Question 1

Exploratory analysis of the data focused on six main categories – Age, Gender, Country, Education, Professional Experience, and Salary, with the mean and medians of each category summarized in Appendix A. Notably, the mean salary for the entire dataset was approximately $49,116, while the median and standard deviation of the salary was $15,000 and $98,090, respectively. This variation in the various descriptive statistics indicate a wide spread in the salary responses (Figure 1, Appendix A). Visualizations of trends between gender and education and age on salary can be seen in Figures 2 & 3 (Appendix A).

Question 2

a) The gender subset of data used in this section is a data frame containing only responses for men and women. The mean, median and standard deviation of the man & woman specific data can be seen in Appendix B, but the results indicate a similar large spread in the overall salary, similar to that seen in the previous section.

b) We assume, as the survey was collected individually, that no one survey response is dependent on another and that the sample was random. however, tests for normality and equal variance indicated that the data were not normally distributed (Figures 4 & 5 in Appendix C) and had unequal variance (Levene's test p-value = 7.3e-12) , meaning that a two sample t-test on the data could not be performed directly.

c) To account for the non-normality and unequal variance in the data, the data were bootstrapped and the normality and equal variance tests were performed on the bootstrapped data. The bootstrapped data had a normal distribution (see Appendix D, Figures 6 & 7) but still had unequal variance (p-value for levene's test <<0.05), therefore a Welch's t-test was conducted, as it doesn't require the equal variance assumption.

d) The Welch's t-test had a p-value of zero (<0.05) and therefore indicates a significant difference between the average salaries of men and women in the data.

e) Based on the information from the distributions in Figure 2 (Appendix A) and Figure 8, the data indicates that women make significantly less than men. This is in-keeping with the information other studies that have examined the relationship between sex or gender and salary [1], [2].

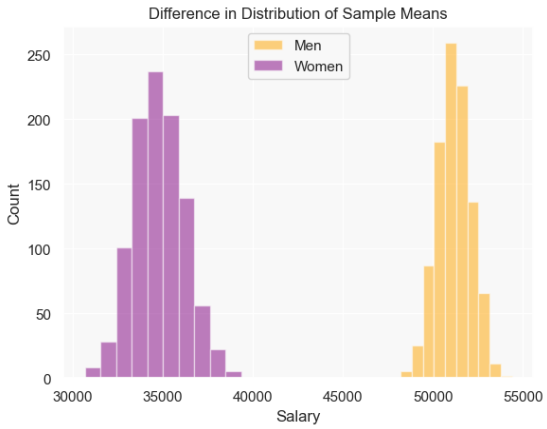


Figure 8: *The distribution of the two sample means calculated in the bootstrapped data. The mean salary in the bootstrapped dataset is approximately $35K for women and $51.2K for men, around a $16K difference in average salaries.*

Question 3

a) The education subset of the data used in this section is a dataframe containing only responses for individuals with a Bachelor's, Master's, or Doctoral degrees. The mean, median and standard deviation of the education specific data can be seen in Appendix E, but similar to sections 1a and 2a, the results indicate a large spread in the overall salary.

b) ANOVA tests assume independence, normality, and equal variance in the data. Similar to the previous section, the normality and equal variance assumptions are violated (see Appendix F) and it can not be performed directly on the data.

c) After bootstrapping , the data had a normal distribution (see Appendix G, Figures 10 & 11) but still had unequal variance (p-value for levene's test = 1.9e-121) , and to account for this a Welch's ANOVA was performed, as it does not require equal variance in the data.

d) The Welch's ANOVA had a p-value of approximately 5.6e-42 (<<0.05) indicating a significant difference in the salaries of individuals with different levels of education.

e) Since Welch's ANOVA does not indicate which variables are significantly different from one another, a post-hoc Games-Howell test was performed. This post-hoc analysis indicates that there is a significant difference between the average salaries for all degree levels (see Appendix H). In general, it appears that individuals with a higher education level make more on average, which is in-keeping with the information seen in Figures 2 (Appendix A) and Figure 12, as well as other studies that have examined the relationship between education and salary [3], [4].

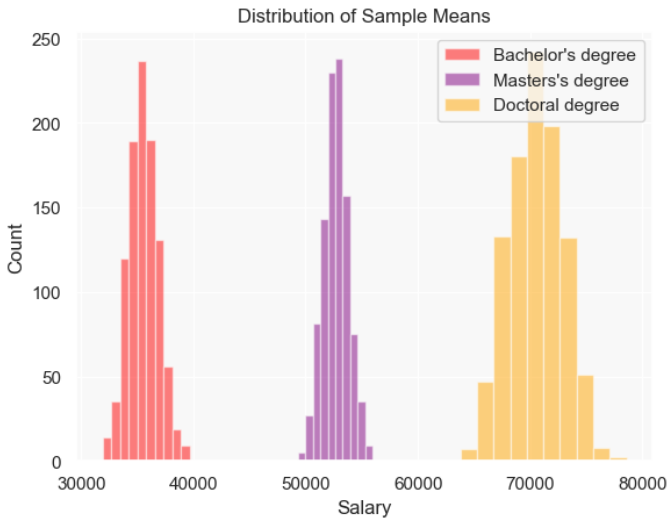


Figure 12: *The distribution of the sample means calculated in the bootstrapped data for different education levels. The mean salary in the bootstrapped dataset is approximately $35.5K for Bachelors degrees, $52.7K for Masters degrees, and $71K for Doctoral degrees.*

References

[1] N. Fleming, "How the gender pay gap permeates science and engineering," *New Scientist*, vol. 237, no. 3167, pp. 22–23, Mar. 2018, doi: 10.1016/S0262-4079(18)30389-0.

[2] C. Woolston, "Scientists' salary data highlight US$18,000 gender pay gap," *Nature*, vol. 565, no. 7737, pp. 527–528, Jan. 2019.

[3] J. Xiao, "Determinants of salary growth in Shenzhen, China: an analysis of formal education, on-the-job training, and adult education with a three-level model," *Economics of Education Review*, vol. 21, no. 6, pp. 557–577, Dec. 2002, doi: 10.1016/S0272-7757(01)00049-8.

[4] "Earn More With a Master's Degree or PhD Degree," *Walden University*. https://www.waldenu.edu/programs/resource/earn-more-with-a-masters-degree-or-phd-degree (accessed Feb. 12, 2023).

Appendix

Appendix A *- Question 1, Figures and Descriptive Statistics*

Figure 1: *Two plots, showing the distribution of the salary responses for the entire dataset. The top plot is the distribution of all salary responses (max of $1M), while the bottom plot shows a more granular distribution of the salaries up to $250K, which is the majority of responses. The red dashed line indicates the mean salary responses ($49,116), and the orange dotted line is the median ($15,000).*

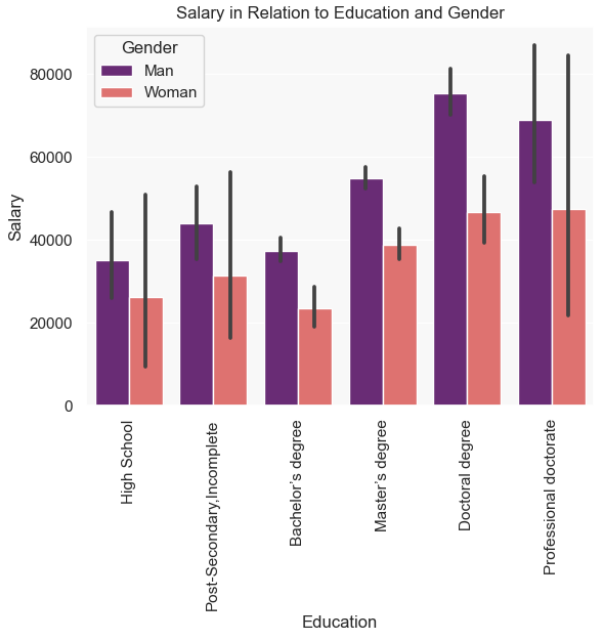


Figure 2: *This plot indicates the average salary recorded for each education level, broken down by gender (men and women). The figure illustrates that, on average, male individuals with multiple degrees make more money than their less educated or female counterparts.*

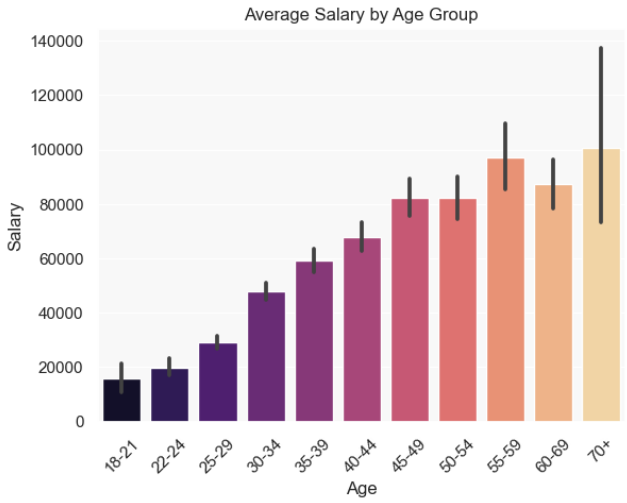


Figure 3: *This plot indicates the average salary recorded for each age group across the entire dataset. A general positive trend of increasing salary with increasing age can be seen.*

Appendix B *- Question 2a, Descriptive Statistics*

Table 1: Descriptive statistics relating to gender and salary.

	Salary							
	count	mean	std	min	25%	50%	75%	max
Gender								
Man	12642.0	51193.600696	99979.274378	1000.0	2000.0	20000.0	60000.0	1000000.0
Woman	2482.0	34816.881547	72017.347888	1000.0	1000.0	7500.0	50000.0	1000000.0

Table 2: Median salary by age group.

Age	*18-21*	*22-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*	*50-54*	*55-59*	*60-69*	*70+*
Salary (USD)	1000	3000	7500	25,000	25,000	40,000	50,000	50,000	60,000	60,000	50,000

Table 3: Median salary by gender.

Gender	***Salary (USD)***
Man	20,000
Woman	7,500

Table 3: Median salary by education.

Education	Salary
Bachelor's degree	7500.0
Doctoral degree	40000.0
High School	7500.0
I prefer not to answer	4000.0
Master's degree	25000.0
Post-Secondary,Incomplete	10000.0
Professional doctorate	20000.0

***Appendix C** - Question 2b, T-Test Assumption Checks*

a)

b)

Figure 4 (a & b): *Two plots showing the salary distribution for men and women. Visually the distribution indicates a right skew to the data (non-normal distribution).*

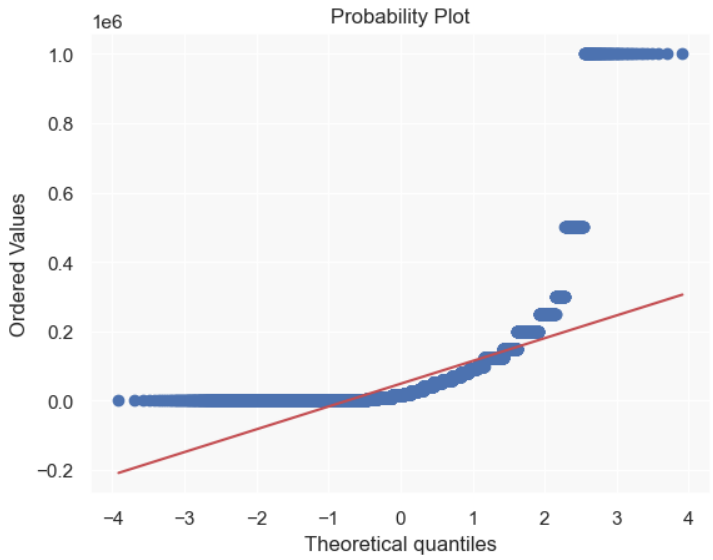
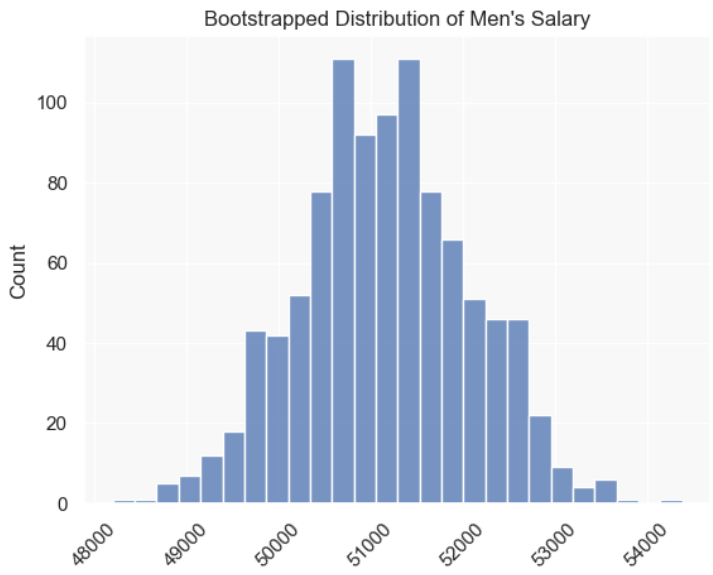


Figure 5: *Q-Q plot demonstrating lack of normality in the data distribution.*

***Appendix D** - Question 2c, T-Test Assumption Check on Bootstrapped Data*

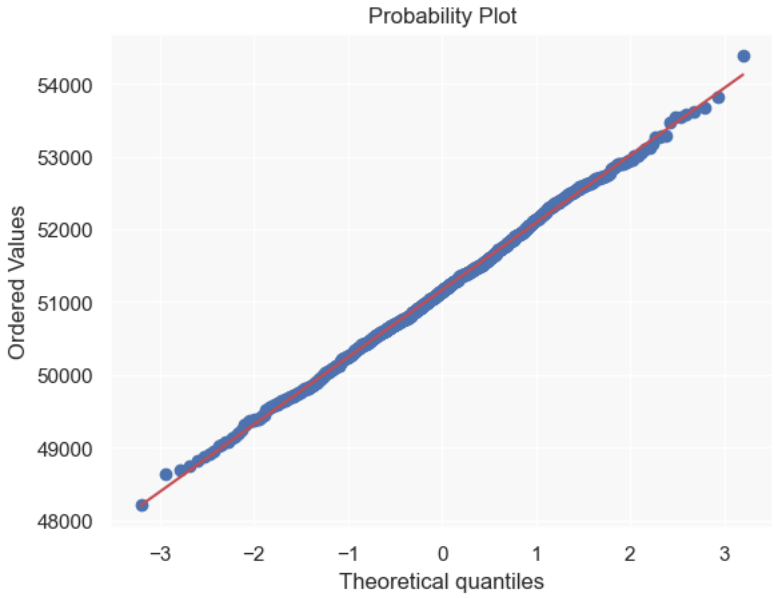
a)



b)

Figure 6 (a & b): *Two plots showing the salary distribution for men and women on the bootstrapped dataset. Visually, the data follows a normal distribution.*

a)



b)

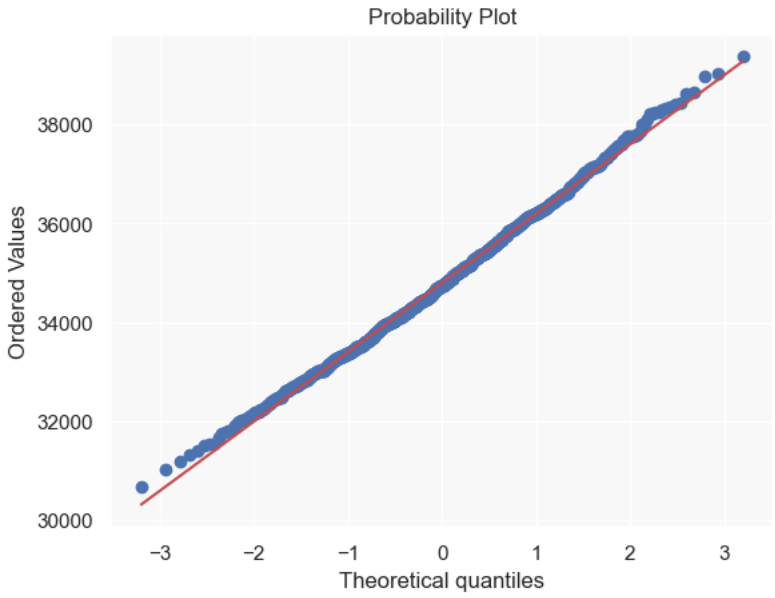


Figure 7 (a&b) : *Q-Q plot confirming the normality of the distributed data for both (a) men and (b) women for the bootstrapped dataset.*

***Appendix E** - Question 3a, Descriptive Statistics*

Table 4: Descriptive statistics relating to education level and salary.

	Salary							
	count	mean	std	min	25%	50%	75%	max
Education								
Bachelor's degree	4777.0	35578.291815	89382.060777	1000.0	1000.0	7500.0	40000.0	1000000.0
Doctoral degree	2217.0	70641.181777	117160.947589	1000.0	4000.0	40000.0	90000.0	1000000.0
Master's degree	6799.0	52706.868657	90928.786678	1000.0	3000.0	25000.0	70000.0	1000000.0

Table 5: Median salary by age group.

Age	*18-21*	*22-24*	*25-29*	*30-34*	*35-39*	*40-44*	*45-49*	*50-54*	*55-59*	*60-69*	*70+*
Salary (USD)	1000	4000	10,000	25,000	30,000	40,000	60,000	60,000	70,000	60,000	55,000

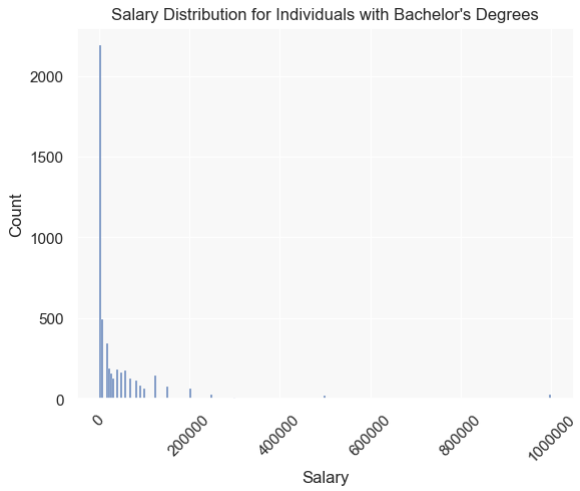
Table 7: Median salary by education.

Education	***Salary (USD)***
Bachelor's Degree	7500
Master's Degree	25,000
Doctoral Degree	40,000

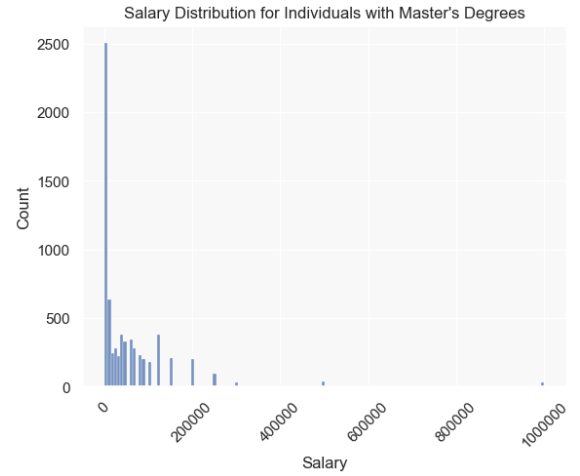
***Appendix F** - Question 3b, ANOVA Assumption Checks*

Levene Test P-Value = 4.5e-35 (<<0.05); data has unequal variance

a)



b)



c)

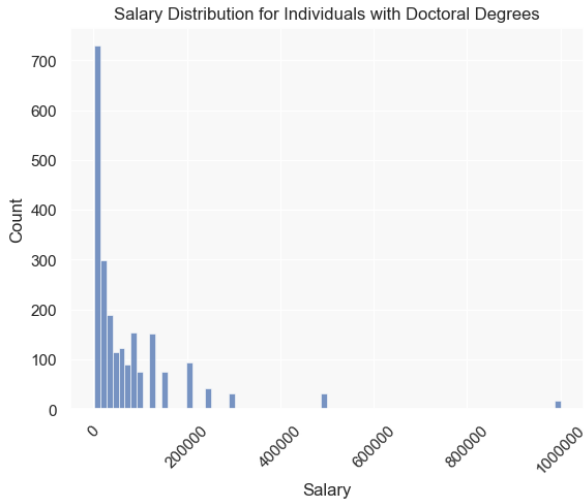


Figure 8 (a-c): *Distribution plots showing the salaries of individuals with (a) Bachelor's, (b) Master's, and (c) Doctoral degrees. Visually the distribution indicates a right skew to the data (non-normal distribution).*

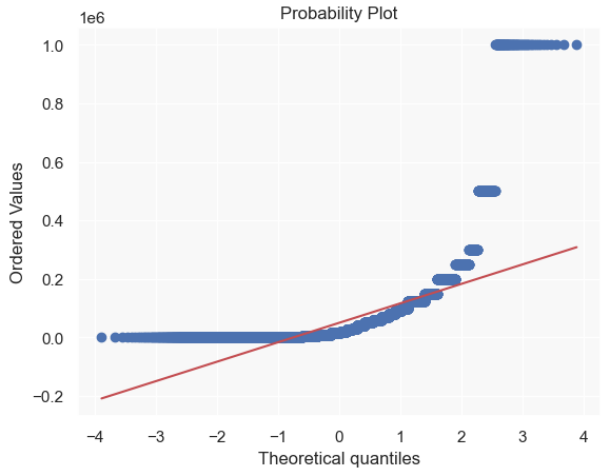
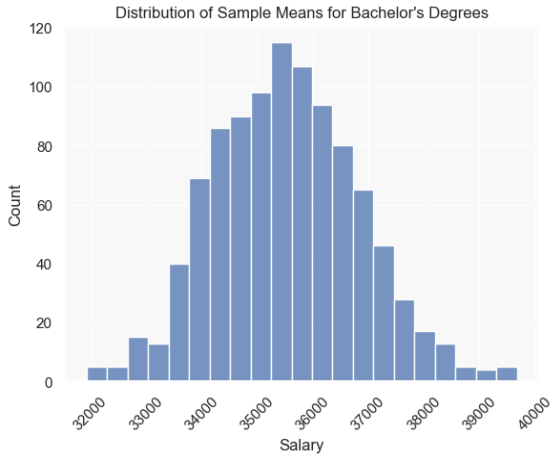


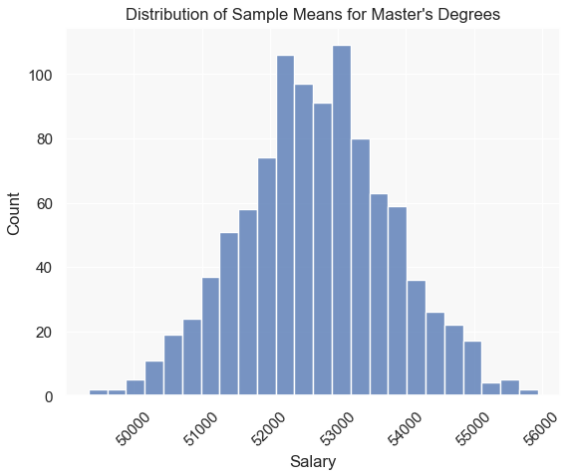
Figure 9: *Q-Q plot demonstrating lack of normality in the data distribution for salary and education level.*

***Appendix G** - Question 3d, ANOVA Assumption Check on Bootstrapped Data*

a)



b)



c)

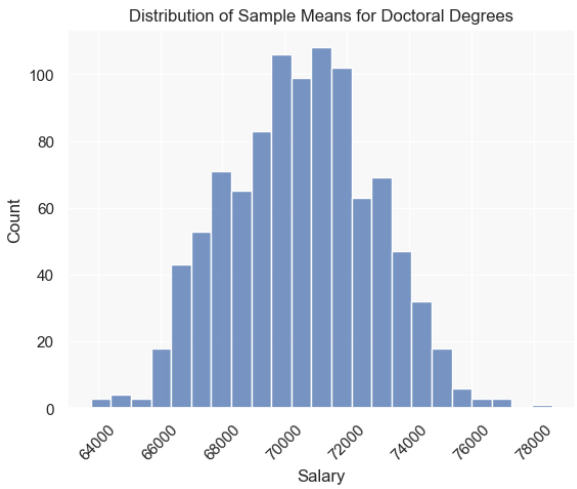
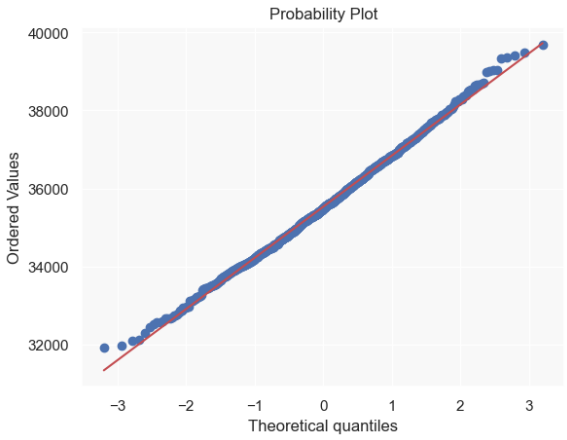
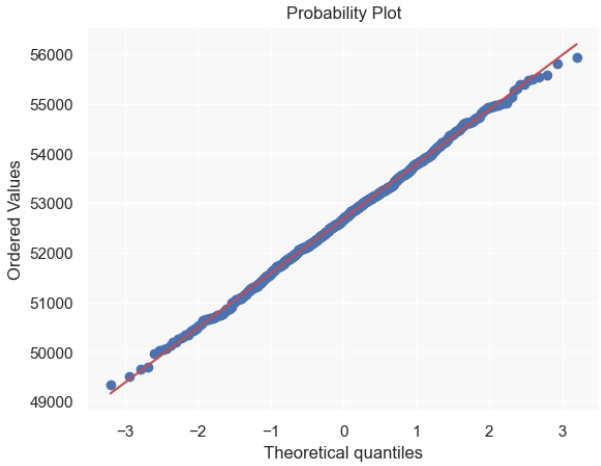


Figure 10 (a - c): *Plots showing the salary distribution for (a) Bachelor's, (b) Master's, and (c) Doctoral degrees, visually indicating a normal distribution.*

a)



b)



c)

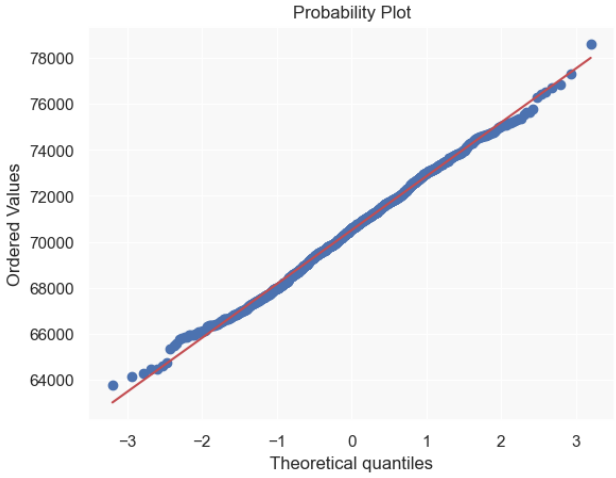


Figure 11 (a-c) : *Q-Q plot demonstrating normality in the data distribution for salary and (a) Bachelor's, (b) Master's, and (c) Doctoral degrees.*

***Appendix H**- Question 3e, Post-Hoc Games-Howell Test Results*

Table 8: Results of post-hoc Games-Howell analysis, demonstrating significant differences between all groups.

	A	B	mean(A)	mean(B)	diff	se	T	df	pval	hedges
0	Bachelor's degree	Doctoral degree	35578.291815	70641.181777	-35062.889962	2804.278894	-12.503353	3457.766716	8.795187e-13	-0.354014
1	Bachelor's degree	Master's degree	35578.291815	52706.868657	-17128.576842	1699.555297	-10.078270	10388.015311	1.602052e-12	-0.189686
2	Doctoral degree	Master's degree	70641.181777	52706.868657	17934.313120	2721.695666	6.589390	3132.569650	1.552711e-10	0.182931