

OIDD 105: Developing Tools for Data Access and Analysis (Fall 2019)

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

(Updated: late August/Near Final)

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 OIDD101, providing a much more extensive treatment of spreadsheet application development (using *Visual Basic for Applications*) and database management (using SQL as implemented in the SQL Server database system – *Transact-SQL*). In addition, we will cover best practices in programming and analytics generally which can carry over to other tools and languages. In the final module, 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 OIDD101 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 some concepts, like object orientation and set-based reasoning, are intellectually challenging. 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. The course counts toward the general OIDD concentration and the Information Systems and Business Analytics tracks.

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. They must be submitted but will be graded lightly.

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

Quizzes: There are two quizzes corresponding to each major section of the class (Quiz I – VBA and Quiz II – SQL) plus a short “checkpoint” quiz in the middle of the VBA section so you can evaluate where you are.

Project: There will be a small final project where you are asked to either a) Build something interesting, or b) Analyze some interesting data.

Course Materials. There are two highly recommended texts:

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

(SQLD) Syyverson and Murach (2016). *Murach's SQL Server 2016 for Developers* (ISBN: 1890774960)

These are trade books and available from a wide variety of sources, including Amazon and other discounters, in both paper and digital form. Neither of these books are free (they are also not very expensive) so please do not use illegal copies.

What does *Highly Recommended* mean? These books are certainly helpful (I use them). However, there is no material in the texts that is not also covered in class. The advantage of getting the books is that it provides another reference and you get a chance to see the material we will be covering in more detail ahead of time.

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" and "Access"). 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).

Amazon Web Services – SQL Server. You will be able to create an account and have your own private database server in the cloud.

Microsoft Azure Data Studio (new for 2019!). This is a relatively new data management tool that, among other things, can connect to SQL Server.

We may be using additional software for the analytics sessions (TBD).

(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.

... but what about Macs?

In the past using Macs was a problem. This is no longer the case – Excel works OK on the Mac (with a few exceptions where you can use the Virtual Labs or the actual labs) and the database client is multiplatform.

Grading and Evaluation.

Problem Sets (20% of final grade). There are four problem set opportunities of which you must complete three (I will take the highest three scores).

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. Two of these are mandatory, and you can skip two of the remaining without penalty.

Class Project (10%). There is a class project which is either a) build something interesting, or b) do an interesting data analysis. There are two deliverables: a proposal due in early November and the final product which includes a 5-page writeup of what you did and supporting code/data/analysis files. Ideal group size is 3 but you are not required to work in a group. I will consider larger groups on an individual basis for larger/more ambitious projects. It is “due” at the end of class, but as long as you get it to me before break there is no late penalty.

Quizzes (50%). There are three in-class quizzes, one on VBA and one on SQL (20% each, non-cumulative). There is also a mini-quiz in the middle of the VBA section so you can get an early assessment of where you stand (this is worth 10%) There is no final exam and we will NOT be using the scheduled final exam time (so go ahead and buy those plane tickets...).

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 40% A's and 60% 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, skip the assignment or submit it early. Note that you are permitted to skip at least one Problem Set and at least two of the class preparatory activities.

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 really easy to spot). 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 presence. I will also, from time to time, take attendance. You are permitted to miss two of these over the course of the semester before it affects your grade (this is in addition to 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 not post code to Piazza as an open message. If you need a quick evaluation of your code, post it as 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. Unless you are taking notes electronically or we are specifically using laptops for a class exercise, computers and tablets are not permitted in class. If you do wish to use your computer to take notes, please come see me for permission. 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 no electronics policy are unpleasant (including a failing grade) because the use of electronics for non-class work is highly damaging to the learning environment for other students.

Preliminary Schedule (Do not rely on this schedule for material. I expect to adhere to the two quiz dates but everything else is subject to change. The actual schedule is on Canvas and will evolve as we go).

Date	Day	Session	Assignments
8/28/2019	Wed	Course Introduction	
9/2/2019	Mon	No Class (labor day)	
9/4/2019	Wed	Excel Review/Introduction to VBA	
9/9/2019	Mon	Programming and Functions (I)	Mandatory SA: Hello Excel!
9/11/2019	Wed	Programming and Functions (II)	PS1: Remembering Excel
9/16/2019	Mon	Algorithms and Complexity	
9/18/2019	Wed	Subroutines (I)	PS2: Function
9/23/2019	Mon	Subroutines and Error Handling	
9/25/2019	Wed	Excel Objects	VBA Mini-Quiz
9/30/2019	Mon	User Interfaces	
10/2/2019	Wed	User Interface Lab*	PS3: Subroutine (Friday)
10/7/2019	Mon	Advanced Topics: RegEx	
10/9/2019	Wed	Optional Session: VBA and the Web	Mandatory SA: Set up AWS (Sunday)
10/14/2019	Mon	SQL: Single Table Queries	
10/16/2019	Wed	Review Session	
10/21/2019	Mon	In-Class Exam	Quiz I
10/23/2019	Wed	SQL: Advanced Queries	
10/28/2019	Mon	SQL: Relational Database Concepts	
10/30/2019	Wed	SQL: Relational Joins	
11/4/2019	Mon	No Class - Day off	Project Proposal (Due EOD)
11/6/2019	Wed	SQL: Complex Joins/Subqueries	
11/11/2019	Mon	SQL:DDL and Scripting	
11/13/2019	Wed	Finish SQL/PS4 Discussion	PS4: Analysis with SQL (due Friday)
11/18/2019	Mon	Adv Topics: VBA-SQL Integration	
11/20/2019	Wed	In-Class Exam	Quiz II
11/25/2019	Mon	Analytics: Network Analysis	
11/27/2019	Wed	No class (schedule shift)	
12/2/2019	Mon	Analytics: Big Data	
12/4/2019	Wed	Analytics: Introduction to Data Science	
12/9/2019	Mon	Course Conclusion	

* - This session has a short assignment (SA) due at the end of the day (you can do it on your own)