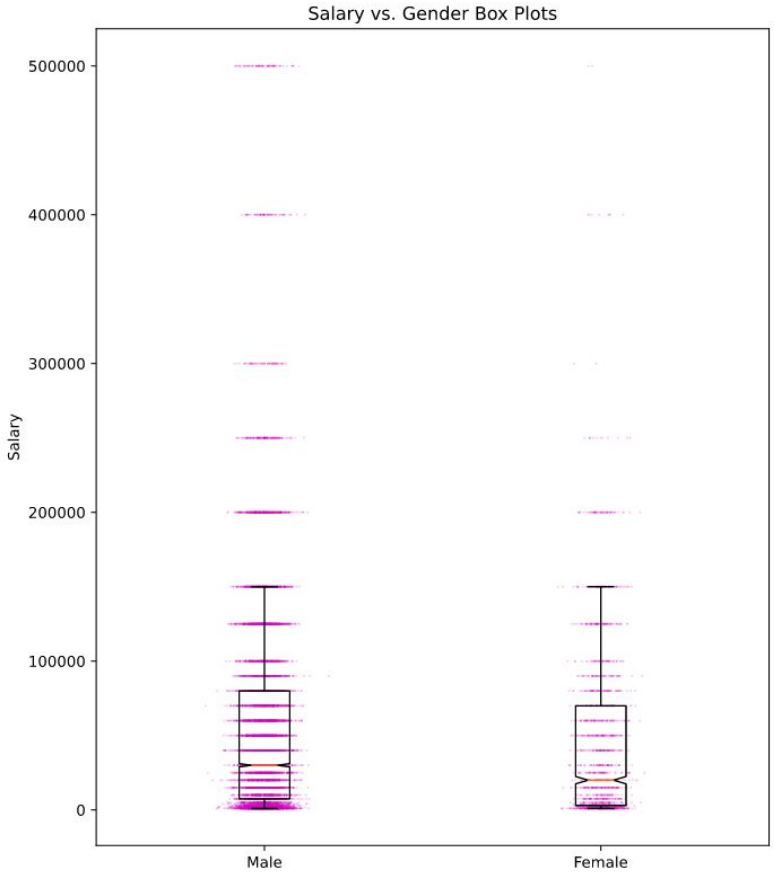
Impact of Gender & Education on Income in DS & ML

"tell a data story about a subset of the data science community represented in this survey, through a combination of both narrative text and data exploration"

Samuel Atkins - October 2020



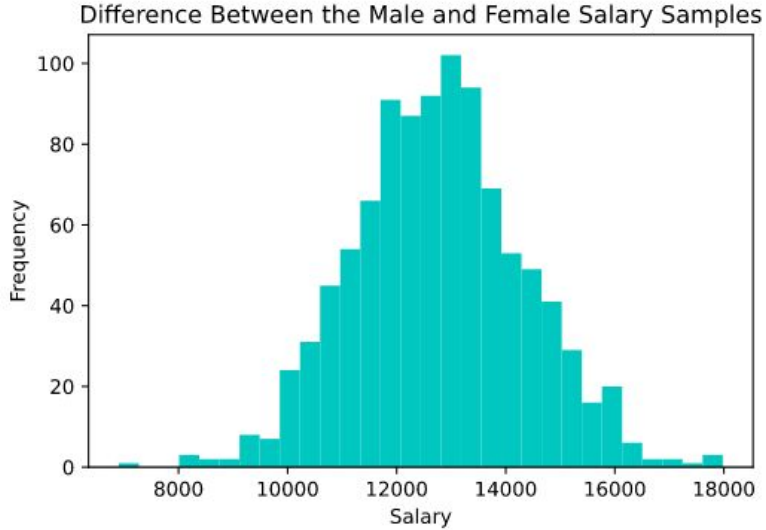
Exploratory Data Analysis

- Data Visualization
- Salary vs. Gender & Salary vs. Education Box Plots
- Descriptive Statistics

	sex	country	age	educ	prof_exp	salary
0	Male	France	22-24	Master's degree	Software Engineer	40000
1	Male	India	40-44	Professional degree	Software Engineer	7500
2	Male	Australia	40-44	Master's degree	Other	300000
3	Male	India	22-24	Bachelor's degree	Other	5000
4	Male	France	50-54	Master's degree	Data Scientist	70000

Comparing Male and Female Salaries in DS & ML

Two-Sample t-Test & Bootstrapping

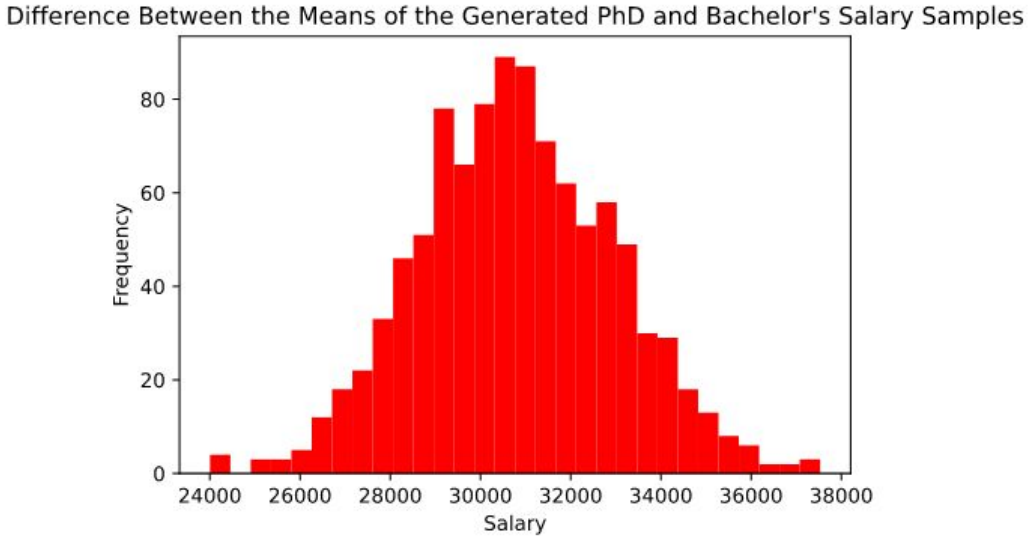



```
In [160]: tc, pc = stats.ttest_ind(males_df["salary"], females_df["salary"])
          print ("t-test: t = %g  p = %g" % (tc, pc))

          t-test: t = 6.90935  p = 5.10894e-12
```

```
In [162]: tc, pc = stats.ttest_ind(male_means, female_means)
          print ("t-test: t = %g  p = %g" % (tc, pc))

          t-test: t = 258.205  p = 0
```

Impact of Education on Income

ANOVA, Bootstrapping, & Estimating the Difference Between the Means




```
In [165]: f_stat, p_val = stats.f_oneway(phd_df.salary, masters_df.salary, ba_df.salary)
          print ("F-stat: F = %g  p = %g" % (f_stat, p_val))

          F-stat: F = 117.839  p = 2.23367e-51

In [170]: f_stat, p_val = stats.f_oneway(phd_means, master_means, ba_means)
          print ("F-stat: F = %g  p = %g" % (f_stat, p_val))

          F-stat: F = 125127  p = 0
```

Conclusion

Difference Between Salaries of Males and Females in DS & ML:

- *Males earn approximately $12,772.14 more on average*

Average Salary Increases for Education Tiers Surpassing a Bachelor's Degree:

- *Bachelor's Degree to Master's degree: $13,740.45 approximate pay increase*
- *Bachelor's Degree to PhD: $30,785.89 approximate pay increase*