You can use Python libraries such as numpy, pandas, sklearn, pytorch, matplotlib etc..

1. Use the dataset in salary.zip to create a linear regression model. Split the dataset into training and test using a 70: 30 split. Find the Mean Squared Error on the test set. If you shuffle the data and then split using the mentioned splits, does it have an impact in the model? Show it by shuffling the data 5 times.

2. Use the dataset in salary.zip to create a linear regression model. Split the dataset into training and test using a 80: 20 split. Find the Mean Squared Error on the test set. As the dataset consists of different scales (units) for different columns, does normalization play a factor in the performance of the model? Use min-max scaling and standard normalization to show the impact of normalization.