

**E4650****BUSINESS ANALYTICS**

Fall Semester 2018

Professor Daniel Guetta

Business analytics refers to the ways in which enterprises such as businesses, non-profits, and governments can 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. For example, Capital One uses sophisticated analytic capabilities to match credit card offerings to customers more accurately than their competition. WalMart uses analytics to monitor and update its inventory in a way that allows it to serve its customers at an exceptionally low cost. In addition, many current and recent startups such as Splunk 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 is becoming an even more critical capability for enterprises of all types and all sizes.

In this course, you will learn how to identify, evaluate, and capture business analytic opportunities that create value. Toward this end, you will learn basic analytic methods and analyze case studies on organizations that successfully deployed these techniques. In the first part of the course, we focus on how to use data to develop insights and predictive capabilities using machine learning, data mining, and forecasting techniques. In the second part, we focus on the use of optimization to support decision-making in the presence of a large number of alternatives and business constraints. Finally, throughout the course, we explore the challenges that can arise in implementing analytical approaches within an organization.

The course emphasizes that business analytics is a practical discipline which requires mastery of both methodology and business applications. The concepts learned in this class should help you identify opportunities in which business analytics can be used to improve performance and support important decisions. It will teach you important tools that can be used to transform

data into high-impact business decisions. Lastly, it should make you alert to the ways that analytics can be used — and misused — within an organization

We have three goals in this course. The first is to help you think critically about data and the analyses based on those data — whether conducted by you or someone else. The second is to enable you to identify opportunities for creating value using business analytics. The third is to teach you essential tools and theory so that you can apply business analytics methods yourself.

---

<b>Class Times</b>	<i>Section 1:</i> Tuesdays, 10am – 1pm <i>Section 2:</i> Tuesdays, 2:30pm – 5:30pm
<b>Room</b>	<i>Section 1:</i> Mudd, 1127 <i>Section 2:</i> International Affairs Building, 403
<b>Schedule</b>	The first class is Tuesday, September 4 <sup>th</sup> consistent with the Business School Schedule.  There will be no class on the following weeks <ul style="list-style-type: none"><li>• Tuesday October 9<sup>th</sup></li><li>• Tuesday October 23<sup>rd</sup>: Fall Break</li><li>• Tuesday November 6<sup>th</sup>: Election Day</li></ul> The last class will be on Tuesday December 11 <sup>th</sup>
<b>Recitations</b>	The TAs will run weekly recitations on Fridays. Please check Canvas for details.
<b>Instructor</b>	Prof Daniel Guetta <a href="mailto:guetta@gsb.columbia.edu">guetta@gsb.columbia.edu</a>  <i>Office hours:</i> Mondays, 10:00am – 11:00am <b>by appointment; please email me to let me know you're coming</b>
<b>Teaching Assistants</b>	Mo Yao ( <a href="mailto:my2540@columbia.edu">my2540@columbia.edu</a> ) Valentina Semanova ( <a href="mailto:vms2148@columbia.edu">vms2148@columbia.edu</a> )

*Office hours:* Check Canvas

## Course Texts

There is no required textbook; all materials will be on our course website. For those who wish to learn more about Business Analytics, the following books are useful resources that will reinforce and expand upon the materials learned in class

- J. Ledolter, *Data Mining and Business Analytics with R*. Wiley, 2013. (Easy to follow with many examples and data sets).
- G. James, D. Witten. T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning with applications in R*. Springer, 2013. (Solid primer, with theory and practical applications; you can download this book [here](#)).
- T. Hastie, R. Tibshirani, J. Friedman. *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (This is a classic book, but requires some mathematical sophistication and goes beyond the material we will be covering. You can download this book [here](#)).

## Software and coding

- The class will mostly use R, which is a powerful and free statistical analysis software package. R is often used in academia and industry, and contains many useful packages that implement many of the techniques we will be discussing in this class.
- The class does not assume prior programming knowledge, and it is not an R programming class. We will provide limited software instruction, in-class demonstration, and code to accompany lectures and assignments. Like any programming language, R is best learned through practice. The learning curve is steep, but students can become proficient in a few weeks with reasonable effort.
- You should install R as soon as possible and familiarize yourself with basic operations. The more proficiency you can gain in R prior to class, the more you will get out of the sessions. Some manuals are very helpful to learn R, e.g., <http://cran.r-project.org/manuals.html>
- R can be downloaded at the link below: <http://www.cran.r->

[project.org/](http://project.org/)

- Rstudio is a recommended free platform for both writing and running R, available at [www.rstudio.org](http://www.rstudio.org). A good tutorial to R can be found here: <http://data.princeton.edu/R>
- We will also provide a business analytics Excel add-in developed at Columbia. The add-in includes special functions for linear regression, logistic regression, k-nearest neighbors, classification and portfolio optimization. The add-in is available on our course website and is free of charge.

### **Pre-class work**

Many of our classes will focus on case studies that will be distributed before class. Students are expected to come to class prepared and ready to discuss the material in those cases, including any assignment questions. See the course calendar for more details.

### **Homework and deliverables**

Hands on experience is important to develop the skills learned in the class. As such,

- Students will be required to submit two individual homework assignments. These will be long, and it is strongly suggested you start working on them far before the deadline.
- Students will be required to work on a project using a real data, in teams of five students. Each team will be required to submit a project proposal and description, a project report, and to present their projects in our last lecture.

See the course calendar for due dates and more details.

To ensure you can use homeworks to learn without undue pressure, they will be given one of two scores:

- **0**: no or very poor attempt
- **10**: good attempt, with a strong effort made at answering every question.

Projects and presentations will be graded using a rubric to be distributed closer to the time.

**Exams**

There will be one midterm exam and a final exam. The midterm exam will occur in class on **Tuesday October 16<sup>th</sup>**. The final will occur on **Sunday December 16<sup>th</sup>, 1pm – 6pm in Uris 301**.

**Grading**

I will calculate a grade for each student using two methods – the method that results in the *higher* percentage will be used to calculate your grade for the course:

- Method 1

Final exam	55%
Project	35%
Class participation	10%

- Method 2

Final exam	30%
Mid-term exam	20%
Homeworks	20%
Project	20%
Class participation	10%

---

# BUSINESS ANALYTICS

## CLASS SCHEDULE (SUBJECT TO CHANGE)

Class	Date	Topics and Assignments
1	Tues, Sept 4 <sup>th</sup>	<b>Introduction</b> <b>Predicting Outcomes I (Nomis)</b> Read before class: <ul style="list-style-type: none"><li>• Nomis A case</li><li>• Nomis B case</li></ul>
2	Tues, Sept 11 <sup>th</sup> ( <i>project milestone 1</i> )	<b>Intro to Statistical Learning</b> <b>Linear Models</b> Due before class: <ul style="list-style-type: none"><li>• Complete the Google form here to submit your project teams (one submission per member only): <a href="https://goo.gl/forms/2MwNQ7HXbH83Grt63">https://goo.gl/forms/2MwNQ7HXbH83Grt63</a></li></ul>
3	Tues, Sept 18 <sup>th</sup>	<b>Quality of Predictions I (Financial Analytics)</b> <b>Resampling Methods</b>
4	Tues., Sept 25 <sup>th</sup>	<b>Subset Selection: the Ride and LASSO</b> <b>Logistic Regression</b>
5	Tues, Oct 2 <sup>nd</sup>	<b>Predicting Outcomes II (Pandora)</b> <b>k-Nearest Neighbors and Tree-Based Methods</b> Read before class: <ul style="list-style-type: none"><li>• Pandora case</li></ul>

---

		<i>No class</i>
	Tues, Oct 9 <sup>th</sup> <i>(project milestone 2)</i>	Due by 11:59pm: <ul style="list-style-type: none"> <li>• Homework 1 (on canvas)</li> <li>• Project proposals (on canvas)</li> </ul>
<b>6</b>	Tues, Oct 16 <sup>th</sup>	<b>Midterm (in class)</b>
	Tues, Oct 23 <sup>rd</sup>	<i>No class (Fall Break)</i>
		<b>Quality of Predictions II (Tahoe Healthcare)</b> <b>Sports Analytics</b>
<b>7</b>	Tues, Oct 30 <sup>th</sup>	Read before class: <ul style="list-style-type: none"> <li>• Tahoe Healthcare Case</li> <li>• “The Sabermetric Revolution”</li> </ul>
	Tues, Nov 6 <sup>th</sup>	<i>No class (Election Day)</i>
		<b>PCA and Partial Least Squares</b>
<b>8</b>	Tues, Nov 13 <sup>th</sup>	<b>Clustering</b>
		<b>Difference in Differences (BOPS)</b> <b>Causal Inference</b>
<b>9</b>	Tues, Nov 20 <sup>th</sup>	Read before class: <ul style="list-style-type: none"> <li>• BOPS case</li> </ul>
		<b>Simulation</b>
<b>10</b>	Tues, Nov 27 <sup>th</sup>	<b>Optimization</b>  Read before class:

- Read GM Case
- Read XGen case

11	Tues, Dec 4 <sup>th</sup>	<p><b>Multiple Objectives</b></p> <p><b>Analytics at Zara</b></p> <p><b>Analytics and Fairness</b></p>
		<p>Due before class:</p> <ul style="list-style-type: none"> <li>• Homework 2 (on canvas)</li> </ul> <p>Read before class:</p> <ul style="list-style-type: none"> <li>• Zara Case</li> </ul>

**Project Presentations**

12	Tues, Dec 11 <sup>th</sup> <i>(project milestones 3 and 4)</i>	<p>Due before class:</p> <ul style="list-style-type: none"> <li>• Final project report (on canvas)</li> <li>• Final project slides (on canvas)</li> </ul>
----	---	---

	Wed, Dec 12 <sup>th</sup> <i>(project milestone 5)</i>	<p>Due by 11:59pm:</p> <ul style="list-style-type: none"> <li>• Complete the Google Form here to evaluate your fellow team members confidentially: <a href="https://goo.gl/forms/mzj9F2MSmLrQkGjZ2">https://goo.gl/forms/mzj9F2MSmLrQkGjZ2</a></li> </ul>
--	---	---

Sunday  
December 16<sup>th</sup>      **Final exam (1pm – 6pm)**  
**Uris 301**