Overview:

Using a set of features, Purpose Financial Analytics Team is asking you to predict which customers will default on a loan.

The Challenge

An applicant that is default on a loan is defined by the **target** variable equaling **'1'**. You are challenged to construct new variables, employ feature-selection methods, and compare different prediction algorithms to approach this dataset.

More specifically, write R or Python code to develop a model on the 'target' variable in the train dataset and score on the associated test set.

Data

You will be provided with two similar datasets that include customer information. One dataset is titled *train.csv* and the other is titled *test.csv*.

Train.csv will contain the details of 20,839 customers and importantly, will reveal whether the customer has defaulted on the loan (i.e., target).

Test.csv will contain similar information but does not disclose the default/not default. It is your task to predict this outcome.

Submission

You will provide us with two files.
1. A CSV file with the test data attached with your prediction named *submission.csv*.
   (A sample submission is provided)
2. A report with the findings and detailing the model methodology applied.
   (You could use this style guide for guidance: https://www.dataquest.io/blog/data-science-project-style-guide/)

** Data Dictionary will not be provided, please use your best judgement as to what a certain variable could mean as it relates to the 'target' variable.