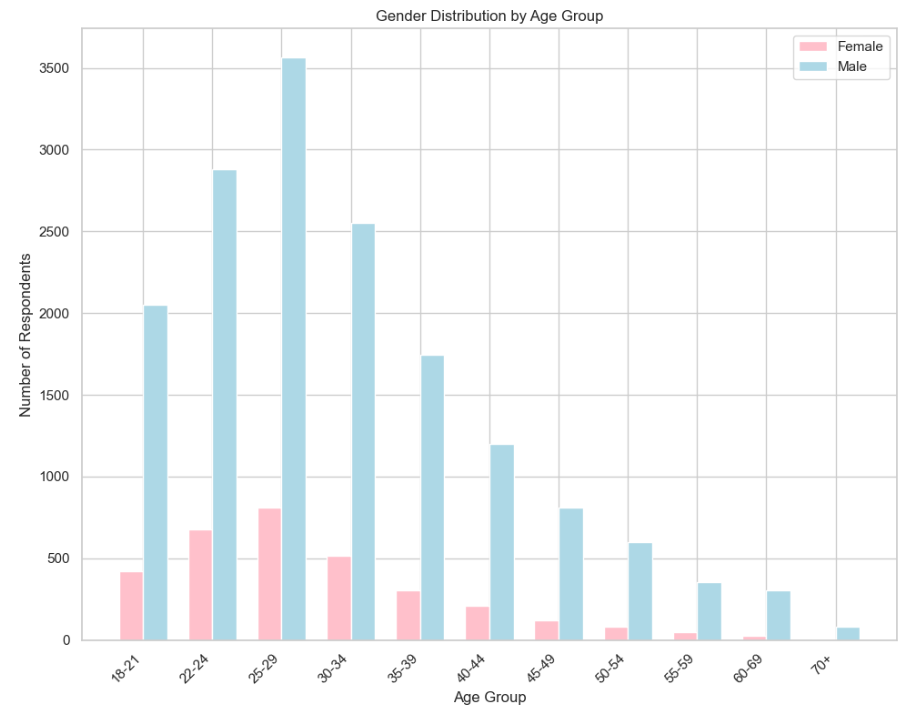
 

An initial analysis of the participant count by country reveals that India and the United States have the highest numbers, followed by Canada, Germany, and the United Kingdom. This is expected since the densely populated countries with a strong technological presence tend to attract higher participation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

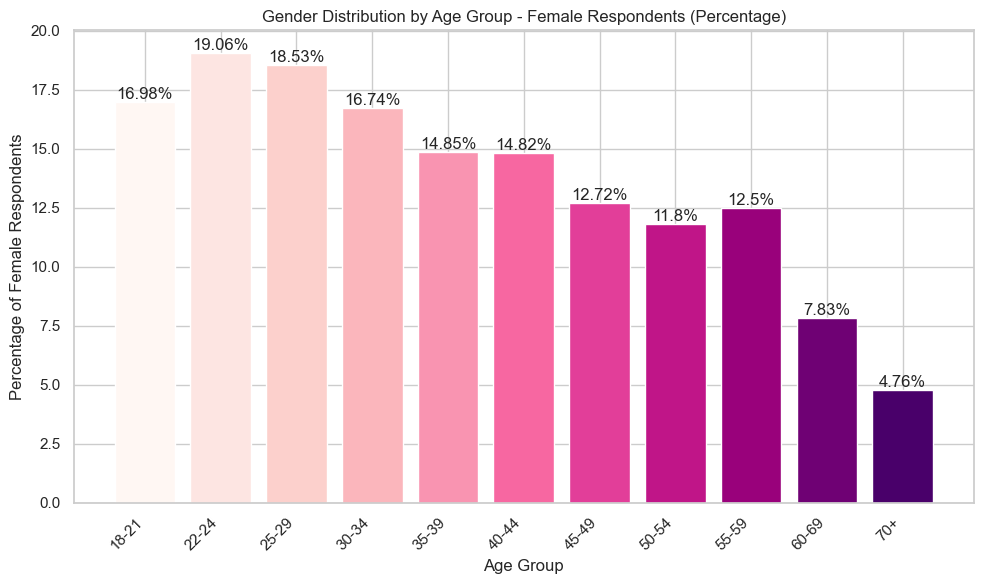
 

## Age Distribution

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.



El porcentaje de mujeres encuestadas dentro de cada uno de los once grupos de edad sigue un patrón similar con algunas diferencias interesantes. En general, parece que hay menos mujeres en grupos de edad más altos, al igual que hay menos mujeres mayores en tecnología. Sin embargo, los porcentajes más altos no se encuentran entre los jóvenes de 18 a 25 años, sino entre los encuestados en los veinte años. Sería interesante averiguar si esto es un cambio general en la cantidad de mujeres en tecnología



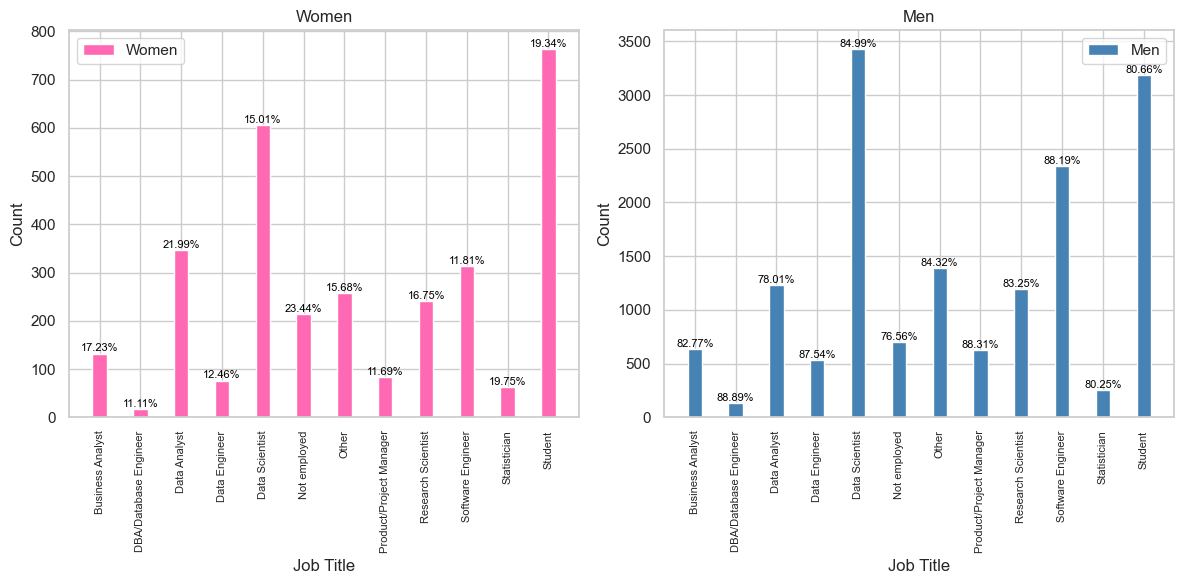
# General Data Analysis [proponer un nombre alternativo a este capítulo]

An in-depth analysis of the data reveals disparities in hiring, promotion, and compensation based on gender.

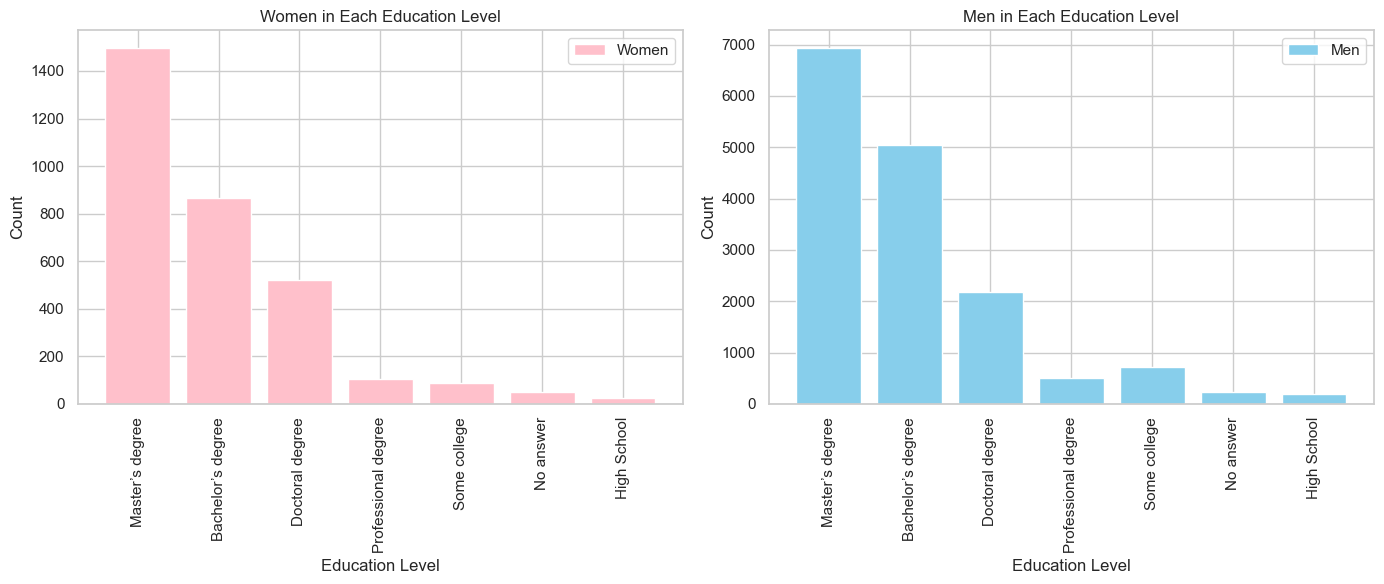
¿Qué hacen para vivir aproximadamente el 16.5% de las mujeres que participaron en la encuesta? ¿Cuántas de ellas están en tecnología o todavía son estudiantes, qué tan grandes son las empresas para las que trabajan y hay diferencias en la compensación en comparación con los hombres o entre diferentes títulos laborales?

## Career and Education

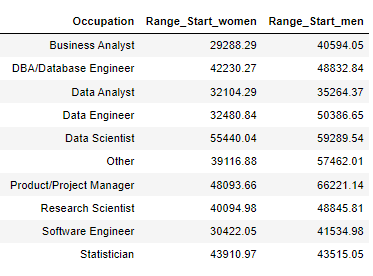
Un examen más detenido de la distribución de los títulos laborales revela que la categoría más grande son los estudiantes, lo cual es probablemente debido al mayor número de encuestados en grupos de edad más jóvenes. Los estudiantes son seguidos por tres posiciones técnicas, Científicos de Datos, Analistas de Datos e Ingenieros de Software. Estos tres grupos, aunque no son la mayoría, constituyen el subconjunto más grande de títulos laborales y confirman que muchas de las encuestadas femeninas en la encuesta de Kaggle están de hecho en carreras STEM. Sin embargo, dentro de los cinco títulos laborales más frecuentes, también hay un gran número de encuestadas femeninas en otros campos.



La mayoría de los participantes en la encuesta han completado al menos o están trabajando en una licenciatura, con un alto número de encuestadas que han pasado a estudios de posgrado. Las encuestadas seleccionaron con más frecuencia un título de maestría como su nivel de educación.



## Salary





# Conclusion

An in-depth analysis of the data reveals disparities in hiring, promotion, and compensation based on gender.

# References

Paul Mooney. (2019). 2019 Kaggle Machine Learning & Data Science Survey. Kaggle. <https://kaggle.com/competitions/kaggle-survey-2019>.

Sheryl Sandberg (2013) ‘Lean In: Women, Work, and the Will to Lead’, Alfred A. Knopf.

Sandberg, S. (2013). Lean In: Women, Work, and the Will to Lead. Knopf.

Chou, T. (2018). "Women in Tech: The Facts." Harvard Business Review. Retrieved from [enlace al artículo].

Pao, E. (2017). Reset: My Fight for Inclusion and Lasting Change. Spiegel & Grau.

Smith, J. (2018). "Women in Tech: Breaking Barriers and Overcoming Challenges." Tech Insights, 5(2), 45-62

Johnson, M. (2019). Programadoras: Mujeres pioneras en la era de la computación. Editorial TechPress

A PMT. (2020). Informe Anual sobre la Participación de Mujeres en la Tecnología. Recuperado de https://www.apmt.org/informe-anual