

Business Analytics 2 Syllabus

Course Description

Business analytics refers to the ways in which enterprises such as businesses, non-profits, and governments use data to gain insights and make better decisions. Business analytics is applied in operations, marketing, finance, and strategic planning among other functions. The ability to use data effectively to drive rapid, precise and profitable decisions has been a critical strategic advantage for companies as diverse as WalMart, Google, Capital One, and Disney. In addition, many current and recent startups are based on the application of analytics to large databases. With the increasing availability of broad and deep sources of information — so-called “Big Data” — business analytics are becoming an even more critical capability for enterprises of all types and all sizes.

You were introduced to the fundamentals of business analytics in your core ‘Business Analytics’ class. In this class, you will continue your study of Business Analytics, and apply these methods to new cases in a broad range of industries. In particular, we will

- Extend and deepen the methods you learnt in Business Analytics. You will learn how to use these methods in more unstructured and diverse situations, and on a broader range of structured and unstructured data (such as text data).
- Introduce more complex, powerful, and flexible methodologies for predictive analytics than those you covered in Business Analytics, such as random forests.
- Introduce new frameworks such as visualization (in Tableau) and regularization that will supplement any analytics work you do.

Much as Business Analytics does, this course emphasizes that the discipline is not theoretical; we will apply these new methodologies in a number of cases, and use them to develop increasingly powerful insights and predictive capabilities. Many of the techniques we will be covering are now considered standard in industry, and developing a good understanding of them will deepen your ability to identify opportunities in which business analytics can be used to improve performance, drive value, and support important decisions.

This course will not require any coding or prior knowledge other than your core Business Analytics and Statistics classes. However, the material presented will require more mathematical sophistication than your core classes.

Course Materials

There is no required textbook for the class. There will be cases, articles and slides, that will all be posted on canvas.

Requirements and Grading

- 40% : successful completion of in-class exercises
- 15% : in-class participation
- 45% : final exam

You are expected to come to class prepared, and ready to discuss the pre-class reading, case or assignment questions.

Software

This course will require the use of Excel, and we will provide a Business Analytics 2 Excel add-in, which we have developed to extend the functionality of Excel to cover the topics in this follow-up elective. This add-in should work on a Mac natively.

Even though this course only requires you to use Excel, the add-in itself will be powered by Python code. Python has quickly become the lingua franca of business analytics, and those hoping to enter analytics-related industries will likely carry out further study to deepen their knowledge of this programming language. The Python code backing this add-in will be made available to you separately, and should you decide to take further courses in Python, you will be able to return to this code and implement the methods you learn in this class directly in Python.

Detailed Session Plan

Session order, cases, and content subject to change.

- **Session 1:** Introduction

In this class, we review logistic regression and ROC curves and introduce the BA2 Excel add-in, as well as some more advanced topics in Business Analytics, including dummy variables for categorical data, interactions, data standardization, and cross-validation.

- **Session 2:** Variable Selection and Regularization

In Business Analytics, we saw that fitting overly complex models could lead to overfitting, and that this pitfall could be diagnosed using training and test sets. We did not, however, discuss *how* to simplify our models when we detect overfitting.

This fundamental problem arises time and time again in a wide range of business problems – there are often thousands of variables that could be used to carry out a given predictive task, but using *all* these variables would lead to severe overfitting. How should a business pick the ‘best’ set of variables to use?

In this lecture, we study this problem and introduce *regularization*, a modern technique for variable selection, in the context of linear and logistic regression.

Case: *Northstar Payment Services*. Northstar is a credit card processor. Every year, the company re-evaluates its pricing strategy, for new and existing clients. We discuss a predictive analytics tool developed by Northstar to predict how a given customer might respond to price changes in a highly customized fashion.

- **Session 3:** Data Visualization in Tableau

Many of the cases we have discussed thus far have featured companies with a very specific problems, and the way they have used analytical techniques to solve these problems.

In real life, things are rarely this clean. Companies are often faced with ill-defined problems that have no single, obvious solution, and a complex data landscape that does not immediately lend itself to easy analysis. In those situations, businesses need to engage in exploratory data analysis to narrow the scope of their problem, and when datasets are large enough, even the simplest of exploratory tasks can be difficult.

In this lecture, we will discuss the art and science of data visualization using a tool called Tableau, and show how companies can use this tool to leverage their data against their most pressing problems.

- **Session 4:** Regression Trees and Random Forests

In this session, we will introduce one of the most powerful and versatile predictive analytics tools in the analytics toolbox – the random forest. Random forests comprise many smaller and simpler models called classification and regression trees, which are weak individually but reinforce each other to produce highly predictive models. Random forests are particularly well suited to problems with many variables. We will also discuss the main shortcoming of random forests – a lack of interpretability – and discuss ways to remedy this shortcoming.

Case: *Lending Club*. Lending Club is a peer-to-peer lender, which allows the general public to invest in micro-loans with varying risks and returns. In this case, we will see how random forest can be used to construct portfolios of Lending Club loans with maximum return.

- **Session 5:** Text Analytics

One of the most impactful ways the data landscape has changed over the last decade is the availability of large-scale unstructured data as well as structured data. Chief among these are textual data. From financial disclosure statements to tweets and news articles, there is an enormous amount of text data now available electronically, and many companies are extracting valuable insights from these vast volumes of unstructured data.

In this class, we will introduce methods that can be used to carry out these analyses.

Case: *Evisort*. Evisort (evisort.com) is a startup that handles automated contract management for companies using text mining methodologies. We discuss the tools Evisort uses to carry out this work, as well as the legal analytics landscape more generally.

- **Session 6:** Demand Forecasting and Course Wrap-up

In our final session, we will consider one of the most common applications of business analytics that encompasses many of the techniques you have studied to date – demand forecasting. This problem arises in a wide range of industries – be it retail, where companies need to forecast sales, healthcare, where hospitals need to forecast patient volumes, services industries, where companies need to plan staffing based on forecasts, or the media, where advertisers need to forecast viewership.

We will study a methodology that is flexible enough to capture many of the forecast components encountered in real-world situations, but simple enough to be interpretable and customizable by key stakeholders.

Case: TBD