



# Model to structure a Project in Data Science

Let's Go

# Data Science Process

- Define the business problem clearly
- Set successful criteria
- Define a clear goal

- Perform predictions
- Show results – Visualizations adapted to the context
- Draw the conclusions & Report your insights
- Tell a story effectively – Who is your Audience?

- Try multiple ML algorithms
- Choose the optimal algorithm
- Choose the best hyperparameters of the chosen algorithm
- Metrics & Error evaluation

Knowledge of Business

Collect the Data

Data Preparation

Exploratory Analysis

Modeling

Evaluate & Communicate

20% of the work

80% of the work

- Identify data sources
- Extract the data into usable format

- Break business problems in data science problems
- Identify ML problem categories
- Understand data points and constraints
- Detect and handle Missing values and Outliers.
- Perform required transformation

- Perform statistical and visual analysis
- Formulate data analytics strategy
- Descriptive analytics

# Define your Problem and Data Science process

---

- **Problem Definition:** To be able to tell a story you should be able to develop the problem statements and frame them. Then use those as an **Introduction** of your report.
- **Data and Process:** Define your goals and describe the approach and the methodologies you used to achieve them. In the report include the **Goals definition**, the **Data sources**, the **Data description** and the **Data Science process** adapted to your specific project.

# Manage your Data: Understand, Clean & Transform

The Data is rarely ready to be analyzed, normally it must be cleaned and transformed but first must be well understood. This step is called **Data pre-process**.

- Remove **useless** & **duplicated** data (columns or maybe rows). Example: would it make sense to have negative mass values in a physic problem?
- Detect and handle **Missing values**
- Detect and handle **Outliers**, be careful those must not be always removed.
- **Unify** and **transform** the variables if necessary. Example: Units transformation  $m \rightarrow cm$ ,  $W \rightarrow mW$  etc.
- **Create new variables**, from existing ones, for example:  $v = v_0 + at$ .

# Exploratory Analysis

---

- Check variable **distributions**: Depending on the type of the variables (numeric, categorical) different kinds of visualizations should be used.
- Look for variables relationships between variables and visualize them, use those that are meaningful.
- Consider **reducing dimensionality** or use Feature Selection: PCA, RFE,...
- Don't forget your **target** and **choose** those visualizations that are **meaningful** and **insightful** for the report. Sometimes less is more.

# Modelling and Performance evaluation

---

- Identify your **target** or dependent variable:
  - Is this **categorical** or **numerical**?
- Try different **Machine Learning Algorithms** based on your target:
  - **Classification**
  - **Regression**
- Choose the algorithm with **best performance**. Try different hyperparameters and avoid **Overfitting** and **Underfitting**.
- **Error evaluation** using visualizations in the report are always interesting, but only those that are meaningful.

# Predictions: Report your results

---

- Once the model based on a Machine Learning algorithm is created the **Predictions** can be performed. Some tips:
  - Visualizations with the units of the variables
  - Well labelled charts (axis, titles, subtitles, etc.)

# Tell a Story effectively from the Data

## Live presentation:

- Who is your Audience?
- Clear message
- Use the data to make the point
- Less detail – use Visuals

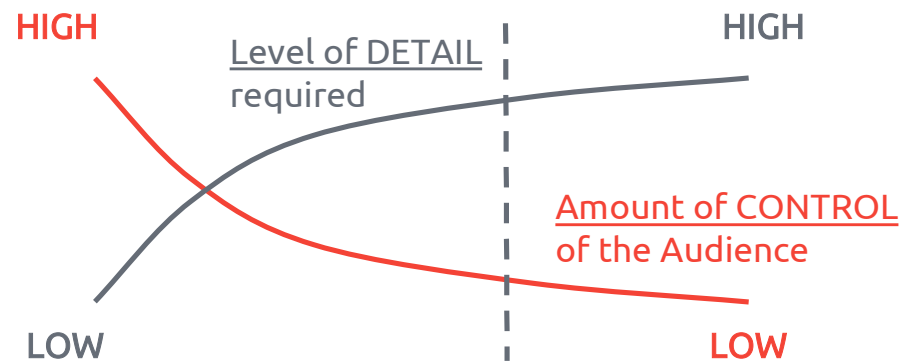
## Written Document – Report:

- Follow the Data Science process structure
- Include the meaningful insights
- More detailed Analysis

## Live presentation



## Written Document



## "Slideument"

Presentation that resembles a document with copious information.



# Conclusions, Recommendations, Next Steps

---

At the end of your Presentation or Report expose your **Conclusions**

- Answer the problem question of the project
- Finish your story successfully: Sum up using a nice and proper Visualization for example
- Point out some possible improvements and/or recommendations for next projects



**Thanks**