Project 1
OpenFoodFacts

OpenFoodFacts can be considerate of a wikipedia for food !
The goal of OpenFoodFacts is to share with everyone a maximum of informations on food products.
It contains more than 800000 products but maybe all products are not perfectly described...
Mainly, for a product, we can find the list of ingredients, nutrition facts and food categories.

1) Define and clean the vocabulary of ingredients, do you find some mistakes ? How do you manage them ? Propose solutions to manage/identify errors.

2) Based on nutrition facts and/or food categories, propose clustering approaches and a visualisation of some categories of products. Find outliers (a product very different from others of the same group). It exists products very similars in terms of nutrition facts but very different in terms of categories or ingredients ?

3) Based on your expertize on this dataset, propose and describe a model (no code required) that would be interesting to enhance the OpenFoodFacts project.

Your report must explain what technics/approachs you use, how you use them and the results obtained. If an approach don't work as planned you can show and explain (It will be very appreciate).
You can work in pairs of students. Your report must contain the names of students involved.
Your report must explain the logic of your approachs and results.
You can write in English or French.
Your report must contain your link to your Colab Notebook. Your Deposit must contain a copy of your Notebook.
Your report must be deposited on DVO before **13 november**.

DataSet (near 4 Go): https://fr.openfoodfacts.org/data with different file formats available.