



# DATA MINING TO-DO LIST

Maddalena Amendola

Daniele Gadler

Riccardo Manetti

Gemma Martini

October 8, 2018

Task Index	Task	Person	Output [Python code + Plots.pdf + explanation]	Remarks	Status	Deadline
1	Determine the quality of the data wrt: - Syntactic Accuracy  - Semantic Accuracy	Maddalena	- Analysis of the syntactic accuracy of attributes in the data  - Analysis of the semantic accuracy of attributes in the data		To-Do	15/10/2018
2	- Detect and examine missing values possibly hidden by default values. and identify possible data normalization  - Check specific application dependent assumptions	Riccardo	- Analysis to identify missing values in the dataset, with an explanation of the identified missing values and eventually plots thereof  - Plot of the distribution of attributes in the dataset and density of attributes in categories		To-Do	15/10/2018
2	- Find outliers via boxplots and understand how to deal with them	Gemma	- Analysis of the outlier data points in the dataset wrt. different attributes and explanation on how to handle them		To-Do	15/10/2018
3	Discover new or confirm expected dependencies or correlations between attributes.	Daniele	- Scatter matrix, scatter plot and bar plots of the attributes in the dataset and textual explanation of the results obtained via the scatter plots	wrt. target attribute	To-Do	15/10/2018
4	Create a report	Gemma/ All	Beautiful Report with Cherubino logo		To-Do	15/10/2018

## 1 Assignment 1 — Data Understanding