Visual Insights of Learning Resource Effectiveness for Lebanese Data Science Students

A PROJECT REPORT

Submitted by

Batoul Hamieh
Samira Jawish
Jana Tarabay

in partial fulfillment of the requirements for the Data Visualization course

Supervisor:

Dr. Yussef Al Attat

LEBANESE UNIVERSITY

FACULTY OF INFORMATION

January 2026

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Standardized in-class teaching methods have been around for ages and over time they have been proven to be quite beneficial and highly impactful to the human's intellectual process of learning and consuming information. After the Covid pandemic hit, new ways and tools have been introduced in order to amplify the in-parallel learning process to almost match the effectiveness of in-person sessions. From Zoom meetings to bootcamps to online platforms for courses such as edX, Coursera, Udemy and many more took advantage of the current worldwide situation to spread the culture of online learning to students. The rapid change in teaching methods wasn't favored at first but then soon became popular allowing people from all over the globe to learn together and share information interchangeably. Institutions saw it as an opportunity to upgrade their education system keeping both in-person and online platforms as a way of adopting new and flexible ways to educate the next generation.

We sought out to study the effectiveness of university courses versus non-university courses for acquiring technical skills targeting data science students. This is a study that is dedicated to understand how dependent are data science students on platforms and educational tools other than their regular university courses and if they are contributing in forming their technical skills that are required in the job-market. Moreover, a study comparing MOOCs with traditional university courses found that MOOCs tend to have lower completion and pass rates and show mixed performance outcomes overall. However, the evaluation relied on broad academic indicators, failing to capture discipline-specific skills such as practical data science competencies or the complementary use of online platforms alongside university instruction.

Data science is a hyper-disciplinary major which requires by definition a multi-skilled individual with technical and analytical skills to execute a fully functional, production ready projects that matches today's technology and innovation. In order to match that level of competence it requires data science students to naturally learn new tools and methodologies that are popular in tech dominant countries so it doesn't come off as odd that they are constantly seeking the next new tutorial on the shiniest framework to come. With that being said they are prone to depend more on learning platforms that match their needs and we assume that the university studies don't quite applies to their job-market responsibilities.

The aim of this study is to understand whether data science students are confident of their university background or it is not reliable anymore with the development of knowledge to rely on traditional ways forcing them to switch to online tools and platforms to meet the needs of a rapidly changing job market.

2. STUDY

The research adopts a survey-based approach, which aligns with established scientific methodology by allowing systematic collection of self-reported data on students' learning experiences, skill acquisition, and perceptions. The survey design is structured to capture both quantitative and qualitative information.

Initial demographic questions, such as academic level, university affiliation, and major, were included to contextualize responses and allow for subgroup analysis. Subsequent questions probe students' prior use of online learning platforms, motivations for using or not using them, and perceived effectiveness relative to university instruction.

Quantitative questions employ Likert-scale ratings to measure the perceived effectiveness of learning, understanding of course content, analytical thinking, and confidence in acquired skills. Qualitative items invite open-ended responses to explore personal recommendations, enriching the dataset with insights that quantitative measures alone cannot capture.

Collectively, the design adheres to the scientific method: it defines measurable variables, systematically collects and organizes data, and lays the foundation for analysis that can support evidence-based conclusions about learning effectiveness in the context of Data Science education in Lebanon.

3. METHODOLOGY

This study employs a comparative, descriptive, and visualization-driven research design to examine differences in learning experiences between university-based instruction and non-university learning channels among Data Science students in Lebanon. The analysis focuses on visual exploration and interpretive comparison rather than statistical modeling or hypothesis testing. Primary comparisons include university learning versus platform-based learning, platform users versus non-users, and differences among various online platforms. The unit of analysis is each learning channel usage instance, with descriptive measures such as counts and percentages forming the basis for visualization.

DATA COLLECTION:

Data were collected through a custom-designed bilingual (Arabic and English) online questionnaire using Google Forms. The survey captured demographics, platform usage, course completion status, Likert-scale evaluations (1–5), and open-ended qualitative feedback. Conditional logic ensured respondents were presented only with relevant questions based on their platform usage. Distribution occurred via university WhatsApp groups and LinkedIn over a two-week period, with voluntary and anonymous participation. Only Data Science students were included, ensuring alignment with the study's scope.

SAMPLING CHARACTERISTICS:

A convenience sampling approach was used, targeting students from multiple public and private Lebanese universities. Respondents outside the Data Science discipline was excluded during data preparation. Ethical considerations were observed through informed consent, anonymity, and responsible data handling.

DATA PREPARATION:

Survey responses were exported to CSV and processed in Python using pandas. Data cleaning ensured structural consistency while retaining missing values arising from survey logic. Binary variables were encoded for analysis, multi-response questions were split into indicator columns, and subsets were created for targeted visualization. Likert-scale responses were maintained as ordinal numeric values without transformation, supporting direct descriptive analysis.

VISUALIZATION AND ANALYTICAL STRATEGY:

Flourish was selected as the primary visualization tool due to its versatility in comparative and interactive visualizations. Cleaned CSV files were uploaded to generate charts that highlight differences in usage patterns, perceived effectiveness, and learning outcomes. Analysis relied on counts, percentages, and visual pattern recognition to identify contrasts and trends across learning channels. No predictive modeling or causal inference was performed, ensuring a strictly descriptive and exploratory framework.

4. FLOURISH TOOL OVERVIEW

Flourish studio is a web-based for creating interactive data visualizations and storytelling experiences without coding. It helps users to transform raw data into engaging charts, maps, scrolly telling narratives and presentations optimized for web and mobile audience.

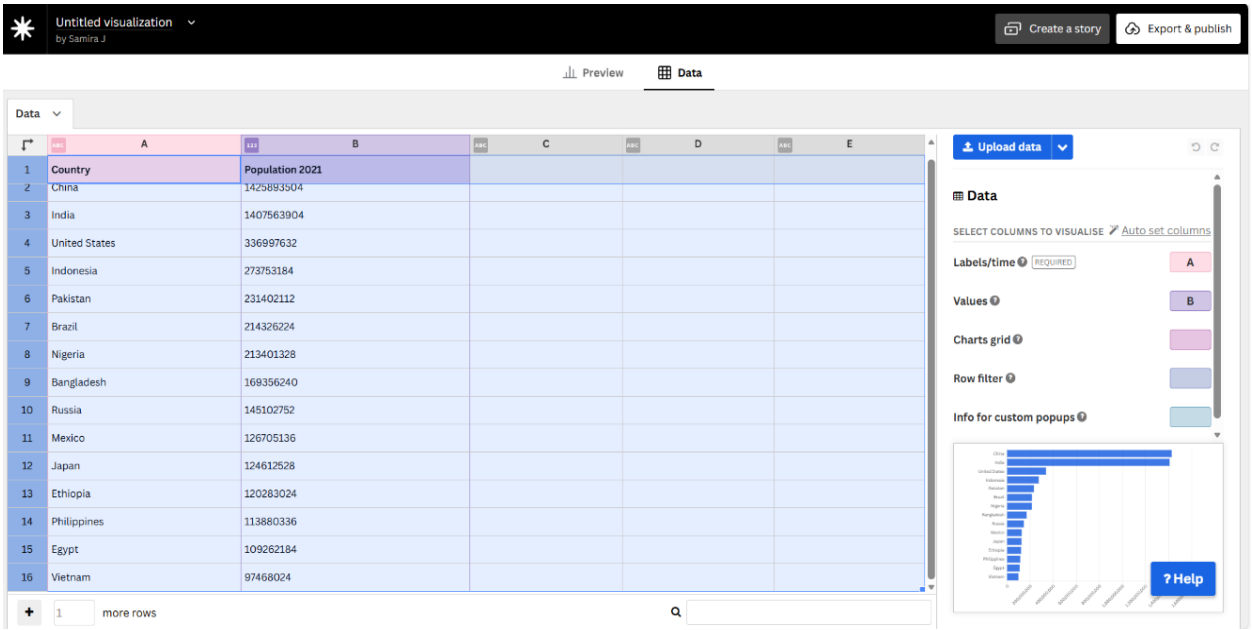
It focuses on intuitive, template-based tools for narrative data visualization, including scrolly telling, animated transitions, and embeddable outputs. No technical skills are required, as users upload data like spreadsheets and customize via drag-and-drop interfaces. Key strengths include storytelling flexibility, publishable robustness for large audiences.

The target group that is designed to serve are primarily journalists, newsrooms, marketers, educators and creators in organizations worldwide who need quick, professional data stories.

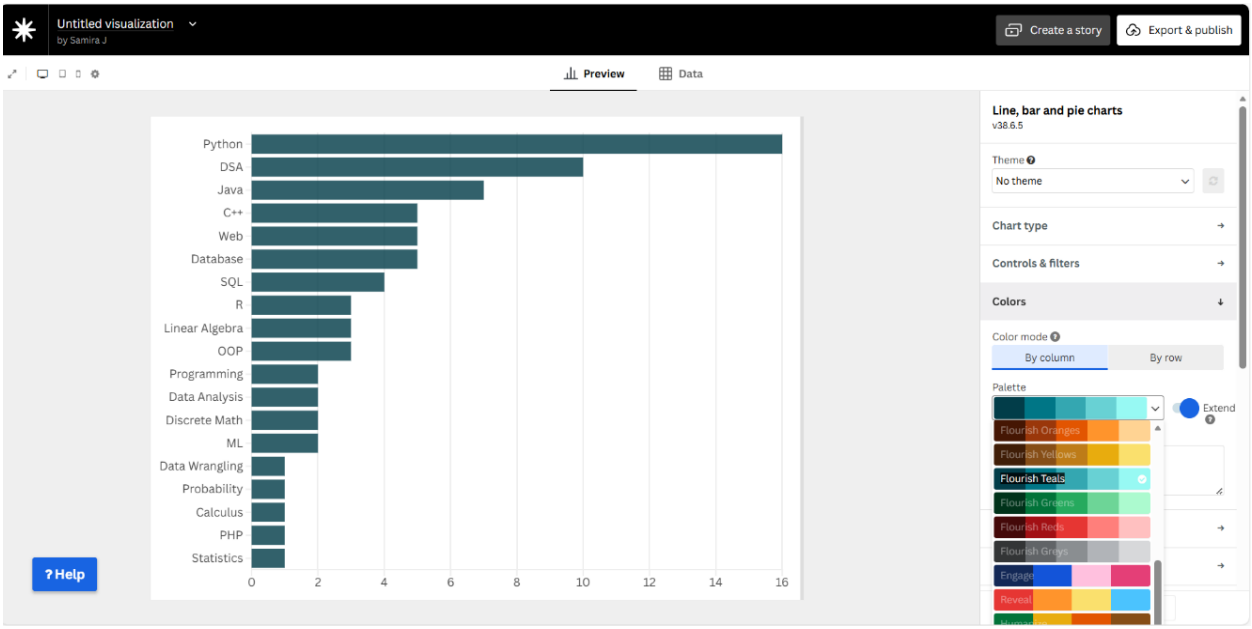
How to Use Flourish Studio: (Bar Chart Example)

Creating a visual in Flourish is a fast and straightforward process that requires minimal setup while delivering highly polished results. The experience begins by accessing Flourish Studio and selecting a visualization template that matches the structure of the data and the analytical goal, whether the aim is to compare categories, show trends over time, or highlight geographic patterns.

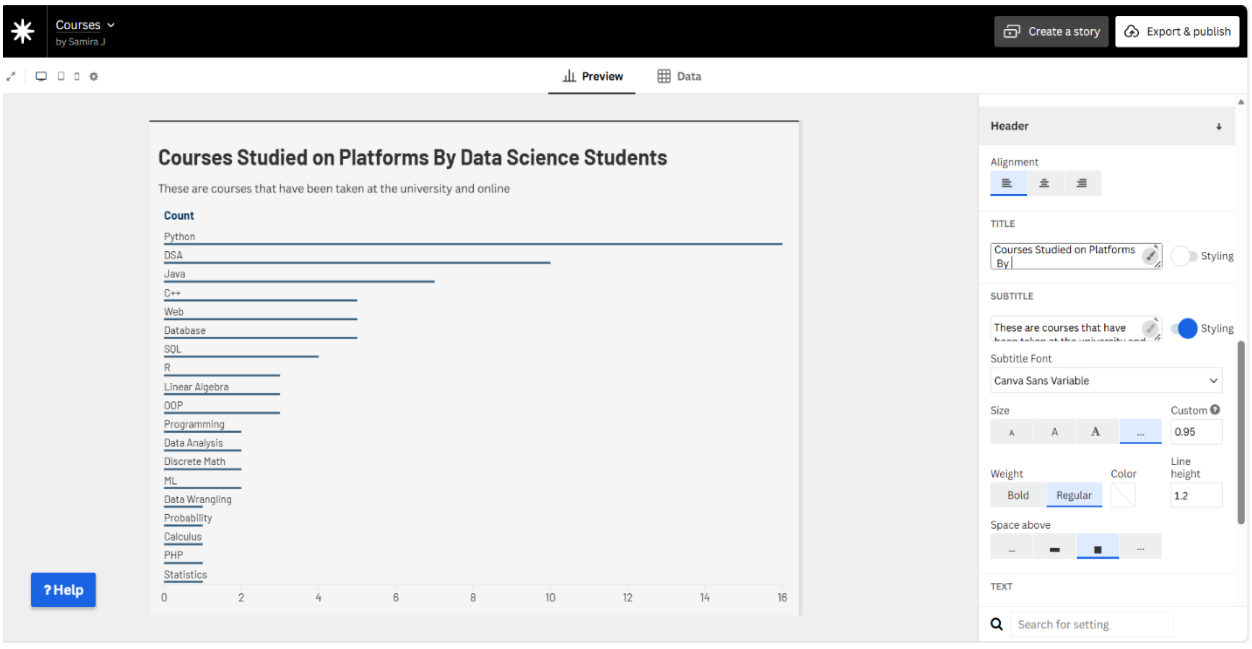
After choosing a template, data are uploaded or pasted into the platform, where organizing variables into clear, well-defined columns ensures accurate data binding.



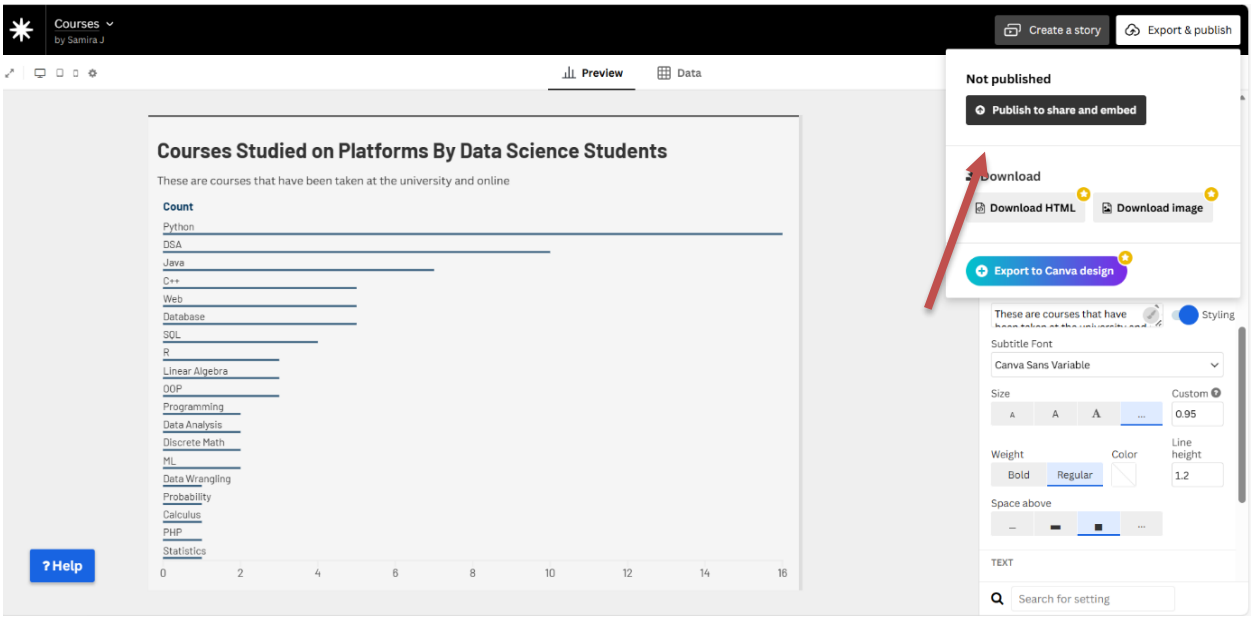
	A	B	C	D	E
1	Country	Population 2021			
2	China	1425893504			
3	India	1407563904			
4	United States	336997632			
5	Indonesia	273753184			
6	Pakistan	231402112			
7	Brazil	214326224			
8	Nigeria	213401328			
9	Bangladesh	169356240			
10	Russia	145102752			
11	Mexico	126705136			
12	Japan	124612528			
13	Ethiopia	120283024			
14	Philippines	113880336			
15	Egypt	109262184			
16	Vietnam	97468024			

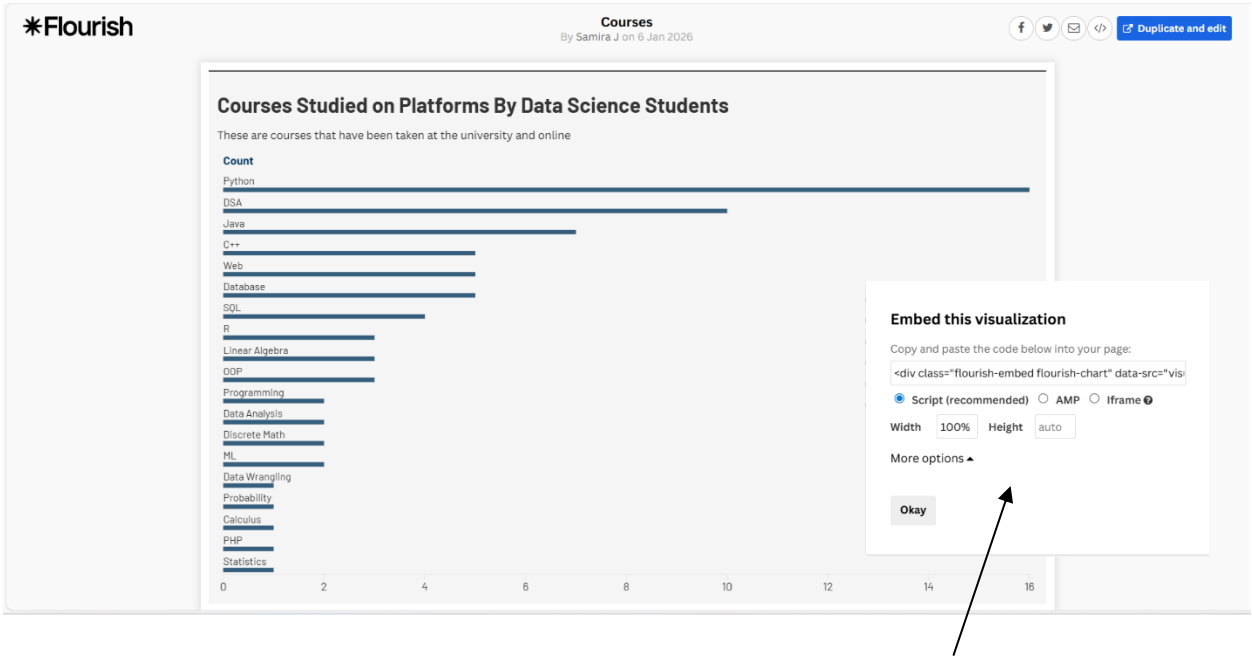


The user then assigns these variables to visual elements such as labels, values, colors, or time dimensions, allowing the visualization to render correctly. Flourish provides flexible customization options for adjusting colors, axes, labels, annotations, and interactive features, enabling the visual to clearly communicate key insights.



Before publishing, the visualization is reviewed to verify accuracy and interpretability, after which it can be shared via a link or embedded, with the option to duplicate and refine the visual as analytical or storytelling needs evolve.





THE ADVANTAGES AND DISADVANTAGES OF USING FLOURISH AS DATA SCIENTISTS:

Flourish Studio offers data scientists a rapid and accessible way to transform structured data into interactive and visually compelling narratives. Its wide range of ready-made templates make it particularly attractive for communicating insights to non-technical audiences. However, while Flourish excels in visualization and storytelling, it also presents limitations in terms of advanced analytical control and customization, making it important for data scientists to carefully evaluate when it is the most appropriate tool to use.

THE ADVANTAGES OF USING FLOURISH STUDIO:

Flourish Studio offers a wide variety of visuals across different data structures which gives the user flexibility to innovate and demonstrate how they want their data to be shaped. For data science students, Flourish Studio provides several practical advantages that support both learning and communication. Its low learning curve allows students to quickly convert cleaned datasets into interactive visualizations without requiring advanced programming skills, enabling them to focus more on data interpretation and insight generation. The platform reinforces good data practices by encouraging structured data formatting and clear variable mapping, which aligns with foundational data science principles. Additionally, Flourish's interactive and visually engaging outputs help students effectively present findings in academic projects, presentations, and portfolios, making complex results more accessible to instructors and non-technical audiences. By bridging technical analysis and visual storytelling, Flourish also helps students develop an essential skill set for communicating data-driven insights in professional contexts.

THE DISADVANTAGES OF USING FLOURISH:

Despite its strengths, Flourish Studio presents several disadvantages for data science students, particularly from an analytical perspective. The platform is not inherently designed to support the transformation of raw data into insights, as it assumes that data are already cleaned, structured, and analytically prepared before being uploaded. This often requires additional preprocessing and restructuring outside the tool, which can limit its usefulness during exploratory analysis. Moreover, Flourish lacks a built-in statistical infrastructure, offering no native support for statistical testing, modeling, or feature engineering, which are core components of data science workflows. As a result, students must rely on external tools such as Python, R, or statistical software to derive insights, using Flourish primarily as a visualization layer rather than a comprehensive analytical environment.

WHY DID WE CHOOSE FLOURISH STUDIO FOR OUR DATA VISUALIZATION PROJECT?

Although Flourish Studio is for non-coders and graphics enthusiasts that want to showcase their data through an attractive visual but we wanted to experiment and understand if we can use a tool non-technical as Flourish Studio and create data diagrams and meaningful charts while aligning with the principles and means of adequate data visualization.

WHO SHOULD USE FLOURISH?

Flourish is best suited for users whose primary objective is to communicate data-driven insights clearly and interactively rather than to perform in-depth analysis within the same environment. It is particularly valuable for data journalists, analysts, educators, and students who already work with cleaned and structured data and need an efficient way to transform results into compelling visual narratives. For data science students and professionals, Flourish is most appropriate as a presentation and storytelling tool used after statistical analysis and modeling have been completed in external platforms such as Python, R, or BI tools. In short, Flourish should be used by individuals and teams who prioritize accessibility, visual impact, and interactivity in sharing insights, rather than those seeking an end-to-end analytical or statistical workflow.

5. DATA VISAULIZATION

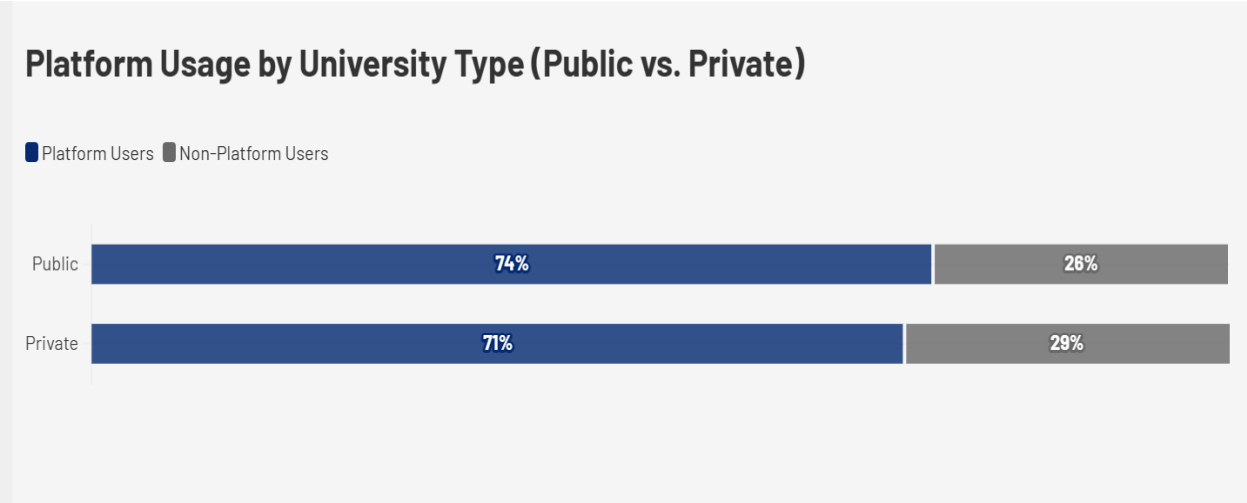


Figure 1

The survey reached data science students across different universities some of them which are public and the others are private. Due to the diverse environments and educational systems offered by them, we suspected that there might be a difference in the attitude of the data science students towards using alternative platforms to support their university studies.

So, we attempted to test that assumption using a proportional bar chart.
To help us compare how many data science students from both public and private universities use platforms as part of their studying process.

The visual showcase that about 74% of data science students in public universities use platforms greater by 3% than those in private universities 71%. However, 26% of students in public education don't use platforms less than that of private institutions 29%.

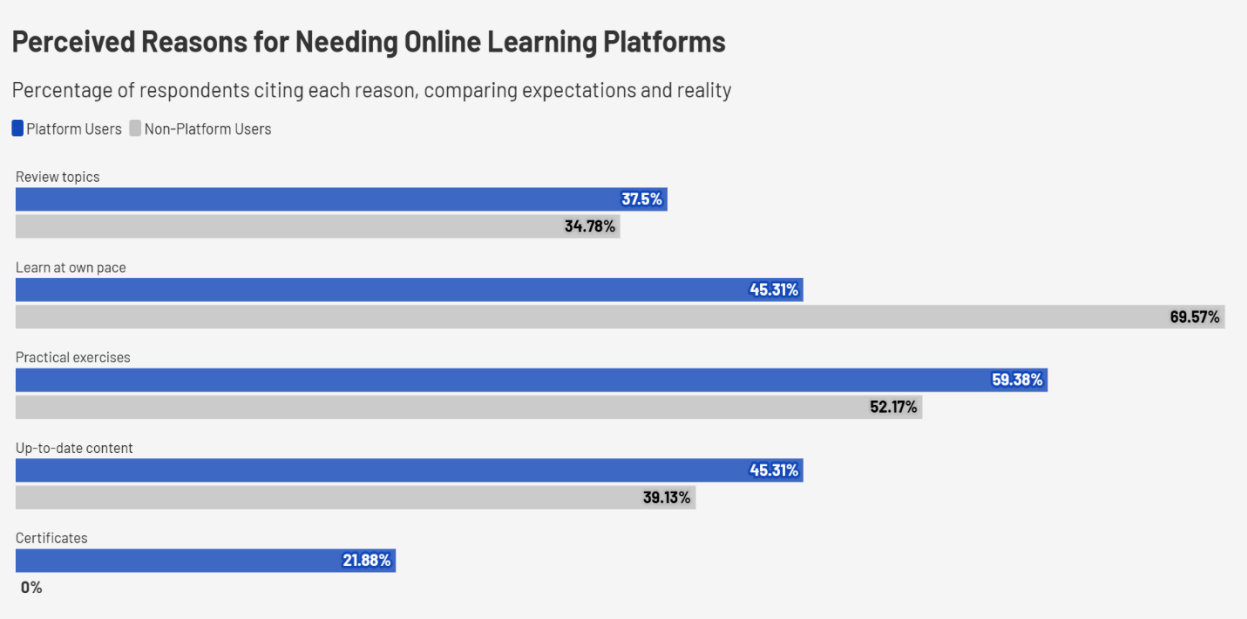


Figure 2

What does the 24.26 percentage-point gap between Platform Users and Non-Platform Users regarding the perceived need to "learn at own pace" reveal about the alignment between expectations and actual user experiences?

Does the 0% interest in certificates among Non-Platform Users, compared with 21.88% interest among active users, suggest that the value of certification is recognized only after platform adoption?

The grouped bar chart shows a clear variation in priorities between current platform users and non-users. Non-users primarily seek the flexibility to “learn at own pace” (69.57%), indicating that time management drive initial interest. In contrast, active users prioritize "practical exercises" (59.38%), while self-pacing is less critical (45.31%), suggesting that hands-on application becomes the dominant need once engaged.

The "certificates" category shows a notable anomaly: non-users report 0% interest, whereas 21.88% of users value credentials. This points to a post-adoption value discovery effect, where the importance of formal recognition emerges after experiencing the platform. Other features, like "Review topics" and "Up-to-date content," show smaller differences, implying that content quality is a baseline expectation across both groups.

Marketing to non-users should emphasize scheduling flexibility and self-paced learning, while retention strategies for active users should focus on practical exercises. Certificates, although not a strong motivator for entry, remain important for ongoing engagement.

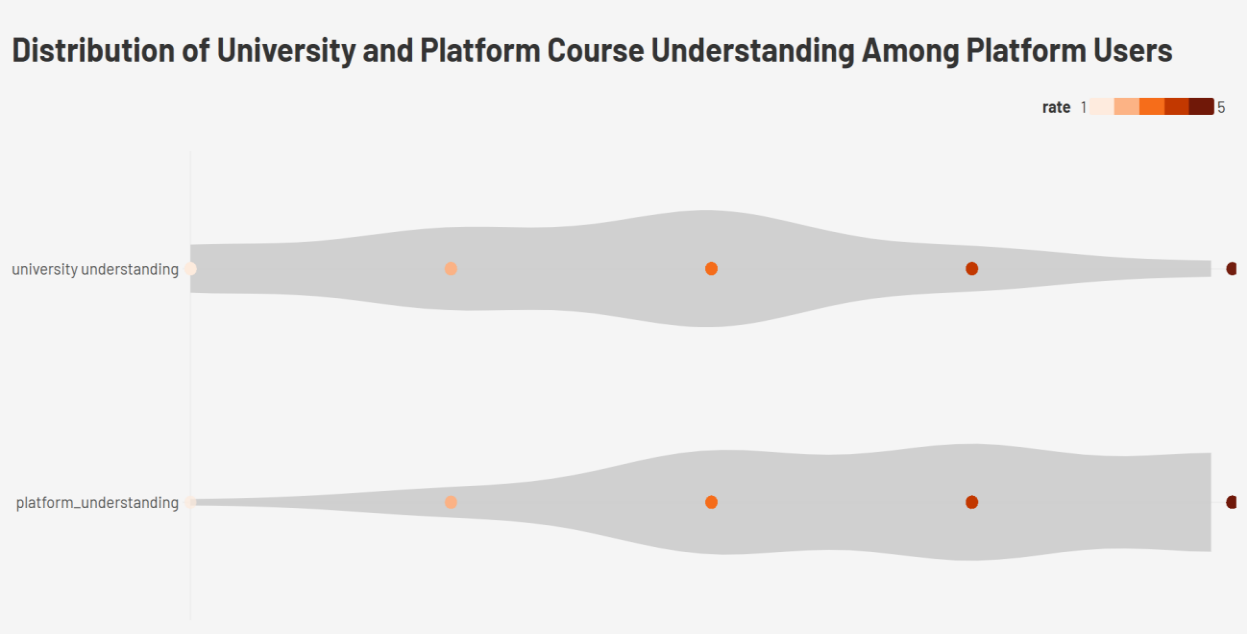


Figure 3

To what extent does the understanding distributions suggest that platform-based learning better achieves high-level mastery than traditional university courses?

The violin plot highlights a marked difference in perceived understanding between university and platform learning. University comprehension shows a centralized distribution across ratings 3-4, reflecting curriculum complexity. In contrast, platform learning is heavily skewed toward rating 5, with minimal low-end 1 rating, indicating that users perceive near-complete mastery. This suggests that platform courses are optimized for high-level comprehension, whereas university courses follow a broader academic spectrum.

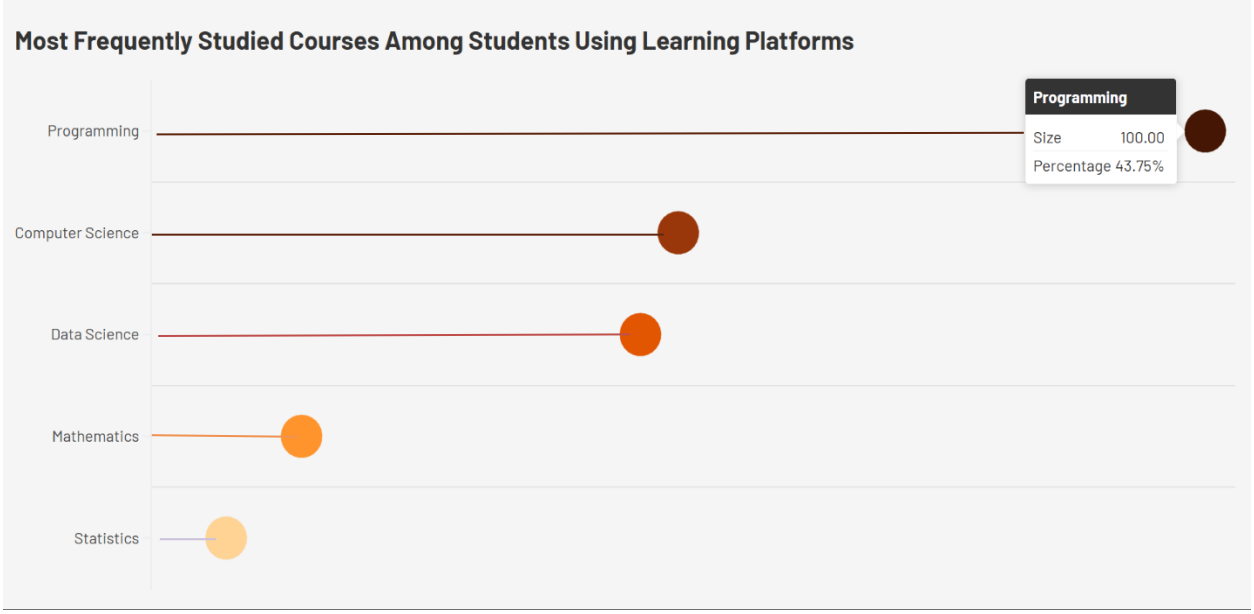


Figure 4

In the survey we tried to explore what type of courses students often studied on platforms alongside the course taken at the university and we received a curriculum of different courses such as Python, Java, C++, Linear Algebra, DSA and many more.

To understand whether students share a common course category that most of them study online. We grouped each course with the compatible category and then proceeded to match the count of the occurrence of the course and then it accumulated to the number of students that may struggle to understand the course in class and depend on platforms to help them improve their grasp of the concept

To display the findings, we used the lollipop chart to showcase the categories that are frequently studied by students online for its easy-to-read layout, direct approach and fast decoding.

We can see that the programming courses are the ones that are studied the most on online platforms by 43.75% other than the regular university classes that they already are taking. Moreover, landing on 2nd place is computer science courses which combines a blend of programming and data science courses.

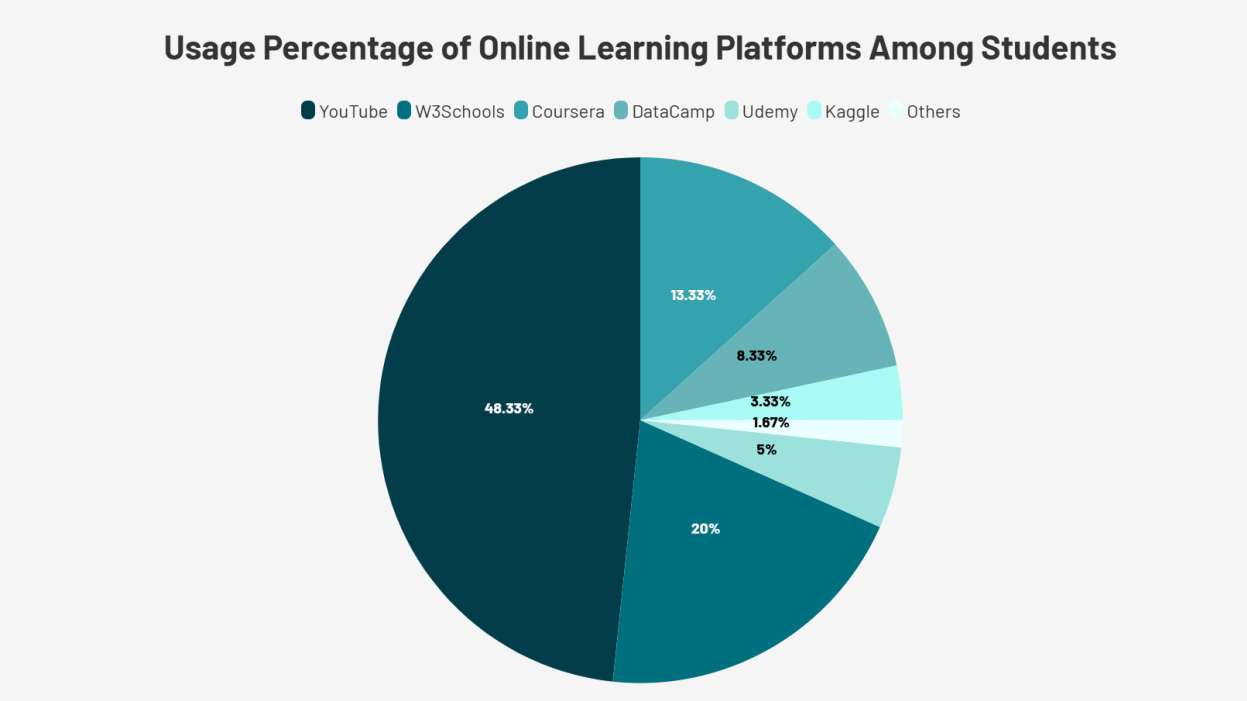


Figure 5

With YouTube and W3Schools accounting for nearly 70% of student usage, what does this reveal about the trend toward low-barrier learning versus formal education?

Does the low 3.33% engagement with Kaggle indicate a preference for foundational over applied data science skills?

The pie chart reveals a significant reliance on high-accessibility learning resources, with YouTube (48.33%) and W3Schools (20%) mapping nearly 70% of the student market. This dominance reflects a strategic preference for visual-based content and rapid-reference documentation, suggesting that students prioritize just-in-time learning over structured academic frameworks.

In contrast, specialized technical platforms like Kaggle occupy a marginal 3.33% share, despite its status as a premier data science and competition community. This low engagement points to a potential expertise gap. This indicates that although generalist tools serve as primary entry points, students still face considerable challenges when moving into specialized, project-oriented learning.

6. CONCLUSION

This study examined the relative effectiveness of university courses versus non-university learning platforms in building technical skills among data science students. The findings indicate a clear positive association between increased use of online platforms and students' self-reported confidence, analytical thinking, and technical competence. Students predominantly rely on non-university platforms for highly technical subjects, such as programming, computer science, data science, machine learning, and mathematics, where hands-on practice and iterative problem-solving are essential. This pattern suggests that traditional university curricula may not always be sufficiently equipped to address the depth and practical intensity required for mastering these skills, prompting students to supplement their formal education with external resources. Notably, programming-related technical skills emerged as the most sought-after competencies developed through online courses, highlighting the central role of applied learning in data science education. Furthermore, the use of Flourish as a visualization and storytelling tool enabled the study to present these insights accurately and transparently, allowing patterns to emerge directly from the data without distortion or over-interpretation. By preserving the authenticity of the information and presenting it clearly, Flourish supported an evidence-based understanding of how students navigate and combine university and non-university learning pathways to develop technical expertise.