**Question 4**: *“How should the project look like, where should I start, how much effort should I put in?”*

**Important dates**

**1st April**: creation of groups and submission of subject (email both group members and subject to Omar and myself)

**10th May**: submission of project (pdf report, code/notebooks, datasets)

**13th May**: Presentation (10-15’ per group)

**Project information**

**Dataset:**

It is up to you to find the dataset. You can use for instance:

* Data you collected by your own (via survey, web scrapping)
* Data you have from company X or Y (but check NDA conditions)
* Data you get from the web (*e.g.* Web APIs)

Rather choose a dataset that is of interest to you (*e.g.* related to your field of study).

**Template/guidelines:**

Imagine you are consultants and you need to promote your skills of data scientists for analysing data and generating value to the business. Stakeholders have a particular question they want to answer, or at least a broad view about it (*e.g.* “we have a real churn problem in our company, our customers are leaving us, what can we do about it?”)

*Typical steps to be followed:*

* Presentation of the business problem, why is it relevant, why should I care?
* What value are you expecting to bring with your analysis?
* How do you translate the business problem into a data problem?
* Data description (data source, transformations, descriptive statistics)
* Feature engineering (can we build new features that are informative to answer the business problem we are facing)
* Modelling (justify the technique, defend methodology used, what are the pros and cons)
* Evaluation of the model from different perspectives (statistical/ML performance, interpretation of results, running time, …)
* What value do you actually bring to the business with your analysis? Could you figure out a way of quantifying the value generated?
* How can your analysis be reused in the future – how would you deploy the model in practice?
* Presentation of results – communicate results to a broad audience

**Additional information**

Report: at least 20 pages of report (although, *quantity is not quality*!)

Presentation of your project: 10-15 minutes