Improving Student Employability with Python and SQL

Hanbo Yu(✉)

Department of Administrative Sciences, Boston University Metropolitan College, Boston, MA 02215, USA
yuhanbo@bu.edu

Abstract. This paper introduces the author's educational approaches and achievements related to analytical and technical skills required by companies during the hiring processes of graduates in MS in business analytics. The author mainly focuses on Python and SQL, the core competencies that many employers currently value significantly. In this paper, the author shares his experiences on curricula development and execution of an elective course, AD 599 Introduction to Python and SQL for Business Analytics at Boston University, such as introducing data structures and algorithms to students and designing standard operating procedures for data projects, improving the employability for students by preparing students for technical job interviews such as job description cracking analysis and real-life data challenge bank building, lessons learned, and prospects of future.

Keywords: Python · SQL · Algorithms · Data structure · Standard Operating Procedure · Employability improvement

1 Introduction to Course AD599

1.1 Introduction to Boston University MS-ABA Program

The Master of Science in Applied Business Analytics (from now on MS-ABA) at Boston University's Metropolitan College (MET) is designed for business analysts, data scientists, data analysts, or others looking for transitions into data-driven roles such as management consultants. It can help people harness the concepts, techniques, and tools to transform available data into business insights or strategies [1]. The MS-ABA ranked No. 1 in Best Business Analytics & Intelligence Programs of 2021, No. 5 in Best Online Master's Business Analytics of 2022, and No. 10 in Best Online Master's in Business Programs (Excluding MBA program) [2]. In recent years, graduates from the MS-ABA program joined famous companies such as Google, Boston Consulting Group, Amazon, Wayfair, Tencent, L'Oréal, P&G, etc., and have made significant achievements and accomplishments in their professional areas.

The MS-ABA program has trained and prepared its students well in their data analytics skills and job searches. The author believes that MS-ABA will perform better in the future as new developments of new courses are coming in.

T. Zlateva and R. Goleva (Eds.): CSECS 2022, LNICST 450, pp. 289–301, 2022.
https://doi.org/10.1007/978-3-031-17292-2_24

1.2 Introduction to AD599 Introduction to Python and SQL for Business Analytics

The course AD599 Introduction to Python and SQL for Business Analytics (referred to as AD599) primarily serves students in the MS-ABA program. Many employers require specific Python and SQL experiences and skill levels in their data analyst/data scientist/business analyst positions. However, not all students in the MS-ABA program have former experiences in Python or SQL. This course, AD599, can give students a preliminary idea about Python and SQL, which can be a good foundation for students to dig deeper in these areas.

The pilot offering of AD599 was in the Summer 2 semester of 2021. Based on students' feedback and recommendations by the initial teaching team, the author reviewed and revised the content. The administrative sciences department approved the content revisions and asked the author to teach the course in the Spring semester of 2022, face-to-face and online.

So far, 70 students have completed this course AD599. Overall, the students' feedback is positive, stating they have learned valuable knowledge, techniques, and skills in Python and SQL needed for managerial decision-making and business analysis applications.

2 Course Content Upgrade and Development

2.1 The First Version of the Course Content Upgrade in November 2021 and Preparation for Online and On-Campus Teaching of AD599 Spring 2022

The author started his job search endeavors before his graduation from the MS-ABA program in January 2021. During his job search journey, the author noticed that Python and SQL working experiences are valued a lot by most employers who were hiring in the labor market. The author did some statistics, randomly picked 500 data analyst/data scientist job descriptions from LinkedIn [3], and found that 77.6% of these jobs required Python knowledge, 85.8% of these jobs required SQL knowledge, and 54.4% jobs required both.

The author then realized that Python and SQL are a crucial part of the core competency of being a data analyst/data scientist and business analyst. As he entered more and more rounds of interviews with various companies, he noticed another requirement for the Python application part. Most data scientist/data analyst jobs do not involve algorithms and data structures in the professional field or in daily life. However, companies still test them, mainly in online assessment, on platforms such as Hackerrank[4] and Coderpad[5]. These companies also leave "data challenges" or "take-home data projects" to candidates, which are common issues that data analysts/data scientists need to solve in daily life.

For the SQL part, the author also noticed that the basic SQL techniques he learned from school were not enough to solve some interview questions. During the author's learning process at Boston University, he did not use one convenient technique in SQL, Common Table Expression (referred to as CTE), and window functions. However, in

the interviews with different companies, he noticed that these were necessary to stand out from all candidates. Therefore, after extensive self-training, he successfully gained these skills and knowledge and finally obtained two job offers, one from a local company in Boston and another from a consulting firm on Wall Street.

The original version of the course contents included twelve lectures in six modules, four individual assignments, four quizzes, five discussion topics, and one final project.

During the initial review of the original course contents of AD599, the author discovered the need to introduce algorithms and data structures, data projects, or CTE and window functions. To improve the employability of students who will take this course, the author included the new content in the revised course materials for the Spring 2022 offerings of AD599.

The author added the following contents to the Python part of lectures in AD599 during the first version of the update and development of course contents: data structures such as Stack, Queue, Linked List, Hash Table, and Trees; Basic Algorithms and related concepts, such as Time Complexity, Linear Search, Binary Search, Bubble Sort, and realization of data structures[6].

The author also added a new part of the data modeling section to the Python-related lectures in AD599. This data modeling section included Introductions to regressions, basic rules and coefficients of regressions, realizations of regression models in Python, and regression model selection by cross-validation score.

The author added CTE and window functions to the SQL part of the lectures. This part included the syntax and use case of CTE and window functions such as Partition By, Over, Order By, Rows Between, Row Number, Rank, Dense Rank, Lead, Lag, N Tile, Count Percentile, etc.

With such significant changes in the content of the lectures, the author also adjusted the assignments, discussion topics, quizzes, and the final project to guarantee that all deliverables from students comply with the lecture notes. In addition to a content redesign of the third individual assignment, a brand new two parts final data project is offered. It is similar to the take-home challenge in real-life technical skills interviews. CTE and window functions are part of the final project. The test banks of all quizzes and questions related to characteristics of data structures and algorithms were included as well. The content of the discussion topics is harmonized with the in-class learning of the students. The new version of the study guide can be found in the Appendix section at the end of the article.

The student feedback at the end of the online delivery of AD599 Spring 2022 is positive - students appreciate the changes and the possibilities that the changes bring.

2.2 The Second Version of the Course Content Upgrade in April 2022 and Preparation for Online and On-Campus Teaching of AD599 Summer 2022

During his teaching process, as the spring semester of 2022 comes to an end and related information and feedback from students are already available, the author noticed that another revision of the course content is possible and submitted for an internal review by the ABA faculty a proposal to upgrade the course contents of AD599 before Summer 2022.

In this version of the course content upgrade, the author added more content related to Algorithms into the Python part of the lectures, such as the QuickSort algorithm, which has a lower time complexity than common sorting algorithms; he also added practical SQL techniques such as Limit, Offset, Between, Like, Union All, Case When clauses, etc., to the SQL part of the lectures, which he believes could be helpful for students to solve SQL interview questions.

To match the lecture notes better, the author also redesigned individual assignments one, two, and four in this upgrade version. He also expanded the four quiz banks even more so more questions are included in the quizzes.

After the approval of this upgrade proposal, compared to the original version of course contents, 50% of lecture contents are upgraded, 100% of individual assignments are redesigned, 100% of quizzes are expanded, 100% of the final project is improved, and 40% of discussion topics are revised.

3 Course Content Execution

3.1 Interactive Teaching Approach

The author taught the course AD599 both on-campus and online. The contents of the on-campus class and online class are identical. The only difference is the teaching speed since the length of the on-campus class is fourteen weeks and of the online course seven weeks. This significant difference requires different teaching approaches.

The author delivered the lecture contents swiftly yet effectively in class for the online session. He also provided extra consultation sessions for students if they had any questions or problems. Though without in-class practices, the students still comprehended the lecture contents well since most of them could deliver satisfactory results in the assignments, quizzes, and the final project.

The author implemented an interactive teaching approach for the on-campus delivery mode. He walks among students and asks and answers their questions in the classroom. Before introducing any lecture topic, he would use a real-life example to intrigue the students. When the students started to get interested and pay attention, he could deliver the core spirits of the topics more efficiently and effectively.

For instance, when discussing a critical application of data analytics, Forecast, the author pivoted an angle with stock price forecast to emphasize the importance of accurate predictions. The author briefly introduced the origin of stocks, which can trace back to the age of sails[7]. Back in history, since people did not have the methods to make accurate forecasts, their stock exchange activities were complete guesses. In contrast, we have multiple approaches to making more accurate forecasts nowadays, and the stock exchange activities are much more reasonable and rational. The students happily embraced this brief introduction of relevant knowledge, it drew their attention, and the whole lecture went well.

The author also used detailed explanations when introducing new concepts. When he introduced the most significant difference between Stack and Queue, he used this skill to make things clear. The queue is known to follow the First-In-First-Out rule (referred to as the FIFO rule). Stack is known to follow the Last-In-First-Out rule

(referred to as the LIFO rule). To compare the FIFO and LIFO rules, the author made a series of calculations for the profit that an imaginary company could get within a fiscal year with different rules. The calculation results were precise, with different rules, and the company's profit within a fiscal year is also not the same. By showing this to students, they immediately realized the difference between rules and better understood the difference between Stack and Queue.

3.2 Teamwork Spirit Building

The author's top priority was always helping students master the necessary skills in Python and SQL, and when that purpose was fulfilled, he sought to help the students a bit more.

The author understands well that teamwork spirit is vital in professional fields. Therefore, the author asked the students to group up as four-person or five-person teams in class and work together. The author also understood that some people might not be able to work well together. In every class session, he gave students chances to swap their groups until they found people with whom they could work. The grouping process entirely depended on the students themselves; the author would not intervene unless some people were left out without any group.

Python and SQL require a lot of practice. Therefore, after all, groups were settled, the author would assign in-class exercises related to topics learned for students to work on as a group. The author would walk around the classroom, approach each group, and check on their progress multiple times for each practice question.

This approach built their teamwork spirit successfully since the final project of AD599 was also group work. Seven out of ten teams delivered results more than satisfactory. The author believes that this teamwork experience can help students in their future working environment and professional fields.

3.3 Standard Operating Procedure of Data Analytics

The author understands the importance of a Standard Operating Procedure (SOP). SOPs can specify working steps to standardize products and quality [8]. In Data Analytics, the products are the results of data analysis, such as insights and forecasts. With the assistance of an SOP, the quality of data analytics projects can be standardized and guaranteed.

In class, the author showed students the SOP of a full data analysis project of Bavarian Motor Works Vehicle Price Forecast. The whole project started with data retrieves and data import through data type determination and conversion, null value detection and replacement, data cleansing, variable selection, train-set, and test-set split, model fitness test with cross-validation scores, model selection, and model building, to the final step which was price forecast of an imaginary vehicle. The students worked in groups and tried to find out the meanings of each step and realized the significance of SOP, which resulted well in their individual assignment three outcomes.

3.4 Retrospective Analytics

It is widely known that most people can make mistakes; it is also not unusual that people make mistakes at the early stages of data analysis projects/consultation projects/data science research when they do not have enough information. The author knows and understands this fact; therefore, he took a proactive approach in teaching this course AD599.

The final project of AD599 is comprised of three parts, phase one, phase two, and the final report. The author emphasized multiple times in class that if students made mistakes in phase one, it was normal and acceptable since they had not learned all skills and techniques in the class; however, as they obtained more knowledge and information, they should look back from the start and check if they had used the wrong methods or made any mistakes. The author called this "retrospective analytics," which resulted well since seven teams out of ten delivered optimal results in Phase two and the final report.

The author believes that making mistakes is normal and not the end of the world. However, realizing one's mistakes and correcting them is the vital part. He shared this idea with students, and from the positive results, most students accepted and understood it.

4 Extra Curricula Activities for Improving Student Employability

4.1 Employability Consultation Service

Other than the attempt to improve students' employability by revising the course contents of AD599 and its proactive execution, the author volunteered and contributed as a consultant in Employability Consultation services organized by the MS-ABA program. The author interviewed multiple companies during his job search process and collected technical interview questions such as online assessments and take-home data challenges from big companies such as Tesla, LinkedIn, Lyft, Airbnb, BNY bank, etc. Instead of wasting these good resources, the author built a technical interview test bank based on his collected material and then shared it freely with students who needed employability consultation services.

The author hosted a consultation session in March 2022. Although not many students attended, the attendees all showed gratitude and appreciation afterward. As more and more students graduate and start their job search, it is imaginable that these students will share their interview questions with the author, and hence, the test bank will grow bigger and stronger and be able to help more students.

4.2 Personal Consultation Sessions for Students

Outside the employability consultation services, the author also provided personal consultation sessions for students from course AD599 on the evening of each Friday. During the sessions, the author offered his advice on how to modify resumes, how to

prepare for behavioral interviews, and how to succeed in technical screenings. He shared his own interview experiences with students and his takeaways and lessons learned. From the feedback of students, these sessions were more than helpful in their job search process.

4.3 Job Description Cracking Analysis

The author helped many students with their job search, among which there is an appropriate example for this article. The author has a teaching assistant who also possesses a master's degree in Data Analytics from another university and is currently seeking a job. This teaching assistant is a competitive individual, and he is interviewing with the located in Boston company New Balance for a Data Analyst, product costing position[9]. Since he did not have too many interview experiences, he asked the author for help.

The author analyzed thoroughly and carefully the job description[9] and found an appropriate angle to "crack" it. He told his teaching assistant that, since this was a product costing data analyst position, showing one's knowledge and understanding of the importance of cost control could make one easily stand out from candidates. There are many links between raw materials and final products on shelves in the modern manufacturing industry, and each link contains a cost. If such costs can be reduced and controlled, the manufacturer can have an advantage in the final sale price. They can either use this advantage to obtain more market share or publish it in their quarterly financial statements to boost stock prices. The teaching assistant took the author's advice and expressed his knowledge and comprehension in this matter. Soon, good news came that the teaching assistant is now in the third round of interviews and very close to an offer. The author feels happy and achieved in this event and will continue providing similar services to students.

5 Lessons Learned and Next Steps

So far, the author's attempts to improve student employability have returned positive results. Most students in the author's class expressed gratitude for this endeavor, which helped them with their academic studies and significantly improved their employability in interviews.

Moreover, the labor market is constantly evolving, and the author will continue improving and refining the course contents, including lectures, quizzes, assignments, etc., and building the technical interview test banks.

The author's effort to improve students' employability will not stop here. As more and more students graduate and get job offers, the precious data on employability will be available. The members of the Administrative Sciences Department of Boston University Metropolitan College, including the author, can gather the data and build a SQL-based database for further analysis. They can use methodologies of statistics, data mining, machine learning, natural language processing, etc., to analyze which key features play the most crucial part in whether a student is hired or not. This project possesses excellent potential and is worth a considerable amount of effort.

Appendix Study guide of AD599

Module 1 Study Guide and Deliverables	
Topics:	• Introduction to the Python language, installation, first programs, workspace options, and current applications. • Script programming, data types, variables, mathematics, list operations, and data structures (stack, queue, linked list, hash table, tree).
Readings:	Module 1 online content
Discussions:	Discussion 0: Introduce yourself
Class Time:	Monday, Jan 24 from 6:00 - 8:45 PM ET Monday, Jan 31 from 6:00 - 8:45 PM ET

Module 2 Study Guide and Deliverables	
Topics:	• Modules, flow control, conditional logic, Text and strings, formatting and printing, loops • Functions, tuples, lists, and dictionaries
Readings:	Module 2 online content
Discussions:	Discussion 1: • Initial post due by Thursday, Feb 17 at 11:59 PM ET. • Respond to at least two of your classmates' posts by Sun-

	day, Feb 20 at 11:59 PM ET.
Assignments:	Assignment 1, due by Sunday, Feb 20 at 11:59 PM ET
Assessments:	Quiz 1, available between Saturday, Feb 19 at 9:00 AM ET and Sunday, Feb 20 at 11:59 PM ET
Class Time:	Monday, Feb 7 from 6:00 - 8:45 PM ET Monday, Feb 14 from 6:00 - 8:45 PM ET

Module 3 Study Guide and Deliverables	
Topics:	• Advanced topics for functions, Introduction to Algorithms, Common Algorithms, Realization of Data Structures • Numpy/Scipy (functions and capabilities), Pandas.
Readings:	Module 3 online content
Discussions:	Discussion 2: • Initial post due by Thursday, Mar 3 at 11:59 PM ET. • Respond to at least two of your classmates' posts by Sunday, Mar 6 at 11:59 PM ET.
Assignments:	Assignment 2, due Sunday, Mar 6 at 11:59 PM ET.
Assessments:	Quiz 2, available between Saturday, Mar 5 at 9:00 AM ET and

	Sunday, Mar 6 at 11:59 PM ET
Class Time:	Tuesday, Feb 22 from 6:00 - 8:45 PM ET (President's day Holiday, no class on Monday) Monday, Feb 28 from 6:00 - 8:45 PM ET

Module 4 Study Guide and Deliverables	
Topics:	• Statsmodel and Sklearn • Regressions, Cross-Validation, and Model Selection
Readings:	Module 4 online content
Discussions:	Discussion 3: • Initial post due by Thursday, Mar 24 at 11:59 PM ET. • Respond to at least two of your classmates' posts by Sunday, Mar 27 at 11:59 PM ET.
Assignments:	Assignment 3, due by Sunday, Mar 27 at 11:59 PM ET.
Assessments:	Quiz 3, available between Saturday, Mar 26 at 9:00 AM ET and Sunday, Mar 27 at 11:59 PM ET
Class Time:	Monday, Mar 14 from 6:00 - 8:45 PM ET Monday, Mar 21 from 6:00 - 8:45 PM ET

Module 5 Study Guide and Deliverables

Topics:	• Database and SQL Introduction, basic syntax and environment, Creating and Populating Databases, basic queries and filtering • Intermediate queries and subqueries, sets and joins, Grouping and Aggregation, Views
Readings:	Module 5 online content
Discussions:	Discussion 4: • Initial post due by Thursday, Apr 7 at 11:59 PM ET. • Respond to at least two of your classmates' posts by Sunday, Apr 10 at 11:59 PM ET.
Assignments:	Assignment 4, due by Sunday, Apr 10 at 11:59 PM ET. Final Project, Phase 1, by Sunday, Apr 10 at 11:59 PM ET.
Class Time:	Monday, Mar 28 from 6:00 - 8:45 PM ET Monday, Apr 4 from 6:00 - 8:45 PM ET

Module 6 Study Guide and Deliverables	
Topics:	• CTE and Window functions • Integration of Python and SQL

Readings:	Module 6 online content
Discussions:	Discussion 5: • Initial post due by Thursday, Apr 21 at 11:59 PM ET. • Respond to at least two of your classmates' posts by Sunday, Apr 24 at 11:59 PM ET.
Assignments:	Final Project, Phase 2, by Sunday, Apr 24 at 11:59 PM ET.
Class Time:	Monday, Apr 11 from 6:00 - 8:45 PM ET Monday, Apr 18 from 6:00 - 8:45 PM ET

Final Presentation	
Topics:	• What is next, and overall wrap-up • Q & A
Assign-ments:	Final report, due Monday, May 2, at 11:59 PM ET Presentation, due Monday, May 2, at 11:59 PM ET
Class Time:	Monday, Apr 25 from 7:00 - 8:30 PM ET

References

1. MS in Applied Business Analytics. Master of Science in Applied Business Analytics|BU MET. https://www.bu.edu/met/degrees-certificates/ms-applied-business-analytics/. Accessed 8 May 2022
2. Best Online Masters in Business Analytics and Intelligence Programs by Online Masters Report 2020 & 2021 & 2022. https://www.bestcolleges.com/features/top-online-masters-in-business-intelligence-programs/
3. Data Analyst jobs in the United States - linkedin.com. https://www.linkedin.com/jobs/search/?currentJobId=3064176256&keywords=data%20analyst&refresh=true. Accessed 8 May 2022
4. Hackerrank. HackerRank. https://www.hackerrank.com/. Accessed 8 May 2022
5. Encrypted online notepad for secret and secure notes. encrypt and keep private notes with password.: Coded pad™. https://www.codedpad.com/. Accessed 8 May 2022
6. Cormen, T.H., et al.: Introduction to Algorithms, 3rd edn. MIT Press, Cambridge (2009)
7. Woodward, D.B.: A new guide to the public funds, or, every man his own stock-broker containing the origin of the funding system, causes of the fluctuation of the prices of stocks, manner of transferring stock
8. Pearce, O.: What are sops? and why does my organization need them? #StayConnected. https://blog.montrium.com/experts/what-are-sops-and-why-does-my-organization-need-them#:~:text=SOPs%20specify%20job%20steps%20that,failure%20or%20other%20facility%20damage. Accessed 8 May 2022
9. Data Analyst product costing in Lawrence, Massachusetts, United States of America: Administration at New Balance. New Balance. https://jobs.newbalance.com/global/en/job/R25113/Data-Analyst-Product-Costing. Accessed 8 May 2022