In [ ]:
```
Name: Ishan Chaskar
URK21CS1181
```

In [ ]:
```
Aim: To develop a Udemy course recommender system with content-based
     recommendation using the scikit-learn
```

In [ ]:
```
Description:

 Recommender systems are among the most popular applications of data science
 today. There are a lot of applications where websites collect data from
 their users and use that data to predict the likes and dislikes of their
 users. This allows them to recommend the content that they like.

 A Recommender System is a software system that provides specific suggestions
 to users according to their preferences. These techniques may provide
 decision-making capabilities to the user. Items refer to any product that
 the recommender system suggests to its user like movies, music, news, travel
 packages, e-commerce products, etc.

 Content-based recommenders suggest similar items based on a particular item.
 This system uses item metadata,such as genre, director, description, actors,
 etc. for movies, to make these recommendations.
 And to recommend that, it will make use of the user& past item metadata.
 A good example could be YouTube, where based on the history,
 it suggests new videos that can be potentially watched.

 Term Frequency(TF)=Number of times term t appears in document d/Total
 number of terms in document d

 Inverse Document Frequency (IDF)=log(Total number of documents in the
 corpus D/Number of documents containing term t)
 TF-IDF(t,d,D)=TF(t,d)×IDF(t,D) Cosine similarity is a metric used to measure
 how similar two non-zero vector
```

In [1]:
```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
df = pd.read_csv('udemy_courses.csv')
df
```

Out[1]:

	course_id	course_title	url	is_paid	price	num_subscribers
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147
1	1113822	Complete GST Course & Certification - Grow You...	https://www.udemy.com/goods-and-services-tax/	True	75	2792
2	1006314	Financial Modeling for Business Analysts and C...	https://www.udemy.com/financial-modeling-for-b...	True	45	2174
3	1210588	Beginner to Pro - Financial Analysis in Excel ...	https://www.udemy.com/complete-excel-finance-c...	True	95	2451
4	1011058	How To Maximize Your Profits Trading Options	https://www.udemy.com/how-to-maximize-your-pro...	True	200	1276
...	...	...	...	...	...	...
3673	775618	Learn jQuery from Scratch - Master of JavaScri...	https://www.udemy.com/easy-jquery-for-beginner...	True	100	1040
3674	1088178	How To Design A WordPress Website With No Codi...	https://www.udemy.com/how-to-make-a-wordpress-...	True	25	306
3675	635248	Learn and Build using Polymer	https://www.udemy.com/learn-and-build-using-po...	True	40	513
3676	905096	CSS Animations: Create Amazing Effects on Your...	https://www.udemy.com/css-animations-create-am...	True	50	300
3677	297602	Using MODX CMS to Build Websites: A Beginner's...	https://www.udemy.com/using-modx-cms-to-build-...	True	45	901

3678 rows × 12 columns

In [ ]:
```
1. Develop a Udemy course recommender system with content-based recommendation
   using the scikit-learn
a. Use the column: 'course_title'
```

In [2]:
```
print('URK21CS1181')
col = df[['course_title']]
col
```

```
URK21CS1181
```

Out[2]:

	course_title
0	Ultimate Investment Banking Course
1	Complete GST Course & Certification - Grow You...
2	Financial Modeling for Business Analysts and C...
3	Beginner to Pro - Financial Analysis in Excel ...
4	How To Maximize Your Profits Trading Options
...	...
3673	Learn jQuery from Scratch - Master of JavaScri...
3674	How To Design A WordPress Website With No Codi...
3675	Learn and Build using Polymer
3676	CSS Animations: Create Amazing Effects on Your...
3677	Using MODX CMS to Build Websites: A Beginner's...

3678 rows × 1 columns

In [ ]:
```
b. Remove the leading and trailing whitespaces in that column
```

In [3]:
```
col['course_title'] = col['course_title'].str.strip()
col['course_title']
```

```
/tmp/ipykernel_2062100/2318620746.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy
  col['course_title'] = col['course_title'].str.strip()
```

Out[3]:
```
0                       Ultimate Investment Banking Course
1         Complete GST Course & Certification - Grow You...
2         Financial Modeling for Business Analysts and C...
3         Beginner to Pro - Financial Analysis in Excel ...
4              How To Maximize Your Profits Trading Options
                                ...
3673      Learn jQuery from Scratch - Master of JavaScri...
3674      How To Design A WordPress Website With No Codi...
3675                          Learn and Build using Polymer
3676      CSS Animations: Create Amazing Effects on Your...
3677      Using MODX CMS to Build Websites: A Beginner's...
Name: course_title, Length: 3678, dtype: object
```

In [ ]:
```
c. Perform feature extraction using Term Frequency Inverse Document Frequency
(TF-IDF).
```

In [4]:
```
print('URK21CS1181')
tf = TfidfVectorizer()
tfidf_matrix = tf.fit_transform(col['course_title'])
print(tfidf_matrix.shape)
```

```
URK21CS1181
(3678, 3716)
```

In [ ]:
```
d. Compute the cosine similarity.
```

In [5]:
```
print('URK21CS1181')
cosine_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)
print(cosine_sim.shape)
```

```
URK21CS1181
(3678, 3678)
```

In [ ]:
```
e. Display the top 'n' suggestions with the similarity score for the given user
   input.
```

In [6]:
```
print('URK21CS1181')
titles = col['course_title']
indices = pd.Series(col.index, index=col['course_title'])
```

```
URK21CS1181
```

In [8]:
```
print('URK21CS1181')
title = input("Enter the columns to suggest: ")
num = int(input("Number of suggestions: "))
idx = indices[title]
```

```
URK21CS1181
```

In [9]:
```
print('URK21CS1181')
sim_scores = list(enumerate(cosine_sim[idx]))
sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)
sim_scores = sim_scores[1:num + 1]
sim_scores
```

```
URK21CS1181
```

Out[9]:
```
[(220, 0.3880382676899578),
 (414, 0.3762140375317894),
 (43, 0.36346910833870283),
 (408, 0.3437200574400444)]
```

In [10]:
```
col_indices = [i[0] for i in sim_scores]
scores = [i[1] for i in sim_scores]
```

In [16]:
```
print('URK21CS1181')
print("Suggestions to " + title + "...")
print("\n")
for rec in range(num):
    print("Suggested:"+titles[col_indices[rec]]+"(score:"+str(scores[rec])+")")
```

```
URK21CS1181
Suggestions to How To Maximize Your Profits Trading Options...


Suggested:Introduction to Options - Strategies For Consistent Profits(score:0.3
880382676899578)
Suggested:How to trade options(score:0.3762140375317894)
Suggested:Options Trading - How to Win with Weekly Options(score:0.363469108338
70283)
Suggested:Trading Options Basics(score:0.3437200574400444)
```

In [ ]:

```
Result:
Hence the aim to develop a Udemy course recommender system with
content-based recommendation using the scikit-learn has been coded and
executed successfully.
```