Case Study Rubric - Texas Death Row Final Statements: Remorse Analysis

DS 4002 – Spring 2025

Submission Format: A GitHub Repository

Individual Assignment

General Description: Submit a link to your GitHub repository containing all materials for your case study on Texas death row final statements. This project invites you to explore the relationship between final words and factors like time on death row, crime type, and execution year. You will use natural language processing techniques to quantify and visualize expressions of remorse and connect your findings to broader patterns in the criminal justice system.

Why am I doing this? This is your opportunity to apply your data science skills—specifically natural language processing and sentiment analysis—to a socially impactful dataset. By investigating emotional expression in the last words of death row inmates, you'll learn how to process unstructured text data, perform sentiment analysis, and visualize trends in human language. You'll also reflect on the ethical dimensions of capital punishment data and how it's used.

What am I going to do? You will analyze last statements from Texas death row inmates and examine how expressions of remorse vary based on factors such as crime type, time on death row, and execution year. You will use a combination of Python libraries including NLTK, VADER, Pandas, and Matplotlib to scrape, clean, analyze, and visualize the dataset. Your findings and reflections will be compiled in a GitHub repository and summarized in a one-page PDF.

Deliverables Include:

- GitHub repository containing:
  - README.md file with project description, repo map, and references
  - DATA folder containing inmate information and last statement data
  - SCRIPTS folder with web scraping, preprocessing, and analysis code
  - OUTPUT folder with remorse scores, plots, and word visualizations
  - A one-page written summary of findings, insights, and takeaways (PDF)

Tips for Success:

- Start by reading about VADER and sentiment analysis basics
- Explore text preprocessing using NLTK (e.g., tokenization, stop word removal)
- Use visual tools like word clouds or boxplots to display remorse trends
- Reflect on what the statements might reveal about the human side of capital punishment

How will I know I have succeeded?

You will meet expectations on this Case Study when your repository contains all required components and follows the rubric criteria below:

Spec Category	Spec Details
Formatting	• A GitHub repo submitted via link containing all materials • Repo contains: ▪ README.md ▪ DATA folder ▪ SCRIPTS folder ▪ OUTPUT folder ▪ One-page written summary in PDF format • Use appropriate formats for code, data, and documents
README.md	• Include an H2 section explaining the project goals and contents of the repository • Provide a summary of your research question and model approach • Include a visual map or outline of folder contents • Include references in IEEE format (cite all sources used or provided)
DATA Folder	• Contain CSV files for inmate data and final words • If applicable, include scraped raw text and processed data (e.g., remorse scores) • Ensure each file is clearly labeled and linked to relevant scripts
SCRIPTS Folder	• Include Python scripts used for: ▪ Data scraping ▪ Preprocessing and cleaning ▪ Sentiment analysis ▪ EDA and visualization • Scripts must be commented and labeled in execution order
OUTPUT Folder	• Include all visualizations (e.g., word clouds, scatterplots, boxplots) • Label plots clearly and link them to analysis questions • Include remorse score data and any exported result tables
Written Summary	• A one-page PDF document • Describe your process: from scraping and preprocessing to sentiment analysis • Briefly interpret your key findings and their broader significance • Reflect on challenges, learnings, and potential extensions of the project

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled from Streifer & Palmer (2020).