Miguel Angel Estrella-Ibarra
12/18/2025
Dr. Mario Bañuelos
MATH 120

Final Project Report

I decided to explore how Ice Cube's lyrics have evolved over the course of his career, focusing on both the words he uses and the major themes that appear across his albums. I wanted to understand whether his language and topics shifted between his early and more recent works, and whether these changes reflected artistic growth or broader societal changes. To do this, I collected raw text files of lyrics from publicly available websites like Genius. My dataset includes his first four albums AmeriKKKa's Most Wanted (1990), Death Certificate (1991), The Predator (1992), and Lethal Injection (1993) as well as four of his most recent albums Everythang's Corrupt (2018), Man vs Machine (2021), Man Down (2024), and Man Up (2025). I was interested in questions such as how the most frequent words differ between early and late albums and how major themes like violence, policing, social conflict, misogyny, or empowerment have changed over time. By answering these questions, I hoped to quantify the evolution of Ice Cube's lyrical style and identify trends that emerge over decades.

Since the raw lyrics are unstructured text, I focused on extracting meaningful variables for analysis. This involved individual tokens, frequencies of each word, and category counts for thematic sets of words, such as those related to violence, police, and weapons. Cleaning the text was a critical step. I wrote Python scripts to lowercase all text, remove punctuation, apostrophes, and hyphens, filter out English stop words using spaCy, replace offensive slurs with neutral placeholders, and tokenize the text using regular expressions. This cleaning process allowed me to generate consistent and comparable word counts across albums. Once cleaned, I created frequency dictionaries, visualized the top words for each album, and compared categories of words across early and late albums. I also developed functions to generate side-by-side visualizations, making it easier to compare trends in themes across albums.

In my initial analysis, I looked closely at AmeriKKKa's Most Wanted and Man Up. The early album was dominated by words and categories related to violence and conflict, including "beat," "weapon," and "murder," whereas the later album displayed a broader thematic range, with more emphasis on empowerment and social commentary. Comparing top words and category frequencies revealed these trends clearly, and the visualizations made it easy to see which themes were more prominent in each period. Although the analysis is ongoing, the initial results suggest that Ice Cube's language has become more varied and nuanced over time.

This project taught me a lot about handling unstructured text data and creating reusable code. Structuring the GitHub repository properly and modularizing functions in cleaning.py and analysis.py was a challenge at first, but it paid off when I could apply the same methods to multiple albums efficiently. Visualizing the data was particularly helpful, as it allowed me to communicate trends at a glance and compare albums in a way that numbers alone could not. I

also had to carefully consider how stop-word removal affects the accuracy of word counts, which required balancing clarity with completeness.

Overall, this project demonstrates that Ice Cube's lyrical style has evolved significantly over his career. Early albums focused on violence and social conflict, while later works show a broader thematic scope and more nuanced language. Cleaning, tokenizing, and analyzing the text allowed me to quantify these changes, and the visualizations made the results easy to interpret. By comparing early and late albums, I gained insight into the shifts in Ice Cube's artistic focus, providing a descriptive narrative of his career through his lyrics. The project highlights the value of careful data cleaning, modular code organization, and thoughtful visualization when working with unstructured textual data, and it sets the stage for expanding this analysis to include all of Ice Cube's albums in future work.