1. Our presentation is about disney movie’s text-mining analysis.
2. Here’s our contents.
3. First, I will have an introduction part.
4. We grew up with Disney animated films and always have special feelings for Disney stories that are full of magic and wonder. As Disney movie enthusiasts with text analysis skills, we want to know how Disney’s content style has changed over time.
5. So our research problem is How have the language usage, sentiment expression, and thematic focus in Disney animated movies evolved over time. We chose to conduct a text mining study of three productions from different time periods: The Lion King (1994)[2], Up (2009)[3], and Zootopia (2016)[4]. We found the scripts of these three movies online and also performed a data cleaning process.
6. Here’s our data cleaning step. In addition to using the classic eight cleaning steps, we also noticed that movie scripts often use EXT. (Exterior) and INT. (Interior) to indicate the context for visualization. Therefore, we used these markers to divide the three movies into chapters.
7. Next part is about text analysis.
8. This is a correlation plot with the lowest term frequency set to 100 and a correlation threshold of 0.9. As it shows, "house," "dog," and "bird" are closely linked, possibly reflecting Up's storyline, while the other words form another strong cluster.
9. We also present the top 15 most frequently occurring words for each of the three movies. Notably, house appears over 200 times in Up. It shows that the movie is strongly related to the topics of home and family.
10. To better analyze the word usage in these movies, we conducted a TF-IDF calculation. Here, we list the top six words with the highest TF-IDF values. However, we can see that these words have higher TF-IDF values because they appear more frequently in a specific movie script. The IDF value of 1.099 indicates that these words are not very common in the other movie scripts. This observation suggests that the primary words used in each movie are distinct.
11. We created comparative plots of word usage proportions for each pair of the three movies.
12. Then we also calculated the correlation coefficients for each pairwise comparison of the movies. The results show that the correlation between the three movies is generally low. The Lion King and Up have the lowest correlation at only 0.3668, which means most contents of these two movies are different.