FRESH ARTS COMMUNITY SURVEY ANALYSIS

A Text and Sentiment Analysis on Open-Response Formatted Questions

Abstract

Understanding the sentiment of artists within the Houston, TX area utilizing free text responses to a survey distributed by the non-profit, Fresh Arts.

Team Gr8 (Team 8): Sara Foster, Micah Moreno, Amir Rastboud

Fresh Arts Community Survey Analysis
Team Gr8 (Team 8)

Contents

Fresh Arts Community Survey Analysis
Team Gr8 (Team 8)

Executive Summary

History of Fresh Arts

Fresh Arts is a non-profit organization aimed at helping advancing the careers of local artists by giving them access to localized resources and career development opportunities regardless of level of experience. The organization was created after the merger of Spacetaker and Fresh Arts Coalition to form the current group, Fresh Arts.

Business Problem

In quarter 4 of 2021 Fresh Arts deployed a survey to local artists in Houston, TX utilizing the Survey Monkey platform. The aim of the survey was to gather information regarding artists' perception of the local Houston arts community. The goal of the survey was to help the group set goals for the next year, and draft a long term plan for the next five to twenty years. This survey asked three open ended questions:

1. What do you feel is your biggest barrier to thriving as an artist?
2. What are arts organizations getting right?
3. What are arts organizations getting wrong?

Over 200 people responded to each of these three questions with fairly robust answers, totaling at 600 free form responses. Survey Monkey generated some insights for Fresh Arts, but was unable to do any sort of sentiment analysis. The most survey monkey is able to provide is basic text analysis of the responses and generate word clouds. This is misleading as some words are repetitive and should be excluded from analysis among other reasons.

The Solution

The goal of our team was to conduct sentiment analysis on these free form responses utilizing AFinn and Google Sentiment analysis. By comparing the two forms of possible sentiment analysis it's possible to understand which type of package should be utilized by Fresh Arts in the future. The next step of this project was to compare how the results align with expectations. By doing this Fresh Arts can better understand the areas in which they need to improve, and what they are doing well in. However, creating a word cloud through the utilization of text analysis helps to visualize the most frequently used terms by respondents. Text analysis enables one to have the ability to exclude certain words which is insightful as most of the survey takers utilized the words art, arts, and artists frequently which is reflected in the Survey Monkey generated word cloud. By eliminating the noise, a clearer picture comes into focus. This project helps Fresh Arts to create a strategic path forward into the future.

Fresh Arts Community Survey Analysis
Team Gr8 (Team 8)

Team Roles

All team members consulted on strategy, analyses, and conclusions throughout the life of the project. However, specific roles were played by each of the team members.

Amir Rastboud

Amir conducted the text analysis through the utilization of Python. This generated word clouds that indicated the different topics present in each question. A final word cloud was created for all three questions. Mr. Rastboud also did AFinn sentiment analysis by creating a script with the AFinn package in Python. In addition to the code, the AFinn sentiment analysis was done manually via an online platform.

Sara Foster

Mrs. Foster built a Python script utilizing Google sentiment analysis to denote a sentiment score and magnitude associated with each free form response. She was also responsible for helping to write content for the slide deck and this holistic report.

Micah Moreno

Micah as an employee of Fresh Arts was responsible for gathering, cleaning, and sanitizing the data to protect the privacy of the artists. She also helped to print the overall results from AFinn and Google sentiment to a csv file for future consumption by Fresh Arts. In addition, she helped to compile the slides and communicate results back to the non-profit.

Sentiment Analysis

The code utilized for the sentiment analysis can be found in the files labeled as **FreshArts_SentimentAnalysis_FINALQ1, FreshArts_SentimentAnalysis_FINALQ2**, and **FreshArts_SentimentAnalysis_FINALQ3**. All three questions were analyzed individually using the AFinn and Google sentiment packages.

AFINN

Methodology

The AFinn package manually scores 2,000 words that were categorized as expressing emotion in English. These words were given certain weights that range from -5 to 5, from negative to positive.

There were two methodologies utilized for AFinn. The first was done by manually scoring each response using a website that evaluated each individual response. This can be seen in **Figure 1** below.
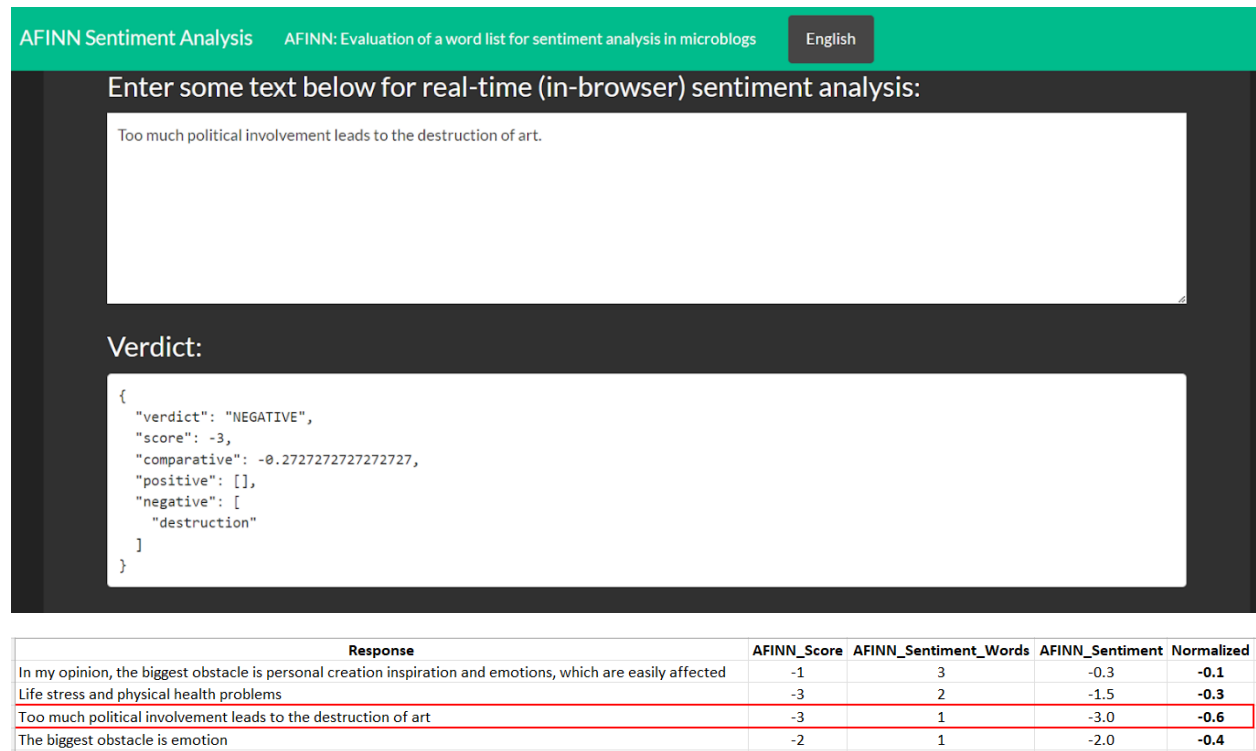

Response	AFINN_Score	AFINN_Sentiment_Words	AFINN_Sentiment	Normalized
In my opinion, the biggest obstacle is personal creation inspiration and emotions, which are easily affected	-1	3	-0.3	-0.1
Life stress and physical health problems	-3	2	-1.5	-0.3
Too much political involvement leads to the destruction of art	-3	1	-3.0	-0.6
The biggest obstacle is emotion	-2	1	-2.0	-0.4

Figure 1: Manual AFinn Sentiment Analysis Using a Website.

Each response was calculated to have a normalized score that was inputted into an excel document by hand as a normalized score. These scores were then averaged for each question utilizing python. The team wanted to do a comparison of a practical website interface that Fresh Arts employees could use if an employee wasn't familiar with coding. The other method conducted for AFinn was with the AFinn package in Python and creating a script as seen in the attached code. Each response was scored before an average was taken for all responses to each question. The scores generated by the manual input into the website and running a customized script were more or less the same as seen in **Figure 2**.

```
Response:   Too much political involvement leads to the destruction of art
AFINN Sentiment Python  : -0.600
AFINN Sentiment Manual  : -0.600
```

Figure 2: Comparison of AFinn Done Manually vs with Customized Python Script.

Results

The AFinn manual and python script produced interesting results as seen below in **Figure 3**.

```
[260 rows x 9 columns]
**********Q1 AVERAGE RESULTS******
------AFINN RESULTS------
AFINN Sentiment Manual: -0.212
AFINN Sentiment Python: 0.055
```

```
[206 rows x 9 columns]
**********Q2 AVERAGE RESULTS******
------AFINN RESULTS------
AFINN Sentiment Manual: 0.97
AFINN Sentiment Python: 0.34
```

```
[204 rows x 9 columns]
**********Q3 AVERAGE RESULTS******
------AFINN RESULTS------
AFINN Sentiment Manual: -0.521
AFINN Sentiment Python: 0.148
```

Figure 3: Results of AFinn Sentiment Analysis for Question 1, 2, and 3.

Google

Methodology

The Google package for sentiment analysis generates a sentiment score and magnitude of sentiment, taking into account the context and entirety of the document. This is vastly different from AFinn which looks at individual words. The sentiment score indicates the overall emotion of a document, whether positive or negative. The magnitude of sentiment denotes how much emotional content exists within the document.

The excel documents for each response were read into separate data frames before each free text response was given a sentiment score and magnitude. These scores were generated from a function within the code that accesses Google's google.cloud package utilizing login credentials. Examples of the results for each individual response can be seen in **Figure 4**.

```
Response:  Life stress and physical health problems
AFINN Sentiment Python  : -0.600
AFINN Sentiment Manual  : -0.300
Google Sentiment  : -0.200
Google Magnitude  :  0.200
====================
Response:  Too much political involvement leads to the destruction of art
AFINN Sentiment Python  : -0.600
AFINN Sentiment Manual  : -0.600
Google Sentiment  : -0.800
Google Magnitude  :  0.800
====================
Response:  The biggest obstacle is emotion
AFINN Sentiment Python  : -0.400
AFINN Sentiment Manual  : -0.400
Google Sentiment  : -0.600
Google Magnitude  :  0.600
====================
```

Figure 4: Google and AFinn Sentiment Scores for Individual Responses.

Results

Each response was averaged for the three questions to generate individualized sentiment scores and magnitudes seen in **Figure 5**.

```
[260 rows x 9 columns]
**********Q1 AVERAGE RESULTS******
------AFINN RESULTS------
AFINN Sentiment Manual: -0.212
AFINN Sentiment Python: 0.055
------GOOGLE RESULTS------
Google Sentiment Score: -0.158
Google Magnitude Score: 0.689
```

```
[206 rows x 9 columns]
**********Q2 AVERAGE RESULTS******
------AFINN RESULTS------
AFINN Sentiment Manual: 0.97
AFINN Sentiment Python: 0.34
------GOOGLE RESULTS------
Google Sentiment Score: 0.317
Google Magnitude Score: 0.541
```

```
[204 rows x 9 columns]
*********Q3 AVERAGE RESULTS******
------AFINN RESULTS------
AFINN Sentiment Manual: -0.521
AFINN Sentiment Python: 0.148
------GOOGLE RESULTS------
Google Sentiment Score: -0.339
Google Magnitude Score: 0.654
```

Figure 5: Google Sentiment Results for Questions 1, 2, and 3.

Overall Results

Both AFinn and Google Sentiment Analysis produced similar and expected results. Questions 1 and 3 had negative sentiment on average, possibly due to the usage of words such as "barrier" and "wrong. Question 2 had a mostly positive sentiment in the free text responses. This may be a result of the positive words used in the question such as "right". The takeaway is that the framing of a question may have undue influence on how a person responds to the prompt.

Text Analysis

Text analysis was done for all 600 free form text responses to all three questions in addition to text analysis for each individual question. Code can be found in the files labeled as **Text** and **Text Analysis**. The words such as art, artist, and Houston were excluded from this as they created too much noise. These were three highly used terms because this survey was about art and artists within the Houston area. By removing these, a clearer picture forms. Five general topics were identified for all three questions with the words comprising topic 4 being the most frequent as seen in **Figure 6**.

Figure 6: Topic #4 Word Cloud for Responses to Questions 1-3.

Analyzing the words in the response to each individual question in **Figures 7-9** give a clearer picture of what matters to the population of artists in Houston.

Figure 7: Word Cloud for Question 1.

Figure 8: Word Cloud for Question 2.

Figure 9: Word Cloud for Question 3.

Already examining these word clouds gives clearer directive than the ones generated by Survey Monkey. Artists seem to value the resources, opportunities, and community that they can find at Fresh Arts. These are three aspects that the non-profit can continue to focus on in the upcoming years.

Conclusion

The first thing to mention is to be wary of survey bias whenever looking at these results. Most can be swayed by the wording of a question or even by their background unintentionally. There is the possibility of an under sampled population, however this would require tying back to the individualized results to see if demographics are a factor in people's sentiment.

Another note is that the text analysis adds another layer of illumination to the sentiment analysis. Understanding how people feel is only part of the problem if one does not know what are their drivers. By doing text and sentiment analysis it is clear that artists want and need a strong sense of community and organizational support. This may be through the resources provided by Fresh Arts or the work opportunities that are generated by the non-profit organization.

The final takeaway is that the programs implemented by Fresh Arts are important to people. They need these programs to derive joy and to feel as if they are succeeding in their artistic careers. Overall, this analysis was much more fruitful and insightful over the generalized analysis done by Survey Monkey.