OIDD 245 Syllabus, Spring 2022

DRAFT COPY. Subject to change.
Last updated, Oct 2021

- **Professor Tambe:** tambe@wharton.upenn.edu
- **TA team**: TBD
- Slack channel is TBD

Course Objectives

The goal of this course is to provide students with applied experience with the world of data science. 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 tech, finance, accounting, management, marketing, and other industries are increasingly being confronted with a new class of problems — how to manage their data assets and to effectively use them for data-driven decision making. 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, digital marketing analyst, financial analyst, strategy 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.

Throughout the course, students will be expected to complete several data projects, which involve acquiring data from outside sources through an API or webscraping and developing data visualizations. 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. There is an emphasis in the course on data-driven "storytelling", which is an increasingly important skill that combines data analysis skills with domain expertise and narrative skills.

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 useful 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, healthcare, 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 for all materials, including to submit assignments and receive grades. We will also use Slack for some course communications.

Required textbooks and software

There is no required textbook. Occasional readings will consist of selected online content which will be posted on the course site. As part of your homework, you may 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. All software will be open source or free on a trial basis.

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 below about participation). The breakdown of points is as follows:

Deliverable	Weight	Points
Data project 1	10%	50
Data project 2	20%	100
Data project 3	25%	125
Individual Homeworks	20%	100
Professionalism	10%	50
Group Labs	15%	75
TOTAL	**100%**	**500**

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.

In corporate America, you will be expected to present your data-driven findings and to make a recommendation based upon them. 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 (please see below for notes on adjustments made for the remote learning format used in Spring 2021).

Active participation requires good preparation—thoughtful completion of home-work 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, 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.

In particular, because this class emphasizes projects and learning-by-doing, attendance is expected. Missing an excessive amount of class without justification will negatively impact your final grade.

Participation grades will include attendance of sessions, interaction offline and online, contributions to Slack and other channels, either in the form of exploring

new topics or answering the questions of others, and interactions with the Professor and/or TAs.

Please note that the resulting score is not a simple high/low measure. The distribution of participation scores in past semesters has routinely been bell-shaped, with scores above 45 only being awarded to students that have distinguished themselves in terms of participation. Most students receive lower scores.

Adjustments for Spring 2021 (may be outdated, Spring 2022 is expected to be in-person)

Because the course is being taught remotely this semester, it will require several adjustments. The good news is that this class has always been "flipped". Most of the learning takes place in a self-driven and self-paced way. It is very interactive, in the sense that sessions require you to continously try exercises as we move through the material. Nevertheless, one thing we will certainly miss is the ability to sit together and learn from one other.

We will be adapting to our new format and to discover new strengths of the format where they can be found. This includes adjusting evaluation metrics, the balance of lecture and exercises, and a greater emphasis on small-group interaction.

Because we have limited physical proximity, the course will work better if I get to know you early on. Although there is no "requirement", I expect to make myself very available early in the semester to learn about you and your goals. This will help the sessions run more smoothly.

Furthermore, I strongly encourage you to reach out and let me know your experiences, suggestions, and recommendations about tools, technologies, and best practices that you have seen used effectively in other classes. I look forward to learning from you.

Zoom expectations (may be outdated, Spring 2022 is expected to be in-person)

Whether or not to require your cameras to be on is not a straightforward decision. Some of you may not be in situations that make it possible for you to leave your camera on. You may have siblings running around in the background, be sitting in a public space, may have limited Internet speed, and so on. On the other hand, *not* requiring cameras to stay on tends to lower overall levels of class engagement.

One silver lining of the remote-learning experience is that Slack and Zoom chat have made it *easier* for some types of students to engage and ask questions. I am optimistic about our collective ability to create a running dialogue as we move through the material. The choice of whether this discussion should take place on Slack or Zoom or on another platform will be made by trial-and-error. However, this class has always been very interactive—students recognize the importance of this area of study and have lots of questions—and I expect that this year, we can match or exceed those levels of interaction.

In terms of asking questions during lecture, I am not averse to Zoom interruptions, but expect that it may be most efficient to use the "raise-hand" feature of Zoom or simply put the question into the chat. We will experiment and make adjustments as necessary.

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.

Points will be deducted from late assignments, labs, or projects at the rate of a 20% penalty for each day the submission is late.

Accounts for Non-Wharton students

If you are a non-Wharton student and need to access public computers, you will need to create a Wharton Class Account. (Once created a class account will then link to their PennKey account and allow you to log into public computers with their PennKey accounts.)

For class account creation, please see: https://apps.wharton.upenn.edu/iam/accountcreator/.