Lecture 2 Assignment: Python Package Development

As a data scientist, you are required to develop a package called **twtr** for twitter data processing and make it ready for the model development. Package is expected to:

1. Clean text (Basic Cleaning like URL / Twitter Handle / Non-ASCII etc)
2. Standardize text

Some standardization techniques discussed at:

https://www.kdnuggets.com/2019/04/text-preprocessing-nlp-machine-learning.html

AIM: To package the twtr package as per the standards.

To be covered:

1. Package should be installable as a package with its dependency (if any).
2. Package should have test cases for its main functionalities.
3. Package repository should have README.md with the sample usage of the package.

You are allowed to use techniques used for twitter data cleaning from Kaggle kernels or blog posts:
E.g. https://www.analyticsvidhya.com/blog/2018/07/hands-on-sentiment-analysis-dataset-python/

Test Dataset: https://www.kaggle.com/crowdflower/twitter-airline-sentiment