STAT 3250 **Assignment 7**

Directions: Please submit your Python code file, formatted as in previous assignments. You will also be uploading the output file described below.

The grader should be able to run your code when it is placed in the same directory as the input data file. Be sure that your code loads any libraries you are using.

Background

An active thread on Reddit is "Random Acts of Pizza" where a user can put in a request for a pizza to be donated by another user. Some requests are granted, others are not. The file `pizza_requests.txt` includes requests from approximately 5600 pizza requests. Data for each request includes a statement describing the nature of the request, and various other pieces of information. The file `README_pizza.txt` gives more information about the data set.

Use the data set to answer the questions below.

1. What proportion of requests were successful? (The requester received pizza.)
2. Find the median account age at the time of request for all requests.
3. Divide the requests into those with account age greater than the median found in the previous question, and those with account age less than or equal to the median. Find a 95% confidence interval for the difference in proportion of successful pizza requests between the two groups. The formula is

$$\hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

   where $\hat{p}_1$ is the proportion of successful requests among older accounts (age greater than the median age), n_1 is the number of older accounts, and $\hat{p}_2$ and n_2 are the same for the newer accounts.
4. Determine the percentage of request texts that mention the word "student" or "children". (Upper or lower case.)
5. Determine the number of requests from Canada.
6. Find a 95% confidence interval for the proportion of successful pizza requests donated anonymously.
7. Find the maximum number of subReddits subscribed to by a single requestor.
8. Determine the number of distinct subReddits among all the requests, and the number of times that each appears. Place a table of the 10 most frequently occurring (in order, starting with most frequent) in your Python code file, organized

```
subReddit01, count01
subReddit02, count02
subReddit03, count03
subReddit04, count04
subReddit05, count05
subReddit06, count06
subReddit07, count07
subReddit08, count08
subReddit09, count09
subReddit10, count10
```

   Also write to file all of the subReddits and their corresponding counts to the file named

   `XXXXX-assignment07-subreddits.txt`

   Replace `XXXXX` with your computing ID.