OIDD 1050: Developing Tools for Data Access and Analysis (Spring 2023)

(0.5 cu, Q3 Only)

Prof. Lorin M. Hitt (lhitt@wharton.upenn.edu)

(Updated: 9 November 2022)

This course introduces the construction and use of data analysis tools that are commonly used for business analysis. The course builds on the spreadsheet and analytical skills developed in OIDD1010, providing a much more extensive treatment of spreadsheet application development (using *Visual Basic for Applications*). In addition, we will cover best practices in programming and analytics generally which can carry over to other tools and languages. Time permitting, we will do an introduction to some advanced analytical methods that show up in complex data analysis tasks and provide a foundation for further study.

The course is intended for students without prior experience in programming, but students must have familiarity with computer-based tools as covered in OIDD1010 or equivalent, or through personal experience. The course is definitely *introductory* in that it does not require prior knowledge of the material. That does not mean it will be *easy* since computers can be unforgiving when you make a programming mistake and understanding some concepts, like object orientation, takes some time. We expect the course to be especially useful for students seeking entry-level analyst positions in data-intensive firms, or those generally seeking to broaden their knowledge and skills in the construction and use of computer-based analytical tools.

Note: Prior to 2022, OIDD105 was taught as a 1 cu course that combined the current material from OIDD105, along with the use of SQL and Databases. Students wishing to get the same "105" experience as in the past, and a full credit, can pair this course with OIDD3150 (either sequentially or concurrently).

Course Format:

Class: Class time will be a mix of introducing and discussing the material and in-class exercises where it is easier to learn "hands on". Some sessions will be moved to the computer labs for that purpose.

Workbook: I have written a workbook that contains a number of short exercises. I expect you to complete these prior to class as they are assigned. Some must be submitted but will be graded lightly.

Problem Sets: There will be three graded Problem Sets of which you must complete. These will have fixed due dates and will be graded seriously.

Quizzes: There are two quizzes corresponding, a shorter "checkpoint" quiz in the middle of the term, and a longer final exam (cumulative).

In the past there was a course project. Given the compressed nature of the class schedule, I don't believe that is feasible at scale, but if you have something you specifically would like to work on we can discuss a modified workload that permits a class project.

Course Materials. There is one highly recommended but optional text:

(PPVBA) Alexander and Kusleika (2016). *Excel 2016 Power Programming with VBA*. (ISBN: 1119067723)

This is a trade book and available from a wide variety of sources, including Amazon and other discounters, in both paper and digital form. While not expensive, it is also not free, so please do not use illegal copies. I will cover everything relevant from the book in class, but sometimes it is useful to have a different perspective. Note that there are multiple versions of this book (2010, 2013, 2016 and 2019) – any will do, although I happen to like this version and it will match page numbers in my reading guide.

There is also a workbook that will provide a number of guided exercises. I will assign some portion of the workbook over the course of the semester and you can use other exercises in the workbook for practice.

Mandatory Computer Resources

While you can use the labs, you will find yourself at a disadvantage in the course if you rely strictly on them for computing resources (disadvantage = measurably lower grade!). In general, if you have a relatively modern computer (preferably laptop) you will be fine.

We will be using:

Office 365 Excel. Available for free through your Penn O365 account or through the Wharton virtual labs. The class will primary use the PC version. The native Mac version of Excel is about 90% compatible. As such, if you plan to use your Mac, you will also need to occasionally use the PC version (the virtual labs are a good solution here).

(optional but recommended). If you are using a laptop, get an external mouse. This will increase your programming productivity significantly (best $5 you will ever spend!). I also highly recommended that programmers use large screens (24" or better). Studies have shown this increases developer productivity.

…But can I use the (physical) labs?

Yes…but… if you rely entirely on the labs you will be spending lots of time in the labs. You will be happier if you have your own machine preferably a laptop.

Grading and Evaluation.

Problem Sets (25% of final grade). There are three problem sets, one each on basic Excel analytics, functions and subroutines.

Pre-Class Preparation (10%). Most class sessions will have some type of preparatory work. I will request submission of some of these prior to class and they will be graded lightly. One of these is mandatory, and you can skip one of the remaining without penalty.

Quizzes (55%). There is a small checkpoint quiz (15%) and a larger exam at the end (40%).

Class Participation (10%). Students are expected to prepare, attend class, actively participate, and make good use of course resources (including the support staff and the instructors out of class time). The class participation grade will reflect our subjective evaluation on these dimensions as well as objective observation of class attendance.

Grade Distribution. There is no pre-specified grade distribution. Historically, we gave approximately 45% A's and 55% B's. Most of the variance in grades is driven by quiz scores (homework scores tend to have modest variation other than missed/late assignments). Grades lower than a "B-" are unlikely if you complete all the assigned work and otherwise follow course guidelines.

Other Course Policies

Regrades. Any requests for regrades should be submitted in writing to your assignment submission folder before the next assignment is due. The request must be labeled clearly and explain why you believe your answer is correct. Please note that we do not consider regrade requests regarding partial credit awarded to incorrect answers (in other words, if your answer is not correct, it is not eligible for regrade consideration).

Deadlines. Assignment deadlines are firm because we often review the assignments in class immediately following the deadline. If for some reason you are not able to complete an

assignment (e.g., you can't get your program to work…) submit what you have by the deadline. If you have a conflict on a deadline date, submit early.

Collaboration. You are free to discuss any and all course material with your fellow students and the course staff, including approaches to the assignments. You can also work together on most assignments in small groups. However, you are not allowed to share code or answers on any graded assignments outside your small work team or copy code for the assignments graded for "being there". You are also not permitted to use materials from prior iterations of OPIM105 or OIDD105 in preparing your written work or to copy code directly from Internet sources (FYI: this is easy to spot, so just don't). All collaborators or should be identified by name in the submitted documents (distinguishing between your work team and anyone you spoke with in preparation of the assignment). If you worked in a group for the assignments, you should submit a common paper for the assignment. Doing an assignment in a group and then creating a private version of the group work violates the "no sharing of code" guideline and is not allowed. You are not required to work in a group. If you don't have a group and would like one, I can facilitate group formation (e-mail me).

Regardless…I strongly discourage "divide and conquer" strategies on assignments where questions are divided among group members or "you drive, I watch" programming where one student writes all the code and the other watches, gets coffee, etc. You cannot learn these skills without actual personal experience. Programmers write code, and you can't write and test code without touching the computer.

Attendance. You are expected to come to class and to be prepared. From time to time, something may happen in class that requires your physical or at least virtual presence. I will also, from time to time, take attendance. You are permitted to miss one of these over the course of the semester before it affects your grade (this is <u>in addition to</u> any University-approved absences such as religious observances). You do not need to tell me why you are missing class or get permission. If you need to miss class due to a religious holiday, I am happy to go over the material by appointment or during office hours or to record a session of the class by request.

Support. There will be office hours by both the instructors as well as undergraduate and graduate teaching assistants. We will be using Piazza, and online discussion tool, for online course questions. A few guidelines about the use of Piazza which will make everyone happier:

- If you have a general question or something about the course material, use Piazza. If you have a personal question, e-mail the instructor.

- Please do post code to Piazza as an open message. If you need a quick evaluation of your code, post it is a private message to instructors. If you have a more complicated question ("why doesn't this work?") that is probably best done in person or by e-mail.
- Please do not make all your questions private. It defeats the purpose of an open discussion forum (the exception is when you need to post code).
- Please do not spam questions on Piazza. If you have lots of questions, come see me or someone on the course staff.
- You can make your questions anonymous to other students but the instructors and TAs can see your real name… so be nice.
- You too can answer questions on Piazza. This is appreciated by the course staff.

Electronics. In some (perhaps most) classes, it is helpful to bring your laptop and you can work in parallel. You may also use your computer to take notes. But…you are not permitted to do ***anything*** except class work on your computer and you are not permitted to make audio or video recordings of class sessions under any circumstances. If you need audio or video of class for some reason, I will arrange it with the school. Cell phones should be turned off or silenced. If you must take a call or respond to a message, please leave the room. Penalties for violating the electronics policy are unpleasant (one of the few ways to earn a failing grade) because the use of electronics for non-class work is highly damaging to the learning environment for other students.

Preliminary Schedule. The actual schedule is on Canvas and will evolve as we go.

Session	Date	Day	Session	Assignments	Location
1	1/11/2023	Wed	Course Introduction/Technology Setup		Zoom
	1/16/2023	Mon	No class		
2	1/18/2023	Wed	Excel Review/Introduction to VBA		Classroom
3	1/23/2023	Mon	Programming and Functions I	Mandatory SA: Hello Excel	Classroom
4	1/25/2023	Wed	Programming and Functions II	PS1: Excel Analytics	Classroom
5	1/30/2023	Mon	Algorithms and Complexity		Classroom
6	2/1/2023	Wed	Subroutines I	PS2: Function Programming	Classroom
7	2/6/2023	Mon	Subroutines and Errors		Classroom
8	2/8/2023	Wed	Excel Obects	VBA Mini-Quiz	Classroom
9	2/13/2023	Mon	User Interfaces		Classroom
10	2/15/2023	Wed	User Interface Lab		Lab
11	2/20/2023	Mon	Advanced Topics: RegEx	PS3 (Friday): Subroutine	Classroom
12	2/22/2023	Wed	Advanced Topics: VBA and the Web		Classroom
13	2/27/2023	Mon	Advanced Topics: Integration		Classroom
14	3/1/2023	Wed	Revew/Course Summary		Classroom