Assessment:	Project Milestone 1
Subject:	Business Intelligence 381
Total:	50

Project Milestone 1

Introduction

This project will investigate various health and demographic datasets to identify meaningful patterns and trends. The details of the project requirements are described in the **Project Outline** document. The project outline underscores the use of CRISP-DM as a structured approach to guide you through the data science project, emphasizing its importance in the context of real-world data analysis and modelling.

Outline

The project is divided into six (6) milestones, where at each stage some deliverables are to be produced in terms of a series of reports that describe the project plan, and the work carried out during the iterative process of data preparation, modelling, evaluation and eventual deployment.

While CRISP-DM breaks down a data mining project life cycle into six phases with each phase consisting of many secondary tasks, the focus of this milestone is on the first two phases. These phases help organisations understand the objectives of the project from a business perspective and identify, collect, and analyse the data sets that can help accomplish the project goals. You may need to gather more information on related projects and scenarios to establish a strong business case for the project.

Understand the Business

Thoroughly understand, from a business (government or non-profit organisations) perspective, what the project really wants to accomplish. Uncover important factors, that can influence the outcome of the project, record the information that is known about the business situation during this phase of the project.

Understand the Data

Start off with the initial data collection and proceed with activities that enable you to become familiar with the data, identifying data quality problems, discovering first insights into the data, and/or detect interesting subsets to form hypotheses regarding hidden information.

Tasks

The entire project consists of the following steps which are then broken into milestones that should be submitted on given due dates.

1. **Business Understanding:**
   - Clearly define the business problem and objectives.
   - Identify stakeholders and their requirements.
   - Determine the success criteria for the recommendation system.
   - Recognize the importance of CRISP-DM as the methodology to guide the project.
2. **Data Understanding:**
   - Explore the provided datasets, including variables like mortality rates, distance to health facilities, facility types, immunisation rates, water sources, literacy, vaccination rates, etc.

- Document data quality problems, missing values, duplicates, and outliers.
- Perform preliminary data visualizations and analysis to gain initial insights.
- Create preliminary dashboards to visualize data distributions, correlations, and initial insights to identify patterns and relationships in the data before moving on to the modelling phase.

Deliverables:

- Project Scoping and Data Exploration Report (in PDF or document format).
- Code (R, R Markdown) used for data analysis and modelling.
- Power BI project file if Power BI was used during this phase of the project.

Grading Criteria:

Criteria	Weight
Background	5
Business Objectives and Success Criteria	5
Inventory of Resources	5
Risks, Assumptions, and Constraints	5
Data Mining Goals and Success Criteria	5
Data Description	5
Data Exploration and Visualization	10
Data Quality Assessment	10
TOTAL	**50**

Additional Information

- This is a group assignment, but the group may not exceed four people.
- All work must be original. Copying another group's work will not be tolerated.
- Includes names of all group members on the Cover Page.
- Submit your project electronically on Moodle (BC Connect) before the due date.
- All writing must be correctly cited and referenced.
- **Plagiarism is a serious offence**. Belgium Campus uses software that can scan for plagiarism and a student caught doing this will get 0 for this assignment.
- No mark will be awarded if the assignment is not uploaded via BC Connect.
- Late assignments will not be accepted; missing the deadline is an automatic 0.