REPORT ON THE ANALYSIS OF UDEMY COURSES DATASET

BY

EVELYN WULLAR

PRESENTED TO

MR FRANCIS KARIKARI (BLOSSOM ACADEMY)

DATE: 18/09/2024

Table of Contents

[Introduction 1](#_Toc177434442)

[Questions (Problems) 1](#_Toc177434443)

[Methodology 1](#_Toc177434444)

[Data Preparation 1](#_Toc177434445)

[Data Visualization 2](#_Toc177434446)

[Power BI Reports 2](#_Toc177434447)

[Dashboard 4](#_Toc177434448)

[Conclusion 4](#_Toc177434449)

[Recommendations for Course Creators at Udemy 5](#_Toc177434450)

# Introduction

The Udemy courses dataset contains 3,682 records of courses from four subjects: Business Finance, Graphic Design, Musical Instruments, and Web Design. This analysis aims to uncover key insights about course popularity, revenue generation, engagement levels, and trends over time to inform decision-making for course creators and Udemy.

# Questions (Problems)

1. Which courses are the most popular based on the number of subscribers?
2. Which courses generate the most revenue?
3. What is the engagement level of each course, measured by the ratio of reviews to subscribers?
4. How are courses distributed across different difficulty levels?
5. Which subjects are the most popular based on the number of courses and subscribers?
6. What are the trends in course publication over time?

# Methodology

## Data Preparation

1. **Data Loading**:
   * The dataset was loaded into Power BI Desktop from a CSV file containing 3,682 records of Udemy courses.
   * The dataset includes columns such as course\_title, url, is\_paid, price, num\_subscribers, num\_reviews, num\_lectures, level, content\_duration, published\_timestamp, and subject.
2. **Data Cleaning**:
   * Ensured that all columns were correctly formatted. For example, price, num\_subscribers, and num\_reviews were formatted as numeric types.
   * Checked for and handled any missing or inconsistent data entries.
3. **Creating Calculated Columns**:
   * **Revenue Calculation**: A new column Revenue was created using the formula:

Revenue = [price] \* [num\_subscribers]

* + **Free courses/ paid courses**: two new columns were created namely: Free courses and Paid courses by separating the is\_paid column. To do this the conditional column was used where the value of Is\_Paid = TRUE was set to 1 and the null values represented FALSE

## Data Visualization

The Power BI Desktop application is used to create various visualizations and the power BI services is used to create a dashboard. Visualizations included bar charts, pie charts, line charts, scatter plots, and tables.

**Visualizations Created**

## Power BI Reports

**Data Overview**

The data comprised of 3,672 total courses out of which 310 are free and 3362 are paid courses. There are a total of 4 subjects and 4 levels and approximately 147,000 lectures. There was a total of 574,000 reviews and the highest is 27,445. Additionally, the highest course duration is 78.5 hours. While there are free courses, the highest price for a course is 200 pounds and the average price for all courses is 66.1 pounds.

**Top 5 Most Popular Courses**: A table is used to display the top 5 courses by the number of subscribers, price and total number of reviews. From the table, it was revealed that “Learn HTML 5 programming from scratch had the most number of subscribers and it is a free course. The course with the highest number of reviews is the web developer Bootcamp.

**Top 5 Courses by Revenue**: From the table, the top 5 courses by revenue are identified as the web developer Bootcamp, The complete web developer course, Pianofloral incredible New way to learn piano and keyboard, angular 4 and Javascript.

**Course Level Distribution**: A pie chart is used to show the distribution of courses across different levels. Courses at “All levels” were more with a percentage of 52.42, followed by “Beginner courses” which were approximately 34.5%, Intermediate courses which were approximately 11% and the lowest being expert courses at 1.58%.

**Subject Popularity**: A clustered column chart is used to display the popularity of different subjects based on the number of subscribers. Of the 4 subjects, web development had the highest number of subscribers, followed by Business Finance, Graphic Design and finally Musical instruments.

**Revenue Generation by Subjects and Courses**: a clustered column chart was used to display revenues generated by different subjects and levels. The charts revealed that the highest Revenue was generated from Web Development with an approximate sum of 7.9 million, with Business Finance coming second, Graphic design third and musical instruments generating the lowest revenue. At the different levels, the highest revenue is generated at “All levels”, followed by the “Beginner level”, “Intermediate” and “Expert”.

Revenue Generated Over time: A line chart was used to visualise the revenue generated with time. The trend revealed that the management of Udemy generated the most revenue in 2015 approximately 0.31 billion pounds. The revenue increased steadily from 2011 to 2014. In 2015, the revenue skyrocketed to the highest within the period and then began to decline again in 2016 and 2017.

**Publishing Trends Over Time**: A line chart was again used to show the number of courses published over time. Just like the revenue, number of books published increased over time from 2011 to 2016, where the highest number of courses were published. After 2016, there was a decline in the number of courses published.

# Dashboard

# Conclusion

The analysis of the Udemy courses dataset has provided several key insights into course popularity, revenue generation, engagement levels, and trends over time. The findings reveal that:

1. The most popular course by the number of subscribers is “Learn HTML 5 Programming from Scratch,” which is a free course.
2. The course with the highest number of reviews is “The Web Developer Bootcamp.”
3. The top 5 revenue-generating courses include “The Web Developer Bootcamp,” “The Complete Web Developer Course,” “Pianofloral: Incredible New Way to Learn Piano and Keyboard,” “Angular 4,” and “JavaScript.”
4. Web Development is the most popular subject, followed by Business Finance, Graphic Design, and Musical Instruments.
5. Web Development generates the highest revenue, followed by Business Finance, Graphic Design, and Musical Instruments.
6. The highest revenue was generated in 2015, with a steady increase from 2011 to 2014, a peak in 2015, and a decline in subsequent years.
7. The number of courses published increased steadily from 2011 to 2016, with a peak in 2016, followed by a decline.

# Recommendations for Course Creators at Udemy

1. Given that Web Development and Business Finance are the most popular and revenue-generating subjects, course creators should consider developing more courses in these areas to attract a larger audience and increase revenue.
2. The success of the free course “Learn HTML 5 Programming from Scratch” suggests that offering free courses can significantly boost subscriber numbers. Course creators can use free courses as a marketing tool to attract subscribers and then upsell paid courses.
3. Courses with higher engagement levels, such as “The Web Developer Bootcamp,” tend to have more reviews and subscribers. Course creators should focus on creating engaging content and encouraging reviews to boost course popularity and credibility.
4. While “All Levels” courses are the most common, there is still a significant demand for “Beginner” and “Intermediate” courses. Course creators should consider offering a range of courses at different difficulty levels to cater to a broader audience.
5. The analysis shows that revenue and the number of courses published peaked in 2015 and 2016, respectively. Course creators should stay informed about industry trends and adjust their strategies accordingly to maximize their reach and revenue.
6. The top revenue-generating courses are not necessarily the most expensive. Course creators should experiment with different pricing strategies to find the optimal price point that maximizes both subscribers and revenue.