

Movie Subtitles Analysis Project

Ashwitha Banoth

200101021

Description

The Movie Subtitles Analysis Project is an exploration of textual data extracted from movie subtitles. The project aims to uncover insights into the thematic elements, dialogue patterns, and emotional undertones present in the movie through various data science techniques.

Features:

1. Exploratory Data Analysis (EDA):

- Initial exploration of the dataset to understand its structure, content, and basic statistics.
- Visualization of subtitle duration, text length, and other relevant attributes.

2. Text Analysis:

- Cleaning and preprocessing of subtitle text data to prepare it for analysis.
- Generation of word clouds to visualize the most common words and phrases.

3. Sentiment Analysis:

- Assessment of the emotional tone of dialogue using TextBlob and VADER sentiment analysis tools.
- Visualization of sentiment trends over time to identify emotional peaks and valleys throughout the movie.

5. Timeline Analysis:

- Visualization of when certain keywords or phrases appear in the movie to correlate dialogue with key plot points.
- Identification of patterns and trends in the occurrence of specific terms over time.

Installation and Setup

Follow these steps to get the project up and running on Google Colab:

Step 1: Prepare the Files

Ensure you have two main files ready:

Code file (.ipynb): This is your Jupyter notebook that contains all the code for the project.

Dataset file (.srt): This is the subtitle dataset used in the analyses.

Step 2:

Upload Files to Google Colab

Step 3: Run the Notebook

After uploading both the notebook and the dataset

Navigate through the cells in the notebook. Each cell contains a portion of the code that can be executed to perform specific tasks as part of the analysis.

Run the cells sequentially from top to bottom to ensure all variables and libraries are loaded correctly and the analysis proceeds without errors.

Usage

The notebook is designed to be self-contained and includes comments and explanations for each part of the analysis. Follow the instructions in each cell to execute the project components, adjust parameters, or input values as necessary to explore different aspects of the data.