



Online Education

By: Suchrita Das

Data Analyst Aspirant | SQL, Excel, & Business Insights | Udemy Dataset Case Study



This presentation provides a data-driven analysis of course performance.

We explore key aspects of Udemy course performance.

Uncovering key insights and strategic opportunities.

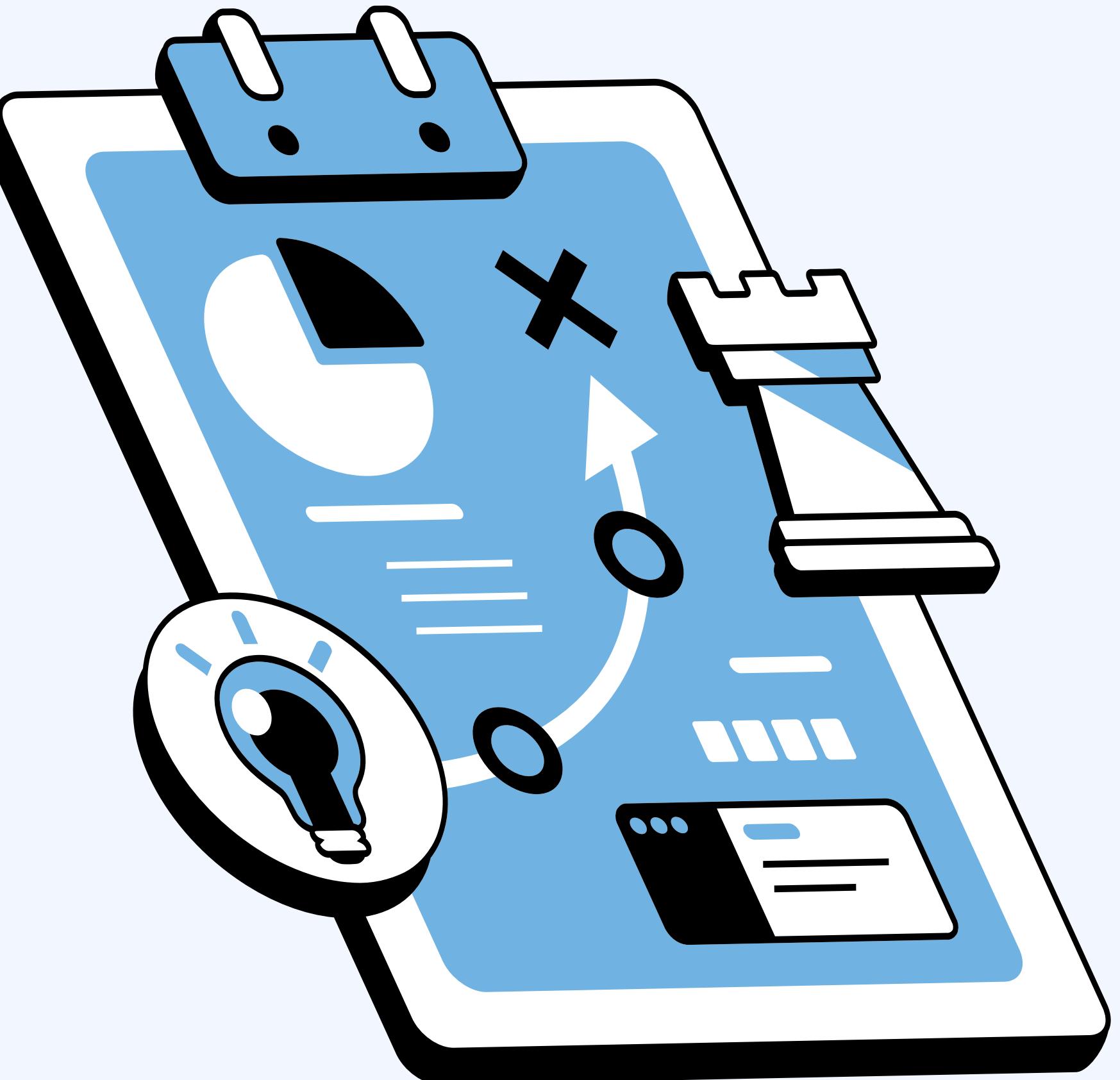
5459	8605	9537	6033	1100	7658
5008	8502	9552	5119	3537	2095
5547	8502	4530	5533	1003	1954
5348	5496	5503	4564	2558	2038
5345	5000	5527	5446	2032	2011
5244	5004	1975	5506	4400	2055
5316	5002	1959	1544	4577	2004
5543	6022	5503	5607	2705	2003
9910	3556	5116	3545	1004	2002
5673	5023	9525	5226	4126	9523
5524	5007	1473	5604	4273	5194
3277	3522	5072	5474	5372	2056
5217	3603	5026	5476	4623	2004
5524	3607	5078	5614	5803	2044
3277	3504	5026	5476	4623	2002
5216	3602	5343	5614	5809	2045

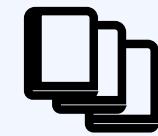
Based on a comprehensive dataset from Udemy.



Content:

- Case Study Context
- Dataset Snapshot
- Database Design Overview
- Subject wise Total Revenue
- Course level wise Revenue generate
- Top 10 courses causing Maximum Revenue Loss
- Engagement based on Duration
- Free vs Paid more Engaging Content
- Under-priced High demand Courses
- Yearly Growth Trends
- Key Findings
- Recommended Actions
- Thank You





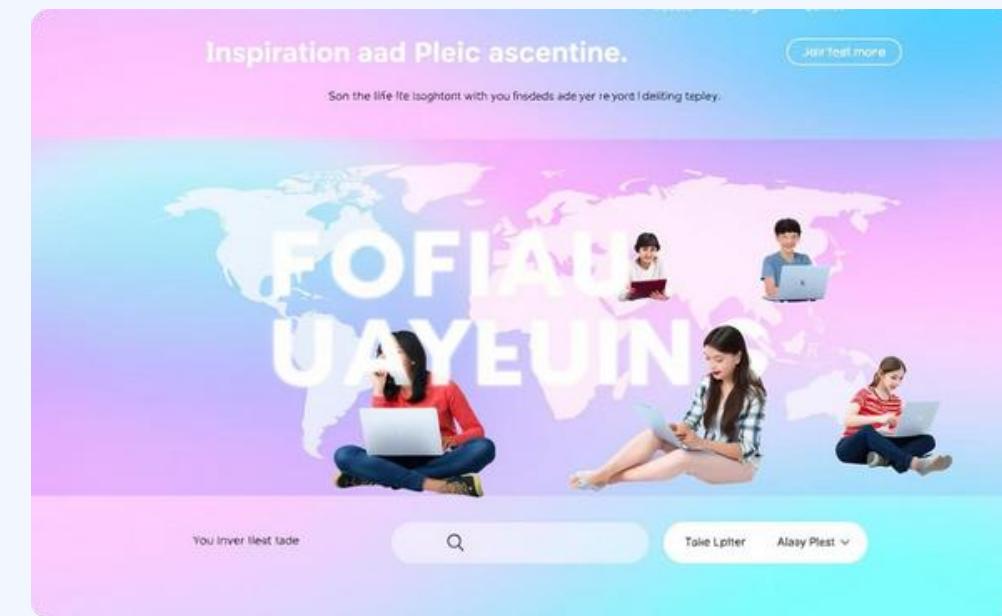
About Udemy – Case Study Context



Strategic Analysis

Our analysis of Udemy course data focuses on identifying opportunities for revenue growth, price optimization, and enhanced engagement.

Leveraging SQL and Excel, PowerPoint I transformed raw course data into clear, actionable insights, empowering instructors and platforms to make more informed decisions.



Udemy's Global Reach

Udemy is a leading online learning platform, serving over 60 million learners with 200,000+ courses across diverse subjects worldwide.



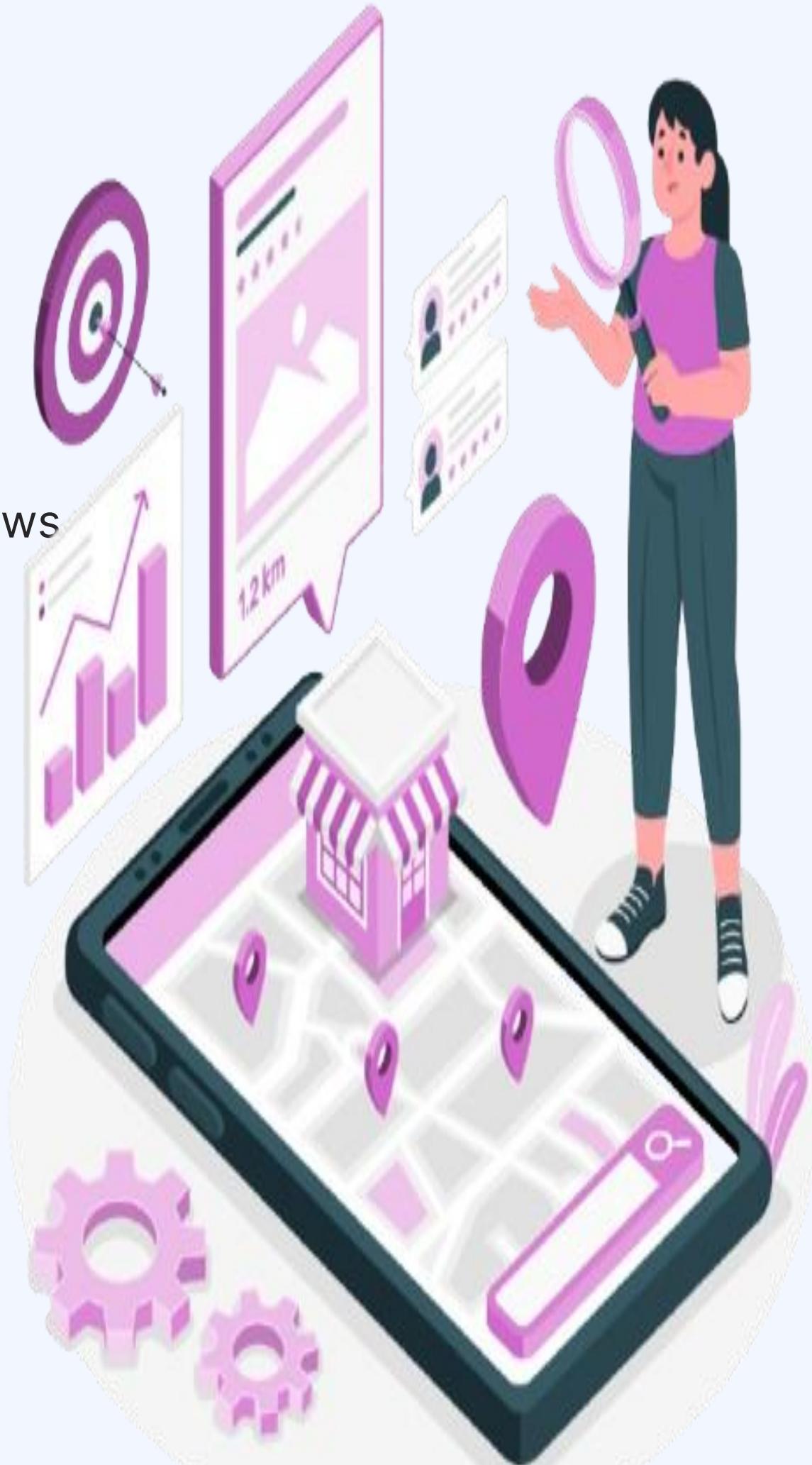
Uncovering Key Insights

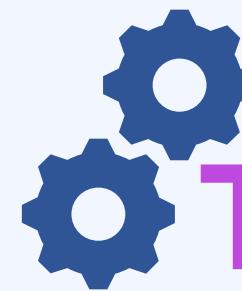
This study investigates the drivers of course success, the impact of pricing on performance, and the subjects and formats that attract the most learners.



Dataset Snapshot

- Source: Kaggle – Udemy Dataset, comprising over 3,600 course records.
- Key Fields Analyzed:
 - **Course Information:** Course ID, Title, URL
 - **Pricing & Engagement:** Is Paid, Price, Number of Subscribers, Number of Reviews
 - **Content Details:** Number of Lectures, Level, Content Duration
 - **Publishing & Subject:** Published Timestamp, Subject
- Dataset Overview:
 - **Total Courses:** 3,672 analyzed
 - **Course Subjects:** 4 major categories – Web Development, Business/Finance, Graphic Design, Musical Instruments
 - **Course Types:** Free and Paid
 - **Levels:** Beginner, Intermediate, Expert, All-Level
 - **Time Range:** Courses published between 2011 and 2017





Tools and Workflow:

1. MySQL Workbench (Structured Query Language)

- Primary tool for data querying, filtering, and joining tables.
- Used to generate insights from raw datasets (e.g., user behaviour, course ratings, revenue trends).

2. Microsoft Excel

- Used to import query results and create interactive charts and visualizations.
- Enabled clean formatting and chart design for reporting.

3. PowerPoint

- Final presentation medium for compiling insights, charts, and conclusions.
- Made the project visually engaging and easy to communicate.

4. Command Prompt (CMD)

- Used to efficiently import large datasets using command-line utilities.
- Enabled faster data handling compared to manual import.





Database Design Overview

```
create database Udemy;
use Udemy;

Table Name: courses
• Primary Key: course_id
create table courses
(course id int primary key,
Course title varchar(255),
URL varchar(2083),
is_paid varchar(10),
price int,
Num subscribers int,
Num reviews int,
Num lectures int,
level varchar(255),
content duration float,
published timestamp varchar(255), subject
varchar(255));
```

The screenshot shows a MySQL Workbench interface with a query editor window. The code in the editor is as follows:

```
1 •  create database udemy;
2 •  use udemy;
3 •  create table courses(
4      course_id int primary key,
5      course_title varchar(255),
6      url varchar(2083),
7      is_paid boolean,
8      price int,
9      num_subscribers int,
10     num_reviews int,
11     num_lectures int,
12     level varchar(255),
13     content_duration float,
14     published_timestamp varchar(255),
15     subject varchar(255)
16 );
17 •  select*from courses;
```

The interface includes a toolbar with various icons for file operations, search, and navigation. A status bar at the top right indicates "Limit to 10000 rows".



Analytical Focus Area:

⌚ Revenue-based Analysis

Subject-wise Total Revenue

- Identify which subject categories generate the most revenue

Course-Level wise Revenue Generation

- Pinpoint high-earning individual courses

Top 10 Free Courses Causing Revenue Loss

- Detect free content that could be monetized



📊 Engagement Metrics Analysis

Engagement vs Course Duration

- Analyse whether course length affects learner engagement

Free vs Paid Course Engagement

- Compare how users interact with free vs paid content



🏷️ Pricing Optimization

Identify Underpriced, High-Demand Courses

- Spot courses with high traffic but low pricing for strategic repricing

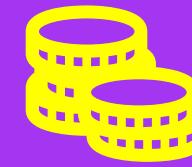
📈 Trend Analysis Year-wise

Revenue & Engagement Trends

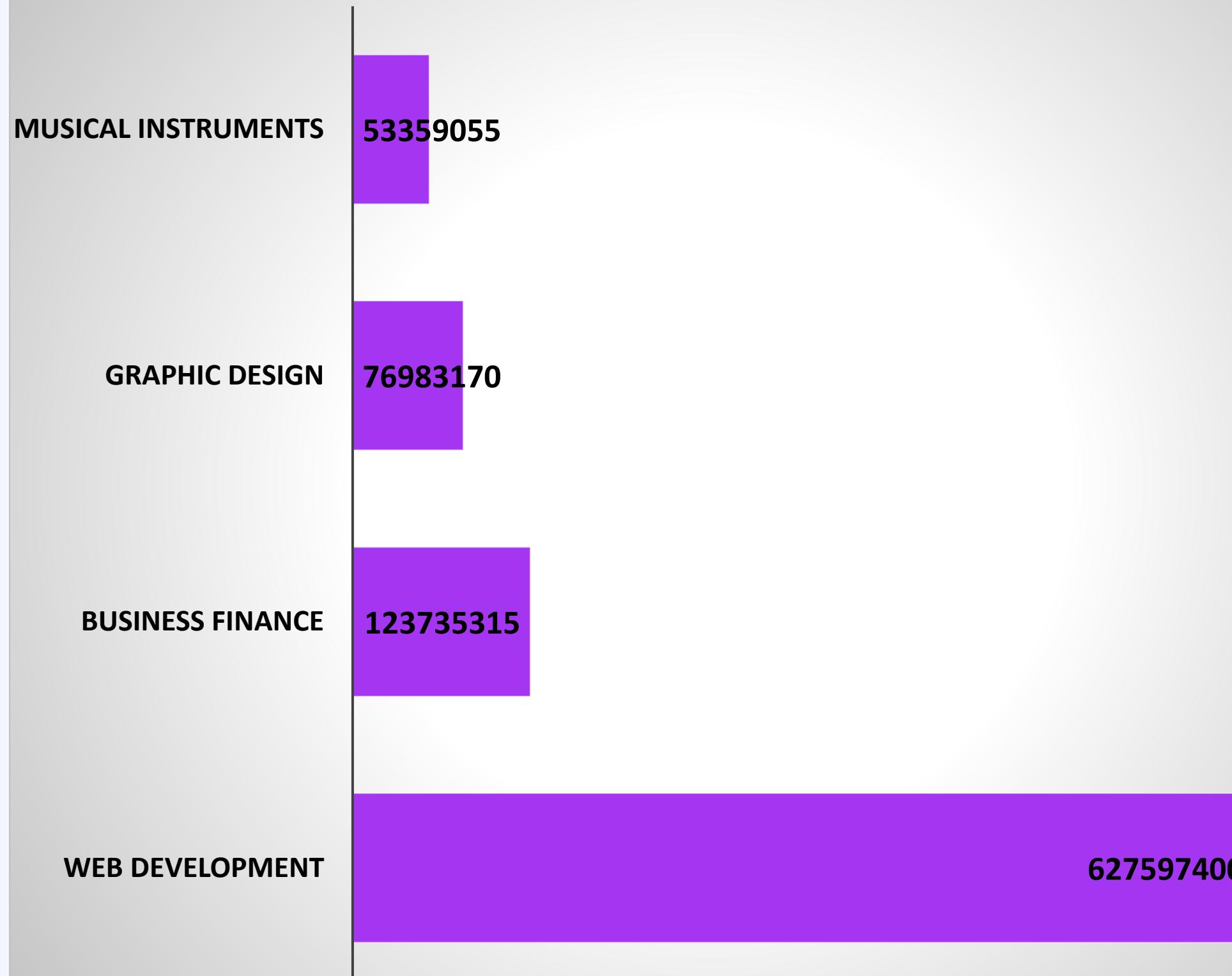
- Identify growth patterns or seasonal spikes
- Predict future performance using past trends



Subject wise Total Revenue



Subject wise Total Revenue



-- Which subject generated the highest total revenue?

```
Create view SubjectRevenue as
select subject,sum(price*num_subscribers) as Total_Revenue
from courses
where price>0 and is_paid= True
group by subject
order by total_revenue desc;

select*from SubjectRevenue;
```

Insight:

Web Development courses generated the highest revenue at 6.27 Cr, nearly triple that of Business Finance. This suggest strong alignment between learner demand and monetization in tech-focused content.

Course level wise Revenue generate



Course level wise Revenue

606793395

219678340

49734135

5469070

All Levels

Beginner Level

Intermediate Level

Expert Level

```
41 -- What is total revenue by course level?  
42 • select level,sum(price*num_subscribers) as total_revenue  
43 from courses  
44 where price>0  
45 group by level  
46 order by total_revenue desc;
```

Insight:

All levels courses generated the highest revenue (606793395), indicating strong learner interest in flexible content that spans across skill stages. Beginner level courses followed distantly(219678340),while Intermediate and Expert levels contributed minimally.

This suggests a shift in learner behaviour toward broader ,inclusive learning paths, and highlights a performance gap in specialized course subjects.

Top 10 Free courses causing Maximum Revenue Loss



25000000

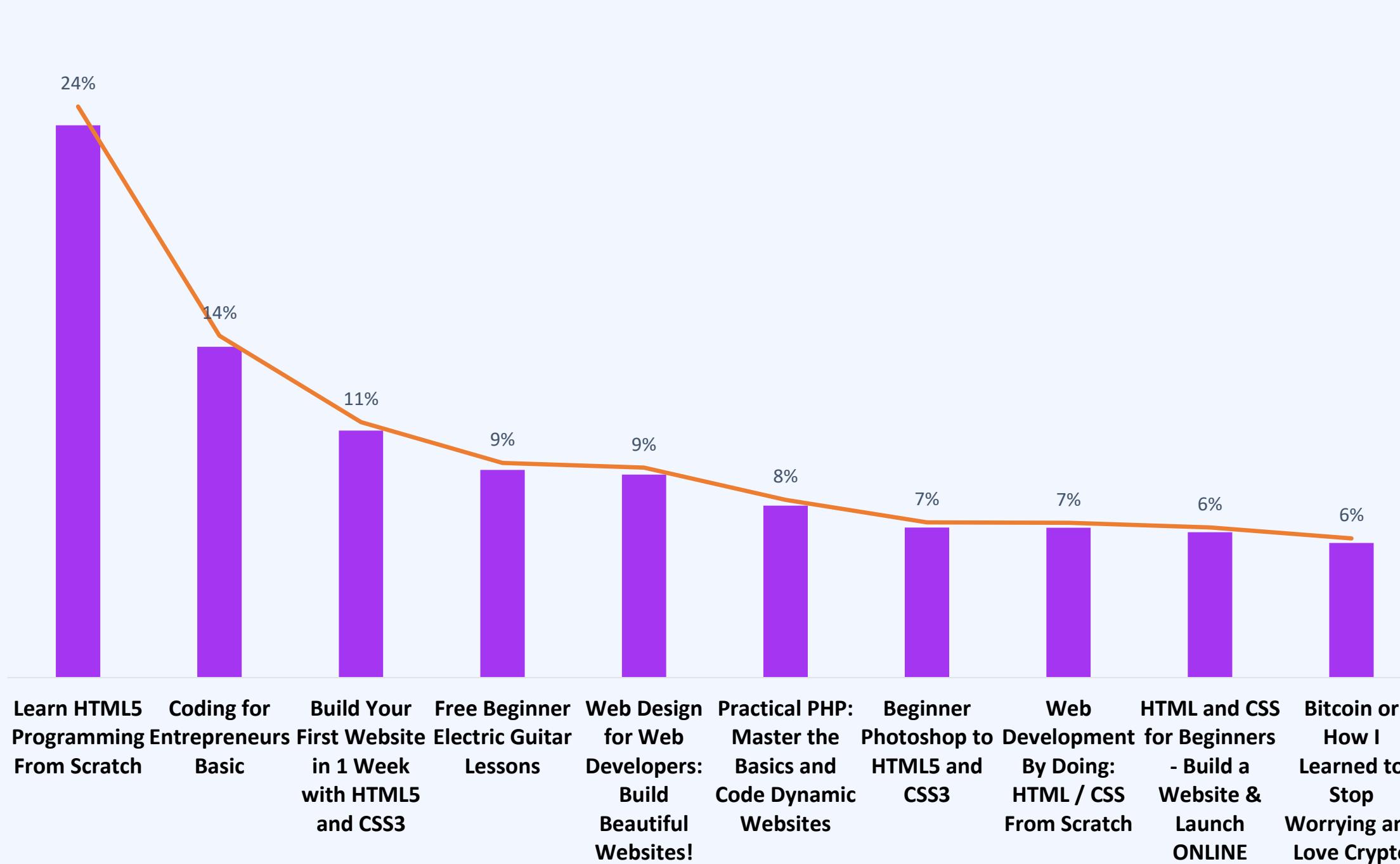
20000000

15000000

10000000

5000000

0



```

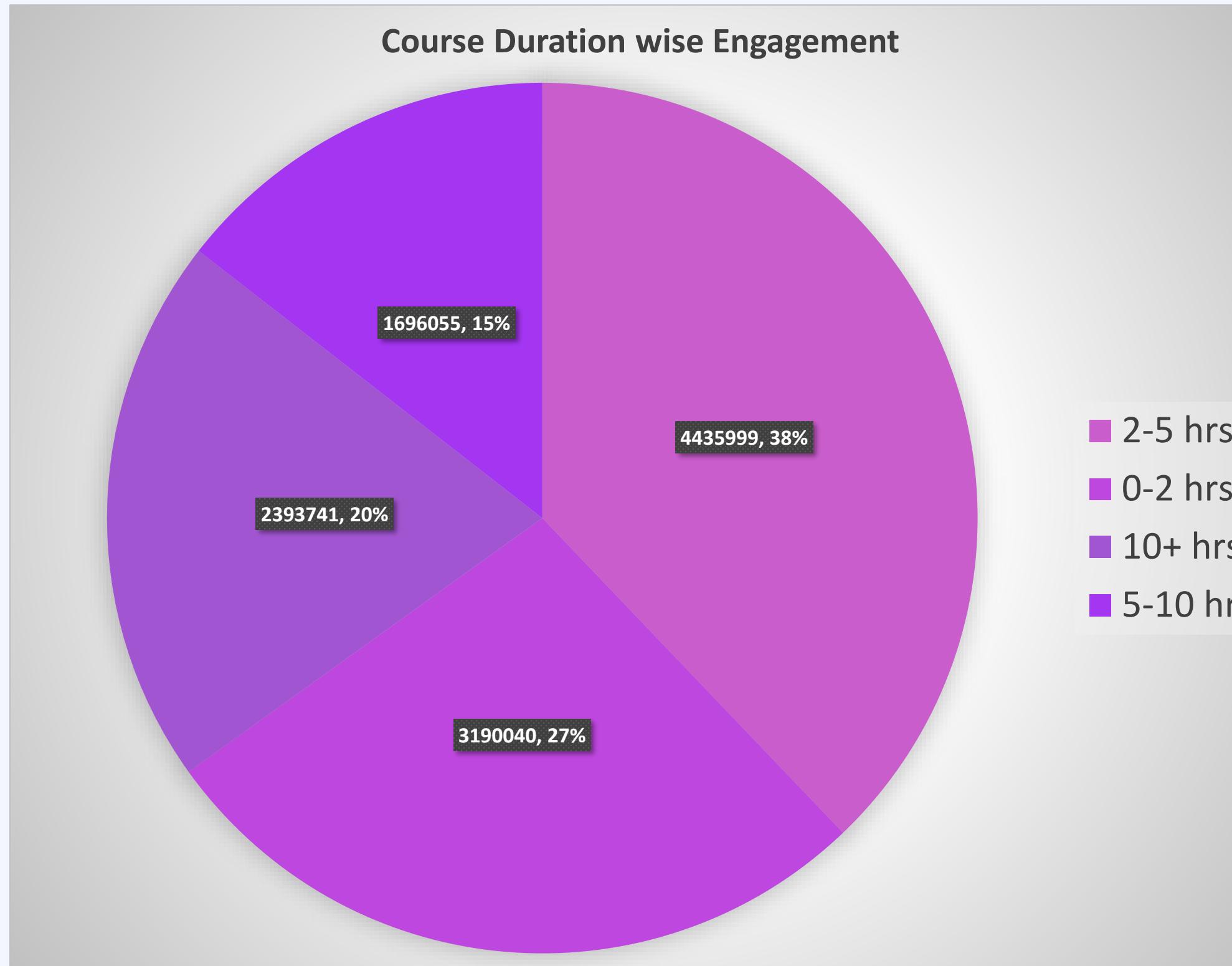
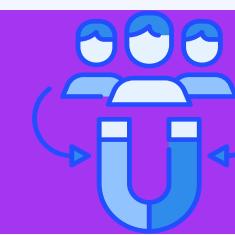
48 -- What's the potential revenue loss from free course?
49 • with avg_price as(
50   select round(avg(price),2) as avg_paid
51   from courses
52   where price>0),
53
54   free_course_loss as(
55     select c.course_title,c.num_subscribers,round(a.avg_paid*num_subscribers,2)as estimated_loss
56     from courses c
57     join avg_price a on 1=1
58     where c.price=0)
59
60   select course_title,num_subscribers,estimated_loss
61   from free_course_loss
62   order by estimated_loss desc
63   limit 10;
/*
```

Insight:

Among the free courses on Udemy , a small number are responsible for a disproportionately high potential revenue loss.

- Over here the course “Learn HTML5 Programming from scratch” is the top contributor to revenue loss (24%).
- The drop-off after the first few courses is steep , suggesting a Pareto distribution
- Most top-loss courses are centered around Web Development , Bitcoin and Website launch , suggesting this categories might be overly saturated with free content.

Engagement based on Course Duration

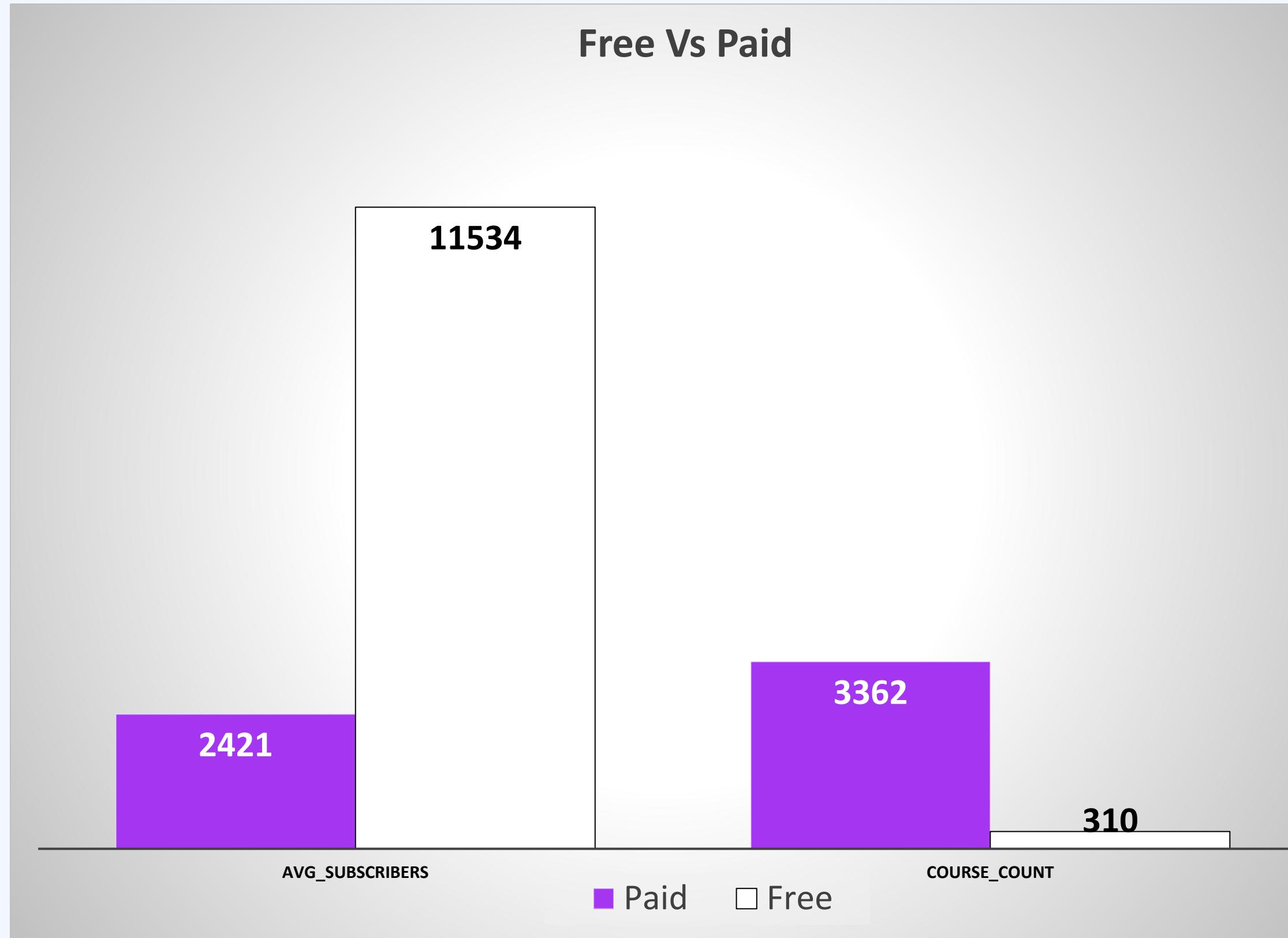


```
30 -- How does content duration relate to total subscribers?  
31 • Ⓜ select case  
32 when content_duration < 2 then "0-2 hrs"  
33 when content_duration between 2 and 5 then "2-5 hrs"  
34 when content_duration between 5 and 10 then "5-10 hrs"  
35 else "10+ hrs"  
36 end as duration_bin,sum(num_subscribers) as total_subscribers  
37 from courses  
38 group by duration_bin  
39 order by total_subscribers desc;  
40
```

Insight:

To maximize user engagement on Udemy, focus on designing courses around 2-5 hr range. This duration hits the optimal balance between user interest, perceived value and time efficiency.

Free Vs Paid more engaging contents



92 -- Free vs Paid course engagement

93 • select is_paid,

94 round(avg(num_subscribers),0) as avg_subscribers,

95 count(*) as course_count

96 from courses

97 group by is_paid;

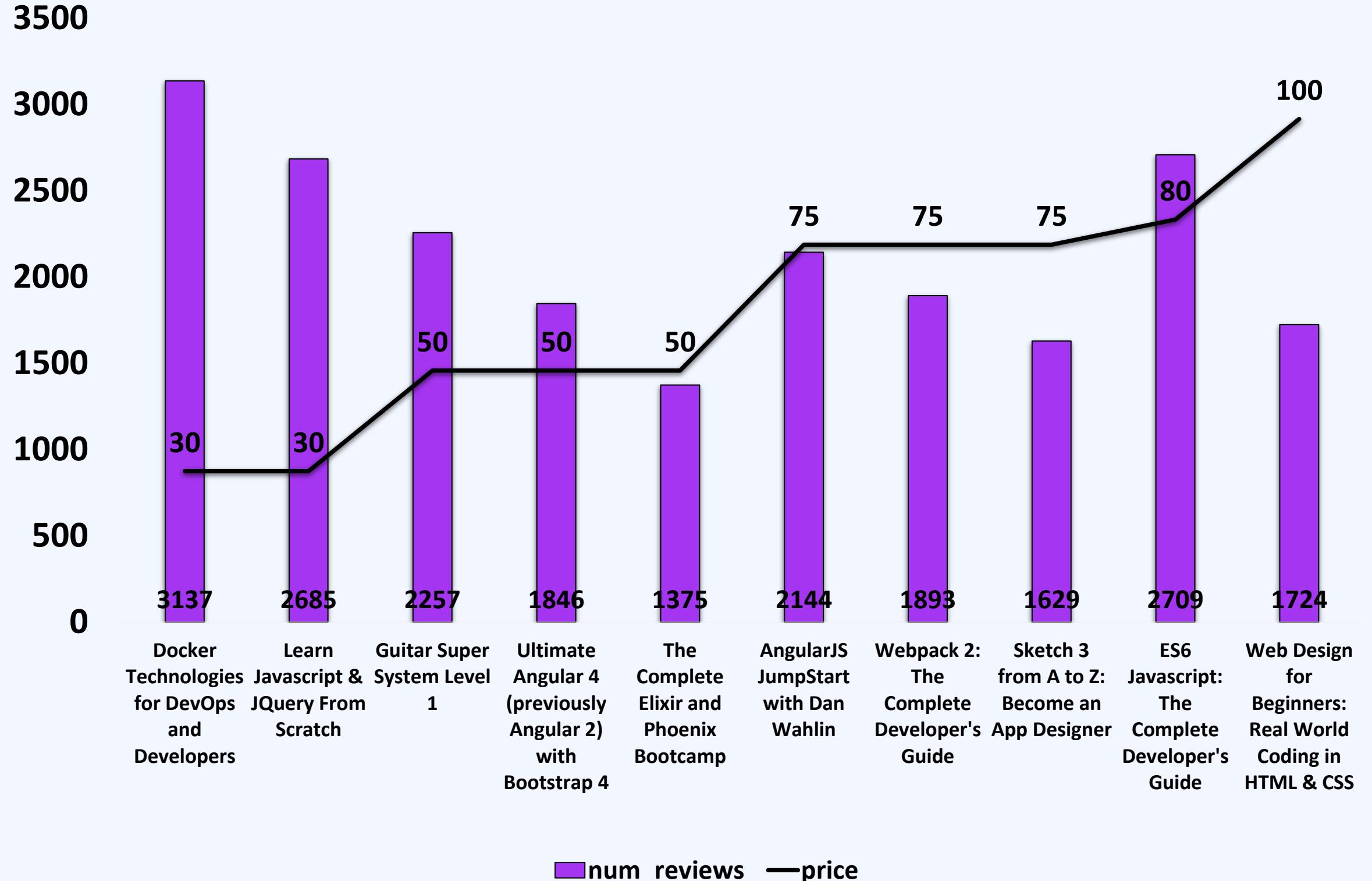
Insight:

1. **Free Courses Have Higher Engagement** Despite being free, these courses have 4.7x more average subscribers than paid ones. This indicates that users are more willing to enroll in free content, making them an excellent gateway for user acquisition.
2. **Paid Courses Are More Numerous** With 3,862 paid courses vs. 310 free ones, the platform is clearly prioritizing monetized content. However, the low engagement for paid courses suggests a potential value or pricing gap.

Identifying Underpriced High-demand Courses for Repricing Strategy



Top Popular Courses with Potential Undervaluation



```

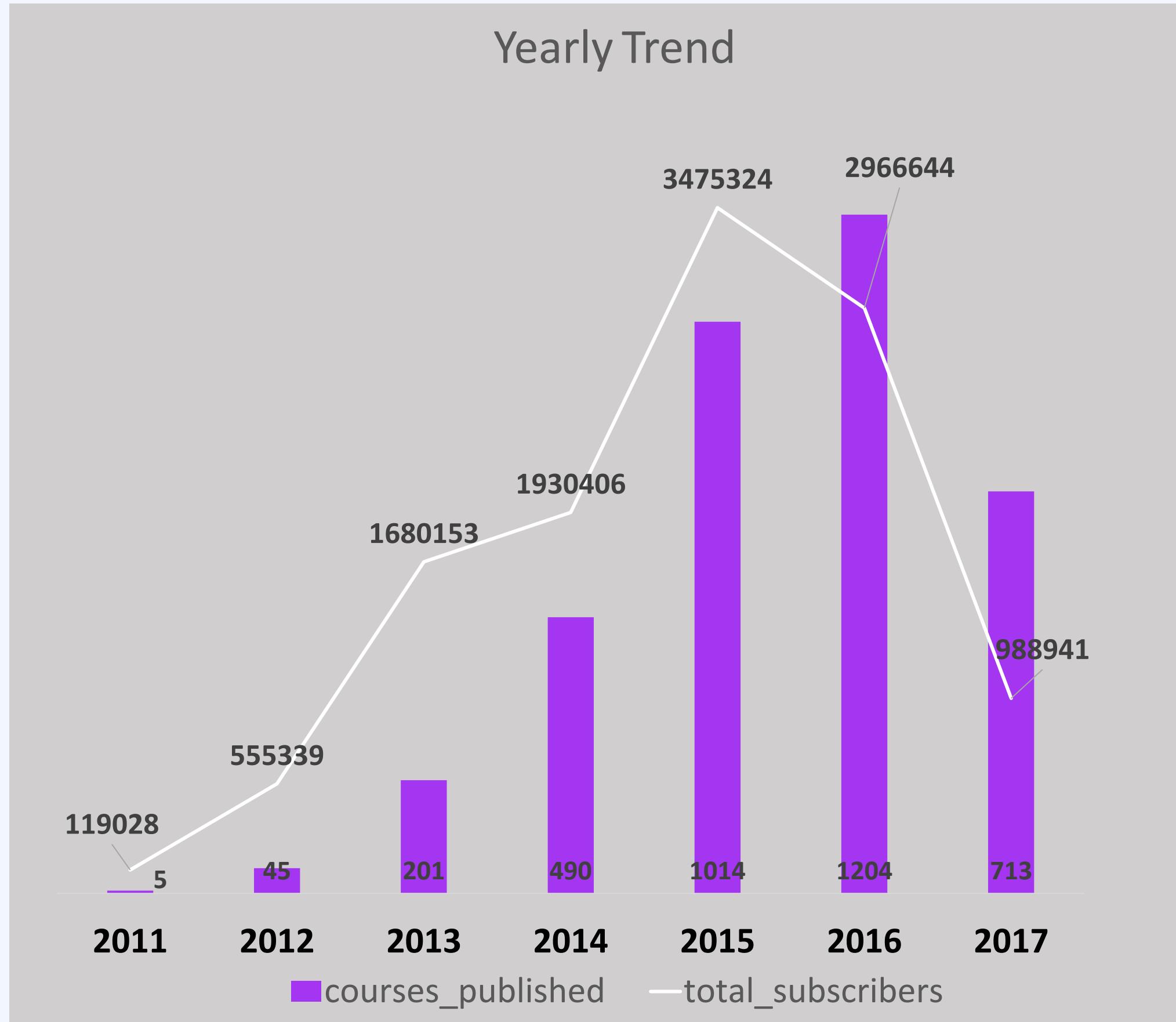
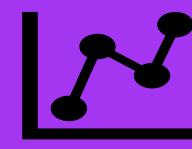
82  -- Which popular courses are underpriced and should be repriced?
83  ## Average reviews among valid paid courses
84 • WITH avg_reviews as(
85      select(num_reviews) as avg_review
86      from courses
87      where is_paid = TRUE and price > 0
88  ),
89
90  ## Get popular courses only if paid and priced properly
91  popular_courses as(
92      select course_title, price, num_reviews
93      from courses, avg_reviews
94      where is_paid = TRUE and price > 0 AND num_reviews > avg_reviews.avg_review
95  ),
96
97  ## Average price of these popular courses
98  avg_price_popular as(
99      select AVG(price) as avg_price
100     from popular_courses
101  )
102
103  ## Final selection of underpriced popular courses
104  select p.course_title, p.price, p.num_reviews
105  from
106      popular_courses p, avg_price_popular a
107  where p.price < a.avg_price
108  order by price ASC;

```

Insight:

The chart identifies courses with a high number of user reviews- a key indicator of popularity –yet priced significantly lower than expected. These include highly rated courses in development, design and tech categories. The pricing does not align with their demand, indicating strong potential for undervaluation and missed revenue opportunities.

Yearly Growth Trends in Course Supply and Audience Engagement



```
71 -- How has course publishing and engagement evolved over time?  
72 • with course_timestamp as  
73 (select  
74 extract(year from published_timestamp) as published_year,  
75 count(course_id) as courses_published,  
76 sum(num_subscribers)as total_subscribers  
77 from courses  
78 where published_timestamp is not null  
79 group by published_year  
80 order by published_year)  
81 select*from course_timestamp;
```

Insight:

Between 2011 and 2015, Udemy experienced exponential growth in both the number of courses published and total subscribers. The course count peaked at 3.4million in 2015. This upward trend indicates strong platform momentum and user engagement during this period. However, from 2016 onward, while course publishing remained high in 2016, a sharp drop in subscribers was observed in 2017, suggesting signs of content saturation, reduced marketing efficiency, or user fatigue.



Key Findings (Business & Learner Insights)

1. Web Development courses generated the highest revenue (₹6.27 Cr), much more than the other subjects.
2. Courses marked as "All Levels" brought in the most income, showing that flexible, open-level courses attract more learners.
3. Free courses, though fewer in number (~8%), led to a major potential revenue loss, especially when popular courses were not monetized.
4. The sweet spot for course duration is between 2–5 hours—these had the best engagement.
5. Paid courses had 5x more subscribers and reviews than free ones—proving that people value paid content more.
6. Some underpriced courses (₹20–₹50) had 30,000+ students, meaning they are highly popular but not priced accordingly—an opportunity for bundling or price adjustment.
7. Year-wise analysis showed that Udemy peaked in 2015, and saw a drop in course performance by 2017—possibly due to too much content or quality dilution.



Recommended Action:

- Identify top underpriced, high-demand courses using review and enrollment data.
- Test small, gradual price increases to see learner response without sudden drop-offs.
- Design attractive bundle offers that group popular courses for better value.
- Monitor changes in enrollment and feedback to adjust pricing strategy.
- Share findings with instructors to align pricing with learner expectations and perceived value.

- LinkedIn: <https://www.linkedin.com/in/suchorita-das>
- Email: suchoritadas15@gmail.com

Thank You for Exploring This Project with Me 

I appreciate your time and interest in reviewing this Udemy course analysis project. I'm always open to feedback, collaboration, or deeper conversations about data, product thinking, or learning platforms.

