CS410 Team Project Progress Report

Overview

As described in the proposal, we aim to explore the sentiments expressed in the Friends TV show scripts. Our focus involves processing the screenplay script text data and conducting an in-depth analysis using text retrieval and sentiment analysis.

We've divided the project into the following sub-tasks, assigned to different team members:
1. Data pre-processing: Clean and organize the raw screenplay script text data.
2. Text retrieval: Retrieve a ranked list of lines based on relevance for a given topic of interest.
3. Sentiment analysis: Analyze sentiments based on the retrieved lines.
4. Data visualization: Develop a Tableau dashboard to consolidate and visualize sentiments related to selected topics.

Progress to date

Below is a summary of the progress we've made on the project:
1. Determined the topics we'd like to dive into, specifically:
   - Who is the most favorable character in the show?
     - Who is the most favorable male?
     - Who is the most favorable female?
   - Who is the happiest character in the show?
     - Who is the happiest male?
     - Who is the happiest female?
   - What is the overall sentiment of lead characters?
     - What is the trend of lead characters' sentiments over time?
     - What is the lead characters' sentiments towards each other?
2. Completed data pre-processing, for example:
   - Identify and consolidate the data sources for Friends scripts across 10 seasons
   - Remove irrelevant information (e.g. the writer of the episode)
   - Separate the speaker of the dialogue from the content of the dialogue
   - Tokenize the dialogue, remove special characters, remove stop words, normalize text (i.e. stemming and lemmatization)

3. Conducted research on how text retrieval and sentiment analysis can be conducted using the existing Python packages. The packages we looked into include Gensim and Hugging Face.
4. Aligned on how analysis results will be aggregated and visualized in Tableau

Pending tasks

Below is a summary of the pending tasks we will complete in order to finalize the analysis:
1. Implement the text retrieval and sentiment analysis to generate the results
2. Build the Tableau dashboard to visualize the analysis results

Challenges

The initial project proposal we submitted was to retrieve Twitter data and conduct sentiment analysis. We encountered challenges when we conducted preliminary research. We realized the task was not feasible because only a very limited number of tweets can be downloaded for free. Therefore, we decided to pivot to conduct sentiment analysis of Friends TV show scripts, which is an interesting topic for all team members.