The purpose of the assignment is to evaluate if the candidate has a basic understanding about the machine learning techniques or not. The expected work hour for the task is 2-3 hrs(excluding documentation and ppt slides). You can find two different types of tasks for the assignment. First task is a programming task and the second is a use case task.

Format of submission:

1. Prepare a slide with maximum 5 slides(excluding intro and ending slides such as thank you). Please make sure that your slides include your thought process, assumptions about the data and the model, model analysis and evaluation and other aspects of data science.
2. Candidates should provide both code and presentation slides.

Questions:

1. Reading from an IoT device from three different locations is provided in the *training.csv*. Your task is to use the regression model to predict target variables. The objective of the task is to have a minimum **RMSE** in the *test.csv.*
2. You are assigned to a project to classify candidates to hire in a job or not. The high-level statistics of the data shows candidates hired for the job have the following statistics: 40% of candidates are Ph.D. 's degree holders, 30% are master's degree holders and 30% are undergraduate. Besides, 20% has work experience of more than 1 yrs, 60% has work experience of more than 5 yrs and 20% has work experience of more than 10 yrs. In addition, 70% were male and 30% were female.

While analysing features, we found gender has the highest impact in accuracy of the model, followed by the work experience, and the education. The accuracy of the trained model on the test dataset(has the same distribution as the training dataset) is 85%.

On your very first day, the AI/ML team lead asked you, shall we deploy the trained model in production or not. What would be your suggestion to her?