Neon Data Scientist Technical Test

Version: neon-data-202111

Objective:

The purpose of this technical test is to test your ability to analyse and solve real business problems, write scalable solution and code (Python code would be preferred, pseudo code is also accepted). You would be expected to explore possible solutions and show your thinking on solving the problem.

Primary objectives of this test as follow:

1. Brief description of steps on how you will target the problem, including techniques, packages, algorithms might be involved.
2. Python code or pseudo code.
3. Optional: results/output from the test.

Task:

Neon values voices from customer. We use surveys to gather feedback on customers' experience using Neon. One of the survey questions is "What contents did you recently watch on Neon?", this is an open-text question, which means customer can type in what they want to say. Please find the sample data for this test from *Survey response sample data.csv*, which contains data as below:

Customer_id	Response
1	Fear the walking dead,Supernatural (huge fan and sad it has finished),The Gentlemen, Outlander
2	A lot! -good doctor -gangs of London - the gentleman -ma -spies in disguise
3	Miss scarlet and the duke,knifes out,Dublin murders
4	History drama-Vikings,Kid friendly-Casper,Sometimes the conversations while watching Neon can get serious but we all end up having fun together, :)
5	The Undoing,Game of thrones,Outlander, Vikings,CB Strike (and most all British dramas) Westworld

As you can see from responses, there are misspellings, emojis, etc. The task will be trying to extract content names out of the responses, this will require the attached *Content sample.csv* data which contains 17 content names relating to this task.

Ideal output for the task will be, for each of the response, list as many as possible content names mentioned in the response, see below as an example

Customer_id	Response	Content names
1	Fear the walking dead,Supernatural (huge fan and sad it has finished),The Gentlemen, Outlander	Fear the Walking Dead Supernatural The Gentlemen Outlander
2	A lot! -good doctor -gangs of London - the gentleman -ma -spies in disguise	…
3	Miss scarlet and the duke,knifes out,Dublin murders	…
4	History drama-Vikings,Kid friendly-Casper,Sometimes the conversations while watching Neon can get serious but we all end up having fun together, :)	…
5	The Undoing,Game of thrones,Outlander, Vikings,CB Strike (and most all British dramas) Westworld	…

Hints:

- Natural Language Processing
- Named Entity Recognition

Attachments:

- Neon Data Scientist Technical Test.pdf
- Survey response sample data.csv
- Content sample.csv

Test output:

Test output will be used evaluate your work, as listed below, point 1 will be required, for point 2 and 3, you can choose one of them.

1. A brief description about how to target the problem, including possible steps, techniques, packages, etc.
2. If you prefer Python code: Git repository contains Python code or Jupyter Notebook (Optional), or attach code into the brief description document as Appendix.
3. If you prefer pseudo code: attach pseudo code into the brief description document as Appendix.