JOHNS HOPKINS
CAREY BUSINESS SCHOOL

Data Analytics
2 Credits

BU.510.650.K2

Fridays 1:30 – 4:30 PM
10/27/2023 – 12/22/2023

Fall 2, 2023
Room 215, Harbor East, Baltimore, MD

Instructor
Mohammad Ali Alamdar Yazdi

Contact Information
yazdi@jhu.edu

Office Hours
Wednesdays 5 - 7 PM
https://jhucarey.zoom.us/j/917963295?pwd=RkxnVWRRcVdTMnpMQ1lSYzg1YTBNdz09

Required Texts & Learning Materials
There is ***no required*** textbook: All required class materials will be available on our Canvas website. However, some books are very useful if you want to learn more about data analytics and its applications. The best way to *learn* is by *doing* (especially for R programming)

Optional Textbook 1 (solid primer, with theory and explanation):
An Introduction to Statistical Learning with Application in R, by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani;
Publisher: Springer (2021); ISBN-13: 978-1071614174;
The book is **free** at: https://www.statlearning.com/

Optional Textbook 2 (a great advanced text):
Elements of Statistical Learning: Data Mining, Inference, and Prediction, by Trevor Hastie, Robert Tibshirani and Jerome Friedman;
Publisher: Springer (2009); ISBN-13: 978-0387848570;
It requires some mathematical sophistication and goes beyond the material we will be covering. The book is **free** at: https://hastie.su.domains/ElemStatLearn/printings/ESLII_print12_toc.pdf

Software:
- We require the **R** Statistical Software, which is **powerful** and **free**. R can be downloaded at the link below: http://www.cran.r-project.org/
- **Rstudio** is a free platform for both writing and running R, available at www.rstudio.org. Some students find it friendlier than basic R (especially in windows OS).
- The learning curve is very steep. Students can become proficient in a few weeks. Some manuals are very helpful to learn R, e.g., http://cran.r-project.org/manuals.html

- I provide limited software instruction, in-class demonstration, and code to accompany lectures and assignments. We do not assume that you have used R in a previous class. However, **this is not a class on R**. Like any language, **R is only learned by doing**. You should install R as soon as possible and familiarize yourself with basic operations.
- Additional resources: YouTube intros to R.

Course Description
This course prepares students to gather, describe, and analyze data, and use advanced statistical tools to make decisions on operations, risk management, finance, marketing, health care management, etc. Analysis is done targeting economic and financial decisions in complex systems that involve multiple partners. Topics include probability, statistics, hypothesis testing, regression, clustering, decision trees, forecasting, and unsupervised learning, etc.

Prerequisite(s)
BU.510.601 OR BU.914.610

Learning Objectives
By the end of this course, students will be able to:

1. Gather sufficient relevant data, conduct data analytics using scientific methods, and make appropriate and powerful connections between quantitative analysis and real-world problems.
2. Demonstrate a sophisticated understanding of the concepts and methods; know the exact scopes and possible limitations of each method; and show capability of using data analytics skills to provide constructive guidance in decision making.
3. Use advanced techniques to conduct thorough and insightful analysis, and interpret the results correctly with detailed and useful information.
4. Show substantial understanding of the real problems; conduct deep data analytics using correct methods; and draw reasonable conclusions with sufficient explanation and elaboration.
5. Write an insightful and well-organized report for a real-world case study, including thoughtful and convincing details.
6. Make better business decisions by using advanced techniques in data analytics.

To view the complete list of the Carey Business School's general learning goals and objectives, visit the Carey website.

Attendance
Attendance and class participation are part of each student's course grade. Students are expected to attend all scheduled class sessions. Failure to attend class will result in an inability to achieve the objectives of the course. Excessive absence will result in loss of points for participation. Regular attendance and active participation are required for students to successfully complete the course.

Class participation is an important part of learning. If you have a question, it's likely that others do as well. I encourage active participation, and course grades will consider students who make particularly strong contributions.

Assignments
Late submissions—including assignments, and exams—will *not* be accepted.

Assignment	Learning Objectives	Weight
Attendance and participation in class discussion		10%
Individual Homework	1, 2, 3, 4, 5, 6	50%
Final Exam	1, 2, 3, 4, 5, 6	40%
Total		100%

Study Groups (not required, but highly recommended)
Many students learn better and faster when working in a group, so I encourage collaborative learning. You can work together in a study group with several students to discuss class materials and homework assignments on a weekly basis. However, each student must write the homework assignment individually; your text should reflect your own understanding of the materials.

Grading
The grade of A is reserved for those who demonstrate extraordinary performance as determined by the instructor. The grade of A- is awarded only for excellent performance. The grades of B+ and B are awarded for good performance. The grades of B-, C+, C, and C- are awarded for adequate but substandard performance. The grades of D+, D, and D- are not awarded at the graduate level. The grade of F indicates the student's failure to satisfactorily complete the course work. For Core/Foundation courses, the grade point average of the class should not exceed 3.35. For Elective courses, the grade point average should not exceed 3.45.

Tentative Course Calendar
Instructors reserve the right to alter course content and/or adjust the pace to accommodate class progress. Students are responsible for keeping up with all adjustments to the course calendar.

Week	Date	Weekly Objectives/Topics	Recommended Reading (Textbook 1)	Assignments
1	10/27	Introduction to Statistical Learning	Chapter 1	
2	11/03	Data Summarization and Visualization	Chapter 2	HW 1
3	11/10	Linear Regression	Chapter 3	HW 2
4	11/17	Decision Trees	Chapter 8	HW 3
-	11/24	*Thanksgiving Break*		
5	12/01	Clustering	Chapter 12	HW 4
6	12/08	Logistic Regression	Chapter 4	HW 5
7	12/15	Model Selection	Chapters 5, 6	HW 6
8	12/22	Final Exam (class time and remote)		

Carey Business School Policies and General Information

Please note that failure to become acquainted with Carey policies will not excuse any student from adhering to these policies.

Canvas Site
A Canvas course site is set up for this course. Each student is expected to check the site throughout the semester as Canvas will be the primary venue for outside classroom communications between the instructor and students. Students can access the course site at https://canvas.jhu.edu/.

Technical Support
24/7 technical support for questions regarding Canvas, Zoom, and other technical issues is available. Please refer to Carey's Academic Resources webpage for contact information and other details.

Students with Disabilities - Accommodations and Accessibility
Johns Hopkins University values diversity and inclusion. We are committed to providing welcoming, equitable, and accessible educational experiences for all students. Students with disabilities (including those with psychological conditions, medical conditions, and temporary disabilities) can request accommodations for this course by providing an Accommodation Letter issued by Student Disability Services. Please request

accommodations for this course as early as possible to provide time for effective communication and arrangements. For further information or to start the process of requesting accommodations, please contact Student Disability Services at the Carey Business School.

Academic Ethics Policy
Carey expects graduates to be exemplary global citizens in addition to innovative business leaders. The Carey community believes that honesty, integrity, and community responsibility are qualities inherent in an exemplary citizen. The objective of the Academic Ethics Policy (AEP) is to create an environment of trust and respect among all members of the Carey academic community and hold Carey students accountable to the highest standards of academic integrity and excellence.

It is the responsibility of every Carey student, faculty member, and staff member to familiarize themselves with the AEP and its procedures. Failure to become acquainted with this information will not excuse any student, faculty, or staff member from the responsibility to abide by the AEP. Please contact the Office of Student Affairs if you have any questions. For the full policy, please visit the Academic Ethics Policy webpage.

Student Conduct Code
The fundamental purpose of the Johns Hopkins University's regulation of student conduct is to promote and to protect the health, safety, welfare, property, and rights of all members of the University community as well as to promote the orderly operation of the University and to safeguard its property and facilities. Please contact the Office of Student Affairs if you have any questions regarding this policy. For the full policy, please visit the Student Conduct Code webpage.

Commitment to Respect
Respectful behavior creates an environment within the Carey Business School where all are valued and can be productive. Carey defines respectful behavior as conduct that, at a minimum, demonstrates consistent courtesy for others, including an effort to understand differences. As such, all in the community agree to the Carey Commitment to Respect, which states that we all strive to show that we value each other's human dignity and our differences, and to choose behavior and language that demonstrates mutual respect. Please visit the Commitment to Respect webpage to learn more about the expectations and resources available.

Classroom Policies for All On-Site and Remote-Live Classes
Carey is committed to maintaining the highest standards of excellence in all forms of instruction. To that end, we have developed policies and procedures for all classes offered in on-site and remote-live formats. These policies will govern all courses occurring in these formats, and all students are expected to familiarize themselves with and adhere to these policies.

Student Success Center
The Student Success Center offers assistance in core writing and quantitative courses. For more information, visit the Student Success Center webpage.

Other Important Policies and Services
Students are encouraged to consult the Student Handbook and Academic Catalog and Student Services and Resources for information regarding other policies and services. For your convenience, there is a singular website students can visit to learn about all JHU and Carey policies.

Copyright Statement
Unless explicitly allowed by the instructor, course materials, class discussions, and examinations are created for and expected to be used by class participants only. The recording and rebroadcasting of such material, by any means, is forbidden. Violations are subject to sanctions under the Academic Ethics Policy.

Appendix. Homework Rubric for Data Analytics Course: Part 1

Assessment Criteria	Not Good Enough (0 ≤ score <6)	Good (6≤ score <9)	Very Good (9 ≤ score ≤10)	Score
Deep understanding of theory and its applications, using qualitative methods to answer business questions	Demonstrate inadequate understanding of some important concepts, methods or their applications, e.g., choose wrong methods, conduct analysis inappropriately, or interpret results incorrectly.	Understand concepts and methods relatively well, analyze data using acceptable methods although not perfect; be able to derive useful information for decision making.	Demonstrate sophisticated understanding for the concepts and methods; know the exact scopes and possible limitations of each method; show capability of using data analytics skills to make right business decision.	
Implementation and interpretation of data analysis techniques	Use wrong techniques to analyze data, present inappropriate interpretations or conclusions.	Choose acceptable methods to analyze data, interpretations are sensible, derive useful results.	Use advanced techniques to conduct thorough and insightful analysis, interpret the results correctly, draw right conclusions based on data analysis .	
Ability of solving real-world problems using quantitative methods	Data is inadequate or unstructured. Use inappropriate methods to analyze data, fail to retrieve useful information. Suggestions are not persuading.	Collect and document just enough data, employ appropriate techniques to retrieve insightful information from data, make reasonable recommendations.	Gather sufficient relevant data, conduct data analytics using scientific methods, make appropriate and powerful connections between analysis and real-world problems, provide constructive guidance in decision making .	
Writing and presenting, especially on organization and communication	Report is inadequately written and poorly organized. Analysis is insufficient. Conclusions are unconvincing.	Report is concise and clearly written. Analyze problems following scientific strategies; provide useful suggestions with detailed explanation.	Report is well organized and insightfully written, includes thorough and thoughtful details. Conclusions are convincing.	
			Total Score	
Comments:				

Appendix. Homework Rubric for Data Analytics Course: Part 2

Assessment Criteria	Not Good Enough (0 ≤ score <6)	Good (6≤ score <9)	Very Good (9 ≤ score ≤10)	Score
Interpretation of Data (qualitative)	Little or no attempt to interpret data; or there are significant errors; or some data are over- or under-interpreted.	Interpret most data correctly; part of conclusions may be suspect; suggestions on future implementation are sound.	Data are completely and appropriately interpreted; there is no over- or under-interpretation; draw convincing conclusions.	
Statistical Analysis (quantitative)	Statistical methods are completely misapplied or applied but with significant errors or omissions. Choose inappropriate methods and make wrong predictions.	Most statistical methods are correctly applied but more could have been done with the data. Predictions are sensible but may deviate from the true results in a large range.	Statistical methods are fully and correctly applied; demonstrate superior data analysis skills; deeply mine the data and obtain useful insights for decision making.	
Critical evaluation of findings	Blindly accept defective results; or recognize defective results but does not know how to fix them.	Recognize defective results and figure out the causes; understand the main sources of errors.	Show deep understanding for the sources of errors; recognize defective results and eliminates the causes.	
Ability to draw proper conclusions and make effective suggestions	Not draw conclusions; draw incorrect conclusions; suggestions are not acceptable .	Draw correct conclusion; suggestions may have potential impact on the future business.	Demonstrate substantial understanding of the problem; conduct deep data analytics using correct methods; draw correct conclusions with sufficient explanation and elaboration.	
			Total Score	
Comments:				