C20.0057 – Spring 2004
MODELING AND DATA MINING WITH SPREADSHEETS

SYLLABUS

INSTRUCTOR : Ravi Mantena
KMC 8-177, Ph: 998-0824, Fax: 995-4228, Email: rmantena@stern.nyu.edu
Office hours: Monday 10 a.m. to 12 p.m., or by appointment

CLASS : TR 9:30 a.m. -- 10:45 a.m. Tisch 202

TUTOR : Gal Oestreicher-Singer (goestrei@stern.nyu.edu)
KMC 8-151, Ph: 998-0456
Office hours: Wednesday 4 p.m. to 6 p.m., or by appointment

COURSE OVERVIEW

As the volume of data that organizations collect continues to grow rapidly, so does managers’ struggle to make sense of it. People who are facile with data and can leverage it into valuable intelligence are in short supply. This course teaches students how to structure and solve business problems using analytical and data-driven models. It provides a hands-on learning experience using the familiar Microsoft Excel modeling environment, while significantly extending its power through add-ins. The aim is two-fold:

- To provide students with powerful spreadsheet modeling skills that enable them to implement and solve a wide range of problems – from capital budgeting and investments to marketing and operations.
- To introduce students to state-of-the-art data mining methods that support decision making by extracting useful knowledge from the increasingly large volumes of raw data that organizations collect through their business processes. Techniques covered in the course will include classification and regression trees, nearest neighbor methods, multiple and logistic regression, neural networks, affinity analysis and clustering.

The course will be taught almost entirely through examples and emphasis will be on creative problem formulation and implementation. Technical details will be kept to a minimum. The course assumes prior knowledge of Microsoft Excel and the fundamentals of finance, marketing and operations at the level of the core freshmen courses at Stern. Prior experience with a programming language or data mining, though useful, is by no means necessary.
COURSE OBJECTIVES

By the end of this course the students will:

- Improve their ability to model and analyze practical business problems using spreadsheets. They will learn to use advanced Excel functions, add-ins and VBA – tools that significantly expand the power of Excel in analyzing business problems and data. The skills learnt will be valuable not only in their jobs and internships, but also in other courses as well.
- Get comfortable with data and improve their skills at using data intelligently to solve business problems.
- Get an intuitive understanding for a number of state-of-the-art data mining methods that can be employed to extract useful knowledge from corporate data. They will recognize the strengths/limitations of these methods and will be able to identify business settings in which these methods can be used effectively.

COURSE WEBSITE AND COMMUNICATION

Website: In this class we will use the web-based “Blackboard” course platform. You can access the course website by logging onto http://sternclasses.nyu.edu using your Stern email username and password, and then choosing the course titled “C20.0057: Modeling and Data Mining with Spreadsheets”. We will use the blackboard site extensively for the course. It will serve as a repository for class materials – current schedule and reading list, assignments and projects, data sets and class handouts will all be available from it.

Announcements: All announcements regarding the course will be posted on the Announcement section of the blackboard course site. No additional email notifications will be sent to the students, save in exceptional circumstances. So please make it a point to check the site every day.

Communication (You to me): You should use the relevant discussion forums on the blackboard site to post your questions. Please refrain from calling or sending emails to me or the tutor unless the communication is of an urgent or personal nature. If you have to send me an email, please include the course number along with a short descriptive phrase about the contents of the email in the subject, and also supply your name in the body of the email. Questions and responses posted on the forums build active discussions and benefit all students. I strongly encourage you to raise issues or post URLs/comments that are likely to be of interest to others in the class. I also encourage you to respond constructively (and responsibly) to each others’ questions and comments. I will check through the discussion boards on a regular basis, and respond to the questions/comments where appropriate. There are currently two discussion boards available from the blackboard site (Click on the Discussion Boards button to access them). They are:
• Administrative issues: Here you will post questions and comments pertaining to the availability or scheduling of readings, assignments, projects or any technical problems (not related to Excel) that you may be facing. Please ensure that you go through the announcements and/or the course syllabus, as well as the discussions already posted there before posting a question or comment of your own.
• Classes and content: Here you will post questions and comments about the material covered in the class, problems related to Excel or other software used in the class, or any other material you feel may be of common interest to other students in the class.

Group specific discussion boards will be set up later in the semester, after you have formed groups (of 4 to 5 people each) for you projects. Once formed, postings on the group boards will only be available to members of the group (you should be aware that I will have access to these postings, although I am unlikely to go through interaction among group members).

COURSE MATERIALS

SOFTWARE: Microsoft Excel XP/2000 spreadsheet software along with the following add-ins.
  • XLMiner (Data Mining add-in used for the second part of the course)
  • Solver (Optimization add-in)
  • Data-analysis Toolpack

All the necessary software is installed in all the undergraduate computer labs (Tisch – L101, LC11-13-15). In addition, both Solver as well as the Data-analysis toolpack come bundled with Microsoft Excel and should hence be available on your personal PCs as well (they may not be available if they were not installed when your Excel was installed, in which case, you’ll need to install them from the Excel CD).

BOOKS:

There are no required print textbooks for this course. However, we will be using materials from a number of books available electronically (some of them through the Bobst library, so you’ll have to access these through the NYU network).

Books used in the class:

Data Mining:

1. ‘Data Mining in Excel with Cases’, by Nitin Patel and Peter Bruce (Electronic access to the book will be provided free of charge to students registered for the course). We will use this book for the second part of the course. This is an as yet unpublished book, and hence will be a little rough around the edges. However, since it is written by the one of the creators of the XLMiner software that we’ll be using for the course, it will suit our purpose quite well.

Reference books (Not mandatory): Some of the better-established textbooks in this field.
  • ‘Seven Methods for Transforming Corporate Data into Business Intelligence’, by Vasant Dhar and Roger Stein, Prentice-Hall 1997. (This is a good non-technical reference on data mining methods).
  • ‘Machine Learning’, by Tom Mitchell, McGraw Hill 1997 (This book is an excellent reference on the technical aspects of the data mining methods)
Excel (and its application for modeling):


Other Electronic references for Excel:

Visual Basic for Applications


Other Electronic references for VBA:

In addition to the extracts from these books, references to electronic resources on the Internet will be posted for some of the sessions. Most of these references will constitute additional (optional) rather than required reading.

COURSE REQUIREMENTS AND GRADING

**BREAKDOWN.** Your final grade for the course will be determined based on your relative performance in the following:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Assignments</td>
<td>35%</td>
</tr>
<tr>
<td>Group Project</td>
<td>30%</td>
</tr>
<tr>
<td>End-Term Exam</td>
<td>30%</td>
</tr>
<tr>
<td>Class Participation</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

**HOMEWORK ASSIGNMENTS AND PROJECT:** All homework assignments, except the group project, are to be done *individually*. Each of these will involve developing a model in Excel, analyzing it and summarizing your findings/recommendations in a short report. You will submit both the report (in Microsoft word or pdf format) and the Excel file. In addition, the group project, which’ll be handed out during the second part of the course and will be done in groups of 4 to 5 people, will also require a short class presentation. Unless expressly stated otherwise, you’ll download and upload your homework assignments to the Assignments area of the course Blackboard website. There will be six to seven homework assignments during the semester (in addition to the group project) – an average of about one every two weeks. You will typically get about two weeks to work on each homework assignment and most homework assignments (unless otherwise stated) will be due by the start of the Tuesday class. Homework assignments handed in after the due data and time will be graded, but will be penalized 25%
of the points. If the assignments are late by more than a week, you’ll lose 50% of the points for that homework.

Since you’ll be submitting both your reports as well as your Excel files as part of your homework assignments, you are strongly encouraged to use some kind of compression utility to pack both files into a single zip file and then upload it to the Blackboard site. This’ll make it faster for you to upload the files to the site (especially if you’re using a dial-up connection), as well as make it easier for me to manage your assignments better. You’ll can find several free compression utilities on the web. A good source is the Tucows website (http://www.tucows.com/comp95_default.html).

END TERM: The final exam will be during exam week. It is likely to be a closed book exam (however, this may change and I may permit some kind of books/notes into the exam). Tentative schedule for the final exam is provided at the end of the course schedule. The final exam time is scheduled university wide by NYU, so it is unlikely to be changed.

ATTENDANCE AND CLASS PARTICIPATION: Class attendance in this course is not mandatory, but is strongly encouraged. Since most of the class materials and discussions will not be directly based on the corresponding material from the textbooks, students who do not attend classes are likely to have considerable difficulty on their homeworks as well as their finals. Simply reading the relevant chapters in the textbooks or talking to their colleagues will typically be insufficient preparation. Students attending class should do their best to arrive on time and stay for the entire duration of the class.

5% of the course grade is based on your participation in the classroom. Therefore you are strongly encouraged to participate constructively in the classroom. For class interaction, it is the quality rather than the quantity of the participation that matters. If you raise constructive and pertinent issues, you can get good class participation grades even for relatively modest participation. However, if your questions/comments are repetitive or unrelated, even a large amount of airtime will not help and will probably cause resentment among your colleagues. Postings to the online discussion boards will generally not be counted towards class participation.

HONOR CODE. Consistent with Stern’s honor code, your exam, your assignments and projects must be solely your own work (unless an assignment is specifically designated a group project, in which case it must be solely the work of the group members whose names appear on the assignment). Any attempt to represent the work of others as your own will be considered plagiarism and will be referred to the Stern discipline committee. Penalties for academic offenses such as plagiarism and cheating can include failure for the course, academic probation, and expulsion.

INSTRUCTION METHOD

This course is very hands-on and will mostly be taught through practical managerial examples drawn from the functional areas of finance and marketing. The main emphasis will be on developing intuition and understanding the when and how of using several modeling techniques – both analytical and data-driven. In-class demonstrations will be used to show how the models are formulated and solved. In addition, as part of your homework assignments and projects, you will get to work on a number of real-world datasets that will be used to solve practical business problems. The following provides a sample of the applications that will be covered in this course.
Finance:

- Break-even analysis
- Valuing European options
- Capital budgeting
- Portfolio analysis
- Calculating cost of capital
- Estimating stock betas
- Calculating stock returns
- Property valuation
- Credit scoring
- Clustering securities

Marketing:

- Sales analysis
- Pricing
- Advertising response
- Product mix decisions
- Customer segmentation
- Market basket analysis
- Predicting purchases
- Direct marketing

**TENTATIVE COURSE SCHEDULE:**

<table>
<thead>
<tr>
<th>Session</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PART 1: MODELING AND DATA ANALYSIS USING SPREADSHEETS</strong></td>
<td></td>
</tr>
</tbody>
</table>
| S1: Tuesday, Jan 20 | Introduction to modeling:  
Models, modeling process, spreadsheet models |
| S2: Thursday, Jan 22 | Sensitivity analysis  
Spinners, data tables, scenarios |
| S3: Tuesday, Jan 27 | Advanced Excel functions |
| S4: Thursday, Jan 29 | Regression analysis |
| S5: Tuesday, Feb 3  | Introduction to optimization  
Objectives & constraints, graphical analysis, Excel Solver |
| S6: Thursday, Feb 5 | Optimization: Linear and non-linear models |
| S7: Tuesday, Feb 10 | Data management:  
Summarizing and querying in Excel |
| S8: Thursday, Feb 12 | Data management:  
Business Intelligence, connecting to external databases |
| S9: Tuesday, Feb 17 | Extending Excel’s functionality:  
Using macros for automation |
| S10: Thursday, Feb 19 | Introduction to VBA:  
Excel object model, Visual Basic editor |
| S11: Tuesday, Feb 24 | Introduction to VBA:  
Branching and looping, debugging |
| S12: Thursday, Feb 26 | Personalizing Excel through user-defined functions |
### PART 2: DATA MINING USING SPREADSHEETS

<table>
<thead>
<tr>
<th>Date</th>
<th>Day</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>S13: Mar 2</td>
<td>Tuesday</td>
<td>Introduction to data mining: Basic terminology and concepts</td>
</tr>
<tr>
<td>S14: Mar 4</td>
<td>Thursday</td>
<td>Introduction to data mining: The data mining process</td>
</tr>
<tr>
<td>S15: Mar 9</td>
<td>Tuesday</td>
<td>Classification: Nearest neighbor methods</td>
</tr>
<tr>
<td>S16: Mar 11</td>
<td>Thursday</td>
<td>Classification: Logistic regression</td>
</tr>
<tr>
<td><strong>Spring Break</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S17: Mar 23</td>
<td>Tuesday</td>
<td>Model evaluation</td>
</tr>
<tr>
<td>S18: Mar 25</td>
<td>Thursday</td>
<td>Model evaluation</td>
</tr>
<tr>
<td>S19: Mar 30</td>
<td>Tuesday</td>
<td>Classification: Decision trees</td>
</tr>
<tr>
<td>S20: Apr 1</td>
<td>Thursday</td>
<td>Classification: Decision trees</td>
</tr>
<tr>
<td>S21: Apr 6</td>
<td>Tuesday</td>
<td>Classification: Naïve Bayes</td>
</tr>
<tr>
<td>S22: Apr 8</td>
<td>Thursday</td>
<td>Classification: Artificial neural networks</td>
</tr>
<tr>
<td>S23: Apr 13</td>
<td>Tuesday</td>
<td>Classification: Artificial neural networks</td>
</tr>
<tr>
<td>S24: Apr 15</td>
<td>Thursday</td>
<td>Prediction: Nearest neighbor, regression trees</td>
</tr>
<tr>
<td>S25: Apr 20</td>
<td>Tuesday</td>
<td>Affinity analysis: Association rules</td>
</tr>
<tr>
<td>S26: Apr 22</td>
<td>Thursday</td>
<td>Clustering: K-means and hierarchical</td>
</tr>
<tr>
<td>S27: Apr 27</td>
<td>Tuesday</td>
<td>Student project presentations</td>
</tr>
<tr>
<td>S28: Apr 29</td>
<td>Thursday</td>
<td>Final review</td>
</tr>
</tbody>
</table>

*End Term Exam: Thursday May 5, 2004 8 a.m. to 9:50 a.m. (Tentative)*