

# Project

Data Analytics Technology 2024  
DEI, FCT, University of Coimbra



## Deliverables and deadlines

### 1. **Part 1:** Pitch – 22 October

- Submit a .pdf file into *Inforestudante* with your presentation slides.
- Deliver a 5-minute presentation in class

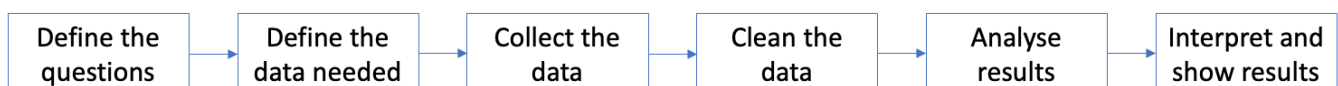
### 2. **Part 2:** Poster – 19 November

- Submit a .pdf file into *Inforestudante* with your presentation slides.
- Deliver a 10-minute presentation in class

### 3. **Part 3:** Report – 6 December and Discussion – 9-10 December

- Submit a .pdf file into *Inforestudante* with your final report (December 6)
- Submit a .zip file with source code (include description of use on the report) (December 6)
- Submit a .pdf file into *Inforestudante* with your presentation slides (December 9-10)
- Deliver a 15-minute presentation/discussion in a session (December 9-10)

Consider the following proposed project workflow:



## Part 1 - Pitch

A pitch is a short presentation to describe an idea, regardless of the type of idea and the degree to which its has been developed. In this case consider a non-technical perspective, essentially think of “selling” an idea to a possible investor. In this case consider you are seeking approval for your idea from the teacher to pursue your project. Your pitch presentation should include:

- Motivation and context
  - Present company/institution/stakeholders
  - Present the business: products, services, location, clients
  - How can a data analytics solution can help in decision making (new information, how can use it, etc.)
- The BIG questions
  - identify the most relevant questions that your solution can answer
  - include not only descriptive approaches but also predictive capabilities
- The Data Analytics Solution
  - A diagram with the proposed solution (end-to-end, from **data sources** to decision support)
- Expected results
  - Present and discuss the results of the analysis
- Summary/Wrap up

Guidelines: more images than text, be objective, remember to “sell” your solution, present it as a complete product, describe how it will work, how it can be implemented, potential costs and benefits, **do not forget to be enthusiastic!**

Important: 5 minutes are strict!

## Part 2 - Poster

The poster can have more information, a more technical approach definition. In this case consider you are presenting to your technical team to obtain feedback on the approach to achieve the results. Your poster should include:

- Major requirements: since this project is all about data, we will focus on a subset of the full range of requirements usually associated with a software project. Make sure to have at least 2 types of analytics (descriptive, diagnostic, predictive, prescriptive, decisive).

- User interface (mockups in e.g. <https://gomockingbird.com> or <https://balsamiq.com/>): in this stage, we will be focusing on the how data can be combined and exhibited on your solution. Use the mockup tool to describe how users will be interacting with the data and how information will be displayed. Make sure to have at least 2 types of interaction, e.g., time-based, location-based.
- Workflow from data sources, ETL, star model, analysis

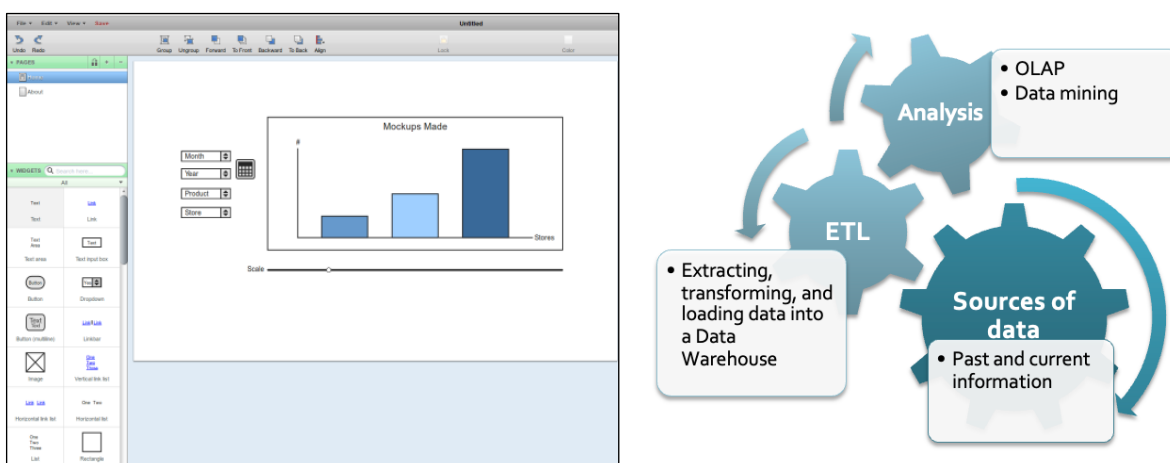


Figure: Examples of interface and workflow images

### Part 3 - Report

The final step before discussion should include the complete data analytics solution, with technical details, rationale for decisions, and analysis results. The project should be described in detail, explaining the choices/options you made, and discussing the results.

The report should include details on:

- ETL (Identify and describe the sources of data; Present the overall ETL plan for your solution)
- Explain how the ETL process is automated in order to allow the future updates of the data in the DW, what is the update strategy for the dimensions and the facts tables and its frequency.
- Data Warehouse (DW) design: star model including facts and dimensions, attributes, and granularity
- Explain the use of the DW to answer the requirements/BIG questions, describe staging area
- Describe and justify software used for a) storing data (e.g., relational database), b) perform the ETL process, and c) do OLAP

- For each major action in the plan explain why it is necessary and how it is implemented
- Present the major challenges to the implementation
- Present the following metrics: size of the source data, size of data on the 1<sup>st</sup> load, size of data on the subsequent loads, time elapsed on the first load, time used for each update
- Identify problems with the source and DW data that were not dealt with (if they exist)
- Describe how the OLAP data is presented to final users, how they can access it and modify the search parameters
- Describe the analyses being performed
- Present and discuss the findings
- Discuss results from the business perspective and explain how the information can be used for Decision Support
- Explain the strategies and techniques used for optimizing the DW and the OLAP queries performance (views, indexes, partitioning, etc.)
- Describe the data mining techniques (already) applied and (if possible) compare results

Guidelines:

- Incorporate information from all the past presentations of the project. You can structure the document using the same order of the intermediate deadlines.
- Make sure to include more detail than you showed in the oral presentations.
- Make sure to comply with the feedback received during the presentations from the teacher and your colleagues.
- Do not leave significant technical details out of the report.
- Make sure to discuss findings and state their relevance in terms of decision support.