**Natural Language Analysis of Build a Better Grinnell Survey Data**

**USER GUIDE**

Made by: Shreyas Agrawal, Jiayi Chen, David Dai, Hong Doan

Spring 2023

1. **Folder Navigation**

The “Landing” Folder

Use the link bellow to access the “Landing” folder:

<https://drive.google.com/drive/folders/1paTqH2P-YRqM01VJwjkkhXyCjFAxOXaY?usp=share_link>

You can find an Executive Summary, a final Slides Deck, a Technical Report, and a User Guide to all the Code Files in the “Landing” folder.

1. The “Datasets” folder.

The “Datasets” folder is located under the “Landing” folder.

In this folder, you will find the dataset used for all the analyses. Specifically, you can find the “formresponses.csv” file here. Later on, you may upload the “formresponses.csv” file during the initialization stage in any analysis. This file can also be updated in case survey responses are updated.

1. The “Individual Question Analysis” folder.

The “Individual Question Analysis” folder is located under the “Landing” folder.

The “Individual Question Analysis” folder includes the “Example Code” that you will need to analyze the word frequencies in responses to each survey question. A detailed explanation of how to navigate the Example Code for the “Individual Question Analysis” can be found in this document below.

While any output can be reproduced using Example Code files, some of the outputs are included in the “Individual Question Analysis” folder, such as the Visuals and the “Words before and after common words” outputs.

1. The “Issue Analysis” folder.

The “Issue Analysis” folder is located under the “Landing” folder.

The “Issue Analysis” folder includes the “Example Code” that you will need to analyze the features of respondent subsets who care about particular issues. A detailed explanation of how to navigate the “Example Code” for the “Issue Analysis” can be found in the document below.

Additionally, you can also find a list of Keywords for each Need as well as some outputs in the “Issue Analysis” folder. The “Subsets of Issues” folder presents the respondent subsets of each issue, and the “Visuals” folder presents some sample outputs that you can produce using the Example Code file.

1. The “Cross Survey Analysis” folder.

The “Cross Survey Analysis” folder is located under the “Landing” folder.

The “Issue Analysis” folder includes the “Example Code” that you will need to conduct the “Cross Survey Analysis”. A detailed explanation of how to navigate the “Example Code” for the “Cross Survey Analysis” can be found in the document below.

In this folder, you can also find some sample outputs that you can produce using the “Example Code” for the “Cross Survey Analysis”.

1. **Individual Question Analysis**

**Purpose:** Analyzing each question of the survey to see what the respondents are saying in their answers to the question, as well as their sentiment toward answering this question.

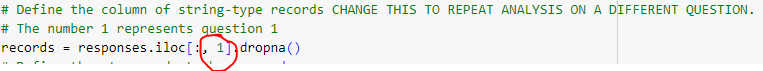
**Information already in the folder:**

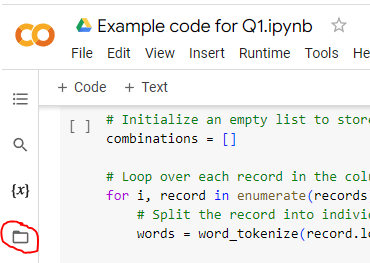
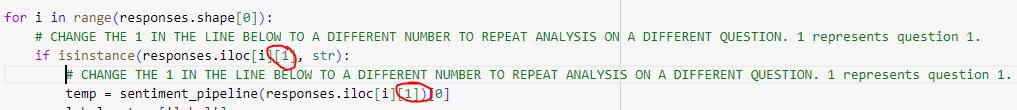
* Navigate to Final Deliverable > Individual Question Analysis & Deliverables > Visuals > Q1 visuals…(google doc)
  + As an example, there are visualizations like a word cloud and a bar chart showing what people are saying in their answers to this question.
  + Visuals highlighting words that appear after or before a certain word, in this case, “community”
  + Following these visuals, there are visualizations showing how people answered this question overall. Numbers on how many people were positive, negative, or neutral in their responses.
* Navigate to Final Deliverable > Individual Question Analysis & Deliverables > Words after common word > aftercommunity.xlsx (Excel file)
  + An Excel file listing which 2 words appear after a certain word, in this case after the word “community”
* This same information is also available for Q8

**Steps to repeat this analysis on a different question:**

We are providing a code template for Question 1: What are things that make you glad to live in Grinnell? Please follow the instructions below to make changes in the correct parts of the codes to repeat the analysis on a different question.

* Open the Code Template titled **“Example code for Q1.ipynb”** located at Final Deliverable > Individual Question Analysis & Deliverables
* Under the **Setting UP** section, run all 8 blocks, and we recommend making **no changes** in this section. After you run the first block only, the “Choose Files” button will become available as shown in the screenshot below. Click this button and select your .csv dataset. Finish running all the blocks in this section.
* After you finish running the Setting UP section, under the section titled “**Question 1: What are things that make you glad to live in Grinnell**”, **change** the first block of the sub-section titled “**Word Cloud, Word Frequency Count**” as follows:
  + In the very first block, there is a line of code that is selecting which column of the dataset to do the analysis on.



* + If this number is 1, it is going to do the analysis on Q1, which in the dataset is Column 2. For example, **Change** this 1 to 8 to repeat the analysis on Question 8.
  + After making this change, run this block and the remaining blocks in this subsection (next 2 blocks) as well, until you see the title “Words before and after a particular word”. This will produce a word cloud and a bar chart showing the most common words in responses to this particular question
  + Under the subsection titled “**Words before and after a particular word**”, make a change to the first block. Depending on which word you want to find words appearing after or before, change every occurrence of the word “community” in this entire block. To do this, select any occurrence of the word community in the code block, press Ctrl+Shift+L to select all occurrences of the word, and replace it with the word you want to repeat the analysis on. Run this block, and it will give you 2 visualizations showing the most common words before and after your specified word.
  + Repeat the previous step for the second block in this subsection which is looking at 2 words after the specified word. Run this block as well, and an Excel file will be produced that will look somewhat like the Excel file at Final Deliverable > Individual Question Analysis & Deliverables > Words after common word > aftercommunity.xlsx (Excel file). To find and download the Excel file, click on the folder icon on the top left bar
  + 
  + Next, under the subsection titled **“Conducting sentimental analysis on Question 1”,** in the very first block, there is a line of code that is selecting which column of the dataset to do the analysis on.
  + 
  + If this number is 1, it is going to do the analysis on Q1, which in the dataset is Column 2. For ex. **Change** is this 1 to be 8 to repeat analysis on Question 8.
  + After making this change, run this block and a visualization showing the sentiment of people responding to this question will be produced as well as which rows are in which sentiment category

1. **Issue Analysis**

**Purpose:** Identifying what the respondents with specific needs think of Grinnell’s strength.

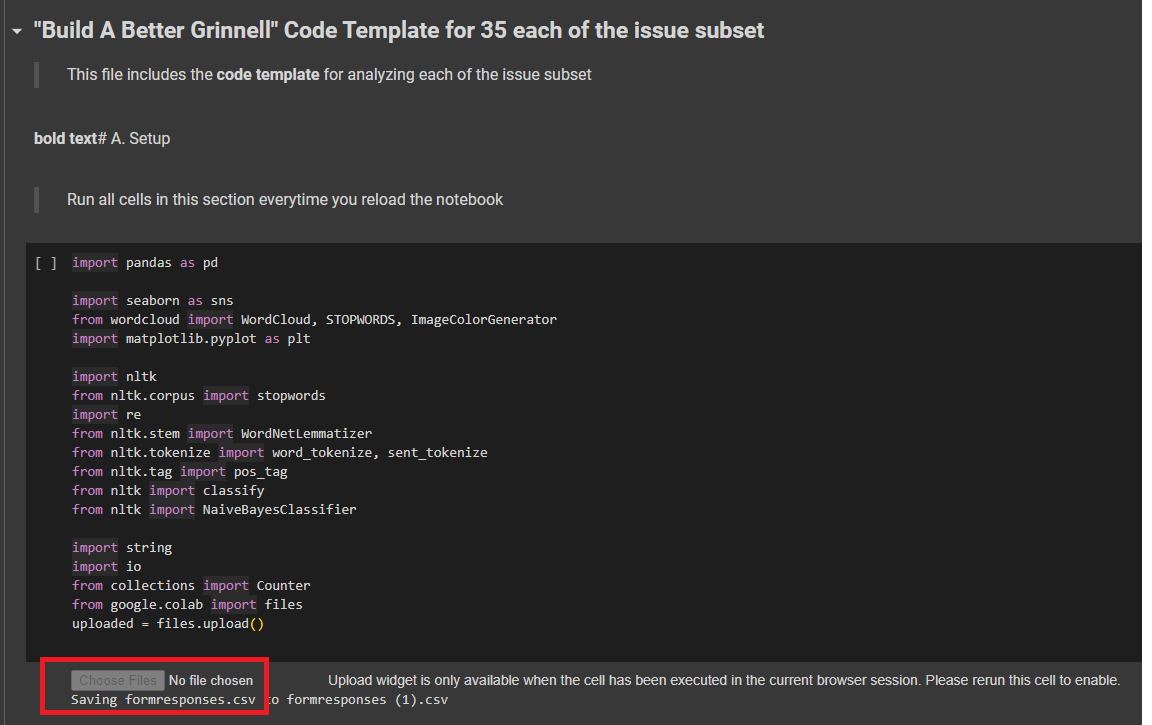
**Information already in the folder:**

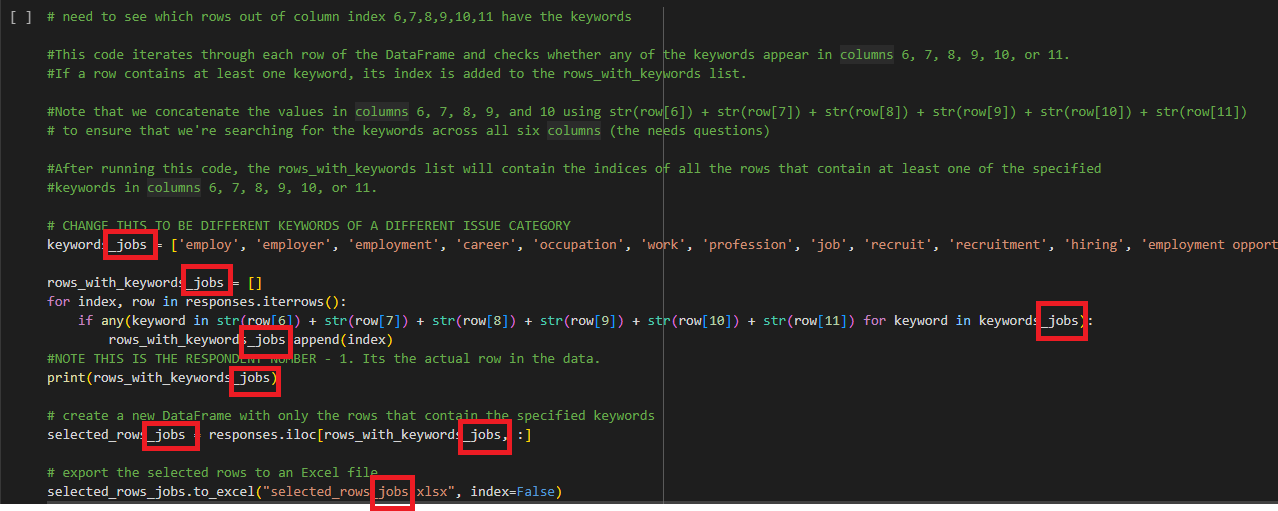
* Navigate to Final Deliverable > Issues Analysis & Deliverables

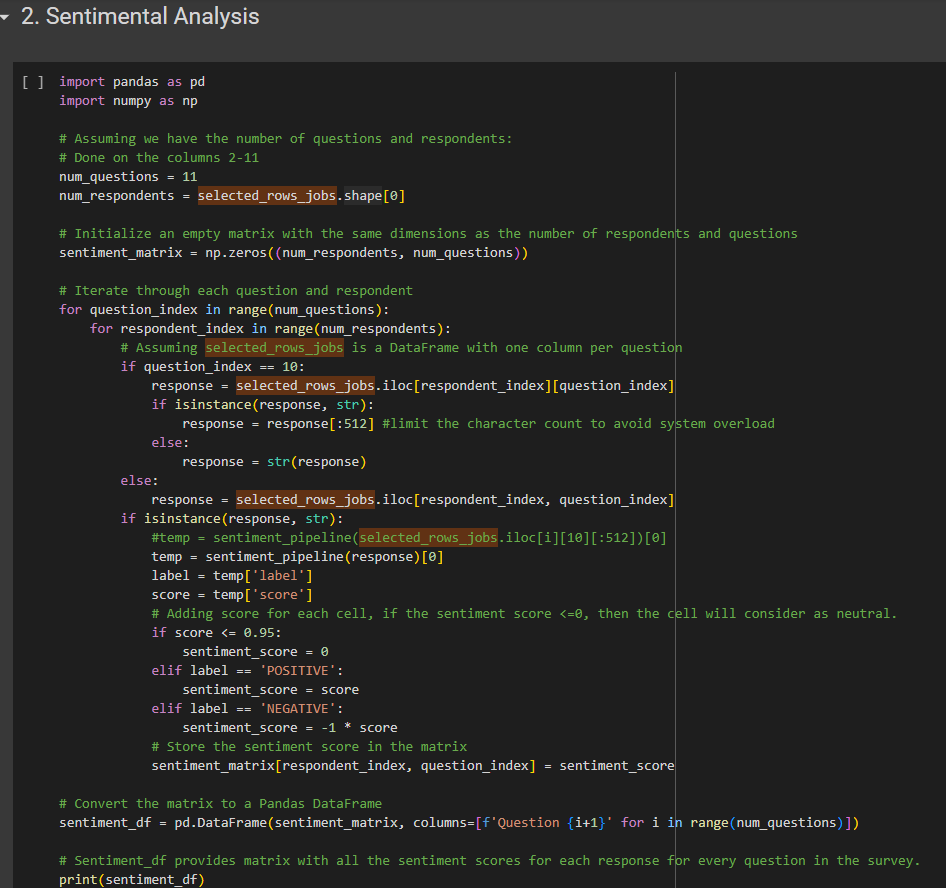
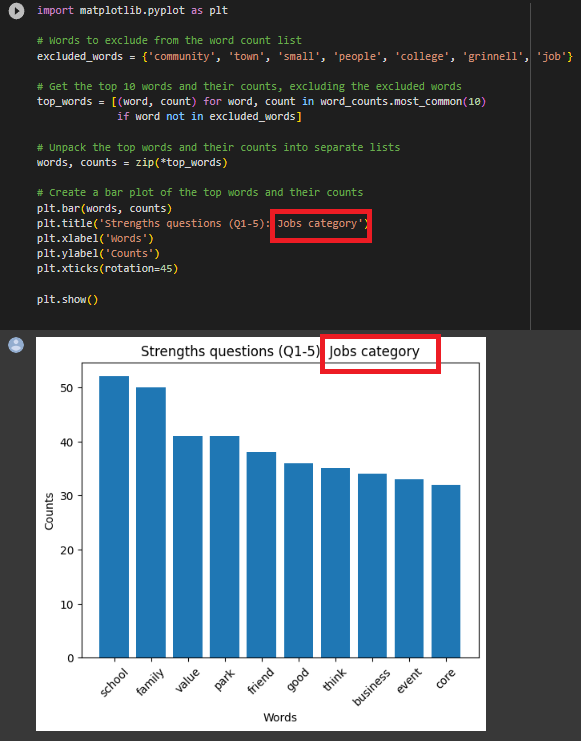
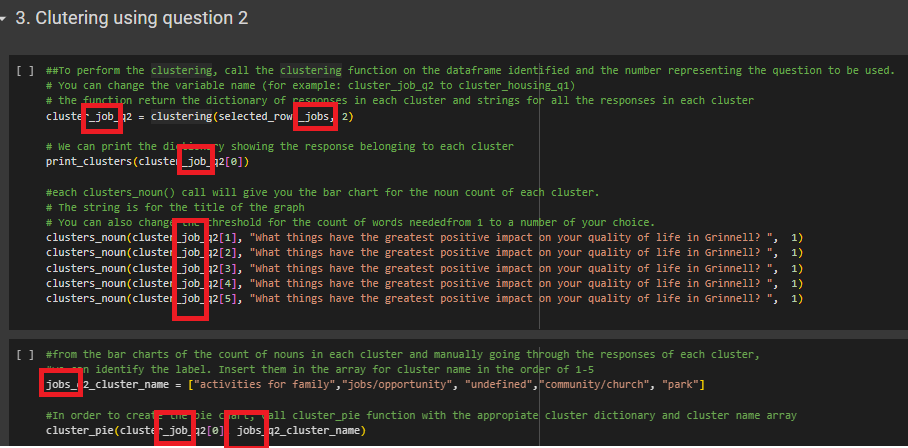
**Steps to repeat this analysis on a different question:**

We are providing a code template for Example code for Jobs issue.

Please follow the instructions below to make changes in the correct parts of the codes to repeat the analysis on a different question.

* Open the Code Template Jobs issue.ipynb atFinal Deliverable > Issue Analysis & Deliverables.
* Under the **A. Setup** section, run all 9 blocks, and we recommend making **no changes** in this section. After you run the first block only, the “Choose Files” button will become available as shown in the screenshot below. Click this button and select your .csv dataset. Finish running all the blocks in this section.
* After you finish running the **A. Setup** section, under the section titled “**B. Example - Issue: Jobs**”, **change** the first block of the sub-section titled “**1. Creating Subsets**” as follows:
  + In the very first block, there is a line of code that subsets the dataset based on the keywords. Replace the keywords to keywords of the group that you are interested in. For example, for housing, it would be “House, housing, apartment, rentals, affordable, rent, lower rent, more housing options, accessible housing, off-campus housing, eco-housing, housing variety.”
  + 
  + An optional step is to change “job” throughout the code to the category that you are interested in. eg replacing every occurrence of jobs with housing. Click Ctrl+Shift+L to select every instance of jobs and then just type housing instead.



* + Next, you can run the entire **“2. Sentimental Analysis”** section without any change. Similarly, you can also change “jobs” here. If you changed it in the previous section, you must change it here too.
  + 
  + You can then run the “What are the respondents of this issue category saying for Q1-5?” without any change.
  + You can change the title of the plot here:
  + 
  + For the “**3. Clustering using question 2**” section, if you changed “jobs” earlier, you should change it here accordingly.
  + 

1. **Overall Survey Analysis**

**Purpose:** Providing an overall survey sentimental analysis for each respondent.

**Information already in the folder:**

* Navigate to Final Deliverable > Across Survey Analysis & Deliverables > respondent\_sentiment(.xlsx) for the full excel file.
* Navigate to Final Deliverable > Across Survey Analysis & Deliverables > respondent\_sentiment\_visualization.png

**Steps to repeat this analysis on a different question:** None. May use Excel Filter to filter or sort the respondents based on their sentimental score and further categorize them into different sentimental categories (positive, neutral, and negative).