

# Business Intelligence Exam Project

This project is designed as data analytics experimental research and development of BI implementation solution. It involves systematic and creative work of finding novel, uncertain, and reproducible results by applying modern BI and artificial intelligence (AI) technologies in a context.

The development workflow goes through *four sprints and milestones*, each of which has an objective, tasks, and deliverables.

## Sprint 1: Problem Formulation

*Foundation of a business case and problem statement*

1. At this stage you brainstorm, browse sources of inspiration and information, collect ideas and discuss business or social domains, where BI and AI can bring a value.
2. Choose one of your ideas and define **context, purpose, research questions, and hypotheses** for a BI problem statement. Write a brief **annotation** of your project, in about four sentences, explaining:
  - which challenge you would like to address?
  - why is it important or interesting research goal?
  - what is the expected solution your project would provide?
  - what would be the positive impact of the solution, which category of users could benefit from it?
3. Prepare the development environment
  - give a title to your project
  - plan and organise the execution of the individual tasks in terms of time, milestones, deliverables, team members engagement
  - prepare the development platform and procedures – Github repository, IDE, software tools
4. Create a **.md file** containing the problem formulation and upload in the Github repository as first release of the project's documentation.

**Recommended deadline: 10/09/24**

## Sprint 2: Data Preparation

*Data collection and pre-processing, exploratory data analysis*

Based on the ideas and assumptions defined at the previous stage:

1. Collect and load **relevant data** from various sources.
2. Clean and integrate the collected data in appropriate data structures. Apply any transformations needed for the integration and the operations - ETL (Extract Transform Load) or ELT (Extract Load Transform).
3. Explore and analyse the data by applying measures from descriptive statistics to discover its basic features. Create charts and diagrams to visualize the features for better understanding and planning.
4. Apply the necessary data engineering to prepare the data for further analysis, ensuring that the data is:
  - a. meaningful – describes relevant and correctly measured features and observations
  - b. sufficient – describes various cases and feature occurrences, decided by testing
  - c. shaped – presented in a structure, appropriate for processing by machine learning algorithms
  - d. cleaned – repaired from missing values and outliers
  - e. scaled – transform data distributions in comparable scales, when necessary
  - f. engineered – analyse all features and select the most informative for further processing

Take use of the instruments created in MP1 and MP2. Export your **first data collection prototype** to the Github repository.

**Recommended deadline: 17/09/24**

## Sprint 3: Data Modelling

### *Applying AI methods and algorithms for building predictive models*

Extend the BI data analysis by implementing machine learning, deep learning, and/or generative methods and algorithms, exercised in class.

1. Select some relevant **methods and algorithms** that could solve the problem. Train predictive or generative data models by use of supervised and unsupervised methods, neural networks, LLMs, or graphs.
2. Select and apply appropriate inference measures for assessing the quality of your models and testing your hypotheses.
3. Test and validate the models applying them on both test data and new data.
4. Explore possibilities for improvement the quality of the models.

Take use of the instruments created in MP3. Publish the new **prototype of your solution** in Github.

**Recommended deadline: 26/09/24**

## Stage 4: Usability Evaluation

### *Present the process and the results of the analysis in human-understandable form*

Extend your solution with explanation and interpretation of the results:

1. Apply visual representation of the data, the analysis process, the applied methods and usage scenarios. Consider the use of **animation and 3D visualisation**, as appropriate.
2. Create a simple visual interface of the application to make it accessible and interactive for other users.
3. Present the visualised prototype to potential users for **usability evaluation**. Take notes and implement the relevant feedback outcomes. Elaborate on the benefits of applying visualisation and explanation techniques for data analytics.

Take use of the instruments created in MP4 and MP5. Revise, complete, and deliver the **final solution** to Github and a link to it in Wiseflow.

**Recommended deadline: 10/11/24**

## Notes

1. This is intended as a group project. The optimal group size is 2-3 students.
2. The exam is individual. Grading is based on both the group project and the individual performance at the exam.
3. The quality of the solution is assessed according various criteria, such as:
  - relevance
  - compliance with the requirements
  - problem statement and work hypotheses correctness
  - argumentation of choices
  - sufficient documenting of procedures and results
  - proper use of resources
  - applying data exploration, engineering, and integration techniques
  - proper implementation of statistics and machine learning methods
  - variety and usability of visualisations
  - explanation and interpretation of results
4. As a pre-requisite for the exam, you need to
  - collect minimum 80% of the study points collected from the semester assignments
  - submit to Wiseflow **one page** of text in pdf format, on which stay
    - the title of the project
    - the names of the team members developing it
    - resume of the problem statement and the solution
    - the link to the Github repository, where the project solution can be found
5. The project solution hosted on Github includes the problem statement, motivation, theoretical foundation, argumentation of choices, design, code, artefacts, outcomes, and implementation instructions, as appropriate.

It is important to enable reproduction of the project functionality and results.