8 . Jupyter Code

```
import numpy as np # linear algebra
import pandas as pd


movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')

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

movies =
movies[['movie_id','title','overview','genres','keywords','cast','crew
']]

import ast

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

movies.dropna(inplace=True)

def convert3(text):
    L = []
    counter = 0
    for i in ast.literal_eval(text):
        if counter < 3:
            L.append(i['name'])
        counter+=1
    return L

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'] == 'Director':
            L.append(i['name'])
    return L

movies['crew'] = movies['crew'].apply(fetch_director)

def collapse(L):
    L1 = []
    for i in L:
        L1.append(i.replace(" ",""))
    return L1

movies['cast'] = movies['cast'].apply(collapse)
movies['crew'] = movies['crew'].apply(collapse)
```

8 . Jupyter Code

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

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

movies['tags'] = movies['overview'] + movies['genres'] +
movies['keywords'] + movies['cast'] + movies['crew']


new =
movies.drop(columns=['overview','genres','keywords','cast','crew'])

from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000,stop_words='english')


vector = cv.fit_transform(new['tags']).toarray()

from sklearn.metrics.pairwise import cosine_similarity

similarity = cosine_similarity(vector)

def recommend(movie):
    index = new[new['title'] == movie].index[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)

import pickle

#import pickle # imports pickle module

# pickle.dump(new,open('movie_list.pkl','wb')) # removed this line
because the variable new was not defined in the user's environment.
#pickle.dump(similarity,open('similarity.pkl','wb')) # changed pikle
to pickle

pickle.dump(new,open('movie_list.pkl','wb'))
pickle.dump(similarity,open('similarity.pkl','wb'))
```