I have faced a good many challenges during this project, some of which being finding a suitable graph to visualise the relationships (or lack thereof) that I wanted to show. Data cleaning was also much more painful than I thought it would be, since I did not realise that the formats of each column would change over the years. I also had troubles trying to find a suitable range for values (for finding outliers or nonsensical values).

I also feel that the way I explained/phrased everything, from results and EDA all the way to the motivations and background or title could have been phrased better, though I am not sure how exactly I could do so at the current moment.

I was not able to conclude much, except that the annual income of those working in computer science related field have decreased over the years, however this conclusion may be off since it was only 3 years. I was also able to conclude that those who are working in computer science related fields were also less likely to have wanted to work in said field.

In general, after comparing with projects of other classmates, I have learnt that the choice of dataset is important. For example, those who used the stack overflow survey are more likely to see trends between different columns of their data as compared to when I used this survey dataset. This could likely be due to respondents of this survey not taking it seriously and answering values that is far off from reality, or the demographic being skewed to self-taught coders rather than those who have learnt computer science in general.

Therefore, if given another chance, I would most likely have chosen to add the stack overflow surveys to my list of data. However, that may also be a bad idea, since based on what I have heard, the formatting of answers of each column change severely over the years, and also had more options so cleaning was much more painful.

Overall, I would say that this project has failed, due to the lack of conclusions or any trends to show. However, I still have learnt many things such as how to remove unnecessary cells in a correlation matrix heatmap.