

[**Alex Kaledin**](https://app.slack.com/team/U023L2M7UUE) [Jul 20th at 5:54 PM](https://skillfactorydst.slack.com/archives/C01843G3S22/p1626792867278500)

осталась одна задача 27.идея такова: сгруппировать по актерам, сложив каст и отсортировать полученные списки по наиболее повторяющимся элементам.def self\_filter(x, l):  
 while x in l:  
 l.remove(x)movie\_pair = data.copy()  
movie\_pair.cast = movie\_pair.cast.str.split('|') # разделяем cast на списки  
movie\_pair = movie\_pair.assign(actor = movie\_pair.cast) # дублируем колонку cast  
movie\_pair = movie\_pair.explode('actor') # разделяем списки актеров  
movie\_pair = movie\_pair.groupby(['actor'])['cast'].sum() #группируем по актерам складывая castfor x in movie\_pair.size:  
 movie\_pair = movie\_pair.apply(lambda x: self\_filter(str(movie\_pair.index[x]), movie\_pair[x])) #  
вот тут надо очистить списки от совпадения с индексом. всю голову сломал, ерунду всякую начал лепить. подскажите пожалуйста, что я делаю не так.pairs = pd.DataFrame(movie\_pair)  
pairs = pairs.assign(pair\_count = pairs.cast.apply(lambda x: dict(Counter(x).most\_common(1)).items())) #получаем колонку с отсортированным количеством случаев совместной работы. главное - зачистить списки от самоповторов.... (edited)

:hourglass_flowing_sand:1

4 replies



[**Владимир Горюнов**](https://app.slack.com/team/U01GC4LTR4Y)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1626797592279500?thread_ts=1626792867.278500&cid=C01843G3S22)

Добрый день!Для решения этого задания рекомендуется пройти по каждой записи в исходном датафрейме - посчитать все возможные парные комбинации из состава актеров и в случае повторения пар - прибавлять значение счетчика по ним.

from collections import Counter

from itertools import combinations

pairs = Counter()

for i in range(0,len(data)):

artists = data.cast[i].split('|')

for j in list(combinations(artists, 2)):

if j not in pairs:

pairs[j] = 1

else:

pairs[j] += 1

pairs.most\_common(5)



[**Alex Kaledin**](https://app.slack.com/team/U023L2M7UUE)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1626798875279800?thread_ts=1626792867.278500&cid=C01843G3S22)

да, этот вариант гораздо проще. однако я бы хотел добить реализацию своей идеи.



[**Sergey Meshcheryakov**](https://app.slack.com/team/U0192UB6HMF)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1626801856280000?thread_ts=1626792867.278500&cid=C01843G3S22)

Добрый вечер.

movie\_pair = data.copy()

movie\_pair.cast = movie\_pair.cast.str.split('|') # разделяем cast на списки

movie\_pair = movie\_pair.assign(actor = movie\_pair.cast) # дублируем колонку cast

movie\_pair = movie\_pair.explode('actor') # разделяем списки актеров

movie\_pair = movie\_pair.groupby(['actor'])['cast'].sum() #группируем по актерам складывая cast

*вот тут надо очистить списки от совпадения с индексом*

def self\_filter(row):

return [element for element in row[1] if element != row[0]]

movie\_pair = movie\_pair.reset\_index()

movie\_pair['cast'] = movie\_pair.apply(self\_filter,axis=1)

movie\_pair

:+1:1



[**Alex Kaledin**](https://app.slack.com/team/U023L2M7UUE)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1626804619280200?thread_ts=1626792867.278500&cid=C01843G3S22)

спасибо!

:handshake:1