

Field	Description	Type
id	Refers to the unique id of the type	integer
game	The name of the game	string
event_type	The type of event	string
created_at	The time when the event happened	timestamp
team_name	The name of the team related to the event	string
player_name	The name of the player that performed the event or to whom it happened	string
related_player_name	The name of the player who is related to the event that <code>player_name</code> performed	string
result	The result after the event happened	string
minute	The game minute in which the event happened	integer
extra_minute	The additional minutes in which the event happened (e.g. for 90+2 this would be 2)	integer
previous_player_event	The <code>event_type</code> of the previous event that happened to the player	string
previous_player_event_at	The timestamp of the previous event that happened to the player	timestamp
seconds_after_previous_event	The seconds after the previous event happened to the player	integer
player_id	The id of the player that performed the event or to whom it happened	integer
related_player_id	The id of the player who is related to the event that <code>player_name</code> performed	integer
participant_id	The id of the team related to the event	integer
fixture_id	The id of the game the event happened in	integer

2 hidden cells

DataFrames and CSVs DataFrame as df

```
SELECT *
FROM game_events.csv;
```

...	↑↓	i..	...	↑↓	game	...	↑↓	event_type	...	↑↓	created_at	...	↑↓	team_n...	...	↑↓	player_name	...	↑↓	related_player_name	...	↑↓	...	↑↓
0	116213644	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:27:00.000			Romania			Bogdan Racovițan			Nicolae Stanciu			null				
1	116213524	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:02:00.000			Ukraine			Volodymyr Brazhko			Taras Stepanenko			null				
2	116213531	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:02:00.000			Romania			Valentin Mihăilă			Florinel Coman			null				
3	116213571	Romania vs Ukraine (2024-06-17)			Yellowcard			2024-06-17T16:07:00.000			Ukraine			Yukhym Konoplyá			null			null				
4	116213583	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:12:00.000			Ukraine			Oleksandr Tymchyk			Yukhym Konoplyá			null				
5	116213533	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:02:00.000			Ukraine			Roman Yaremchuk			Mykola Shaparenko			null				
6	116213534	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:02:00.000			Ukraine			Andriy Yarmolenko			Viktor Tsygankov			null				
7	116213507	Romania vs Ukraine (2024-06-17)			Goal			2024-06-17T15:57:00.000			Romania			Denis Drăguș			Dennis Man			3-0				
8	116213593	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:15:00.000			Romania			George Pușcaș			Denis Drăguș			null				
9	116213525	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:02:00.000			Romania			Ianis Hagi			Dennis Man			null				
10	116213596	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:15:00.000			Romania			Adrian Rus			Marius Marin			null				
11	116213620	Romania vs Ukraine (2024-06-17)			Yellowcard			2024-06-17T16:19:00.000			Romania			Răzvan Marin			null			null				
12	116213493	Romania vs Ukraine (2024-06-17)			Goal			2024-06-17T15:53:00.000			Romania			Răzvan Marin			null			2-0				
13	116213437	Romania vs Ukraine (2024-06-17)			Goal			2024-06-17T15:29:00.000			Romania			Nicolae Stanciu			Dennis Man			1-0				
14	116213630	Romania vs Ukraine (2024-06-17)			Substitution			2024-06-17T16:23:00.000			Ukraine			Ruslan Malinovskyi			Georgiy Sudakov			null				
15	116479144	Germany vs Denmark (2024-06-29)			Yellowcard			2024-06-29T21:59:00.000			Germany			Julian Nagelsmann			null			null				

Rows: 838

↗ Expand

## Exploratory data analysis

DataFrames and CSVs DataFrame as df1

```
SELECT
  COUNT(1) AS rows,
  COUNT(DISTINCT(id)) AS unique_id
FROM game_events.csv ;
```

index	...	↑↓	rows	...	↑↓	unique_id	...	↑↓
0			838			838		

Rows: 1

↗ Expand

## Summary stats

Let's look at the separate tables and their columns using the SUMMARIZE function.

## SUMMARIZE SELECT

```

*
FROM game_events.csv;

```

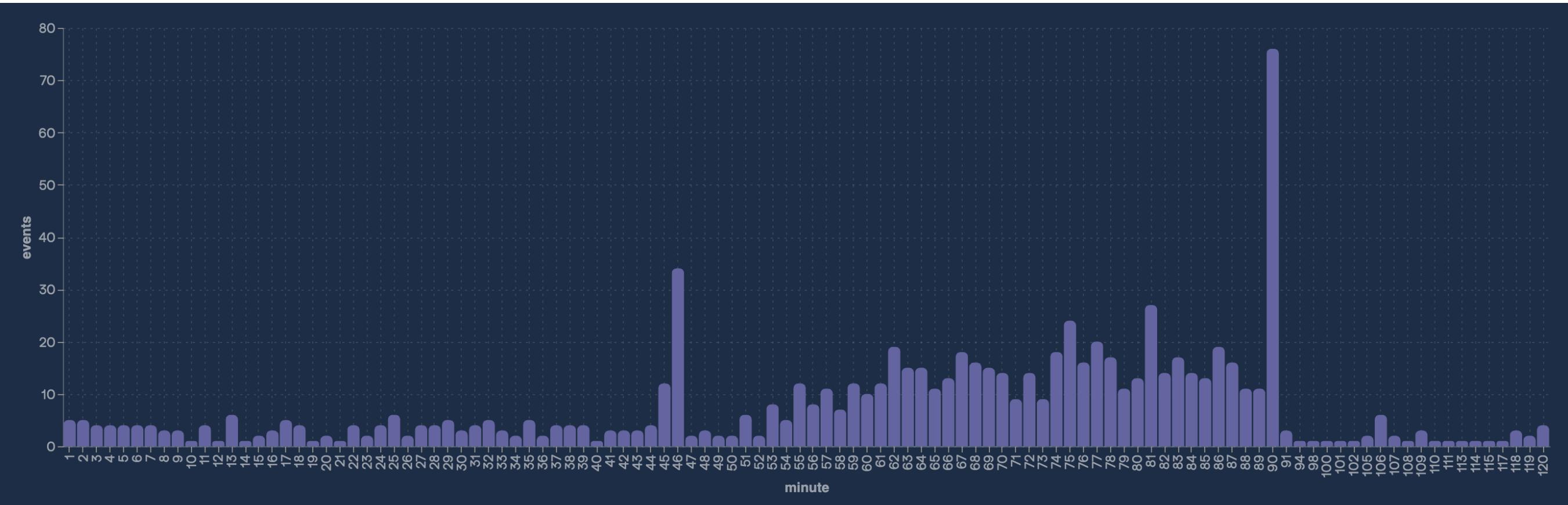
...	↑↓	column_name	...	↑↓	col...	...	↑↓	min	...	↑↓	max	...	↑↓	approx...	...	↑↓	avg	...	↑↓	std	...	↑↓	q	...
0		id			BIGINT			116112635			116635472			831			116364186.64319809			148824.0936745581			1162444	
1		game			VARCHAR			Albania vs Spain (2024-06-24)			Ukraine vs Belgium (2024-06-26)			48			null			null			null	
2		event_type			VARCHAR			Goal			Yellowcard			9			null			null			null	
3		created_at			TIMESTAMP			2024-06-14 21:10:00			2024-07-06 22:36:00			624			null			null			null	
4		team_name			VARCHAR			Albania			Ukraine			24			null			null			null	
5		player_name			VARCHAR			Abdülkerim Bardakçı			Ž. Karničnik			413			null			null			null	
6		related_player_name			VARCHAR			A. Bah			İsmail Yüksek			296			null			null			null	
7		result			VARCHAR			0-0			5-3			24			null			null			null	
8		minute			BIGINT			1			120			108			64.85202863961814			24.88769642123774			52	
9		extra_minute			BIGINT			0			11			12			0.32577565632458233			1.250119824996759			0	
10		previous_player_event			VARCHAR			Goal			Yellowcard			8			null			null			null	
11		previous_player_event_at			TIMESTAMP			2024-06-14 21:25:00			2024-07-06 22:22:00			112			null			null			null	
12		seconds_after_previous_event			DOUBLE			0.0			6600.0			57			1595.593220338983			1780.5658096085092			360.0	
13		player_id			BIGINT			268			37656179			378			5716804.0690072635			11773280.234363215			34634	
14		related_player_id			BIGINT			268			37656179			258			6411015.117647059			12482800.246435147			34926	
15		participant_id			BIGINT			18583			18873			24			18679.655131264917			53.09167172779943			18645	

Rows: 17

↗ Expand

## Distributions

```
-- Distribution of "minute" values
SELECT minute,
       COUNT(*) AS events
FROM game_events.csv
GROUP BY 1
ORDER BY 1 ;
```



2 hidden cells

```
-- Events in minute 90
```

```
SELECT
    event_type,
    COUNT(1) AS events
FROM game_events.csv
WHERE minute = 90
GROUP BY 1
ORDER BY 2 DESC;
```

...	↑↓ even...	...	↑↓	...	↑↓
0	Yellowcard			32	
1	Substitution			29	
2	Goal			12	
3	Redcard			2	
4	VAR			1	

Rows: 5

[↗ Expand](#)

Distribution of extra\_minute

```
-- Distribution of "extra_minute"
```

```
SELECT
    extra_minute,
    COUNT(1) AS events
FROM game_events.csv
GROUP BY 1;
```

...	↑↓	extra...	...	↑↓	...	↑↓
0		0		763		
1		11		1		
2		5		7		
3		9		1		
4		2		16		
5		10		1		
6		8		4		
7		7		2		
8		1		13		
9		4		12		
10		6		5		
11		3		13		

Rows: 12

[↗ Expand](#)

— 7 hidden cells —

```
WITH buckets AS (
    SELECT DISTINCT
        minute,
        minute / 5,
        ceil(minute / 5),
        ceil(minute / 5) * 5 as ceil_bucket
    FROM 'game_events.csv'
    ORDER BY minute
)
```

```
SELECT * FROM buckets
```

...	↑↓	...	↑↓	("minute")	...	↑↓	ceil(("minute" / 5))	...	↑↓	cei...	...	↑↓
0		1		0.2			1			5		
1		2		0.4			1			5		
2		3		0.6			1			5		
3		4		0.8			1			5		
4		5		1			1			5		
5		6		1.2			2			10		
6		7		1.4			2			10		
7		8		1.6			2			10		
8		9		1.8			2			10		
9		10		2			2			10		
10		11		2.2			3			15		
11		12		2.4			3			15		
12		13		2.6			3			15		
13		14		2.8			3			15		
14		15		3			3			15		
15		16		3.2			4			20		

Rows: 110

↗ Expand

1 hidden cell

```
COPY (
    WITH goal_events_with_minute_string AS (
        SELECT
            *,
            -- Use lpad for better sorting behavior (e.g. 1 -> ' 1')
            lpad(minute, 3, ' ') AS minute_str_sortable,
        FROM 'game_events.csv'
        WHERE event_type = 'Goal'
    )

    SELECT
        *,
        CASE
            WHEN extra_minute = 0 THEN minute_str_sortable
            ELSE concat(minute_str_sortable, ' + ', extra_minute)
        END AS game_minute,
        CASE
            -- We simplify overtime and move it into a "OT" bucket
            WHEN minute > 90 THEN 'OT'
            -- For extra minutes, we use 45+, 90+, etc.
            WHEN extra_minute != 0 THEN concat(minute_str_sortable, ' + ')
            -- Use lpad for better sorting behavior (e.g. 1 -> ' 1')
            ELSE lpad(
                (ceil(minute / 5) * 5)::int,
                3,
                ' '
            )
        END AS game_period_bucket
    FROM goal_events_with_minute_string
) TO 'goals.csv' (HEADER, DELIMITER ',');
```

```
SELECT * FROM 'goals.csv'
```

...	↑↓	↓...	↑↓	game	...	↑↓	e...	...	↑↓	created_at	...	↑↓	team_n...	...	↑↓	player_name	...	↑↓	related_player_name	...	↑↓	...	↑↓	...	↑↓	extra...	...	↑↓
0	116213507	Romania vs Ukraine (2024-06-17)			Goal					2024-06-17T15:57:00.000			Romania			Denis Drăguș			Dennis Man			3-0		57				
1	116213493	Romania vs Ukraine (2024-06-17)			Goal					2024-06-17T15:53:00.000			Romania			Răzvan Marin			null			2-0		53				
2	116213437	Romania vs Ukraine (2024-06-17)			Goal					2024-06-17T15:29:00.000			Romania			Nicolae Stanciu			Dennis Man			1-0		29				
3	116479263	Germany vs Denmark (2024-06-29)			Goal					2024-06-29T22:08:00.000			Germany			Jamal Musiala			Nico Schlotterbeck			2-0		68				
4	116478811	Germany vs Denmark (2024-06-29)			Goal					2024-06-29T21:53:00.000			Germany			Kai Havertz			null			1-0		53				
5	116521748	Spain vs Georgia (2024-06-30)			Goal					2024-06-30T22:15:00.000			Spain			Nico Williams			Fabián Ruiz			3-1		75				
6	116520536	Spain vs Georgia (2024-06-30)			Goal					2024-06-30T21:39:00.000			Spain			Rodri			Nico Williams			1-1		39				
7	116521204	Spain vs Georgia (2024-06-30)			Goal					2024-06-30T21:51:00.000			Spain			Fabián Ruiz			Lamine Yamal			2-1		51				
8	116520002	Spain vs Georgia (2024-06-30)			Goal					2024-06-30T21:18:00.000			Georgia			Robin Le Normand			Otar Kakabadze			0-1		18				
9	116521999	Spain vs Georgia (2024-06-30)			Goal					2024-06-30T22:23:00.000			Spain			Dani Olmo			Mikel Oyarzabal			4-1		83				
10	116145822	Italy vs Albania (2024-06-15)			Goal					2024-06-15T21:01:00.000			Albania			Nedim Bajrami			null			0-1		1				
11	116145951	Italy vs Albania (2024-06-15)			Goal					2024-06-15T21:11:00.000			Italy			Alessandro Bastoni			Lorenzo Pellegrini			1-1		11				
12	116146022	Italy vs Albania (2024-06-15)			Goal					2024-06-15T21:16:00.000			Italy			Nicolò Barella			null			2-1		16				

13	116635174	Netherlands vs Turkey (2024-07-06)	Goal	2024-07-06T22:16:00.000	Netherlands	Mert Müldür	Denzel Dumfries	2-1	76	0	0
14	116634386	Netherlands vs Turkey (2024-07-06)	Goal	2024-07-06T21:35:00.000	Turkey	Samet Akaydin	Arda Güler	0-1	35	0	0
15	116635127	Netherlands vs Turkey (2024-07-06)	Goal	2024-07-06T22:10:00.000	Netherlands	Stefan de Vrij	Memphis Depay	1-1	70	0	0

Rows: 108

Expand

1 hidden cell

DataFrames and CSVs DataFrame as

```
WITH counts as (
    select
        game_period_bucket,
        count(1) as cnt
    from 'goals.csv'
    group by 1
    order by 1
)

select
    *
from counts
```

...	↑↓	game_period_...	...	↑↓	...	↑↓
0	5				4	
1	10				4	
2	15				6	
3	20				8	
4	25				5	
5	30				7	
6	35				4	
7	40				4	
8	45				1	
9	45 +				3	
10	50				2	
11	55				8	
12	60				7	
13	65				2	
14	70				10	
15	75				4	

Rows: 21

Expand

SELECT

```
game,  
player_name,  
game_minute  
FROM goals.csv  
ORDER BY game_minute  
LIMIT 10;
```

...	↑↓	game	...	↑↓	player_name	...	↑↓	ga...	...	↑↓
0	Austria vs Turkey (2024-07-02)	Merih Demiral			1					
1	Italy vs Albania (2024-06-15)	Nedim Bajrami			1					
2	Georgia vs Portugal (2024-06-26)	Khvicha Kvaratskhelia			2					
3	Belgium vs Romania (2024-06-22)	Youri Tielemans			2					
4	Netherlands vs Austria (2024-06-25)	Donyell Malen			6					
5	Belgium vs Slovakia (2024-06-17)	Ivan Schranz			7					
6	Poland vs Austria (2024-06-21)	Gernot Trauner			9					
7	Germany vs Scotland (2024-06-14)	Florian Wirtz			10					
8	Italy vs Albania (2024-06-15)	Alessandro Bastoni			11					
9	Croatia vs Albania (2024-06-19)	Qazim Laci			11					

Rows: 10

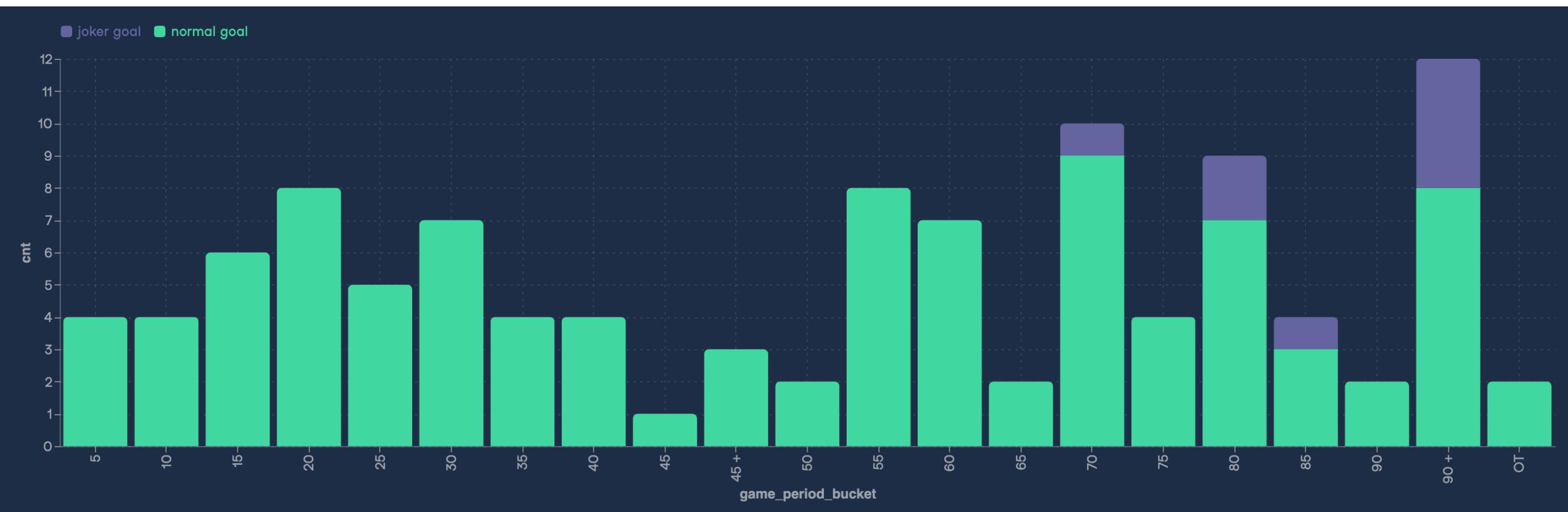
↗ Expand

1 hidden cell

```

WITH counts as (
  select
    game_period_bucket,
    case
      -- Joker goal if player was substituted within 15 minutes prior to scoring
      when
        previous_player_event = 'Substitution'
        and seconds_after_previous_event <= 900
      then 'joker goal'
      else 'normal goal'
    end as goal_type,
    count(1) as cnt
  from 'goals.csv'
  group by 1, 2
  order by 1
)

select
  *
from counts
  
```



1 hidden cell

SELECT

```
game,  
COUNT(1) AS goals  
FROM goals.csv  
GROUP BY 1  
ORDER BY 2 DESC  
LIMIT 10;
```

...	↑↓ game	...	↑↓	...	↑↓
0	Germany vs Scotland (2024-06-14)			6	
1	Netherlands vs Austria (2024-06-25)			5	
2	Spain vs Georgia (2024-06-30)			5	
3	Poland vs Austria (2024-06-21)			4	
4	Hungary vs Switzerland (2024-06-15)			4	
5	Turkey vs Georgia (2024-06-18)			4	
6	Croatia vs Albania (2024-06-19)			4	
7	Romania vs Netherlands (2024-07-02)			3	
8	Spain vs Croatia (2024-06-15)			3	
9	Portugal vs Czech Republic (2024-06-18)			3	

Rows: 10

↗ Expand

```
WITH goals_per_game as (
    select
        fixture_id,
        count(1) as goals
    from 'goals.csv'
    group by 1
    order by 1
)

select
    goals,
    count(1) as cnt
from goals_per_game
group by 1
order by 2
```

...	↑↓	...	↑↓	...	↑↓
0		6		1	
1		5		2	
2		4		4	
3		1		7	
4		3		13	
5		2		15	

Rows: 6

[↗ Expand](#)