# **Introduction**

The following research question will guide the focus of this study: ‘To what extent, if at all, has the percentage of women in tech increased over the time span of 2014 to 2018?’. To answer this research question, analysis will be conducted on a data set that accurately represents this development.

Despite a visible push in support of progress for women in tech throughout the past 50 years, a very clear gender divide persists. At present, there are two times as many men in tech as women (Guadalupe, 2018). Currently, there is a very strong drive in the modern society for the increase of women in stem and computing-related areas. Especially the past year, 2020, clearly marked a new wave of feminist mobilization (Bonnie Chiu, 2020), supported through protests rearguing gender equality and empowerment of women. Although it seems like a lot of progress has been made, the numbers show the opposite. For example, in 2015, 57% of all professional work was held by women, and only 25% of all these jobs where computing (Eger, 2016). In this report, this hypothesis I set out will be analysed and visualised using Python, Pandas, Seaborn and the datasets provided.

# **Findings**

I imported and looked over the dataset, then started with two simple bar graphs:

*Fig 2: What is the female % vs. male % in the average US population?*

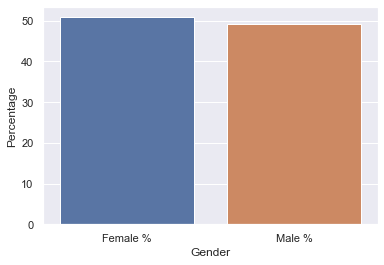


Fig 1: What is the female % vs. male % in the companies in the dataset?

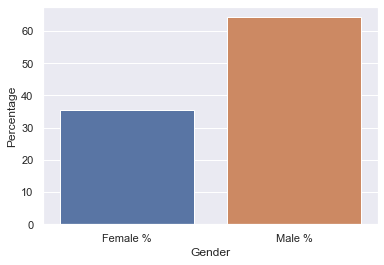


Fig. 1 and Fig. 2 clearly show that women are underrepresented in tech companies, since there is almost half number of women than there are men in tech jobs (Fig 1), yet women represent just over half of the general US population (Fig 2). It is very visible that the mean female percentage (35.36%) is much lower than the male percentage (64.53%).

This very clear outline of the issue lead me into looking into year-by-year changes between 2014 and 2018:

Figure 3: How has the female % changed year on year?

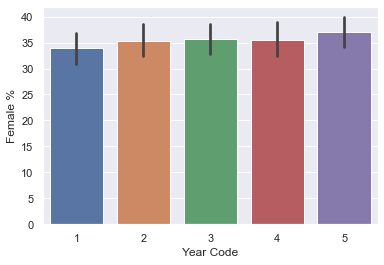


Figure 3 plots the percentage of women year by year. It shows a steady but minute increase. The graph itself has vertical black lines on each of the years, the higher the black line, the more the data in that year deviates from the mean (the standard deviation). A higher standard deviation could be indicative of a wider spread of results.

Figure 4: Has the change in female % increased or decreased each year?

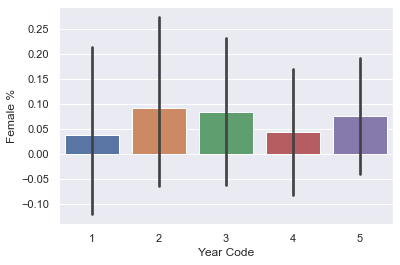


Figure 4 plots the **change** in percentage each year for every year. There is no concrete increasing positive trend, although it is always positive, it fluctuates. This looks promising from an equality perspective since the change in percentage is always positive, However, the percentage change is so small each year that it is almost meaningless in terms of actual parity.

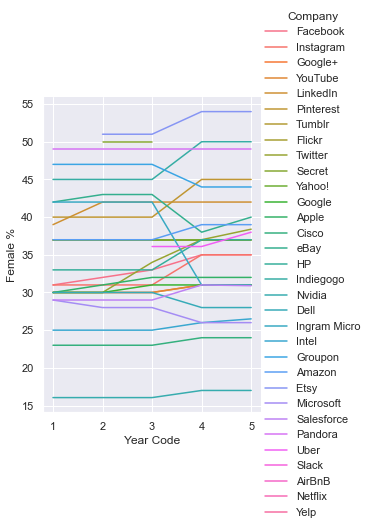
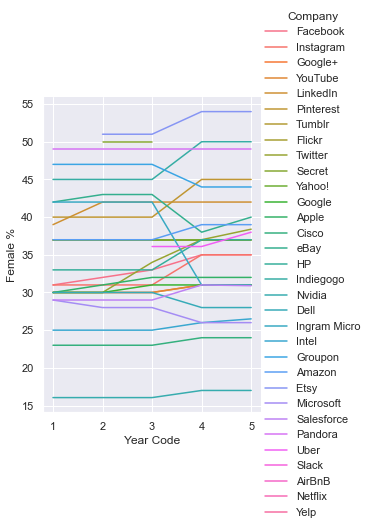
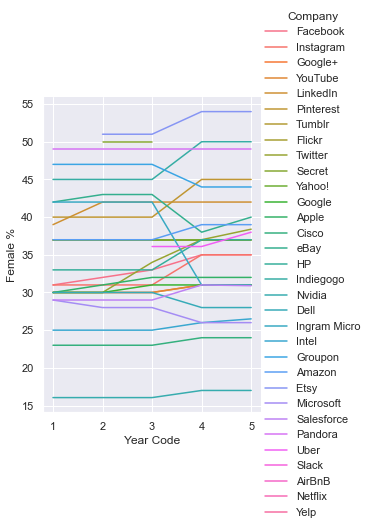


Figure 5: What companies have changed their percentage in female employees?

To lead on from this, I was interested in what specific companies have changed their percentage of female employees over the five-year span. The line graph in Figure 5 clearly showed the year-to-year development of all tech and social media companies, visualising an increase of female workers over the years for most companies such as Instagram and LinkedIn. However, there are companies such as Microsoft and Dell where the percentage of female workers has decreased.

I then went on to target two specific companies that have increased and decreased their percentage of female employees: Instagram and Microsoft:

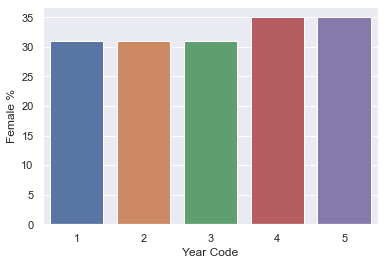


Figure 6: Has Instagram increased or decreased its female diversity over the years?

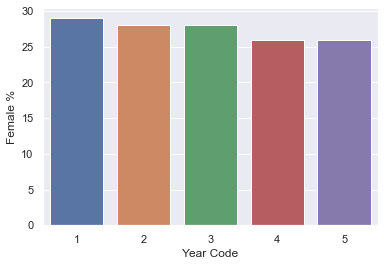


Figure 7: Has Microsoft increased or decreased its female diversity over the years?

Clearly it can be seen that Instagram has made small but positive progress as oppose to Microsoft, who have decreased their percentage of female employees.

# **Methodology**

The initial dataset I was given was from 2018 only contained one year of data, therefore it wasn’t sufficient to answer my research question. So I successfully found a similar data set in terms of structure and content on informationisbeautiful.net which is a site known for well sourced and reliable content.

Firstly, the data needed to be cleaned up in Excel. To do this, I collected all the years’ worth of data into one sheet and created a new column for the year that data was in (a “Year Code”); I also made a column for the “type” of data, (e.g. if it was a Tech or Social Media Company); and finally I replaced dashes (“-”) with zeroes (0) and fields that contained “<1” were replaced with 0.01 to ensure that Python understood these cells as intergers.

Following the clean-up, the data set was analysed within Python, since it is a powerful general purpose programming language, easily human-readable since it’s a high-level language (Meyers, 2012). Pandas and Python were chosen for this project due their power and simplicity, as well as their use of objects such as data frames that allow for features such as sorting/filtering for particular columns, rows, or values.

Furthermore, creating graphs was essential because they are a powerful way to show numerical data, correlations, and changes over time that would otherwise be very hard to visualize (Waskom, 2020).

# **Conclusion and evaluation**

After analysing the Diversity in tech datasets from 2014 to 2018, I concluded that the percentage women in tech are generally increasing, however results have shown that for most companies the number of women in tech have gone up at an extremely small rate. These are very interesting results because they suggest that although there is a forceful push for it through all this diversity and feminism action, there is very little real world change in this sector.

All in all, the study has been successful in answering the research question, however there was no data from the early 21st century, which would have made the comparison even more interesting. If this analysis were to be repeated or expanded in the future, a potential research question could be whether the percentage of women in tech will increase in the future, or perhaps an understanding of **why** women take less jobs in tech.

In conclusion, even though progress has been made, gender disparity in the tech industry is present and women are strongly underrepresented in this field.

# **Bibliography**

Ashcraft, C., McLain, B. and Eger, E. (2016). WOMEN IN TECH: THE FACTS. See what’s changed and what hasn’t.. 1st ed. national center for women and information technology, chapter 1-2.

Bowen, G. (2009). Document Analysis as a Qualitative Research Method. Qualitative Research Journal, [online] 9(2), pp.27-40. Available at: https://www.emerald.com/insight/content/doi/10.3316/QRJ0902027/full/html [Accessed 22.04.2021].

Del Carpio, Lucia and Guadalupe, Maria. (2018). More Women in Tech? Evidence from a Field Experiment Addressing Social Identity. IZA Discussion Paper No. 11876, Available at SSRN: <https://ssrn.com/abstract=3273715>[Accessed 5th April 2021]

Docs.python.org. (2019). Sorting HOW TO — \_Python 3.7.3 documentation. [online] Available at: <https://docs.python.org/3/howto/sorting.html> [Accessed 5th April 2021].

Frier, S., (2020). Bloomberg - Are you a robot?. [online] Bloomberg.com. Available at: <https://www.bloomberg.com/news/articles/2020-07-29/snap-commits-to-doubling-women-and-minorities-on-staff> [Accessed 10 April 2021].

J. Elkner, A. B. Downey, and C. Meyers, (2012) How to Think Like a Computer Scientist: Learning with Python [[Available online](http://openbookproject.net/thinkcs/python/english3e/)] [Accessed 5th April 2021].

McCandless, D. (2018). Diversity in Tech — Information is Beautiful. [online] Information is Beautiful. Available at: <https://www.informationisbeautiful.net/visualizations/diversity-in-tech/> [Accessed 10 April 2021].

Renee B Adams, Brad M Barber, Terrance Odean. (2018). STEM Parents and Women in Finance Financial Analysts Journal, volume 74, issue 2.

Selvaraj, N. ( 2020). The Gender Gap in Data Science: What the Data Says. [online] Medium. Available at: <https://betterprogramming.pub/the-gender-gap-in-data-science-what-the-data-says-2a74892655f1> [Accessed 10 April 2021].

Waskom, M. (2021). seaborn: statistical data visualization — seaborn 0.11.1 documentation. [online] Seaborn.pydata.org. Available at: <https://seaborn.pydata.org/index.html> [Accessed 10 April 2021].