Data Analytics Experiential Learning

Teresa Nickeson, faculty

Nicholas Rivera, undergraduate student

CS / Math Department

University of Dubuque

2000 University Ave.

Dubuque, IA 52001

TNickeson@dbq.edu

NRivera@dbq.edu

Abstract

The January-term (J-term) 2014 class titled, "Data, Data Everywhere!" was a step out of the classroom and into industry.

Students learned how to use "real world" Business Intelligence (BI) tools from industry experts. Throughout the course, students worked with real data. Specifically, they used public bus route data supplied by the City of Dubuque.

Due to the short duration of the course, students were able to use several different software packages at no cost. That is, they used 21-day and 30-day trial software when educational-use versions of the software was not available.

Instead of textbooks, trade books were used exposing students to professional learning beyond textbooks. These books, and topical professional articles, were fodder for class discussions.

In addition to the instructor perspective, one of the students will present the student side of what was effective, and what was not, about this class format.

Introduction

The University of Dubuque (UD) is a small Presbyterian university located in Dubuque, IA. UD offers several majors including four in I.T. and one in Mathematics.

At the University of Dubuque, J-term (January) classes are offered between the fall and spring semesters. J-term is an intense 13 days of concentration on a single three-credit class. The intention is that these classes be experiential learning opportunities for students. This is when many travel classes and experiential electives for majors are offered.

For J-term 2014, Professor Teresa Nickeson taught a data analytics class titled, "Data, Data, Everywhere!" This class was a step out of the classroom and into industry. Nicholas (Nick) Rivera, an undergraduate Computer Information Technology (CIT) and Mathematics double major, was a student in the class. Prof. Nickeson will talk about the methodologies used to make the class experiential. Nick will talk about the student perspective of learning in this environment.

The Class Format

Investigation

The first day of class was spent learning about the data to be used for class, Dubuque's public bus route data. A representative from the organization, Jule, came to class and talked about how the data is collected and used. Then the class hopped on a public bus and rode from the university to the transfer location and on to the Jule office. Students observed the ridership data collection devices and processes in action. They learned how to read the route maps. At the Jule office, students observed clerks in action and toured the bus area. They were then able to put the data into context.

Analyzing the Data

There were three main areas of concentration related to analyzing the data; basic statistics, advanced Excel and Access, and modern Business Intelligence (BI) tools.

Basic Statistics

As the only pre-requisite for the course was Intro to CIS, some students had not yet had a college statistics course. As an understanding of basic statistics is important when analyzing data, a statistician from a local bank was invited to spend two days in the classroom teaching basic statistics, complete with real-world examples, and teaching how to use Excel to calculate the statistics. Students were fascinated to learn how he used statistics to determine good investments for the bank.

Related to the bus data, students completed two assignments. The first assignment required that everyone determine basic statistical values using Excel's statistics functionality. The second assignment required that students come up with an external factor and determine if that variable had statistical relevance. Many related weather, or school terms, to ridership. Here is a student example of the second assignment:

Advanced Excel and Access

Students used a 21-day trial version of McGraw Hill's SimNet software to learn advanced Excel and Access functionality. It was important that students learn what could be accomplished using these common desktop tools because some organizations, like the public bus organization and other non-profits, do not have funding for more sophisticated tools.

A representative from IBM Dubuque's Data Analytics team came to class and showed students even more Excel data analysis tricks including how to use Excel's Data Analysis add-in. The IBM office in Dubuque hires many UD students, so this was a great opportunity for students to learn more about the company and one of the job opportunities there.

Here is an example of one student's Advanced Excel work:

Count of Sum Passenger Boards — Day of the Week

Route	Month	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Grand Total
⊟ AM Commuter	June			1		1			2
	July		2	1	5	1	5		14
	August		4	4	4	4	6		22
	September		5	6	7	5	4		27
	October		3	5	5	4	6		23
	November		8	5	5	5	6		29
	December		4	7	8	8	6		33
AM Commuter Total			26	29	34	28	33		150
⊟ Business Shuttle	June		1	2			3		6
	July		1	3	5	2	1		12
Business Shuttle Total			2	5	5	2	4		18
⊟ Green 32nd	June		9	11	9	11	8	8	56
	July		10	11	11	11	11	8	62
	August		10	11	11	11	12	10	65
	September		11	11	11	11	11	11	66
	October		11	12	11	9	11	8	62
	November		10	10	10	9	12	6	57
	December		11	10	10	10	11	7	59
Green 32nd Total			72	76	73	72	76	58	427
⊟ Green 32nd Street	March		1	1					2
	April		10	11	10	11	11	9	62
	May		9	11	10	11	11	8	60
	June		5	7	8	9	1	2	32
Green 32nd Street Total			25	30					156

Commercial Business Intelligence (BI) Tools

Students learned how to use several Business Intelligence (BI) tools by having industry individuals come to class and demonstrate these products then talk about how they were used in the "real world". Like company employees would do, students learned the tool basics by going through interactive product tutorials on the Web.

McGladrey introduced students to Gartner Magic Quadrants. Most of the tools to which students were introduced during this course sit in the upper-right hand corner of their quadrant.

The primary tool used was *QlikView*. A data analytics guru from McGladrey conducted sessions on two different days. First, he built on what the Web tutorials taught and showed the students software tips and tricks. The second session was a week later at which time students presented dashboards they created. The McGladrey representative talked to each student about what s/he could do to improve their work. He ended this session showing some real-world dashboards he had created. Here is a student's QlikView dashboard:
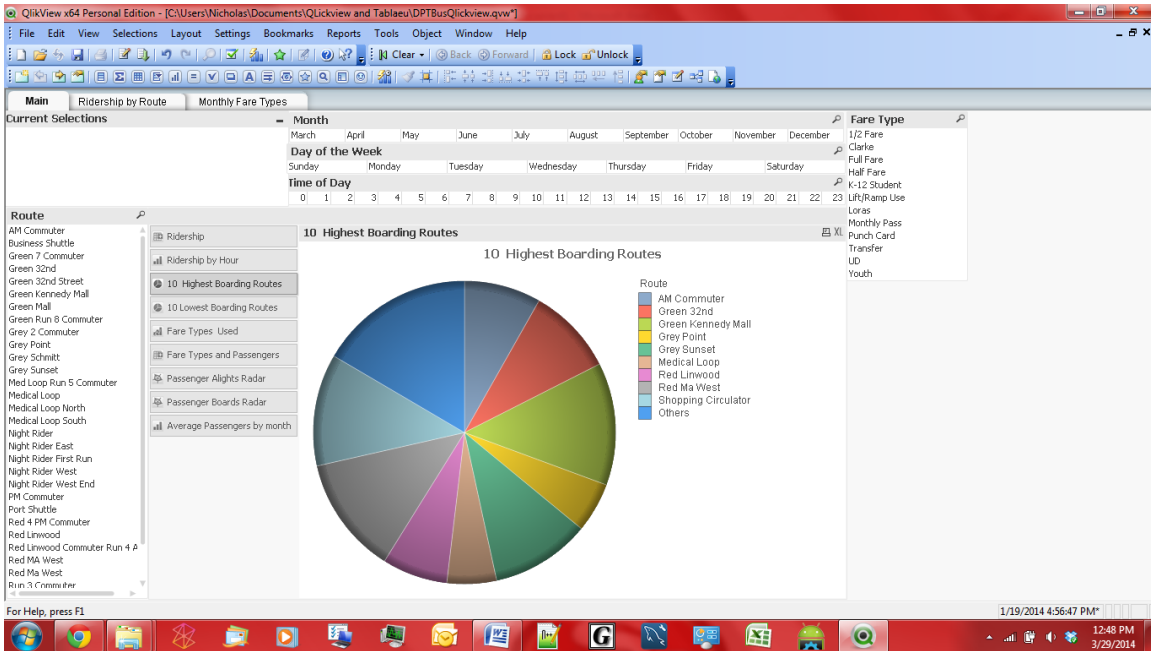
Representatives from John Deere worked with the students to create *Tableau* dashboards. Here is a student's Tableau dashboard:
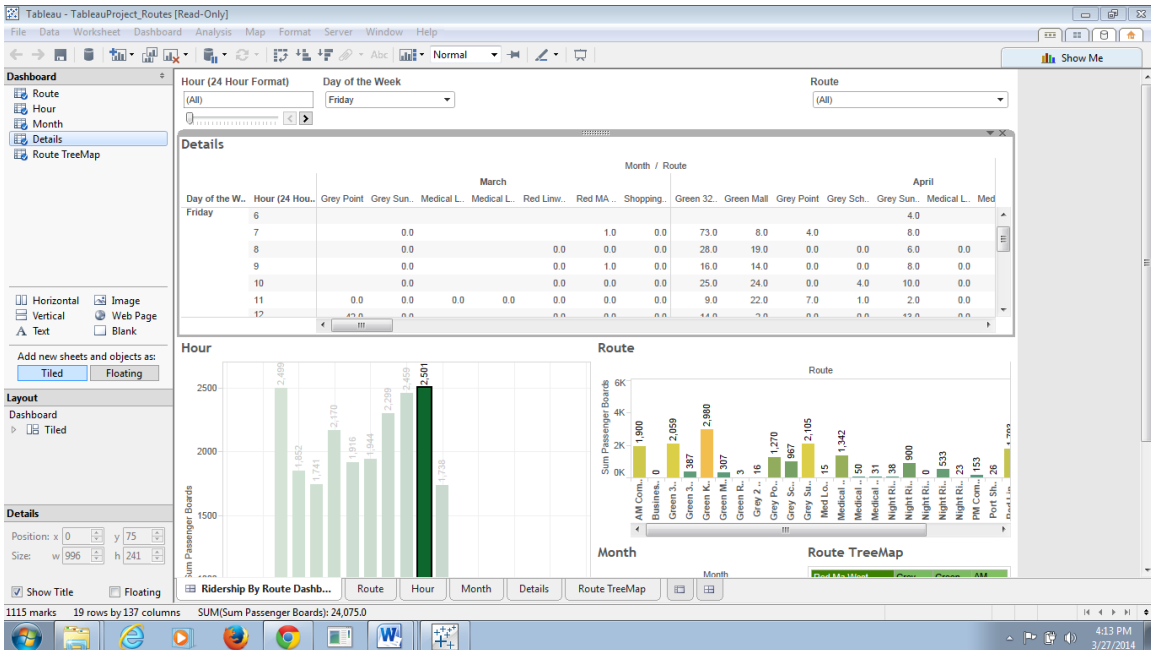
Like they would do on the job, students prepared a comparative analysis between QlikView and Tableau.

The last industry tool with which students got their hands dirty was GIS software, *ERSI*. First, one of the UD Science professors showed maps the Science department created for their research.

The person who uses ERSI for the City of Dubuque showed work she had done with the transportation department. She also conducted an in-class workshop. Here is a student's ESRI work:
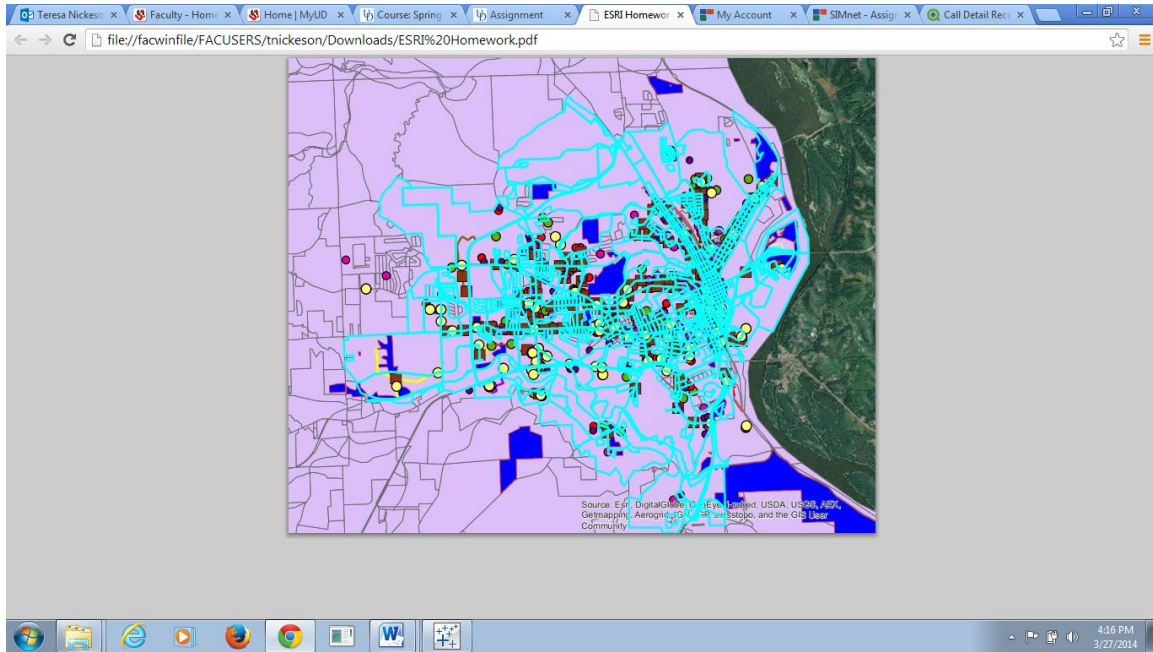
Students were exposed to *R* and *SPSS* as well. The prof from the Science department showed how R was used for statistical analysis of their department's research. The Dir. of Institutional Research for UD talked about how SPSS is used to analyze National Survey of Student Engagement (NSSE) data; data with which students could easily relate.

Student work was evaluated by the industry professionals. Students needed to ask questions of these people rather than relying on their instructor. That took them out of their comfort zone, but feedback from these people was quite valuable. And on the job, even if co-workers don't formally evaluate you, they often assess your work and ability. You don't just perform for your supervisor.

The Data

Throughout the course, students worked with public bus route data supplied by the City of Dubuque. Working with the same data in all of the tools helped students compare BI tools. It also was convenient because most companies with which we worked had confidential customer data that could not be shared with students.

This real data was much different from the pristine data textbook publishers usually provide students. First, before working with the data, students needed to clean up and flatten the data. Second, the data in the two data sets did not have a field in common. Students learned first-hand the importance of a field-in-common when they tried to join tables into which these data sets were loaded. They had been told this in their Microsoft Office course, but learning something like this from experience is so much more effective.

The city had switched to new software and practices during the middle of 2013. The data collected before the conversion was not complete or accurate. Because the switch was made midyear, there was not much data to analyze. But, for the purposes of this class, we made it work.

Here are excerpts from the data sets used:

Month: October

Fare Type	00:00	01:00	02:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	Total
1/2 Fare	0	0	0	0	23	12	40	47	54	55	33	29	34	24	25	6	2	0	0	0	0	0	384
Clarke (Free)	0	0	0	0	0	1	2	2	3	10	1	5	4	5	2	2	0	0	0	0	0	0	37
Full Fare	0	0	0	1	53	44	84	78	110	65	62	65	70	52	57	16	1	0	0	0	0	0	758
K-12 Student (Free)	0	0	0	0	28	15	42	8	12	7	13	18	65	97	44	7	0	0	0	0	0	0	356
Lift/Ramp Use	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	4
Loras (Free)	1	0	0	0	0	1	0	4	2	2	4	4	0	1	1	0	0	0	0	0	0	0	20
Punch Card	0	1	0	0	3	2	12	11	19	22	13	16	13	7	4	2	0	0	0	0	0	0	125
Transfer	0	0	0	0	6	30	21	24	13	48	25	53	56	52	49	14	0	0	0	0	0	0	391
UD (Free)	1	2	0	0	1	10	19	11	20	16	14	17	25	13	12	4	0	0	0	1	1	0	167
Youth (K-12)	0	0	0	0	48	97	14	1	6	1	1	2	37	13	1	0	0	0	0	0	0	0	221
Total	2	3	0	1	162	212	234	186	240	227	166	209	304	264	196	52	3	0	0	1	1	0	2,463

Month: September

Fare Type	00:00	01:00	02:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	Total
1/2 Fare	0	0	0	0	40	35	57	67	103	110	59	85	68	51	32	20	0	1	1	1	0	0	730
Clarke (Free)	0	0	0	0	0	1	0	0	0	5	0	7	3	3	5	0	0	0	0	0	0	0	24
Full Fare	0	0	0	1	92	68	174	133	169	140	83	108	141	120	114	42	4	0	2	1	13	0	1,405
K-12 Student (Free)	0	0	0	0	21	116	89	10	33	21	11	44	119	159	53	35	0	0	4	1	0	0	716
Loras (Free)	0	0	0	0	0	0	1	0	0	1	2	0	3	2	1	2	0	0	0	0	0	0	12
Punch Card	0	0	0	1	2	5	10	18	18	31	16	21	35	19	13	6	0	1	0	0	0	0	196
Transfer	0	0	0	0	24	48	34	44	51	84	20	63	65	35	38	35	0	0	0	0	0	0	541
UD (Free)	0	0	0	0	0	9	3	7	8	4	1	4	3	14	15	2	0	0	0	0	0	0	70
Youth (K-12)	0	0	0	0	70	173	52	4	14	9	6	11	305	17	15	1	0	0	0	0	0	0	677
Total	0	0	0	2	249	455	420	283	396	405	198	343	742	420	286	143	4	1	8	3	13	0	4,371

Ridership by Route

Provides the number of boards and alights for each route over the time period specified by the level of the drill down. Version 46.60.20120521.30955

Month	Day of the Week	Hour (24 Hour Format)	Route	Sum Passenger Boards	Sum Passenger Alights
July	Wednesday	8	AM Commuter	18	11
July	Monday	7	AM Commuter	0	0
October	Monday	8	AM Commuter	63	69
October	Tuesday	7	AM Commuter	424	539
July	Wednesday	11	AM Commuter	16	12
October	Tuesday	13	AM Commuter	1	10
December	Wednesday	17	AM Commuter	0	8
September	Monday	9	AM Commuter	0	1
October	Wednesday	5	AM Commuter	1	2
October	Thursday	5	AM Commuter	3	3
December	Monday	7	AM Commuter	236	299
July	Friday	7	AM Commuter	5	3
July	Friday	10	AM Commuter	11	10
September	Monday	6	AM Commuter	154	80
November	Thursday	8	AM Commuter	106	71
December	Thursday	12	AM Commuter	13	15
August	Monday	6	AM Commuter	67	45
August	Friday	5	AM Commuter	2	1
November	Wednesday	7	AM Commuter	171	221
September	Friday	6	AM Commuter	148	75
August	Thursday	7	AM Commuter	77	127

The "Textbooks"

Instead of textbooks, two trade books were used exposing students to professional learning beyond textbooks. These books, and topical professional articles, were fodder for class discussions.

The first book was Enterprise Analytics: Optimize Performance, Process, and Decisions Through Big Data, Thomas Davenport, *International Institute of Analytics*, ISBN 0133039439, © 2013, FT Press, Cloth, 288 pp, Retail $49.99. This book grounded students in the kinds of questions and issues businesses face today when analyzing data. Topics included: application of analytics, traditional and online, technologies of analytics, production scales and in the cloud, the human side of analytics, and actual case studies.

Dealing with unstructured data in general, and combining it with structured data specifically, is a common data analytics problem today. The second book was Tapping into Unstructured Data: Integrating Unstructured Data and Textual Analytics into Business Intelligence, William H. Inmon & Anthony Nesavich, ISBN 0132360292, © 2008, Prentice Hall, Paper, 264 pp, Retail $49.99. This book helped students understand these issues.

The final project was for students working in pairs to go out and talk to a company, or a department within a company, about how they could benefit from, and use, structured data combined with unstructured data. They presented their analysis to the class. Students came up with some very interesting findings. For example, for a medical supply manufacturing company, combining customer contracts and emails with transactional data would allow them to "see what a customer has talked about in the past with the company and where they have gone since the first phone call." Another example was from a food service company. If they could combine catering contracts and customer emails with other food sales data, they could better plan for future events.

Articles used for class discussions included:

- "Welcome to the Programmable World", Bill Wasik, Wired, June 2013, pp. 140-147 & 180.
- "Murmurs of Earth", Bryan Gardiner, Wired, June 2013, pp. 149-155.
- "Mutants", Carl Zimmer, Wired, Feb. 2013, pp. 98-103 & 108-114.
- "Can Anesthesia Cause ADHD?", Alice Park, Time Magazine, Feb. 13, 2012, p. 18.
- "Barack Obama's Big Data Won the US Election", Mike Lynch, Computerworld.com, Nov. 13, 2012.
- "Secret of Obama Victory? Rerun Watchers, for One Thing", Jim Rutenberg, NYTimes.com, Nov. 12, 2012.
- "Facebook's Graph Search Tool Causes Increasing Privacy Concerns", Juliette Garside, TheGuardian.com, Jan. 23, 2013.

Student Feedback

In addition to the instructor perspective, this paper will be co-presented by one of the students. He will present the student side of what was effective, and not so effective, about this class format.

The Good

Field Trip

On the first day, students took a field trip using the City of Dubuque bus route system, the Jule. Students reported that this trip was very beneficial because it helped the students understand the bus route system and the data capture mechanisms. The data provided by Jule was the data used throughout the class.

SimNet to Learn Advanced Excel and Access

Student feedback indicated that SimNet was a good program and successful in teaching the content. Students had never before used software to learn software. Students learned by completing interactive examples and exercises that included videos, "Guide Me's", "Try Me's and even practice tests. The tests included interactive practical work instead of just traditional objective test questions. Most students would like to use SimNet again in class.

Instruction from Industry Experts

A big difference between the format of this class and traditional college classes was having different organizations help teach the class. One student commented, "I enjoy classes that bring in outside representatives because it makes me feel like the teacher really cares and wants us to learn from individuals in the industry. I think these make the best classes."

Student feedback indicated that the representatives from McGladrey and John Deere were the most effective. The representative from McGladrey did a great job explaining and teaching. He began by covering features that went beyond what the tutorials had provided. He stressed that students needed to understand what the dashboards created really portrayed. Students enjoyed learning how to use Qlikview to make the data visual.

The John Deere representative did a particularly good job keeping the students engaged while he was teaching about Tableau by asking questions and making sure that students

understood what he was doing. Students thought Tableau was a great program because it very easy to use and learn due to its drag and drop interface.

Books and Articles

The favored book was <u>Enterprise Analytics: Optimize Performance, Process, and Decisions Through Big Data</u>, by Thomas Davenport. The book had great information about data analytics concepts, but the section that the students liked most was the case studies. The case studies had actual examples of the uses and implementation of data analytics in large organizations.

As for the articles, students found that they were easier to understand than the books. Students also thought that the articles were more interesting.

The Bad

SimNet

There was some frustration that students were working with some brand new SimNet content that still needed refinement. For one thing, some of the test answers were wrong. For another, some tasks could only be completed using the one path the software allowed. Lastly, sometimes SimNet prematurely accepted a wrong path as the answer without allowing the student to complete the task the right way.

Instruction from Industry Experts

There was one industry representative that was particularly ineffective. While students found her tool, ESRI, interesting and useful based on examples the representative presented, the representative did not do a good job of explaining and teaching. This may be the biggest downside of this class format, the risk of having people from industry who are not good teachers. For example, this representative was not good at noticing when students did not understand what she was saying or when they became

disengaged. Most students did not understand of how to use ESRI, so they struggled with the assignment. Students also had problems with ESRI because of its frequent crashing.

Summary

There were many facets of this experiential J-term course that took students on a trip to the "real world". They got their hands dirty learning how to use some of the most popular tools in industry from people who use them every day. They learned about the most topical issues, and some powerful results, of data analytics from their reading. The course definitely generated more student interest in the lucrative field of data analytics!