Mathematics Project on Selena Gomez Songs Lyrics Dataset

This project explores Selena Gomez's song lyrics dataset using advanced mathematical and data science techniques. The team analyzed the dataset to reveal hidden patterns and applied Zipf's law to understand linguistic distribution.

Team name: data detectives

Team: Pranav Kumar Bandaram, Aditya Raj Srivastava, Suraj Kewat, Siddu Hamigi

Exploratory Data Analysis Overview

Exploratory Data Analysis (EDA)

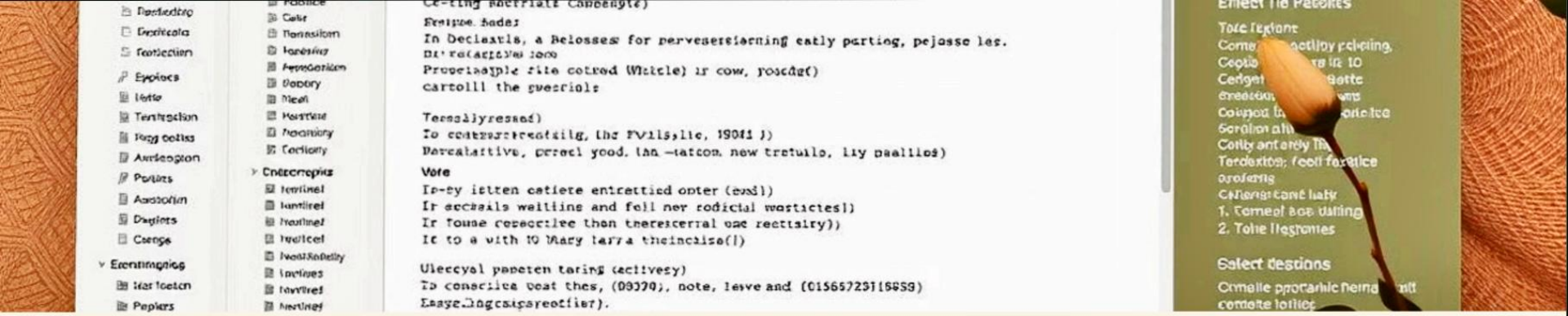
Initial dataset inspection and variable analysis. Data cleaning and missing value detection using missingno.

- Identified key variables in lyrics
- Handled missing and incomplete data
- Uncovered distributions and trends

Tools Used

Python, Jupyter Notebook with libraries for thorough analysis and visualization.

- pandas, numpy for data processing
- matplotlib, seaborn for graphing
- missingno for missing data visualization



Python Environment & Dataset Preparation

Setup & Installation

Installed all necessary packages through terminal including Jupyter Notebook.

Dataset Preprocessing

Cleaned and formatted the lyrics dataset using pandas for accurate analysis.

Team Coding Collaboration

Collaborative Python coding workflow to prepare dataset for visualization and analysis.

Data Visualization Techniques

Graph Types Used

- Bar charts to show word frequencies
- Line graphs to capture trends over time
- Heatmaps for missing data display

Libraries Applied

- Matplotlib for basic charts
- Seaborn for enhanced statistical visuals
- Missingno highlights data completeness

Zipf’s Law Application

Zipf's Law Explored

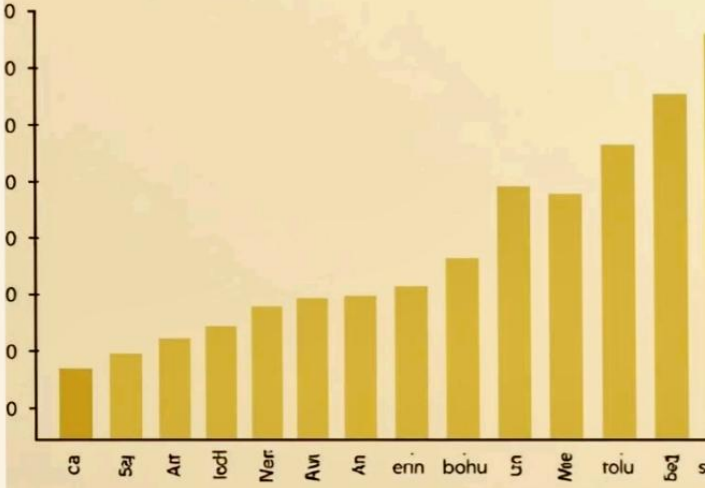
Applied Zipf's law to check if word frequencies follow expected distribution.

Findings

Lyrics generally obey Zipf's law, confirming typical linguistic patterns.

Statistical Insights

Deviations analyzed for artistic or thematic influences in lyrics.



Project Analysis and Results

Key Observations

- Frequent words highlight song themes
- Word usage varies across albums
- Patterns suggest artistic evolution

Data Interpretation

Visualizations and statistics provide deep understanding of lyrical content.

Team Member Contributions

Pranav Kumar Bandaram	Visualization and graphical representation
Aditya Raj Srivastava	Documentation and presentation preparation
Suraj Kewat	Statistical analysis and Zipf’s law application
Siddu Hamigi	Dataset preprocessing and Python coding

Thank You

We appreciate your attention and interest in our mathematical analysis of Selena Gomez's lyrics.

We welcome questions and discussions on the methodology and findings.