

Business Analytics Group Project

1. Teams

The project is a group assignment. Team size is limited to 5 people to enable full participation for all. If you cannot fill a team of 5 people, you may register a team of 4 and fifth team member may be assigned to your group. If you do not have a team or have a partial team, email your TA before Oct 28. The TAs will complete and finalize the teams.

2. Important Milestones

1. Oct 28

Form a team of 5 people and sign up at a link that will be provided by your TA.

2. November 9

Submit a one-page description of your project on Canvas. You should describe the problem(s) you are tackling, why it is interesting, what data you will use, and a general plan of attack. The person in the group whose last name is first in alphabetic order will submit this under their name.

3. Dec 7

Submit your presentation slides, 4-6 page report (with appendices if needed), R code, and other supplementary materials via Canvas. The person in the group whose last name is first in alphabetic order will submit this under their name.

4. Dec 7-14

Deliver your presentation. Each team should be ready to present on either of the two dates. We will choose teams in random order, in real time. You may have 1-5 people present depending on what your group wants to do. Questions may be asked during the presentation, so everyone should understand all the components of the project.

3. Project Details

The project is for each group to identify a business problem and answer it by analyzing data using techniques from class. Using one or more of the regression, classification, simulation, and optimization tools we learned in class, you should generate valuable insights from the data. Your project should make an interesting case to a potential stakeholder (for example,

the marketing director at your company, a reporter to the NYTimes). The project will be evaluated according to whether you showed proficiency of the subject, and you correctly used the analytical tools to derive interesting insights.

We encourage you to be creative in finding interesting and relevant questions, finding the right data to answer your question, and being creative in implementing the right analytics necessary. You can work on questions that came out from your previous experience, as long that they are not submitted for credit in another class.

In addition to the presentation itself, the main deliverables will be a set of presentation slides. The amount of time for each presentation will be announced. The content is more important than formatting): a 4-6 page report with appendices as needed, and supporting files (datasets, R code) to fully document your work.

Students should work closely as a group in defining their project, collecting data, doing the analytics, developing the project presentation, and writing the report. In explaining your project, be concise and clear. Keep in mind that plots, tables, and other visual representations can be effective in conveying your ideas and findings.

Three main criteria to keep in mind when looking for a good project:

1. **Value opportunity**—Where is the potential to use analytics to capture value? Specifically, what benefit can be achieved in terms of added value and/or cost reductions 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.
Note that this value may not be monetary, for example you may want to improve the workings of a charitable organization.
2. **Analytical Rigor**— 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?
3. **Clarity and Interpretability**— Can you explain the landscape you are working in and how your problem fits into this landscape? Is the problem clearly defined? Are your final outputs easily interpretable? Is it easy to understand what you did to the data and how it was done? Can someone reproduce your results based on your presentation and report? Can someone use your results to capture value?

We encourage you to find a topic that will be personally relevant to you. You can base your project idea on current or previous work experience. You may want to think of various topics you encountered as a student at Columbia (the class blog has many interesting references). 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. Rather, 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.

4. Data Sets References

Groups are encouraged to collect data by themselves. Some references of notable sources of data are listed below. These are just examples, and groups are encouraged to search beyond this list, as there many sources.

1. UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/datasets.html>
2. New York City open data: <https://nycopendata.socrata.com/>
3. U.S. open government data: <http://www.data.gov/>
4. World Bank: <http://data.worldbank.org/>
5. US Census Bureau: <http://www.census.gov/main/www/access.html>
6. Google trends: <http://www.google.com/trends>
7. Google finance: <https://www.google.com/finance>
8. Million song dataset: <http://aws.amazon.com/datasets/6468931156960467>
9. Data from academic papers, for example see data links in Esther Duflo's website: <http://economics.mit.edu/faculty/eduflo/papers>