Women In Tech

By: Claudia Feochari
COMP 499- Summer 2019

Raising The Glass Ceiling for Women in STEM

According to Engineers Canada, only 20% of Engineering Students across Canada are Females.

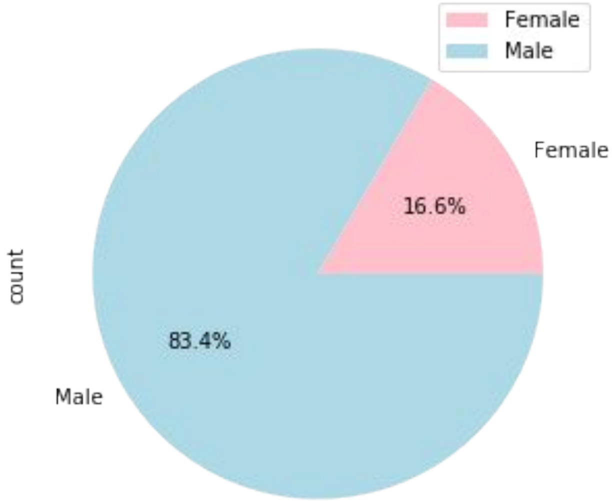
Dr. Gina Parvaneh Cody donated $15 million to the Faculty of Engineering Concordia University.

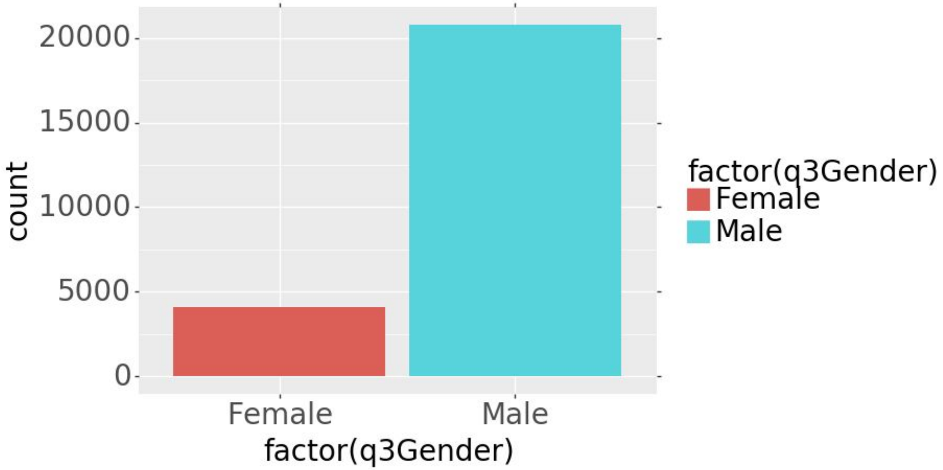
She did it as she hopes that barriers are broken down for women who want to pursue careers in STEM. She was the first woman who got a PHD in Engineering in 1989 at Concordia University.

Concordia will use part of the money for the creation of a equity, diversity and inclusion programming and increase research in Data Analytics and Ai.

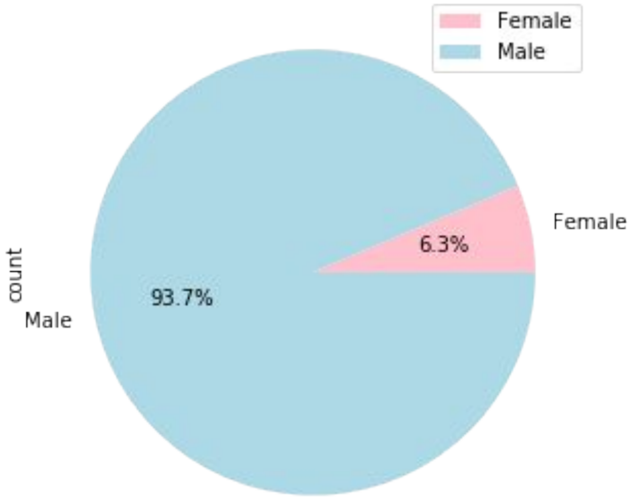
Source : https://www.cbc.ca/news/canada/montreal/concordia-gina-cody-engineering-school-1.4835855

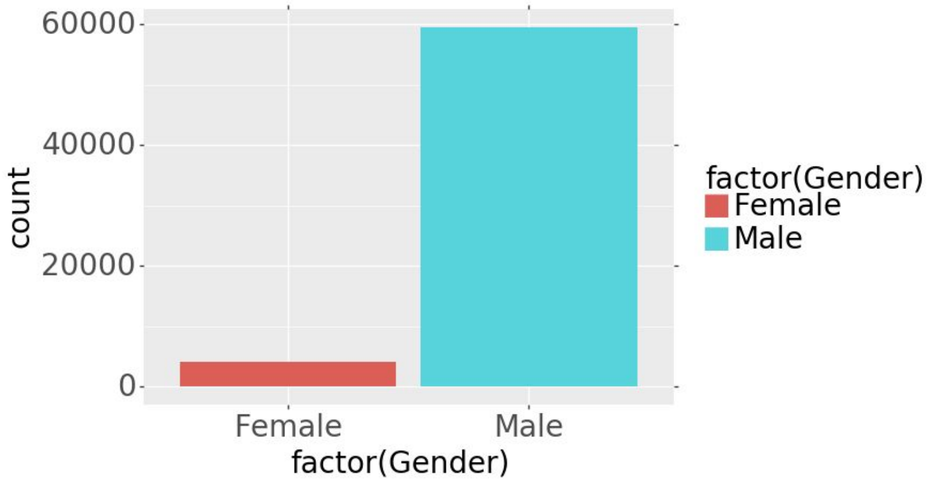
HackerRank Survey 2018





StackOverflow Survey 2018





So What Is the Point?

Due to Dr. Cody's recent donation to Concordia University, this got me curious to see what types of skills women in STEM possess as well as their main points of interest in the tech industry and to discover if inequalities still exist.

It also got me curious to discover how women stack up relative to men in the tech industry and to see if inequality between these two groups still exist.

This data analysis might also be useful for hiring managers as it would give them a better idea to understand this minority group better.

This data might also be relevant for gender equality and inclusivity groups that create specialized programs to try to appeal to women the best that they can.

Questions

At What Age do Women compared to Men Start Learning how to code ? Why this disparity?

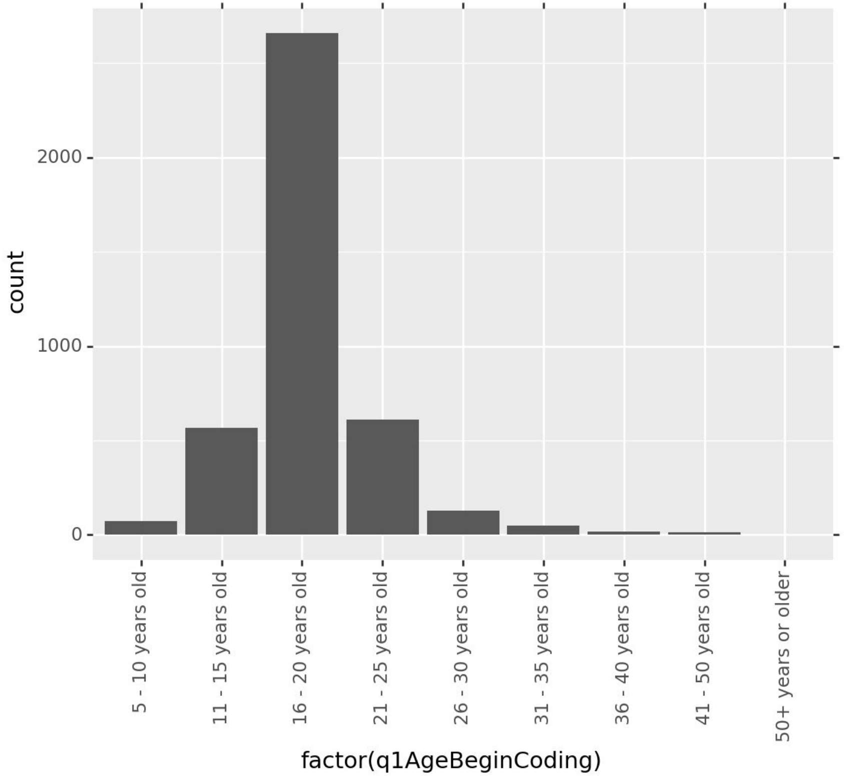
What type of technologies do women prefer to code in ? Do they prefer Front-end / Back-end or Full-Stack? What Coding Languages are most popular for Women?

What level of education is optimal for women to have in order for them to reduce their level of unemployment.

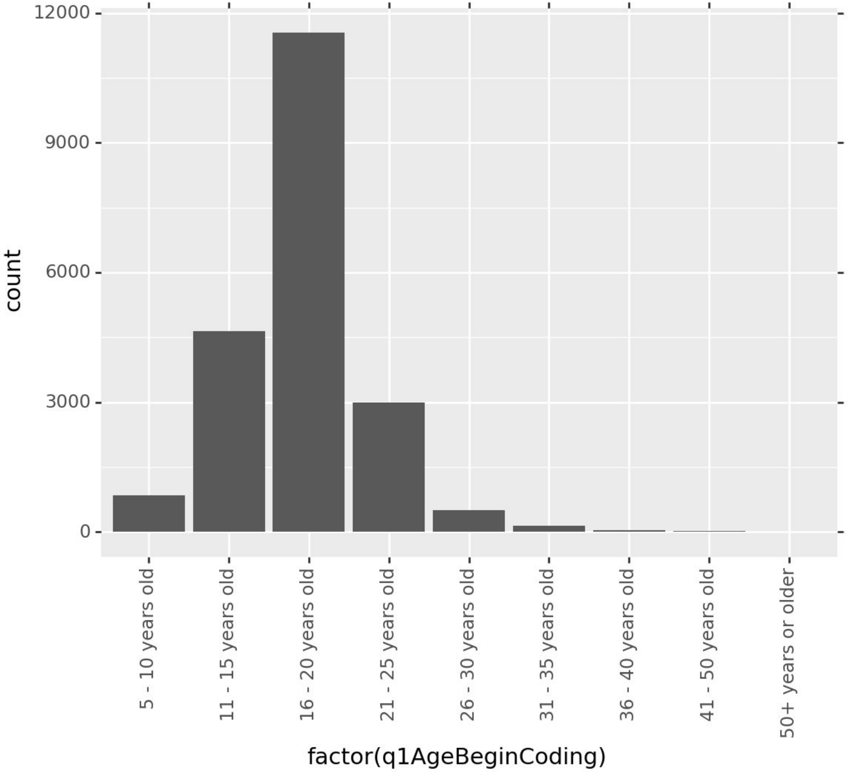
How much do women make on average? How is this compared to men? Do men still really make more than women?

Age Distribution When Participants First Learn to Code

Age Distribution When Women First Learn to Code



Age Distribution When Men First Learn to Code



What can we say about our Data?

Women typically learn to code at a later age than their male counterparts as the graph shows that a greater proportion of men learn to code from 11-15 years old than women.

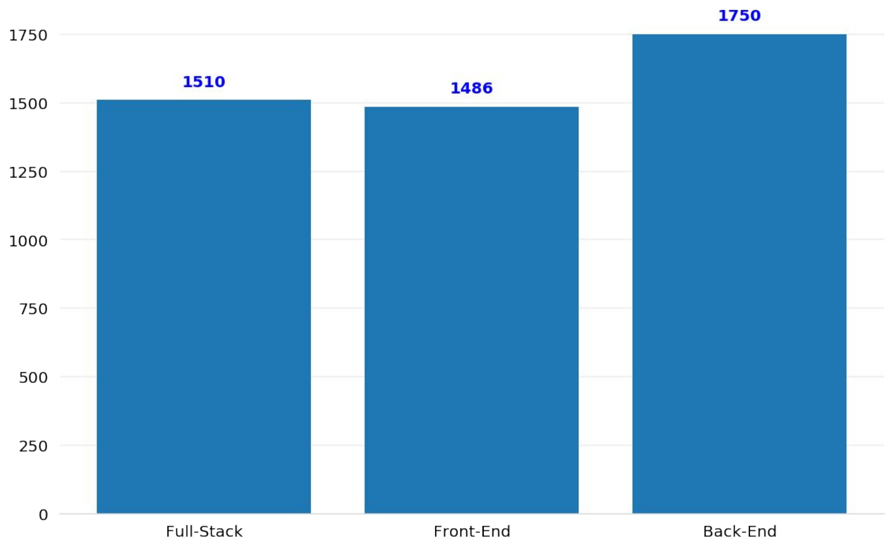
This may be caused by the fact that these early pre-teen and teenage years women may be influenced by gender stereotypes to steer away from male dominated interests like STEM related technologies such as learning how to code.

In the future it is predicted this will increase as new inclusivity programs for female students are being implemented in elementary and high schools. **Kids Code Jeunesse**, a non-profit I work for is implementing these programs.

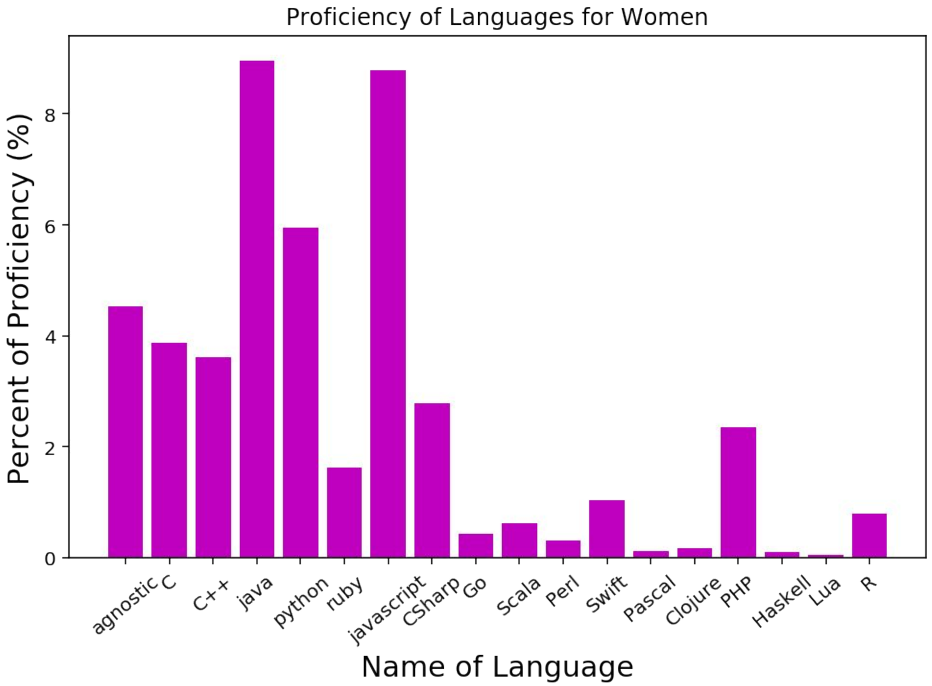
From 16-20 years old, men and women are equally proportional in terms of learning how to code then. This may be due to the fact that around this age, females are preparing to go to college or start their first years in college.
Since college is a more liberating environment for people in general as the social pressures of high school being gone, women might be more prone to try something new without the fear of being judged (aka learning how to code) and pursuing STEM related majors.

Programming Languages Women Prefer

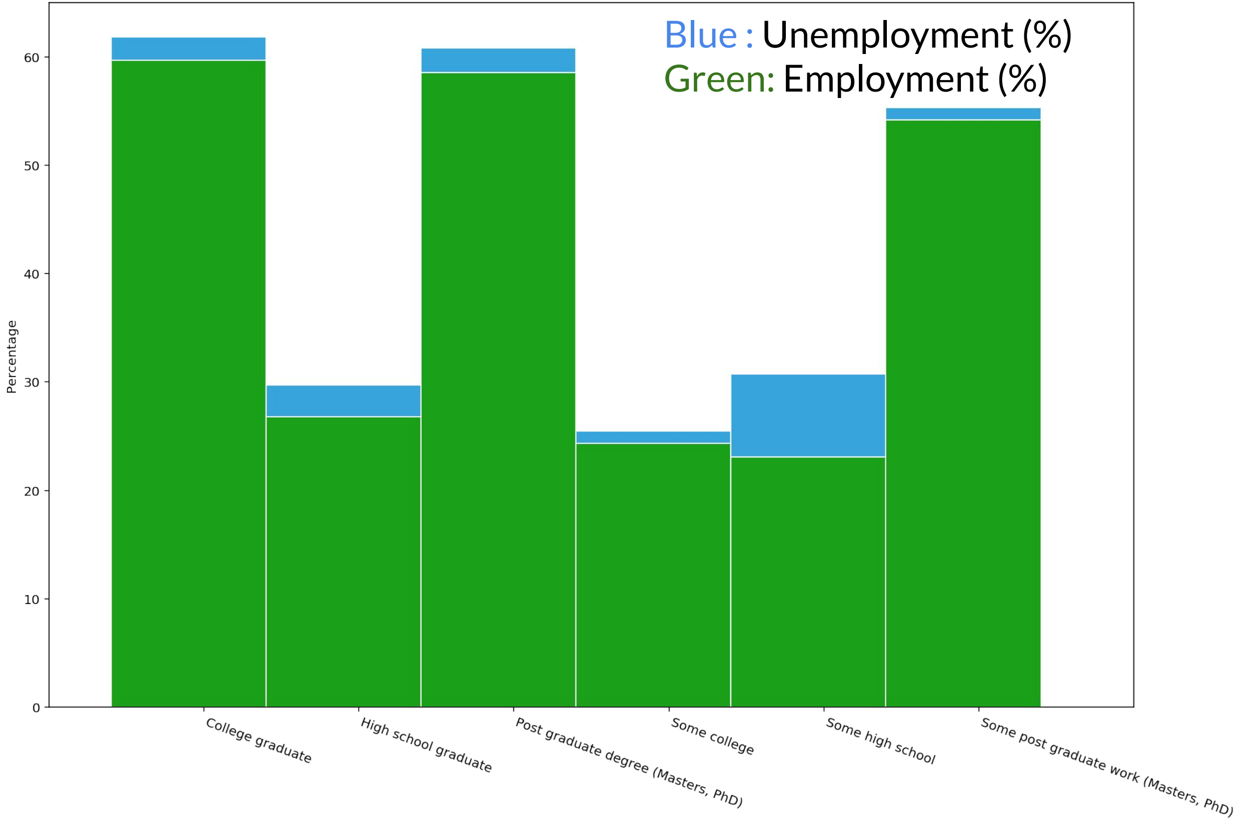
Women Prefer working in Back-End over Front-End and Full-Stack



The Proficiency Level for Women is mostly geared towards Java and JavaScript



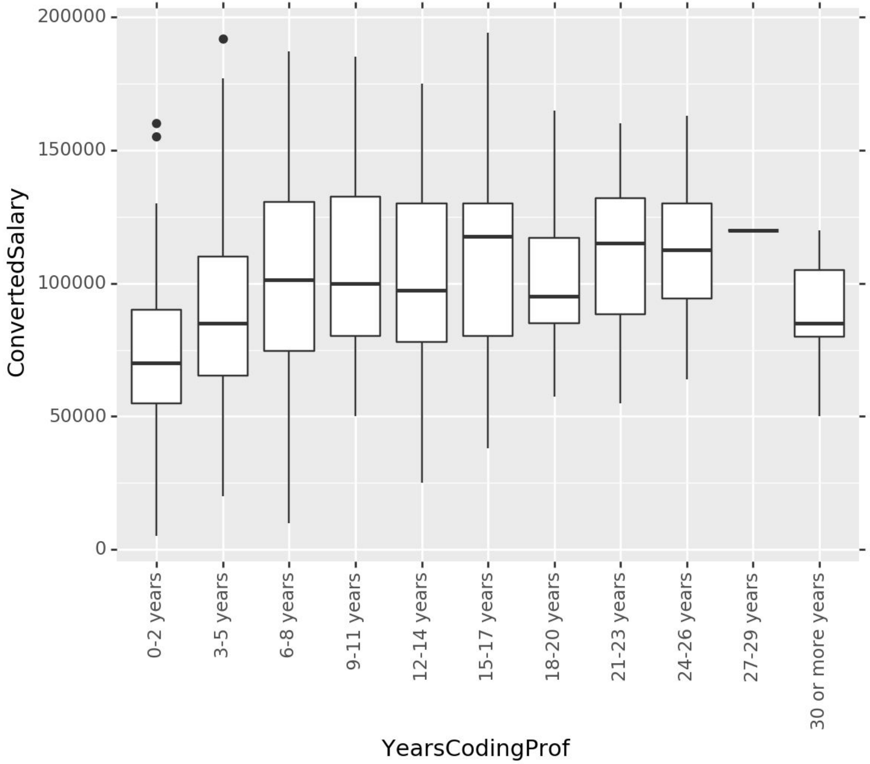
Does Level of Education Matter?



According to this graph, if women want to be ensured that they are the least likely to be unemployed, they should do only some college or finish college and only do some post graduate work and not complete their post graduate degree.

As expected, if a female does not finish high school, she will most likely be unemployed as employers usually want their employees to have some level of basic education

Years of Experience and Annual USD Salary Earned

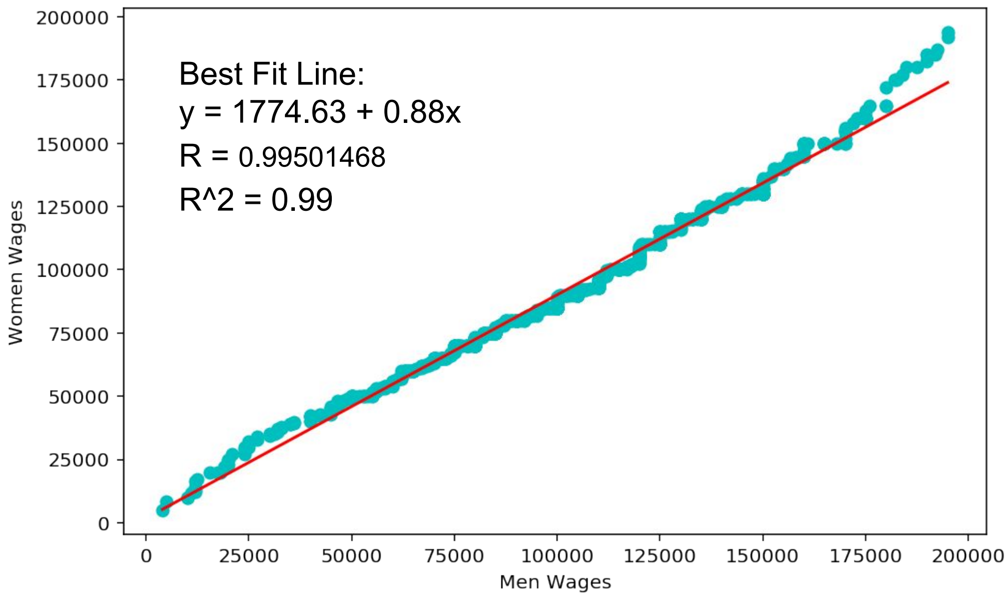


As expected in the majority of the cases in this graph, the number of years a female has of professional coding experience in general leads her to have a higher salary.

Although as the boxplot shows, the more experience a woman may have especially over 30 years may hinder her as she will be paid less as she may be deemed “over-qualified” by certain companies who do not want to pay extra for females with too much experience

Do women really get paid less?

n=700 Sample Size



- Men make an average of $99566.16 USD
- Women make an average of $89710.97 USD

Random sample of 700 men and 700 women was taken and their salaries were ordered in ascending order
On Average, women only earned 90% of what their Male counterparts are paid in USD so the inequality in the pay gap between the two genders still exists

Machine Learning Predictions

I wanted to explore more about the salary of the participants of the StackOverflow Survey.

The data that was used was a subset containing a sample size of n=200 of female participants with different worldwide currencies.

The sample included values such as the type of currency and the Salary Type

	Unnamed: 0	Gender	Salary	Currency	Salary Type	Respondent	Label
0	53479	Female	30597	Euros (€)	Yearly	75820	0
1	50145	Female	54000	U.S. dollars ($)	Yearly	71157	1
2	68591	Female	15600	Mexican pesos (MXN$)	Monthly	97374	0
3	28911	Female	25000	U.S. dollars ($)	Yearly	41064	1
4	46663	Female	52000	U.S. dollars ($)	Yearly	66190	1

Female Payment Sample Data (n=200)

After the random sample was taken, a column containing the Labels for each row in the form of 0's and 1's was added.

Since I wanted to use U.S. Dollars ($) for my prediction model, the machine learning algorithm needed to match only the values that contained U.S. Dollars ($) as the currency type.

Labelling a row with the number 1 indicated that the row was a match and labelling a row with the number 0 indicated that the row was not a match.

So, the rows that contained U.S. Dollars ($) with any type of Salary Type were labelled with 1's and those containing the Weekly or Monthly salary were labelled with 0's.

Training Data Set

Training Data Set. This is a dataset that is used by the algorithm to learn the patterns visible in the dataset.

	Unnamed: 0	Gender	Salary	Currency	Salary Type	Respondent	Label
158	56238	Female	40800	U.S. dollars ($)	Monthly	79792	1
191	78532	Female	120000	U.S. dollars ($)	Yearly	79113	1
50	76614	Female	80000	U.S. dollars ($)	Yearly	57106	1
0	53479	Female	30597	Euros (€)	Yearly	75820	0
94	64632	Female	40000	U.S. dollars ($)	Yearly	91773	1

Prediction Models

6 Types of prediction models were evaluated for their effectiveness for the model that was trained. The Best model was Chosen to be Decision Tree as it had the highest score as the number of folds increased . (Fold 5 it was 62.5%)

	Name	Matcher	Num folds	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean score
0	DecisionTree	<py_entitymatching.matcher.dtmatcher.DTMatcher object at 0x1d0c9e2278>	5	0.375000	0.625000	0.555556	0.421053	0.625000	0.520322
1	RF	<py_entitymatching.matcher.rfmatcher.RFMatcher object at 0x1d0c902470>	5	0.352941	0.555556	0.625000	0.428571	0.428571	0.478128
2	LogReg	<py_entitymatching.matcher.logregmatcher.LogRegMatcher object at 0x1d0c902240>	5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	LinReg	<py_entitymatching.matcher.linregmatcher.LinRegMatcher object at 0x1d0c902160>	5	0.000000	0.333333	0.000000	0.181818	0.000000	0.103030
4	NaiveBayes	<py_entitymatching.matcher.nbmatcher.NBMatcher object at 0x1d0c902278>	5	0.750000	0.500000	0.560000	0.400000	0.400000	0.522000
5	knn	<py_entitymatching.matcher.nbmatcher.NBMatcher object at 0x1d0c902208>	5	0.750000	0.500000	0.560000	0.400000	0.400000	0.522000

Supervised Learning : Decision Tree Prediction

Precision : 78.26% (18/23)

Recall : 45.0% (18/40)

F1 : 57.14%

False positives : 5 (out of 23 positive predictions)

False negatives : 22 (out of 77 negative predictions)

As you can see in the table, the Decision Tree Algorithm can somewhat predict that the data we want is U.S dollars($) and not some other form of currency. Although there are some errors in prediction, the overall result is very precise at 78.26%

	Unnamed: 0	Gender	Salary	Currency	Salary Type	Respondent	Label	predicted
18	15030.0	Female	40392.0	Euros (€)	Monthly	21394	0.0	0.0
170	12535.0	Female	58000.0	U.S. dollars ($)	Yearly	17871	1.0	0.0
107	43886.0	Female	65000.0	U.S. dollars ($)	Yearly	62282	1.0	0.0
98	15187.0	Female	72500.0	U.S. dollars ($)	Yearly	21610	1.0	0.0
177	17832.0	Female	70000.0	U.S. dollars ($)	Yearly	25358	1.0	0.0
182	18538.0	Female	31800.0	Singapore dollars (S$)	Monthly	26331	0.0	0.0
5	20203.0	Female	66000.0	U.S. dollars ($)	Yearly	28708	1.0	0.0
146	24503.0	Female	69036.0	Danish krone (kr)	Monthly	34853	0.0	0.0
12	33464.0	Female	145000.0	U.S. dollars ($)	Yearly	47527	1.0	1.0
152	24400.0	Female	36720.0	Euros (€)	Monthly	34718	0.0	0.0
61	52817.0	Female	11365.0	Indian rupees (₹)	Yearly	74868	0.0	0.0
125	68051.0	Female	36084.0	Russian rubles (₽)	Monthly	96615	0.0	0.0
180	16981.0	Female	115500.0	U.S. dollars ($)	Yearly	24165	1.0	1.0
154	40227.0	Female	187000.0	U.S. dollars ($)	Yearly	57119	1.0	1.0
80	69360.0	Female	22030.0	Euros (€)	Yearly	98463	0.0	0.0
7	42072.0	Female	9912.0	Euros (€)	Monthly	59682	0.0	0.0
33	38850.0	Female	41612.0	Euros (€)	Yearly	55123	0.0	0.0
130	71161.0	Female	200000.0	U.S. dollars ($)	Yearly	101078	1.0	1.0
37	40358.0	Female	16040.0	Indian rupees (₹)	Yearly	57294	0.0	0.0

Machine Learning Prediction (See Predicted Column)

Conclusion

The data displayed in this presentation will help recruiters decide which jobs would be best for women to succeed in.

Inclusivity and Non-Profit groups like Kids Code Jeunesse use this data to create workshops for which is what I do at work at Kids Code Jeunesse where I work as a Junior Data CRM Integration Specialist and deal with data cleaning and analysis.

There still exists inequality for women in tech as they get paid less overall than their male counterparts.

Let's hope that Dr. Cody’s $15 million dollar for the next generation as well as KCJ will inspire more women to pursue STEM related programs as well as bridge the gender gap for women.

Any Questions?

kids code
jeunesse