Online Learning Platform Comparison Analysis Using Python

B RAGHUNATH

Problem Statement:

With the growing popularity of online education, learners are often overwhelmed by the multitude of platforms available, each offering unique features and benefits. Choosing the right platform can be challenging due to differences in pricing, course offerings, and the perceived value of certificates. This project seeks to compare key factors, such as course variety, instructor quality, pricing, and certificate credibility, and provide a data-driven approach to help users make informed decisions about which platform best aligns with their educational and career goals.

Proposed Methodology:

Data Collection: Publicly available datasets from platforms like Kaggle (e.g., Udemy course datasets, Coursera course data).

Key Data Points to Gather:

Course Duration,Course Title, Category,Review Count and Rating: To analyse the variety of courses available on each platform. Instructor Ratings To evaluate the quality of instructors and platform chosen for these are Udemy and Coursera.

Attributes for the dataset:

Course Category

Course Name

Duration (in hrs)

Number of Students Enrolled

IF Free/Not

Price(in $)

Course Level

Rating

Review Count

Data Exploration and Cleaning:

The dataset, sourced from Kaggle, comprises a diverse range of courses across subjects like Business Finance, Web Development, Musical Instruments, and Graphic Design. Initial data exploration revealed the need for data cleaning and preprocessing to ensure data quality.

Key Data Points to Gather:

Following are some of the key attributes for analysis of course data:

Course Category, Duration (in hrs),Number of Students Enrolled, IF Free/Not, Price(in $),Course Level,Rating,Review Count.

Also to compare both the platforms precisely key metrics such as **average rating** weighted by **number of reviews** (which gives more weight to courses with higher reviews), is used to calculate the weighted average rating and then visualize it.

Key insights for data Analysis:

The insights can be classified into categories for better organization:

1. Analysis based course enrolment

- Course enrolments Analysis Based on Category
- enrolments by Course Level
- enrolment Based on Free vs Paid

2. Analysis on Review and Popularity

- Review Count vs Rating
- Course Popularity by Review Count
- Most Popular Courses Word Cloud

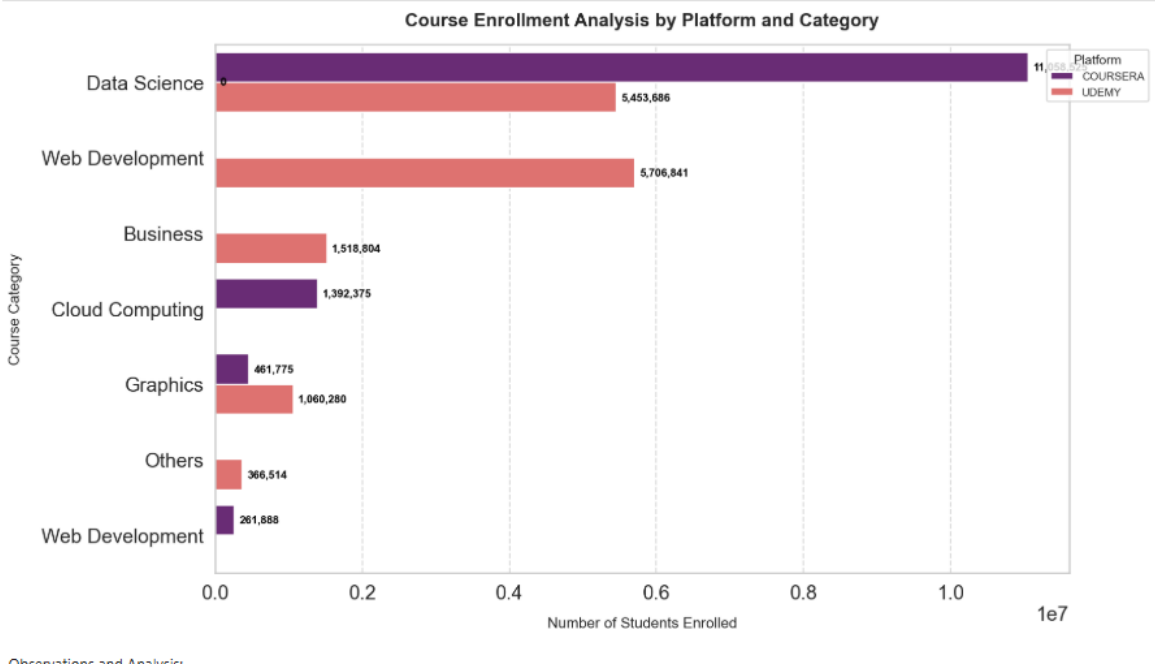
3. Analysis on Pricing and Ratings

- Free vs Paid enrolment by Platform
- Rating by Price Category
- Course enrolment by Price Category
- Weighted Average Rating by Platform

4. Analysis on Course Duration

- Correlation Between Course Duration and Rating
- Average Duration by Course Category

I)Analysis based course enrolment:



Observations: Highest Enrolments on Each Platform:

COURSERA->Data Science-> 11058525

UDEMY-> Web Development->5706841

Categories Consistently High Across Both Platforms:

Course Category Number of Students Enrolled Platform

Data Science 058525-> COURSERA

Web Development 706841-> UDEMY

Data Science 5453686-> UDEMY

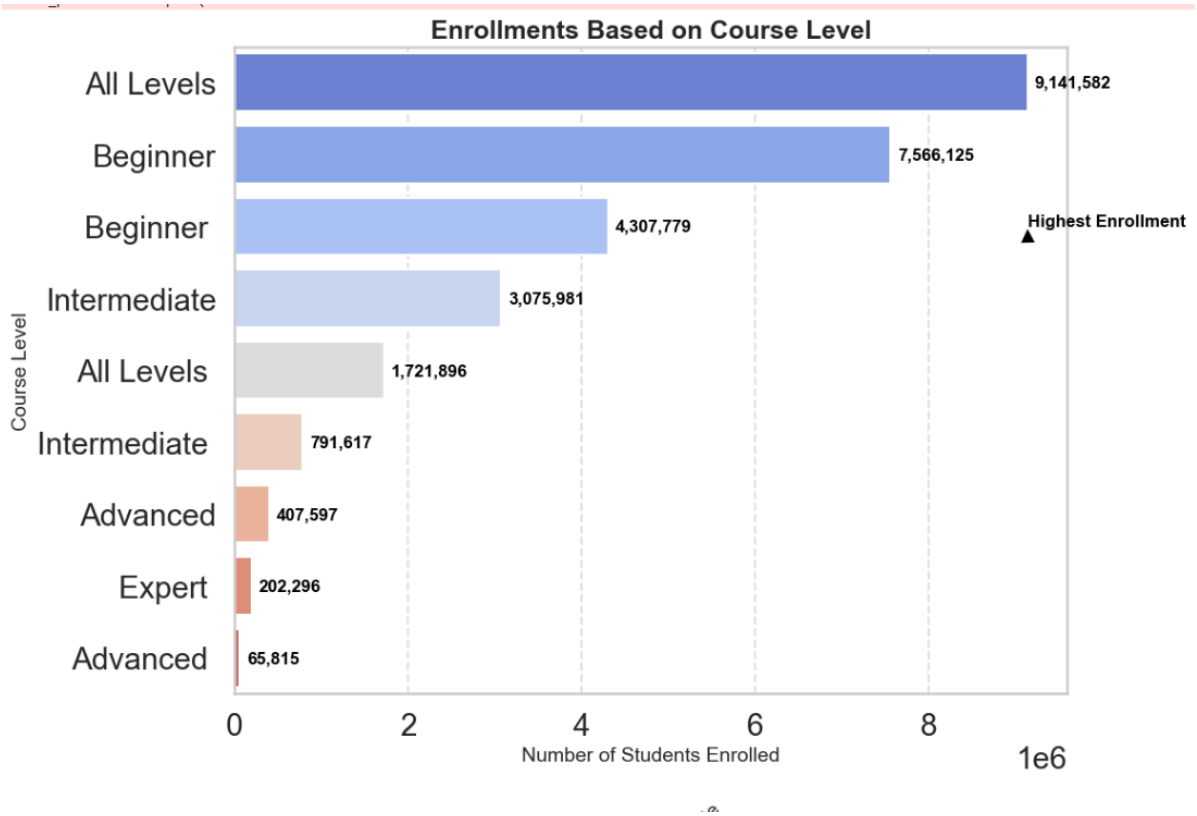
1.Which course categories have the highest student enrolments on each platform? Outcome: Identify the most popular categories among learners.

Answer: Coursera: The course category with the highest student enrolments is Data Science, as it shows a significantly larger bar than any other category. Udemy: The most popular category is Web Development, followed closely by Data Science.

2.Do certain categories consistently have higher enrolments across both platforms? Outcome: Highlight universal interest areas for learners.

Answer: Data Science consistently has high enrolments across both platforms, indicating it is a universally popular category for learners. Categories like Graphics and Others have lower enrolments across both platforms, showing less universal interest compared to Data Science or Web Development.

ii)Enrollment Analysis Based on Course Level:

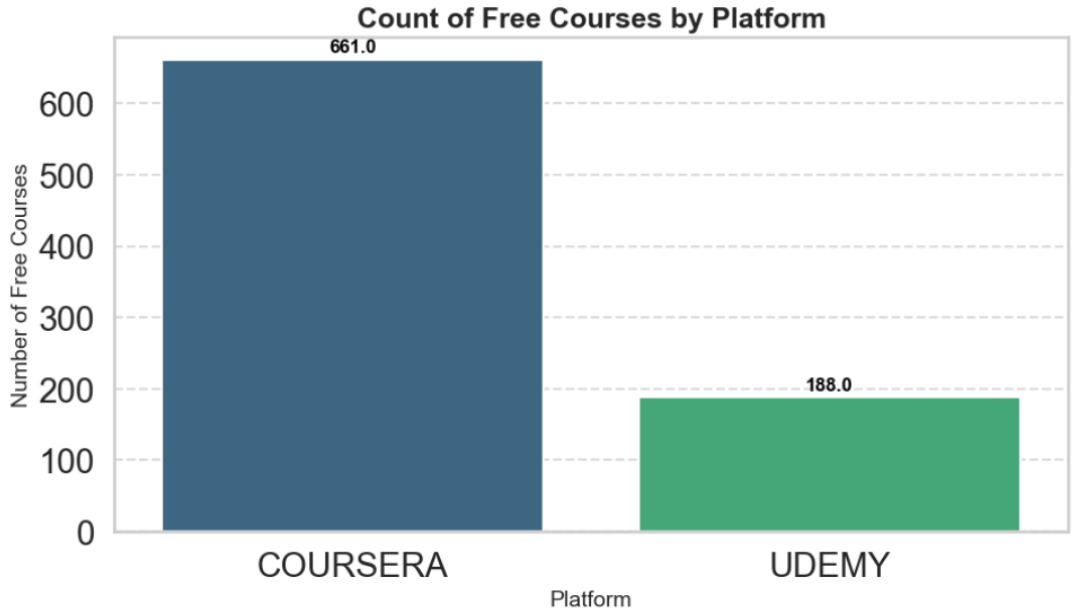


Observations: Additional Insights: • All Levels courses have a significantly higher enrolment compared to Beginner, Intermediate, Advanced, and Expert courses. • Beginner courses have the second-highest enrolment, followed by Intermediate courses. • Advanced and Expert courses have the lowest enrolments. The Analysis suggests that courses designed for learners of all skill levels, from beginners to advanced, are the most popular choice among students.

3.Which course level attracts the most students?

Answer: Based on analysis both the bar chart and the pie chart titled "Enrolments by Course Level," it's clear that All Levels courses attract the most students.

iii)Analysis of Enrollment Based on Free vs Paid across both the platforms



Observations:

COURSERA: 661 free courses

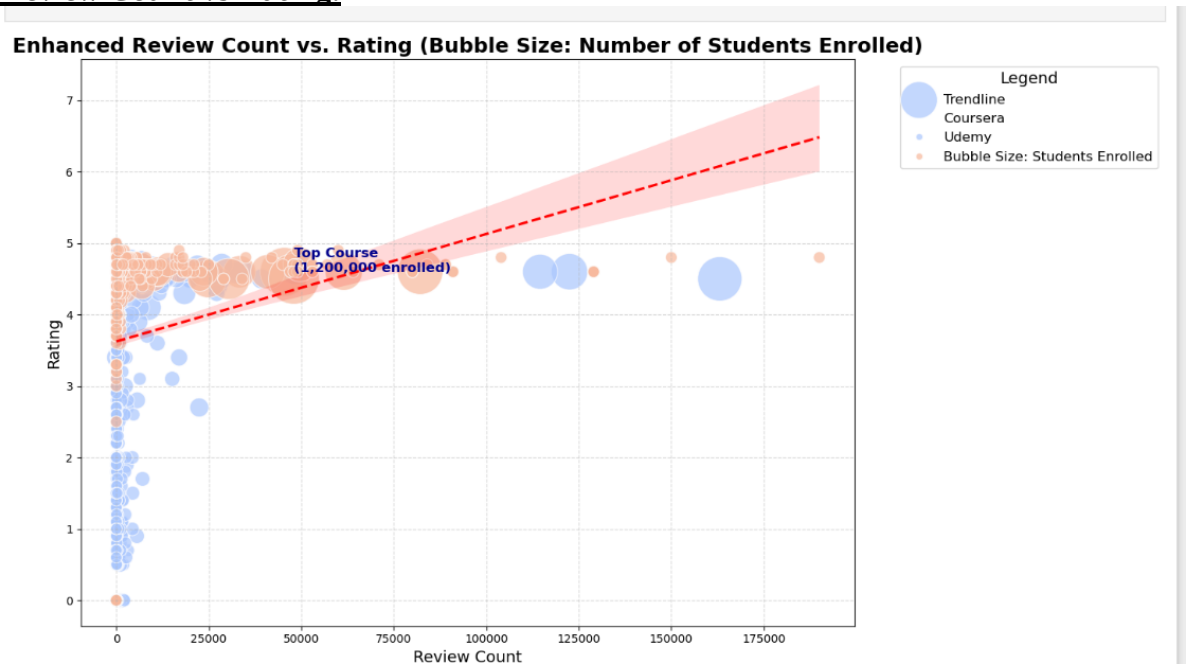
UDEMY: 188 free courses

4.Which platform offers the most free courses?

Answer: Based on analysis it is clear that course era offers more free course than udemy.

II)<u>Review and Popularity Analysis:</u>

i)Review Count vs Rating:



Observations:

The dotted black line in your plot represents a regression line, which is a statistical model that shows the trend or relationship between two variables. In this case, the dotted black line is a linear regression line that fits the relationship between: Review Count (on the x-axis) and Course Rating (on the y-axis).

The dotted black regression line is moving upward, it indicates a positive correlation between Review Count and Course Rating.

Specifically: As the number of reviews increases, the course rating tends to increase as well. This suggests that courses with more reviews generally receive higher ratings.

Further Observations: A higher number of reviews could imply that more learners are interacting with the course, and as more users rate it, the average rating could be higher. This might indicate that courses with a larger audience or wider popularity (more reviews) tend to have a higher overall quality, or at least appear to have higher ratings.

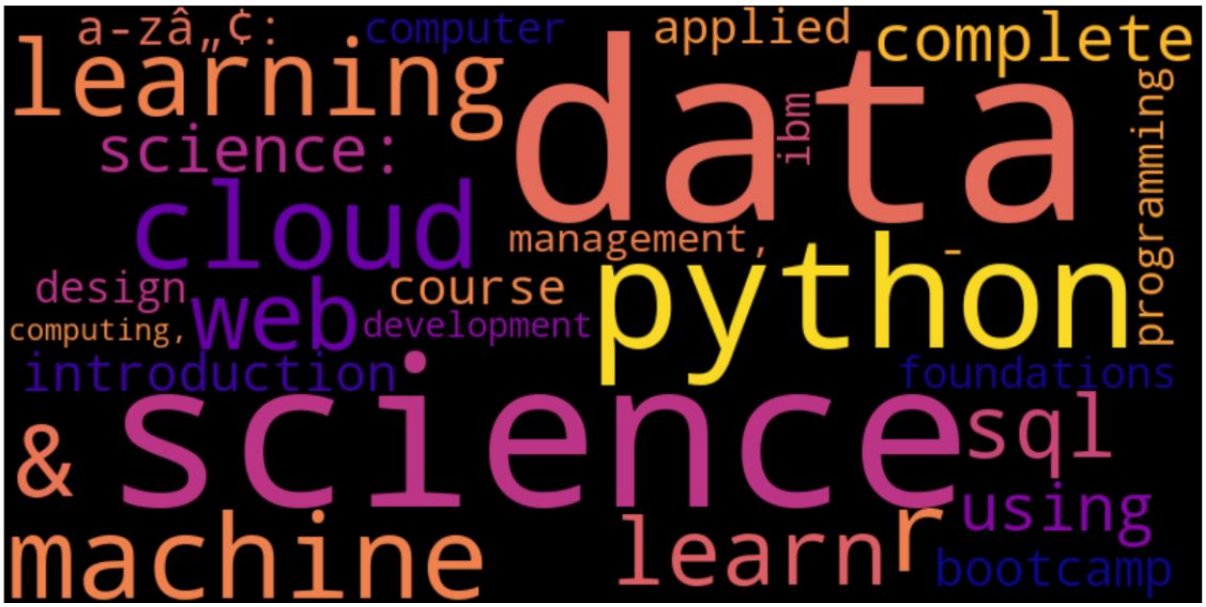
5)How does the number of reviews relate to the overall course rating?

Answer: There is a positive correlation: as the number of reviews increases, the course rating tends to be higher. More reviews often reflect better course quality.

6)Do courses with more enrolled students tend to have higher ratings or reviews?

Answer: Courses with more enrolled students generally have more reviews and higher ratings, indicating that popular courses tend to be well-rated.

iii)Most Popular Courses Word Cloud:

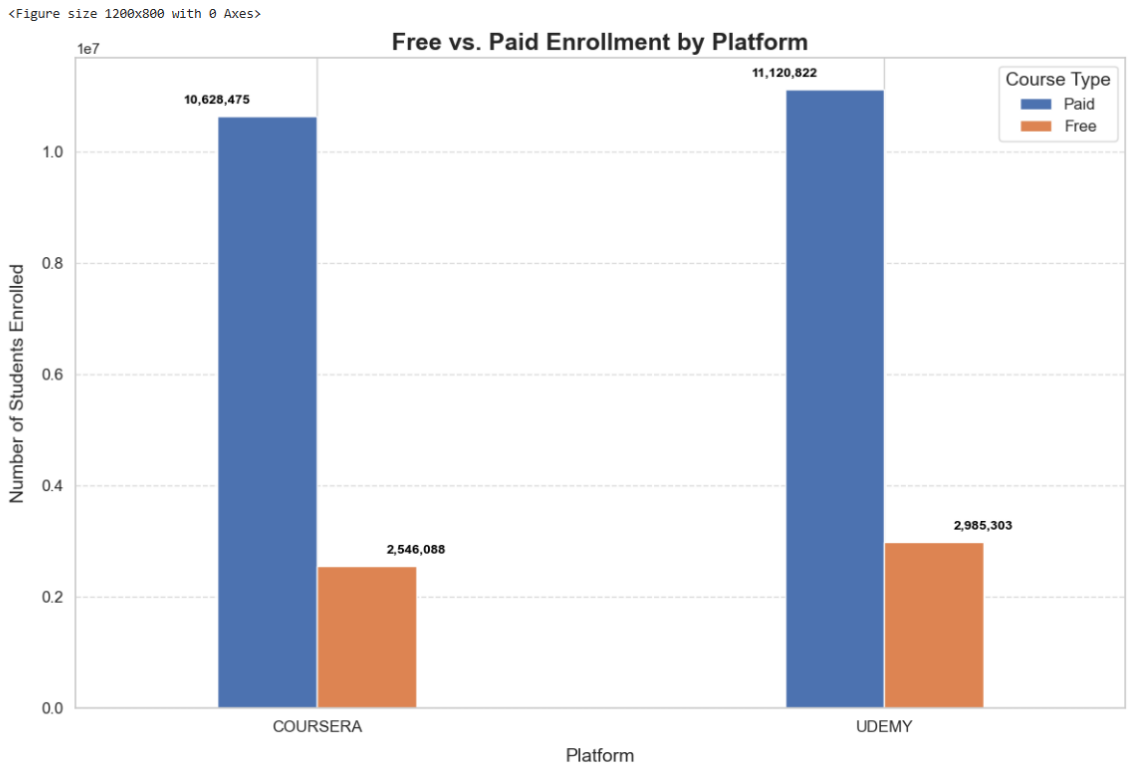


7. What is the most popular course keyword based on student enrolments?

Answer: The most popular course keyword based on student enrolments is "data".

III)Analysis on Pricing and Ratings:

i)Free vs Paid Enrolment:



Observations:

Enrolment Summary by Platform and Course Type:

COURSERA: 10,628,475 students enrolled in Paid courses.

COURSERA: 2,546,088 students enrolled in Free courses.

UDEMY: 11,120,822 students enrolled in Paid courses.

UDEMY: 2,985,303 students enrolled in Free courses.

8.Which platform sees higher student interest in free courses versus paid ones?
Outcome: Compare enrolments for free vs. paid courses for each platform.

Answer: Udemy has a higher proportion of free enrolments compared to paid ones, indicating greater student interest in free courses on this platform.

i)Rating By Price Category :

9. **Which price category has the highest average rating?**

Answer: Based on the bar chart titled "Average Ratings by Price Category," it appears that the Free price category has the highest average rating Followed by High Price Rating Shows that more number of students have enrolled in free category.

10.Do higher-priced courses tend to receive better ratings on average?

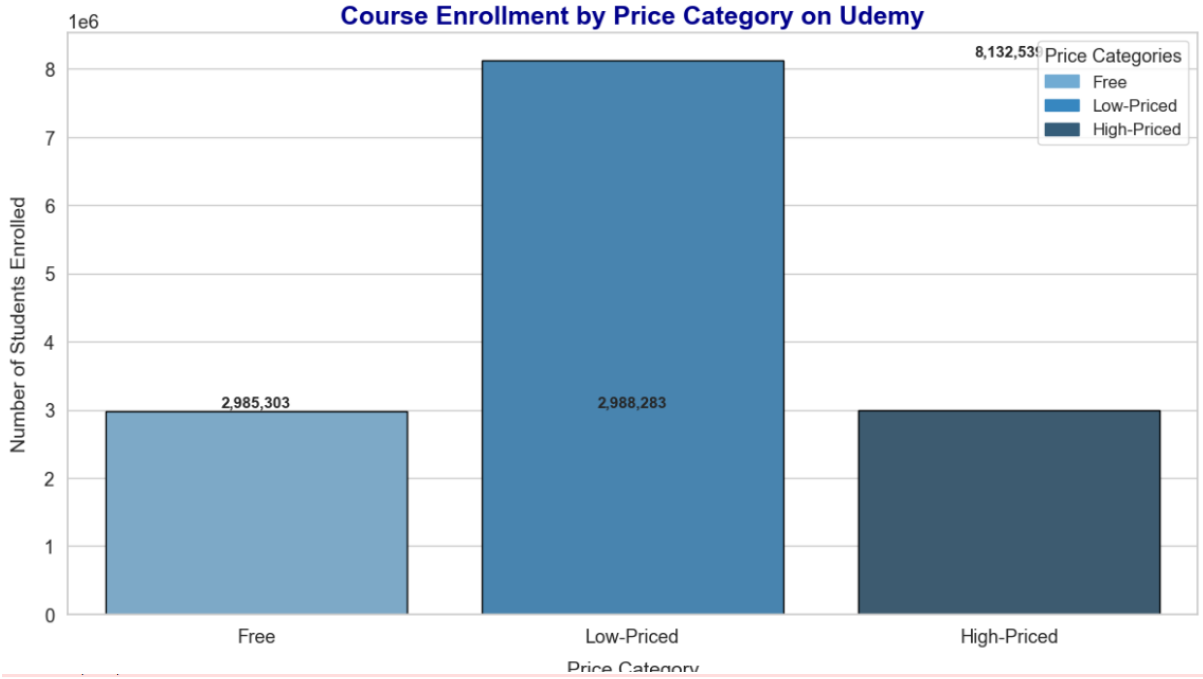
Answer: From the boxplot: Higher price categories generally have higher medians for ratings. This suggests that higher-priced courses might be receiving better ratings on average.

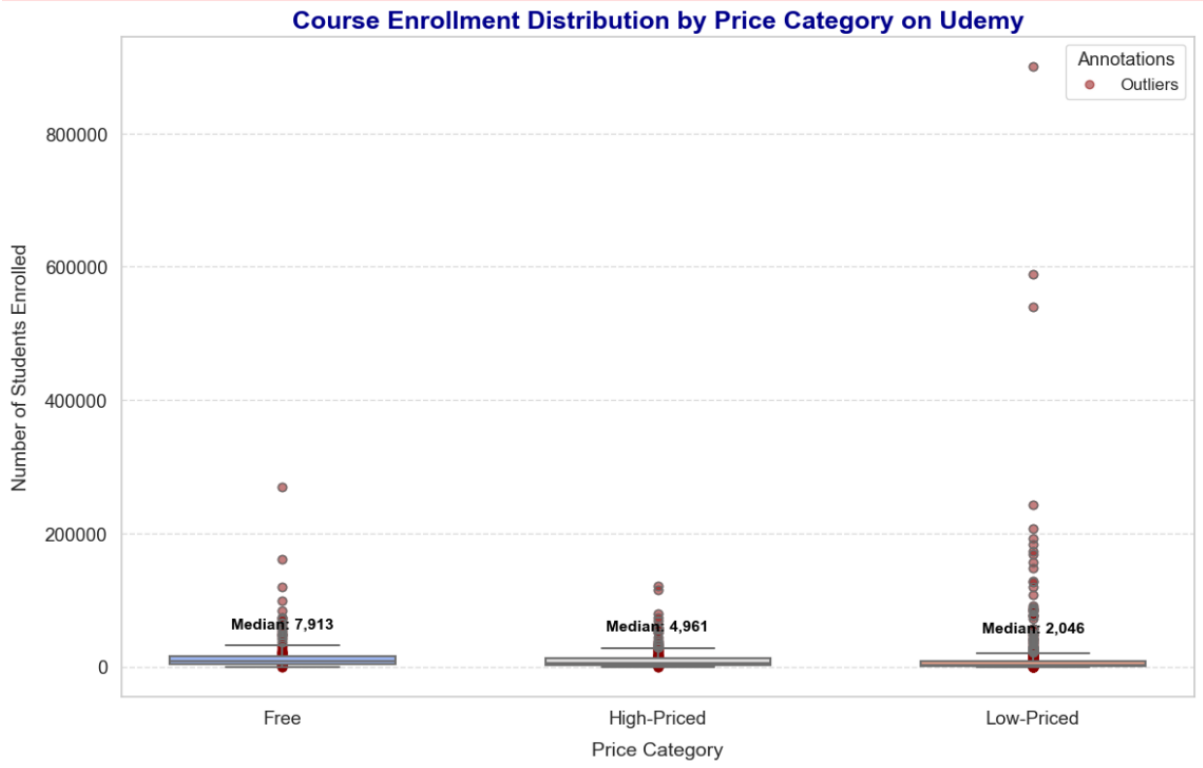
The lowest price category has the smallest median rating and a larger number of outliers, which indicates more variability in user satisfaction. However, the differences in medians are not dramatic, so other factors (e.g., course content, quality, or instructor reputation) might also influence ratings.

10.Which price category shows the widest range of ratings for courses?

Answer: The range of ratings for each category can be determined by the length of the box (interquartile range, IQR) and the whiskers: The widest range appears to be in the second or third price categories, where the boxes and whiskers are longer compared to other categories. The lowest price category has the smallest range of ratings, suggesting that user experiences are relatively concentrated within this group.

iii) Course Enrollment by Price Category(Udemy)





Observations:

In the 'Free' category, there are 188 courses with a total enrolment of 2,985,303 students. The median enrolment is 7,913. There are 20 outliers in this category.

In the 'Low-Priced' category, there are 586 courses with a total enrolment of 8,132,539 students. The median enrolment is 2,046. There are 64 outliers in this category.

In the 'High-Priced' category, there are 285 courses with a total enrolment of 2,988,283 students. The median enrolment is 4,961. There are 22 outliers in this category.

11. **Which price category attracts the highest number of student enrollment on Udemy?**

Answer: Free courses attract the highest number of student enrollment on Udemy. Explanation: • In the box plot, the median line for the "Free" category is the highest among all three categories. • The median represents the middle value, so a higher median indicates that a larger number of students are enrolled in free courses.

12. Do higher-priced courses see significantly fewer enrollment compared to free or low-priced ones?

Answer: Yes, higher-priced courses generally see significantly fewer enrolments compared to free or low-priced ones. Explanation: • The box for the "High-Priced" category is the lowest among all three categories. This indicates that the majority of enrolments for higher-priced courses are lower than the majority of enrolments for free and low-priced courses. • Additionally, the "High-Priced" category has a few outliers with very high enrolments. These outliers might be due to popular courses or instructors, but they don't represent the typical enrolment pattern for higher-priced courses.

Further Observations:

Free Courses:

Key Findings:

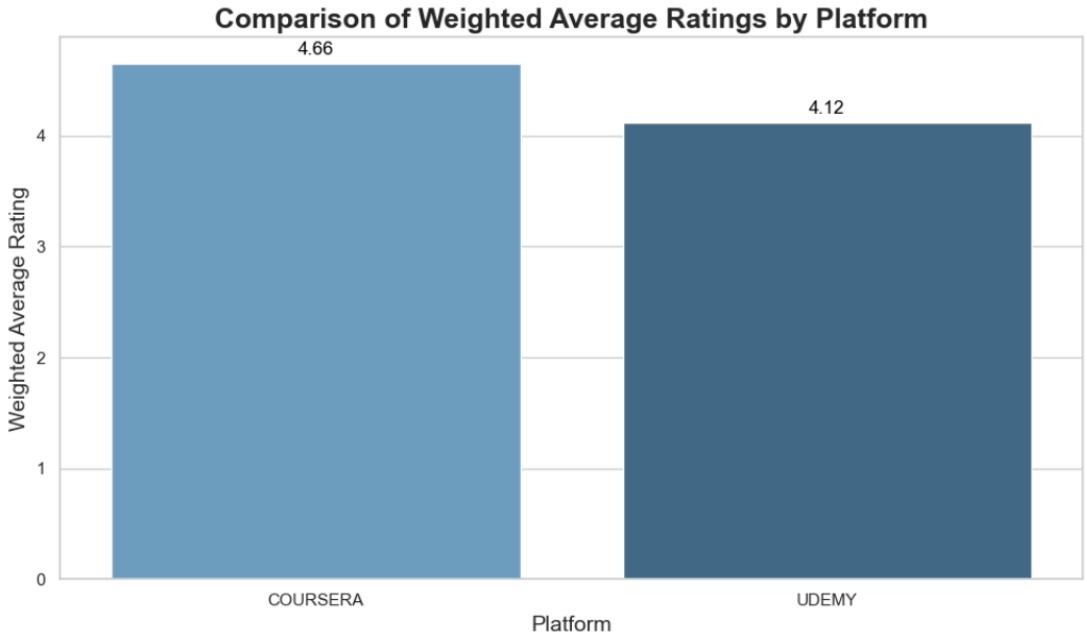
1. **Free Courses**:
    - **Highest Median Enrollment**: 7,913 students per course.
    - **Wide Enrollment Range**: Some courses attract over 200,000 students, with extreme outliers near 800,000.
    - **Conclusion**: Free courses are highly attractive, likely due to the zero cost barrier, making them accessible to a broader audience.
2. **High-Priced Courses**:
    - **Median Enrollment**: 4,961 students per course, lower than free courses.
    - **Outliers**: Fewer high-enrollment outliers compared to free courses, but some still perform exceptionally well.
    - **Conclusion**: While not as widely popular, high-priced courses with strong reputations or quality content can still attract significant enrollment.
3. **Low-Priced Courses**:
    - **Lowest Median Enrollment**: 2,046 students per course.
    - **Narrow Range**: Fewer outliers and generally less enrollment variability.
    - **Conclusion**: Low-priced courses may not appeal as much, potentially because they are perceived as lower value compared to free or premium courses.
4. **Overall Observations**:

- **Popularity of Free Courses**: Free offerings dominate enrollment figures, likely due to their accessibility and broad appeal.
- **Price vs Enrollment**: A higher price often corresponds with lower enrollment, but standout courses can defy this trend.
- **Student Perception**: Students may equate price with value, favoring free or premium options over mid-tier pricing.

Implications:

- **For Instructors**:
  - Offering free courses can help maximize reach and build a broad audience.
  - High-priced courses should focus on delivering premium content or leveraging strong branding to justify their cost.
  - Low-priced courses may need better marketing or clearer value propositions to stand out.
- **For Udemy**:
  - Encourage more free or freemium models to attract users to the platform.
  - Provide tools to enhance the perceived value of low-priced courses, such as testimonials, certifications, or bundling options.
- **For Students**:
  - Consider quality and reviews rather than price alone to ensure the best learning experience.

iii) Weighted Average Rating based on platform



Observations:

Weighted Average Ratings by Platform:

COURSERA: Weighted Average Rating: 4.66

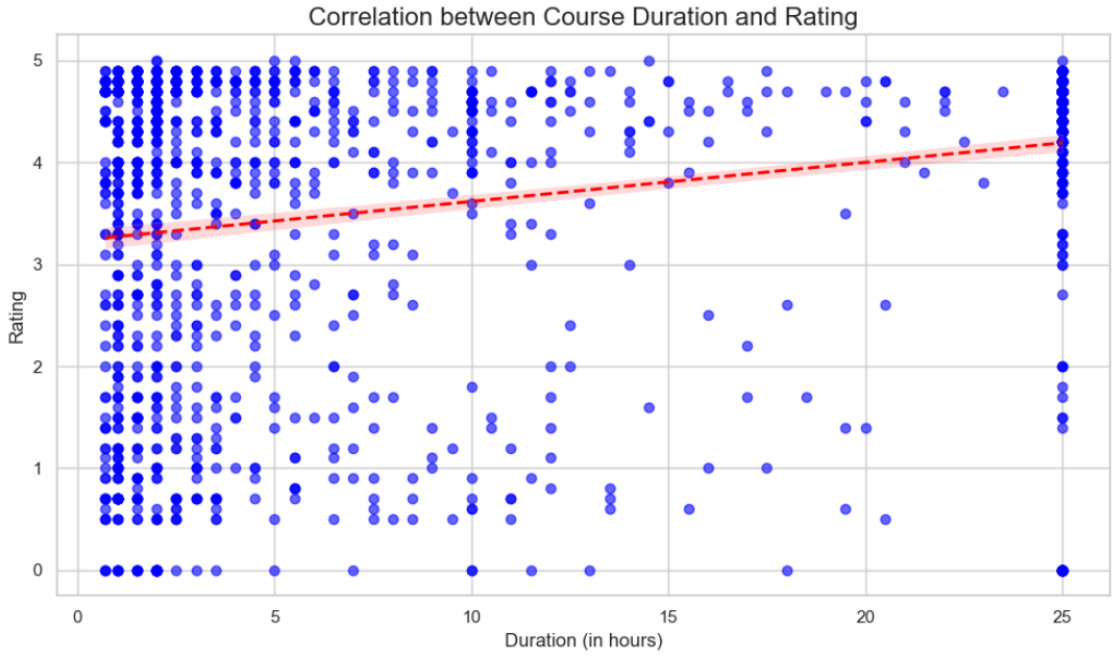
UDEMY: Weighted Average Rating: 4.12

Insights:

COURSERA has a higher weighted average rating (4.66) compared to UDEMY (4.12). This suggests that, on average, courses on Coursera are rated higher than those on Udemy when taking the number of reviews into account.

The weighted average rating reflects not just the course ratings but also the number of reviews, meaning that Coursera's courses have more impactful ratings due to higher review volumes, despite Udemy possibly having higher ratings for individual courses.

IV) Course Duration Based Analysis:



13.Is there a noticeable trend of higher ratings for courses with longer durations?

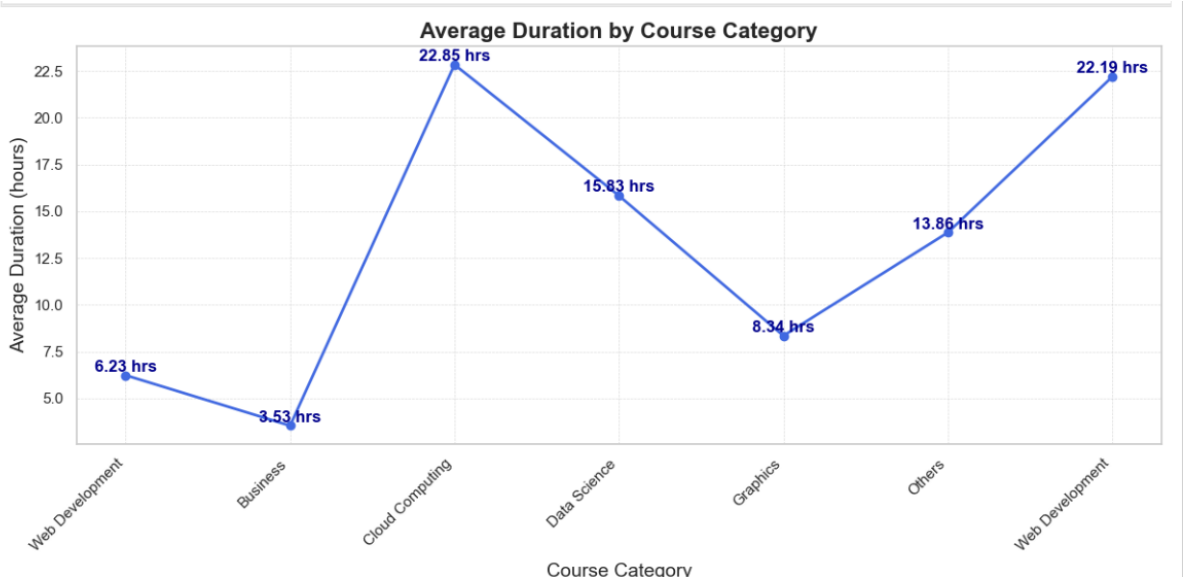
Answer: By examining the scatter plot, you should look for a pattern where courses with longer durations appear to have higher ratings. If there is a noticeable upward trend in the data points, it indicates a correlation between longer course duration and higher ratings. Do shorter courses also achieve high ratings, or is duration a key factor in quality perception? The scatter plot will also show if shorter courses, despite their brief duration, can achieve high ratings. If you see points in the lower-left quadrant of the plot (short courses with high ratings), it indicates that factors beyond just duration contribute to course quality.Upon examining the scatter plot, we observe that there is a slight tendency for courses with longer durations (close to 25 hours) to have higher ratings (4 and above). However, the trend is not very strong, and ratings are spread across various durations. This suggests that while there is some association, duration alone is not a perfect predictor of course ratings.

14. Do shorter courses also achieve high ratings?

Answer: Yes, shorter courses (less than 5 hours) also achieve high ratings (4 and above). This demonstrates that shorter courses can still deliver quality content, and duration is not the only factor influencing ratings. It's likely that factors such as content quality, instructor expertise, and course relevance play a substantial role.

Conclusion: While there is a slight trend showing that longer courses tend to have higher ratings, this is not a strong, definitive rule. Shorter courses can still receive high ratings, indicating that duration is not the only factor in determining the perceived quality of a course. Other factors such as the quality of the content, the instructor, and the relevance of the course material likely play a significant role in shaping student ratings. Regression Line: The red dashed line shows the overall trend. A positive slope would indicate a positive correlation between course duration and ratings.

ii) Average Duration by Course Category:



Observations:

In the ' Web Development' category, the average course duration is 6.23 hours.

In the 'Business' category, the average course duration is 3.53 hours.

In the 'Cloud Computing' category, the average course duration is 22.85 hours.

In the 'Data Science' category, the average course duration is 15.83 hours.

In the 'Graphics' category, the average course duration is 8.34 hours.

In the 'Others' category, the average course duration is 13.86 hours.

In the 'Web Development' category, the average course duration is 22.19 hours.

15)Which is the course with longest durations?

The course category with the longest average duration in the dataset is Cloud Computing, with an average course duration of 22.85 hours.

Other Observations:

Comparison of Course Categories by Average Duration: Cloud Computing: 22.85 hours (Longest average duration) Web Development: 6.23 hours (First mention) and 22.19 hours (Second mention). The average for Web Development appears to have two different values, possibly due to duplicate entries in the dataset. Data Science: 15.83 hours Graphics: 8.34 hours Others: 13.86 hours Business: 3.53 hours (Shortest average duration) Key Insight: Cloud Computing offers the longest average course duration at 22.85 hours. The Web Development category has two different values, but the higher of the two (22.19 hours) is quite close to Cloud Computing. Business courses have the shortest average duration, at 3.53 hours. Conclusion: The Cloud Computing category has the longest average course duration, and Web Development has a highly variable duration range, with both a short (6.23 hours) and long (22.19 hours) duration. Other categories like Data Science, Graphics, and Others also offer substantial course durations but are generally shorter than Cloud Computing.

Final Conclusion

This project offers valuable insights into the dynamics of online learning platforms, highlighting patterns and relationships that benefit students, platforms, and instructors alike:

1. **For Students:**
    - Learners gain clarity on course popularity, pricing, and ratings, helping them choose courses tailored to their needs and budgets.
    - Categories like *Web Development* and *Data Science* emerge as excellent starting points for skill development.
    - High-quality, affordable paid courses are readily available, ensuring accessibility for a wide audience.
2. **For Platforms:**
    - Platforms can optimize their offerings by increasing free and beginner-level courses to attract more users.
    - Mid-range pricing appears most appealing, offering a balance of affordability and quality.
    - Monitoring user engagement metrics (e.g., review counts, ratings) helps platforms identify trends and refine their content strategy.
3. **For Instructors:**
    - Insights into trending topics like *Python* and *Cloud Computing* can guide content creation to target areas of high demand.
    - Offering comprehensive courses with longer durations and higher interactivity can lead to better ratings and learner satisfaction.
    - Instructors can leverage mid-range pricing to maximize enrolments without compromising revenue.

Future Work:

Future Work Plans

1. **Advanced Analytics**:
    - Incorporate machine learning techniques to predict course success based on attributes like content type, duration, and price.
    - Analyze sentiment from course reviews to better understand user satisfaction.
2. **Cross-Platform Comparison:**
    - Expand the analysis to include other platforms such as EdX or LinkedIn Learning for a more comprehensive study.
    - Evaluate course completion rates and their impact on ratings and reviews.
3. **Student-Centric Features**:
    - Develop personalized course recommendations based on user behavior, preferences, and goals.
    - Explore accessibility features like multilingual courses and regional pricing for global reach.
4. **Instructor and Platform Improvements:**
    - Suggest optimal pricing and marketing strategies for instructors.
    - Use insights to design targeted advertising campaigns for platforms.

By focusing on these areas, future work can further enhance the learning experience, provide more business opportunities for instructors, and strengthen platform engagement.