Week 12 Final Project

Pankaj K Yadav

Bellevue University

The question for this assignment was to analyze and find answers for

1. Whether School grades define your salary
2. Is there a pay gap between male and female salaries

Github link : <https://github.com/pankajsyadav/EDA2024>

Outcome of your EDA?

After cleaning all the data and going all the analysis, the outcome was that the salaries are impacted by grades but not significantly. The model was able to predict the salaries but only up to some extent and I agree with the outcome.

Your placement in a job is influenced by other factors too and your percentage is just one of the qualifications companies looks for. Your aptitude & behavioral test scores play significant roles in your interview which were missing from the data.

Analysis did confirm that there are some gaps between salaries of males and females where males earn a higher salary compared to their female counterparts.

What do you feel was missed during the analysis?

The data was limited and was missing some crucial variables which could have helped gain more insight and helped build a better predictive model. Whether the student attended a job interview or not, the selection criterion, Scores like behavioral and aptitude scores, the prior salaries in case the student was working. If the student is planning to go and do business or join back their old organization, as these students could skew the results.

Also, due to time constraints I skipped few tests like normalcy tests, chi-squared, checking biases and sampling analysis which could have been included in my analysis to get a better and robust model.

Were there any variables you felt could have helped in the analysis?

As mentioned above variables that could provide more information on students behavioral and aptitude scores, if student did not attend the job interview, plan to work or go with startup/ businesses, selection criteria for interviews, e.g if all students were allowed or only students with certain work experience or percentages were allowed to sit in the interview. These columns could have provided more insights and helped build a better predictive model.

Were there any assumptions made you felt were incorrect?

Yes, first the assumption about students which higher grades gets higher salaries was itself wrong. Higher grades do help in the process of placement but there were no significant results showing that the higher grades increase your probability to end up getting higher salaries.

What challenges did you face, what did you not fully understand?

The PMF, if you check in my code, are very confusing. The placed ( student offered job) vs students who were not placed shows a box type step plot. Also , as I learnt almost everything using the proprietary codes provided in thinkstat2 book, it was very difficult for me to think out of it and write something by my own. Even for generation of plots I used thinkplot whereas I was pretty comfortable with pyplot.