

Project Description

1 Group Project

Students are encouraged to form groups of up to 4 members. We will send out a project sign-up sheet soon. The project requires each group to identify an industrial problem of interest and analyze data using one or more techniques learned in this course. The main deliverables include

- A one-page final report describing your problem and solution;
- A set of slides for in-class presentation;
- Sufficient appendices and supporting files (datasets, Python code) to fully document your work.

Groups will present their project during the class time on Friday, 10 May. There are two main criteria to consider when looking for a good project:

- **Value Opportunity:** Where is the potential to use analytics to capture value? Specifically, what benefits can be achieved in terms of added value and/or reduced costs using analytics? What data is available that can be leveraged to achieve these benefits? The ideal project setting would be one in which there is a large value opportunity and novel data that can be leveraged to capture this value.
- **Solution Quality:** Does the proposed data, model, and methodology do a good job of capturing the value identified above? That is, are the data available appropriate and sufficient? Is the methodology used appropriate? Are the models and methods applied correctly? Are the models well-validated? Is the model performance good enough to deliver the anticipated benefits?

It would be best to base your project idea on current or previous experience. Alternatively, you may want to think in terms of your current experience as a student at HKUST and what analytics you and your fellow classmates would find useful. Lastly, your project could be based on an idea you have for a new business venture based on analytics. In all cases, we are not looking for a fully functioning system. Instead, think of your project as a proof-of-concept prototype. Specifically, we are looking for a good problem idea, and then sample data and analyses that are sufficient to validate the potential of your idea. View your project and presentation as something you might use to demonstrate your idea to a potential investor, consulting client, or senior executive.

2 Data Sets

Students are encouraged to collect data by themselves. Students can collect data from any public sources such as, but not limited to:

1. Kaggle datasets: <https://www.kaggle.com/datasets>
2. UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/datasets.html>

3. New York City open data: <https://nycopendata.socrata.com/>
4. U.S. open government data: <http://www.data.gov/>
5. World Bank: <http://data.worldbank.org/>
6. US Census Bureau: <http://www.census.gov/main/www/access.html>
7. Google trends: <http://www.google.com/trends>
8. Google finance: <https://www.google.com/finance>
9. Topsy social media data: <http://topsy.com/>
10. Million song dataset: <http://aws.amazon.com/datasets/6468931156960467>
11. Data from your work experience
12. Survey data collected by your group (<http://www.qualtrics.com/> is a good free tool for conducting surveys)

3 Analysis and Report

Students should work closely as a group in defining their project, collecting data, analysis, and developing the project presentation. In explaining your project, be concise and clear. Keep in mind that plots, tables, and other visual representations of the problem, data, and methodology are effective in conveying your ideas and findings. The project written report should contain:

1. **One-Page Main Body:** A summary overview of your problem, methodology and results. This is effectively an executive summary of your work, sufficient to explain and “sell” your idea and analysis.
2. **Appendices:** describing in more detail the data, models, analyses and conclusions.
3. **Supporting Files:** Additional supporting files containing sample data and any Python code required to fully define and explain your approach.

The combination of the main body and appendices should be a self-contained document; that is, someone should be able to read through the document and understand all components of your problem, data, and analysis. You may want to have a title for your project report. All the group members’ names should be included.

4 In-Class Presentation

In addition to the written report and submitted code, the students should present slides to succinctly present their findings. Each group will have approximately 15 minutes for their presentation, including time for Q&A. Please note, the exact time may adjust slightly depending on the total number of groups presenting.

5 Project Submission

Each group must submit the slide deck, final report, the appendices and any supporting files on Canvas before 23:59:59 on 9 May.

6 Project Evaluation

All members of a group will receive the same score for the project, irrespective of individual contributions. This policy is designed to encourage effective teamwork and equitable distribution of tasks. The project will be evaluated through a combined approach:

- 50% by Instructor and TAs: This portion of the evaluation will be conducted by myself and the teaching assistants, focusing on the project's technical accuracy, creativity, and adherence to the guidelines provided.
- 50% by Class Peer Review: For this segment, each of you will be responsible for submitting scores for the projects presented by other groups. We will calculate the median of these scores to determine the peer review component of each project's evaluation.

Please prepare to participate actively in this peer review process, as your input is crucial for a fair and balanced evaluation of all projects. Further details on the submission and scoring process will be provided as we approach the project presentation dates.