Sonnet Insurance Focus Groups: An Analysis using Word Clouds

March 23, 2020

Sonnet Insurance Focus Groups: Word Clouds

Alexei Marcilio March 19, 2020

Here we will create word clouds from the transcribed text of Sonnet Insurance employee focus groups. We use Python to create word clouds from this text. This is a useful way to quickly analyze focus group discussions as the text need only be transcribed. Some valuable insights can be gained before further in depth analysis begins.

Contents

1 Introduction

In February of 2020 Sonnet Insurance, in an effort to better understand employee satisfaction, hosted several employee focus groups. One group was held in Montreal and the other in Toronto. A separate focus group of team leaders was also conducted.

A word cloud is a pictorial display of a group of words depicted in various sizes. Words will appear bigger and bolder depending on the frequency they occur. Word Clouds are often used to display the frequency of words used by focus group participants. Focus groups collect data in an open-ended way in which participants are encouraged to give textual answers. Word clouds allow for a quantitative way to display this textual data.

This paper presents the analysis of the three focus groups conducted by Sonnet Insurance using word clouds, and describes the methods by which these results were created.

2 Analytic Techniques

Python is now the most popular programming language used in data science due to the many libraries that can be utilized for specific tasks. Here will will leverage the library "wordcloud" in order to create word clouds from text files.

The following methods were used:

1. Each of the google docs which contained transcribed text from the three focus groups were saved as text files after the questions were removed.
2. A function was created in Python which utilized wordcloud and take three parameters, including an input and output file.
3. Additional stopwords were added based on the preliminary results. Stopwords are common words, such as "the" and "and" which are excluded as they have no meaning.
4. Each dataset was loaded and a word cloud image (png file) was created for each focus group, one for Toronto, one for Montreal and one for Team Leaders.
5. A combined dataset and word cloud comprised of the Montreal and Toronto focus groups was created to assess overall sentiment.

2.1 Libraries

Let's import all the necessary libraries. STOPWORDS are the default stopwords - words which we do not want to include in the results.

[9]:
```
# Start with loading all necessary libraries
import matplotlib.pyplot as pPlot
import numpy as npy
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
from PIL import Image
```

2.2 Combine Files

We create a file for all employee discussion text by combining the Montreal and Toronto files.

[10]:
```
# We run a shell command to combine the files
!cat tor_focus.txt mon_focus.txt > emp_focus.txt
```

2.3 Import Files

We open each of the four files for reading. Let's also convert all the words to lowercase, so that every word is only counted once.

[11]:
```
# Open each of the three files
toro_data = open("tor_focus.txt", "r").read()
mont_data = open("mon_focus.txt", "r").read()
lead_data = open("tl_focus.txt", "r").read()
```

```
emp_data = open("emp_focus.txt", "r").read()
# Convert all files to lowercase
toro_data = toro_data.lower()
mont_data = mont_data.lower()
lead_data = lead_data.lower()
emp_data = emp_data.lower()
```

2.4 Additional Stopwords

We run through preliminary results and notice common words that should be excluded. We do that here by adding additional stopwords to our list.

[12]:
```
# Add additional stopwords
stop_words = ["people", "don", "time", "name", "thing", \
              "one", "feel", "call", "say","tl", "know", \
              "us", "will", "work","make", "say", "think" \
             "take", "go", "company", "someone", "day", "year", \
              "jen", "think", "take", "lot", "now", "makes", \
              "said", "going", "come", "things"] + list(STOPWORDS)
```

2.5 Python Function

We create a python function that will accept three parameters. The first two indicate the input file, which is the transcribed focus group text, and the second is the name of the output file, in case we need to create a hard copy of the image. The third function is a boolean value.

If True then the function will only print a png image to file, otherwise, if False, it will only display the image. Sometimes a hard copy of an image is necessary for rendering a PDF document from a LaTeX file for example.

[13]:
```
# Python function to display word clouds
def create_cloud(input_file, output_file, yes_to_print):

    wordcloud = WordCloud(max_words=70, background_color="white", \
                          width=1600, height=800,
                          stopwords=stop_words).generate(input_file)
    if not yes_to_print:
        pPlot.figure(figsize=(12, 8))
        pPlot.imshow(wordcloud, interpolation="bilinear")
        pPlot.axis("off")
        pPlot.show()
    else:
        pPlot.savefig(output_file)
        wordcloud.to_file(output_file)
```

2.6 Create Word Cloud Images

We run the function for each file creating a word cloud image for each of the three focus groups. These will be used for other applications.

[14]:
```
# Create the word clouds
create_cloud(toro_data, "torontoCloud.png", True)
create_cloud(mont_data, "montrealCloud.png", True)
create_cloud(emp_data, "employeeCloud.png", True)
create_cloud(lead_data, "leadersCloud.png", True)
```

```
<Figure size 432x288 with 0 Axes>
```

3 Analysing Word Clouds

Let's take a look at each of the images. We should remember that it can be a challenge to interpret word clouds because they emphasize the frequency of the words, not really their importance.

Word clouds will often not accurately show the content of the text if different words are used for the same meaning, for example "large", "huge", "giant", and so on will dilute the point being made. They will also not provide any context, so we may lose the meaning of certain words. Therefore due to these limitations word clouds are most suited for exploratory qualitative analysis.

3.1 Toronto Word Cloud

First let's take a look at the Toronto Focus Group word cloud.

[15]:
```
create_cloud(toro_data, "torontoCloud.png", False)
```

Toronto Focus Group Word Cloud

The first thing we notice about the Toronto Word Cloud is that the word good is the largest word. This means this was said the most often. That would seem like a good sign, as this can only be construed as a positive thing.

"Change" and "want" are the second and third largest words which would seem to indicate that things aren't all perfect. If change was brought up frequently in the focus group it probably indicates most people are not complacent and are willing to change the status quo.

3.2 Montreal Word Cloud

Second we take a look at the Montreal Focus Group word cloud.

[16]:
```
create_cloud(mont_data, "montrealCloud.png", False)
```

Montreal Focus Group Word Cloud

The first thing we notice about the Montreal Word Cloud is that the word "toronto" is front and center. It's interesting that Toronto was mentioned so often during the focus group. Perhaps there's some resentment towards the Toronto team from the Montreal employees.

The word "need" is used very often. It's interesting that the word "need" is used so frequently among the Montreal employees, in contrast to the Toronto employees who tended to use the word "want" frequently. In fact the word "need" is hardly used at all at the Toronto office.

Training is on their mind more at the Montreal location.

3.3 Employee Word Cloud

Now let's take a look at the Employee Focus Group word cloud. This is the word cloud created from the combined text of the Montreal and Toronto focus groups.

[17]:
```
create_cloud(emp_data, "employeeCloud.png", False)
```

Employee Focus Group Word Cloud

Here we can see that the word "good" is prominent and also the word "want". "Need" follows behind. This could be due to the fact the Toronto focus group included more people. Employees use the word "good" quite a bit, but are looking for change. We can see this with the frequency of words such as "want", "need" and "change".

3.4 Leader Word Cloud

Lastly we examine the Leader's Focus Group word cloud.

[18]:
```
create_cloud(lead_data, "leaderCloud.png", False)
```

Leader's Focus Group Word Cloud

It's interesting that the Leader's word cloud is very different from that of the employee's word cloud. "Need" is emphasized, which is similar to Montreal, but the word "see" is the most common word. It could be that the frequent use of the word "see" represents some frustration among leaders of the company who believe that employees don't see either what is happening, or the vision of the company leaders wish to present.

4 Conclusion

We have used analytic techniques to create word clouds for each of the Sonnet Insurance focus groups and one for a combined employee focus group. Word clouds are sometimes used as an initial step in analyzing focus group text as they can quickly provide a quantitative overview.

Although overall sentiment is often easy to interpret from a word cloud it is more difficult, or impossible to determine more than a high level take on the data. Domain knowledge can often help as employees who are familiar with the company might be able to better interpret these images.

We can see that at Sonnet Insurance there seems to be more employee satisfaction in Toronto than in Montreal, and that many employees are looking for something different as the words "change", "need" and "want" are the most frequently used words.