recommendation-system

March 10, 2025

[ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

[ ]:
```
import pandas as pd
file_path='/content/Movie.csv'
df = pd.read_csv(file_path)
```

[ ]:
```
df.head()
```

[ ]:

	userId	movie	rating
0	3	Toy Story (1995)	4.0
1	6	Toy Story (1995)	5.0
2	8	Toy Story (1995)	4.0
3	10	Toy Story (1995)	4.0
4	11	Toy Story (1995)	4.5

[ ]:
```
df
```

[ ]:

	userId	movie	rating
0	3	Toy Story (1995)	4.0
1	6	Toy Story (1995)	5.0
2	8	Toy Story (1995)	4.0
3	10	Toy Story (1995)	4.0
4	11	Toy Story (1995)	4.5
…	…	…	…
8987	7087	GoldenEye (1995)	3.0
8988	7088	GoldenEye (1995)	1.0
8989	7105	GoldenEye (1995)	2.0
8990	7113	GoldenEye (1995)	3.0
8991	7117	GoldenEye (1995)	3.0

```
[8992 rows x 3 columns]
```

[ ]:
```
def cosine_similarity(movie1, movie2):
  return df.loc[movie1, movie2]
```

```
[ ]: def correlation_similarity(movie1, movie2):
       return df.loc[movie1, movie2]
```

```
[ ]: movies_df=pd.read_csv(file_path)
```

```
[ ]: movies_df[0:5]
```

```
[ ]:    userId             movie  rating
     0       3  Toy Story (1995)     4.0
     1       6  Toy Story (1995)     5.0
     2       8  Toy Story (1995)     4.0
     3      10  Toy Story (1995)     4.0
     4      11  Toy Story (1995)     4.5
```

```
[ ]: len(movies_df.userId.unique())
```

```
[ ]: 4081
```

```
[ ]: len(movies_df.movie.unique())
```

```
[ ]: 10
```

```
[ ]: user_movies_df = movies_df.pivot(index='userId',
                                    columns='movie',
                                    values='rating').reset_index(drop=True)
```

```
[ ]: user_movies_df
```

```
[ ]: movie  Father of the Bride Part II (1995)  GoldenEye (1995)  \
     0                                     NaN               NaN
     1                                     NaN               NaN
     2                                     NaN               NaN
     3                                     NaN               4.0
     4                                     NaN               NaN
     …                                     …                 …
     4076                                  4.0               NaN
     4077                                  3.5               NaN
     4078                                  NaN               3.0
     4079                                  NaN               NaN
     4080                                  NaN               NaN

     movie  Grumpier Old Men (1995)  Heat (1995)  Jumanji (1995)  Sabrina (1995)  \
     0                          NaN          NaN             3.5             NaN
     1                          4.0          NaN             NaN             NaN
     2                          NaN          NaN             NaN             NaN
     3                          NaN          3.0             NaN             NaN
     4                          NaN          NaN             3.0             NaN
```

```
…                          …           …              …              …
4076                         NaN         NaN            NaN            NaN
4077                         NaN         NaN            NaN            NaN
4078                         4.0         5.0            NaN            3.0
4079                         NaN         NaN            NaN            NaN
4080                         NaN         NaN            4.0            4.0

movie  Sudden Death (1995)  Tom and Huck (1995)  Toy Story (1995)  \
0                        NaN                  NaN               NaN
1                        NaN                  NaN               NaN
2                        NaN                  NaN               4.0
3                        NaN                  NaN               NaN
4                        NaN                  NaN               NaN
…                        …                    …                 …
4076                     NaN                  NaN               NaN
4077                     NaN                  NaN               4.0
4078                     1.0                  NaN               4.0
4079                     NaN                  NaN               5.0
4080                     NaN                  NaN               4.5

movie  Waiting to Exhale (1995)
0                            NaN
1                            NaN
2                            NaN
3                            NaN
4                            NaN
…                            …
4076                         NaN
4077                         NaN
4078                         NaN
4079                         NaN
4080                         NaN

[4081 rows x 10 columns]
```

[ ]:
```
user_movies_df.index = movies_df.userId.unique()
user_movies_df
```

[ ]:
```
movie  Father of the Bride Part II (1995)  GoldenEye (1995)  \
3                                      NaN               NaN
6                                      NaN               NaN
8                                      NaN               NaN
10                                     NaN               4.0
11                                     NaN               NaN
…                                      …                 …
7044                                   4.0               NaN
7070                                   3.5               NaN
```

```
7080                                NaN               3.0
7087                                NaN               NaN
7105                                NaN               NaN

movie  Grumpier Old Men (1995)  Heat (1995)  Jumanji (1995)  Sabrina (1995)  \
3                          NaN          NaN             3.5             NaN
6                          4.0          NaN             NaN             NaN
8                          NaN          NaN             NaN             NaN
10                         NaN          3.0             NaN             NaN
11                         NaN          NaN             3.0             NaN
…                          …            …               …               …
7044                       NaN          NaN             NaN             NaN
7070                       NaN          NaN             NaN             NaN
7080                       4.0          5.0             NaN             3.0
7087                       NaN          NaN             NaN             NaN
7105                       NaN          NaN             4.0             4.0

movie  Sudden Death (1995)  Tom and Huck (1995)  Toy Story (1995)  \
3                      NaN                  NaN               NaN
6                      NaN                  NaN               NaN
8                      NaN                  NaN               4.0
10                     NaN                  NaN               NaN
11                     NaN                  NaN               NaN
…                      …                    …                 …
7044                   NaN                  NaN               NaN
7070                   NaN                  NaN               4.0
7080                   1.0                  NaN               4.0
7087                   NaN                  NaN               5.0
7105                   NaN                  NaN               4.5

movie  Waiting to Exhale (1995)
3                            NaN
6                            NaN
8                            NaN
10                           NaN
11                           NaN
…                            …
7044                         NaN
7070                         NaN
7080                         NaN
7087                         NaN
7105                         NaN

[4081 rows x 10 columns]
```

[ ]:
```
user_movies_df.fillna(0, inplace=True)
user_movies_df
```

```
[ ]: movie  Father of the Bride Part II (1995)  GoldenEye (1995)  \
     3                                     0.0               0.0
     6                                     0.0               0.0
     8                                     0.0               0.0
     10                                    0.0               4.0
     11                                    0.0               0.0
     …                                     …                 …
     7044                                  4.0               0.0
     7070                                  3.5               0.0
     7080                                  0.0               3.0
     7087                                  0.0               0.0
     7105                                  0.0               0.0

     movie  Grumpier Old Men (1995)  Heat (1995)  Jumanji (1995)  Sabrina (1995)  \
     3                          0.0          0.0             3.5             0.0
     6                          4.0          0.0             0.0             0.0
     8                          0.0          0.0             0.0             0.0
     10                         0.0          3.0             0.0             0.0
     11                         0.0          0.0             3.0             0.0
     …                          …            …               …               …
     7044                       0.0          0.0             0.0             0.0
     7070                       0.0          0.0             0.0             0.0
     7080                       4.0          5.0             0.0             3.0
     7087                       0.0          0.0             0.0             0.0
     7105                       0.0          0.0             4.0             4.0

     movie  Sudden Death (1995)  Tom and Huck (1995)  Toy Story (1995)  \
     3                      0.0                  0.0               0.0
     6                      0.0                  0.0               0.0
     8                      0.0                  0.0               4.0
     10                     0.0                  0.0               0.0
     11                     0.0                  0.0               0.0
     …                      …                    …                 …
     7044                   0.0                  0.0               0.0
     7070                   0.0                  0.0               4.0
     7080                   1.0                  0.0               4.0
     7087                   0.0                  0.0               5.0
     7105                   0.0                  0.0               4.5

     movie  Waiting to Exhale (1995)
     3                            0.0
     6                            0.0
     8                            0.0
     10                           0.0
     11                           0.0
     …                            …
     7044                         0.0
```

```
7070                    0.0
7080                    0.0
7087                    0.0
7105                    0.0

[4081 rows x 10 columns]
```

[ ]:
```
from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import cosine, correlation
```

[ ]:
```
from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import cosine, correlation
```

[ ]:
```
vec1 = [5, 4, 0, 0]
vec2 = [4, 5, 3, 0]

correlation_distance = correlation(vec1, vec2)
print(correlation_distance)
```

```
0.2372709564964851
```

[ ]:
```
user_sim = 1 - pairwise_distances( user_movies_df.values,metric='cosine')
```

[ ]:
```
user_sim
```

[ ]:
```
array([[1.        , 0.        , 0.        , …, 0.        , 0.        ,
        0.55337157],
       [0.        , 1.        , 0.        , …, 0.45883147, 0.        ,
        0.        ],
       [0.        , 0.        , 1.        , …, 0.45883147, 1.        ,
        0.62254302],
       …,
       [0.        , 0.45883147, 0.45883147, …, 1.        , 0.45883147,
        0.47607054],
       [0.        , 0.        , 1.        , …, 0.45883147, 1.        ,
        0.62254302],
       [0.55337157, 0.        , 0.62254302, …, 0.47607054, 0.62254302,
        1.        ]])
```

[ ]:
```
user_sim.shape
```

[ ]:
```
(4081, 4081)
```

[ ]:
```
user_sim_df = pd.DataFrame(user_sim)
```

[ ]:
```
user_sim_df.iloc[0:5,0:5]
```

[ ]:

	0	1	2	3	4
0	1.0	0.0	0.0	0.0	1.0
1	0.0	1.0	0.0	0.0	0.0
2	0.0	0.0	1.0	0.0	0.0
3	0.0	0.0	0.0	1.0	0.0
4	1.0	0.0	0.0	0.0	1.0

[ ]:
```
user_sim_df
```

[ ]:

	0	1	2	3	4	5	6	\
0	1.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	
1	0.000000	1.000000	0.000000	0.000000	0.000000	0.390567	0.707107	
2	0.000000	0.000000	1.000000	0.000000	0.000000	0.650945	0.000000	
3	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	
4	1.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	
…	…	…	…	…	…	…	…	
4076	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
4077	0.000000	0.000000	0.752577	0.000000	0.000000	0.489886	0.000000	
4078	0.000000	0.458831	0.458831	0.619422	0.000000	0.701884	0.567775	
4079	0.000000	0.000000	1.000000	0.000000	0.000000	0.650945	0.000000	
4080	0.553372	0.000000	0.622543	0.000000	0.553372	0.765455	0.391293	

	7	8	9	…	4071	4072	4073	\
0	0.000000	0.000000	0.000000	…	0.000000	0.000000	1.000000	
1	0.615457	0.000000	0.000000	…	0.000000	0.000000	0.000000	
2	0.492366	1.000000	0.874157	…	0.000000	1.000000	0.000000	
3	0.615457	0.000000	0.388514	…	0.800000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	…	0.000000	0.000000	1.000000	
…	…	…	…	…	…	…	…	
4076	0.000000	0.000000	0.000000	…	0.000000	0.000000	0.000000	
4077	0.370543	0.752577	0.657870	…	0.000000	0.752577	0.000000	
4078	0.889532	0.458831	0.568212	…	0.344124	0.458831	0.000000	
4079	0.492366	1.000000	0.874157	…	0.000000	1.000000	0.000000	
4080	0.306519	0.622543	0.544201	…	0.000000	0.622543	0.553372	

	4074	4075	4076	4077	4078	4079	4080
0	0.707107	0.000000	0.000000	0.000000	0.000000	0.000000	0.553372
1	0.000000	0.000000	0.000000	0.000000	0.458831	0.000000	0.000000
2	0.707107	0.000000	0.000000	0.752577	0.458831	1.000000	0.622543
3	0.000000	0.989949	0.000000	0.000000	0.619422	0.000000	0.000000
4	0.707107	0.000000	0.000000	0.000000	0.000000	0.000000	0.553372
…	…	…	…	…	…	…	…
4076	0.000000	0.000000	1.000000	0.658505	0.000000	0.000000	0.000000
4077	0.532152	0.000000	0.658505	1.000000	0.345306	0.752577	0.468511
4078	0.324443	0.648886	0.000000	0.345306	1.000000	0.458831	0.476071
4079	0.707107	0.000000	0.000000	0.752577	0.458831	1.000000	0.622543
4080	0.831497	0.000000	0.000000	0.468511	0.476071	0.622543	1.000000

```
[4081 rows x 4081 columns]
```

[ ]:
```
user_sim_df.index = movies_df.userId.unique()
```

[ ]:
```
user_sim_df
```

[ ]:
```
          3         6         8         10        11        12        13    \
3     1.000000  0.000000  0.000000  0.000000  1.000000  0.000000  0.000000
6     0.000000  1.000000  0.000000  0.000000  0.000000  0.390567  0.707107
8     0.000000  0.000000  1.000000  0.000000  0.000000  0.650945  0.000000
10    0.000000  0.000000  0.000000  1.000000  0.000000  0.000000  0.000000
11    1.000000  0.000000  0.000000  0.000000  1.000000  0.000000  0.000000
…        …         …         …         …         …         …         …
7044  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
7070  0.000000  0.000000  0.752577  0.000000  0.000000  0.489886  0.000000
7080  0.000000  0.458831  0.458831  0.619422  0.000000  0.701884  0.567775
7087  0.000000  0.000000  1.000000  0.000000  0.000000  0.650945  0.000000
7105  0.553372  0.000000  0.622543  0.000000  0.553372  0.765455  0.391293

          14        16        19    …      6975      6979      6993  \
3     0.000000  0.000000  0.000000  …  0.000000  0.000000  1.000000
6     0.615457  0.000000  0.000000  …  0.000000  0.000000  0.000000
8     0.492366  1.000000  0.874157  …  0.000000  1.000000  0.000000
10    0.615457  0.000000  0.388514  …  0.800000  0.000000  0.000000
11    0.000000  0.000000  0.000000  …  0.000000  0.000000  1.000000
…        …         …         …    …     …         …         …
7044  0.000000  0.000000  0.000000  …  0.000000  0.000000  0.000000
7070  0.370543  0.752577  0.657870  …  0.000000  0.752577  0.000000
7080  0.889532  0.458831  0.568212  …  0.344124  0.458831  0.000000
7087  0.492366  1.000000  0.874157  …  0.000000  1.000000  0.000000
7105  0.306519  0.622543  0.544201  …  0.000000  0.622543  0.553372

          7030      7031      7044      7070      7080      7087      7105
3     0.707107  0.000000  0.000000  0.000000  0.000000  0.000000  0.553372
6     0.000000  0.000000  0.000000  0.000000  0.458831  0.000000  0.000000
8     0.707107  0.000000  0.000000  0.752577  0.458831  1.000000  0.622543
10    0.000000  0.989949  0.000000  0.000000  0.619422  0.000000  0.000000
11    0.707107  0.000000  0.000000  0.000000  0.000000  0.000000  0.553372
…        …         …         …         …         …         …         …
7044  0.000000  0.000000  1.000000  0.658505  0.000000  0.000000  0.000000
7070  0.532152  0.000000  0.658505  1.000000  0.345306  0.752577  0.468511
7080  0.324443  0.648886  0.000000  0.345306  1.000000  0.458831  0.476071
7087  0.707107  0.000000  0.000000  0.752577  0.458831  1.000000  0.622543
7105  0.831497  0.000000  0.000000  0.468511  0.476071  0.622543  1.000000

[4081 rows x 4081 columns]
```

[ ]:
```
user_sim_df.columns = movies_df.userId.unique()
user_sim_df
```

[ ]:
```
          3         6         8         10        11        12        13    \
3     1.000000  0.000000  0.000000  0.000000  1.000000  0.000000  0.000000
6     0.000000  1.000000  0.000000  0.000000  0.000000  0.390567  0.707107
8     0.000000  0.000000  1.000000  0.000000  0.000000  0.650945  0.000000
10    0.000000  0.000000  0.000000  1.000000  0.000000  0.000000  0.000000
11    1.000000  0.000000  0.000000  0.000000  1.000000  0.000000  0.000000
…          …         …         …         …         …         …         …
7044  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
7070  0.000000  0.000000  0.752577  0.000000  0.000000  0.489886  0.000000
7080  0.000000  0.458831  0.458831  0.619422  0.000000  0.701884  0.567775
7087  0.000000  0.000000  1.000000  0.000000  0.000000  0.650945  0.000000
7105  0.553372  0.000000  0.622543  0.000000  0.553372  0.765455  0.391293

          14        16        19    …      6975      6979      6993  \
3     0.000000  0.000000  0.000000  …  0.000000  0.000000  1.000000
6     0.615457  0.000000  0.000000  …  0.000000  0.000000  0.000000
8     0.492366  1.000000  0.874157  …  0.000000  1.000000  0.000000
10    0.615457  0.000000  0.388514  …  0.800000  0.000000  0.000000
11    0.000000  0.000000  0.000000  …  0.000000  0.000000  1.000000
…          …         …         …  …         …         …
7044  0.000000  0.000000  0.000000  …  0.000000  0.000000  0.000000
7070  0.370543  0.752577  0.657870  …  0.000000  0.752577  0.000000
7080  0.889532  0.458831  0.568212  …  0.344124  0.458831  0.000000
7087  0.492366  1.000000  0.874157  …  0.000000  1.000000  0.000000
7105  0.306519  0.622543  0.544201  …  0.000000  0.622543  0.553372

          7030      7031      7044      7070      7080      7087      7105
3     0.707107  0.000000  0.000000  0.000000  0.000000  0.000000  0.553372
6     0.000000  0.000000  0.000000  0.000000  0.458831  0.000000  0.000000
8     0.707107  0.000000  0.000000  0.752577  0.458831  1.000000  0.622543
10    0.000000  0.989949  0.000000  0.000000  0.619422  0.000000  0.000000
11    0.707107  0.000000  0.000000  0.000000  0.000000  0.000000  0.553372
…          …         …         …         …         …         …         …
7044  0.000000  0.000000  1.000000  0.658505  0.000000  0.000000  0.000000
7070  0.532152  0.000000  0.658505  1.000000  0.345306  0.752577  0.468511
7080  0.324443  0.648886  0.000000  0.345306  1.000000  0.458831  0.476071
7087  0.707107  0.000000  0.000000  0.752577  0.458831  1.000000  0.622543
7105  0.831497  0.000000  0.000000  0.468511  0.476071  0.622543  1.000000

[4081 rows x 4081 columns]
```

[ ]:
```
user_sim_df.iloc[0:5,0:5]
```

[ ]:
```
      3    6    8    10   11
3   1.0  0.0  0.0  0.0  1.0
6   0.0  1.0  0.0  0.0  0.0
8   0.0  0.0  1.0  0.0  0.0
10  0.0  0.0  0.0  1.0  0.0
11  1.0  0.0  0.0  0.0  1.0
```

[ ]:
```
import numpy as np
np.fill_diagonal(user_sim, 0)
user_sim_df.iloc[0:5, 0:5]
```

[ ]:
```
      3    6    8    10   11
3   0.0  0.0  0.0  0.0  1.0
6   0.0  0.0  0.0  0.0  0.0
8   0.0  0.0  0.0  0.0  0.0
10  0.0  0.0  0.0  0.0  0.0
11  1.0  0.0  0.0  0.0  0.0
```

[ ]:
```
user_sim_df.idxmax(axis=1)
```

[ ]:
```
3         11
6        168
8         16
10      4047
11         3
        …
7044      80
7070    1808
7080     708
7087       8
7105    4110
Length: 4081, dtype: int64
```

[ ]:
```
movies_df[(movies_df['userId']==6)]
```

[ ]:
```
      userId                    movie  rating
1          6         Toy Story (1995)     5.0
3725       6  Grumpier Old Men (1995)     3.0
6464       6           Sabrina (1995)     5.0
```

[ ]:
```
movies_df[(movies_df['userId']==11)]
```

[ ]:
```
      userId              movie  rating
4         11  Toy Story (1995)     4.5
7446      11  GoldenEye (1995)     2.5
```

[ ]:
```
movies_df[(movies_df['userId']==156)]
```

[ ]:
```
      userId                                movie  rating
56       156                    Toy Story (1995)     5.0
2589     156                      Jumanji (1995)     5.0
3741     156             Grumpier Old Men (1995)     2.0
4411     156            Waiting to Exhale (1995)     3.0
4557     156  Father of the Bride Part II (1995)     3.0
5237     156                           Heat (1995)     4.0
6480     156                       Sabrina (1995)     4.0
7247     156                  Sudden Death (1995)     3.0
7470     156                     GoldenEye (1995)     4.0
```

[ ]:
```
movies_df[(movies_df['userId']==6) | (movies_df['userId']==168)]
```

[ ]:
```
      userId                     movie  rating
1          6         Toy Story (1995)     5.0
60       168         Toy Story (1995)     4.5
3725       6  Grumpier Old Men (1995)     3.0
6464       6           Sabrina (1995)     5.0
```

[ ]:
```
user_1=movies_df[(movies_df['userId']==6)]
```

[ ]:
```
user_2=movies_df[(movies_df['userId']==1)]
```

[ ]:
```
user_2.movie
```

[ ]:
```
2569    Jumanji (1995)
Name: movie, dtype: object
```

[ ]:
```
user_1.movie
```

[ ]:
```
1              Toy Story (1995)
3725    Grumpier Old Men (1995)
6464              Sabrina (1995)
Name: movie, dtype: object
```

[ ]:
```
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-50-b85925c54080> in <cell line: 0>()
----> 1 pd.merge(user_1,user_2,on="movie1")

/usr/local/lib/python3.11/dist-packages/pandas/core/reshape/merge.py in␣
 ↪merge(left, right, how, on, left_on, right_on, left_index, right_index, sort,␣
 ↪suffixes, copy, indicator, validate)
    168         )
    169     else:
```

```
--> 170         op = _MergeOperation(
    171             left_df,
    172             right_df,

/usr/local/lib/python3.11/dist-packages/pandas/core/reshape/merge.py in␣
 ↪__init__(self, left, right, how, on, left_on, right_on, left_index,␣
 ↪right_index, sort, suffixes, indicator, validate)
    792             left_drop,
    793             right_drop,
--> 794         ) = self._get_merge_keys()
    795
    796         if left_drop:

/usr/local/lib/python3.11/dist-packages/pandas/core/reshape/merge.py in␣
 ↪_get_merge_keys(self)
   1295                         rk = cast(Hashable, rk)
   1296                         if rk is not None:
-> 1297                             right_keys.append(right.
 ↪_get_label_or_level_values(rk))
   1298                         else:
   1299                             # work-around for␣
 ↪merge_asof(right_index=True)

/usr/local/lib/python3.11/dist-packages/pandas/core/generic.py in␣
 ↪_get_label_or_level_values(self, key, axis)
   1909             values = self.axes[axis].get_level_values(key)._values
   1910         else:
-> 1911             raise KeyError(key)
   1912
   1913         # Check for duplicates

KeyError: 'movie1'
```

[ ]: