Assignment 2 - Study of Top 3 Disney Live-Action Movies among different social media platforms

Group 19 member:

Phankawee Mukdaprakorn (S4053635) (Tai)
Nadia Du (S4039758)
Wan Yi Lam (S4060745) (Sirius)

Meeting & Working Schedules

Date	Goals	Achievements	Discussions
2/5/25	- Reading through the overall assignment together - Discover some interesting topics - Deciding the final topic and problem question	- We all understood an overall of this assignment - Getting a few potential topics - Listed some questions that need clarification from Lida	- We read the Assignment Specifications together and made sure all of us had a clear understanding. - We found some questions that needed clarification. - We drafted the initial plan, e.g., the

	- Define the questions we want to ask Lida		structure of our report. - We did not decide on the final topic we wanted to do because of the time limitations.
6/5/25	- Study methods to collect data from Social Media platforms - List questions for the next consultation session - Sharing the knowledge from assignment 1	Potential Social Media platforms - YouTube-(Sirius) - Twitter – Nadia - Reddit- (Sirius) - Instagram (Tai) - Thread (Nadia) - Facebook (Tai)	-We shared about how we did our own assignment and talked about which parts we might want to apply in assignment 2.
9/5/25	- Collect the data after consultation - Exploration of Gephi	- Share results of what we found on listed social media platform - Confirm two social media platform – Reddit and YouTube - Explore the function of Gephi for graph analysis - Consult with Lida at 13.45	- We shared the result of how to scrape the data from social media e.g., YouTube, Facebook, X, Thread, Instagram. - We talked about the limitation of each social media platform, likes we need to pay if we want to scrape the data from the Mega Co. - We discussed about Graph software for creating the graph, e.g., NetworkX package, Gephi.
13/5/25	- Plan for the next consultation - Deciding the final topic question - Finalised the targets movies	- We tried to scrape the data from YouTube video & Reddit submission - Confirm the topic and problem question	- After quick searching on Reddit and YouTube, we found that there is lack of discussion about those worst 3 movies.

		- Set the target task for each member: - Tai > How to collect the data from YouTube and choose to go deeper with Alice in Wonderland - Sirius > How to collect the data from YouTube choose Maleficent - Nadia > How to retrieve the data from Reddit and choose Snow white	- So, we decided to change the target movies for our analysis, which are the top three movies with the largest total number of ratings on IMDB, to make sure sufficient data could be scraped from both social platforms.
17/5/25	- Reconfirm the Report topic & section - Set the target tasks for each member - Finalised how we retrieve the data from Reddit and YouTube	- Finalised those 3 Disney movies with largest number of ratings in IMDB - Allocated tasks to each member - Decided all members applied the python codes from Tai to scrape the data from YouTube - For the data from Reddit, we decided to apply our own codes based on assignment 1 - <u>Consultation with Angel at 13.45</u>	- We discussed research questions. - We listed some questions to clarify in consultation. - We shared how we can construct the graph in Gephi. - We decided to retrieve the data of each movie from Reddit and YouTube separately - So, we finalised the necessary data type that we need to make sure our data is under the same criteria.
19/5/25	- Online meeting - Sharing the data size of each movie from Reddit and YouTube - Types of graphs that we needed for the report	- Decided focus on graph analysis, community detection, sentiment analysis and topic modelling for each movie first	- We discussed the content of submissions for each movie. - We shared opinions about our analysis that make comparisons of submission and comments of each

			movie from Reddit and YouTube.
20/5/25	- Update the progression of drawing replyGraph - Set the calculation standard for the centrality and community detection for each movie	- Shared the know-how of using Gephi and decided on the layout and templates of the graphs we'll use - Make sure we applied the same parameters of calculation for different movies	- We checked each other codes and output data. - We shared own difficulties and tried to resolve together. - We discussed and helped detecting anomalies. - We examined our data together and found that some data with number of nodes larger than the number of edges, so we investigated how many submissions we have retrieved from platform to consider if it is possible.
23/5/25	- Set the writing structure of the final report - Confirmation of types of graphs we needed in the report	- <u>Consultation with Angel at 14.00</u> - Decided to mainly focus on interesting part of each movie in the analysis - Confirmed we might try different perspectives to analysis each movie	- We shared some graphs of communities with Angel and asked for suggestions. - Also, we reconfirmed the node represented unique user ID on that social platform and the edge was each comment and reply from a user.
28/5/25	- Documentation for the whole project - Finalised the formatting and the report	- Combined our analysis parts together - Discussed the analysis results together	- Tai focused on formatting the report, e.g. scale of writing, denotation of figures and tables - Nadia focused on drafted the introduction and conclusion part

			- Sirius focused on summarized the group documentation - We have summarized the coding - We wrote about own reflections
29/5/25	- Clarification for final submission	- Consultation with Angel	

Section in Details

Initial plan-structure of the report

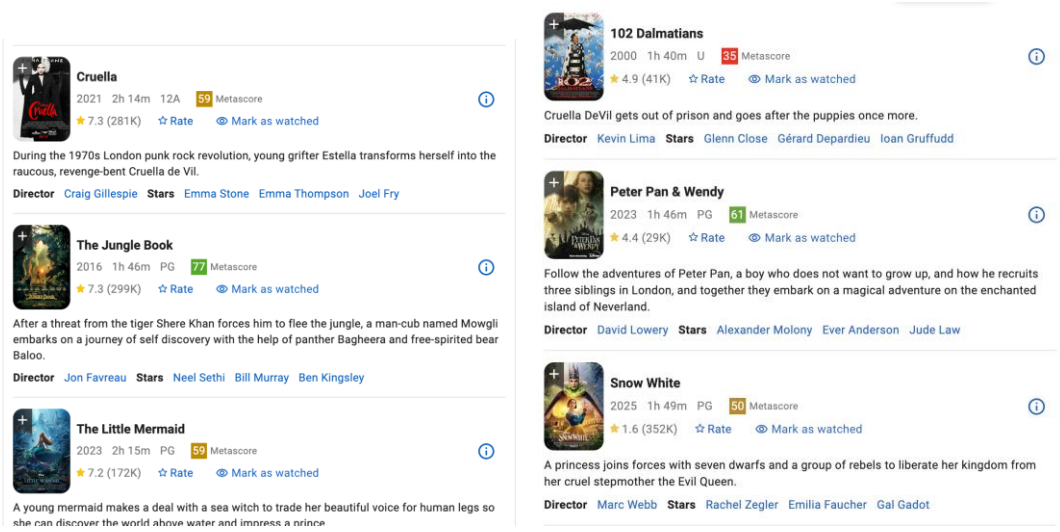
- Problem question > Person and famous topic in the community
- Defining platform to scrapethe data
    - Learn how to scrape the data from each platform
- Data Preprocessing
- Data exploration
    - Graph analysis
    - Community analysis
- Sentiment Analysis
- Topic Modelling
- Can do time series analysis

Potential Topics

1. Films with Many Superstars eg., **Harry Potter (Potter head),** (https://medium.com/data-science/basic-nlp-on-the-texts-of-harry-potter-sentiment-analysis-1b474b13651d)
2. **Marvel**,
3. **Taylor Swift** (https://edition.cnn.com/2024/12/08/business/taylor-swift-eras-tour-economy/index.html) --> Swiftie community
4. **Disney movie** > Main character (for graph analysis) and sentiment analysis about that movie / maybe we can compare this to another disney movie? (the really famous one)
    - Eg. Why the box-office of Snow white drop tremendously comparing to other princess series (https://www.boxofficemojo.com/title/tt6208148/, https://www.boxofficemojo.com/title/tt2771200/)
    - https://www.imdb.com/list/ls000069513/
    - Top 10 Best and Worst (Live-action) Disney movies (to narrow the scope)
        - https://www.imdb.com/list/ls062279726/?sort=user_rating%2Cdesc
        - **See if the stars impact the movie popularity**

Best 3:

Worst 3:



(IMDB, 2025)

5. Trending tech companies in social media eg. Apple Amazon Facebook
6. Amanda Nguyen - The Blue Origin NS-31 sub-orbital spaceflight took place on April 14, 2025, making Nguyen the first woman of Vietnamese heritage to fly into space

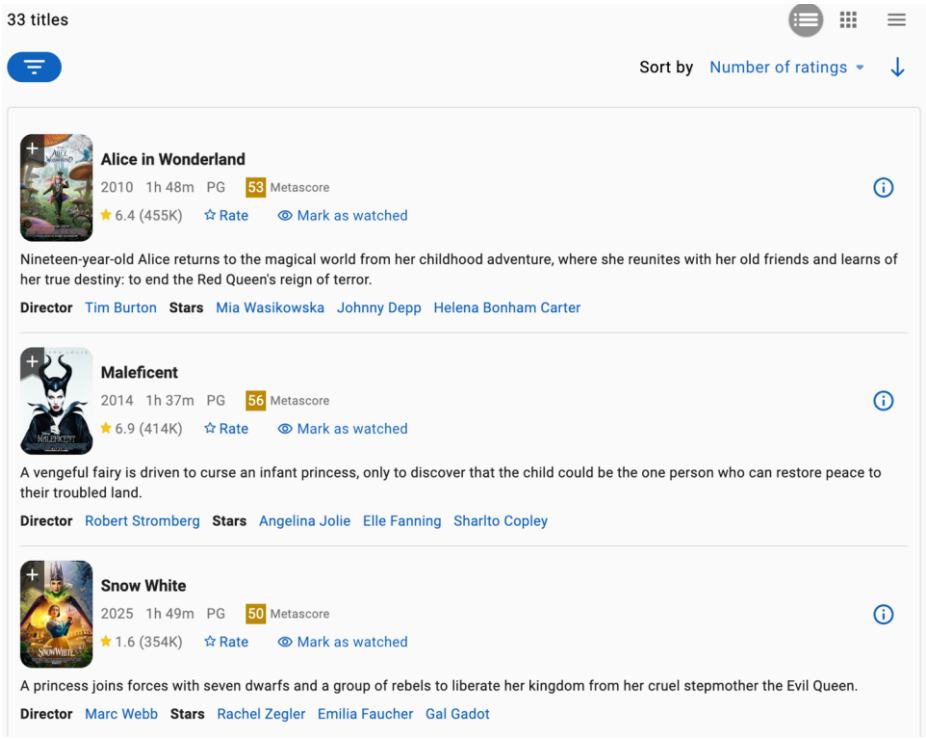
Consultation Question on 9/5/25

1. Confirm the topic (Disney – top 3 and worse 3 movies)
2. Size of the dataset (how large)
3. What graph analysis can do, apart from following/follower/ connection --> reply graph
4. What kind of information do we need to retrieve from social media in order to analysis the graph and detect the community. Users who comment all the time., user proactive

Clarification from Lida on 9/5/25

1. reddit community
2. influent of the user does not have high degree
3. link YouTube to another link > built the graph communities
4. reddit reply graph, reply graph on YouTube doesn't work that well
5. 100K -200 K posts and comments from Reddit and YouTube
6. What promotes the movie?
7. Mainly focus on graph analysis

Research Question



(IMDB, 2025)

1. **Study the success factors of Disney-Liveaction movies (Top 3 Number of ratings: Alice in Wonderland, Maleficent, Snow White):** This consists of good and bad movies.
2. **What makes Alice in wonderland and Maleficent better than Snow white**
3. **Official trailer > our methodology**
4. If there is any community (or fanbase) among these movies?
5. If there any influencers among different movies?

Social Media Platforms

YouTube:
https://developers.google.com/youtube/v3/quickstart/python
Reddit:
List of the potential subreddits
'shittymoviedetails', 'disney', 'disneyprincess', 'movies', 'MovieSuggestions'

Disney Live-action Movies

Alice in Wonderland (2010) – Tai
Maleficent (2014) – Sirius
Snow White (2025) – Nadia

Data Size from Reddit and YouTube

Regarding to Reddit, we did not define the specific number of submissions since we found that the submissions with the number of replies are vast for different movies. Hence, we attempted to retrieve as more relevant submissions as possible for the purpose of fulfilling the minimum requirements of the data size.
Conversely, we examined the official movie trailer for each movie initially and found the average number of comments and replies from trailer are over than 5000. If we set the number of videos to more than 10, the size of the data will be very large. Due to the limitation of computation ability, we decided to narrow down the number of videos to 3 for each movie. Additionally, we decided to scrape the same type of videos, including official movie trailer, movie review and any related video of that movie for the last one to make sure the data remains objective.

All Tasks

Coding Part

Data Collection

- Scrape YT DATA: Snow White
- Scrape YT DATA: Maleficent
- Scrape YT DATA: Alice in Wonderland
- Scrape Reddit DATA: Snow White
- Scrape Reddit DATA: Maleficent
- Scrape Reddit DATA: Alice in Wonderland

Graph Analysis part (Same code)

- Centralities and community from movies
    - Follow Week 8 and 9 workshop Snow White: Find the centralities and community detection
    - Follow Week 8 and 9 workshop Maleficent: Find the centralities community detection
    - Follow Week 8 and 9 workshop Alice in Wonderland: Find the centralities community detection
- Reply Graph

Data Preprocessing for sentiment Analysis and Topic modelling
Sentiment Analysis and Topic modelling part

Measure for calculation

Base on Katz – for building the reply graph
Base on Louvain – for the community detection graph

Reporting part

- Introduction
- Methodology
- Analysis Results
    - Alice in Wonderland

        - Graph Analysis
        - Sentiment Analysis and Topic Modelling
    - Maleficent
        - Graph Analysis
        - Sentiment Analysis and Topic Modelling
    - Snow white
        - Graph Analysis
        - Sentiment Analysis and Topic Modelling
- Conclusion