Project Report on F.R.I.E.N.D.S

Manas P Panse

Table of contents

Motivation

FRIENDS hasa been a cornerstone of popular culture for decades now. With it's witty dialogues, intricate & relatable characters, it continues to captivate audiences globally. This project aims to explore the dynamics of spoken dialogue, emotional expressions across all 10 seasons. The motivation for this analysis stems from my desire to understand the narrative flow and character development over time, from a completely data-driven point-of-view, apart

from the fact that this is my favourite show :). Additionally, by examining viewership & ratings trends, I aim to assess whether the show's popularity and perceived quality followed a predictable pattern of fluctuated based on certain events or arcs within the series.

Project Overview

This analysis leverages 3 primary datasets :

1. `friends.csv` : Contains every line of dialogue spoken across all 10 seasons, with key variables such as speaker, season, episode, scene, and utterance.
2. `friends_emotions.csv` : Assigns an emotional label to the dialogues, enabling sentiment analysis and tracking of character emotions.
3. `friends_info.csv` : Provides contextual metadata about each episode, including directed_by, written_by, us_views_millions, and imdb_rating.

Basic EDA (Exploratory Data Analysis)

```
# # Importing Necessary Library
import pandas as pd

# Importing Datasets
friends_df = pd.read_csv('data/friends.csv')
emotions_df = pd.read_csv('data/friends_emotions.csv')
info_df = pd.read_csv('data/friends_info.csv')
```

FRIENDS

```
friends_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 67373 entries, 0 to 67372
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   text       67373 non-null  object
 1   speaker    67097 non-null  object
```

```
 2   season     67373 non-null  int64
 3   episode    67373 non-null  int64
 4   scene      67373 non-null  int64
 5   utterance  67373 non-null  int64
dtypes: int64(4), object(2)
memory usage: 3.1+ MB
```

FRIENDS EMOTIONS

```
emotions_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12606 entries, 0 to 12605
Data columns (total 5 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   season     12606 non-null  int64
 1   episode    12606 non-null  int64
 2   scene      12606 non-null  int64
 3   utterance  12606 non-null  int64
 4   emotion    12606 non-null  object
dtypes: int64(4), object(1)
memory usage: 492.6+ KB
```

FRIENDS INFO

```
info_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 236 entries, 0 to 235
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   season             236 non-null    int64
 1   episode            236 non-null    int64
 2   title              236 non-null    object
 3   directed_by        236 non-null    object
 4   written_by         236 non-null    object
```

```
 5   air_date           236 non-null    object
 6   us_views_millions  236 non-null    float64
 7   imdb_rating        236 non-null    float64
dtypes: float64(2), int64(2), object(4)
memory usage: 14.9+ KB
```

Key Analysis Sections

Viewership & IMDb Ratings Over Time

1. Objective - Analyze the average IMDb ratings and US viewership for each episode over time.
2. Approach -
    - Calculate the average IMDb rating and average US viewership for each season.
    - Then plot these trends separately across seasons to see if there's a consistent increase/decrease in ratings & viewership over time.
3. Insight - Understand whether the show improved in terms of both quality and popularity over time or were there any dips.

Character Dialogue Domination

1. Objective - Identify which character dominates the dialogue in each season and episode.
2. Approach -
    - Aggregate the number of lines spoken by each character per season & episode.
    - Visualize the results using bar charts to plot the dominant character.
3. Insight - Uncover which characters take center stage in different story arcs and how their roles evolve across seasons.

Emotional Tone Analysis

1. Objective - Determine which character uses the most positive or optimistic language throughout the series.
2. Approach -
    - Perform sentiment analysis using models : TextBlob for simplicity and BigBird for nuanced contextual analysis.
    - Compare polarity scores by character.
    - Visualize emotional tone distribution using donut charts.

3. Insight - Explore how each character's emotional expression contributes to the show's tone and audience appeal.

NLP Tools & Models

Initially, the `friends_emotions.csv` dataset was intended for emotional labeling, but it provides a limited coverage of emotionally-labelled dialogues (approx. 12k). Instead, sentimental analysis was conducted. The models that were used are :

TextBlob

This is a Python library that makes NLP tasks easier. It provides simple & easy-to-use tools for text processing. It includes functionalities such as part-of-speech tagging, sentimental analysis, translation, etc.

BigBird

BigBird is a transformer-based model for NLP, developed to handle long sequences of text that traditional transformer models like BERT, might struggle with due to their quadratic time complexity. It is based on the transformer architecture and is designed to process large text sequences more efficiently.

Results

This was a really interesting project. I was very much surprised by the results I obtained from my analysis. Before reading this, try to look at the

1. In the `Viewership & IMDb Ratings Over Time` section, there are two things which stand out :

    - Viewership Trend - There is a major spike in viewership immediately during the Season 2 of the series, which might be due to the episode, `The One After the Superbowl`, that was the lead-out program of `Super Bowl XXX` on January 28th, 1996. It is the **MOST-WATCHED EPISODE** of the entire series, while also being the **HIGHEST-RATED SUPER BOWL LEAD-OUT PROGRAM OF ALL TIME**, a title which it still holds.

- IMDb Ratings Trend - The show never dips below **8.30**, which is a great sign. There are 2 MAJOR spikes that happen during the season 5, and the last season. The fifth season does have a compelling storyline, and the final season is really the ribbon that is tied in sorf of a bow on a box of everything. It really can't be explained in words.

2. In the `Character Domination` section, there are also two things which stand out :

   - Per Season - The most dialogues spoken by a character per season have Chandler, portrayed by `Matthew Perry`, Ross portrayed by `David Schwimmer`, and Rachel portrayed by `Jennifer Aniston` in them. Since, this plot could cause misconceptions among new viewers questioning if these characters are the spotlight of the show, I went a step ahead.
   - Per Episode - This is the most beautiful plot I've made so far. It gives justive to the main cast i.e the 6 FRIENDS. The thing that stood out to me is that **2 GUEST APPEARANCES** made the cut. That's right! Guest stars who starred as `Paul Stevens` and `Amy Green` captured their place on this beautiful piece of art by having the most spoken dialogue in their episode. Considering this their short story arcs, it definitely feels worth it that I analyzed this dataset.

3. In the `Emotional Tone Analysis` section, there are things which interest me :

   - TextBlob - This model completed the analysis within few minutes. The results it gave me, show that the character with most optimistic dialogue is Phoebe Buffay, portrayed by `Lisa Kudrow`. As someone who has watched the series many many many many many times, and continues to do so on a regular basis, I have to say that's true, atleast in my opinion. The second most optimistic character based on their dialogue is Rachel Green, and the third spot is stood on by Ross Geller. The rest in order are Chander Bing, Joey Tribbiani, portrayed by `Matt LeBlanc` and Monica Geller, portrayed by `Courtney Cox`. And while I am in sync with #1, #2 is a detabatable character for me, instead of Rachel, I'd have Chandler there. #3 is also a position I agree with. #4 would be Rachel, #5 would be Joey, and #6 would be Monica. Now, these rankings don't make any characters more/less lovable and funny; this is just my opinion that I felt like sharing. You can have your rankings.
   - BigBird - This model took a while to process (around 6 hours). The model returned the characters' polarities alomst equal to each other with very minute differences. The rankings it gave me are, #1 is Monica, #2 is Rachel, #3 is Chandler, #4 is Joey, #5 is Phoebe, and #6 is Ross. While these results are really neck-to-neck and the model analyzes very patiently, this might look correct. But, the bottom line is that, no matter how advanced an AI model is, at least in today's day and age & as of writing this, **An AI model won't be able to replace *YOU*. Because when you watch something, you don't just listen to the dialogues, you connect it with your own though processes, emotions, celebrity sentiments, etc.**

and this is something that any AI model won't be able to do. The conclusions you draw in your head include all of this information, and not just dialogues.

Future

While I want to do more work on this dataset, I am currently limited by my knowledge and computational capabilities. After acquiring more knowledge about the field of NLP & Data Science, sometime in the future, I wouldn't mind diving back into this for a more in-depth analysis, and I hope that I do it way more professionally.