

## Priorities and requirements

This documents outlines a few adjustments due to the progress all groups made so far. It indicates the minimum and a few optional requirements to be implemented.

### 1. Data processing

This is the first and one of the most important steps in a Machine Learning case. In real life, you will be confronted with incomplete, messy and difficult datasets. Learning how to properly deal with this is a very important part of Machine Learning.

#### **Adjustment:**

- For now, this task only entails data cleaning and preparation.

**This step is a minimum requirement.**

Tips:

- Use the notebook on GitHub to get started (for the crime dataset).

### 2. Data visualization

A major part of understanding data lies in its visualization. This is the next step to learn how you can gain insights from your data.

**This step is a minimum requirement.**

Tips:

- Check kaggle kernels for some tips on how to visualize:
  - <https://www.kaggle.com/PROPPG-PPG/hourly-weather-surface-brazil-southeast-region/kernels>
  - <https://www.kaggle.com/inquisitivecrow/crime-data-in-brazil/kernels>

### 3. Insights

Data processing and visualization leads to insights about your data.

I.E. for the weather data set:

- Temperature data
  - Monthly average
  - Min temperature, max temperature per month
- Precipitation data
  - Monthly/yearly average
  - Min precipitation, max precipitation

Crime dataset:

- What regions of the city have a high density of crimes?
- Has there been an increase or decrease in crime throughout the years?
- Which populations are most likely to be perpetrators?

**Adjustment:**

- Since the crime data is in portuguese, we'll keep the insights high level, and won't go too much into detail: (i.e. the difference between violent and nonviolent crime).

**This step is a minimum requirement.**

**Tips:**

- Use clustering to see if you can identify high density regions (check github/week12 for ideas)
- Keep in mind that the model you train for prediction for the weather dataset should use features it will be able to use for the coming year (i.e. "tmin" and "tmax" won't be available for the year you're trying to predict)

#### 4. Prediction and exploration

- Generate predictions for the weather in th coming year using techniques from the course (Random Forests, Neural Networks, Linear regression).
- Optimize and compare your models.
- Find correlations in the data.

**This step is a minimum requirement.**

**Adjustments:**

- Prediction about crime will become an **extra**. If you can gain insights through clustering, you can use these clusters to obtain a classification problem.

#### 5. Website and recommendation

Keep this extremely simple. Even the CSS part is optional.

**Priorities:**

- Data insights and graphs.
- Very basic interaction and simple prediction for weather data

**This step is a minimum requirement.**

## Evaluation

- Data analysis / ML code
- Simple REST API and front end
- Presentation + short demo & visualizations, insights
- Written report (~5 pages)
  - README.md
  - Short summary
  - Retrospective++ of the project
    - Mad, Sad, Glad?
    - Pitfalls?
  - Personal reflection of each member

**Deadline:**

- Last commit: 28/05/2019 at midnight!
- Presentation: 29/05/2019

AI & Robotics Final Presentation of the group project	
Week 13:	
	Wednesday May 29, 2019
13u00 - 13u30	
13u40 - 14u10	
14u20 - 14u50	
15u00 - 15u30	Group 4
15u40 - 16u10	Group 1
16u20 - 16u50	Group 5
17u00 - 17u30	Group 2
17u40 - 18u10	Group 3