

Data Analysis Portfolio

AM is a Data Analyst Intern with experience in Data Consolidation, Cleaning and Visualization. AM specializes in Python and Tableau analytic technologies to develop strong analytic insight that meet business need.

Professional Background

During the first half of the year 2021, I discovered a need to proffer insight from data to meet business needs of both individuals and organisations by suggesting feasible plans to embark on as a result of analysis on information provided.

During the past months I participated in Entry Level virtual internship program as a data analyst intern to develop my analytic skill set for a future in the field of Data and Business Analysis. While in the program I participated in the UdeMY Course Review project where I discovered that Web Development skill is a sought after skill of a lot online subscribers.

Though being a student, I was able to meet up with the deadlines given for projects and assignment submission. Through my experience as a data analyst I have demonstrated the eagerness to learn new technologies and tools to aid the in analysing data. Starting out with Excel, for data handling and wrangling and then Tableau for visualization.

I am excited at the potential opportunity to translate my newly developed

skills and experience into gaining informed insights for decision making at Capstone. I want to help keep Capstone Analytics accurate, progressive and optimal in providing technological solutions.

Table of Contents

| | |
|---|-----|
| Professional Background | 1 |
| Table of Contents | 3 |
| Udemy Project Description | 4 |
| The Problem | 5 |
| Design | 6 |
| Findings | 7 |
| Finding 1 | 7 |
| Figure 1: Total Subscriptions by Category | 7 |
| Figure 2: Users by Skill Level | 8 |
| Table 1: Free/Paid Users | 8 |
| Finding 2 | 9 |
| Table 2: Sum Subscribers by Subject | 9 |
| Table 3: Average Subscribers by Subject | 9 |
| Figure 3: Average Subscribers by Course | 110 |
| Finding 3: | 11 |
| Table 4: Average Cost per Level for each Subject | 11 |
| Figure 4: Average Cost per Level for each Subject | 12 |
| Finding 4: | 13 |
| Table 5: Average Rating per Level for each Subject | 13 |
| Figure 5: Average Rating per Level for each subject | 13 |
| Finding 5 | 14 |
| Table 6: Average Content Duration per Subject | 14 |
| Figure 6: Average Content Duration per Subject | 14 |
| Analysis | 15 |
| Conclusions | 16 |
| Appendix | 17 |



Udemy Project Description

Provided with four different set of unclean data generated from Udemy's course review, in comma separated values (.csv) format. I was to employ analytic skills to develop insight from the data for the purpose of understanding where opportunities to increase revenue for the organization may lie and also to track the performance of the courses.

I initially loaded and consolidated the data using Google Docs and the following operations were performed: Data Quality Assessment, Data Cleaning and Manipulations and Visualization in a bid to understand the data.

Based on the insight gain during analysis, courses related to Web Development has the highest interaction between subscribers of Udemy and resources provided.

The conclusion was that better investment of resources to provide services related to Web Development skill would be an opportunity to increase company's revenue.

The Problem

The business problem was to discover opportunities to increase revenue for the organization based on recent business and course review.

The entire project has a timeline of seven (7) to at most twelve (12) days.

The data required for such a task would be data containing the following information: the courses being offered, relative number of subscribers, relative cost, level of relevancy based on subscribers ratings, duration, category of subjects for the course, etc. The data should be provided in tabular form for example in excel portable format.

Design

Data quality assessment test was carried out to identify the correctness, completeness, relevancy and reliability of the data provided. And then the cleaning process in order to prepare the data for analysis. The cleaning process involved removing duplicates, removing blank cells, and also finding and replacing inconsistent data.

- 1 The data initially was unsorted incomplete and inconsistent.
- 2 Duplicate rows were deleted, there were no duplicate columns.

For Visualization, Tableau and Excel viz tools. because they provide easy to understand and dynamic visuals and they are easy of usage.

Findings

Interestingly, data has a lot of information but to find the information that meet the business or analysis goal is not easy. While searching through the Udemy course review data given, it was worthy of attempt to group all courses offered in their subject categories. Then by sorting based on categories, operations like population and rating were performed and visualized.

Below shows the transition and the end note insight was that web development has a lot of influence on cost and number of subscription on various difficulty levels.

Findings1 : Categories and Information about Total number of subscriber

Based on the analysis, it was discovered that Web Development courses has the highest number of subscribers, followed by Graphic Design, Business Finance and least Musical Instrument.

Table 1: Free/Paid Users

| <i>Free/Paid</i> | COUNTA of Free/Paid |
|------------------|---------------------|
| Free | 310 |
| Paid | 3362 |
| Grand Total | 3672 |

Figure 1: Total Subscriptions by Category

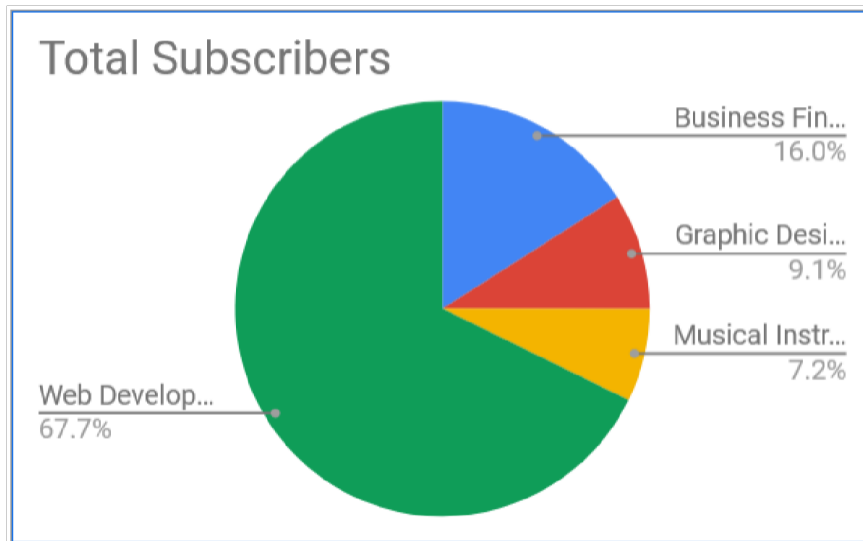
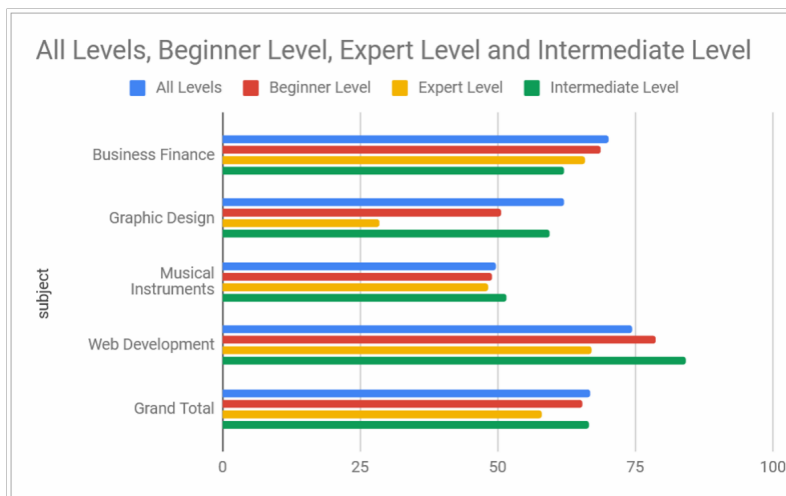


Figure 2 Users by Skill Level



Finding 2 : Web Development has the highest number of subscribers

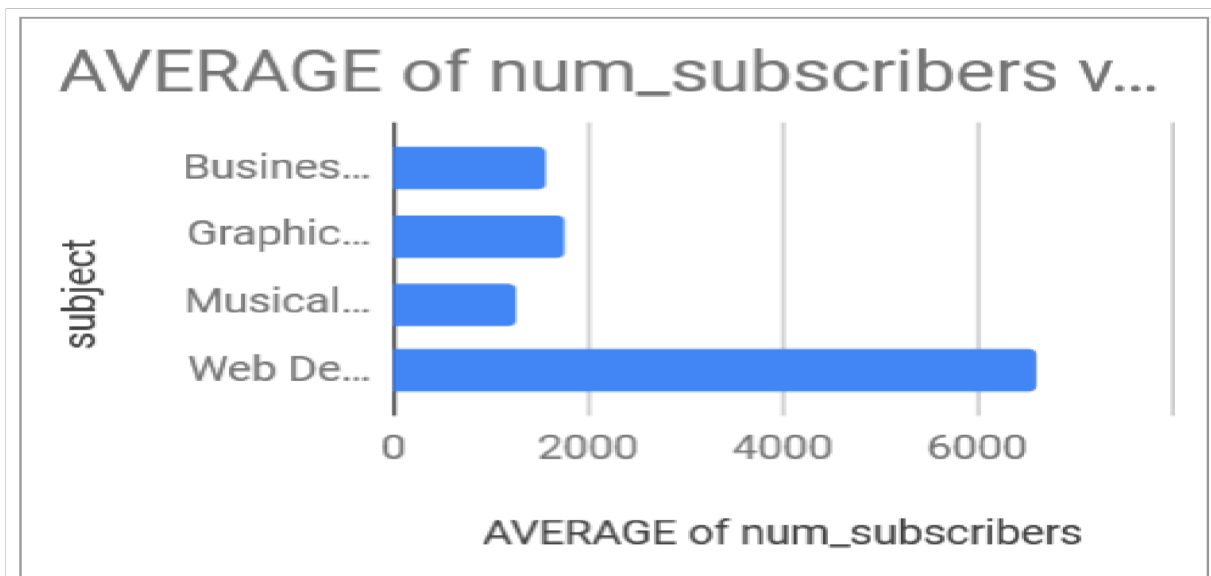
Table 2: Sum Subscribers by Subject

| <i>subject</i> | SUM of num_subscribers |
|---------------------|------------------------|
| Business Finance | 1868711 |
| Graphic Design | 1063148 |
| Musical Instruments | 846689 |
| Web Development | 7937287 |
| Grand Total | 11715835 |

Table 3: Average Subscribers by Subject

| <i>subject</i> | AVERAGE of num_subscribers |
|---------------------|----------------------------|
| Business Finance | 1569.026868 |
| Graphic Design | 1766.026578 |
| Musical Instruments | 1245.130882 |
| Web Development | 6619.922435 |
| Grand Total | 3190.586874 |

Figure 3: Average Subscribers by Course



Findings 3 : On average Web Development Intermediate Level cost most and therefore generate more revenue for the company.

Table 4: Average cost per Level for each subject

| subject | All Levels | Beginner Level | Expert Level | Intermediate Level | Grand Total |
|----------------|------------|----------------|--------------|--------------------|-------------|
| Business | | | | | |
| Finance | 70.20 | 68.73 | 65.80 | 62.01 | 68.69 |
| Graphic Design | 62.12 | 50.68 | 28.57 | 59.41 | 57.89 |
| Musical | | | | | |
| Instruments | 49.58 | 48.98 | 48.33 | 51.60 | 49.56 |
| Web | | | | | |
| Development | 74.55 | 78.83 | 67.14 | 84.25 | 77.04 |
| Grand Total | 66.75 | 65.31 | 58.02 | 66.64 | 66.10 |

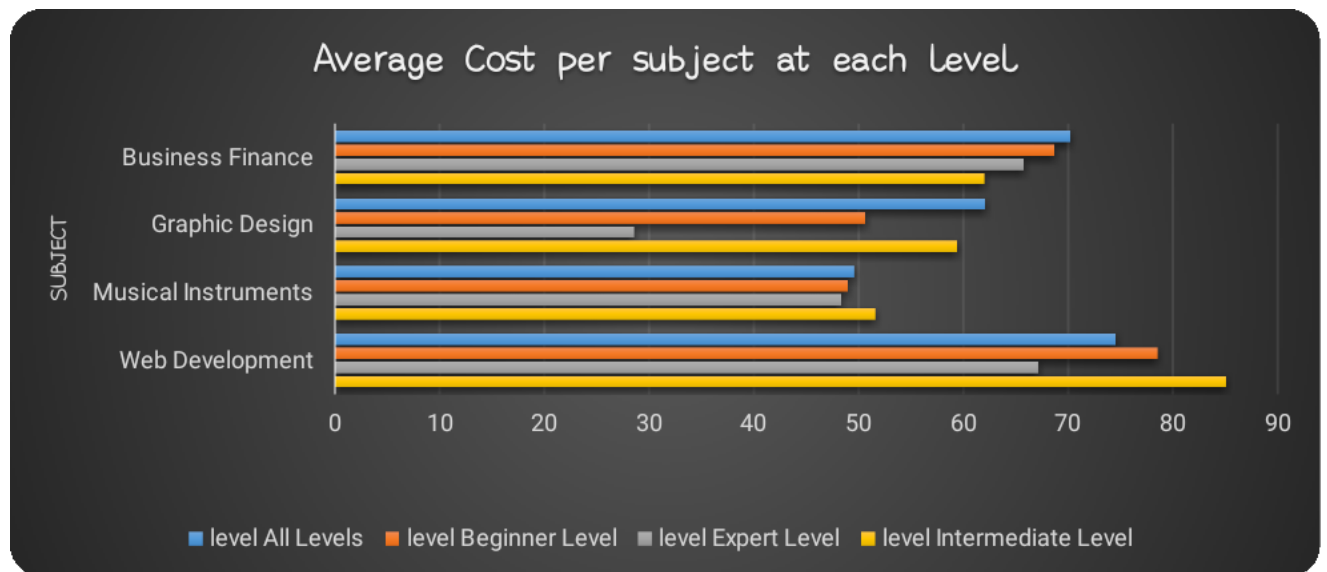


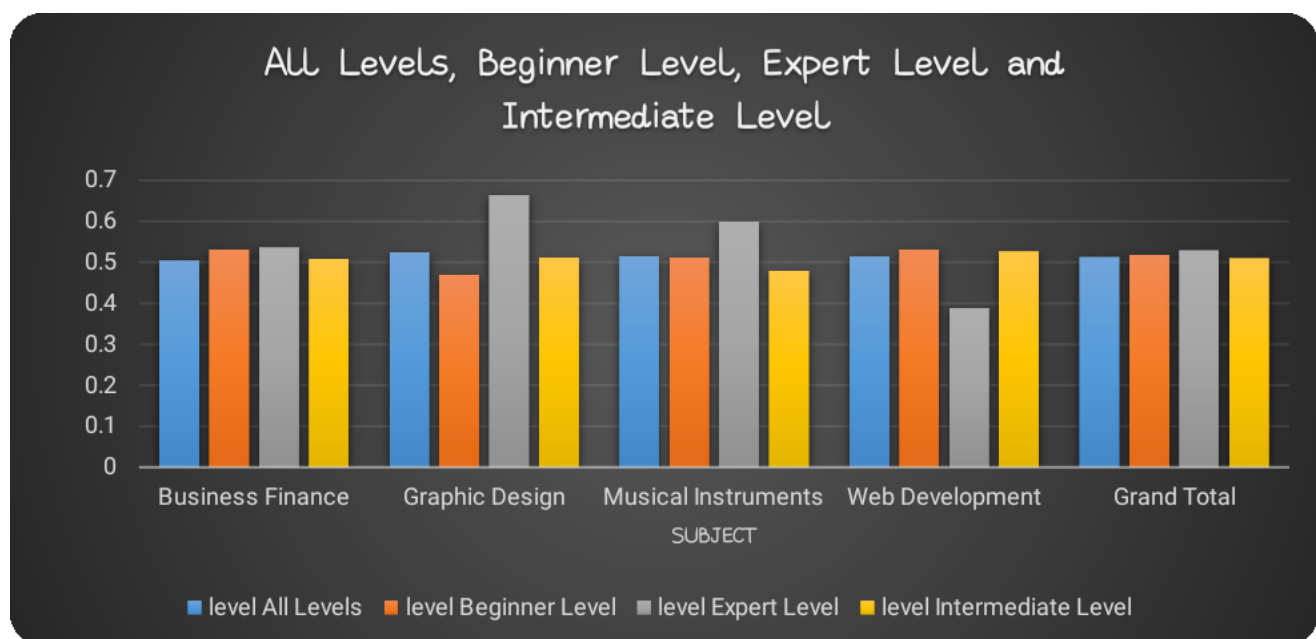
Figure 4: Average cost per level for each subject

Findings 4 : Graphic Design courses relatively has the highest rating.

Table 5: Average rating per level for each Subject

| <i>AVERAGE of Rating</i> | <i>Level</i> | | | | |
|----------------------------|-------------------|-----------------------|---------------------|---------------------------|--------------------|
| <i>Subject</i> | <i>All Levels</i> | <i>Beginner Level</i> | <i>Expert Level</i> | <i>Intermediate Level</i> | <i>Grand Total</i> |
| <i>Business Finance</i> | 0.50 | 0.53 | 0.54 | 0.51 | 0.52 |
| <i>Graphic Design</i> | 0.53 | 0.47 | 0.66 | 0.51 | 0.51 |
| <i>Musical Instruments</i> | 0.52 | 0.51 | 0.60 | 0.48 | 0.51 |
| <i>Web Development</i> | 0.51 | 0.53 | 0.39 | 0.53 | 0.52 |
| <i>Grand Total</i> | 0.51 | 0.52 | 0.53 | 0.51 | 0.52 |

Figure 5: Average rating per level for each Subject

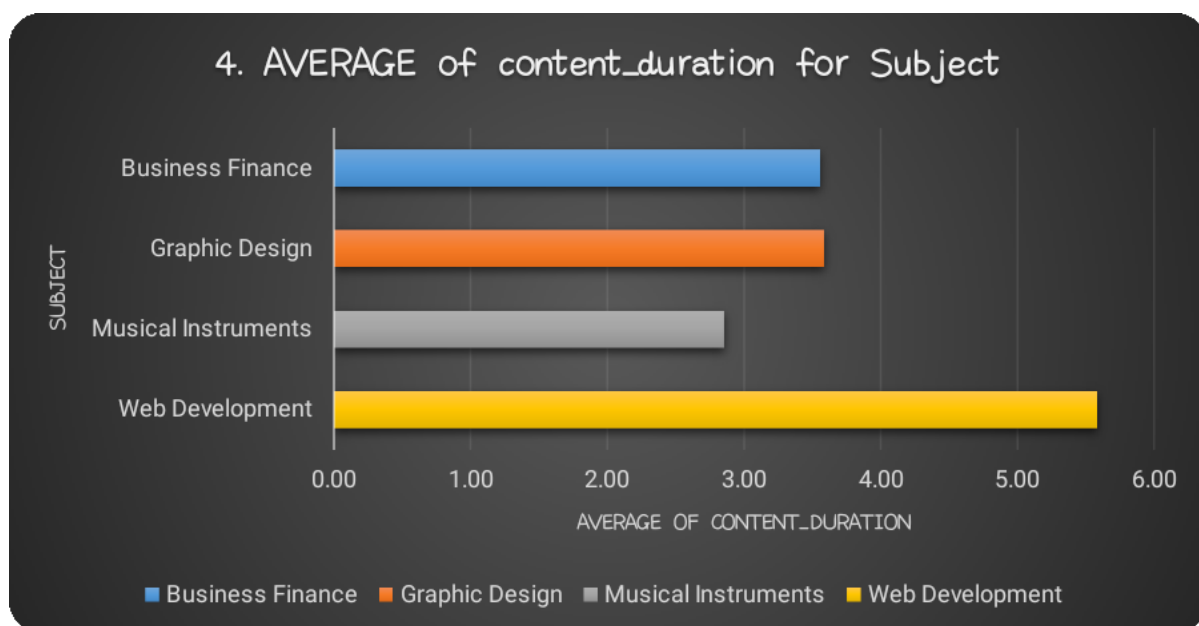


Findings 5 : On Average Web Development has the highest content duration.

Table 6 : Average content duration for each subject

| <i>subject</i> | <i>Average Content Duration</i> |
|----------------------------|---------------------------------|
| <i>Business Finance</i> | <i>3.56</i> |
| <i>Graphic Design</i> | <i>3.59</i> |
| <i>Musical Instruments</i> | <i>2.85</i> |
| <i>Web Development</i> | <i>5.59</i> |

Figure 6 : Average content duration for each subject



Analysis

The analysis therefore shows Web development category generally is the most sought after because it has the highest number of subscribers, the longest content during, the highest average cost and because it has a lot of resources available.

Concl usions

In a bid to understand the course review and to see if there lies any opportunity to increase revenue investing more in web development skill because it has the highest number of subscribers generates the most income could increase the companies revenue.

Appendix

Please remember that you need to have sharing access enabled when sharing links!

Google Sheets Data Set for Udeny Project

https://docs.google.com/spreadsheets/d/1Eqnc0Jh0udWkpR9wrx2Y10RIokZKslYgkGTqY72_grl0/edit?usp=sharing

Tableau Data Visualization For Udeny Project

[\[Link here\]](#)

Google Sheets Data Set for [chosen] Project

https://docs.google.com/spreadsheets/d/1Eqnc0Jh0udWkpR9wrx2Y10RIokZKslYgkGTqY72_grl0/edit?usp=sharing

Tableau Data Visualization For [chosen] Project

[\[Link here\]](#)