OIDD 105: Developing Tools for Data Access and Analysis (Spring 2020)
Lorin M. Hitt (lhitt@wharton.upenn.edu)

(Updated: mid-November)

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 significant quizzes - Quiz I for VBA and Quiz II for SQL. The VBA quiz will be during class time. Quiz II will be during final exam time.

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. It works on both PCs and Macs.

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.

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.

Quizzes (60%). There are two major quizzes which each represent 25% of the final grade. There will be also 1 – 2 mini-quizzes that represent progress checks. These will be 5% each.

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 very 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 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. This class uses computers for everything. Computers are welcome in the class… but… you must commit to not using your computers for anything other than notetaking or class work. Penalties for violation of this guideline will be severe (including making it impossible to get credit for "competing all the work" necessary to earn the guaranteed minimum grade).

Regardless, 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.

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
1/15/2020	Wed	Course Introduction	
1/20/2020	Mon	No Class (MLK Day)	
1/22/2020	Wed	Excel Review/Introduction to VBA	
1/27/2020	Mon	Programming and Functions (I)	Mandatory SA: Hello Excel!
1/29/2020	Wed	Programming and Functions (II)	
2/3/2020	Mon	Algorithms and Complexity	
2/5/2020	Wed	Subroutines (I)	PS1: Function
2/10/2020	Mon	Subroutines and Error Handling	
2/12/2020	Wed	Excel Objects	
2/17/2020	Mon	User Interfaces	
2/19/2020	Wed	User Interface Lab*	
2/24/2020	Mon	Advanced Topics: RegEx	PS2: Subroutine
2/26/2020	Wed	Optional Session: VBA and the Web	
3/2/2020	Mon	Review Session	
3/4/2020	Wed	Quiz I: VBA	Quiz I
3/9/2020	Mon	Spring Break	
3/11/2020	Wed	Spring Break	
3/16/2020	Mon	SQL: Intro SQL	
3/18/2020	Wed	SQL: Computational Queries	
3/23/2020	Mon	Day Off - Preparing for Projects	
3/25/2020	Wed	Day Off - SQL Mini-Quiz	
3/30/2020	Mon	SQL: Relational Database Concepts	
4/1/2020	Wed	SQL: Relational Joins	
4/6/2020	Mon	SQL: Complex Joins/Subqueries	
4/8/2020	Wed	SQL:DDL	Quiz II
4/13/2020	Mon	SQL:Scripting	PS3: Database
4/15/2020	Wed	SQL: Advanced Scripting/Integration	
4/20/2020	Mon	Analytics: Network Analysis	
4/22/2020	Wed	Analytics: Big Data	
4/27/2020	Mon	Analytics: Introduction to Data Science	Optional PS4: Scripts/Integration
	TBD	Review Session (location TBD)	
	TBD	Quiz II (During Final Exam Slot)	