Business Intelligence Exam Project

This project is designed as experimental research and software 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 with the rest of the team some 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 this challenge interesting or important research goal?
   - how could your project contribute to a solution of the challenge?
   - what would be the positive impact of the solution and which category of users can benefit from it?
3. Prepare the development environment
   - give a title to your project – the title is important, as it is the shortest summary of the project
   - plan and organise the execution of the project tasks in terms of time, milestones, and teamwork
   - prepare the development platform and procedures – Github repository, IDE, additional 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.

Sprint Completed: 23/09/25

Sprint 2: Data Preparation

Data collection, pre-processing, and exploration

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

1. Collect and load relevant data from various authentic sources. Avoid relying on cleaned sources only!
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.
4. As a part of the exploration, create charts and diagrams to visualize the features for better understanding of data and planning of operations.
5. Ensure the quality of the data in terms of:
   - relevance – the data consists of meaningful and correctly measured features and observations
   - sufficiency – the sample is large enough and includes various cases and feature occurrences
   - structure – the data is presented in one or more structures, appropriate for BI processing and analysis
   - completeness – the data is repaired from missing values, outliners, and unnecessary attributes
   - dimensionality – the data is transformed into proper scales and dimensions, where it is necessary

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

Sprint Completed: 25/09/25

Sprint 3: Data Modelling

Applying AI methods and algorithms for predictive analytics

Extend the exploratory data analysis with predictive analytics by implementing machine learning algorithms and generative models, exercised in class.

1. Select some relevant methods and algorithms that could provide probable solution of the problem. Train predictive data models by use of supervised and unsupervised machine learning.
2. Test and validate the models applying them on both test data and new data.
3. Select and apply appropriate inference measures for assessing the quality of your models and testing your hypotheses.
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.

Sprint Completed: 02/10/25

Sprint 4: Business Application

Presenting the process and the results of the analysis in human-understandable form

Enable deployment and operability of your solution:

1. Create web application with simple visual interface to make the solution accessible, understandable, and interactive for non-technical users. Consider use of and LLMs and natural language dialogues.
2. Apply visual representation of the source data, the analysis process, the implemented methods and usage scenarios. Consider the use of dashboards and data stories with text, 2D and 3D diagrams, images, geo-maps, animation and video, as appropriate.
3. Provide explanation and interpretation of the results. Elaborate on the benefits of applying visualisation and explanation techniques for data analytics.
4. Present the visualised prototype to others for usability evaluation. Take notes and implement the relevant feedback outcomes.

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

Project Completed: 06/10/25

Notes

1. The project is a teamwork. The optimal size of a team is 3-4 students.
2. The exam is individual. Grading is based equally on both the project quality and the individual performance at the exam.
3. The quality of the solution is assessed according multiple criteria, such as:
   - BI relevance
   - compliance with the requirements
   - problem statement and work hypotheses correctness
   - data collection and proper use of resources
   - variation of data exploration, engineering, and integration techniques
   - argumentation of choices
   - proper implementation of statistics and machine learning methods
   - sufficient documentation of procedures, operations, and results
   - variety and usability of visualisation, explanation, and interpretation of results

4. As a pre-requisite for the exam, you need to
   a. collect minimum 80% of the study points achieved by the semester assignments
   b. 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 is hosted on GitHub and includes the problem statement, motivation, theoretical foundation, argumentation of choices, design, code, artefacts, outcomes, and implementation instructions.
6. It is important to enable reproduction of the project functionality and results.

T. Dimitrova

instructor