

Sentiment Analysis

Analyse automatique de la tendance d'un texte en fonction de son contenu

Le corpus de texte rassemble les transcripts des sketches de 12 comiques Américain. L'objectif est de faire une analyse de type "Sentiment Analysis" sur ce corpus pour mettre en évidence les différences entre les personnes.

Ce projet met en oeuvre les techniques courantes du Natural Language Processing (NLP)

- Tockenisation,
- Stop words
- Stemming / lemmatization
- Regex
- Document-Term Matrix - word counts in matrix format
- Exploratory Data Analysis (EDA)

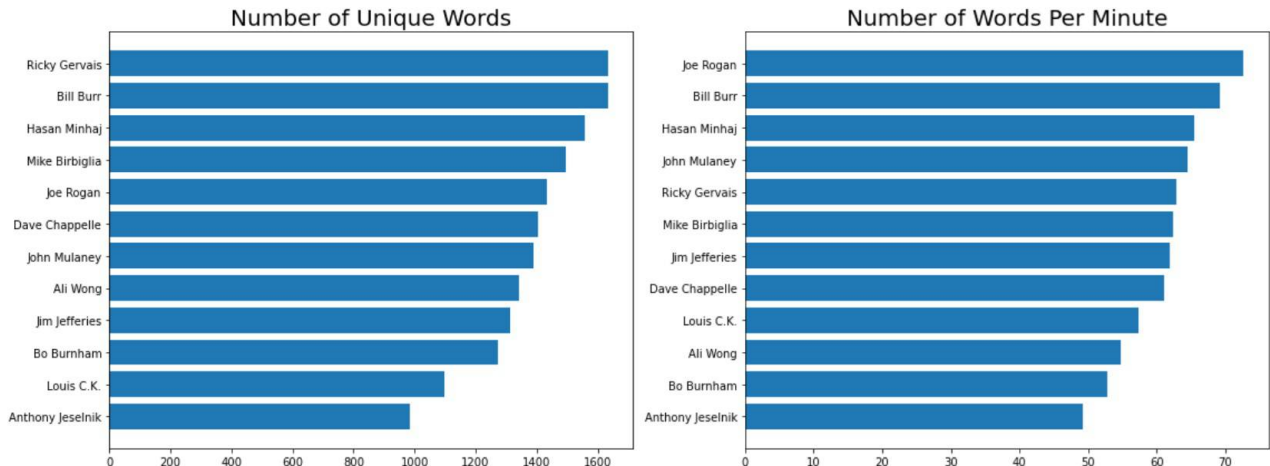
Technologies : Nltk, Textblob.sentiment

Corpus de mots pour chaque acteur :

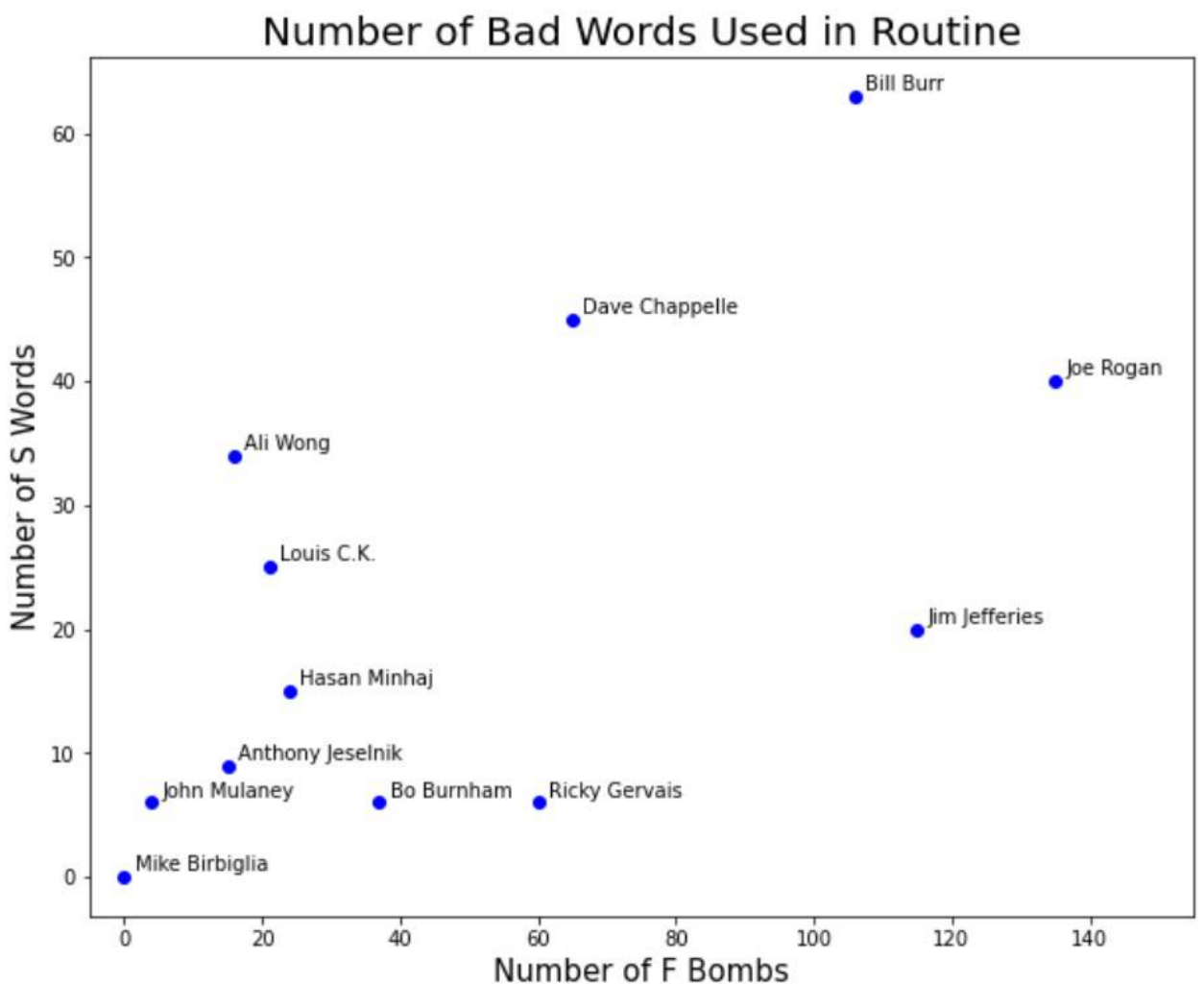


EDA : exploration des données

Diversité du vocabulaire et vitesse



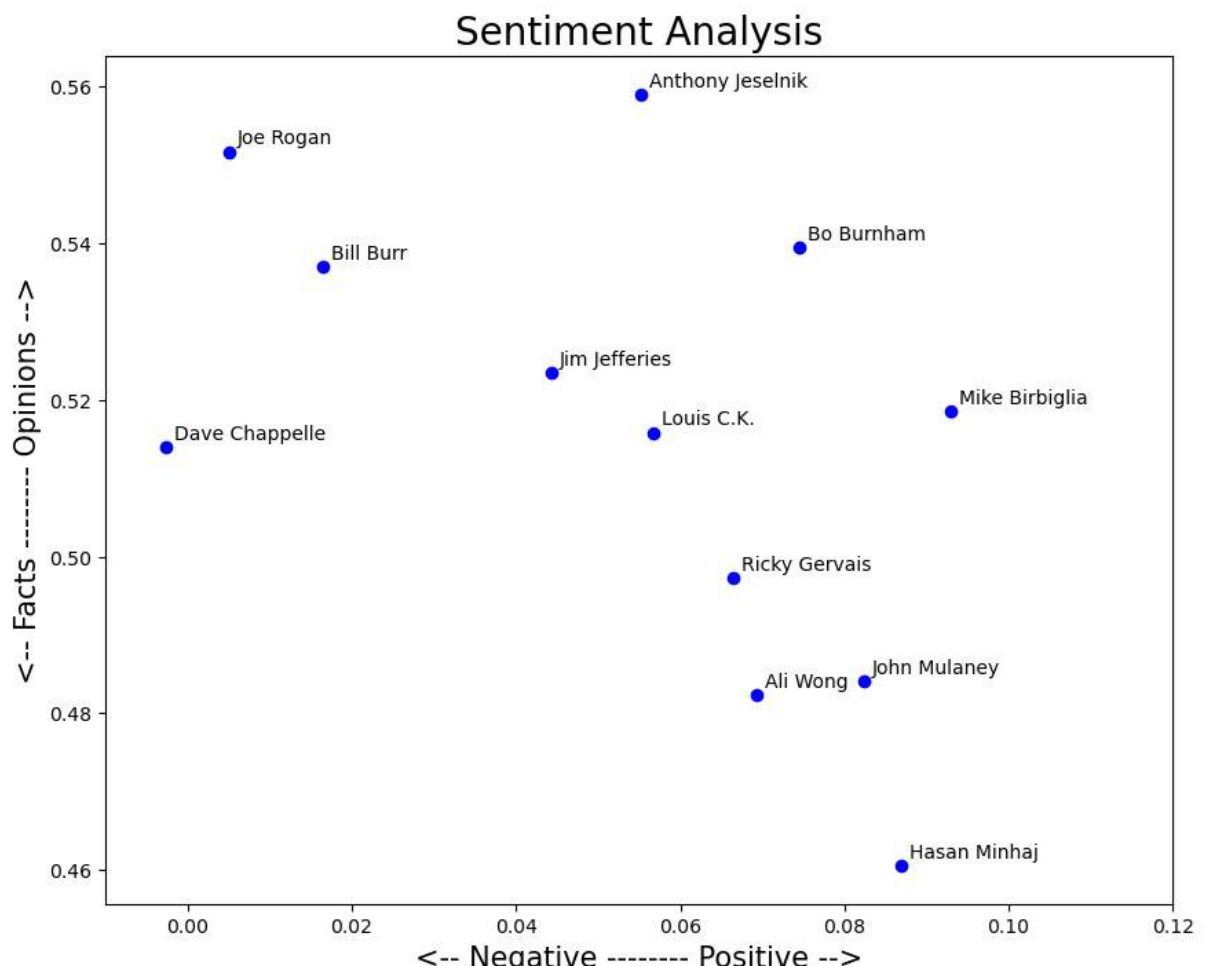
F-Word et S-Word (Bad words)



Sentiment-Analysis

Textblob.sentiment nous donne deux axes, la polarité et la subjectivité :

- Polarity: How positive or negative a word is. -1 is very negative. +1 is very positive.
- Subjectivity: How subjective, or opinionated a word is. 0 is fact. +1 is very much an opinion.



Evolution dans le temps de cette analyse (les sketches dans le corpus commencent par les plus anciens jusqu'au plus récent)

