**[Eduard Nasybulin](https://app.slack.com/team/U022K1NA5BQ" \t "_blank)** [Jul 21st at 6:30 PM](https://skillfactorydst.slack.com/archives/C01843G3S22/p1626881418289200)

Привет! Я в итоге понял как делать 27 задание но тем не менее хотел бы уточнить что не так с моими первоначальными попытками. Проблема лишь в том что идет перегруз данных или есть что-то ещё? Я подозреваю что проблема в том что  
 (data.cast.str.contains(i[0], na=False)) and (data.cast.str.contains(i[1], na=False) (Попытка 2)  
выдаёт  серию True/False вместо одного значения.  
Можно ли как то исправить?Попытка 1 (без использования combinations):actors\_list = []  
for s in data.cast:  
    actors\_list.extend(s.split("|"))  
actors\_set = set(actors\_list)collect = {}  
for i in actors\_set:  
    for j in actors\_set:  
        if i != j:  
            collect[i+j] = data[data.cast.str.contains(i) and data.cast.str.contains(j)].imdb\_id.count()  
            print(collect)  
max(collect, key=collect.get)Попытка 2 (с использованием combinations):actors\_list = []  
for s in data.cast:  
    actors\_list.extend(s.split("|"))  
actors\_set = set(actors\_list)actors\_comb\_list = combinations(actors\_set, 2)co\_actor\_list = []  
for i in actors\_comb\_list:  
    if (data.cast.str.contains(i[0], na=False)) and (data.cast.str.contains(i[1], na=False)):  
        co\_actors\_list.append(i)  
pd.DataFrame(co\_actors\_list).value\_counts().head(1)Спасибо! (edited)

2 replies



[**Владимир Горюнов**](https://app.slack.com/team/U01GC4LTR4Y)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1626882775289400?thread_ts=1626881418.289200&cid=C01843G3S22)

Добрый вечер!У вас ошибка в условии выборки - два условия складываются через побитовые операции

data[data.cast.str.contains(i) & data.cast.str.contains(j)]

И способ очень ресурсоемкий и неоптимальный - в двух вложенных циклах каждый раз осуществляется проход по всему датафрейму, расчет очень много времени занимаетИ во втором случае вы рассчитываете ВСЕ возможные попарные комбинации по всем актерам, а это как вы понимаете очень много. При том что большинство попарных комбинаций - никогда не встречаются в этом датасете.Так что лучше посчитать только пары, которые реально играли вместе.



[**Eduard Nasybulin**](https://app.slack.com/team/U022K1NA5BQ)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1626883686289800?thread_ts=1626881418.289200&cid=C01843G3S22)

Понятно, спасибо. Но в принципе нижеследующий код даст результат если долго ждать?actors\_list = []  
for s in data.cast:  
    actors\_list.extend(s.split("|"))  
actors\_set = set(actors\_list)collect = {}  
for i in actors\_set:  
    for j in actors\_set:  
        if i != j:  
            collect[i+j] = data[data.cast.str.contains(i) & data.cast.str.contains(j)].imdb\_id.count()  
max(collect, key=collect.get)