Full name and semester: Durga Sindhu Animalla, Spring 2023

Title: “Analyzing Cultural Perspectives on the Tragedy of Harambe's Death through Sentiment Analysis and Topic Modeling” (3822 words)

1.Introduction

In the digital age, public opinions and reactions to events unfold quickly, encompassing a wide range of perspectives. A significant topic of public discourse is animal welfare and high-profile incidents involving animals in captivity, which often evoke intense emotions and debates. A prominent example is the 2016 death of Harambe, a gorilla which got shot and killed at the Cincinnati Zoo to protect a child who had fallen into its enclosure, sparked global conversations on animal rights, zoo safety, and parent’s responsibility. This event has been a topic of debate as some people view it as harmless fun and some view it as disrespectful mockery and stands as a suitable case for examining animal-human interactions as people have exhibited different reactions over the social media platforms.

This paper aims to employ sentiment analysis as a methodological tool to investigate the emotional resonance of people towards Harambe's death. Additionally, it seeks to uncover the themes and topics present in these comments through topic modeling, which can help identify the key subjects and aspects discussed by the public in relation to this event. Previous studies have investigated the emotional responses using sentimental analysis(Hutto & Gilbert, 2014) and have focused on situations where an animal poses an immediate threat to human safety, such as when an animal becomes aggressive or escapes, necessitating lethal action to safeguard human life (Carter et al., 2017). Additionally, research has been made to examine the concept of animal rights from a psychological standpoint, assuming that animals exhibit cognitive and emotional capabilities similar to humans and, as a result, should be afforded the same moral considerations (Monsó, S., Benz-Schwarzburg, J., & Bremhorst, A., 2018). In light of these findings, employing sentiment analysis to examine the public's reactions to Harambe's death offers valuable insights on cultural attitudes towards animals, ultimately contributing to a better understanding of society's stance on animal welfare, human-animal interactions, and the moral implications of such events.

2.Research Question

The purpose of this research is to investigate and address the following questions:

- How did the public react to the killing of a gorilla at Cincinnati Zoo after a child fell into its enclosure, and what were the dominant sentiments expressed by the users on the social media platform Reddit during the period of the incident?
- What are the predominant themes and topics that can be identified in the public discussions regarding this incident?

3. Methods

3.1 Data:

The primary aim of this paper is to tackle the research questions that were mentioned earlier. The dataset provided pertains to the comments on the post made on Reddit in the year 2016, specifically when the incident occurred.

<u>Data Selection Criteria:</u>

- Choosing the subreddit post to scrape the comments: Utilizing the PRAW module, a search is carried out through the Reddit API using the keyword "Cincinnati Zoo," where the incident occurred, and the results are sorted by the number of comments in decreasing order. From the retrieved results, the top-ranked subreddit post with the highest number of comments is chosen for extracting the comments indicating high level of user engagement and diverse range of perspectives for our study.
- Data Filtering and Selection: Initially, the dataset contained 5837 comments from the subreddit post. The data underwent filtering which involved the removal of deleted comments and selecting comments solely posted in the year 2016. Following these procedures, the total number of comments was narrowed down to 2,201. This data was filtered to ensure the uniqueness of users by only considering each user's first comment and discarding subsequent replies. This approach maintained the distinctiveness of each author and captured a broader range of audience perspectives, rather than focusing on a single author's input.
- Libraries and modules: Comments were scraped from Reddit using the 'praw' module, and the 'nltk' library was employed to obtain the essential tools for working with natural language data, including sentiment analysis and text processing. A data frame was created and manipulated using the 'pandas' library to store and handle the processed comments data. The text data was preprocessed by utilizing the 'stopwords' module within the NLTK library, which allowed for the removal of common words to decrease the noise in the data and achieve more meaningful patterns.
- Data Attributes: With the scraping of data from Reddit after the preprocessing, the comments were saved in a .csv file where the dataset consisted of Author of the comment, Date of the comment, Comment itself and Subreddit from which the comment was extracted.
- Overall, In the next steps, Sentimental Analysis and Topic modeling will provide us with a better understanding of the opinions, attitudes, and feelings of users towards the incident, which can be used to improve our understanding of the incident.

4. Analysis:

- Before extracting data from Reddit, a thorough manual analysis of the data was conducted to assess its structure. The data was found to consist of unstructured comments containing irrelevant information, misspelled words, slang, emojis, and multiple responses from the same author. Consequently, there was a need for data cleaning and preprocessing to ensure the data was usable for analysis. The PRAW module in Python was used to identify the top subreddits by searching for the keyword "Cincinnati Zoo" sorted by the highest number of comments. The selected subreddit post will undergo data cleaning and preprocessing steps. Here is an example of the raw data that is provided prior to processing:

Raw data before preprocessing:

username	Date	Comment	Subreddit
argentgrove	5/29/2016 1:15	Sad for zoos and animals, didn't a man try to commit suicide by a lion recently and the lion had to be killed too?	news

Table.4.0.1 Example of Raw data before preprocessing

- After selecting the post, the comments were extracted using the Reddit API. To ensure data accuracy, comments from duplicate authors and those not from 2016 were filtered out, along with comments containing shareable links. Preprocessing was then done on the comments by eliminating URLs, punctuation, and stopwords and converting them to lowercase. As a result, the data was saved in two separate CSV files, one for the raw comments and one for the processed comments. Sentimental_analysis_rawdata.csv consists of the raw comments, while the Sentimental_analysis_processeddata.csv consists of the processed and filtered data. The processed data after preprocessing is given by an example below:

Processed data after preprocessing:

username	Date	Comment	Subreddit
argentgrove	5/29/2016 1:15	sad zoos animals didnt man try commit suicide lion recently lion killed	news

Table.4.0.2 Example of processed data after preprocessing

The raw data provided in Table 4.0.1 contains words such as "to," "be," and "a," which are commonly occurring and do not add much value for our analysis. To improve the quality of our data for further analysis, we removed these uninformative words and eliminated any extraneous punctuation. In addition, we converted the text to lowercase, as shown in Table 4.0.2. These modifications are an important part of the preprocessing stage and are necessary to enhance the quality of our data for subsequent analysis: Sentimental Analysis.

4.1. Sentimental Analysis:

For the sentimental analysis, VADER and EMPATH were employed as sentiment analysis methods that provide a rich and detailed understanding of the emotions expressed within the comments.

a) VADER:

- To gauge public opinion on the killing of Harambe by the Zoo Staff, VADER was selected as one of the sentiment analysis tools due to its widespread use and proven efficacy. VADER (Valence Aware Dictionary and sEntiment Reasoner) is Sentiment Lexicon that assigns a sentiment score based on their inherent positivity, negativity, or neutrality. After incorporating set of heuristic rules, VADER calculates polarity scores for each sentiment and also computes a compound score representing the overall sentiment of the text. These scores range from -1 (most negative) to 1 (most positive), with 0 indicating a neutral sentiment.

- Polarity Scores and the VADER Overall Sentiments: The SentimentIntensityAnalyzer class from the nltk.sentiment module is utilized to produce polarity scores for each entry in the preprocessed comments list. By manually defining threshold values, the overall VADER sentiment is determined according to the compound score computed through a sentiment analysis tool. The thresholds can be set in any desired sequence, depending on user preference. Below are the threshold values currently set for the Overall Sentiment category.

Coumpound Score	Overall VADER Sentiment Category
Compound Score <= -0.5	Very negative
-0.5 < Compound Score <=-0.05	Negative
-0.05 <Compound Score<=0	Neutral
0<Compound Score<0.05	Neutral
0.05<=Compound Score<0.5	Positive
Compound Score>=0.5	Very Positive

Table.4.1.a. Categorization of VADER Overall Sentiment by Compound Score

b) EMPATH:

- To make this research more interesting and gain deeper insights into the distribution of emotions within each main category, we utilized EMPATH. EMPATH is a lexicon-based method that categorizes text into predefined categories based on the occurrence of particular words. The lexicon comprises over 200 categories and thousands of related words, and the method assigns a score to each category based on the frequency of the words present in the text.
- Predefined Categories: In this study, we selected a set of predefined categories that represent various emotions, such as love, hate, joy, play, positive_emotion, trust, aggression, anger, fear, pain, and negative_emotion, to provide a comprehensive representation of the emotional content within the comments as these predefined categories include both positive and negative sentiments. Users can customize the categories based on their preferences. For each comment, the method analyzes the words and calculates a score for each predefined category. The result is a set of scores, one for each category, that represent the emotional content of the comment. By analyzing the scores of each category, we gain insights into the distribution of emotions within each main category.

Predefined Categories
love, hate, joy, play, positive_emotion, trust, aggression, anger, fear, pain, and negative_emotion

Table.4.1.b.1 Predefined Categories for EMPATH

- Calculation of EMPATH Overall Sentiments: In contrast to VADER, EMPATH does not produce a compound score for each comment. Rather, scores are generated for all predefined categories, for each comment. Consequently, If the sentiment score for a particular predefined category is the highest among all other categories, then that predefined category would be considered as the EMPATH Overall sentiment of that comment. If the maximum score of predefined categories is '0', then we assign "Neutral" to the EMPATH Overall Sentiment for the given comment.

Sentiment Scores	Overall EMPATH Sentiment
If the Sentiment score for a predefined category, let's say, x, is highest among all others	x
If the maximum Sentiment score is '0'	'Neutral'

Table.4.1.b.2 Predefined Categories for EMPATH

The combined results of VADER and EMPATH analyses are stored in a CSV file named 'Sentimental_analysis_VADER_EMPATH.csv', which also contains other preprocessed information collected from Reddit, such as the author's name, date, comment, and subreddit.

Author	Date	Comment	Subreddit	VADER_neg	VADER_neu	VADER_pos	VADER_compound	EMPATH_love	EMPATH_hate	EMPATH_joy	EMPATH_pla	EMPATH_	EMPATH_t	EMP	EMPATH_r	Overall_VADER_s	Overall_EMPATH_sentimen
argentgrov	5/29/2016 1:15	sad zoos ar	news	0.636	0.364	0	-0.9324	0	0	0	0	0	0	0	0.166667	Very Negative	Negative_emotion
Knittingpas	5/29/2016 3:39	remember	news	0.099	0.582	0.319	0.762	0	0	0	0.12	0	0	0	0.08	Very Positive	Play
soupcansa	5/29/2016 5:39	totally lost	news	0.319	0.518	0.163	-0.8036	0	0	0	0	0.0625	0	0	0.09375	Very Negative	Negative_emotion
PateranTik	5/29/2016 6:02	worked aq	news	0.145	0.649	0.206	0.1531	0	0	0	0	0	0	0	0	Very Positive	Neutral
deleted	5/29/2016 2:21	removed	news	0	1	0	0	0	0	0	0	0	0	0	0	Neutral	Neutral
bearchyllz	5/29/2016 4:37	little sister	news	0.254	0.746	0	-0.7778	0	0	0	0	0	0	0	0.125	Very Negative	Negative_emotion
hammilithc	5/29/2016 6:22	shame leas	news	0.508	0.492	0	-0.4767	0	0	0	0.25	0	0	0	0	Negative	Play
tenderbool	5/29/2016 6:58	upset happ	news	0.322	0.601	0.077	-0.9625	0.01587302	0	0	0.031746	0.03175	0.015873	0	0.063492	Very Negative	Negative_emotion

Table.4.1.b.3 Sample Comments from Sentimental_analysis_VADER_EMPATH.csv

c) Determining the most dominant Sentiment using VADER and EMPATH:

After assigning the overall sentiment category to each comment using both VADER and EMPATH, we computed the average sentiment scores for the five categories, namely Very Positive, Positive, Neutral, Negative, and Very Negative for VADER. Similarly, we calculated the scores for predefined categories using EMPATH. The sentiment category with the highest average score was deemed as the most dominant sentiment in both VADER and EMPATH analyses. The below table of distribution sentiments across VADER is calculated in .csv file:

Type of Sentiment	Count
Very Positive	703
Positive	36
Neutral	429
Negative	383
Very Negative	651

Table.4.1.c.1 VADER Sentiment percentages from Sentimental_analysis_VADER_EMPATH.csv

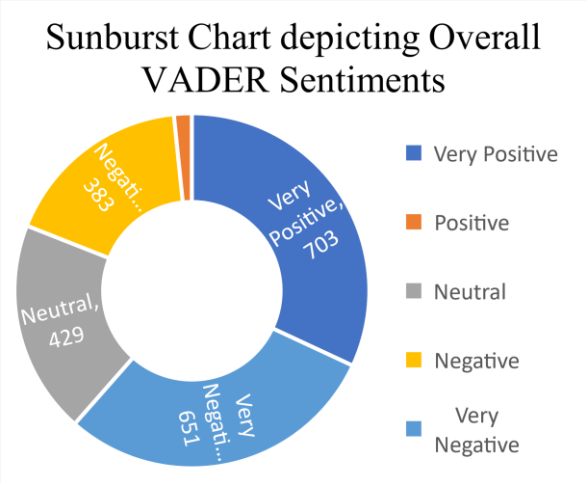

Fig.4.1.c.1 Sunburst Chart depicting Overall VADER Sentiments

Based on the figures and tables presented above, it is evident that the overall sentiment category with the "Very Positive" sentiment contains a higher number of comments, However, the Results section will discuss about the average compound scores produced for each sentiment category and the highest average compound score sentiment will become the dominant sentiment in VADER approach.

Similarly, the below table shows the percentage of distribution sentiments across EMPATH is calculated in .csv file:

Type of Sentiment	Count	Sentiment Percentage
Aggression	41	1.862%
Anger	10	0.454%
Fear	13	0.590%
Hate	135	6.130%
Joy	1	0.045%
Love	64	2.906%
Negative_emotion	436	19.800%
Neutral	752	34.150%
Pain	59	2.679%
Play	476	21.616%
Positive_emotion	145	6.584%
Trust	70	3.178%

Table.4.1.c.2 EMPATHSentiment percentages from Sentimental_analysis_VADER_EMPATH.csv

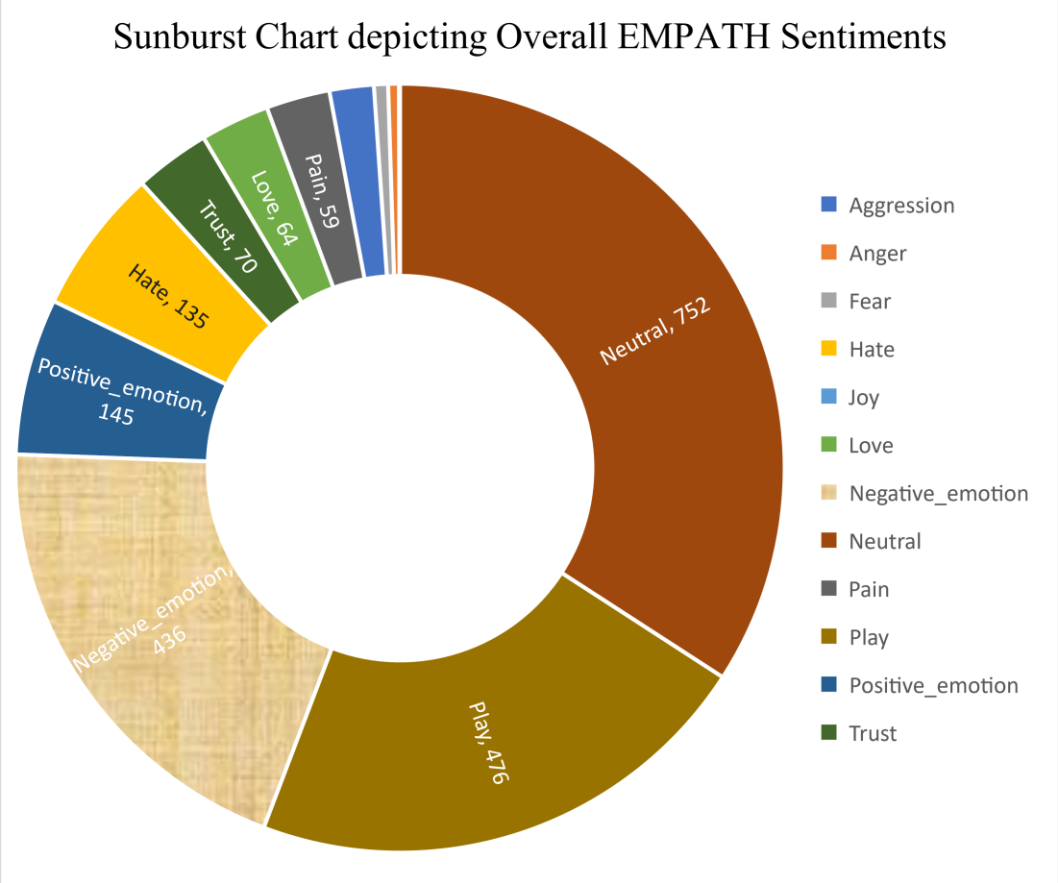

Fig.4.1.c.2 Sunburst Chart depicting Overall EMPATH Sentiments

Similarly, based on the figures and tables presented above for EMPATH Sentiments, it is evident that the overall sentiment category with the "Neutral" sentiment contains a higher number of comments, indicating that the dominant sentiment identified by EMPATH is "Neutral". The Results section will include the average scores and dominant score in EMPATH obtained through the code.

d) Finding Top words and Visualizing Sentiment Distribution:

To obtain a better understanding of the frequently used words in the comments, we identified the top words and their frequency for each sentiment category. To visually represent this data, we utilized a WordCloud, which displays a visual representation of the top n most commonly used words for each comment that was categorized as "Very positive" and "Very Negative" by the VADER Overall Sentimental Analysis. The size of each word in the Word Cloud is directly proportional to the frequency of its occurrence in the text. The word 'animal' has high occurrence in positive category and relatively good occurrence in negative category. and the words 'gorilla', 'parent' and 'kid' has higher occurrence in negative category than in positive. The word 'human' has more occurrence in positive category. The wordcloud is shown below in Fig4.1.d.1 for both sentiment categories:

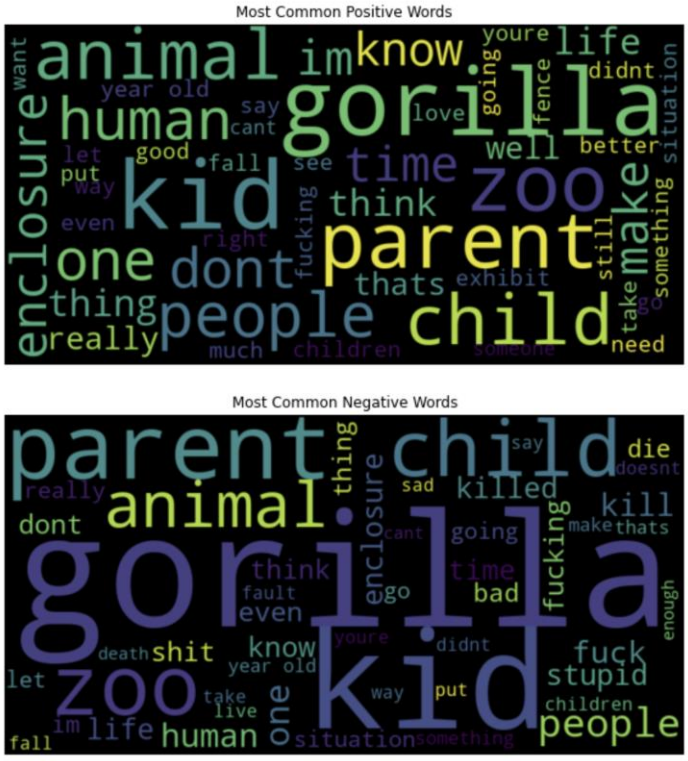

Fig.4.1.d.1 Sunburst Chart depicting Overall EMPATH Sentiments

e) Other Visualizations:

Seaborn and matplotlib modules were used to create bar plots and heatmaps to visually represent the distribution of sentiment within the comments analyzed using VADER and EMPATH methods. The bar plots show the number of comments that fall into each sentiment category for both VADER and EMPATH analyses. The heatmaps show the sentiment distribution by subreddit for each sentiment category using VADER and EMPATH analyses. The results provide valuable insights into the distribution of sentiment within the comments, enabling us to identify patterns and trends in the data. The visualizations are made using the modules.

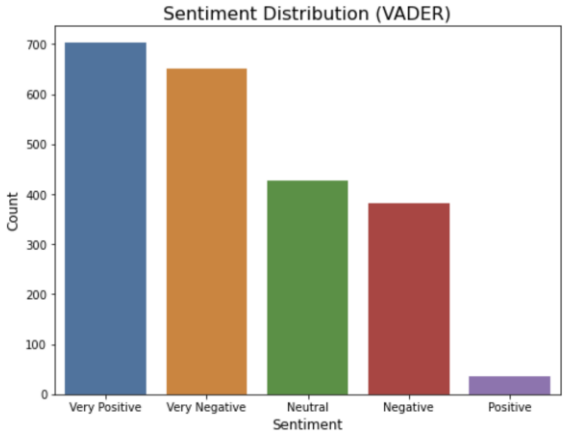

Fig.4.1.e.1 Sentiment Distribution in VADER

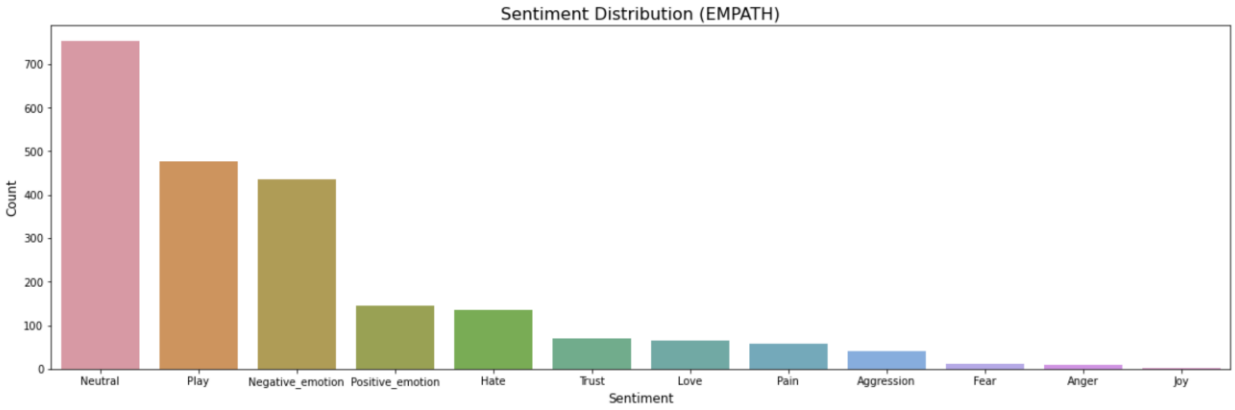

Fig.4.1.e.2 Sentiment Distribution in EMPATH

4.2. Topic Modeling:

Our study aims to uncover the themes and topics present in the comments through topic modeling. To achieve this, we utilized topic modeling through the LDA and Gensim library to uncover the underlying thematic structure and main topics discussed within the text. This process involved creating dictionaries, corpora, and training LDA models, resulting in valuable insights for making informed decisions and better recommendations.

Sentiment distribution across the dominant topics: In order to examine the sentiment distribution across the dominant topics, we employed the LDA model to analyze the corpus with a specific number of topics. The LDA model identified patterns in the data and assigned words to topics based on their co-occurrence in the comments. After the model was trained, the most relevant topic for each comment was determined and added to the dataset as a new column indicating the dominant topic.

Aggregation of Sentiment Distribution Across Topics: In addition to determining the dominant topic for each comment using the LDA model, we also used the VADER method to calculate the sentiment score for each comment. This allowed for a comprehensive analysis of the sentiment distribution across the dominant topics. The sentiment score for each comment was added to the dataset as a new column, as shown in the example table below in addition to the Sentiment category generated by VADER Overall Sentiment on the provided threshold in the code:

Comment	Dominant Topic	Compound Score	Sentiment Category
1	Topic 3	0.75	Very Positive
2	Topic 2	-0.20	Negative
3	Topic 1	0.40	Positive
4	Topic 3	-0.90	Very Negative
5	Topic 2	0.80	Very Positive

Table.4.2. Example table for Topic Modeling functionality to examine the sentiment distributions across dominant topics

The Results section includes graphs and charts for each topic, providing necessary information for understanding the sentiment distribution across topics and aiding in the interpretation of the data. These visuals offer insights into the key subjects and aspects discussed by the public in relation to Harambe's death, ultimately contributing to a better understanding of society's stance on animal welfare and human-animal interactions. pyLDAvis library to visualize the topic modeling results obtained from the LDA model trained on the corpus.

5.Results

Our analysis of the sentiment and topics related to the killing of Harambe by Cincinnati Zoo staff has revealed a range of perspectives among Reddit users.

- After filtering and preprocessing 2,201 comments, our analysis using VADER found that the most prevalent overall sentiment category was "Very Positive." followed by the sentiment “Neutral” Sentiment.

Sentiment category	Sentiment Count	Average Sentiment Count
Very Positive	703	0.319
Positive	36	0.016
Neutral	429	0.195
Negative	383	0.174
Very Negative	651	0.016

Table.5.1. Average Sentiment Count for the VADER obtained from code

To gain further insights, we conducted EMPATH analysis to determine the most prevalent overall sentiment category by utilizing predefined categories. The overall sentiment category was determined by selecting the maximum value of the predefined category in each column. Surprisingly, the results obtained through EMPATH analysis differed from those obtained using VADER. According to EMPATH analysis, the most prevalent overall sentiment category was "Neutral." This variation in results is expected, as the predefined categories in EMPATH are limited to 12 categories to maintain consistency in the codebook, while EMPATH has over 2000 categories. This could be one of the reasons for the differing results. The sentiment counts and average sentiment counts obtained through EMPATH analysis are presented in the following table:

Sentiment category	Average Sentiment Count
Love	0.029
Hate	0.061
Trust	0.032
Pain	0.026
Anger	0.005
Joy	0.000
Fear	0.006
Aggression	0.019
Neutral	0.342

Play	0.216
Negative_emotion	0.198
Positive_emotion	0.066

Table.5.2. Average Sentiment Count for the EMPATH obtained from code

The results in Table 5.1 show that there were many positive comments regarding the Harambe incident. However, when analyzing the data with EMPATH, as shown in Table 5.2, the overall prevalent sentiment category was found to be "Neutral," even though multiple predefined sentiments were detected.

- To identify the predominant themes and topics of the incident, Topic Modeling was performed on the processed dataset. This analysis was complemented by the Sentiment analysis results obtained from VADER, providing a comprehensive layout of valuable insights for the public, zoo staff, and parental responsibility in making informed decisions going forward in similar situations. The resulting topic words for each topic and the sentiment distribution for each sentiment category are calculated and presented below for a better understanding.

```
Topic: 0
Words: 0.019*"parents" + 0.017*"kid" + 0.012*"zoo" + 0.012*"child" + 0.010*"dont" + 0.010*"know" + 0.009*"year" + 0.008*"could" + 0.008*"get" + 0.008*"kids"

Topic: 1
Words: 0.014*"people" + 0.008*"animals" + 0.008*"like" + 0.007*"zoo" + 0.007*"take" + 0.007*"one" + 0.006*"reason" + 0.006*"go" + 0.005*"let" + 0.005*"probably"

Topic: 2
Words: 0.015*"like" + 0.013*"youre" + 0.011*"kid" + 0.010*"people" + 0.007*"get" + 0.007*"time" + 0.007*"dont" + 0.007*"life" + 0.007*"one" + 0.006*"really"

Topic: 3
Words: 0.028*"gorilla" + 0.015*"would" + 0.015*"gorillas" + 0.014*"humans" + 0.014*"human" + 0.012*"kill" + 0.011*"species" + 0.010*"life" + 0.009*"one" + 0.008*"peop

Topic: 4
Words: 0.031*"gorilla" + 0.018*"child" + 0.015*"parents" + 0.013*"zoo" + 0.009*"zoos" + 0.009*"fault" + 0.009*"animal" + 0.008*"enclosure" + 0.008*"think" + 0.008*"do
```

Fig.5.1 Topic words for each topic

Sentiment	Negative	Neutral	Positive	Very Negative	Very Positive
Dominant_topic					
0	0.172161	0.163614	0.023199	0.300366	0.340659
1	0.150685	0.188356	0.010274	0.294521	0.356164
2	0.158209	0.256716	0.011940	0.241791	0.331343
3	0.197917	0.226562	0.020833	0.294271	0.260417
4	0.185484	0.180108	0.005376	0.336022	0.293011

Fig.5.2 Proportion of Comments for Each Sentiment category for each dominant topic

- The above figure Fig.5.2. represents that, for dominant topic 0, 17.2% of comments were classified as negative, 16.4% as neutral, 2.3% as positive, 30.0% as very negative, and 34.1% as very positive.
- The following chart in Fig.5.3 is provided below to better comprehend how sentiments vary across the identified topics:

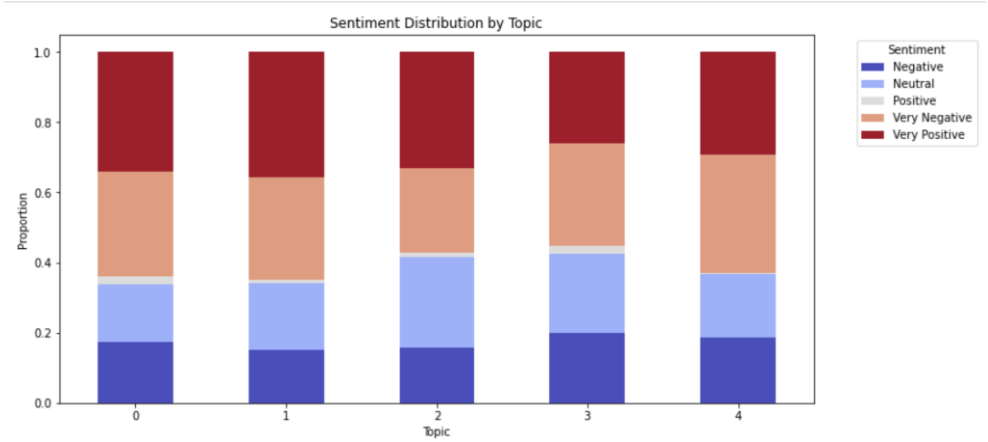

Fig.**5.3 Visualization of Proportion of Comments for Each Sentiment category for each dominant topic**

pyLDAvis display interactive visualization, which allows for an exploration of the topics and their associated keywords, as well as their interrelationships. It is shown below:

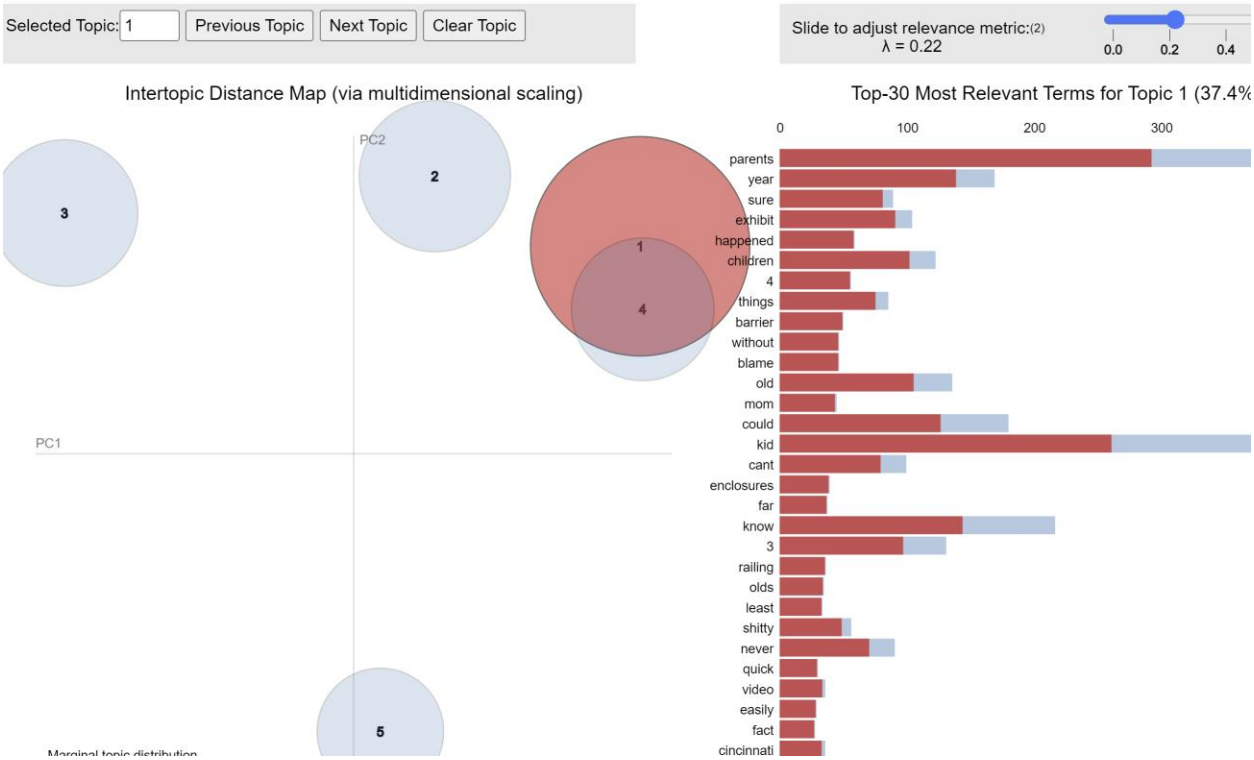

Fig.5.4 pyLDAvis interactive display

- The lambda value, which can be set to a lower value (e.g. 0.1) or a higher value (e.g. 0.9), determines the prominence of unique or frequent words for each topic. By adjusting the lambda value, a deeper understanding of each topic's subtleties can be gained. The proximity of the circles on the chart indicates the similarity of the topics.
- The analysis of comments has revealed that the main topics and themes discussed were related to animal welfare, zoo safety, and parental responsibility with regards to children and animals in

captivity, as shown in Figure 5.1. Subtopics within these themes included discussions about whether Harambe posed a threat to the child, and animal welfare.

- From the Sentimental analysis and Topic modeling methods performed, the research questions were addressed.

6.Conclusion

In conclusion, our analysis of sentiment and topics related to the killing of Harambe by Cincinnati Zoo staff has provided valuable insights into public perception of the incident. Our use of VADER and EMPATH allowed us to gain a comprehensive understanding of prevalent sentiment categories, revealing that while negative comments were present, many users expressed positive sentiments. Our analysis using VADER revealed that the most prevalent sentiment category among the comments was "Very Positive," whereas the EMPATH analysis identified "Neutral" as the dominant sentiment category. Additionally, the topic modeling analysis showed that the primary themes discussed were related to animal welfare, zoo safety, and parental responsibility, with specific subtopics focused on the circumstances surrounding the incident. Overall, these findings can be used to make informed decisions and recommendations for future situations involving animal welfare and safety in zoos. Our analysis also revealed many other useful insights, as explained in the Analysis section of this report.

7.Limitations

Similar to other research endeavors, the current study is not without its limitations. First, the dataset utilized in our research was limited to a single subreddit and 2,201 comments. Although this provided a varied sample of comments, enabling us to draw meaningful conclusions, it may not be representative of a broader audience. In terms of sentiment analysis, we employed VADER and EMPATH. While these tools yielded valuable insights, the accuracy of VADER always depends on many factors like data quality and quantity, linguistic complexity, and textual context. Moreover, VADER might struggle to accurately discern language intricacies, like sarcasm and irony, which results in affecting sentiment analysis precision. Similarly, data accuracy can be compromised in EMPATH, as it relies on predefined categories that might not sufficiently capture the all the human emotions and language. Additionally, these predefined categories have a finite range, while human emotions and language exhibit vast diversity. Furthermore, the model can only detect topics in the data based on word patterns and associations but cannot discern the true meaning or context of those words. Ultimately, topic modeling necessitates human interpretation to evaluate the relevancy and accuracy of the identified topics. Therefore, it is important to know the limitations when analyzing with the Sentimental analysis and topic modeling.

8.References

- Monsó, S., Benz-Schwarzburg, J., & Bremhorst, A. (2018). Animal Morality: What It Means and Why It Matters. The journal of ethics, 22(3), 283–310. https://doi.org/10.1007/s10892-018-9275-3

- Dubois, S., Fenwick, N., Ryan, E. A., Baker, L., Baker, S. E., Beausoleil, N. J., Carter, S., Cartwright, B., Costa, F., Draper, C., Griffin, J., Grogan, A., Howald, G., Jones, B., Littin, K. E., Lombard, A. T., Mellor, D. J., Ramp, D., Schuppli, C. A., & Fraser, D. (2017). International consensus principles for ethical wildlife control. Conservation biology : the journal of the Society for Conservation Biology, 31(4), 753–760.https://doi.org/10.1111/cobi.12896

- Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225. https://doi.org/10.1609/icwsm.v8i1.14550