

Online Courses Website (Tool For Price Range Prediction and Analysis of Consumer Behaviour)

A Thesis Submitted in partial
fulfilment of the requirements for the
degree of Bachelor of Technology

in

Data

Science

By



Ahnaan Merchant

(SAP ID 70091019038)

under the guidance of
Prof. Jay Krishna Joshi

Data Science and Artificial Intelligence

Department of Data

Science NMIMS

University Mumbai

Certificate

This is to certify that the thesis of the Capstone Project entitled “**Online Courses Website (Tool For Price Range Prediction and Analysis of Consumer Behaviour)**” is a bonafide work of “**Ahnaan Merchant (SAP ID: 70091019038)**” submitted to the NMIMS University in partial fulfilment of the requirement for the award of the degree of “**Bachelor of Technology**” in “**Data Science**”.

Acknowledgement

It is with a feeling of great pleasure that I would express my most sincere heartfelt gratitude to **Prof. Jaykrishna Joshi** for his steady and able guidance throughout my capstone project tenure. I would also like to extend my sincere thanks to **Prof. Siba Panda** for his guidance and encouragement, and for providing the necessary facilities in the department. I am also thankful to all the staff members of the department of Data Science and Artificial Intelligence, especially **Prof. Ekta Shah** for her constant inspiration and help.

Table of Contents

Sr No.	Topic	Page No.
1.	ABSTRACT	5
2.	INTRODUCTION	6-7
3.	WEB DEVELOPMENT	8-9
4.	RESEARCH METHODOLOGY	10-11
5.	DATA EXPLORATION	12-13
6.	CONSUMER BEHAVIOUR ANALYSIS	14-18
7.	TOOL BUILDING USING MACHINE LEARNING	19-20
8.	FUTURE SCOPE AND CONCLUSION	21
9.	REFERENCES	22

Abstract

Throughout the pandemic, we as students struggled with online lectures as our medium of learning was snatched away from us. Our professors expected us to pay attention to online lectures and with all our sincerity we tried our best initially. But the learning environment never felt the same with family members constantly disturbing you with household chores. Eventually we started giving up. The pace of online lectures did not match our daily schedule. But in this darkness, we found the streetlight our fathers talked about.

Websites like Udemy and Coursera helped us learn throughout the pandemic with Coursera even providing free courses without certifications. With the internet access with us always, we could learn at our own pace.

While exploring these websites and learning the essentials of data science, I came across a method through which anyone can teach on these websites. But these courses even if not trending were expensive. Clearly there was a gap in understanding the consumers. Here I thought about my capstone idea.

Why not develop a tool for teachers on these websites which can help predict the price range of their courses using data science?

Introduction

In websites like Udemy, the instructor of the course determines the list price of the course. In addition, both instructors and Udemy can offer discounts on courses at any time. Udemy's goal is to help match students to the right course at the right price, and occasionally that may result in different promotions for the same course. For example, promotions may be run that target different regions, markets, or prospective learners. In addition, instructors also have the ability to run independent promotions for their courses if they wish.

Further, when purchasing courses via their mobile app, depending on your device, location, and what country your Udemy account is registered in, your payment options may include Apple App Store and Google Play. Udemy does not have control over the price tier structure provided by these platforms. As a result, the course price listed on the mobile apps may differ slightly from the course price posted on Udemy.com. This is because the pricing matrix for these mobile providers will select the price tier that is closest to the course price set by the Udemy instructor, but they do not always match up exactly.

Additionally, courses on the mobile app may have a different price listed during a sale, because there is no way to enter coupon codes on the Udemy mobile apps, like there is on the web version of Udemy.com.

Teaching quality is more consistent on Coursera – though a number of Udemy's instructors are also high-profile professionals. Udemy's certificates of completion are not accredited, Coursera's are. Coursera is the more expensive of the two.

Every company has different strategy for showing different prices to different people. It's a trick to grab new customers by cutting prices. Most of the marketers create many plans to increase sales.

The idea behind this project was to build and replicate a website similar to the ones we have been using throughout the pandemic and use that website to obtain the necessary resources to build the tool. The tool would basically take some inputs from a user who wants to teach on our website and use

those inputs to predict a price range and display it to the user which he can possibly use. This recommendation system will help these users who want to teach on our website to decide the price they want to keep for their courses.

This website will have its own UI and own branding which will cater to its users. The users can use this website with the best of functionalities. The website will be responsive, that is, it can be used on any device and the user won't face any issues in terms of UI/UX. We will expand more on this website and its uses later in this report.

All websites on the internet (the secured ones) collect data through various sources. Some use cookies, some use forms and there are many other ways through which data can be collected. Udemy uses cookies, which are small text files stored by your browser, to collect, store, and share data about your activities across websites, including on Udemy. They allow us to remember things about your visits to Udemy, like your preferred language, and to make the site easier to use.

Coursera will use your information to provide and operate their Live2Coursera services and will make certain information available to other Live2Coursera participants: For example, we will allow students to view the content uploaded by their instructors and will allow instructors to know which students have viewed their lectures.

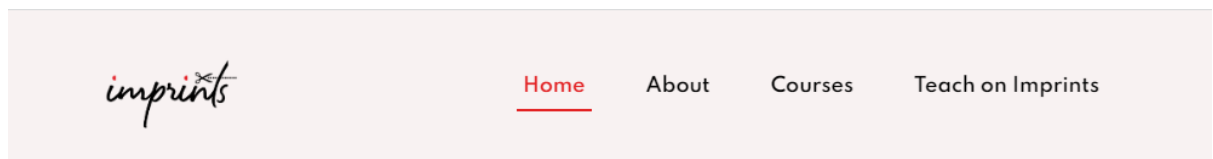
Data collection essentially helps many websites to improve their services. The development of our tool also uses data collected from people.

Web Development

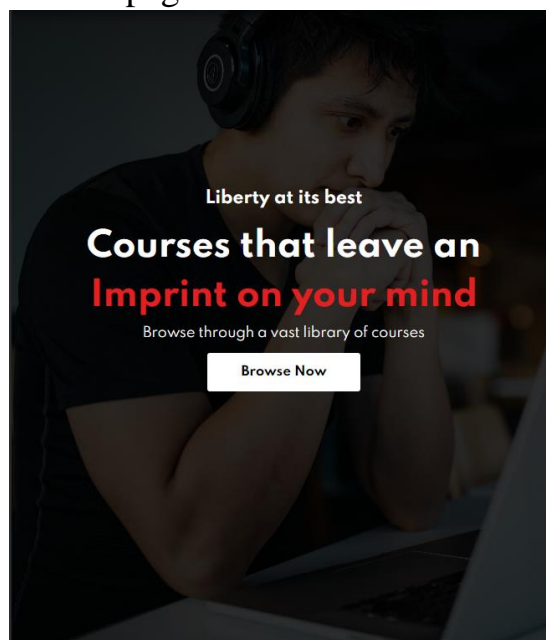
The first and foremost part of our project is the development of our website. Our brand name is Imprints – Liberty at its best.



This website is a multi-page website with the following pages: Home, About, Courses, Teach on Imprints.

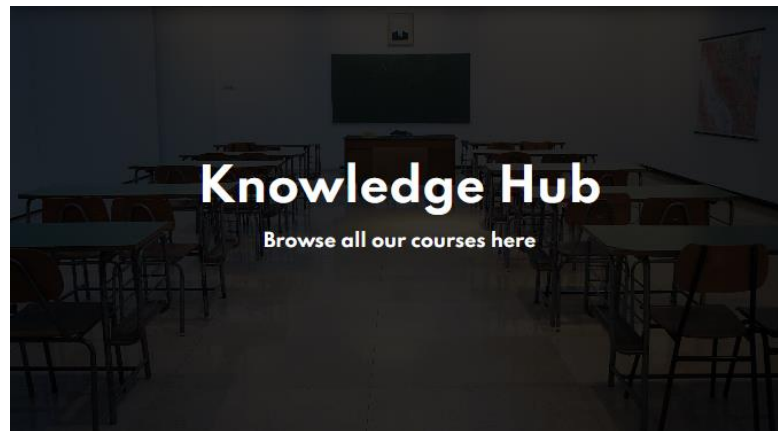


Home Page: It consists of a brief of every page on our website. We will talk about how we use this page in data collection further in this report.



About Page: It gives a brief of what our brand is which talks about cutting away from the traditional ways of learning.

Courses Page: Here our user can browse through a library of courses.



Teach on Udemy: We will talk about this page later in this report.

The development of this website was with the help of HTML, CSS and Javascript. It is a fully responsive page that can be viewed on any device.

The colour scheme and the UI of this website matches our logo. Using our logo, we create the content of our website.

Web Development is an essential part of this project as this website can be used and has been used for many purposes.

Research Methodology

The main aim of this project was to understand the consumer behaviour on websites that provide online courses which could be done by analysing the data. At the same time, I wanted to use that data to develop my tool.

So, I had to collect the data accordingly. We used a google form which has been linked in our website to collect data from users.

**We value our customers and guide our experts
to design their courses accordingly**

Click Here
to give your valuable input

The most important thing was to think about the questions that can satisfy the aim of this project.

Eventually I came up with the following questions to collect our data and use it in our dataset to build our tool.

Are you a student? *

- ☐ Yes
- ☐ No

Which format do you prefer your courses in? *

- ☐ Video
- ☐ Audio
- ☐ Document
- ☐ None of the above

Who do you prefer to take courses from? *

- ☐ Organizations
- ☐ Individual Instructors

How many hours per week do you spend on your courses *

- 1 2 3 4 5 6 7 8 9 10
- ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

How many courses do you take at a time? *

- 1 2 3 4 5
- ☐ ☐ ☐ ☐ ☐

Which of these are you likely to choose as your field of interest *

Business

The prices of courses varies depending on the above factors. Depending on your selected options above, what price range do you think should be of your choice of course? *

- ☐ ₹ 499 - ₹ 599
- ☒ ₹ 699 - ₹ 799
- ☐ ₹ 899 - ₹ 999
- ☐ ₹ 1000 +



Data Exploration

Eventually our dataset would look like this:

	Student	Format	Instructors	Hours per week	No. of courses	Categories	Price Range
0	Yes	Video	Individual Instructors	7	3	Health & Fitness	499 - 599
1	Yes	Video	Individual Instructors	5	2	Data Science	499 - 599
2	Yes	Video	Individual Instructors	8	2	Data Science	499 - 599
3	Yes	Document	Individual Instructors	10	2	Data Science	499 - 599
4	No	Audio	Individual Instructors	5	1	Personality Development	1000 +

Student: Determining whether our user is a student or not will definitely help predicting the price range.

Format: This feature had 3 formats in which people prefer to learn from namely: Video, Audio, Document or None of the above.

Instructors: Determining whether our users prefer to take courses from organizations or individual instructors.

Hours per week: This feature tells us how many hours per week our users dedicate to their courses.

No. of courses: The number of courses our users can take at a time.

Categories: Our users' field of interest. Note that these categories are directly taken from Udemy and there are 8 of these diverse categories which can also be seen on our website on the Courses Page.

Business
Finance & Accounting
IT & Software
Personality Development
Design
Photography & Video
Health & Fitness
Data Science

Price Range: There are 4 price ranges and this feature will be our target variable further when we build our tool.

This data collected from our users has a shape of 109, 8 out of which 7 features are our input features and the price range feature is our target variable. Yes, the data collected is comparatively small in size but has been collected from real users and has helped us further in this project. Since, this data has been collected by real users and the collection method was designed in a way where there can be no null values, our data has 0 missing values.

There could have been other features that we could have created but their answer would have been obvious and hence create an imbalanced feature. For example: Do you prefer courses with certifications? The answer would have been an obvious yes because everybody wants proof of what they are learning.

Features like hours per week and no. of courses at a time were collected keeping in mind the realistic aspects like no one can possibly learn from more than 5 different courses or no can dedicate more than 10 hours a week for a course.

We will talk further about this data and the feature importance in the final output section of this report.

Consumer Behaviour Analysis

Using our data, we will look for patterns in how our users think. Analysing the consumer behaviour is essential for every company. Understanding consumer behaviour is important for businesses because it can help them to make better decisions about their products and services. By understanding why people purchase certain products and how they use them, businesses can adapt their offerings to better suit the needs and wants of their target market.

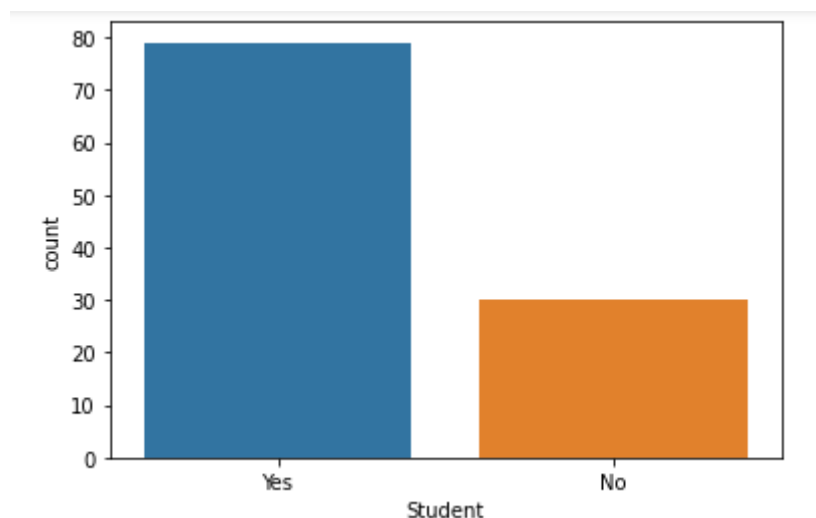
In our case, it is important to analyse the consumer behaviour to design our courses accordingly and help our users who want to teach on our website.

About Us

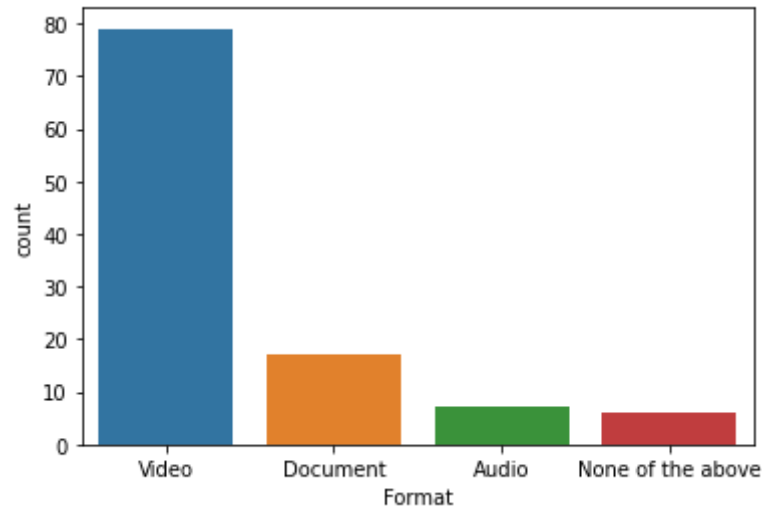
Cut away from the traditional ways of learning and explore the world of knowledge on your own pace. Imprints gives you the opportunity to learn according to your topic of interest at an extremely cheap rate. All our courses are developed by experts and every course you see here will definitely leave an imprint on your mind.

[Read More](#)

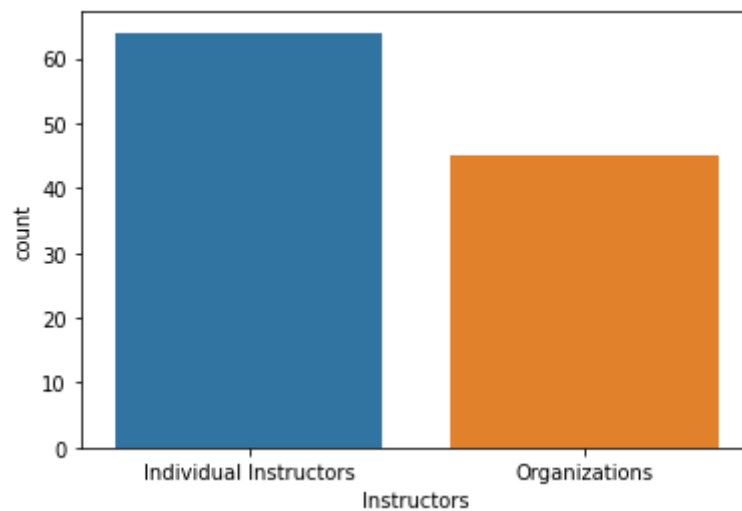
Let's look at the following charts and look for patterns in our consumer behaviour.



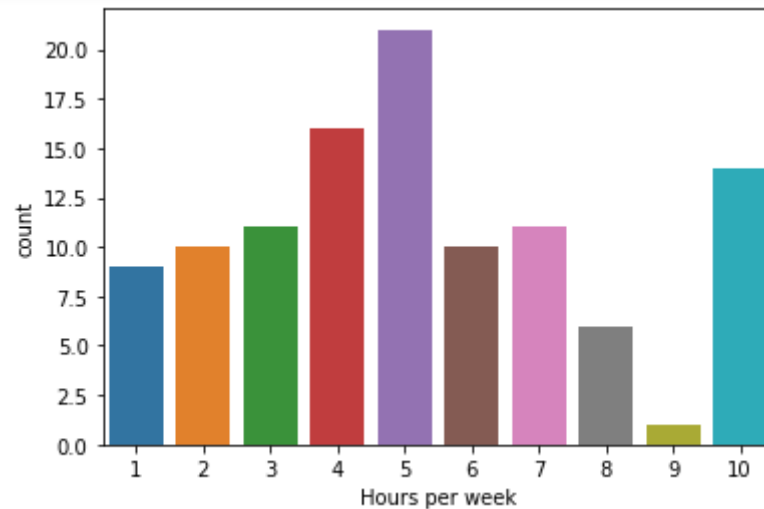
Inference: The above chart clearly shows us that the number of users who are students is more than the number of users who are not students.



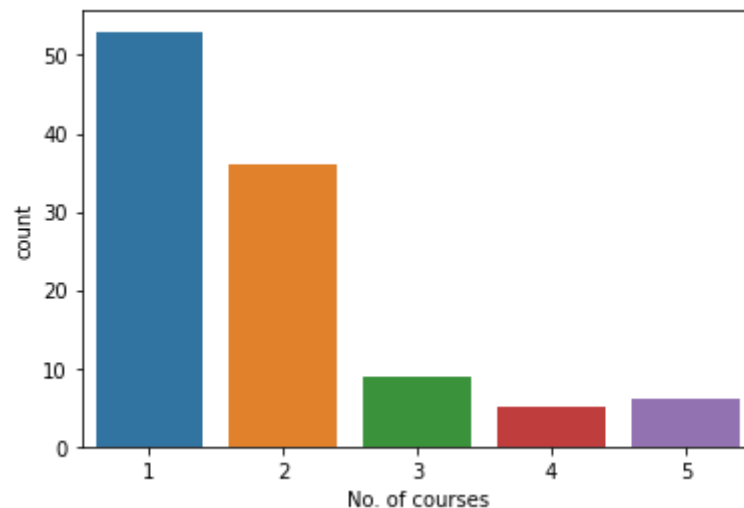
Inference: Clearly, people prefer Video format over other mediums.



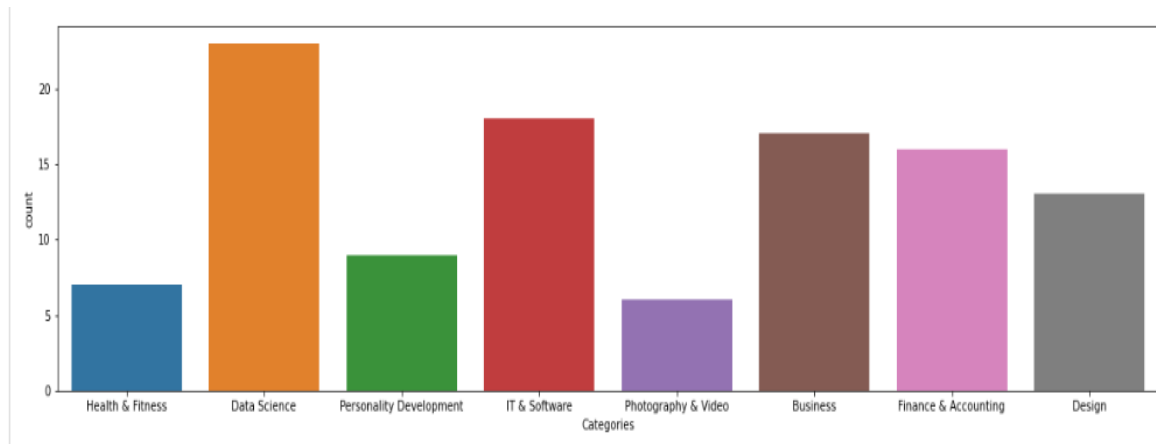
Inference: The count of users who prefer individual instructors rather than organizations is more than the users who prefer it from organizations. We can conclude from this that people want to learn through these online courses.



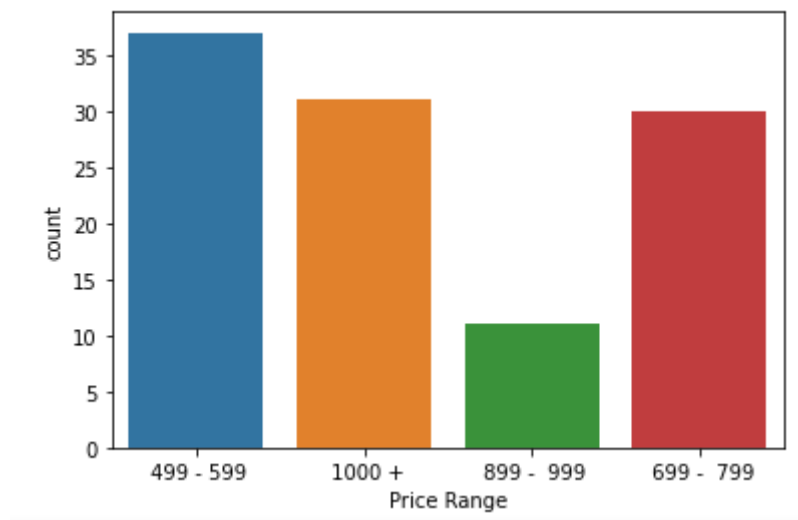
Inference: We can see a normal distribution type chart from 1 to 9 and at the tail there is an increase in the count.



Inference: We can see the count dropping as the number of courses increases which is a human trait and we can conclude that many users prefer taking 1 or 2 courses over multitasking.



Inference: With the rise in demand for technology, we can see here that tech related categories are in high-demand over other categories.



Inference: It would be obvious for a data filled by users to prefer courses that are cheap like 499-599, 699-799 and clearly our users prefer that. But surprisingly, many users are willing to pay more than 1000 Rs. If the courses are worth it.



Inference: The correlation between hours per week and no. of courses is positive but is not high and it makes sense because it doesn't if the number of courses taken by users are more, they will only be able to dedicate a limited amount of time.

Tool Building Using Machine Learning

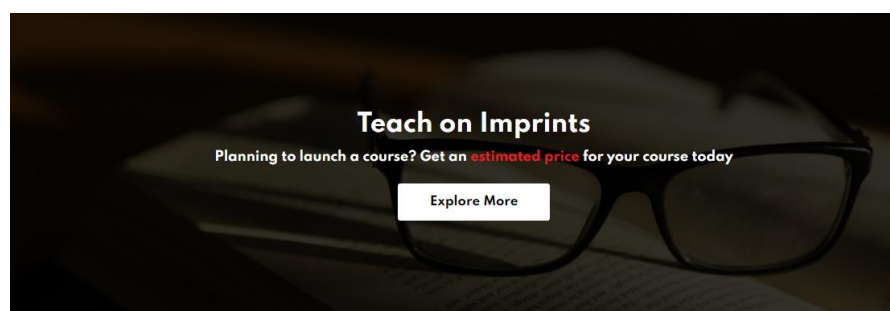
We label encode our categorical variables to pass it through our machine learning model. The model that we are going to use is Decision Tree Classifier. The decision tree classifier creates the classification model by building a decision tree. Each node in the tree specifies a test on an attribute, each branch descending from that node corresponds to one of the possible values for that attribute. We use the criterion gini in our model. Gini impurity is an important measure used to construct the decision trees. Gini impurity is a function that determines how well a decision tree was split. Basically, it helps us to determine which splitter is best so that we can build a pure decision tree. Gini impurity ranges values from 0 to 0.5.

On applying this model to our encoded data, we get the accuracy as 55%. To check why we get such a low accuracy we check the feature importance.

```
model.feature_importances_  
array([0.24851702, 0.07873119, 0.08935661, 0.24159372, 0.  
       0.34180145])
```

We can see that Student, Hours Per Week and Categories are the most important features. But it's not that the other features are insignificant, it's that our data is quite insufficient and is flawed by human error. We will talk about improvements further in the report.

We use this model to build a tool for our website where anyone who wants to teach on our website will fill a form and get an estimated price range for their course.



Are you making your course for students?

- ☐ Yes
☐ No

Which format will your courses mostly be in?

- ☐ Video
☐ Audio
☐ Document
☐ None of the above

Are you providing this course on behalf of an organization or are you an individual instructor?

- ☐ Organization
☐ Individual Instructor

How many hours do you expect your students to dedicate per week?

(Choose between 1 to 10)

How many courses do you expect your students to study at a time?

(Choose between 1 to 5)

Choose a category for your course

Predict

These questions take the same input as our data collection form and hence can be used to predict the estimated price.

The final prediction will look like this:

```
model_run("Yes","Video","Organizations",7, 2,"Data Science")  
array([' 699 - 799'], dtype=object)
```

Future Scope and Conclusion

The insufficiency of data and human error leads to a not so accurate model but the same data gives us some great insights on consumer behaviour patterns which can be used to improve the services on various online courses website like Udemy and Coursera.

After collecting sufficient data to improve the accuracy of our tool we can also implement it into our website by building a web app using Flask – a python library. It will be an end-to-end UI for users who want to teach to get an accurate price estimation of these courses. This will help them design better courses for users. The accurate price estimation will help these teachers get more students as well because the price range will fit the bracket of the customers.

We can expand our survey by asking more relevant questions and increasing the number of features to improve the model accuracy as well.

But the consumer behaviour analysis can be used for improving various services on multiple online courses providers to understand who their customers are, what they want, what are they willing to pay, etc.

References

<https://www.w3schools.com/>

[https://www.sciencedirect.com/topics/computer-science/decision-tree-classifier#:~:text=The%20decision%20tree%20classifier%20\(Pang,possible%20values%20for%20that%20attribute.](https://www.sciencedirect.com/topics/computer-science/decision-tree-classifier#:~:text=The%20decision%20tree%20classifier%20(Pang,possible%20values%20for%20that%20attribute.)

<https://www.futurelearn.com/info/blog/how-to-understand-consumer-behaviour#:~:text=Understanding%20consumer%20behaviour%20is%20important,wants%20of%20their%20target%20market.>

<https://in.coursera.org/about/privacy>

<https://www.udemy.com/terms/privacy/#:~:text=We%20use%20cookies%2C%20which%20are,the%20site%20easier%20to%20use.>

<https://palletsprojects.com/p/flask/#:~:text=Flask%20is%20a%20lightweight%20WSGI,popular%20Python%20web%20application%20frameworks.>