**SQL Assignment - 2**

**Title**: **Counting Organizations**

This application will read the mailbox data (mbox.txt) and count the number of email messages per organization (i.e. domain name of the email address) using a database with the following schema to maintain the counts.

CREATE TABLE Counts (org TEXT, count INTEGER)

When you have run the program on mbox.txt upload the resulting database file above for grading.

If you run the program multiple times in testing or with different files, make sure to empty out the data before each run.

The data file for this application is mbox.txt

The expected output from this task is;

“

Enter file name: mbox.txt

Counts:

iupui.edu 536

umich.edu 491

indiana.edu 178

caret.cam.ac.uk 157

vt.edu 110

uct.ac.za 96

media.berkeley.edu 56

ufp.pt 28

gmail.com 25

et.gatech.edu 17

”

Evaluation Scheme:

1. Read the mbox.txt file in python - 20 marks
2. Create a logic for cleaning the junk words and symbols - 40 marks
3. Insert the count of domain names into a database table - 30 marks
4. Returning the expected output of the domain counts - 10 marks