Problem Statement:

You are given a dataset that represents the job postings for different positions with the offered salary. Your objective is to use the available information in the job posting to try to predict the salary for the position.

The dataset will come in a form of CSV file and will have 900,000 observations and 9 fields including the target variable.

Expected Output:

1- A Jupyter notebook divided into two parts
   a. The first part includes the code used to solve the above problem.
   b. The second part should be an independent section that includes a function that takes the path of a CSV file with an identical structure to the one provided to you for training and returns the predictions for the data sample in the file.

2- A PDF report that includes at least the following sections:
   a. **Data Exploration & preparation:** includes the results of conducting basic EDA on the data as well as the steps taken “if any” to clean and prepare the data for modeling
   b. **Modeling:** includes the comparison of the results of applying different models on the data with a brief explanation of the results and choosing the best model to fit the data.
   c. **Key findings:** includes an analysis of the results of the chosen model to extract useful insight
   d. **Next Step:** includes your suggestions for improving the model accuracy

Assessment Criteria:

The criteria for assessing the output will be based on below points in a descending order of priority:

1- Model accuracy based on the MSE metric
2- Report structure and clarity of information
3- A clean, understandable and well-structured code
4- A working function for predicting unseen data