CIS434 Final Take-Home Project

Download 3 data files from Blackboard. Every line in each file corresponds to a customer tweet sent to an airline (i.e., @ an airline) although sometimes a tweet may mention multiple airlines.

1. The file *complaint.txt* contains samples of customer tweets that are considered by my former TA as complaints.

2. The file *noncomplaint.txt* contains samples of customer tweets that are not considered by my former TA as complaints.

3. The file *customertweets.csv* contains samples of customer tweets that may or may not be complaints, although most of them are complaints. Your task is to find out as many noncomplaints as you can from this file.Your grade depends both on the number of noncomplaints you identify, and whether they are indeed noncomplaints.

A customer tweet is considered a complaint if the tweet expresses anger, frustration, or disappointment towards an airline, and is considered a non-complaint otherwise. Such an evaluation should be clear in most cases, but may be hard in some ambiguous scenarios. For example,

- A tweet may mention multiple airlines, lashing out at one while praising or swearing to fly with the other. Do not include these as non-complaints.

- A tweet may contain too little information, for example, due to text truncation. Include these as non-complaints.

You need to submit 2 files with file names formatted as **firstname_lastname.csv** and **firstname_lastname.pdf**. Suppose a student's first name is John and last name is Smith.

1. john_smith.csv: a CSV file of **non-complaint** tweets consisting of **two columns**. Please note that unless your file conforms to the following requirements, you will receive **0** grade.

   - **Only** include your identified non-complaint tweets in this CSV file.
   - **Only** include 2 columns in this CSV file and the 2 columns **must be separated by comma** and be in the following order.
     - The first column contains the *id* field from the **original** csv file.
     - The second column contains your identified **non-complaint** tweet.
   - Please do not put quotation marks around any field.

2. john_smith.pdf: a PDF file describing how you completed the task, along with any code you used.