

Написание сложных формул на языке DAX Power Pivot с использованием ChatGPT

Я увлекаюсь игрой [Fantasy Premier League](#). Помимо спортивного азарта соперничества с 10M игроков по всему миру, игра создаёт почву для серьёзных статистических исследований. Особенно после того, как я открыл для себя мир [API](#). Большие данные, получаемые с помощью API, я помещаю в Excel, в модель данных Power Pivot. Дальнейший анализ веду с использованием формул на языке DAX (Data Analysis Expressions). Я только осваиваю этот язык, а практические задачи требуют продвинутых навыков. На помощь приходит ChatGPT.

Я взял итоги выступления игроков в сезоне 2022/23, поместил их в модель данных и вывел на лист Excel в виде сводной таблицы. Ниже показан фрагмент таких данных:

	A	B	C	D	E	F	G	H
1	Element	Name	Team	Pos	event	Cost	Pts	
2	283	Salah	LIV	MID	1	13	12	
3					2	13	2	
4					3	13	8	
5					4	13	3	
6					5	13	10	
7					6	13	3	
8					9	12,8	5	
9					10	12,7	2	
10					11	12,7	8	
11					12	12,7	3	
12					13	12,8	2	
13					14	12,8	7	
14					15	12,8	15	
15					16	12,8	2	
16					17	12,8	12	
17					18	12,8	2	
18					19	12,9	2	
19					20	12,9	2	
20	318	Haaland	MCI	FWD	1	11,5	13	
21					2	11,6	5	
22					3	11,7	6	
23					4	11,7	17	
24					5	11,8	17	
25					6	11,9	9	
26					8	12	6	
27					9	12,1	23	
28					10	12,2	6	
29					11	12,2	2	
30					13	12,2	13	
31					14	12,2	0	
32					15	12,1	5	
33					16	12,2	1	
34					17	12,2	13	
35					18	12,2	6	
36					19	12,2	2	
37					20	12,2	8	

Рис. 1. Фрагмент сводной таблицы

В столбцах указаны: ID игрока, его имя, команда, за которую он выступает, позиция, номер тура, стоимость и набранные очки. В исходных данных отсутствует важный параметр для анализа – форма. Строго говоря, форма – среднее количество очков, набранных игроком за последние 30 дней. Для целей анализа я упрощаю определение, и хочу рассчитать форму, как среднее число очков, набранных в предыдущих четырех турах. Таким образом в первых четырех турах форму определить нельзя, а в пятом туре форма равна среднему числу очков, набранных в турах 1–4. И т.д.

Рассчитать форму я хочу не на листе Excel, используя приведенную сводную таблицу, а в модели данных. Мало того, что я не знаю, какой использовать код, так еще и строки в модели данных не отсортированы.

	was_home	Name	Team	Pos	Element	event	fixture
1	ИСТИНА	Haaland	MCI	FWD	318	29	287
2	FALSE	Haaland	MCI	FWD	318	14	137
3	FALSE	Haaland	MCI	FWD	318	38	373
4	FALSE	Salah	LIV	MID	283	29	74
5	FALSE	Haaland	MCI	FWD	318	22	215
6	FALSE	Salah	LIV	MID	283	29	287
7	ИСТИНА	Haaland	MCI	FWD	318	4	37
8	ИСТИНА	Haaland	MCI	FWD	318	20	68
9	FALSE	Salah	LIV	MID	283	10	91
10	FALSE	Salah	LIV	MID	283	31	304
11	ИСТИНА	Haaland	MCI	FWD	318	13	125
12	FALSE	Salah	LIV	MID	283	36	358
13	FALSE	Haaland	MCI	FWD	318	1	10
14	FALSE	Haaland	MCI	FWD	318	36	356
15	ИСТИНА	Haaland	MCI	FWD	318	23	228
16	FALSE	Haaland	MCI	FWD	318	25	241
17	ИСТИНА	Haaland	MCI	FWD	318	15	145
18	FALSE	Haaland	MCI	FWD	318	30	298

Рис. 2. Фрагмент модели данных в Power Pivot

Ну что ж... обратимся к ChatGPT.

Вопрос

Напиши формулу на языке DAX. В столбце A номера категорий. В столбце B номера периодов. В столбце C объем продаж. Нужно добавить столбец D, в котором подсчитать скользящее среднее по столбцу C за 4 предшествующих периода. Если строк за предыдущие периоды меньше четырех, то в столбце D указываем 0. Строки в таблице не отсортированы.

Ответ

Для расчета скользящего среднего в столбце D по столбцу C с учетом предыдущих четырех периодов в Power BI или Excel с использованием языка DAX, вы можете использовать следующую формулу:

Moving Average =

```

VAR CurrentCategory = [Category] // Здесь [Category] – это ссылка на столбец с категориями
VAR CurrentPeriod = [Period] // Здесь [Period] – это ссылка на столбец с номерами периодов
VAR PreviousPeriods = FILTER(ALL('YourTable'), [Category] = CurrentCategory && [Period] < CurrentPeriod)
VAR CountPreviousPeriods = COUNTROWS(PreviousPeriods)
RETURN
    IF(
        CountPreviousPeriods < 4,
        0, // Если предыдущих периодов меньше 4, то возвращаем 0
        AVERAGEX(TOPN(4, PreviousPeriods, [Period], ASC), [Sales]) // Иначе считаем среднее за последние 4 периода
    )

```

В этой формуле мы сначала определяем текущую категорию и текущий период, а затем создаем временную таблицу PreviousPeriods, которая фильтруется так, чтобы включать только строки с предыдущими периодами для текущей категории. Затем мы подсчитываем количество предыдущих периодов.

Используя функцию IF, мы проверяем, если количество предыдущих периодов меньше 4, то возвращаем 0. В противном случае мы используем функцию AVERAGEX для расчета среднего

значения продаж ([Sales]) за последние 4 периода, отсортированных по номерам периодов в возрастающем порядке.

Замените 'YourTable', [Category], [Period] и [Sales] на соответствующие имена вашей таблицы и столбцов данных. Эта формула будет рассчитывать скользящее среднее в столбце D для каждой строки в вашей таблице, учитывая условие на количество предыдущих периодов.

Доработка кода

После подстановки названий таблицы и столбцов код не заработал, но моих знаний хватило, чтобы его поправить.

=

```
VAR CurrentCategory = [Element]
```

```
VAR CurrentPeriod = [event]
```

```
VAR PreviousPeriods = FILTER(ALL(ElementSummary);
```

```
[Element] = CurrentCategory && [event] < CurrentPeriod)
```

```
VAR CountPreviousPeriods = COUNTROWS(PreviousPeriods)
```

```
RETURN
```

```
IF(
```

```
CountPreviousPeriods < 4;
```

```
0;
```

```
AVERAGEX(TOPN(4; PreviousPeriods; [event];DESC); [total_points])
```

```
)
```

Я убрал название Moving Average, и начал формулу со знака равно. Заменял разделитель запятую на точку с запятой. Изменил тип сортировки периодов, заменив аргумент ASC на DESC. Вуаля!

	A	B	C	D	E	F	G	H	I
1	Element	Name	Team	Pos	event	Cost	Pts	Form	
2	283	Salah	LIV	MID	1	13	12	0	
3					2	13	2	0	
4					3	13	8	0	
5					4	13	3	0	
6					5	13	10	6,25	
7					6	13	3	5,75	
8					9	12,8	5	6	
9					10	12,7	2	5,25	
10					11	12,7	8	5	
11					12	12,7	3	4,5	
12					13	12,8	2	4,5	
13					14	12,8	7	3,75	
14					15	12,8	15	5	
15					16	12,8	2	6,75	
16					17	12,8	12	6,5	
17					18	12,8	2	9	
18					19	12,9	2	7,75	
19					20	12,9	2	4,5	
20	318	Haaland	MCI	FWD	1	11,5	13	0	
21					2	11,6	5	0	
22					3	11,7	6	0	
23					4	11,7	17	0	
24					5	11,8	17	10,25	
25					6	11,9	9	11,25	
26					8	12	6	12,25	
27					9	12,1	23	12,25	
28					10	12,2	6	13,75	
29					11	12,2	2	11	
30					13	12,2	13	9,25	
31					14	12,2	0	11	
32					15	12,1	5	5,25	
33					16	12,2	1	5	
34					17	12,2	13	4,75	
35					18	12,2	6	4,75	
36					19	12,2	2	6,25	
37					20	12,2	8	11	

Рис. 3. Форма игрока выведена в сводную таблицу