DATA ANALYTICS: TOOLS FOR BIG DATA
Master course (ECTS: 7)

Lectured by:
Mamani Hamed, University of Washington, Michael G. Foster School of Business, USA

Aims of the course:
Data analytics is a set of techniques that enterprises use to gain insight from their data and make better decisions. Many firms in a variety of industries use these techniques: Google, Amazon, Target, Coca-Cola, WalMart, Capital One. These techniques are also applicable to the many functional areas of business, such as operations, marketing, accounting, finance, etc. Furthermore, the modern abundance of data, so-called “Big Data,” underscores the value that analytics can provide a firm, be it non-profit, for-profit, or government.

This course introduces data analytic techniques via quantitative tools and sophisticated software (R, Rattle and Tableau). These techniques are drawn from machine learning, data mining, and optimization. Note that this is not a technical or theoretical course. This course does not aim to produce experts in statistical analysis; rather, the aim is to provide students competency to interact with and manage a team of analytics professionals. Furthermore, this is not a technical or theoretical course; we will instead focus on the application of analytics techniques to real business situations, with the aim of creating insight and value.

The course is divided into the following six main modules under three topic areas of Descriptive Analytics, Predictive Analytics, and Prescriptive Analytics:

1. Introduction and Visualization (Descriptive and Inferential Analytics)
2. Supervised Learning: Prediction (Predictive Analytics)
3. Supervised Learning: Classification (Predictive Analytics)
4. Unsupervised Learning: Clustering (Descriptive and Predictive Analytics)
5. End to End Analytics: Optimization and Learning (Prescriptive Analytics)
6. Projects

Our course goals are the following:
1. Students should be able to think critically about data analysis, which includes selecting the right type of analysis for a given task.
2. Students should be able to identify opportunities of applying data analytics, in real business settings.
3. Students should be well equipped to become data-savvy managers.
To achieve the above goals, lectures will cover the major concepts and analytical tools. Cases and practice problems will allow you to analyse different industry settings, analyse different company strategic problems, and identify key issues related to data and modelling.

**Course syllabus:**

Session 1:
Introduction to Data Analytics and R
Introduction to Visualization

Session 2:
**Descriptive Analytics:**
Visualization: Tableau I

Session 3:
**Descriptive Analytics:**
Visualization: Tableau II
Data Manipulation
Data Preprocessing and Cleaning

Session 4:
**Predictive Analytics:**
Prediction: Linear Regression (Supervised Learning)
- Predicting Wine Quality
- Sports Analytics

Visualization Assignment Due

Session 5:
**Predictive Analytics:**
Classification: Logistic Regression (Supervised Learning)
- Predicting Quality in Healthcare
- Framingham Heart Study

Linear Regression Assignment Due

Session 6:
**Predictive Analytics:**
Prediction and Classification: CART (Supervised Learning)
- Predicting Supreme Court Decisions
- Healthcare Cost Prediction

Logistic Regression Assignment Due
Session 7:
**Descriptive Analytics:**
Clustering (Unsupervised Learning)
- Netflix and Recommendation Systems
- Hearth Attack Prediction
*CART Assignment Due*

Session 8:
- Guest Speaker
- Homezilla Case Study
*Clustering Assignment Due*

Session 9:
**Team Presentations**

**Bibliography:**

**Teaching methods:**
Computer Software
We will be using Tableau, R, and Rattle in this class. These software packages are both available in the computer labs. Student licenses to Tableau will be provided; R and Rattle are free.
- Tableau can be downloaded at http://www.tableau.com/products/trial
- R can be downloaded at https://cran.fhcrc.org

**Assignments**
Most days a case will be assigned. Students are expected to work either individually or in groups on the cases. However, each student must submit his/her own solution. These assignments are due at the beginning of class on their posted due date and are electronically submitted through the class website.

**Group Project**
The final module of this course will consist of a group project where teams will apply the techniques of the course to real data. Further details will be released later in the course.
Examination methods:
Grades will be calculated applying the following allocation in a simple weighted average:

1. Assignments  50%
2. Group Project  25%
3. In Class Assignments  20%
4. Participation  5%