Coursera Data Analysis

Sandra ROCHE

December 20, 2023

1 Introduction

Coursera was founded in 2012 by Stanford University computer science professors Andrew Ng and Daphne Koller. Ng and Koller started offering their Stanford courses online in fall 2011, and soon after left Stanford to launch Coursera. Princeton, Stanford, the University of Michigan, and the University of Pennsylvania were the first universities to offer content on the platform.

In 2014, Coursera received both the Webby Winner (Websites and Mobile Sites Education 2014) and the People's Voice Winner (Websites and Mobile Sites Education) awards.

In March 2021, Coursera filed for an IPO. The nine-year-old company brought in roughly $ 293 million in revenue for the fiscal year ended December 31 — a 59% growth rate from 2019, according to the filing. Net losses widened by roughly $20 million yearly, reaching $66.8 million in 2020. Coursera spent $107 million on marketing in 2020.

Currently, on the website https://www.coursera.org/, Coursera announced 5800 courses from more than 300 universities and companies like IBM and Google. Does data set confirm these information ? What time students spend to these training ? What type of certification is proposed ? What is the Top 10 of university and company ?

2 Methodology

2.1 Data set

The data set contains 1000 lines corresponding to different courses with their description and several parameters link to them like certification type, time, an so on.

The data set is provided by the Kaggle website https://www.kaggle.com/datasets/tianyimasf/coursera-course-dataset which contains a lot of data sets on multiple thematic.

2.2 Variables

Some variables contained in the data set are explained in the Table 1. Additional one could be found (course_url, course_skills, course_summary, course_description) but they could not be used in the present analysis since there are unique and depends on each course.

Variables	Explanation
course_title	Title of the course
course_organization	University or company
course_certificate_type	Type of certification
course_rating	Rate provided from students
course_review_num	Number of reviews for the course
course_time	Estimated time to complete the program
course_difficulty	Level difficulty for the course
course_students_enrolled	Number of student enrolled

Table 1: Variables explanation

2.3 Method

First, data set is imported from Kaggle. At the beginning, data are analyzed from a quality point of view including search of missing values, duplicate data and typography issue. Then, some variables distributions are observed.

Next, an analysis is performed to explore the data set and find the Top 10 of university and companies that proposed courses on the Coursera platform, the level of the courses, certification type and time spend by the students.

Finally, limitations and discussion are proposed at the end of the report.

3 Results

The data set contains 4% of missing values and none duplicate data. Moreover, no typography issue has been found in variables. Thus, only 958 data (N=958) could be explored representing only 16% of the total courses proposed by Coursera on the website.

First is explored the satisfaction degree from students after they ended courses. This corresponding to a rate between 0 and 5 and the Figure 1 shows that 99% of rates are comprised between 4 and 5 meaning that students are very satisfied at the end.
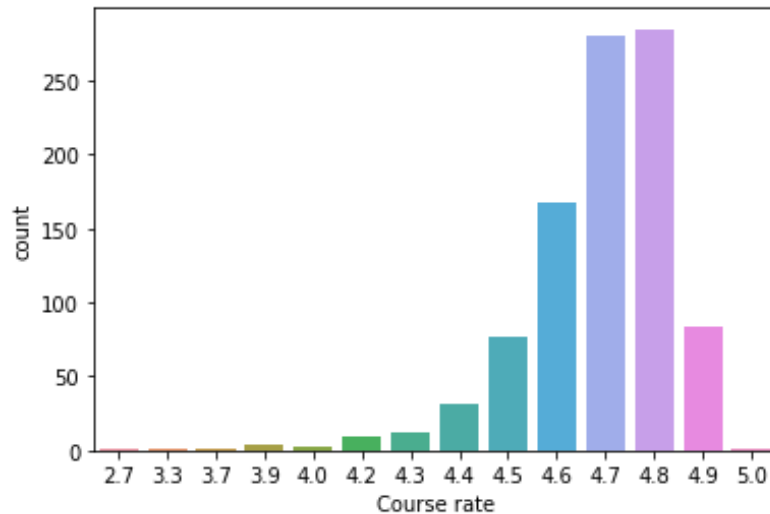

Figure 1: Course rate from students

This satisfaction degree is not link to a course certificate type or to a course difficulty, all courses have a good rate from students.

Next, the number of courses per difficulty and per certificate type are explored.

As show in the Figure 2, most of the courses (70%) proposed by the platform are for beginner meaning for people that want to acquire new knowledge. These courses required no degree or prior experience. Moreover, only 3% are advanced courses.
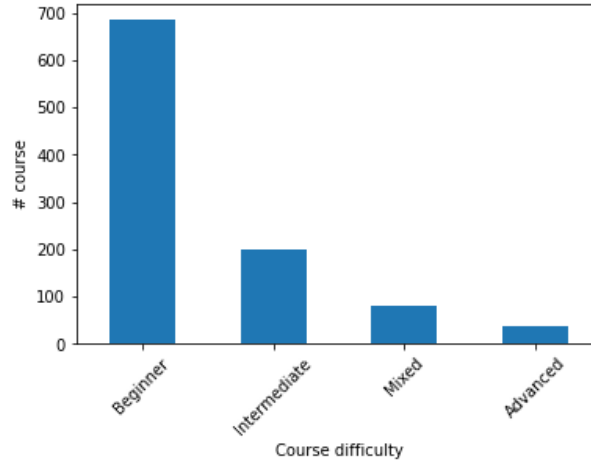

Figure 2: Number of courses proposed per difficulty

The Figure 3 shows that majority of course are simply courses (51%) or specialization (40%). Professional certificate and guided project represented a small part of the courses proposed by the Coursera platform (9%).
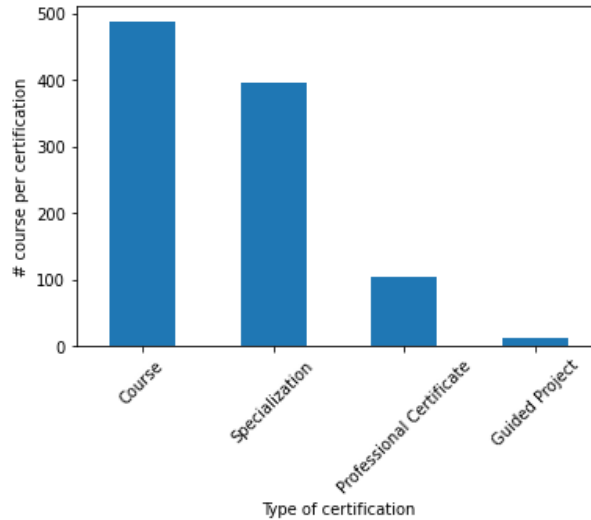

Figure 3: Number of courses proposed per certification type

Moreover, as shown in the Figure 4, 62% of the students are enrolled in courses while 26% are enrolled in specialization. When people registered them self to the platform they want to discover new skills and learn new knowledge or obtain specialization.
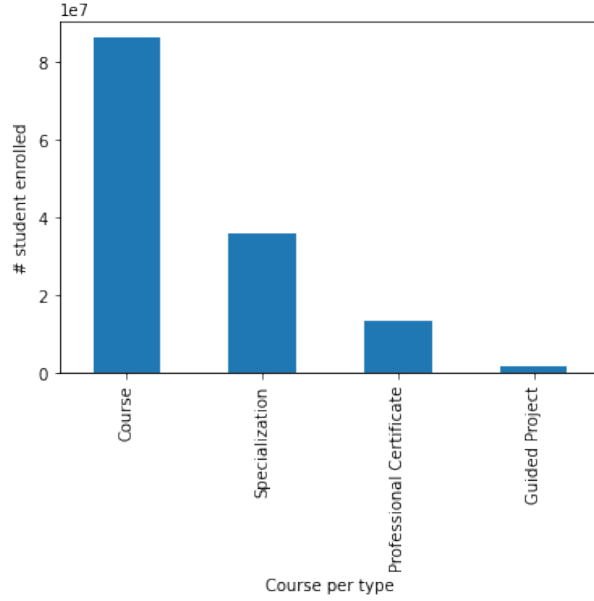
Figure 4: Number of students enrolled per certification type

Another variable of interest is the estimated time to complete the program that is summarized in the Table 2. Majority of the courses proposed has an estimated time between 3 and 6 months and the Figure 5 shows that concerns courses, specialization and professional certificate. Only the guided project could be reached in less than 2 hours and the level is beginner or intermediate.

Estimated time	# courses	Proportion
3 - 6 Months	460	48%
1 - 3 Months	333	34%
1 - 4 Weeks	196	20%
Less Than 2 Hours	11	1%

Table 2: Proportion of courses per estimated time

A last analysis is performed to find the Top 10 of universities and companies that proposed courses on the Coursera platform. The Figure 6 shows that Google proposes 10% of the courses, IBM 8% and University of Pennsylvania 6%. If Princeton, Stanford, the University of Michigan, and the University of Pennsylvania were the first universities to offer content on the platform today only the University of Pennsylvania is still in the Top 10 and behind Google and IBM.
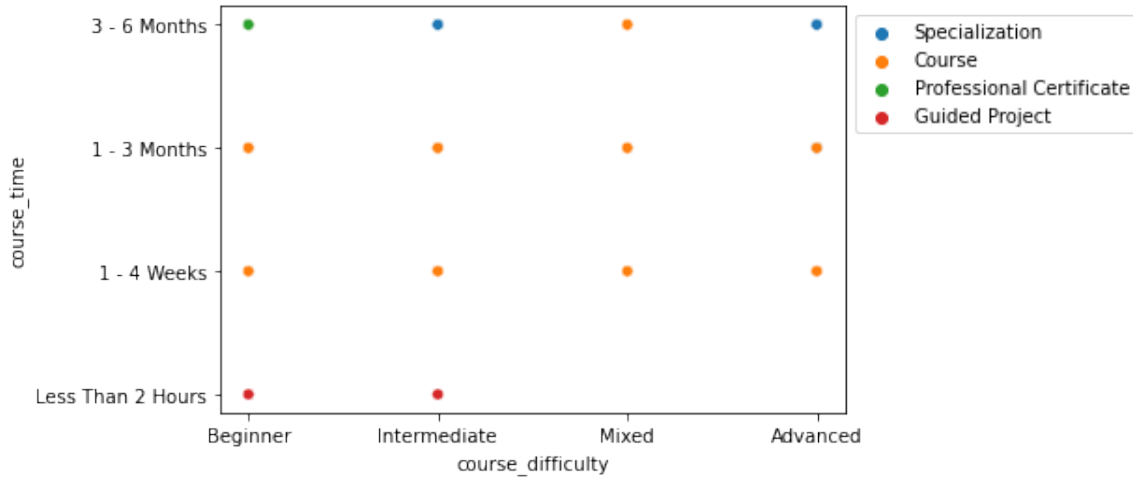
Figure 5: Time and difficulty per certification type

4 Discussion

The data set contains only a sample representing 16% of the total courses proposed on the Coursera platform which is a very small part. However, the analyzed data show that the majority of people are enrolled in courses and that's corresponding to the majority of courses proposed on the platform. They want to discover new skills and learn new knowledge and take a lot of time for that.

The Learner Outcomes Report for 2023 is an extensive collaboration between Coursera and Huron Consulting Group. This report compiles insights gleaned from over 55000 learners across 190 countries. Its purpose is to gauge how Coursera assists learners in achieving their professional and personal goals. The report underscores the influence of online learning. Key findings include:

- 77% of learners, and an astonishing 91% in developing economies, report career benefits.

- 95% of learners, with 99% in developing economies, experience personal growth and confidence.

- For unemployed learners, 30% secured employment after completing a course on Coursera.

- 1 in 4 Entry-Level Professional Certificate completer successfully transitioned to new jobs.

Finally, learners can leverage Coursera's diverse range of courses and programs to enhance their career prospects, acquire new skills, and boost personal confidence. The platform provides a gateway to growth and development, empowering individuals to navigate an ever-changing world.
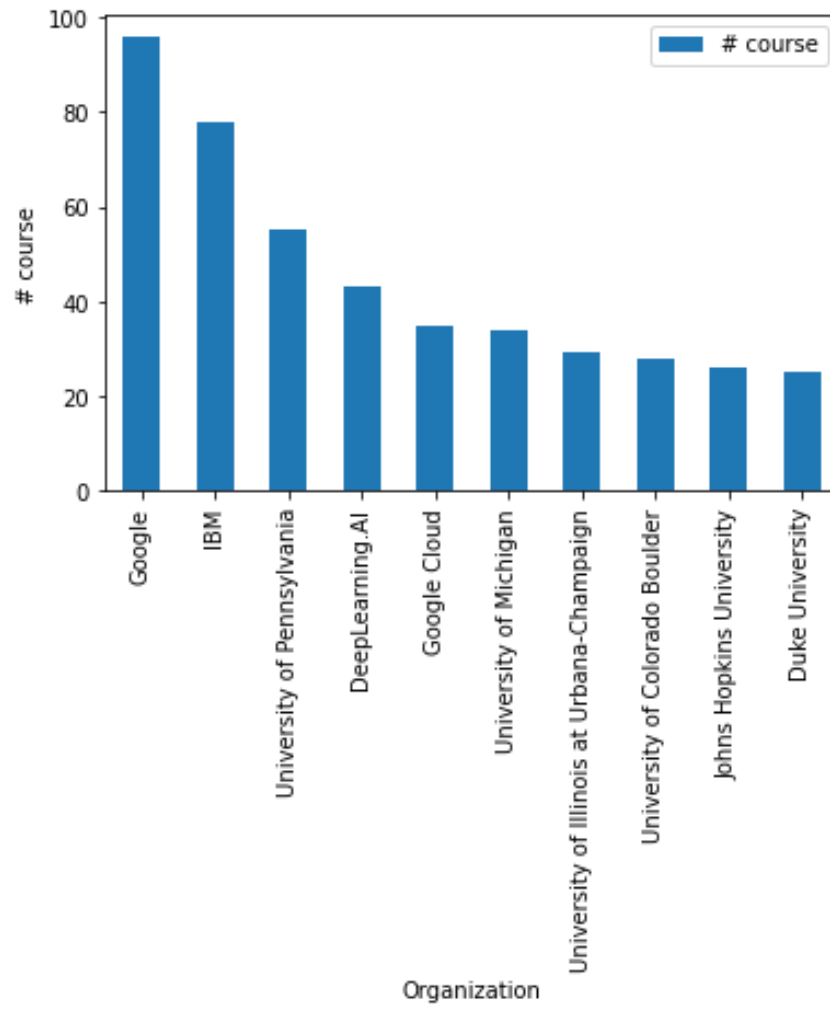
Figure 6: Number of courses proposed per organization (Top 10)