

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
Pass Task P8	Armaan Chetal (s222183675)	s222183675@deakin.edu.au

2. TASK DESCRIPTION

ITEM	RESPONSE
What was the objective(s) of the task?	The objective to apply basic natural language processing techniques to analyze user replies to Covid-19 news on the platform Twitter. This involves cleaning and preprocessing text data, performing frequency analysis, and conducting sentiment analysis to gain insights into user engagement and sentiment toward the news content from various publishers.
What kind of data did you work with?	The task involves working with a dataset named replies.csv, which contains information about user replies to Covid-19 news posted on the social media platform Twitter. This diverse set of data allows for a comprehensive analysis of the user replies, including text content analysis, sentiment analysis, and understanding the engagement dynamics of the replies in relation to the authors and news publishers.
Briefly describe the data science task you worked on.	The data science task involves analyzing user replies to Covid-19 news on a social media platform using NLP techniques. Data is loaded from a CSV file containing reply details. Text cleaning includes removing mentions, hashtags, URLs, punctuation, and converting text to lowercase. Word frequency analysis identifies the top 5 words per news publisher. Sentiment analysis classifies replies as positive, negative, or neutral, visualizing sentiments across publishers. The task aims to understand user reactions to Covid-19 news, including sentiment towards publishers and common discussion topics. Through NLP, it provides insights into public perception and engagement with pandemic-related content on social media.

3. TECHNICAL SKILLS

ITEM	RESPONSE
What technical skills did you use during the task?	During the task, I utilized skills in data manipulation with Pandas, text preprocessing using regular expressions, NLP techniques like sentiment analysis with TextBlob, statistical analysis for word frequency, and data visualization with Matplotlib and Seaborn. These skills enabled the extraction of insights from user replies to Covid-19 news on social media, including sentiment distribution and common discussion topics among news publishers.
list any challenges or obstacles you faced while working on the task and how you overcame them.	One challenge was handling diverse text formats in replies, requiring robust cleaning methods to ensure uniformity. Overcoming this, I implemented comprehensive text preprocessing using regular expressions. Another challenge was the presence of irrelevant words in replies affecting sentiment analysis. To address this, I curated a custom list of stopwords and excluded words tailored to the context, refining sentiment classification accuracy..

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

ITEM	RESPONSE
What steps did you take to clean and prepare the data?	To clean and prepare the data, I used regular expressions to remove mentions, hashtags, URLs, and punctuation, ensuring uniformity. Lowercasing standardized text format. I used NLTK's English stopwords list to remove common stopwords and defined a custom list to filter out context-specific words like "covid". These steps ensured data consistency and relevance for subsequent analysis
Did you encounter any issues with the data during this process? How did you address these issues with the data?	NA

5. DATA ANALYSIS (ENTER NA WHEN NOT APPLICABLE)

ITEM	RESPONSE
How did you analyze the data?	The data analysis involved extracting words using regular expressions, grouping data by news publishers, and aggregating word lists. Word frequency analysis identified the top 5 words per publisher. Sentiment analysis classified replies as positive, negative, or neutral. Visualization depicted sentiment distribution across publishers.

ITEM	RESPONSE
Did you use any visualization techniques to better understand the data?	Yes, I utilized visualization techniques to gain insights from the data. Specifically, I created a stacked bar plot to illustrate the distribution of sentiment (positive, negative, or neutral) across different news publishers. This visualization enabled a clear comparison of how users reacted to Covid-19 news from various sources, facilitating a better understanding of sentiment trends and engagement levels across publishers.
What insights did you gain from this analysis?	From the analysis, several insights emerged. Firstly, we identified common discussion topics across news publishers, such as "vaccine" and "people". Secondly, sentiment analysis revealed varying reactions towards Covid-19 news, with some users expressing positivity, while others showed negativity or remained neutral. Additionally, the analysis highlighted engagement levels, with certain publishers eliciting more responses.

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"> For pass tasks: with guidance, no time limit. For credit tasks: with limited guidance, no time limit. For distinction tasks: independently, no time limit. For high distinction tasks: independently, in a limited time. 	Yes

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

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

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.	NA

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: ARMAAN CHETAL