```
In [1]:
                      numpy
                                   np
                                    pd
                      pandas
In [3]:
           movies
                     = pd.read csv
In [4]:
           movies.head
Out[4]:
                   budget
                                  genres
                                                                            homepage
                                                                                                   keywords
                                                                                                       [{"id":
                                [{"id": 28,
                                                                                                       1463,
                                  "name":
                                                                                                     "name":
               237000000
                                "Action"},
                                                           http://www.avatarmovie.com/
                                                                                          19995
                                                                                                     "culture
                                {"id": 12,
                                                                                                     clash"},
                                  "nam...
                                                                                                      {"id":...
                                                                                                       [{"id":
                                [{"id": 12,
                                                                                                         270,
                                                                                                     "name":
                                 "name":
           1 300000000
                                             http://disney.go.com/disneypictures/pirates/
                                                                                             285
                            "Adventure"},
                                                                                                   "ocean"},
                             {"id": 14, "...
                                                                                                   {"id": 726,
                                                                                                        "na...
                                                                                                       [{"id":
                                [{"id": 28,
                                                                                                        470,
                                  "name":
                                                                                                     "name":
              245000000
                                "Action"},
                                           http://www.sonypictures.com/movies/spectre/
                                                                                         206647
                                                                                                      "spy"},
                                {"id": 12,
                                                                                                   {"id": 818,
                                  "nam...
                                                                                                     "name...
                                                                                                       [{"id":
                                [{"id": 28,
                                                                                                        849,
                                  "name":
                                                                                                     "name":
              250000000
                                "Action"},
                                                     http://www.thedarkknightrises.com/
                                                                                          49026
                                                                                                         "dc
                                {"id": 80,
                                                                                                   comics"},
                                  "nam...
                                                                                                       {"id":
                                                                                                       853,...
                                                                                                       [{"id":
                                [{"id": 28,
                                                                                                         818,
                                  "name":
                                                                                                     "name":
              260000000
                                "Action"},
                                                    http://movies.disney.com/john-carter
                                                                                          49529
                                                                                                   "based on
                                {"id": 12,
                                                                                                     novel"},
                                  "nam...
                                                                                                      {"id":...
In [6]:
           credits
                          pd.read csv
In [7]:
           credits.head
```

Out[7]:		movie_id	title	cast	crew
	0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "	[{"credit_id": "52fe48009251416c750aca23", "de
	1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa	[{"credit_id": "52fe4232c3a36847f800b579", "de
	2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr	[{"credit_id": "54805967c3a36829b5002c41", "de
	3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"credit_id": "52fe4781c3a36847f81398c3", "de
	4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "c	[{"credit_id": "52fe479ac3a36847f813eaa3", "de
In [11]:	mo	vies1 = r	movies.merge(c	redits,on='title')	
In [12]:	mo	vies1.hea	ad()		

Out [12]

2]:		budget	genres	homepage	id	keywords	(
	0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":	
	1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na	
	2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name	
	3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,	
	4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":	

5 rows × 23 columns

Now we will ask ourselves which columns to pick ( will they be useful for genreating tags or not)

- 1. Genres
- 2. id
- 3. keywords
- 4. original language (here we have a little imbalanced data) so will discard this
- 5. original title or title
- 6. Overview ( as content based LSTM we are using , we will pick this)
- 7. popularity (same as above reason, we will discard this), companies, countries, release date, revenue, runtime, spoken lang, status, tagline, vote average, vote count,
- 8. Cast n Crew ( we will pick)

```
In [13]: movies2 = movies1[['movie_id',
    'title','overview','genres','keywords','cast','crew']]
In [15]: movies2.head()
```

Out[15]:		movie_id	title	overview	genres	keywords	cast	
	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", "	"52fe48009251416
	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	"52fe4232c3a3684
	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	"54805967c3a3682
	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	"52fe4781c3a3684
	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	"52fe479ac3a3684

Now we will do data preprocessing which includes (filling missing values, removal of duplicates)

```
Out[26]:
In [27]:
         movies3.duplicated().sum
Out[27]:
In [28]:
         movies3.iloc[0].genres
Out[28]:
In [29]:
                 ast
In [32]:
             convert(obj):
              for i in ast.literal_eval(obj):
                  l.append(i['name'])
In [34]:
         movies3
                               movies3
                                                 apply(convert
In [36]:
         movies3
                               = movies3
                                                      .apply(convert
In [37]:
         movies3.head(
```

movie\_id title overview genres keywords Out[37]: cast [culture In the [Action, clash, [{"cast\_id": 22nd Adventure, future, 242, [{' century, a 0 19995 "52fe48009251416c7 Avatar Fantasy, space "character": paraplegic Science war, "Jake Sully", Marine is Fiction] space di... colon... [ocean, Captain Pirates of [{"cast\_id": drug Barbossa, the [Adventure, abuse, long "character": 285 Caribbean: Fantasy, "52fe4232c3a36847f exotic believed At World's "Captain Action] island, to be End east india Jack Spa... dead, ha... A cryptic [spy, [{"cast\_id": message based on from [Action, novel, [{' 2 206647 "54805967c3a36829k Spectre Bond's Adventure, secret "character": past Crime] agent, "James Bond", "cr... sends him sequel, o... mi... [{"cast\_id": [dc Following comics, [Action, 2. The Dark the death [{' "character": Crime, crime 3 49026 Knight of District "52fe4781c3a36847 "Bruce Drama, fighter, Rises Attorney Thriller] terrorist, Wayne / Harve... secret i... Ва... John [based on [{"cast\_id": Carter is a [Action, novel, 5, war-John Adventure, mars, "character": 4 49529 "52fe479ac3a36847 weary, Carter Science medallion, "John former Fiction] space Carter", military travel... "c... ca...

In [41]:

movies3.head

movie\_id title overview genres keywords Out[41]: cast [culture In the [Action, clash, [Sam 22nd Adventure, future, Worthington, [{ century, a 0 19995 Fantasy, "52fe48009251416c Avatar space Zoe Saldana, paraplegic Science war, Sigourney Marine is Fiction] space Weaver] di... colon... [ocean, Captain Pirates of [Johnny drug Barbossa, Depp, the [Adventure, abuse, long 285 Caribbean: Fantasy, Orlando "52fe4232c3a36847 exotic believed At World's Bloom, Keira Action] island, to be End east india Knightley] dead, ha... A cryptic [spy, message based on [Daniel from [Action, novel, Craig, [{ 2 206647 "54805967c3a36829 Spectre Bond's Adventure, secret Christoph past Crime] agent, Waltz, Léa sends him sequel, Seydoux] o... mi... [dc Following [Christian [Action, comics, The Dark the death Bale, [{ Crime, crime 3 49026 Knight of District Michael "52fe4781c3a36847 Drama, fighter, Rises Attorney Caine, Gary Thriller] terrorist, Harve... Oldman] secret i... John [based on Carter is a [Taylor [Action, novel, Kitsch, Lynn war-John Adventure, mars, 4 49529 Collins, "52fe479ac3a36841 weary, Carter Science medallion, Samantha former Fiction] space military Morton] travel... ca...

Out[45]:		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 [46]:		vies3['		]=movies3	['overvie	w'].apply	(lambda x	:		

```
In
```

In [47]: movies3.head(

movie\_id title overview keywords Out[47]: genres cast crew [culture [In, the, [Action, **[Sam** clash, 22nd, Adventure, Worthington, future, [James 0 19995 Avatar Zoe Saldana, century,, a, Fantasy, space war, Cameron] Science Sigourney paraplegic, space Fiction] Weaver] Marin... colon... [ocean, [Captain, Pirates of [Johnny drug [Adventure, the Barbossa,, Depp, [Gore abuse, 1 285 Caribbean: Fantasy, Orlando long, Verbinski] exotic At World's Bloom, Keira believed, Action] island, east Knightley] End to, be, d... india ... [A, cryptic, [spy, based message, on novel, [Daniel Craig, [Action, from, secret Christoph [Sam 2 206647 Spectre Adventure, Waltz, Léa Bond's, agent, Mendes] Crime] past, sequel, Seydoux] send... mi... [dc comics, [Following, [Action, [Christian The Dark crime the, death, Crime, Bale, Michael [Christopher 3 49026 Knight fighter, of, District, Drama, Caine, Gary Nolan] Rises terrorist, Attorney... Thriller] Oldman] secret i... [John, [based on [Taylor Carter, is, [Action, novel, Kitsch, Lynn Adventure, [Andrew John a, warmars, 4 49529 Collins, Carter Science Stanton] medallion, weary,, Samantha Fiction] former, space Morton] mili... travel... In [51]: movies3 =movies3 i.replace .apply i in x] keywords'].apply(lambda x : movies3 keywords']=movies3| i.replace "") for i in x]) cast']=movies3['cast'].apply(lambda x : [i.replace( ]=movies3['crew'].apply(lambda x : [i.replace(

```
In [52]: movies3.head()
```

Out[52]:		movie_id	title	overview	genres	keywords	cast	
	0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, 	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[Jam
	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]	[Go
	2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6,	[DanielCraig, ChristophWaltz, LéaSeydoux]	[5
	3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden	[ChristianBale, MichaelCaine, GaryOldman]	[Christo
	4	49529	John Carter	[John, Carter, is, a, war- weary,, former, mili	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p	[TaylorKitsch, LynnCollins, SamanthaMorton]	[And
In [53]:	mo	vies3[ˈ	tags'l =	movies3	'overview'l	+ movies3	genres'l +	

```
In [53]: movies3['tags'] = movies3['overview'] + movies3['genres'] +
    movies3['keywords'] + movies3['cast'] + movies3['crew']
In [54]: movies3.head()
```

```
movie_id
                                 title
Out[54]:
                                        overview
                                                           genres
                                                                         keywords
                                                                                                 cast
                                          [In, the,
                                                                     [cultureclash,
                                                          [Action,
                                            22nd,
                                                                            future,
                                                                                    [SamWorthington,
                                                        Adventure,
            0
                   19995
                               Avatar century,, a,
                                                                         spacewar,
                                                                                          ZoeSaldana,
                                                                                                         [Jam
                                                          Fantasy,
                                       paraplegic,
                                                                      spacecolony,
                                                                                    SigourneyWeaver]
                                                   ScienceFiction]
                                          Marin...
                            Pirates of
                                         [Captain,
                                                                           [ocean,
                                  the
                                                       [Adventure,
                                                                                        [JohnnyDepp,
                                       Barbossa,,
                                                                        drugabuse,
            1
                     285
                          Caribbean:
                                                          Fantasy,
                                                                                       OrlandoBloom,
                                                                                                           [Go
                                            long,
                                                                      exoticisland,
                           At World's
                                                                                       KeiraKnightley]
                                         believed,
                                                           Action]
                                                                    eastindiatrad...
                                        to, be, d...
                                 End
                                       [A, cryptic,
                                        message,
                                                                              [spy,
                                                          [Action,
                                                                                         [DanielCraig,
                                            from,
                                                                     basedonnovel,
            2
                                                                                                            [S
                 206647
                                                                                       ChristophWaltz,
                             Spectre
                                                        Adventure,
                                          Bond's,
                                                                      secretagent,
                                                                                         LéaSeydoux]
                                                           Crime]
                                             past,
                                                                     sequel, mi6, ...
                                           send...
                                                                        [dccomics.
                                       [Following,
                            The Dark
                                                    [Action, Crime,
                                                                                        [ChristianBale,
                                                                      crimefiahter.
                                       the, death,
            3
                  49026
                               Knight
                                                           Drama,
                                                                                        MichaelCaine,
                                                                                                        [Christe
                                       of, District,
                                                                          terrorist.
                                Rises
                                                          Thriller]
                                                                                         GaryOldman]
                                        Attorney...
                                                                       secretiden...
                                           [John,
                                                                    [basedonnovel,
                                        Carter, is,
                                                          [Action,
                                                                                         [TaylorKitsch,
                                                                             mars,
                                John
                                           a, war-
                  49529
                                                       Adventure,
                                                                         medallion,
                                                                                          LynnCollins,
                                                                                                         [And
                               Carter
                                           weary,,
                                                   ScienceFiction]
                                                                       spacetravel,
                                                                                     SamanthaMorton]
                                          former,
                                                                               р...
                                            mili...
In [57]:
            movies_final = movies3[['movie_id','title','tags']]
            movies final['tags']=movies final['tags'].apply(lambda x : "
              .join(x)
In [59]:
            movies final
                                 tags']=movies final['tags'].apply
            x.lower
In [61]:
            movies final.head
```

tags	title	movie_id		Out[61]:	
in the 22nd century, a paraplegic marine is di	Avatar	19995	0		
captain barbossa, long believed to be dead, ha	Pirates of the Caribbean: At World's End	285	1		
a cryptic message from bond's past sends him o	Spectre	206647	2		
following the death of district attorney harve	The Dark Knight Rises	49026	3		
john carter is a war-weary, former military ca	John Carter	49529	4		

Now for the model, we gotta calucalte similarity between texts, such that our engine will be able to recommend easily, but how to check similarity, i guess we can by similarity score, but here similarity can't be calculate as it is textual data ( string form ) [tags] So for this we will do vectorisation of text ( bag of words ) We won't consider stop words here

```
In [75]:
              sklearn.feature extraction.text import CountVectorizer
              CountVectorizer(max_features=5000,stop_words='english')
In [76]:
             ort nltk
              nltk.stem.porter import PorterStemmer
         ps=PorterStemmer()
             stem(text):
             y=[]
             for i in text.split():
                 y.append(ps.stem(i))
                        .join(y)
In [77]:
         movies final
                               = movies final
                                                      l.apply(stem
           able/user_guide/indexing.html#returning-a-view-versus-a-copy
In [78]:
         cv.fit transform(movies final[
                                          tags']).toarray().shape
Out[78]:
In [79]:
         vectors = cv.fit transform(movies final
                                                          ]).toarray
```

06/12/2022,18:26 MRE
In [80]: vectors[0]
Out[80]: array([0, 0, 0, ..., 0, 0])

In [81]:

cv.get\_feature\_names

```
Out[81]:
```

Higher dimensionality leads to failure of euclidian distance or manhola distances, so here we will use cosine similarity

```
In [82]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [84]:
         cosine similarity(vectors)
Out[84]:
In [87]:
         simi=cosine similarity(vectors)
In [90]:
         simi[1]
Out[90]:
In [92]:
         movies final[movies final]
Out[92]:
            movie_id
                      title
                                                         tags
              19995 Avatar in the 22nd century, a parapleg marin is dispa...
In [102...
              recommend(movie):
              movie index =
         movies final[movies final['title']==movie].index[0]
              distances = simi[movie index]
              movies_list = sorted(list(enumerate(distances)),reverse=Erue,
         key = lambda x:x[1])[1:6]
              for i in movies list:
                  print(movies final.iloc[i[0]].title)
In [107...
         recommend(
In [108...
            port pickle
         pickle.dump(movies final,open('movies.pkl','wb'))
In [109...
         pickle.dump(simi,open(
```

Now code the below lines on Pycharm for deploying your streamlit app locally and then via streamlit cloud, heroku, deta, vercel, you can upload it publically

```
In []:
        # def fetch poster(movie id):
```

```
# with col1:
# st.text(names[0])
# st.image(posters[0])

# with col2:
# st.text(names[1])
# st.image(posters[1])

# with col3:
# st.text(names[2])
# st.image(posters[2])

# with col4:
# st.text(names[3])
# st.image(posters[3])

# with col5:
# st.text(names[4])
# st.image(posters[4])
```