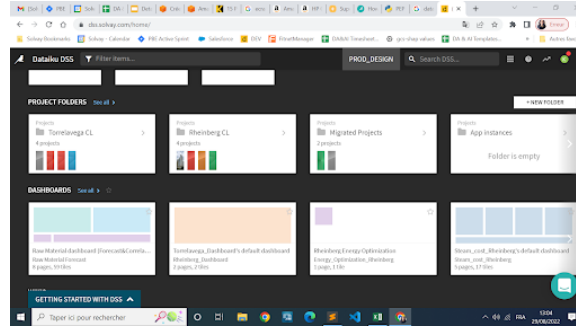


Start a Data Science Project

Step 1 - Create a new folder for the project

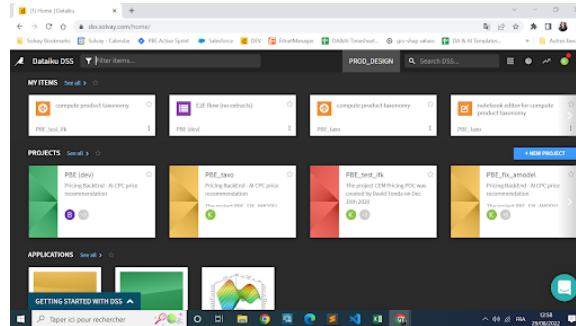
This folder contains all the different projects you will need.

You must store the projects here. This ensures the DSS space is correctly organised for all the Data Scientists in Solvay.



Step 2 - Create a new project on DSS

- Create a project on Dataiku containing its name.
- Keep it short to avoid bugs.
- Create a master project and a dev project for beginning.



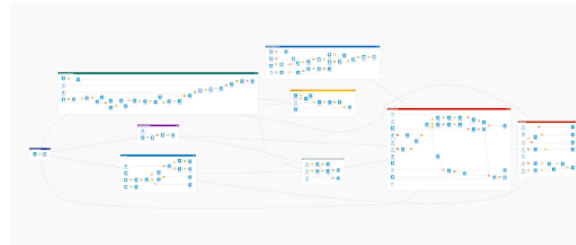
Step 3 - Organising the Flow

Have a structured flow by separating the different steps in dedicated zones :

- Data Extraction
- Data Preparation
- Data Exploration
- Modelling
- Units tests

Tip: You can have several zones for the same step if needed.

For better visibility you should use the same colour for each step.



Structure of recipes

Short description:

Have a business explanation of the things done in the recipe when you hover it. (so we don't need to open a recipe to understand what's happening or where is the root of an error)

Clear title:

Short text explaining why this recipe is needed and what are the expected outputs

Comment every step of the code

(keep in mind that everyone should understand what the part of the code is doing)

Functions:

Respect the pep8 convention (details what is in input and also expected outputs), you can store all your functions in a dedicated Library.

Units tests:

Regroup the units tests in a zone of the flow, [here's](#) the documentation.

Global documentation: [here](#)

QA / code review:

Before merging a branch, each data scientist must have his code been reviewed by another data scientist.

Code Review

Objectives:

- Identifying errors
- Opportunities for development