# TEAM PROJECT DESCRIPTION

The team project is intended to allow you to put into practice some of the techniques from class on real world data and to explain those results to a managerial audience. It is composed of **two deliverables**: a proposal and the final report.

To start your project, your team will be asked to obtain a dataset and propose a collection of analysis and data management techniques to apply to that data. An ideal dataset will have sufficient observations and variables to support a wide range of analysis. A list of types of analysis you might consider are presented below. Note that the predictive techniques in this course are limited to predicting continuous variables.

You may need to select more than one dataset but ultimately your data should be sufficient to support five distinct business-relevant analyses/activities. It is my hope that the datasets will come from your own organizations that you can share within your team for analysis. I encourage you to attempt to find a dataset that is relevant to your organization as you may find useful insights by analyzing your own organization's data. If this is not possible, there are websites with publicly available data. Please be cautious about using Kaggle data, and ensure the data has sufficient information for a modelling project.

**Proposal (5%)**The proposal is a short document that presents the techniques you propose to use and the data you plan to apply it to. Its principal purpose is to facilitate a discussion between you and I. The document should contain about half a single-spaced page for each technique you plan to use (i.e., about 2.5 pages) and an executive summary of what you are trying to accomplish in the project. For each technique, you should address the following:

* Specifically what you propose to do (e.g. if you propose to do a test, you must describe it.)
* Why this technique is appropriate for the data.
* What problems you might have/limitations you may encounter, and what you plan to do about them.

To be successful in preparing your proposal you may need to do some research on the techniques before you develop the proposal. The textbook will provide some guidance on each of these techniques, moreover, Professor Scott will be available in office hours to discuss the project. During the 6th session of the course we will do team breakouts to discuss your proposals.

As a final note, be aware that using user-supplied data is often significantly more challenging than using data from published sources, which has typically been cleaned/tested/assessed to some degree by the providers. To be fair to those who use this 'real' data, this fact will be recognized in the evaluation of the projects.

The proposal will be worth 5% of your final course grade. It will primarily be assessed on the basis of its embedding a degree of sophisticated thinking and preparation for the project.

**Final report (35%)**   
The final report will be due before the next set of classes start. The report should contain an executive summary designed for an executive-level audience along with a technical summary of your analysis. The final report will be worth 35% of your final grade.

In the final report you will present both a technical and managerial description of each of the five techniques you have completed along with their individual findings as appropriate. You may choose to organize your report in such a manner to provide distinct technical and managerial descriptions if that is easier since normally a report has only one type of audience.

The format of the report is flexible and can be submitted as any of the following. This is not an inclusive list, and I encourage you to use the format best suited to telling your story and solving your specific business problem. To the extent that you are using real data from your organization, you can submit whatever format would be acceptable to your organization (and feel free to use this final project in your day job).

* A business memo in Word covering both business and technical descriptions
* A PowerPoint presentation focused on recommendations and the business problem with a technical appendix (in Word, PowerPoint, or some other format)
* A Tableau Dashboard/Story with enough context for the business problem, and a technical appendix

**Ultimately, the report should address one or more specific business problem(s) with recommendations or insights, provide insight into the underlying data, contain output from your analyses, and include any artifacts you believe are important in convincing management to take your recommendation.**

Your grade in the final project will be based primarily on your understanding of the business and providing meaningful recommendations that are *well justified with analysis.* Your model accuracy (or lack thereof) will not have implications on your final grade as long as your technical appendix justifies the results. Sometimes, it is useful enough to know you can’t predict something accurately (at least with current data and circumstances). I will not be grading your code, though you are welcome to submit it.

***NOTES REGARDING THE TEAM PROJECT***

* **You are strongly encouraged to begin identifying potential data sets as soon as possible.**
* When thinking about data, be aware that you are individually responsible for the confidentiality obligations associated with your data. You will not be required to submit the data with your analysis, but you will be expected to share the data within the team. Issues around confidentiality can sometimes be addressed by eliminating key identifiers, introducing random rounding, etc.
* When it comes to data, size matters, and within reason, bigger is better. If your dataset is sufficiently large you may consider dividing it into two sets and using one for model development and the other for model testing.
* If you have several potential data sets, you may use different ones for different elements of the project.
* **Please do not use techniques focused on in other classes – for example, discrete choice modelling.**
* In aiming for five activities, you can consider some of the following. This is not an inclusive list.

|  |  |
| --- | --- |
| * Significant merging/joining data in SQL/R | * Visualizing data for insight generation |
| * Dealing with missing/incomplete data | * Significant data exploration |
| * Feature engineering | * Multiple/compound hypothesis tests |
| * Confidence intervals | * Significant cleaning of data |
| * Managing problems with error terms | * Predictive modelling |
| * Joining external (e.g. library) data | * Dashboarding |

Two sample project templates from prior years are provided below:

1. Significant cleaning of data
2. Significant merging/joining of data OR Feature Engineering
3. Predictive Modelling
4. Hypothesis Testing
5. Executive Dashboard

Or,

1. Significant cleaning of data
2. Significant data exploration OR Visualization for insight development
3. Feature engineering
4. Predictive Modelling
5. Executive dashboard