

# Zipf's Law in Nicki Minaj Lyrics

---

Fundamentals of AI Project

***Student Names:***

**Daksh Saini**

**Tanubhav Katiyar**

**Shubh Arya**

**Rishi Seth**

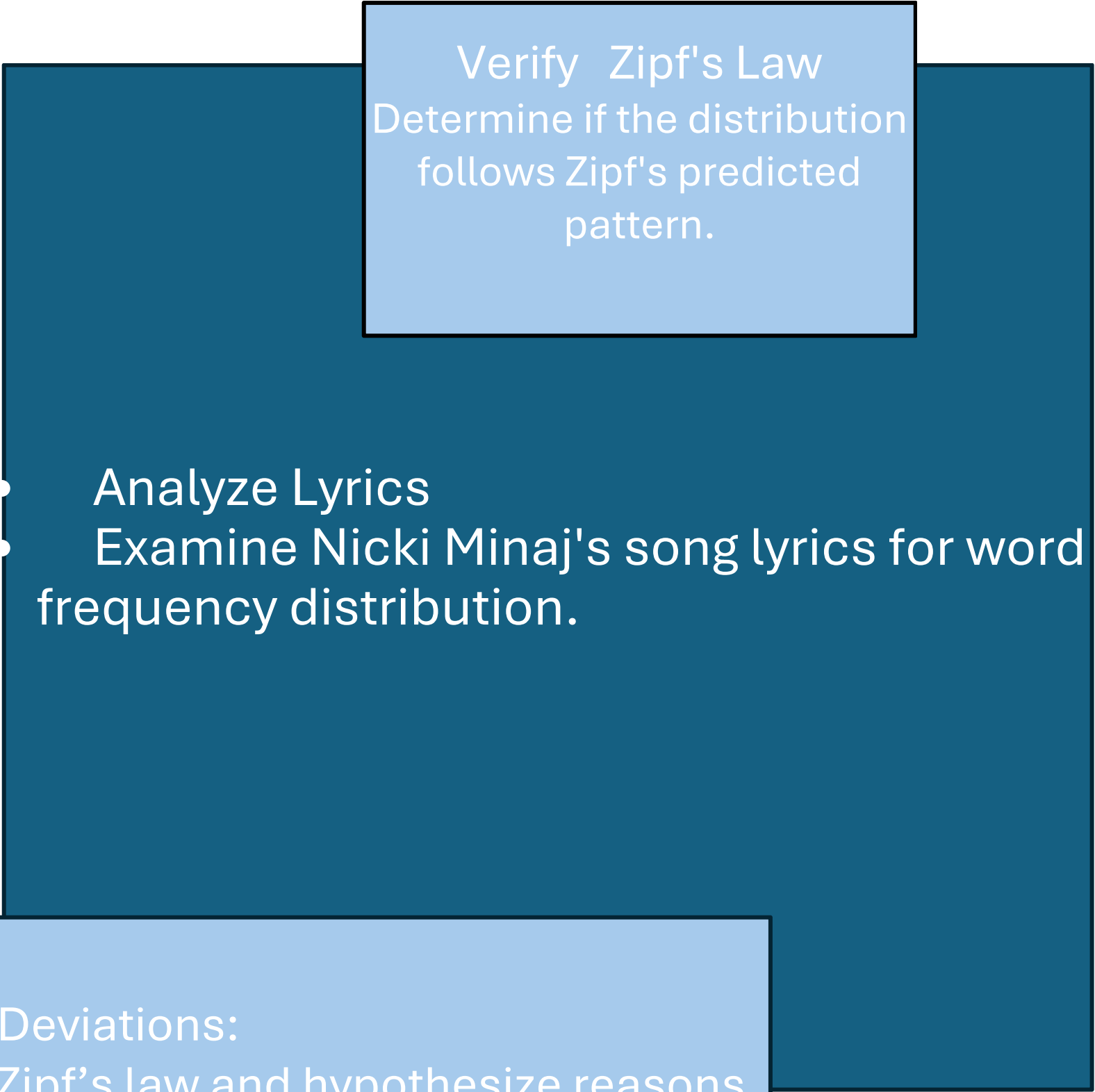
# Team Contributions

- Daksh Saini : Worked on writing and organizing the Jupyter Notebook, implemented Zipf's Law visualizations, and contributed to the research on word frequency analysis. Also helped refine the final results.
- Shubh Arya: Focused on data preprocessing, regex-based text cleaning, and plotting log-log graphs. Contributed significantly to exploring Zipf's Law and structuring the notebook flow.
- Rishi Seth: Designed and structured the PowerPoint presentation. Summarized technical content into easy-to-understand slides and helped with visual storytelling of the project.
- Tanubhav Katiyar: Contributed to the presentation design, formatted charts and explanations, and ensured the project message was communicated clearly. Also assisted in final content polishing.





# Project Objective



- Analyze Lyrics
- Examine Nicki Minaj's song lyrics for word frequency distribution.

Verify Zipf's Law  
Determine if the distribution follows Zipf's predicted pattern.

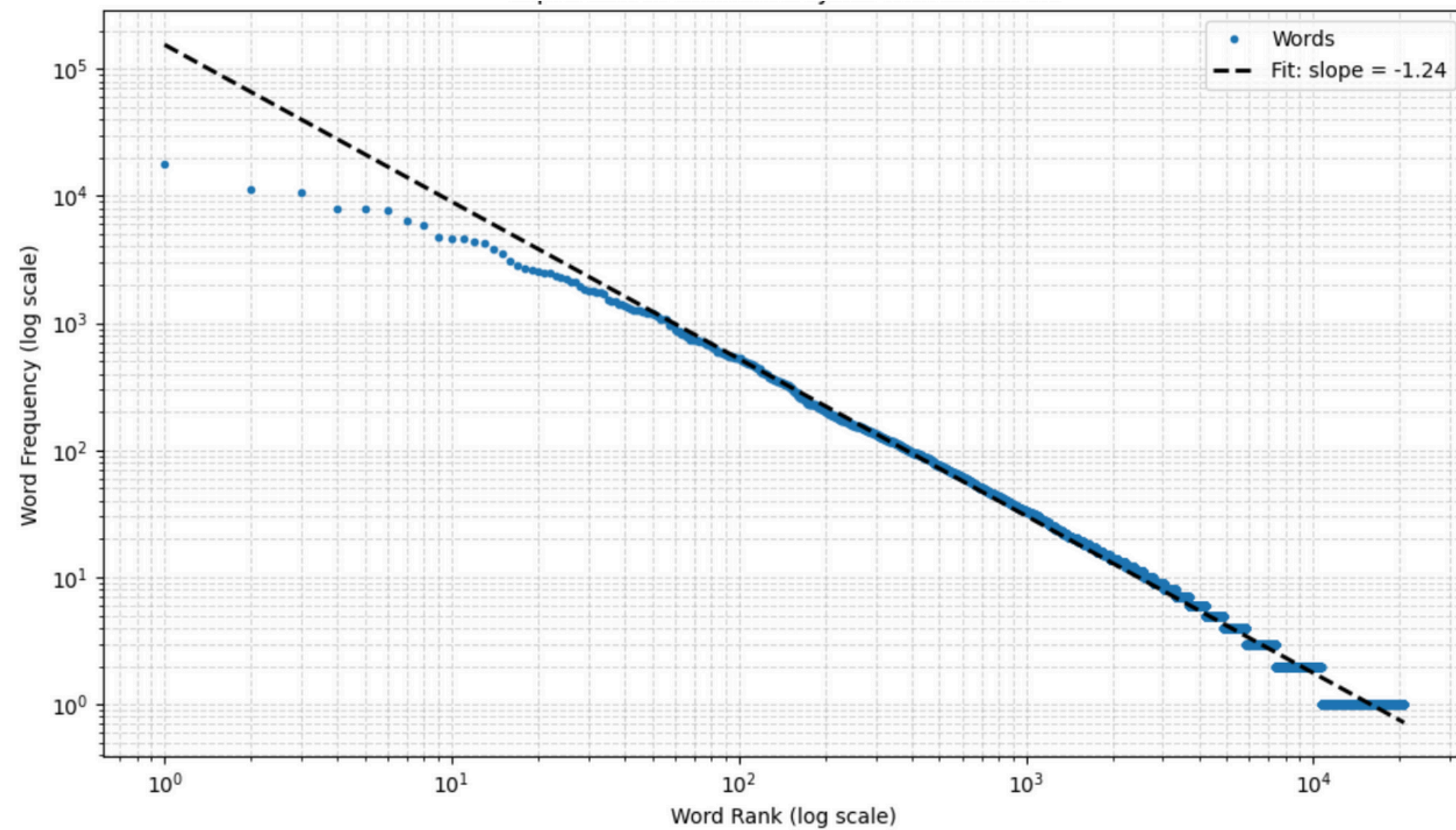
Explore Deviations:  
Investigate any departures from Zipf's law and hypothesize reasons.



# Data Collection and Preprocessing :

- **Gather Data:** Compiled a comprehensive dataset from Eminem.csv file as a college project.
- **Clean Lyrics:** Removed extraneous characters, punctuation, and formatting irregularities to standardize input.
- **Tokenize and Filter:** Split lyrics into individual words (tokens) and excluded common stop words to focus on meaningful terms.





## Methodology



### Frequency Calculation

Count occurrences of each unique word in the cleaned lyrics corpus.



### Rank Assignment

Order words from most to least frequent to assign ranks accordingly.



### Plot & Fit

Visualize rank versus frequency on a log-log scale and fit a power law model.



### Analyze Exponent

Estimate top 50 words frequency to evaluate adherence to Zipf's Law characteristics.

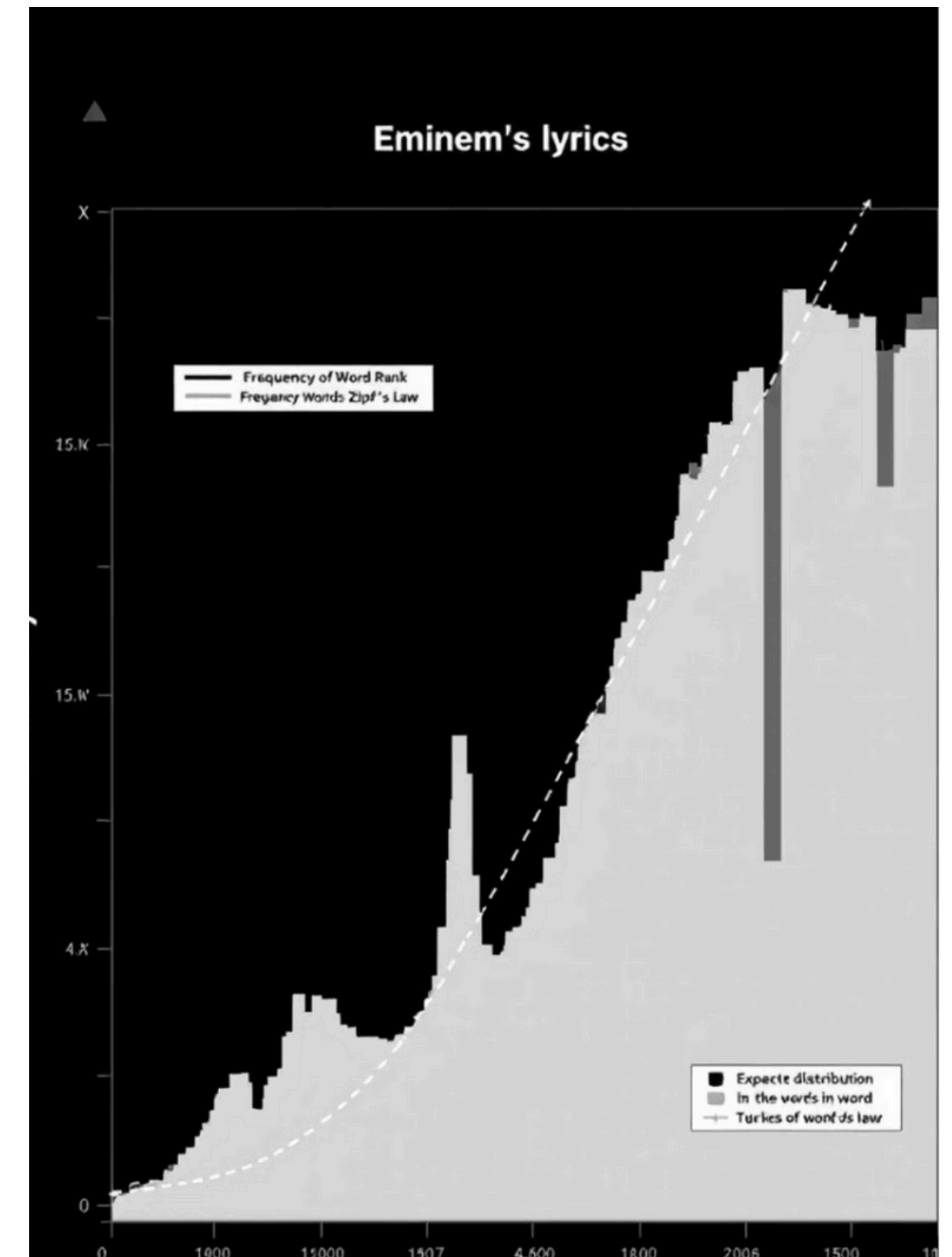
# Results and Analysis

## Rank-Frequency Plot :

The distribution generally follows a Zipfian trend but shows slight deviations at high and low ranks.

## Exponent Measurement :

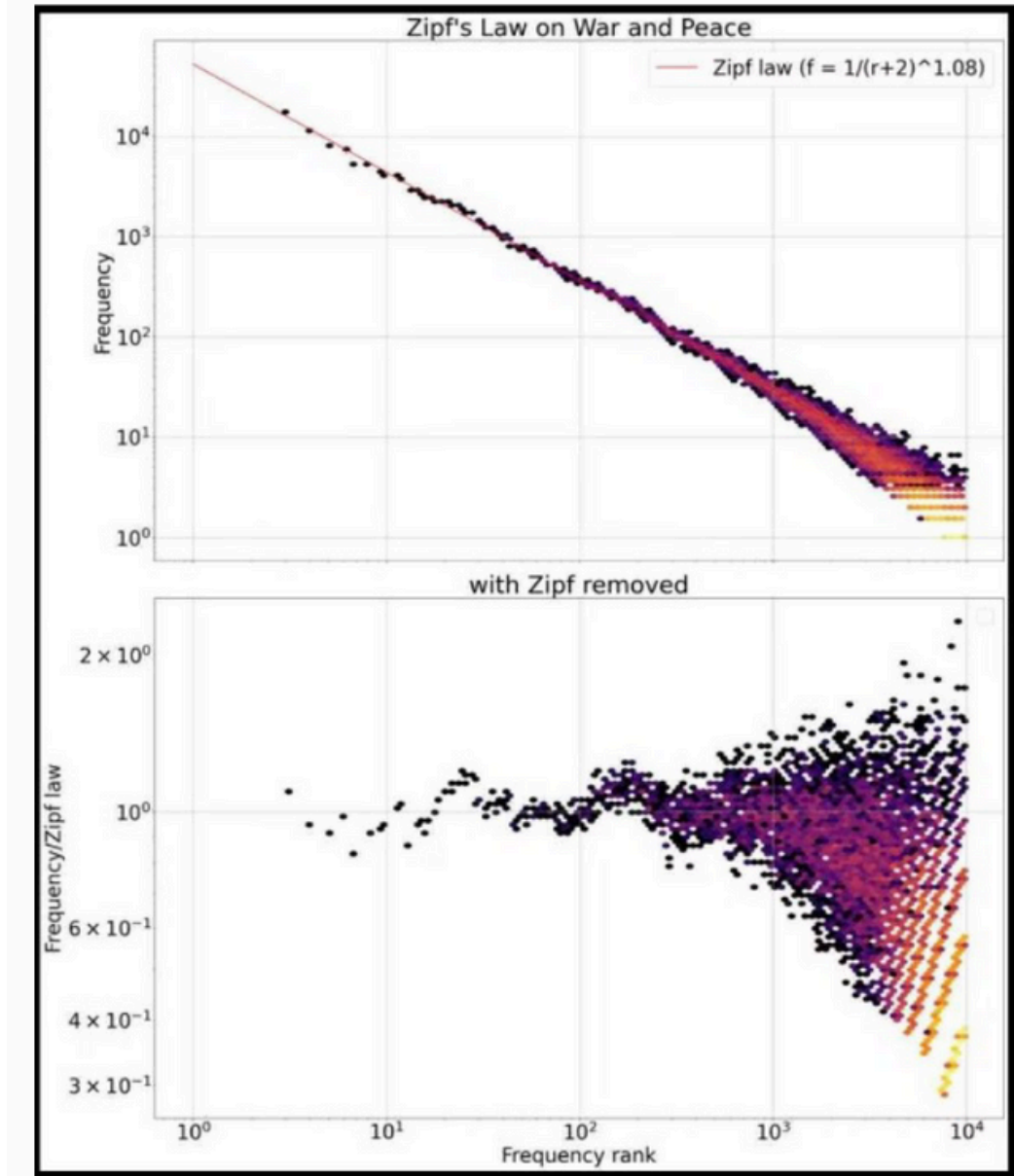
The log-log plot approximates a straight line, supporting Zipf's Law.- Lyrics slightly deviate due to artistic repetition and unique vocabulary. Deviation Factors Eminem's stylistic choices, such as use of slang and complex rhyme structures, may cause deviations. Below mentioned in the iPython notebook link are the top 50 most occurring words in order.





# Top 20 most repeating words and Zipf's Application using graph :

Top 20 Words in Lyrics:		
1.	i	17786
2.	the	11157
3.	you	10819
4.	and	8008
5.	a	7970
6.	to	7781
7.	it	6311
8.	s	5821
9.	t	4814
10.	m	4600
11.	my	4576
12.	me	4347
13.	in	4296
14.	that	3804
15.	of	3498
16.	like	3064
17.	on	2829
18.	your	2719
19.	this	2601
20.	with	2559



# Conclusion

## **1. Summary**

Eminem's lyrics largely conform to Zipf's Law, reaffirming its relevance in stylized natural language.

## **2. Implications**

This supports AI and NLP models using Zipfian assumptions in handling artistic text data.

## **3. Future Work**

Comparative study on other artists' lyrics to explore genre and personal style effects on Zipf's distribution.





THANK YOU