Zipf's Law in Song Lyrics: A Case Study on Lady Gaga

Presented by Team Hackdev:

Krit Garg - krit.garg01@adypu.edu.in

Deepak Pathik - deepak.pathik@adypu.edu.in

Harsh Hirawat - harsh.hirawat@adypu.edu.in

Harshit Singh - harshit.singh@adypu.edu.in

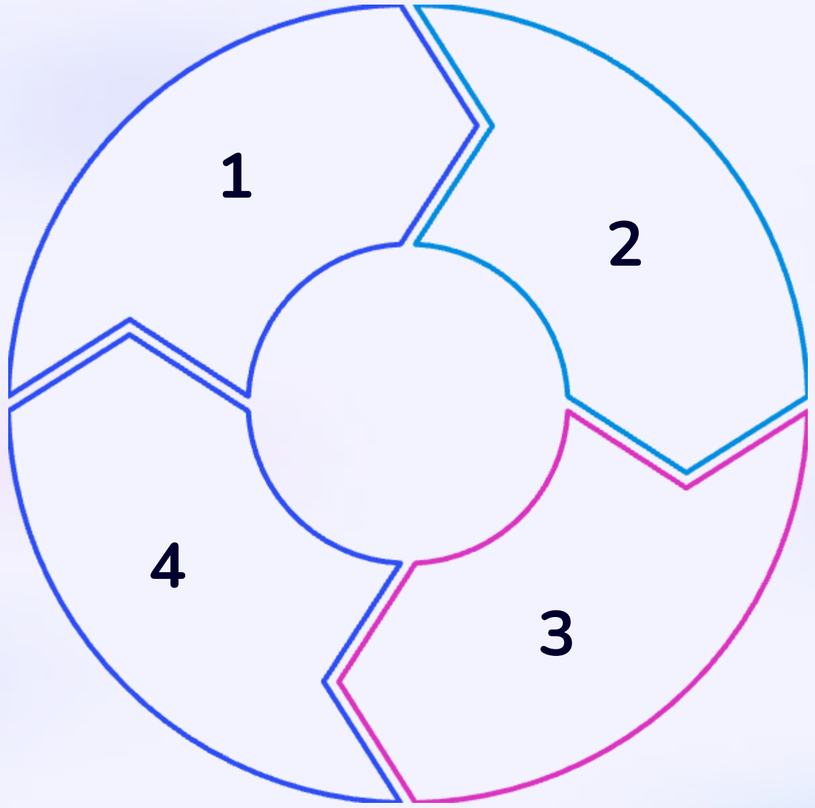
Introduction to Zipf's Law

Zipf's Law Defined

The frequency of a word is inversely proportional to its rank in a given text.

Project Objective

To analyze Lady Gaga's song lyrics through the lens of Zipf's Law.



Mathematical Formula

The frequency of the word ranked r is given by: $f(r) \propto 1/r$.

Applications

This phenomenon is observed in natural languages and various other contexts.

Dataset Overview

Source

Dataset: LadyGaga.csv

Word Counts

136,516 total words

Songs Analyzed

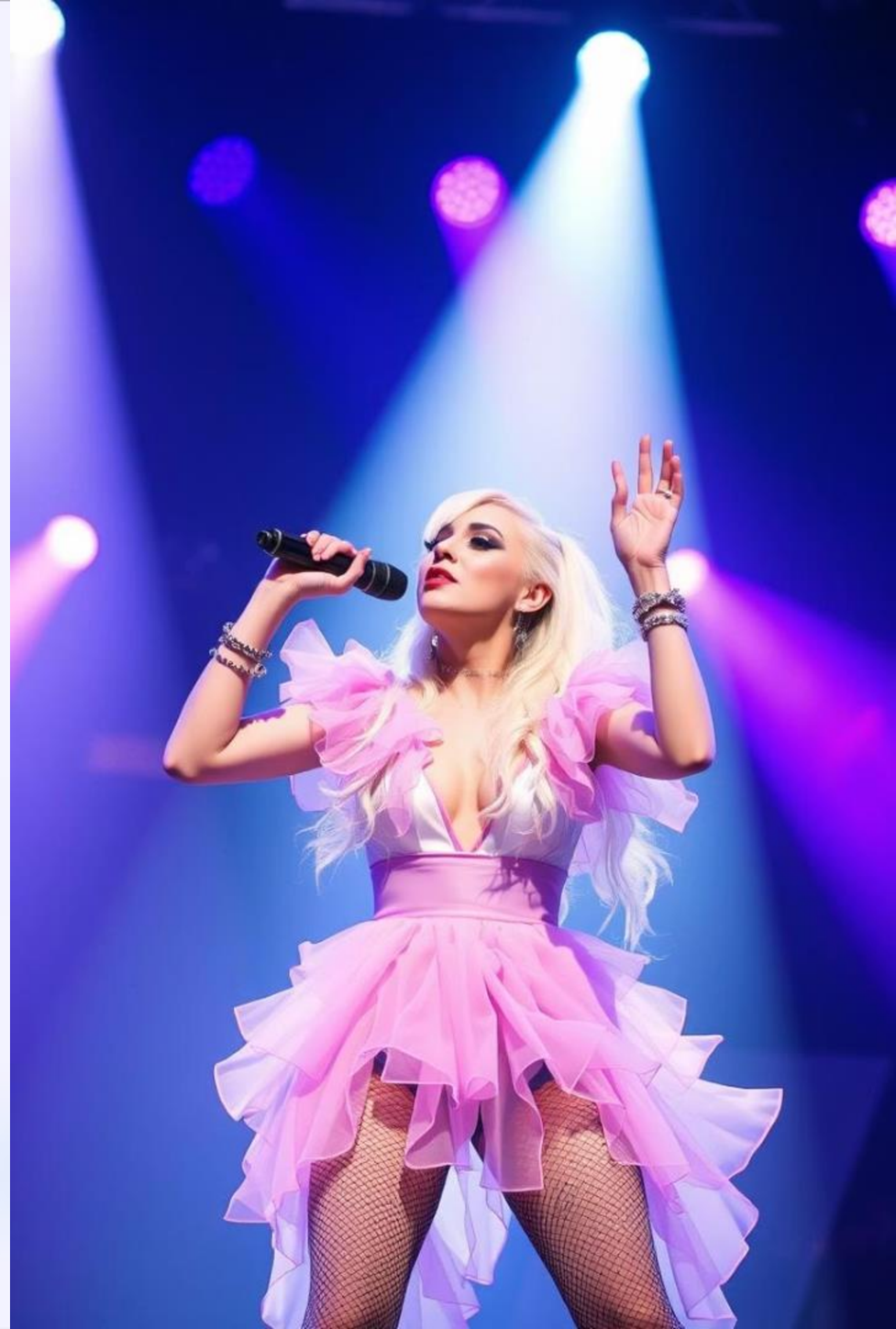
395 songs included

Unique Words

4,975 distinct words

Attributes

- Artist
- Album
- Date
- Title
- Year
- Lyric



Data Preprocessing

Normalize

Lowercase all lyrics for uniformity

Tokenize

Split text into individual words

Clean

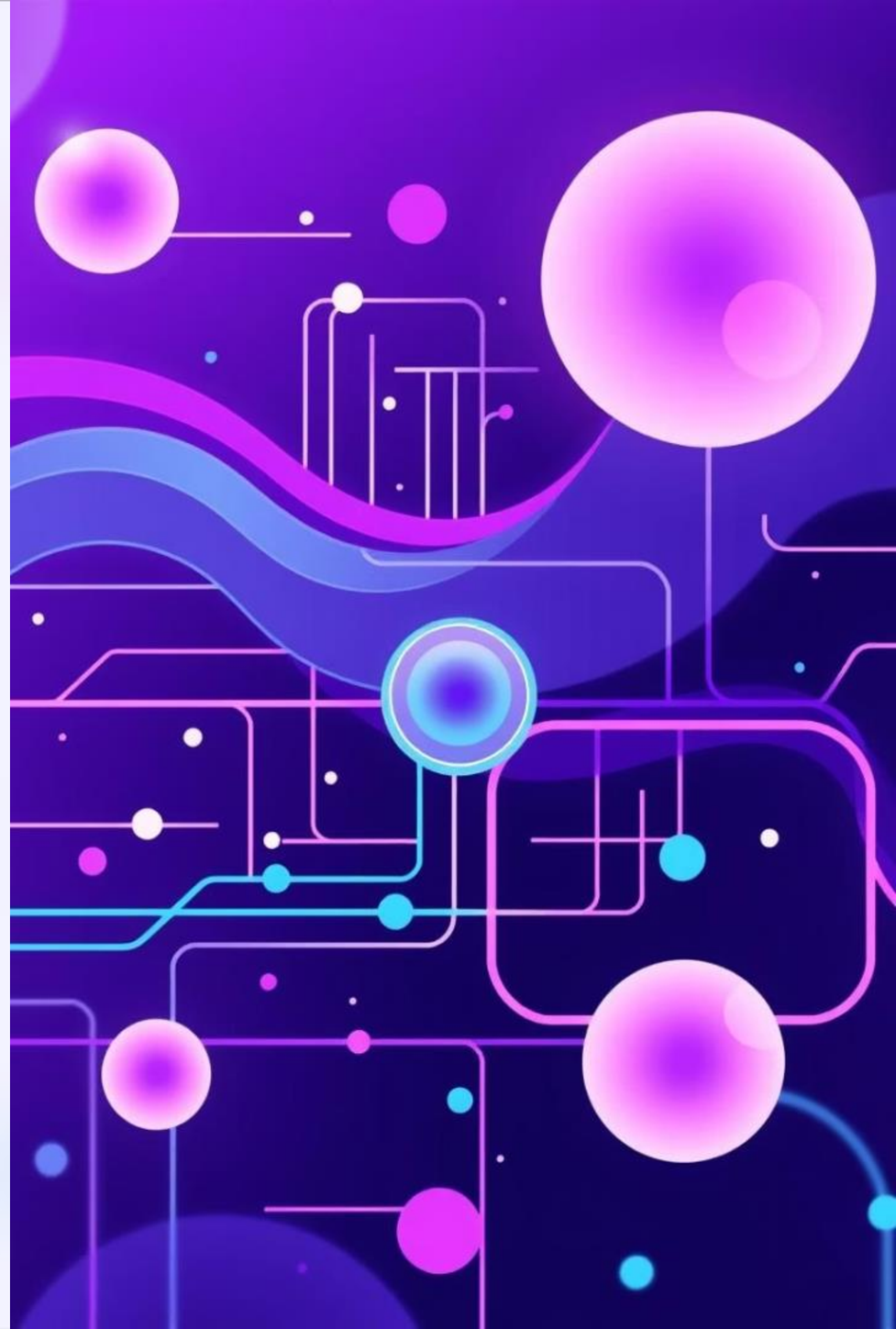
Remove empty and non-alphabetic entries

Aggregate

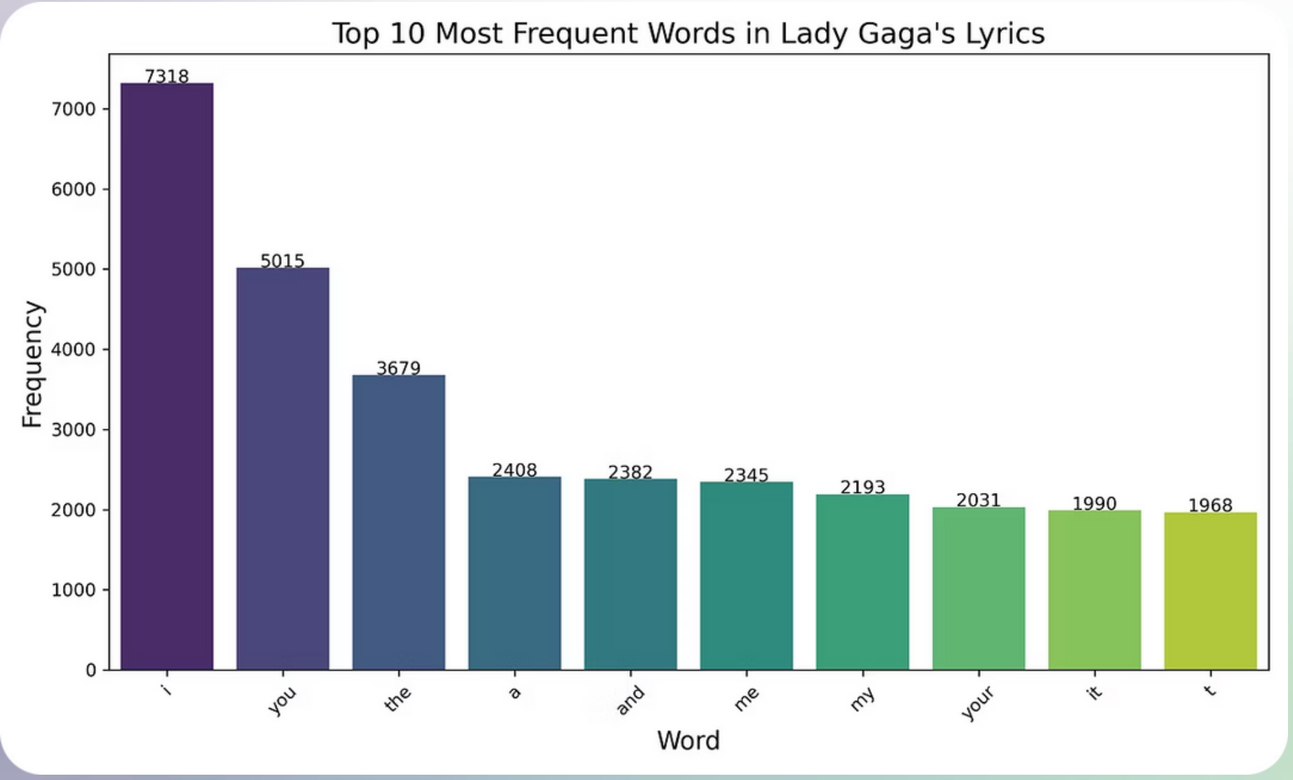
Combine lyrics into one corpus

Rank & Count

Calculate word frequencies and ranks



Frequency Analysis



Most Frequent Words

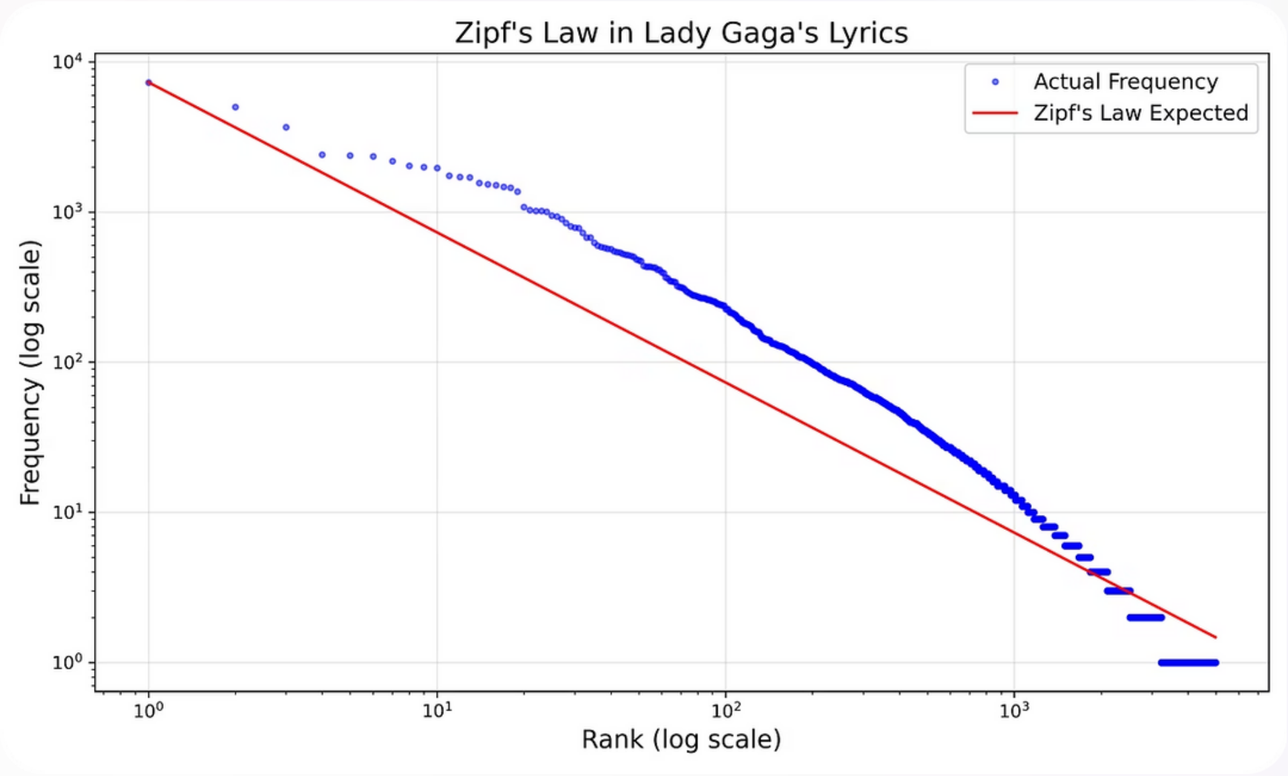
- 'I', 'you', 'love', 'like'
- Common pronouns & verbs dominate

Frequency Distribution

Top words occur thousands of times

Significance

Reflects thematic and stylistic choices



Zipf's Law Visualization

This log-log plot compares the actual word frequencies found in Lady Gaga's song lyrics with the theoretical expectations predicted by Zipf's Law. The plot visually demonstrates how word usage follows an inverse relationship between a word's rank and its frequency of occurrence. By observing the alignment between the empirical data and the Zipfian curve, we can better understand the linguistic patterns and the natural distribution of words in her lyrics.

Key Observations

Zipf's Law Fit

Lyrics closely follow Zipf's distribution overall

Dominant Words

'I', 'you', and 'love' occur most frequently

Artistic Influence

Repetition skews distribution of rare words

Pop Lyric Patterns

Lady Gaga's vocabulary is typical of pop music

Thematic Words

Certain themes appear more frequently than usual

Tools and Technologies

Python

Primary programming language for analysis

pandas

Efficient data manipulation and handling

matplotlib & seaborn

Used for creating visualizations and plots

Other Libraries

- collections.Counter for frequency counting
- Regular Expressions for text cleaning

Individual Contributions

Harsh Hirawat

Preprocessing, tokenization, frequency analysis

Deepak Pathik

Zipf's Law computations, log-log visualization

Harshit Singh

Visualization, data insights, report writing

Krit Garg

Presentation design, content structure, review

Thank You!

We appreciate your attention and interest.