

Project for SD 701

Louis Jachiet

Proposition 1

Movie Recommendation

Use the Movie Lens dataset (or any other dataset) to provide Movie Recommendation.

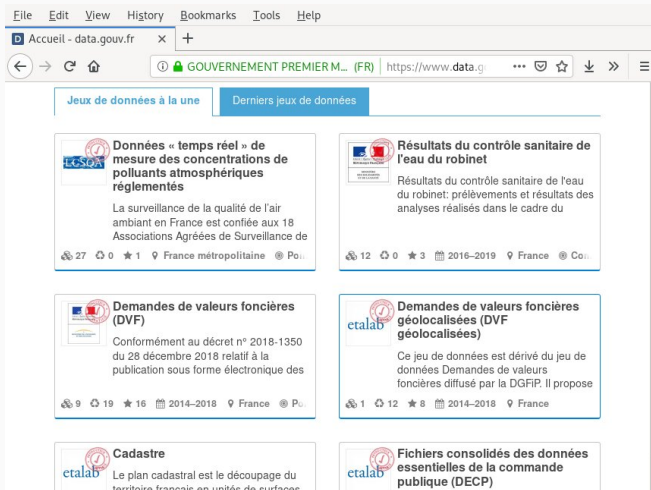
Ideally, your code should scale even with:

- billions of users
- billions of movies
- hundreds of ratings per user

Proposition 2

Use Data Mining On Open Data

Create an interesting Data Mining pipeline using an open dataset.



Output

You should send a zipped folder with:

1. A PDF (1-3 pages) explaining what you did, how you did it, how you could improve your project.
2. The code to run what you did.

All by the 15th November.

Three sessions will be dedicated to help you in your task.

02/10 familiarization with Spark

16/10 or 23/10 hands on the project

06/11 exam + help on the project