

## Group Project Instructions

### Objective:

For this group assignment, students will need to select and analyze a dataset using a data mining method (classification or clustering), and then show how the results of such analysis can produce “actionable” insights, i.e., useful results that an organization can use to improve their operation, products, or services.

### Main Steps:

1. Find a real-world problem and a dataset that can help to answer this problem. You can browse the following open data repositories to find a dataset of interest:
  - Popular open data repositories:
    - Kaggle datasets: <https://www.kaggle.com/datasets>
    - UC Irvine Machine Learning Repository: <http://archive.ics.uci.edu/ml/index.php>
    - Amazon’s AWS datasets: <https://registry.opendata.aws/>
  - Meta portals (websites that list open data repositories):
    - Google Dataset Search: <https://datasetsearch.research.google.com/>
    - Data Portals: <http://dataportals.org/>
    - OpenDataMonitor: <http://dataportals.org/>
    - Quandl: <https://www.quandl.com/>
  - Other pages listing popular open datasets:
    - Wikipedia’s list of Machine Learning datasets: [https://en.wikipedia.org/wiki/List\\_of\\_datasets\\_for\\_machine-learning\\_research](https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research)
    - Quora.com: <https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>
    - The datasets subreddit: <https://www.reddit.com/r/datasets/>
2. Retrieve data and perform necessary data cleaning and preparation in Excel or RapidMiner.
3. Use RapidMiner to mine data for useful insights using either a classification or clustering approach.
4. Prepare a final report that will describe the problem, data preparation stage, your approach and results. As part of the report, discuss how your results can help the organization in question to improve their operation, products, or services. The written project report (~10 pages) should be submitted electronically via D2L by one of the group members. The report template is provided along with this file (“*Group Project Report Outline.pdf*”).

Report\_<last\_name1>\_<last\_name2>\_<last\_name3>.pdf

5. Present the project results in class. The presentation should be based on your final report. Each group will have 5 minutes to present the report. All group members must participate in the presentation. The presentation should be accompanied by up to six PowerPoint slides saved as pdf and submitted electronically through Blackboard before the deadline. The first slide

should contain the project title and the names of the group members. Use the following name format for the pdf file:

Presentation\_<last\_name1>\_<last\_name2>\_\_<last\_name3>.pdf

6. After submitting the report and presentation before the due date, complete and submit the peer evaluation form (5% of the overall mark). The form is posted on Blackboard along with this file. Each student must submit the form individually.

**Assessment:**

All group members will receive the same grade. However, the instructor reserves the right to amend this grading policy under special circumstances. If you are experiencing any significant difficulties with your group members, let the instructor know as soon as possible.

The late submission penalty is 4% per hour.

Here is the breakdown of the grade for this assignment:

- Report: 70%
- Presentation: 25%
- Peer evaluation: 5%

Plagiarism in any form will be treated as an academic misconduct. For more information about academic integrity, please refer to “Policy on Academic Honesty” described in the course outline.