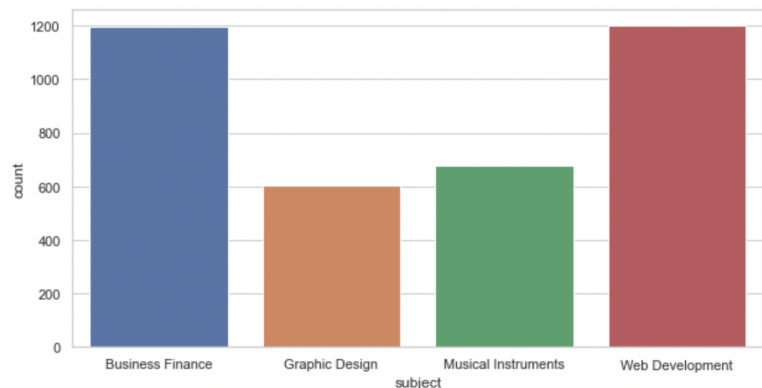
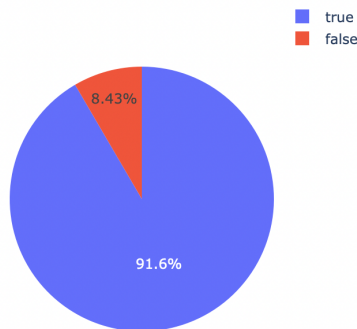


Data Visualization and statistical analysis on Udemy courses

Udemy is an online educational platform which provides courses on various subjects. Some courses are paid and some are free. And there are different levels for all the courses, few courses are intermediate, beginner and expert levels. Each course has different price. Udemy data also provides the number of reviews and subscribers for each course as well. So, let's begin and see some tables and graphs to explore and understand the data well for good analysis.

The Purpose of this report is to gain some business insights for the development of Udemy, based on some basic data analysis and statistical analysis. And to see how the business can grow!

Free Vs Paid courses



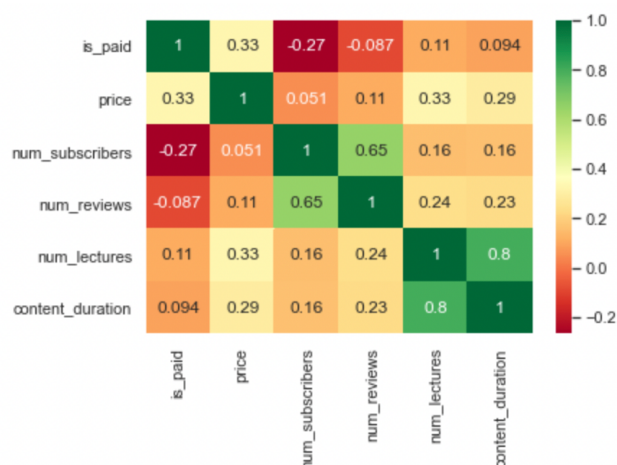
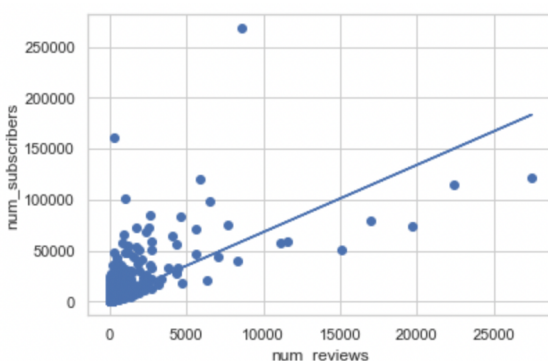
- Based on the pie diagram, we can see that there are 91.6% courses are paid and remaining 8.43% of courses are free courses.
- Based on second graph, we can say that 'Business Finance' and 'Web Development' courses are in high demand with 1200 courses on each subject.

If the business requirement is to check how the company can expand its growth, and to increase the number of subscribers, then we must check if there is any correlation between number of subscribers and number of reviews in Udemy data set. To answer this question, we should see whether if there is positive correlation between them.

Correlation test is used in data analysis for better comprehension of the two variables.

- Positive correlation speaks about the increase in one variable implies the increase in the other.
- Negative correlation speaks about the decrease in one variable implies the decrease in the other.
- Zero correlation says that the two variables are not related in any way.

Let's perform scatterplot and Heat map to find Pearson's correlation between subscribers and reviews.



There is a positive correlation (65%) for "subscribers and reviews", means the more reviews, the more the subscribers will be. And for "number of lectures and content duration" Pearson correlation is about 80%, means the more lectures the more the content duration will be.