

The background features several abstract geometric shapes. In the top left, there is a thin, dark grey curved line. In the top right, there is a large, dark grey circle. In the bottom left, there is a large, dark grey circle. In the bottom right, there is a thin, dark grey curved line. The main title is centered in a large, bold, dark grey sans-serif font.

LYRICS ANALYSIS

Tamer Hosny


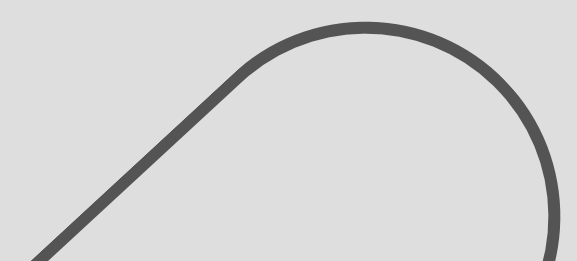
TEAM PROJECT

- 1 Mahmoud Rady
- 2 Youssef Mahmoud
- 3 Hamza Hesham
- 4 Mohamed Soffar

Under the supervision of: Dr. Walid Gomaa



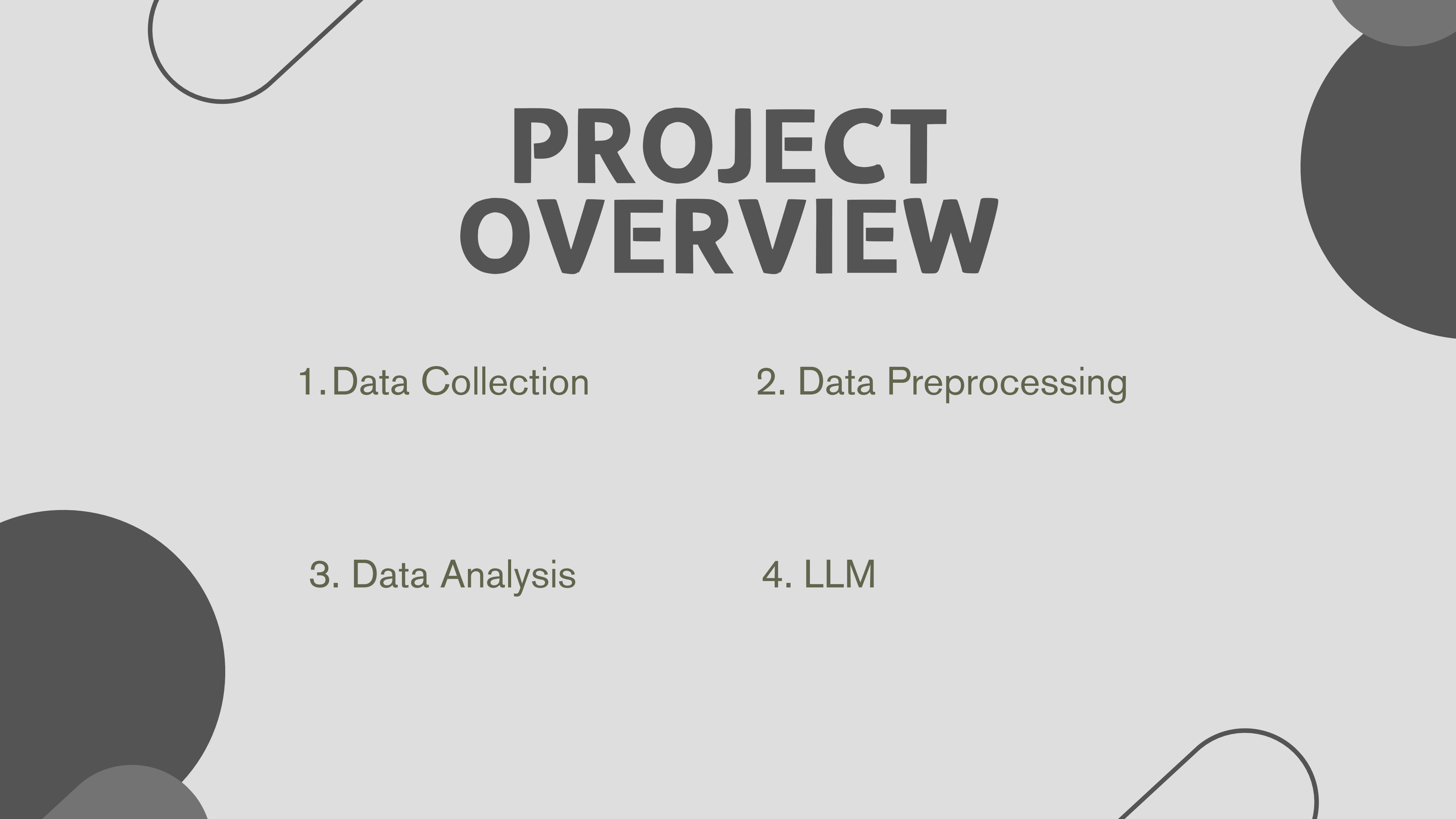
INTRODUCTION

- This project applies deep learning and natural language processing to analyze the lyrics of Tamer Hosny, a prominent figure in Arabic pop music.
 - The goal is to uncover patterns in emotion, sentiment, and collaboration within his songs.
 - By leveraging data-driven techniques, we aim to gain new insights into the artistic and linguistic features that define his work.
- 
- 



BACKGROUND

Tamer Hosny is a leading figure in Arabic pop music, known for his emotionally resonant lyrics and influential collaborations. Understanding the patterns and themes in his lyrics can provide deeper cultural and artistic insights.



PROJECT OVERVIEW

1. Data Collection

2. Data Preprocessing

3. Data Analysis

4. LLM

DATA COLLECTION

- Dataset Size:
 - Collected a total of 140 songs by Tamer Hosny.
- Diversity of Contributors:
 - 31 different composers
 - 54 different lyricists
 - 3515 Distinct words
- Data Source & Method:
 - Song lyrics and metadata were acquired using web scraping techniques from Genius.com, a widely used platform for music lyrics.

DATA PREPROCESSING

- Key Steps:
 - Text Normalization
 - Stopword Removal
 - Tokenization

TEXT NORMALIZATION

- Removed diacritics, punctuation, and numbers using arabicStemR.
- Standardized Arabic character variations (e.g., ه → ة, ي → ي) to resolve ambiguities.
- Applied CAMEL-Tools for advanced normalization (Unicode, Alef/Yeh unification).

STOPWORD REMOVAL

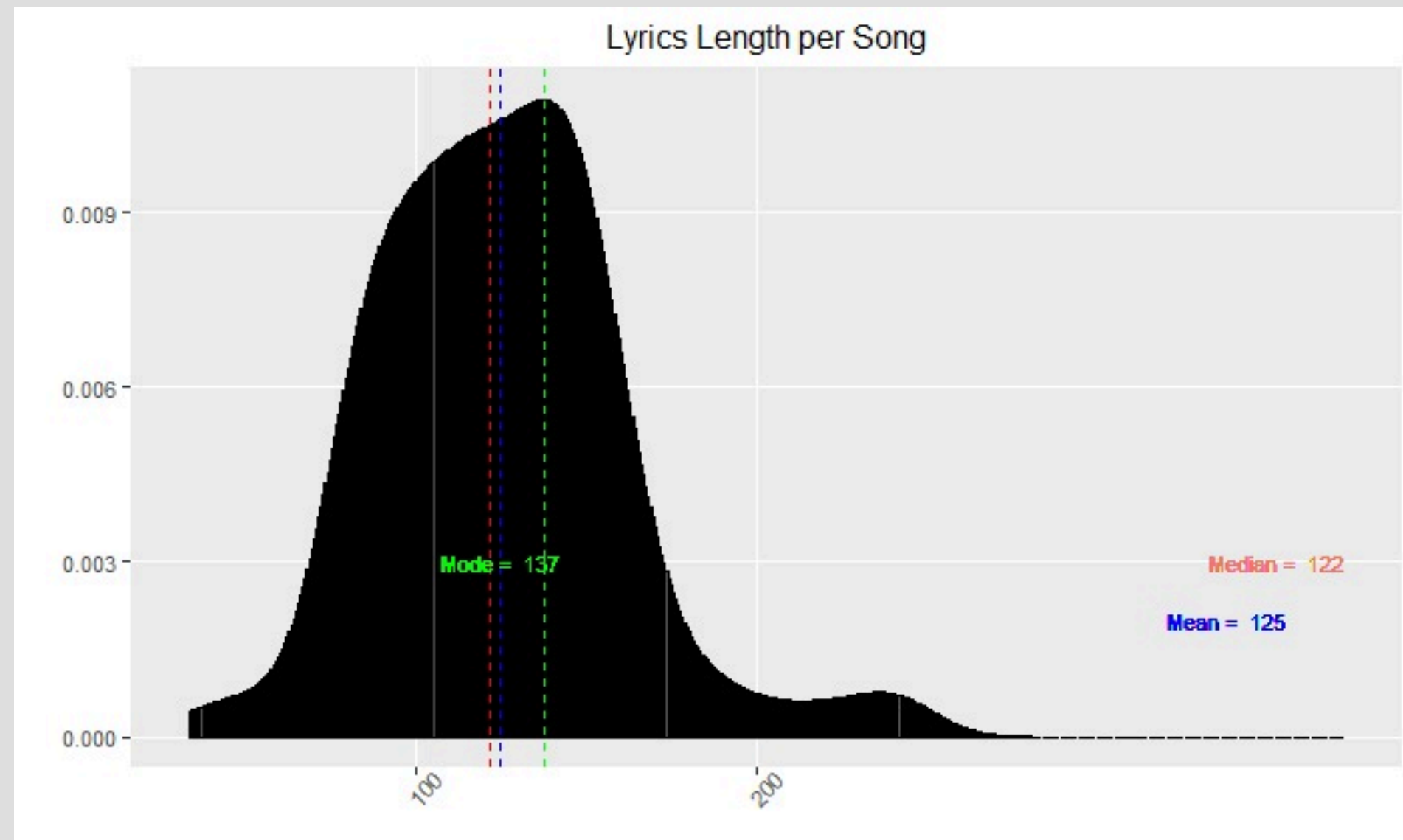
- Filtered 800+ Arabic stopwords (e.g., لا, لأن) and dialect-specific terms using a custom list.
- The custom list was a combination of most well-known Arabic Stopwords Libraries: NLTK, AraNLP.

TOKENIZATION

- Split lyrics into 1-grams using tidytext in R.
- Handled compound Arabic phrases (e.g., "على قد") as single tokens where applicable.

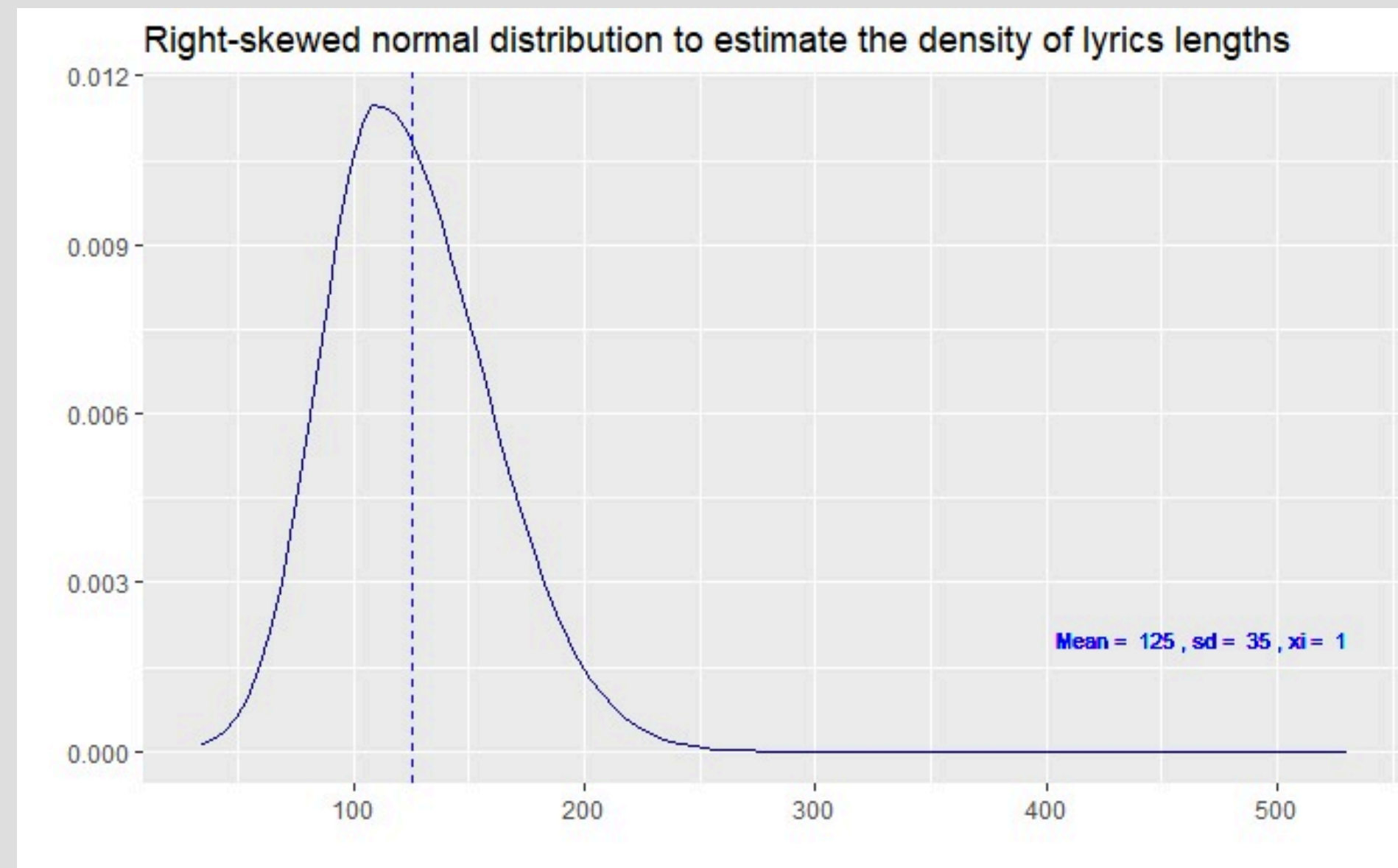
DATA ANALYSIS

- Mean, Mode & Median for Lyrics Length per song.



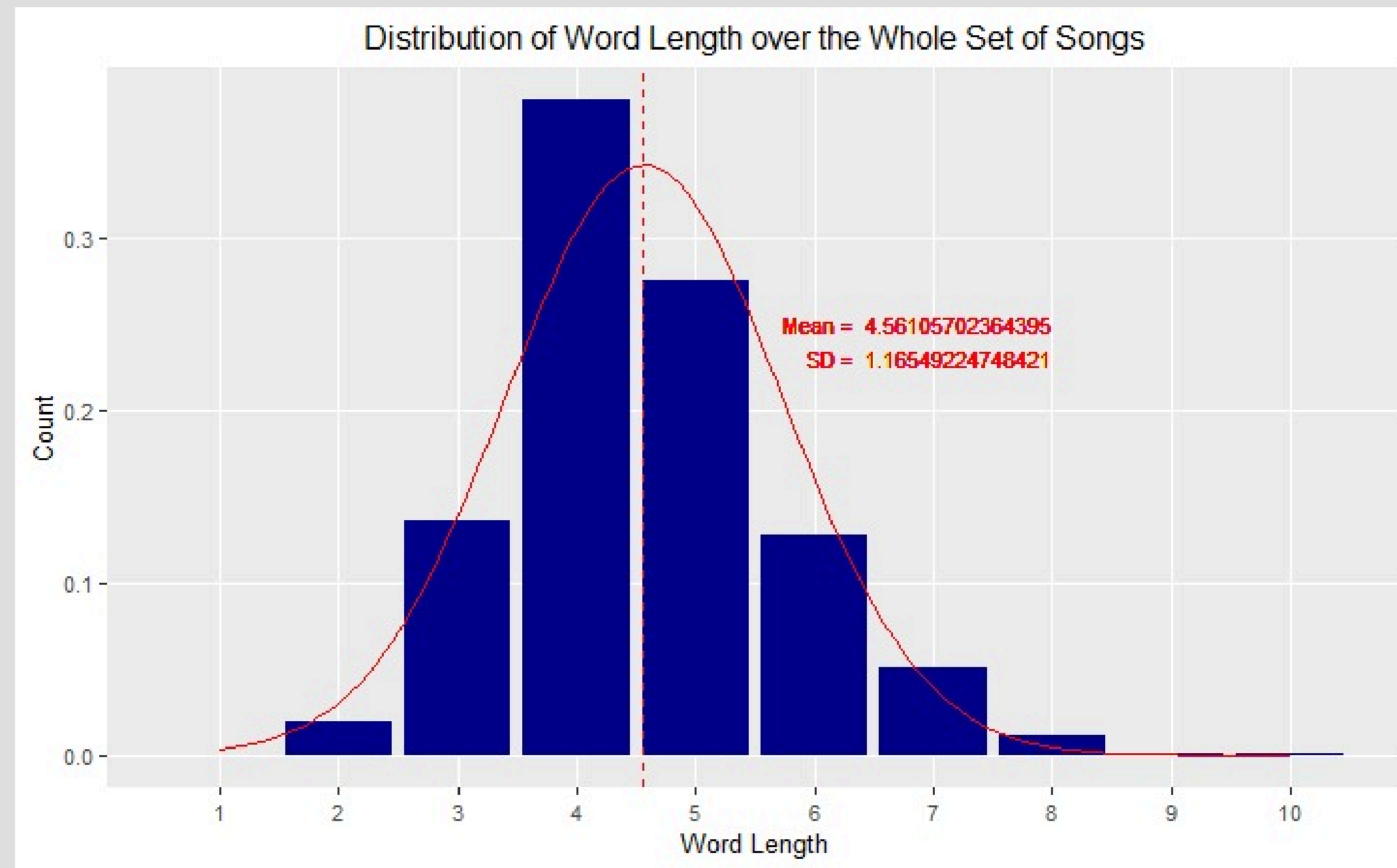
DATA ANALYSIS

Estimating Density of Lyrics Length:



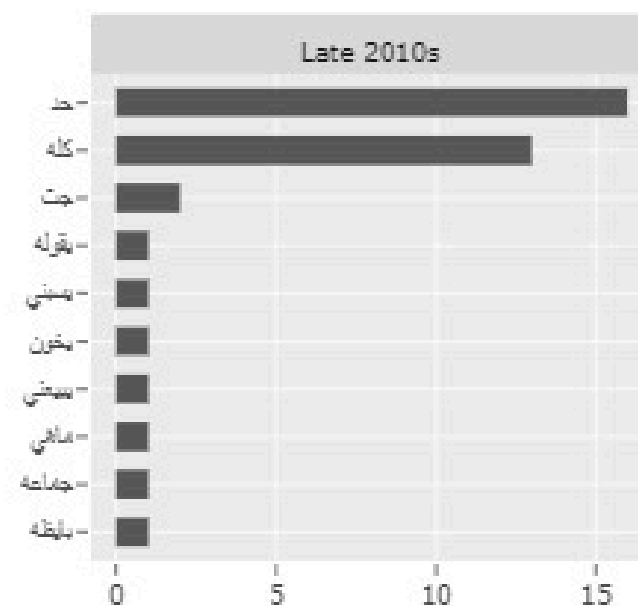
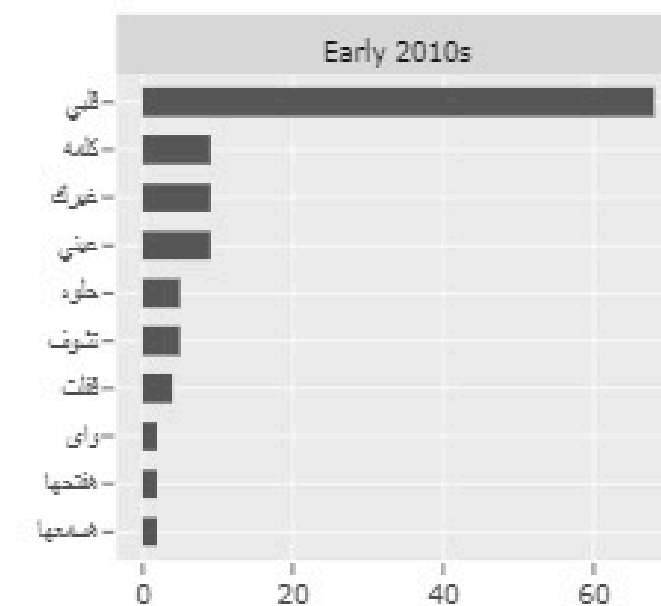
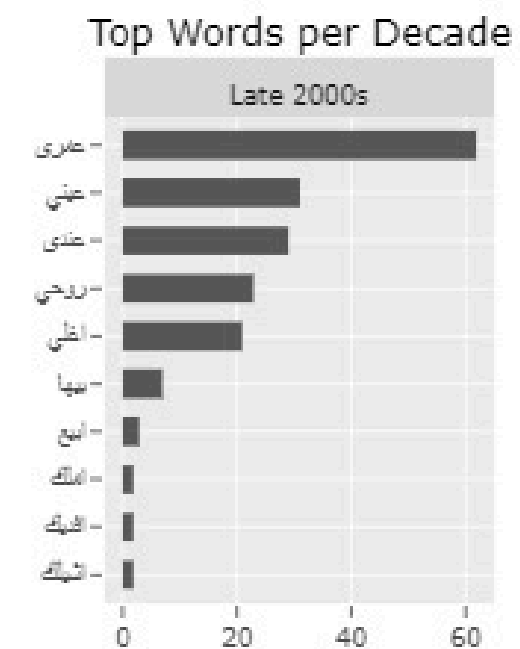
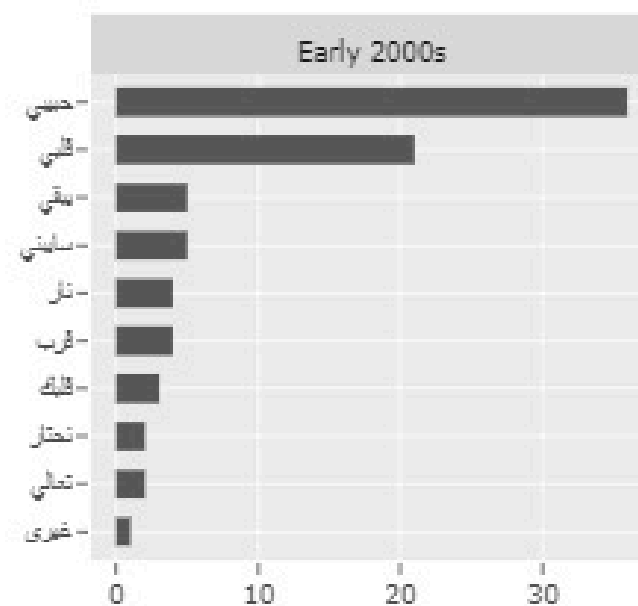
DATA ANALYSIS

Distribution of Word Length:



DATA ANALYSIS

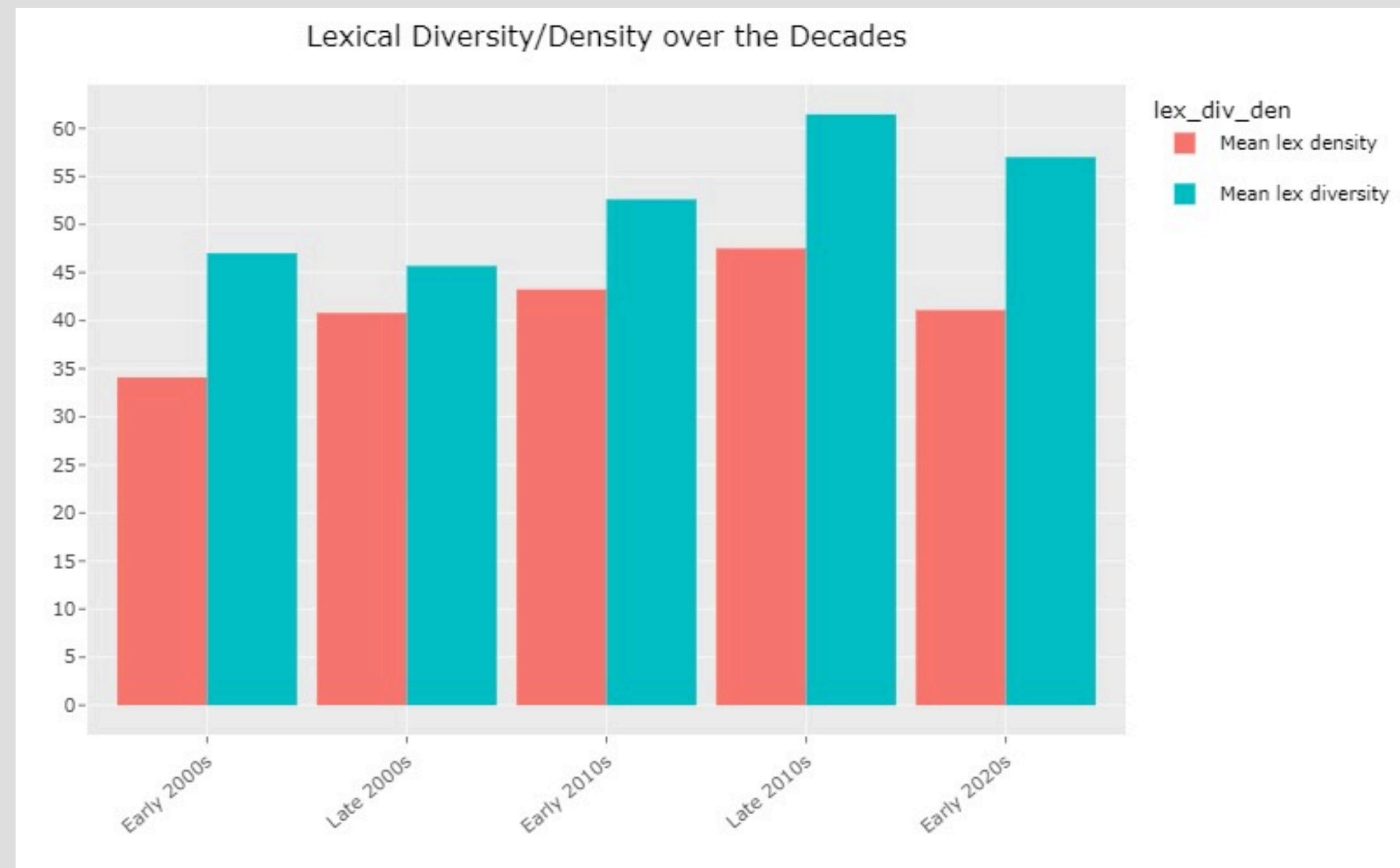
Top words per Decade:



Word Count

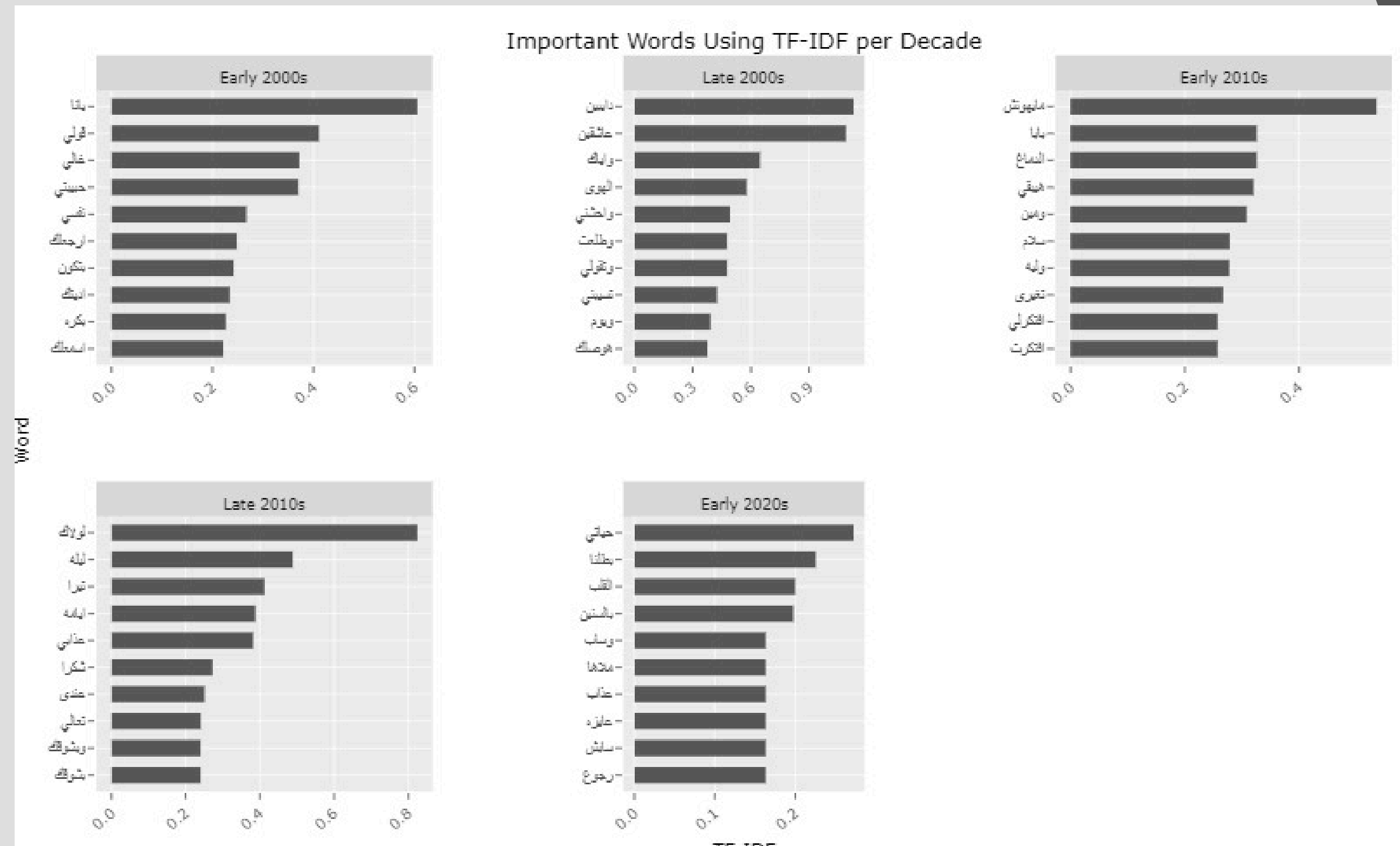
DATA ANALYSIS

Lexical Diversity/ Density over a Decade:



DATA ANALYSIS

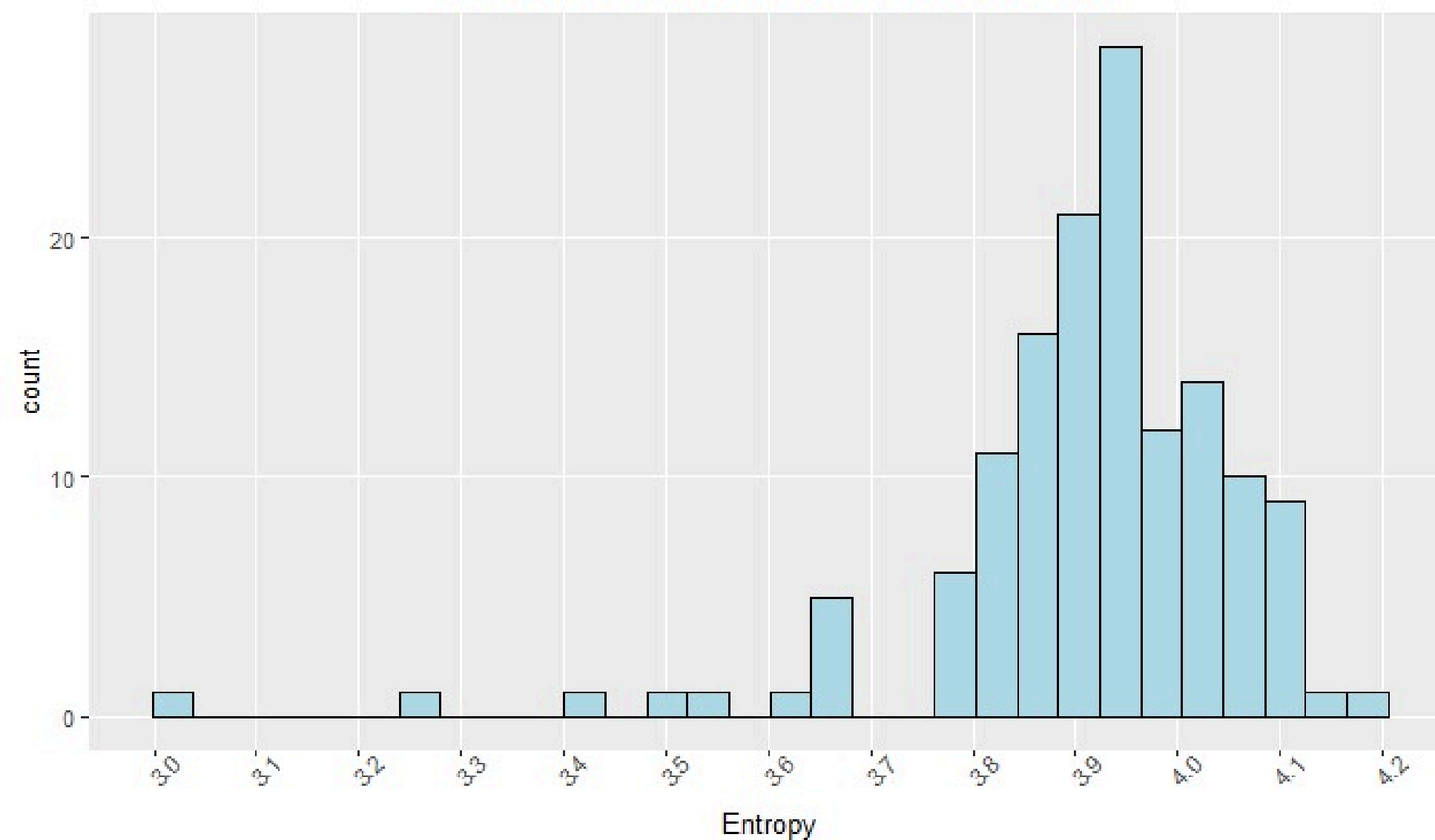
TF-IDF Results:



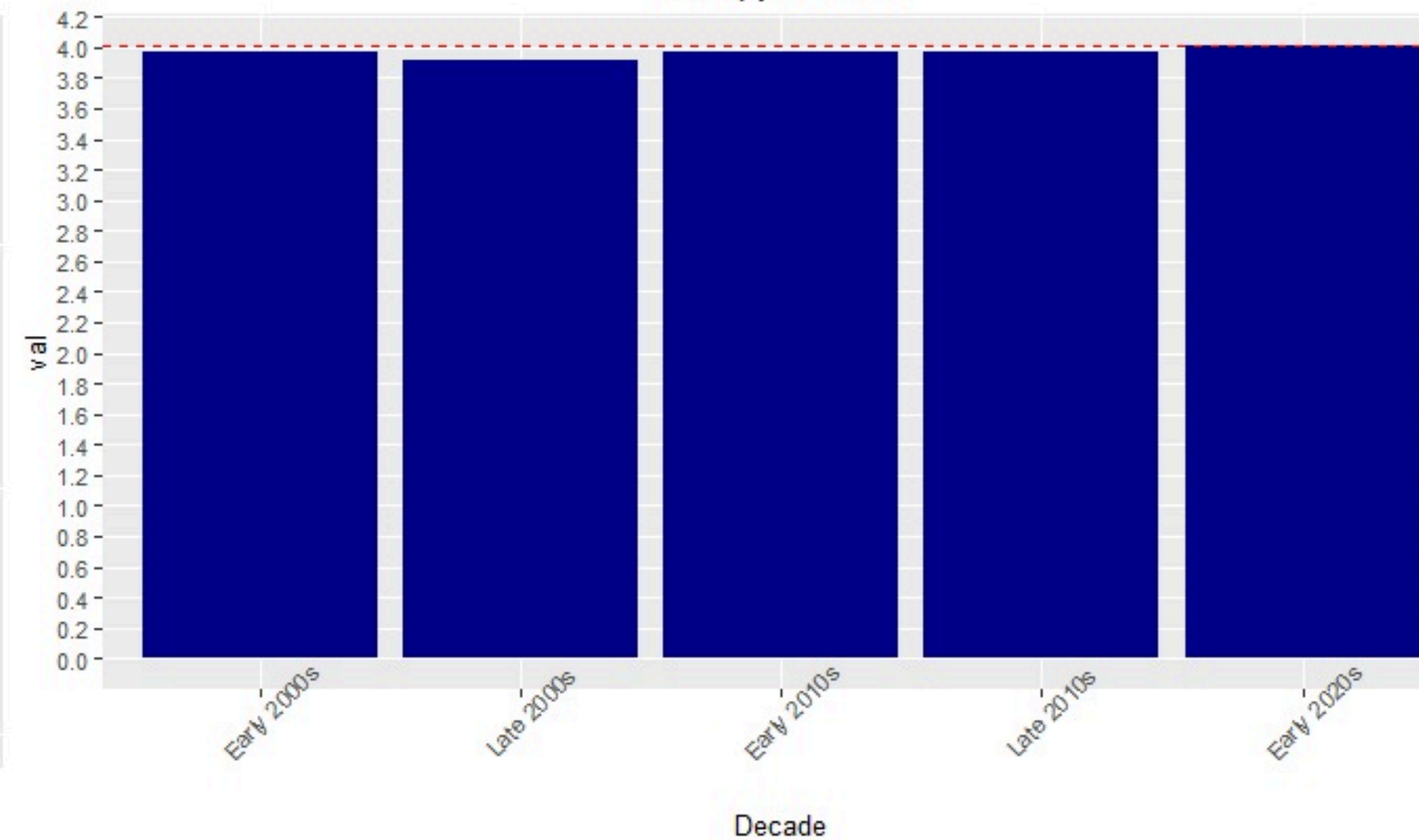
DATA ANALYSIS

Entropy Results:

Histogram of entropy values of individual songs

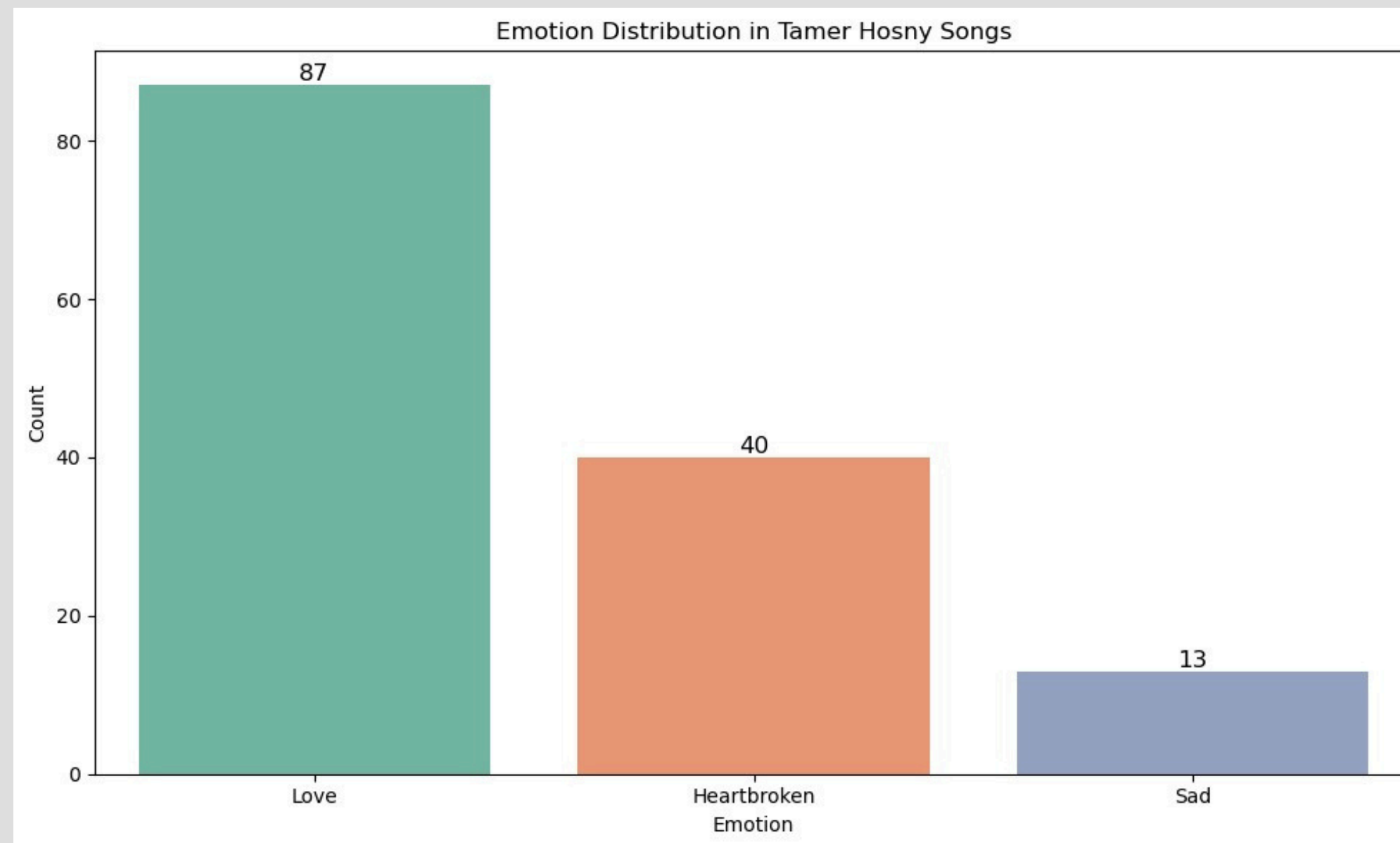


Entropy/Decade



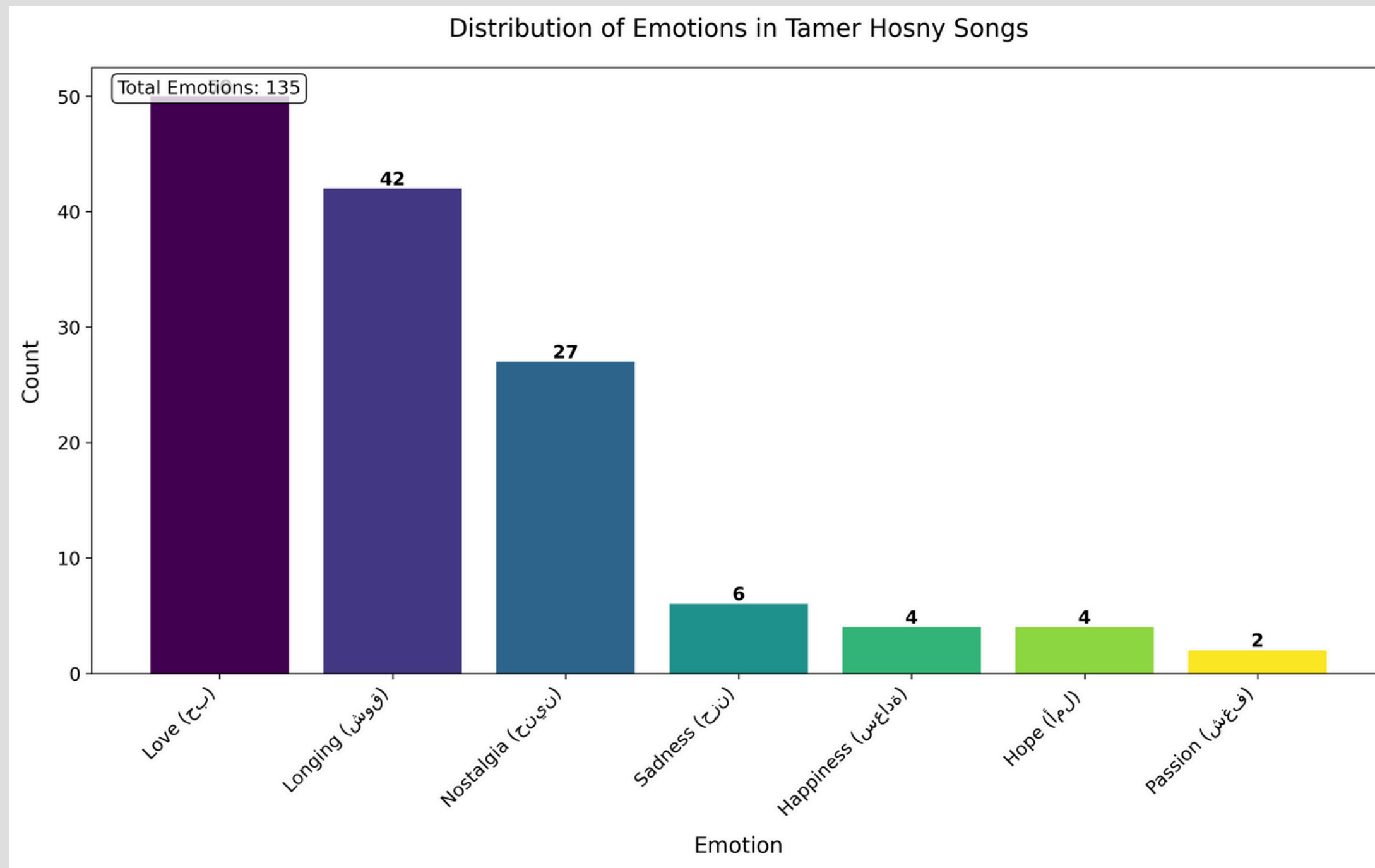
LLM ANALYSIS USING TEXT

Emotion Classification. Model used Qwen3:4B



LLM ANALYSIS USING VOICE

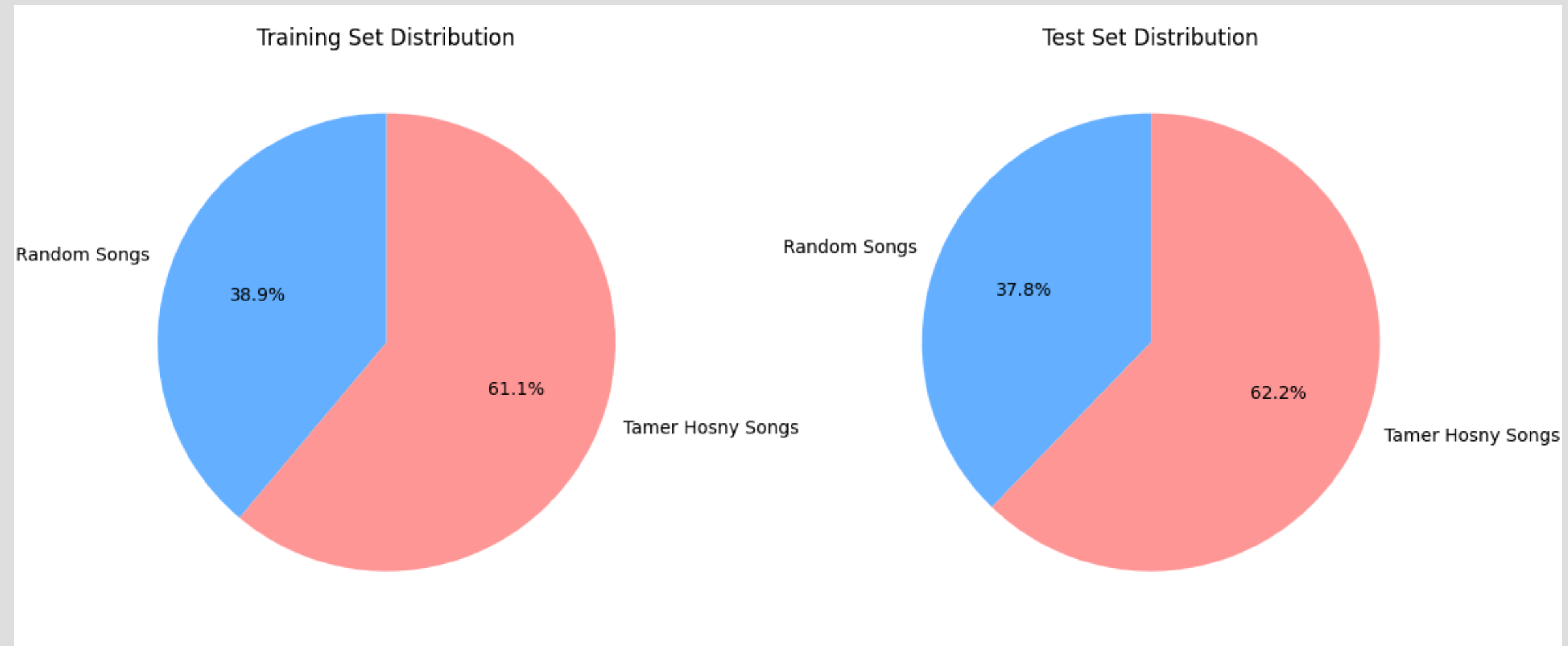
Emotion Classification. Model used Qwen3:4B



CNN-BASED SINGER CLASSIFICATION: IDENTIFYING TAMER HOSNY'S VOICE

- Developed and trained a Convolutional Neural Network (CNN) to classify audio tracks based on the singer's identity.
- Focused on binary classification: distinguishing between Tamer Hosny and other artists.
- Processed audio data into spectrograms for input into the CNN, enabling effective learning of vocal features.
- Achieved 80% accuracy on the test set, demonstrating the model's ability to identify Tamer Hosny's voice.

DATA SPLITTING DISTRIBUTION



CONFUSION MATRIX

