



Reflections on the Applied Business Analytics Student Writing Project

Greg Page^{1,2}(✉) and Huey Fern Tay^{1,2}

¹ Princeton University, Princeton, NJ 08544, USA

² Boston University, Boston, USA

{gpage, hftay}@bu.edu

Abstract. In this paper, the authors describe a student article writing project run by a Professor in Applied Business Analytics at Boston University’s Metropolitan College. Through this project, students write and self-publish articles on various topics related to data science. The project enables students to build and refine their data modeling skills, and to build a portfolio of content related to the data science field. The authors trace the history of the project, point out some of its notable successes and setbacks, and identify some thoughts about how the project should continue to evolve.

Keywords: Data science · Data mining · Data science education

1 Introduction

Beginning in the Fall of 2020, Greg Page, a Senior Lecturer in Applied Business Analytics (ABA) at Boston University’s Metropolitan College, queried several students in a Marketing Analytics class to ask whether any were interested in collaborating on articles related to data science. Three students from the section stepped forward, and expressed a desire to involve themselves in such a project.

Those three students – Jiawen Huang, Qiuyue Wang, and Abhishek Kumar – worked with Professor Page to release the first “wave” of articles. The students published each of these articles to Medium, a web-based, self-publishing platform. For the most part, these articles consisted of more in-depth analysis and refinement of topics introduced in ABA specialization courses.

Page’s idea to collaborate with ABA students on data science articles stemmed from an experience during the previous summer. In June and July of 2020, Page worked with Yiting Wang, a recent ABA graduate, on two data science articles that were posted to Medium. The first of these, written by Wang, explained the point scoring system in a video game called Animal Crossing [1]. The second of these, written by Page, was titled, “What’s the story with those dynamite sticks on my bar plot?” [2].

Seeing the positive experiences across the summer and fall of 2020, and the potential for future benefits as well, Page institutionalized this group in Spring 2021. He opened it to all members of the ABA community, dubbing it the ABA Writers’ Group. Since that

time, the group has met continuously, on a weekly basis, via Zoom. During the twice-annual ABA webinars, Page and the current group of student participants deliver a slide presentation to current and incoming ABA students. This webinar serves to update the ABA community about the group's efforts and progress, and to advertise the group to students who may be interested in participating.

The topics covered by this paper include:

- The student article writers' process, from idea generation to publishing;
- The structure and pacing of the weekly Zoom meetings;
- Examples of successful student outcomes;
- Challenges and setbacks associated with the project; and
- Conclusions and final thoughts from the authors

2 The Operational Concept, from Brainstorming to Publishing

During the weekly Zoom sessions, Professor Page cycles through the student participants, in order of their arrival time to the meeting. The "check-in" period offers the student a chance to offer updates about recent progress and/or to identify any recent roadblocks or sticking points. Once a student's check-in time is completed, the student is free to either stay in the meeting, to observe the other topics and ideas, or to depart. Throughout the meetings, Professor Page maintains the meeting notes via a Google Doc that is shared with all student participants.

Students are encouraged to generate their own article ideas – when students have done this, they are more likely to take ownership of their projects, and less likely to perceive it as something akin to a class assignment. When students attend for the first time, and are searching for a topic, Page first encourages them to think about their interests. For instance, a student passionate about movies could analyze a cinema revenue dataset, or a student with a keen interest in finance could analyze stock market returns. Kegui "Gary" Zhuo is one student who combined his strong passion for NBA basketball with some data manipulation skills by using the popular R package `dplyr` [3] for his analysis.

A second option for any student is to choose a single function or concept, and to demonstrate it using either Python or R. Shang Ding's "Cross-Validation Techniques: k-fold cross-validation vs Leave-One-Out Cross-Validation [4]" encapsulates this approach. This student had first gained exposure to the cross-validation concept while taking AD654: Marketing Analytics, an ABA specialization course. Prior to writing the article, however, she had not yet acquired firsthand experience with the technique known as leave one out cross-validation.

If a student wishes to participate but is completely at a loss for a domain area or a data science angle, Page suggests starting points for the student to explore.

Students have the option to write an article entirely independently, or with some level of direct involvement from Professor Page. A student choosing the former option could of course pick any topic to explore. In such a case, Professor Page's role would be purely 'behind-the-scenes', and the article's by-line would belong entirely to the student. For a student choosing the latter option, the byline would be split (by the student, with Professor Page).

It is essential that Professor Page clearly outlines the roles and expectations from the outset. For this project to work, it is essential that students “buy in” and take ownership of the exploration and article drafting. This is essential for two reasons: first, because to use an alternative, in which the professor was truly the one performing the coding, drafting, and conclusions, would not lay a genuine foundation for a student-led writing project; second, this would not be sustainable. A truly student-led writing project can scale up or down easily as the number of participants changes.

These stories are published primarily through Medium (a self-publishing platform), or through specialized sub-channels of Medium, such as Towards Data Science, Codex, and Nerd For Tech. Others are posted to R-Bloggers, a site that aggregates content related to the R programming language and statistical environment.

Students can write a single article on a single topic as a “one-off” sort of project, or could use a series of articles to explore a topic in more depth.

Huey Fern Tay, an online ABA student with a background in journalism and marketing communications, did both.

Tay wrote a series of articles on variable imputation after joining the project upon completion of ADR100 in Summer 2021. She started with an explanation of the Last Observation Carried Forward, setting the story within a fictional scenario involving the local weather in Melbourne during the Australian Open. She then wrote about imputation with a categorical average, followed by linear regression model imputation using the `simputation` [5] library in R, emphasizing the `impute_lm()` function which is used to replace missing values in a dataset with the results of a linear model, built with the values of known, related variables as inputs [6].

Those articles have reverberated around the data science community, helping to build a digital footprint around Tay as an authority on data science. Tay’s articles have been shared on Twitter by several data scientists, and by the Towards Data Science platform itself.

The final installment of the imputation series revolved around the perils of zero imputation as well as the occasions when the method is appropriate. This piece was subsequently cited in a Masters thesis by a graduate student from Utah State University.

3 Establishing a Regular, Predictable Rhythm – with “All Volunteers, No Hostages”

Although the meeting time block may change from semester to semester, once it is set, it remains firmly in place throughout the term. The maintenance of a regular, steady meeting time reduces the friction that would otherwise arise if meetings with individual student authors had to be scheduled separately. New student writers are often reminded that they can choose to attend these sessions as their schedules permit, rather than view the meeting time block as a fixed, firm obligation.

To reinforce this, Page even asks the students not to RSVP when they are unable to attend. Discouraging student RSVP emails is a way of reminding students that the sessions are meant to be loose, free-wheeling, and completely optional.

An important tenet of the ABA Writers’ Group is that all the members are there voluntarily. Students are regularly reminded that the writing project is not a class, or

even an extracurricular obligation. Professor Page often reminds the students that if the writing group begins to feel like a burden, or obligation, they should simply take time away from it, and return when their schedule is freer.

Attrition is fairly common – approximately 1 in 3 students who visit do not return for a second meeting. While this attrition rate may sound high at first, there are some positive ways to view it. Chief among these is that it shows that students understand the messaging around the project – those who lose interest in article writing do not stay affiliated out of a sense of obligation.

Meeting attendance fluctuates considerably, sometimes based on the natural rhythm of the academic semester. Meetings have included as many 12 separate ABA students and recent graduates, and have also been as small as just a single student meeting directly with Professor Page.

4 Student Success Vignettes

Students can benefit from participation in the writing project in several ways.

First, students will gain more expertise on the topic that they select for their article. Generating the code required to demonstrate a concept, and then articulating the steps and results in words, forces a student to carefully assess the concepts involved. This benefit can apply to both the domain knowledge involved, and to the data science concepts illustrated in the article.

In addition, student writers can expand on the coding and modeling skills that they learn in their ABA courses. Given the rapid pace of all ABA courses – whether on-campus or online – there is often not enough time in the semester to cover every single topic in considerable depth. A student who encounters a theme or concept of interest in class can supplement that knowledge with extra material through the writing project, such as Tay, whose piece about detecting multicollinearity was inspired by the final project in her AD699 Data Mining course.

As students explore the datasets, they train themselves to troubleshoot data quirks that need to be addressed before the start of an analysis (e.g. missing values, illogical values, the data format of multi-response surveys). The process of problem solving helps students build confidence around those decisions. Furthermore, students can begin to establish an article portfolio that may be of interest to future employers. The ‘Animal Crossing’ analysis written by Yiting Wang is a case in point. Her analysis involved relatively straightforward techniques such as ‘groupby’, ‘mean’, and ‘sort’ but her subject matter left an indelible impression on readers, including a member on the interview panel at Wayfair, a Boston-based home furnishings company. Yiting got the job.

Recent graduates of the program can use the article writing as a way to continue to bolster their data science skills while they are otherwise immersed in the job hunt. During the long, stressful, and often discouraging process that students encounter with the job search, students often experience intensifying feelings of frustration. Involvement with a project such as this one may not only alleviate some of those feelings, but could also offer them a topical, relevant answer to the question, “What have you been doing lately to sharpen your analytics skill set?”.

Student participants may also gain some networking benefit by participating in the weekly check-ins. Naturally, some conversational topics arise during these sessions,

beyond the immediate focus on the articles. These side topics sometimes include job interview experiences, including takeaways about employers' expectations, and the skills valued by potential employers. Students armed with this 'market intelligence' gain an edge in a competitive market by being better prepared for the job interview process. One of the student authors found that during interviews with Meta, and with a food delivery app service, several of the fundamental data handling questions that arose overlapped with other students' article topics.

Finally, the practice with writing and editing may be helpful for English language learners, especially as they prepare for job interviews and corporate recruiting, as the descriptive experiences assist them in structuring their thought process and explanatory skills.

5 Limiting the Scope

To help students narrow their scope, Page reminds students to consider publicly available datasets that can be found on data science websites such as Kaggle where datasets can be filtered according to analytics techniques or genre. He also reminds them to consider the abundance of data from the realms of sports, government, and finance – such information is often easily obtainable.

Sometimes, students begin with extremely ambitious ideas – they want to write something far-reaching that explains a topic in great depth. However, an effective article often succeeds due to its focus, rather than its breadth. Most data science blog posts are concise, and they serve a single, important purpose. They enable a student in a university course, or a working professional, to answer a question about how to accomplish some particular purpose. Students who wish to tackle a topic in considerable depth are encouraged to break their content up into multiple parts.

6 Challenges and Frustrations

Students new to analytics generally do not appreciate the benefit and value that can come with descriptive techniques. Often, when a student arrives at the Writers' Group meetings for the first time, and is in search of an idea, Professor Page will suggest an article that demonstrates a single concept, function, or type of data visualization.

An example of such an article is Xiangyu Wang's "Exploring the World of Hexagonal Bin Plots [7]." In this article, the author describes the problem that arises when scatterplots are built with enormous datasets – there is so much overplotting that the points become impossible to distinguish from one another. Hexagonal bin plots address this problem by dividing the graph into distinct regions, and then filling those regions based on the count of observations that fall within that area. Xiangyu goes through several modification options in the article, including adjustments to the hexagon sizes, color scheme alterations, and even the decision of whether denser areas should be filled with lighter or darker shades [8].

An article like the one described above is ideal for someone who initially attempts a scatter plot and wonders how to address severe overplotting that cannot be solved by transparency or jitter adjustments. It also might be a great find for a student in a

University course tasked with making a hexagonal bin plot, but not sure where to start, or which ‘levers’ to pull to achieve the desired outcome.

Even after seeing an article such as the one described here, students often resist the idea of writing about a specific type of visualization. Most likely, this is because such articles lack the “punch” of something like an advanced tree algorithm, neural network, or matrix decomposition.

Some students may approach the writing project with a degree of wariness; after all, if they are new to analytics, what can they contribute to the field? How would it provide unique insight or value? To address these questions, Page often reminds students to reframe the issue – rather than expect to write something groundbreaking and Nobel Prize-worthy from the outset, they can simply pick a single idea, and tackle it. At a bare minimum, the resulting article will help the student to better understand the chosen topic, and could be a “nice to have” item on a resume.

The language barrier may also present a challenge. Students may begin with a strong sense of motivation, but then run out of steam when the time comes to draft the article. To the best degree possible, Page can become involved when he senses that this is creating a barrier – he can ask the student to generate a first draft, and then to “pass the torch” to him for a review and a round of edits.

Ideally, the process flow would look something like this:

1. Professor Page works with the student to help shape the article concept and scope;
2. The student explores the topic and/or any related datasets independently;
3. The student checks in during the weekly sessions for guidance and/or the chance to “sanity check” any ideas or concepts being explored in the draft;
4. If the student reaches a complete impasse, Page can take the student’s existing code and narrative under review, for a round of edits;
5. The student makes a final round of reviews and edits, before sending the article back to Page for review;
6. After Page gives the code and written content a final “okay”, the student posts the article directly to a destination site.

7 Publicizing the Work

There is not a single “go-to” location for publishing the students’ work. When the project started, all of the articles were simply posted to Medium. Xiaotong “Claire” Ding was the first student to publish to r-bloggers, with a detailed description of `nearZeroVar()`, a function from R’s `caret` package [9]. Ginna Gomez was the first student to post to Towards Data Science, back in May of 2021. A specialized channel within Medium, Towards Data Science is focused specifically on its eponymous field. Articles submitted to Towards Data Science are screened by a team of editors prior to approval; the site, therefore, carries more prestige than Medium.

After an article is published, there is not a single standard used for publicizing the students’ work. When the project started, Professor Page shared all of the student articles in the ABA LinkedIn group page. This is an internal page, accessible only to members of the ABA community. Page has also shared some articles on his own LinkedIn page, helping gain more exposure for the group.

Professor Page has registered the domain GoodWaveData.com. This summer, he plans to link all of the existing published body of student work to this site. He intends to develop this site into a hub for topics related to learning, and teaching, data science.

In one case, a student author created a [website](#) to showcase her work.

Going forward, student authors can be encouraged to share their content through the ePortfolio platform offered by Digication. This platform offers students a standard, visually appealing way to tell their story and to share their work.

8 Conclusions

From its beginnings in Fall 2020, the writing project's meetings have been entirely virtual. With COVID-19 now in its endemic phase, and in-person meeting restrictions lifted, it may be worth experimenting with face-to-face sessions in addition to Zoom. No matter what, the Zoom format should be maintained – that is what enables remote learners to participate in the process.

Among the cases in which the articles had a significant, positive impact on a student's job search, the common theme among them is the popularity of the subject matter, rather than any specific analytics insights featured in the article. One of the best such examples is Lujia Wang's 2021 piece, "Data Insights from Reddit WallStreetBets" posts [10]. Wang released this post on March 12, 2021, when the hoopla surrounding the Reddit WallStreetBets forum was still fresh in the minds of many investors and stock market analysts. The article received considerable social media "buzz" in the form of shares, comments, and likes. Wang spoke to a class at the University of Washington about her findings, after the course professor saw her article on LinkedIn. She was asked about the article during her hiring process with Applied Materials, a semiconductor company based in San Jose, California, where she now works.

While the benefits that can accrue to participating students have been noted above, it should also be noted that Professor Page benefits from the experience, too. The topic development and data exploration bolster his subject matter knowledge and help him to develop and refine his instructional style. This semester, he embedded two of Tay's articles into assignment prompts for AD699. The writing project also enables Page to spot truly remarkable talent among members of the ABA community, which can help the department to identify future facilitators, adjunct instructors, and project partners. Recognizing the need for a marketing analytics textbook that balances substantive Python instruction, statistical rigor, and contemporary marketing concepts, Page reached out to Tay this spring to suggest a collaboration on this project. They are now co-writing a marketing analytics textbook, set for release in digital and print editions by Labor Day of this year.

References

1. Wang, Y.: Animal Crossing: Secret of Miles Rewards with Python Visualization. (24 Jun 2020). <https://medium.com/self-training-data-science-enthusiast/animal-crossing-secret-of-miles-rewards-with-python-visualization-a441aa6238f5>

2. Page, G.: What's the Story with those 'Dynamite Sticks' on my Barplot?. Medium (30 Jun 2020). https://medium.com/@greg.page_41455/whats-the-story-with-those-dynamite-sticks-on-my-bar-plot-c9820e4788e9
3. Wickham, H., François, R., Henry, L., Müller, K.: dplyr: A Grammar of Data Manipulation (2021). R package version 1.0.7. <https://CRAN.R-project.org/package=dplyr>
4. Ding, S.: Cross-Validation Techniques: k-fold cross-validation vs Leave-One-Out Cross-Validation. Medium (21 May 2021). <https://shangding.medium.com/cross-validation-techniques-k-fold-cross-validation-vs-leave-one-out-cross-validation-2dcfb7d5deb3>
5. van der Loo, M. Imputation: Simple Imputation (2021). R package version 0.2.7. <https://CRAN.R-project.org/package=imputation>
6. Tay, H.F.: Using R's imputation to handle missing values. Towards Data Science (25 Jul 2021). <https://towardsdatascience.com/how-much-money-did-we-make-again-using-imputation-to-balance-the-books-f8e5c7b74f51>
7. Wang, X.: Exploring the World of Hexagonal Bin Plots. Medium (15 Jan 2021). <https://xiaogyuw.medium.com/exploring-the-world-of-hexagonal-bin-plots-1562c65f3f94>
8. Wickham, H.: ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York (2016)
9. Ding, X.C.: Function with Special Talent from the caret package in R: NearZeroVar (04 Sep 2021). <https://www.r-bloggers.com/2021/09/function-with-special-talent-from-caret-package-in-r-nearzerovar/>
10. Wang, L.: Data Insights from Reddit's Wall Street Bets Posts. Medium (12 Mar 2021). <https://lujiaw.medium.com/data-insights-from-reddit-wallstreetbets-posts-8f95dbd59a6c>