## *General instructions*

The project has to be conducted in teams of 3-5 persons (**maximal number of groups is 11**). The scope of the project is not fixed, and each group can decide to implement what they want. The project can cover any aspect of process oriented data science and we encourage the involvement of external participants (the company you work on, an association you are part of, etc.).

**IMPORTANT:** Before starting the work on the project, it is NECESSARY that the teachers approve the project, to ensure that the amount of work is proportionate to the group size. To foster group formation, use this spreadsheet to communicate your group (if you still do not belong to any group, at the end you can add your ideas for the project to gather interest from others).

It has to be an **Analysis of a real event log**, to extract some knowledge out of it:

- You can decide to analyze a log coming **from your own company/organization.**

- You can use a log publicly available. You can see here an example from Fluxicon blog; based on Statsbomb data set: https://fluxicon.com/blog/2019/10/process-mining-meets-football-how-does-a-football-team-possess-the-ball-on-the-pitch/

  *A set of event logs can be found at https://www.tf-pm.org/resources/logs but note not all of them have business questions associated; therefore, you might need to do some research to understand where the logs are coming from and what is relevant in the specific domains.*

- Possible options, where data and relevant business questions are provided, are coming from **previous editions of BPI Challenges**, including:
  - https://www.tf-pm.org/competitions-awards/bpi-challenge/2020
  - https://www.tf-pm.org/competitions-awards/bpi-challenge/2019
  - https://www.win.tue.nl/bpi/2018/challenge.html
  - https://www.win.tue.nl/bpi/2017/challenge.html
  - https://www.win.tue.nl/bpi/2016/challenge.html
  - https://www.win.tue.nl/bpi/2015/challenge.html
  - https://www.win.tue.nl/bpi/2013/challenge.html

For most of them, there is a .dsc project file that you can directly import into Disco:

- 2020 University reimbursement process:
https://files.fluxicon.com/Datasets/BPIC/BPI-Challenge-2020.dsc
- 2019 Purchasing process:
https://files.fluxicon.com/Datasets/BPIC/BPI-Challenge-2019.dsc
- 2018 Government funding application process:
https://files.fluxicon.com/Datasets/BPIC/BPI-Challenge-2018.dsc
- 2017 Loan Application Process (from 2012) revisited:
https://files.fluxicon.com/Datasets/BPIC/BPI-Challenge-2017.dsc
- 2016 Customer journey Dutch Employee Insurance Agency:
https://www.win.tue.nl/bpi/doku.php?id=2016:challenge
- 2015 Dutch municipality processes:
https://files.fluxicon.com/Datasets/BPIC/BPI-Challenge-2015.dsc
- 2013 ITSM Process Volvo IT:
https://files.fluxicon.com//Datasets/BPIC/BPI-Challenge-2013.dsc

*Please note that in these cases, you can also have access to the reports that people contributed as solutions. You are very much encouraged to read and study them (as we have been doing in Lab Sessions with BPI Challenge 2012) and therefore you need to explain (in the written report)* ***how your analysis is improving all these solutions****.*

**Take a look beforehand to the selected event log and scope of problem and take into account for instance that :**

- **Disco's Academic license is limited to handle 5 million rows (events)**
- **Apromore's academic cloud is limited to handle event logs of up to 500 000 rows (events) and up to 25 columns (event attributes)**

## *What to submit*

The deliverable must include a report where each group describes what they have done in a critical way (i.e., motivate all their decisions and document and explain all results); all the code produced (also in case of analysis I expect that some pre-processing needs to be done using some scripts); the slides of the presentation that will be given. Further instruction below.

The final submission comprises two deliverables:

- A *group submission* consisting of one .zip file containing (please, be precise with the naming of each element):
    - A report in a file called **report.pdf** (the report has to be in PDF):

        - Has to be **<u>maximum 15 pages long</u>**;
        - Has to provide a complete description of the context, the tools being used, and why the group chose them. In case of analysis of a real event log indicate:
            - Domain and data description;
            - Problem description;
            - Method and tools used to solve the problem;
            - Results;
            - Reflections on the project./Conclusions/Future lines of work (*Potential and results have to be commented and discussed in details and proper comparison with state of the art has to be discussed as well)*
        - Include the information **of the contribution of each member**, including who performed what as part of the project and **who wrote each section/subsection of the report.**
    - The final presentation in a file called **presentation.pdf** (the presentation has to be exported in PDF).
    - In case the group wrote some software (e.g., processing scripts in Python/R/... or entire software in Java/C#/...), the complete source code should be in a file called **source.zip**.
- An *individual submission* consisting of a file called **individual-report.pdf**, which has to be up to 1 pages long, where each student has to comment and reflect about:
    - What are the main “take-home messages” of the course, including the most important points in process mining, what they feel as most relevant and impactful in their future career, which elements of the course they feel to have familiarity with (about 1 page);
    - What are their individual opinions and comments about the project, which tools each student used and felt most comfortable with, which were the challenges, what they learnt from running the project, and how the group work was conducted.

## *The final presentation*

Each group will need to present their project in a 5-8 minutes (exact time depends on the number of groups and will be communicated later) presentation.

During the presentation, it is required that:

- Each member of the group presents something;
- The presentation has not to be a “slides” version of the report. The focus has to be on the elements that are better explained in person. Still, it has to cover all elements of the project, including;
    - Domain and data description;
    - Problem description;
    - Method and tools used to solve the problem;
    - Results;
    - Reflections on the project./Conclusions/Future lines of work
- Each member must be able to answer questions about any part of the project;
- The presentation might include a brief live demonstration of the system if you think this can contribute to the better understanding of the project.

## *Evaluation*

It will be taken into account the submission of the group and individual report. **Projects not submitted in racó (Practicals) or submitted after the deadline will not be accepted.**

It will be taken into account for the submitted report:

- Complexity of the problem
- Domain and data description
- Problem description
- Methdology and tools
- Reflections on the project./Conclusions/Future lines of work

It will be taken into account for the presentation:

- Structure and contents
- Quality of the slides
- Presentation skills
- Suitability for audience
- Ability to cope with questions about the context
- Ability to cope with technical questions