все ли здесь ок? смущает, что 2 компании в итоге

2 files

:white_check_mark:1

10 replies



[**Sergey Meshcheryakov**](https://app.slack.com/team/U0192UB6HMF)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625775857209300?thread_ts=1625775256.209200&cid=C01843G3S22)

Их две, потому что в исходном датасете они в одной ячейке. Соответственно, после разделения, каждой из них соответствует один и тот же фильм с самым длинным названием.

:+1:1



[**Крупень Сергей**](https://app.slack.com/team/U022A2ECASH)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625810395210900?thread_ts=1625775256.209200&cid=C01843G3S22)

У 20 века фокс должны быть и другие фильмы. Средняя длинна названия соответственно у него должна быть меньше и в топ не попасть.

:+1:1



[**Sergey Meshcheryakov**](https://app.slack.com/team/U0192UB6HMF)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625813697211100?thread_ts=1625775256.209200&cid=C01843G3S22)

Добрый день. Вопрос же не про среднюю длину названий фильмов, выпускаемых студиями )  
А в результате выше две компании в топе, потому что они обе сняли фильм, который имеет самое длинное название.

data = pd.read\_csv('movie\_bd\_v5.csv')

data['original\_title\_len'] = data['original\_title'].apply(len)

data[data['original\_title\_len']==data['original\_title\_len'].max()]\

[['production\_companies','original\_title']]

(edited)



[**Крупень Сергей**](https://app.slack.com/team/U022A2ECASH)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625816650217300?thread_ts=1625775256.209200&cid=C01843G3S22)

Ну…. Не согласен. Тогда формулировка расплывчатая. Ясно же, что идёт речь про единственное число «какая студия» и множественное число «названия своим фильмам». Какой ещё аггригирующий признак такой группировке можно применить? Не сказано же, «Макс»? Не сказано. Да и не могло, формулировка не позволяет  
Если хотели  получить кинокомпанию, которая сняла (единственное число - поэтому, максимум) фильм с самым длинным названием, то тут ещё как то могло получиться, что таких компаний несколько.



[**Марина Архипцева (эксперт)**](https://app.slack.com/team/U01NGMHLZFY)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625817526217700?thread_ts=1625775256.209200&cid=C01843G3S22)

Добрый день! Соглашусь с [@Крупень Сергей](https://skillfactorydst.slack.com/team/U022A2ECASH) , вопрос звучит так "24. Какая студия дает самые длинные названия своим фильмам по количеству символов?" - не самое длинное название, а самые длинные названия, то есть надо проанализировать все фильмы студии и посчитать среднее.



[**Sergey Meshcheryakov**](https://app.slack.com/team/U0192UB6HMF)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625817847217900?thread_ts=1625775256.209200&cid=C01843G3S22)

Про среднее все таки это следующий вопрос )) Ну а в данном вопросе, стоило бы уточнить формулировку, потому что и max() и mean() дает одну и ту же студию, ввиду того, что Four By Two Productions сняла только один фильм



[**Марина Архипцева (эксперт)**](https://app.slack.com/team/U01NGMHLZFY)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625820157221400?thread_ts=1625775256.209200&cid=C01843G3S22)

Нет ) следующий вопрос про количество слов, а тут количество символом, но и там, и тут среднее.



[**Sergey Meshcheryakov**](https://app.slack.com/team/U0192UB6HMF)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625821043221700?thread_ts=1625775256.209200&cid=C01843G3S22)

Думаю, не принципиально в чем мерить длину в символах или словах. Речь об инструментах измерения(max или mean) :) не соглашусь, что в 24 задаче речь идет о средней величине.  
Вот ответ на вопрос:  
*Какая студия дает самые длинные названия (как в задаче 24)*

df = pd.DataFrame({'company':['company1', 'company1', 'company2', 'company2'],'len\_of\_title':[10,4,9,8]})

df.groupby('company')['len\_of\_title'].max()

company

company1 10

company2 9

А вот ответ на вопрос:  
*Какая студия дает в среднем самые длинные названия (аналогично задаче 25)*

df.groupby('company')['len\_of\_title'].mean()

company

company1 7.0

company2 8.5

(edited)



[**Марина Архипцева (эксперт)**](https://app.slack.com/team/U01NGMHLZFY)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625825319222200?thread_ts=1625775256.209200&cid=C01843G3S22)

И всё таки нет. И там, и там подразумевается среднее, поэтому и в вопросе множественное число. Принципиально - подсчет слов отличается от подсчета символов )



[**Sergey Meshcheryakov**](https://app.slack.com/team/U0192UB6HMF)  [2 months ago](https://skillfactorydst.slack.com/archives/C01843G3S22/p1625825719222400?thread_ts=1625775256.209200&cid=C01843G3S22)

*Принципиально - подсчет слов отличается от подсчета символов*  
Я имел ввиду, что в споре максимум мы ищем или среднее, уже не важно символы или слова )И все же. Если в вопросе множественное число, почему тогда в вариантах ответа, правильный -Four By Two Productions - студия, которая сняла только один фильм?