

« Grand Débat » synthesis

The goal of this project is to generate a synthesis of the datas created by french citizens to answer some political open questions.

Like the datas, the project is open : you don't know what you will find yet !

All datas are available on french open data web site :

<https://www.data.gouv.fr/fr/datasets/donnees-ouvertes-du-grand-debat-national/>

Three mains points that your project must contain :

Pretreatments

You need to manage datas on french language and adapt all your pretreatements.

You need also to manage a certain amount of noise in the answers, people made mistakes.

unsupervised analysis

most representative words, tokens, tf-idf, resumes, LDA, Language models... all techniques seen can be applied in order to represent information in a new way.

Correlations with announcements

Subject is open, but in order to precise a result, you will compare your analysis with the annouces of Emmanuel Macron :

Here, is a open summary (made manually by some reddit users) which details all topics covered by the Emmanuel Macron :

https://www.reddit.com/r/france/comments/bhb76w/r%C3%A9sum%C3%A9_des_annonces_demmanuel_macron/

Your main goal is to put in perspective what you can find with your analysis and show a corelation between some topics and announcements but also decorelations in other ones.