Step- 1 :import Libraries

```
In [1]: import pandas as pd
```

Import and read data

```
In [2]: data = pd.read_csv("E:\SATHVIKA\coapps\wiki_movie_plots_deduped.csv")
```

df.head()

```
In [3]: data.head()
```

Out[3]:

	Release Year	Title	Origin/Ethnicity	Director	Cast	Genre	
0	1901	Kansas Saloon Smashers	American	Unknown	NaN	unknown	https://en.wikipedia.org/wiki/Ka
1	1901	Love by the Light of the Moon	American	Unknown	NaN	unknown	https://en.wikipedia.org/wiki/L
2	1901	The Martyred Presidents	American	Unknown	NaN	unknown	https://en.wikipedia.org/wiki/T
3	1901	Terrible Teddy, the Grizzly King	American	Unknown	NaN	unknown	https://en.wikipedia.org/wik
4	1902	Jack and the Beanstalk	American	George S. Fleming, Edwin S. Porter	NaN	unknown	https://en.wikipedia.org/wiki/Ja

df.Shape()

```
In [4]: data.shape
```

Out[4]: (34886, 8)

Step 2: Choose only some portion of data

In [5]:
```python
data2 = data.loc[(data['Origin/Ethnicity']=='Telugu') & (data['Release Year
len(data2)
```

Out[5]: 104

In [6]:
```python
my_data = pd.DataFrame(data2)
```

In [7]:
```python
my_data.tail()
```

Out[7]:

	Release Year	Title	Origin/Ethnicity	Director	Cast	Genre	
32797	2017	Logesh Kanagaraj	Telugu	Unknown	Sundeep Kishan, Sri, Regina Cassandra	unknown	https:
32798	2017	Doss Ramaswamy	Telugu	Unknown	Nayanthara, Sulile Kumar, Harish Uthaman, Tham...	unknown	https://en.
32799	2017	James Erskine	Telugu	Unknown	Sachin Tendulkar, Arjun Tendulkar, Mahendra Si...	unknown	https://en.v
32800	2017	ARK Saravanan	Telugu	Unknown	Aadhi, Nikki Galrani, Anandaraj	unknown	https://en.wikipe
32801	2017	Ravi Udyawar	Telugu	Unknown	Sridevi, Nawazuddin Siddiqui, Akshaye Khanna, ...	unknown	http

In [8]:
```python
finaldata = my_data[["Title", "Plot"]]
```

In [9]:
```python
finaldata = finaldata.set_index('Title')
```

```
In [10]: finaldata.head(50)
```

Title	Plot
Chitram Bhalare Vichitram	Raja (Naresh), Sudhakar (Subhalekha Sudhakar),...
Abbayitho Ammayi	Abhi (Naga Shourya) is a desperate youth who w...
Nenu Sailaja	The film starts with Hari (Ram Pothineni) sitt...
Nannaku Prematho	Abhiram (Jr. NTR), the youngest son of London-...
Express Raja	The film begins with the introduction of Inumu...
Dictator	Chandu (Balakrishna Nandamuri) works at Dharma...
Soggade Chinni Nayana	Ram "Ramu" Mohan is a socially-inept cardiolog...
Lacchimdeviki O Lekkundi	Devi (Lavanya) works in a bank as cashier, and...
Seethamma Andalu Ramayya Sitralu	Sri Ram (Raj Tarun) is a young village dude, w...
Speedunnodu	Sobhan (Bellamkonda Sreenivas) is a son of pre...
Vennello Hai Hai	Susheel(Ajmal) is a software engineer working ...
Krishna Gadi Veera Prema Gaadha	Krishna (Nani) is a soft natured person who is...
Garam	Varala Babu (Aadi) who is poor at studies woul...
Malupu	A young woman tries to call the emergency hotl...
Kshanam	Rishi, a San Francisco-based investment banker...
Eluka Majaka	Balu (Vennela Kishore) falls in love with Swap...
Guntur Talkies	Hari (Siddu) and Giri (Naresh) live in a slum ...
Shourya	Shourya (Manchu Manoj) and Netra (Regina Cassa...
Tuntari	When a group of five IT professionals decide t...
Tulasi Dalam	Satvik (Nishchal Deva) who is an orphan moving...
Seesa	Seesa is a thriller film all about human emoti...
Run	Sanjay (Sundeep Kishan) loses his job. He is m...
Oopiri	Vikramadhitya is a wealthy entrepreneur who ow...
Attack	Guru Raj (Prakash Raj) the man of the masses, ...
Savitri	Savitri (Nanditha) is crazy about getting marr...
Sardaar Gabbar Singh	The story tells us about gabbar singh[pawan ka...
Eedo Rakam Aado Rakam	The story runs between two good friends Arjun ...
Sarrainodu	Gana (Allu Arjun) is a youth who beats crooked...
Raja Cheyyi Vesthe	Raja ram (Nara rohit) is a aspiring film maker...
Supreme	Jagruthi Foundation was founded by a royal dyn...
Brahmotsavam	The movie starts with Ajay's (Mahesh Babu) ent...
A Aa	A Aa revolves around the lives of 23-year old ...
Right Right	An accident interlinks the life of a driver an...
Gentleman	Aishwarya, the daughter of a wealthy entrepren...
Oka Manasu	Sandhya (Niharika Konidela), a young medical s...
Nayaki	Nayaki is a horror movie . Nayaki is about Gay...
Pellichoopulu	Prasanth (Vijay Deverakonda) goes to Chitra's ...
Jakkanna	A man's (Sunil) attempts to repay a kindness e...

Title	Plot
Manamantha	Manamantha follows the lives of 4 people from ...
Srirastu Subhamastu	Sirish (Allu Sirish) falls in love with Ananya...
Babu Bangaram	ACP Krishna (Venkatesh) who is a kind-hearted ...
Thikka	The film is a confusion kidnap drama; Aditya (...
Janatha Garage	Sathyam (Mohanlal) is a small workshop owner i...
Nirmala Convent	Samuel (Roshan Meka), son of a poor farmer, Da...
Majnu	The plot revolves around Aditya (Nani), who wo...
Hyper	Surya (Ram) is a carefree youth who lives with...
Abhinetri	Krishna Kumar (Prabhu Deva) is a happy go luck...
Mana Oori Ramayanam	The film is set against the backdrop of Rama N...
Premam	The film opens in the year 2000, where Vikram ...
Ism	Javed Ibrahim (Jagapati Babu), is a high-profi...

```
In [11]: finaldata["Plot"]["Malupu"]
```

Out[11]: "A young woman tries to call the emergency hotline to save her before she is murdered in a flat in Chennai. Someone is shown covering up the crime.\r\nSaga (Aadhi), a regular, middle-class college student from Chennai arrives in Mumbai to meet Mudaliar, the city's underworld don and protector of the poor. At around the same time, Mudaliar is rushed to the hospital after he is almost killed in an assassination attempt. A mob assembles before his house and demands he show up to prove he is alive. When he finally does appear, the crowd is grateful for his survival and pray for his well-being.\r\nSaga and his three wealthy friends purposely skip their finals so that they can continue to stay in college and enjoy themselves rather than face adulthood. Despite being warned by his parents to stay away from his friends as they are a bad influence on him, Saga hangs out with them. He finally decides to turn over a new leaf after Kayal (Nikki Galrani), the girl he falls in love with, convinces his father that Saga can be a responsible adult if given the chance. His parents then put him in charge of his sister's wedding arrangements before leaving for their hometown to meet relatives.\r\nWith his family out of town, Saga follows his three friends for a night in the city to celebrate New Year's Eve. At a restaurant, two of his friends are drunk and get into a fight with a girl named Priya and her boyfriend, Surya. When the police come to resolve the ruckus, they realise that Saga's third friend is the son of the police commissioner and decides to arrest Surya instead. Priya protests and is publicly humiliated by the police inspector, without knowing who she is. She then warns Saga and his friends to go into hiding as they will not live to see the light of day.\r\nBy midnight, Saga and his friends come to know that Priya is actually the daughter of Mudaliar. Her brother, Guna, who runs Mudaliar's businesses in Chennai is after them for revenge. Saga's friends try to convince him to follow them into hiding as arranged by their wealthy and influential fathers, but he insists on staying to arrange for his sister's wedding. Later that morning, Saga is chased down by Guna's men while in town. He is finally saved by Deva, Guna's henchman who realises that Saga saved his brother's life once. Deva advises Saga to go to Mumbai to meet Mudaliar in person and explain everything since he is a sensible man, unlike Guna.\r\nSaga goes to Mudaliar's home and claims that Guna sent him. When it is discovered he is lying, he is tortured since he is suspected for the assassination attempt on Mudaliar. Saga finally gets to meet Mudaliar and explains the truth of what happened that night in Chennai. However, Mudaliar drops a bombshell by saying that after Priya bailed out Surya from the police station that night, both of them went missing and have not been seen since. Since Saga and his friends were the last to see the couple alive, they are the prime suspects. Saga refuses to reveal his friends' whereabouts until Mudaliar threatens to have his family killed. Mudaliar then gives Saga some time to find his friends or else his family will die.\r\nWith Kayal and Deva's help, Saga finds out where his friends are hiding. In between, Saga is being attacked by a person in a purdah. As Saga cannot see his face, he is unable to catch his attacker. Saga goes into hiding as he suspects that he is being targeted by the father of one of his friends, in order to keep their identities safe. The attacker then tries to kill Deva and runs away. In the process, he begins to suspect that their fathers are trying to frame him for Priya's disappearance to save their sons. This forces Saga to reveal his friends' hideouts to Mudaliar, who releases his family. Saga then visits Deva in the hospital who is receiving treatment. Deva hands Saga a paper that has a symbol and says the attacker had this symbol tattooed in his hand. After much thinking, Saga realises that he saw such a tattoo on Surya's forearm and he might know about Priya's whereabouts and goes to his flat. Once there, he finds Surya's frozen body being kept by Surya's psychotic ex-girlfriend, Nila who also has the same tattoo on her hand. It is then revealed that she killed both Surya and Priya on New Year's Eve out of jealousy. Before Saga can take Nila to meet Mudaliar, she is killed by Deva's brother, leaving Saga with no alternative to convince the don of the truth. At around the same time, Mudaliar and Guna find Priya's body washed

up on the beach.\r\nOnce his men have rounded up Saga and his friends, Mud
aliar orders Guna and their men to kill them. However, the friend who trie
d to hit on Priya and caused the whole misunderstanding finally admits his
mistake to Mudaliar and pleads for his friends' release in return for his
life. Moved by the young man's deed, Mudaliar lets the four friends go as
he realises they couldn't have killed his daughter as they were willing to
sacrifice their life for their friends."

Step 3: Preprocess the data

In [12]:
```python
import nltk
```

In [13]:
```python
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\home\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Out[13]: True

In [14]:
```python
nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     C:\Users\home\AppData\Roaming\nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
```

Out[14]: True

In [15]:
```python
nltk.download('wordnet')
```

```
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\home\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

Out[15]: True

In [16]:
```python
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\home\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[16]: True

In [17]:
```python
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords
lemmatizer = WordNetLemmatizer()
```

```
In [18]:   def preprocess_sentences(text):
               text = text.lower()
               words = nltk.word_tokenize(text)
               my_sent=[lemmatizer.lemmatize(word) for word in words if word not in st
               finalsent = ' '.join(my_sent)

               finalsent = finalsent.replace("n't", " not")
               finalsent = finalsent.replace("'m", " am")
               finalsent = finalsent.replace("'s", " is")
               finalsent = finalsent.replace("'re", " are")
               finalsent = finalsent.replace("'ll", " will")
               finalsent = finalsent.replace("'ve", " have")
               finalsent = finalsent.replace("'d", " would")

               return finalsent

           finaldata["new_plot"]= finaldata["Plot"].apply(preprocess_sentences)
           finaldata.head()
```

Out[18]:

Title	Plot	new_plot
Chitram Bhalare Vichitram	Raja (Naresh), Sudhakar (Subhalekha Sudhakar),...	raja (naresh) , sudhakar (subhalekha sudhak...
Abbayitho Ammayi	Abhi (Naga Shourya) is a desperate youth who w...	abhi (naga shourya) desperate youth want fun...
Nenu Sailaja	The film starts with Hari (Ram Pothineni) sitt...	film start hari (ram pothineni) sitting beac...
Nannaku Prematho	Abhiram (Jr. NTR), the youngest son of London-...	abhiram (jr. ntr) , youngest son london-base...
Express Raja	The film begins with the introduction of Inumu...	film begin introduction inumu (dhanraj) , pe...

```
In [19]:   !pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in c:\users\home\anaconda3\lib
\site-packages (0.24.1)
Requirement already satisfied: numpy>=1.13.3 in c:\users\home\anaconda3\li
b\site-packages (from scikit-learn) (1.20.1)
Requirement already satisfied: scipy>=0.19.1 in c:\users\home\anaconda3\li
b\site-packages (from scikit-learn) (1.6.2)
Requirement already satisfied: joblib>=0.11 in c:\users\home\anaconda3\lib
\site-packages (from scikit-learn) (1.0.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\home\anaco
nda3\lib\site-packages (from scikit-learn) (2.1.0)
```

Step 4: Use TF-IDF to vectorize preprocessed movieplots

```
In [20]:   from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [21]: tfidf = TfidfVectorizer()
         tfidf_movieid = tfidf.fit_transform((finaldata["new_plot"]))
```

```
In [22]: from sklearn.metrics.pairwise import cosine_similarity
         similarity = cosine_similarity(tfidf_movieid, tfidf_movieid)
```

```
In [23]: indices = pd.Series(finaldata.index)
         def recommendations(title, cosine_sim = similarity):
           try:
             index = indices[indices == title].index[0]
             print(index)
             similarity_scores = pd.Series(cosine_sim[index]).sort_values(ascending
             print(similarity_scores)
             top_10_movies = list(similarity_scores.iloc[1:11].index)
             print(top_10_movies)
             recommended_movies = [list(finaldata.index)[i] for i in top_10_movies]
             return recommended_movies
           except:
             print("No movie name found")
```

Step 5: Final predictions

```
In [24]: recommendations("Malupu")
```

```
13
13    1.000000
72    0.078639
73    0.074301
30    0.069783
34    0.063669
        ...
83    0.010938
94    0.010430
32    0.005092
20    0.004148
57    0.002554
Length: 104, dtype: float64
[72, 73, 30, 34, 9, 99, 74, 42, 50, 98]
```

```
Out[24]: ['Keshava',
          'Rarandoi Veduka Chudham',
          'Brahmotsavam',
          'Oka Manasu',
          'Speedunnodu',
          'Logesh Kanagaraj',
          'Andhhagadu',
          'Janatha Garage',
          'Sahasam Swasaga Sagipo',
          'Jay. K']
```

```
In [26]: recommendations("Express Raja")
```

```
         4
4          1.000000
92         0.234612
21         0.164629
28         0.155028
72         0.119526
              ...
25         0.006935
64         0.006551
35         0.005901
19         0.001605
32         0.000000
Length: 104, dtype: float64
[92, 21, 28, 72, 43, 100, 16, 84, 65, 59]
```

```
Out[26]: ['Raja the Great',
          'Run',
          'Raja Cheyyi Vesthe',
          'Keshava',
          'Nirmala Convent',
          'Doss Ramaswamy',
          'Guntur Talkies',
          'Anando Brahma',
          'Winner',
          'Khaidi No. 150']
```