GROUP 4 – UDEMY COURSES SUBSCRIPTION

Udemy, being a massive online open course (MOOC) platform, offers both free and paid courses. Udemy has created a business model, where anybody can create a course on its platform, and people can subscribe to these courses for learning. This business model has allowed Udemy to have hundreds of thousands of courses. Udemy looks forward to making user experience better for both course authors and course subscribers.

Phase 1 – Ask Questions (Define Objectives)

How can Udemy strategically optimize its course offerings by analyzing market demand and other key factors to maximize revenue?

Goal – To increase revenue by at least 15% by the next quarter.

Possible Questions for analysis:

1. Which has the greatest number of subscriptions; Paid or unpaid courses?
2. Does the number of lectures have any effect on the number of subscriptions?
3. Does the course duration have any effect on the number of subscriptions?
4. Which subject is making the most sales (Paid Courses)?
5. Does the course price affect the number of subscriptions in any way?
6. What percentage of the revenue is contributed by each subject?
7. Is there a relationship between the number of reviews and the number of subscriptions?
8. Is there a relationship between the number of lectures and the course duration?

Phase 2 – Data Collection and Preparation (Data Sources)

Data needed for this analysis was retrieved from internal sources of the Udemy, specifically, the organization's database. A single table with the following columns was obtained:

- course_id – The Course ID
- course_title – The Course Title
- url – The URL to access the Course
- is_paid – Boolean, indicating if the course is free or paid
- price – The Price of the Course
- num_subscribers – Number of subscribers
- num_reviews – Number of reviews
- num_lectures – Number of lectures
- level – Course difficulty (Beginner, Intermediate, Expert and All Levels)

- content_duration – Duration of all course materials in the course
- published_timestamp – Date that the course was published.
- Subject – The Subject under which the Course is categorized (Web Development, Business Finance, Graphic Design and Musical Instrument)

Phase 3 – Data Processing / Transformation (Data Cleaning)

Data Cleaning Procedures

The following actions were taken to clean the data.

1. Removed duplicates from the course_id field.
2. Four rows had their course titles concatenated with the other fields of the various rows.
   The split-column function was used on the course_title row to separate the inconsistent data. The delimiter was double quotes (“). Then, the column which contained the inconsistent data was deleted.
3. Performed the trim function on all text fields.
   This was done to remove extra spaces between texts.
4. Removed all records of courses that had zero course_duration and zero num_of_lectures, because they are considered outliers. The operation was performed by sorting the course_duration column in ascending order and removing the top row.
5. Changed the published_timestamp data type from Date/Time to Date.
6. Changed the price data type from whole number to fixed decimal.
7. Extracted the initials for the subjects to shorten the text length.
   First, the subject column was duplicated. The duplicated subject column was split using space as the delimiter. Then, the first characters were extracted from the values in the split columns. Finally, the first characters were concatenated into a new column.

Phase 4 – Analyze Data

Scrutinizing the clean data gave the idea for the various visualizations we needed to use and what data aggregations to use. The dashboard was set to be split into four pages;

1. Summary – Visualizes a general overview of the data
2. Revenue – Focuses on visualizing data on paid courses only
3. Subscription – Visualizes number of subscribers and reviews
4. Course Content – Visualizes the number of lectures and course content

Visualizations

1. Clustered Column Chart
2. Clustered Bar Chart
3. Donut Chart
4. Line Chart
5. Table
6. Funnel Chart
7. Scatter Plot
8. Narrations
9. **Slicers**

   Slicers were used for filters on:

   - Date
   - Subject
   - Levels
   - Paid or Unpaid

Analysis – Aggregations

Calculated Columns

Subscriptions Table

1. "total_price" column [represents the total amount realised per course]

```
total_price = subscriptions[Subscribers] * subscriptions[price]
```

2. "Reviews_%_Sub" column [represents the percentage of reviews to subscribers]

```
Reviews_%_Sub = DIVIDE(SUM('subscriptions'[Reviews]), SUM('subscriptions'[Subscribers]))
```

3. "%_Subscribers" column [represents the percentage of subscribers of each course to total subscribers]

```
%_Subscribers = DIVIDE(subscriptions[Subscribers], SUM(subscriptions[Subscribers]))
```

4. "%_Reviews" column [represents the percentage of reviews of each course to total reviews]

```
%_Reviews = DIVIDE(subscriptions[Reviews], SUM(subscriptions[Reviews]))
```

5. "%_Revenue" column [represents the percentage of revenue of each course to total revenue]

```
%_Revenue = DIVIDE(subscriptions[total_price], SUM(subscriptions[total_price]))
```

Calculated Tables

A Dax function was used to summarise the subscriptions table grouped by the subject column for only paid courses, to be used to visualize revenue. The function first of all filtered out the values TRUE for the is_paid column and then grouped the data using the subject column as the grouping factor.

```
   summary_table_paid =

SUMMARIZE(
    FILTER(
        subscriptions,
        subscriptions[is_paid] = TRUE()
    ),
    subscriptions[Subject],
    subscriptions[Subject_Initials],
    subscriptions[is_paid],
    "Total_Courses", COUNTAX(FILTER(subscriptions, subscriptions[is_paid] = TRUE()),
subscriptions[Subject]),
    "Total_Price", SUM(subscriptions[total_price]),
    "Total_Subs", SUM(subscriptions[Subscribers]),
    "Subs_%_Amt", DIVIDE(SUM(subscriptions[Subscribers]), SUM(subscriptions[total_price]))
)
```

Measures

1. Highest number of subscriptions by subject

```
_MaxSubscribers =
MAXX(
    SUMMARIZE(
        subscriptions,
        subscriptions[Subject],
        "Total_Subscribers", SUM(subscriptions[Subscribers])
    ),
    [Total_Subscribers]
)
```

2. Least number of subscriptions by subject

```
_MinSubscribers =
MINX(
    SUMMARIZE(
        subscriptions,
        subscriptions[Subject],
        "Total_Subscribers", SUM(subscriptions[Subscribers])
    ),
    [Total_Subscribers]
)
```

3. Highest total price by subject

```
_MaxPrice = MAX(summary_table_paid[total_price])
```

4. Least total price by subject

```
_MinPrice = MIN(summary_table_paid[Total_Price])
```

5. Percentage of the number of courses anytime filters are used.

```
%_Courses =
DIVIDE(
    COUNTA(subscriptions[course_id]),
    CALCULATE(COUNT(subscriptions[course_id]), ALL(subscriptions))
)
```

6. Year over Year Percentage Change

```
total_price YoY% =
IF(
    ISFILTERED('subscriptions'[published_date]),
    ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date column."),
    VAR __PREV_YEAR =
        CALCULATE(
            SUM('subscriptions'[total_price]),
            DATEADD('subscriptions'[published_date].[Date], -1, YEAR)
        )
    RETURN
        DIVIDE(SUM('subscriptions'[total_price]) - __PREV_YEAR, __PREV_YEAR)
)
```

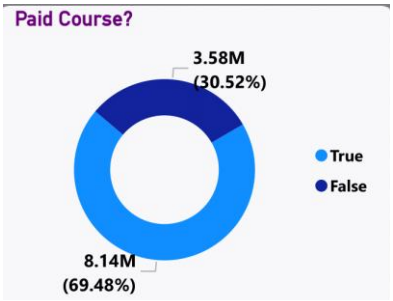
Phase 5 – Share Insights

Here are some observations, relationships, and insights based on the analysis of the **subscriptions dataset** for Udemy courses:

1. Paid vs. Unpaid Subscriptions

- **Observation**:

  The number of subscribers to **paid courses** (8.14M, 69.48%) significantly exceeds those of **unpaid courses** (3.58M, 30.52%). [*Ref.: "Paid Course?" – Summary Dashboard*]



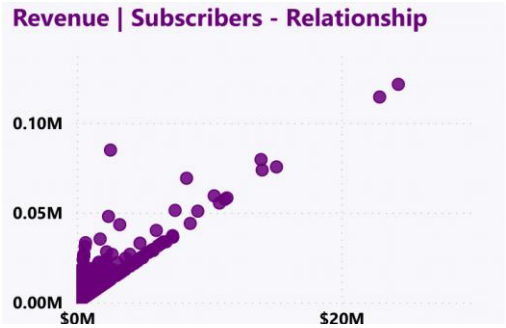
- **Insight**:

  Despite the difference in numbers, the fact that both paid and unpaid courses attract subscriptions suggests that the price (whether a course is paid or free) has **no direct impact on user interest** in subscribing.

2. Relationship between Paid Subscriptions and Revenue

- **Observation**:

  Paid course subscriptions drive revenue, showing a **positive correlation** between paid subscriptions and overall earnings. [*Ref.: "Revenue | Subscribers – Relationship" – Revenue Dashboard*]



- **Insight**:

  Increasing subscriptions for paid courses will result in higher revenue, emphasizing the importance of marketing strategies to attract more subscribers to paid offerings.

3. Relationship between Reviews and Subscribers

- **Observation**:

  There are **574.20K reviews** correlated with a total of **11.72M subscribers**. This is evident in the trend analysis visual displayed on the Subscriptions Dashboard. [*Ref.: "Trend Analysis– Subscribers | Revenue" – Summary Dashboard*]

- **Insight**:

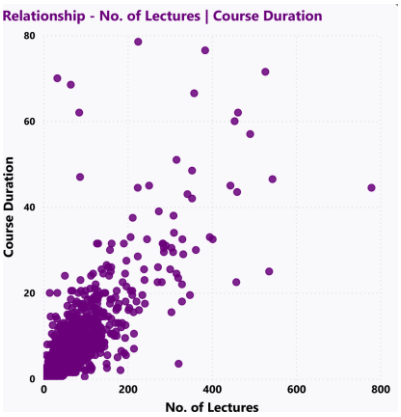
  A strong positive relationship exists between **reviews and subscriptions**. The more reviews a course has, the more likely it will attract additional subscribers.

5. Course Duration and Number of Lectures

- **Observation**:

  There is a **positive relationship** between **course duration and number of lectures**, meaning longer courses often have more lectures. [*Ref.: "Relationship - No. of Lectures | Course Duration" – Summary Dashboard*]

Subject	Sum of num_lectures	Sum of content_duration_(Hrs)	Sum of Subscribers
Web Development	63,357	6,709.75	7,937,287
Business Finance	38,663	4,237.22	1,868,711
Musical Instruments	26,055	1,940.98	846,689
Graphic Design	19,320	2,158.45	1,063,148

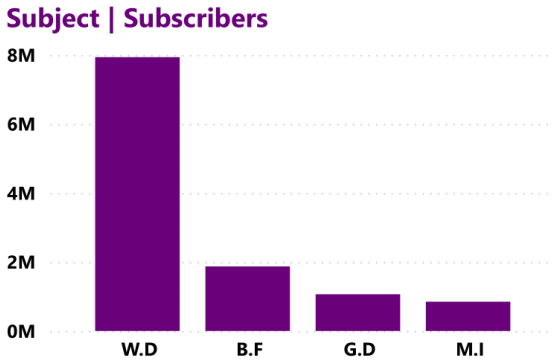


- **Insight**:

  Judging from the table above, neither course duration nor the number of lectures has a notable impact on **subscriber numbers**.

6. Subject with the Highest Subscribers and Revenue

- **Observation**:
  - **Web Development** has the highest number of subscribers (7.94M – 68%) and revenue ($627.60M – 71%).
  - This success is linked to its high number of reviews (429.50K – 75%).



Subject	Sum of Reviews	Sum of Subscribers
Business Finance	75,902	1,868,711
Graphic Design	37,070	1,063,148
Musical Instruments	31,724	846,689
Web Development	429,500	7,937,287

- **Insight**:

  Web Development's performance highlights the **impact of reviews** on both subscriptions and revenue.

7. Course Quantity vs. Performance (Business Finance vs. Web Development)

- **Observation**:
  - Business Finance has **1,190 courses** (32%) compared to Web Development's **1,199 courses** (33%).
  - However, its reviews (75.90K – 13%), subscribers (1.87M – 16%), and revenue ($123.74M – 14%) are significantly lower.

- **Insight**:

  The number of courses alone does not contribute to better performance in terms of reviews, subscribers, or revenue.

8. Reviews-Subscribers Proportion

- **Observation**:

  Across all subjects, the relationship between **number of reviews and subscribers** maintains an average proportion of **4%**. [*Ref.: "Reviews | Subscribers – Proportion" – Subscriptions Dashboard*]

Reviews | Subscribers - Proportion

Subject	Sum of Reviews	Sum of Subscribers	Reviews_%_ Sub
Business Finance	75,902	1,868,711	4.06%
Graphic Design	37,070	1,063,148	3.49%
Musical Instruments	31,724	846,689	3.75%
Web Development	429,500	7,937,287	5.41%

- **Insight**:

  The relationship between the four subjects and their various reviews and subscriptions is perfectly proportionate thus, there are no anomalies in the data.

Recommendation:

Since reviews boost subscriptions, **encouraging user reviews** can be a strategic method to increase subscriptions and, consequently, revenue.

Phase 6 – Act (Course of Action)

A. Boost User Engagements for Reviews

The following strategies can be implemented to boost reviews:

1. **Incentivize Reviews**: Offer small rewards, such as discount coupons for future courses or access to exclusive content, for users who leave detailed and honest reviews.
2. **Simplify the Review Process**: Make it easy and quick for users to leave reviews by providing a streamlined interface with options for both star ratings and short feedback.
3. **Personalized Follow-Up**: Send personalized thank-you messages and gentle reminders via email or in-app notifications, emphasizing the value of their feedback for improving the platform.
4. **Highlight the Impact of Reviews**: Show users how their reviews help other learners make better choices and contribute to enhancing course quality.
5. **Feature Top Reviewers**: Recognize and reward frequent or detailed reviewers by showcasing them in a "Top Reviewer" section or offering exclusive benefits.
6. **Leverage Social Proof**: Display the number of reviews and average ratings prominently to encourage participation by illustrating how other users are engaging with the review system.

B. Targeted Marketing for High-Growth Categories:

1. Allocate more resources to market courses in categories like **Web Development** where return on investment (ROI) is evident.
2. Diversify offerings in underperforming categories with creative course topics to appeal to wider audiences.