
THE MEANING OF LAST WORDS: DEATH ROW INMATES IN TEXAS

Kenneth Hudson

khud1010@gmail.com

Abstract

Since 1982, Texas has kept a record of last statements given by death row inmates about to receive their sentence. Past studies have used manual methods of extracting the sentiment behind these last statements and separating them into distinct classes. However, this study aims to focus on the term-frequency inverse document frequency text vectorization method to find any meaning in the clusters, which are formed via a k-means algorithm. We also look at the race and age of inmates and stratify the data into new sub-corpi. These are then investigated again in order to uncover any differences that might exist between these demographics as they face the end of their lives.

Keywords:

Sentiment Analysis, Tf-Idf, k-means clustering, death row, inmates

Table of Contents

Abstract.....	i
Table of Contents.....	ii
Introduction and Problem Statement.....	1
Literature Review.....	1
Data.....	3
Methods.....	5
Results.....	6
Discussion.....	8
Conclusions.....	8
Directions for Future Work.....	9
Data Availability.....	9
Code Availability.....	9
References.....	9
Appendix A.....	11

Introduction and Problem Statement

This paper investigates the last words of Texas inmates on death row, who chose to give a final statement, through sentiment analysis via term frequency inverse document frequency (tf-idf), accompanied by k-means clustering. Not only will sentiment analysis be done on each individual's last words, but demographics, like race and age, will also be compared in order to find trends in the text data, if any exist.

The death penalty goes as far back as 1700 BCE when Hammurabi's Code was introduced, officially codifying the first laws for the Babylonian people. As time moved on, many nations would change the way in which executions were performed, but the punishment for a heinous crime was the same. In modern times we have seen a wave of human rights activists calling for the abolishment of the death penalty. As of the writing of this paper, there are 144 countries that have abolished the death penalty in law or in practice, while 55 countries have retained the death penalty as a form of punishment for crimes (Death Penalty Information Center n.d.). The most notable of these countries are China and the United States.

While there are U.S. federal laws in which the death penalty is invoked, 23 states have abolished the practice with 6 more pausing executions (Death Penalty Information Center n.d.). Texas is one of the 21 states who still practices the death penalty and since 1982, there have been 587 inmates executed, the most recent being in February 2024 (Texas Department of Criminal Justice). Before an inmate's execution, they are asked if they wish to give a final statement. This right of last words has been common practice in America and the world at large, but Texas has put all of these last statements online as an open source dataset of death row inmates. The website not only has the last statements, but also some basic demographics about each inmate. (https://www.tdcj.texas.gov/death_row/dr_executed_offenders.html). Each inmate has the right to speak their last message before they leave this world. This paper attempts to discover what someone, who is usually viewed as the worst of society, has to say when faced with the ultimate ending.

Literature Review

There are a few different papers that address the topics related to this dataset. In 2005, Nathan Heflick wrote 'Sentenced to Die: Last Statement and Dying on Death Row', in which he discussed six themes among these last statements. Afterlife belief, claims of innocence, activism, love, forgiveness, or silence (lack of response) were the common themes seen through his qualitative content analysis approach. Sharon Foley and Brendan Kelly followed up this work in 2018 when they studied the dataset and found nine themes, some matching Heflick's a decade

prior. Schuck and Ward (2008) identified even more themes and, more importantly, defined a common structure to last messages in which a self-reference is made, followed by messages to their family or the victim's family and their feelings on the current situation (whether accepting or rejecting). Many others also conducted such experiments with a new lens, but they mostly relied on manual examination of documents and did not look at differences between demographics.

David Lester and John Gunn co-authored a paper in the *Journal of Ethnicity in Criminal Justice* that takes on the task of exploring whether or not ethnicity has any affect on last statements made by death row inmates (2013). They used a program that could analyze and count words with positive or negative meanings. This technique allowed them to segment the statements into nine different content types: affective emotion processes, positive emotions, positive feelings, discrepancy, references to others, past tense verbs, religion, physical states and functions, and sexuality. They found these differences were only slight and they did not outweigh the similarities between ethnic groups. This study aims to do a similar analysis to Lester and Gunn, however, we will use a different analysis approach (tf-idf and k-means) and go deeper by using age as a new demographic.

While these past works utilize various methods to draw conclusions about the text corpus, they mostly relied on manual observations or some programming script that allows words to be counted and categorized. This paper will study the corpus using term-frequency inverse document frequency text vectorization to find terms with a high tf-idf score and then perform k-means clustering to find similar documents. The theoretical benefits of using term weighting to aid with information retrieval were introduced by Karen Sparck Jones in a 1972 paper, "A statistical interpretation of term specificity and its application in retrieval". In it, the use of inverse document frequency (idf) proved to have noticeable improvement in performance due to improved term-specificity (Jones 2004). The idea behind idf is that terms that appear many times over a corpus are not great discriminators when attempting to find similar documents in the corpus. However, it wasn't until 1988 the Salton and Buckley tested automatic text retrieval systems that they combined tf and idf in order to optimize their text retrieval system. The product of tf and idf gives a weight, or importance, to specific terms in a document by penalizing those that show up more often throughout the corpus. In simpler terms, words that show up often in a few documents will have more importance than those that show up many times in all of the documents of a corpus. More recent models, like doc2vec and its variations, are powerful tools but tf-idf was used in the paper because of its ability to work well on smaller, more focused corpora of text.

Data

The Texas Department of Criminal Justice (TDCJ) has kept a record of death row inmates' last words going back to 1982. As of the date when this study was conducted, there were 587 inmates executed throughout Texas' correctional system. The data for this study was obtained via a web scraping python library (BeautifulSoup). Along with these last words, some basic demographics are also included for each inmate as well as their execution date. The web scrape used the same url formatting used by the TDCJ web page, which has a base url with the inmates last and first names as the last route of the url. While this logic worked in most cases, some inmates had no last words and a generic page was seen. Other issues that arose in data collection was a difference in inmate names between the death row information table and the page where the last words are located. Some of these names were either nicknames or simple misspellings and led to no record being found.

After the web scrape, the resulting data included 459 inmates with unique last words and 50,000 total words in the corpus. The documents then went through a simple preprocessing function that removes punctuation, non-alphabetic characters, words shorter than 3 letters, and stop words while also converting all the letters to lowercase. This excluded two inmates as their last statements matched one of the removal parameters above. The final dataset consists of 4 racial groups: White, Black, Hispanic, Other (Figure 1). Inmates had an average age of 40.21 years and ranged from 24 to 78 years old at the time of their execution. This dataset consists of both males and females, but the TDCJ does not have any of their genders displayed nor is it necessary to track for the goal of this paper. Out of the entire corpus, the term 'love' is mentioned over twice as much as the next most common word (family). Figure 2 shows the top 10 terms by frequency in the corpus and many of these same terms are also listed in the top 10 tf-idf scores (Figure 3). So while tf-idf ignores certain words like 'the', it finds that other very common words are important to this corpus. Listed below are four examples of last statements by inmates to give some insight into the variation between last statements:

- Yes sir. I would like to thank God, my dad, my Lord Jesus savior for saving me and changing my life. I want to apologize to my in-laws for causing all this emotional pain. I love y'all and consider y'all my sisters I never had. I want to thank you for forgiving me. Thank you warden.
- I'm ready to go home.
- I hereby declare, Robert Steven Everett and Nicholas Velasquez, guilty of crimes against me, Douglas Alan Feldman. Either by fact or by proxy, I find them both guilty. I hereby sentence both of them to death, which I carried out in August 1998. As of that time, the

State of Texas has been holding me illegally in confinement and by force for 15 years. I hereby protest my pending execution and demand immediate relief.

- I'm innocent. I had nothing to do with my family's murders. I want to thank everyone who has supported me. I hope they continue to fight. You know who you are. That's all. Thank you, Warden.

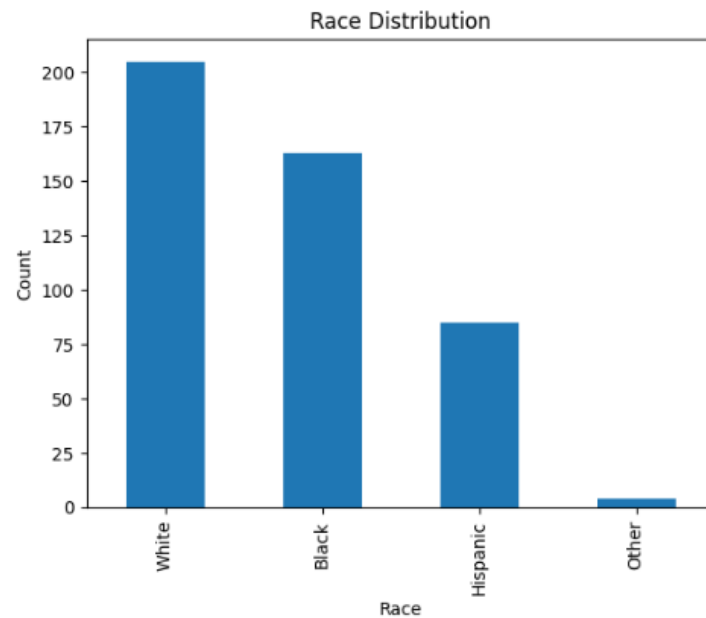


Figure 1: Race distribution in executed inmates in Texas

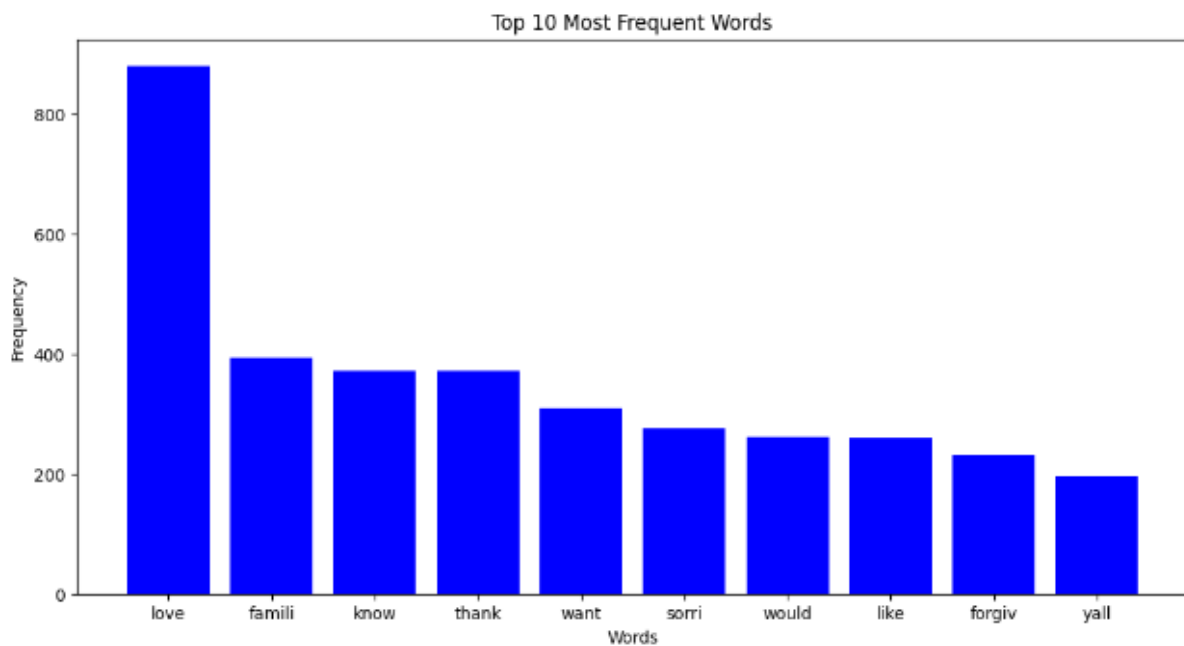


Figure 2: Top 10 most frequent (stemmed) words in the death row inmate corpus

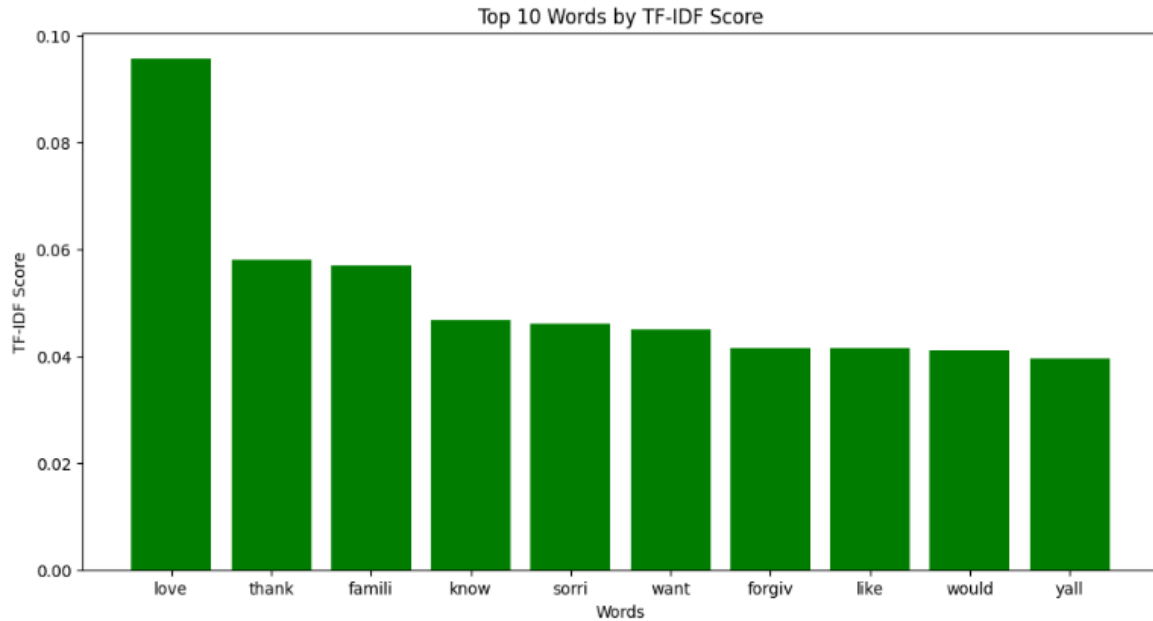


Figure 3: Top 10 words ranked by tf-idf score

Methods

Initially, the entire pre-processed corpus of data was passed to a tf-idf vectorizer and the resulting matrix was then fitted on a k-means clustering model, where number of clusters (k) equals 4. K-means clustering finds similar documents by comparing each one's multi-dimensional text vector to all the other document vectors and giving a score back. Both of these functionalities come from scikit-learn's feature extraction and cluster libraries, respectively. This was done in order to get a sense of what terms were used throughout the entire corpus of data. The next step was stratifying the dataset by race and then running the corpus of each represented race through the same procedure seen above. Finally, age groups were set up in ten year increments (Figure 4) and again each new corpus was passed through the initial procedure.

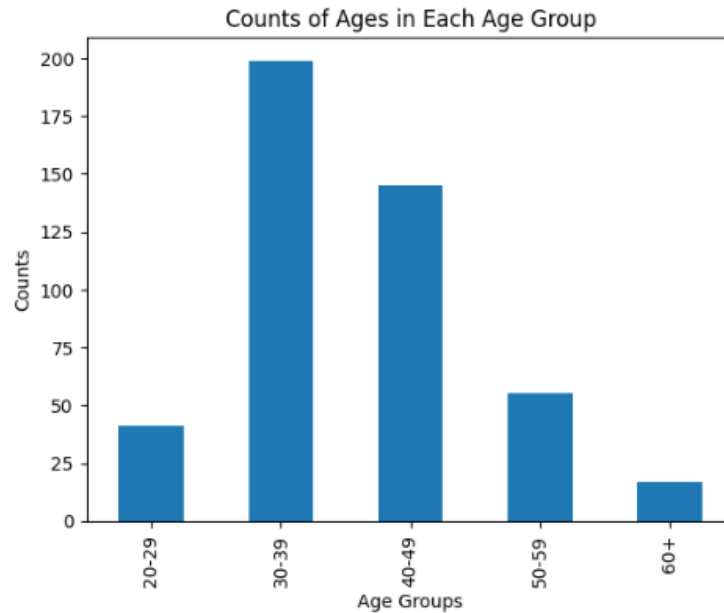


Figure 4: Number of inmates in each age group

Results

For the entire corpus, the four clusters generated after k-means fitting can be seen as representing different sentiments for the documents of that cluster. Below is a list of these clusters and a few of the top terms from each (the top terms helped to guide the naming of each cluster):

1. **Forgiveness and Religion:** sorry, forgive, lord, peace, pain, hope
2. **Love for Family:** love, thank, family, would, like, support
3. **Apology to Victim's Family:** love, know, tell, want, family, sorry
4. **Acceptance of Death:** (no) statement, ready, love, know, warden, innocent

There are certain commonalities between a few of these clusters as it is a more focused and niche corpus of data. As mentioned above, love is by far the most common term in the corpus, so it would make sense that it appears in more than one cluster. Table 1 (below) shows the distribution of documents per cluster.

Cluster	Number of Documents	Percentage
Forgiveness and Religion	116	25.27%
Love for Family	178	38.78%

Apology to Victim's Family	44	9.59%
Acceptance of Death	121	26.36%

Table 1: Cluster distribution of all documents in corpus

The same experiments were run on the four different race corpi and the five age groups tested. Tables for these are found in the Appendix but shall be discussed here. These other tables are new instances of tf-idf vectorization/clustering and were matched to the above clusters based on the common words (above) found in each cluster. The 'Other' class from the race stratified subset is excluded because it only has four total documents present, making it so each document represents one cluster. After manual review of these documents, the sentiment behind each was determined to not align with the k-means clustering results. The main comparison will therefore be done on the three remaining races: black, hispanic, and white. While many inmates discussed love, family, and forgiveness, there are still some differences between clusters of races and age groups.

Race

Based on the k-means clusters, black inmates were more likely to seek forgiveness and speak on religion in their last statement. Whereas both hispanic and white inmates were more inclined to apologize to the victim's family. Interestingly, only 5.83% (12/206) of white inmates were associated with the acceptance of death cluster. This is a stark difference from the other two races, which both had around 25% of their respective documents in this cluster. Finally, given the catholic background of many hispanic people, the hispanic inmates were surprisingly the least likely to mention religion in their last statements.

Age Group

The distribution of ages in Figure 3 shows that of the five age groups, a majority of the inmates were between the ages of 30 and 50 at the time of their execution. Each age group had slightly different distributions between the four clusters, although inmates over 60 years old only represented a small portion (3.72%) of the sample population. The youngest group (20-29 year olds) had a specific focus on the forgiveness and family (both their own and the victims) clusters but only a few documents associated with the acceptance of death cluster. The 30-39 year olds had half (50.25%) of the population in the forgiveness and religion cluster and conversely very few (12.06%) were related to the cluster apologizing to the victims family. Inmates aged 40-49 had a fairly even distribution between all clusters besides the forgiveness and religion cluster which had half the number of documents as the second lowest cluster. Those executed between ages 50-59 saw more documents in the apology to the victim's family cluster, much like was seen in the youngest group (20-29 year olds). Finally, inmates over 60 years old mostly clustered

towards forgiveness and religion but there were a few documents in each cluster that were accurate representations of that cluster's sentiment. These age groups all show that there are slightly different meanings behind their last statements and they are not all uniform distributions.

Discussion

Although slight variations between races were observed in this paper, the overall impression received was that each race had more similarities than actual differences in the sentiment of an inmate's last statement. This was also observed by Lester and Gunn in their paper (2018). Many sought to apologize to those around them, whether that be the victim's family or their own. Black inmates and white inmates had a similar number of documents in the forgiveness and religion cluster, meaning both races look to the afterlife in hopes of being forgiven by their deity. The more compelling analysis is when looking at the differences in age groups. The youngest group of inmates, those 20-29 years old, were the least accepting of their imminent death. This could be due to the lack of time spent in the world or the animalistic desire to live a long life. The inmates in the two largest age groups had similar distributions between clusters, besides the forgiveness and religion cluster. The 30-39 year old age group had a large number of documents in that cluster and it might have skewed the data due to possible mis-classification of certain documents. Although after manual inspection of a few documents, it was agreed that the cluster placement was correct for those documents. As the age groups get older, there appears to be more acceptance of death and a willingness to ask for forgiveness (from the victim's family, their own family, or even asking their god). That is until the final age group of inmates over 60 years old. The sample size of these inmates most likely affected the results as they represent less than 5% of the total population. However, it was interesting to see that the forgiveness and religion cluster had the most documents in it for that age group. These differences appear significant and might coincide with how one might feel about their death when they are younger versus having lived many decades.

Conclusions

Tf-idf and a k-means clustering algorithm have grouped the data in four clusters that represent four different sentiments of Texas' death row inmate's last words. There appears to be only slight differences between races when it comes to sentiment of last words. Black inmates were more likely to speak on religion, but all races usually spoke of forgiveness and love for family. While race did not show much about differences between these inmates, the age groups did in fact shed some light on these demographics. The younger inmates appeared to be less willing to accept death and more willing to apologize to those around and also express love for

their own family. Most of the executed inmates fell within the 30-49 year old age range at time of death and these inmates had a relatively even distribution across three of the four clusters, excluding forgiveness and religion. It was hypothesized that the category had some issues associating the proper documents when given a larger corpus due to the common theme of forgiveness. The oldest age group had little data but still provided insight when comparing their distributions to those of the younger age groups. Age appears to play a role in what someone thinks about as they face death.

Directions for Future Work

The parsing of the data via the web scrape could use some refinement as many (100+) of the inmates on the website have their last statements scanned in as a pdf. This could not be scraped with the current code and would either need to be manually inputted into the dataset or some other program would need to be used to extract info from the pdf files.

Data Availability

Texas Department of Criminal Justice Website:

https://www.tdcj.texas.gov/death_row/dr_executed_offenders.html

Code Availability

Github:

<https://github.com/khud1010/NLP-Project>

References

Death Penalty Information Center. n.d. "History of the Death Penalty." *Death Penalty Information Center*. Accessed May 5, 2024.

<https://deathpenaltyinfo.org/curriculum/high-school/about-the-death-penalty/history-of-the-death-penalty>.

Death Penalty Information Center. n.d. "State by State." *Death Penalty Information Center*. Accessed May 5, 2024. <https://deathpenaltyinfo.org/states-landing>.

- Foley, S. R., & Kelly, B. D. 2018. "Forgiveness, spirituality and love: thematic analysis of last statements from Death Row, Texas (2002–17)." *QJM: Monthly Journal of the Association of Physicians*, 111(6), 399–403. <https://doi.org/10.1093/qjmed/hcy062>
- Heflick, Nathan A. 2005. "Sentenced to Die: Last Statements and Dying on Death Row." *Omega: Journal of Death and Dying*, 51(4), 323–336. <https://doi.org/10.2190/96X8-FLUT-TCLH-EL71>
- Jones, Karen Sparck. 2004. "A statistical interpretation of term specificity and its application in retrieval." *Journal of Documentation*, 60(5), 493–.
- Lester, David, & Gunn, John F. III. 2013. "Ethnic differences in the statements made by inmates about to be executed in Texas." *Journal of Ethnicity in Criminal Justice*, 11(4), 295–302.
- Salton, Gerard, & Buckley, Christopher. 1988. "Term-weighting approaches in automatic text retrieval." *Information Processing & Management*, 24(5), 513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- Schuck, Andreas R. T., & Ward, Janelle. 2008. "Dealing with the inevitable: strategies of self-presentation and meaning construction in the final statements of inmates on Texas death row." *Discourse & Society*, 19(1), 43–62. <http://www.jstor.org/stable/42889177>
- Texas Department of Criminal Justice. 2024. "Texas Death Row Information." Death Row Information. https://www.tdcj.texas.gov/death_row/dr_executed_offenders.html.

Appendix A

Tables for Race group corpus:

Cluster	Number of Documents	Percentage
Forgiveness and Religion	55	33.74%
Love for Family	41	25.15%
Apology to Victim's Family	25	15.34%
Acceptance of Death	42	25.77%

Table 2: Black inmates

Cluster	Number of Documents	Percentage
Forgiveness and Religion	8	9.30%
Love for Family	11	12.79%
Apology to Victim's Family	43	50.00%
Acceptance of Death	24	27.91%

Table 3: Hispanic inmates

Cluster	Number of Documents	Percentage
Forgiveness and Religion	59	28.64%
Love for Family	48	23.30%
Apology to Victim's Family	87	42.23%
Acceptance of Death	12	5.83%

Table 4: White inmates

Tables for Age group corpus:

Cluster	Number of Documents	Percentage
Forgiveness and Religion	14	22.95%
Love for Family	17	27.87%
Apology to Victim's Family	26	42.62%

Acceptance of Death	4	6.56%
---------------------	---	-------

Table 5: 20-29 year old inmates

Cluster	Number of Documents	Percentage
Forgiveness and Religion	100	50.25%
Love for Family	32	16.08%
Apology to Victim's Family	24	12.06%
Acceptance of Death	43	21.61%

Table 6: 30-39 year old inmates

Cluster	Number of Documents	Percentage
Forgiveness and Religion	17	12.78%
Love for Family	41	30.83%
Apology to Victim's Family	34	25.56%
Acceptance of Death	41	30.83%

Table 7: 40-49 year old inmates

Cluster	Number of Documents	Percentage
Forgiveness and Religion	7	14.00%
Love for Family	8	16.00%
Apology to Victim's Family	25	50.00%
Acceptance of Death	10	20.00%

Table 8: 50-59 year old inmates

Cluster	Number of Documents	Percentage
---------	---------------------	------------

Forgiveness and Religion	7	43.75%
Love for Family	3	18.75%
Apology to Victim's Family	4	25.00%
Acceptance of Death	2	12.50%

Table 9: 60+ year old inmates