# The University of Texas at Dallas

Naveen Jindal School of Management

MIS 6324/BUAN 6324 Business Analytics with SAS

# Final Project

## Objectives

- Demonstrate the use of various analytical techniques on a real world problem

## Instructions

250 points of the final grade will be determined by your performance on a group term paper. The primary objective of the term paper is to encourage you to explore and think about potential applications of the techniques you will learn in this class. This group project offers you an opportunity to apply your BI knowledge to real-life data and to mine managerially-relevant insights.

The final project will consist of three different submissions:

1. A one-page proposal
2. A progress report
3. A final report

All three deliverables should be submitted via eLearning. eLearning will stop accepting submissions after the due time, and no late submissions will be accepted. Each report should be a single Microsoft Word document with the group number and all group member names clearly identified. For each report, submit one report per group. A professional quality report is expected - messy or hard-to-read reports will be penalized.

## Mining Real Data

This project allows your group to walk through the data mining process using real data that is of some interest to you. Data can come from two possible sources:

1. **First-hand data** directly from your job or your connections. If you choose this option, make sure you have approval from the data owner. This option is strongly recommended.
2. **Second-hand data** may be obtained from a variety of online sources. Some possible sources of data are listed below:
    - **Kaggle** (https://ww.kaggle.com) – This site hosts data mining competitions. You do not have to compete in order to gain access to the data sets.
    - **UCI Machine Learning Repository** (https://archive.ics.uci.edu/ml/) – Free data sets hosted by the University of California Irvine.
    - **Datahub** (https://datahub.io/dataset) – An online repository of user-submitted data sets.
    - **SQLBelle Blog** (https://sqlbelle.com/2015/01/16/data-sets-for-bianalyticsvisualization-projects/) – A large list of data sets from a variety of sources on the web.

You should pick a topic/dataset that is not only interesting, but also understandable to you. You should also have a clear idea about your data mining objective, which should be closely related to the methods covered in this course. It is acceptable to use methods not covered in the course, but the primary focus should remain one (or more) of the methods we have discussed.

## Deliverables

**Proposal:** Clearly indicate whether you'll use first-hand or second-hand data, and how you propose to obtain the data. A brief description of the data should be included so it is clear that you have done enough work to ensure that you will get the data, and know what to do with it. If you wish, you may attach the list of all properties (though this is not required). I'll make a quick assessment based on this proposal whether this project will be feasible and interesting enough. I'll raise flags if I have any concerns.

**Progress Report:** At the beginning of this deliverable, describe your data mining objective and the data. Data should be available and submitted along with this report, so the TA or I can look at the data if needed. Your team should have conducted at least one successful trial of data mining on this dataset (so I know you can do it), and should report your rough findings (so I know you are on the path to get something interesting).

**Final Report:** This should be a professionally prepared report that has the following parts: cover page, executive summary, project motivation/background, data description, your BI model, Enterprise Miner diagrams used, your findings and managerial implications/conclusions, references. Feel free to add other sections if needed. If you feel any of the above required parts should not be included in the final report, you need to get my prior approval by explaining why. There is no page limit (but be reasonable). What really matters is whether you successfully discovered useful knowledge from a dataset, and whether you presented it well to readers - if you can do so in a concise way, that is preferred. However, missing important parts will result in penalties.