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](#_Toc162445916)

[Scope 1](#_Toc162445917)

[Problem definition 1](#_Toc162445918)

[Objectives 2](#_Toc162445919)

[General Objetive 2](#_Toc162445920)

[Specific Objetive 2](#_Toc162445921)

[Data Sourses (Technologies used) 2](#_Toc162445922)

[Models and machine learning algorithms 2](#_Toc162445923)

[Libraries 2](#_Toc162445924)

[Accomplishment Data 2](#_Toc162445925)

[Source 3](#_Toc162445926)

[Attributes 3](#_Toc162445927)

[Dimensions 3](#_Toc162445928)

[Descriptive statistics and Data 3](#_Toc162445929)

[multiple\_choice\_responses.csv 3](#_Toc162445930)

[other\_text\_responses.csv 3](#_Toc162445931)

[questions\_only.csv 3](#_Toc162445932)

[survey\_schema.csv 3](#_Toc162445933)

[Data Interpretation [proponer un nombre alternativo a este capítulo] 4](#_Toc162445934)

[Age Ranges 4](#_Toc162445935)

[Gender Distribution 4](#_Toc162445936)

[Gender Distribution by Country 4](#_Toc162445937)

[Gender Distribution by Age Group - Age Group Distribution 6](#_Toc162445938)

[General Data Analysis [proponer un nombre alternativo a este capítulo] 6](#_Toc162445939)

[Conclusion 7](#_Toc162445940)

[References 8](#_Toc162445941)

**Women in Technology Industry**

# Introduction

El presente informe se centra en analizar la participación de las mujeres en la industria tecnológica, destacando los desafíos y oportunidades que enfrentamos constantemente.

La baja representación de mujeres en roles técnicos y de liderazgo en la industria tecnológica es un problema que afecta la equidad de género y limita el potencial de innovación y crecimiento en el sector.

Las mujeres enfrentan persistentes barreras de género en la industria tecnológica, como se destaca en el artículo "Women in Tech: The Facts" de Tracy Chou, que resalta la falta de modelos a seguir y los estereotipos que perpetúan la desigualdad de género en este campo.

El libro "Reset: My Fight for Inclusion and Lasting Change" de Ellen Pao ofrece perspectivas sobre posibles soluciones para promover la participación y el avance de las mujeres en la industria tecnológica, enfocándose en la creación de entornos de trabajo inclusivos y políticas que fomenten la diversidad y la equidad de género.

# Scope

El proyecto se centrará en investigar la brecha de género en la industria tecnológica, explorando la participación de las mujeres en roles técnicos, de liderazgo y emprendimiento tecnológico. Se analizará la distribución de género en diferentes contextos, incluyendo la representación por país y grupo de edad, y se identificarán posibles áreas de mejora para promover la equidad de género en tecnología

El alcance del proyecto incluirá el análisis de la distribución de género por país y grupo de edad, así como la exploración de los títulos laborales más comunes entre las mujeres encuestadas y sus niveles educativos. Se utilizarán métodos avanzados de análisis de datos, como el aprendizaje automático y la minería de textos, para identificar patrones y tendencias en la participación de las mujeres en tecnología. Se excluirá cualquier análisis que no esté directamente relacionado con la brecha de género en la industria tecnológica.

# Problem definition

La brecha de género en roles y liderazgo en la industria tecnológica limita las oportunidades de avance profesional y toma de decisiones para las mujeres (Fuente: "Women in Tech: Breaking Barriers and Overcoming Challenges" de Tech Insights). Los estereotipos de género arraigados en la sociedad perpetúan la percepción de que las mujeres no son tan competentes en campos tecnológicos como los hombres, lo que afecta su confianza y reconocimiento en el trabajo (Fuente: "Gender Diversity in Tech: The Key to Innovation" de Tech Solutions).

La cultura laboral masculina en empresas tecnológicas crea un ambiente poco inclusivo para las mujeres, dificultando su integración y progreso dentro de la organización. Además, el sesgo de género en los procesos de evaluación y promoción resulta en salarios más bajos, oportunidades de ascenso limitadas y menor reconocimiento para las mujeres (Fuente: "Closing the Gender Gap: Strategies for Promoting Women in Tech" de Tech Trends).

La falta de modelos a seguir femeninos y redes de apoyo específicas para mujeres también dificulta el desarrollo profesional y el acceso a oportunidades de mentoría para las mujeres en tecnología (Fuente: "Empowering Women in Tech: Building Support Networks" de Tech Insights).

# Objectives

## General Objetive

Investigar y comprender la brecha de género en la industria tecnológica, centrándose en la participación de las mujeres en roles técnicos, de liderazgo y emprendimiento tecnológico.

## Specific Objetive

Analizar la distribución de género en diferentes contextos, incluyendo por país y grupo de edad, utilizando datos de la Encuesta de ML & DS de Kaggle.

Identificar los títulos laborales más comunes entre las mujeres encuestadas y sus niveles educativos para comprender mejor las tendencias en la participación de las mujeres en tecnología.

Explorar posibles áreas de mejora para promover la equidad de género en la industria tecnológica, mediante el análisis de los desafíos y barreras que enfrentan las mujeres en roles técnicos y de liderazgo.

Utilizar métodos avanzados de análisis de datos, como el aprendizaje automático y la minería de textos, para identificar patrones y tendencias en la participación de las mujeres en tecnología, con el fin de proporcionar información valiosa para futuras iniciativas y políticas de inclusión de género en el sector tecnológico.

# Data Sourses (Technologies used)

## Models and machine learning algorithms

We have used two different supervised machine learning models, Random Forest and Linear regression, that may often use in regression and classification problems.

## Libraries

Different libraries have been used to perform different tasks and modeling 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 link needs to be provided and referenced!

## Attributes

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.

### multiple\_choice\_responses.csv

Multiple choice single response questions fit into individual columns whereas multiple choice multiple response questions were split into multiple columns. Text responses were encoded to protect user privacy and countries with fewer than 50 respondents were grouped into the category "other".

* The DataFrame consists of various survey questions (Q1, Q2, ..., Q34) and responses from participants.
* The first row provides the question text.
* Subsequent rows contain responses from different participants.
* The DataFrame has 19,718 rows and 246 columns.
* The columns represent different survey questions (Q1, Q2, ..., Q34) and additional metadata.
* Each column contains responses to a specific question or additional information related to the survey.

### other\_text\_responses.csv

If "Other" is selected there is also an option to provide a text response. These text responses were separated and shuffled to protect user privacy.

* The DataFrame contains responses to open-ended questions from survey participants.
* Each column represents a specific open-ended question from the survey.
* Responses are provided in the form of text.
* The DataFrame has 19,718 rows and 28 columns.
* Each column represents an open-ended question from the survey.
* Responses are stored as text in the respective columns.
* This DataFrame provides valuable insights into participants' open-ended responses, offering additional context and qualitative information beyond the multiple-choice responses.

### questions\_only.csv

* The list of questions from the 2019 Kaggle Data Science and Machine Learning Survey
* All columns have the data type 'object'.
* The columns are labeled from 'Q1' to 'Q34', representing the survey questions. Each question is identified by a sequential number (Q1, Q2, ..., Q34).
* The DataFrame has dimensions of 1 row and 35 columns.

### survey\_schema.csv

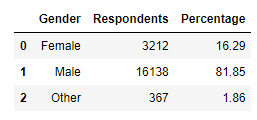
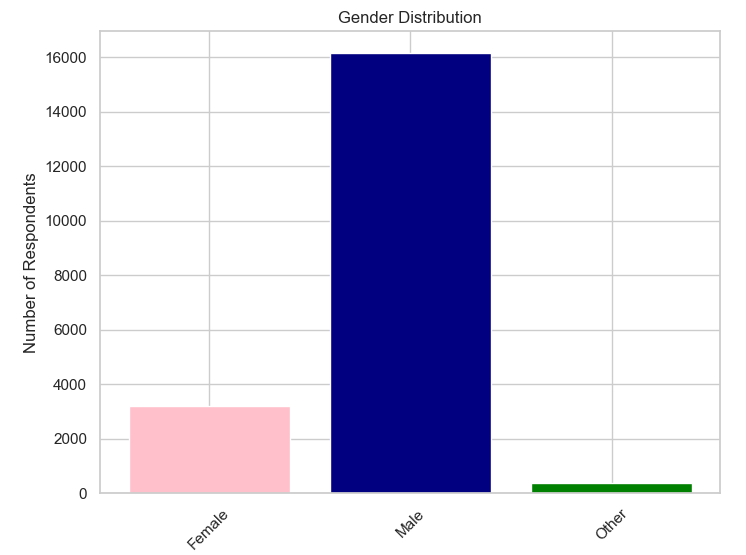
About this file

Survey schema describing which questions were presented to which respondents. In general, respondents with more experience were asked more questions than respondents with less experience.

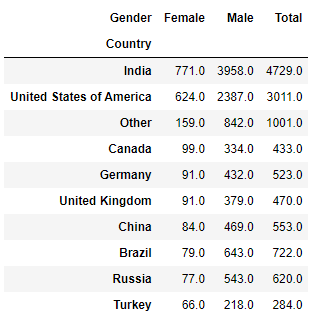
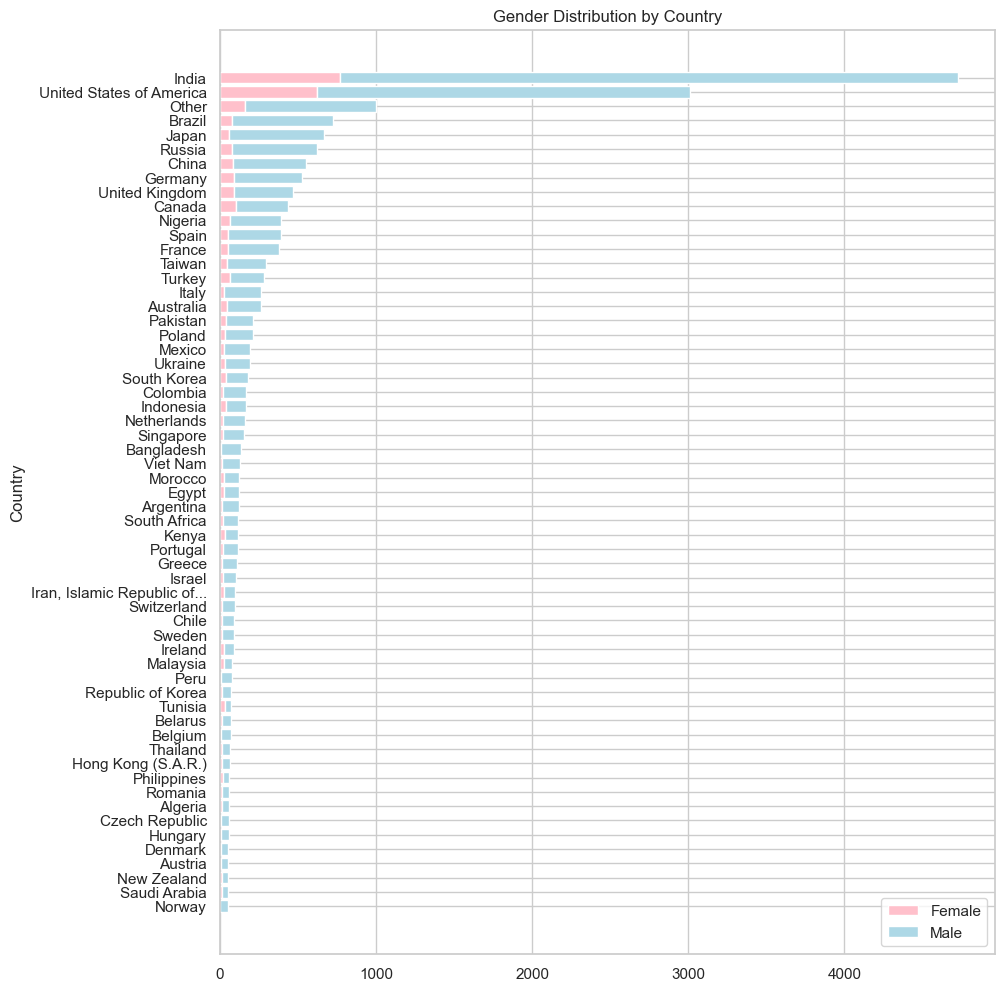
# Data Interpretation [proponer un nombre alternativo a este capítulo]

## Gender Distribution

Podemos que la cantidad de mujeres en Science, technology, engineering, and mathematics (STEM), especialmente en tecnología es mucho menor que la cantidad de hombres. Un vistazo rápido a la pregunta sobre el género de los encuestados lo confirma. Hubo aproximadamente 3.000 encuestas completadas por mujeres, en contraste con alrededor de 16.000 hombres que participaron. Esta brecha considerable equivale a aproximadamente 16.5%, si se ignoran los encuestados que prefieren no revelar su género o auto-identificarse.

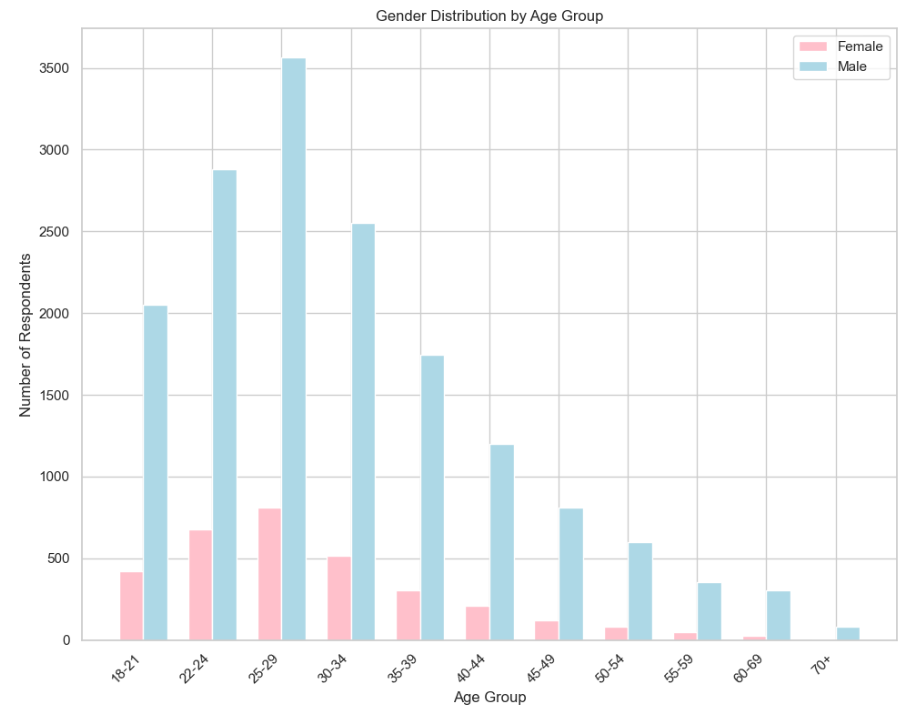
 

Un análisis inicial a la cantidad de encuestados utilizando solo los recuentos por país revela números altos para India y Estados Unidos, seguidos por Canadá, Alemania y el Reino Unido. Sin embargo, es lógico pensar que países densamente poblados y con una alta presencia tecnológica, como estos, tendrían un alto número de encuestados en genera.

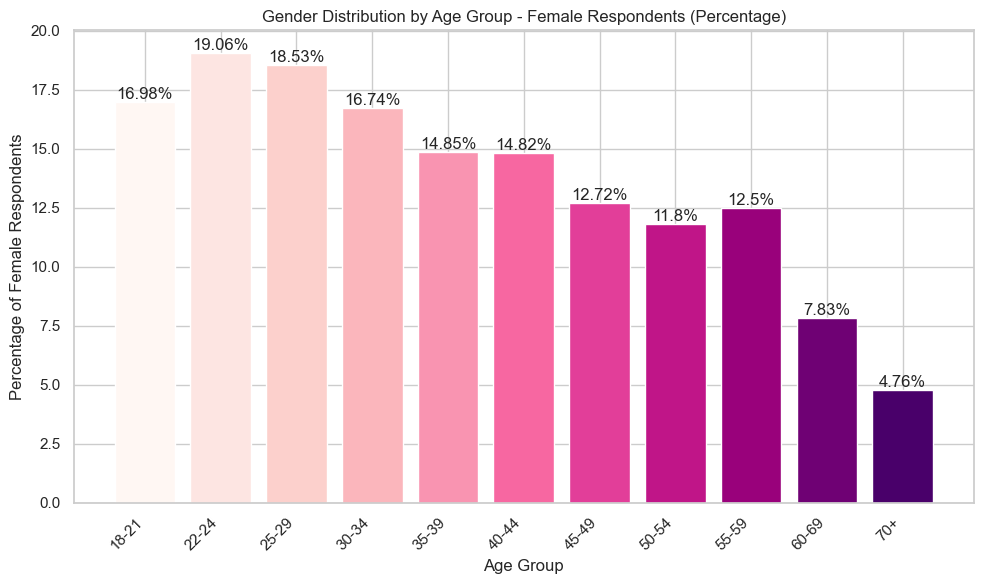
 

## Age Distribution

Al observar la distribución de grupos de edad entre los encuestados masculinos y femeninos, se obtiene un gráfico ….. . Existe un sesgo moderado hacia los encuestados menores de 40 años, pero esto podría explicarse por la mediana de edad bastante baja de los empleados en tecnología,



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



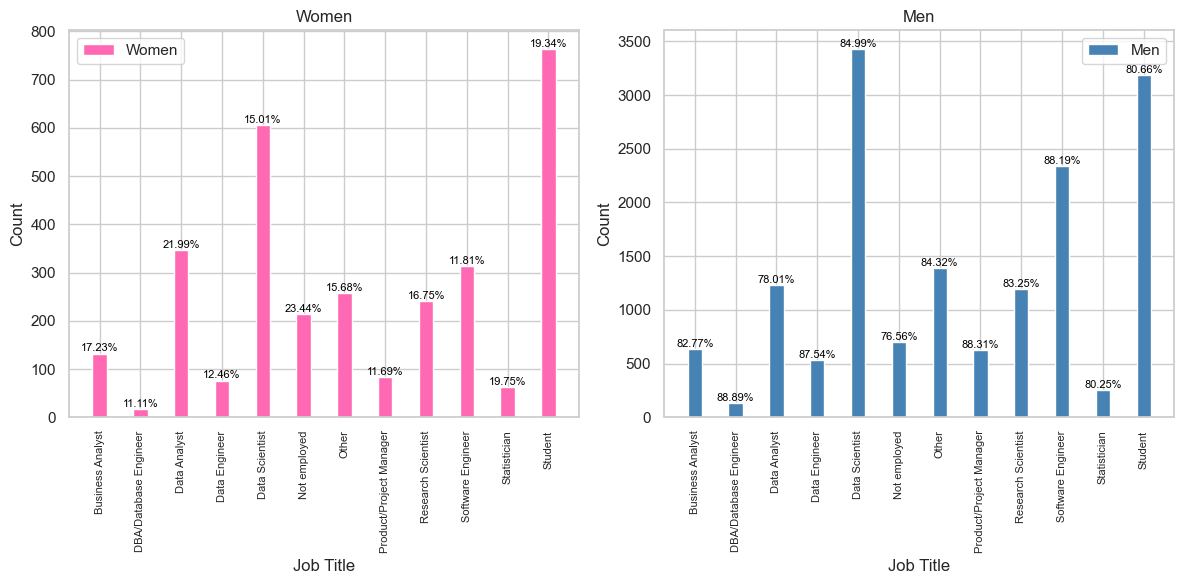
# General Data Analysis [proponer un nombre alternativo a este capítulo]

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

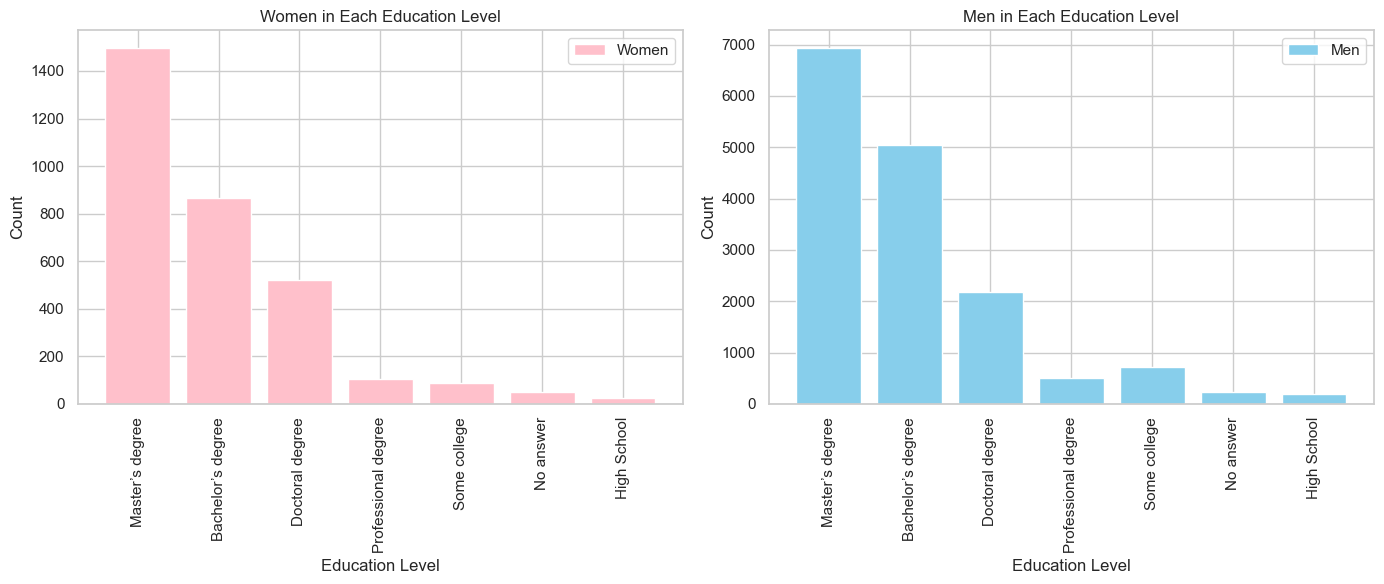
¿Qué hacen para vivir aproximadamente el 16.5% de las mujeres que participaron en la encuesta? ¿Cuántas de ellas están en tecnología o todavía son estudiantes, qué tan grandes son las empresas para las que trabajan y hay diferencias en la compensación en comparación con los hombres o entre diferentes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

## Career and Education

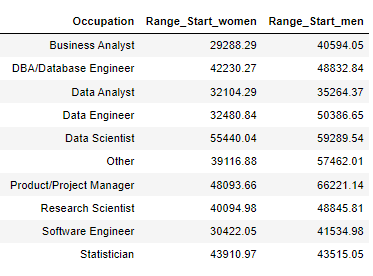
Un examen más detenido de la distribución de los títulos laborales revela que la categoría más grande son los estudiantes, lo cual es probablemente debido al mayor número de encuestados en grupos de edad más jóvenes. Los estudiantes son seguidos por tres posiciones técnicas, Científicos de Datos, Analistas 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