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	Krystal Nguyen (223212228)	s223212228@deakin.edu.au

2. TASK DESCRIPTION

ITEM	RESPONSE
What was the objective(s) of the task?	The objective of this task was to exploring the relationship between sentiment and publishers by analyzing sentiment distribution in video titles and comments.
What kind of data did you work with?	The task involved analyzing two datasets: one works with videos and the another works with comments. I worked with many types of data: text, numbers, dates.
Briefly describe the data science task you worked on.	The task performs data preprocessing and analysis on these datasets, applying sentiment analysis to the video titles and comments, and aggregating the data to calculate sentiment distributions for different publishers, months, and sentiment categories.

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, merging datasets, data preparation (date extraction), data analysis (calculating size counts), and data visualization (creating stacked bar charts).
list any challenges or obstacles you faced while working on the task and how you overcame them.	One challenge I faced was dealing with missing values in the datasets (February had no comments). I overcame this by creating a list of dictionaries representing the new rows for month 2. This can be concatenated with the existed data frame to ensure all channels have data for each month.

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

ITEM	RESPONSE
What steps did you take to clean and prepare the data?	I created a list of dictionaries representing the new rows for month 2 and concatenated with existed data frame. Grouping video titles by publisher, month, and sentiment using the groupby() function, merges the comments_df with the videos_df.
Did you encounter any issues with the data during this process? How did you address these issues with the data?	I got many syntax errors during this process, I solved this by analyzing the errors and search into it (using code library and chat gpt).

5. DATA ANALYSIS (ENTER NA WHEN NOT APPLICABLE)

ITEM	RESPONSE
How did you analyze the data?	The code calculates the sentiment distribution for video titles and comments. After group the video titles and comments by publisher, month, and sentiment using the groupby() function, I could calculate the size (count) of each sentiment category within each group. These calculations could make better understanding distribution of sentiment across different publishers and months.
Did you use any visualization techniques to better understand the data?	Yes, I visualized the sentiment distribution of video titles and comments using stacked bar charts. Also, bar charts were created to show the proportion of positive, neutral, and negative sentiments for each month and channel, providing a visual representation of the sentiment trends.
What insights did you gain from this analysis?	From the analysis, I can guess the content or new topics covered by 4 channels that may evoke negative/positive/neutral reactions.

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? • 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?	YES
Have you executed all the cells in your attached notebook and ensured there is no error? *Please note your submission will not be flagged as complete if your attached notebook contains any error.*	YES

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.	https://youtu.be/K1-lXU6-Eys

9. AKNOWLEDGEMENT

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 INFORMAIOTN.

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: KRYSTAL NGUYEN