# Business Intelligence I

## Lab 1 – Introduction + Project Overview

Academic Year 2023-2024

Rui Monteiro    Pedro Batista    Bruno Jardim    Duarte Rodrigues

rmonteiro@novaims.unl.pt    lbatista@novaims.unl.pt    bjardim@novaims.unl.pt    duarte.rodrigues@novaims.unl.pt

**Instituto Superior de Estatística e Gestão de Informação**
Universidade Nova de Lisboa

# Lab Plan

- Lab calendar

- Lab methodology

- Lab software and resources

- BI I (practical) Project overview

# Lab Calendar

# Lab Calendar

## Lab general topics (subject to small changes):

**Lab 1: Program Overview (12 Sep.)**
Overview of BI I Labs, presentation of Project Guidelines and review of course software

**Lab 2: Data Warehousing (10 Oct.)**
Business Needs, designing a Data Warehouse, implementing a DW schema

**Lab 3: Practical steps in DW (17 Oct.)**
Building a Data Warehouse, implementing SCDs, Date dimensions, further SQL

**Lab 4: Introduction to ETL (31 Oct.)**
Understanding ETL, first steps with Data Integration and Transformation tools

**Lab 5: Continuing with ETL (7 Nov.)**
Developing a Full Load of the Staging Area, with integral load of dimensions and facts

**Lab 6: Improving the ETL (14 Nov.)**
Enhancing the SA load with logs, use of variables, monitoring for errors

**Lab 7: Completing full ETL (21 Nov.)**
Additional ETL to load the DW, testing full integral loads to SA and DW with SCDs

**Lab 8: Making ETL dynamic (28 Nov.)**
Use of project parameters, adding logic to allow for both integral and incremental loads

**Lab 9: Review of Lab Work (5 Dec.)**
Review and consolidation of work covered in all the Labs, demonstration of PowerBI

# Lab Calendar (cont.)

- Further details will be made available with each Lab

- Please keep in mind the following important dates:
  - We expect to make detailed Project Guide (Handout) available on 10$^{th}$ Oct. (2$^{nd}$ Lab)
  - Last week of October is Trimester break/Exam period (no mid-term exam in BI 😎)
  - First **INTERMEDIATE PROJECT DELIVERY** is due 23h59 on **19$^{th}$ of November**
  - Last **FINAL PROJECT DELIVERY** is due 23h59 on **23$^{rd}$ of December**
  - Both dates (and times) are non-negotiable; see further details on "Project section"

# Lab Methodology

# Lab Methodology

- All labs (regardless of teacher/class/time) will follow the same methodology:

  - We will explain context of WHAT IS GOING TO BE DONE

  - We will demonstrate DOING IT using a special (customized) use-case (IMSALES)

  - We expect EVERY STUDENT to follow along, in the lab; we help fix any problems

  - We make each Lab solution (solved use-case) available in Moodle, after each Lab

  - We STRONGLY RECOMMEND each student replicates – by themselves – the techniques and skills demonstrated in the Lab, in their own time

  - The PRACTICAL PROJECT basically uses these same techniques, skills and mindset to solve a different challenge (a new data problem, chosen by your Group)

# Lab Methodology (cont.)

- All labs will make use of a fictional company, to demonstrate all of the BI techniques and skills you will be learning:



IMSales is a trading company (business-to-business) that imports and exports **specialty foods from all around the world**. IMSales does not manufacture its own products, these are manufactured by their suppliers and sold to third-party retailers (Clients).

IMSales Headquarters are in Seattle, USA and there is a second office in London, UK that works as a business representation.

# Lab Methodology (cont.)

Each individual Lab teacher will have their own (distinct) Office Hours (2 hours each week).

These will be **published and announced in Moodle.**

We will also have a dedicated **Lab & Project Help forum** in Moodle, where all students are encouraged to place their questions and problems (both for Labs and the Project).

Please note that emails sent directly to teachers – *unless to do with personal (private) questions/issues* – may not be answered (or will be answered quite late…).

The ONLY OFFICIAL (monitored) channel is MOODLE.

# Lab Methodology (cont.)



| | | | | |
|---|---|---|---|---|
| ☆ | Lookups for Loading of Fact Table in DW not working | 28 de dezembro de 2 | Pedro Batista<br>29 de dezembro de 2 | 3 |
| ☆ | Error CSV file datatype four byte signed integer | 28 de dezembro de 2 | Pedro Batista<br>28 de dezembro de 2 | 1 |
| ☆ | Incremental Load - Last load date variable | 19 de dezembro de 2 | Pedro Batista<br>21 de dezembro de 2 | 1 |
| ☆ | Error during Incremental Load | 19 de dezembro de 2 | Pedro Batista<br>19 de dezembro de 2 | 1 |
| ☆ | SQL_Script_Create_Table | 19 de dezembro de 2 | Pedro Batista<br>19 de dezembro de 2 | 1 |
| ☆ | Surrogate key and Business Key (STG and DW) | 18 de dezembro de 2 | Pedro Batista<br>18 de dezembro de 2 | 1 |
| ☆ | Primary key for fact table | 20 de novembro de 2 | Pedro Batista<br>21 de novembro de 2 | 3 |
| ☆ | 11th Question Unavailable - FAQ | 21 de novembro de 2 | Pedro Batista<br>21 de novembro de 2 | 1 |
| ☆ | Project language | 20 de novembro de 2 | Pedro Batista<br>20 de novembro de 2 | 1 |
| ☆ | Project questions - datasets, hierarchies, dimensions | 10 de novembro de 2 | Pedro Batista<br>10 de novembro de 2 | 1 |
| ☆ | This week's class - recording? | 30 de outubro de 20 | Pedro Batista<br>7 de novembro de 20 | 3 |
| ☆ | Dataset size limit for the project | 4 de novembro de 20 | Pedro Batista<br>4 de novembro de 20 | 1 |
| ☆ | Whatsapp group creation for group formation? | 6 de setembro de 20 | 20 de outubro de 20 | 33 |
| ☆ | Can we entrich our starting dataset, by bringing in new data? | Pedro Batista<br>18 de outubro de 20 | Pedro Batista<br>18 de outubro de 20 | 0 |
| ☆ | P1 & P4 with prof. Rui Monteiro on Saturday Oct 15 | 14 de outubro de 20 | Rui Monteiro<br>15 de outubro de 20 | 3 |
| ☆ | Exam Mode | 12 de outubro de 20 | Pedro Batista<br>13 de outubro de 20 | 2 |
| ☆ | Saturday's class - recording? | 7 de outubro de 202 | Pedro Batista<br>11 de outubro de 20 | 6 |
| ☆ | [FAQ] Getting an error when using diagrams in SSMS? See this post, please | Pedro Batista<br>11 de setembro de 2 | Pedro Batista<br>13 de setembro de 2 | 4 |
| ☆ | Error in Setup | 7 de setembro de 20 | Pedro Batista<br>8 de setembro de 20 | 3 |

*A sample of previous Forum questions*

# Lab Software and Resources

# Lab Software and Resources

The essence of the BI work we will be executing will be Cloud-based (as explained in Lecture).

As such, there is very little to install, on your part.

The one exception is **Azure Data Studio** (as also used for the "SQL and Databases Crash Course"; only required by the time of the 2nd Lab – 10th of October; see next slides).

Almost no prior knowledge or skills are required for Business Intelligence... again, the only (minor) exception is a **basic understanding of SQL / Databases**. Obviously, the Crash Course is already sufficient; alternatives are available – see next slides.

# Lab Software and Resources (cont.)

**Azure Data Studio**
(available for Windows + MacOS + Linux)

Download and install from Microsoft:
https://learn.microsoft.com/en-us/sql/azure-data-studio/download-azure-data-studio

(on Windows, select User Installer)

# Lab Software and Resources (cont.)



Azure Data Studio with Dark Theme

# Lab Software and Resources (cont.)

- Where can I learn more about databases and SQL:

  - Wikipedia:
  https://en.wikipedia.org/wiki/Database
  https://en.wikipedia.org/wiki/SQL

  - W3Schools | Khan Academy | … (online, hands-on tutorial on SQL):
  http://www.w3schools.com/sql/
  https://www.khanacademy.org/computing/computerprogramming/sql
  http://www.sqlcourse.com/

  - Microsoft Virtual Academy
  http://www.microsoftvirtualacademy.com/product-training/sql-server

  - Online tool for 'quick and dirty' experiments
  http://sqlfiddle.com/

# Lab Software and Resources (cont.)

Cloud services used in Lab is (new) Microsoft offer:
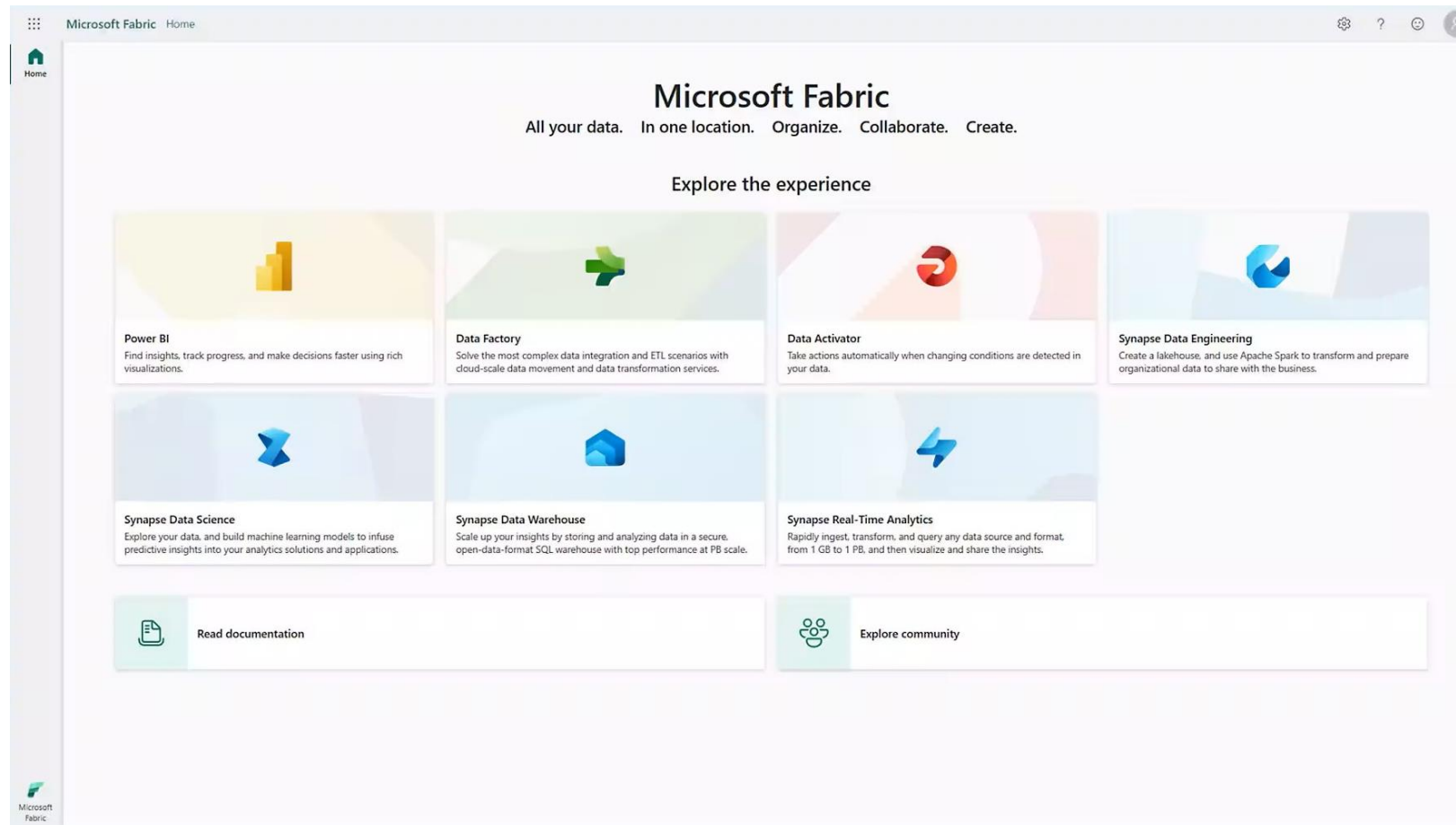
# Lab Software and Resources (cont.)

Complete + integrated Data Architecture (more details in later labs):
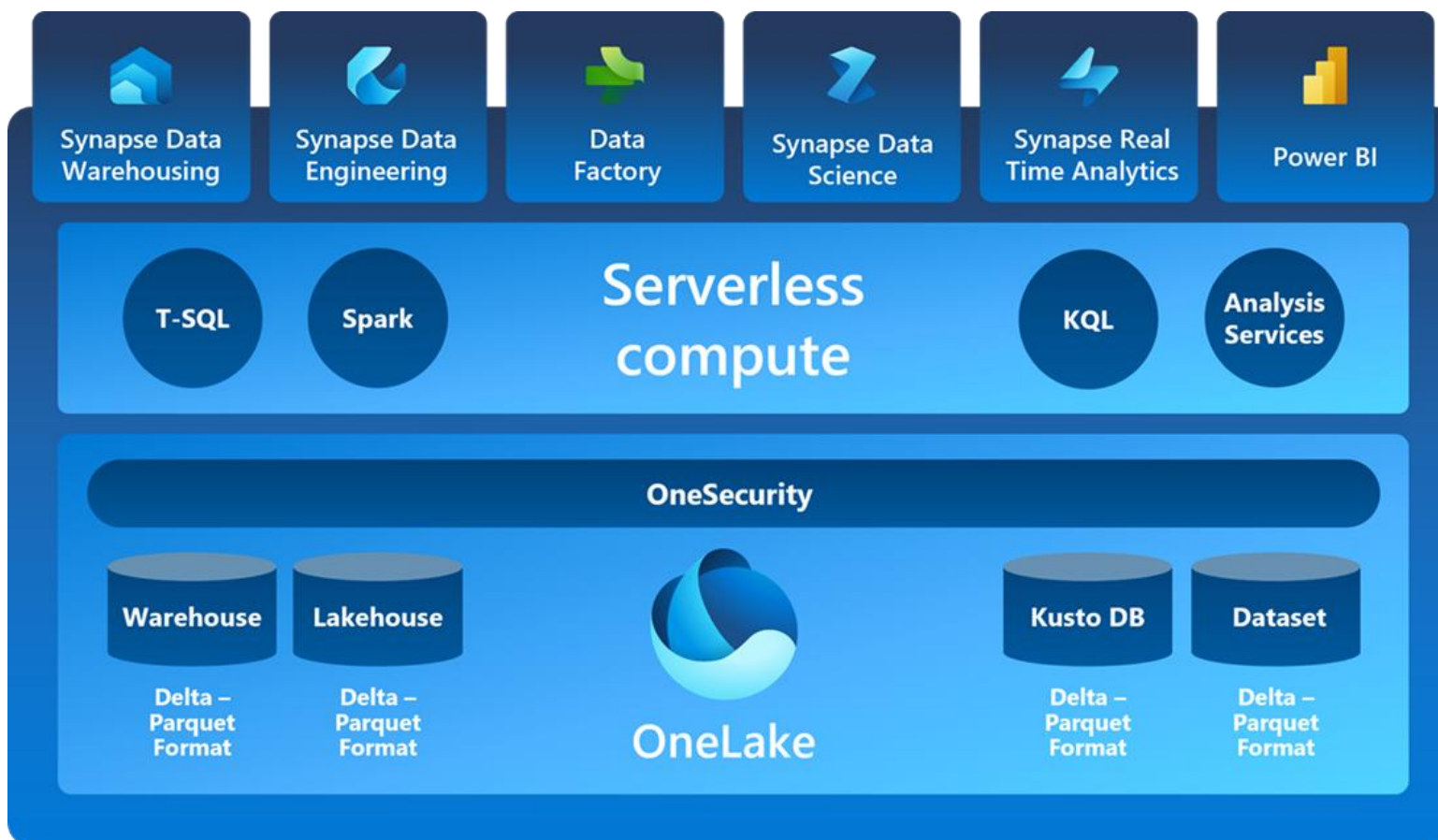
# BI I Project Overview

# BI I Project Overview

Main Goal of Project: **demonstrate you can build a BI solution!**



| Relational DB | | BD Staging Area | | Organizational Value |
| --- | --- | --- | --- | --- |
| Files | | Data Warehouse | OLAP | Reporting |
| **INGEST** | **PROCESS / TRANSFORM** | **STORE /ORGANIZE** | **OPTIMIZE / ANALYZE** | **VISUALIZE / STORYTELLING** |
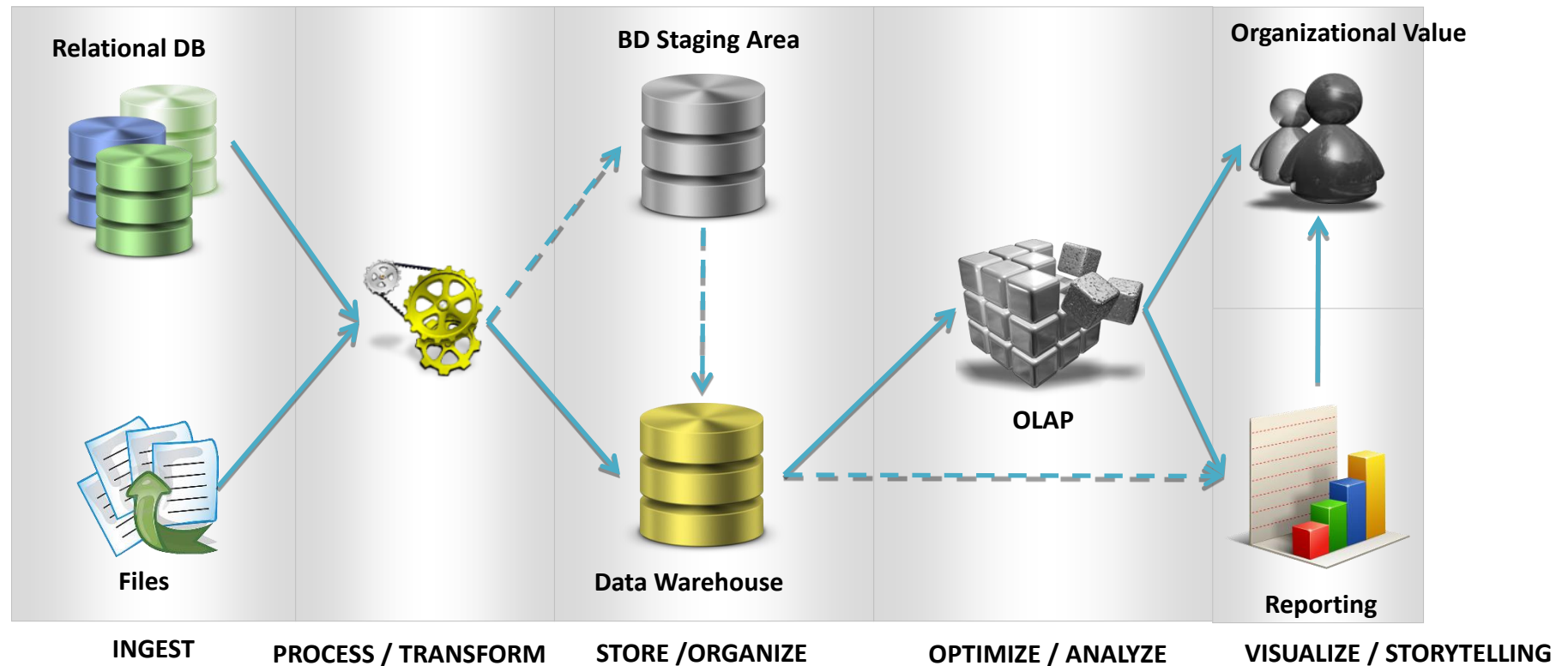
# BI I Project Overview (cont.)

- Remember: full ( definitive) details will be made available in the "**BI I Project (Handout) Guide**", to be made available on date of 2$^{nd}$ Lab (10$^{th}$ of October), in Moodle.

- Some preliminary indications:

  - Group project (in real-life, BI projects are ALWAYS carried out by teams with several members)

  - Allowed Group size will be between ONE (1) and THREE (3) members

  - In special cases – if Project is based on original and real-life context, and demonstrably complex – Lab teacher may approve larger groups (4 or 5 members)

  - There will be a Moodle resource created to make Group selection easier, but the ONLY Group that officially counts is the list of students on Project Report cover.

# BI I Project Overview (cont.)

- Project Roadmap (in general terms):
  1. Global context and explaining / illustrating the "starting data problem"
  2. Translating and documenting the Business Problem into Business (data) Needs
  3. Data Discovery and Profiling of original (operational) data sources (typically, relational databases)
  4. Data Warehouse planning, design and development
  5. Staging Area and development of transformation (data cleansing) and processing routines – Staging Area ETL
  6. Data Warehouse ETL (loading DW with data from Staging Area)
  7. Conclusions and Lessons Learnt

# BI I Project Overview (cont.)

- Delivered in two (2) distinct but complementary parts:

First Part :                    **midnight 19th of November 2023**

*Only the Design and Implementation of Data Warehouse (using SQL script) but must submit initial part of*

*the Report (so we understand what you are thinking and what you have done!)*

Second Part  :            **midnight 23rd of December 2023**

*Project Report (complete version), the Staging and DW databases and all of the ETL work*

# BI I Project Overview (cont.)

- Other things to be aware of (and may be made clearer when Project Guide is published; if needed, teachers will explain in Lab)
    - There will be penalizations if Project guidelines are NOT FOLLOWED (these will include, for example, Group sizes, late deliveries, wrong or missing components in the project delivery, etc.)
    - Further details – as already mentioned – will be detailed in Project Guide
    - Up to the point of planning and designing your Data Warehouse, your Lab teacher will be happy to advise on Project work; after that stage, your teacher is happy to clear or answer technical doubts BUT WILL NOT be able to make decisions (or do actual work) about your project on your behalf – you must be able to demonstrate ETL skills, when creating and loading your Data Warehouse.
    - Delivery methods (how you share your work with teachers) will be set at later date

# BI I Project Overview (cont.)

- **On what Data Problem should I work? How to find data?**
  - We always recommend that your Group works a real-life problem, if at all possible!
  - We have been recommending this action for many years, and it always seems to bring the best results (typically, Groups working on real-life contexts are more motivated, have a deeper understanding of work to be done and get better grades!)
  - If this is not possible, we will be pointing to online (internet) resources where you can get data that is close to real-life problems (typically, not as good but still quite solid choice – and you still get to choose what business domain to work on)
  - As (last) alternative, we hope to have – around 2$^{nd}$ or 3$^{rd}$ lab – collections of artificial datasets that represent synthetic "organizations" having a data problem that you can attempt to fix – quite similar to what you can get by searching through online resources, but curated and slightly refined/customized by the teachers

# Questions and doubts

# Thank you!

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa