RUTGERS
Rutgers Business School
Newark and New Brunswick

Statistics
Course Number: 22:960:646
Course Title: Data Analysis and Visualization

COURSE DESCRIPTION

Data analytics and visualization is an emerging field concerned with analyzing, modeling, and visualizing complex high dimensional data. This course will introduce state-of-the-art modeling, analysis and visualization techniques. It will emphasize practical challenges involving complex real world data and include several case studies and hands-on work with the R/Python programming language.

COURSE MATERIALS

Textbook: Business Intelligence: A Managerial Perspective on Analytics (3rd Edition) Paperback – by Ramesh Sharda, Dursun Delen, and Efraim Turban.

Textbook: An Introduction to Statistical Learning with Applications in R by Gareth James, Daniell Witten, Trevor Hastie, Robert Tibshirani: Springer This book is downloadable free from Stanford University: http://www-bcf.usc.edu/~gareth/ISL/

Reference: Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic, Wiley

Software: R Studio, R 3.4.3. Check the course site for download instructions.

Check Blackboard (blackboard.rutgers.edu) and your official Rutgers email account regularly.

Additional technical readings for R and Python will be available on Blackboard

---

LEARNING GOALS AND OBJECTIVES

- This course is designed to provide students with the foundations necessary for understanding and extending the current state of the art in data visualization.

- Apply quantitative modeling and data analysis techniques to the solution of real world business problems, communicate findings, and effectively present results using data visualization techniques.
- Demonstrate knowledge of statistical data analysis techniques utilized in business decision making.
- Apply principles of Data Science to the analysis of business problems.
- Employ cutting edge tools and technologies to analyze Big Data.

This course combines lectures, in class examples, lab exercises, individual assignments, and exams, to achieve the stated objectives. Class sessions will comprise (1) lecture/discussion of various data analytics and visualization concepts, (2) instructor demonstrations of these data analytics and visualization concepts, and (3) student lab sessions working with these data analytics and visualization concepts.

The purpose of this pedagogical approach is to introduce and reinforce ideas and skill sets so that students can master these on their own after class hours. To bring this knowledge to a highly proficient, professional level, students will have to spend time and effort outside of class using the system.

PREREQUISITES
None

ACADEMIC INTEGRITY
Students are responsible for understanding the RU Academic Integrity Policy (https://slwordpress.rutgers.edu/academicintegrity/wpcontent/uploads/sites/41/2014/11/AI_Policy_2013.pdf). All assignments and individual projects are to be done individually and must be the result of your own work. Do NOT turn in a copy or duplicate of someone else's work nor work that does NOT represent your individual effort. Violation of this policy will result in a grade of zero for the assignment and/or a failing course grade.

ATTENDANCE AND PREPARATION POLICY
- **Class Policies**
  This syllabus and the class schedule are subject to modification. Modifications may be communicated in class, on the course website, or via email. This may include changes to assignments, dues dates, exams, material covered, etc. The course schedule will be updated throughout the semester and can be viewed/printed on the course website.
- **Attendance, and Tardiness:** You are strongly encouraged to attend all class meetings. If you are tardy or absent, it is your responsibility to obtain class notes and announcements (including assignment or schedule changes) from other students and/or your instructor. Make-up exams are not available (except for excused absences).
- **In-Class Participation:** Coming to class fully prepared helps push the conversation much further. You are encouraged and expected to ask questions and enter into discussions on relevant issues as part of the class. However please respect other students and your professor by minimizing unnecessary personal conversations, text messaging, and other disturbances in class.
- Professional conduct is expected from all students.

EXAM DATES AND POLICIES
Midterm Exam:
Final Exam:
During exams, the following rules apply:
- If you have a disability that influences testing procedures, provide me an official letter from the Office of Disability Services at the start of the semester.
- No cell phones or other electronics are allowed in the testing room.

GRADING POLICY
ALL deliverables for this class have weightings which means they contribute differently towards your final grade. The weightings of each course deliverable is as follows:

Deliverables	**Weights**
Homework Assignment	15%
Midterm	20%
Final Project	30%
Final Exam	30%
Participation	5%

****Late submission policy: No submission is accepted after the due date.***

SUPPORT SERVICES

If you need accommodation for a *disability*, obtain a Letter of Accommodation from the Office of Disability Services. The Office of Disability Services at Rutgers, The State University of New Jersey, provides student-centered and student-inclusive programming in compliance with the Americans with Disabilities Act of 1990, the Americans with Disabilities Act Amendments of 2008, Section 504 of the Rehabilitation Act of 1973, Section 508 of the Rehabilitation Act of 1998, and the New Jersey Law Against Discrimination. https://ods.rutgers.edu

If you are a military *veteran* or are on active military duty, you can obtain support through the Office of Veteran and Military Programs and Services. http://veterans.rutgers.edu/

If you are in need of *mental health* services, please use our readily available services.
Rutgers University-Newark Counseling Center: http://counseling.newark.rutgers.edu/

If you are in need of *physical health* services, please use our readily available services.
Rutgers Health Services – Newark: http://health.newark.rutgers.edu/

If you are in need of *legal* services, please use our readily available services: http://rusls.rutgers.edu/

COURSE SCHEDULE

Topic	Items Due

Lecture 1: Introduction: Data Science for Business ▪ The Ubiquity of Data Opportunities ▪ Examples ▪ Data Science, Engineering, and Data-Driven Decision Making ▪ Data Processing and "Big Data" ▪ From Big Data 1.0 to Big Data 2.0 ▪ Data and Data Science Capability as a Strategic Asset ▪ Data-Analytic Thinking ▪ Chemistry Is Not About Test Tubes: Data Science Versus the Work of the Data Scientist Introduction to R - Basics	
Lecture 2: Business Problems and Data Science Solutions ▪ From Business Problems to Data Mining Tasks o Business Understanding o Data Understanding o Data Preparation o Modeling o Evaluation o Deployment ▪ Analytic techniques and technologies ▪ Answering business questions with these techniques Lab 1: Visualizing categorical data (bar and column charts, pie charts, and radial diagrams) Visualizing Continuous data (regression line, box and whiskers plot, density plot, correlation plot)	Team preferences
Lecture 3: Introduction to Predictive Modeling ▪ Linear Regression ▪ Classification o Logistic Regression o Support Vector Machines o Decision Trees ▪ Example: Addressing the churn problem	
Lecture 4: Fitting a Model to Data ▪ What is a good model? o Overfitting o Generalization ▪ Evaluating Classifiers o Plain Accuracy and Its Problems o Confusion Matrix	Lab Exercise 1

○ Problems with Unbalanced Classes ○ Problems with Unequal Costs and Benefits ▪ Generalizing Beyond Classification ○ Using Expected Value to Frame Classifier Evaluation ▪ Evaluation, Baseline Performance, and Implications for Investments in Data Lab 2: Distributions (histogram, pyramid, and box plots) Scatter plots	
Lecture 5: Similarity, Neighbors, and Clusters ▪ Similarity and Distance ▪ Nearest-Neighbor Reasoning ○ Example: Whiskey Analytics ○ How Many Neighbors and How Much Influence? ○ Issues with Nearest-Neighbor Methods ▪ Clustering ○ Hierarchical clustering ▪ Example: Whiskey Analytics ○ Nearest Neighbors Revisited: Clustering Around Centroids ▪ Example: Clustering Business News Stories ○ Understanding the Results of Clustering ▪ Stepping Back: Solving a Business Problem Versus Data Exploration	
Midterm Exam	
Spring Break	
Lecture 6: Decision Analytic Thinking ▪ Targeting the Best Prospects for a Charity Mailing ○ The Expected Value Framework: Decomposing the Business Problem and Recomposing the Solution Pieces ○ A Brief Digression on Selection Bias ▪ Churn Example Revisited with Even More Sophistication ○ The Expected Value Framework: Structuring a More Complicated Business Problem ○ Assessing the Influence of the Incentive ○ From an Expected Value Decomposition to a Data Science Solution Lab 3: Heat maps and dendrograms	Lab Exercise 2 Team Assignment 1
Lecture 9: Visualizing Model Performance ▪ Ranking Instead of Classifying ▪ Profit Curves ▪ ROC Graphs and Curves ▪ The Area Under the ROC Curve (AUC)	

▪ Cumulative Response and Lift Curves ▪ Example: Performance Analytics for Churn Modeling	
Lecture 8: Evidence and Probabilities ▪ Example: Targeting Online Consumers with Advertisements ▪ Combining Evidence Probabilistically o Joint Probability and Independence o Bayes' Rule ▪ Applying Bayes' Rule to Data Science o Conditional Independence and Naive Bayes o Advantages and Disadvantages of Naive Bayes ▪ A Model of Evidence "Lift" ▪ Example: Evidence Lifts from Facebook "Likes" o Evidence in Action: Targeting Consumers with Ads	Lab Exercise 3
Lecture 7: Other Data Science Tasks and Techniques ▪ Co-occurrences and Associations: Finding Items That Go Together o Measuring Surprise: Lift and Leverage o Example: Beer and Lottery Tickets o Associations Among Facebook Likes ▪ Profiling: Finding Typical Behavior ▪ Link Prediction and Social Recommendation ▪ Data Reduction, Latent Information, and Movie Recommendation ▪ Bias, Variance, and Ensemble Methods ▪ Data-Driven Causal Explanation and a Viral Marketing Example Lab 4: Maps	
Lecture 10: Data Science and Business Strategy ▪ Thinking Data-Analytically ▪ Achieving Competitive Advantage with Data Science ▪ Sustaining Competitive Advantage with Data Science ▪ Formidable Historical Advantage ▪ Unique Intellectual Property ▪ Unique Intangible Collateral Assets ▪ Superior Data Scientists ▪ Superior Data Science Management ▪ Attracting and Nurturing Data Scientists and Their Teams ▪ Be Ready to Evaluate Proposals for Data Science Projects ▪ A Firm's Data Science Maturity	
Student Presentations	Team Assignment 2 Lab Exercise 4

Final Exam	