UDEMY USERS' ANALYSIS

by Gonçalo Bernardo

May 2023

DATA ANALYSIS PORTFOLIO

I believe everyone, from one way or another, has a story to tell and I'm not an exception. Since I can remember I have a passion for numbers and math in general. When I was a child, I used to spend long hours reading books about math and I really enjoyed applying my knowledge in a real context, making statistics with my toys, analyzing prices and discounts at the supermarkets, and so on.
Later in school I was encouraged to proceed with my studies specifically in math or applied math, statistics, but instead, I decided to take a master in mechanical engineering, which confirmed I should have chosen the "math path" because after working a few years in mechanical industry I had to quit and I jumped to a software developer carrier for other few years when finally I realized I really missed the numbers, the mathematical problems, the graphs, the calculations. I wasn't happy even if programming was giving me joy.
That's when I finally found data analysis and science would fit all my passions at once, so I started a new project to my life creating a learning center to teach math, giving more time to study data analysis, so I could, in a few months, have my dream job: data analyst.

PROFESSIONAL BACKGROUND

Having a degree in mechanical engineering didn't give me access to my dream job and after five years taking a MsC, I have worked for two years as a fellow researcher by programming algorithms in matlab to reconstruct surfaces from point clouds obtained by laser scanning.
Then, for a few years I worked as a quality assurance engineer and mechanical designer, but it didn't make me happy, so I quit my job to do a java and SQL bootcamp, which I finished in 4 months and started working as a IT consultant.
During three years I have worked as a full-stack software programmer and I have participated in the development of some projects, most of the time in the banking field. Those jobs gave me joy, especially because programming fits great in the type of tasks I love to do, and my teams were always supportive and my senior coworkers had taught me a lot.
However, something was missing and I should have been obvious from the beginning, but it took some time to realize that it was math: my biggest passion.
That's why I decided to change my path by opening my own business: a math learning center, which gives me more flexibility to manage my time and study data analysis and science so as soon as possible I proceed to my second part of my plan: become a data analyst.

TABLE OF CONTENTS

DATA ANALYSIS PORTFOLIO **2**
PROFESSIONAL BACKGROUND **2**
TABLE OF CONTENTS **3**
A1. UDEMY PROJECT DESCRIPTION **4**
A2. PROBLEM DEFINITION **5**
A3. DATA DESIGN **6**
A4. FINDINGS **7**
4.1 What are the total numbers of subscribers in each subject? 7
4.2 How does the average content duration/price/number of students vary across different subjects? 8
4.3 How many courses are free and paid for each subject? 10
4.4 What is the average price of web development courses at different levels? 10
4.5 What are the 20 most popular courses? 11
4.6 Does content duration impact the price of the course? 13
A5. ANALYSIS **14**
A6. CONCLUSIONS **16**
A7. RECOMMENDED ACTIONS **17**
B1. SCHOOL GRADES PROJECT **18**
B2. PROBLEM DEFINITION **19**
B3. DATA DESIGN **20**
B4. FINDINGS **21**
4.1 What are the average grades of each school? 21
4.2 Which school has the higher general rate of alcohol consumption? 22
4.3 How do fathers' education and the quality of family relationships influence students' average? 23
4.4 How living in a rural or urban region influences students' average grades? 25
4.5 Some students intend to get higher education. Does it really have an effect on their grades? 26
4.6 What are the main characteristics of the best 10 and worst 10 students? 27
B5. ANALYSIS **29**
B6. CONCLUSIONS **31**
B7. RECOMMENDED ACTIONS **32**
DATASET SOURCE **33**

1. UDEMY PROJECT DESCRIPTION

In this section, it is explained the description of the project, which can be observed in table 1.1.

Table 1.1 - Project description using SMART Method.

Situation:	Described in detail in section A3, the analyzed data set was collected to better understand the pricing and subscriptions numbers of Udemy courses, considering four different subjects. Initially our data set had some duplicated information, as well as gaps and wrongly attributed data types and it was used google sheets to clean it, organize it and create some graphics to better observe the data.
Task:	Section A2 gives more details about the key questions to be analyzed and answered, which pretend to identify possible trends and patterns related with the subject of the courses, the prices and if it is paid or free. Among others key questions, we search for information about the most popular courses and their subject or the effect of course levels in the prices.
Action:	In order to achieve the final results, and implement our analysis, the data was cleaned using google sheets and to perform the visualizations, it was used google sheets as well as tableau. Section A4 shows the main findings of the project.
Result:	The results are discussed in section A5 with more detail, but, generally speaking, it was found that web development courses were the most popular and that subject has 4x of business and financial courses' subscribers, and has almost 2x of content duration.

2. PROBLEM DEFINITION

The problem is more clearly defined if some questions to be answered are defined, so next we have the main questions we want to discuss along the report and try to find patterns and trends to better understand what kind of subject, levels and prices would give to the company the best market strategies.
So, we have considered the questions below:

- What are the total numbers of subscribers in each subject?
- How does the average content duration/price/number of students vary across different subjects?
- How many courses are free and paid for each subject?
- What is the average price of web development courses at different levels?
- What are the 20 most popular courses?
- Does content duration impact the price of the course?

3. DATA DESIGN

Our data was extracted and stored as a csv file, which was read and transformed by importing it with google sheets.
Initially the raw data had some gaps, i.e., blank spaces and/or rows, as well as, some duplicated information and incoherent subject type, which had to be corrected by using text substitution tool from google sheets.
The content duration columns had some issues related with data type which was supposed to be numerical but was assumed to be a date. Using google sheets' formulas the problem was solved in order to make it possible to proceed with the project.
All the cleaning was performed by google sheets and several graphics were created as well as by isolating the pretended columns or by constructing pivot tables in google sheets.
Additionally some visual content was created through tableau software, by importing our cleaned data and analyzing for each key question what would be the best visualizations to better "scan" our data.

4. FINDINGS

4.1 What are the total numbers of subscribers in each subject?

Concerning the first question, three circular graphs were created (fig. 4.1 to 4.3), which show the total subscribers by subject.

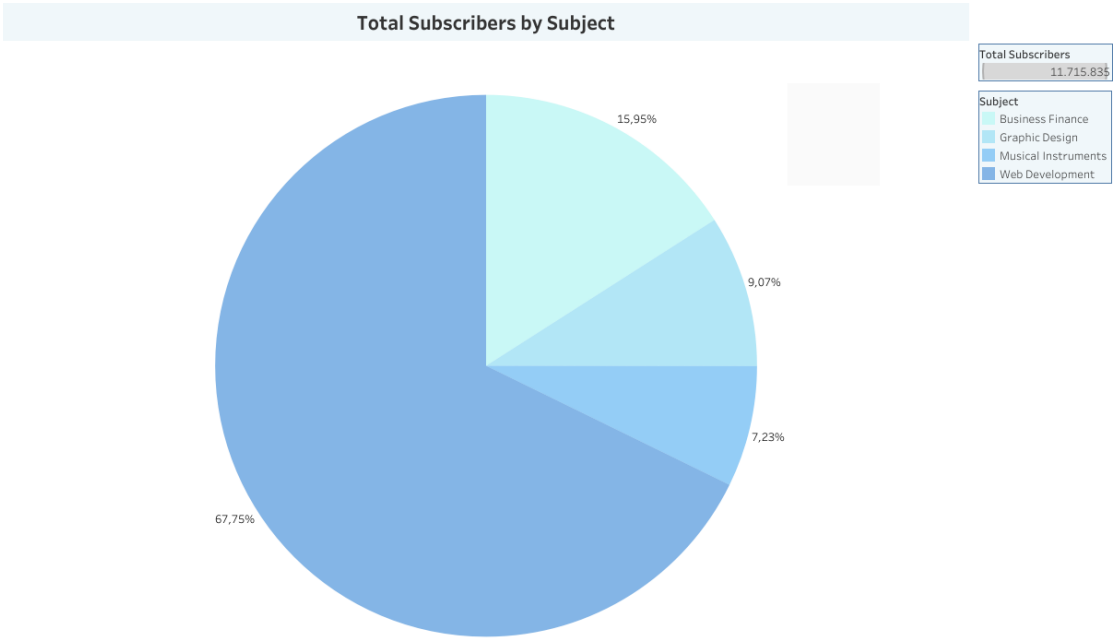

Fig. 4.1 - Total subscribers by subject.

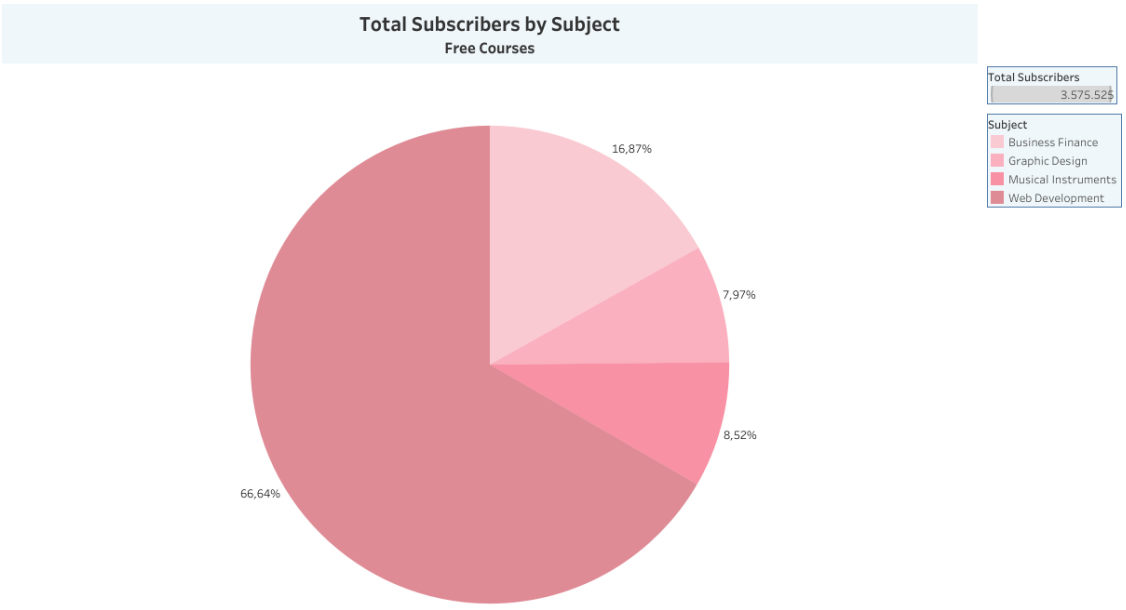

Fig. 4.2 - Total subscribers by subject (considering only free courses).

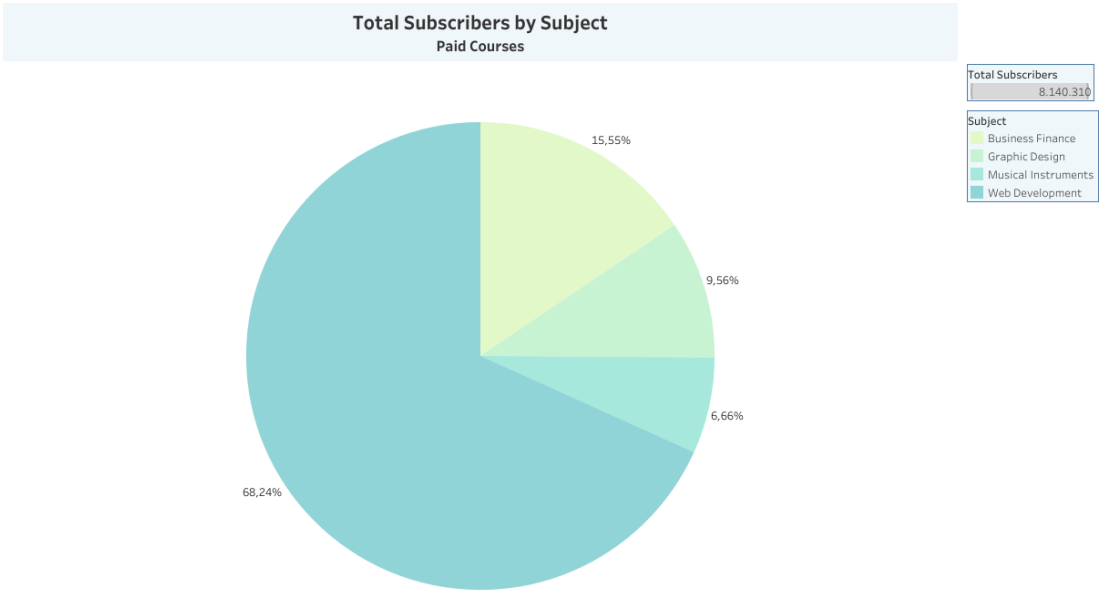

Fig. 4.3 - Total subscribers by subject (considering only paid courses).

4.2 How does the average content duration/price/number of students vary across different subjects?

The second question was studied by creating the bar graphs 4.4, 4.5 and 4.6 which shows the average content duration by subject, the average price by subject and number of students by subject, respectively.

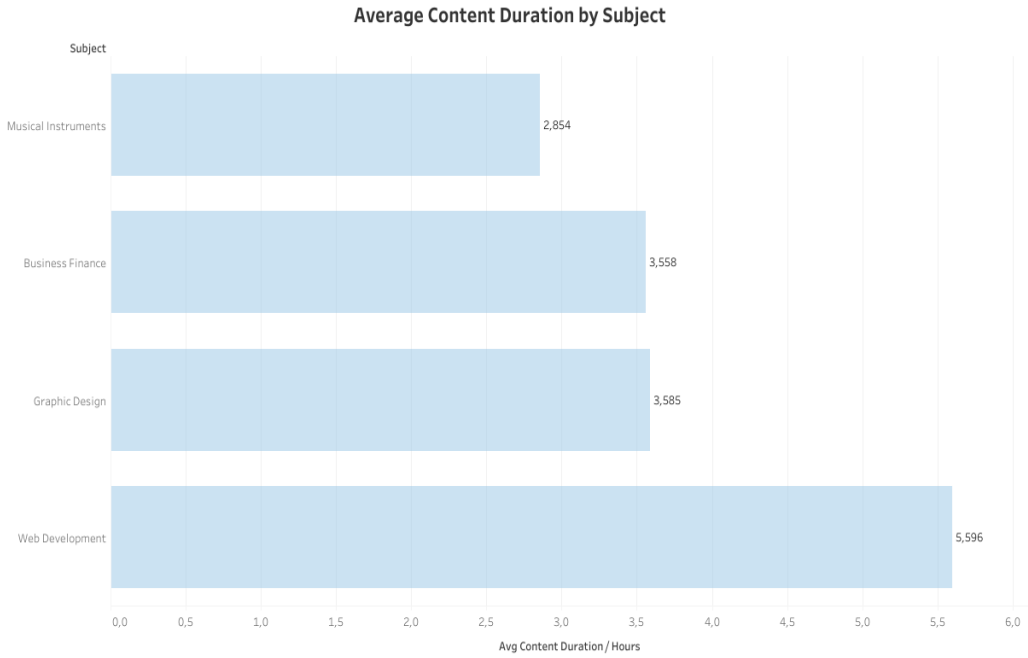

Fig. 4.4 - Average content duration by subject [Hours].

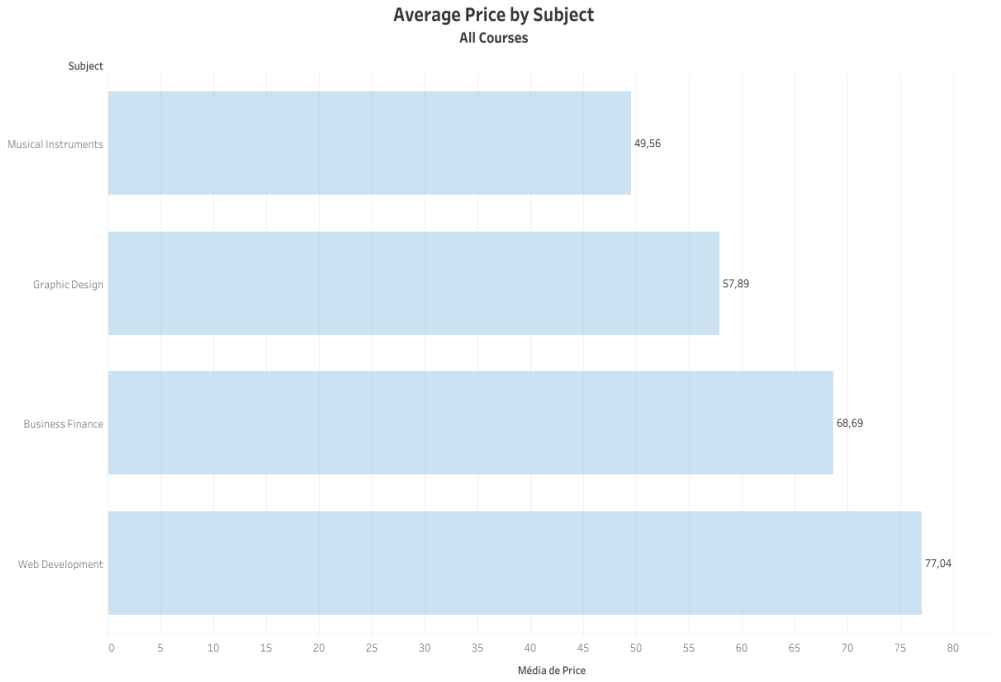

Fig. 4.5 - Average price by subject [€].

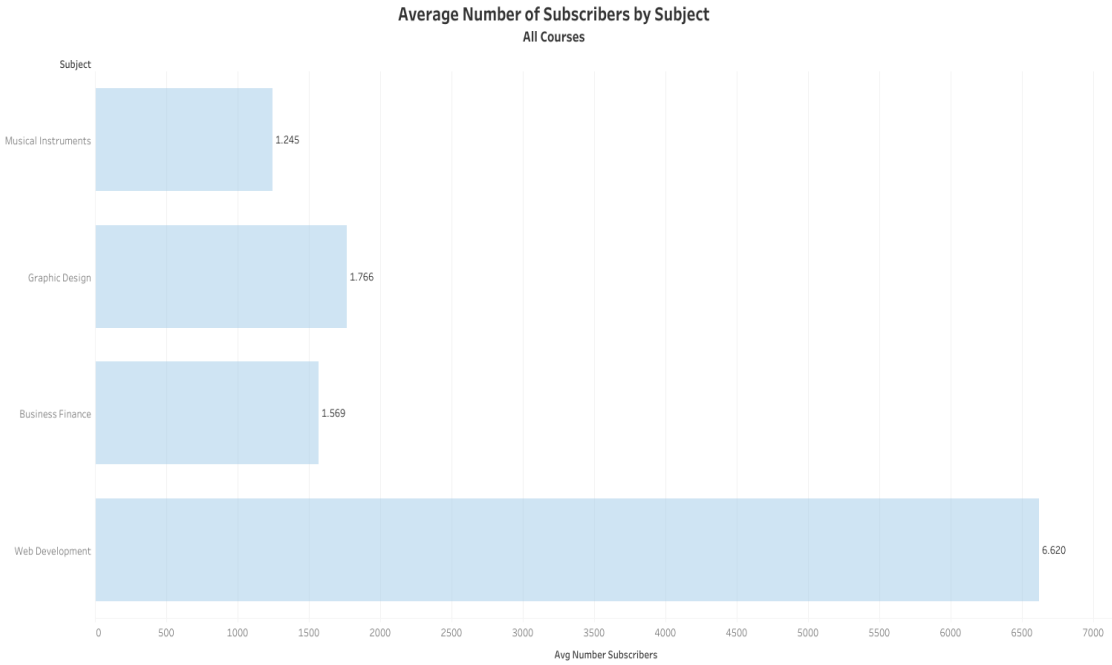

Fig. 4.6 - Average number of subscribers by subject.

4.3 How many courses are free and paid for each subject?

The third question was studied by creating a column graph to compare the percentage of free and paid courses of each subject. Next figure 4.7 shows the results.

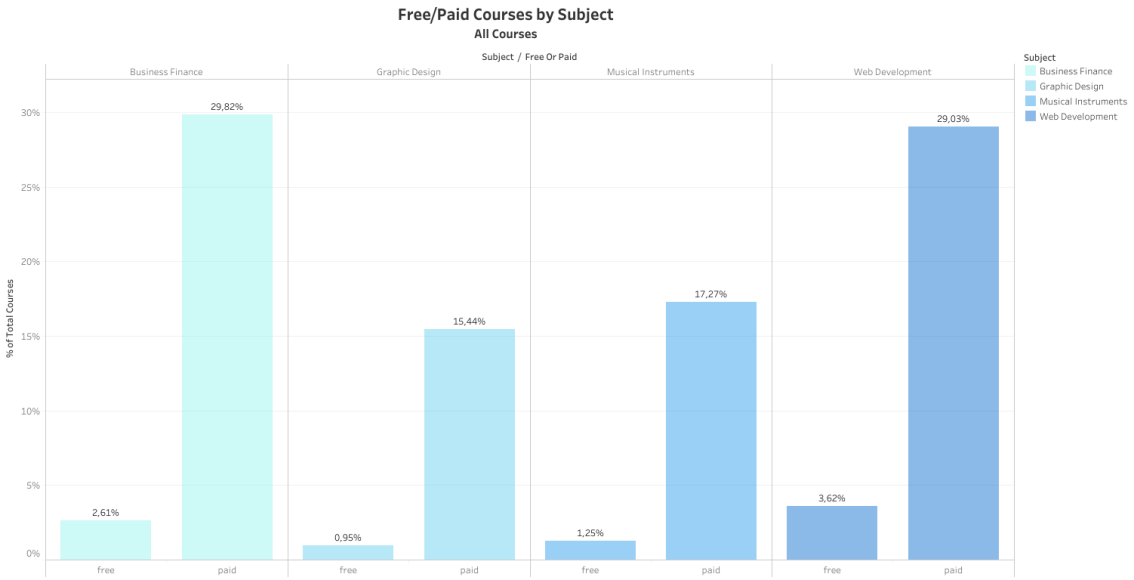

Fig. 4.7 - Free and paid courses by subject (among all courses).

4.4 What is the average price of web development courses at different levels?

The average price of web development courses at different levels was analyzed through a bar graph (fig. 4.8), which is shown below.

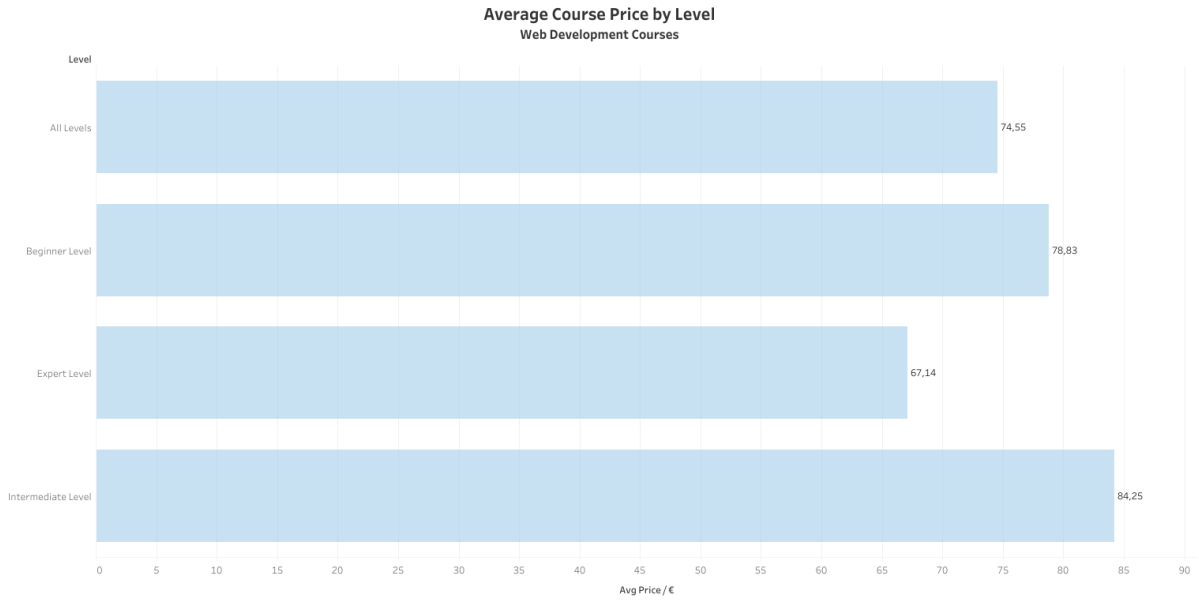

Fig. 4.8 - Average course price by level.

4.5 What are the 20 most popular courses?

Another relevant question is related to the trend or pattern we can find among the most popular courses. This subsection shows some detailed data related to the top 20 popular courses, i.e., the courses that have the higher number of subscribers. Tables 4.1 to 4.4, as well as, figure 4.9 show us that information.

Table 4.1 -The 20 most popular courses and their levels, free/paid information, content duration and total subscribers.

Subject	Course Title	Level	Free or Paid	BF or NBF	Content Duration	Total Subscribers
Web Development	Learn HTML5 Programming From Scratch	Beginner Level	free	BF	10,5	268923
	Coding for Entrepreneurs Basic	Expert Level	free	NBF	3,5	161029
	The Web Developer Bootcamp	Beginner Level	paid	NBF	43	121584
	Build Your First Website in 1 Week with HTML5 and CSS3	All Levels	free	NBF	3	120291
	The Complete Web Developer Course 2.0	Beginner Level	paid	NBF	30.5	114512
	Web Design for Web Developers: Build Beautiful Websites!	All Levels	free	NBF	3	98867
	Learn Javascript & JQuery From Scratch	All Levels	paid	NBF	2	84897
	Practical PHP: Master the Basics and Code Dynamic Websites	Intermediate Level	free	NBF	6,5	83737
	JavaScript: Understanding the Weird Parts	All Levels	paid	NBF	11,5	79612
	Angular 4 (formerly Angular 2) - The Complete Guide	Beginner Level	paid	NBF	22	73783
	Beginner Photoshop to HTML5 and CSS3	All Levels	free	NBF	2	73110
	Web Development By Doing: HTML / CSS From Scratch	All Levels	free	NBF	1	72932
	HTML and CSS for Beginners - Build a Website & Launch ONLINE	All Levels	free	NBF	6	70773
	Become a Web Developer from Scratch	All Levels	paid	NBF	27.5	69186
	Quickstart AngularJS	Beginner Level	free	BF	1,5	64128
	Learn Responsive Web Development from Scratch	All Levels	free	NBF	4,5	59639
	Learn and Understand AngularJS	Beginner Level	paid	NBF	7	59361
Musical Instruments	Free Beginner Electric Guitar Lessons	All Levels	free	NBF	4,5	101154
	Pianoforall - Incredible New Way To Learn Piano & Keyboard	Beginner Level	paid	NBF	30	75499
Business Finance	Bitcoin or How I Learned to Stop Worrying and Love Crypto	All Levels	free	NBF	8	65576

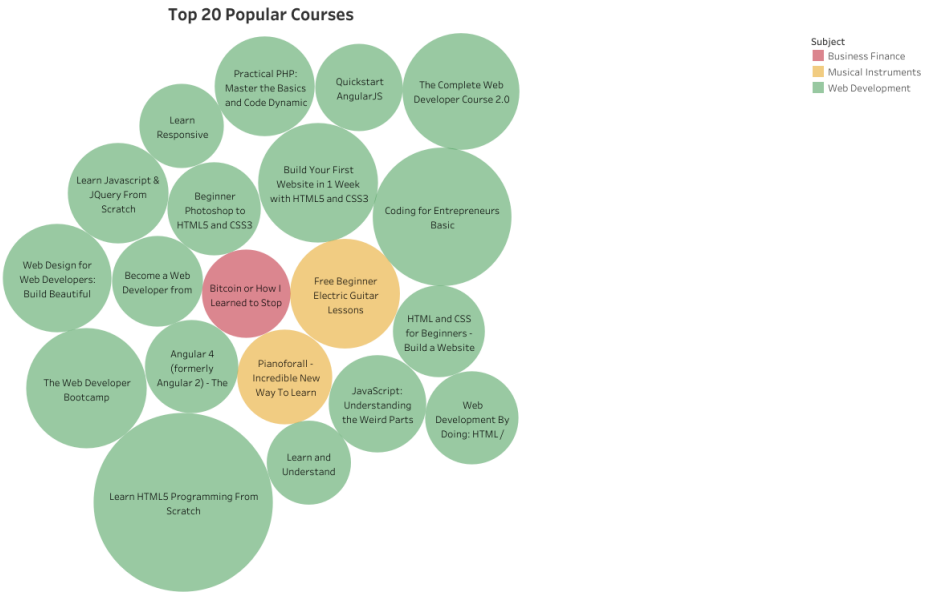

Fig. 4.9 - Top 20 popular courses.

Table 4.2 - Comparison between course subject and their level (among top 20).

Subject	Total	Level				Total (%)
		Beginner	Intermediate	Expert	All levels	
Business Finance	1	0,0%	0,0%	0,0%	5,0%	5,0%
Musical Instruments	2	5,0%	0,0%	0,0%	5,0%	10,0%
Web Development	17	30,0%	5,0%	5,0%	45,0%	85,0%
	20	35,0%	5,0%	5,0%	55,0%	100,0%

Table 4.3 - Comparison between course paid/free courses and their subject (among top 20).

Subject	Total	Free	Paid	All levels
Business Finance	1	5,0%	0.0%	5,0%
Musical Instruments	2	5,0%	5,0%	10,0%
Web Development	17	50,0%	35,0%	85,0%
	20	60,0%	40,0%	100,0%

Table 4.4 - Comparison between paid/free courses and their level (among top 20).

Level	Total	Free	Paid	All levels
All Levels	11	40,0%	15,0%	55,0%
Beginner Level	7	10,0%	25,0%	35,0%
Intermediate Level	1	5,0%	0.0%	5,0%
Expert Level	1	5,0%	0.0%	5,0%
	20	60,0%	40,0%	100,0%

4.6 Does content duration impact the price of the course?

The last question to be discussed was the impact of content duration in the course price. Figure 4.10 shows a scatter plot showing that information.

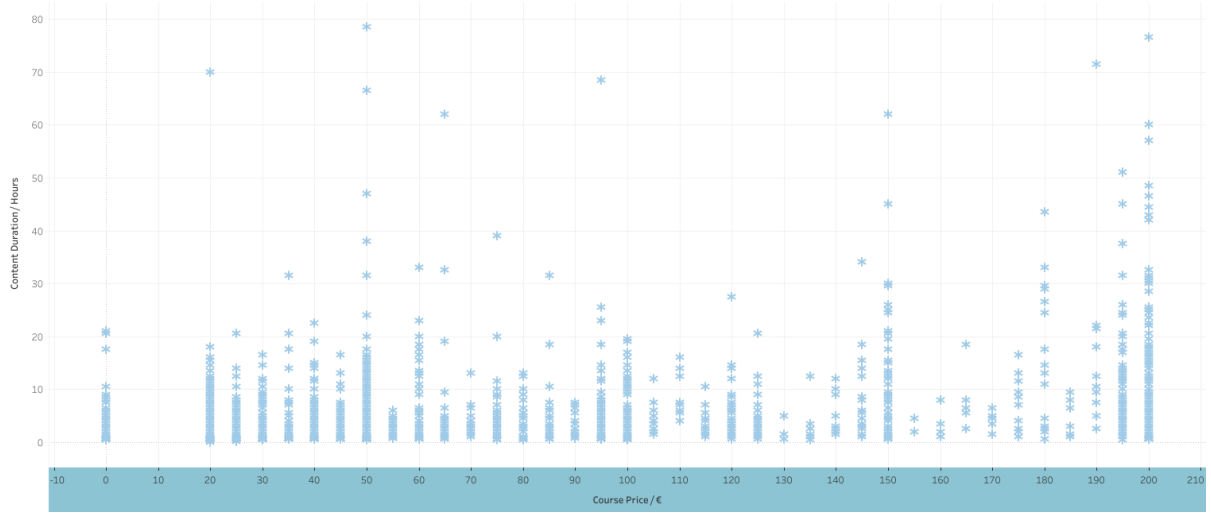

Fig. 4.10 - Content duration compared [Hours] with course price [€].

5. ANALYSIS

This section pretends to make some analysis considering the previous findings.
From the figures of subsection 4.1 we clearly can conclude that the proportion of web development subscribers' number compared with other courses is high, and the fact that the courses are paid/free, doesn't have a relevant effect on that proportion.
According to figure 4.4 in section 4, one can say that the average content duration is more likely to be high in web development courses, followed by business and financial courses and graphic design courses, which have a similar value.

Regarding the prices observed in figure 4.5, web development courses, on average, seem to be more expensive than the other subjects. Musical instruments courses have the lowest price, and contrary to web development courses, we can conclude the content duration has some influence in the courses prices, besides not being so clear in the case of graphic design courses and business and financial courses.

By observing all the three bar graphs of subsection 4.2, we can say that it is clear web development courses are the most popular courses and it must influence both content duration and prices, which are higher for web development courses.

However, the difference in the number of students across the distinct subjects is clearly the most relevant data collected because, with no doubt, we can say that the average number of students is much greater in web development courses.

Figure 4.7 from subsection 4.3, compares the data in terms of percentage and clearly, for all courses, the percentage of paid courses is greater than free courses.
The percentages of paid courses are around 90% for all courses, so we can conclude that the pattern is similar for all courses.

Concerning the average course price among the different levels, we can see they are similar in general, but we must highlight the fact that the average value for intermediate courses are higher than the others.

From figure 4.8, one can see the average prices among the different levels are similar in general, but we can highlight the fact that the average value for intermediate courses are higher than the others.

According to table 4.1, one can say that the most popular courses are mostly web development courses (87%) and there are many graphic design courses.

By analyzing tables of subsection 4.5, we can say that, generally, the amount of paid courses is higher compared with the free ones. However, among the most popular

courses, we have 60% that are free and intermediate and expert levels courses represent only 10% of the 20 top courses. More than half are “All levels” courses.
Courses that are free and beginner level at same time are also just two.

Finally, by observing the scatter plot in subsection 4.6, we can see that a few courses costing around 200€ have more than 40 hours of content and the longest free course has 20 hours of content.

However, we can’t conclude that the content duration has a strong relation with the price of the courses, once we can find several paid courses that are short as well the opposite.

6. CONCLUSIONS

Concerning our data and analysis, we are able to proceed to some conclusions.
One of the most relevant is the fact that 68% of the subscribers com web development courses and the top 20 courses are for that topic as well.

Regarding business and financial courses, we can conclude that even the amount of total courses is similar to the amount of web development courses, the content duration is almost two times higher compared to the second subject and the number of subscribers is four times higher in case of web development courses.

Another relevant conclusion is related to the prices. The intermediate level course tends to be more expensive than the other level and the expert level courses are the cheapest courses.

7. RECOMMENDED ACTIONS

After our analysis, we can proceed to make some recommendations concerning the conclusions we were able to do.
In that way, we can divide the topic in to distinct types of recommendations, mentioned below:

Product recommendations:

- Focus on getting more paid course by creators in web development subject, especially all levels and beginner levels, once among the 20 most popular courses, 90% have such levels;

Marketing recommendations:

- Our data tells us that our web development courses are most popular and people are willing to pay for it. Marketing and advertising campaigns should therefore redirect the attention for those kinds of courses by highlighting the relevance of having knowledge of this subject.
- All levels of web development should be advertised too because our data also says that most of the 20 top courses have such levels, which correspond to general courses letting people learn topics from scratch.