

ETM 58D BUSINESS ANALYTICS

Spring 2019

"For every leader in the company, not just for me, there are decisions that can be made by analysis. These are the best kinds of decisions. They're fact-based decisions."

Amazon's CEO, Jeff Bezos.

Instructor: Mustafa Gökçe Baydoğan (mustafa.baydogan@boun.edu.tr)

Lecture hours: Monday 19:00-21:50

Office Hours: by appointment

Course website: To be decided

Reading materials and resources:

- *An Introduction to Statistical Learning with Applications in R*, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer, NY, 2013
(available online: <http://www-bcf.usc.edu/~gareth/ISL/>)
- *An Introduction to Data Science* (2013) version 3 by Stanton
- *Business Analytics for Managers* (2011) by Jank
- Udacity Course – Data Analysis with R: <https://classroom.udacity.com/courses/ud651>
- *Big Data @ Work: Dispelling the Myths, Uncovering the Opportunities* by Davenport
- *R for Data Science* by Wickham
- DataCamp

About the course:

In Harvard Business Review issue (January 2006), Prof. Thomas Davenport and his colleagues mention about a new form of competition based on the extensive use of analytics, data, and fact-based decision making. Currently employed competitive strategies by almost every industry are claimed to rely extensively on data analysis to predict the consequences of alternative courses of action, and to guide executive decision making. Extensive interviews with executives from successful firms show that companies require decision makers who understand the value of analytics, can identify opportunities and know how best to apply data analytics to enhance business performance. The main motivation of such requirement is the creation of massive collections of data as a byproduct of the activities of businesses, governments, and individuals. To improve decision-making, decision-makers and systems rely increasingly on intelligent approaches to analyze data systematically. Automation of analytical and decision-making processes is necessary in many cases with the increase in the volume of data and the speed with which new data are generated.

This course has a simple objective. You are expected to perform an end-to-end reproducible data analysis from raw data to the communication of model outputs. The aim is to learn how data analysis approaches can be used to improve decision making.

Warning: Course load is heavy. Free R software will be used for examples.

Prerequisite background:

A working knowledge of basic statistical methods. A formal course in engineering statistics is the official prerequisite. A previous course in empirical modeling such as regression analysis or design of experiments is recommended. A formal education in one of the programming languages is a plus.

Course objectives:

After taking this course you should:

- Proven ability to manipulate data sets and creating summary tables
- Proven ability to visualize data with the proper choice of tools (e.g. histogram, scatterplot, pie charts)
- Proven ability to code with R and related packages (e.g. data.table)
- Proven ability to perform input/output operations
- Proven ability to perform reproducible research
- Proven ability to apply data mining algorithms (e.g. regression, logistic regression) and interpret the output
- Proven ability to communicate findings of analyzed data in a coherent and understandable way

Grading Criteria

Your course grade is determined from a final exam (35%), 5-6 homework (total of 30%) and group project (35%) which will be developed continuously during the semester.

Requirements:

- A group final project is used to apply techniques from the course on a larger data set. Here, the learning is put to use as a sequence of steps is developed and implemented for a provided data set and problem objective. In the project, when one attempts to develop and implement a model, the complexities of an analysis can arise. Still, the modeling fundamentals provide the guidance for an effective solution.
- The final exam is comprehensive and it will be take-home.
- Academic integrity is expected. Your work is to be your own.

Topics to be covered

Exploratory Data Analysis (EDA)

Basics of Base R

Preprocessing of data. Feature selection and feature reduction.

- ❖ Working with dirty data: handling different data types (nominal, ordinal, etc.), imputation of missing data
- ❖ Toward analytical engineering: variable selection, feature engineering

Descriptive analytics

Predictive analytics

Fundamentals of model selection

- ❖ How do I produce a focused segmentation? What is a model? Basic terminology
- ❖ How do I know my model is any good?