Movie Title Sentiment Analysis by Genre Project Overview

Introduction

This project examines how the sentiment conveyed by movie titles varies across genres. Using the IMDb box office dataset, each title is scored for polarity (from –1 negative to +1 positive) to reveal naming patterns and emotional framing choices in different genres.
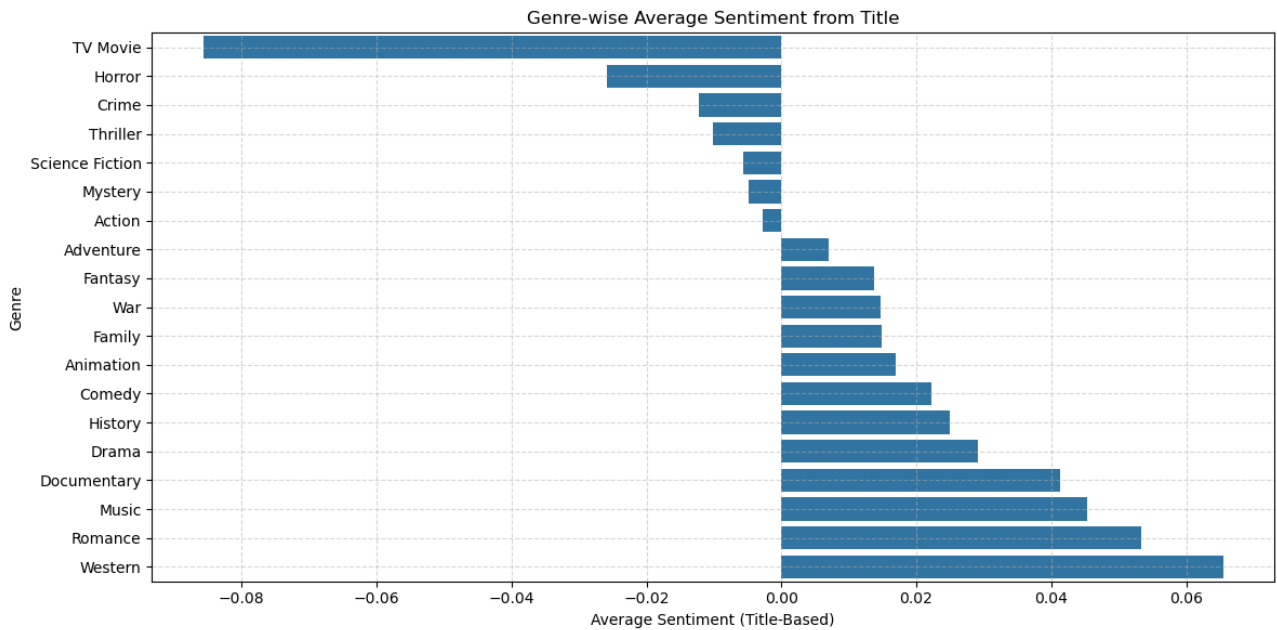
Dataset

- **Source**: CSV of top movies by worldwide gross, including title and genre metadata.

- **Key Fields**:

  - Movie Title

  - Comma-separated Genres

Preprocessing & Methodology

1. **Data Cleaning**

   - Remove entries missing title or genre.

2. **Genre Explosion**

   - Split multi-genre entries so each row represents one title–genre pair.

3. **Title Sentiment Scoring**

   - Compute a polarity score for each title on a –1 to +1 scale.

4. **Aggregation**

   - Calculate the average title polarity for each genre.

Visualizations

- **Genre-wise Average Sentiment Bar Chart**

  - Bars sorted by mean polarity, highlighting which genres use more positive or negative language in titles.

Genre-wise Average Sentiment from Title

- **Title Sentiment Distribution Box Plot**

  - Boxplots per genre showing median, interquartile range, and outliers of title polarity scores.

Distribution of Title Sentiment by Genre

Results & Output

- **Average Polarity by Genre**

  - Highest in Family and Animation (titles skew positive).

  - Lowest in Horror and Thriller (titles skew negative).

- **Distribution Observations**

  ○ Comedy and Action exhibit wide sentiment ranges, indicating mixed title tones.

  ○ Adventure and Sci-Fi cluster around neutral polarity.

Limitations

- **Title-Only Focus**: Does not consider full review text, plot details, or audience reception.

- **Lexicon Constraints**: Basic polarity scoring may misinterpret proper nouns or nuanced language.

- **Genre Overlap**: Multi-genre films contribute equally to each genre's aggregate, potentially diluting signals.

Future Work

1. **Keyword Frequency Analysis**: Identify the most common positive or negative terms per genre.

2. **Advanced Sentiment Models**: Apply transformer-based analyzers to capture subtler language cues.

3. **Metadata Correlation**: Explore links between title sentiment and box office performance or critic scores.

4. **Temporal Trends**: Track how title sentiment by genre evolves over decades.