🎶 Zipf's Law Meets Pop: Exploring Word Frequencies in Justin Bieber's Lyrics

Team Name : Pani_pee_liya

- Archisman Nath Choudhury | 2024-B-29012005
- Yash Pratap Singh Solanki | 2024-B-18112003
- MILAN KUMAR | 2024-B-12092004
- Aaryan Sahu | 2024-B-18092006

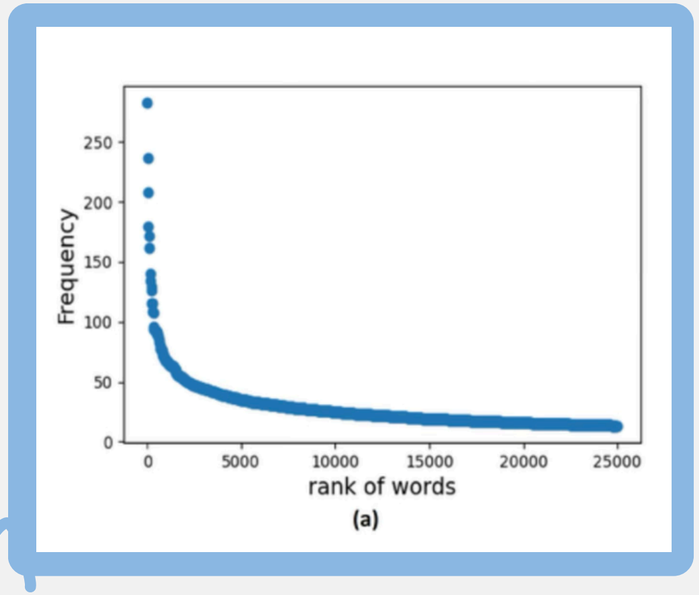
Zipf's Law :-

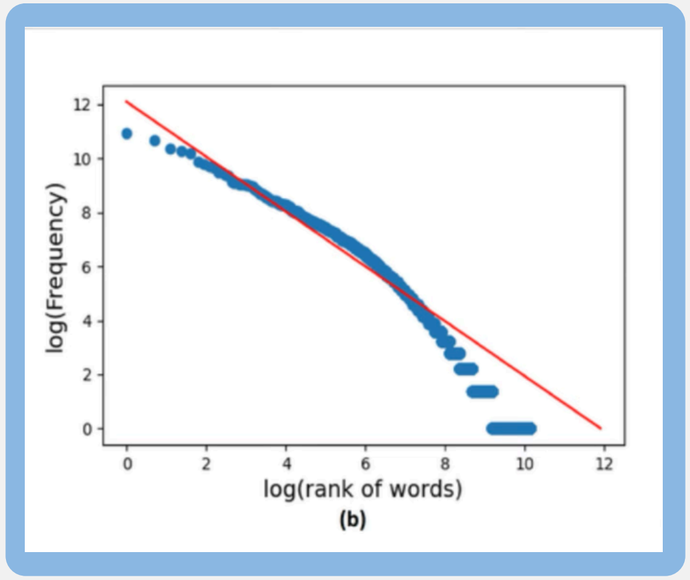
- States that in any large sample of natural language, the frequency of any word inversely proportional to its rank.
- The 2nd most frequent word occurs \~half as often as the most frequent, and so on.
- Can this mathematical law apply to something as artistic as pop lyrics ???

Objective

To examine if Justin Bieber's lyrics follow Zipf's Law, analyze whether word frequency is inversely proportional to its rank in the frequency table.

- If we plot word frequency versus rank, we should observe a rapidly declining curve.
- If we plot the same data on a log-log scale, the points should form a roughly straight line — indicating a power-law distribution.





BASIC EDA OF OUR DATASET

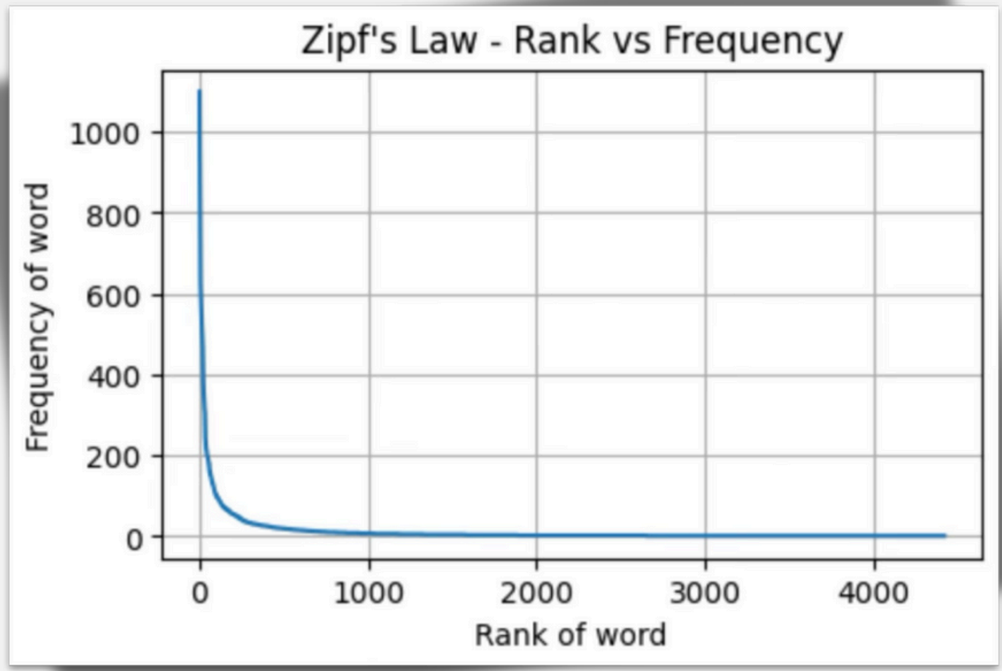
- Top 10 frequent words:

```
[('love', 1099),
 ('girl', 977),
 ('know', 893),
 ('like', 842),
 ('baby', 794),
 ('bieber', 641),
 ('go', 627),
 ('youre', 615),
 ('one', 608),
 ('never', 578)]
```

- 347 songs analyzed.
- Over 57,235 total words.
- 4420+ unique words.

Zipf's Law Visualization

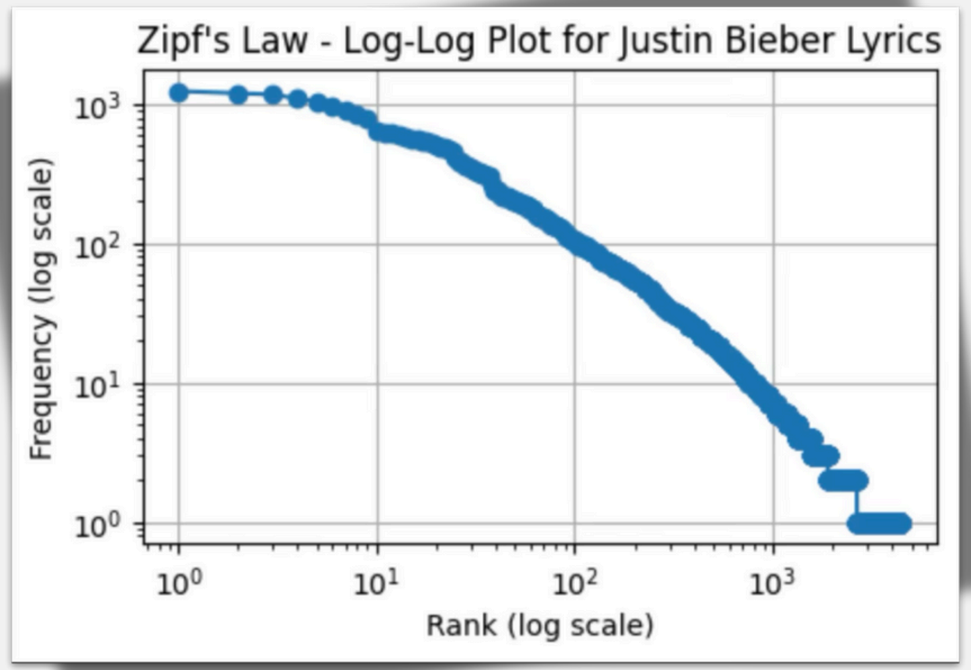
- Graph of actual word frequency vs rank :



- Curve demonstrates steep drop initially, then gradual tail.

- “The data exhibits a near-Zipfian distribution, typical of natural language.”

INSIGHTS



- Common emotional themes: love, loss, self-worth.
- High rank words dominated by stopwords.
- Lyrics reflect human language patterns, even in music.

Challenges

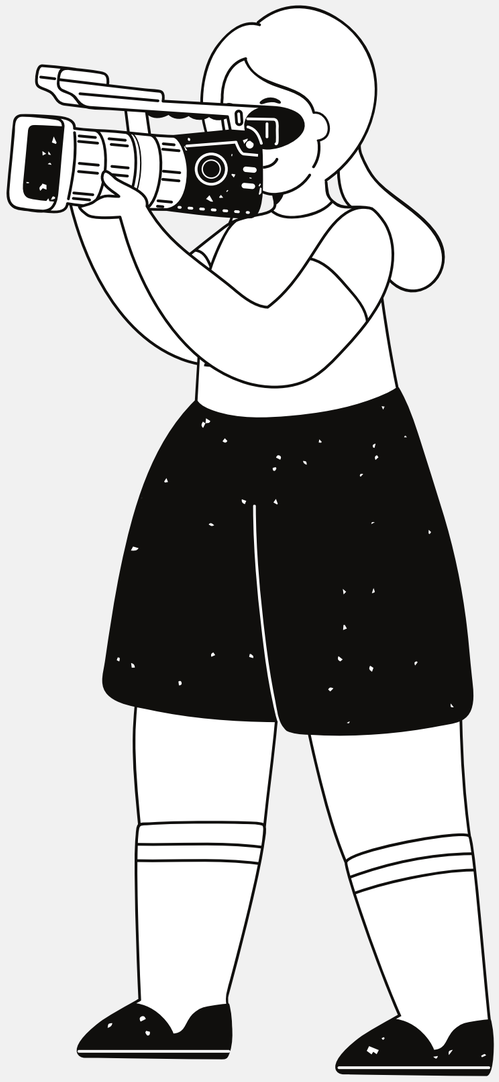
Cleaning lyrical text and removing noise

Handling stopwords and contractions (e.g., I'm, don't)

Learnings

Natural language follows beautiful mathematical patterns

Hands-on Zipf's Law implementation & EDA



Roles and Responsibilities

• **Data Exploration and Visualization Lead**	Yash Pratap Singh Solanki
• **Implementation and Coding Expert**	Archisman Nath Choudhury
• **Content and Research Coordinator**	Aaryan Sahu
• **Presentation and Documentation Designer**	Milan Kumar

Thank you very much!

We loved exploring Zipf's Law through pop music

We would love to have your feedback.