A MajorProject Report on

MOVIE RECOMMENDATION SYSTEM USING MACHINE LEARNING

Submitted by

V.S.PAVITHRA -R180258
M.SOUMYA -R180919

Submitted to

IIIT RK Valley Idupulapaya, Vempalli, YSR
Kadapa, Andhra Pradesh, India PIN 516330.

Under the guidance of

Ms. V.SRAVANI

Assistant Professor

as a part of

Partial fulfillment of the degree of Bachelor of Technology in Computer Science and Engineering

CERTIFICATE

This is to certify that the report entitled “MOVIE RECOMMENDATION SYSTEM USING MACHINE LEARNING” submitted by V.S.Pavithra bearing ID.No. R180258 and M.Soumya bearing ID.No R180919 in partialfulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out by them under my supervision andguidance.The report has not been submitted previously in part or in full to this or another University or Institution for the award of any degree or diploma.

Project Guide,

V.Sravani,
Assistant Professor,
Dept of CSE,
RGUKT,RK.VALLEY.

Head of Department,

N.Sathyanandaram,
Dept of CSE,
RGUKT,RK.VALLEY.

Acknowledgement

I would like to express my sincere gratitude to **Ms.V.SRAVANI** , my project internal guide for valuable suggestions and keen interest throughout the progress of my course of research.I am grateful to **N.Sathyanandaram** , HOD CSE , for providing excellent computing facilities and a congenial atmosphere for progressing with my project.At the outset, I would like to thank Rajiv Gandhi University of Knowledge Technologies , for providing all the necessary resources for the successful completion of my course work.

ABSTRACT

The rapid growth of digital platforms and streaming services has led to an overwhelming abundance of movies, making it challenging for users to discover content that aligns with their preferences. This project proposes the development of an Intelligent Movie Recommendation System to address this issue and enhance the user experience in navigating the vast landscape of available films.

This analysis mainly focuses on the domain of Machine Learning example of Movie Recommendation System with a approach of finding the similarity scores between the two contents in the content based filtering. It helps us in finding the distance between the two vectors and their angle by the help of the cosine similarity formula and their magnitude of relative scores. In this model, we have considered two texts and plotted them in the form of a graph from where we have plotted the points taking two-dimension plane of x-y plane only.We have taken in consideration two fields (i.e.) in xplane we have considered the word London of the text whereas in y-plane we have considered Paris word of the texts. This model analysis the points to find the two texts similarity scores by using Python shell and distance between two vector model approach more effectively and precisely..

Many of the recommendation systems we are seeing today in our environment such as in the YouTube for example if I see lots of news regarding the GK and Current Affairs then it offers me the related videos according to it with different subscribers. It gains popularity by application rating and at the same time enhances the customer experience. This policy of recommendation system is really helpful in giving optimum results to an application profitability and to make the organisation more connected. We can also see the recommendation work in online food applications such as Zomato, Food Panda and Swiggy which offers their customers the food restaurants which supplies their taste food. They learn upon the behaviour of the customer from the previous orders and tries to impress them with the latest add-ons of their favourite cuisines and stuffs.

Table Of Contents

1.Introduction

Machine Learning is that field of study that offers computers the aptitude to find out while not being expressly programmed. ML is one of the foremost exciting technologies that one would have ever stumbled upon because it is obvious from the name, it provides the pc that creates it additional like humans: the power to learn. Machine Learning is actively getting used these days, maybe in many places than one would expect. Machine Learning is employed in net search engines, email filters to delineated spam, websites to create individualized recommendations, banking software systems to sight uncommon transactions, and much of apps on our phones like voice recognition.

Recommender systems are systems that are designed to suggest things to the user that support many alternative factors. These systems predict the foremost possible product that the users are possibly to buy and are of interest to firms like Netflix, Amazon etc. use recommender systems to assist their users to spot the right product or movies for them. The recommender system deals with an outsized volume data present by filtering the foremost necessary information supported by the information provided by a user’s preference and interest. It finds out the match between user and item and imputes the similarities between users and ratings for recommendation. Both the users and the services provided have benefited from these sorts of systems, the standard and decision-making method has additionally improved through these sorts of systems.

Recommender systems used in a various form of areas together with movies, music, news, books, analysis articles, search queries, social tags, and merchandise normally. Recommendation System is a filtration program whose prime goal is to predict the movie to a user towards a domain specific item. In our case, this domain-specific item is a movie, so the most focus of our recommendation system is to filter and predict solely those movies that a user would favor given some information concerning the user him or herself. There are many alternative ways that to create movie recommendation system however we've selected the content base recommender system in order that user will simply get the foremost similar movies on the user's interest. As our recommender system recommends the top high five movies as like movie that user is selected

1.1.Problem Statement:

The goal of the project is to recommend a movie to the user.Providing related content out of relevant and irrelevant collection of items tousers of online service providers.

1.2.Objective of the Projects:

- Improving the Accuracy of the recommendation system
- Improve the Quality of the movie Recommendation system
- Improving the Scalability.
- Enhancing the user experience

1.3.Methodology:

Content-Based Filtering:

Content-based filtering methods are done based on user characteristics. This method is used in situations where data is known on an item such as name, location, or description and not on the user. It predicts the items based on user's information and completely ignores contributions from other users as with the case of collaborative techniques. It uses the data that is provided by the user either explicitly or implicitly. When the user provides more content-based filtering mechanisms actions on the recommendations such as content- based recommender the engine becomes more and more accurate.

1. Collecting the data sets:

Collecting all the required data set from Kaggle web site in this project we require movie.csv, ratings.csv,users.csv.

2. Data Analysis:

Make sure that that the collected data sets are correct and analyzing the data in the csv files i.e. checking whether all the column Felds are present in the data sets.

3. Algorithms: in our project we have only two algorithms one is cosine similarity and other is single valued decomposition are used to build the machine learning recommendation model.

4. Training and Testing the model: once the implementation of algorithm is completed. we have to train the model to get the result. We have tested it several times the model is recommend different set of movies to different users.

5. Improvements in the project: In the later stage we can implement different algorithms and methods for better recommendation.

2.DETAILED STUDY OF RECOMMENDATION SYSTEMS:

2.1 Types of Recommendation System

Systems that propose things, create playlists, find matches, and do much more are known as recommender systems. User-item interactions and characteristic information are key components of recommender systems' operation. Information about the user and the items constitutes characteristic information, whereas information about user-item interactions includes ratings, the volume of purchases, user likes, and many other things. Based on this, a collaborative filtering, content-based filtering, or hybrid filtering approach can be used to create the recommendation system.

Collaborative Filtering:

Filtering by collaboration. This algorithm finds people who share similar tastes and uses their feedback to suggest the same to a different person who has those interests. Utilizing data from rating profiles for various persons or things it creates suggestions. It has been incorporated into a variety of programmes, including Spotify, Netflix, and YouTube. It is a common strategy and a component of the hybrid system.

Content-Based Filtering:

User attributes are used to inform content-based filtering techniques. When information about an item, such as its name, location, or description, but not about the user, is known, this method is employed. As with collaborative approaches, it entirely disregards the input from other users and guesses the things based on the user's information. It makes either explicit or implicit use of the user-provided data. The accuracy of the engine increases as more content-based filtering mechanisms, such as content-based recommenders, are provided by the user.

Hybrid Annroach:

Combining collaborative filtering with content-based filtering or any other strategy is known as a hybrid approach. By establishing independent forecasts for content-based and collaborative-based approaches before integrating them, hybrid systems can be put into practice. It improves the recommender systems' performance and accuracy.

This project is made with the approach of Content-Based Filtering.

3.System Requirements Specification

This involves both the hardware and software requirements needed for the project and detailed explanation of the specifications.

3.1 Hardware Requirements

1. A PC with Windows/Linux OS
2. Processor with 1.7-2.4gHz speed
3. Minimum of 8gb RAM
4. 2gb Graphic card

3.2 Software Requirements

1. Jupyter notebook
2. Python libraries- Numpy, Pandas, Scikit-learn

Jupyter Notebook:

The Jupyter Notebook App is a server-customer application that permits altering and running note pad records by means of an internet browser. The Jupyter Notebook App can be executed on a nearby work area requiring no web access as portrayed in this report or can be introduced on a remote server and got to through the web. A scratch pad part is a computational motor that executes the code contained in a Notebook record. When you open a Notebook report, the related part is consequently propelled. Note that the RAM isn't discharged until the part is closed down, the Notebook Dashboard is the part which is indicated first when you dispatch Jupyter Notebook App. The Notebook Dashboard is essentially used to open note pad archives, and to deal with the running portions. The Notebook Dashboard has different highlights like a record director, in particular exploring organizers, renaming and erasing documents.

NumPy:

NumPy is a general-purpose array-processing package. It provides a high-performance multi-dimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

Pandas:

Pandas is one of the most widely used python libraries in data science. It provides high-performance, easy to use structures and data analysis tools. Unlike NumPy library which provides objects for multi-dimensional arrays, Pandas provides in-memory 2d table object called Data frame.

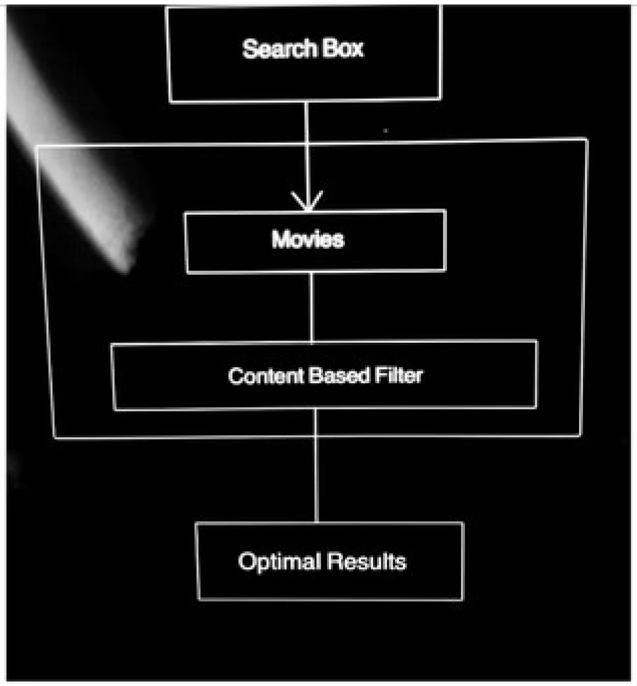
Scikit-learn:

scikit-learn is a powerful library in Python that helps you do machine learning easily. It provides tools and functions for various tasks like building models to make predictions, clustering data, and more. Whether you want to predict if an email is spam or analyze patterns in your data, scikit-learn has tools to make the process simpler. It's widely used by data scientists and researchers to work with data and build machine learning models.

4.SYSTEM ANALYSIS AND ARCHITECTURE:

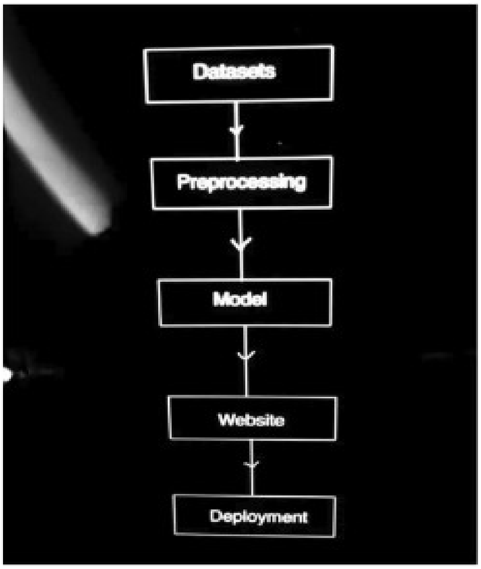
4.1.System Architecture of Proposed System:

Based on content based filtering approaches used in the project, each search will recommend a set of 5 movies to a particular user. When the user will search for a movie, top 5 most similar movies in terms of similar content will get recommended to the user.



4.2 Dataflow:

Initially load the data sets that are required to build a model the data set that are required in this project are movies.csv, ratinfg.csv, users.csv all the data sets are available in the Kaggle.com. Preprocessing of the data set will take place and then by applying content-based filtering, users will get recommended top 5 most similar movies according to their search.



5.IMPLEMENTATION

Recommending movies to users can be done in multiple ways using content-based filtering and collaborative filtering approaches. Content-based filtering approach primarily focuses on the item similarity i.e., the similarity in movies, whereas collaborative filtering focuses on drawing a relation between different users of similar choices in watching movies. filtering approaches. Content-based filtering approach primarily focuses on the item similarity i.e., the similarity in movies, whereas collaborative filtering focuses on drawing a relation between different users of similar choices in watching movies. Based on the plot of a movie that was watched by the user in the past, movies with a similar plot can be recommended to the user. This approach comes under content-based filtering as the recommendations are done only based on the user's past activity.

5.1.Datapreprocessing:

DATASET USED:A kaggle dataset which was scraped from wikipedia and contains plot summary of movies.
In machine learning, we use “pandas” and “numpy” libraries for pre-processing purpose. Machine Learning Architecture

NUMPY: NumPy stands for ‘Numerical Python’ or ‘Numeric Python’. it's Associate in Nursing ASCII text file module of Python that provides quick mathematical computation on arrays and matrices. Numpy will be foreign into the notebook using:

```
>>>import numpy as np
```

PANDAS: Pandas is one in every of the foremost wide used python libraries in information science. It provides superior, straightforward to use structures and information analysis tools. Hence, with 2nd tables, pandas square measure capable of providing several further functionalities like making pivot tables, computing columns supported different columns
and plotting graphs. Pandas will be foreign into Python using:

```
>>>import pandas as pd
```

Choosing a model and strategy is extremely vital method wherever in we've victimisation libraries and machine learning techniques specifically Scikit Learn, NLTK (Natural Language Toolkit), and victimisation formula referred to as circular function similarity.

SCIKIT- LEARN: Scikit-learn (Sklearn) is that the most helpful and sturdy library for machine learning in Python. It provides a variety of economical tools for machine learning and applied math modeling as well as classification, regression, clump and spatial property reduction via a consistence interface in Python. This library, that is basically written in Python, is made upon NumPy, SciPy and Matplotlib.Rather than that specialize in loading, manipulating and summarising information, Scikit-learn library is targeted on modeling the info. Stop words area unit simply a listing of words you don’t wish to use as options. You'll be able to set the parameter stop words=’english’ to use a integral list. Alternatively, you'll be able to set stop words adequate to some custom list. This parameter defaults to none.

NLTK: NLTK (Natural Language Toolkit) Library could be a suite that contains libraries and programs for applied math language process. it's one in all the foremost powerful NLP libraries, that contains packages to form machines perceive human language associated reply to that with an acceptable response. Stemming and Lemmatization in Python NLTK square measure text standardization techniques for language process. These techniques square measure wide used for text preprocessing. The distinction between stemming and lemmatization is that stemming is quicker because it cuts words while not knowing the context, whereas lemmatization is slower because it is aware of the context of words before process. Stemming could be a methodology of standardization of words in language process. It is a way during which a collection of words in a very sentence square measure born-again into a sequence to shorten its search. during this methodology, the words having identical which means however have some variations consistent with the context or sentence square measure normalized.

Stemming and Lemmatization in Python NLTK area unit text normalisation techniques for language process. These techniques area unit wide used for text preprocessing. The distinction between stemming and lemmatization is that stemming is quicker because it cuts words while not knowing the context, whereas lemmatization is slower because it is aware of the context of words before process. Stemming could be a methodology of normalisation of words in language process. it's a method within which a collection of words in an exceedingly sentence area unit reborn into a sequence to shorten its operation. during this methodology, the words having an equivalent which means however have some variations in step with the context or sentence area unit normalized

5.2. Cosine Similarity:

Cosine similarity is a metric used to measure the similarity between two non-zero vectors in an inner product space. In the context of machine learning and information retrieval, cosine similarity is often applied to measure the similarity between two documents or vectors representing items, such as in a recommendation system.The movie plots are transformed as vectors in a geometric space. Therefore, the angle between two vectors represents the closeness of those two vectors. Cosine similarity calculates similarity by measuring the cosine of the angle between two vectors.

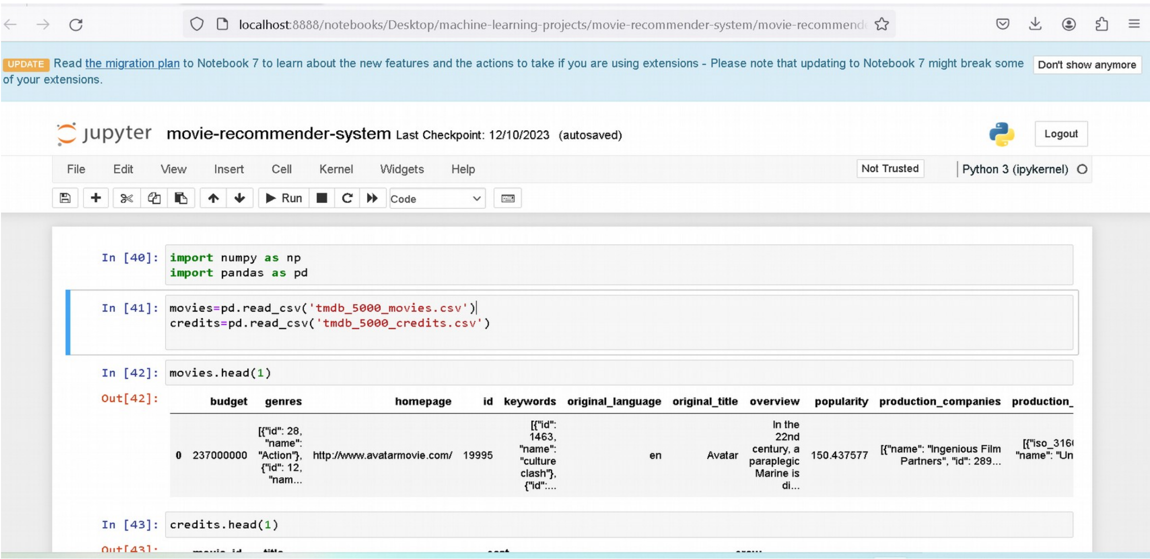
5.3.Vectorization:

Vectorization is crucial because most machine learning algorithms, including those used in recommendation systems, operate on numerical data. By transforming non-numeric features into numerical vectors, you enable the model to understand and learn patterns from the data, ultimately improving the effectiveness of the recommendation system.Vectorization is used to speed up the Python code without using loop. Using such a function can help in minimizing the running time of code efficiently. Various operations are being performed over vector such as dot product of vectors which is also known as scalar product as it produces single output, outer products which results in square matrix of dimension equal to length X length of the vectors, Element wise multiplication which products the element of same indices and dimension of the matrix remain unchanged

6.Experimental Setup:

Back-end (Jupyter notebook): For backend, we used Jupyter notebook to generate a local host Api and the resultant Api is fetched in front to display the result.

We used python language for making our movie recommendation system.



localhost:8888/notebooks/Desktop/machine-learning-projects/movie-recommender-system/movie-recommend

UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions.

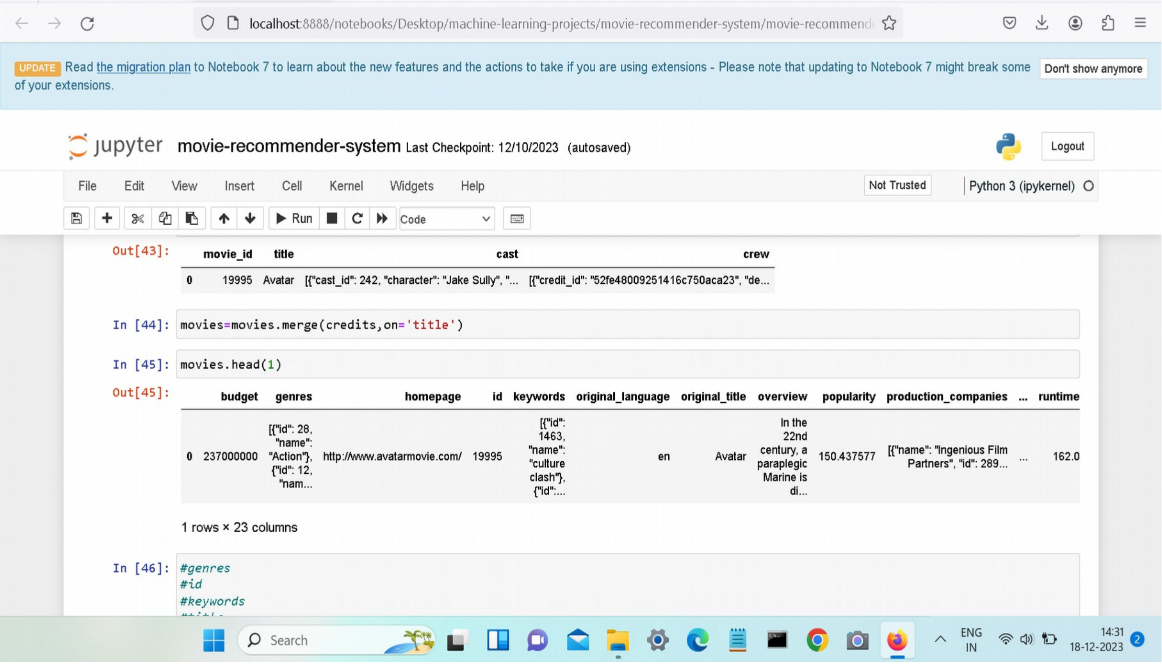
jupyter movie-recommender-system Last Checkpoint: 12/10/2023 (autosaved)

```
In [40]: import numpy as np
         import pandas as pd

In [41]: movies=pd.read_csv('tmdb_5000_movies.csv')
         credits=pd.read_csv('tmdb_5000_credits.csv')

In [42]: movies.head(1)
```

Out[42]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies	production_
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingenious Film Partners", "id": 289...	[{"iso_316 "name": "Un

```
In [43]: credits.head(1)
```



localhost:8888/notebooks/Desktop/machine-learning-projects/movie-recommender-system/movie-recommend

UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions.

jupyter movie-recommender-system Last Checkpoint: 12/10/2023 (autosaved)

Out[43]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...

```
In [44]: movies=movies.merge(credits,on='title')

In [45]: movies.head(1)
```

Out[45]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies	...	runtime
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingenious Film Partners", "id": 289...	...	162.0

1 rows × 23 columns

```
In [46]: #genres
         #id
         #keywords
```

7.RESULT AND DISCUSSION

Our movie recommendation system is based on content based filtering. Its few advantages and disadvantages are listed below-

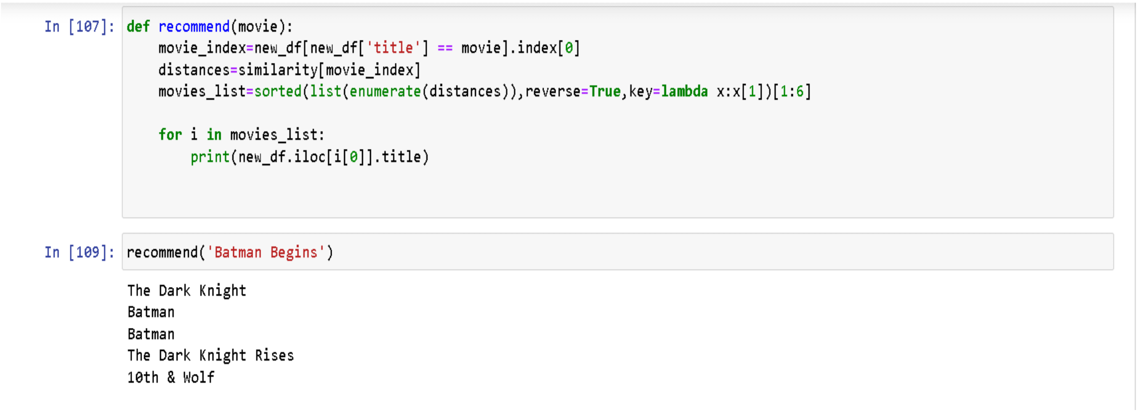
Advantages

- The model doesn't need any data about other users, since the recommendations are specific to this user. This makes it easier to scale to a large number of users.
- The model can capture the specific interests of a user, and can recommend niche items that very few other users are interested in.

Disadvantages

- Since the feature representation of the items are hand-engineered to some extent, this technique requires a lot of domain knowledge. Therefore, the model can only be as good as the hand- engineered features.
- The model can only make recommendations based on existing interests of the user. In other words, the model has limited ability to expand on the users' existing interests.

SCREENSHOT OF THE RESULT:



```
In [107]: def recommend(movie):
              movie_index=new_df[new_df['title'] == movie].index[0]
              distances=similarity[movie_index]
              movies_list=sorted(list(enumerate(distances)),reverse=True,key=lambda x:x[1])[1:6]

              for i in movies_list:
                  print(new_df.iloc[i[0]].title)

In [109]: recommend('Batman Begins')

          The Dark Knight
          Batman
          Batman
          The Dark Knight Rises
          10th & Wolf
```

8.TESTING:

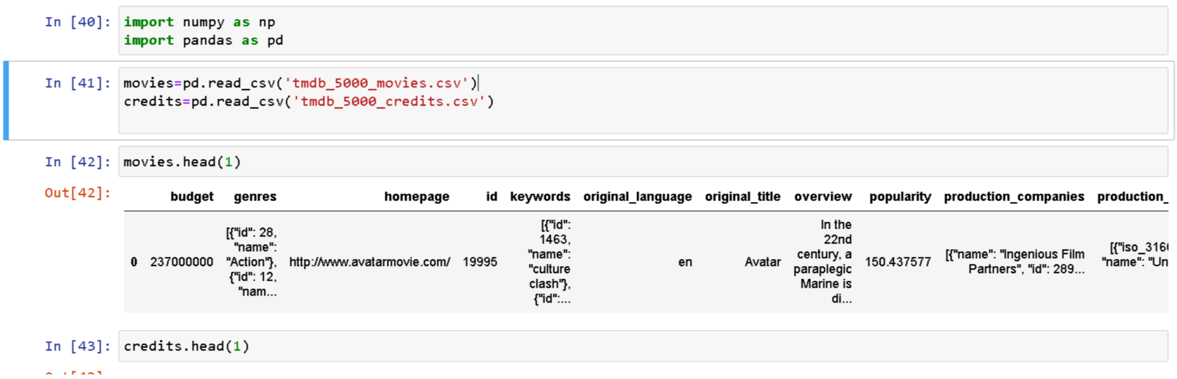
The goal of system testing, which consists of a variety of tests, is to completely exercise the computer-based system. Despite the fact that each test has a distinct objective, they all aim to ensure that all system components have been correctly integrated and carry out their assigned functions. The testing procedure is actually used to ensure that the product performs exactly as it is intended to. The testing phase aims to accomplish the following objectives: - To validate the project's quality. To identify and fix any leftover mistakes from earlier steps. To confirm that the software is a viable fix for the original issue. To guarantee the system's operational reliability. Some of the important testing methodologies are: Unit Testing, Integration Testing and System Testing.

Unit Testing - Unit testing is the first level of testing and is often performed by the developers themselves.

Integration Testing - After each unit is thoroughly tested, it is integrated with other units to create modules or components that are designed to perform specific tasks or activities.

System Testing - System testing is a black box testing method used to evaluate the completed and integrated system, as a whole, to ensure it meets specified requirements.

9.CODE SNAPSHOT



In [40]:
```
import numpy as np
import pandas as pd
```

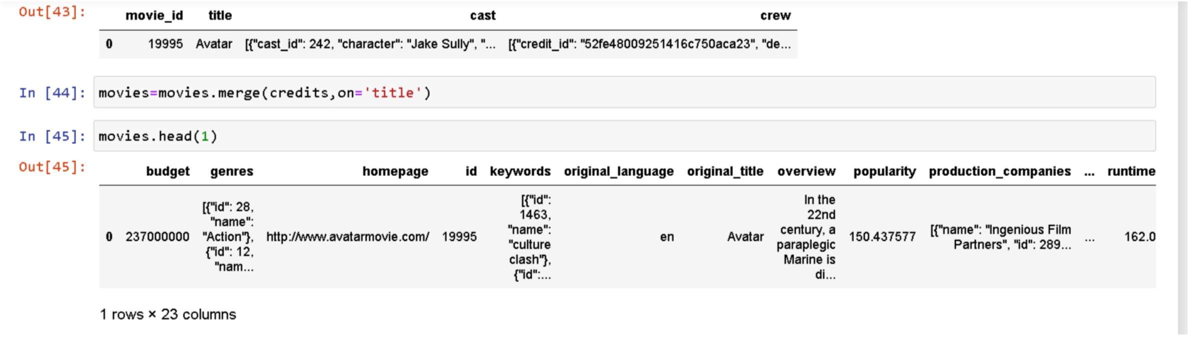
In [41]:
```
movies=pd.read_csv('tmdb_5000_movies.csv')
credits=pd.read_csv('tmdb_5000_credits.csv')
```

In [42]:
```
movies.head(1)
```

Out[42]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies	production_
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingenious Film Partners", "id": 289...	[{"iso_316 "name": "Un

In [43]:
```
credits.head(1)
```



Out[43]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...

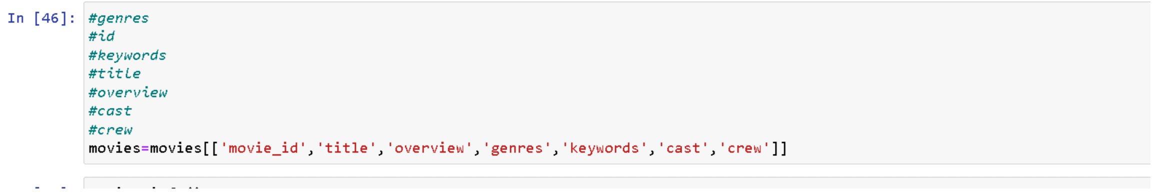
In [44]:
```
movies=movies.merge(credits,on='title')
```

In [45]:
```
movies.head(1)
```

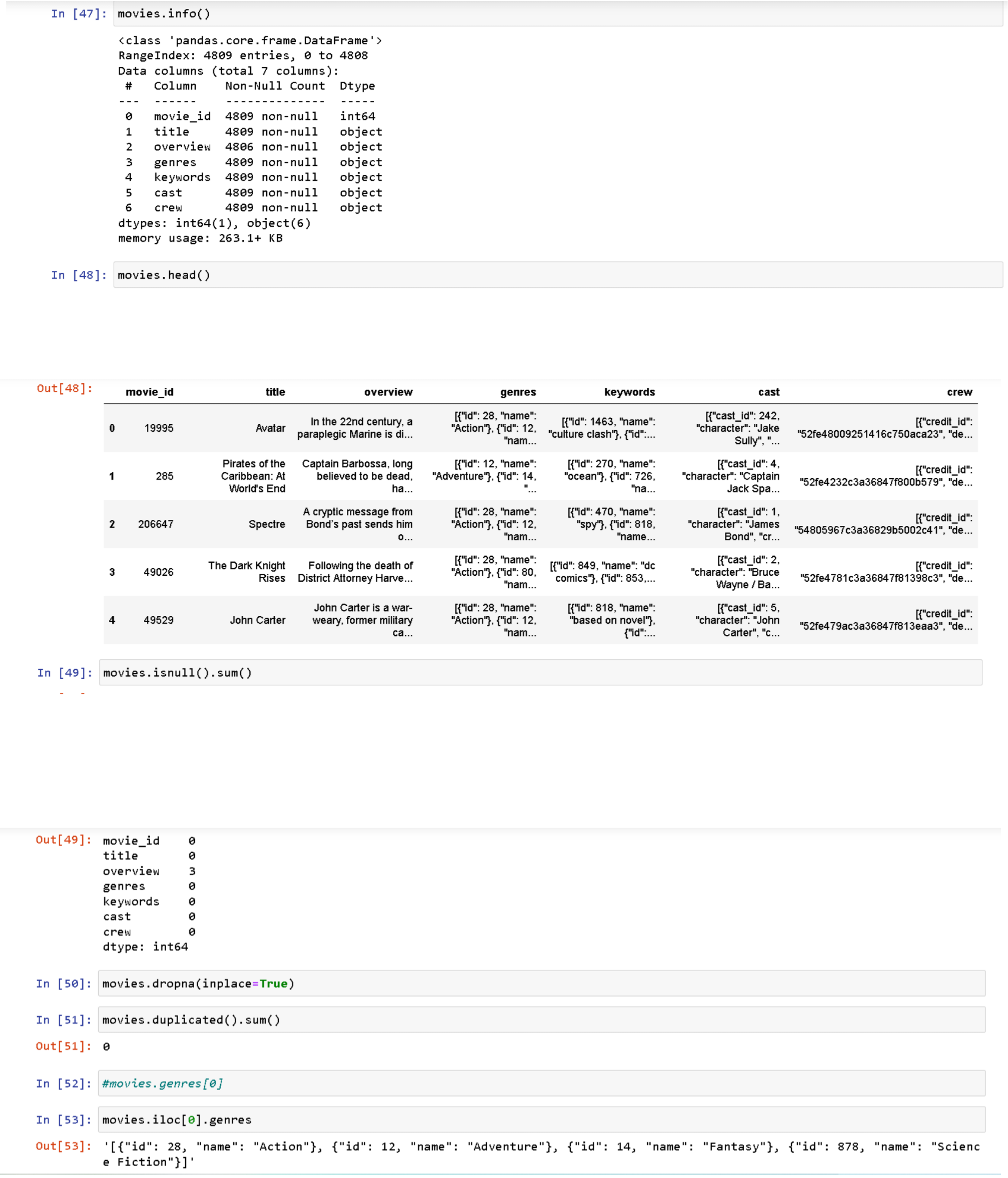
Out[45]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies	...	runtime
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingenious Film Partners", "id": 289...	...	162.0

1 rows × 23 columns



In [46]:
```
#genres
#id
#keywords
#title
#overview
#cast
#crew
movies=movies[['movie_id','title','overview','genres','keywords','cast','crew']]
```



In [47]:
```
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4809 entries, 0 to 4808
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   movie_id  4809 non-null   int64 
 1   title     4809 non-null   object
 2   overview  4806 non-null   object
 3   genres    4809 non-null   object
 4   keywords  4809 non-null   object
 5   cast      4809 non-null   object
 6   crew      4809 non-null   object
dtypes: int64(1), object(6)
memory usage: 263.1+ KB
```

In [48]:
```
movies.head()
```

Out[48]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 1463, "name": "culture clash"}, {"id":...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [49]:
```
movies.isnull().sum()
```

Out[49]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [50]:
```
movies.dropna(inplace=True)
```

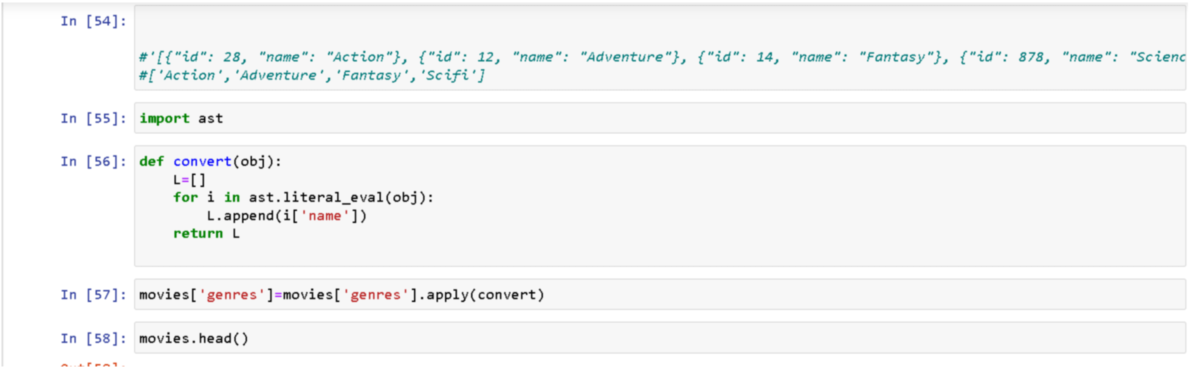
In [51]:
```
movies.duplicated().sum()
```

Out[51]:
```
0
```

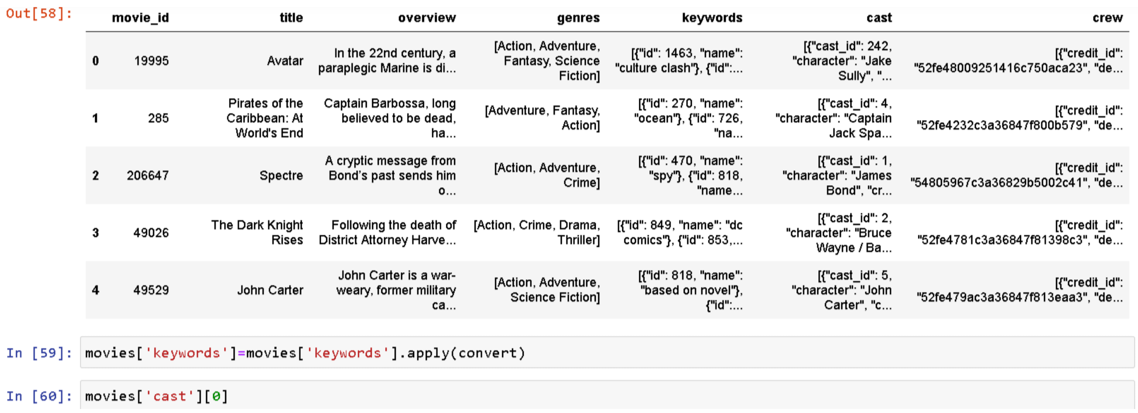
In [52]:
```
#movies.genres[0]
```

In [53]:
```
movies.iloc[0].genres
```

Out[53]:
```
'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Scienc
e Fiction"}]'
```



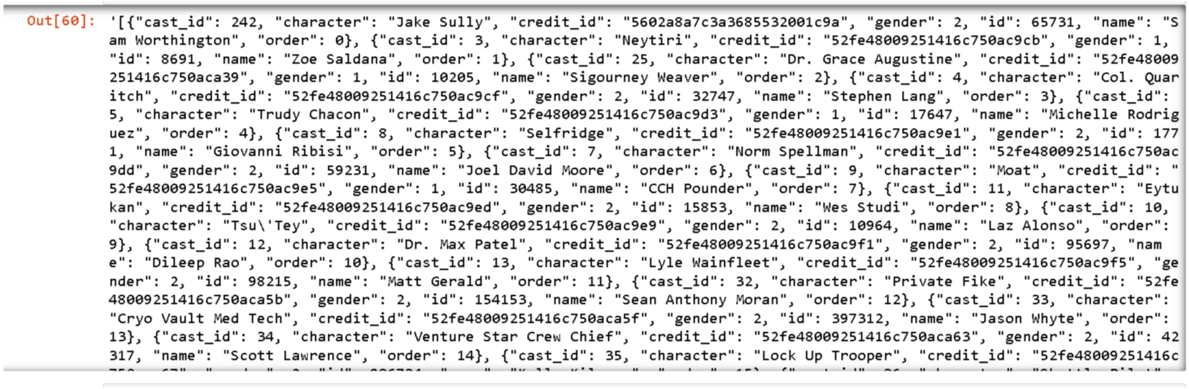
```
In [54]: 

         #'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Scienc
         #['Action','Adventure','Fantasy','Scifi']

In [55]: import ast

In [56]: def convert(obj):
             L=[]
             for i in ast.literal_eval(obj):
                 L.append(i['name'])
             return L

In [57]: movies['genres']=movies['genres'].apply(convert)

In [58]: movies.head()
```



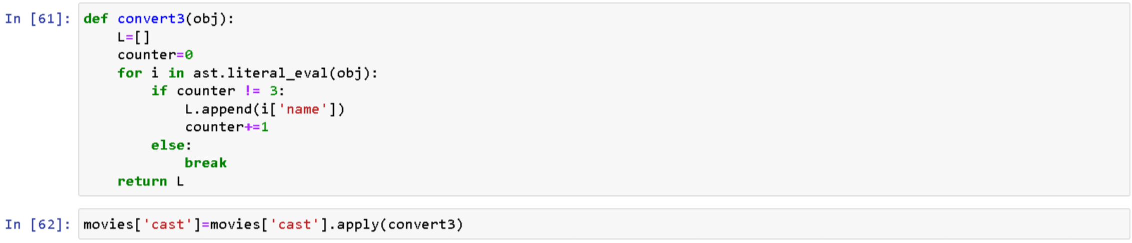
Out[58]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[{"id": 1463, "name": "culture clash"}, {"id":...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

```
In [59]: movies['keywords']=movies['keywords'].apply(convert)

In [60]: movies['cast'][0]
```

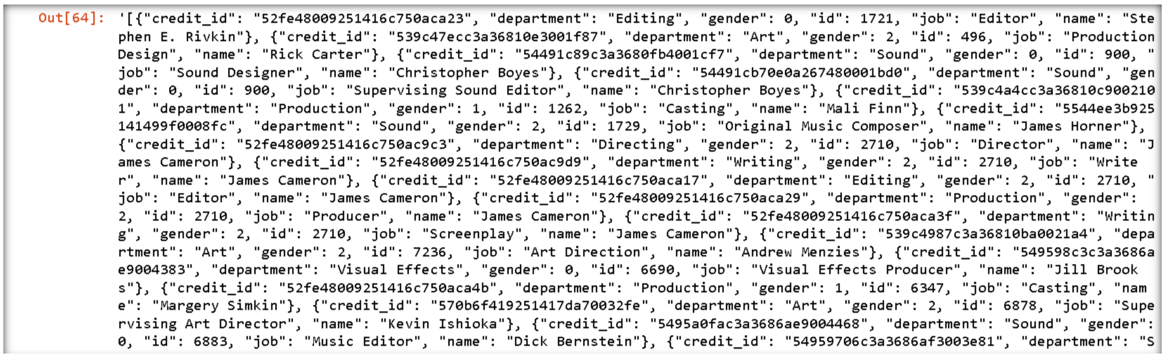


```
Out[60]: '[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a", "gender": 2, "id": 65731, "name": "S
         am Worthington", "order": 0}, {"cast_id": 3, "character": "Neytiri", "credit_id": "52fe48009251416c750ac9cb", "gender": 1,
         "id": 8691, "name": "Zoe Saldana", "order": 1}, {"cast_id": 25, "character": "Dr. Grace Augustine", "credit_id": "52fe48009
         251416c750aca39", "gender": 1, "id": 10205, "name": "Sigourney Weaver", "order": 2}, {"cast_id": 4, "character": "Col. Quar
         itch", "credit_id": "52fe48009251416c750ac9cf", "gender": 2, "id": 32747, "name": "Stephen Lang", "order": 3}, {"cast_id":
         5, "character": "Trudy Chacon", "credit_id": "52fe48009251416c750ac9d3", "gender": 1, "id": 17647, "name": "Michelle Rodrig
         uez", "order": 4}, {"cast_id": 8, "character": "Selfridge", "credit_id": "52fe48009251416c750ac9e1", "gender": 2, "id": 177
         1, "name": "Giovanni Ribisi", "order": 5}, {"cast_id": 7, "character": "Norm Spellman", "credit_id": "52fe48009251416c750ac
         9dd", "gender": 2, "id": 59231, "name": "Joel David Moore", "order": 6}, {"cast_id": 9, "character": "Moat", "credit_id": "
         52fe48009251416c750ac9e5", "gender": 1, "id": 30485, "name": "CCH Pounder", "order": 7}, {"cast_id": 11, "character": "Eytu
         kan", "credit_id": "52fe48009251416c750ac9ed", "gender": 2, "id": 15853, "name": "Wes Studi", "order": 8}, {"cast_id": 10,
         "character": "Tsu\'Tey", "credit_id": "52fe48009251416c750ac9e9", "gender": 2, "id": 10964, "name": "Laz Alonso", "order":
         9}, {"cast_id": 12, "character": "Dr. Max Patel", "credit_id": "52fe48009251416c750ac9f1", "gender": 2, "id": 95697, "nam
         e": "Dileep Rao", "order": 10}, {"cast_id": 13, "character": "Lyle Wainfleet", "credit_id": "52fe48009251416c750ac9f5", "ge
         nder": 2, "id": 98215, "name": "Matt Gerald", "order": 11}, {"cast_id": 32, "character": "Private Fike", "credit_id": "52fe
         48009251416c750aca5b", "gender": 2, "id": 154153, "name": "Sean Anthony Moran", "order": 12}, {"cast_id": 33, "character":
         "Cryo Vault Med Tech", "credit_id": "52fe48009251416c750aca5f", "gender": 2, "id": 397312, "name": "Jason Whyte", "order":
         13}, {"cast_id": 34, "character": "Venture Star Crew Chief", "credit_id": "52fe48009251416c750aca63", "gender": 2, "id": 42
         317, "name": "Scott Lawrence", "order": 14}, {"cast_id": 35, "character": "Lock Up Trooper", "credit_id": "52fe48009251416c
```

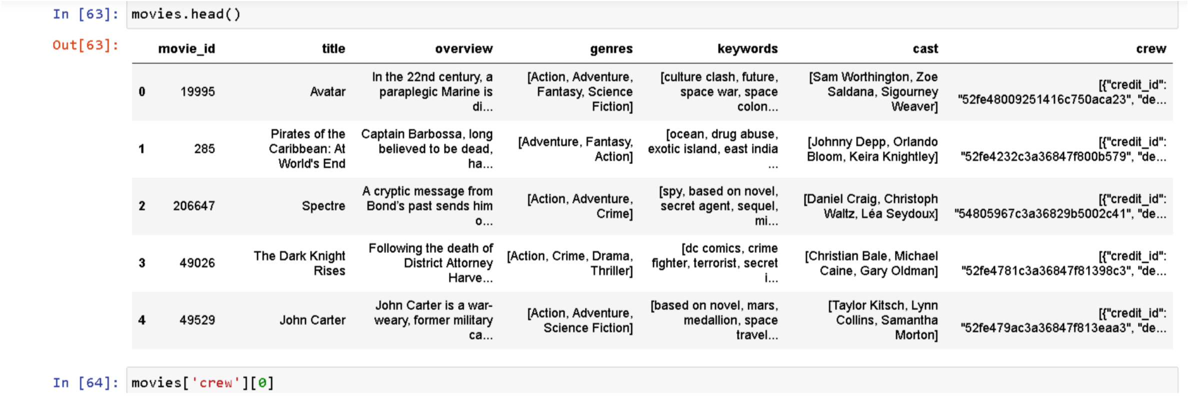


```
In [61]: def convert3(obj):
             L=[]
             counter=0
             for i in ast.literal_eval(obj):
                 if counter != 3:
                     L.append(i['name'])
                     counter+=1
                 else:
                     break
             return L
```

```
In [62]: movies['cast']=movies['cast'].apply(convert3)
```



```
Out[64]: '[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "id": 1721, "job": "Editor", "name": "Ste
         phen E. Rivkin"}, {"credit_id": "539c47ecc3a36810e3001f87", "department": "Art", "gender": 2, "id": 496, "job": "Production
         Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900, "
         job": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Sound", "gen
         der": 0, "id": 900, "job": "Supervising Sound Editor", "name": "Christopher Boyes"}, {"credit_id": "539c4a4cc3a36810c900210
         1", "department": "Production", "gender": 1, "id": 1262, "job": "Casting", "name": "Mali Finn"}, {"credit_id": "5544ee3b925
         141499f0008fc", "department": "Sound", "gender": 2, "id": 1729, "job": "Original Music Composer", "name": "James Horner"},
         {"credit_id": "52fe48009251416c750ac9c3", "department": "Directing", "gender": 2, "id": 2710, "job": "Director", "name": "J
         ames Cameron"}, {"credit_id": "52fe48009251416c750ac9d9", "department": "Writing", "gender": 2, "id": 2710, "job": "Write
         r", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca17", "department": "Editing", "gender": 2, "id": 2710, "
         job": "Editor", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca29", "department": "Production", "gender":
         2, "id": 2710, "job": "Producer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca3f", "department": "Writin
         g", "gender": 2, "id": 2710, "job": "Screenplay", "name": "James Cameron"}, {"credit_id": "539c4987c3a36810ba0021a4", "depa
         rtment": "Art", "gender": 2, "id": 7236, "job": "Art Direction", "name": "Andrew Menzies"}, {"credit_id": "549598c3c3a3686a
         e9004383", "department": "Visual Effects", "gender": 0, "id": 6690, "job": "Visual Effects Producer", "name": "Jill Brook
         s"}, {"credit_id": "52fe48009251416c750aca4b", "department": "Production", "gender": 1, "id": 6347, "job": "Casting", "nam
         e": "Margery Simkin"}, {"credit_id": "570b6f419251417da70032fe", "department": "Art", "gender": 2, "id": 6878, "job": "Supe
         rvising Art Director", "name": "Kevin Ishioka"}, {"credit_id": "5495a0fac3a3686ae9004468", "department": "Sound", "gender":
         0, "id": 6883, "job": "Music Editor", "name": "Dick Bernstein"}, {"credit_id": "54959706c3a3686af3003e81", "department": "S
```

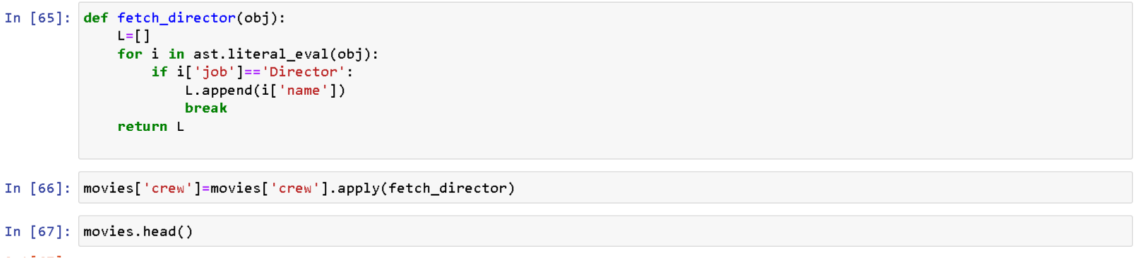


```
In [63]: movies.head()
```

Out[63]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman]	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

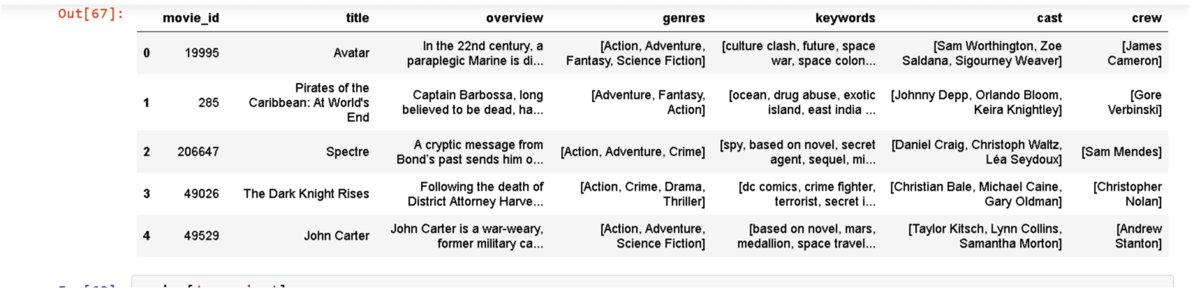
```
In [64]: movies['crew'][0]
```



```
In [65]: def fetch_director(obj):
             L=[]
             for i in ast.literal_eval(obj):
                 if i['job']=='Director':
                     L.append(i['name'])
                     break
             return L
```

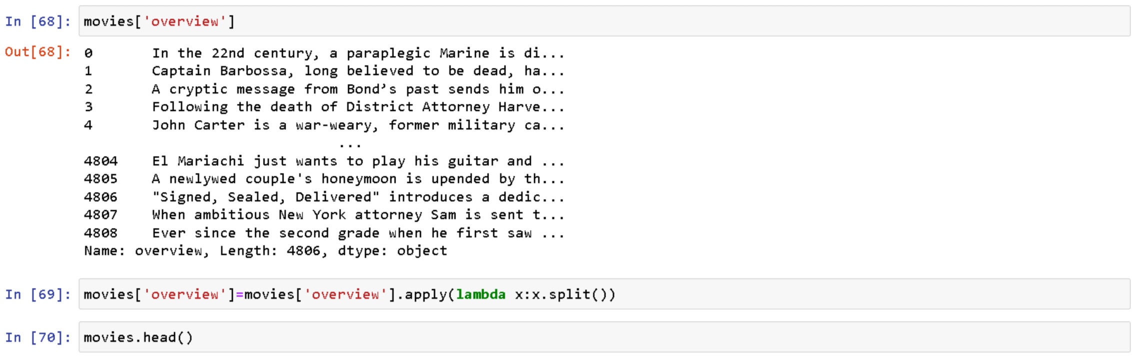
```
In [66]: movies['crew']=movies['crew'].apply(fetch_director)
```

```
In [67]: movies.head()
```



Out[67]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman]	[Christopher Nolan]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

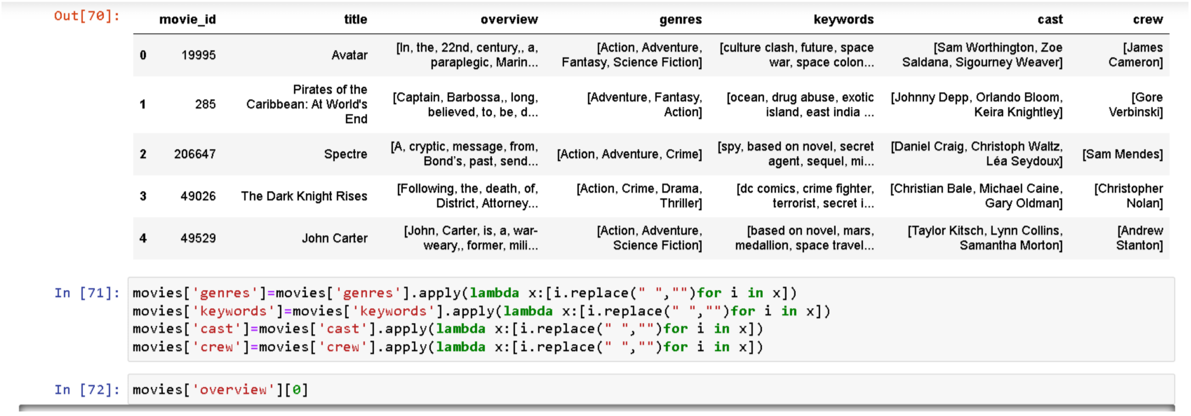


```
In [68]: movies['overview']
```

```
Out[68]: 0       In the 22nd century, a paraplegic Marine is di...
         1       Captain Barbossa, long believed to be dead, ha...
         2       A cryptic message from Bond's past sends him o...
         3       Following the death of District Attorney Harve...
         4       John Carter is a war-weary, former military ca...
                                       ...
         4804    El Mariachi just wants to play his guitar and ...
         4805    A newlywed couple's honeymoon is upended by th...
         4806    "Signed, Sealed, Delivered" introduces a dedic...
         4807    When ambitious New York attorney Sam is sent t...
         4808    Ever since the second grade when he first saw ...
         Name: overview, Length: 4806, dtype: object
```

```
In [69]: movies['overview']=movies['overview'].apply(lambda x:x.split())
```

```
In [70]: movies.head()
```

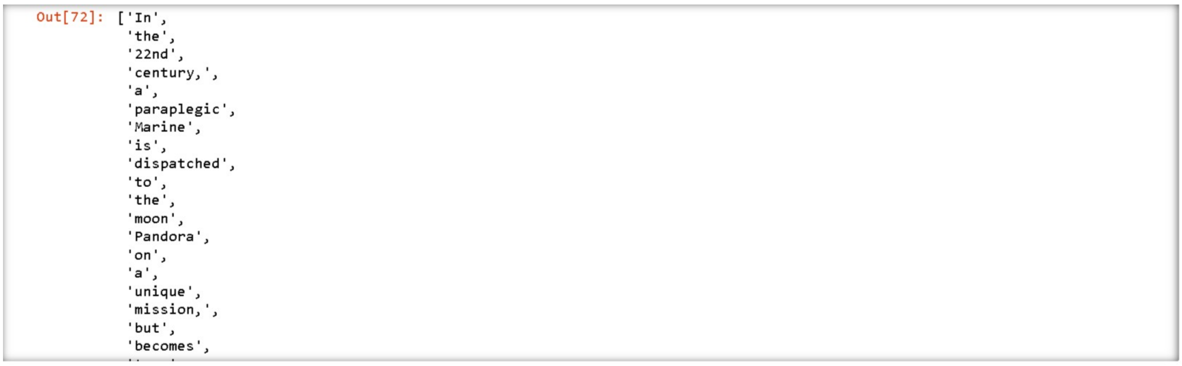


Out[70]:

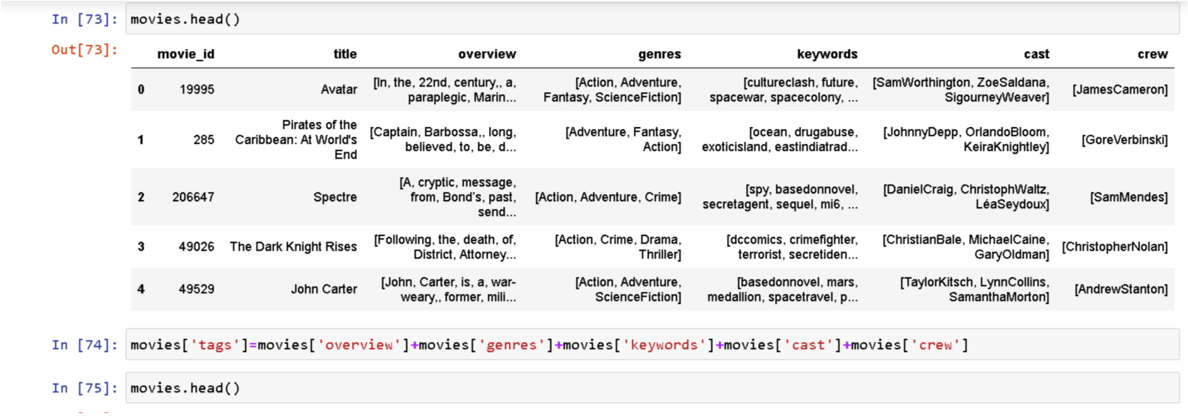
	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman]	[Christopher Nolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

```
In [71]: movies['genres']=movies['genres'].apply(lambda x:[i.replace(" ","")for i in x])
         movies['keywords']=movies['keywords'].apply(lambda x:[i.replace(" ","")for i in x])
         movies['cast']=movies['cast'].apply(lambda x:[i.replace(" ","")for i in x])
         movies['crew']=movies['crew'].apply(lambda x:[i.replace(" ","")for i in x])
```

```
In [72]: movies['overview'][0]
```



```
Out[72]: ['In',
          'the',
          '22nd',
          'century,',
          'a',
          'paraplegic',
          'Marine',
          'is',
          'dispatched',
          'to',
          'the',
          'moon',
          'Pandora',
          'on',
          'a',
          'unique',
          'mission,',
          'but',
          'becomes',
```



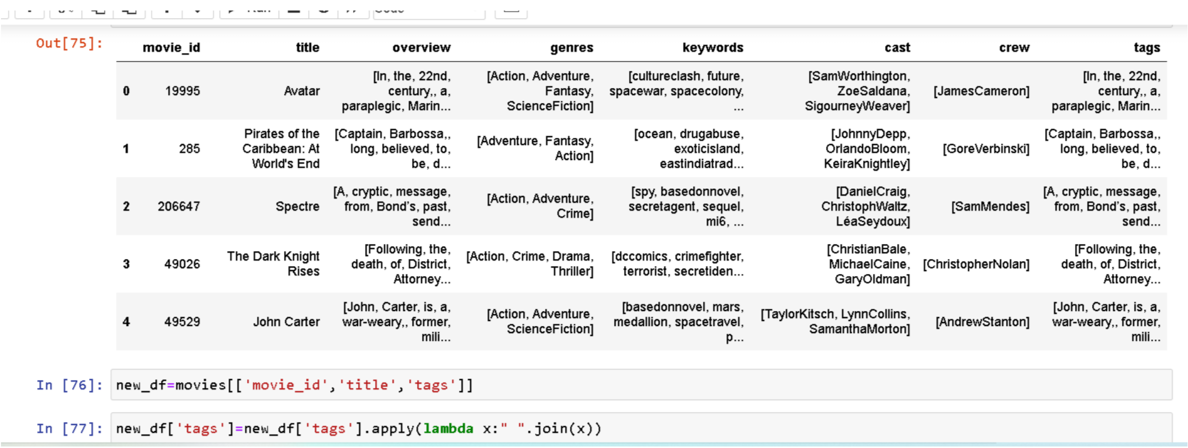
```
In [73]: movies.head()
```

Out[73]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreVerbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewStanton]

```
In [74]: movies['tags']=movies['overview']+movies['genres']+movies['keywords']+movies['cast']+movies['crew']
```

```
In [75]: movies.head()
```

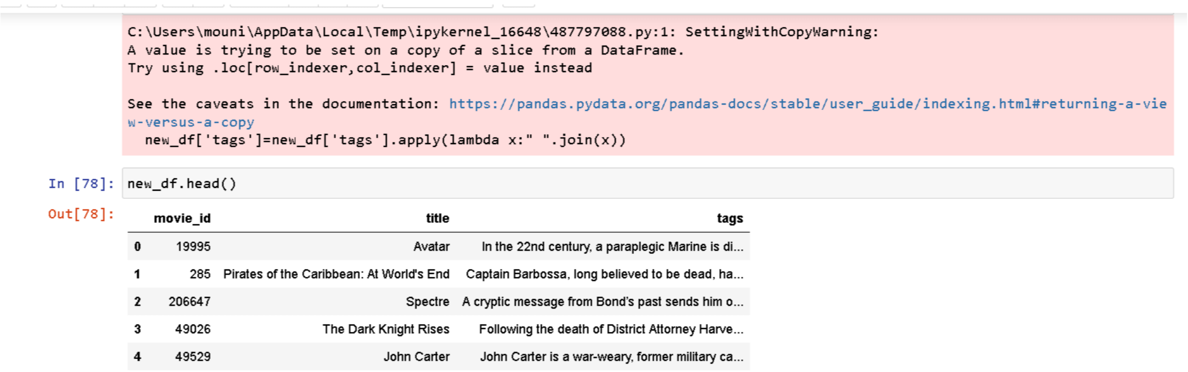


Out[75]:

	movie_id	title	overview	genres	keywords	cast	crew	tags
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]	[In, the, 22nd, century,, a, paraplegic, Marin...
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreVerbinski]	[Captain, Barbossa,, long, believed, to, be, d...
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]	[A, cryptic, message, from, Bond's, past, send...
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]	[Following, the, death, of, District, Attorney...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewStanton]	[John, Carter, is, a, war-weary,, former, mili...

```
In [76]: new_df=movies[['movie_id','title','tags']]
```

```
In [77]: new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x))
```

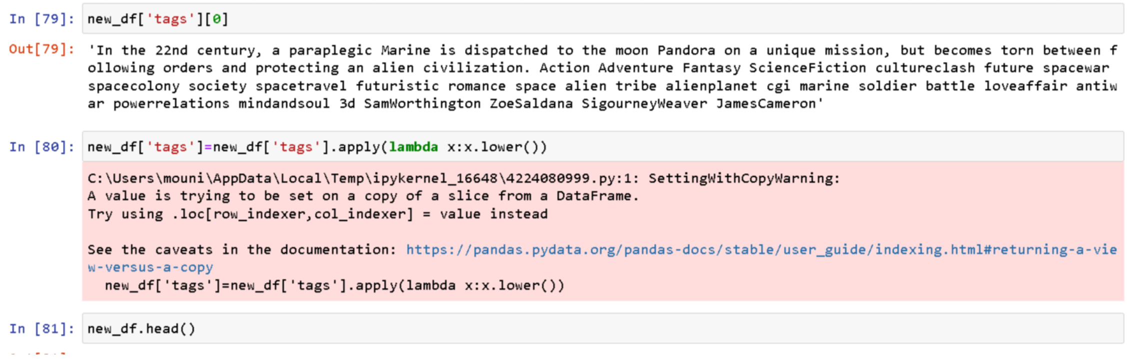


```
C:\Users\mouni\AppData\Local\Temp\ipykernel_16648\487797088.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-vie
w-versus-a-copy
  new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x))
```

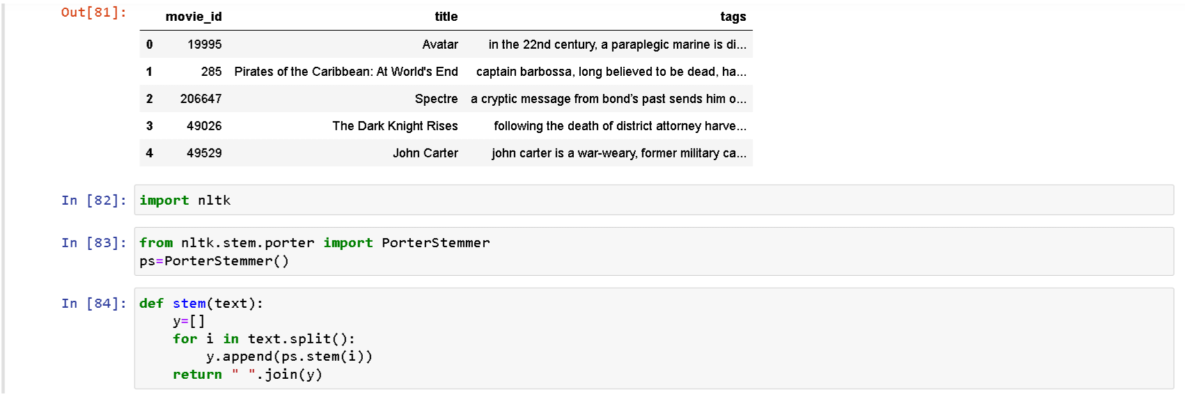
```
In [78]: new_df.head()
```

Out[78]:

	movie_id	title	tags
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...
2	206647	Spectre	A cryptic message from Bond's past sends him o...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...
4	49529	John Carter	John Carter is a war-weary, former military ca...



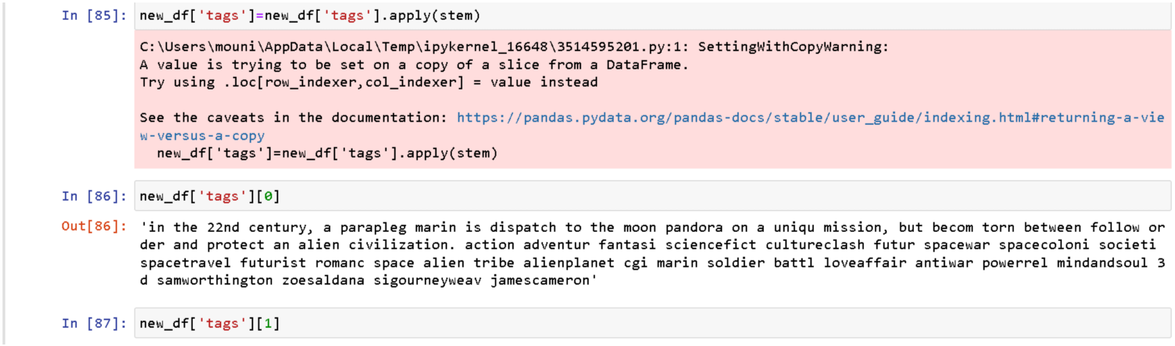
```
In [79]: new_df['tags'][0]

Out[79]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between f
         ollowing orders and protecting an alien civilization. Action Adventure Fantasy ScienceFiction cultureclash future spacewar
         spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle loveaffair antiw
         ar powerrelations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver JamesCameron'

In [80]: new_df['tags']=new_df['tags'].apply(lambda x:x.lower())

         C:\Users\mouni\AppData\Local\Temp\ipykernel_16648\4224080999.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-vie
         w-versus-a-copy
           new_df['tags']=new_df['tags'].apply(lambda x:x.lower())

In [81]: new_df.head()
```



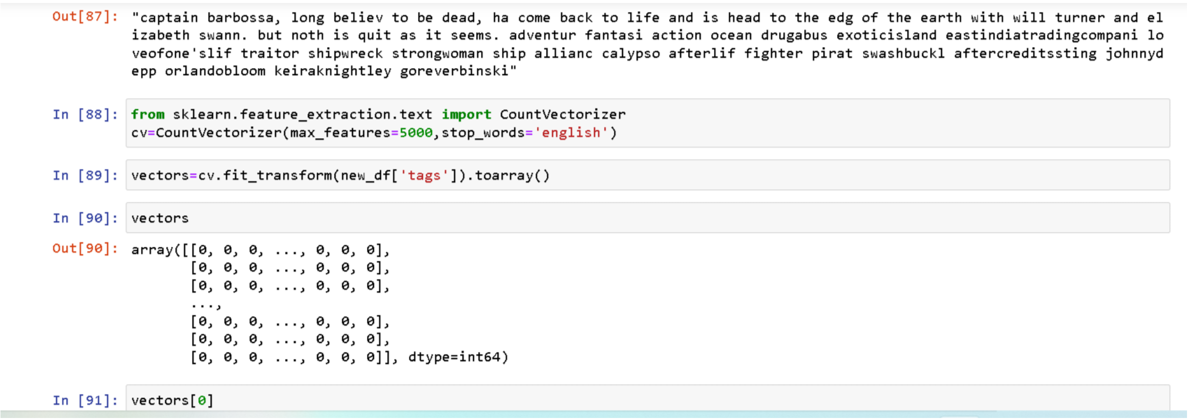
Out[81]:

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a paraplegic marine is di...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believed to be dead, ha...
2	206647	Spectre	a cryptic message from bond's past sends him o...
3	49026	The Dark Knight Rises	following the death of district attorney harve...
4	49529	John Carter	john carter is a war-weary, former military ca...

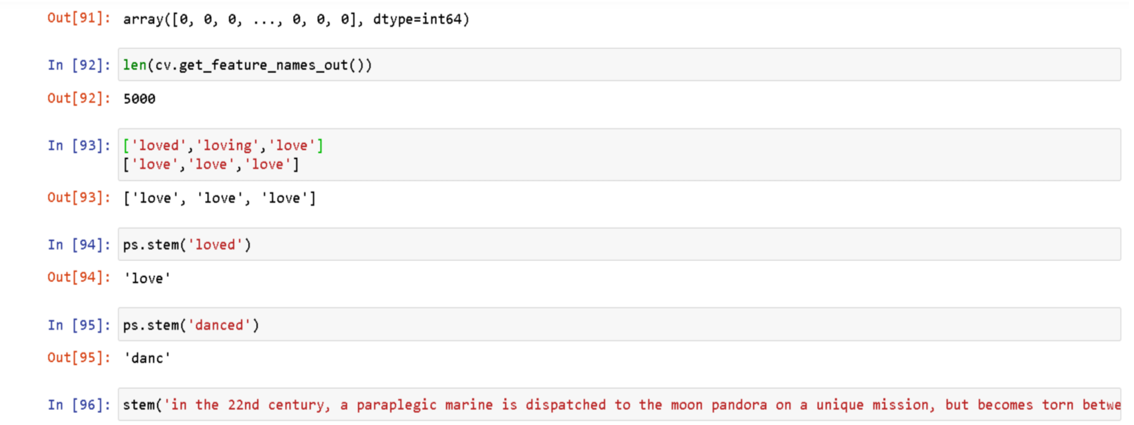
```
In [82]: import nltk

In [83]: from nltk.stem.porter import PorterStemmer
         ps=PorterStemmer()

In [84]: def stem(text):
             y=[]
             for i in text.split():
                 y.append(ps.stem(i))
             return " ".join(y)
```



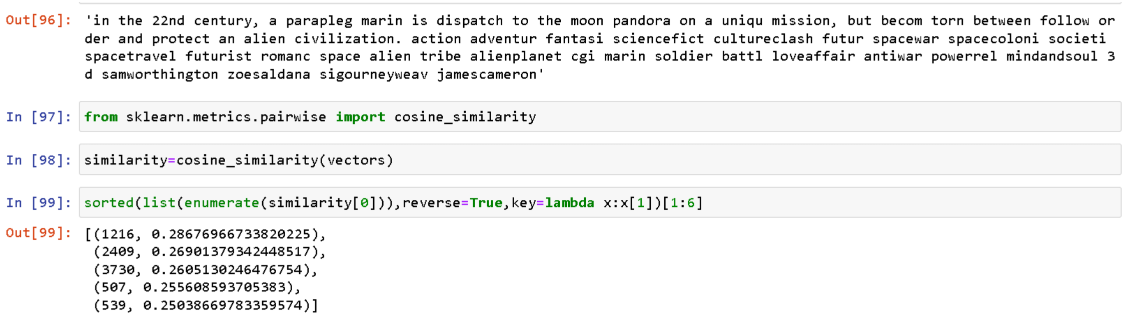
```
In [85]: new_df['tags']=new_df['tags'].apply(stem)

         C:\Users\mouni\AppData\Local\Temp\ipykernel_16648\3514595201.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-vie
         w-versus-a-copy
           new_df['tags']=new_df['tags'].apply(stem)

In [86]: new_df['tags'][0]

Out[86]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu mission, but becom torn between follow or
         der and protect an alien civilization. action adventur fantasi sciencefict cultureclash futur spacewar spacecoloni societi
         spacetravel futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul 3
         d samworthington zoesaldana sigourneyweav jamescameron'

In [87]: new_df['tags'][1]
```



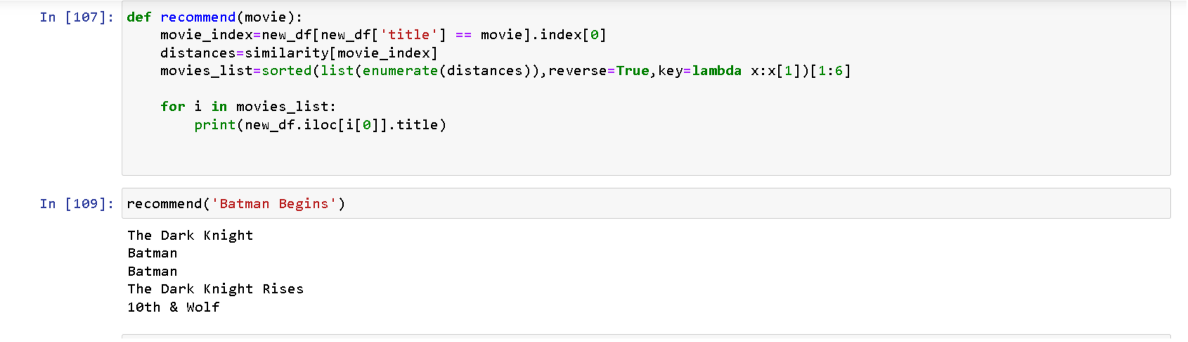
```
Out[87]: "captain barbossa, long believ to be dead, ha come back to life and is head to the edg of the earth with will turner and el
         izabeth swann. but noth is quit as it seems. adventur fantasi action ocean drugabus exoticisland eastindiatradingcompani lo
         veofone'slif traitor shipwreck strongwoman ship allianc calypso afterlif fighter pirat swashbuckl aftercreditssting johnnyd
         epp orlandobloom keiraknightley goreverbinski"

In [88]: from sklearn.feature_extraction.text import CountVectorizer
         cv=CountVectorizer(max_features=5000,stop_words='english')

In [89]: vectors=cv.fit_transform(new_df['tags']).toarray()

In [90]: vectors

Out[90]: array([[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)

In [91]: vectors[0]
```



```
Out[91]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)

In [92]: len(cv.get_feature_names_out())

Out[92]: 5000

In [93]: ['loved','loving','love']
         ['love','love','love']

Out[93]: ['love', 'love', 'love']

In [94]: ps.stem('loved')

Out[94]: 'love'

In [95]: ps.stem('danced')

Out[95]: 'danc'

In [96]: stem('in the 22nd century, a paraplegic marine is dispatched to the moon pandora on a unique mission, but becomes torn betwe
```



```
Out[96]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu mission, but becom torn between follow or
         der and protect an alien civilization. action adventur fantasi sciencefict cultureclash futur spacewar spacecoloni societi
         spacetravel futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul 3
         d samworthington zoesaldana sigourneyweav jamescameron'

In [97]: from sklearn.metrics.pairwise import cosine_similarity

In [98]: similarity=cosine_similarity(vectors)

In [99]: sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[1:6]

Out[99]: [(1216, 0.28676966733820225),
          (2409, 0.26901379342448517),
          (3730, 0.2605130246476754),
          (507, 0.255608593705383),
          (539, 0.25038669783359574)]
```



```
In [107]: def recommend(movie):
              movie_index=new_df[new_df['title'] == movie].index[0]
              distances=similarity[movie_index]
              movies_list=sorted(list(enumerate(distances)),reverse=True,key=lambda x:x[1])[1:6]

              for i in movies_list:
                  print(new_df.iloc[i[0]].title)

In [109]: recommend('Batman Begins')

          The Dark Knight
          Batman
          Batman
          The Dark Knight Rises
          10th & Wolf
```

10.CONCLUSION AND FUTURE SCOPE

In this project, we have made the movie recommender system which recommends movies based on the content that is the most similar to what the user searches on the site. We have used cosine similarity to find out which top 5 movies would be the closest to what the user searches. This system tries to save up the time for the users who want to watch a movie similar to watch what they had watched before. We have used movie datasets from Kaggle.com which contained information like genre, cast, movie title, keywords in the movie, language and any more to preprocess the data and filter and gather all the information that would to be required to find out all the relevant data for content-based filtering.

Future scope

The development of a web interface stands as the next significant milestone for our movie recommendation system. While the core recommendation algorithm has been successfully implemented and validated in a Jupyter Notebook environment, transitioning to a user-friendly web interface opens up new avenues for accessibility and user engagement. The primary focus will be on creating an intuitive and visually appealing web interface that seamlessly integrates with our existing recommendation system Users will have the opportunity to interact with the recommendation engine effortlessly, providing input and receiving personalized movie suggestions Implement a user-friendly input mechanism allowing users to express their preferences easily. This may include options for entering keywords, selecting genres, or any other relevant criteria for movie recommendations..