

Analytics and the Digital Economy-INTRO

OIDD 215, Spring 2018

DRAFT SYLLABUS - Subject to change

Instructor

- **Professor Prasanna (Sonny) Tambe**
- tambe@wharton.upenn.edu, JMHH 558

Teaching Assistants

- To be determined

Office hours

- To be determined

Course Objectives

The goal of this Introductory course is to provide students with hands-on experience with the world of data science projects. In doing so, a course objective is to ensure that students who complete the course are comfortable in any business or policy environment where data are extensively used to inform strategic decision-making. Students should leave the course with an understanding of what is required to build data products, and with the confidence that they have the skills necessary to acquire, analyze, and communicate insights in a data rich environment.

The course is oriented around hands-on in-class exercises, homeworks, and labs. Students will be expected to leave the class with a level of proficiency in modern data analysis tools. Broadly, here's what you'll learn from the course, and why these things are important:

- First, data rich firms in finance, tech, management, marketing, and other industries are increasingly being confronted with a new class of problems — how to manage their data assets. It is important for modern managers to understand the opportunities and challenges introduced by these data assets so that they can credibly communicate about these issues with others in the firm. We will cover many of these issues, so that you will be able to think about the opportunities and challenges that arise when firms try to use data to solve business problems.
- Second, as a future manager, marketing analyst, financial analyst, consultant, etc. it is increasingly likely that you will require the skills to acquire data, to model and analyze data, and to communicate the analysis of your data. These tools are likely to provide a powerful “edge” in the future labor market, and in this class, you will gain hands-on experience with some of the more popular data analysis tools, including R and Tableau, and you will gain experience in breaking down data sets and communicating your findings through both visualizations and presentations. You will also learn about some of the key challenges that arise when working with data.

Projects throughout the course will reinforce your learning of how to use data analysis to solve business problems.

Course Overview

Over the last decade, there has been a dramatic rise in the use of tech skills and data analytic thinking to solve business problems in many domains, including finance, HR, policy, and strategy. As a result, the modern “analytic leader” increasingly requires the use of technology, statistics, and data analysis skills to facilitate business analysis. This includes knowing how to a) effectively frame data-driven questions, b) analyze data, and c) use a new generation of tools that is becoming available to acquire, analyze, interpret, and communicate insights derived from data. Students that take this course will engage with the world of data analysis using tools such as Tableau and R that are becoming increasingly popular in industry.

The Intro segment of the course is designed for students with limited experience with data analysis projects, and while familiarity with R, via courses such as STAT 405 or STAT 470, will be ideal preparation, students with other

programming exposure can pick up the required skills via review sessions and self-instruction. The second 0.5 CU, Advanced, course will extend students' experience to industry applications of text mining and machine learning and require students to work with more unstructured data. In contrast to the first course, the Advanced module will rely heavily on R and will require the completion of STAT 405, STAT 470, or equivalent preparation.

Throughout the semester, each week of the course will be devoted to analysis of a data set from a particular industry (e.g. HR, sports, fashion, real estate, music, education, politics, restaurants, non-profit work), which we will use to answer business questions by applying analytic techniques. Beyond applications of data tools and methods, a learning goal of this course is exposure to how data is changing decision-making in different industries. The course is *extremely* hands on, and each week focuses on the application of a particular set of tools or analytic methods. Limited time will be devoted to lectures. Most class time will be devoted to supervised work on weekly data projects. Through these exercises, students are expected to become proficient at applying data to business decisions and at effectively analyzing big data sets to inform decisions about business problems using data analysis tools.

Course web site

We will be using Canvas to submit assignments and receive grades. All course information will be posted on the course website. Course communication will be primarily through [Slack](#).

Required textbooks and software

There is no textbook. Occasional readings will consist of selected online content which will be posted on the course site. As part of your homework, you will also be expected to complete some online courses that supplement what we do in class. The majority of the homework requirements involve working on data analysis projects.

Deliverables and grading

During this course, you will be assigned a number of hands on data projects which you will spend time on both in class and out of class. You will have one written exam that will take place during class time. You are expected to participate in classroom discussions (there is more information about participation below). The breakdown of points is as follows:

Data Labs	25%
Exam	25%
Final Project	25%
Professionalism + Participation + Slack	25%

With each project, you will be provided with a set of guidelines. You can expect to use various data analysis tools extensively, including R and Tableau. We may also, to a limited extent, explore the use of Python and SQL for data analysis/visualization.

In corporate America, you will be expected to present your analytic findings and make a recommendation. Therefore homework deliverables may include short, informal analyses and an accompanying recommendation.

Group projects will be completed in small groups (two to three students, no more than three). You may also be asked to evaluate the contribution of each of your team members after the group project.

The classroom presentation and discussion presents a unique opportunity for you to develop and enhance your confidence and skills in articulating a personal position, sharing your knowledge, and reacting to new ideas. All of you have personal experience that can enhance our understanding of this subject, and we want to encourage you to share that experience.

Participation and Professionalism

This course, like many other courses at Wharton, uses learning methods that require active involvement (e.g. attendance, participation in discussions, and in-class exercises). Not only is this the best way to learn, but it also develops your communication and presentation skills. Regular attendance, participation, presentations, and in general, presenting yourself professionally are all very important, and are an important part of your grade. Active participation requires good preparation—thoughtful completion of homework before class

is essential. We recognize that expressing viewpoints in a group is difficult, but it is an important skill for you to develop. We will do what we can to make this as easy as possible. Remember though that only regular and insightful contributions will be rewarded.

The grade we assign for your class participation and attendance is a careful, subjective assessment of the value of your input to classroom learning. We keep careful track of attendance, your contributions towards each class session, and these contributions can include (but are not restricted to) raising questions that make your classmates think, providing imaginative yet relevant analysis of a situation, contributing background or a perspective on a classroom topic that enhances its discussion, providing thoughtful feedback on the presentations of other students, and simply answering questions raised in class. A lack of preparation, missing classes without justification, negative classroom comments, or improper behavior (such as talking to each other, sleeping in the classroom or walking in and out of the class while the lecture is in process) can lower this grade.

Grading Guidelines

At Wharton, we strive to create courses that challenge students intellectually and that meet the Wharton standards of academic excellence. If you believe that an assignment or project grade you received was unjustified, you can appeal the grade. To appeal the grade you must write a one-page explanation as to the reason for your appeal and hand it along with your graded assignment back to the TA responsible for that assignment. Please think twice before appealing a grade: the TA will completely re-grade the assignment, which may increase your grade, but may also lower it (e.g., if the TA catches more mistakes the second time around). If after re-grading you feel that your grade was again unjustified, you can appeal the grade with the instructor.

Overview of Course Schedule for Intro Module (Q3)

Session	Topic	Date	Due
1	Introduction to the course	Jan 10	
2	Using Tableau	Jan 17	Install Tableau
3	Lab 1A: Citibike	Jan 22	
4	Scraping data I	Jan 24	Lab 1A
5	Lab 2A: Moneyball	Jan 29	
6	Lab 2A: Moneyball	Jan 31	Install R-studio
7	Mini-review of R with applications I	Feb 5	Lab 2A
8	Mini-review of R with applications II	Feb 7	
9	Lab 3A: Baby names	Feb 12	
10	Lab 3A: Baby names	Feb 14	Lab 3A
11	In class assesement	Feb 19	
12	Using API's: Facebook, Google, and Uber	Feb 21	
13	Scraping data II	Feb 26	
14	Visualizing data	Feb 28	Final Project

Details of Individual Sessions

Session 1: Introduction to course

- Session Objectives: Introduce course and course goals. Assist students with course setup.
- Due before leaving class today:
 - Follow directions for installing Tableau
 - Fill out survey
 - Join Slack team for the class
 - * Upload a picture of yourself
 - * Add your proper name

Session 2: Introduction to Tableau

- Session Objectives: This session presents an overview of how to use the Tableau software package for data analysis and visualization, including a discussion of how these new data visualization and analysis skills are changing the labor market they are likely to face and working on some hands-on applications during class.
 - Read/review:
 - Make sure Tableau is working on your laptop before class.
 - *Big data or pink slip*
 - *What Data Scientists Do*
-

Session 3: Lab 1 (Citibike)

- Session Objectives: This session asks students to work on a lab assignment in which they use Tableau along with the NYC Citibike data to answer strategic questions that Citibike faces. In the lab, students are specifically asked how to modify Citibike's service/product offering in a way that addresses one of Citibike's current business challenges.
 - Citibike lab
 - Read/review:
 - *9 languages for crunching data*
 - *R finds fans in data analysts*
-

Session 4: Scraping data I

- Session Objectives: This session is a primer on the use of web-based tools that can be used to scrape data from the web (e.g. import.io) and clean the data (e.g. Google Refine). We also engage in a discussion of the opportunities and challenges associated with responsible web scraping.

- Read/review:
 - *Using Google Refine to Clean Data*
 - *Janitor work is key*
 - *A legal primer on web scraping*
 - Citibike lab is due
 - Install OpenRefine before class
-

Session 5: Lab 2A (Moneyball)

- Session Objectives: This lab is a hands-on application of the web scraping and wrangling tools discussed in the prior sections. Students are asked to collect salary, draft, and performance data for current NBA players from the web using tools for web scraping and to use Tableau to develop a draft-day Tableau dashboard that would be useful to general managers.
 - Moneyball lab details
 - Read/review:
 - *Welcome to Smarter Basketball*
 - Review material in R for beginners, pp. 3-18
-

Session 6: Lab 2A (Moneyball)

- Session Objectives: This session continues the in-class lab work from previous session.
 - Moneyball lab details
 - Read/review:
 - Install R-Studio if it is not already installed
-

Session 7: Mini-review of R with applications I

- Session Objectives: This session provides a quick refresher of basic R commands (for those without an R background, we cover the basics of what we will need for Lab 3A). It covers the R-studio IDE, the use of scripts to organize code, basic mathematical operations, assignment of values, reading in external data, and where to turn for online help.
 - Read/Review:
 - Review material in Chap 1, 2, & 3 of Try R Code School
-

Session 8: Mini-review of R with applications II

- Session Objectives: This session is a continuation of our refresher of the R language. It covers vectors, data frames, functions, merging, and sorting.
 - Read/Review:
 - Moneyball lab
 - Review material in Chap 4 & 5 of Try R CodeSchool
-

Session 9: Lab 3A (Baby names)

- Session Objectives: This session is an analysis of immigration waves and the changing popularity of baby names in the US. Students are asked to produce a number of visual analyses of trends in baby names.
- Begin Baby names lab

Session 10: Lab 3A (Baby names)

- Session Objectives: This session is a continuation of Lab 3A.
 - Continue Baby names lab
-

Session 11: In-class assessment

- Session Objectives: Exam is meant to assess students' mastery of basic applied data science tasks.
 - Guidelines for exam preparation
-

Session 12: Using API's: Facebook, Google, Yelp, and Uber

- Session Objectives: This session provides an introduction to some of the packages offered by leading technology firms that allow developers to access and use their data.
 - Read/review:
 - Review the list of R packages
-

Session 13: Scraping data II

- Session Objectives: Acquire and scrape data through the use of packages like *rvest* that enable the scraping and parsing of web information.
-

Session 14: Visualizing data

- Session Objectives: An overview of some basic data visualization concepts. We work through a variety of exercises involving mapping and chart creation.
- Read/review:
 - Review pp. 36-48 of 'R for Beginners'
 - Final projects are due