

# **Data Analysis Portfolio**

Tech enthusiast interested in solving problems. Proficient in SQL, Python, Tableau and PowerBI. An adaptable person who likes problem solving. A thrive to create something beautiful yet powerful is what keeps me awake at night.

## **Professional Background**

Masters student majored in Computer Application who is interested in working with Data to solve problems. Experience in working with a company (National Instruments) as a Software Testing Intern while pursuing the Bachelors. Also worked with a number of teams for doing projects for college events and collaborated with people across social media to perform some fun project exercises.

# Table of Contents

<b>Professional Background</b>	<b>1</b>
<b>Table of Contents</b>	<b>2</b>
<b>Udemy Project Description</b>	<b>2</b>
<b>The Problem</b>	<b>4</b>
<b>Design</b>	<b>5</b>
<b>Findings</b>	<b>6</b>
<b>Conclusions</b>	<b>9</b>
<b>ICC Test Batting Project Description</b>	<b>10</b>
<b>Data Design</b>	<b>11</b>
<b>Analysis</b>	<b>11</b>
<b>Result</b>	<b>12</b>

# Udemy Project Description

The Udemy dataset which contains the course URL, Rating, Prices, Subject, Date Published, duration, price of the course etc...

With the data's that are available to us, we can use it to find some insights which can be used in the future to drive some business solutions.

We can find answers to specific questions like Top 10 courses people purchase, the area of expertise or the level of course the customer needs which can then be utilized to add more number of courses.

Before proceeding, there were a few number of incomplete data, there were some typo in some of the course name etc... These kind of small errors have been found out and corrected before proceeding with the analysis.

# The Problem

The Udemy dataset can be used to answer some of the business problems like which are the top courses customers purchases. Knowing this can help the business understand the taste of the customers of courses and which can then be used to add more courses similar to it or we can even recommend courses to other users which can help in giving the users a better experience while trying to purchase courses from Udemy.

About 3, 4 days have been spent on cleaning the data, making sure the data is consistent, analysing the pattern and finding useful insight and making readable visuals.

# Design

The dataset contains only a few missing values and some of the values have some spelling mistake associated with them. The missing values rows are deleted and those having spelling mistakes are found out and corrected it manually.

Once, the cleaning of the data was completed, the next step was to analyse the data and then according to the need, it is then shown visually. For visualization, Tableau and Excel is being used. As most of the visuals can be easily shown using Excel, most of the charts are in Excel.

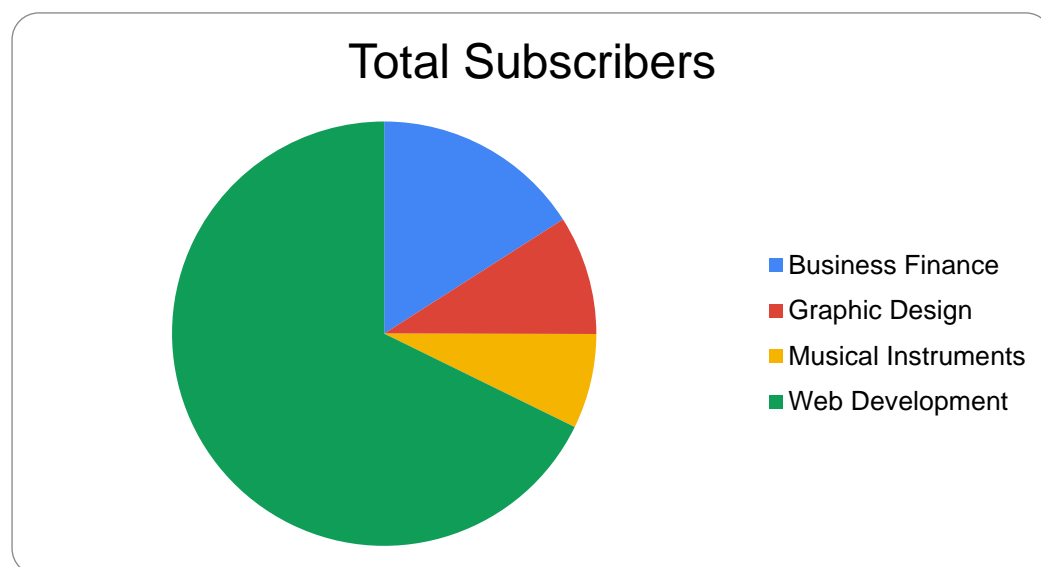
# Findings

The 20 Most Popular Course – can be found out which on analysing, we can find the subject in which they come under, the price of the course which can be used to give recommendation since most people prefer those particular subjects, the courses which comes in those prices so that the company would benefit if more money would be spend on those type of courses in the future.

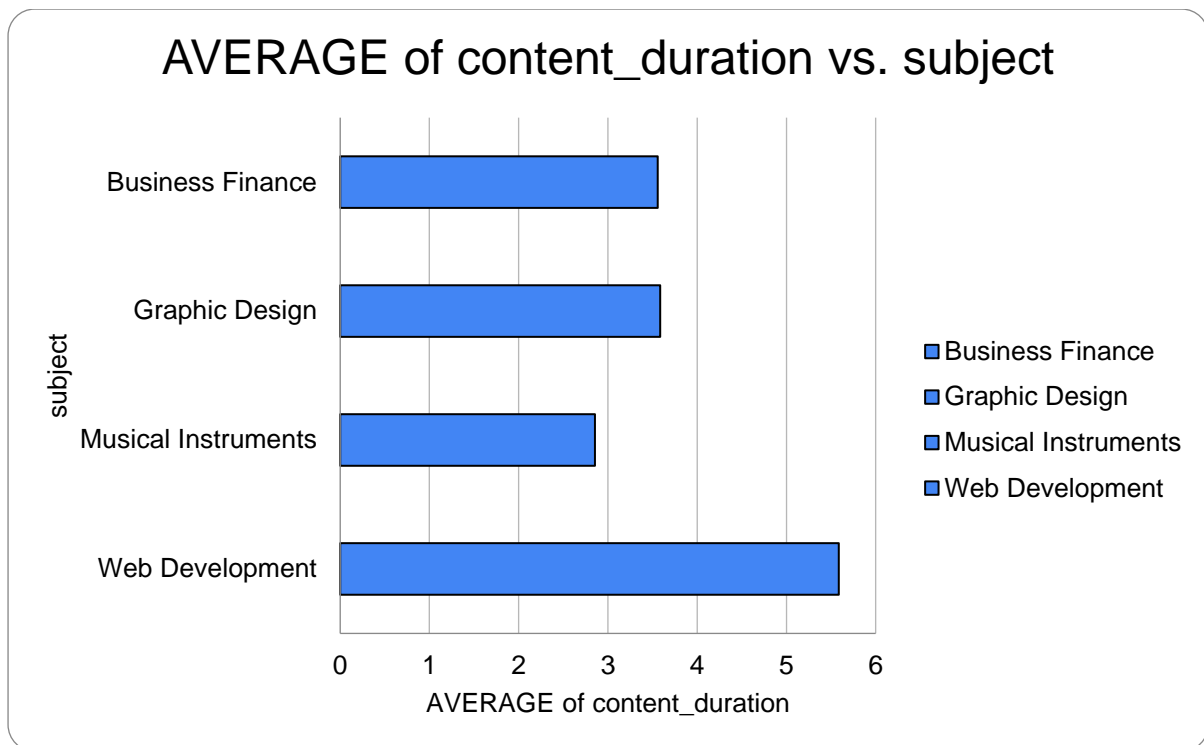
Course	Rating	Level	Free/Paid	Duration	Subject
Learn HTML5 Programming From Scratch	0.82	Beginner Level	Free	10.5	Web Development
Coding for Entrepreneurs Basic	0.69	Expert Level	Free	3.5	Web Development
The Web Developer Bootcamp	0.89	Beginner Level	Paid	43	Web Development
Build Your First Website in 1 Week with HTML5 and CSS3	0.78	All Levels	Free	3	Web Development
The Complete Web Developer Course 2.0	0.55	Beginner Level	Paid	30.5	Web Development
Free Beginner Electric Guitar Lessons	0.88	All Levels	Free	4.5	Musical Instruments
Web Design for Web Developers: Build Beautiful Websites!	0.82	All Levels	Free	3	Web Development
Learn Javascript & JQuery From Scratch	0.79	All Levels	Paid	2	Web Development
Practical PHP: Master the Basics and Code Dynamic Websites	0.85	Intermediate Level	Free	6.5	Web Development
JavaScript: Understanding the Weird Parts	0.69	All Levels	Paid	11.5	Web Development
Pianoforall - Incredible New Way To Learn Piano & Keyboard	0.96	Beginner Level	Paid	30	Musical Instruments
Angular 4 (formerly Angular 2) - The Complete Guide	0.90	Beginner Level	Paid	22	Web Development
Beginner Photoshop to HTML5 and CSS3	0.94	All Levels	Free	2	Web Development
Web Development By Doing: HTML / CSS From Scratch	0.39	All Levels	Free	1	Web Development
HTML and CSS for Beginners - Build a Website & Launch ONLINE	0.57	All Levels	Free	6	Web Development
Become a Web Developer from Scratch	0.61	All Levels	Paid	27.5	Web Development
Bitcoin or How I Learned to Stop Worrying and Love Crypto	0.56	All Levels	Free	8	Business Finance
Quickstart AngularJS	0.96	Beginner Level	Free	1.5	Web Development
Learn Responsive Web Development from Scratch	0.55	All Levels	Free	4.5	Web Development
Learn and Understand AngularJS	0.87	Beginner Level	Paid	7	Web Development

Here, we can find that most of the most popular courses are of Web Development. Hence, it is understood that users prefer web development in 90% of the cases. Therefore, more courses on those topics would yield more income or more subscriber to the company.

Also, it is found out that most of the subscriptions are done by beginners (since most of the course levels are either beginner level or those of all level).



The above diagram proves that most subscribers prefer Web Development followed by Business Finance, Graphic Design and finally Musical Instruments.



### Average Duration vs Subject

The courses in Web Development have more hours of duration on an average. Considering that it is a subject having very wide amounts of concepts it is understandable.



# Conclusion

Therefore, we are able to find that Web Development is one of the highly sought-after courses in Udemy. Customers prefer almost all levels of courses from beginner to advanced almost equally. Hence, it is important to have courses in all of the levels.

For the content duration, the duration of the course would be better if it doesn't become so lengthy except for Beginner Level courses as they need to be lengthy because so many concepts are essential to be explained in a timely manner.

In order to gain more popularity some of the beginner level courses from each of the subjects can be given for free because lot of the beginner might not be willing to spend more money as they don't have a thorough knowledge of the subject to start with.

# ICC Test Batting Project Description

Test Cricket is a form of Cricket in which 2 teams play each other for a total of 4 innings which span for 5 days. The team scoring more runs and taking all the wickets of the opposite team by the end of 5<sup>th</sup> day is considered the winner of the match. If both the team fails to do so by the end of 5<sup>th</sup> day, it can result in a “Draw”, which means there is no clear winner in the match.

The ICC Test cricket Batting Figures Dataset consist of runs scored by the Batsmen from 1877 to 2019 December.

## Columns

Player – Name of the player

Span – duration in which player played

Mat – number of matches played

Inn – number of innings batted

NO – number of times batsman remained not out

Runs – total runs scored

Highest Score – highest score in career

Average – runs scored per innings (minus the number of times batsman remained not out)

Hundreds

Fifties

# Design

The Dataset contains majority of the column in the way in which we require. For some batsman with only 1 or 2 matches played, most of the columns remains empty (only less than 10 players). Hence, we can delete the rows containing those players.

# Analysis

We can answer variety of questions from the dataset. Some of the questions we are trying to answer are:

1. Batsman with
  - a. highest runs
  - b. high score
  - c. best average
  - d. most ducks (0 per innings)
  - e. most fifties
  - f. most hundreds
2. Top 10 players (based on runs scored)
3. Comparison of certain players scores

# Results

On analysing the dataset, we were able to answer the above questions :

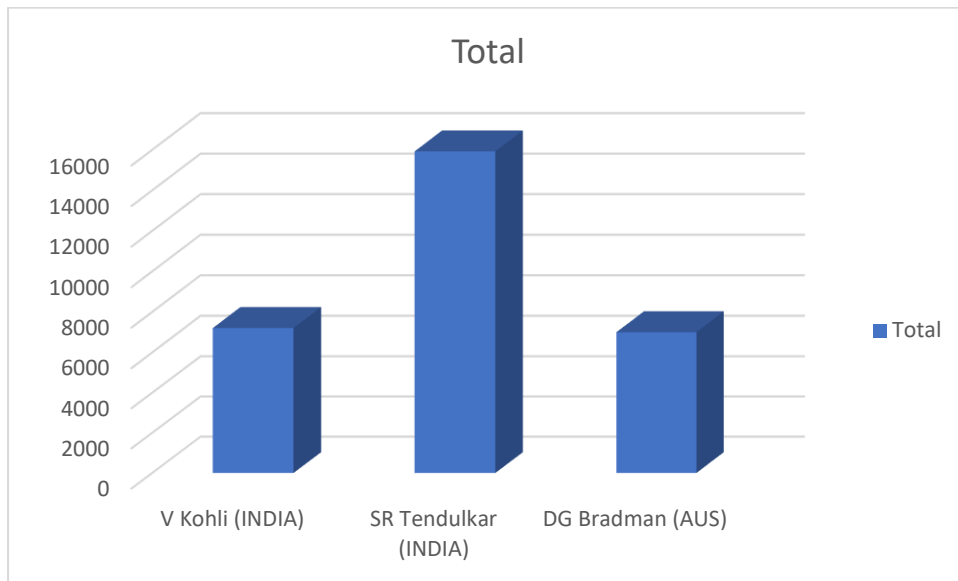
1. Batsman with

- a. highest runs – Sachin Tendulkar
- b. high score – Brian Lara
- c. best average – Donald Bradman
- d. most ducks (0 per innings) – CA Walsh
- e. most fifties – Sachin Tendulkar
- f. most hundreds – Sachin Tendulkar

2. Top 10 players (based on runs scored)

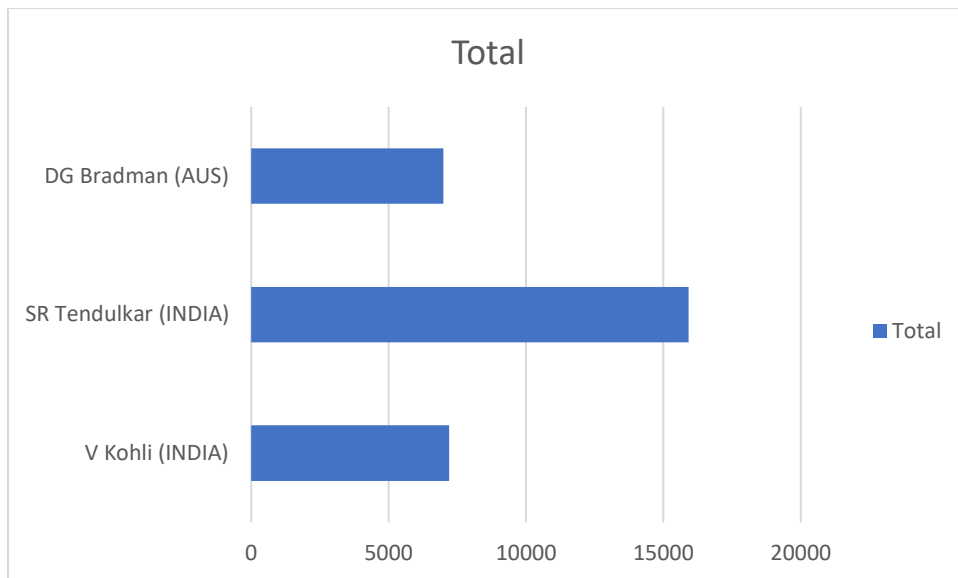
<a href="#">SR Tendulkar (INDIA)</a>	1989-2013	200	329	33	▼ <b>15921</b>	248*	53.78
<a href="#">RT Ponting (AUS)</a>	1995-2012	168	287	29	<b>13378</b>	257	51.85
<a href="#">JH Kallis (ICC/SA)</a>	1995-2013	166	280	40	<b>13289</b>	224	55.37
<a href="#">R Dravid (ICC/INDIA)</a>	1996-2012	164	286	32	<b>13288</b>	270	52.31
<a href="#">AN Cook (ENG)</a>	2006-2018	161	291	16	<b>12472</b>	294	45.35
<a href="#">KC Sangakkara (SL)</a>	2000-2015	134	233	17	<b>12400</b>	319	57.4
<a href="#">BC Lara (ICC/WI)</a>	1990-2006	131	232	6	<b>11953</b>	400*	52.88
<a href="#">S Chanderpaul (WI)</a>	1994-2015	164	280	49	<b>11867</b>	203*	51.37
<a href="#">DPMD Jayawardene (SL)</a>	1997-2014	149	252	15	<b>11814</b>	374	49.84

### 3. Comparison of runs scored by Sachin and Kohli



## Comparison of runs scored and average of Sachin, Kohli and Bradman

### Average



## Run scored

