

## Business Analytics<sup>2</sup> 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, on complex real-life datasets, 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. For those of you who will work closely with data science and product teams, the deep knowledge we will develop in this class will prove invaluable.

*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 and slides, that will all be posted on canvas.

## Requirements and Grading

- 45% : final exam (multiple choice, no Excel)
- 20% : homeworks (successful completion and write-up of in-class exercises, and additional practice exercises; one per module)
- 20% : annotated slides
- 15% : in-class participation

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

Before class begins, you will be required to install the BA2 add-in and complete a short survey. Details will be posted on Canvas. Anyone who has not completed the pre-work at least three days before class begins will be removed from the class.

Attendance at the first class is compulsory, because we will be familiarizing ourselves with the add-in which we will use during the rest of the class.

## 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, without the need for a virtual machine.

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 Class Plan

Due to the advanced nature of the material covered in this class, we will focus on quality over quantity, with a strong focus on making sure you understand the concepts in depth before we move on. Based on past iterations of this class, we will most likely only cover modules 1 through 4.

We will cover modules in the following order. You will be expected to read each case before each module.

- **Pre-work:** Before class begins, you will be required to install the BA2 add-in and complete a short survey. Details will be posted on Canvas. Anyone who has not completed the pre-work at least three days before class begins will be removed from the class.

Attendance at the first class is compulsory, because we will be familiarizing ourselves with the add-in which we will use during the rest of the class.

- **Module 1:** Introduction

In this class, we introduce the BA2 Excel add-in. We review linear regression, including advanced topics including dummy variables for categorical data, interactions, and data standardization. We introduce the bias-variance trade-off, a fundamental concept in Business Analytics, and cross-validation, a key tool for model selection.

*Case:* Analyzing Performance in New York City Public Schools

- **Module 2:** Powerful Predictions; Regression Trees and Random Forests

In this session, we will introduce one of the most powerful, versatile, and popular predictive analytics tools used by businesses today – 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:* Data Driven Investment Strategies for Peer-to-Peer Lending – the Case of Lending Club

- **Module 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.

*Case:* Understanding Citibike: Data Visualization and Exploration in Tableau and Python

- **Module 4:** 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 realizing there are valuable insights to be gleaned from this mass of data.

Unfortunately, valuable as these data might be, they are more difficult to analyze than structured data. In this module, we will study techniques that can be used to extract meaning and value from textual data.

**Case:** Evisort: An A.I.-Powered Startup Uses Text Mining to Become Google for Contracts

- **Module 5:** 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:** Pricing at Northstar Payment Services

- **Module 6:** Demand Forecasting and Time Series Data

In our final module, 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:** Demand Estimation at Micky’s Market