UDEMY DATASET ANALYSIS

BY:

Abdulrahman Hamzat

TABLE OF CONTENT

ABSTRACT

This project focused on analyzing Udemy course data (2011 – 2017) to provide actionable insights aimed at improving the company's business model. The analysis covered subscription data across four subjects: web development, musical instruments, business finance, and graphic design with a primary objective of identifying trends that can drive revenue growth.

Data cleaning and Exploratory Data Analysis (EDA) was conducted in Google Sheets, and the explanatory visualizations were created using Tableau Public to convey the findings.

Key findings revealed that web development courses have the highest number of subscribers and highest average prices, while musical instruments lag in both areas. The top 20 most popular courses also overwhelmingly belong to the web development category. And no direct correlation was found between course price and content duration.

The company should focus on increasing paid course offerings in web development and promoting business finance courses through targeted marketing.

SECTION ONE

PROJECT DESCRIPTION

This data analytics project was completed during my enrollment in EntryLevel's Data Analyst Level 1 course.

For this project, I took on the role of a data analyst working at Udemy - an education tech company. I was tasked to generate insights which will help to improve the organization's business model and increase the next quarterly earnings. The findings I generate would help the manager to make smart, data driven decisions.

SECTION TWO

PROBLEM DEFINITION

- The business problem is to better understand the pricing and subscriptions data of Udemy courses for four different subjects: web development; musical instruments; business finance; graphic design, as well as identifying trends and patterns.

- This will help the management to identify for which subject more courses should be created, and also help to create targeted strategies to increase the company revenue.

- The data analysis report was expected to be provided within 4 weeks wherein I sought to investigate the following questions:

  1. What is the total number of subscribers in each subject?

  2. How does the average content duration /price/number of students vary across different subjects?

  3. How many courses are free and paid for each subject?

  4. What is the average price of web development courses at different levels?

  5. Does content duration impact the price of the course?

  6. What are the 20 most popular courses?

SECTION THREE

DATA DESIGN

- The dataset contains subscription information of Udemy courses collected from 2011 - 2017. It includes the level of courses, rating, number of reviews, number of subscribers, price, and published time for each course.

- My data cleaning efforts include removing blank cells and incomplete data, correcting invalid and inconsistent data using google sheets. Next, the cleaned data was visualized using Tableau, and all visuals were organized in a dashboard.

- I used Google sheets pivot chart and chart tools to perform exploratory data analysis in order to assess the dataset both statistically and visually.

SECTION FOUR

FINDINGS

4.1   Finding #1

Web development has the highest number of subscribers (67.9%) and musical instruments has the least (7.2%).

subject	SUM of num_subscribers
Web Development	7981935
Business Finance	1868711
Graphic Design	1063148
Musical Instruments	846689
Grand Total	11760483

Fig 4.1: Pivot table showing total count of subscribers for each subject.

4.2   Finding #2

Web development has the highest for all average values of content duration, subscriber counts, and price. While musical instruments have the least average values.

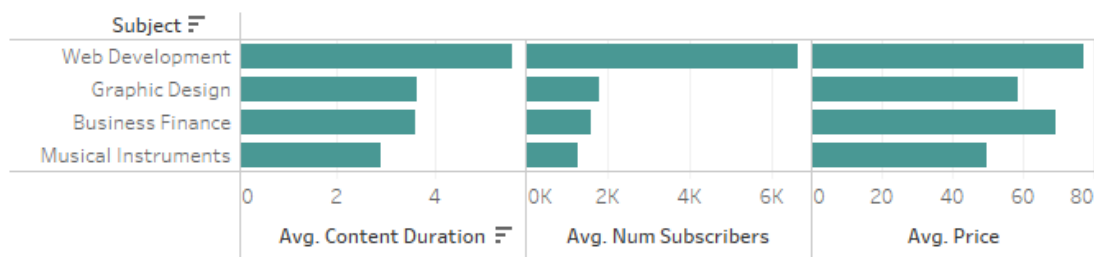Average Content Duration/ Number of Subscribers/Price across all subjects

Fig 4.2: Tableau viz. of average content duration, number of subscribers, and price across all subjects.

4.3   Finding #3

Total paid courses are about ten times more than the total free courses.

COUNTA of free or p	free or paid		
subject	free	paid	Grand Total
Web Development	134	1069	1203
Business Finance	96	1059	1155
Musical Instruments	46	623	669
Graphic Design	35	549	584
Grand Total	**311**	**3300**	**3611**

Fig 4.3: Pivot Table showing count of free and paid courses for each subject.
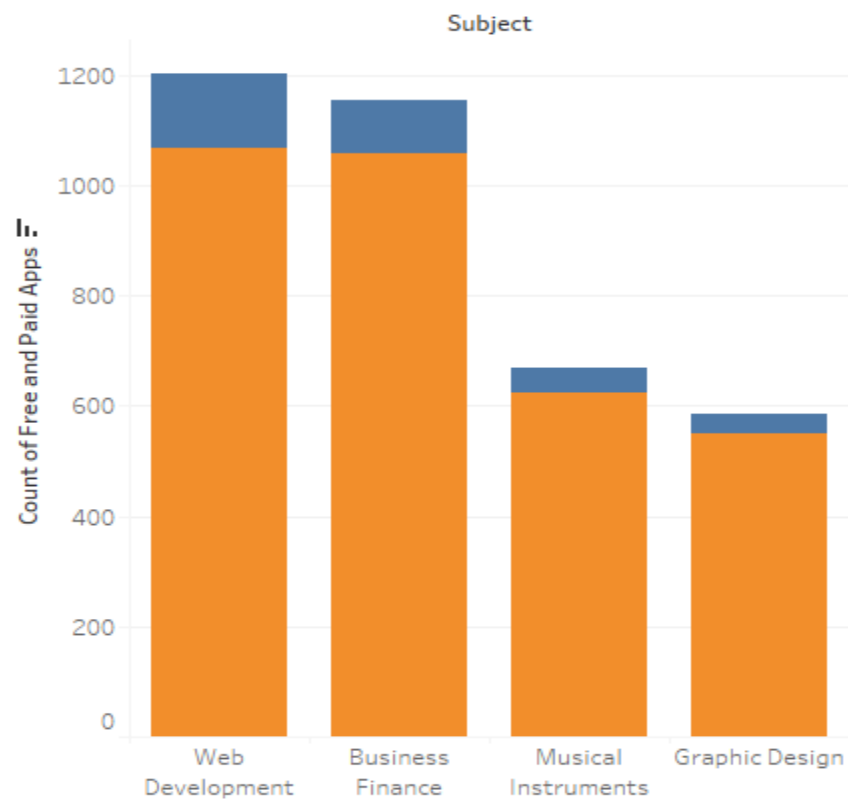
Count of free and paid courses for all subjects

Fig 4.4: Tableau viz. of free and paid courses for each subject.

4.4   Finding #4

The highest average price by levels of web development courses goes to the "intermediate level", while the "Expert level" has the least average price.

AVERAGE of price	subject
level	Web Development
Intermediate Level	85.07
Beginner Level	78.54
All Levels	74.55
Expert Level	67.14
Grand Total	77.04

Fig 4.5: Pivot table showing average price of web development courses at different levels.

4.5   Finding #5

As at the time this data was collected, the most popular course – **"Learn HTML5 Programming from Scratch"** - has a subscriber count of 268,923. While the 20th popular course – **"Learn and Understand Angular JS"** - has a subscriber count of 59,361. Seventeen out of the twenty most popular courses belong to web development.

Plot of top 20 courses with their respective number of subscribers

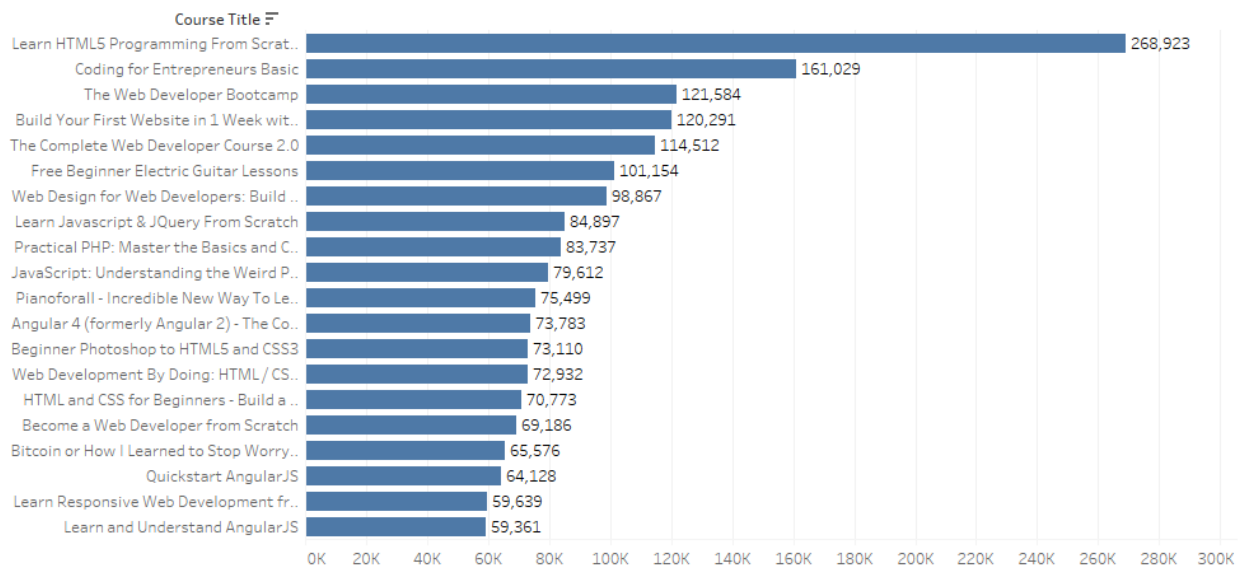| Course Title ⇉	
Learn HTML5 Programming From Scrat..	268,923
Coding for Entrepreneurs Basic	161,029
The Web Developer Bootcamp	121,584
Build Your First Website in 1 Week wit..	120,291
The Complete Web Developer Course 2.0	114,512
Free Beginner Electric Guitar Lessons	101,154
Web Design for Web Developers: Build ..	98,867
Learn Javascript & JQuery From Scratch	84,897
Practical PHP: Master the Basics and C..	83,737
JavaScript: Understanding the Weird P..	79,612
Pianoforall - Incredible New Way To Le..	75,499
Angular 4 (formerly Angular 2) - The Co..	73,783
Beginner Photoshop to HTML5 and CSS3	73,110
Web Development By Doing: HTML / CS..	72,932
HTML and CSS for Beginners - Build a ..	70,773
Become a Web Developer from Scratch	69,186
Bitcoin or How I Learned to Stop Worry..	65,576
Quickstart AngularJS	64,128
Learn Responsive Web Development fr..	59,639
Learn and Understand AngularJS	59,361

Fig 4.6: Tableau viz. of the 20 most popular courses.

4.6   Finding #6

It is observed that for all ranges of prices, most of the course content durations fall between 0 - 20 hours. Also, no direct correlation is observed between course content durations and course prices.
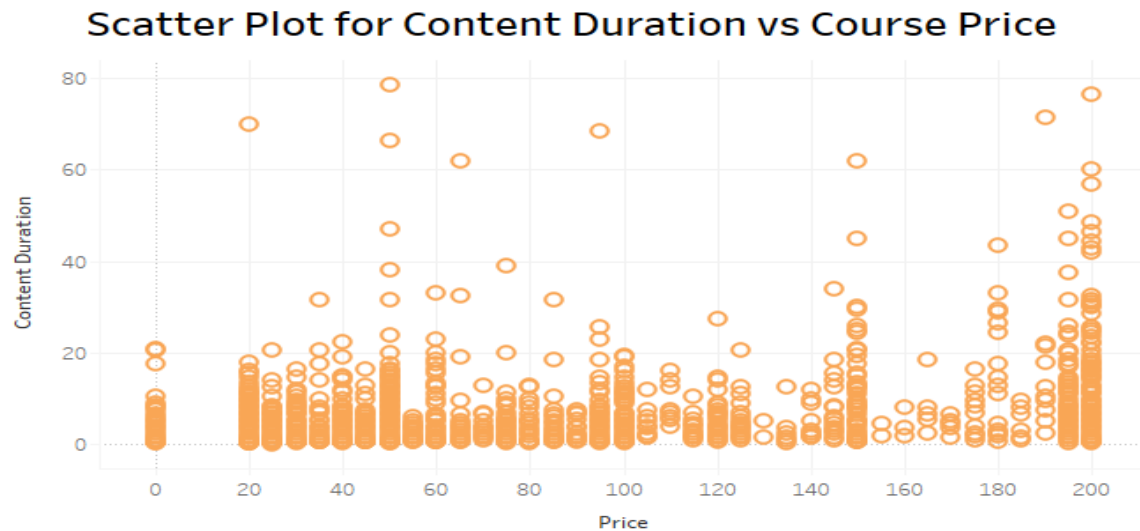
Scatter Plot for Content Duration vs Course Price

Fig 4.7: Tableau viz. of content duration against price.

7

SECTION FIVE

SUMMARY

- Almost 68% of subscribers come from Web Development and the top 5 courses are for web development as well.

- Business Finance and web development have almost equal number of total courses (1155 vs 1203), but web development courses are almost 2x in content duration and have 4x the average subscribers.

- Both web development and business finance have almost equal average prices, and their total number of paid courses are almost the same.

- Web development intermediate level courses have the highest average price, followed by the beginner level, all level, then the expert level in decreasing order.

- Overall web development courses have the highest average price, followed by business finance, graphic design, then musical instruments.

- There isn't much evidence to say that the price is directly related to course content duration.

SECTION SIX

ACTIONS AND RECOMMENDATIONS

- Focus on getting more paid courses by creators in web development in order to boost revenue. Currently, our data tells that web development has the highest number of subscribers.

- Our data tells us that our web development courses are most popular and people are willing to pay for it. Marketing and advertising campaigns should therefore be focused on business finance courses because it has almost the same average price and total number of paid courses with web development, yet it is less popular.

APPENDIX

Project's Dashboard Link:

https://public.tableau.com/app/profile/abdulrahman.hamzat1331/viz/UdemyDatasetDashboard/Dashboard1