Department of Decision and Information Science

BZAN 6355 – Advanced Programming Big Data Analytics
Cross-listed with
BZAN 4397 – Selected Topics in Business Analytics: Adv Progrm Bus Anlyt

Spring 2021
Online Synchronous Hybrid [1] Meetings
Important: Class must NOT meet every week. In weeks that it does: Wed 6-9

Meetings on MS Teams*
Team name for this course: BZAN4397-23604-2021-SP.C-1817

Joining the online class meetings in MS Teams does not require a password. You automatically have permission by enrollment to this course; dropping this course will automatically remove your permission to join meetings for this course.

Familiarize yourself with the MS Teams interface. Join scheduled class meetings from inside the Team, find the meeting item, and join. Set up your devices ready before meeting (required: Internet access, camera, and mic)

All Teams meetings will be recorded and uploaded to MS Stream.

MS Teams technical difficulties are not an acceptable excuse for missing synchronous meetings!

Professor
Dr. Xiao Ma ("Xiao" is pronounced as /sh-ih-ow/) Email: xma@bauer.uh.edu
Office: MH #280H Office phone: 713-743-4725 (not in effect)
Office hours: MS Teams meeting (By Appointment only; send request via email)
** This course is a core course of the Track: Data Management, and an elective of the Track: Analytics.*

TA:
TBD Email:
Office & Office hours:

****
The information contained in this course syllabus is subject to change anytime. Students are expected to be aware of any additional course policies presented by the Professor during the course.
****

[1] "Hybrid"–sometimes called "blended" or "flipped"–course is such that some traditional face-to-face (F2F) "seat time" has been replaced by online learning activities. Even with an online synchronous delivery, a hybrid course is feasible. Note that: (1) hybrid course is not online asynchronous course; you **must NOT skip the "synchronous" meetings** for which the timings are critical and thereby adapted to your learning process. (2) hybrid course does **not** have a fixed weekly schedule like a traditional course; that means, **class must NOT meet every week**.

Prerequisites

For Master's students enrolled in BZAN 6355:

- BZAN 6351 Basic Programming for Bus Anlyt (grade B or above)
- BZAN 6352 (or 6310) Quantitative Foundations for Bus Anlyt (grade B or above)

For undergraduate students enrolled in BZAN 4397:

- Senior standing
- BZAN 4397 Basic Programming for Bus Anlyt (grade B- or above)
- STAT 3331 Statistical Analysis for Business Applications I (grade B or above)
- MIS 3376 DBM-1 (grade C or above)

Computer Equipment

Use your own machine (laptop, notepad, surface, remote desktop) as your best computing resource.

Course Objective

The primary course goals are for students to be able to do the following upon completion of the course:

- Explain and differentiate the meaning of business intelligence, machine learning, predictive analytics
- Use programming to transfer, describe, manipulate, transform, wrangle with, and visualize data
- Understand the principles and several classic models of predictive analytics and data mining
- Use programming to build classic predictive analytics models and achieve decent performance results
- Be able to run complete projects; analyze, evaluate, and interpret results; and make management recommendations based on the results

Textbooks

***Required*:**

1. Wes McKinney, ***Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2nd ed.)***, O'Reilly, 2017. ISBN-13: 978-1491957660. (referred to as McKinney below)
2. Foster Provost and Tom Fawcett, ***Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking***, O'Reilly, 2013. ISBN-13: 978-1449361327. (referred to as Provost below)

Recommended Reading

Daniel T. Larose and Chantal D. Larose, ***Discovering Knowledge in Data: An Introduction to Data Mining (2nd ed.)***, Wiley, 2014. ISBN-13: 978-0-470-90874-7 (hardback). (referred to as Larose below)

Course Description

Immersed in the most popular open-source programming language – Python – this course will solidify the fundamental programming concepts and introduce advanced programming techniques of interactive visualization and predictive analytics. There will be plenty of opportunities for students to practice cutting-edge Python programming skills for diagnosing and manipulating raw data. This heavily hands-on course will emphasize confidence with the Python language grammar, programming flow, and mastery of a wide range of useful data operations. Beyond, students will learn to understand, apply, interpret the results of, and criticize the capabilities of several widely used predictive models on real-life datasets. It will also cover skills of online data collection, such as building web crawlers and extracting numeric and textual data. The essential focus of the course is building problem-solving analytics projects involving such skills as data management, data transformation, interactive visualization, predictive modeling, and critical thinking.

Teaching Methods

1. Modularized learning, hybrid-form of meetings: For each module, students will read materials, practice programming skills with tutorial(s) and sample problems, and learn to expand their knowledge and skills of data analytics, by following guidelines and tutorials available on Blackboard on their own.

Students will also attend class meetings scheduled irregularly throughout the semester where critical course-, concept-, and programming-related contents are covered or reviewed live. Another integral purpose of class meetings is *hack-shop*, sometimes individual-based, others in-group. The dates for class meetings is clearly marked in the *Tentative Course Outline* on the last pages of this syllabus (for MS Teams, meeting items will be scheduled ahead of each); note that the schedule is subject to change as the course progresses, and notification of any change will be posted on Blackboard.

Complementary to the recorded meetings to be made available throughout the semester, technical recordings from previous semester(s) hosted on YouTube are also available via Blackboard.

2. **"Flipped classroom"**: This course will operate in a "flipped classroom" style – students digest module contents, read textbooks and materials, and practice hands-on exercises before attending meetings. In a "flipped classroom," students lead the weekly learning cycle. Professor provides deeper explanations and lead focused (group) discussions on challenging concepts. Professor demonstrates complex hands-on skills and provides additional explanations and tips.

3. Expectations: Because students lead the weekly learning cycle, and meetings will only focus on the contents that need more explanation and discussion, students are expected to be well prepared for each meeting. They are expected to be very familiar with module contents, read textbooks and materials, and practice hands-on exercises before attending each meeting.

4. Q&A collaboration: Make use of MS Teams to raise and collaborate on questions throughout the semester with your Professor and classmates. Students are encouraged to post questions related to the course, meetings, and assignments on MS Teams, under Class Notebook tab, "_Collaboration Space" thread, "Frequent Q and A" section, "Frequent Q&A …" page. When a different page for a separate topic is desired, feel free to create a new one, and name it properly.

5. **In-class hack-shop**: The professor will turn some of the class meetings into in-class hack-shop. A *hack-shop* is a session of fun and intensive work: for each hack-shop, a problem is defined, context introduced, some dataset(s) made available, and the students will work to solve the problem, independently or collaboratively as required by the Professor accordingly. You will be asked to turn your work in on Blackboard to showcase your product at the end of the hack-shop. Both your participation and the quality of your product are graded. **In-class hack-shops make up a significant portion of the final grade!**

6. Assignments: Problems and readings might be periodically assigned to help support and supplement the textbook. Assignments must be turned in on time to receive the maximum points possible. Late submissions are not tolerable.

7. Exams: Exams will be open book/note and will test required readings, tutorial practices, and programming skills. Review sheets will be provided in advance to the exam day. The final exam will not be cumulative, but may require a good understanding of some fundamental concepts covered in previous exams. All relevant concepts and points will be noted on the review sheets.

8. Project Description: The course project will give you an opportunity to apply most of the techniques you learned in this course. I will make available to you a set of data on which your project is based. The data will be what we consider "big." You will be required to get an understanding of the variables in your data set. You will use the software applications for the course to perform various analytics techniques on your data. In your project, you will attempt to show how well you can apply the techniques, define multiple problems, and solve them using appropriate techniques. You will need to

interpret all your results. And, provide managerial recommendations to help a stakeholder make good business decisions, which you will also define.

9. Blackboard: Most materials will be distributed on Blackboard. It is assumed that students know how to access the content on Blackboard.

10. Announcements: Announcements regarding the course such as schedule changes, homework, projects, and so on will be made in class meetings during the first 10 minutes, as well as on Blackboard. Students are responsible for being in class meetings on time to hear the announcements and for regularly checking the course website.

11. Contacting the Professor and Teaching Assistant (TA): You can reach me by email or telephone. Email will be the most productive channel of communication to resolve the vast majority of inquiries, as such, it is recommended that students send their questions via email. If a TA is available to this course (dictated on the cover page), when emailing your question and replying any follow-up message, you must include both your Professor and your TA as the recipients, using the "Reply All" function of the email service, in order to make us maximally informed about the recent development of the resolution process. If you only included the Professor or the TA as the only recipient of your email, it is not guaranteed your inquiry will be answered, and you will be responsible for any negative consequence (e.g. late submission penalty for being stuck on a problem needing the Professor's and/or the TA's help).

    The best email address to reach the Professor and/or the TA are included on the cover page of syllabus.

12. Grading:
    1. In-class hack-shops – 22%
    2. Homework – 18%
    3. Exam #1 – 15%
    4. Exam #2 – 15%
    5. Project – 30%

    Final course letter grade follows the numeric-letter grade system shown in the table below.

Raw Score	Letter Grade
> or = 92	A
> or = 89, but < 92	A-
> or = 86, but < 89	B+
> or = 83, but < 86	B
> or = 80, but < 83	B-
> or = 77, but < 80	C+
> or = 74, but < 77	C
> or = 70, but < 74	C-
> or = 67, but < 70	D+
> or = 63, but < 67	D
> or = 60, but < 63	D-
< 60	F

Course Policies
Late submissions will NOT be accepted! Cheating or plagiarism will result in severe academic sanction!

Missed Meetings: The student is responsible for obtaining material, which may have been distributed in class when he/she was absent. This can be done through contacting a classmate who was present or by contacting the Professor during his office hours or other times. Missed or late exams cannot be made up under any circumstances, unless an official excuse is provided. **Any uncoordinated, unexcused missed exam will result in a score of 0 for that exam.**

Academic Dishonesty: Plagiarism and cheating are serious offenses and may be punished by failure on exam, paper or project; failure in course; and or expulsion from the University. For more information, refer to the "Academic Honesty Policy" accessible here (http://www.uh.edu/provost/policies/honesty/). The University of Houston Academic Honesty Policy is strictly enforced by the C. T. Bauer College of Business. No violations of this policy will be tolerated in this course. Students are expected to be familiar with this policy.

Need for Assistance: If you have any condition, such as a physical or learning disability, which will make it difficult for you to carry out the work as outlined in this document, or which will require academic accommodations, please notify me as soon as possible. I will recommend that you contact the Center for Students with Disabilities. The contact person is Justin Dart in the CSD building #568, room 110. The numbers for the CSD office are Ph: 713-743-5400; TDD: 713-749-1527; Fax: 713-743-5396 or email: uhcsd@central.uh.edu. Also available to you is *Counseling and Psychological Services (CAPS),* which can help students who are having difficulties managing stress, adjusting to college, or feeling sad and hopeless. You can reach CAPS (www.uh.edu/caps) by calling 713-743-5454 during and after business hours for routine appointments or if you or someone you know is in crisis. In addition, there is no appointment necessary for the "Let's Talk" program, which is a drop-in consultation service at convenient locations and hours around campus. http://www.uh.edu/caps/outreach/lets_talk.html.

Inclement Weather or Technical Problems
In case of inclement weather or technological problems that prevent the University from providing access to course materials, the Professor will notify students as soon as possible and provide instructions on how the course will proceed. You may send the Professor an email inquiry.

Main Modules

Module 1.	Python programming environment setup (4 technical tutorial documents)		P
	The data mining process, Intro to data mining & business analytics	T	
Module 2.	[2.0] Launch .ipynb file as slideshow		P
	[2.1] Python language basics		P
	[2.2] Python built-in data structures and custom functions		P
Module 3.	[3.1] Fundamental packages of data analytics I: NumPy		P
	[3.2] Fundamental packages of data analytics II: Pandas		P
Module 4.	[4.1] Data preparation and cleansing		P
	[4.2] Data wrangling, join, append, and merge		P
Module 5.	[5] Visualization and Exploratory Data Analysis (EDA)		P
Module 6.	[6] Hypothesis testing and regression model	T	P
Module 7.	[7.1] Foundation of predictive modeling: classification and overfitting	T	P
	[7.2] Algorithms of predictive modeling I: logistic regression	T	P
Module 8.	[8] Algorithms of predictive modeling II: *k*-Nearest-Neighbor	T	P
Module 9.	[9] Predictive model evaluation & implementation	T	P
Module 10.	[10] Scoring new dataset: steps & caveats		P

* Notes about symbols: P indicates "practical" hands-on topic/material; T indicates "theoretical" conceptual topic/material.

Tentative Course Outline. This outline is tentative and subject to change anytime in the event of unexpected class disruptions.

(Week #) Date	Module (see module topics in section Main Modules)	Required Textbook Readings	Due
(1) 1/20/2021	Class Meeting Study Module 1.1~1.4; Conceptual Module 1	McKinney 1.2~1.4 Provost 2 ("The Data Mining Process")	Hack-shop 1 due end of meeting
(2) 1/27	Study Module 2.0~2.2;	McKinney 2.3; 3.1~3.2	Homework 1 due Saturday midnight
(3) 2/3	Class Meeting Study Module 3.1	McKinney 4.1~4.3; 4.6	Hack-shop 2 due end of meeting
(4) 2/10	Study Module 3.2	McKinney 5	
(5) 2/17	Class Meeting Study Module 4.1~4.2	McKinney 6.1; 7; 8.2~8.3	Hack-shop 3 due end of meeting Homework 2 due Saturday midnight
(6) 2/24	Study Module 5	McKinney 9.1	
(7) 3/3	Class Meeting Review exam #1 Study Conceptual Module 6; Module 6	Provost 3 (pp. 43-48), 4 (Regression pp. 81-90, 95-97) McKinney 13.1, 13.3	Homework 3 due Saturday midnight
(8) 3/10	**Exam #1** covers Modules 1~5 Finish study Module 6	Provost 3 (pp. 43-48), 4 (pp. 81-90, 95-97) McKinney 13.1, 13.3	
3/15 – 3/20	Enjoy Spring Holiday!		
(9) 3/24	Class Meeting Review Exam #1 results & Module 6 Study Conceptual Module 7; Module 7.1	Provost 3 (pp. 48-62); 5 (Overfitting pp. 113-115, [Figure 5-1 particularly useful]; 118-123; Cross-validation pp. 126-130); McKinney 13.4	
(10) 3/31	Conceptual Module 7 *cont'd*; Study Module 7.2	Provost 4 (Logistic Regression pp. 97-103)	Homework 4 due Saturday midnight
(11) 4/7	Study Conceptual Module 8; Module 8	Provost 6 (Nearest-Neighbor Reasoning pp. 141-157; optional technical pp. 157-163)	
(12) 4/14	Class Meeting Study Conceptual Module 9; Module 9	Provost 7	Hack-shop 4 due end of meeting
(13) 4/21	Study Module 10	McKinney 6.3~6.4	
(14) 4/28	Class Meeting Review exam #2		Final Project Deliverables due Saturday midnight
(15) 5/5	**Exam #2** covers Modules 6~10		

Policies Related to COVID-19 Pandemic

Recording of Class (required for all courses)

Students may not record all or part of class, livestream all or part of class, or make/distribute screen captures, without advanced written consent of the instructor. If you have or think you may have a disability such that you need to record class-related activities, please contact the Center for Students with DisABILITIES. If you have an accommodation to record class-related activities, those recordings may not be shared with any other student, whether in this course or not, or with any other person or on any other platform. Classes may be recorded by the instructor. Students may use instructor's recordings for their own studying and notetaking. Instructor's recordings are not authorized to be shared with *anyone* without the prior written approval of the instructor. Failure to comply with requirements regarding recordings will result in a disciplinary referral to the Dean of Students Office and may result in disciplinary action.

Excused Absence Policy (required for all courses)

Regular class attendance, participation, and engagement in coursework are important contributors to student success. Absences may be excused as provided in the University of Houston Undergraduate Excused Absence Policy and Graduate Excused Absence Policy for reasons including: medical illness of student or close relative, death of a close family member, legal or government proceeding that a student is obligated to attend, recognized professional and educational activities where the student is presenting, and University-sponsored activity or athletic competition. Additional policies address absences related to military service, religious holy days, pregnancy and related conditions, and disability.

Syllabus Changes (required for all courses)

Due to the changing nature of the COVID-19 pandemic, please note that the instructor may need to make modifications to the course syllabus and may do so at any time. Notice of such changes will be announced as quickly as possible through (*specify how students will be notified of changes*).