Stephanie Benavidez

Data Mining DSC550-T301

Week 12 Assignment

Milestone 4

Gender Pay Gap

The topic I thought I would use to discuss throughout this project is the “Gender Pay Gap.” A research study (The Gender Pay Gap Breakdown, n.d.) states that in 2019, “women earn $0.79 for every $1.00 man earn. Thus, when broken down by racial and ethnic groups the pay was even more significant. According to Dr. Chamberlain, in his opening remarks to congress, he stated, “Pay equality is about fairness. But it also helps build a more dynamic and prosperous U.S. economy. It clears the way for Americans to make their best contribution to our prosperity – a win-win for workers and employers that can also be a pro-growth agenda for America (Fluker, 2021).”

           In this project, I hope I can differentiate in pay between a male and female worker with the same education. Regardless of race or ethnicity, a workplace should pay their co-workers equally. Our world is evolving and changing. It is no longer set for just one gender working in the workplace. As Dr. Chamberlain stated in his closing statement, “The best thing you can do is to assure that when men and women show up in jobs, they believe they will be treated fairly, and they will not get the short end of the stick once they arrive (Fluker, 2021).” After everything that has happened in the last two years, I think Americans are wanting more equality and not to be treated differently than someone with a similar upbringing. As a future employer, it is good to have the knowledge and understanding of what requirements will be needed for the position being presented. As well as preparing questions to be presented in an interview that you may have with the company you are interested in.

After all the racial and inequality events that have happened within the past years, companies should have their Human Resources Department do more research on how to pay all employees equally regardless of race or ethnicity. Most Human Resource Departments have analysis programs that help decipher what an employee's compensation is by their current work experience, school degrees, and certifications they have acquired. They then take that information and do a market analysis for that position within the area to validate compensation. However, there are still more companies that don’t necessarily do those things. This is where the analysis from the project would be good to bring to the director of Human Resources or inform the board to show why it is beneficial for the company.

With this project, the questions that I would like to be able to prove are as follows:

* Which job title had the highest salary?
* Which job had the highest bonus? Was it the same title as the highest salary?
* Out of the highest salary and bonus, which gender reflected that salary?
* Did the opposite gender have the same schooling as the gender that had the highest salary?
* Lastly, are there any job titles that had the lowest salary? Lowest bonus? Which gender reflected that salary?

After pulling in all the information and running the analysis in each milestone. I discovered errors and invalid results due to not knowing how to code the results sought in this project. The first step was to find what information you wanted to use as the source for your project. The second step was to clean up the data from the dataset used for the project. For me, those two steps, I have had more practice with these and was able to get that information cleaned. The third step was to prepare visualizations with graph analysis. In this step, I discovered errors with my coding to produce the incorrect graph analysis that I was trying to solve. To solve the discovered errors, I used the application Power BI, to show the correct visualizations.

In the first graph analysis, the question being answered is: Which job title had the highest salary? As shown below, the visualization shows in the scatter plot that the highest salary in the dataset being used is a Manager with a salary of around $180,000.

Graphical user interface, text

Description automatically generated

I compared the first graph analysis with the visualization created in Power Bi, as shown below, and the visualization showed the same results as coded.

Chart

Description automatically generated

In the second graph analysis, the question being answered is: Which job had the highest bonus? Was it the same title as the highest salary? In the original visualization, the scatter plot shows that the job title that has the highest bonus is Software Engineer with a bonus of over $11,000, and was not the same job title as the first graph analysis.

Application

Description automatically generated with low confidence

I compared the second graph analysis with the visualization created in Power Bi, as shown below, and the visualization showed the same results as coded.

Chart

Description automatically generated

In the third graph analysis, the question being answered is: Out of the highest salary and bonus, which gender reflected that salary? In the original visualization, the scatter plot is only showing the job title: Manager in the graph analysis. The graph shows that both males and females are considered for this position.

Graphical user interface, text, application

Description automatically generated with medium confidence

Though this graph analysis didn’t answer the in-depth answer to that question. Since I am still new to coding, I wanted to see if I could get that in-depth answer that I was looking for in the Power BI visualization.

Graphical user interface, chart, bar chart

Description automatically generated

As shown above, in the visualization a male manager makes more than a female manager. The difference between those base salary amounts is approximately $19,000.

In the fourth graph analysis, the question being answered is: Did the opposite gender have the same schooling as the gender that had the highest salary? With the prior analysis, the highest salary is for a male manager. Though a male and females have different salaries, did they have the same education? In my original graph analysis, the code didn’t answer the in-depth question, as shown below.

Chart, bar chart

Description automatically generated

The graph analysis shows the education but doesn’t show the answers if the male or female for the manager position had the same education. Since the coding for the graph analysis didn’t work, I wanted to see if I could get that in-depth answer that I was looking for in the Power BI visualization.

Background pattern

Description automatically generated with low confidence

As shown above, the visualization is filtered to just show the job title: manager. In the analysis, you can see that the difference in education is approximately a 1% difference between genders. Though the base salary may be increased for a male manager, the female manager has the same education as a male.

In the first milestone of this project, it stated to do four visualizations, which were completed as shown above. However, there was still one question that wasn’t answered throughout the milestones but the visualization provided the answer to the question. The question is as follows: Lastly, are there any job titles that had the lowest salary? Lowest bonus? Which gender reflected that salary? From the first graph analysis, it allowed us to see that the lowest salary is a marketing associate with a salary between $120,000 and $140,000 and was a male. The job title with the lowest bonus was a data scientist making a bonus of about $10,000 and a female.

In conclusion, the analysis did prove that there is a difference between the pay of a male and a female. Though I wasn’t able to prove the model accuracy for the data, the visualization was able to tell the story that was being questioned. For the model accuracy for the data to be used, it will need to be reworked through the coding. In the original graph analysis, the last two analyses will need to be re-coded so that it truly shows the analysis that is being done, like the visualizations in Power BI. The challenge that I have faced in this project is coding. If the way the code is written is not correct, then it allows your data not to represent the information that you are needing. There is still more learning and practice that will need to be done for all of the information to be accurate.

# References

Fluker, D. (2021, April 29). *Glassdoor Shines Light on Gender & Racial Pay Gap Before Congress*. Retrieved from www.glassdoor.com: https://www.glassdoor.com/blog/glassdoor-shines-light-on-gender-racial-pay-inequality/

*The Gender Pay Gap Breakdown*. (n.d.). Retrieved from www.glassdoor.com: https://www.glassdoor.com/blog/lp/equalpay/