l

# This Python 3 environment comes with many helpful analytics libraries installed

# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python # For example, here's several helpful packages to load

import numpy as np # linear algebra

import paudas as pd # data processing, CSV file I/0 (e.g. pd.read\_csv)

# Input data files are available in the read-only "../input/" directory

# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os

for dirname, \_, filenames in os.walk('/kaggle/input'): for filename in filenames:

print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using “Save & # You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session

/kaggle/input/tmdb-movie-metadata/tmdb\_5000\_movies.csv

/kaggle/input/tmdb-movie-metadata/tmdb\_5000\_credits.csv

novles = pd . read\_c sv( ’ /kaggIe/Input/tndb-novle -metadata/tmdb 5066 movies . csv ' )

credits = pd.read\_csv('/kaggle/input/tmdb-movie-metadata/tmdb 5000 credits.csv')

movies.head()

budget **genres** homepage id keywords original\_language original\_title overview populari

[{"id": 28,

"name":

0 237000000 "Action"},

{"id”: 12,

"nam...

[("id": 12,

<http://www.avatarmovie.com/> 19995

[{"id'.

1463,

"name". "culture clash"),

{"id":...

[{"id . 270,

en Avatar

In the 22nd century, a paraplegic Marine is

di...

Captain Barbossa,

150.437\*

1 300000000

"name":

"Adventure"},

{"id": 14, "...

[{"id”: 28,

"name":

<http://disney.go.com/disneypictures/pirates/> 285

"name".

"ocean"},

{"id": 726,

"na...

Pirates of the long

en Caribbean: At believed 139.082E World's End to be

dead, ha...

A cryptic message

from

2 245000000 "Action"}, <http://www.sonypictures.com/movies/spectre/> 206647 ("id": 12,

"nam...



{"id": 818,

"name...

en Spectre Bond’s 107.3767

past

sends him

[("id": 28,

"name":

3 250000000 "Action"},

{"id": 80,

"nam...

[("id": 28,

"name":

<http://www.thedarkknightrises.com/> 49026

g t ’84

"namd

comics"),

{"id":

853,...

[{"id":

818,

"name .

Dark Knight

Rises

Following

the death

of District 112.312.° Attorney

Harve...

John

Carter is a

war-

4 260000000 "Action"},

("id”: 12,

"nam...

http:IImovies.disney.com/john-carter 49529 "based

on novel"), ("id":...

en John Carter weary, 43.926C

former military ca...

movies.shape

(4803, 20)

movies = movies.merge(credits,on='title')

credits.head()

movie\_id

## title cast

cre i

|  |  |  |
| --- | --- | --- |
|  | | [{"cast\_id": 242, [{"credit id”: |
| 0 19995 | Avatar | "character": "Jake Sully", "52fe48009251416c750aca23", "de... |
|  | Pirates of the | [("cast\_id'. 4, [{"credit id": |
| 285 | Caribbean: At World's  End | "character": "Captain "52fe4232c3a36847f800b579", "de...  Jack Spa... |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | [("cast id": 1, |  | [("credit\_id": |
| movies = movies.menge(cnedits,on='title') | | |  |  |  |
| movies.head() | | |  |  |  |
| # budget | | |  |  |  |
| # homepage | | |  |  |  |
| # Id | | |  |  |  |
| # original\_language | | |  |  |  |
| # original\_title | | |  |  |  |
| # popularity | | |  |  |  |
| # production\_comapny | | |  |  |  |
| # production\_countries | | |  |  |  |
| # release-date(not sune) | | |  |  |  |
| budget **genres** | | | homepage | 1d | keywords or1gJ |
| 0 | 237000000 | [("id": 28,  "name":  "Action"},  ("id": 12,  "nam... | <http://www.avatarmovie.com/> | 19995 | [{"id": 1463,  "name": "culture clash"),  ("id":... |

300000000

[("id": 12,

"name":

"Adventure"},

{"id": 14, "...

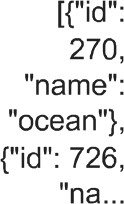
<http://disney.go.com/disneypictures/pirates/> 285

[("id": 28,

"name":

2 245000000 "Action"}, <http://www.sonypictures.com/movies/spectre/> 206647 ("id": 12,

"nam...

[{"id'. 470,

"name":

"•py"1.

("id": 818,

"name...

[{"id": 28,

"name":

3 250000000 "Action"},

("id": 80,

"nam...

[("id": 28,

"name":

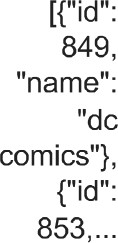
4 260000000 "Action"},

("id": 12,

"nam...

5 rows 29 columns

<http://www.thedarkknightrises.com/> 49026

[{"id'. 818,

"name'. <http://movies.disney.com/john-carter> 49529 "based

on novel"),

movies = movies[['movie\_id', 'title','overview', 'genres','keywords','cast', 'crew']]

movies . head ( )

movie\_id **title overview**

genres keywords

cast



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,

[("cr

0 19995

Avatar

"character": "52fe48009251416c75C "Jake

Sully", "...

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| import ast | Pirates of Barbossa, the long | [("id": 12, | [{"id":  270, | [("cast id":  4, | [{"cr |
| 1 285 Caribbean: believed  de-r- convert(text) • At WOrld's to be  L \_ [] End dead,  fdr i in ast.literal\_eval(text):  L.append(i['name']) return L | | "Adventure"),  ("id": 14,"... | "ocean"},  {"id": 726,  "na. | "character":  "Captain Jack Spa... | ”52fe4232c3a36847f80 |
| movies.dropna(inplace=True) | | | | | |
| movies['gennes'] = movies['genres'].apply(convert)  movies.head() | | | | | |

rtiov1e\_1d t1t1e overv1ew genres keywords

cast



In the 22nd century, a paraplegic Marine is

di...

[Action, Adventure, Fantasy, **Science** Fiction]

[("id'. 1463,

"name'.

"culture

clash"),

{"id":...

[("cast id":

242,

[{"cre‹

0 19995

Avatar

"character": "52fe48009251416c750a "Jake

Sully", "...

Captain Pirates of Barbossa,

the long [Adventure,

1 285 Caribbean: believed Fantasy,

At World’s to be Action] End dead,

[("id": 270,

"name":

"ocean"),

("id": 726,

"na...



[{"cast id":

4, [{"cre‹

"character": "52fe4232c3a36847f800l "Captain

Jack Spa...

movies['keywords'] = movies['keywords'].apply(convert)

movies.head()

movie\_id t1t1e overv1ew genres keywords cast

1. 19995 Avatar

In the 22nd century, a paraplegic Marine is

di...

[culture

[Action, clash, [{"cast\_id":

Adventure, future, 242, [{"crew Fantasy, space "character": "52fe48009251416c750a Science war, "Jake

Fiction] space Sully", "... colon...

Captain [ocean,

Pirates of Barbossa, drug [{"cast\_id":

the long [Adventure, abuse, 4, [{"crew

1. 285 Caribbean: believed Fantasy, exotic "character”: "52fe4232c3a36847f800 At World’s to be Action] island, "Captain



import ast

ast.literal\_eval('[{“id“: 28, "uame”: ”Action“}, {“id": 12, "uame": "Adventure“}, {“id": 14, "name" : "Fantasy" }, {"Id" : 878, "name" : "Sc lenc

[{'id': 28, 'name': 'Action'},

{'id': 12, 'name': 'Adventure'},

{'id': 14, 'name': 'Fantasy'},

{'id': 878, 'name': 'Science Fiction'}]

def convert3(text): L = []

counter = 0

for i in ast.literal\_eval(text): if counter < 3:

L.appeud(i['name']) counter+=1

return L

movies['cast'] = movies['cast'].apply(convert) movies.head()

movie\_id title overview genres keywords **cast**



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,

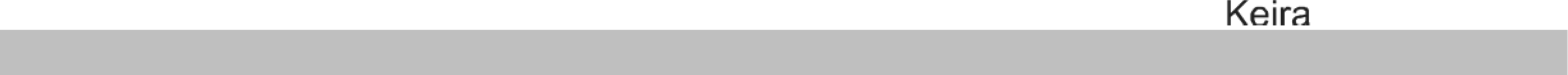
Zoe Saldana, Sigourney Weave...

[(" \* "52fe48009251416c750

Captain [ocean,

Pirates of Barbossa, drug the long [Adventure, abuse,

1 285 Caribbean: believed Fantasy, exotic AtWorld’s to be Action] island,



[Johnny Depp, Orlando Bloom,

[(" ° "52fe4232c3a36847f80

|  |  |  |
| --- | --- | --- |
|  |  |  |
| movies['cast'] = movies['cast'].apply(lambda x:x[0:3]) | | |
| def fetch\_director(text): L = []  for i in ast.literal\_eval(text): if i['job'] == 'Dinector':  L.append(i['name]) return L | | |
| movies['crew'] = movies['crew'].apply(fetch\_director) | | |
| #movies['ovenview'] = movies['ovenview'].apply(lambda x:x.split()) movies.sample(5) | | |

nov1e\_1d t1t1e overv1ew genres keywords

Nada, a [Action, [dystopia, down-on-his- Horror, social

**cast** crew

[Roddy

Piper, Keith [John

3792 8337 They Live

luck Science commentary,

construction Fiction, alien invasion, worker, ... Thriller]

David, Meg Carpenter] Foster]

[Amy

1070 64328

When Kermit The the Frog and Muppets the Muppets

learn tha...

An American

[Comedy, Family, Music]

[musical, the Adams,

muppets, Jason [James

robot, oil Segel, Bobin] tycoon, stud... Emily

Blunt]

[Paul

def collapse(L):

L1 = []

for i in L:

L1.append(i.replace(“ “,““)) return L1

movies['cast'] = movies['cast'].apply(collapse) movies['cnew'] = movies['cnew'].apply(collapse) movies['genres'] = movies['genres'].apply(collapse) movies['keywords'] = movies['keywords'].apply(collapse)

movies.head()

movie\_id **title overview**

In the

## genres

keyuiords

[cultureclash,

cast

0 19995 Avatar

22nd [Action,

century, a Adventure,

paraplegic Fantasy, Marine is ScienceFiction]

|  |
| --- |
| movies['overview'] = movies['overvi4W!].apply(lambda x:x.split())” |
| Captain  movies [ ' tags ' ] = nov&lT3@&99r@ @W. movies [ ' genres ' ] + mov 'keyword s ' ] + movies['cast'] + movies['crew']  the long [Adventure, drugabuse, [JohnnyDepp, |
| 1 285 Caribbean: believed Fantasy, exoticisland, OrlandoBloom, [Gore\  new = novles . drop( c 4Btrl/l¥Bf | overv1Vg ' genres '  'I jqiraKnightley]  #new. head( ) End dead, |
| new['tags'] = new['tags'].apply(lambda x: “ “.join(x)) new.head() |

future, [SamWorthington,

spacewar, ZoeSaldana, [JamesC spacecolony, SigourneyWeaver]

## mov1e\_1d

**t1t1e**

tags

0 19995 Avatar In the 22nd century, a paraplegic Marine is di...

28s Pirates of the Caribbean: At World's Captain Barbossa, long believed to be dead,

End ha...

|  |  |  |  |
| --- | --- | --- | --- |
| 2 | 206647 | Spectre | A Cryptic message from Bond’s past sends him |
| 3 | 49026 | The Dark Knight Rises | Following the death of District Attorney Harve... |

from sklearn.feature\_extraction.text import CountVectorizer cv = CountVectorizer(max\_features=5000,stop\_words='english')

vector = cv.fit\_tnansform(new['tags']).toarnay()

vector.shape

(4842, 5000)

from sklearn.metrics.pairwise import cosine\_similarity

similarity = cosine\_similarity(vector)

similarity

array([[1. , 0.09107651, 0.06071767, ..., 0.02492224, 0.02817181,

0. ],

[0.09107651, 1.

0. ],

, 0. 66451613,

0.02648136, 0.

[0.06071767, 0.06451613, 1.

0. ],

, 0.02648136, 0.

[0.02492224, 0.02648136, 0.02648136,

0.04783649],

1. 0.07372098,

|  |  |  |  |
| --- | --- | --- | --- |
| [0.02817181, 0. | , 0. | , | 0.07372098, 1. |
| 0.05407381], |  |  |  |
| [0. , 0. | , 0. |  | 0.04783649, 0.05407381, |

new[new['title'] == 'The Lego Movie'].index[0]

744

def recommend(movie):

index = new[new['title'] == movie].iudex[0]

distances = sorted(list(enumerate(similarity[index])),reverse=True,key = lambda x: x[1]) for i in distances[1:6]:

print(new.iloc[i[0]].title)

Double-click (or enter) to edit

# reconnend( 'Avatar' )

Titan A. E . Small Soldiers

# Independence Day Ender ' s Gane

Aliens vs Predator: Requiem

import pickle

pickle.dump(new,open('movie\_list.pkl','wb')) pickle.dump(similarity,open('similarity.pkl','wb'))