Working with Data

Data has been split into two groups:

* training set (file)- [Train.xlsxView in a new window](https://olympus.greatlearning.in/courses/2662/files/169133/download?wrap=1)
* test set (File)- [Test.xlsxView in a new window](https://olympus.greatlearning.in/courses/2662/files/169132/download?wrap=1)

The training set is used to build your machine learning model. For the training set, we provide the salary (also known as the variable CTC) for each participant.

The test set should be used to see how well your model performs on unseen data. For the test set, it is your job to predict the outcome variable (CTC) for each participant.

This is an example of what a submission file should look like: [Submission Format.xlsxView in a new window](https://olympus.greatlearning.in/courses/2662/files/168647/download?wrap=1)

Data Dictionary

Data description: [data\_description\_doc.xlsxView in a new window](https://olympus.greatlearning.in/courses/2662/files/168646/download?wrap=1)