

# Data Science Workshop 2015/16: Project Guidelines

---

## Project phases:

1. Pick a business goal/problem for either customers or for the platform (Microsoft).
2. Formulate hypotheses
3. Build a data model (preprocessing)
4. Implement/Choose the correct ML algorithms and methods over the data model to output results.
5. Evaluate your model statistically.
6. Present your results in a friendly, useful UI e.g: Auto generated report, a dashboard, preferably but not limited to MS Excel ( You can also use iPython Notebook or a web development platform of your choice).

---

## Requirements:

1. The project should use a significant amount of the data.
2. The UI needs to support updating, or a fast regeneration of the results, given new data
3. The insights should be meaningful to the customers/providers.

---

## Project administrations:

1. Teams will use Trello to divide and manage their tasks, and be ready to present it if asked to.
2. Teams will use GIT for collaborative development, and be ready to present it if asked to

---

## Submission:

1. Demo presentation of the work.
2. Your code, including the UI commented and well structured.
3. Document (up to 6 pages) describing thoroughly your work, explaining your choices of models , algorithms and methods and your results.
4. A usable interface accessible online.

---

## Grading:

### **Usefulness**

- Interesting goal
- Meaningful insights
- Useful UI

### **Data model**

- Preprocessing efficiency
- Data representation and model

### **Algorithms and methods**

- Algorithm choices
- Implementation
- Code cleanliness, readability and structure (OOP)

### **Evaluation techniques**

- How did you evaluate your results
- Correctness of statistical models