

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
HD Task	Hoang Long Tran (s223128143)	s223128143@deakin.edu.au

2. TASK DESCRIPTION

ITEM	RESPONSE
What was the objective(s) of the task?	The objective of this task was to analyze the COVID 19 video data to find any meaningful relationship between the sentiment of video title and the sentiment of its corresponding comments. The same applies for the video likes and duration. I use correlation and regression analysis to find if there is any relationship between the 3 predictors and the response variable.
What kind of data did you work with?	The data I work with was YouTube video data on COVID 19. Most of the necessary data has already been cleaned and merged for me, I just need to find the sentiment score for the video title and the comments.
Briefly describe the data science task you worked on.	The data science task I work on was to find if there is any meaningful relationship between 3 predictors (video title sentiment score, video likes and duration) and the comment sentiment score from COVID 19 YouTube video. I use correlation and regression analysis to find meaningful relationships.

3. TECHNICAL SKILLS

ITEM	RESPONSE
What technical skills did you use during the task?	There are many technical skills involved. The first was to understand and interpret the scatterplots and correlation coefficient. Understand and interpret the regression model output like the R-Squared and p-value. Plot some plots related to regression and check if the assumption is met or not.
list any challenges or obstacles you faced while working on the task and how you overcame them.	The most difficult challenge I faced during this task was to understand what regression analysis is. I need to understand the steps involve in a regression analysis, how to output what I found, how to check for the

ITEM	RESPONSE
	assumption of the model, did it violate any assumptions and how to deal with those assumption when it is violated.

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

ITEM	RESPONSE
What steps did you take to clean and prepare the data?	Most of the data has been cleaned and prepared for me. One thing I did to prepare the data was to use a model to predict the sentiment score for video titles and comments. I did not use the TextBlob libraries because I feel like it gives not very accurate result.
Did you encounter any issues with the data during this process? How did you address these issues with the data?	I found issues with the sentiment score from the TextBlob, so I use a pretrained model to predict the sentiment. Similar to some of my previous answers, the issue I faced the most was understanding what regression analysis is. To understand it, I have watched many videos and created a few jupyter notebook for testing.

5. DATA ANALYSIS (ENTER NA WHEN NOT APPLICABLE)

ITEM	RESPONSE
How did you analyze the data?	I first check the correlations. Then I will regression analysis to find the relationship between the predictors and the response variable. There are methods involves which I have explained in the video.
Did you use any visualization techniques to better understand the data?	Yes, I used many plots. The first plot is scatterplot to look at the correlation between variables. The second plot is leveraging plot, to look for potential outliers in our model. The third plot is normal QQ, to see if the residuals are normally distributed or not. The fourth plot is fitted vs residual plot to check for homoskedasticity. The final plot is the partial regression plot to see what effect the predictor has on the response variable when the other predictors have been isolated.
What insights did you gain from this analysis?	I conclude that the video title sentiment score has a positive relationship with the comment sentiment score. The duration has a negative relationship with the comment sentiment score. The like count has no meaningful relationship. Here are some of my assumptions after the conclusion. The positive relationship between video title sentiment and comment sentiment may indicate that the tone set by the title greatly affects the comments. What I mean by this is during the time of uncertainties, people tend to find positive things to uplift their mood. The negative relationship between duration and comment sentiment suggests that longer videos might overwhelm the users by their length and cause anxiety. This might make the users to leave down fear comments.

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?	Yes
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>	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.	Link: https://www.youtube.com/watch?v=lOGot4lMSkM

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 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: **HOANG LONG TRAN**

A handwritten signature in black ink, appearing to be 'Long', written over a horizontal line.