```python
import pandas as pd
import numpy as np
import neattext.functions as nfx
import seaborn as sn

from sklearn.feature_extraction.text import TfidfVectorizer,CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity,linear_kernel
```

```python
df = pd.read_csv('udemy_courses.csv')
df.head()
```

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147	23	51	All Levels	
1	1113822	Complete GST Course & Certification -	https://www.udemy.com/goods-and-services-tax/	True	75	2792	923	274	All Levels	

```python
dir(nfx)
```

```
['BTC_ADDRESS_REGEX',
 'CURRENCY_REGEX',
 'CURRENCY_SYMB_REGEX',
 'Counter',
 'DATE_REGEX',
 'EMAIL_REGEX',
 'EMOJI_REGEX',
 'HASTAG_REGEX',
 'MASTERCard_REGEX',
 'MD5_SHA_REGEX',
 'MOST_COMMON_PUNCT_REGEX',
 'NUMBERS_REGEX',
 'PHONE_REGEX',
 'PoBOX_REGEX',
 'SPECIAL_CHARACTERS_REGEX',
 'STOPWORDS',
 'STOPWORDS_de',
 'STOPWORDS_en',
 'STOPWORDS_es',
 'STOPWORDS_fr',
 'STOPWORDS_ru',
 'STOPWORDS_yo',
 'STREET_ADDRESS_REGEX',
 'TextFrame',
 'URL_PATTERN',
 'USER_HANDLES_REGEX',
 'VISACard_REGEX',
 '__builtins__',
 '__cached__',
 '__doc__',
 '__file__',
 '__generate_text',
 '__loader__',
 '__name__',
 '__numbers_dict',
 '__package__',
 '__spec__',
 '_lex_richness_herdan',
 '_lex_richness_maas_ttr',
 'clean_text',
 'defaultdict',
 'digit2words',
 'extract_btc_address',
 'extract_currencies',
 'extract_currency_symbols',
 'extract_dates',
 'extract_emails',
 'extract_emojis',
 'extract_hashtags',
 'extract_html_tags',
 'extract_mastercard_addr',
 'extract_md5sha',
 'extract_numbers',
 'extract_pattern',
 'extract_phone_numbers',
```

```
        'extract_postoffice_box',
        'extract_shortwords',
        'extract_special_characters'
```

```
df['course_title'].iloc[1:5]
```

```
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
Name: course_title, dtype: object
```

```
df['Clean_title'] = df['course_title'].apply(nfx.remove_stopwords)
```

```
df['Clean_title'] = df['Clean_title'].apply(nfx.remove_special_characters)
```

```
df['Clean_title'].iloc[1:5]
```

```
1    Complete GST Course  Certification  Grow Practice
2    Financial Modeling Business Analysts Consultants
3           Beginner Pro  Financial Analysis Excel 2017
4                    Maximize Profits Trading Options
Name: Clean_title, dtype: object
```

```
countvect = CountVectorizer()
```

```
cv_mat = countvect.fit_transform(df['Clean_title'])
```

```
cv_mat
```

```
<3683x3564 sparse matrix of type '<class 'numpy.int64'>'
        with 18364 stored elements in Compressed Sparse Row format>
```

```
cv_mat.todense()
```

```
matrix([[0, 0, 0, ..., 0, 0, 0],
        [0, 0, 0, ..., 0, 0, 0],
        [0, 0, 0, ..., 0, 0, 0],
        ...,
        [0, 0, 0, ..., 0, 0, 0],
        [0, 0, 0, ..., 0, 0, 0],
        [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

```
df_cv_words = pd.DataFrame(cv_mat.todense(),columns=countvect.get_feature_names())
```

```
df_cv_words
```

	000005	001	01	02	10	100	101	101master	102	10k	...	zend	zero	zerotohero	zf2	zinsen	zoho	zombie	zu	zuhause	zur
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3678	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3679	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3680	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3681	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3682	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

3683 rows × 3564 columns

```
cosine_sim_mat = cosine_similarity(cv_mat)
```

```
cosine_sim_mat
```

```
array([[1.        , 0.20412415, 0.        , ..., 0.        , 0.        ,
        0.        ],
       [0.20412415, 1.        , 0.        , ..., 0.        , 0.        ,
```

```
        0.      ],
       [0.      , 0.      , 1.      , ..., 0.      , 0.      ,
        0.      ],
       ...,
       [0.      , 0.      , 0.      , ..., 1.      , 0.      ,
        0.23570226],
       [0.      , 0.      , 0.      , ..., 0.      , 1.      ,
        0.      ],
       [0.      , 0.      , 0.      , ..., 0.23570226, 0.      ,
        1.      ]])
```

```
course_index = pd.Series(df.index,index = df['course_title']).drop_duplicates()
```

```
course_index
```

```
course_title
Ultimate Investment Banking Course                          0
Complete GST Course & Certification - Grow Your CA Practice  1
Financial Modeling for Business Analysts and Consultants     2
Beginner to Pro - Financial Analysis in Excel 2017           3
How To Maximize Your Profits Trading Options                 4
                                                          ...
Learn jQuery from Scratch - Master of JavaScript library  3678
How To Design A WordPress Website With No Coding At All    3679
Learn and Build using Polymer                              3680
CSS Animations: Create Amazing Effects on Your Website     3681
Using MODX CMS to Build Websites: A Beginner's Guide       3682
Length: 3683, dtype: int64
```

```
temp = df[df['course_title'].str.contains('Python')]
temp.head()
```

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	con
14	1196544	Python Algo Trading: Sentiment Trading with News	https://www.udemy.com/hedge-fund-strategy-trad...	True	200	294	19	42	All Levels	
30	1170894	Python Algo Stock Trading: Automate Your Trading!	https://www.udemy.com/algorithmic-stock-tradin...	True	95	1165	21	41	Beginner Level	
41	1035472	Python for Finance: Investment	https://www.udemy.com/python-for-	True	195	3811	278	103	All Levels	

```
top6 = temp.sort_values(by = 'num_subscribers',ascending=False).head(6)
```

```
top6
```

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content
2502	16646	Web Programming with Python	https://www.udemy.com/web-programming-with-pyt...	True	50	35267	217	53	All Levels	
2575	47963	Coding for Entrepreneurs: Learn Python, Django...	https://www.udemy.com/coding-for-entrepreneurs/	True	195	23412	799	251	All Levels	
2533	391546	Learn Python and Django: Payment Processing	https://www.udemy.com/learn-django-code-accept...	True	70	17714	198	23	All Levels	
3205	822444	Python and Django Full Stack Web	https://www.udemy.com/python-and-	True	200	11832	1883	191	All	

```
index = course_index['How To Maximize Your Profits Trading Options']
```

```
scores = list(enumerate(cosine_sim_mat[index]))
scores
```

```
[(0, 0.0),
 (1, 0.0),
 (2, 0.0),
 (3, 0.0),
 (4, 1.0),
 (5, 0.20412414523193154),
 (6, 0.20412414523193154),
 (7, 0.1889822365046136),
 (8, 0.3779644730092272),
 (9, 0.0),
 (10, 0.20412414523193154),
 (11, 0.5),
 (12, 0.0),
 (13, 0.17677669529663687),
 (14, 0.35355339059327373),
 (15, 0.0),
 (16, 0.0),
 (17, 0.1889822365046136),
 (18, 0.22360679774997896),
 (19, 0.0),
 (20, 0.30151134457776363),
 (21, 0.20412414523193154),
 (22, 0.1889822365046136),
 (23, 0.0),
 (24, 0.0),
 (25, 0.0),
 (26, 0.0),
 (27, 0.0),
 (28, 0.0),
 (29, 0.1889822365046136),
 (30, 0.35355339059327373),
 (31, 0.0),
 (32, 0.0),
 (33, 0.3779644730092272),
 (34, 0.0),
 (35, 0.3779644730092272),
 (36, 0.30151134457776363),
 (37, 0.0),
 (38, 0.0),
 (39, 0.0),
 (40, 0.0),
 (41, 0.0),
 (42, 0.0),
 (43, 0.5669467095138407),
 (44, 0.35355339059327373),
 (45, 0.0),
 (46, 0.35355339059327373),
 (47, 0.0),
 (48, 0.35355339059327373),
 (49, 0.4082482904638631),
 (50, 0.22360679774997896),
 (51, 0.0),
 (52, 0.1889822365046136),
 (53, 0.0),
 (54, 0.4082482904638631),
 (55, 0.0),
 (56, 0.0),
 (57, 0.20412414523193154),
```

```
sorted_score = sorted(scores,key = lambda x:x[1],reverse=True)
sorted_score
```

```
 (378, 0.5),
 (514, 0.5),
 (647, 0.5),
 (738, 0.5),
 (947, 0.5),
 (991, 0.5),
 (811, 0.45226701686664544),
```

```
(100, 0.4082482904638631),
(167, 0.4082482904638631),
(205, 0.4082482904638631),
(295, 0.4082482904638631),
(353, 0.4082482904638631),
(402, 0.4082482904638631),
(411, 0.4082482904638631),
(802, 0.4082482904638631),
(953, 0.4082482904638631),
(956, 0.4082482904638631),
(1002, 0.4082482904638631),
(8, 0.3779644730092272),
(33, 0.3779644730092272),
(35, 0.3779644730092272),
(102, 0.3779644730092272),
(109, 0.3779644730092272),
(113, 0.3779644730092272),
(157, 0.3779644730092272),
(186, 0.3779644730092272),
(363, 0.3779644730092272),
(434, 0.3779644730092272),
(510, 0.3779644730092272),
(650, 0.3779644730092272),
(798, 0.3779644730092272),
(900, 0.3779644730092272),
(14, 0.35355339059327373),
(30, 0.35355339059327373),
(44, 0.35355339059327373),
(46, 0.35355339059327373),
(48, 0.35355339059327373),
(75, 0.35355339059327373),
(89, 0.35355339059327373),
(149, 0.35355339059327373),
(153, 0.35355339059327373),
(200, 0.35355339059327373),
(346, 0.35355339059327373),
(361, 0.35355339059327373),
```

```python
sorted_indices = [i[0] for i in sorted_score[1:]]

sorted_values = [i[1] for i in sorted_score[1:]]

sorted_values
```

```
[0.5773502691896258,
 0.5669467095138407,
 0.5303300858899106,
 0.5303300858899106,
 0.5303300858899106,
 0.5303300858899106,
 0.5303300858899106,
 0.5,
 0.5,
 0.5,
 0.5,
 0.5,
 0.5,
 0.5,
 0.5,
 0.5,
 0.5,
 0.5,
 0.45226701686664544,
 0.4472135954999579,
 0.4472135954999579,
 0.4472135954999579,
 0.4472135954999579,
 0.4472135954999579,
 0.4472135954999579,
 0.4472135954999579,
 0.4330127018922194,
 0.4082482904638631,
 0.4082482904638631,
 0.4082482904638631,
 0.4082482904638631,
 0.4082482904638631,
 0.4082482904638631,
 0.4082482904638631,
 0.4082482904638631,
 0.4082482904638631,
 0.4082482904638631,
```

```
        0.4082482904638631,
        0.4082482904638631,
        0.4082482904638631,
        0.4082482904638631,
        0.4082482904638631,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
        0.3779644730092272,
```

```
recommended_result_df = df.iloc[sorted_indices]
```

```
recommended_result_df
```

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	c
410	889066	Trading Options Basics	https://www.udemy.com/trading-options-basics/	True	200	8	0	8	Beginner Level	
43	627540	Options Trading - How to Win with Weekly Options	https://www.udemy.com/work-from-home-setup-you...	True	115	7489	1190	25	Intermediate Level	
96	474928	Intermediate Options trading concepts for Stoc...	https://www.udemy.com/intermediate-options-tra...	True	40	2000	30	9	All Levels	
138	1243448	Forex Trading with Fixed 'Risk through Options...	https://www.udemy.com/forexoptions/	True	200	611	4	26	Beginner Level	
195	919906	Trading Options For Consistent Returns: Option...	https://www.udemy.com/trading-options-for-income/	False	Free	4077	281	20	Beginner Level	
...	...	...	...	...	...	...	...	...	...	

```
recommended_result_df['Similarity_Score'] = np.array(sorted_values)
```

```
recommended_result_df
```

```
<ipython-input-43-f632f55da6e8>: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/stable/user_guide/indexing.html#returning-a-view-versus-a-co
```
  recommended_result_df['Similarity_Score'] = np.array(sorted_values)
```

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	c
410	889066	Trading Options Basics	https://www.udemy.com/trading-options-basics/	True	200	8	0	8	Beginner Level	
43	627540	Options Trading - How to Win with Weekly Options	https://www.udemy.com/work-from-home-setup-you...	True	115	7489	1190	25	Intermediate Level	
96	474928	Intermediate Options trading concepts for Stoc...	https://www.udemy.com/intermediate-options-tra...	True	40	2000	30	9	All Levels	
138	1243448	Forex Trading with Fixed 'Risk through Options...	https://www.udemy.com/forexoptions/	True	200	611	4	26	Beginner Level	
195	919906	Trading Options For Consistent Returns: Option...	https://www.udemy.com/trading-options-for-income/	False	Free	4077	281	20	Beginner Level	
...	...	...	...	...	...	...	...	...	...	
3678	775618	Learn jQuery from Scratch - Master of JavaScri...	https://www.udemy.com/easy-jquery-for-beginner...	True	100	1040	14	21	All Levels	
3679	1088178	How To Design A WordPress	https://www.udemy.com/how-to-make-a-wordpress...	True	25	306	3	42	Beginner Level	

```
use_df = recommended_result_df[['Clean_title','Similarity_Score']]
use_df
```

	Clean_title	Similarity_Score
410	Trading Options Basics	0.577350
43	Options Trading Win Weekly Options	0.566947
96	Intermediate Options trading concepts Stocks O...	0.530330
138	Forex Trading Fixed Risk Options Trading	0.530330
195	Trading Options Consistent Returns Options Basics	0.530330
...	...	...
3678	Learn jQuery Scratch Master JavaScript library	0.000000
3679	Design WordPress Website Coding	0.000000
3680	Learn Build Polymer	0.000000
3681	CSS Animations Create Amazing Effects Website	0.000000
3682	MODX CMS Build Websites Beginners Guide	0.000000

3682 rows × 2 columns

```
def recommend_course(title,numrec = 10):

  course_index = pd.Series(
      df.index, index=df['course_title']).drop_duplicates()
```

```
        index = course_index[title]

        scores = list(enumerate(cosine_sim_mat [index]))

        sorted_scores = sorted(scores, key=lambda x: x[1], reverse=True)

        selected_course_index = [i[0] for i in sorted_scores[1:]]

        selected_course_score = [i[1] for i in sorted_scores[1:]]

        rec_df = df.iloc[selected_course_index]

        rec_df['Similarity_Score'] = selected_course_score

        final_recommended_courses = rec_df[[
            'course_title', 'Similarity_Score', 'url', 'price', 'num_subscribers']]

        return final_recommended_courses.head(numrec)
```

```
ans = recommend_course('Trading Options Basics',20)

ans
```

```
<ipython-input-49-8694092f04cb>:18: 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/stable/user_guide/indexing.html#returning-a-view-versus-a-cc
```
  rec_df['Similarity_Score'] = selected_course_score
```

	course_title	Similarity_Score	url	price	num_subscribers
97	Options Trading 101: The Basics	0.866025	https://www.udemy.com/options-trading-101-the-...	Free	1514
195	Trading Options For Consistent Returns: Option...	0.816497	https://www.udemy.com/trading-options-for-income/	Free	4077
864	Basics of Trading	0.816497	https://www.udemy.com/basicsoftrading/	40	99
66	Options Trading Basics (3-Course Bundle)	0.774597	https://www.udemy.com/learn-options-trading-co...	180	10100
802	Trading: Basics of Trading for Beginners	0.707107	https://www.udemy.com/money-basics-of-trading-...	140	1309
956	Options Basics & Trading With Small Capital! -...	0.707107	https://www.udemy.com/options-basics-trading-w...	30	120
43	Options Trading - How to Win with Weekly Options	0.654654	https://www.udemy.com/work-from-home-setup-you...	115	7489
96	Intermediate Options trading concepts for Stoc...	0.612372	https://www.udemy.com/intermediate-options-tra...	40	2000
138	Forex Trading with Fixed 'Risk through Options...	0.612372	https://www.udemy.com/forexoptions/	200	611
444	The Advantages of ETF Options and Index Option...	0.612372	https://www.udemy.com/learn-etf-options-and-in...	60	52
803	Options Spreads Bundle- the heart of Options ...	0.612372	https://www.udemy.com/options-spreads-explained/	120	623
4	How To Maximize Your Profits Trading Options	0.577350	https://www.udemy.com/how-to-maximize-your-pro...	200	1276
11	Trading Options With Money Flow	0.577350	https://www.udemy.com/trading-options-using-mo...	200	1380
68	How I Make Consistent Returns Trading Options	0.577350	https://www.udemy.com/how-i-make-15-per-month-...	200	2310
135	Forex Trading for Beginners - Basics	0.577350	https://www.udemy.com/complete-forex-trading-f...	50	989
330	Professional Options Trading, Simplified	0.577350	https://www.udemy.com/optionstrading/	200	631
378	Foundation of Options Trading and Investing	0.577350	https://www.udemy.com/options-trading-course-a...	20	6374
514	Winning Options Trading System	0.577350	https://www.udemy.com/winning-options-trading-...	105	367
647	Trading News Using Binary Options	0.577350	https://www.udemy.com/trading-news/	145	1186
738	Advanced Options Trading Course	0.577350	https://www.udemy.com/advanced-option-trading-...	40	1252

```
ans.shape
```

```
(20, 5)
```

```
ans.columns
```

```
Index(['course_title', 'Similarity_Score', 'url', 'price', 'num_subscribers'], dtype='object')
```

```
df.head()
```

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147	23	51	All Levels	
1	1113822	Complete GST Course & Certification - Grow You...	https://www.udemy.com/goods-and-services-tax/	True	75	2792	923	274	All Levels	
2	1006314	Financial Modeling for Business	https://www.udemy.com/financial-	True	45	2174	74	51	Intermediate	

```
df.to_csv('UdemyCleanedTitle.csv',index = None)
```

Start coding or generate with AI.