Dr. Eick

COSC 3337 *“Data Science I”* Fall 2021

7-week Long Data Science Group Project

Fourth Draft

Last updated: November 8,8a

Deadlines: Groups who present on Tuesday, Nov. 16, have to submit their deliverables end of the day *Wednesday, Nov. 17*; groups who present on Thursday, Nov. 18, have to have to submit their deliverables end of the day *Tuesday, November 16!*

There will be a 7-week long group project (typically 3-4 students per group) September 27 to November 18 in Fall 2021 in which groups:

1. identify a dataset; maybe multiple datasets
2. determine what question(s) should be answered using this dataset
3. employ Data Science/ML/Statistics/DM/Data Visualization techniques to answer the identified question(s)
4. summarize the project findings in a report
5. give short presentation about the project topic, methods used and project findings.

We will approve almost any project theme as long the presented project goals make sense; however, projects which center on exploratory data analysis will be not approved due to their overlap with Problem Set1 Tasks.

Schedule and Deliverables[[1]](#footnote-1):

September 27-October 9: Groups meet and solve tasks a and b, and submit a 1-2 page project description to Dr. Eick and the TAs[[2]](#footnote-2). (10% group project credit will be allocated to coming up with a project description)

October 10-…: Project description will be reviewed, feedback will be provided and the project will either a) be approved b) conditionally approved with suggests on how to improve the project description c) disapproved—we hope this will not happen often!

October 29: 1.5-2 page Project Status report is due (10% group project credit will be allocated to this report)

November 16/17—see above: Project Final Report and Code submission is due; report should have 6-8 pages; organization: 1. Introduction and Problem Statement 2. Methods 3. Results and Discussion 4. Conclusion; reports should be single space and use an 11pt or 12pt font. The code (or codes) should be submitted with the Final Report.

November 16+18: Group Project Student Short Presentations (about 6-7 minutes per group); we will divide the class into 2 disjoint groups; each group member should be present during his/her group’s presentation. We will broadcast the presentations via MS Teams.

Some links concerning project ideas and potential datasets:

[16 Data Science Projects with Source Code to Strengthen your Resume - DataFlair (data-flair.training)](https://data-flair.training/blogs/data-science-project-ideas/)

[6 Interesting Data Science Project Ideas & Examples | Springboard Blog](https://www.springboard.com/blog/data-science/data-science-projects/)

[21 Data Science Project Ideas: Good, Interesting, And Cool Ideas (intellspot.com)](https://www.intellspot.com/data-science-project-ideas/)

[12 Cool Data Science Projects Ideas for Beginners and Experts | by Claire D. Costa | Towards Data Science](https://towardsdatascience.com/12-cool-data-science-projects-ideas-for-beginners-and-experts-fc75b5498e03)

[14 Data Science projects to improve your skills - KDnuggets](https://www.kdnuggets.com/2020/12/14-data-science-projects-improve-skills.html)

**Academic Honesty:** You will find a lot of Data Science project descriptions on the internet; reusing their content in the Data Science group project without proper citing them represents a serious academic honesty violation[[3]](#footnote-3). However, it is perfectly fine if your project builds on top of work others have done and software others have written, as long as you disclose and properly cite these sources. Moreover, we also ask you to disclose all software libraries and packages you used in your project.

1. All deliverables are due end of the day; e.g. 11:59p [↑](#footnote-ref-1)
2. Moreover, if you have doubts about your group project topic, feel free to visit Mathew, Sadat and Dr. Eick during their respective office hours to discuss your project ideas. [↑](#footnote-ref-2)
3. And we will conduct a websearch to find projects similar to yours on the internet. [↑](#footnote-ref-3)