

```
In [1]: import pandas as pd
import numpy as np

In [3]: df = pd.read_csv('book.csv', encoding='latin1')

In [4]: df.head()

Out[4]:
   Unnamed: 0  User.ID  Book.Title  Book.Rating
0           1  276726      Classical Mythology      5
1           2  276729      Clara Callan      3
2           3  276729  Decision in Normandy      6
3           4  276736  Flu: The Story of the Great Influenza Pandemic...      8
4           5  276737      The Mummies of Urumchi      6

In [5]: book_df=df.rename({"User.ID":"userid", "Book.Title":"bookttl", "Book.Rating":"bookrtg"})
book_df

Out[5]:
   Unnamed: 0  userid  bookttl  bookrtg
0           1  276726      Classical Mythology      5
1           2  276729      Clara Callan      3
2           3  276729  Decision in Normandy      6
3           4  276736  Flu: The Story of the Great Influenza Pandemic...      8
4           5  276737      The Mummies of Urumchi      6
...         ...     ...         ...     ...
9995        9996  162121  American Fried: Adventures of a Happy Eater.      7
9996        9997  162121      Cannibal In Manhattan      9
9997        9998  162121      How to Flirt: A Practical Guide      7
9998        9999  162121      Twilight      8
9999       10000  162129      Kids Say the Darndest Things      6

10000 rows x 4 columns

In [6]: #number of unique users in the dataset
len(book_df.userid.unique())

Out[6]: 2182

In [7]: len(book_df.bookttl.unique())

Out[7]: 9659

In [8]: book_df.duplicated(subset=None, keep='first')

Out[8]:
0      False
1      False
2      False
3      False
4      False
...
9995     False
9996     False
9997     False
9998     False
9999     False
Length: 10000, dtype: bool

In [9]: book_df.drop_duplicates(subset=None, keep='first', inplace=False, ignore_index=False)

Out[9]:
   Unnamed: 0  userid  bookttl  bookrtg
0           1  276726      Classical Mythology      5
1           2  276729      Clara Callan      3
2           3  276729  Decision in Normandy      6
3           4  276736  Flu: The Story of the Great Influenza Pandemic...      8
4           5  276737      The Mummies of Urumchi      6
...         ...     ...         ...     ...
9995        9996  162121  American Fried: Adventures of a Happy Eater.      7
9996        9997  162121      Cannibal In Manhattan      9
9997        9998  162121      How to Flirt: A Practical Guide      7
9998        9999  162121      Twilight      8
9999       10000  162129      Kids Say the Darndest Things      6

10000 rows x 4 columns

In [10]: user_book_df = book_df.pivot_table(index='userid',
      columns='bookttl',
      values='bookrtg').reset_index(drop=True)

In [11]: user_book_df

Out[11]:
   bookttl  Jason, Madison &amp;  Stories;Merril;1985;McClelland &amp;  Other Repairing PC Drives &amp;  '48  'O Au No Keia: Voices from Hawai'i's Mahu and Transgender Communities  ...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR  01-01-00: A Novel of the Millennium  1,40: Mon Thing: The PS: Me C
0      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
1      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
2      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
3      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
4      NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
...     ...         ...         ...         ...         ...         ...         ...
2177     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
2178     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
2179     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
2180     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
2181     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN

2182 rows x 9659 columns

In [12]: user_book_df.index = book_df.userid.unique()

In [13]: user_book_df

Out[13]:
   bookttl  Jason, Madison &amp;  Stories;Merril;1985;McClelland &amp;  Other Repairing PC Drives &amp;  '48  'O Au No Keia: Voices from Hawai'i's Mahu and Transgender Communities  ...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR  01-01-00: A Novel of the Millennium  1,40: Mon Thing: The PS: Me C
276726     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
276729     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
276736     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
276737     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
276744     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
...     ...         ...         ...         ...         ...         ...         ...
162107     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
162109     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
162113     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
162121     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN
162129     NaN      NaN      NaN      NaN      NaN      NaN      NaN      NaN

2182 rows x 9659 columns

In [14]: #Impute those NaNs with 0 values
user_book_df.fillna(0, inplace=True)

In [15]: user_book_df

Out[15]:
   bookttl  Jason, Madison &amp;  Stories;Merril;1985;McClelland &amp;  Other Repairing PC Drives &amp;  '48  'O Au No Keia: Voices from Hawai'i's Mahu and Transgender Communities  ...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR  01-01-00: A Novel of the Millennium  1,40: Mon Thing: The PS: Me C
276726     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
276729     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
276736     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
276737     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
276744     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
...     ...         ...         ...         ...         ...         ...         ...
162107     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
162109     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
162113     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
162121     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0
162129     0.0      0.0      0.0      0.0      0.0      0.0      0.0      0.0

2182 rows x 9659 columns

In [16]: #Calculating Cosine Similarity between Users
from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import cosine, correlation

In [17]: user_sim = 1 - pairwise_distances( user_book_df.values,metric='cosine')

In [18]: user_sim

Out[18]: array([[1., 0., 0., ..., 0., 0., 0.],
      [0., 1., 0., ..., 0., 0., 0.],
      [0., 0., 1., ..., 0., 0., 0.],
      ...,
      [0., 0., 0., ..., 1., 0., 0.],
      [0., 0., 0., ..., 0., 1., 0.],
      [0., 0., 0., ..., 0., 0., 1.]])

In [19]: #Store the results in a dataframe
user_sim_df = pd.DataFrame(user_sim)

In [20]: user_sim_df

Out[20]:
   0  1  2  3  4  5  6  7  8  9  ...  2172  2173  2174  2175  2176  2177  2178  2179  2
0  1.0  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
1  0.0  1.0  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
2  0.0  0.0  1.0  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
3  0.0  0.0  0.0  1.0  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
4  0.0  0.0  0.0  0.0  1.0  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
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
2177  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  1.0  0.0  0.0
2178  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  1.0  0.0
2179  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  1.0
2180  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
2181  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

2182 rows x 2182 columns

In [21]: #Set the index and column names to user ids
user_sim_df.index = book_df.userid.unique()
user_sim_df.columns = book_df.userid.unique()

In [22]: user_sim_df.iloc[0:5, 0:5]

Out[22]:
   276726  276729  276736  276737  276744
276726    1.0    0.0    0.0    0.0    0.0
276729    0.0    1.0    0.0    0.0    0.0
276736    0.0    0.0    1.0    0.0    0.0
276737    0.0    0.0    0.0    1.0    0.0
276744    0.0    0.0    0.0    0.0    1.0

In [23]: np.fill_diagonal(user_sim, 0)
user_sim_df.iloc[0:5, 0:5]

Out[23]:
   276726  276729  276736  276737  276744
276726    0.0    0.0    0.0    0.0    0.0
276729    0.0    0.0    0.0    0.0    0.0
276736    0.0    0.0    0.0    0.0    0.0
276737    0.0    0.0    0.0    0.0    0.0
276744    0.0    0.0    0.0    0.0    0.0

In [24]: #Most Similar Users
user_sim_df.idxmax(axis=1)[0:5]

Out[24]:
276726    276726
276729    276726
276736    276726
276737    276726
276744    276726
dtype: int64

In [25]: book_df[(book_df['userid']==8) | (book_df['userid']==162129)]

Out[25]:
   Unnamed: 0  userid  bookttl  bookrtg
1           2  276729      Clara Callan      3
2           3  276729  Decision in Normandy      6
2399        2400      8      Ancient Celtic Romances      5
2400        2401      8  The Western way: A practical guide to the West...      5
2401        2402      8      Wings      5
2402        2403      8      Truckers      5
2403        2404      8  Keepers of the Earth Teachers Guide      6
2404        2405      8      The Celts Activity Book      6
2405        2406      8      The Art Of Celtia      7

In [26]: book_df[(book_df['userid']==6) | (book_df['userid']==162121)]

Out[26]:
   Unnamed: 0  userid  bookttl  bookrtg
9990        9991  162121      The Cloister Walk      7
9991        9992  162121      Open Water      5
9992        9993  162121      The Evolution of Jane      8
9993        9994  162121      AT PARADISE GATE      8
9994        9995  162121  I Should Have Stayed Home: The Worst Trips of ...      8
9995        9996  162121  American Fried: Adventures of a Happy Eater.      7
9996        9997  162121      Cannibal In Manhattan      9
9997        9998  162121      How to Flirt: A Practical Guide      7
9998        9999  162121      Twilight      8

In [27]: book_df[(book_df['userid']==4) | (book_df['userid']==162129)]

Out[27]:
   Unnamed: 0  userid  bookttl  bookrtg
9999       10000  162129  Kids Say the Darndest Things      6

In [28]: user_1=book_df[book_df['userid']==6]

In [29]: user_2=book_df[book_df['userid']==11]

In [32]: user_1.bookttl

Out[32]: Series([], Name: bookttl, dtype: object)

In [33]: user_2.bookttl

Out[33]: Series([], Name: bookttl, dtype: object)

In [34]: pd.merge(user_1,user_2,on='bookttl',how='outer')

Out[34]:
   Unnamed: 0_x  userid_x  bookrtg_x  Unnamed: 0_y  userid_y  bookttl  bookrtg_y
0           0.0      276726      3.0      0.0      276729      3.0      3.0
1           0.0      276729      6.0      0.0      276729      6.0      6.0
2           0.0      276729      6.0      0.0      276729      6.0      6.0
3           0.0      276729      6.0      0.0      276729      6.0      6.0
4           0.0      276729      6.0      0.0      276729      6.0      6.0
5           0.0      276729      6.0      0.0      276729      6.0      6.0
6           0.0      276729      6.0      0.0      276729      6.0      6.0
7           0.0      276729      6.0      0.0      276729      6.0      6.0
8           0.0      276729      6.0      0.0      276729      6.0      6.0
9           0.0      276729      6.0      0.0      276729      6.0      6.0
10          0.0      276729      6.0      0.0      276729      6.0      6.0
11          0.0      276729      6.0      0.0      276729      6.0      6.0
12          0.0      276729      6.0      0.0      276729      6.0      6.0
13          0.0      276729      6.0      0.0      276729      6.0      6.0
14          0.0      276729      6.0      0.0      276729      6.0      6.0
15          0.0      276729      6.0      0.0      276729      6.0      6.0
16          0.0      276729      6.0      0.0      276729      6.0      6.0
17          0.0      276729      6.0      0.0      276729      6.0      6.0
18          0.0      276729      6.0      0.0      276729      6.0      6.0
19          0.0      276729      6.0      0.0      276729      6.0      6.0
20          0.0      276729      6.0      0.0      276729      6.0      6.0
21          0.0      276729      6.0      0.0      276729      6.0      6.0
22          0.0      276729      6.0      0.0      276729      6.0      6.0
23          0.0      276729      6.0      0.0      276729      6.0      6.0
24          0.0      276729      6.0      0.0      276729      6.0      6.0
25          0.0      276729      6.0      0.0      276729      6.0      6.0
26          0.0      276729      6.0      0.0      276729      6.0      6.0
27          0.0      276729      6.0      0.0      276729      6.0      6.0
28          0.0      276729      6.0      0.0      276729      6.0      6.0
29          0.0      276729      6.0      0.0      276729      6.0      6.0
30          0.0      276729      6.0      0.0      276729      6.0      6.0
31          0.0      276729      6.0      0.0      276729      6.0      6.0
32          0.0      276729      6.0      0.0      276729      6.0      6.0
33          0.0      276729      6.0      0.0      276729      6.0      6.0
34          0.0      276729      6.0      0.0      276729      6.0      6.0
35          0.0      276729      6.0      0.0      276729      6.0      6.0
36          0.0      276729      6.0      0.0      276729      6.0      6.0
37          0.0      276729      6.0      0.0      276729      6.0      6.0
38          0.0      276729      6.0      0.0      276729      6.0      6.0
39          0.0      276729      6.0      0.0      276729      6.0      6.0
40          0.0      276729      6.0      0.0      276729      6.0      6.0
41          0.0      276729      6.0      0.0      276729      6.0      6.0
42          0.0      276729      6.0      0.0      276729      6.0      6.0
43          0.0      276729      6.0      0.0      276729      6.0      6.0
44          0.0      276729      6.0      0.0      276729      6.0      6.0
45          0.0      276729      6.0      0.0      276729      6.0      6.0
46          0.0      276729      6.0      0.0      276729      6.0      6.0
47          0.0      276729      6.0      0.0      276729      6.0      6.0
48          0.0      276729      6.0      0.0      276729      6.0      6.0
49          0.0      276729      6.0      0.0      276729      6.0      6.0
50          0.0      276729      6.0      0.0      276729      6.0      6.0
51          0.0      276729      6.0      0.0      276729      6.0      6.0
52          0.0      276729      6.0      0.0      276729      6.0      6.0
53          0.0      276729      6.0      0.0      276729      6.0      6.0
54          0.0      276729      6.0      0.0      276729      6.0      6.0
55          0.0      276729      6.0      0.0      276729      6.0      6.0
56          0.0      276729      6.0      0.0      276729      6.0      6.0
57          0.0      276729      6.0      0.0      276729      6.0      6.0
58          0.0      276729      6.0      0.0      276729      6.0      6.0
59          0.0      276729      6.0      0.0      276729      6.0      6.0
60          0.0      276729      6.0      0.0      276729      6.0      6.0
61          0.0      276729      6.0      0.0      276729      6.0      6.0
62          0.0      276729      6.0      0.0      276729      6.0      6.0
63          0.0      276729      6.0      0.0      276729      6.0      6.0
64          0.0      276729      6.0      0.0      276729      6.0      6.0
65          0.0      276729      6.0      0.0      276729      6.0      6.0
66          0.0      276729      6.0      0.0      276729      6.0      6.0
67          0.0      276729      6.0      0.0      276729      6.0      6.0
68          0.0      276729      6.0      0.0      276729      6.0      6.0
69          0.0      276729      6.0      0.0      276729      6.0      6.0
70          0.0      276729      6.0      0.0      276729      6.0      6.0
71          0.0      276729      6.0      0.0      276729      6.0      6.0
72          0.0      276729      6.0      0.0      276729      6.0      6.0
73          0.0      276729      6.0      0.0      276729      6.0      6.0
74          0.0      276729      6.0      0.0      276729      6.0      6.0
75          0.0      276729      6.0      0.0      276729      6.0      6.0
76          0.0      276729      6.0      0.0      276729      6.0      6.0
77          0.0      276729      6.0      0.0      276729      6.0      6.0
78          0.0      276729      6.0      0.0      276729      6.0      6.0
79          0.0      276729      6.0      0.0      276729      6.0      6.0
80          0.0      276729      6.0      0.0      276729      6.0      6.0
81          0.0      276729      6.0      0.0      276729      6.0      6.0
82          0.0      276729      6.0      0.0      276729      6.0      6.0
83          0.0      276729      6.0      0.0      276729      6.0      6.0
84          0.0      276729      6.0      0.0      276729      6.0      6.0
85          0.0      276729      6.0      0.0      276729      6.0      6.0
86          0.0      276729      6.0      0.0      276729      6.0      6.0
87          0.0      276729      6.0      0.0      276729      6.0      6.0
88          0.0      276729      6.0      0.0      276729      6.0      6.0
89          0.0      276729      6.0      0.0      276729      6.0      6.0
90          0.0      276729      6.0      0.0      276729      6.0      6.0
91          0.0      276729      6.0      0.0      276729      6.0      6.0
92          0.0      276729      6.0      0.0      276729      6.0      6.0
93          0.0      276729      6.0      0.0      276729      6.0      6.0
94          0.0      276729      6.0      0.0      276729      6.0      6.0
95          0.0      276729      6.0      0.0      276729      6.0      6.0
96          0.0      276729      6.0      0.0      276729      6.0      6.0
97          0.0      276729      6.0      0.0      276729      6.0      6.0
98          0.0      276729      6.0      0.0      276729      6.0      6.0
99          0.0      276729      6.0      0.0      276729      6.0      6.0
100         0.0      276729      6.0      0.0      276729      6.0      6.0
101         0.0      276729      6.0      0.0      276729      6.0      6.0
102         0.0      276729      6.0      0.0      276729      6.0      6.0
103         0.0      276729      6.0      0.0      276729      6.0      6.0
104         0.0      276729      6.0      0.0      276729      6.0      6.0
105         0.0      276729      6.0      0.0      276729      6.0      6.0
106         0.0      276729      6.0      0.0      276729      6.0      6.0
107         0.0      276729      6.0      0.0      276729      6.0      6.0
108         0.0      276729      6.0      0.0      276729      6.0      6.0
109         0.0      276729      6.0      0.0      276729      6.0      6.0
110         0.0      276729      6.0      0.0      276729      6.0      6.0
111         0.0      276729      6.0      0.0      276729      6.0      6.0
112         0.0      276729      6.0      0.0      276729      6.0      6.0
113         0.0      276729      6.0      0.0      276729      6.0      6.0
114         0.0      276729      6.0      0.0      276729      6.0      6.0
115         0.0      276729      6.0      0.0      276729      6.0      6.0
116         0.0      276729      6.0      0.0      276729      6.0      6.0
117         0.0      276729      6.0      0.0      276729      6.0      6.0
118         0.0      276729      6.0      0.0      276729      6.0      6.0
119         0.0      276729      6.0      0.0      276729      6.0      6.0
120         0.0      276729      6.0      0.0      276729      6.0      6.0
121         0.0      276729      6.0      0.0      276729      6.0      6.0
122         0.0      276729      6.0      0.0      276729      6.0      6.0
123         0.0      276729      6.0      0.0      276729      6.0      6.0
124         0.0      276729      6.0      0.0      276729      6.0      6.0
125         0.0      276729      6.0      0.0      276729      6.0      6.0
126         0.0      276729      6.0      0.0      276729      6.0      6.0
127         0.0      276729      6.0      0.0      276729      6.0      6.0
128         0.0      276729      6.0      0.0      276729      6.0      6.0
129         0.0      276729      6.0      0.0      276729      6.0      6.0
130         0.0      276729      6.0      0.0      276729      6.0      6.0
131         0.0      276729      6.0      0.0      276729      6.0      6.0
132         0.0      276729      6.0      0.0      276729      6.0      6.0
133         0.0      276729      6.0      0.0      276729      6.0      6.0
134         0.0      276729      6.0      0.0      276729      6.0      6.0
135         0.0      276729      6.0      0.0      276729      6.0      6.0
136         0.0      276729      6.0      0.0      276729      6.0      6.0
137         0.0      276729      6.0      0.0      276729      6.0      6.0
138         0.0      276729      6.0      0.0      276729      6.0      6.0
139         0.0      276729      6.0      0.0      276729      6.0      6.0
140         0.0      276729      6.0      0.0      276729      6.0      6.0
141         0.0      276729      6.0      0.0      276729      6.0      6.0
142         0.0      276729      6.0      0.0      276729      6.0      6.0
143         0.0      276729      6.0      0.0      276729      6.0      6.0
144         0.0      276729      6.0      0.0      276729      6.0      6.0
145         0.0      276729      6.0      0.0      276729      6.0      6.0
146         0.0      276729      6.0      0.0      276729      6.0      6.0
147         0.0      276729      6.0      0.0      276729      6.0      6.0
148         0.0      276729      6.0      0.0      276729      6.0      6.0
149         0.0      276729      6.0      0.0      276729      6.0      6.0
150         0.0      276729      6.0      0.0      276729      6.0      6.0
151         0.0      276729      6.0      0.0      276729      6.0      6.0
152         0.0      276729      6.0      0.0      276729      6.0      6.0
153         0.0      276729      6.0      0.0      276729      6.0      6.0
154         0.0      276729      6.0      0.0      276729      6.0      6.0
155         0.0      276729      6.0      0.0      276729      6.0      6.0
156         0.0      276729      6.0      0.0      276729      6.0      6.0
157         0.0      276729      6.0      0.0      276729      6.0      6.0
158         0.0      276729      6.0      0.0      276729      6.0      6.0
159         0.0      276729      6.0      0.0      276729      6.0      6.0
160         0.0      276729      6.0      0.0      276729      6.0      6.0
161         0.0      276
```