This is a two credit module that examines the opportunities and industry disruption in an era of massive, high velocity, unstructured data and new developments in data analytic. We treat some strategic, ethical, and technical dimensions of big data. The technical foci of the course include data structures, data warehousing, Structured Query Language (SQL), and high-impact visual displays. The principal objective of the course is to help students build understanding of data as an essential competitive resource, and acquire advanced computer skills through cases and hands-on applications. Assignments and classroom time will be devoted to both to analysis of current developments in analytics and to gaining experience with current tools.

Required Reading

- Readings as posted on LATTE site.

Learning Goals and Objectives

Upon successful completion of this module, students will:

- Think of data as a strategic resource in business.
- Understand the logic of complex data queries in the context of on-line business research sources.
- Be familiar with current developments in Big Data, business intelligence, and competitive analytics.
- Be able to design a relational database structure suited to a business enterprise.
- Understand the relationships between human cognitive processes and effective informational visualization.

Prerequisites

Students should have some prior background in accounting and statistical techniques at a level comparable to that provided by FIN 212a and ECON 210f(1).
Nobel Laureate Herbert Simon wrote “What information consumes is rather obvious: it consumes the **attention** of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.” This was true when Simon wrote it in 1969 and it is all the more true in our current age of Big Data.

This course is designed to provide students with an understanding of some techniques for managing information abundance and for controlling the costs and benefits of information processing in decision contexts. As such, we start with typical decision-making situations in business settings and work towards making data-driven decisions. Readings focus on the theory of decision-making, data structure, and analytic models. In addition, articles and cases illustrate typical decision problems and the application of the techniques we will study.

Communications

We’ll make regular use of LATTE. All lecture notes, handouts, assignments, and supporting materials will be available via LATTE, and any late-breaking news will reach you via email. Please check your Brandeis email and the LATTE site regularly to keep apprised of important course-related announcements.

Other Course Technology

In addition to software available on the IBS computer clusters, we will also use web-based resources made available through the Teradata Student Network (TSN). This site, sponsored by Teradata, the Walton School of Business at the University of Arkansas, IBM, MicroStrategy, SAS and others, is a gateway to articles, cases, software tools and real corporate data. Details about use of TSN will be provided separately on our LATTE site.

Throughout the course you have the choice of using the public computer clusters at IBS and/or your personal laptops. If you do use a laptop, the class schedule below indicates dates when it will be useful to have it with you. Three key tools that we will use in this module are:

- **Teradata SQL Assistant**: Teradata SQL Assistant will provide exposure to writing Structured Query Language code to interrogate a large database compiled by Dillard’s Department Stores, a large US retail chain.

- **R**: R is a free software environment for statistical computing and graphics, and is widely used by both academia and industry. In this module we will use R mainly for data visualization. The advantage of the R software is that it can work on both Windows and Mac-OS. It is ranked no. 1 in the KDnuggets 2013 poll on top languages for analytics, data mining, and data science. RStudio is a user friendly environment that has become popular.
  
  R Software: [http://www.r-project.org/index.html](http://www.r-project.org/index.html)
  
  
  Youtube link for downloading & installing R and RStudio: [https://www.youtube.com/watch?v=PHi-6GWfMBQ&index=1&list=PL34t5iLFzdDv8tjZboegN6tmyh2-zr_T](https://www.youtube.com/watch?v=PHi-6GWfMBQ&index=1&list=PL34t5iLFzdDv8tjZboegN6tmyh2-zr_T)

- **LucidCharts**: LucidCharts is a cloud-based tool that enables users to design a coherent data structure as part of the process of designing a database.
### Student Contributions

Class participation is important in this course both as a means of developing understanding and as an indicator of student progress. Participation can take many forms, and each student is expected to contribute actively, freely, and effectively to the classroom experience by raising questions, demonstrating preparedness and proficiency in the analysis of problems and cases, and explaining the implications of particular analyses in context. Homework-based discussion and presentations are an important part of participation. **To this end, regular class attendance is required, and students should use name cards.** Even if you must arrive late or leave early, be here.

With assistance from the TA, I will evaluate the quality of your contributions in class each evening, as well as the quality of your contributions via email, LATTE discussion, etc. These will all be factored together in determining your ultimate Contributions grade (see below). In general, absence from class reduces your contribution grade.

### Written Assignments, Projects, and individual teach-a-topic presentation

Students will complete five written assignments during the course. Three of these will be brief analyses, requiring modest analysis and writing. One assignment will be individual and **remaining two may be completed with one or two partners**, and each student should expect to briefly discuss one of these analyses in class.

Two other written assignments will be “Projects” requiring more significant time and analysis. The projects will be prepared in teams of four students, and will include written and computer-based elements. Owing to the size of the class this term, students will have only limited opportunities to present parts of their projects orally in the course.

Each student will be required to teach a topic by working in a team of maximum three students and choosing topics related to the class (Schedule of these presentations will be done on the first day of class). This presentation should be for about 6 to 8 minutes. Students can choose and present a summary of any three articles of their choice from the following link:

[http://flip.it/f5XfI](http://flip.it/f5XfI)

All assignments should be submitted via LATTE upload prior to the start of class. Papers should be professional in appearance and use clear, grammatically correct business English. Analytical work (graphs, tables, and other output) should be incorporated seamlessly into the written document, showing readers exactly and only what you want them to see.

### Evaluation

Your final grade in the course will be computed using these weights:

<table>
<thead>
<tr>
<th>Component</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online quizzes &amp; Class Discussions</td>
<td>10%</td>
</tr>
<tr>
<td>Teach-a-topic</td>
<td>10%</td>
</tr>
<tr>
<td>Brief analyses (3)</td>
<td>30%</td>
</tr>
<tr>
<td>Projects (2)</td>
<td>50%</td>
</tr>
</tbody>
</table>

**TOTAL** 100%
Academic Integrity

You are expected to follow the University’s policies on academic integrity (see http://www.brandeis.edu/studentaffairs/srcs/ai/index.html). Instances of alleged dishonesty will be forwarded to the Office of Campus Life for possible referral to the Student Judicial System. Potential sanctions include failure in the course and suspension from the University.

Disabilities

If you are a student with a documented disability on record at Brandeis and wish to have a reasonable accommodation made for you in this class, please see me immediately.

Study Groups

Working with one or two partners is an excellent way to gain understanding of this subject. I encourage small groups to work on assignments, with a few caveats:

- Be sure that you are neither carrying nor being carried by the group; each member of the group is entitled to learn.
- Except for the group project, each student is responsible for turning in original memos and problem sets.
- Each group member retains the right to “go it alone.” Joining a group is not a marriage. Similarly, teams are encouraged to dismiss underperforming members.

Workload expectation

As BUS 211F is a two-credit course, you are expected to spend a minimum of 9 hrs of study time per week in preparation for class (readings, assignments, discussion sections, projects, research, etc.).

Course Outline

Note: for each session, you should complete the assigned reading before coming to class. See the list of deliverables on below; detailed assignments will be posted each week, and all assignments and handouts will also be available on our LATTE site.

<table>
<thead>
<tr>
<th>Session Date</th>
<th>Topics and Readings</th>
<th>Upload to LATTE before class</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session 1</strong></td>
<td><strong>Jan 15</strong></td>
<td></td>
</tr>
<tr>
<td>Introduction to Big Data Analytics</td>
<td><strong>READINGS:</strong> Davenport HBR article “Competing on Analytics” McKinsey Report Exec Summary (LATTE)</td>
<td><em>(none)</em></td>
</tr>
<tr>
<td></td>
<td>a. Course introduction and objectives</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Introduction to BIG data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Getting started with R/RStudio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. Team Activity</td>
<td></td>
</tr>
<tr>
<td><strong>Session 2</strong></td>
<td><strong>Jan 22</strong></td>
<td></td>
</tr>
<tr>
<td>Big Data &amp; Data Exploration</td>
<td><strong>Laptops will be useful</strong> <strong>READINGS:</strong> R and Data Mining: Examples and Case Studies (Ch 1-3) Princeton – Intro to RStudio</td>
<td>Quiz 1 &amp; 2 TAT PPT</td>
</tr>
<tr>
<td></td>
<td>a. Teach-a-topic Presentation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. RStudio</td>
<td></td>
</tr>
<tr>
<td>Session Date</td>
<td>Topics and Readings</td>
<td>Upload to LATTE before class</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
</tr>
</tbody>
</table>
| **Session 3** Jan 29 | **SQL / RStudio**  
  **READING:** Russell section on SQL (Chap 3)  
  Selected Teradata material: SQL, Dillards  
  Laptops will be useful  
  a. Concepts of coding  
  b. SQL: one language, many dialects  
  c. Fundamental commands: SELECT, FROM, WHERE, JOIN, computation of a new column (AS), case sensitivity, relational operators  
  d. Teradata interface  
  e. RStudio | Analysis 1 |
| **Session 4** Feb 5 | **Data Exploration/ Visualization**  
  Laptops will be useful  
  READING: R and Data Mining: Examples and Case Studies (Ch 1-3)  
  Princeton – Intro to RStudio  
  a. Data exploration/ visualization with R - Categorical variables  
  b. Data exploration/ visualization with R - Numeric variables  
  c. Data exploration/ visualization with R - Relationships | Analysis 2 |
| **Session 5** Feb 12 | **Visualizing Text**  
  Laptops will be useful  
  a. Debrief Project-1  
  b. Visualizing text in R | Project 1 |
| **Session 6** Feb 26 | **Cluster Analysis**  
  Laptops will be useful  
  a. Hierarchical clustering  
  b. Non-hierarchical clustering  
  c. K-means clustering  
  d. Project-2 | Quiz 3  
  Analysis 3 |
| Mar 4 | **No Class Session this week**  
  - Final project due before this date  
  - Early submissions appreciated | Quiz 4  
  Project 2 |

**Brief Description of Assignments (complete assignment details to be provided in class):**

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis 1</td>
<td>As explained in class.</td>
</tr>
<tr>
<td>Analysis 2</td>
<td>Visualization-1</td>
</tr>
<tr>
<td>Analysis 3</td>
<td>Visualization-2</td>
</tr>
<tr>
<td>Project 1</td>
<td>SQL</td>
</tr>
<tr>
<td>Project 2</td>
<td>Cluster analysis</td>
</tr>
</tbody>
</table>