**Capstone Project – Problem Statement**

Great Learning PGP - ARTIFICIAL INTELLIGENCE & MACHINE LEARNING culminates with a Capstone Project. Automatic Ticket Assignment is given as the final capstone project

This Project involves “Natural Language Processing” and consists of analyzing ticket short description and description and other relevant features to arrive at a model to automatically assign a ticket to one of the Groups

**Objective**

The objectives of the project:

* Load and clean files for analysis
* Exploratory analysis on data present in the files
* Build frequency plots and word clouds to pictorially represent the data
* Understand the specifics of each text column and be able to tokenize the words
* Complete Feature Engineering to decide on features required for model
* Complete Hypothesis testing, if required
* Build Models aimed at predicting Assignment Group
* Finalize on one best and accurate model

**Summary of Data**

* Data set has one Label and three independent Features
* There are 74 unique values, Label Assignment Group can take
* Data set has some features with null values, and they are in minority
* Data set has some duplicate entries and can be removed
* 75% of the Groups have less than 100 tickets assigned
* Dataset is Highly skewed with most tickets assigned to GR0
* Many Assignment groups have less representation in this Dataset

**Summary of Findings**

**Step 1: Read, clean and ready the data for analysis**

We read the source files and sample the data to form as base for the Auto Assignment prediction exercise. The sub-steps involved in the exercise are:

* Read files
* Do basic processing like removing punctuations, whitespaces, conversion of all words to lower case

**Step 2: Exploratory analysis**