**Dashboards Manual**

Last Revised: 20 Oct 2022

[**01 Introduction**](#_8ljenajaa4dp) **2**

[**02 Installation**](#_60vzclgezq5) **3**

[**03 Good To Know**](#_qknycjcp3978) **3**

[3.1 Understanding Folder Structure](#_okcjegdtpvvw) 3

[3.2 Good File Naming Convention](#_m5hbtga5w5r6) 4

[3.3 Graph Functionalities](#_3vqsiq7wqsxw) 4

[**04 Getting Started**](#_smobuqtih69h) **6**

[4.1 Dashboard: Topic Modeling](#_amhw01h9oco) 6

[4.2 Dashboard: Text Summarisation](#_jfb1twruruh) 7

[4.3 Dashboard: Sentiment Analysis](#_w46l9j3djs6y) 8

[4.4 Dashboard: Social Media](#_w1edjfmbzuwi) 9

# **01 Introduction**

A brief overview of each of the dashboards are as described below:

| **Dashboard** | **Description** |
| --- | --- |
| Topic Modeling | Performs clustering of text data into similar groups for ease of identifying topics  Models utilized: roBERTa |
| Text Summarisation | Performs summarisation of key points in text data  Models utilized: TextRank and BART |
| Sentiment Analysis | Performs analysis of sentiments of text data  Models utilized: distilbert |
| Social Media Analysis | Displays time-series visualizations of social media data |

The 4 dashboard tools as listed above have the overall objective of improving research efficiency by providing insights into text data via natural language processing models.

# **02 Installation**

Proceed to this link and follow the instructions on the page to set up the application.

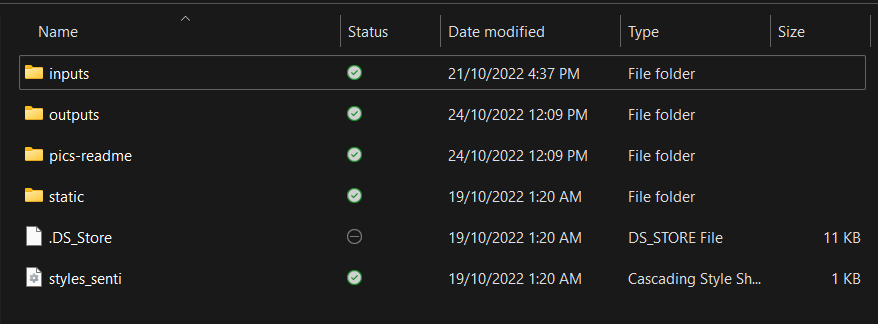
Note: There are different sets of instructions for Windows and Macs. Please follow the correct instructions based on the system you are using.

# **03 Good To Know**

Here is some information that is good to know before you start using the application.

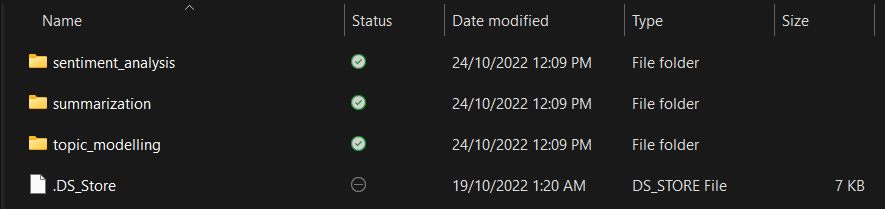
## 3.1 Understanding Folder Structure

Inside the downloaded file, you will see many folders and subfolders. The only key folder you need to take note of is the “**assets**” folder. Upon entering this folder, you will see the following contents.

****

The “**inputs**” folder is where you should put your input files for running any of the dashboards so that you can easily find your input files. However, this is not mandatory. When using the application and you are prompted to select the input file, you are free to select a file from any location as long as it is on your computer.

The “**outputs**” folder is where you can find all the output files - the results from running the different dashboards. Upon entering this folder, you will see the following contents. It contains subfolders which contain the output files for the different dashboards.



The output files found here are either excel files used for plotting the graphs you see in the dashboards (for sentiment\_analysis and topic\_modelling) or text files (for summarization). In the event that you wish to have these results files in another location on your computer, please copy and paste the files. Do not remove the file from this folder.

## 3.2 Good File Naming Convention

Having good file naming convention is important as the output files will be named using your input file name. Good file names ensure that users can easily understand what each file contains and also avoid having two files with the same name.

Guidelines:

1. Describe the file contents in a few words
2. Provide a date - this can be the date that you are doing analysis or the date that the data is collected

| Examples of Good File Names | Examples of Poor File Names |
| --- | --- |
| “Interview\_Covid\_2022-10-26.txt”  “Interview\_Race\_2022-02-28.txt”  “September\_Survey\_2022-10-15.xlsx”  “October\_Survey\_2022-10-15.xlsx” | “Interview.txt”  “Covid dataset.xlsx”  “Survey.xlsx” |

## 3.3 Graph Functionalities

While using the dashboards, you will see many graphs for you to analyze. These graphs come with some functionalities that can help you in your analysis. These functionalities can be seen on the top right when you hover your mouse over the graph.

****

For a more comprehensive guide on how to use the functionalities, head to this [link](https://plotly.com/chart-studio-help/zoom-pan-hover-controls/).

Here are some key functionalities that you should use:

* Download graph (icon #1): clicking on this will download the graph as a png



* Zooming (icon #2): an alternative to this will be pinching the touchpad or scroll the mouse



# **04 Getting Started**

In this section, there are step-by-step procedures on how to use the different dashboards.

1. Set up the application - <insert public repo link>
2. Start the application. <insert pic of app>
3. Select the Dashboard and refer to the procedures below

## **4.1 Dashboard: Topic Modeling**

This tool serves to identify the topics covered in the input file. The dashboard will display the proportion of the topics and the keywords under those topics.

Input File Requirement

* File Format: .xlsx
* Column required: “text”

Procedures

1. Select “Topic Modeling” in the dropdown.
2. Select the input file that meets the above input file requirement.
3. Click “Confirm”.
4. While waiting, look at the Terminal portion to see the progress.
5. Press “Ctrl” and click the link in the Terminal simultaneously to open the Dashboard in the browser.

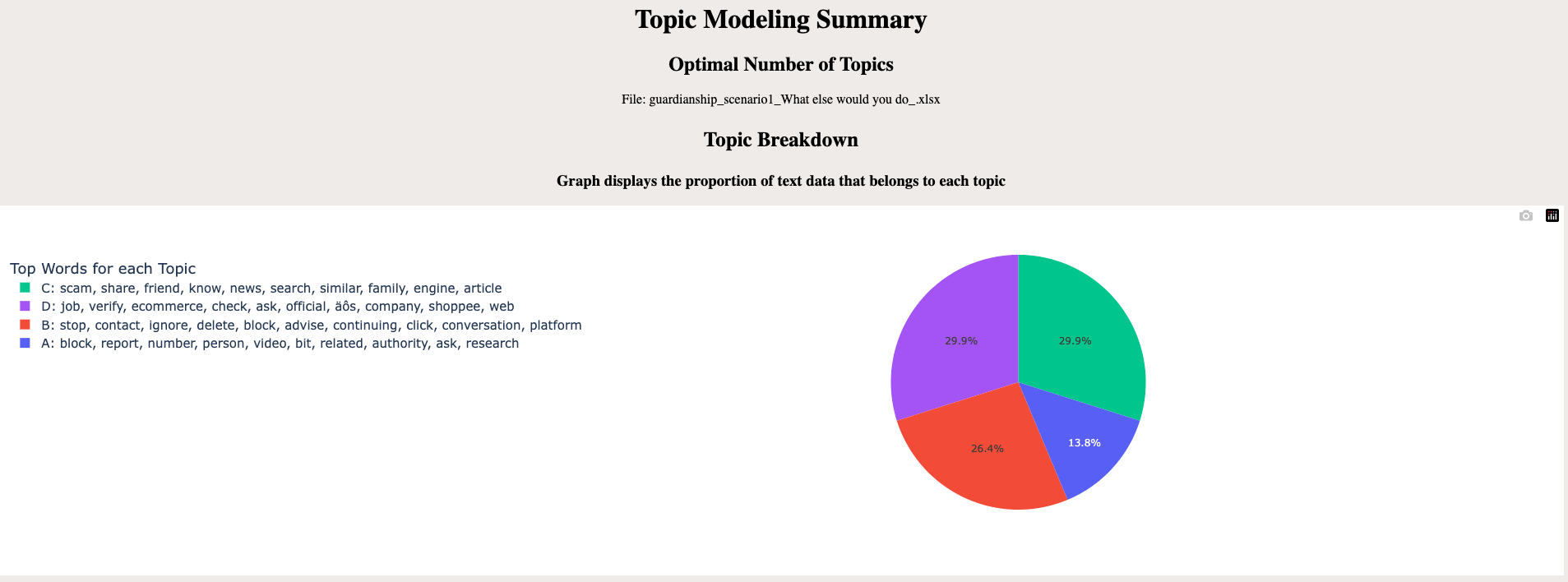
Topic Modeling Interpretation

*Pie Chart*

In the top half of the dashboard, users will see a pie chart containing a breakdown of the optimal number of topics detected based on the text input. Hovering over the pie chart will give an exact count of the number of text documents belonging to each topic.

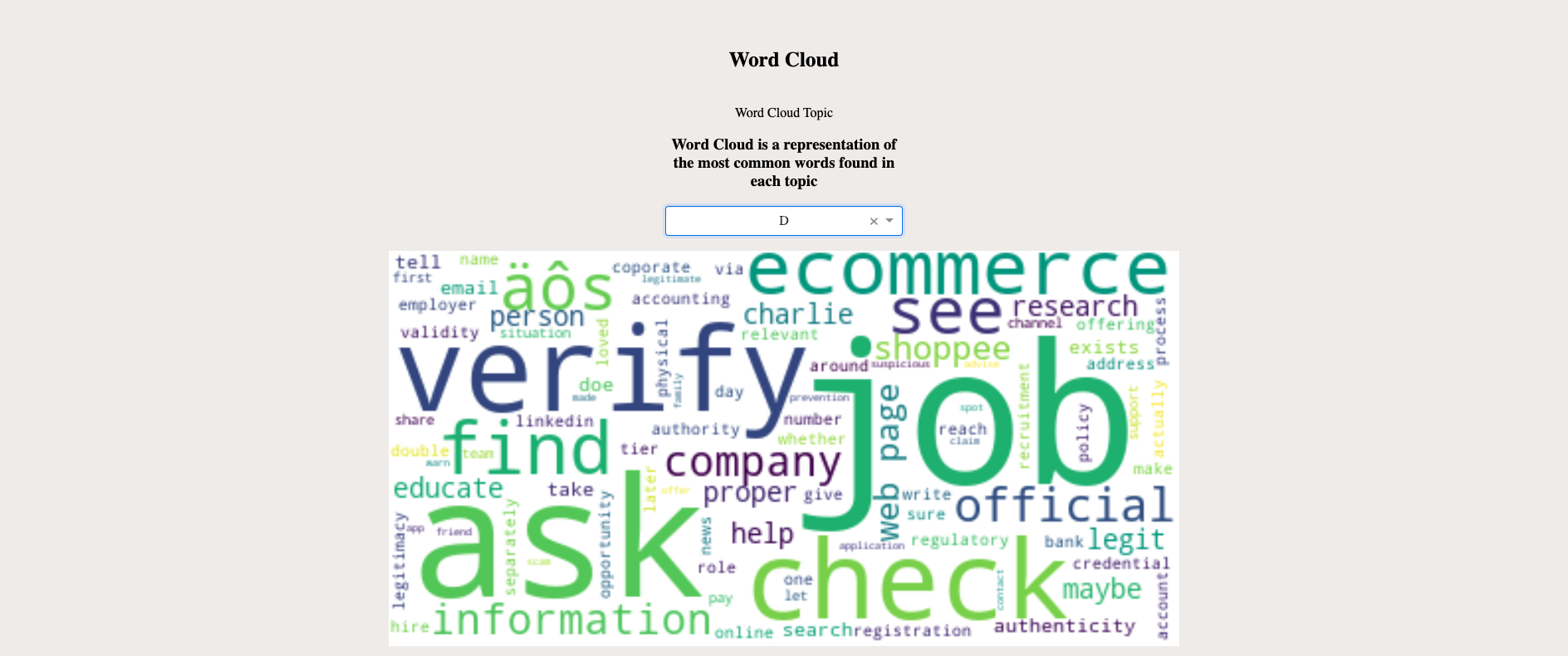
*Legend*

Top Words for each Topic: Represents the words which are most representative of each topic and are likely to describe what each topic represents



*Word Cloud*

At the bottom of the dashboard, users will be able to see the word cloud for each topic to get a visual representation of which words are the most common in each topic. Users will be able to use the filter to choose the topic they are interested in, which corresponds to the topics found in the pie chart



*Documents File*

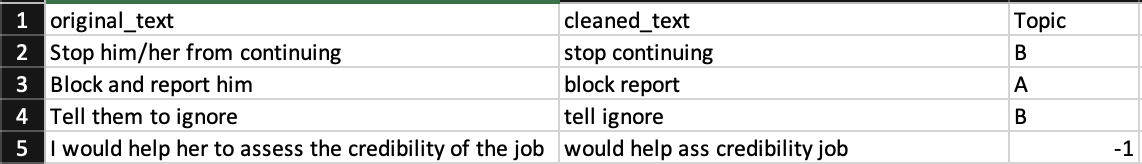
To get a better understanding of which topic each particular text belongs to, users can look at the output generated in the output/topic\_modeling folder.

| Input File Name | Output File names |
| --- | --- |
| “September\_Survey\_2022-10-15.xlsx” | “September\_Survey\_2022-10-15\_documents.csv”  “September\_Survey\_2022-10-15\_chart.csv”  “September\_Survey\_2022-10-15\_wordcloud.csv” |

Users can look at the documents csv to get a complete view of which text belongs to which topic. In this instance, users can look at the “September\_Survey\_2022-10-15\_documents.csv” file.

Eg: Rows 2 & 4 belong to topic **B** while row 3 belongs to topic **A**

Topic **-1**: Implies that this text data does not nicely fit into any of the other topics and may need further intervention to more accurately understand them

****

## **4.2 Dashboard: Text Summarisation**

This tool serves to summarize long text.

Input File Requirement

* File Format: .txt
* General Guideline for Minimum number of words: 200
* Note that having new lines in the text file is okay. Just refrain from having emojis.
* Please input the specific content that you wish to summarize. For example, if you want to summarize the answer to one specific interview question, please input only that portion with the answer to that specific interview question.

Procedures

1. Select “Text Summarization” in the dropdown.
2. Ensure that the File Explorer is displaying “All Files” (so that you can see the text files)
3. Select the input text file that meets the above input file requirement.
4. Click “Confirm”.
5. While waiting, look at the Terminal portion to see the progress.
6. The output summary text file can be found in the assets/output/summarization folder. (Refer to Section 2A of this guide if unsure)

## **4.3 Dashboard: Sentiment Analysis**

This tool serves to analyze the sentiments and aspects across the topics found in the text data provided.

Input File Requirement

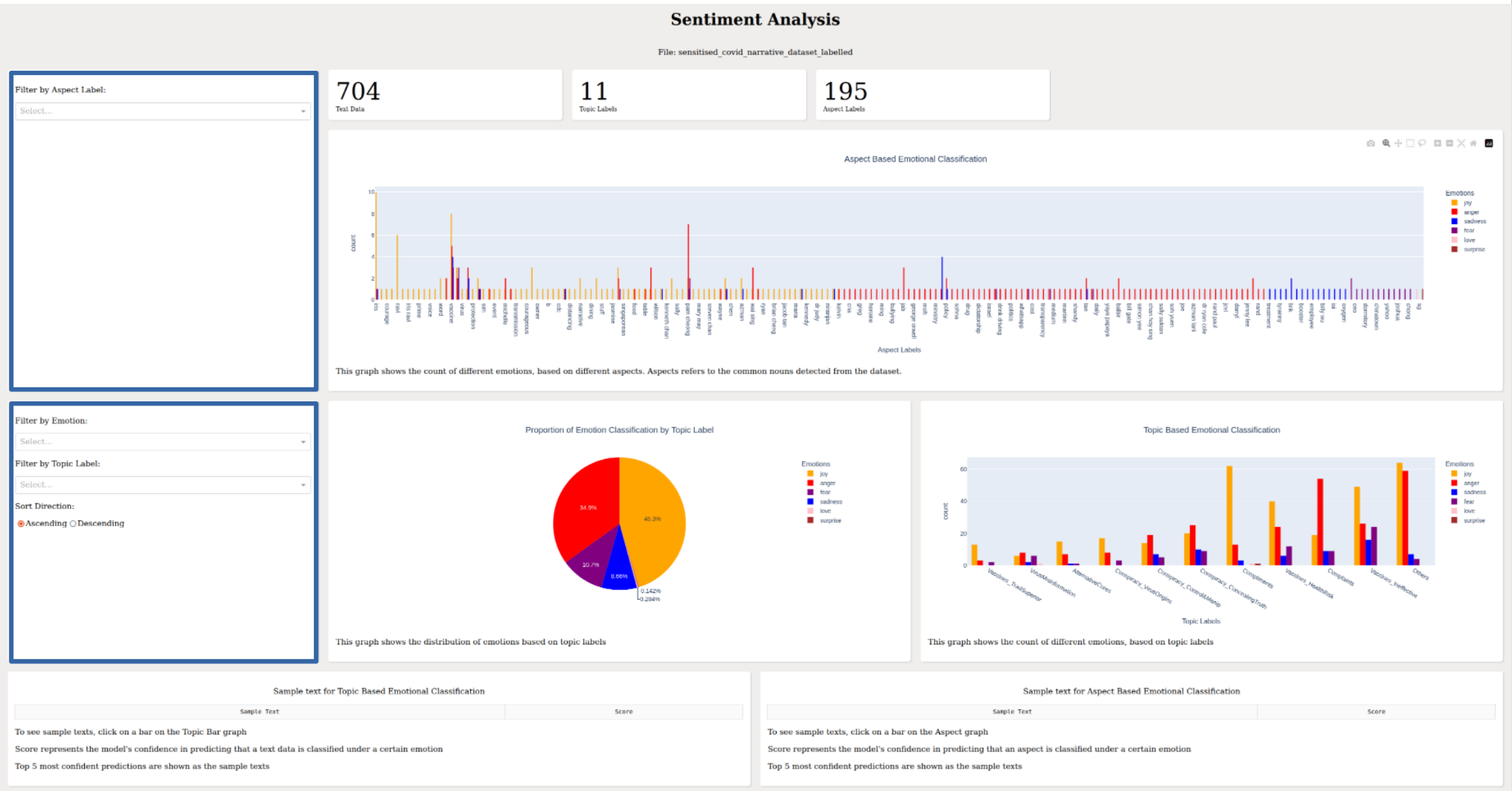
* File Format: .xlsx
* Columns required: “text”, “label”

Procedures

1. Select “Sentiment Analysis” in the dropdown.
2. Select the input file that meets the above input file requirement.
3. Click “Confirm”.
4. While waiting, look at the Terminal portion to see the progress.
5. Press “Ctrl” and click the link in the Terminal simultaneously to open the Dashboard in the browser.

Dashboard Interpretation

The sentiment analysis dashboard will look similar to the picture attached below. Note that the boxed up portions on the left of the dashboard are for users to customize the results they see.



*Top Input Box*

The top input box allows users to filter the data for the aspect based bar graph.

1. Filter by aspect label: Select the aspects that you want to analyze in the bar chart. (Do not select any aspects if you want to see all aspects in the bar chart)

*Second Input Box*

The bottom input box allows users to filter the data for the Topic based bar graph and/or pie chart.

1. Filter by Emotion: Select the emotion that you want to analyze in the bar chart
2. Filter by Topic Label: Select the topic label that you want to analyze. NOTE: This filter will change all 3 graphs accordingly to the input in this filter
3. Sort Direction: Rank the topic labels in the Topic Based bar Chart either in ascending or descending order

*Descriptive Analytics*

Descriptive Analytics gives the user a rough idea on how the dataset looks like

1. Text Data: Number of text data observations in the dataset
2. Topic Labels: Number of unique Topic labels given in the dataset
3. Aspec Labels: Number of unique Aspect labels identified by our classifier

*Aspect Based Bar Chart*

This graph shows the count of each emotion for every aspect label

*Pie Chart*

This graph shows the proportion of emotion classification

*Topic Based Bar Chart*

This graph shows the count of each emotion per topic label

*Sample texts*

To view the sample texts, click on a bar in either the Topic Based Bar Graph or the Aspect Based Bar graph. The top 5 most confident predictions of an emotion made by our classifier will be shown.

## **4.4 Dashboard: Social Media**

This tool serves to analyze social media data collected over time. Note that there are 2 main types of social media content - Posts and Comments that can be used with this tool. The procedures of using this tool for the 2 types of content will be the same. The only key difference to note is the Input File Requirement.

Input File Requirement

| **Posts** | **Comments** |
| --- | --- |
| * File Format: .xlsx * Columns required   + “content\_type” = ‘Post’   + “text”   + “time”   + “label” | * File Format: .xlsx * Columns required   + “content\_type” = ‘Comment’   + “text”   + "time”   + “label”   + “likes”   + “comments” |

Procedures

1. Select “Social Media Analysis” in the dropdown.
2. Select the input file that meets the correct input file requirement.
3. Click “Confirm”.
4. While waiting, look at the Terminal portion to see the progress.
5. Press “Ctrl” and click the link in the Terminal to open the Dashboard in the browser.

Dashboard Interpretation

The social media dashboard will look similar to the picture attached below. Note that the boxed up portions on the left of the dashboard are for users to customize the results they see.

*First Input Box*

The first input box allows users to filter the data seen in Graphs A and B.

1. Topic Labels: select the topics you wish to filter for by clicking options in the dropdown (do not select any if you wish to see all topics in graphs)
2. Aggregation Period: select from Yearly, Monthly and Daily to see the trend across different years, months and days respectively
3. Reaction Type: select whether you wish to observe the trends for Likes or Comments (this is only available when user uploads Post data)
4. Date Range: select the dates you are interested to look at, alternatively you can use the zoom function of the graph to observe the graph for the date period of interest

*Second Input Box*

The second input box allows users to adjust their rules for identifying trending topics. Note that if you wish to use only one of the metric as your rule, set the other metric to a 0 value.

1. Metric 1: Increase in Number of Posts/Comments under that Topic from previous month

This metric looks at the absolute increase in the number of Posts under that topic from one month to the next. Input a value above 0.

1. Metric 2: Percentage Increase in Proportion of Posts/Comments under that Topic from the previous month

This metric looks at the relative popularity of a topic as compared to all topics. Input a value between 0 and 100.

*Output A*

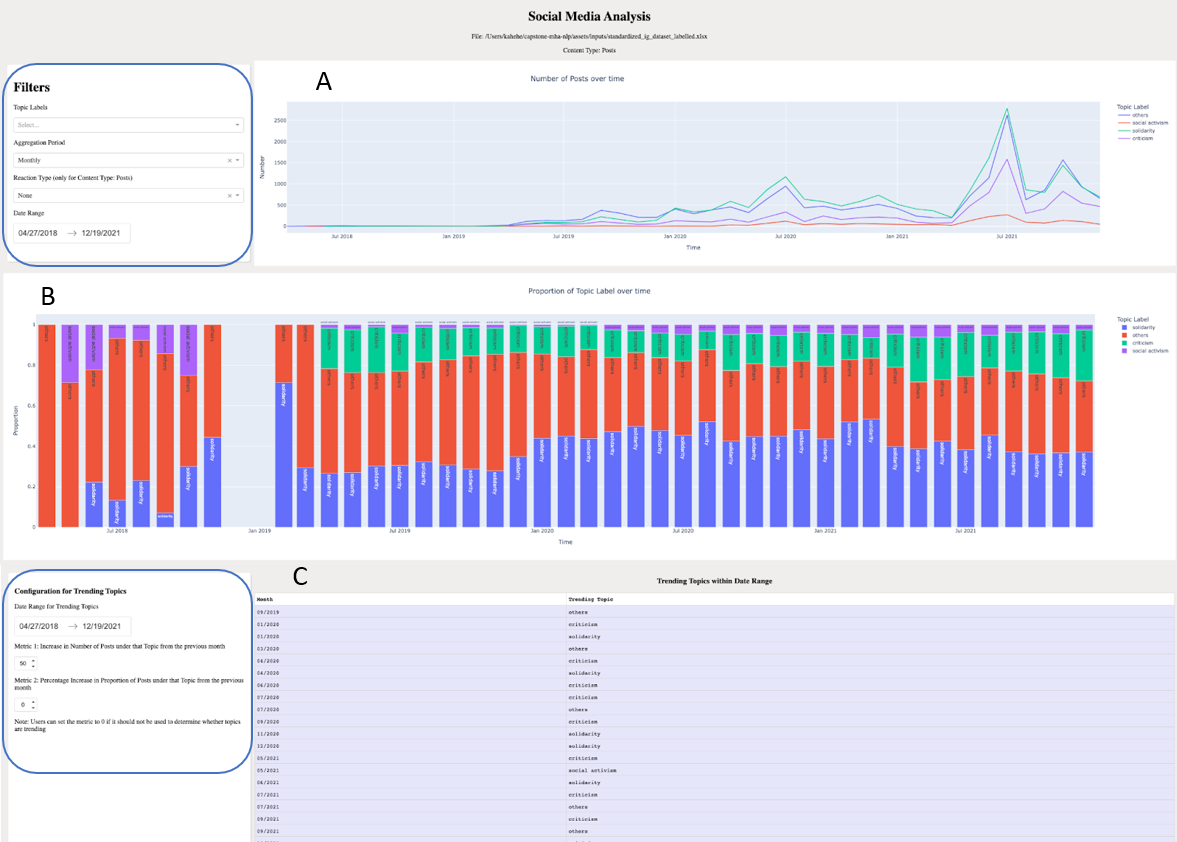
This graph shows the absolute number of posts/comments under each topic over time.

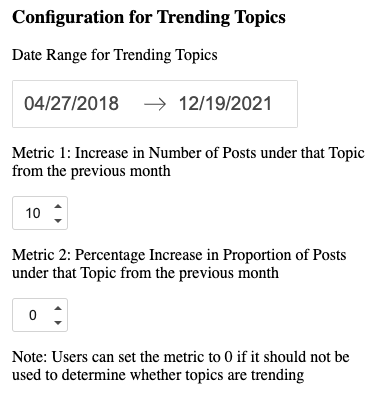
*Output B*

This graph shows the proportion of the various topics over time.

*Output C*

The table shows the topics trending on which month based on the metric rules.





**TO DO**

**Need explain everything!! Eg the functionality of each component, and what it is supposed to tell.**

**User manual should be v clear about where to select files etc**

* **User manual need include the deactivate**
* **Inform users how to look at old outputs!**
* **Note that the old output files should NOT be deleted. Esp if the output files may be referred to again. Eg output files of sentiment analysis shld not be deleted if they dw to run absa again.**
* **Rename the repository (dont put MHA) → make the repository public**

**Topic Modeling**

* **Need to explain the dashboard → can write a paragraph to explain what does topic modeling do,what is wordcloud and what do the topic 0,topic 1 mean. Users are confused what is 0 and 1 etc eg topic 0is first topic, topic 1 is second topic etc**

**Social Media**

**Sentiment Analysis**

* **Can state what are aspect labels and what is topic label.**
* **Sample text: can state that the top texts with highest score are chosen to be displayed**
* **Sample text: Explain what the score means**
* **Sample text: Let user know how to use the sample text → must click on the bar**
* **Sample Text: State where the sample text came from, eg on click on certain emotions/aspects in the graphs above**
* **User Manual: Let user not to delete the output → can view the dashboard again**
* **User Manual: Recommend user not to change the output files**
* **User Manual: Recommend user to name input files differently**

**Text Summarization**

* **User Manual: Steps on where to input file and where to find output file shld be clearly written in manual**
* **Why only the top part of input text is summarised? Bottom part not captured..?**

**Diagnostic Tool**

* **Screenshot the runtime ribbon HAHAHA**
* **Code block 4 maybe can put in screenshot → since user need do sth**
* **User manual: Maybe can be less technical (don’t put BERT and everything better), more layman terms**
  + **Eg just say it compares the original model with the new model and outputs the differences. Give a tldr**
* **User manual: Maybe more point form?**

**General Guideline for User Manual**

* **Provide TLDR?**
* **More screenshots**
* **Less technical, use layman terms!**
* **Point forms**