**Project Statement**

**Dataset**

You have an Eco-hotel Data Set (https://archive.ics.uci.edu/ml/datasets/Eco-hotel) in the format of comma separated file (dataset.csv), it can be read and open in excel easily.

There are customer reviews in each sentence. For example, the third row has the following sentence.

"The hotel is fantastic built by the sea, living together with nature. Environment it is great as well as people and service.

We full enjoyed the place, and facilities.

Thanks for the ""cidreira"" and ""madalenas"" tea at reception"

**To Do**:

You have one-page pdf document, which explains the term TF-IDF. You will read it carefully, and then calculate the TF-IDF for the given dataset. Also, you will ignore/remove the stop words (the file consists of stop words is also attached with this email), means, you don't need to calculate TF-IDF for stop words.

Important: You will not use NLTK toolkit for this purpose, or any built-in libraries (like scikit-learn)