BSDS3001 Social Data Science Foundations

Assessment 3 – Group Project (*25% of total course marks*)

A.   Background

This assessment aims to develop students' interdisciplinary thinking for addressing social and global challenges with computing skills and to enhance their ability in planning research work and proposal writing. Groups can consist of 3 or 4 students. Each group is required to develop a research proposal that investigates an important social or global problem/phenomenon using social data science theories and methods.

It is essential to ensure that the theme of the project aligns with at least one of the United Nations' Sustainable Development Goals (SDGs). The research proposal should demonstrate the group's knowledge of concepts and methods of social inquiry. As social data science is an interdisciplinary field, the proposal should clearly illustrate how the proposed research integrates different disciplines throughout the document.

B.   Deliverables

(I)   Research Proposal (*15% of total course marks*)

(II)   Short Project Presentation (*5% of total course marks*)

(III) Peer Evaluation (*5% of total course marks*)

(I)   Research Proposal

The proposal should contain the following parts:

Abstract

The abstract is a summary of the research proposal, typically containing the problem statement, the rationale for the study, the research questions, the research design, the expected results, and the significance of the proposed study. This section provides readers with an overview of the proposed research. Please limit the length of the abstract to 400 words.

Introduction

This section sets the context for your proposed research and should attract the readers' interest. The group should explain the background of the study, starting from a broad picture and narrowing down to the research questions. To help readers obtain a general understanding of the proposed research, the Introduction should include:

●   The background and context of the study, including which disciplines are involved and which concepts in each discipline are relevant.

●   The problem statement and research questions.

●   The connection with one of the SDGs and its justification(s).

●   The significance of the study (e.g., theoretical and practical).

B. Deliverables (cont'd)

(I)   Research Proposal (cont'd)

The proposal should contain the following parts: (cont'd)

Literature Review and Theoretical Ground

This section should provide a critical review of related existing work and relate the literature to your proposed research. Are the sources from different disciplines, and how are they related? It should articulate the research gaps in the existing literature. Usually, the research questions are proposed to fill or bridge these research gaps.

Your group is expected to apply social theories to guide your exploration of the raised problem/questions. As social sciences are broad, you may use any theories related to the problem you are investigating, drawing from your current or previous courses, readings, and the literature. Note that you may combine related theories from different disciplines.

Methodology

This section should provide the details of the data acquisition/collection and explain how the dataset(s) is/are able to address the research question(s) listed in the introduction. Key attributes/features should be introduced. This section also proposes the social data science methods that may be applied to address the research question(s). Your group is expected to propose techniques and/or methods from data science and statistics, which may reflect the interdisciplinarity of the research.

Current and Expected Results

Each group is expected to present some results of data analysis. You are expected to develop data visualizations using Tableau Desktop. Depending on the problem and data availability, the amount of results will vary across groups. Given the time constraints, it may not be possible to obtain all results. Please describe the expected results and possible implications of the expected results for related disciplines.

The length of the research proposal should be **2,500 words**, depending on the group size. The proposal should be submitted through Moodle on or before **8 December 2025**.

B.   Deliverables (cont'd)

(II)   Short Project Presentation

Each group will give a short presentation of 10 minutes to introduce their research proposal. This is an opportunity for each group to receive feedback from the instructor and other students. Each group decides how many presenters will be involved. The presentation needs to cover the following:

- The project background, the problem statement, the connection between the project and SDG(s), and the project's significance;
- A literature review and methodology (data acquisition and key attributes);
- Expectations of results and reflection (No data visualizations are required during the presentation).

The presentations should be correct, clear, with proper visual aids, and with good timing. The presentation will be conducted on **28 November or 1 December 2025**.

(II)   Peer Evaluation

Each group member is invited to complete the confidential peer assessment forms at the end of the project for assessing the performance of team members (Due Date: **8 December 2025**). The form will be available on Moodle. It asks questions on quality of work, timeliness of delivery, and reliability.

C.   Submission

(I)   Research Proposal

- Times New Roman 12pt font
- 1" margins and 1.5 line spacing
- APA 7th Edition
- Save it as BSDS3001_GP_ <<Group ID>>.**docx**
- Submit to The Moodle (Turnitin)
- Due date is 23:59, **8 December 2025**. No late submission will be accepted.

(II)   Short Project Presentation

- APA 7th Edition
- Save it as BSDS3001_GP_ << Group ID>>.**pptx**
- Submit to The Moodle (Turnitin)
- Due date is 23:59, **27 November 2025**. No late submission will be accepted.

(III)   Peer Evaluation

- Peer Evaluation Form will be available in due course.
- Due date is 23:59, **8 December 2025**.

D.   Evaluation

(I)   Research Proposal

The proposal will be evaluated based on the grading criteria of the Assessment Feedback Proforma (Appendix 2).

(II)   Short Project Presentation

The presentation will be evaluated in terms of:

(a) complete coverage of required content;

(b) fluency of presentation including eye contact with the audience;

(c) clear and proper visual aid; and

(d) good timing (within 10 minutes).

(III)   Peer Evaluation

Each student's score will be averaged across the scores given by his/her groupmates. Those who submit no or incomplete evaluations will have their own scores deducted.

B. Project Ideas

Possible ideas, but not limited to the following, include:

- City Management
- Climate
- Economy
- Education
- Employment and Labor
- Environment
- Food
- Health
- Housing
- Social Welfare
- Transport

C. Open-source Datasets

Site	URL
Hong Kong – Open Data	https://data.gov.hk/
United Kingdom – Open Data	https://data.gov.uk/
Australia – Open Data	https://data.gov.au/
United States – Open Data	https://www.data.gov/
EU Open Data Portal	https://data.europa.eu/en
South Korea Open Government Data portal	https://www.open.go.kr/com/main/mainView.do
GSS General Social Survey | NORC	https://gss.norc.org/
The International Social Survey Programme	http://www.issp.org/
Google's Dataset Search	https://datasetsearch.research.google.com/
Registry of Research Data Repositories	https://www.re3data.org/
Kaggle *(only newer ones)*	https://www.kaggle.com/datasets
Harvard Dataverse	https://dataverse.harvard.edu/
FiveThirtyEight (Journalism)	https://data.fivethirtyeight.com/
UNICEF Data	https://data.unicef.org/resources/resource-type/datasets/
World Bank Open Data	https://data.worldbank.org/
OECD Data	https://data.oecd.org/
United Nations Data	http://data.un.org/
GHO data repository - World Health Organization	https://www.who.int/data/gho
UK Data Service	https://ukdataservice.ac.uk/
National Archive of Criminal Justice Data	https://www.icpsr.umich.edu/icpsrweb/content/NACJD/index.html
HealthData.gov	https://healthdata.gov/browse?limitTo=datasets
Quandl	https://data.nasdaq.com/search

Remarks:

(a) It is your responsibility to ensure the credibility of the data source. Please consult your teacher if you have any concern.

(b) The data set is suggested to have 15+ attributes and 300+ instances.

(c) Avoid collecting data by questionnaires.