In [2]:
```python
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [3]:
```python
!pip install seaborn
```

```
Collecting seaborn
  Obtaining dependency information for seaborn from https://files.pythonhosted.org/packages/83/11/00d3c3dfc25ad54e731d91449895a79e4bf2384dc3ac01809010ba88f6d5/seaborn-0.13.2-py3-none-any.whl.metadata
  Downloading seaborn-0.13.2-py3-none-any.whl.metadata (5.4 kB)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from seaborn) (1.24.3)
Requirement already satisfied: pandas>=1.2 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from seaborn) (2.0.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from seaborn) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.1.0)
Requirement already satisfied: cycler>=0.10 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.41.1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (9.5.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: tzdata>=2022.1 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\srini\appdata\local\programs\python\python310\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
Downloading seaborn-0.13.2-py3-none-any.whl (294 kB)
   ---------------------------------------- 0.0/294.9 kB ? eta -:--:--
   ---------------------------------------- 0.0/294.9 kB ? eta -:--:--
   ---------------------------------------- 0.0/294.9 kB ? eta -:--:--
   ---------------------------------------- 0.0/294.9 kB ? eta -:--:--
   ---------------------------------------- 0.0/294.9 kB ? eta -:--:--
   - -------------------------------------- 10.2/294.9 kB ? eta -:--:--
   - -------------------------------------- 10.2/294.9 kB ? eta -:--:--
   ---- ----------------------------------- 30.7/294.9 kB 217.9 kB/s eta 0:00:02
   ----- ---------------------------------- 41.0/294.9 kB 245.8 kB/s eta 0:00:02
   ----- ---------------------------------- 41.0/294.9 kB 245.8 kB/s eta 0:00:02
   ---------- ----------------------------- 81.9/294.9 kB 305.0 kB/s eta 0:00:01
   ---------- ----------------------------- 81.9/294.9 kB 305.0 kB/s eta 0:00:01
   ------------ --------------------------- 92.2/294.9 kB 261.7 kB/s eta 0:00:01
   ------------ --------------------------- 92.2/294.9 kB 261.7 kB/s eta 0:00:01
   ------------ --------------------------- 92.2/294.9 kB 261.7 kB/s eta 0:00:01
   --------------- ------------------------ 112.6/294.9 kB 242.7 kB/s eta 0:00:01
   --------------- ------------------------ 112.6/294.9 kB 242.7 kB/s eta 0:00:01
   --------------- ------------------------ 112.6/294.9 kB 242.7 kB/s eta 0:00:01
   --------------- ------------------------ 112.6/294.9 kB 242.7 kB/s eta 0:00:01
   ---------------- ----------------------- 122.9/294.9 kB 184.7 kB/s eta 0:00:01
   ---------------- ----------------------- 122.9/294.9 kB 184.7 kB/s eta 0:00:01
   ---------------- ----------------------- 122.9/294.9 kB 184.7 kB/s eta 0:00:01
   ---------------- ----------------------- 122.9/294.9 kB 184.7 kB/s eta 0:00:01
   ------------------- -------------------- 143.4/294.9 kB 163.7 kB/s eta 0:00:01
   ---------------------- ----------------- 163.8/294.9 kB 182.0 kB/s eta 0:00:01
   ----------------------- ---------------- 174.1/294.9 kB 180.8 kB/s eta 0:00:01
   ----------------------- ---------------- 174.1/294.9 kB 180.8 kB/s eta 0:00:01
   -------------------------- ------------- 194.6/294.9 kB 184.3 kB/s eta 0:00:01
   -------------------------- ------------- 194.6/294.9 kB 184.3 kB/s eta 0:00:01
   -------------------------- ------------- 194.6/294.9 kB 184.3 kB/s eta 0:00:01
   -------------------------- ------------- 194.6/294.9 kB 184.3 kB/s eta 0:00:01
   --------------------------- ------------ 204.8/294.9 kB 166.0 kB/s eta 0:00:01
   --------------------------------- ------ 245.8/294.9 kB 190.7 kB/s eta 0:00:01
   --------------------------------- ------ 245.8/294.9 kB 190.7 kB/s eta 0:00:01
   --------------------------------- ------ 245.8/294.9 kB 190.7 kB/s eta 0:00:01
   ---------------------------------- ----- 256.0/294.9 kB 185.1 kB/s eta 0:00:01
   ---------------------------------- ----- 256.0/294.9 kB 185.1 kB/s eta 0:00:01
   ---------------------------------- ----- 256.0/294.9 kB 185.1 kB/s eta 0:00:01
   ---------------------------------- ----- 256.0/294.9 kB 185.1 kB/s eta 0:00:01
   ---------------------------------- ----- 256.0/294.9 kB 185.1 kB/s eta 0:00:01
   --------------------------------------- - 286.7/294.9 kB 173.4 kB/s eta 0:00:01
   ---------------------------------------- 294.9/294.9 kB 176.8 kB/s eta 0:00:00
Installing collected packages: seaborn
Successfully installed seaborn-0.13.2
[notice] A new release of pip is available: 23.2.1 -> 24.0
[notice] To update, run: python.exe -m pip install --upgrade pip
```

import seaborn as sns

In [4]:
```python
data = pd.read_csv(r"C:\Users\srini\rec\Coursera.csv")
data.head(5)
```

Out[4]:

	Course Name	University	Difficulty Level	Course Rating	Course URL	Course Description	Skills
0	Write A Feature Length Screenplay For Film Or ...	Michigan State University	Beginner	4.8	https://www.coursera.org/learn/write-a-feature...	Write a Full Length Feature Film Script In th...	Drama Comedy peering screenwriting film D...
1	Business Strategy: Business Model Canvas Analy...	Coursera Project Network	Beginner	4.8	https://www.coursera.org/learn/canvas-analysis...	By the end of this guided project, you will be...	Finance business plan persona (user experien...
2	Silicon Thin Film Solar Cells	École Polytechnique	Advanced	4.1	https://www.coursera.org/learn/silicon-thin-fi...	This course consists of a general presentation...	chemistry physics Solar Energy film lambda...
3	Finance for Managers	IESE Business School	Intermediate	4.8	https://www.coursera.org/learn/operational-fin...	When it comes to numbers, there is always more...	accounts receivable dupont analysis analysis...
4	Retrieve Data using Single-Table SQL Queries	Coursera Project Network	Beginner	4.6	https://www.coursera.org/learn/single-table-sq...	In this course you'll learn how to effectively...	Data Analysis select (sql) database manageme...

In [5]:
```python
data.shape
```

Out[5]: (3522, 7)

In [6]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3522 entries, 0 to 3521
Data columns (total 7 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Course Name         3522 non-null   object
 1   University          3522 non-null   object
 2   Difficulty Level    3522 non-null   object
 3   Course Rating       3522 non-null   object
 4   Course URL          3522 non-null   object
 5   Course Description  3522 non-null   object
 6   Skills              3522 non-null   object
dtypes: object(7)
memory usage: 192.7+ KB
```

In [7]:
```python
data.isnull().sum()
```

Out[7]:
```
Course Name           0
University            0
Difficulty Level      0
Course Rating         0
Course URL            0
Course Description    0
Skills                0
dtype: int64
```

In [8]:
```python
data['Difficulty Level'].value_counts()
```

Out[8]:
```
Difficulty Level
Beginner          1444
Advanced          1005
Intermediate       837
Conversant         186
Not Calibrated      50
Name: count, dtype: int64
```

In [10]:
```python
data['Skills'].value_counts()
```

Out[10]:
```
Skills
Google Cloud Platform  Big Data  Cloud Infrastructure  google storage  Cloud Storage  bigquery  Cloud Platforms  Kubernetes  Cloud Computing  Google App Engine information-technology cloud-computing    8
art  listening  emotions  Sound Design  music  signature  Audio Production  Writing  microphone  sound arts-and-humanities music-and-art    4
Python Programming  python libraries  analysis  semantics  Computer Programming  syntax  Data Analysis  Pandas  python syntax and semantics  Numpy data-science data-analysis    4
analysis  Machine Learning  Data Mining  Artificial Neural Networks  Regression Analysis  Data Analysis  data reporting  Human Learning  Regression  Big Data data-science data-analysis    4
Machine Learning  artificial general intelligence  application domain  ethics of artificial intelligence  Human Learning  language  Artificial Neural Networks  Computer Vision  ethics  Deep Learning data-science machine-learning    4
                                                                                                                ..
music  re-define  h.e.a.r.  publicity  Social Media  popular music  Planning  Payments  label  Audio Recording arts-and-humanities music-and-art    1
microbiology  nutrition  metagenomics  probiotics  probiotic  biology  food allergy  allergy  microbiota  thought life-sciences basic-science    1
Management Accounting  cost benefit analysis  Finance  Profit Analysis  Financial Accounting  Accounting  Leadership and Management  contribution margin  cost allocation  cost business finance    1
tableau software  Data Visualization Software  Computer Graphics  visualization software  Data Analysis  Data Visualization  software  Parallel Coordinates  graphs  analysis data-science data-analysis    1
bigtable  bigquery  SQL  Google Cloud Platform  role-based access control  Cloud Storage  Kubernetes  identity management  Cloud Computing  Cloud Platforms computer-science software-development    1
Name: count, Length: 3424, dtype: int64
```

In [11]:
```python
data['Course Rating'].value_counts()
```

Out[11]:
```
Course Rating
4.7               740
4.6               623
4.8               598
4.5               389
4.4               242
4.9               180
4.3               165
4.2               121
5                  90
4.1                85
Not Calibrated     82
4                  51
3.8                24
3.9                20
3.6                18
3.7                18
3.5                17
3.4                13
3                  12
3.2                 9
3.3                 6
2.9                 6
2.6                 2
2.8                 2
2.4                 2
1                   2
2                   1
3.1                 1
2.5                 1
1.9                 1
2.3                 1
Name: count, dtype: int64
```

In [12]:
```python
data['University'].value_counts()
```

Out[12]:
```
University
Coursera Project Network                      562
University of Illinois at Urbana-Champaign    138
Johns Hopkins University                      110
University of Michigan                        101
University of Colorado Boulder                101
                                             ...
GitLab                                          1
Yeshiva University                              1
University of Glasgow                           1
Laureate Education                              1
The World Bank Group                            1
Name: count, Length: 184, dtype: int64
```

In [13]:
```python
data['Course Name']
```

Out[13]:
```
0       Write A Feature Length Screenplay For Film Or ...
1       Business Strategy: Business Model Canvas Analy...
2                           Silicon Thin Film Solar Cells
3                                    Finance for Managers
4            Retrieve Data using Single-Table SQL Queries
                              ...
3517    Capstone: Retrieving, Processing, and Visualiz...
3518                     Patrick Henry: Forgotten Founder
3519    Business intelligence and data analytics: Gene...
3520                                  Rigid Body Dynamics
3521    Architecting with Google Kubernetes Engine: Pr...
Name: Course Name, Length: 3522, dtype: object
```

In [15]:
```python
data = data[['Course Name','Difficulty Level','Course Description','Skills']]
```

In [16]:
```python
data.head(8)
```

Out[16]:

	Course Name	Difficulty Level	Course Description	Skills
0	Write A Feature Length Screenplay For Film Or ...	Beginner	Write a Full Length Feature Film Script In th...	Drama Comedy peering screenwriting film D...
1	Business Strategy: Business Model Canvas Analy...	Beginner	By the end of this guided project, you will be...	Finance business plan persona (user experien...
2	Silicon Thin Film Solar Cells	Advanced	This course consists of a general presentation...	chemistry physics Solar Energy film lambda...
3	Finance for Managers	Intermediate	When it comes to numbers, there is always more...	accounts receivable dupont analysis analysis...
4	Retrieve Data using Single-Table SQL Queries	Beginner	In this course you'll learn how to effectively...	Data Analysis select (sql) database manageme...
5	Building Test Automation Framework using Selen...	Beginner	Selenium is one of the most widely used functi...	maintenance test case test automation scree...
6	Doing Business in China Capstone	Advanced	Doing Business in China Capstone enables you t...	marketing plan Planning Marketing consumpti...
7	Programming Languages, Part A	Intermediate	This course is an introduction to the basic co...	inference ml (programming language) higher-o...

In [17]:
```python
# Removing spaces between the words (Lambda funtions can be used as well)

data['Course Name'] = data['Course Name'].str.replace(' ',',')
data['Course Name'] = data['Course Name'].str.replace(',,',',')
data['Course Name'] = data['Course Name'].str.replace(':','')
data['Course Description'] = data['Course Description'].str.replace(' ',',')
data['Course Description'] = data['Course Description'].str.replace(',,',',')
data['Course Description'] = data['Course Description'].str.replace('_','')
data['Course Description'] = data['Course Description'].str.replace(':','')
data['Course Description'] = data['Course Description'].str.replace('(','')
data['Course Description'] = data['Course Description'].str.replace(')','')

#removing paranthesis from skills columns
data['Skills'] = data['Skills'].str.replace('(','')
data['Skills'] = data['Skills'].str.replace(')','')
```

In [18]:
```python
data.head(7)
```

Out[18]:

	Course Name	Difficulty Level	Course Description	Skills
0	Write,A,Feature,Length,Screenplay,For,Film,Or,...	Beginner	Write,a,Full,Length,Feature,Film,Script,In,thi...	Drama Comedy peering screenwriting film D...
1	Business,Strategy,Business,Model,Canvas,Analys...	Beginner	By,the,end,of,this,guided,project,you,will,be,...	Finance business plan persona user experienc...
2	Silicon,Thin,Film,Solar,Cells	Advanced	This,course,consists,of,a,general,presentation...	chemistry physics Solar Energy film lambda...
3	Finance,for,Managers	Intermediate	When,it,comes,to,numbers,there,is,always,more,...	accounts receivable dupont analysis analysis...
4	Retrieve,Data,using,Single-Table,SQL,Queries	Beginner	In,this,course,you'll,learn,how,to,effectively...	Data Analysis select sql database management...
5	Building,Test,Automation,Framework,using,Selen...	Beginner	Selenium,is,one,of,the,most,widely,used,functi...	maintenance test case test automation scree...
6	Doing,Business,in,China,Capstone	Advanced	Doing,Business,in,China,Capstone,enables,you,t...	marketing plan Planning Marketing consumpti...

In [19]:
```python
data['tags'] = data['Course Name'] + data['Difficulty Level'] + data['Course Description'] + data['Skills']
```

In [20]:
```python
data.head(10)
```

Out[20]:

	Course Name	Difficulty Level	Course Description	Skills	tags
0	Write,A,Feature,Length,Screenplay,For,Film,Or,...	Beginner	Write,a,Full,Length,Feature,Film,Script,In,thi...	Drama Comedy peering screenwriting film D...	Write,A,Feature,Length,Screenplay,For,Film,Or,...
1	Business,Strategy,Business,Model,Canvas,Analys...	Beginner	By,the,end,of,this,guided,project,you,will,be,...	Finance business plan persona user experienc...	Business,Strategy,Business,Model,Canvas,Analys...
2	Silicon,Thin,Film,Solar,Cells	Advanced	This,course,consists,of,a,general,presentation...	chemistry physics Solar Energy film lambda...	Silicon,Thin,Film,Solar,CellsAdvancedThis,cour...
3	Finance,for,Managers	Intermediate	When,it,comes,to,numbers,there,is,always,more,...	accounts receivable dupont analysis analysis...	Finance,for,ManagersIntermediateWhen,it,comes,...
4	Retrieve,Data,using,Single-Table,SQL,Queries	Beginner	In,this,course,you'll,learn,how,to,effectively...	Data Analysis select sql database management...	Retrieve,Data,using,Single-Table,SQL,QueriesBe...
5	Building,Test,Automation,Framework,using,Selen...	Beginner	Selenium,is,one,of,the,most,widely,used,functi...	maintenance test case test automation scree...	Building,Test,Automation,Framework,using,Selen...
6	Doing,Business,in,China,Capstone	Advanced	Doing,Business,in,China,Capstone,enables,you,t...	marketing plan Planning Marketing consumpti...	Doing,Business,in,China,CapstoneAdvancedDoing,...
7	Programming,Languages,Part,A	Intermediate	This,course,is,an,introduction,to,the,basic,co...	inference ml programming language higher-ord...	Programming,Languages,Part,AIntermediateThis,c...
8	The,Roles,and,Responsibilities,of,Nonprofit,Bo...	Intermediate	This,course,provides,a,more,in-depth,look,at,t...	Planning Peer Review fundraising strategic ...	The,Roles,and,Responsibilities,of,Nonprofit,Bo...
9	Business,Russian,Communication.,Part,3	Intermediate	Russian,is,considered,to,be,one,of,the,most,di...	Russian market economics tax exemption coop...	Business,Russian,Communication.,Part,3Intermed...

In [21]:
```python
data['tags'].iloc[1]
```

Out[21]:
```
'Business,Strategy,Business,Model,Canvas,Analysis,with,MiroBeginnerBy,the,end,of,this,guided,project,you,will,be,fluent,in,identifying,and,creating,Business,Model,Canvas,solutions,based,on,previous,high-level,analyses,and,research,da
ta.,This,will,enable,you,to,identify,and,map,the,elements,required,for,new,products,and,services.,Furthermore,it,is,essential,for,generating,positive,results,for,your,business,venture.,This,guided,project,is,designed,to,engage,and,ha
rness,your,visionary,and,exploratory,abilities.,You,will,use,proven,models,in,strategy,and,product,development,with,the,Miro,platform,to,explore,and,analyse,your,business,propositions.,,We,will,practice,critically,examining,results,f
rom,previous,analysis,and,research,results,in,deriving,the,values,for,each,of,the,business,model,sections.Finance  business plan  persona user experience  business model canvas  Planning  Business  project  Product Development  prese
ntation  Strategy business business-strategy'
```

In [22]:
```python
new_df = data[['Course Name','tags']]
new_df.head(10)
```

Out[22]:

	Course Name	tags
0	Write,A,Feature,Length,Screenplay,For,Film,Or,...	Write,A,Feature,Length,Screenplay,For,Film,Or,...
1	Business,Strategy,Business,Model,Canvas,Analys...	Business,Strategy,Business,Model,Canvas,Analys...
2	Silicon,Thin,Film,Solar,Cells	Silicon,Thin,Film,Solar,CellsAdvancedThis,cour...
3	Finance,for,Managers	Finance,for,ManagersIntermediateWhen,it,comes,...
4	Retrieve,Data,using,Single-Table,SQL,Queries	Retrieve,Data,using,Single-Table,SQL,QueriesBe...
5	Building,Test,Automation,Framework,using,Selen...	Building,Test,Automation,Framework,using,Selen...
6	Doing,Business,in,China,Capstone	Doing,Business,in,China,CapstoneAdvancedDoing,...
7	Programming,Languages,Part,A	Programming,Languages,Part,AIntermediateThis,c...
8	The,Roles,and,Responsibilities,of,Nonprofit,Bo...	The,Roles,and,Responsibilities,of,Nonprofit,Bo...
9	Business,Russian,Communication.,Part,3	Business,Russian,Communication.,Part,3Intermed...

In [23]:
```python
new_df['tags'] = data['tags'].str.replace(',',' ')
```

```
C:\Users\srini\AppData\Local\Temp\ipykernel_31524\3366302733.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/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df['tags'] = data['tags'].str.replace(',',' ')
```

In [24]:
```python
new_df.rename(columns = {'Course Name':'course_name'}, inplace = True)
```

```
C:\Users\srini\AppData\Local\Temp\ipykernel_31524\2809988695.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df.rename(columns = {'Course Name':'course_name'}, inplace = True)
```

In [25]:
```python
new_df['tags'] = new_df['tags'].apply(lambda x:x.lower()) #lower casing the tags column
```

```
C:\Users\srini\AppData\Local\Temp\ipykernel_31524\172956468.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/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df['tags'] = new_df['tags'].apply(lambda x:x.lower()) #lower casing the tags column
```

In [26]:
```python
new_df.head(10)
```

Out[26]:

	course_name	tags
0	Write,A,Feature,Length,Screenplay,For,Film,Or,...	write a feature length screenplay for film or ...
1	Business,Strategy,Business,Model,Canvas,Analys...	business strategy business model canvas analys...
2	Silicon,Thin,Film,Solar,Cells	silicon thin film solar cellsadvancedthis cour...
3	Finance,for,Managers	finance for managersintermediatewhen it comes ...
4	Retrieve,Data,using,Single-Table,SQL,Queries	retrieve data using single-table sql queriesbe...
5	Building,Test,Automation,Framework,using,Selen...	building test automation framework using selen...
6	Doing,Business,in,China,Capstone	doing business in china capstoneadvanceddoing ...
7	Programming,Languages,Part,A	programming languages part aintermediatethis c...
8	The,Roles,and,Responsibilities,of,Nonprofit,Bo...	the roles and responsibilities of nonprofit bo...
9	Business,Russian,Communication.,Part,3	business russian communication. part 3intermed...

In [27]:
```python
new_df.shape
```

Out[27]: (3522, 2)

In [28]:
```python
#TEXT Vectorization
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000,stop_words='english')
vectors = cv.fit_transform(new_df['tags']).toarray()
```

In [29]:
```python
#STEMMMING process
import nltk #for stemming process
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
#defining the stemming function
def stem(text):
    y=[]

    for i in text.split():
        y.append(ps.stem(i))

    return " ".join(y)
```

In [30]:
```python
new_df['tags'] = new_df['tags'].apply(stem)
```

```
C:\Users\srini\AppData\Local\Temp\ipykernel_31524\3213734980.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/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df['tags'] = new_df['tags'].apply(stem)
```

In [31]:
```python
#similarity measure
from sklearn.metrics.pairwise import cosine_similarity
similarity = cosine_similarity(vectors)
```

In [52]:
```python
# Assuming you have already defined your DataFrame new_df and similarity matrix similarity

# Define your recommend function
def recommend(course):
    # Find the index of the course in the DataFrame
    course_index = new_df[new_df['course_name'] == course].index[0]

    # Retrieve distances/similarities for the given course
    distances = similarity[course_index]

    # Sort distances/similarities and get the top 6 (excluding the course itself)
    course_list = sorted(list(enumerate(distances)), reverse=True, key=lambda x: x[1])[1:7]

    # Print the recommended courses
    for i in course_list:
        print(new_df.iloc[i[0]].course_name)
```

In [54]:
```python
# Test the recommend function
recommend('Finance,for,Managers')
```

```
Fundamentals,of,financial,and,management,accounting
Introduction,to,Finance,The,Basics
The,Language,and,Tools,of,Financial,Analysis
Corporate,finance,Know,your,numbers,2
Finance,for,Non-Financial,Professionals
Operations,Management,Analysis,and,Improvement,Methods
```

In [55]:
```python
recommend('Business,Strategy,Business,Model,Canvas,Analysis,with,Miro')
```

```
Product,Development,Customer,Persona,Development,with,Miro
Product,and,Service,Development,Empathy,Mapping,with,Miro
Product,Development,Customer,Journey,Mapping,with,Miro
Analyzing,Macro-Environmental,Factors,Using,Creately
Business,Strategy,in,Practice,(Project-centered,Course)
Innovating,with,the,Business,Model,Canvas
```