IAE Nantes
Économie & Management
Pôle Sociétés
Nantes Université

Nantes Université

Master 2 Econométrie et Statistique, parcours

Econométrie Appliquée

NLP : Analyse de sentiments de la saga Star Wars

Mathis GIRAUD – Yohan TESSON

2 Juin 2024

Sommaire

I. Introduction

En 1977 est sorti le premier Star Wars. Son réalisateur, George Lucas ne s'attendait pas à un tel engouement du public et pourtant, le film sera le plus rentable de son époque et devenant la 2ème saga la plus rentable au monde avec plus de 10 milliards de dollars de recettes. Ce succès peut s'expliquer par des effets spéciaux qui ont révolutionné le cinéma. À la suite du succès des deux premières trilogies, Disney rachète les droits pour 4 Milliards de dollars. La dernière trilogie finira par rapporter 4 milliards de dollars au studio. Cependant, les coûts de production entre le premier et le dernier Star Wars ont été multiplié par 25 pour une augmentation des recettes de 25%. Nous pouvons ajouter à cela que le premier film a accueilli près de 220 millions de spectateurs dans le monde en salle contre près de 130 millions de spectateurs pour le dernier film. Nous pouvons donc nous interroger sur l'intérêt et l'engouement du public pour cette saga, qui, malgré des coûts de production de plus en plus importants, n'arrive pas attirer plus de spectateurs en salle.

L'objectif de ce rapport est donc, à l'aide du NLP (Natural Language Processing) et des avis Allo Ciné sur les 9 films de la saga, de faire une analyse de sentiments afin de constater ou non, une baisse d'intérêt sur les récents films.

II. Description des données

1. Importation des données

Pour ce projet, les données sont issues du site « Allo Ciné ». Nous avons récupéré les commentaires des 9 films de la saga Star Wars, donnés par des utilisateurs. Dans cette base, nous avons aussi la note de l'utilisateur entre 0,5 et 5. Pour chaque commentaire, il y a un nombre de « like » ou « dislike » qui sont donnés par les autres utilisateurs lorsqu'ils sont d'accord ou non avec le commentaire posté. Nous allons donc regarder le nombre de commentaires pour chaque film est donné dans le tableau ci-dessous :

Tableau du nombre de commentaires par film

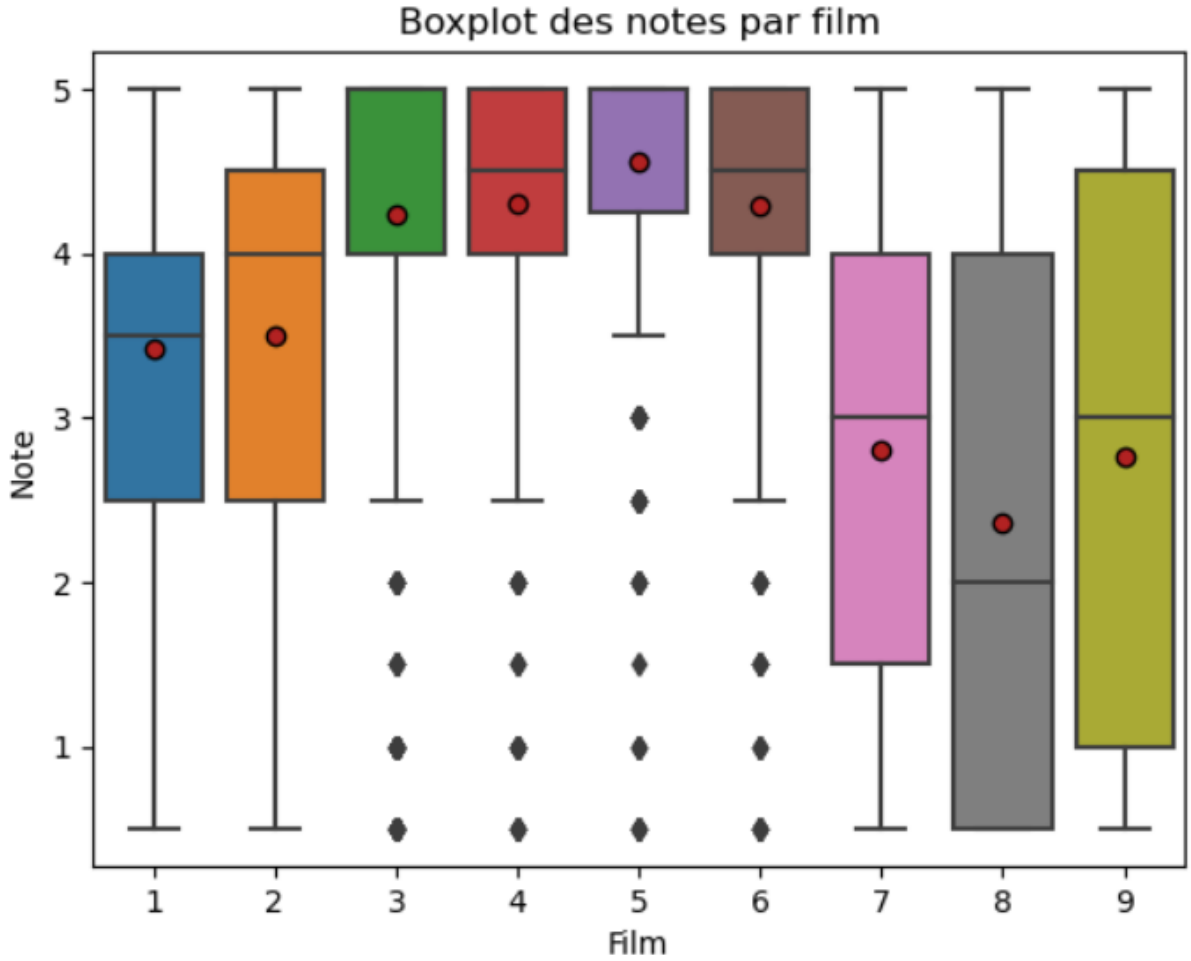
Film	Nombre de commentaires
1	1 442
2	1 112
3	2 247
4	1 122
5	1 055
6	912
7	4 711
8	4 594
9	2 740

Source : Dossier Mathis GIRAUD Yohan TESSON, NLP

Les derniers films sont les plus commentés ce qui peut être expliqué par le fait que se sont les films les plus récents (sortis entre 2015 et 2019). A l'inverse, les films 4-5-6 sont ceux avec le moins de commentaires car ce sont les films qui ont tournés en premier (entre 1977 et 1983).

2. Statistiques descriptives

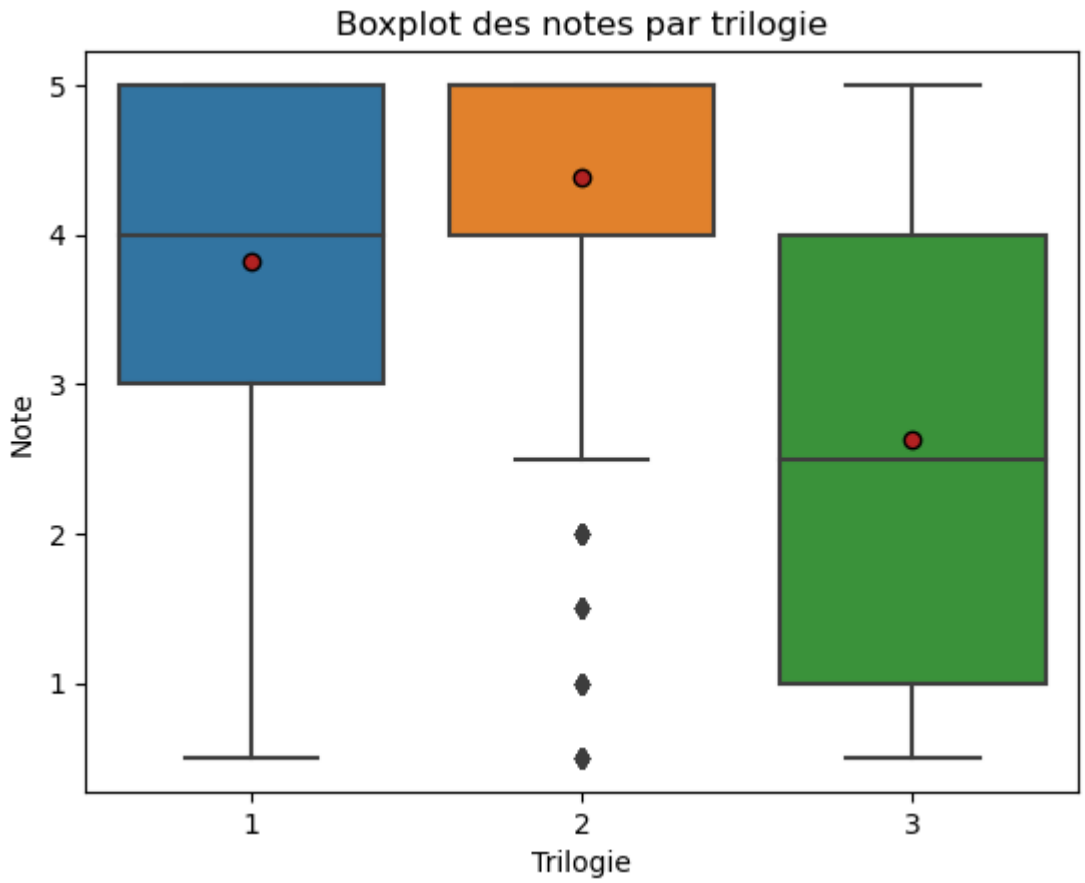
Nous allons maintenant présenter quelques statistiques descriptives de la base. Nous commençons par la distribution des notes, données par les utilisateurs sur les différents films.



Source : Dossier Mathis GIRAUD Yohan TESSON, NLP

Nous constatons que les 3 derniers films ont obtenus des notes moyennes inférieures aux films précédents. En outre, les premiers films sortis au cinéma, qui correspondent aux films 4-5-6, ont obtenus des notes moyennes entre 4,3 et 4,6, avec moins de 25% des notes inférieurs à 4. A l'inverse, pour les 3 derniers films, 75% des notes ont une note inférieure à 4. Les films 1 et 2, qui sortent 20 ans après le succès des premiers films, ont des notes moyennes de 3,5. Nous pouvons nous rendre compte que globalement, les avis sont plutôt homogènes et positifs pour les films 3-4-5-6. Cela se reflète graphiquement avec des points pour les notes inférieures ou égales à 2, ce qui nous indique ces notes sont « rares » et peu données. A l'inverse, pour les autres films, les avis sont très hétérogènes avec aucune note « rare ».

Nous allons maintenant représenter le boxplot par trilogie.



Source : Dossier Mathis GIRAUD Yohan TESSON, NLP

En moyenne, la dernière trilogie a obtenu une note de 2.6 contre 3.8 et 4.4 pour les 2 premières trilogies. En outre, la 2ème trilogie (qui correspond aux 3 premiers films Star Wars) compte 75% des notes supérieures à 4 contre 50% pour la première trilogie et seulement 25% pour la dernière. Nous pouvons donc imaginer qu'il existe des différences significatives sur la satisfaction du public sur les films Star Wars.

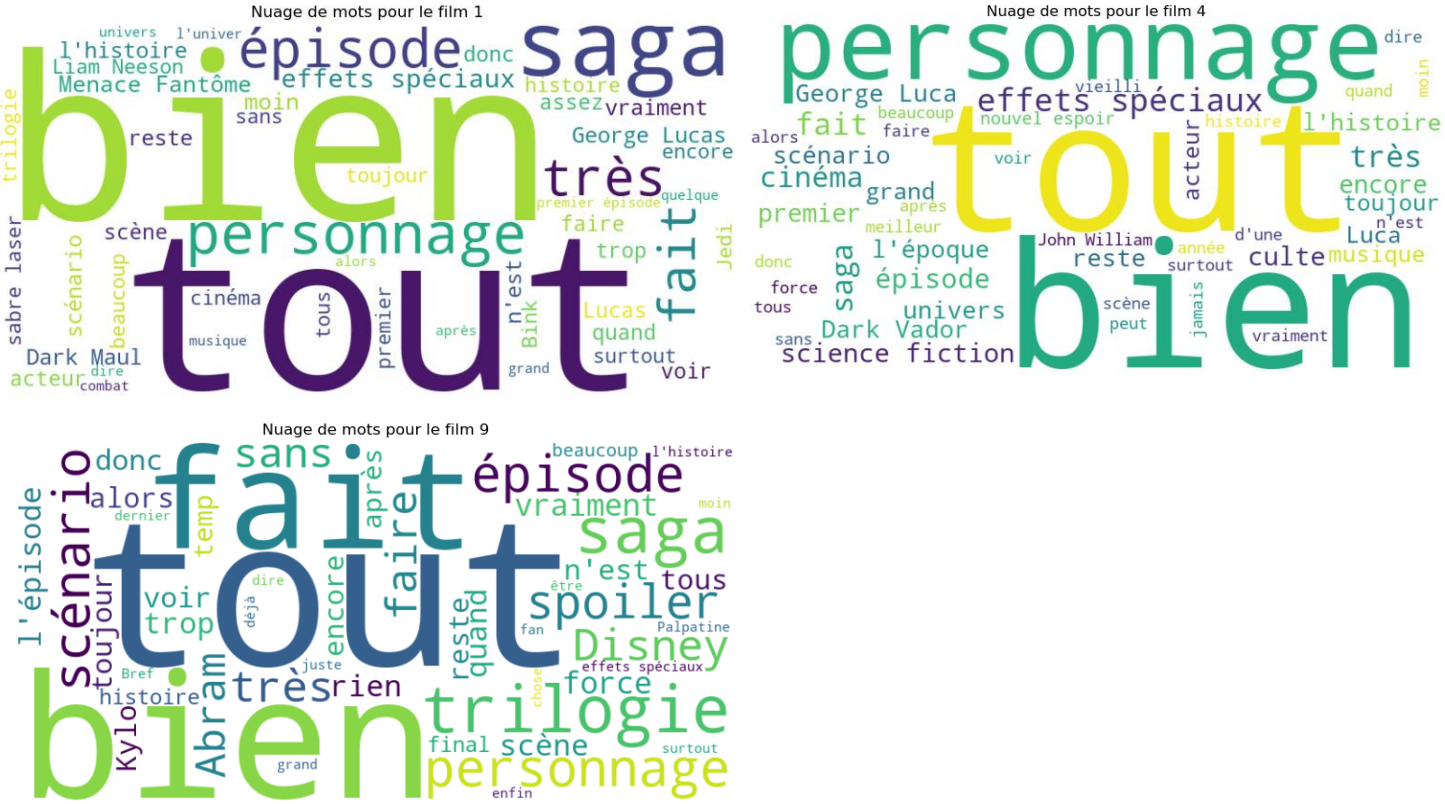
Après avoir fait une analyse descriptive des notes données par les utilisateurs, nous allons analyser les commentaires.

III. Etude des commentaires

1. Nuage de mots

Dans cette partie, nous ferons une analyse des avis des spectateurs. Dans un premier temps, nous allons représenter un nuage de mots pour chaque film. Cette représentation permet de faire ressortir les mots les plus utilisés dans les commentaires de chaque film. Nous représentons ci-dessous un nuage de mots d'un film de chaque trilogie, les autres nuages sont en annexe.

Nuage de mots de films de chaque trilogie

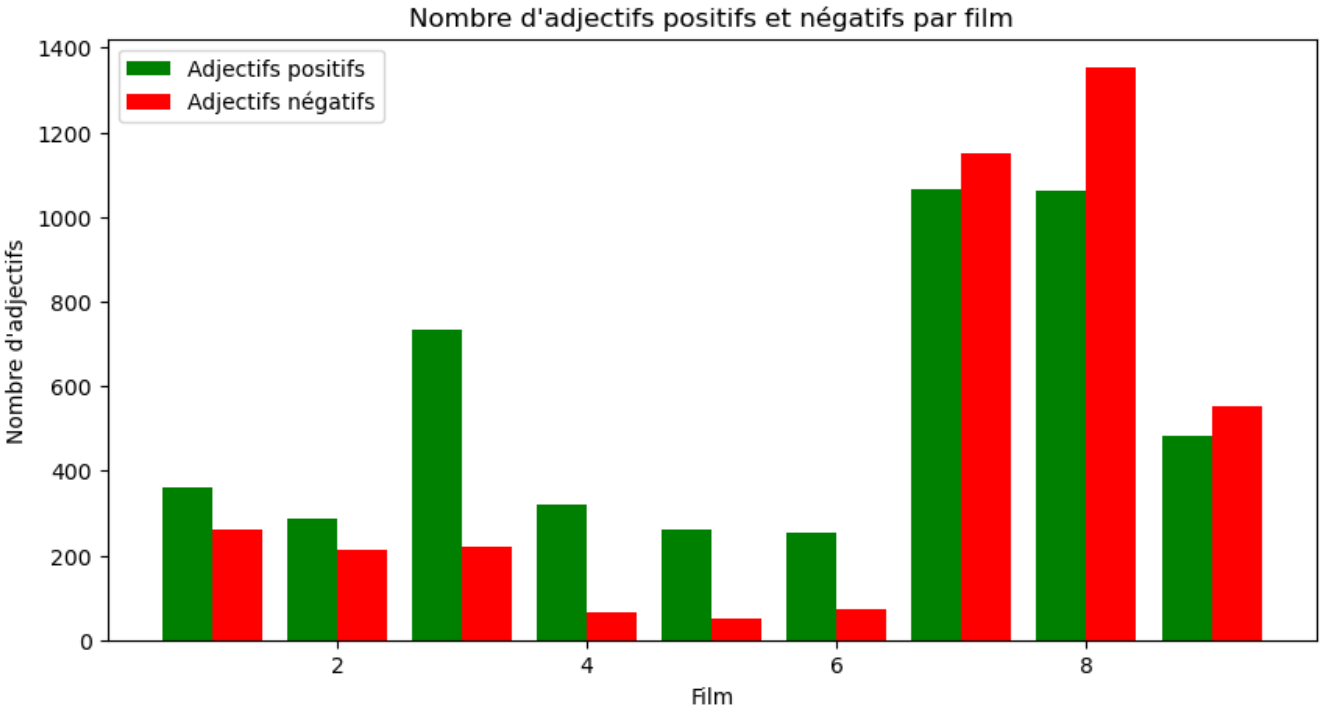


Source : Dossier Mathis GIRAUD Yohan TESSON, NLP

A la vue de ces nuages de mots, ce sont souvent les mêmes mots qui ressortent, que ce soit pour le scénario, la science fiction, les effets spéciaux… Il y a relativement peu d'adjectifs qui sont régulièrement utilisés et qui se dégagent dans les nuages de mots. De même pour les nuages de mots en annexe, nous ne pouvons constater de grosses différences dans les mots utilisés par les utilisateurs pour décrire les films.

2. Nombre d'adjectifs positifs et négatifs

A la suite des nuages de mots qui n'ont pas donné de résultats probants, nous allons compter le nombre d'adjectifs positifs et négatifs par film. Cela nous donnera une plus fine idée de l'avis des utilisateurs sur les différents films.



Source : Dossier Mathis GIRAUD Yohan TESSON, NLP

Nous ne comparerons pas le nombre d'adjectifs positifs et négatifs mais le rapport entre les deux car tous les films n'ont pas le même nombre de commentaires. Les deux premiers films ainsi que les trois derniers ont un rapport d'adjectifs positifs/négatifs très proche voire, plus d'adjectifs négatifs pour la dernière trilogie. Ainsi, avec ce graphique, nous pouvons confirmer que les films les plus récents ont été moins apprécié par le public que les 3 films sortis en premiers.

3. Analyse de sentiments

À la suite de l'analyse du nombre d'adjectifs positifs et négatifs, nous allons passer à l'analyse de sentiments sur les différents films. L'analyse de sentiments consiste à déterminer le sentiment général d'un texte grâce au NLP. Cette analyse de sentiments est réalisée avec l'outil Vader sur Python. Cet outil attribue un score de sentiment à chaque mot, ce qui permet de donner un sentiment général d'un texte. Le tableau ci-dessous recense le sentiment général, la proportion de commentaires positifs, négatifs et neutres, par film.

Tableau de sentiments des films Star Wars

Sentiment	Film 1	Film 2	Film 3	Film 4	Film 5	Film 6	Film 7	Film 8	Film 9
Général	-0.27	-0.20	-0.24	-0.20	-0.17	-0.21	-0.25	-0.34	-0.22
Neutre	0.91	0.92	0.92	0.93	0.92	0.92	0.92	0.91	0.92
Positif	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Négatif	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.05

Source : Dossier Mathis GIRAUD Yohan TESSON, NLP

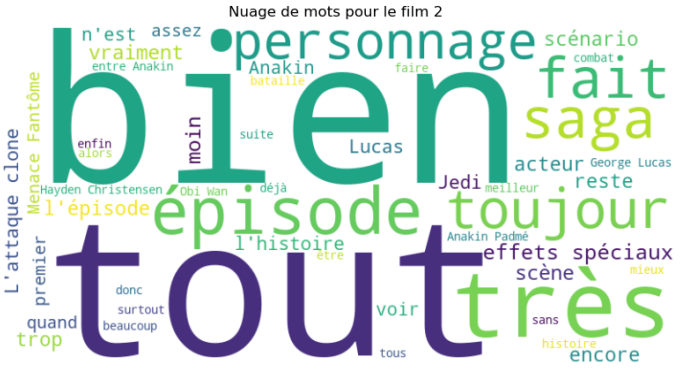
A première vue, il n'existe pas de grandes différences sur le sentiment général des avis donnés par les utilisateurs sur les différents films. En effet, cela peut s'expliquer par le fait qu'à 90%, le sentiment des avis est « neutre ». Seulement 3% sont détectés comme positifs et 5-6% comme négatifs. Ainsi, l'outil Vader détermine un sentiment général négatif, quelque soit le film. Malgré cela, les films 1-7-8 sont les films ayant un sentiment moyen le plus négatif. Cela reflète bien les représentations des boxplots précédemment et de l'analyse du nombre d'adjectifs positifs et négatifs.

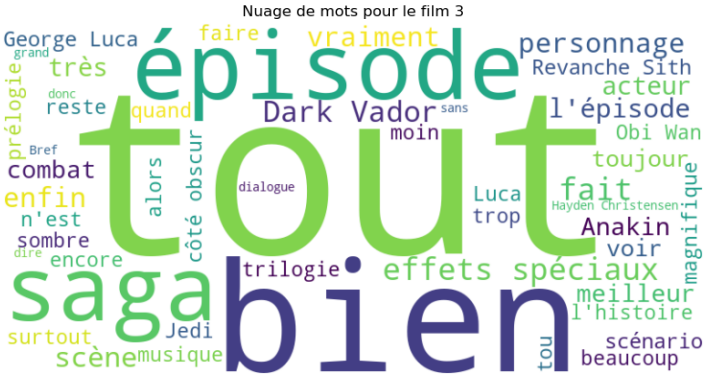
IV. Conclusion

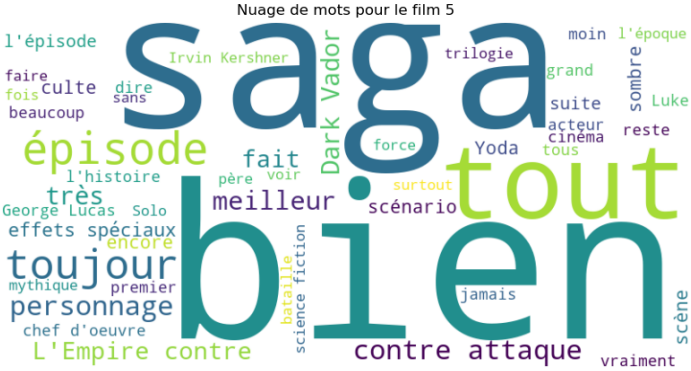
L'objectif de ce projet était de faire une analyse de sentiments sur les 9 films de la saga Star Wars. Nous avons commencé par représenter graphiquement les notes données par les utilisateurs aux différents films et avons constaté de meilleures notes pour la 2ème trilogie et des notes plus faibles pour les derniers films. Dès lors, nous avons réalisé des nuages de mots pour faire ressortir les mots les plus donnés par les utilisateurs. Cependant, cette analyse n'a pas donné de résultats probants. C'est pour cela que nous avons décidé de représenter graphiquement le rapport nombre d'adjectifs positifs et négatifs. Cette fois-ci les résultats étaient plus intéressants et confirmaient notre première impression des boxplots. Enfin, nous avons fait une analyse de sentiments pour analyser les commentaires. De manière générale, les avis ont été plutôt neutres.

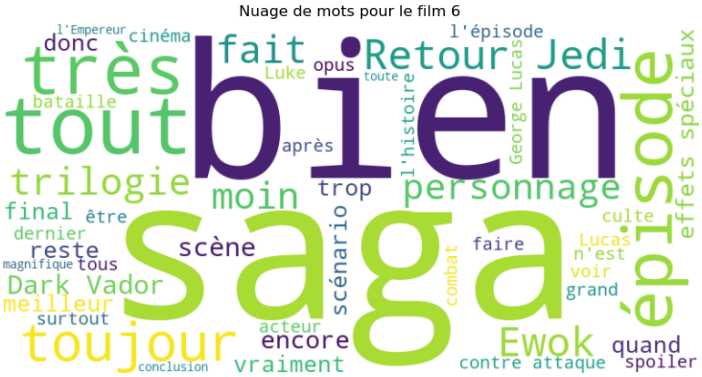
V. Annexes

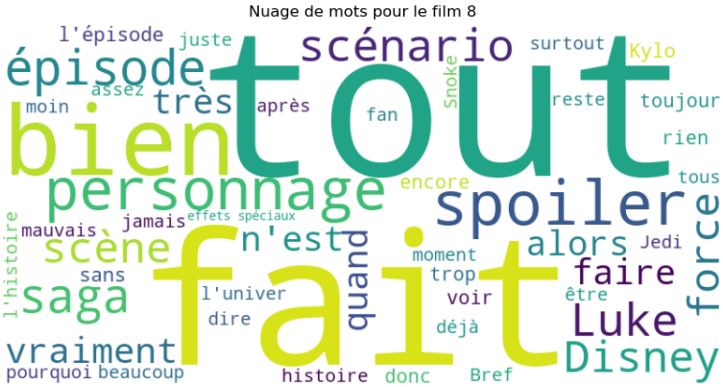
Annexe 1 : Nuage de mots











VI. Bibliographie

Star Wars Franchise Box Office History - The Numbers (the-numbers.com), The numbers

Ilyes Talbi, NLP avec python, Introduction au NLP avec Python pour l'analyse de sentiments (larevueia.fr)

Lavanya Geetha, Vader, Vader : un guide complet de l'analyse des sentiments en Python | par Lavanya Geetha | Douleur moyenne (medium.com)

Jeremy Robert, NLP Twitter - Analyse de Sentiment - DataScientest

Table des matières