Motherhood in the Workforce

Kue Ho
Springboard Data Science Career Track

OUTLINE

- Introduction
- Data Source and Wrangling
- Exploratory Data Analysis
  - Motherhood and gender pay gap
  - Unemployment gap
  - Motherhood in job sectors
- Statistical Analysis
- Predictive Analysis
  - Feature selection
  - Predict income levels
  - Predict job sector
- Summary
- Future Improvements

Introduction

One societal norm is for people to have kids as they get older. According to the CDC, in 2014 the average age of first-time mothers increased to 26.3[(1)]; however, women seem to carry most of the costs of parenthood. They spend nine or so months carrying and developing the fetus to term and then have to go through labor, delivery and recovery. Do the cost for women end there, or are there ways parenthood impacts women in the workforce more than their male counterparts?

The gender pay gap has been proven and well-studied. Pay disparity between men and women is a real thing. The question is does having children and all variables that are involved add to this disparity? Is the difference in sex the only reason why men and women get paid differently, or are there other factors such as family that play into it? With that in mind, are there characteristics of companies that are more friendly towards woman, particularly those with families?

[(1)] https://www.cdc.gov/nchs/data/databriefs/db232.pdf

Target Audience

- Individuals who desire children and are looking to pursue a career in STEM (anyone from high schoolers to college students who are thinking about their future as well as individuals who are thinking of a career change).
- Nonprofits who are interested in this realm of data and information.
- Companies who are looking to increase the gender and personality diversities of their workforce and stay competitive in the business world.
- Recruiters and job hunting agencies/resources (such as LinkedIn, Indeed, etc.) would be interested in this project as they work to sell individuals on the companies they work for/are recruiting for.
- HR departments would also be interested as they seek to change their company's work culture and create space for the ever diversifying workforce.

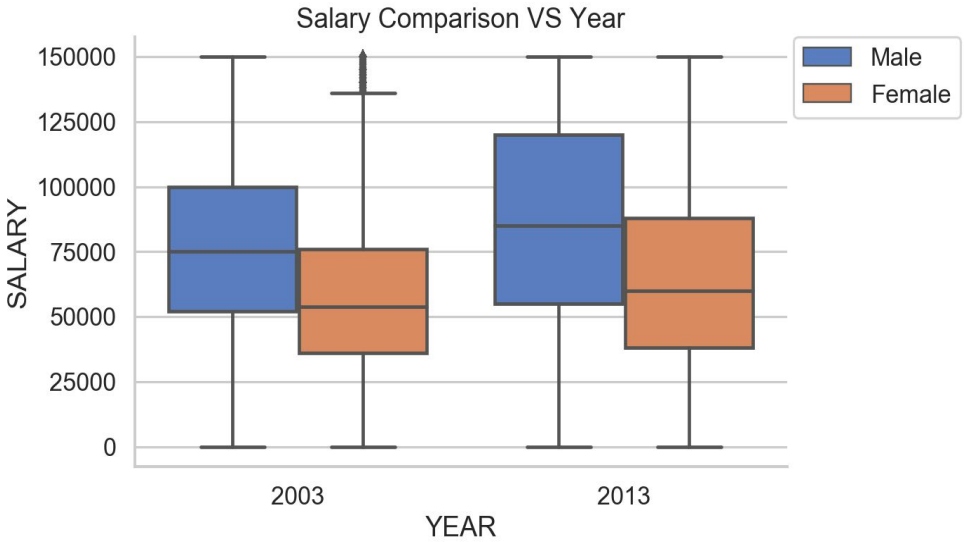
Data Source and Wrangling

- Survey from IPUMS-Highered website: https://highered.ipums.org/highered/
- Survey data focused on scientists and engineers
- Relevant variables such as demographics, employment, education and number of children were selected from 2003 and 2013
- Columns were filtered and null values dropped.
- Final data size: 206,703 entries with 18 columns; size 28.4 MB

Exploratory Data Analysis

Gender pay gap - 10 year difference

The boxplot shows the median and IQR for both men and women in the years 2003 and 2013. Although the salary has increased for both men and women, there still exists a gap. At initial glance, we see that men get paid more than women. In 2003 and 2013, we see that max salary for women was barely median salary for men. Furthermore, an increase in salary change favored men.



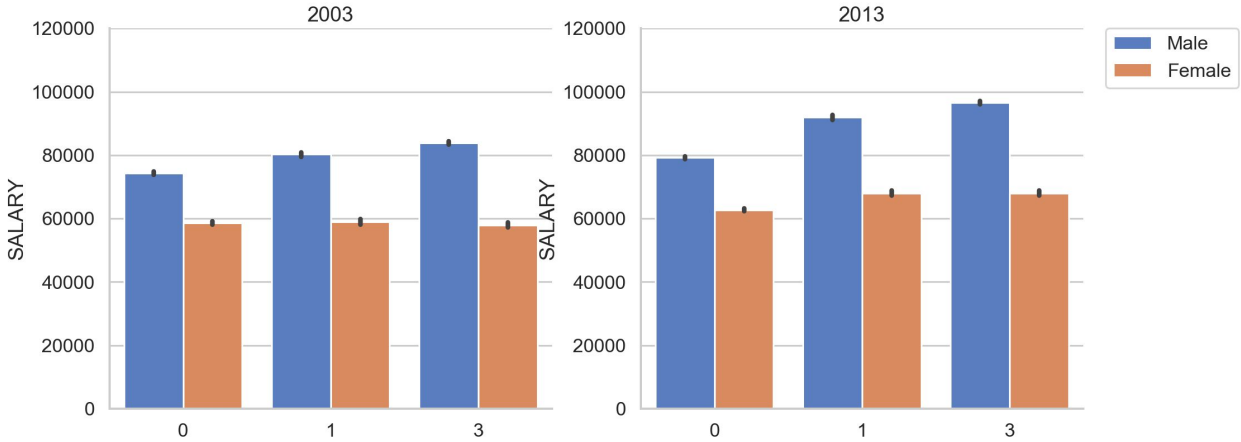
**Note that salary can be affected by location, which is missing from this dataset.

Exploratory Data Analysis - cont.

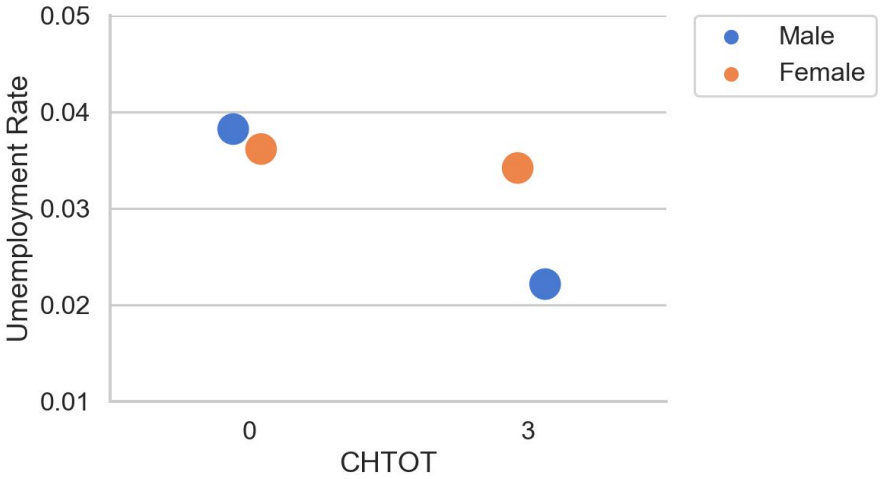
Motherhood pay gap - 10 year difference

The boxplots show that the salary for women who have children do not significantly exceed women without children. On the other hand, men experience a salary increase as the number of children they have increases. This begs the question, why are women with children, who are theoretically older and have more experience still making the same as women who are younger and do not have children?

**Note that salary can be affected by location, which is missing from this dataset.



Exploratory Data Analysis - cont.



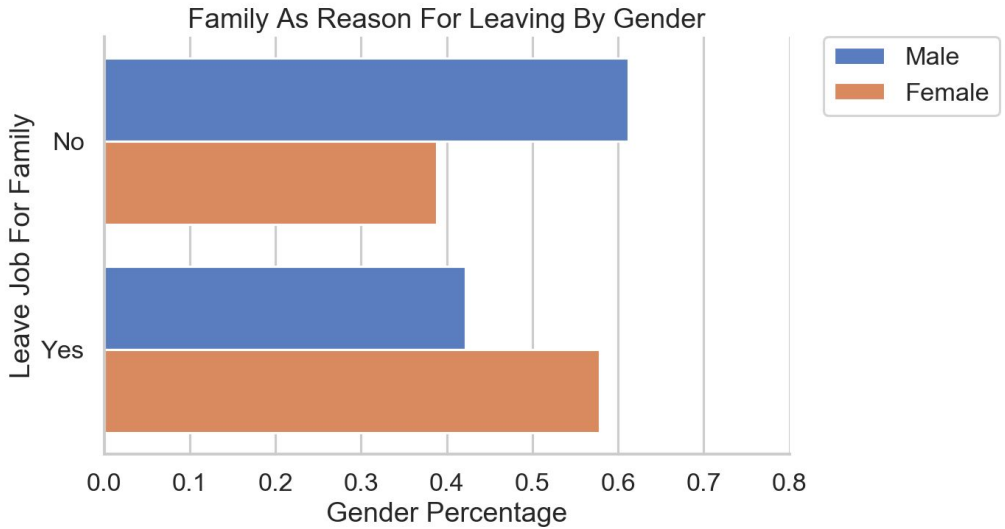
Unemployment Gap

Men and women with no children share the same unemployment rate. This could be due to a number of factors (such as new graduates); however, men and women with children share different rates of unemployment. Men with children have a lower unemployment rate than women with children. This could be due to the fact that most women leave the field to take care of their children whereas most men do not.

Exploratory Data Analysis - cont.

Unemployment reason - Family

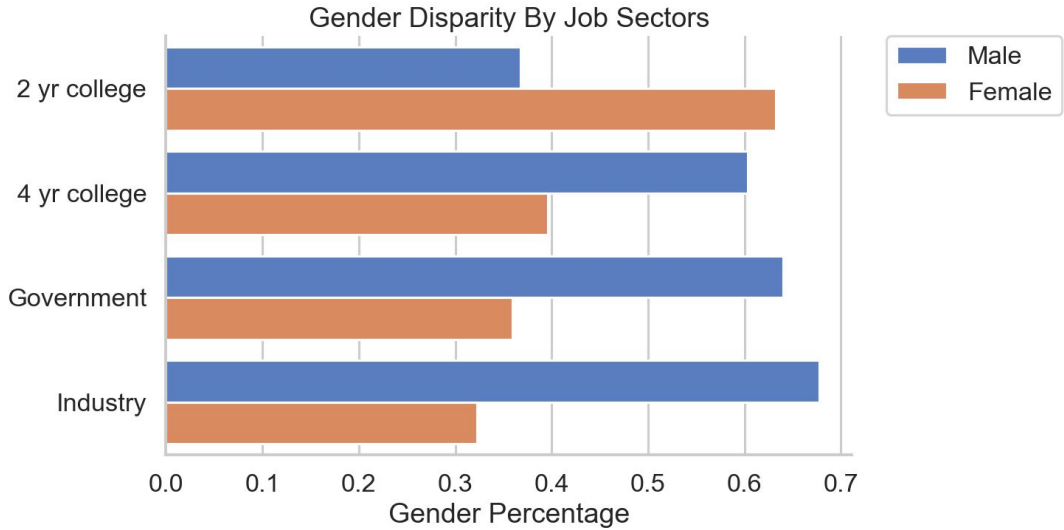
A greater percentage of women leave their jobs due to family reasons (i.e. childcare, maternity, etc.). Women who leave the workforce for family reasons for “x” number of years will not be able to catch up to a man who started at the same salary range.



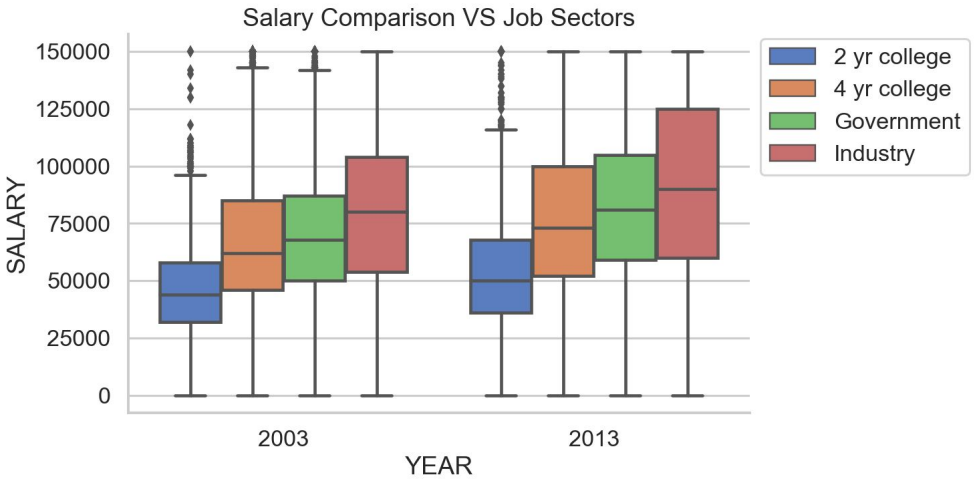
Exploratory Data Analysis - cont.

Motherhood in job sectors

Women with children tend to work at 2 year colleges. This is most likely due to work schedule and flexible responsibilities. Men with children are mostly seen in other sectors, mainly industries.



Exploratory Data Analysis - cont.



Salary in job sectors

When salaries of each sectors are compared, 2 year colleges pay the least amount while industries pay the most. With all things considered, these plots show that women with children are paid less than men with children because they stay in job sectors that pay less.

Statistical Analysis

Feature Examined	Difference	P-value from t-test	Conclusion
Gender Pay Gap	32.71%	0.0	Significant
With Children Pay Gap	38.74%	0.0	Significant

Feature Examined with having children	Correlation	P-value from t-test	Conclusion
Salary	0.11	0.0	Significant
Unemployment	-0.72	2.82e-26	Significant

Predictive Analysis

Classification steps

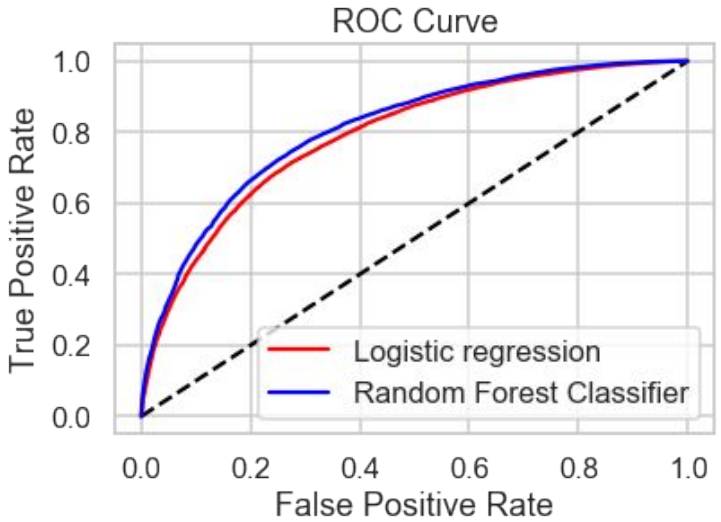
1. Prepare target and feature columns
    a. Use only employed with children portion of survey data
2. Create dummy variables
3. Train test split and check for oversampling
4. Rank and drop features with recursive feature elimination
5. Parameter optimization
6. Fit and predict with classification model
7. Model comparison

Classification Models

Predict Salary Level

- Random forest classifier has the highest accuracy score compared to logistic regression and proves the best model for predicting salary of people with children.
- ROC curve provides a visual of how much better random forest classifier is in predicting salary for this dataset.
- Using coefficient of attributes by logistic regression, the top three features with greatest impact on salary is employment sector, degree, and employment size.

Model	Accuracy score on train	Accuracy score on test
Logistic Regression	0.724	0.737
Random Forest Classifier	0.782	0.750



Classification Model

Job Sector	Train Score	Test Score
2 yr college or other	0.932	0.896
Government	0.884	0.852
4 yr college or medical	0.863	0.836
Industry	0.758	0.726

Predict employment sector

- Similar data process as classification of salary level
- Binarize target column so there are n classification given n employment sectors
- Run random forest classifier and loop through n employment sectors.

Classification Model

Predicting employment sector

- Average accuracy of model: 0.858
- Greatest impacts on choosing an employment sector are company size and degree type. Company sizes are generally really small (11-24) or really big (5000-25000). Greatest degree of impact is having a masters.
- The classification used total number of children data. This can help mothers decide or prepare for job changes.
- Combined with salary prediction, these classification models can become a part of the recommendation system for job searching sites.

Summary

- Gender pay gap is significantly increased for women with children.
- Motherhood does make a difference in the STEM workforce. Unemployment rate shares a statistically significant negative correlation with having children. Women are the most likely to leave the workforce for family reasons.
- Using random forest classifier, the salary prediction reported an accuracy score of 0.750 although there are other missing variables, such as location, that can affect salary range. The employment sector prediction reported an average accuracy score of 0.858 with 2 year colleges or other being the most predictable sector.
- Machine learning models provide a reference for mothers in the STEM field to decide and prepare themselves for career changes. These models can contribute to recommendation systems for job searching websites.

Future Improvements

- Explore length of maternity leave and its impact on salary
- Explore length of employment to salary
- Obtain location specific data and explore motherhood wage gap more closely