OIDD 105: Developing Tools for Data Access and Analysis (Fall 2018)

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

(Updated: August/preliminary)

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[1], 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. There will be a sign-in sheet for class.

[1] This course is not intended for students with an extensive computer science background. Students in this category are strongly encouraged to take an alternative class such as OIDD311 (Business Computer Languages) instead of OIDD105. Please contact me immediately if you think you fall into this category and we can figure out the appropriate course of action.

Workbook: I have written a workbook that contains a number of short exercises. I expect you to complete these prior to or shortly after class. They must be submitted but will be graded for "being there". You can miss two of these without penalty.

Exercises: There will be four graded exercises (one minor exercise required, and best two of the remaining 3). These will have fixed due dates and will be graded seriously.

Quizzes: There are two required quizzes at about the 1/3 and 2/3 time in the class (Quiz I – VBA and Quiz II – SQL).

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. Note that the digital version of Murach uses a non-standard copy protection scheme ("Locklizard") and is therefore not recommended. 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!). If you have a PC with Windows 7 or better you are in good shape. If you have a relatively new Mac, you can make it work with a little effort.

We will be using (*Updated*):

Office 365 for Windows ("Excel" and "Access"). Available for free through your Penn O365 account or through the Wharton virtual labs. The native Mac version is not acceptable and cannot be used because it does not have a full implementation of VBA.

Amazon Web Services – SQL Server. You will be able to create an account and have your own private database server in the cloud. You can also use this server to run SSMS (see below) if you have a Mac.

Microsoft SQL Server Management Studio ("SSMS") (free, preferred, Windows-only). Windows users can install it directly. Mac users can use it in the labs, on their AWS server instance, or on a remote desktop for a dedicated server for this class. There are also native SQL clients you can use if these are not satisfying (but they are not free - I recommend *Navicat Essentials for SQL Server,* available directly from Navicat for $99).

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… Excel and Access run just fine in both the physical and virtual labs. We will also have SSMS installed in at least one lab. However, if you rely entirely on the labs you will be spending lots of time in the labs. Since I have found that this can be a problem, the first exercise will involve setting up access to AWS for SQL Server and Excel (either your own machine or the virtual lab).

… but what about Macs?

In the past using Macs was a problem. This is no longer the case – you now have plenty of options including the virtual labs and AWS. There may be a little setup time and it might not be free, but these solutions are good. We will be providing extra support in the first week to make sure this goes smoothly. Regardless, make sure you get these issues settled early on or you will be unhappy.

Grading and Evaluation.

Graded Exercises (20% of final grade). There is one required activity (Exercise 1) and four additional Exercise opportunities (I will take the best 2 of 3, so you can skip one or do them all to

try to get a better score). Exercises may be done in groups of no more than 3 (except Exercise 1 which is individual).

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 (full credit for being there on time). You will be able to skip at least two for full credit.

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 just after mid semester the final product which includes a 5-page writeup of what you did and supporting code/data/analysis files. Ideal group size is 3. 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 two in-class quizzes, equally weighted and non-cumulative. The first is on VBA and the second is on Database. 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.

<u>Other Course Policies</u>

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 exercise 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…We 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, 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. Unless you are taking notes electronically, 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. If you need to do so for some reason, I will arrange it with the school. Cell phones should be turned off or put in airplane mode. If you must take a call or respond to a message, please leave the room. Anyone using an electronic device in class without permission will not receive credit for the class (no warnings). I have tried many permutations of this policy and this is the only one that seems to avoid the inevitable distraction created by electronics in the classroom.

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
8/29/2018	Wed	Course Introduction	
9/3/2018	Mon	No Class (labor day)	
9/5/2018	Wed	Excel Review/Introduction to VBA	
9/10/2018	Mon	Programming and Functions (I)	Ex 1: Get the Tools
9/12/2018	Wed	Programming and Functions (II)	
9/17/2018	Mon	Algorithms and Complexity	
9/19/2018	Wed	Subroutines (I)	Ex 2: Function
9/24/2018	Mon	Subroutines and Error Handling	
9/26/2018	Wed	Excel Objects	
10/1/2018	Mon	User Interfaces	
10/3/2018	Wed	User Interface Lab*	Ex 3: Subroutine (Friday)
10/8/2018	Mon	Review Session	
10/10/2018	Wed	Quiz I (in class)	Quiz I
10/15/2018	Mon	SQL: Single Table Queries	
10/17/2018	Wed	SQL: Advanced Queries	
10/22/2018	Mon	SQL: Relational Database Concepts	Project Proposal (Due Friday)
10/24/2018	Wed	SQL: Relational Joins	
10/29/2018	Mon	SQL: Complex Joins/Subqueries	
10/31/2018	Wed	SQL:DDL and Scripting	Ex 4: Analysis with SQL (due Friday)
11/5/2018	Mon	Class Canceled (Review Session in Evening)	
11/7/2018	Wed	Quiz II (in class)	Quiz II
11/12/2018	Mon	Adv Topics: Regular Expressions	
11/14/2018	Wed	Adv Topics: XML and Web Services	
11/19/2018	Mon	Adv Topics: VBA-SQL Integration	
11/21/2018	Wed	No class (schedule shift)	
11/26/2018	Mon	Adv Topics: Time and Space	
11/28/2018	Wed	Analytics: Network Analysis	
12/3/2018	Mon	Analytics: Big Data	
12/5/2018	Wed	Analytics: Introduction to Data Science	
12/10/2018	Mon	Course Conclusion	Exercise 5: Analytics/Projects "due"
* - This session has an assignment due at the end of the day (you can do it on your own)			