

## **Data Analysis Portfolio**

This is a testimony to my affinity to stitch narrations and stories based on the data that I process and observe. I love looking for interesting trends and pouring into the statistics to make the narrations more interesting. Apart from crunching the numbers, I emphasize on writing as simple and comprehensible stories as possible to the ease of the general, non-technical consensus.

My preference for story writing isn't constricted to just data analysis, as I also incline towards writing user personas, strategy documents for my UX projects.

### **Professional Background**

I am a senior year student, currently pursuing my undergraduate in the field of Computer Science Engineering. Other than the classroom education, I prefer self learning and a hands-on learning approach to learn new things. I am keenly interested in improving my expertise in the field of data analysis, working from the ground-up and learning from the experts.

Simultaneously, I would prefer to contribute significantly for the project or organisation that I would have a chance to work with. I am a newbie entering this field of data analysis, having no professional experience as of now, however, to compensate, I have done a couple of projects through which I have tried to touch upon the majority of steps in this process.

### **Table Of Contents**

<b>Udemy Project Description</b>	<b>3</b>
<b>The Problem</b>	<b>4</b>
<b>Design</b>	<b>5</b>
<b>Required Findings and Insights</b>	<b>6</b>
<b>Analysis</b>	<b>13</b>
<b>Conclusions</b>	<b>16</b>
<b>Capstone Project Description</b>	<b>18</b>
<b>Design</b>	<b>19</b>
<b>Findings</b>	<b>20</b>
<b>Analysis</b>	<b>21</b>
<b>Conclusions</b>	<b>23</b>
<b>Appendix</b>	<b>24</b>

The project comprises a dataset about the courses offered by Udemy – an online learning platform that enables users to upskill through a MOOC style learning pattern. The courses are in four broad subjects namely, web development, musical instruments, graphic design and business finance. The data further provides stats about the number of subscribers, rating per course (range : 0-1), total duration, pricing, date of release etc. Through this dataset, several functions have also been implemented (IF, IF AND, LEFT, RIGHT, VLOOKUP etc) and top 20 subscribed courses have been fetched. Furthermore, to implement data visualization, pivot tables are used with the cleaned and complete dataset to represent various trends through graphs. Finally, conclusions are made on the basis of those trends.

## **The Problem**

*“To present the data on course revenue, understand where opportunities to increase revenue may lie, and track the performance of courses.”*

Total time to complete this project is three weeks, during which the data has to be cleaned, meaningful insights are to be made and a narration is to be framed as well.

**Design**

The dataset provided initially has been cleaned as – empty rows are removed, every record has been validated for its uniqueness, columns such as num\_subscribers, price etc have been equipped with a filter that displays entire data accordingly with ease. After the initial preprocessing, the prepared dataset is used to make visualizations (through pivot tables option) several graphs visualize trends like average subscribers per subject, average pricing, average rating per subject etc. Finally, Tableau, a powerful BI tool has been used for data visualization, where all viz have been uploaded.

## **Findings**

Cleaned dataset and visualizations on excel:


 Udemy Courses Consolidated Data (Portfolio Project)

Tableau visualizations (dashboard):

[https://public.tableau.com/views/UdemyCourseDataAnalysis/Dashboard1?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/UdemyCourseDataAnalysis/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link)

**Few insights from the cleaned dataset:**

- i Maximum subscribers of free courses is also from web dev.
- ii There are 11 courses that have a rating of 1 (excellent), however, to determine if the courses are performing well, their reviews are too few. Thus, along with a higher rating, the number of subscribers should also be considered.
- iii The best performing course is from web dev, ratio of subscription to rating high. However, the price is 0 dollars, so if that is upped a bit, e.g. made 10 dollars (apt for a beginner level course), the revenue can be increased significantly.
- iv Similarly, the second most subscribed course, also from web dev, should be upped to 20 dollars as it is an expert level course.

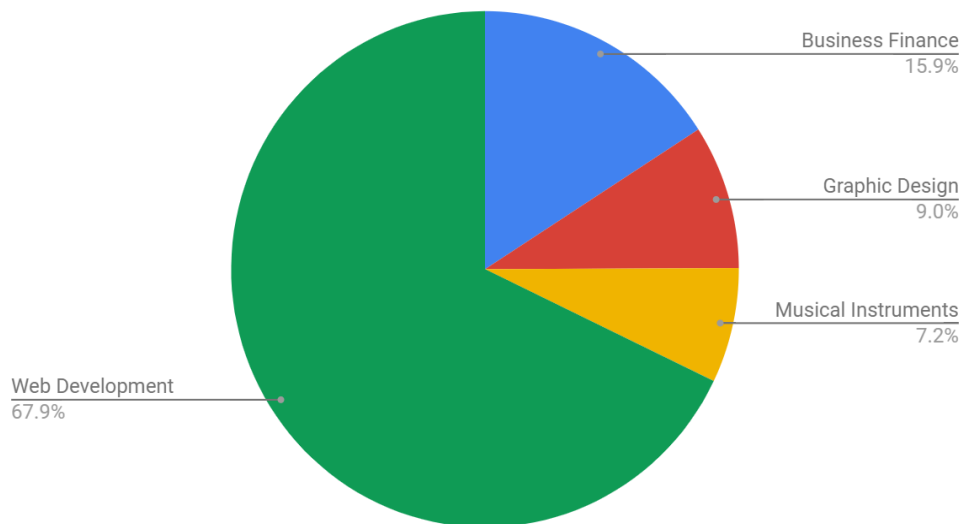
### Required Findings

The main insights to be made included:

Total number of subscribers for each subject (Pie Chart)

<i>subject</i>	SUM of num_subscribers
Business Finance	1868711
Graphic Design	1063148
Musical Instruments	846689
Web Development	7981935
<b>Grand Total</b>	<b>11760483</b>

Total number of subscribers for each subject

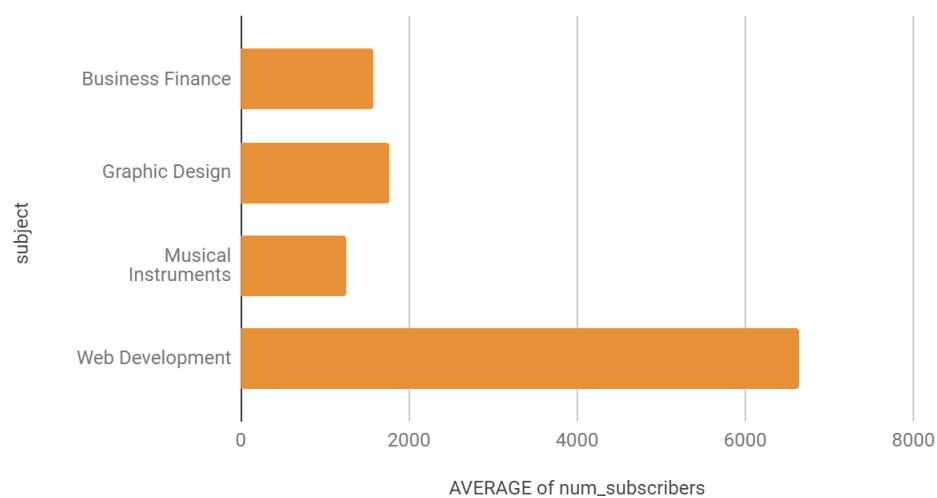


Average number of subscribers for each subject (Bar Chart)

<i>subject</i>	AVERAGE of num_subscribers
----------------	----------------------------

Business Finance	1569.026868
Graphic Design	1766.026578
Musical Instruments	1245.130882
Web Development	6635.024938
<b>Grand Total</b>	<b>3199.260881</b>

Average number of subscribers for each subject



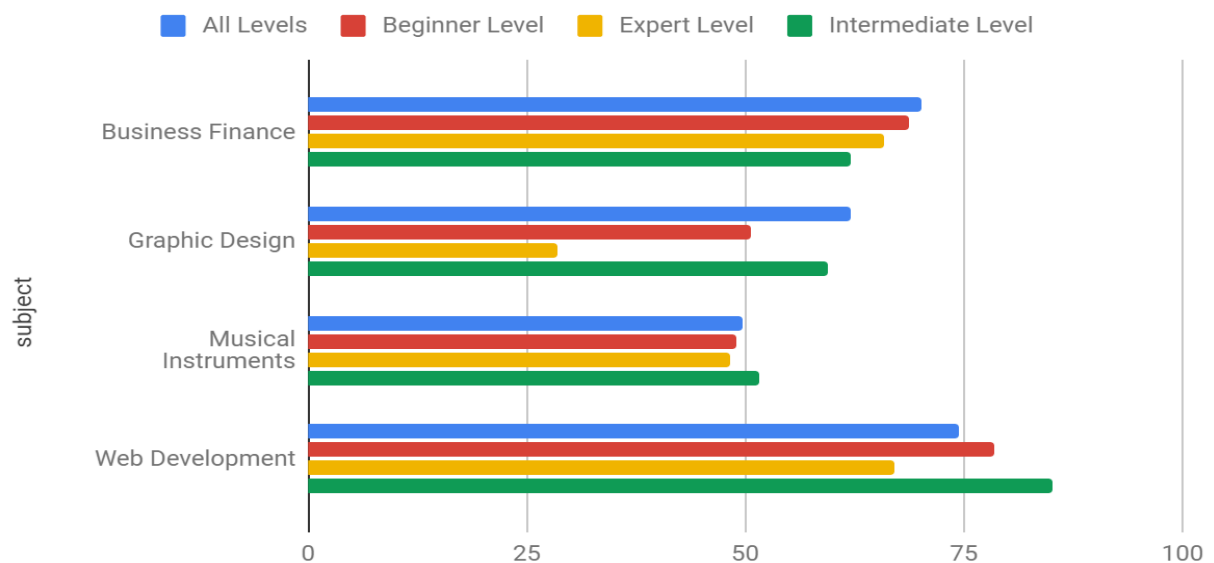
Average cost per subject at each level (Bar Chart)

*AVERAGE of price*    *level*



<i>subject</i>	All Levels	Beginner Level	Expert Level	Intermediate Level
Business Finance	70.19747235	68.73433584	65.8	62.01492537
Graphic Design	62.11940299	50.67934783	28.57142857	59.40789474
Musical Instruments	49.58333333	48.98496241	48.33333333	51.6025641
Web Development	74.54976303	78.54265403	67.14285714	85.07462687
<b>Grand Total</b>	<b>66.75324675</b>	<b>65.24390244</b>	<b>58.01724138</b>	<b>66.94312796</b>

### Average cost per subject at each level

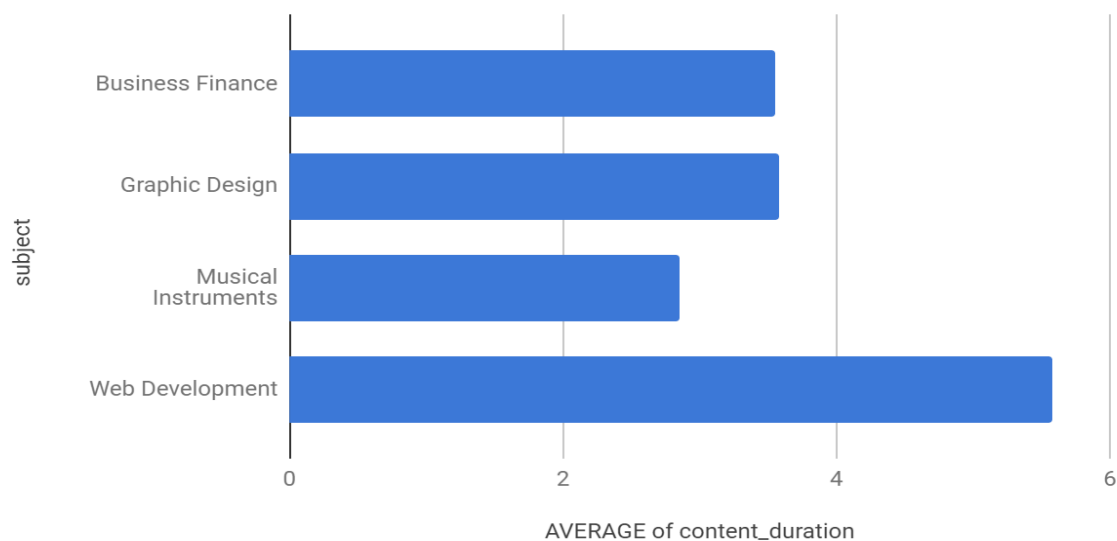


### Average content duration for each subject (Bar Chart)

<i>subject</i>	AVERAGE of content_duration
----------------	-----------------------------

Business Finance	3.557696613
Graphic Design	3.585465116
Musical Instruments	2.854387255
Web Development	5.586658354
<b>Grand Total</b>	<b>4.096137106</b>

Average content duration for each subject

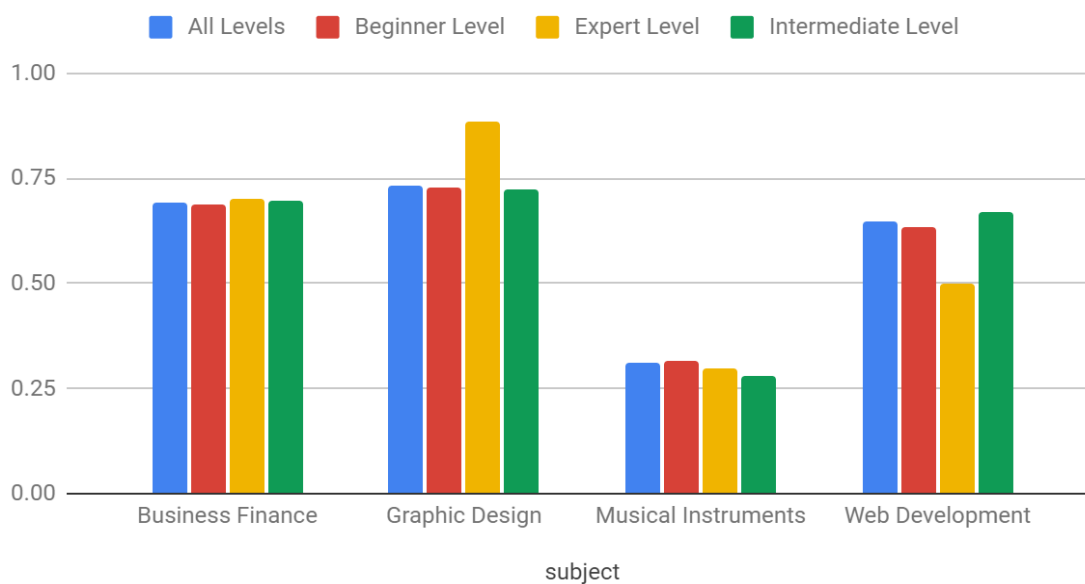


Average rating per subject for each level (Column Chart)

AVERAGE of Rating	level			
subject	All Levels	Beginner Level	Expert Level	Intermediate Level

Business Finance	0.6915165877	0.6862155388	0.7	0.6953731343
Graphic Design	0.7309850746	0.7270652174	0.8842857143	0.7215789474
Musical Instruments	0.312037037	0.3138345865	0.2966666667	0.281025641
Web Development	0.6454660348	0.6353791469	0.4985714286	0.6708955224
<b>Grand Total</b>	<b>0.6193714286</b>	<b>0.5973170732</b>	<b>0.5901724138</b>	<b>0.6157345972</b>

### Average rating per subject for each level



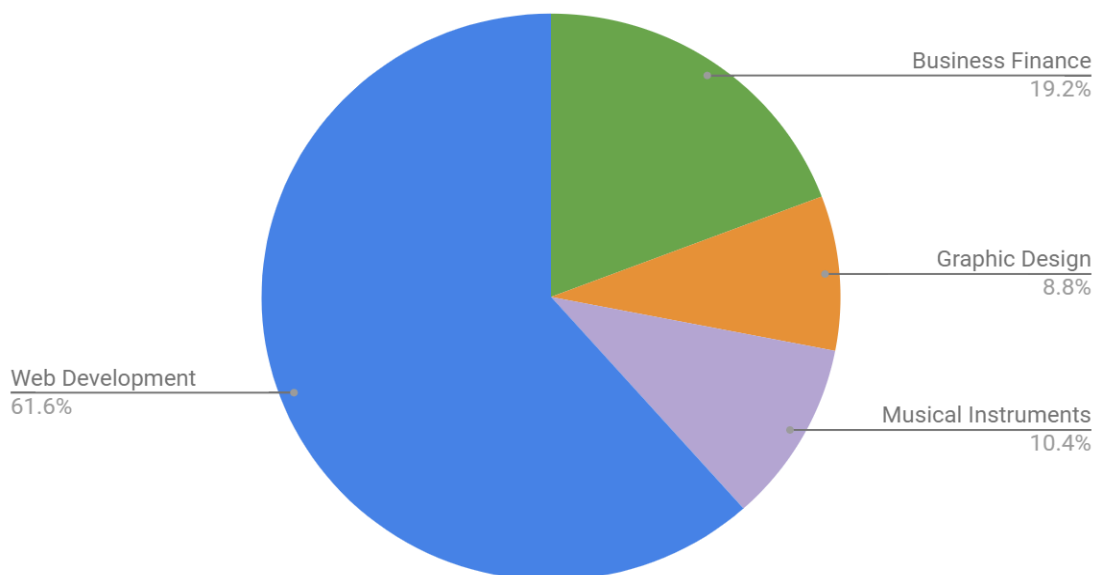
Other relevant insight(s) [Total subscribers of the free courses from all subjects]

<i>SUM of num_subscribers</i>	<i>level</i>			
<i>subject</i>	All Levels	Beginner Level	Expert Level	Intermediate Level

Business Finance	390849	140186	15955	56238
Graphic Design	179107	95600		10114
Musical Instruments	210626	62045	61	32003
Web Development	1251319	806244	170862	165342
<b>Grand Total</b>	<b>2031901</b>	<b>1104075</b>	<b>186878</b>	<b>263697</b>

Here, there is no free course of expert level in graphic design

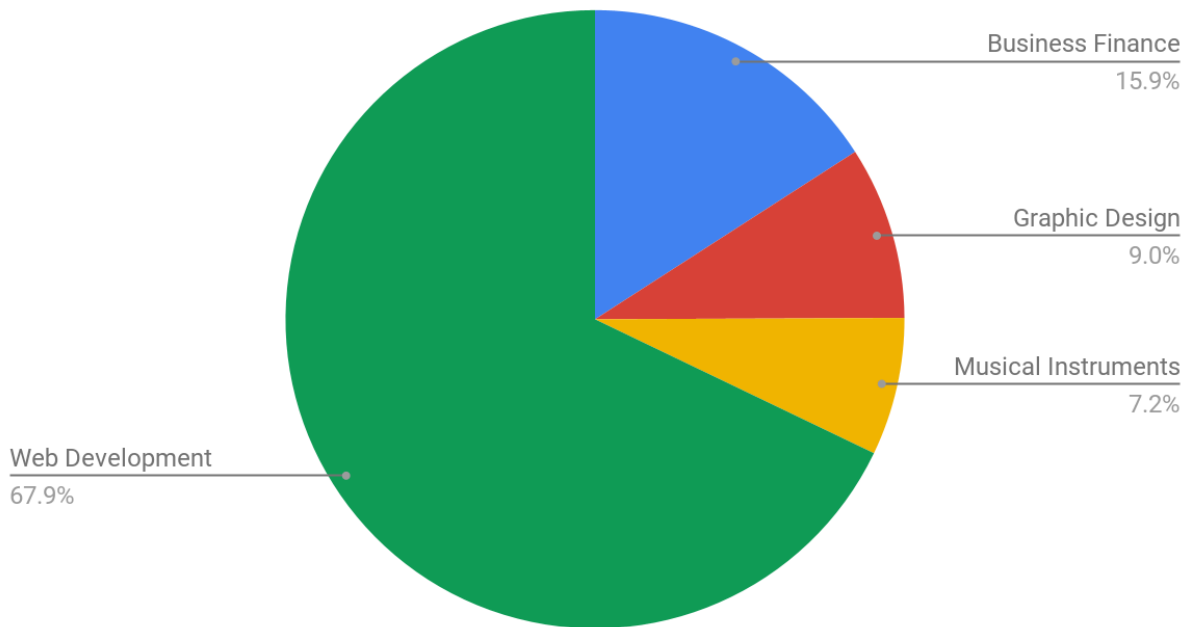
Total subscribers of free courses for all levels (subject wise)



### Analysis

**Problem 1: Why is web development highly subscribed?**

Total number of subscribers for each subject



Looking at the dashboard, a trend that is noticeable - *web development* is the most popular subject (more than 67% subscriptions), given the high demand in today's tech savvy world, people are inclining towards this field of computer science, mainly due to the high salaries and even greater demands. The most popular *and* free course from this subject is - **Learn HTML 5 Programming From Scratch**, this is a basic requirement for this subject. The popular web dev courses are all about learning the skill of creating static and responsive websites, useful for a wide range of public, from tech workers to business owners.

course title	no. of subscribers	level	free/paid	beginner free	duration	date of publishing
Learn HTML5 Programming From Scratch	268923	Beginner Level	Free	Free for beginners	10.5	2013-02-14
Coding for Entrepreneurs Basic	161029	Expert Level	Free	Paid/other levels	3.5	2013-06-09
The Web Developer Bootcamp	121584	Beginner Level	Paid	Paid/other levels	43	2015-11-02
Build Your First Website in 1 Week with HTML5 and CSS3	120291	All Levels	Free	Paid/other levels	3	2014-04-08
The Complete Web Developer Course 2.0	114512	Beginner Level	Paid	Paid/other levels	30.5	2016-03-08
Free Beginner Electric Guitar Lessons	101154	All Levels	Free	Paid/other levels	4.5	2012-06-15
Web Design for Web Developers: Build Beautiful Websites!	98867	All Levels	Free	Paid/other levels	3	2015-04-13
Learn Javascript & JQuery From Scratch	84897	All Levels	Paid	Paid/other levels	2	2013-10-10
Practical PHP: Master the Basics and Code Dynamic Websites	83737	Intermediate Level	Free	Paid/other levels	6.5	2014-07-19
JavaScript: Understanding the Weird Parts	79612	All Levels	Paid	Paid/other levels	11.5	2015-03-12
Pianoforall - Incredible New Way To Learn Piano & Keyboard	75499	Beginner Level	Paid	Paid/other levels	30	2014-08-07
Angular 4 (formerly Angular 2) - The Complete Guide	73783	Beginner Level	Paid	Paid/other levels	22	2016-02-11
Beginner Photoshop to HTML5 and CSS3	73110	All Levels	Free	Paid/other levels	2	2012-07-27
Web Development By Doing: HTML / CSS From Scratch	72932	All Levels	Free	Paid/other levels	1	2013-09-25
HTML and CSS for Beginners - Build a Website & Launch ONLINE	70773	All Levels	Free	Paid/other levels	6	2015-03-19
Become a Web Developer from Scratch	69186	All Levels	Paid	Paid/other levels	27.5	2011-11-19
Bitcoin or How I Learned to Stop Worrying and Love Crypto	65576	All Levels	Free	Paid/other levels	8	2013-04-20
Quickstart AngularJS	64128	Beginner Level	Free	Free for beginners	1.5	2014-11-22
Learn Responsive Web Development from Scratch	59639	All Levels	Free	Paid/other levels	4.5	2013-12-09
Learn and Understand AngularJS	59361	Beginner Level	Paid	Paid/other levels	7	2014-09-24

(In Pic: The top 20 subscribed courses)

Appositely, only four courses from the other disciplines feature in the top 20.

The top 20 subscribed courses are mostly beginner and all levels. The paid courses are usually of longer duration – around 10 hours or more, or are of intermediate or higher levels.

### ***Problem 2: Why are the other three subjects not as highly subscribed?***

The courses from a subject like business finance for instance, are in general only taken by those who either need to learn it for their own use (help in running their business, for school or assignments) and are rarely taken as a source of leisure or out of sheer interest. But, since the demand of learning business is not as acute as that of web dev, it is the second most subscribed subject. Now, for the remaining subjects, their courses are merely a source of extracurricular learning, or taken as a hobby. People learning musical instruments from an online course would probably only see themselves playing for leisure. Graphic designing is more popular between the two as, along with web dev, people tend to use this skill in embellishing their projects.

***Problem 3: How can more revenue be generated for the courses?***

Courses which have a **higher** number of subscribers but are *free*, must be priced at a **reasonable price**, in accordance with their difficulty level. The beginner courses can start from \$10 dollars and the expert can go even upto \$100, however, the relevance of courses must also be regarded. **The courses which are devoid of any learners should be considered to be removed from the platform in order to cut down the prices and improve our profit margins.** The pricing for the courses can be modified a bit, i.e. **adding various pricing tiers, monthly payments** of a small amount ensure that the learners do not shy away from the courses simply due to the money factor. Eg- a course called **“The Web Developer Bootcamp”** is of \$200 and has a good 121k learners, can have pricing as 50 dollars per month for 4 months of course duration. Also, lucrative offers like **one-on-one mentor review, portfolio and resume reviews (only for relevant courses), additional training material** can be provided to our courses of graphic design and web dev. Several tiers can be decided with a few prices that would provide few of these features. **Courses for all subjects should provide learners with additional learning tips, follow ups and links to various blogs pertinent to the nature of their course mailed to them periodically.** This ensures that the learners are **on target** to complete their course, they get more encouragement to finish it and a lower urge to procrastinate or return it.

The course named **“The Complete Web Development Course - Build 15 Projects”** has only 7k subscribers, mainly because it has a duration of 70+ hours, the second lengthiest course in the dataset. This and other such **lengthy courses can be optimized** by providing the content *one module at a time, say weekly, with a deadline for module assignments every weekend.* This ensures that the learner is not overwhelmed by seeing a large number of modules altogether. An inmail reminder of a new module and assignment deadline can help engage the learner too.

## Conclusion

The course dataset has been cleaned and processed, and several insights have been drawn regarding the same. The trends have been represented visually, all using Excel and Tableau. Web dev and business finance are two hugely popular niches and measures mentioned above should be considered to give a lift to the revenue that's coming in from their courses. Course pricing should be altered a bit, adding multiple payment options like upfront, installments or even timed trials then pay. The learners should be introduced to these tiers with various features that help to gain more sales and subscriptions. The courses should also be revealed module by module, to keep the learner engaged and energized. Regular contact must be done with the learners to ensure they are motivated to learn and be consistent with their progress. Regular mails can also act as a reminder to their modules and assignments.

Thus, the udemy course dataset has been cleaned, processed and various insights have been drawn from the trends. These are then visually represented, devoid of any technical jargon, for the ease of understanding of the clients and non technical users.



## **Capstone Project Description**

### **Ramen Reviews Data Analysis**

Penne for your thoughts, an international noodle firm has several subsidiaries across 30+ nations. It's brands work reasonably well, selling out various kinds of ramen noodles ranging from cup to tray style noodles. However, India is one nation where only pack styles are sold and reviewed in majority.

The dataset available at our dispense contains fields like reviews, brand name, variety offered, style, country sold in and stars received.

Our problem stood in identifying how the firm can take some actions to increase sales across all noodle styles in this country. We had to survey various brands that sell out here, and the ones which have lower ratings. Our focus remained on improving sales as well as reviews of the maximum number of brands. Through this dataset, several functions have also been implemented (IF, IF AND, LEFT, RIGHT etc). Furthermore, to implement data visualization, pivot tables are used with the cleaned and complete dataset to represent various trends through graphs. Finally, conclusions are made on the basis of those trends.

## Project Description

*“To present the data on ramen reviews, understand where opportunities to increase sales may lie, and track the performance of brands.”*

Total time to complete this project is one week, during which the data has to be cleaned, meaningful insights are to be made and a narration is to be framed as well.

## Design

The dataset provided initially has been cleaned as - empty rows are removed, every record has been validated for its uniqueness, columns such as stars, country etc have been equipped with a filter that displays entire data accordingly with ease. After the initial preprocessing, the prepared dataset is used to make visualizations (through pivot tables option) several graphs visualize trends like total reviews per style of ramen, number of brands per country, total reviews per country and ramen review breakdown of India. Finally, Tableau, a powerful BI tool has been used for data visualization, where all viz have been uploaded.

## Findings

Cleaned dataset and visualizations on excel:


 Ramen Reviews

Tableau visualizations (dashboard):

[https://public.tableau.com/views/RamenReviews/Dashboard1?:language=en-US&publish=yes  
&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/RamenReviews/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)

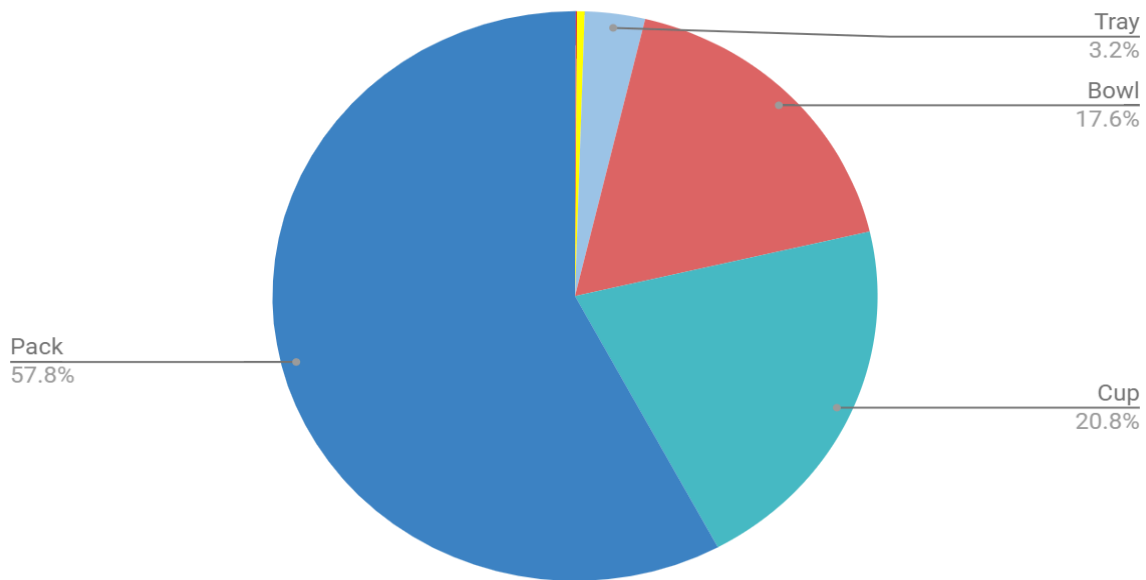
**Few insights from the cleaned dataset:**

- i**      Maximum ramen reviews are generated from Japan.
- ii**     The overall reviews are roughly proportional to the sales generated across the globe, which signify that the pack style noodles are the top sellers consistently in almost all the nations. The bar style on the other hand has a meagre share of 0.03% total reviews, implying a similar sale ratio.
- iii**    Nigeria has only one brand of ramen noodles, while Japan has more than 300 brands of ramen.
- iv**     Penne for your thoughts is yet to reach a majority of the African, European and the South American continent.

## Analysis

*Problem 1: Why is the pack style ramen highly sold?*

Total Reviews By Style



The pack style ramen accounts for more than half of the reviews and proportionate sales of the ramen across 30+ nations. This is mainly due to the fact that the style of noodles do not really matter to people unless in exceptional cases like travellers, hikers etc. who prefer hassle free preparation and easy to have noodles. The firm should also focus if the packs sold have the same varieties or offer some classic flavours which are preferred by the majority.

***Problem 2: Why do few nations majorly show sales of only one type of noodles?***

SUM of Review	Style			
Country	Cup	Pack	Grand Total	
India		1951	45238	47189

Countries like India, Nigeria, Netherlands, Mexico, Myanmar etc usually imply a sale of not more than two types of ramen noodles. This is due to the fact that the variety offered is covered by those styles and the lesser sold types either are a bit too pricey or not as appealing as the top sellers.

***Problem 3: What can be done to improve sales of various noodle types across client nations?***

The prices of non pack style ramens can be made a bit customer friendly along with producing identical varieties as offered in the pack style, making preferred flavours available in various portions, thus, boosting sales. This however, should be done after further analysis of highly sold varieties. Then, a survey should be done to know more about the palettes of people from minority client nations, to expand the business venture.

## **Conclusion**


The review dataset has been cleaned and processed, and several insights have been drawn regarding the same. The trends have been represented visually, all using Excel and Tableau. The pack style noodles are sold the most across various nations. Japan generated the most reviews and probably more sales too. While the brands reviewed imply a greater outreach in the Asian and North American nations, the firm is yet to expand its reach to South American, African and European nations. The ramen styles which are sold less should be boosted by offering tested varieties and those which are preferred by the general consensus.

Thus, the review dataset has been cleaned, processed and various insights have been drawn from the trends. These are then visually represented, devoid of any technical jargon, for the ease of understanding of the clients and non technical users.

## Appendix

### Coursework Project links (Udemy Dataset)

1. Cleaned and processed dataset:

 Udemy Courses Consolidated Data (Portfolio Project)

2. Tableau dashboard link:

[https://public.tableau.com/views/UdemyCourseDataAnalysis/Dashboard1?:language=en-US&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/UdemyCourseDataAnalysis/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link)

### Capstone Project (Ramen Reviews)

1. Cleaned and processed dataset:

 Ramen Reviews

2. Tableau dashboard link:

[https://public.tableau.com/views/RamenReviews/Dashboard1?:language=en-US&publish=yes&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/RamenReviews/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)