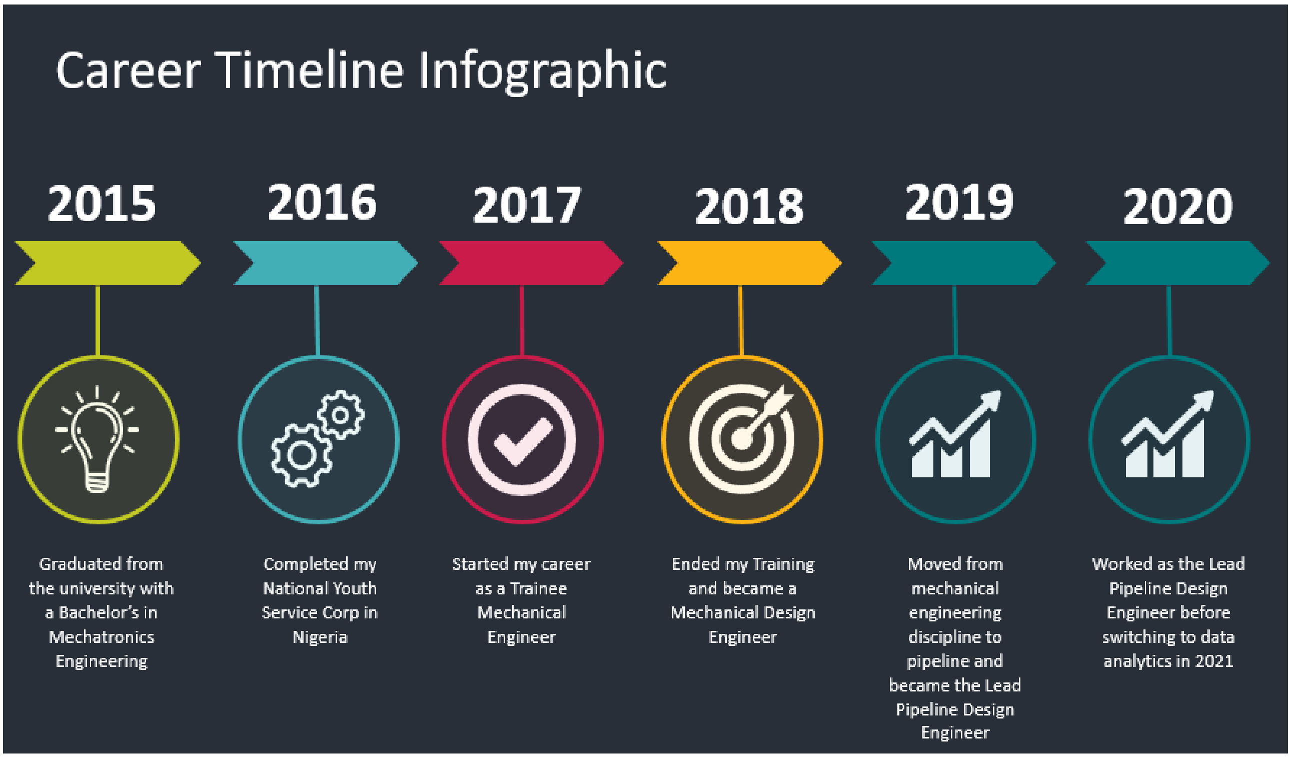
Data Analysis Portfolio

Professional Background

Sam is a result-oriented and self-driven professional looking for job opportunities as a data analyst. He possesses strong technical skills rooted in substantial training in his previous role as a design engineer in the oil and gas industry, He started as a Trainee Mechanical Engineer and rose up the career ladder to being the Lead Pipeline Design Engineer in three years where he led a team of design engineers. He analyzed environmental data that was used to develop equipment drawings and sizing calculations. As someone that has always worked with data, he recognizes the importance and power of data and the valuable insights that can be derived from data. He is passionate about working with large amounts of data, turning the data into information, information into insight and insight into business decisions.

He is proficient in the use of Python, Advanced Excel, MySQL and BI tools such as Tableau for data visualization. He is able to explain the most complicated analysis in simple, straightforward, and actionable terms. He possesses knowledge of machine learning techniques (clustering, decision trees etc.) and advanced statistical techniques (regression, properties of distribution, statistical tests etc.). He is currently growing his career as a data analyst and some of his achievements in the field of data analytics includes executing two data analytics projects and five data science/machine learning projects.

He graduated in July 2015 and holds a Bachelor's in Mechatronics Engineering from the Bells University of Technology. He is an avid biker, a voracious reader and an analyst at heart.



Career Timeline Infographic

2015

Graduated from the university with a Bachelor's in Mechatronics Engineering

2016

Completed my National Youth Service Corp in Nigeria

2017

Started my career as a Trainee Mechanical Engineer

2018

Ended my Training and became a Mechanical Design Engineer

2019

Moved from mechanical engineering discipline to pipeline and became the Lead Pipeline Design Engineer

2020

Worked as the Lead Pipeline Design Engineer before switching to data analytics in 2021

Table of Contents

Analysis of Different Udemy Courses and How to Increase Revenue from the Popular Courses and Track Performance of the Courses.

1. Udemy Project Description

You have been asked by your manager, Head of Curriculum at Udemy, to present the data on course revenue, and you have been provided with data on courses from different topics to understand where opportunities to increase revenue may lie, and track the performance of courses. Your manager has suggested encouraging Web Development courses to charge more because she believes that these are the most popular courses. She needs to send a report to the CEO in the next three weeks on how they will increase their next quarterly earnings.

The purpose of this report is to address the root cause of why the company is not generating enough revenue from its various courses and recommend solutions to improve on its bestselling courses while also uncovering other opportunities within its services (other courses) that it can further explore to achieve an overall robust increase in its next quarterly earnings (revenue).

2. What is the Business Problem?

The business seeks to improve on its bestselling point (courses) while also uncovering other opportunities within its services (other courses) that it can further explore to achieve an overall robust increase in its next quarterly earnings (revenue).

2.1. How long do you have to work on this project?

The manager has three weeks until she reports to the CEO, however, unlike the Manager, the Data analyst has two weeks maximum to work on this project. The data analyst will use the remaining one-week interval to make the presentation of his findings and solution to the Manager. The one-week span that is left out for the Manager should allow the Manager to review and understand what was done and to better be grounded on the findings before her onward presentation to the CEO in the third week.

2.2. What data should be collected to understand this problem? How should it be presented?

The following data needs to be collected to understand the problem.

List of all courses/curricula at Udemy

- List of all Udemy courses.
- List of bestselling courses at Udemy relative to all other courses at Udemy.
- List of other courses authored by the author of the bestselling courses at Udemy.
- List of all Udemy subscribers.
- List of the subscribers of the bestselling courses at Udemy relative to all other courses.
- The geographical location of the subscribers of the bestselling courses at Udemy relative to all other courses.
- Organization, work, the age range, and gender of the subscriber of the

bestselling courses relative to all other courses at Udemy.

- Subscription amount/fee/rate/range of the bestselling courses relative to all other courses at Udemy.
- Medium of purchase of the bestselling courses at Udemy relative to all other courses.
- Time of purchase/subscription of bestselling courses relative to all other courses.
- Duration of training/tutorials of the bestselling courses at Udemy.
- List of the educational bodies that Udemy provides with premium services.
- Data of how much is spent on putting the courses together.
- What are the most used keyword search?
- Data of how much is spent on putting the courses together

The data should be presented in tabular format.

2.3. What questions would you ask to better understand the business problem?

- Has the company recorded a drop in the Company's revenue in recent times?
- What was the company's last quarterly revenue?
- What is the company's next quarterly revenue target?
- What is the current company expenses?
- Does the company have sought-after/ in-demand courses/curriculum not currently available at Udemy?
- Is there any promotional offers list attached to the courses, particularly the bestselling courses?
- What is the company's current database strength

3. Data Design

There are several factors to consider when analyzing a dataset to find insight. For this project, some of the steps that was taken to clean the data are listed below:

- Removed duplicates
- Removed blank cells
- Ensured the headers are properly named
- Used Find and Replace to ensure the web development subjects are concise with the other subjects.

Some of the visualization tools that was used to clean and visualize the data are Excel and Tableau. The reason the visualization was done with Tableau is because Tableau is a powerful visualization tool and can be shared across the world and between work colleagues if needed, it is the most widely adopted visualization tool across companies worldwide.

4. Findings

After cleaning and visualizing the data, two columns were created to indicate courses that are free or paid courses and free beginner courses or paid beginner courses.

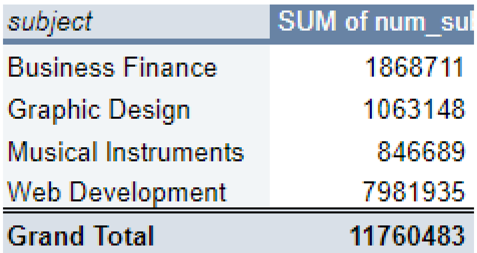
I also looked at the top 20 most subscribed courses that generated the most revenue for Udemy, the level of the course, if the courses were free or paid courses, if they were free beginner courses, the content duration, the date the course was introduced and the ratings of the various courses.

For this project, I performed my analysis using the following criteria:

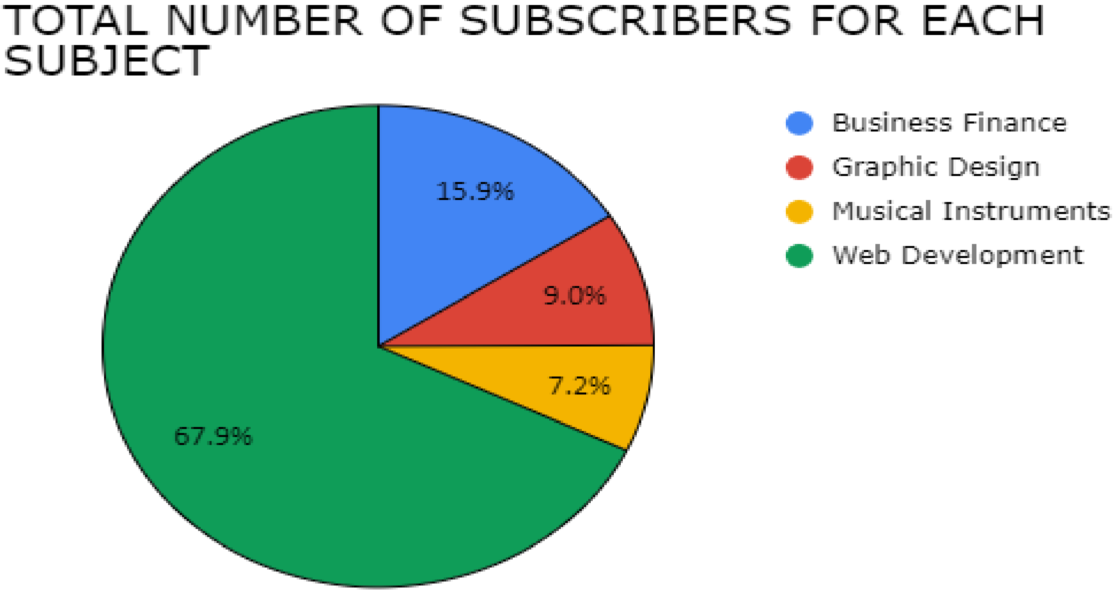
- Total number of subscribers for each subject.
- Average number of subscribers for each subject.
- Average cost per subject at each level.
- Average content duration for each subject.
- Average rating per subject for each level.
- Average number of lectures per subject.
- Average number of reviews per subject.

4.1. Total number of subscribers for each subject

Looking at the total number of subscribers for each subject criteria, the total number of subscribers for web development is almost 75% of the total subscribers among the four courses.

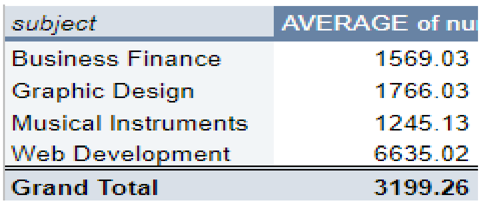


subject	SUM of num_su
Business Finance	1868711
Graphic Design	1063148
Musical Instruments	846689
Web Development	7981935
Grand Total	**11760483**

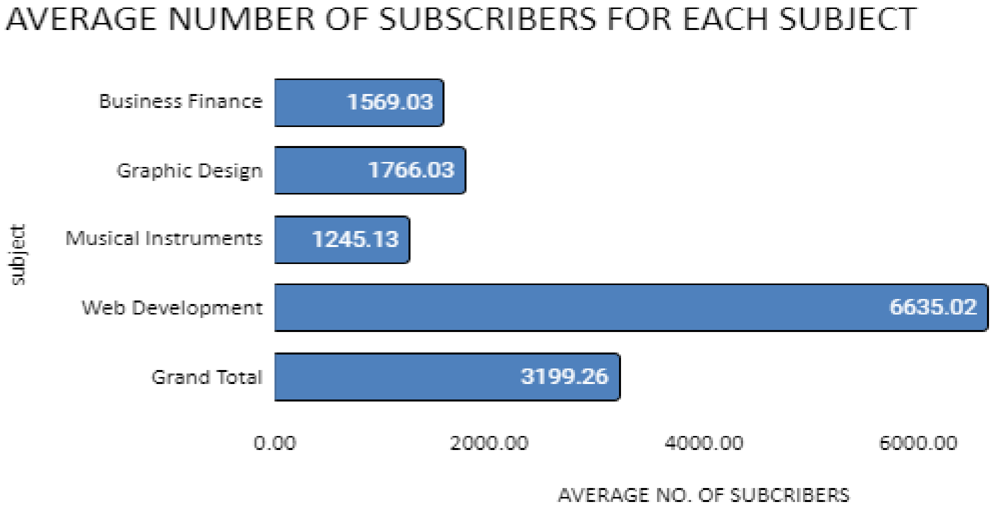


4.2. Average number of subscribers for each subject.

Web development has the highest average number of subscribers for each subject among the four courses.

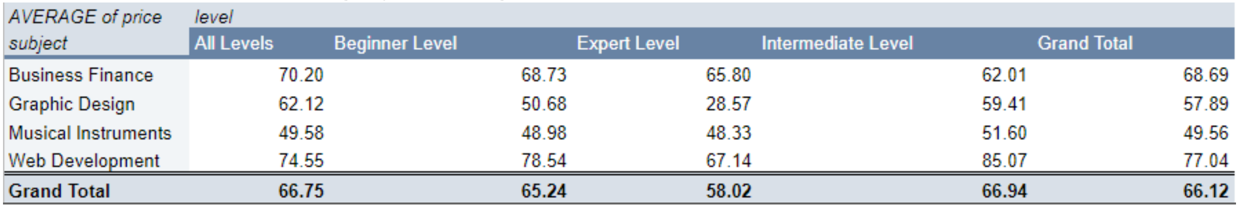


subject	AVERAGE of nu
Business Finance	1569.03
Graphic Design	1766.03
Musical Instruments	1245.13
Web Development	6635.02
Grand Total	**3199.26**

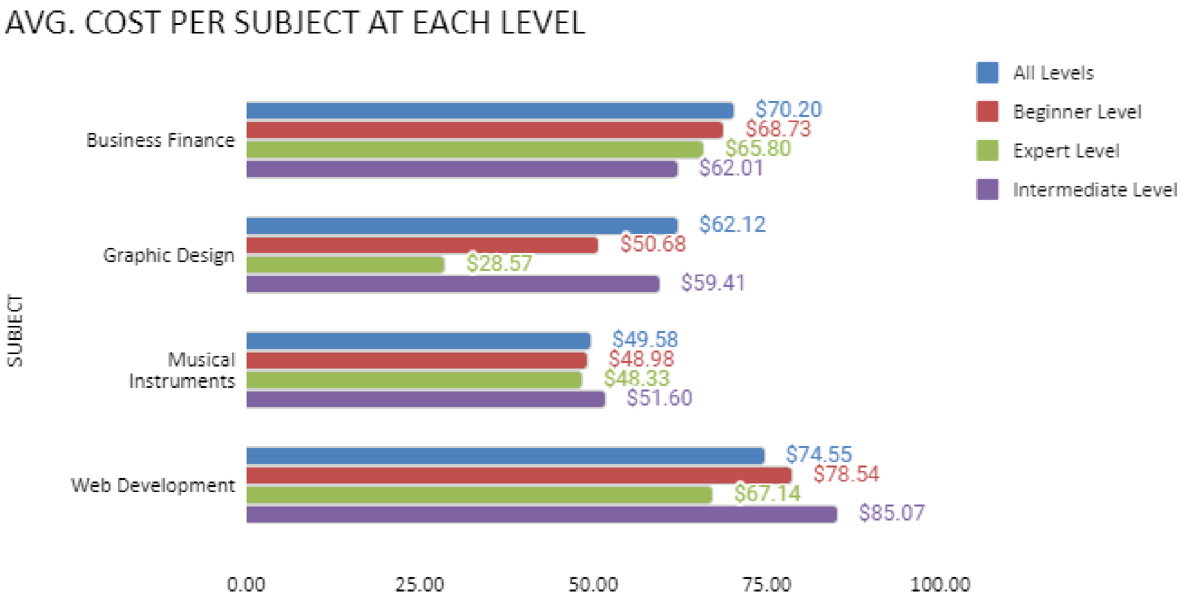


4.3. Average cost per subject at each level.

The average cost per subject at each level is higher for web development courses, followed by business finance courses, graphic design and musical instruments courses.

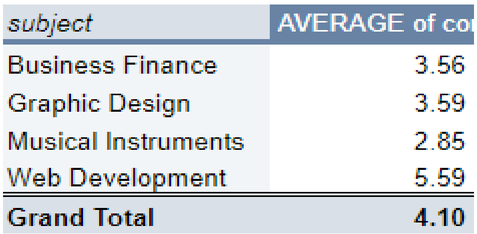


AVERAGE of price	*level*				
subject	All Levels	Beginner Level	Expert Level	Intermediate Level	Grand Total
Business Finance	70.20	68.73	65.80	62.01	68.69
Graphic Design	62.12	50.68	28.57	59.41	57.89
Musical Instruments	49.58	48.98	48.33	51.60	49.56
Web Development	74.55	78.54	67.14	85.07	77.04
Grand Total	**66.75**	**65.24**	**58.02**	**66.94**	**66.12**

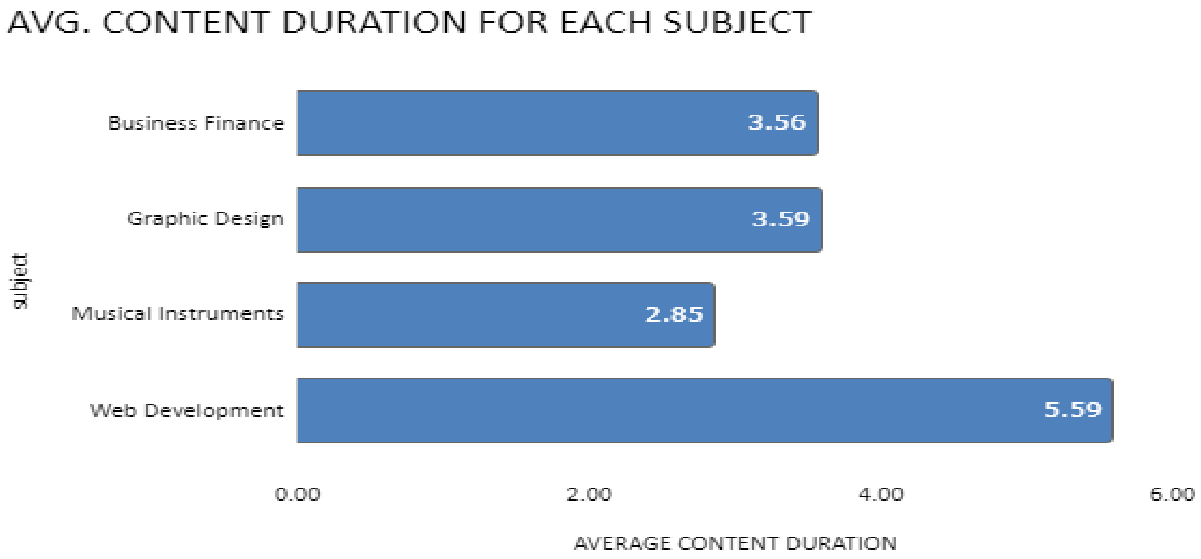


4.4. Average content duration for each subject.

The average content duration for each subject is higher for web development courses, then graphic design courses, business finance courses and musical instrument courses.

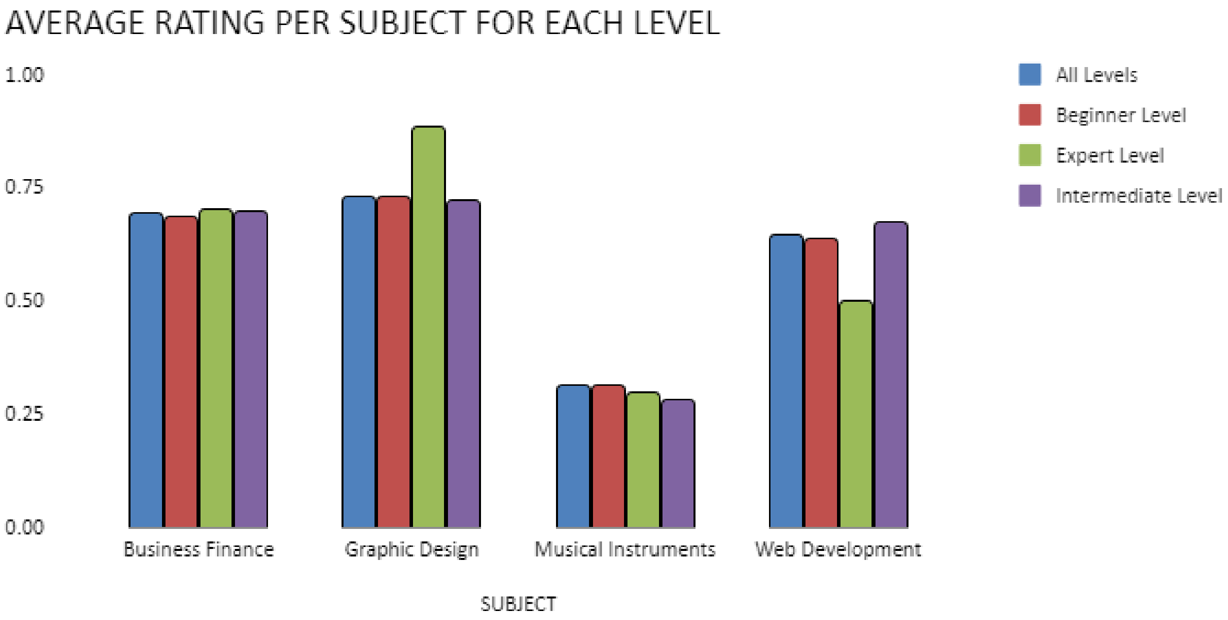


subject	AVERAGE of co
Business Finance	3.56
Graphic Design	3.59
Musical Instruments	2.85
Web Development	5.59
Grand Total	**4.10**



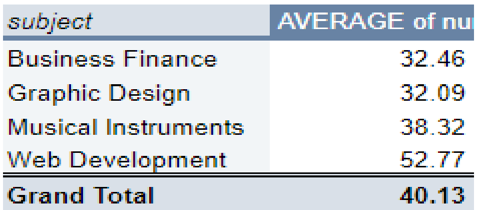
4.5. Average rating per subject for each level.

The average rating per subject for each level is higher for graphics design courses, then business finance courses, web development courses and musical instruments courses.

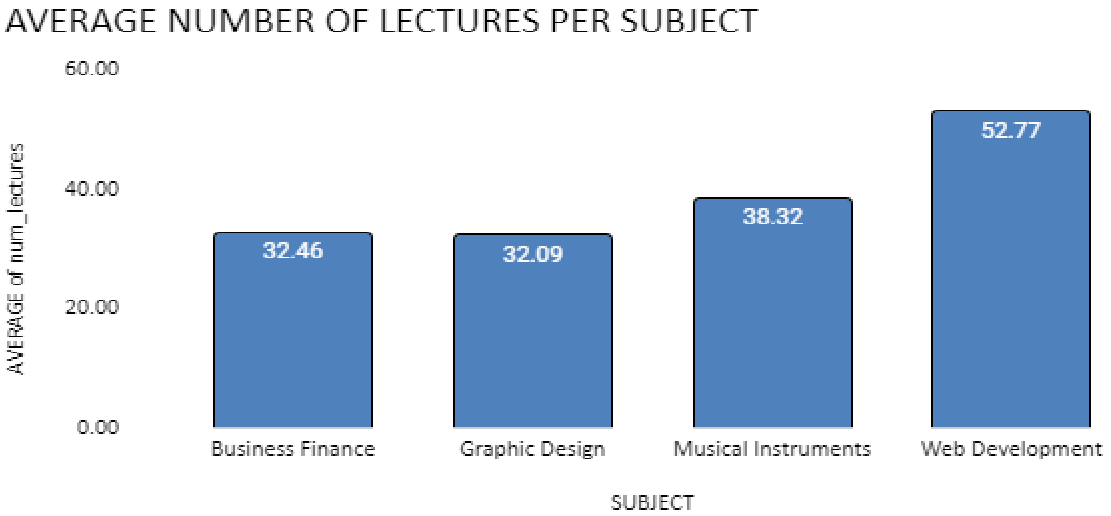


4.6. Average number of lectures per subject.

The average number of lectures per subject is higher for web development courses, then musical instruments courses, business finance courses and graphic design courses.

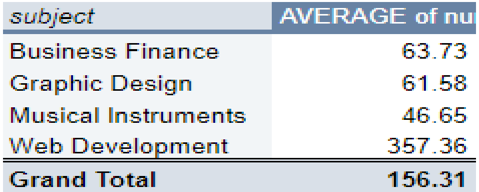


subject	AVERAGE of nu
Business Finance	32.46
Graphic Design	32.09
Musical Instruments	38.32
Web Development	52.77
Grand Total	**40.13**

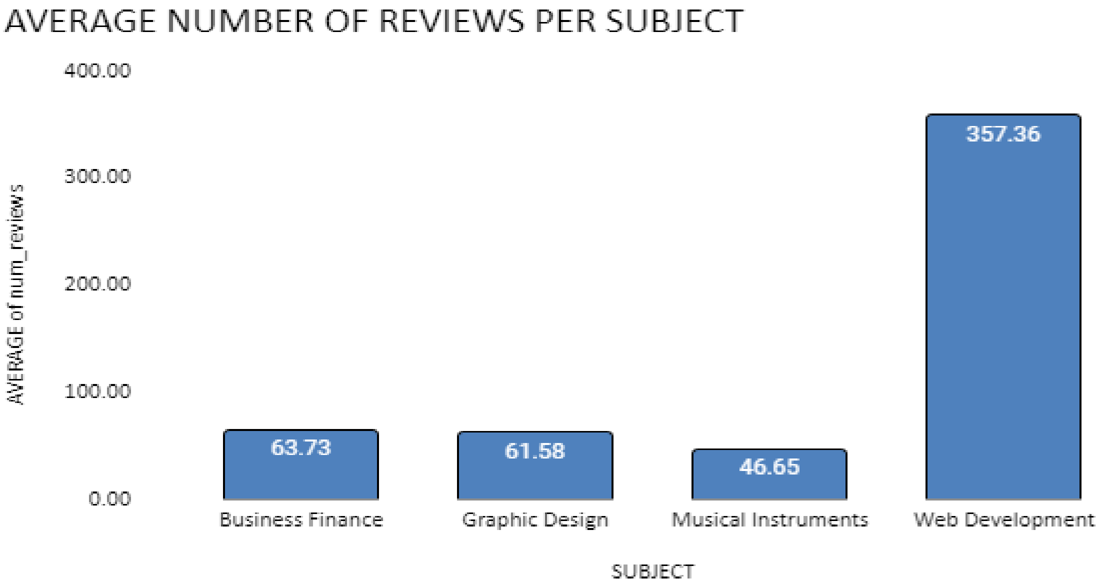


4.7. Average number of reviews per subject.

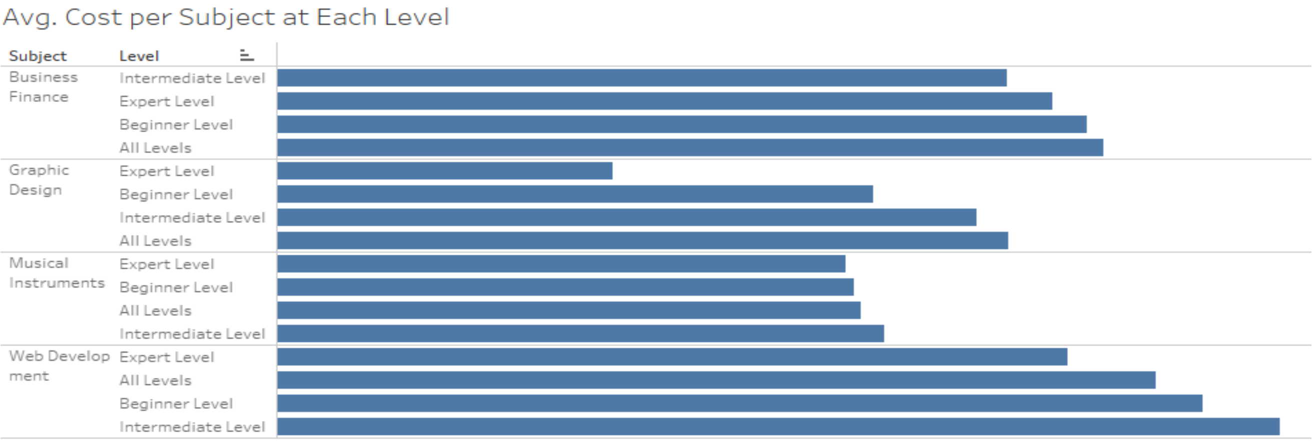
The average number of reviews per subject is higher for web development courses, then business finance courses, graphic design courses and musical instruments courses.

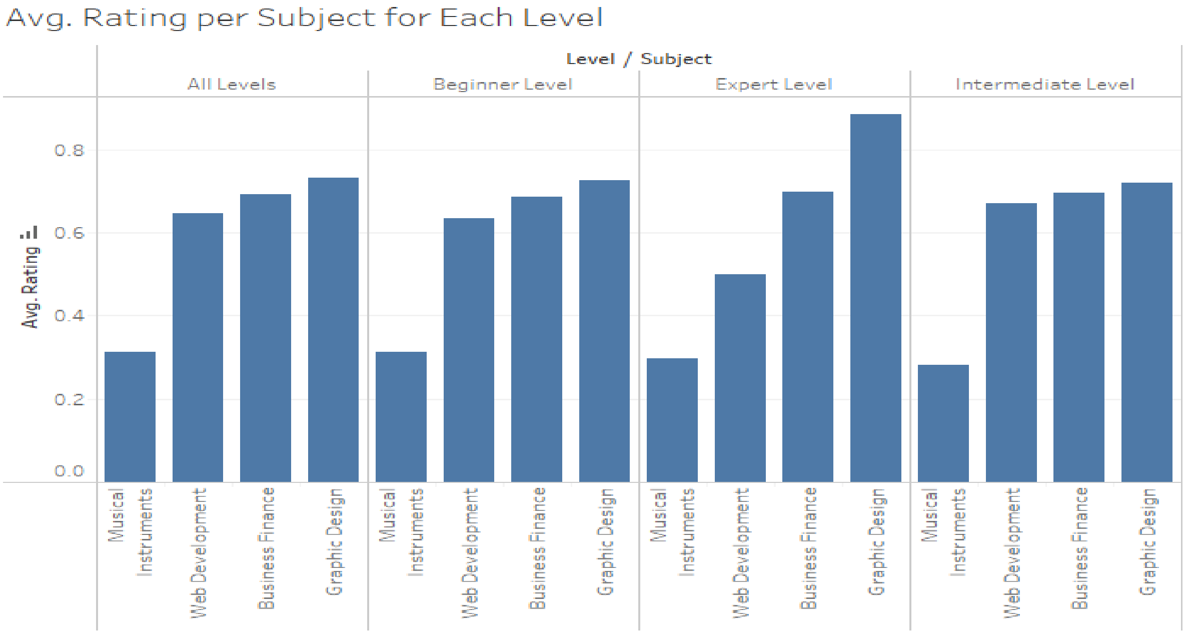


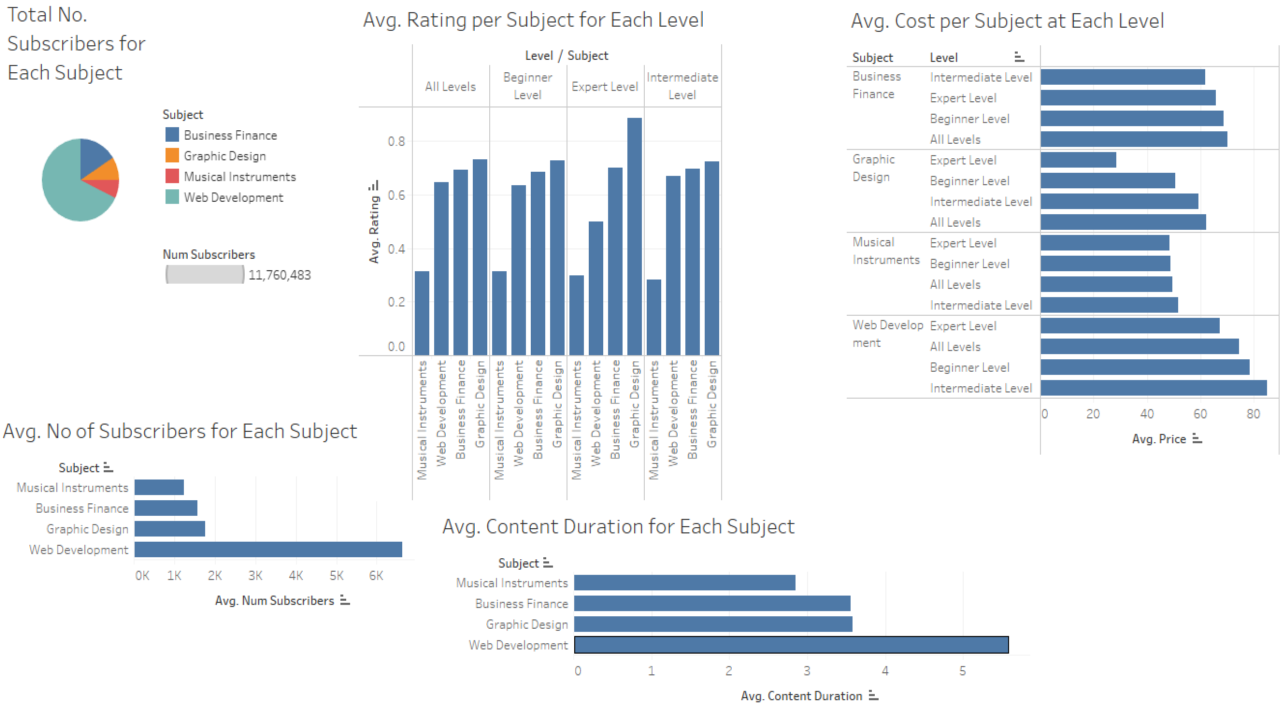
subject	AVERAGE of nu
Business Finance	63.73
Graphic Design	61.58
Musical Instruments	46.65
Web Development	357.36
Grand Total	**156.31**



4.8. Tableau Visualization







5. Data Analysis

Using the findings explained in Section 3 and applying the 5 Whys method to identify where opportunities may lie to increase revenue and track performance of courses, after digging into the data, I was able to obtain identify the root cause of why the company is not generating enough revenue from the various courses, they are listed below:

1. Web development has the highest number of subscribers, the average cost per subject at each level is higher for web development than the other courses but it has the third highest quality rating among the four courses analyzed.
2. The average content duration for each subject is higher for web development than the other courses but it has a lower average rating among the four courses.
3. The low web development course rating compared to business finance and graphic design courses could be because of a higher subscriber rate of web development courses than the other courses.
4. Graphics design and business finance courses on average have a higher quality rating at all levels than web development but they cost less than web development courses.

6. Conclusions

From the result of the analysis, I can make the following recommendations:

1. Engage the content creators to reduce the number of hours for web development courses since the longer the course, the lower the rating.
2. Increase the cost of beginner, intermediate level web development courses since they are more popular among the subscribers and have a much higher rating than the expert level web development course.
3. Increase the cost of business finance and graphic design courses because they have much higher rating on average across all the levels than web development and musical instruments courses. This indicates that the quality of business finance and graphic design courses are of a much higher quality and the consumers are satisfied with the contents of the courses.

7. Appendix

Click or Tap on the any of the links below to view the resources:

[1]. Data Analysis Visualization on Udemy Courses

[2]. Google Data Sheet Udemy Courses

Analysis of Covid 19 Data from John Hopkins University

1. Covid-19 Capstone Project Description

Corona viruses are a large family of viruses, which may cause illness in humans. In humans, several corona viruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Severe Acute Respiratory Syndrome (SARS). COVID-19 is an infectious disease; the new virus and disease were unknown before the outbreak began in Wuhan, China. In December 2019, a new respiratory illness began to spread throughout Wuhan, China, a city of 11 million people in Hubei province. The virus, known as COVID-19, quickly infected tens of thousands of people over the ensuing weeks. China imposed major restrictions on travel and work, and by the end of February 2020, cases of COVID-19 had slowed inside the country while spiking in other countries including South Korea, Italy, Iran and all over the world.

In March 2020, the World Health Organization recognized the breakout as a global pandemic — the first since 2009.

The purpose of this report is to address the root cause of how Covid 19 quickly spread across the globe in a short period and what could have been done to prevent it from happening. The data that was analyzed for this project report was last updated on 4th of June 2020. This is a short report on the analysis and insights that was derived from the Covid 19 data visualization.

2. Data Design

There are several factors to consider when analyzing a dataset to find insight. For this project, some of the steps that was taken to clean the data are listed below:

- Removed duplicates
- Removed blank cells
- Ensured the headers are properly named

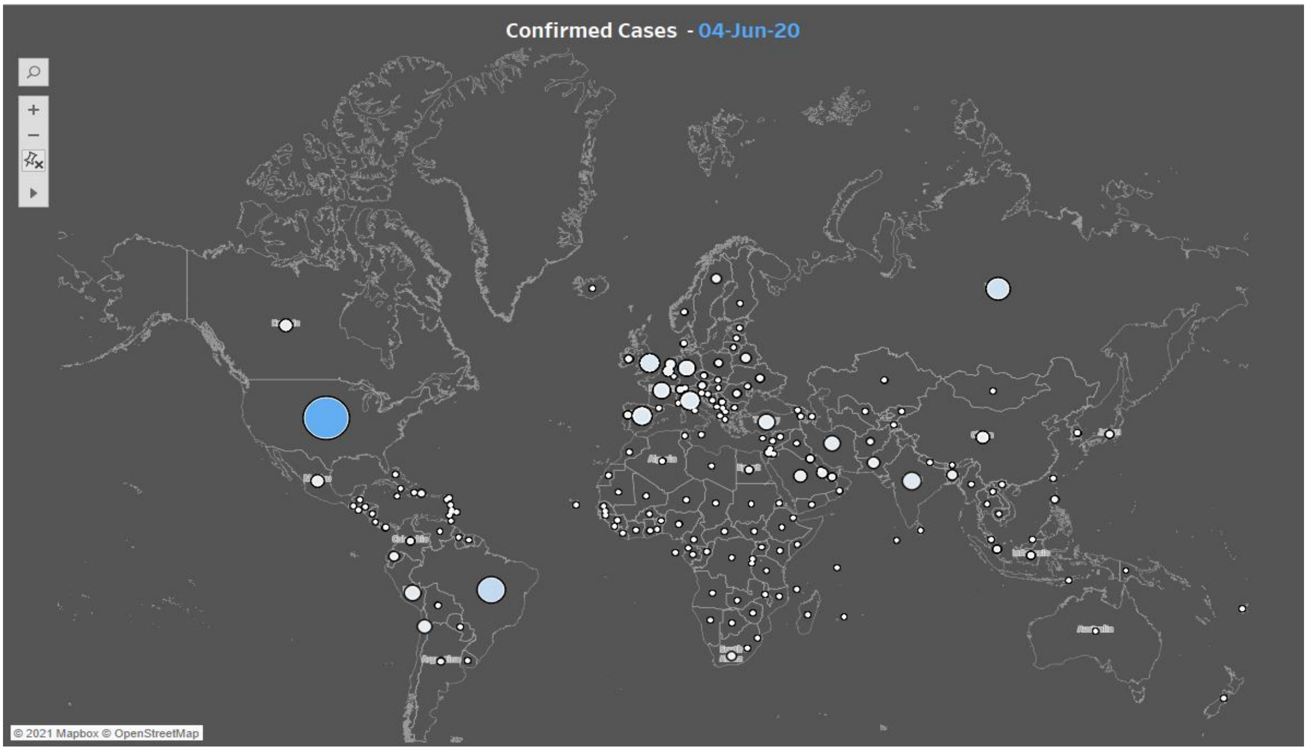
The visualization tool that was used to clean and visualize the data is Tableau. The reason the visualization was done with Tableau is because Tableau is a powerful visualization tool and can be shared across the world and between work colleagues if needed, it is the most widely adopted visualization tool across companies worldwide.

3. Findings

This section reports on findings in this exploratory stage of the project. The attachments below demonstrates, for the selected subset of countries, the diversity at the outset of the pandemic and illustrates changes over time. This analysis will help to visualize the top 10 affected countries and total number of cases recorded after Covid-19 was detected in wuhan china and subsequent spread across the globe.

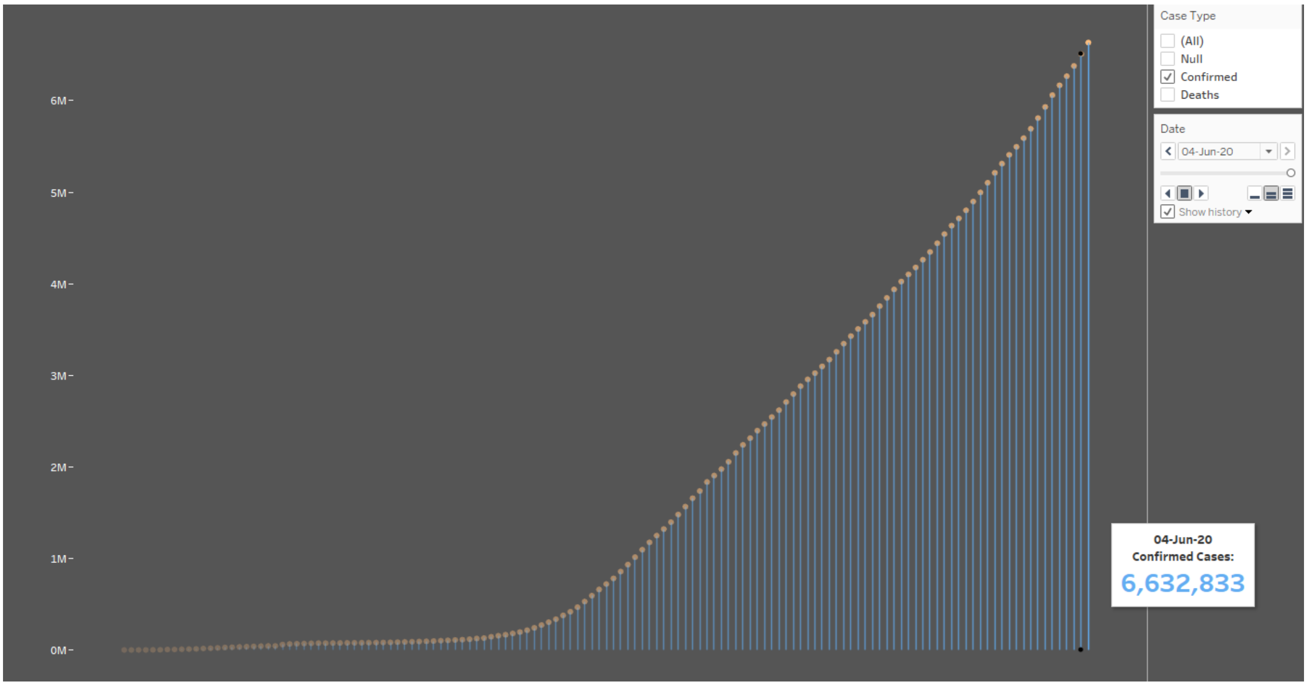
3.1. Map of Countries Affected

As at 4th June 2020, Looking at the map of affected countries, the Covid-19 virus had spread across the globe and the number of recorded confirmed cases as at 4th of June 2020 was 6,632,833.



3.2. Line Chart of Total Number of Confirmed Cases

The attachment below is a line chart of the total number of confirmed cases worldwide as at 4th of June 2020.



Case Type

□ (All)
□ Null
☑ Confirmed
□ Deaths

Date

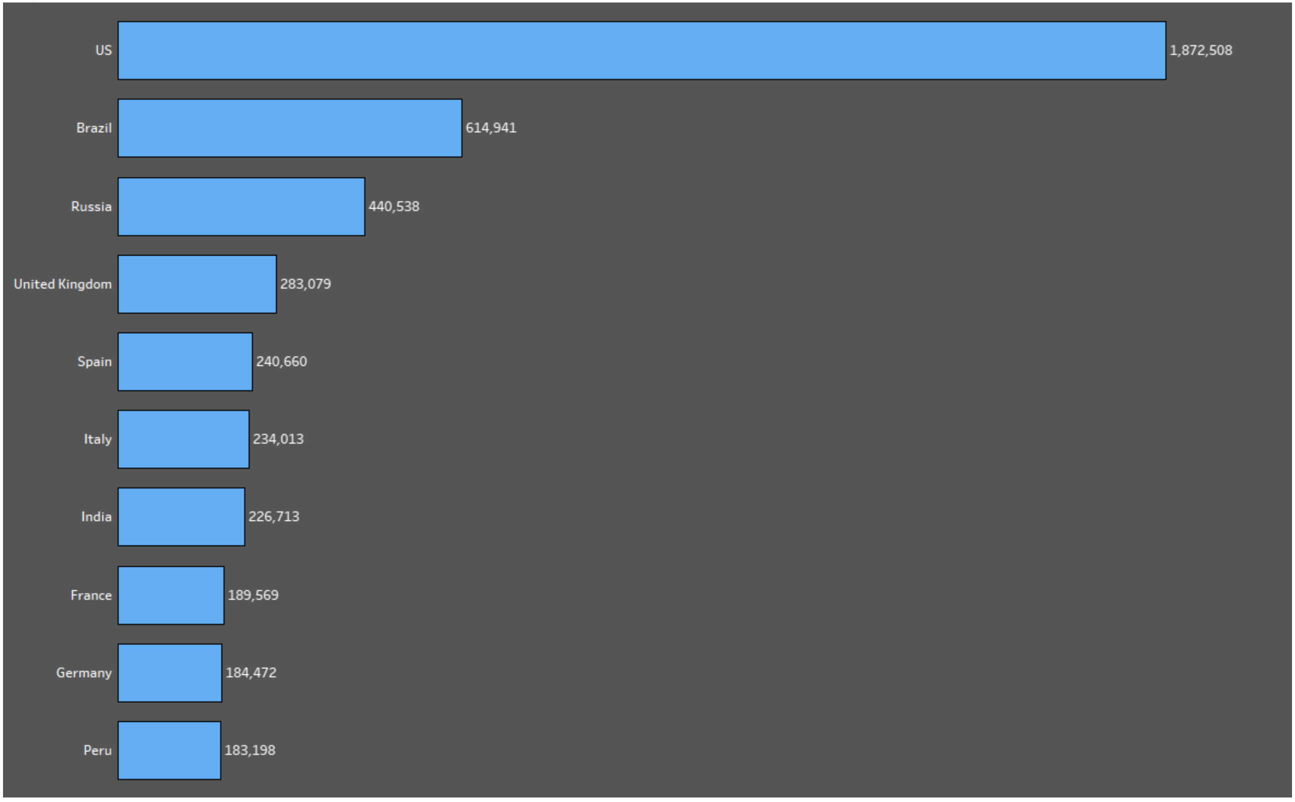
04-Jun-20

☑ Show history

04-Jun-20
Confirmed Cases:
6,632,833

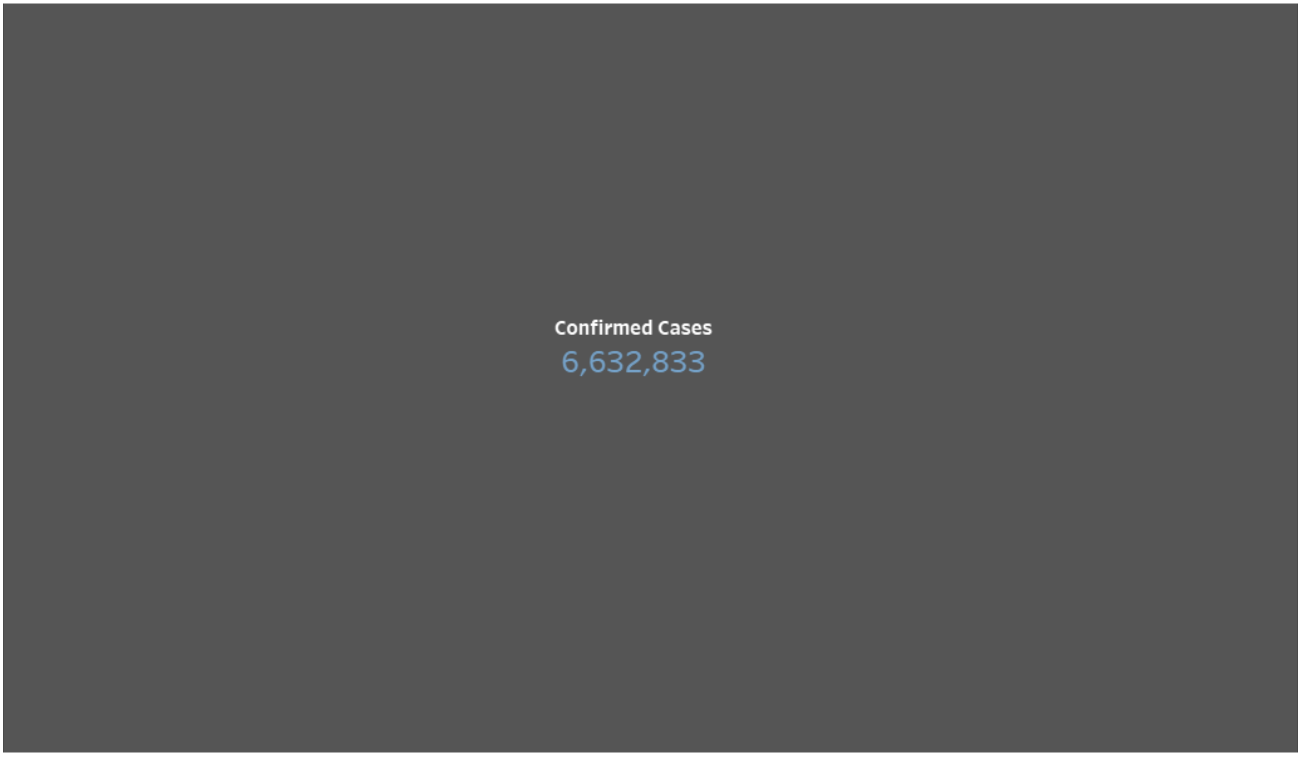
3.3. Bar Chart of Top 10 Countries with Highest Covid-19 Related Cases

As at 4th of June 2020 when this dataset was last updated, the top 10 countries with the highest number of confirmed cases worldwide can be seen below in the bar chart.



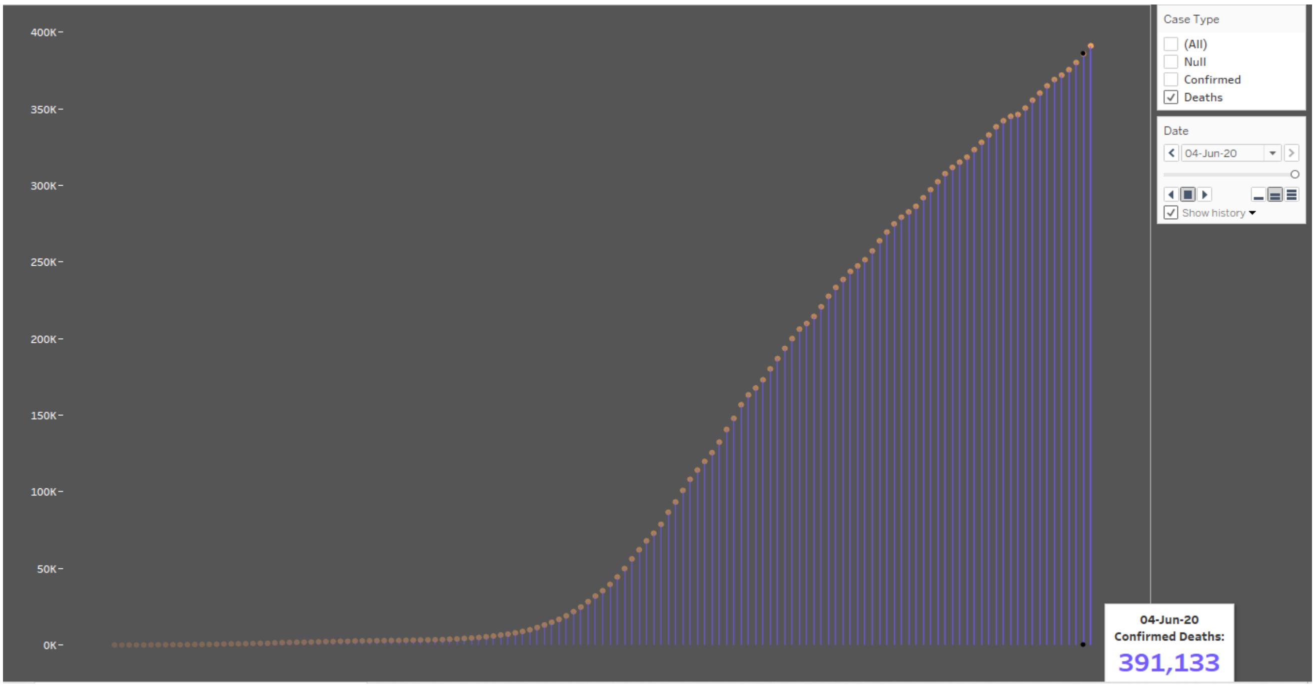
3.4. Confirmed Cases Worldwide

The total number of recorded confirmed cases as at 4th of June 2020 was 6,632,833.



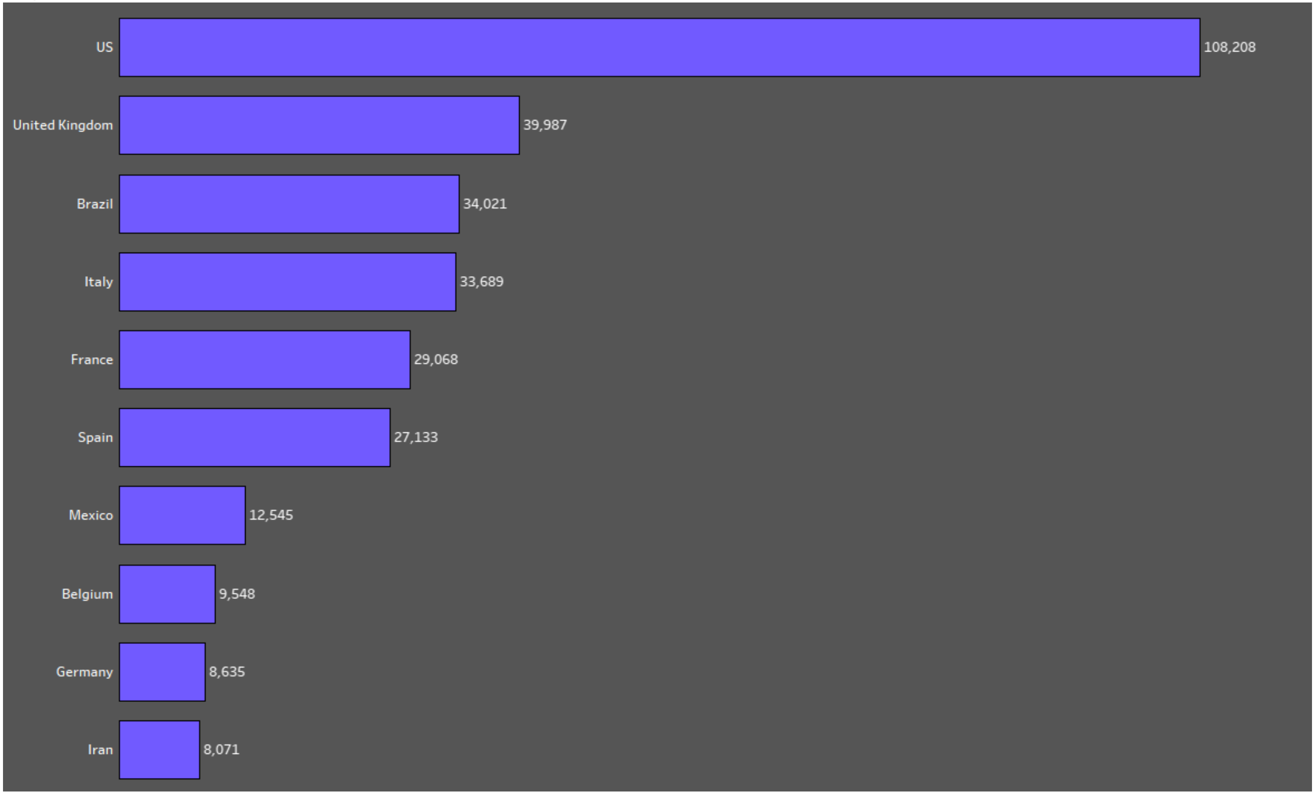
3.5. Line Chart of Countries with Confirmed Deaths

The attachment below is a line chart of the total number of confirmed deaths worldwide as at 4th of June 2020.



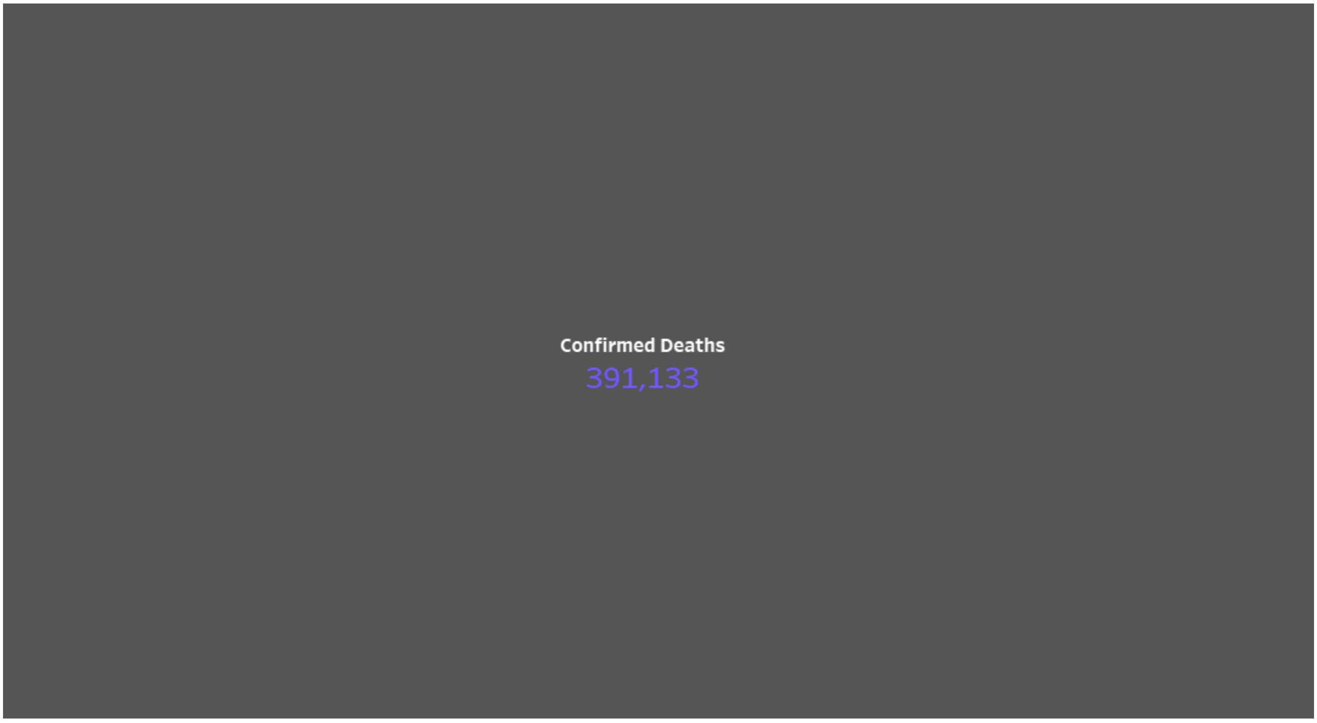
3.6. Bar Chart of Top 10 Countries With the Highest Covid 19 Related Deaths

As at 4th of June 2020 when this dataset was last updated, the top 10 countries with the highest number of Covid-19 related deaths worldwide can be seen below in the bar chart.



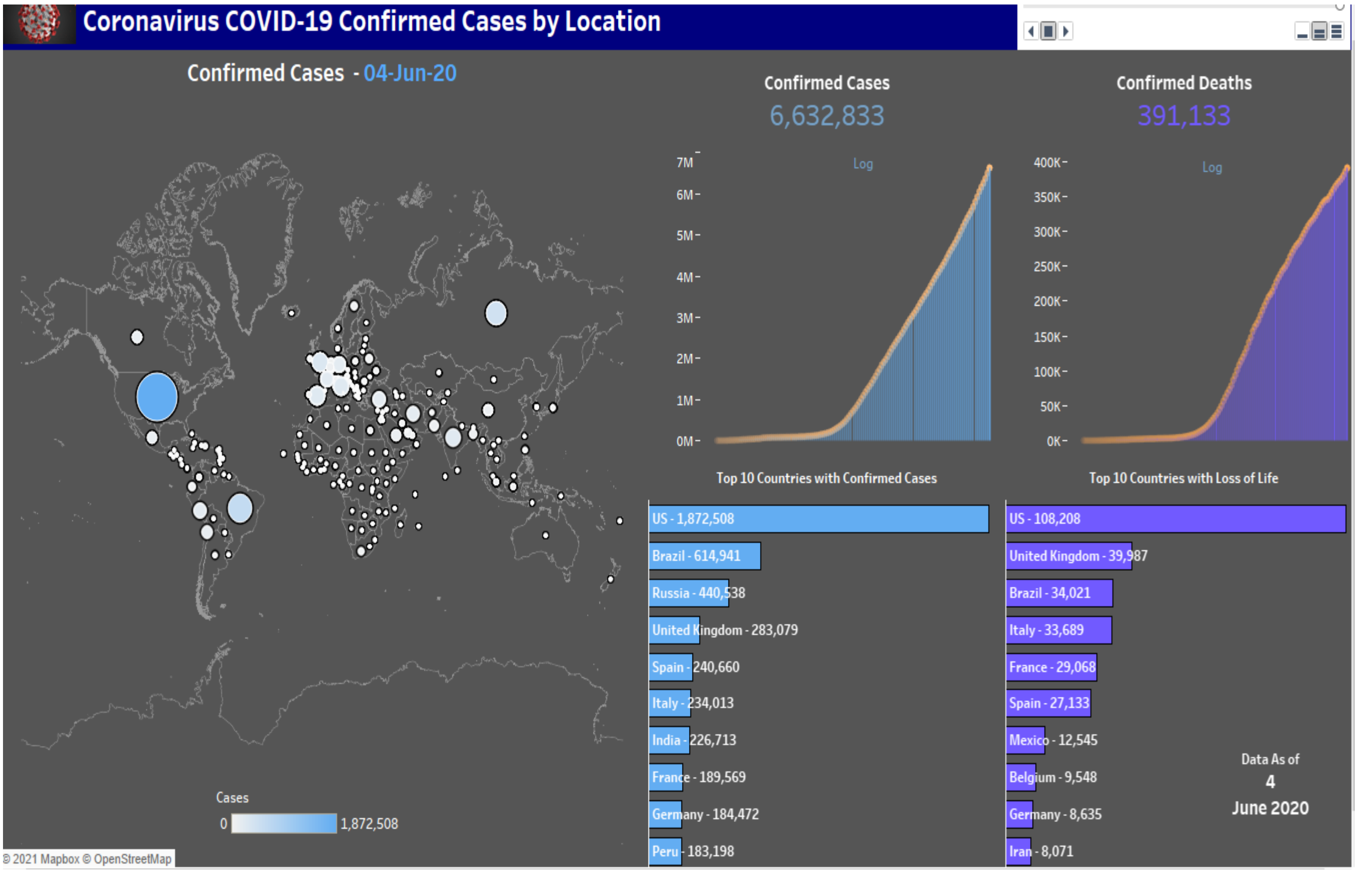
3.7. Confirmed Deaths Worldwide

As at 4th of June 2020 when this dataset was last updated, the total number of confirmed deaths worldwide was 391,133.



Confirmed Deaths
391,133

3.8. Dashboard of Countries Affected and the Total Number of Confirmed Cases and Deaths Worldwide



Coronavirus COVID-19 Confirmed Cases by Location

Confirmed Cases - 04-Jun-20

Confirmed Cases
6,632,833

Confirmed Deaths
391,133

Top 10 Countries with Confirmed Cases

- US - 1,872,508
- Brazil - 614,941
- Russia - 440,538
- United Kingdom - 283,079
- Spain - 240,660
- Italy - 234,013
- India - 226,713
- France - 189,569
- Germany - 184,472
- Peru - 183,198

Top 10 Countries with Loss of Life

- US - 108,208
- United Kingdom - 39,987
- Brazil - 34,021
- Italy - 33,689
- France - 29,068
- Spain - 27,133
- Mexico - 12,545
- Belgium - 9,548
- Germany - 8,635
- Iran - 8,071

Data As of
4
June 2020

© 2021 Mapbox © OpenStreetMap

4. Data Analysis

Using the findings explained in Section 3 and applying the 5 Whys method to identify why the virus spread so quickly around the world, I was able to obtain identify the root cause of why Covid 19 was able to spread worldwide at a rapid pace, they are listed below:

1. Health care services around the world were not prepared for a virus of this magnitude.
2. The country where the virus originated from did not contain the virus early enough.
3. Precautionary measures were not implemented early enough and that hastened the spread of the virus.
4. The WHO and virology specialists under estimated how fast the virus could spread and how infectious covid-19 is.

5. Conclusions

From the result of the analysis, we can make the following conclusion:

- Technologically advanced countries were the most affected by the Covid-19 virus.
- Third world countries were least affected by the Covid-19 virus.
- The US had the highest Covid-19 confirmed cases and related deaths worldwide.

6. Appendix

Click or Tap on the links below to view the dataset and visualization:
[1]. Covid-19 Tableau Covid-19 Data Dashboard Visualization
[2]. Covid-19 Dataset