

Project

Logistics

Please write your report **using the paper formatting instructions** at the end of this document. Upload your report to canvas. Please name your report “*Project_Lastnames-of-group-members.pdf*”, and please clearly state the full names of the group members at the beginning of the report.

The reports are due on **August 12th, before 10am**.

In addition, each group must **present their results in a 10-minute talk..** We will set aside 2 minutes for questions after every talk.

Project setup. You may investigate any dataset and try to solve problems of your choice. The goal is to look at an interesting problem and provide an answer backed by a dependable analysis.

To find datasets, you may consider looking at NYC Open Data or Kaggle, both of which provide several instances of publicly downloadable datasets. If you had some particular question in mind, try searching Google as well; there are many different sources of data, and you might be able to find something that fits your purposes.

Taboo datasets. Please don’t choose datasets on *Austin Animal Shelter* or *Airline Delays*. I’ve heard too many presentations about these recently!

Deliverables. There are two deliverables: the report and the presentation.

Report. Write a report with the following:

- **Description of project goals:** This should include two parts:

1. *Description:* What dataset are you investigating? What question are you trying to answer in this project? For example, “*the dataset of restaurant inspections in New York, where we look for patterns of misbehavior among different types of restaurants.*”
 2. *Importance of the problem:* What is special about this problem, or this dataset? You could briefly mention why you picked this particular problem, but it is also useful to point out why your work would be useful or interesting to a broader audience. Does it have some economic impact? Is there a large community of people who would be happy that you solved this problem?
- **Exploratory analysis:** Present some basic statistics or intuitions about your dataset. This is your opportunity to show any patterns or abnormalities you found in the dataset, even if they were not directly related to the main goals of the project.
 - **Solution and insights** Discuss your solution to the problem, and any insights gained from this process. For example, if your problem was a classification problem, describe (a) the features you decided to use and why, (b) which classifier(s), (c) a summary of the results obtained, and (d) an analysis of the model learnt by the classifier (e.g., which features were most important? which ones were useless? did you find anything surprising?)

This report must be uploaded to Canvas.

Presentation. Each group will have to present their work in class. Budget for 10 minutes. Ideally, the presentation should be structured like the report, but that is just a guideline. Make it interesting.

Paper Format Requirements

- Please create a Word document, write the names of group members on the upper right of the page, and **save as a PDF file**.
- Please use a **maximum of four pages**. The four pages are inclusive of everything except:
 - references to articles/Internet-URLs, and
 - plots, images, or data tables.

You can reference these plots from the main body of the report. This page limit is designed to improve the clarity of your analyses and recommendations, and your ability to get them across to an executive audience succinctly. **Anything beyond four pages will not be read or graded!**

- Please use **single-line spacing, 1 inch margins on all sides, Verdana font type, and 12-point font size.**
- Please use **proper subtitles for each section** of the report.
- If you are over the four-page limit, please do not use character condensation or other word processing tricks to fit your write-up into four pages. Please focus on writing that is more concise.
- Please adhere to grammar, spelling, and punctuation rules.