Data-Driven Insights from Online Course Analysis

Analyzing Online Course Performance
to Drive Product Development

Rahma Ardita Putri

Analyzing Online Course Performance to Drive Product Development

In the world of online education, **figuring out what makes a course successful is crucial**. To really make a difference, we need to dive into metrics like platform performance, course popularity, category distribution, pricing effectiveness, and course durations. This case study is all about **giving the product development team insight and actionable recommendations** to improve course offerings and create a better learning experience for users.

Introduction to the Dataset

123 course_id ▼	ABC course_name ▼	ABC category ▼	123 Duration ▼	123 enrolled_students ▼	123 Completion_Rate ▼	ABC platform ▼	123 Price ▼	123 Rating (out of 5) ▼
1	Course_1	Office Tools	21	4,217	50.6468270009	Coursera	38.7974245935	4.8112516098
2	Course_2	Office Tools	57	4,238	82.2402398041	edX	160.6509909079	3.8293293935
3	Course_3	Technology	52	2,700	55.7290284644	LinkedIn Learning	123.5037809575	4.8519500804
4	Course_4	Office Tools	69	4,308	58.6647294335	LinkedIn Learning	116.775703882	3.9137318442
5	Course_5	Technology	43	4,792	62.5981474786	Udemy	96.2466956296	4.9219680986
6	Course_6	Finance	66	3,792	96.6719004303	edX	127.5099201581	3.2522834618
7	Course_7	Marketing	29	2,607	71.8304412082	Coursera	155.5411167649	4.1018068319
8	Course_8	Design	21	964	99.0762723988	edX	79.9409443327	3.0887304873
9	Course_9	Programming	57	1,246	78.8176224357	LinkedIn Learning	174.6774799209	4.4984712756
10	Course_10	Design	59	489	58.8960677572	edX	82.0948190093	3.3418675021

The dataset consists of information about **online courses** from various platforms, containing a total of **10,000 rows.** Each row represents a single course and includes the following attributes:

- Course_ID
- Course_Name
- Category
- Duration (hours)
- Enrolled_Students
- Completion_Rate (%)
- Platform
- Price ($)
- Rating (out of 5)

Tools: DBeaver

Business Problems

1 How do **completion rates** differ across platforms?

3 What makes spesific **platform** has **lowest total enrollment** ?

5 Which **categories** are **performing well** in terms of average rating and completion rate?

2 What **top platform based on enrolled students,** what the most popular courses on that platform, and what makes them stand out?

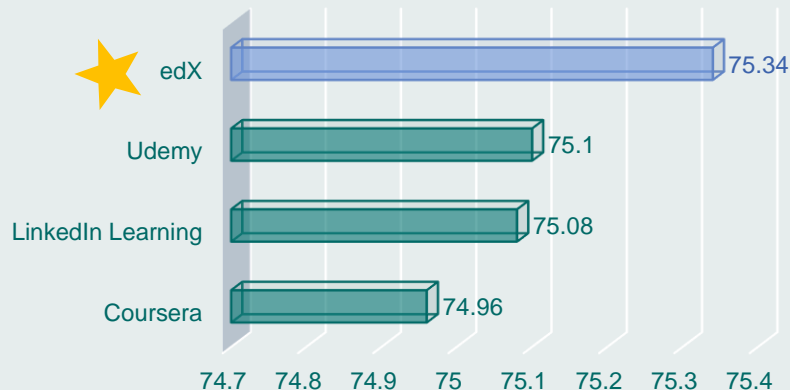**4** How does course **pricing affect user ratings**, and **total enrollment?**

6 How can we **optimize course durations across categories** to maximize student engagement and enrollment?

Average Completion Rate by Platform

1

Average Completion Rate (%)

edX — 75.34
Udemy — 75.1
LinkedIn Learning — 75.08
Coursera — 74.96

74.7 74.8 74.9 75 75.1 75.2 75.3 75.4
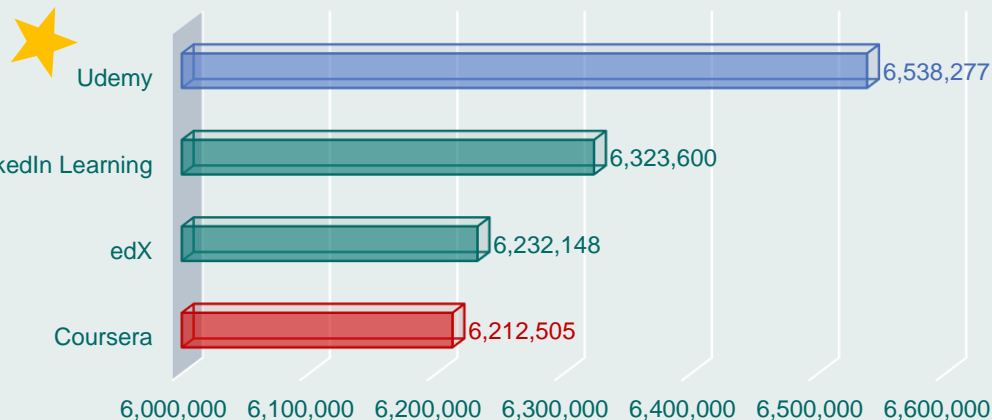
Query

```
select platform,
    round(avg(`Completion_Rate (%)`),2) as Avg_Completion_Rate
from online_courses_uses ocu
group by platform
order by Avg_Completion_Rate desc;
```

edX has the highest average completion rate at **75.34%**, which is **0.34% higher** than **Udemy** at **75.1%**, the closest competitor. This suggests that edX could be attributed to its **well-structured courses, and high levels of interactivity,** which likely contribute to its higher completion rates.

Platform Analysis by Enrollment Students

2

Enrolled Students

Platform	Enrolled Students
Udemy	6,538,277
LinkedIn Learning	6,323,600
edX	6,232,148
Coursera	6,212,505

Udemy leading platform in terms of student engagement because **Udemy** has significantly student enrollment compared to other platform with a total of **6,538,277** enrollment student.

The product development team wants to know
1) **Which courses** on Udemy with the highest total enrollment and **which categories** drive the most engagement?
2) **Why Coursera has the lowest** total enrollment among platforms?

Query

```
select platform , sum(enrolled_students) as total_enrolled_students
from online_courses_uses ocu
group by platform
order by total_enrolled_students desc;
```

Which courses and categories on Udemy have the highest total enrollment?

Course Name	Category	Total Enrolled Students
Course_4759	Data Science	4996
Course_2753	Marketing	4995
Course_4223	Data Science	4993
Course_4332	Finance	4986
Course_5233	Data Science	4985

Query

```
select course_name, category, sum(enrolled_students) as total_enrolled_students
from online_courses_uses ocu
where platform = 'Udemy'
group by course_name, category
order by total_enrolled_students desc
limit 5;
```
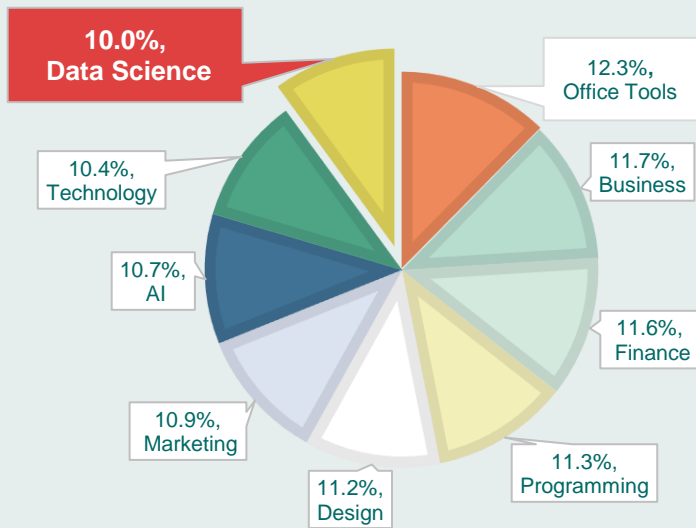
Data Science courses dominate the top five, indicating **growing interest** in this field because of **high demand for data skills** nowadays. The top courses has very close enrollment number, highlighting a **strong and competitive market for Data Science content.**

3 Why Coursera has the lowest total enrollment?

TOTAL COURSE IN TECH CATEGORY

10.0%, Data Science
10.4%, Technology
10.7%, AI
10.9%, Marketing
11.2%, Design
11.3%, Programming
11.6%, Finance
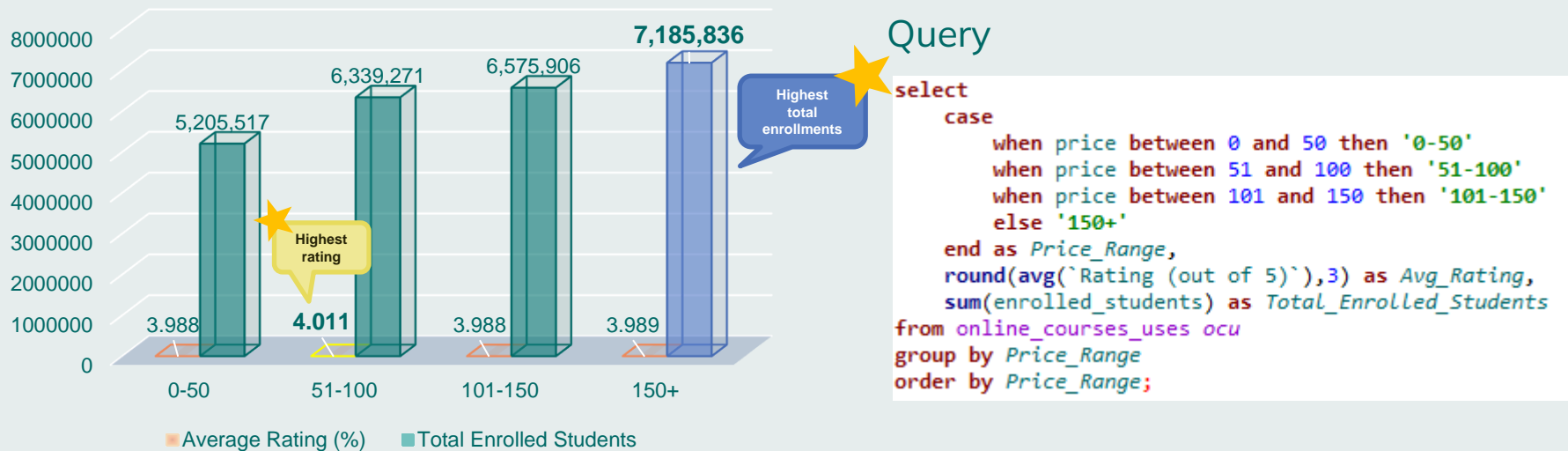11.7%, Business
12.3%, Office Tools

Query

```sql
SELECT category, COUNT(course_id) AS total_courses
FROM online_courses_uses
WHERE platform = 'Coursera'
GROUP BY category
ORDER BY total_courses DESC;
```

Data Science stands out with the **lowest** number of available courses in Coursera (10%). In contrast, Data Science emerged as the top category on Udemy, top platform based on enrollment. This suggests a **potential relationship between course availability and platform popularity**. Specifically, Coursera's limited offerings in Data Science might contribute to its lower overall enrollment.

Price Range Analysis: Average Rating vs Enrollments

Query

```
select
    case
        when price between 0 and 50 then '0-50'
        when price between 51 and 100 then '51-100'
        when price between 101 and 150 then '101-150'
        else '150+'
    end as Price_Range,
    round(avg(`Rating (out of 5)`),3) as Avg_Rating,
    sum(enrolled_students) as Total_Enrolled_Students
from online_courses_uses ocu
group by Price_Range
order by Price_Range;
```

Courses in the $51-$100 price range have the **highest average rating** of **4.01%**. In contrast, **the $150+ price range**, despite having a lower average rating, it attracts **the highest total number of enrolled students**. This suggests that while higher-priced courses might have a larger student base, **they may not achieve the same level of satisfaction as moderately priced courses**.

Top Performing Category

Average rating **higher than 4**

Output

Category	Total Enrollments	Average Rating	Average Completion Rate
Data Science	2,734,806	4.02	75.34
Marketing	2,949,738	4.01	74.69
Office Tools	2,882,713	4	74.64

Query

```sql
SELECT
    category,
    sum(enrolled_students) as Enrolled_Students,
    round(avg(`Rating (out of 5)`),2) AS Average_Rating,
    round(avg(`Completion_Rate (%)`),2) AS Average_Completion_Rate
FROM online_courses_uses
GROUP BY Category
HAVING avg(`Rating (out of 5)`) > 4
ORDER BY Average_Rating desc, Enrolled_Students desc;
```
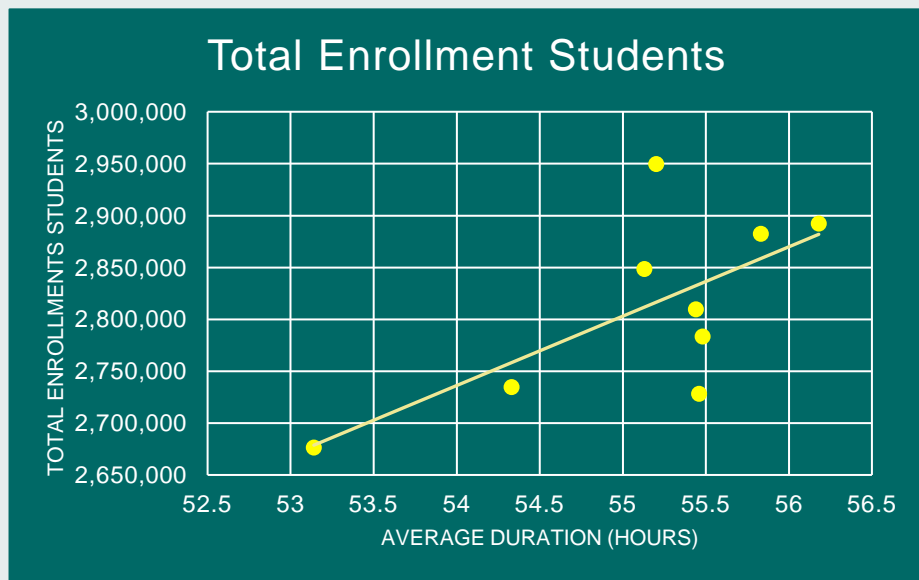
Although **Data Science** has the **lowest number of enrolled students** among the categories, at 2,734,806, it **achieves the highest average rating and completion rate**, with scores of 4.02 and 75.34%, respectively. This could indicates **high-quality course content** in these category.

Analyze Course Durations and Enrollment

Total Enrollment Students

The analysis shows that categories with average course durations **55.2 hours** tend to have **highest total enrollments (2,949,738 students).** This indicates a **strong correlation** between course duration and student engagement, suggesting that courses with durations near this average are **likely resonating well with learners.**

Query

```sql
select
    Category,
    round(avg(`Duration (hours)`), 2) as Average_Duration,
    sum(Enrolled_Students) as Total_Enrolled_Students
from online_courses_uses ocu
group by Category
order by Total_Enrolled_Students desc;
```

Recommendation Strategies

1. Investigate the elements contributing edX's high completion rate and **Udemy's** high enrollment. **Consider applying successful elements** from top-performing platform into new and existing platform to boost completion rates and enrollments.

2. **Expand Coursera's offerings in high-demand category**, particularly Data Science, **Conduct market research** to align course availability with student interests and industry trends.

3. Enhance value and quality of course by **adding premium features or additional content on $150+ price** to boost rating and maintain high enrollment numbers.

4. **Invest in developing high-quality Data Science course** across all platforms. Also apply the success factors from Data Science courses to improve content in other categories.

5. **Aim for an average duration around 55.2 hours** across categories while maintaining content quality

Thank You!

Rahma Ardita Putri

rahmaarditap@gmail.com