

# SYLLABUS

**Technology Fundamentals for Business Analytics**

**MGMT6963**

**3 Credits**

**Fall/2015 Monday 6:00 – 8:50**

**Room Location: Pittsburgh Building 4114**

**Websites:**

**Prerequisites or Other Requirements**: None.

***INSTRUCTOR***

**Instructor Name:** Jason Kuruzovich **Office Location:** Pitt. 4th Floor (Sev. Center)

**Tel. No.:** 518-698-9910 **Email Address:** kuruzj@rpi.edu

**Office Hours:** Thusday 10:00 AM-12:00 PM

***TEACHING ASSISTANT***

**TA Name(s):** TBD **TA Office Location:**

**TA Office Hours:** **TA(s) Email Address:**

## COURSE DESCRIPTION:

The widespread proliferation of IT-influenced economic activity leaves behind a rich trail of micro-level data, enabling organizations to use analytics and experimentation in both strategy and operations. This course provides a hands-on introduction to the concepts, methods and processes of business analytics. We will learn how to obtain data and draw business inferences from data by asking the right questions and using the appropriate tools.

## COURSE GOALS/OBJECTIVES

The goal of this course will be to provide the technical foundation to enable students to become data scientists.

## STUDENT LEARNING OUTCOMES

Through this course you are expected to:

1. Demonstrate an understanding of analytics based problem solving and analytics thinking.
2. Be able to extract, match, transform, and clean data from a variety of sources.
3. Conduct structured statistical analyses to answer business related questions.

## COURSE ASSESSMENT MEASURES

**Class Participation and In-Class Assignments**

Participation in class is absolutely necessary. I expect this to be a hands on class where you will be actively working on problems, demonstrating solutions, and helping your colleagues. This may also include in-class assignments designed to push absorption of the material.

**Assignments (Weekly)**

You will have assignments associated with many classes. Be sure to check the course website and submit assignments via the LMS> Preparing the case questions in advance is expected to help you participate and contribute to class discussion. Assignments up to 24 hours late will have their grade reduced by 25%; assignments up to one week late will have their grade reduced by 50%. After one week, late assignments will receive no credit. Please turn in your assignment early if there is any uncertainty about your ability to turn it in on time.

**Kaggle Competition / Organizational Assignments (2)**

There will be two individual assignments to be completed as part of the class.

*Assignment 1 – Solution Assessment and Analysis.*

[*https://www.kaggle.com/c/titanic-gettingStarted*](https://www.kaggle.com/c/titanic-gettingStarted)

The goal of the first assignment is to address an existing Kaggle sample project and identify and diagnose 3 different solutions to the project, research any methods used, assess the relative quality of each, and make a recommendation on which is the best to utilize for the context. We will go through the process of how to go through the process of identifying and analysing solutions in class.

The Titanic is one of the most well developed and documented projects. We will all go through this.

*Assignment 2 – Solution Creation.* In this assignment you will select a challenging problem from Kaggle or from our relationships with different organizations and develop a unique solution to the problem. The solution may employ methods of analysis examined in the class or others you have investigated.

**For complete details, see the course web page.**

**GRADING CRITERIA**

All grading is out of 100%. Grade breakdown and feedback will be given through the course learning management system.

|  |  |
| --- | --- |
| **Component** | **Weight** |
| Participation and Homework Assignments and Quizes | 30% |
| Kaggle/Organizational Assignment (Due 13th Class) | 30% |
| Midterm (7th Class) | 25% |
| Final Exam (14th Class) | 25% |

## ATTENDANCE POLICY

A maximum of 2 unexcused absences are allowed. After that 10% of the overall participation grade will be removed for each missed class.

Labs are designed to be worked on during class time in a collaborative environment in which you each help one another. We have very different levels of technical expertise in the class, and it is important to work together on the labs to help one another in a classroom environment. While you may be able to complete the work at home by yourself, in doing so you will be robbing your classmates of your expertise. I will consider helping behaviours during the class as part of the participation component of the class.

Therefore, the imperative clearly stated: each participant attends class fully prepared, willing and able to offer constructive criticism, provide goal-oriented analytic and synthetic insights, and encourage investigative dialectic. You earn your grade on participation through consistent, daily contribution. Merely "COMING TO CLASS” is not sufficient, but is necessary.

Simply put: Do not miss class hours or group meetings! Understandably, there are circumstances (e.g., job interviews, family matters, extracurricular activity, etc.) that may cause you to miss class; nevertheless, excessive absences will reduce your class participation grade. Notify the instructor and group IN ADVANCE of any planned absences (especially students who participate in extracurricular activities as representatives of RPI.)

## OTHER COURSE POLICIES

After the add deadline, assignments up to 24 hours late will have their grade reduced by 25%; assignments up to one week late will have their grade reduced by 50%. After one week, late assignments will receive no credit. Please turn in your assignment early if there is any uncertainty about your ability to turn it in on time.

## ACADEMIC INTEGRITY

Student-teacher relationships are built on trust. For example, students must trust that teachers have made appropriate decisions about the structure and content of the courses they teach, and teachers must trust that the assignments that students turn in are their own. Acts that violate this trust undermine the educational process. The Rensselaer Handbook of Student Rights and Responsibilities and the Graduate Student Supplement (For Graduate courses) define various forms of Academic Dishonesty and you should make yourself familiar with these. In this class, all assignments that are turned in for a grade must represent the student’s own work. In cases where help was received, or teamwork was allowed, a notation on the assignment should indicate your collaboration. Submission of any assignment that is in violation of this policy will result in a ***grade of zero will be given on the first assignment where a violation is detected, and the infraction will be reported to the Associate Dean for Academic Affairs. If there is a subsequent infraction the student will receive a grade of F for the course*.**

If you have any question concerning this policy before submitting an assignment, please ask for clarification.

## COURSE CALENDAR

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Date** | **Topic** | **Readings/Notes** |
| 1 | 8/29 | The Rise of Analytics and the Tools of the Data Scientist | 1. Join Analytics Dojo 2. Analytics Platform 3. Rise of Analytics |
|  | 9/5 | Labor Day |  |
| 2 | 9/12 | Data Types and Structures | 1. Concepts – Data Structures 2. Python – Overview and Packages 3. Python – Data structures 4. R – Overview and Packages 5. R – Data Structures |
| 3 | 9/19 | Data Sources I | 1. Python – Accessing Data 2. R – Accessing Data |
| 4 | 9/26 | Programming Logic | (1) [Subsetting with R](http://adv-r.had.co.nz/Subsetting.html) (2) [Data Aggregation with R](http://www.stat.ubc.ca/~jenny/STAT545A/block04_dataAggregation.html) (3) [Merging Data in R](http://www.princeton.edu/~otorres/Merge101R.pdf)  Lab 3 Due (Blackboard) |
| 5 | 10/3 | Transforming Data | (1) [Missing data in R](http://www.statmethods.net/input/missingdata.html) (2) [The Danger of Overfitting](http://blog.kaggle.com/2012/07/06/the-dangers-of-overfitting-psychopathy-post-mortem/) (3) [Dealing with Missing Data](http://www.bu.edu/sph/files/2014/05/Marina-tech-report.pdf) (4) [Cross Validation in Python](http://scikit-learn.org/stable/modules/cross_validation.html)  Lab 4 Due (Blackboard) |
|  | 10/10 | Columbus Day - No Class | Data Science for Business Chapters 2-3  Columbus Day/Class Tuesday Rather than Monday |
| 6 | 10/11 | (Tuesday) Linking and Aggregating Data | Data Science for Business Chapters 4  Kaggle 1 Assignment Due 10/23; 10/23 is Drop Deadline |
| 7 | 10/17 | Missing Data and Feature Creation | None. |
| 8 | 10/24 | Midterm | [David McCandless TED Talk](http://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization?language=en); [Introduction to Tableau](http://www.tableau.com/learn/tutorials/on-demand/getting-started?signin=121f83d099689c4afb5b35f8d7906e0d) |
| 9 | 10/31 | Intoduction to Visualization | Chapter 10: Representing and Mining Text; [Mining Twitter](https://rawgit.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition/master/ipynb/html/__Chapter%201%20-%20Mining%20Twitter%20(Full-Text%20Sampler).html) (we will now cover second part)  Lab 6 Visualization Due at 9:00 PM |
| 10 | 11/7 | Introduction to Modeling | Chapter 7: Decision Analytics Thinking I: What is a good model?; Chapter 8: Visualizing Model Performance |
| 11 | 11/14 | Introduction to Modeling II | None – Continue to work on |
| 12 | 11/21 | Text Mining and Unstructured Data | [Spark Overview](http://www.infoq.com/articles/apache-spark-introduction)  Kaggle 2 Assignment Due 12/1 11:59 PM |
| 13 | 11/28 | Introduction to Big Data with Spark |  |
| 14 | 12/5 | Spark Continued |  |