THE DATAs CHALLENGE 2022

**1. How you proceeded with the data. Your thought process before you started working!**

When I opened the text file, my first thought is that the content is so long, I need to find the point of the article and condense it into a few sentences. Besides, when it comes to earning call speech, it is useful for investors to know what the most frequently mentioned thing by CEO is, and the sentiment of CEO in this speech. Therefore, there are three main analysis parts which are: A. Show the most frequency noun mentioned by CEO. B. Calculate the sentiment of each sentences. C. Condense the speech into a summary.

**2. Tools, languages, software's used**

In this project, I mainly use NLTK to process the data, and use matplotlib to visualize the data, the languages I use is Python.

**3. Answers to these 2 simple questions: (you may need to read the text a little bit)**

a. The company you think this is related to?

The company is Twitter, it is related to consumer service. It is able to show what's happening immediately, being able to get people their news immediately, about anything that's new and noteworthy, and also to show all the commentary and conversations around it.

b. Financial year CEO is discussing?

The year of 2016 and 2017.

I organize the sentences including financial year into a dictionary so that if I want to see the content of financial year CEO discussing, I can search it quickly.

﻿一張含有 文字 的圖片

自動產生的描述

Fig. 1 Sentences include financial year

**4. The visualizations, insights you extracted from data – images, thoughts, anything!**

A. Show the most frequency noun mentioned by CEO.

 

Fig. 2 Frequency of proper noun Fig. 3 Frequency of noun

B. Calculate the sentiment of each sentences.



Fig. 4 CEO sentiment analysis

一張含有 文字 的圖片

自動產生的描述C. Condense the speech into a summary.

Fig. 5 Speech summary

**5. What more you think can be done in future on this data!**

Improve the accuracy of the analysis results.

**6. Any code (if you have written) should be uploaded in a public GitHub repository.**

<https://github.com/YuChing1204/THE-DATAs-CHALLENGE-2022.git>