***Step 4-a\_ last step:***

Now, let’s get closer to the real world.

Here are the changes we need to implement to make this project more real:

1. In the previous step, you simply copied and pasted the two lists that I gave you for positive and negative words. I told you that those lists are not comprehensive. Now, instead of those lists, I want you to use comprehensive lists of positive and negative words.

For this purpose, I have provided you two files: negatives.txt and postives.txt.

**Hint:** In previous steps, you copied and pasted the given lists into your programs in order to create lists of negative\_words and positive\_words. Now, instead of copying and pasting, you need to create these two lists from the two given txt files. You need to read the data from the files and then convert them into an appropriate format that your current program can work with.

1. Most of you, during step 3, took care of special characters. Well done! However, different groups used different special characters. For example, some of you just chose

[‘(’ , ‘)’, ‘@’, ‘/’]

and some of you picked

[‘,’ , ‘.’ , ‘’’ , ‘’’ , ‘!’ , ‘?’ , ‘&’, ‘/’, ‘\*’ , ‘#’ ].

Although taking care of special characters in text mining is a critical step, consistency is a must. Therefore, in order to be consistent, in this final step let’s take care of the following special characters:

SC = [',' , '.' ,'”' , '’' , '"' , "'" , '?' , '!' , '&', '@', '\*' , '#']

Just copy and paste this list into your program. Do not add or delete anything from this list.

1. In real world, usually when analysts conduct a sentiment analysis, they are usually interested to know about the positive or negative words that have been used to find meaningful patterns. So, let’s provide this opportunity for our user.

For this purpose, your program needs to provide two lists for the mentioned positive and negative words. So, the output of your program should be

1. Count of Positive Words
2. Count of Negative Words
3. Length of the text
4. Valence
5. List of Mentioned Positive words
6. List of Mentioned Negative words

***Guideline:***

To run your programs, in the previous steps, I ran two tests:

the painting was displaying a man bleeding and, yet gorgeously presenting love

goood, grace! and #graceful and help! bleeding? blur

None of them made any sense. But I just wanted to check if your programs were running correctly.

After your have completed step 4, you can check the result of your program by running these two tests. See if your program gives you the correct answers.

***Instruction:***

Once you are sure that your program is working, name your file as

*yourFirstName\_yourTeammateFirstName\_step4-a.py*

then submitted as the third part of your project.

***Knowledge needed:***

To do this part of the project, you need to have enough knowledge about lists and strings, and you must know how to work with file. (Chapters 6, 7 and 8).

***Challenge yourself:***

***Step4-b-Optional – no extra credit- no penalty***

The above step is a very simplified version of sentiment analysis. Due to COVID-19 situation and the conversion of this course into an online course, I thought it would be just fair if I simplify this project as much as possible. However, if you like to challenge yourself, I would suggest adding the details below to your project. DON’T WORRY, IF YOU MAKE A MISTAKE ON THIS CHALLENGE, YOU WILL NOT BE PENALZIED AS LONG AS YOU SUBMIT THE STEP4-A. If you decided to accept the challenge you properly would need my help (or Mr. google’s help) to learn how to work with a cvs file. 😉 So, don’t shy away, send me an email and we can set up a meeting to help you.

Your program should conduct a sentiment analysis for 20 posts that have been extracted from a social media platform. These 20 posts have been extracted and imported into a .cvs file. Your program requires to conduct the sentiment analysis for each post and for each post it should calculate

1. Count of Positive Words
2. Count of Negative Words
3. Length of the text
4. Valence
5. List of Mentioned Positive words
6. List of Mentioned Negative words

The 20 posts are presented in the table below. Look at the posts and get a better sense towards each of the comments/posts. The file of these posts is presented to you in the project module as project\_gab\_data.csv .

|  |
| --- |
| **posts\_text** |
| Check@gabuser: On now - @Fusion scores first points #FirstFinals @overwatchleague @umich @umsi Michigan Athletics made out of emojis. #GoBlue |
| BUNCH of things about crisis respons? available July 8th? scholarship focuses on improving me? in North America! A s? and frigid temperatures |
| FREE ice cream with these local area deals: chance to pitch yourself to hundreds of employers? Msi student Name Here is spending her summer |
| AWAY from the Internet of Things attacks? right and the left? See? use DIY language to discuss other projects not traditionally viewed as masculine |
| IN City Name!? from @gabuser has some amazing datasets and tools for collective action. - UMSI a? grateful for? equipping elderly African-American |
| ENTREPRENEURSHIP in #City.? a great social media job opportunity for our UMSI alumni looking to work at City Name Public Schools! #goblue |
| BRINGING #datascience to community researchers with a team of researchers have been helping people in 'lean' economies learn entrepreneurial |
| WHY not pay you?? endure crushing pressures and frigid temperatures but the most difficult part of? faculty and staff across the city |
| A bunch of things about crisis respons? ? but the implications are much bigger t? for some but a financial burden for others. @umichdpss and |
| @THEEMMYS nomination for Best Lead Actor in a library role focused on Digital Services and Projects? As a bonus you get to work at t? by @cab938 |
| THIS headline has been inescapable this summer. Now the infectious chart-topper from 'Scorpion' gets a vi? Seventh Victim 'Scorpion' gets a |
| OF wine with a shoe? Yes but it ain't pretty. Its First Amendment rights. That it claims will for? and non-binary musicians shaping the sound |
| HAVE detained six people in history to ride to orbit in private space taxis next year if all to faithful them is that they simply can't afford |
| PET Name. She is 1 year old Shiba Inu. When her caregiver is not at home Name likes to have? I tell them about my website and ask if my |
| YOU?RE welcome! He was a mix of many breeds. He is a 2 year old Yellow Lab. When he was a mix of breeds but has the bark of a road!! |
| Name. He is wild and playful. He likes to chase and play with his stuffed elephant! the Dir. Of Human Resources @gabuser. He is a |
| BORDER Terrier puppy. Name is loving and very protective of the people she loves. Name2 is a 3 year old Maltipoo. Name3 is an 8 year old Corgi. |
| REASON they did not rain but they will reign beautifully couldn't asked for a crime 80 years in the Spring Name's Last Love absolutely love |
| HOME surrounded by snow in my Garden. But City Name people musn't: such a good book: RT @gabuser The Literature of Conflicted Lands after a |
| Fake news everywhere! Horrible and terrifying @newsfollower we should not follow social media. It is corrupted and bad! |

1. Table 1: Table of the posts

All you need to do is to read this cvs file and iterate through each row separately (for iteration we know we need a loop) to get access/ read each line. Once you can access each line you need to conduct your sentiment analysis from step 4-a and implement it as a one major program. So, when you conduct the entire program, your program must provide 20 sets of results, for each post, 1 set of output.

Be aware that the first line / row is a title and it is not a post. So, you need to skip this line in your iteration.

***Instruction:***

Once you are sure that your program is working, name your file as

*yourFirstName\_yourTeammateFirstName\_step4-b.py*

then submitted as the third part of your project.

***Knowledge needed:***

To do this part of the project, you need to have enough knowledge about lists and strings. You must also know how to work with txt and cvs files. (Chapters 6, 7 and 8).

***Challenge yourself:***

***Step4-c-Optional – no extra credit- no penalty***

I have another challenge for you. In this step you need to save your output into a new file. This way, your user can use this file and export to into other applications for further analysis.

To do so, the result of sentiment analysis that your program has created for each post should be saved into a new file. This new file is cvs.

For each post, your program should save only 3 values (out of the 6 values of sentiment analysis):

1. Count of Positive Words
2. Count of Negative Words
3. Valence

The output of this program is a two-dimensional list with 21 elements. The first element represents the header and the other 20 elements represent the information about each data point (each post). This list is a 2-d list, therefore each element itself is a list of three elements. The list is represented in a table format for a better undertraining. You need to create such a table in your CSV file. Name this file samantest\_resulting\_data.csv

Your csv file should populated values similar to table 2.

|  |  |  |
| --- | --- | --- |
| Positive Score | Negative Score | Net Sentiment Score |
| 0 | 0 | 0 |
| 2 | 2 | 0 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |
| 2 | 0 | 2 |
| 2 | 0 | 2 |
| 2 | 0 | 2 |
| 0 | 3 | -3 |
| 0 | 2 | -2 |
| 4 | 0 | 4 |
| 0 | 1 | -1 |
| 1 | 0 | 1 |
| 2 | 0 | 2 |
| 1 | 0 | 1 |
| 1 | 0 | 1 |
| 2 | 1 | 1 |
| 3 | 0 | 3 |
| 3 | 1 | 2 |
| 1 | 1 | 0 |
| 0 | 4 | -4 |

Table2: table representation of the requested 2-D list

:

***Instruction:***

Once you are sure that your program is working, name your file as

*yourFirstName\_yourTeammateFirstName\_step4-c.py*

then submitted as the third part of your project.

***Knowledge needed:***

To do this part of the project, you need to have enough knowledge about lists and strings. You must also know how to work with txt and cvs files. (Chapters 6, 7 and 8).