Predict item6 for Alice

Mehrdad Mohammadian

Data

Title	item1	item2	item3	item4	item5	item6	corr with alice
alice	5	2	3	3	0		
user1	2	0	4	0	4	1	-1
user2	2	1	3	0	1	2	0.33
user3	4	2	3	2	0	1	0.9
user4	3	3	2	0	3	1	0.19
user5	0	3	0	2	2	2	-1
user6	5	3	0	1	3	2	0.65
user7	0	5	0	1	5	1	-1

Code:

```
import pandas as pd
data = {"item1": [5,2,2,4,3,0,5,0],
        "item2": [2,0,1,2,3,3,3,5],
        "item3": [3,4,3,3,2,0,0,0],
        "item4": [3,0,0,2,0,2,1,1],
        "item5": [0,4,1,0,3,2,3,5],
        "item6": [0,1,2,1,1,2,2,1],
        "corr with alice" : [None,-1,0.33,0.90,0.19,-1,0.65,-1]
        }
data = pd.DataFrame(data, index=["alice", "user1", "user2", "user3", "user4", "user5", "user6", "user7"])

K=3
r_bar_a = sum(data.loc['alice'][0:5]) / 5
max_corrs = data.sort_values('corr with alice', ascending=False)['corr with alice'][0:K]
max_corrs = list(zip(max_corrs.index, max_corrs.values))

numerator = 0
denominator = 0
for i in max_corrs:
    numerator += ( data['item6'][i[0]]  -  ( sum(data.loc[i[0]][0:6]) / 6  ) ) * i[1]
    denominator += i[1]

pearson = r_bar_a + (numerator/denominator)
print(f'predicted item6 for alice: {pearson}')

-------------- OUTPUT --------------
predicted item6 for alice: 2.093794326241135
```

Formula:

$$P_{a,i} = \bar{r_a} + \frac{\sum_{u=1}^{k}(r_{ai} - \bar{r_u}) \times sim(a,u)}{\sum_{u=1}^{k} sim(a,u)}$$

Calculation:

$$k = 3$$

$$P_{alice} = (13/5) + \frac{(1-2) \times 0.9 + (2-2.33) \times 0.65 + (2-1.5) \times 0.33}{0.9 + 0.65 + 0.33} = 2.0937$$

The End