Course Name and Number: DATA 621 - Business Analytics and Data Mining
Credits: 3 cr.
Prerequisites: DATA 606 - Statistics and Probability for Data Analytics; DATA 607 - Data Acquisition and Management

Course Description:

This course develops the foundations of predictive modeling by introducing the key concepts of applied regression modeling and its extensions. The main topics covered in this course include: simple and multiple linear regression, variable selection and shrinkage methods, binary logistic regression, weighted least squares, robust regression, generalized least squares, count regression, multinomial logistic regression, generalized linear models, panel regression, and nonparametric regression. The course is heavily weighted towards practical application using the R statistical programming language and data sets containing missing values and outliers. The course also addresses issues of exploratory data analysis, data preparation, model development, model validation, and model deployment.

Course Learning Objectives:

By the end of the course, students should be able to:
- Demonstrate a practical understanding of the theoretical concepts behind applied regression modeling.
- Analyze and select appropriate types and combinations of models given particular business situations.
- Develop applied regression modeling techniques to address different types of data.
- Use R statistical software to build and deploy specific models based on real-world business problems.

Program Learning Outcomes/Competencies addressed by the course:

- Business Understanding. Students will learn how applied regression modeling techniques can add value to existing business analytics.
- Data Programming. Use industry standard statistical programming tools.
- Foundational Math and Statistics. Emphasis on probability, statistics, and computational methods.
- Data Culture. Students will learn how applied regression modeling can enhance business capabilities and extend the value of existing data.
- Data Understanding. Students will learn how to explore data to find new patterns.
- Predictive Modeling. Selecting predictive modeling techniques, building and assessing models.
- Model Implementation. Students will learn to implement models for the various applied regression modeling techniques covered in the course.
- Presentation. Students will deliver presentations of their assignment and project results.

How is this course relevant for IS and data analytics professionals?

Regression modeling skills are crucial, high-value skills in today's data-driven business environment where real-world decision-making processes are complex. The ability to leverage rapidly expanding data sets to obtain new insights is at the heart of predictive data analytics.

How does this course work?

The course is conducted entirely online via Blackboard. Each week, the student will complete assigned readings from the required textbooks, watch lecture videos, complete optional (but recommended) textbook exercises, complete homework assignments, and conduct a final group project. Students are expected to complete all deliverables by their assigned due dates.

Assignments and Grading:

Homework Assignments - There will be 5 homework assignments (15% each, or 150 points each) used to re-enforce course concepts and provide implementation experience. - Students will work on the homework assignments in collaboration with the members of their Critical Thinking Group. Only one homework solution should be submitted per group.	**75%**	**750 points**
Final Group Project - Students will form a Critical Thinking Group and conduct a final course project using regression modeling techniques covered in class to solve a real-world problem. - Only one project report should be submitted per group.	**25%**	**250 points**
TOTAL	**100%**	**1000 points**

Grading Scale: Your grade will be based on your final weighted average score and the letter grade will be assigned according to the following table.

Quality of Performance	Letter Grade	Range %	GPA/ Quality Pts.
Excellent - work is of exceptional quality	A	93 - 100	4
	A-	90 - 92.9	3.7
Good - work is above average	B+	87 - 89.9	3.3
Satisfactory	B	83 - 86.9	3
Below Average	B-	80 - 82.9	2.7
Poor	C+	77 - 79.9	2.3
	C	70 - 76.9	2
Failure	F	< 70	0

Discussion Board Etiquette: There is *no graded discussion board participation requirement*. However, a discussion board forum will be available each week to allow students to freely exchange ideas about the course content. It is imperative to remain respectful of all viewpoints and positions and, when necessary, agree to respectfully disagree.

Attendance: This course will not meet at a particular time each week, except for the scheduled course meetups. All course goals, session learning objectives, and assessments are supported through classroom elements that can be accessed at any time. If you are unable to attend a scheduled synchronous meeting, you will need to make alternative arrangements with the instructor.

Late Policy: Unless otherwise noted, all work is due on the assigned day by 11:59 PM (Eastern Time). This includes homework assignments and the final project. *Late work is not accepted, unless pre-coordinated with the instructor.*

Required Textbooks:
- *A Modern Approach to Regression with R,* by Simon J. Sheather. ISBN 978-0-387-09607-0 (**MARR**)
- *Linear Models with R*, by Julian J. Faraway. ISBN 978-1-4398-8733-2 (**LMR**)
- *Extending the Linear Model with R*, Julian J. Faraway. ISBN 1-58488-424-X (**ELMR**)

Recommended Textbooks:

- *Applied Regression Analysis and Generalized Linear Models* (2016), by John Fox. ISBN 978-1-4522-0566-3
- *A Guide to Modern Econometrics* (2012), by Marno Verbeek. ISBN 978-1-119-95167-4
- *Introductory Econometrics: A Modern Approach* (2013), by Jeffrey M. Wooldridge. ISBN 978-1-111-53104-4
- *Using R for Introductory Econometrics* (2016), by Florian Heiss. ISBN 978-1-523-28513-6

Relevant Software: The primary software environment is the R statistical programming language, which can be downloaded for *free* from http://www.r-project.org. RStudio is the recommended interface for the R statistical programming language software, which can also be downloaded for *free* at http://www.rstudio.org.

My Contact Information:

Instructor Name:	Dr. Nathan Bastian (Prof. B)
E-mail Address:	nathaniel.bastian@sps.cuny.edu
Response Times:	24 – 48 hours
Office Hours / Sync:	By appointment / Tuesdays from 8:00 – 9:00pm EST
Phone / Skype:	(570) 809-3619 / nathan.bastian

Office hours are conducted by appointment via GoToMeeting, Skype or phone. You are encouraged to ask questions on the "Ask Your Instructor" forum on the course discussion board where other students will be able to benefit from your inquiries. For the most part, you can expect me to respond to questions by email within 24 to 48 hours. If you do not hear back from me within 48 hours of sending an email, please resend your message. I will be checking in on the course regularly, just about every day and likely several times each day. You can expect me to grade and return assignments within 10 days. Please do not hesitate to ask if you have questions or concerns.

Meetup Information:

Please join my meeting from your computer, tablet or smartphone.

- **Web-link:** https://global.gotomeeting.com/join/173213461

You can also dial in using your phone:

- **United States:** +1 (669) 224-3412
- **Access Code:** 173-213-461

Course Outline:

Unit	Meetup	Topic	Readings	Key Tasks
Week #1 6/6 – 6/12	6/7 8-9pm	Simple Linear Regression: Estimation, Inference, Prediction and Explanation	*MARR* – Ch. 1, 2 *LMR* – Ch. 1 – 5	- HW #1 Assigned
Week #2 6/13 – 6/19	None	Simple Linear Regression: Diagnostics and Transformations / Multiple Linear Regression: Missing Data, Diagnostics and Transformations	*MARR* – Ch. 3, 5, 6 *LMR* – Ch. 6, 7, 9, 13, 14	- HW #1 Due - HW #2 Assigned

Week #3 6/20 – 6/26	6/21 8-9pm	Variable Selection, Shrinkage Methods, and Binary Logistic Regression	*MARR* – Ch. 7, 8 *LMR* – Ch. 10, 11 *ELMR* – Ch. 2	- HW #2 Due - HW #3 Assigned
Week #4 6/27 – 7/3	None	Weighted Least Squares, Robust Regression, and Generalized Least Squares	*MARR* – Ch. 4, 9 *LMR* – Ch. 8	- HW #3 Due - HW #4 Assigned
Week #5 7/4 – 7/10	None	Count Regression and Multinomial Logistic Regression	*ELMR* – Ch. 3, 5	- HW #4 Due - HW #5 Assigned
Week #6 7/11 – 7/17	7/12 8-9pm	Generalized Linear Models and Panel Regression	*ELMR* – Ch. 6, 7, 9	- HW #5 Due
Week #7 7/18 – 7/21	7/19 8-9pm	Nonparametric Regression	*ELMR* – Ch. 11	- Project Report Due

ACCESSIBILITY AND ACCOMMODATIONS

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see: http://sps.cuny.edu/student_services/disabilityservices.html

ONLINE ETIQUETTE AND ANTI-HARASSMENT POLICY

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see: http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf

ACADEMIC INTEGRITY

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see: http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf

STUDENT SUPPORT SERVICES

If you need any additional help, please visit Student Support Services: http://sps.cuny.edu/student_resources/