Nom / Prénom : _____   **email** :
_____

1. Data Science :

Provide a definition that seems the most appropriate to you, in regards of your understanding

_____

_____

_____

2. Car classification

For this exercise, you will use:
- o *Data Science Experience*
- o *Python 3.X*
- o *Machine learning*

Create a python notebook (EMGrenoble_studentname.ipynb) and send the solution (export notebook) with your comments back to us at :
**christophe.burgaud@fr.ibm.com**

First, you must connect to your account, then create a new empty project. Click on the "Asset" button (1) and then "add to project" button (2) and add a notebook.
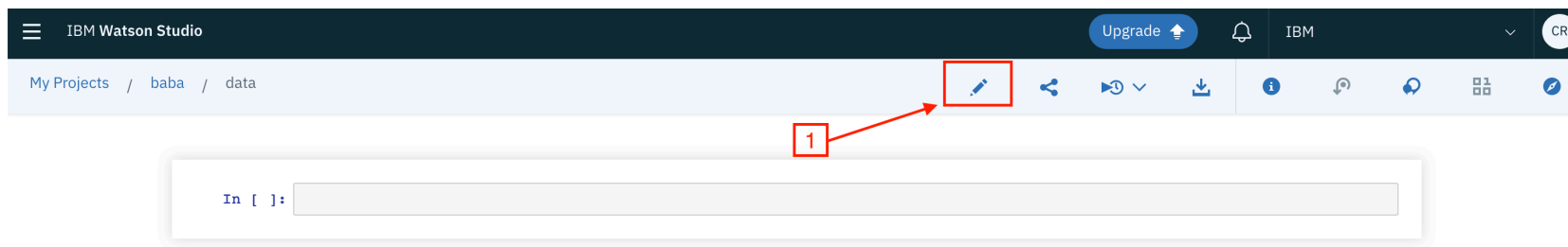
Come back to your project (3)
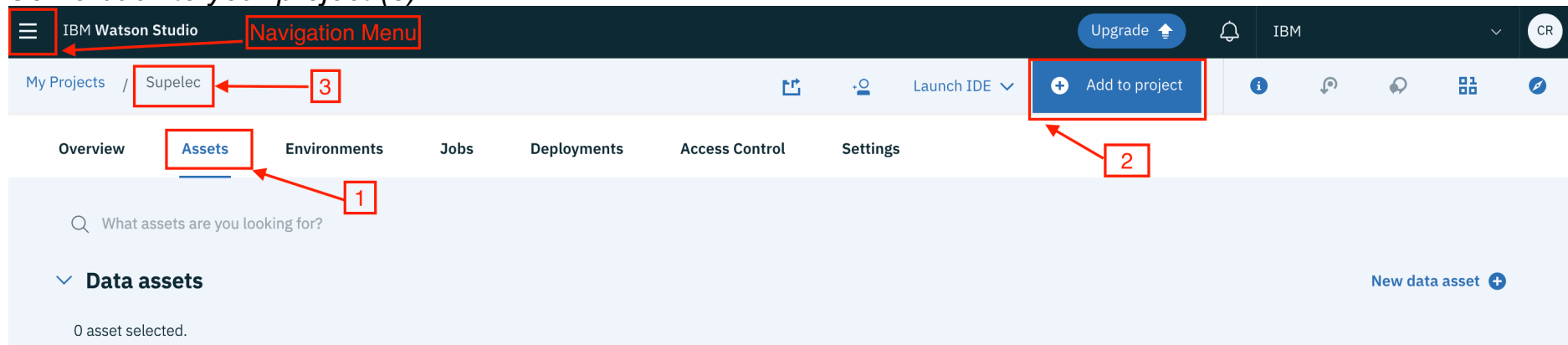
*Use the dataset "**Car performance data"**, to find it you need to go to "Navigation Menu" then to "Gallery". Write "car performance" into the search bar and select "data set" in the drag-down menu "All filters".*
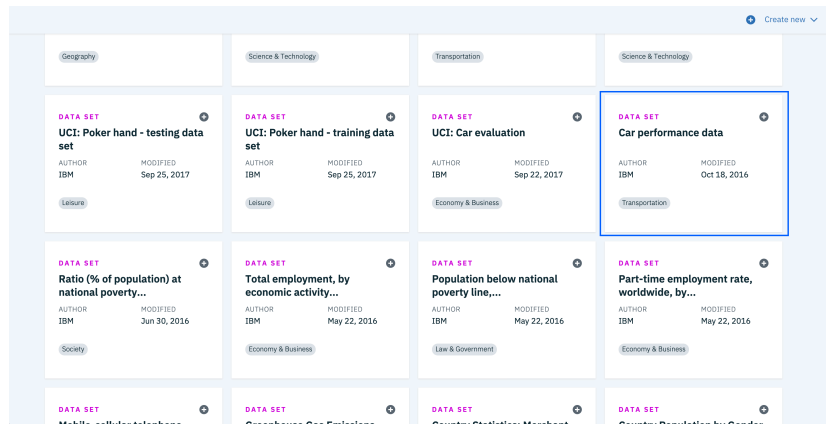
Click on it then click on the "add to project" button, select your project and click on "View project"

*Go to your notebook and click on the pen button (1)*

You must use Python and data science libraries: sciki-learn, Pandas, XgBoost… to analyse the dataset and create a model to predict car consumption efficiency (MPG or miles per gallon)

- *Non-Efficient car: car with MPG <=23*

- *Efficient car: car with MPG >23*

Suggested steps:

a) *Read the data into a pandas DataFrame by using integrated tools.*
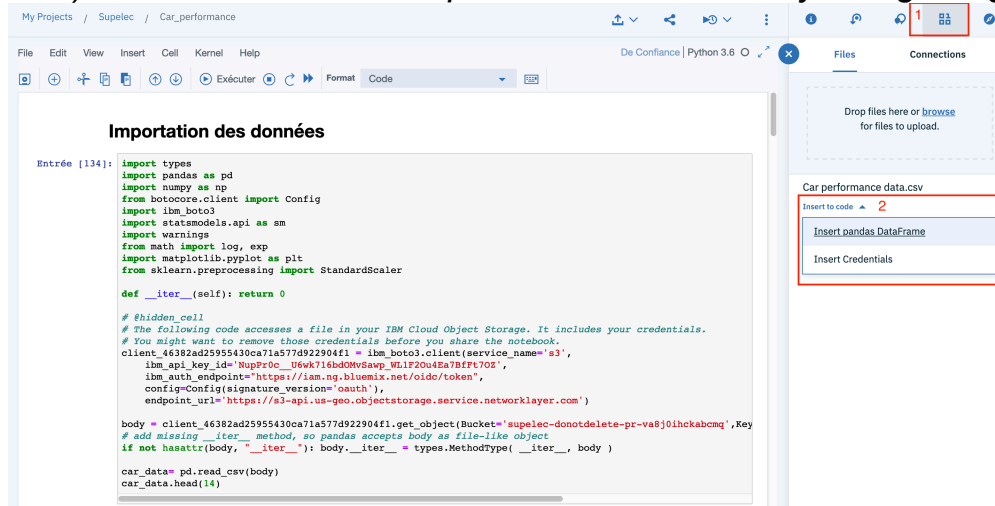
b) *Visualize the columns: distributions, empty values, correlation, descriptive statistics …*

c) *Feature engineering:*
- *create the target column (car efficiency)*
- *format the predictive columns (features)*
- *…*

d) *Create the model*

- *Create the classifier*
- *Train the classifier on a training set*
- *Validate the classifier on the testing set using classification accuracy*