

# SIT112 - TASK COMPLETION REPORT

**Important note 1:** Please **do not** include Python code in this report. It would be acceptable though to make brief references to function names or different parts of the task's notebook when needed.

**Important note 2:** Please **do not** leave any response field empty; if not applicable, enter NA.

## 1. TASK SUMMARY

TASK NAME (ABBREVIATION)	YOUR NAME (STUDENT ID)	YOUR DEAKIN EMAIL
High Distinction Task HD	Pham Hoang Son (223240555)	s223240555@deakin.edu.au

## 2. TASK DESCRIPTION

ITEM	RESPONSE
What was the objective(s) of the task?	The objective of this task was to test the hypothesis that video titles that contain certain words are more likely to elicit sentiment in the comments
What kind of data did you work with?	The task involved analyzing one dataset: covid 19 data from 4 news publishers: CNN, BBC, Fox and MSNBC
Briefly describe the data science task you worked on.	The data was cleaned using NLP techniques, and prepared for analysis, and three word clouds were created to visualize the relationship between words frequencies and sentiment category. Also, a bar chart was made to further analyze the data

## 3. TECHNICAL SKILLS

ITEM	RESPONSE
What technical skills did you use during the task?	I used several technical skills during the task, including data cleaning (using NLP techniques to clean the video titles), merging datasets, data preparation (converting Counter type and dictionary into data frames), data analysis, and data visualization (creating a word clouds and a bar chart).
list any challenges or obstacles you faced while working on the task and how you overcame them.	One challenge I faced was dealing with the dictionary type when trying to convert it to a data frame. I looked for the approach on the StackOverFlow forum and overcame the problem.

## 4. DATA CLEANING AND PREPARATION (ENTER NA WHEN NOT APPLICABLE)

ITEM	RESPONSE
What steps did you take to clean and prepare the data?	I removed the stop words, punctuations and special characters from the video titles and extracted singular words from them.
Did you encounter any issues with the data during this process? How did you address these issues with the data?	This was the first time I have used the NLP techniques, so I had quite many difficulties. However, by reading the reference on the nltk website, I could manage the task.

## 5. DATA ANALYSIS (ENTER NA WHEN NOT APPLICABLE)

ITEM	RESPONSE
How did you analyze the data?	I analyzed the data by counting the number of times each word appears among the sentiment categories and compare them.
Did you use any visualization techniques to better understand the data?	Yes, I used 3 word clouds for three sentiment categories(positive, negative and neutral) and a bar chart
What insights did you gain from this analysis?	From the analysis, I found that a video title that contain certain words has a tendency to receive a specific sentimental response from the viewers. This analysis states the hypothesis as being true.

## 6. BASIC REQUIREMENTS FOR THE TASK

ITEM	RESPONSE (YES/NO)
Are you confident to execute the Python code in this task and explain the output?	Yes
Are you confident to explain what each line of code does and how it contributes to the solution(s)?	Yes
Are you confident to rewrite or modify the code after completing this task? <ul style="list-style-type: none"> <li>For pass tasks: with guidance, no time limit.</li> <li>For credit tasks: with limited guidance, no time limit.</li> <li>For distinction tasks: independently, no time limit.</li> <li>For high distinction tasks: independently, in a limited time.</li> </ul>	Yes

## 7. CODE ATTACHMENT (NOT APPLICABLE TO THE PASS TASKS – ENTER NA)

ITEM	RESPONSE (YES/NO/NA)
Have you attached the notebook file that contains your solutions (Python code) for this task?	YES

ITEM	RESPONSE (YES/NO/NA)
Have you executed all the cells in your attached notebook and ensured there is no error?	YES
<i>Please note your submission will not be flagged as complete if your attached notebook contains any error.</i>	

## 8. VIDEO ATTACHMENT (NOT APPLICABLE TO THE PASS/CREDIT TASKS: NA)

ITEM	RESPONSE (THE VIDEO LINK/NA)
Provide the link to the video recording that presents your completed task. This is only for Distinction and High Distinction tasks. Enter NA for Pass/Credit tasks.	<a href="https://deakin.au.panopto.com/Panopto/Pages/Viewer.aspx?id=f97c6a84-f97c-44af-9401-b17300c68dc2">https://deakin.au.panopto.com/Panopto/Pages/Viewer.aspx?id=f97c6a84-f97c-44af-9401-b17300c68dc2</a>

## 9. ACKNOWLEDGEMENT

BY SUBMITTING THIS REPORT, I ACKNOWLEDGE THAT:

- MY RESPONSES ARE ACCURATE AND ARE MY OWN WORDS.
- I HAVE MET ALL THE BASIC REQUIREMENTS OF THE TASK (LISTED IN SECTION 6).
- I HAVE READ AND FULLY UNDERSTOOD THE ASSESSMENT GUIDELINE OF THE UNIT.
- THIS REPORT DOES NOT EXCEED 3 PAGES.
- THIS REPORT DOES NOT INCLUDE CODE EXCEPT BRIEF REFERENCES TO FUNCTION NAMES OR DIFFERENT PARTS OF THE TASK'S NOTEBOOK.
- MY SUBMISSION DOES NOT CONTAIN ANY CREDENTIALS (E.G., PASSWORD, API KEY, ETC) OR PERSONAL INFORMATION.

**IMPORTANT NOTE 3:** IF YOU HAVE ANSWERED NO TO ANY OF THE QUESTIONS IN SECTIONS 6, PLEASE RECONSIDER SUBMITTING YOUR REPORT; ASK HELP FROM YOUR TUTOR.

ADD YOUR NAME AND SIGNATURE HERE: PHAM HOANG SON

