RIHANNA.CSV

Proof by Zipf’s Law

Team name : πrates911

Team members-

1. Soham Saranga
2. Subhan Kumar Rai
3. Siddharth Pareek
4. Shorya Taneja

WHAT IS ZIPF'S LAW?

- ZIPF'S LAW IS A STATISTICAL PRINCIPLE THAT SAYS:

IN A GIVEN DATASET OF NATURAL LANGUAGE, THE FREQUENCY OF ANY WORD IS INVERSELY PROPORTIONAL TO ITS RANK IN THE FREQUENCY TABLE

- **IN SIMPLE TERMS:**

THE MOST FREQUENT WORD APPEARS TWICE AS OFTEN AS THE SECOND MOST FREQUENT WORD,

THREE TIMES AS OFTEN AS THE THIRD MOST FREQUENT, AND SO ON.

FORMULA: FREQUENCY $\propto$ 1 / RANK

Borcelle

CODE BLOCK


```
[ ] from google.colab import drive
    drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
[ ] fileLoc = "drive/MyDrive/MATHS.csv"
```

```
import pandas as pd
import matplotlib.pyplot as plt
import re
from collections import Counter

df = pd.read_csv(fileLoc)
```

```
lyrics_text = " ".join(df['Lyric'].dropna().astype(str)).lower()

words = re.findall(r'\b\w+\b', lyrics_text)

word_freq = Counter(words)
frequencies = sorted(word_freq.values(), reverse=True)

ranks = range(1, len(frequencies) + 1)

plt.figure(figsize=(10, 6))
plt.loglog(ranks, frequencies, marker='.', linestyle='none', color='blue')
plt.title("Zipf's Law - Word Frequency vs Rank (Rihanna Lyrics)")
plt.xlabel("Rank of Word (log scale)")
plt.ylabel("Frequency of Word (log scale)")
plt.grid(True, which="both", ls="--")
plt.tight_layout()
plt.show()
```

// INSIGHTS ON CODE IMPLEMENTATION

🎯 Slide 3: Code Implementation – Visualizing Zipf's Law

📥 Step 1: Load Dataset

- The CSV file (MATHS.csv) containing Rihanna's song lyrics is loaded from Google Drive using pandas.

🔤 Step 2: Prepare the Text

- All lyrics are combined into one long lowercase string.
- re.findall() extracts all words using regular expressions.

📊 Step 3: Count Word Frequencies

- collections.Counter counts occurrences of each word.
- Frequencies are sorted in descending order.

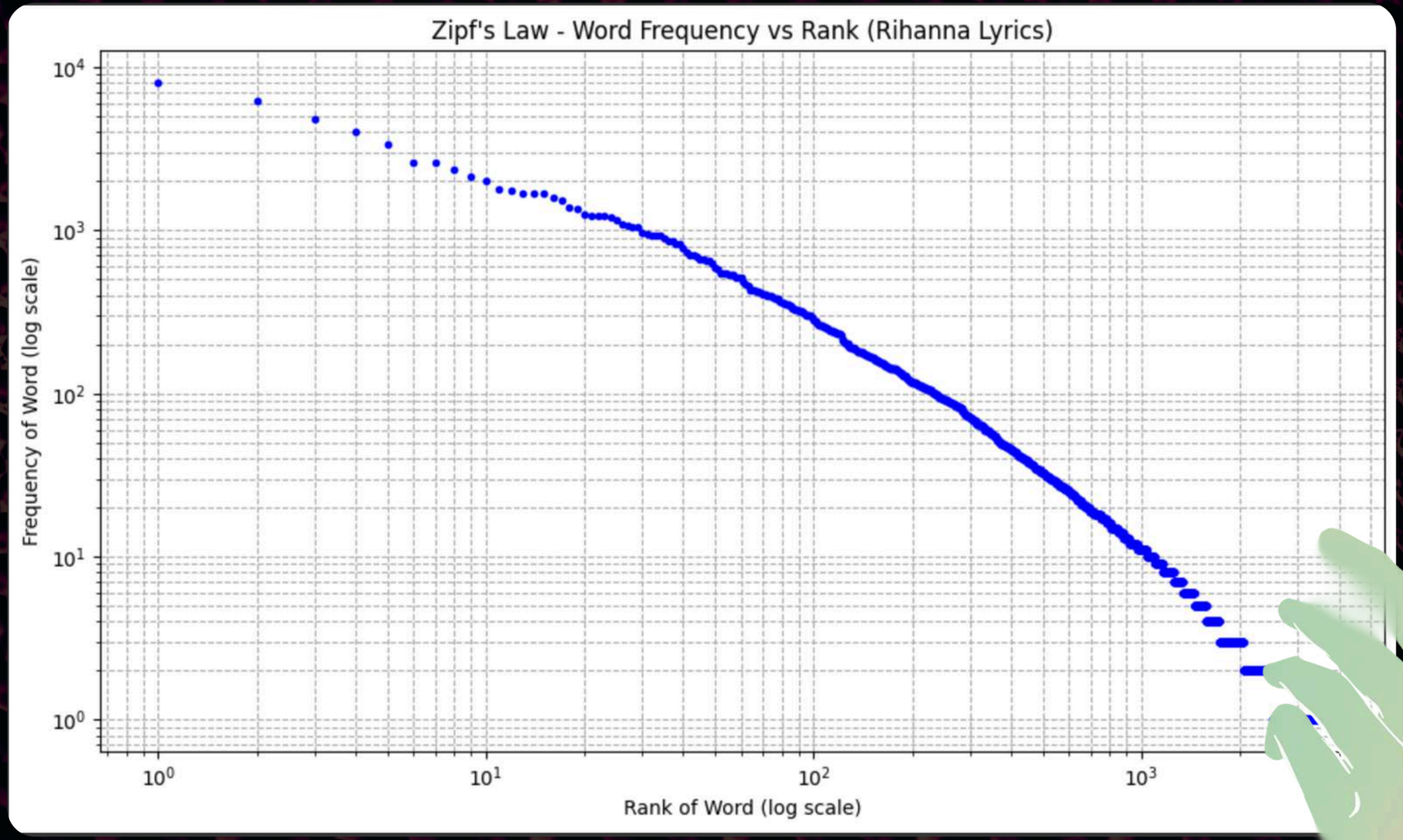
📈 Step 4: Plot Zipf's Law

- Ranks are assigned to each unique word by frequency.
- A log-log plot is created using matplotlib to show:
  - X-axis: Word rank
  - Y-axis: Word frequency
- This visually tests Zipf's Law on Rihanna's lyrics.

🧠 Result:

The plot shows a typical downward curve, confirming that Rihanna's lyrics follow Zipf's distribution pattern.

GRAPH OUTPUT



INSIGHTS FROM GRAPH

- EVIDENCE OF ZIPF’S LAW:

THE GRAPH EXHIBITS A POWER-LAW DISTRIBUTION TYPICAL OF NATURAL LANGUAGE.

THIS CONFIRMS THAT RIHANNA'S LYRICS FOLLOW THE LINGUISTIC FINGERPRINT SEEN IN MOST NATURAL TEXTS.

- SONGWRITING PATTERNS:

A SMALL NUMBER OF COMMON WORDS (LIKE "I", "YOU", "LOVE") APPEAR VERY FREQUENTLY.

UNIQUE OR RARE WORDS APPEAR ONLY A FEW TIMES, LIKELY FOR EMOTIONAL OR CREATIVE EMPHASIS.

- IMPLICATION:

THIS STRUCTURE REFLECTS HOW HUMAN COMMUNICATION IS OPTIMIZED:
FREQUENT WORDS MAKE EXPRESSION EFFICIENT, WHILE RARE WORDS MAKE IT MEANINGFUL

TEAM MEMBERS AND CONTRIBUTION

1. **Soham Jayesh Saranga :** I handled the visualization of word frequency data by generating a log-log plot to demonstrate the relationship between word rank and frequency. This helped us verify the presence of Zipf’s Law in Rihanna’s lyrics.

2. Subhan Kumar Rai : I was responsible for preparing and cleaning the lyrics data, extracting words from the text, and calculating word frequencies. This laid the groundwork for analyzing the linguistic distribution in Rihanna’s lyrics as required by Zipf’s Law.

TEAM MEMBERS AND CONTRIBUTION

3. Shorya Taneja : I focused on explaining the core concept of Zipf’s Law in a way that is easy to understand, using real-world examples and visual metaphors. I also created the introductory slides and theory section to set the foundation for how Zipf’s Law connects to song lyrics.

4. Siddharth Pareek : I worked on designing the visual output slides, including the log-log graph, its interpretation, and insights. I helped translate the results into meaningful takeaways about Rihanna’s lyric patterns, emphasizing how Zipf’s Law appears in real music data.

CONCLUSION

“Even pop lyrics obey mathematical patterns.”

Zipf's Law reveals hidden order in human expression.
With simple tools (Pandas + Matplotlib), we uncovered a fundamental rule of language encoded in music.