

## CIS 9660 Term Project (Total Points 22)

### Deliverables:

| Component  | Points | Due Date   |
|--|--------|--|
| Project Proposal   | 1      | Monday 10/17 11:59 pm via blackboard                 |
| Progress Report  | 1      | Monday 11/14 11:59 pm via blackboard                 |
| Final report deliverables (R code, data, and one clean report) | 12     | Tuesday 12/06 11:59 pm via blackboard                |
| Oral presentation  | 8      | Thursday 12/08 6:05 pm or 12/01 (possible) via VOCAT |

Each group only needs to submit one set of proposal, report, final deliverables (R code, data, and one clean report). Late submission is not accepted for any component of project. Unauthorized collaborative work will result in appropriate disciplinary action.

### Description:

The goal of the project is to develop the appropriate data mining methods for data analysis using R. During this project, you can demonstrate how well you have acquired the course material. In addition, you can gain practical experiences to be prepared for the future career.

1. Project Proposal (1 pts): **Monday 10/17 11:59 pm via blackboard**

You are encouraged to find a data set that you are interested in. You can use your own data if you already have an ongoing project. Or you can search one from online data repository, e.g. UCI machine learning repository<sup>1</sup>, Kaggle<sup>2</sup>, or NYC OpenData<sup>3</sup>.

Please create a team name. Identify interesting problem(s) that can be addressed by that data set. It can be an unsupervised, a regression, or a classification problem. Turn in a project proposal including the following items:

- What is your team name?
- Which data set you intend to analyze? Where is your data source? Provide the data description, including the number of observations, the number of variables, variables description and data type.
- What is (are) the question(s) you intend to address?
- Which variable is the target variable?
- Which methods you plan to use in this analysis? (Feel free to select any appropriate data mining methods in your analysis.)
- Who is your target audience?

Do not include R code and analysis results in the proposal. The proposal should be no more than 1-page long.

---

<sup>1</sup> <https://archive.ics.uci.edu/ml/index.html>

<sup>2</sup> <https://www.kaggle.com/datasets>

<sup>3</sup> <https://opendata.cityofnewyork.us/data/>

2. Progress Report (1 pts): **Monday 11/14 11:59 pm via blackboard**

Describe your progress on the project so far. You do not need to attach R code.

- What have you completed so far? If you are addressing a different problem from what you originally proposed, introduce your new data and the target variable.
- Describe your exploratory data analysis results.
- What is the plan for remaining tasks?

3. Final report deliverables (12 pts): **Tuesday 12/06 11:59 pm via blackboard. Turn in data, R code, and one report via blackboard**

- One final report (excluding figures, tables and reference) should be about 3-5 pages with 11-point font. Append all your figures and tables at the end of the report with the appropriate captions. Use them in your report where necessary to support your arguments, e.g. “Based on Figure 1,...”.
- Do not show R code in the report. The report only conveys your work and results.
- The report can be typed using any word processor you prefer, and it should not contain R raw outputs except figures.
- Cite any references used in your project in the reference section.
- Turn in R code to Blackboard as a separate file. The report will not be graded without code.
- Submit the used data set to Blackboard. All results in the report should be able to be reproduced using R code and data.

4. Oral presentation (8 pts): **Thursday 12/08 6:05 pm via VOCAT. Class time on 12/01 may be used for presentations, depending on the final total number of groups.**

- We use VOCAT to upload and view final presentation. The website for VOCAT is: [https://baruch.vocat.io/users/sign\\_in](https://baruch.vocat.io/users/sign_in). You can use your Baruch netID (the ID & password for your Baruch email login) to access VOCAT.
- Upload your video presentation to VOCAT before the class. The presentation will be evaluated by all class members.
- The presentation is expected to be around 15 minutes, depending on the final number of teams.
- The final presentation must be saved as video. A presentation link won't work on VOCAT.
- You can record the presentation via camera, or software such as Zoom or just voice-over narration on a PowerPoint presentation (i.e. video without person is fine). Microsoft PowerPoint 2016 and later versions provide this function **Slide Show-→ Record Slide Show**. If you use voice-over narration on a PowerPoint, please make sure that you create a video using **File->Export->Create a Video->Create Video**. The video size is suggested to be below 2GB.

### Final Report Content

A professional report should contain multiple sections from introduction to conclusion. The content of the final report should include the following items.

- Introduce the problem(s) addressed in this analysis and the data set used.
- Perform exploratory data analysis to explore the entire data set. Preprocess the data set when necessary. Reduce variable(s), if necessary, and justify your choices.

- Select appropriate data mining methods to develop models. Implement **multiple** models until you reach a “final or optimal” model. Compare model performances. Describe the model development and selection process.
- Adopt appropriate methods and measures for model evaluation. Use figures or tables to show model performance.
- State your final model clearly and describe your findings. Use figures, tables or formatted outputs to support your arguments.
- Make conclusions following logically from results and findings. Discuss its implications to the target audience.