Analytics and the Digital Economy

OIDD 245, Spring 2019

Instructor

- Professor Prasanna (Sonny) Tambe

- tambe@wharton.upenn.edu, JMHH 558

Teaching Assistants

- TBD

Office hours

- TBD

Course Objectives

The goal of this 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. We focus on working with large, unstructured data sources and gain experience with introductory machine learning concepts. Students will spend time inside and outside of the classroom combining data and code to develop data products for a number of new industries, including finance, the restaurant industry, and health care.

At the end of the course, students will be expected to complete an advanced data project, which involves acquiring data from an online web property (e.g. Uber, Facebook) through an API and developing an interactive data visualization. Students who complete this course should have the necessary tools to begin building a portfolio of data science projects that they can share online with future employers through platforms such as GitHub.

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 are 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 first half 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 half of the course will extend students' experience to industry applications of text mining and machine learning and require students to work with more unstructured data.

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 are expected to participate in classroom discussions (there is more information about participation below). The breakdown of points is as follows:

Data Labs	15%
Individual Homeworks	20%
Assessment Exam	20%
Data Projects	30%
Professionalism + Participation	15%

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 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

Session	Topic	Date	Due
1	**Introduction to course**	Jan 10	Lab 0: Setup
2	**Getting started with Tableau**	Jan 17	
3	**Lab 1: Citibike**	Jan 22	
4	**Web scraping**	Jan 24	
5	**Data wrangling**	Jan 29	Lab 1: Citibike
6	**Lab 2: Moneyball**	Jan 31	
7	**R review with applications: Part I**	Feb 5	
8	**R review with applications: Part II**	Feb 7	Lab 2: Moneyball
9	**Lab 3: Baby Names + R Markdown**	Feb 12	
10	**Lab 3: Baby Names**	Feb 14	
11	**In class exam**	Feb 19	
12	**Packages and API's**	Feb 21	
13	**Data visualization in R**	Feb 26	
14	**Web scraping (redux) + wrapup**	Feb 28	Data project 1
	(SPRING BREAK)		
15	**Datathon 1: In-class challenge**	Mar 12	
16	**Tidying data (tidyverse)**	Mar 14	
17	**Introduction to machine learning**	Mar 19	
18	**Snow day**	Mar 21	
19	**Datathon 2: In-class challenge**	Mar 26	
20	**Introduction to text mining**	Mar 28	HW 1: Rats!
21	**Lab 4: Yelp reviews**	Apr 2	
22	**Lab 4: Yelp reviews**	Apr 4	
23	**Applications of machine learning**	Apr 9	Lab 4: Yelp Reviews
24	**Lab 5: Peer to peer lending**	Apr 11	Final project proposals
25	**Datathon 3: In-class challenge**	Apr 16	
26	**Coding in the cloud: Github and AWS**	Apr 18	HW 2: News Analytics
27	**Final project presentations**	Apr 23	
28	**Final project presentations + wrapup**	Apr 25	Final projects