

Easy as Py: A First Course in Python with a Taste of Data Analytics

Mark Frydenberg
mfrydenberg@bentley.edu

Jennifer Xu
jxu@bentley.edu

Computer Information Systems Department
Bentley University
Waltham, MA 02452

Abstract

Python is a popular, general purpose programming language that is gaining wide adoption in beginning programming courses. This paper describes the development and implementation of an introductory Python course at a business university open to students in a variety of majors and minors. Given the growing number of career opportunities in analytics, the instructors felt that including a module on Data Analytics would add relevance and interest in the course. A survey given at the end of the semester shows students found this topic to be relevant to their future uses of Python. The paper also discusses challenges in teaching a first programming course to students with varying levels of programming experience.

Keywords: Python, Data Analytics, programming languages, programming courses

1. INTRODUCTION

Python is a first programming language in many secondary schools and universities (Siegfried, Siegfried, & Alexandro, 2016), and industry has embraced it as well. A recent study (Cass, 2017) showed that Python was the #1 language among top US Computer Science departments for teaching introductory programming courses (Guo, 2014). Companies such as Google, Facebook, Instagram, Netflix, and Dropbox all make use of Python because of its simplicity and ease of deployment and code maintenance, and variety of available code libraries. These features allow developers to focus on writing code for their specialized applications. (Reynolds, 2018)

In addition to Python's popularity in creating applications for the Internet of Things (Frydenberg, 2017) and bioinformatics (Kortsarts, Morris, & Utell, Janine, 2010), recent years have seen a major increase in data science

and analytics applications implemented in Python.

Python in the Business Curriculum

While many engineering and information science departments offer Python in an engineering context, Python in the business curriculum serves Computer Information Systems (CIS) and Information Technology (IT) majors, minors, and students in other business programs by giving them the opportunity to learn a popular language that is widely used in business.

Appendix 1, Table 1 shows the list of the top 12 MIS programs (US News & World Report, 2018) and based on information in the course catalogs on their websites as of June 1, 2018, which schools require Python, and which offer it as an elective for Information Systems majors. Only MIT requires Python for business analytics majors, and only Indiana University Bloomington offers it as an elective. Many of these schools

have engineering, computer science, or information science departments that offer a Python elective, but their course design likely focuses on programming concepts and algorithms rather than business applications. As business universities increase offerings in data science, analytics, and related fields, an understanding of Python will be essential.

Bentley University, a business university in Massachusetts, offered one section of Python as an experimental elective through its CIS Department for the first time during the fall 2017 semester, and three sections taught by two instructors during the spring 2018 semester. The instructor teaching the two spring sections sat in on the fall course so that both instructors could have a common basis of experience when modifying and preparing the course for the spring semester. Graded assignments and exams were the same across all three sections in the spring.

This paper explores considerations, assignments, and results of the implementation of an introductory Python course in a business school curriculum as well as student reactions to learning the language. The course fulfilled an Arts and Science elective.

The following questions guided this study:

- What factors motivate students to take (and universities to offer) Python as part of the business curriculum?
- In addition to programming fundamentals, which topics and applications of Python are considered relevant in a business context?
- What value do business students receive from taking an introductory Python course?

2. DESIGNING A PYTHON COURSE FOR BUSINESS STUDENTS

Undergraduate CIS majors at Bentley University are required to take a semester course in Java programming. Adding another language as an elective provides additional opportunities for students to strengthen their development skills, which will open more employment opportunities. Offering Python as an elective for CIS minors provides similar benefits.

Because Python is a very popular and widely used language in data-intensive disciplines, the course is also beneficial to Mathematical Sciences majors and minors. The introductory Python course has no prerequisites other than IT 101, a required course on digital literacy, covering technology

concepts, Excel, and designing basic web pages with HTML, that is typically taken during the students' first year.

Topics

This course presented topics found in most introductory Python programming courses (McMaster, Sambasivam, Rague, & Wolthuis, 2017; Topi et al., 2010), including:

- Variables, Data Types, and Expressions
- Loops and Selection Statements
- Strings and text files
- Lists and Dictionaries
- Functions
- Classes and Objects

Several introductory Python textbooks (Downey, 2016; Lambert, 2018) also offer chapters presenting applications that include graphics processing and user interface development. In planning the spring 2018 course, the instructors recognized that teaching even the most basic applications of data analytics would be of interest to business students. To accommodate this additional topic, the instructors chose to replace the chapters on graphics processing and user interface development with a module introducing the basic data analytics capabilities of Python.

As this is an introductory course, the instructors omitted advanced topics such as higher order functions (map, reduce, filter, lambda), inheritance and polymorphism, even though they were covered in the course textbook (Lambert, 2018).

Course Structure

The Python course met for two 80-minute sessions each week during the Fall 2017 and Spring 2018 semesters. Each class session included instructor-led demonstrations, lectures or presentations, and often, short, in-class exercises that reinforced the topics being presented. Students relied on instructor office hours and assistance from tutors in the University's CIS learning center when they needed assistance.

The course assignments included a standardized midterm and final exam, comprised of multiple choice, trace the output, fill in the missing code statements, and coding questions. In addition to short coding problems completed in class which counted toward their class participation scores, students completed seven major programming assignments for homework, as summarized in Appendix 1, Table 2.

3. INTRODUCING DATA ANALYTICS IN A FIRST PYTHON COURSE

Given the widespread use and highly promoted applicability of Python to data analytics, the instructors added a unit on this topic during the Spring 2018 semester. Even though most introductory Python text books do not include this content, accomplishing basic data analytics tasks are within reach of beginning programming students.

To use the pandas (an acronym derived from "Python and data analysis"), matplotlib (2D plotting) and numpy (scientific computing and numerical analysis) libraries for data analytics in a Python application, an understanding of objects and collections is useful. Providing a taste of data analytics in an introductory Python course becomes possible, as it builds on earlier topics and makes the course content more relevant for students.

Basic Data Analytics Capabilities with pandas, matplotlib, and numpy

The instructors spent three class sessions on data analytics. Much of the first session was spent having students install pandas, matplotlib and numpy libraries using the Miniconda distribution of Python (Continuum Analytics, 2018) onto their laptops. Miniconda provides a minimal Python installation containing additional libraries and packages. Even with step-by-step instructions, some students found the install process to be cumbersome. Students running macOS were challenged by the need to run installation commands in a bash shell, as they were not familiar with using a text-based command line interface to interact with an operating system.

During the first and second classes on Data Analytics, the instructors demonstrated programs that use the pandas DataFrame. A DataFrame is a two-dimensional data structure with predefined methods to sort, filter, and rearrange columns, create pivot tables, and perform other calculations and operations. Examples included reading data from a file and storing it to a DataFrame, printing data with and without column headings (and contrasting how to accomplish the same task without using the pandas module), sorting data by one or more columns, finding maximum and minimum values. After having written for loops to iterate through lists and dictionaries and a Grid class (Lambert, 2018, p. 330) to process data earlier in the course, students found interacting with a pandas DataFrame to be much more intuitive for processing two-dimensional data.

The instructors also demonstrated how to use the matplotlib library to create simple line, bar, and pie charts from data stored in a DataFrame. Students completed short in-class activities that mirrored the demonstrations to develop their competency in accomplishing these tasks.

A Taste of Data Analytics: Twitter Analytics

The remainder of the second and third sessions introduced an application for analyzing Twitter data using pandas. The instructors shared an example based on Mayo (Mayo, 2017) showing how to read and analyze a file of Tweets obtained from a user's Twitter account. Determining the total number of Tweets, most popular retweets and likes is accomplished by storing the tweets in a DataFrame, and sorting or summarizing the appropriate columns using methods of that class.

The corresponding homework assignment had students use the pandas, numpy, and matplotlib modules to analyze popular hashtags or mentions from a file of Tweets. A hashtag is any word in a Tweet that begins with a # symbol. A mention is any word in a Tweet that begins with an @ symbol. Students could analyze a file provided by the instructors containing Tweets from a university Twitter account or create a file containing their own Tweets. The assignment required students to import the Tweets into a DataFrame, process the text of each Tweet to create a dictionary of hashtags or mentions and their frequency of use, sort the results alphabetically and by frequency, and plot the results on a horizontal bar chart. The complete assignment description and sample output are shown in Appendix 2.

4. METHODS AND RESULTS

This study takes a quantitative approach using a survey instrument. In addition, student comments and reflections provide qualitative examples to support the results and conclusions.

Seventy of 75 students enrolled in the three spring 2018 sections voluntarily completed an anonymous online survey at the start and end of the semester to share their learning, interest, and impressions of the Python course. The incentive for completing the survey was three extra credit points added to their final exam scores. The discussion that follows presents results from that survey.

The classes were gender balanced, with 51% males, and 49% females. Fifty-four percent of the students were seniors, 36% juniors, 8% sophomores, and 1% freshmen. Of the students

enrolled, 48% had taken or were currently taking a first programming course in Java, 15% had taken or were currently taking a second Java programming course, and 37% had taken or were currently taking an introductory web development course covering HTML and JavaScript. Seventeen students claimed to have taken a programming class in high school. Appendix 1, Table 3 shows the variety of students' majors and minors enrolled in the course. 43% of the students registered were CIS majors and 35% were CIS minors. Of the non-CIS majors, 50% were other non-CIS business majors and 7% were arts and science majors across 17 undergraduate programs of study.

As a response to the first research question, "What factors motivate students to take (and universities to offer) Python as part of the business curriculum?" students cited a desire to increase their future career opportunities and interest in programming as the two most common reasons for taking this class. Sample responses included "I enjoy the challenge of programming" and "My internship for this coming summer asked me if I could get into a Python programming class to be more helpful on the trading floor."

Students with little or no prior coding experience found the course to be extremely difficult. Short group projects assigned during class time created an active learning experience which enabled those with prior experience to assist their classmates for whom programming was new. Students recognized the different skill levels of their classmates and suggested that the CIS department consider offering both introductory and intermediate level Python courses in the future.

Of the assignments described in Appendix 1, Table 2, students found the first two assignments (About Me and Calculations) to be relatively easy, as expected. These programs required writing simple print statements and performing sequential calculations. The Buzz simulation described in (Offenholley, 2012) required converting values between numbers and string representations, conditional statements and loops as well as formatting data using format strings. This assignment proved to be more challenging as students were not used to integer division or modulus operations. The Account and Donor Management projects were even more challenging because of their complexity and the need for conceptual familiarity required when writing code to iterate through collections and files. The Battleship game proved to be very difficult because it required both an

understanding of objects and classes, and the ability to develop relatively complex logic to place ships randomly in a grid. Students found the Data Analytics assignment at the end of the semester to be a welcome change because of its inherent relevance and much more manageable scope. The assignment required writing far fewer lines of code than the three previous assignments, and its solution closely mirrored examples shown in class.

Usefulness and Relevance of Course Topics

Students reflected on the course topics as related to their usefulness in understanding programming, as well as on their relevance to future careers. Appendix 1 Figures 1 (a) and (b) summarize these results. It is interesting to note that students found the data analytics module to be the most relevant while they considered all the other topics more useful in contributing to their overall knowledge of programming. This may be because creating an application using pandas and related libraries does not develop new programming skills, but rather, provides a relevant context in which to apply Python skills developed earlier in the semester. While mastering programming concepts provide the foundation for building Python applications, students found building applications to be more pertinent to their future careers.

In response to the second research question regarding which topics and applications of Python are considered relevant in a business context, students mentioned data analytics, dictionaries, lists and functions most frequently. One student said, "Pandas and matplotlib were cool, I wish we could have done less restrictive projects using those libraries."

Python and Employability

To answer the third research question on the value that business students receive from taking an introductory Python course, students reflected in the end-of-semester survey on the importance of knowing how to write code even if it is not a job requirement, the ability to write code as it relates to their commitment to become IT professionals, the importance of knowing Python when applying for future jobs or internships, the extent to which employers value employees who have Python skills, the extent to which of having Python skills increases their value in the job market, compared to students without Python skills, and their abilities to tackle real-world problems and projects in their future work. Most students somewhat or strongly agreed with each of these and related statements, as shown in Appendix 1, Figure 2.

Business students in majors other than CIS found value in the course as they will apply their knowledge to in their future careers. A data analytics major wrote, "Python will be very relevant in my future career." Another student said, "Midway through the course I looked up how to apply analytics to Python and saw the pandas module. I didn't know how to download it or go about it, so I waited to see if we would get to it at the end of the semester. We ended up getting to it but had 2 classes on it. If we went more in depth with Pandas I think knowing that could help me more in my job as an investment analytics associate at a media agency this August." Another student commented, "I am a marketing major. Often, I need to understand the context of discussions around a product/topic. Python can [be used to] design web-crawlers and extract those data automatically. With this technology, I do not have to passively observe discussions anymore."

6. CONCLUSIONS AND FUTURE WORK

Teaching an introductory Python course with no prerequisites to a classroom of students from multiple business majors, each with varying experiences in developing code, was a challenge. Grouping students with prior programming knowledge with their less-experienced classmates was one way to bridge the gap between students majoring or minoring in CIS and other business disciplines. By completing several smaller in-class programming problems, students were prepared to work on larger homework projects.

Replacing graphics processing and user interface development with a module on pandas and data analytics was a favorable change in the course during the spring 2018 semester. Using these tools to analyze Twitter data has more immediacy and relevance than a more abstract or contrived textbook example. Based on student evaluations after teaching the course, students would prefer additional exposure to data analytics topics. In future semesters, the course will increase coverage of data analytics topics, replacing advanced topics such as recursion, and polymorphism.

As business curricula increase their major and minor offerings to include programs of study related to data analytics, such as Data Analytics, Finance and Technology, and Auditing Analytics), the demand for Python instruction will continue to increase. Introducing data analytics-related examples to the introductory Python course will

make its content more relevant and offer wider appeal to the variety of students enrolled.

7. ACKNOWLEDGEMENTS

The authors acknowledge Professor Wendy Lucas for proposing this course at Bentley University, and for shepherding it through the curriculum committee for approval, and student tutors for their careful review of the assignments.

8. REFERENCES

- Cass, S. (2017, July 18). The 2017 Top Programming Languages. *IEEE Spectrum: Technology, Engineering, and Science News*. Retrieved May 23, 2018, from <https://spectrum.ieee.org/computing/software/the-2017-top-programming-languages>
- Continuum Analytics. (2018). Miniconda — Conda. *Miniconda - Conda*. Retrieved May 24, 2018, from <https://conda.io/miniconda.html>
- Downey, A. (2016). Think Python, 2nd Edition. *Green Tea Press*. Retrieved June 4, 2018, from <http://greenteapress.com/thinkpython2/html/index.html>
- Frydenberg, M. (2017). Ding Dong, You've Got Mail! A Lab Activity for Teaching the Internet of Things. *Information Systems Education Journal*, 15(2), 20.
- Guo, P. (2014, July 7). Python Is Now the Most Popular Introductory Teaching Language at Top U.s. Universities. Retrieved May 23, 2018, from <https://cacm.acm.org/blogs/blog-cacm/176450-python-is-now-the-most-popular-introductory-teaching-language-at-top-u-s-universities/fulltext>
- Kortsarts, Y., Morris, R., & Utell, Janine. (2010). Interdisciplinary Introductory Course in Bioinformatics. *Information Systems Education Journal*, 8(27). Retrieved June 3, 2018, from <http://isedj.org/8/27/index.html>
- Lambert, K. (2018). *Fundamentals of Python: First Programs* (2nd ed.). Cengage. Retrieved June 4, 2018, from <https://www.cengage.com/c/fundamentals-of-python-first-programs-2e-lambert/9781337560092>

- Mayo, M. (2017, March). A Beginner's Guide to Tweet Analytics with Pandas. Retrieved May 23, 2018, from <https://www.kdnuggets.com/2017/03/beginners-guide-tweet-analytics-pandas.html>
- McMaster, K., Sambasivam, S., Rague, B., & Wolthuis, S. (2017). Java vs. Python Coverage of Introductory Programming Concepts: A Textbook Analysis. *Information Systems Education Journal*, 15(3), 4.
- Offenholley, K. (2012). Gaming Your Mathematics Course: The Theory and Practice of Games for Learning. *Journal of Humanistic Mathematics*, 2(2), 79–92.
- Reynolds, J. (2018, February 8). 8 World-Class Software Companies That Use Python. *Real Python*. Retrieved June 1, 2018, from <https://realpython.com/world-class-companies-using-python/>
- Richardson, L. (2017, August 11). Beautiful Soup: We called him Tortoise because he taught us. Retrieved June 4, 2018, from <https://www.crummy.com/software/BeautifulSoup/>
- Siegfried, R. M., Siegfried, J., & Alexandro, G. (2016). A Longitudinal Analysis of the Reid List of First Programming Languages. *Information Systems Education Journal*, 14(6), 47.
- Topi, H., Valacich, J. S., Wright, R., Kaiser, K., Nunamaker, J. F., Sipior, J., & de Vreede, G. J. (2010). IS 2010 Curriculum Guidelines for Undergraduate Degree Programs in Information Systems. Association for Computing Machinery. Retrieved June 4, 2018, from <https://www.acm.org/binaries/content/assets/education/curricula-recommendations/is-2010-acm-final.pdf>
- US News & World Report. (2018). The 10 Best Colleges for Management Information Systems. Retrieved June 3, 2018, from <https://www.usnews.com/best-colleges/rankings/business-management-information-systems>

Appendix 1. Additional Tables and Figures

Table 1. Python at Top 12 MIS Programs (according to US News)

Rank	Name	Python is required as a Core Course for IS Majors	Python is offered an elective for IS Majors
1	Massachusetts Institute of Technology	Python is one of eight required courses for business analytics major (MIT has no IS undergraduate program)	No
2	Carnegie Mellon University	No	Python is only offered at graduate level (MSBA)
3	University of Arizona	No	No
4	University of Minnesota Twin Cities	2 object-oriented courses (can be Python or other OO language)	No
5	University of Texas Austin	1 app development course	No
6	Georgia Institute of Technology	Java required, not Python	No
7	Indiana University Bloomington	No	Yes
8	University of Maryland College Park	No	No
9	University of Pennsylvania	No	No
10	Georgia State University	1 object-oriented course (can be Python or other OO language)	No
11	University of Michigan Ann Arbor	N/A	N/A
12	New York University	1 object-oriented course (can be Python or other OO language)	No

Table 2. Homework Assignments (Spring, 2018)

#	Topics	Description
1	Displaying information, using an IDE	Print information about you
2	Expressions and Data Types	Calculate unit prices of food items based on quantity purchased
3	Control Structures	Buzz Game (test for numbers containing or divisible by 7) (Offenholley, 2012)
4	Strings, Text Files, Lists, OS Module	User Account Manager – store usernames, passwords, allow users to add/edit/delete account information
5	Dictionary of Lists	Manage a list of donors and donation amounts; determine most generous donor, and total donations.
6	Classes and Objects	Hide ships on a Battleship game grid; a player must locate them all within a specified number of turns
7	Introduction to Data Analytics	Analyze a file of Tweets to determine most popular hashtags; create a horizontal bar chart showing hashtags and frequency

Table 3. Majors and Minors enrolled in 3 sections.

The number of students from each major is shown in bold. Minors of students with identified majors shown are indented.

Accountancy	1	Natural Sciences	1	Computer Information Systems	2
None Specified	1	Philosophy	1	Liberal Studies	1
Actuarial Science	7	Psychology	1	Computer Information Systems	1
Business Studies	4	None Specified	15	Management	2
Computer Information Systems	1	Corporate Finance and Accounting	1	Computer Information Systems	1
Data Technologies	1	Computer Information Systems	1	Mathematical Sciences	1
English	1	Data Analytics	1	Managerial Economics	1
Business Studies	3	Business Studies	1	Computer Information Systems	1
Computer Information Systems	2	Economics-Finance	4	Marketing	5
None Specified	1	Computer Information Systems	3	Computer Information Systems	4
Computer Information Systems	31	None Specified	1	None Specified	1
Data Technologies	1	Finance	8	Mathematical Sciences	2
Finance	4	Accountancy	1	Computer Information Systems	1
Information and Process Management	2	Computer Information Systems	6	Sociology	1
Information Design and Corporate Communication	1	Mathematical Sciences	1	Public Policy	1
Leadership	1	Information Design and Corporate Communication	1	Computer Information Systems	1
Management	2	Computer Information Systems	1	None Specified	1
Marketing	1	Information Systems Audit and Control	2	None Specified	1
Mathematical Sciences	1			Grand Total	72

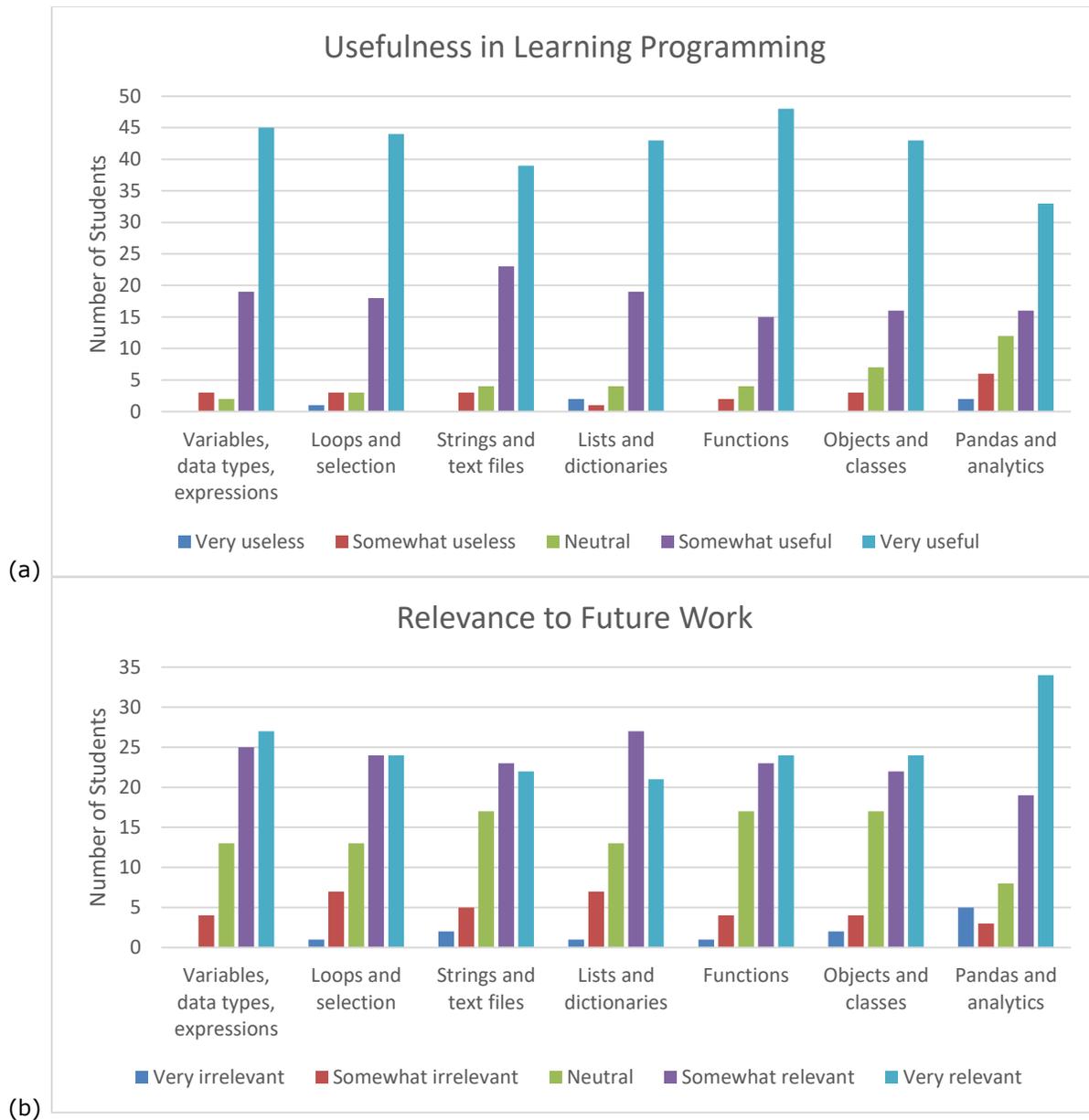


Figure 1. Python Topics: (a) Usefulness in Learning Programming and (b) Relevance to Future Work.

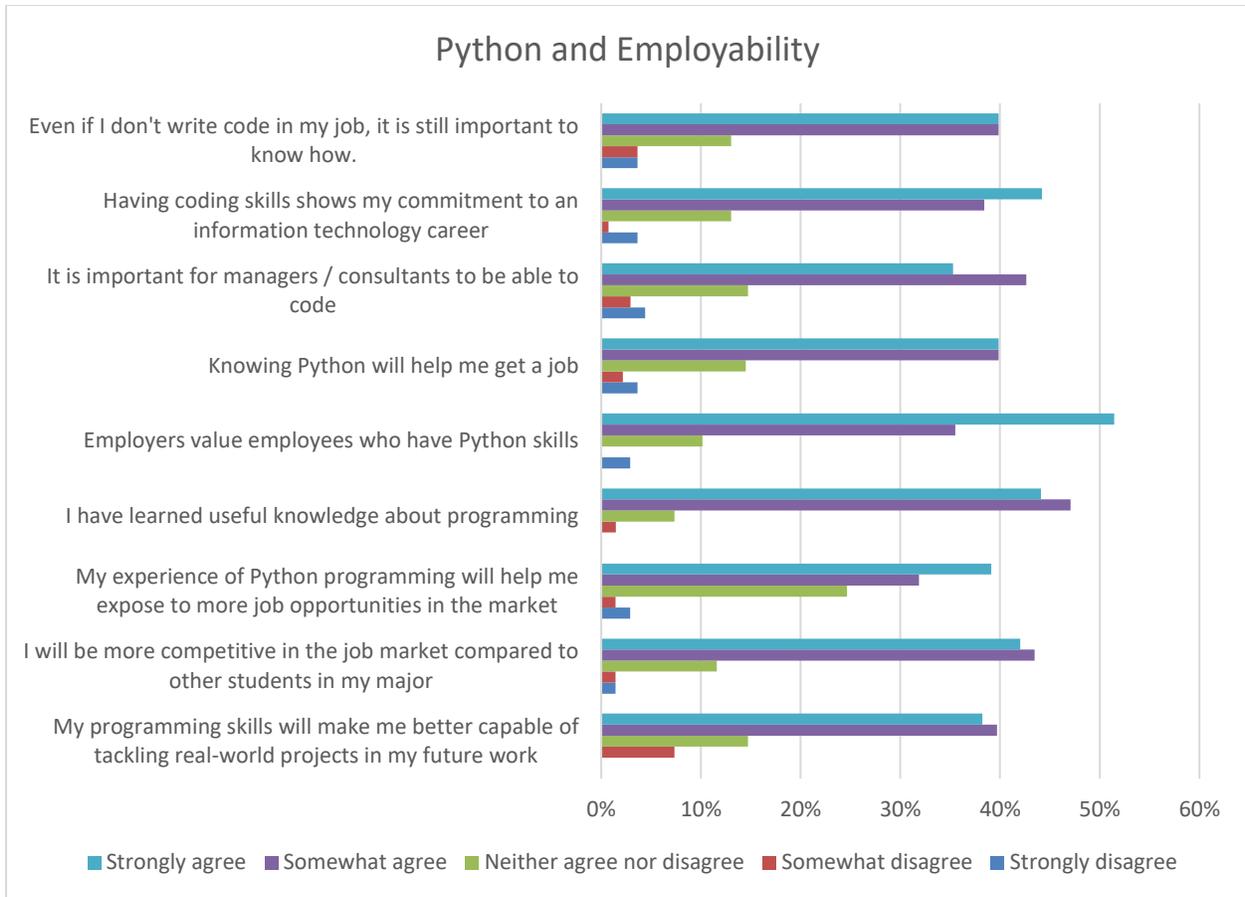


Figure 2. Perceived Value of Python Skills as related to Employability

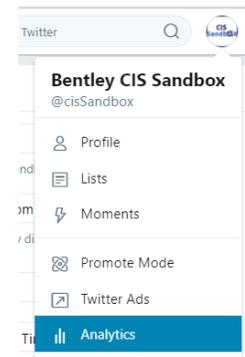
Appendix 2. Twitter Analytics Assignment Description

A **hashtag** is any word in a Tweet that begins with a # symbol, a **mention** is any word in a Tweet that begins with an @ symbol. In this assignment, you will use the pandas, numpy, and matplotlib modules to analyze hashtags or mentions found in Tweets during a period of several months and exported from Twitter.

Export Your Tweets

You can use the sample Twitter data file supplied containing Tweets from a university Twitter account or export your own Tweets to analyze. Follow these steps to export your own Tweets:

1. Sign in to your Twitter account.
2. Click **your profile icon** at the top right corner of the page.
3. Click **Analytics** from the dropdown menu below your profile icon.
4. On the Analytics page, click **Tweets** at the top of the page



5. Select a date range of activity (choose a large enough date range so you have Tweets containing several hashtags or mentions),
6. Click the **Export Data** button to export your Twitter data.
7. Save the file as tweets.csv in the folder with the python program you are about to write.

Analyze Your Tweets

1. Follow the example given in class to load Twitter data from the tweets.csv file into a pandas DataFrame, removing irrelevant columns.
2. Allow the user to specify whether to analyze hashtags or mentions by typing h or m.
3. Analyze the **Tweet text** column of the data frame to create a dictionary of hashtags/mentions and corresponding frequencies (number of times each appears). For simplicity, the DataFrame will have two columns: Hashtag and Frequency, regardless of whether you are analyzing hashtags or mentions. As you process the data, convert all words to lower case so that, for example, hashtags #PYTHON and #python would be considered the same.
4. Create a pandas DataFrame containing each hashtag and its corresponding frequency.
5. Print the DataFrame.
6. Sort the DataFrame alphabetically by hashtag. Print the sorted DataFrame.
7. Sort the DataFrame in decreasing order by frequency. Print the sorted DataFrame.
8. Create a horizontal bar chart that plots each hashtag and its frequency. Be sure to set x and y labels for the axes, ticks for values and hashtags, and a title for your plot.

See https://matplotlib.org/gallery/lines_bars_and_markers/barh.html or <https://medium.com/python-pandemonium/data-visualization-in-python-bar-graph-in-matplotlib-f1738602e9c4> for examples of how to create a horizontal bar chart.

Sample Output

Analyze [H]ashtags or [M]entions? h

	Hashtag	Frequency
0	#bentley2022	6
1	#ohbot	1
2	#teslatuesday	1
3	#tiltbrush	3

Sorted Alphabetically

	Hashtag	Frequency
0	#bentley2022	6
1	#ohbot	1
2	#teslatuesday	1
3	#tiltbrush	3

Sorted Descending by Frequency

	Hashtag	Frequency
0	#bentley2022	6
3	#tiltbrush	3
1	#ohbot	1
2	#teslatuesday	1

