Background: In Assignment 3, you would be given one data set and perform data cleaning and data exploration, and provide plan for feature engineering. Submit one ipynb files containing your analysis and answers to questions. The total marks of this assignment is 20 points. However, it would be counted as 10 in your final marks. For instance, if you achieve 10/20, you will get 5/10.

dataset1.csv contains a list of data-related job posts crawled from a job hosting site in 2020. Columns are self-explained. Given the dataset, write code and answer the following three questions.

(1) (4 marks) Propose two questions (non-predictive) that you believe are interesting to explore and can be answered using the provided dataset (at least one question should be answered using hypothesis test). Briefly describe why you think those questions are interesting to whom. You can answer this question in a markdown cell of your ipynb file.

(2) (6 marks) Analyze the quality of data (all columns) and report statistics of missing data and outliers. Briefly describe your step and findings.

(3) (6 marks) For the two questions you proposed in the first subquestion, what are the null hypothesis and alternative hypothesis? Perform statistical test to answer your question and report your findings.

(4) (4 marks) Briefly describe your feature engineering plan (at least two columns should be involved). No coding needed.