

# GENDER DISPARITIES IN DATA SCIENCE: A STATISTICAL AND EXPLORATORY ANALYSIS OF KAGGLE'S ML SURVEY

- SRIKANTH RANGANATHAN

**Objective:** The objective of this report is to analyze the trends in the given dataset and understand women's representation in DS and ML, and the effect of education on the salary/income of a person.

## Question 1: Performing EDA with the given dataset.

**Answer:** To aid toward my objective, I have plotted three bar plots that visualize the overall trend in income and job roles in data science and machine learning based on Gender and level of education. Bar plots have been used because I'm personally a big fan of them and believe they can give meaningful insights with a very simple style and design. I used "Plotly" because of its design.

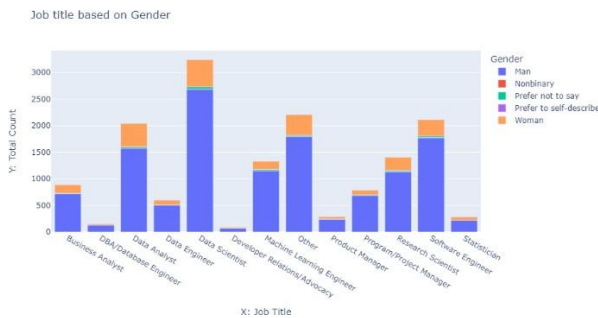


Fig. 1

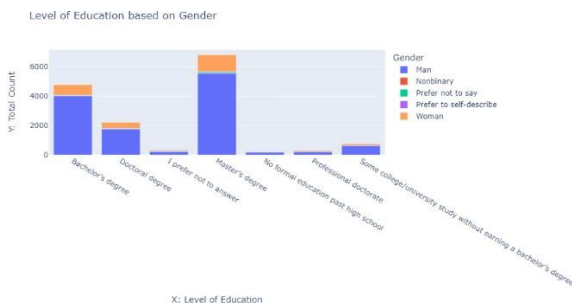


Fig. 2

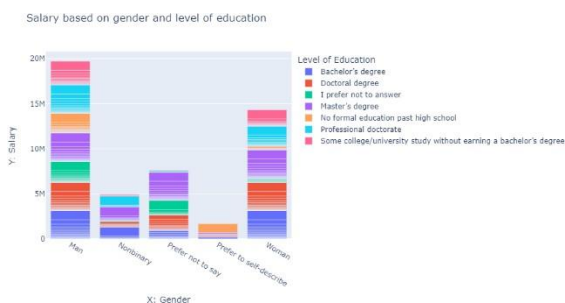


Fig. 3

It is evident from the figures that the population of men dominates other genders in university-level education and income levels. It is important to note that a significantly greater number of responses have been recorded by men compared to others.

## Question 2: Estimating the difference between the average salary (Q25) of men vs women (Q2).

**Answer: Part a)** Let us first look at the descriptive statistics of the salaries of men and women. The describe() function was used as it produces all the relevant descriptions of the data such as mean, standard deviation, etc.

Report: Women		Report: Men	
count	2482.000000	count	12642.000000
mean	34816.881547	mean	51193.600696
std	72017.347888	std	99979.274378
min	1000.000000	min	1000.000000
25%	1000.000000	25%	2000.000000
50%	7500.000000	50%	20000.000000
75%	50000.000000	75%	60000.000000
max	1000000.000000	max	1000000.000000

There is an evident difference in the mean salaries of men and women of 16,376.719149 USD.

**Part b)** It is clear from the data that the mean of men's and women's salaries differ substantially. The difference between the mean salary of men and women is 16,376.719149 USD. No statistical testing is required to prove the significant difference between the mean salaries of men and women. Hence, **no t-test is required**.

**Part c)** For this part, I have bootstrapped the mean of both men's and women's salaries as a Pandas series into two different variables with n=1000. I have plotted 3 histograms representing the mean salaries of men and women, and the difference in mean salaries of men and women.

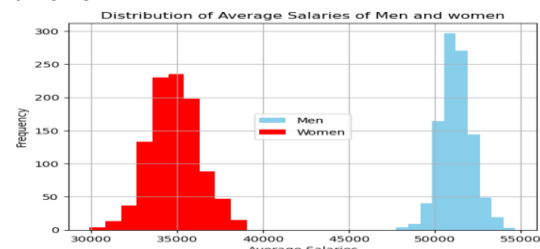


Fig. 4

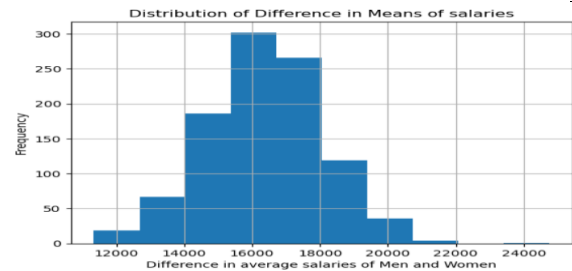


Fig. 5

"Matplotlib" is used to plot the histograms as they are very simple to produce.

**Part d)** There is **no need to perform a t-test** for the bootstrapped data as the difference in means of average salaries of men and women are not equal, which is evident from the plot above. **The differences are statistically significant and visually apparent.** The mean of bootstrapped differences of means of salaries of men and women is 16,314.45664812596 USD. The same for the dataset values was 16,376.719149 USD.

**Part e)** Overall, the income levels of men are far higher than that of women. There is a lesser deviation in salaries with women, but a substantially higher deviation in that of men. Although education has a big impact on salaries, the average salary that men receive regardless of education level is far higher than what women receive.

**Question 3: Comparing the average salaries of people from 3 different education groups: bachelor's degree, Master's degree, and Doctoral degree.**

**Answer: Part a)** Three education groups are selected and the income levels in the respective groups are analyzed. The descriptive statistics of the three groups are shown below:

Report: Bachelor's degree			Report: Master's degree			Report: Doctoral degree		
Q25			Q25			Q25		
count	4777.000000		count	6799.000000		count	2217.000000	
mean	35578.291815		mean	52706.868657		mean	70641.181777	
std	89382.060777		std	90928.786678		std	117160.947589	
min	1000.000000		min	1000.000000		min	1000.000000	
25%	1000.000000		25%	3000.000000		25%	4000.000000	
50%	7500.000000		50%	25000.000000		50%	40000.000000	
75%	40000.000000		75%	70000.000000		75%	90000.000000	
max	100000.000000		max	100000.000000		max	100000.000000	

**Part b)** There is **no need to perform an ANOVA test** on the 3 groups as the difference in means of salaries of the 3 groups is significant and clearly visible. **The means of salaries of the 3 groups are not equal and the difference is statistically significant.** Hence, no ANOVA is needed. Bachelor's Degree: 35578.291815 USD, Master's Degree: 52706.868657 USD, Doctoral Degree: 70641.181777 USD. The plots from the bootstrapped values further indicate the same.

**Part c)** In this part, we will Bootstrap average salaries for Bachelor's, Master's, and Doctoral Degree holders and plot the distribution of the same while also plotting differences in means of salaries of all three groups.

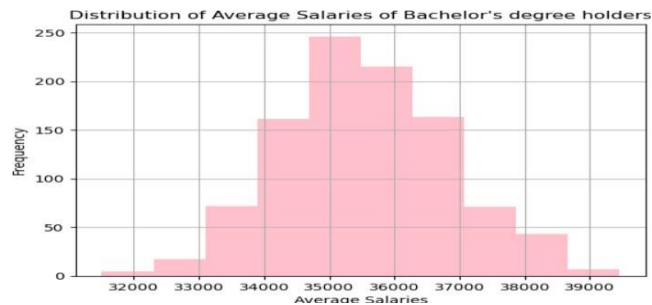


Fig. 6

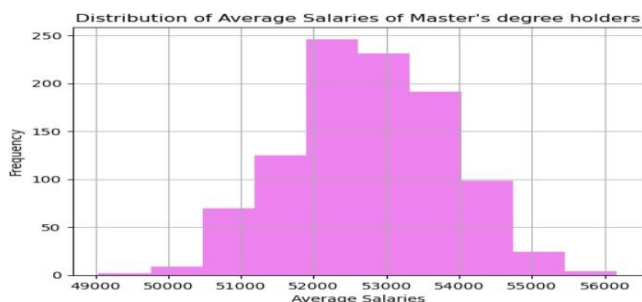


Fig. 7.

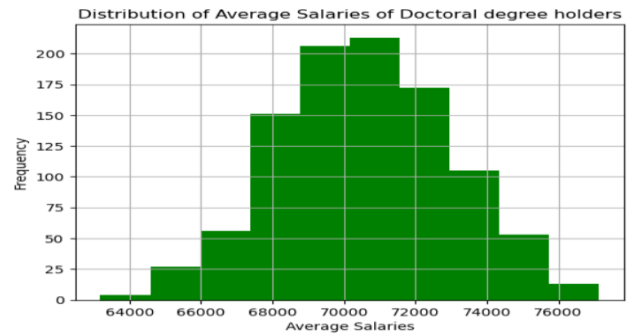


Fig. 8

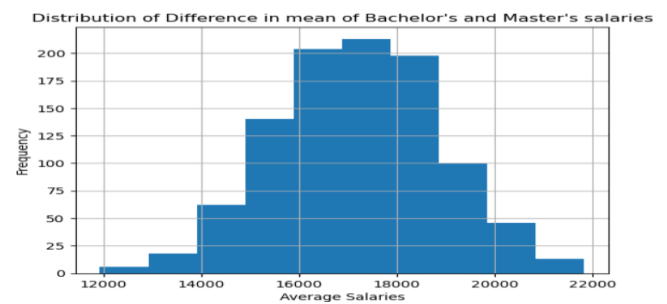


Fig. 9

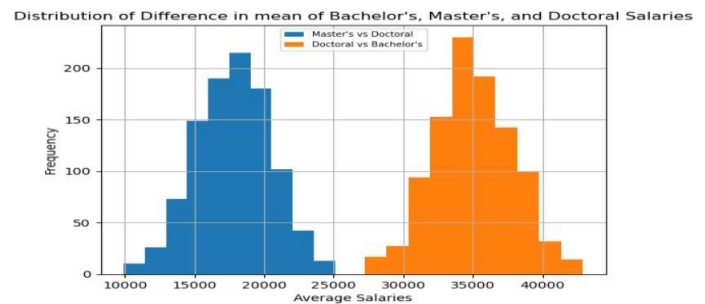


Fig. 10

**Part d)** Yet again, there is no need to perform an ANOVA test here as the difference in bootstrapped means of salaries based on level of education is significantly different and **the difference is statistically significant.** The mean of the average salaries for 3 different groups is not equal. Hence, **no ANOVA is needed.**

**Part e)** The mean salaries of the 3 education groups are significantly different with the Doctoral degree holders earning the highest, then the master's and bachelor's degree holders. The difference in means of the mean salaries of bachelor's and master's holders overlaps but is not the case whatsoever with Doctoral degree holders.

**Conclusion:** Women are found the highest in Job Titles such as "Data Scientist" and "Data Analyst", but very few in other titles. There is a huge salary discrepancy between men and women in the same Job titles with the same level of education. Our plots and analysis point out that women get paid a lot less to do the same job than men even when women have similar qualifications or more. The level of education does have a notable impact on income levels across all genders, but the sizeable differences in salaries across genders persist regardless of the level of education.