# Uncovering linguistic pattern using Zipf's Law

## Team: QuadForce

Team Members:

Dhruv Sareen - dhruv.sareen@adypu.edu.in

Ayush Anand - ayush.anand@adypu.edu.in

Ved Bhadani - ved.bhadani@adypu.edu.in

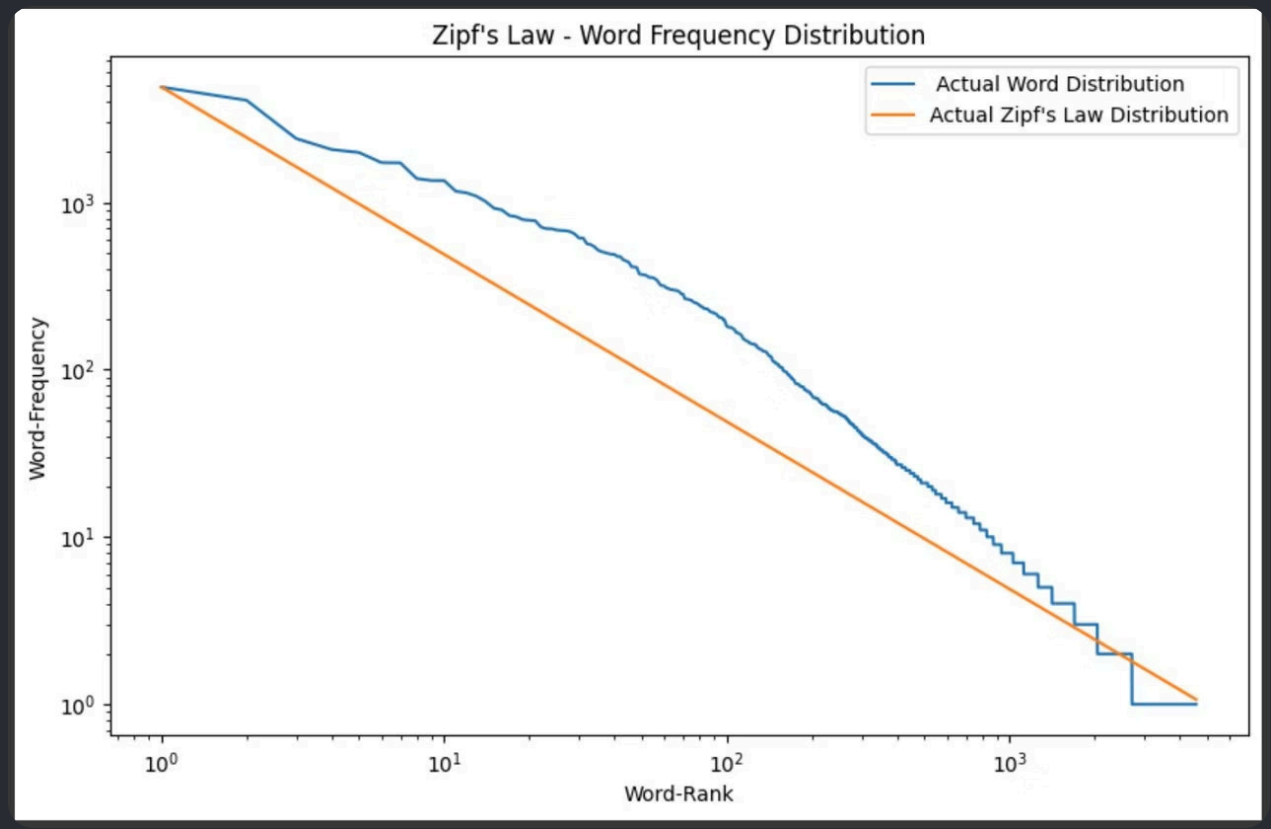
Ovais Arif Koite - ovais.koite@adyou.edu.in

# Project Overview

## Zipf's Law:

**Zipf's Law** is a statistical principle that suggests in many natural datasets, particularly in human language, the frequency of an item is inversely proportional to its rank in a frequency table. This means that the most common word in a language appears about twice as often as the second most common word, three times as often as the third, and so on. For example, in English, the word "the" is the most frequently used word, followed by "of," "and," and others, with their frequencies decreasing roughly in proportion to their rank. This pattern, observed by linguist George Zipf, appears not only in languages but also in phenomena like city populations, website traffic, and income distributions, highlighting a common structure in complex systems.

## Result & Findings




- The graph uses a **log-log scale** to show word rank vs. frequency.
- The **orange line** represents the ideal Zipf's Law (perfect inverse relation).
- The **blue line** is the actual word frequency from real text.
- Real data **closely follows Zipf's Law** but deviates slightly.
- **Top-ranked words** are more frequent than predicted.
- **Low-ranked words** drop off faster than the ideal curve.
- Overall, **natural language exhibits a Zipfian-like distribution**.

# Individual Contributions

| Team Member | Key Contributions |
| --- | --- |
| **Dhruv Sareen** | **- Implemented Zipf's Law in pure Python**<br>**- Created the log-log word frequency plot**<br>**- Cleaned and preprocessed lyrics data** |
| **Ayush Anand** | **- Created bar chart of songs per year**<br>**- Handled EDA visualization**<br>**- Analyzed results and wrote insights** |
| **Ovais Arif Koite** | **- Identified long-tail and anomalies in word use**<br>**- Contributed to insights section** |
| **Ved Bhadani** | **- Structured and formatted the notebook**<br>**- Made a full presentation on Zipf's Law, wrote every team member contributions.** |

# Thank You