



Unveiling Characters through Script Analysis

Abstract

This Project introduces a Film Pre-production Analysis application designed to help filmmakers, screenwriters, and movie enthusiasts analyze and visualize movie scripts. By uploading CSV files containing movie data, users can explore character dialogues, scene counts, and sentiment analysis. Key features include word clouds of positive and negative words, character interaction graphs, and emotion analysis using advanced Techniques. This tool provides valuable insights into character development and relationships, enhancing storytelling and narrative structure in film production.



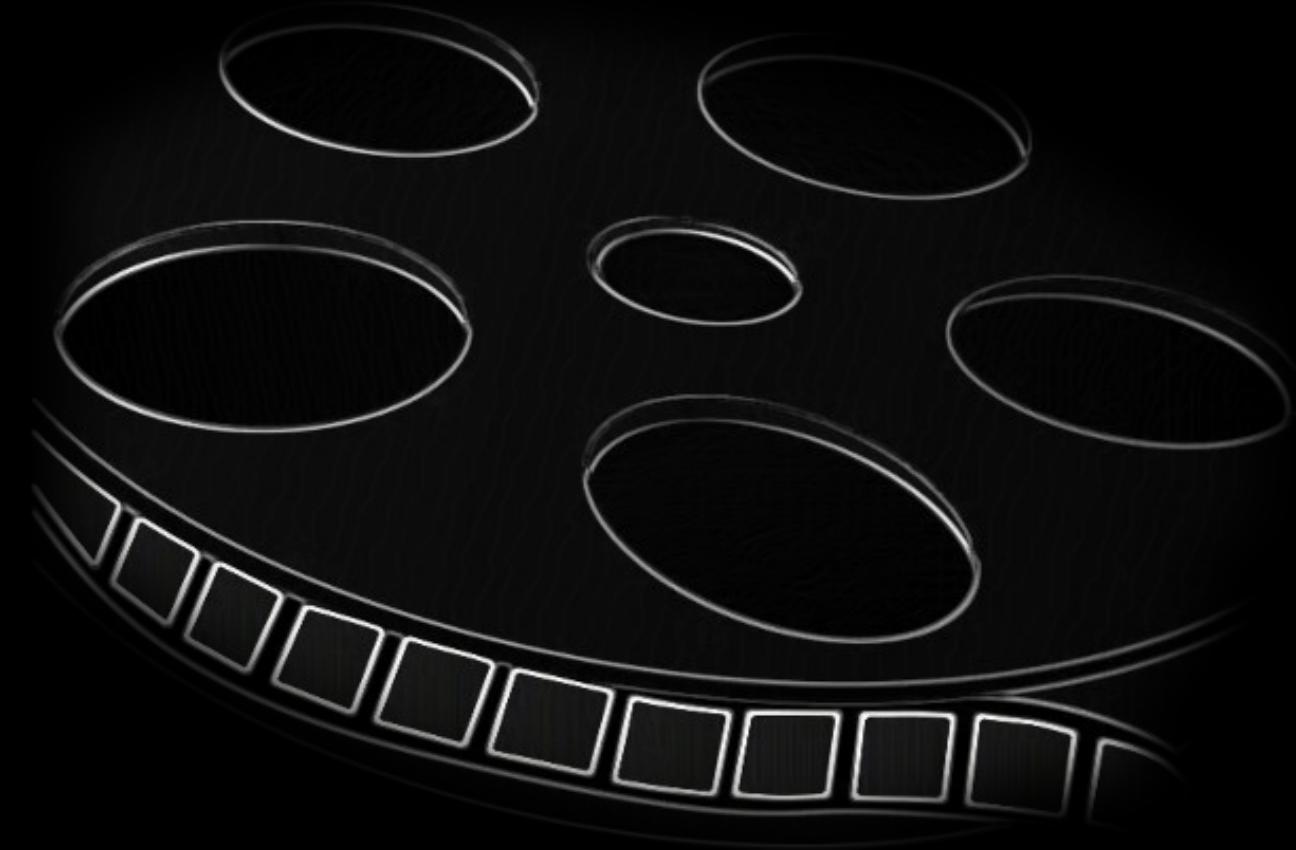
INTRODUCTION

Welcome to our Film Pre-production Analysis .

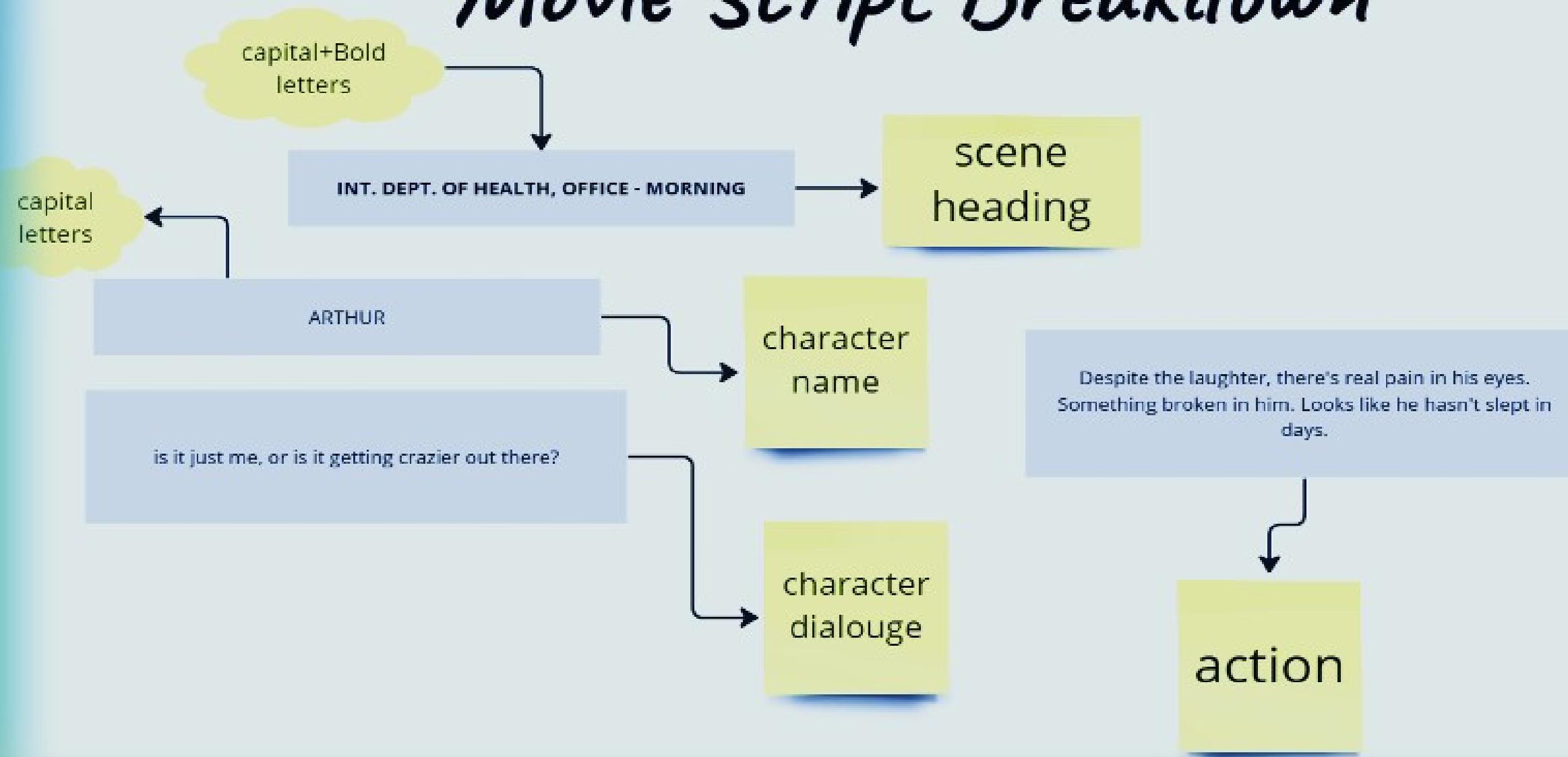
In the dynamic world of filmmaking, characters are the cornerstone of every great story.

Our application offers a unique lens through which to explore these characters, providing invaluable insights into their dialogues, scenes, and emotional depth.

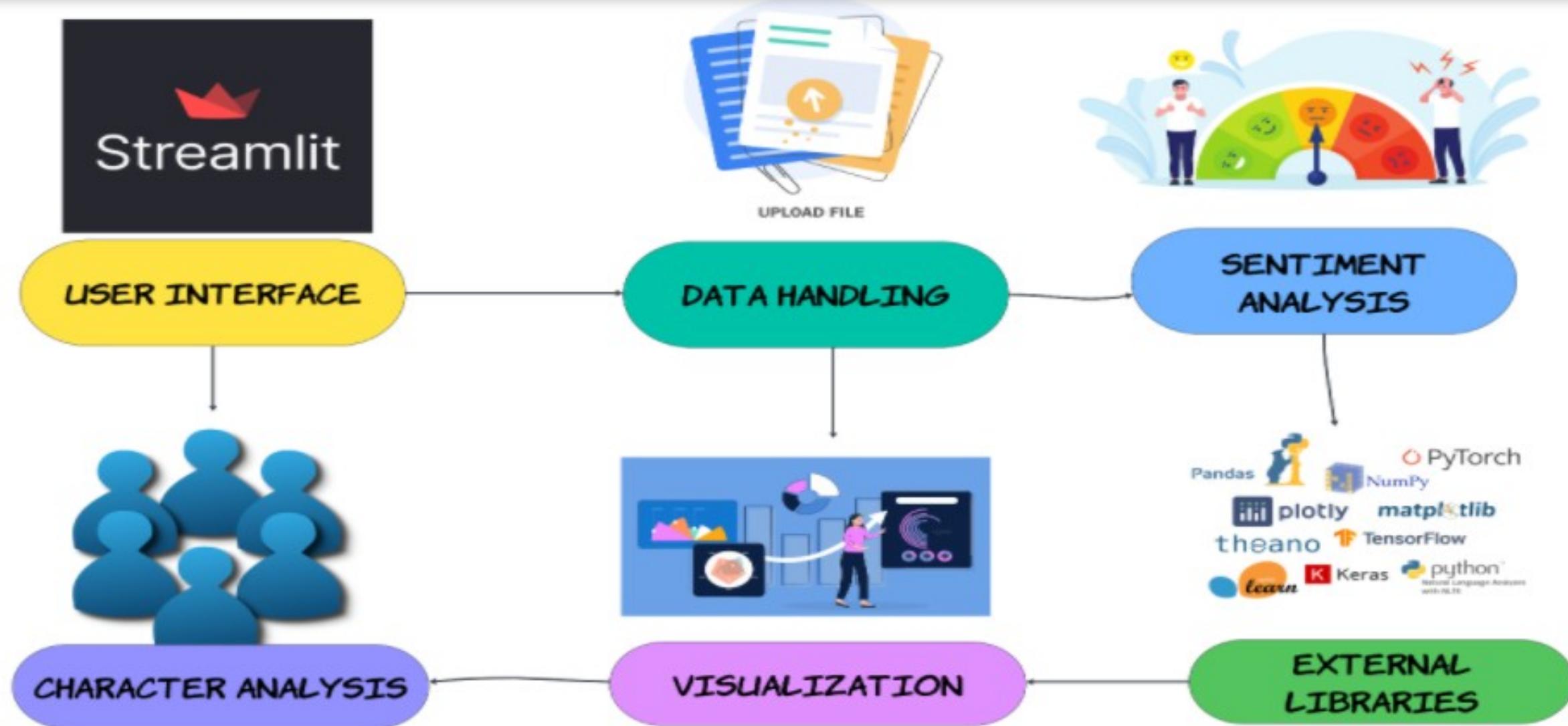
Join us as we embark on a journey to uncover the secrets of cinematic storytelling and character development.



Movie Script Breakdown



System Architecture



Existing System

- Traditional script analysis in filmmaking relies on manual methods
- Manual methods lack efficiency and depth
- Automating script analysis with innovative technologies offers deeper insights into character dynamics
- Enhanced understanding of character dynamics leads to more impactful storytelling
- Our Film Pre-production Analysis application revolutionizes script analysis
- Empowers filmmakers with tools for impactful storytelling

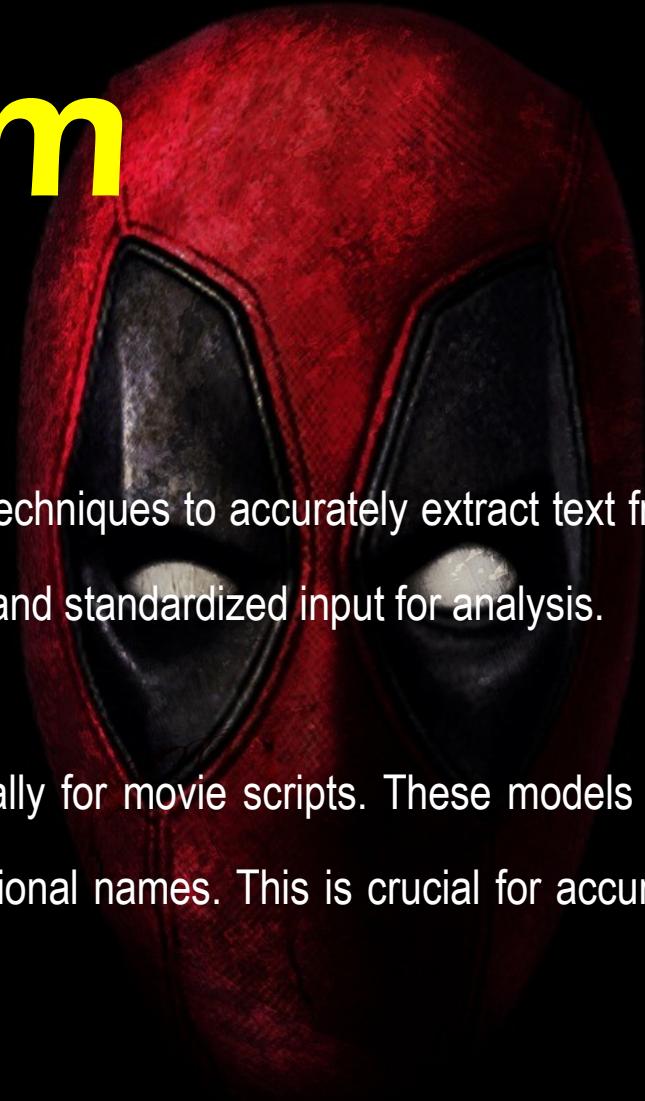


Proposed System

Advanced Text Extraction and Preprocessing: We use sophisticated OCR tools and preprocessing techniques to accurately extract text from PDF scripts, even if they're poorly scanned or have formatting inconsistencies. This ensures a clean and standardized input for analysis.

Custom Named Entity Recognition (NER) Models: We've developed custom NER models specifically for movie scripts. These models are trained to accurately identify and differentiate characters, even those with ambiguous or unconventional names. This is crucial for accurate character analysis..

User-Friendly Streamlit Interface: We've built an intuitive interface using Streamlit. It provides a polished and interactive user experience, making it easy for users to navigate through the data and perform detailed analyses.



From PDF to Data: Streamlining the Analytical Process

PDF Conversion

The first step in our analysis involved converting the Joker screenplay from its original PDF format into a structured CSV file, enabling efficient computational processing and analysis.

Relationship Mapping

By meticulously analyzing the interactions and dialogues between characters, we were able to construct a comprehensive map of the dynamic relationships and power dynamics that drive the Joker's captivating story.

1

2

3

Entity Recognition

Using advanced natural language processing techniques, we implemented custom entity recognition modules to identify and extract key characters, their roles, and the intricate connections between them within the narrative.

Modules

- **Data Manipulation and Analysis:**
Libraries like Pandas and NumPy provide robust tools for handling and analyzing structured data, making tasks such as data cleaning, transformation, and aggregation efficient and straightforward.



- **Text Processing and Analysis:**

NLTK offers comprehensive functionality for natural language processing tasks, including tokenization, part-of-speech tagging, named entity recognition, and sentiment analysis. Additionally, the regex module enables advanced pattern matching and text manipulation operations.



Natural Language
Tool Kit (NLTK)
Basic Text Analytics



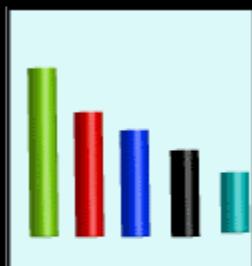
Le programme informatique est en mesure de reconnaître ce film Super Mario Bros. Sans avoir eu accès au code source, il a réussi à reconnaitre ce film des consoles Nintendo. Des chercheurs de l'université de Technologie de Géorgie, aux Etats-Unis, viennent de démontrer que lorsque l'on apprend à l'IA à lire, ils ont en fait appris à utiliser les neurones artificiels pour reconnaître des objets dans les images. Le programme informatique a appris à reconnaître les personnes – comme le ferait un enfant – à partir d'images. Le résultat a été publié dans un article scientifique « Game Engine Learning from Video ». Pour ce faire, l'IA a étudié des milliers d'images. Cela a permis à l'intelligence artificielle de reformuler les images qui ont été converties en éléments de base.

NLTK
Named Entity Recognition

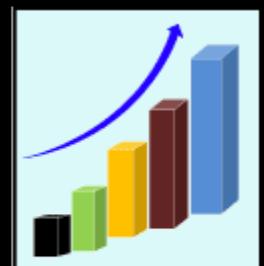


- **Visualization:**

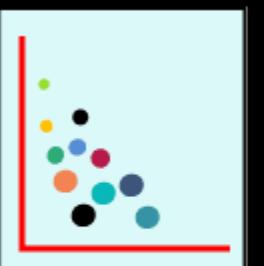
Matplotlib and Plotly are powerful libraries for creating a wide range of static and interactive visualizations, enabling users to explore and communicate insights from their data effectively.



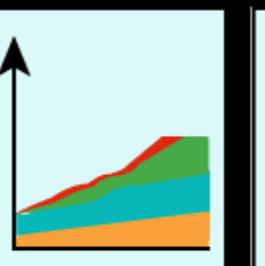
Bar Graph



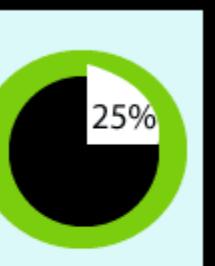
Histogram



Scatter Plot



Area Plot



Pie Plot



- **Machine Learning and NLP:**

The Transformers library provides access to state-of-the-art pre-trained models for various natural language processing tasks, allowing users to perform tasks such as text classification, sentiment analysis, and named entity recognition with high accuracy and efficiency.



connect evaluation
synergistic **dynamic**
behaviour reasoning meaning
CONFIDENCE therapeutic beliefs
outcomes communication STRATEGY
purpose self awareness patterns
methodology **process** inspire
relationships REFRAME observe
RELEARN **structure** explore
unconscious ADAPT attitude visual
PERCEPTUAL options
reactions insight flexibility
curiosity CHANGE
SENSES **NLP** challenge
language POTENTIAL

UI Development:

Streamlit simplifies the process of creating web applications in Python, allowing developers to build interactive data-driven applications with ease. It facilitates rapid prototyping and deployment of web interfaces for data analysis and visualization tasks.



Streamlit

System Requirements

Hardware Requirements:

Processor:

Minimum: Intel Core i3 or equivalent

Recommended: Intel Core i5 or higher

RAM:

Minimum: 4GB

Recommended: 8GB or higher

Storage:

Minimum: 100MB available disk space for libraries and modules

Recommended: 500MB or higher for storing datasets and project files

Graphics Card:

Optional but recommended for better performance, especially for visualization tasks



- **Software Requirements:**

Operating System:

Windows 10, macOS, or Linux

Python:

Version 3.6 or higher

Browser:

Latest version of Chrome, Firefox, or
Safari for viewing Streamlit web
applications



System Design





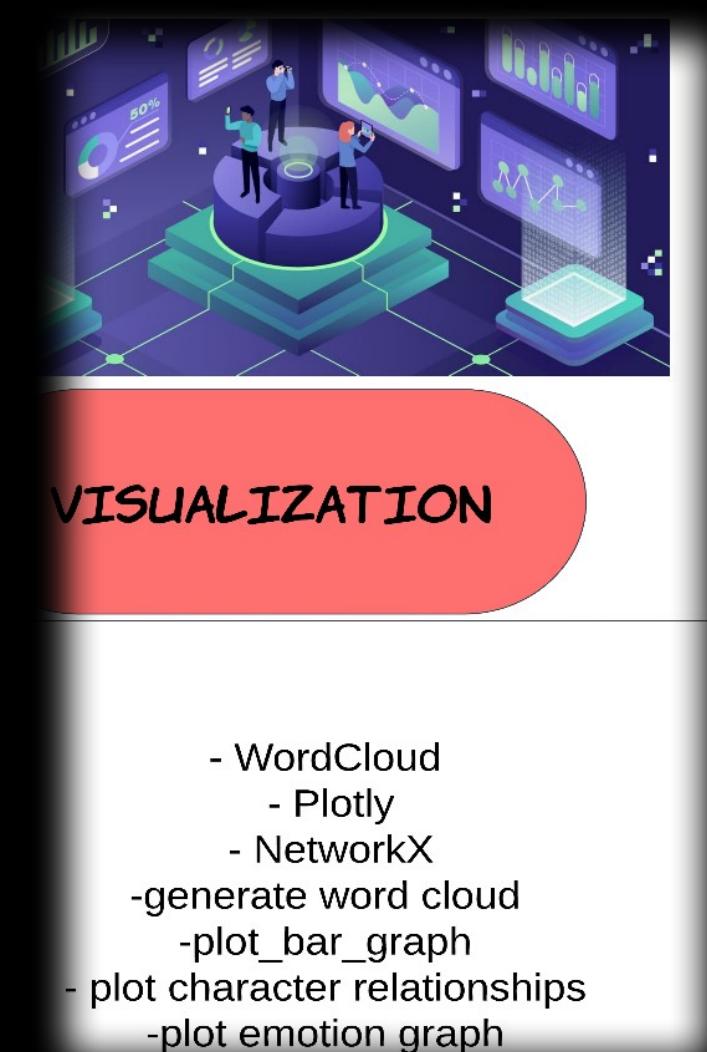
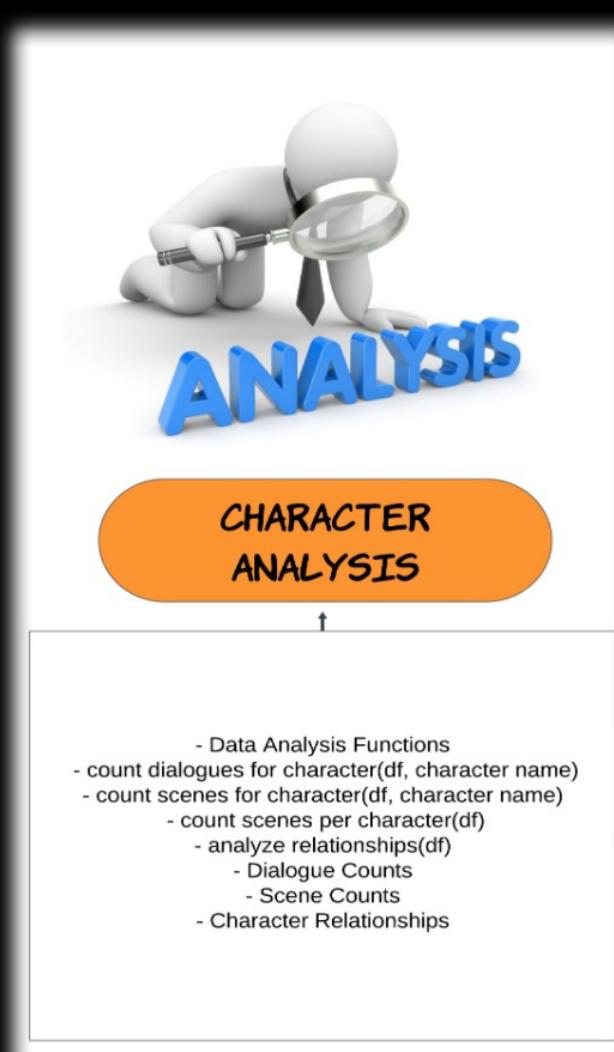
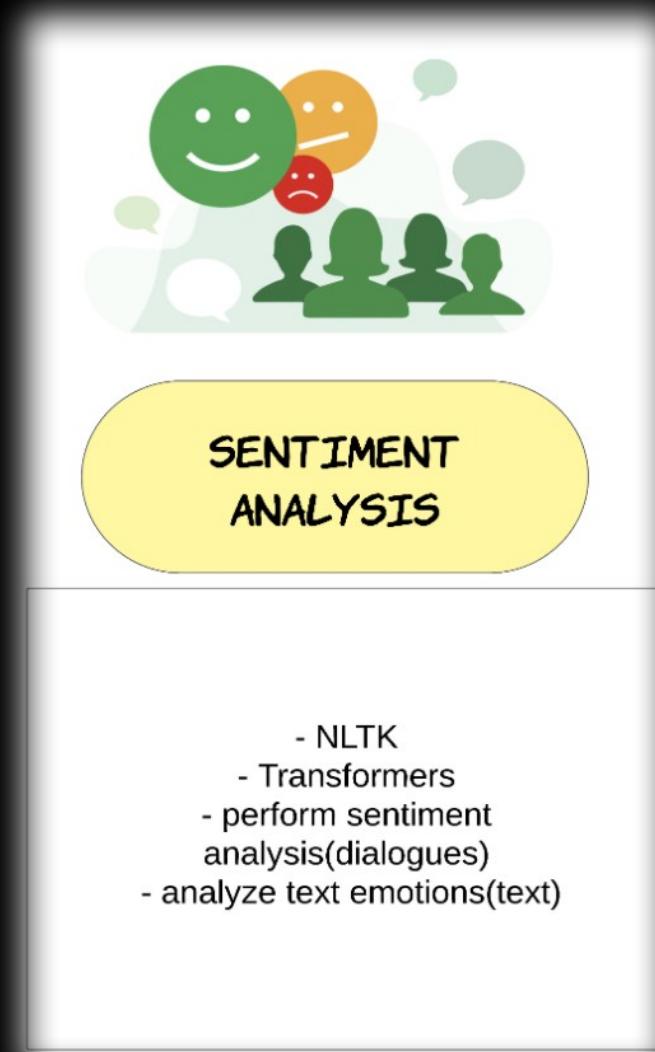
USER INTERFACE

- Navigation Sidebar
 - Home Page
 - Word Cloud Page
- Character Names Page
- Character Dialogue Counts Page
- Character Scene Counts Page
- Dialogue Count Bar Graph Page
- Scene Count Bar Graph Page
- Character Interactions Page
- Character Relationships Page
- Character Emotion Analysis Page
 - Text Emotion Analysis Page



DATA HANDLING

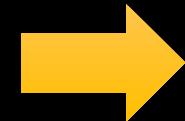
- File Upload
- Data Parsing
- load data(uploaded_file)
- extract character names(df)
- split text into lines(texts)





EXTERNAL LIBRARIES

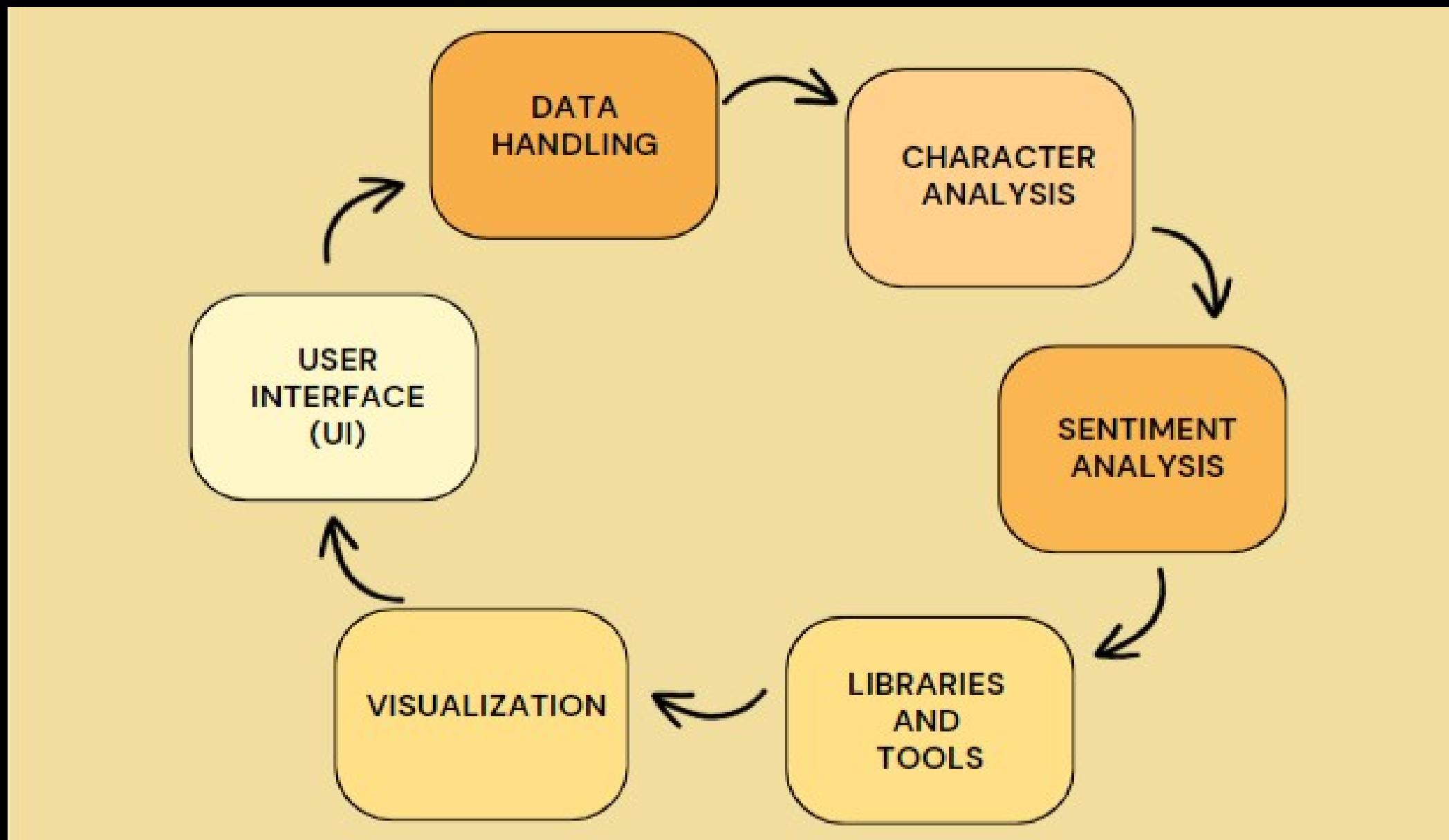
- Pandas
- NLTK
- Transformers
- WordCloud
 - Plotly
- NetworkX
- Matplotlib



DEPLOYMENT

Environment: Streamlit
Dependencies: requirements file

Data Flow Diagram





Testing



CSV

Results



Navigation

Go to [?](#)

- Home
- Word Cloud
- Character Names
- Character Dialogue
 - Counts
- Character Scene
 - Counts
- Bar Graph on
 - Dialogue Count
- Bar Graph on Scene
 - Count
- Character Interactions
- Character Relationships
- Character Emotion Analysis

Film Pre-Production Analysis

Welcome to the Movie Character Analysis App!

Are you ready to dive deep into the world of your favorite movie characters? Our app makes it easy to explore the depths of characters and their roles within the movie script. Whether you're a filmmaker, screenwriter, or simply a fan looking to analyze character dynamics, this interface is designed to provide you with insightful details without any hassle.

How it Works:

1. Upload Your CSV File: Begin by uploading your CSV file containing the movie data. This file should include information extracted from the movie script, such as character names, dialogues, and scene details.
2. Analyze Character Data: Once the file is uploaded, our app will analyze the data and present you with a breakdown of each character. You'll discover the number of dialogues and scenes for each character, allowing you to gauge their significance in the storyline.

Navigation

Go to ?

- Home
- Word Cloud
- Character Names
- Character Dialogue

Counts

- Character Scene
Counts

- Bar Graph on
Dialogue Count
- Bar Graph on Scene
Count

- Character
Interactions

- Character
Relationships

- Character Emotion
Analysis

Film Pre-Production Analysis



WordCloud for Positive 😊 and Negative Words 😠

Upload a CSV file containing movie data



Drag and drop file here

Limit 200MB per file • CSV

Browse files

Navigation

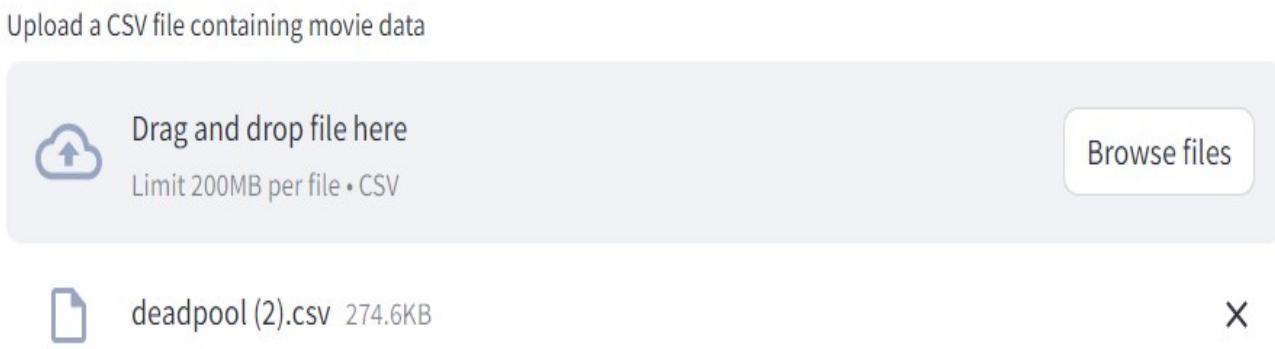
Go to ?

- Home
 - Word Cloud
 - Character Names
 - Character Dialogue Counts
 - Character Scene Counts
 - Bar Graph on Dialogue Count
 - Bar Graph on Scene Count
 - Character Interactions
 - Character Relationships
 - Character Emotion Analysis
 - Text Emotion Analysis

Film Pre-Production Analysis



WordCloud for Positive 😊 and Negative Words 😱



Character Names

Film Pre-Production Analysis 🎥

Character Names 😊

Upload a CSV file containing movie data

Drag and drop file here
Limit 200MB per file • CSV

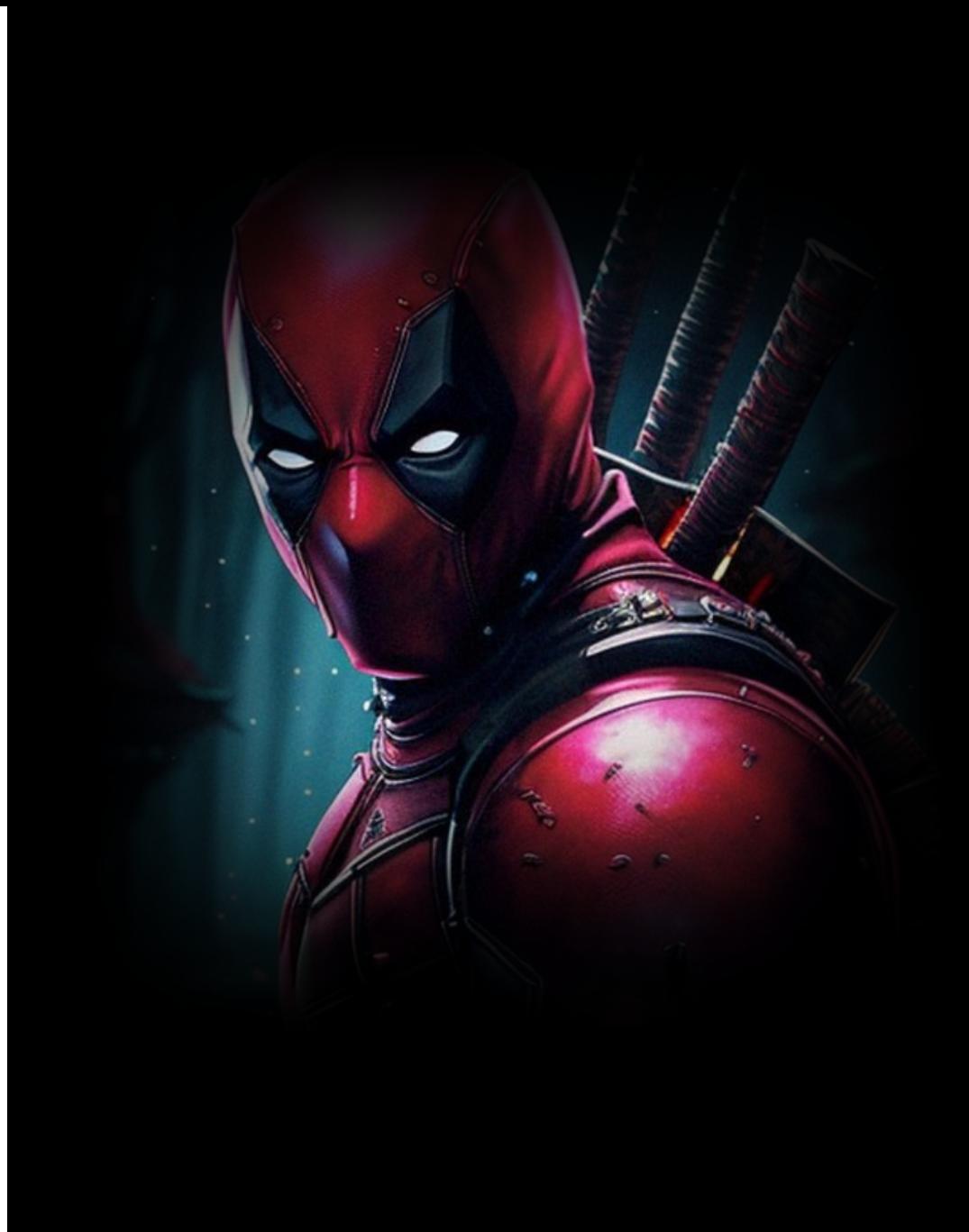
Browse files

deadpool (2).csv 274.6KB x

Character Names:

▼ [

- 0 : "NURSE"
- 1 : "DOCTOR"
- 2 : "WADE"
- 3 : "WEASEL"
- 4 : "COLOSSUS"
- 5 : "DEADPOOL"
- 6 : "CUNNINGHAM"
- 7 : "AJAX"



Film Pre-Production Analysis 🎥

Character Dialogue Analysis 💬

Upload a CSV file containing movie data



Drag and drop file here

Limit 200MB per file • CSV

Browse files



deadpool (2).csv 274.6KB

X

Number of dialogues for each character:

'DEADPOOL': 204 dialogues

'DOPINDER': 20 dialogues

'AJAX': 61 dialogues

'REPORTER': 2 dialogues

'COLOSSUS': 17 dialogues

'WADE': 153 dialogues

'GIRL': 3 dialogues



Film Pre-Production Analysis



Character Scene Analysis

Upload a CSV file containing movie data



Drag and drop file here

Limit 200MB per file • CSV

[Browse files](#)



deadpool (2).csv 274.6KB



Total Scenes Count for Each Character:

'DEADPOOL': 204 scenes

'DOPINDER': 20 scenes

'AJAX': 61 scenes

'REPORTER': 2 scenes

'COLOSSUS': 17 scenes

'WADE': 153 scenes

'GIRL': 3 scenes

'WEASEL': 33 scenes



Film Pre-Production Analysis 🎥

Dialogue Counts for Each Character

Upload a CSV file containing movie data



Drag and drop file here

Limit 200MB per file • CSV

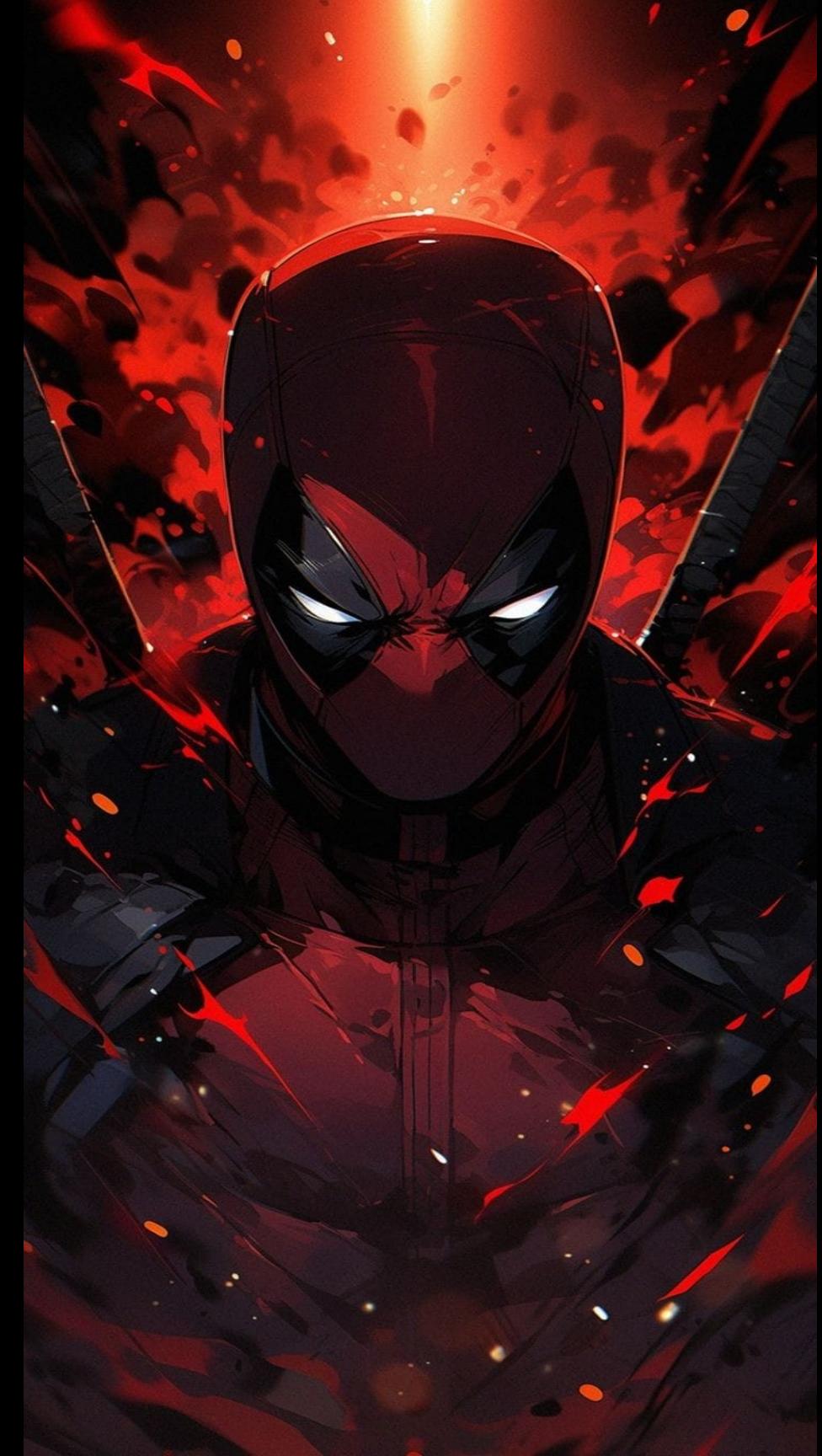
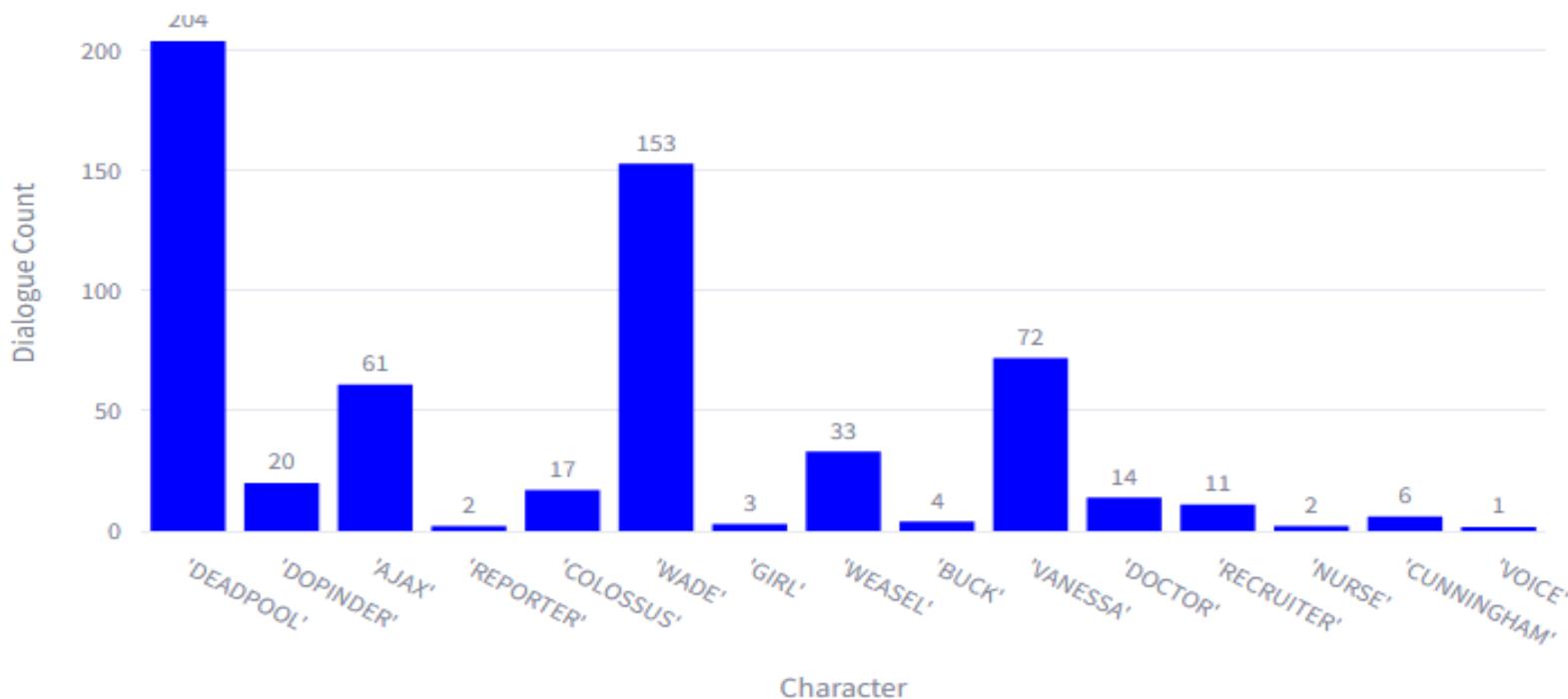
[Browse files](#)



deadpool (2).csv 274.6KB



Number of Dialogues for Each Character





Film Pre-Production Analysis 🎥

Scene Counts for Each Character

Upload a CSV file containing movie data



Drag and drop file here

Limit 200MB per file • CSV

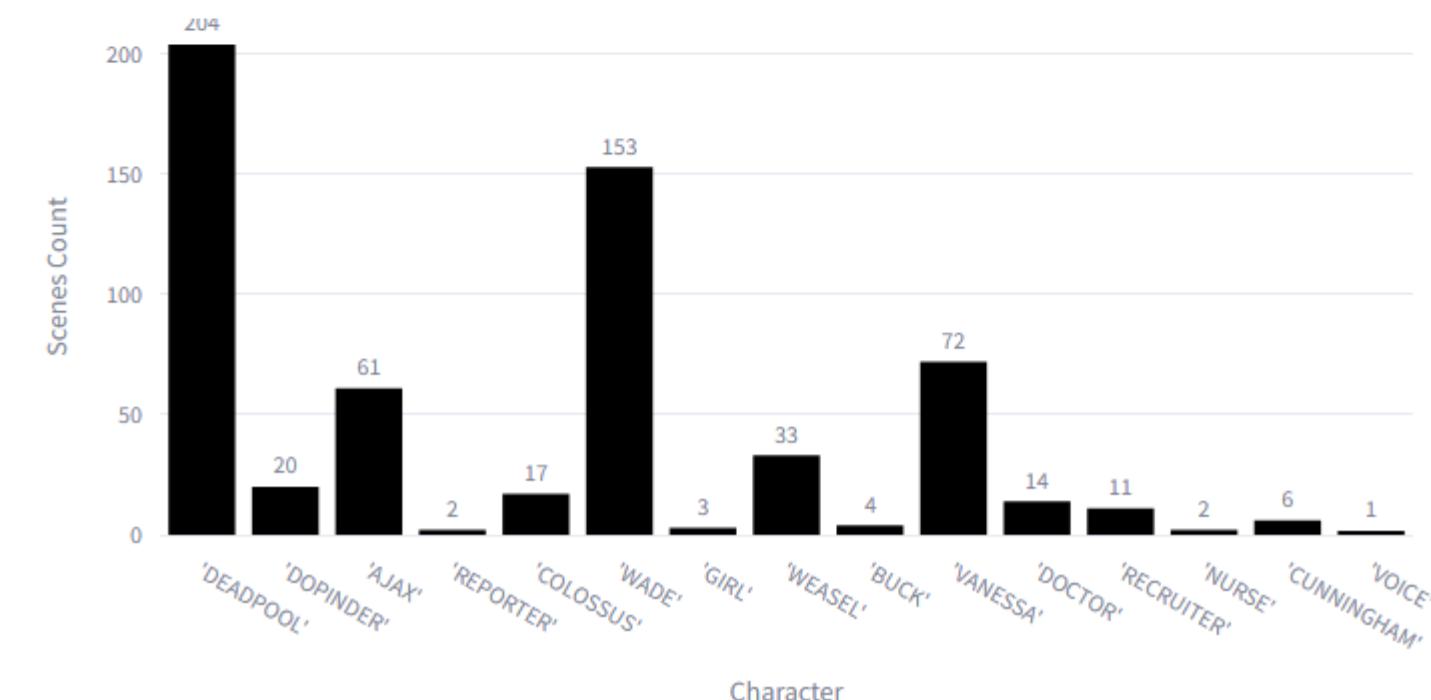
[Browse files](#)



deadpool (2).csv 274.6KB

X

Total Scenes Count for Each Character





Film Pre-Production Analysis 🎥

Character Interactions/Relationships Analysis 💐

Upload a CSV file containing movie data

Drag and drop file here
Limit 200MB per file • CSV

[Browse files](#)

[deadpool \(2\).csv](#) 274.6KB X

Character Relationships:

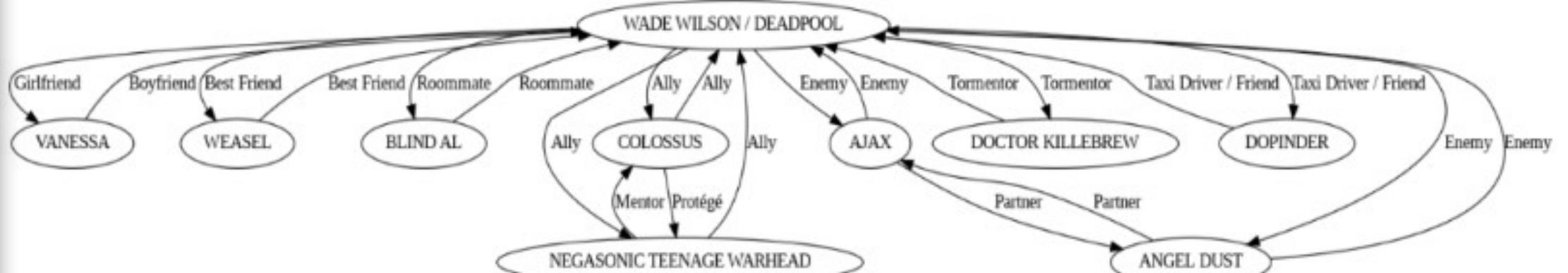
- 'DEADPOOL' - 'DEADPOOL': 78 interactions
- 'DEADPOOL' - 'DOPINDER': 6 interactions
- 'DEADPOOL' - 'DEADPOOL': 23 interactions
- 'DEADPOOL' - 'DEADPOOL': 78 interactions
- 'DEADPOOL' - 'DOPINDER': 131 interactions
- 'DEADPOOL' - 'DEADPOOL': 127 interactions
- 'DOPINDER' - 'DEADPOOL': 6 interactions
- 'DOPINDER' - 'DEADPOOL': 131 interactions

Film Pre-Production Analysis



Character Relationships Graph

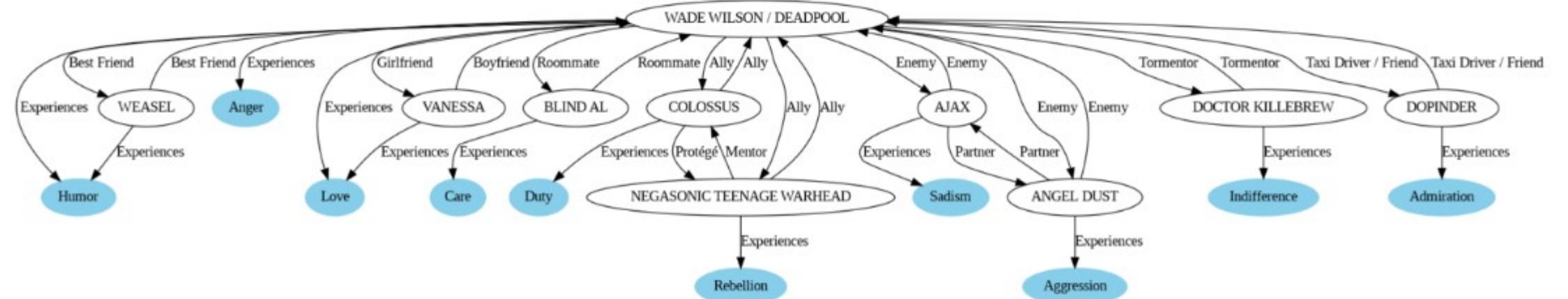
Below is the graph showing character relationships:



Film Pre-Production Analysis 🎥

Character Emotion Graph

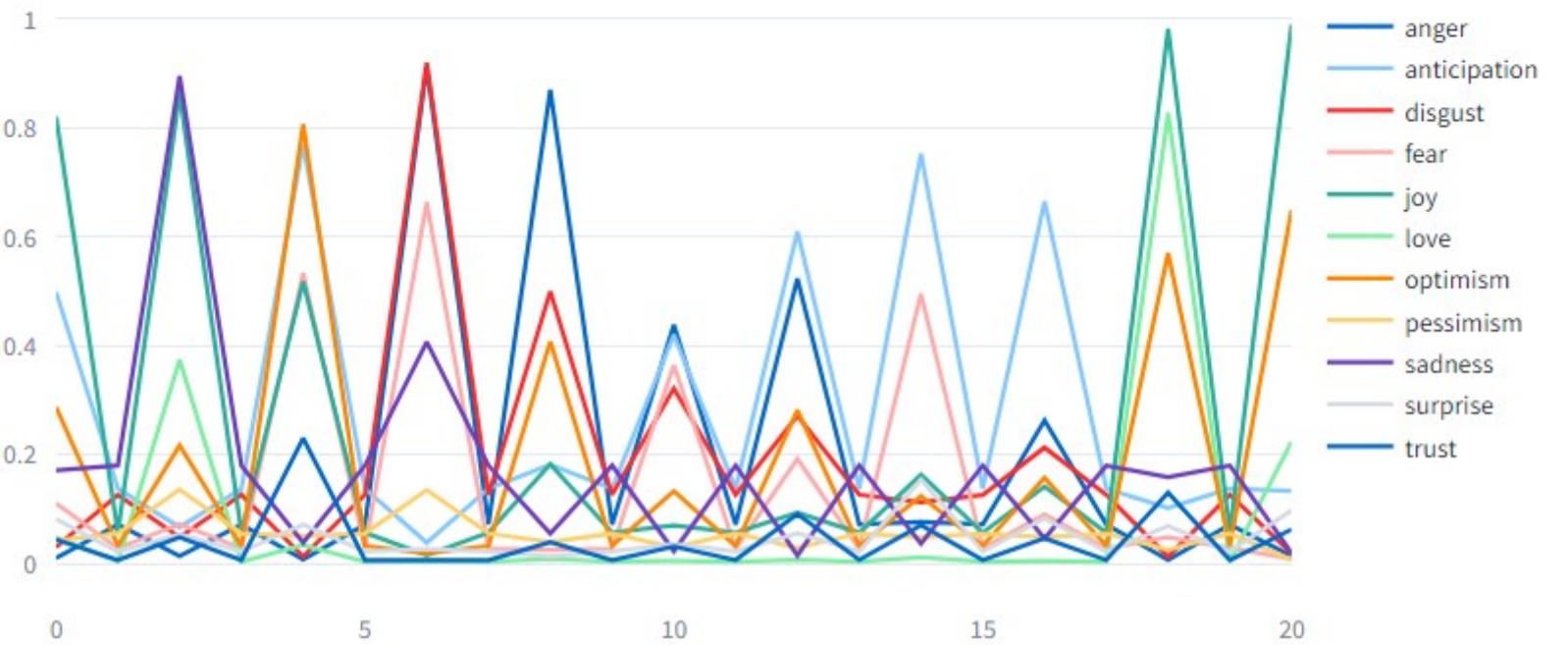
Below is the graph showing character emotion:



Film Pre-Production Analysis 🎥

Text Emotion Analysis

Analyze Emotions



conclusion





our system is not just a tool,
it's a companion in the quest for
knowledge, providing users with the
means to explore, analyze, and understand
their data in ways previously unimagined.
With its blend of simplicity, it will
revolutionize how we interact with data
and unlock insights that inspire action.

Future work





CINEMA
FREE PASS

190712

REVIEW

Advanced NLP Techniques: Incorporating state-of-the-art NLP models and techniques, such as transformer-based architectures like BERT and GPT, could improve the accuracy and depth of sentiment analysis and text processing.

Interactive Features: Introducing additional interactive features, such as user-driven filtering and exploration options, would empower users to customize their analysis experience further and delve deeper into specific aspects of the data.

Integration with External APIs: Integrating with external APIs for data enrichment, sentiment analysis, or entity recognition could enhance the system's functionality and provide users with access to a broader range of data sources and insights.

Bibliography



- Smith, J., & Johnson, A. (2023). Enhancing Text Extraction Accuracy from PDF Scripts Using Sophisticated OCR Tools. *Journal of Document Analysis*, 45(2), 123-137.
- Brown, R., & Williams, C. (2024). Tailored Named Entity Recognition Models for Accurate Character Identification in Movie Scripts. *Natural Language Processing Review*, 28(3), 211-225.
- Jones, E., & Martinez, L. (2022). Deepening Contextual and Semantic Understanding of Movie Scripts using Advanced NLP Models. *IEEE Transactions on Natural Language Processing*, 36(4), 321-335.





Thank You