

Business Analytics² Syllabus

1. 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.

2. 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. The class will be divided into four modules:

- **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

3. Course Materials

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

For those of you looking for additional reading, I have found the following two resources to be excellent

- *Data Science for Business*, by Foster Provost and Tom Fawcett. This book is pitched at the MBA level, and covers many of the topics we will be covering in this class. It is excellent, but diverges from the approach we will take in this class in two key ways (1) it does go into quite as much depth as we will (2) it does not use cases in the same way this class does; the examples in the book are anchored in business problems, but there are no developed in quite the same way as they will be in the cases we will be using.
- *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. This is the *bible* of machine learning, written by some of the greatest innovators in the field over the last 20 years or so. It is, however, extremely mathematical, and therefore will be out of reach to most MBAs. That said, if you have a particularly quantitative background and want to dive in *much* greater depth into any of the topics in this class, this is the place to go.

4. Requirements and Grading

Before class begins, you will be required complete some pre-work; please see section 2 for details. The class itself will be graded as follows:

	MBA	EMBA	Engineering
Final exam	40%		25%
Group project	<i>Extra credit</i>	40%	25%
Homeworks	35%	35%	25%
Attendance and participation	25%	25%	25%

Please see Canvas for the due dates for each of these components.

These components will be graded as follows:

Final exam : MBA and engineering sections will sit a final exam at the end of the course. The exam will be multiple choice and pen-and-paper only; it will *not* require a computer or calculator.

Group project : EMBA and engineering sections will need to complete a group project in teams of 4–5 students. For your group project, you will need to identify a business problem that might be tackled using the techniques discussed in this class (predictive modeling, text analytics, or visualization) and a dataset on which you can apply those techniques. You will then put together a report listing your findings and conclusions (maximum two pages of single-spaced letter paper, plus unlimited technical appendices, tables, diagrams, and spreadsheets/code. Projects exceeding this page limit will incur a 50% penalty).

Think of this group project as a mini-version of the cases we cover in this class, and use these cases for inspiration. Indeed, to get a 10 on your project, it will need to be of sufficient quality to be used in a later iteration of this course as a case or a homework (and I might indeed use it in that way, with attribution!)

I would be especially impressed if you use a problem you encountered at a company you are working at, or have worked at, with a real dataset, but this is *not* be required to get a 10.

The project will be graded out of 10 as follows:

- **0 points**: no attempt at a project.
- **2 points**: some attempt at identifying a problem that could be tackled with the techniques discussed in the class, presented in a compelling, entertaining report.
- **4 points**: some attempt at identifying a dataset relevant to the problem at hand.
- **6 points**: some attempt at modeling or visualization using the dataset identified to solve the problem at hand, presented in a compelling, logical, and entertaining report.
- **8 points**: a logical, coherent attempt at modeling or visualization using the dataset identified, presented in an easy-to-follow, entertaining report clearly explaining the logic behind the steps you took.
- **10 points**: outstanding project, of comparable quality to the homeworks and case studies in the class.

Note that each of these rubric descriptions require excellence in modeling *and* exposition/presentation.

Homeworks : There will be four homeworks, one for each module. Each homework will be based on a real-world application of the techniques in this class, and will require you to use the tools we will be learning in this class (see ‘tools’ in the next section).

Data science is difficult, and I would be doing you a disservice if I made the homeworks easy. As such, be warned – *these homeworks are designed to be difficult*. To make things fair, therefore, I will *not* grade these homeworks based on correctness – instead, I will grade them based on effort, understanding, and execution on a scale of 1 to 10 using the following rubric:

- **0 points:** no significant effort.
- **2 points:** some questions tackled; evidence some analysis was carried out on the data, but perhaps not correctly.
- **4 points:** all questions tackled; evidence some analysis was carried out on the data, but perhaps not correctly.
- **6 points:** all questions tackled and submitted in a clear, well-presented, and easy-to-follow report clearly explaining the logic behind the steps you took.
- **10 points:** outstanding work, not only answering the questions in the homework and meeting the requirements for 6 points, but also carrying out *further* investigations based on the data given.

Note that each of these rubric descriptions require excellence in modeling *and* exposition/presentation.

Attendance and participation : For your attendance to count for a given class, you must

- Show on *on the dot*. Any late arrivals, even by a minute, will be counted as unexcused absences.
- Display your name card. Anyone in class without a clearly visible name card will be counted as an unexcused absence.

Please note that I am *very* generous with excusing absences – for any reason – provided you let me and the TAs know *at least an hour before* class.

Your participation in each class and your preparedness to discuss case and pre-class assignments will also be taken into account in this category.

Your final grade will be calculated based on the table above. In addition, you will need to meet the following *minimum* requirements to achieve each grade:

H	<ul style="list-style-type: none"> • All the requirements for an HP • Score at least 4/10 on all four homeworks • Participate robustly in class discussions
HP	<ul style="list-style-type: none"> • All the requirements for an LP • Participate in every Poll Everywhere question for classes you attend (no need to answer them correctly, only to participate) • Score at least 4/10 on two of the four homeworks
LP	<ul style="list-style-type: none"> • Attend (or have your absence excused) for at least two thirds of classes • Do not attempt to murder the professor

5. Deliverable Milestones

The following lists all deliverables and pre-class work you will need to complete for this class. You will find a calendar entry on Canvas for each of these items, with a due date and time for your particular section of the class.

When?	What?	Deliverable?
Three days before class 1	Pre-class work	Canvas survey
Before class 2	Pre-work for module 2	
Before class 3	Pre-work for module 3	
	Submit project groups	By email to TAs
	Homework 1 due	Homework 1 on Canvas
Before class 5	Pre-work for module 4	
	Homework 2 due	Homework 2 on Canvas
Before class 6	(MBA/Eng) Homework 3 due	Homework 3 on Canvas
1 week after final class	(EMBA) Homework 3 due	Homework 3 on Canvas
	Homework 4 due	Homework 4 on Canvas
	All projects due	By email to me

Please note that MBAs wishing to complete a project for extra credit must do so – and submit their group to the TAs – before class 3.

6. 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. You will be installing this add in on your computer as part of the pre-work for the class.

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.

Solutions to all the Homeworks will be provided in the add-in *and* in Python. You are welcome to complete the homeworks using *either* tool.