

Introdução à Manipulação e Visualização de Dados


Rodrigo Lima Peregrino

29 de junho de 2022

Overview



1. Introdução
2. Principais operações
3. Power Query
4. Python
5. Breves comentários sobre SQL

The background of the slide is composed of two large, overlapping geometric shapes. On the left, there is a teal-colored shape that tapers towards the bottom left. To its right is a light gray shape that also tapers towards the bottom left, creating a white triangular area in the center where the text is located.

Introdução

Links legais/importantes



- ▶ Kaggle
- ▶ Datacamp
- ▶ Highcharts
- ▶ Matplotlib
- ▶ Plotly
- ▶ Shiny
- ▶ Streamlit
- ▶ Power BI

The background of the slide is composed of two large, overlapping geometric shapes. A teal-colored shape occupies the top-left corner, while a light gray shape occupies the bottom-left corner. The rest of the slide is white. The text is centered in the white area.

Principais operações

Principais operações



- ▶ Projetar
- ▶ Filtrar
- ▶ Agrupar
- ▶ Pivotear
- ▶ Concatenar (union all)
- ▶ Mesclar (merge, join)

Operações básicas



Projetar

- ▶ "Pegar" atributos em seu conjunto de dados
- ▶ Escolher as colunas, tabelas, etc. de interesse

Filtrar

- ▶ "Pegar" a parcela dos dados que satisfaça uma determinada condição
- ▶ Escolher as linhas/população de interesse

Agrupamento



Juntar registros de acordo com uma chave, podendo gerar sumarizações.

Agrupamento



Juntar registros de acordo com uma chave, podendo gerar sumarizações.

short_name	overall	wage_eur	age	club_name
L. Messi	93	320,000	34	Paris Saint-Germain
R. Lewandowski	92	270,000	32	FC Bayern München
Cristiano Ronaldo	91	270,000	36	Manchester United
Neymar Jr	91	270,000	29	Paris Saint-Germain
K. De Bruyne	91	350,000	30	Manchester City
J. Oblak	91	130,000	28	Atlético de Madrid
K. Mbappé	91	230,000	22	Paris Saint-Germain
M. Neuer	90	86,000	35	FC Bayern München
M. ter Stegen	90	250,000	29	FC Barcelona
H. Kane	90	240,000	27	Tottenham Hotspur
N. Kanté	90	230,000	30	Chelsea
K. Benzema	89	350,000	33	Real Madrid CF
T. Courtois	89	250,000	29	Real Madrid CF
H. Son	89	220,000	28	Tottenham Hotspur
Casemiro	89	310,000	29	Real Madrid CF
V. van Dijk	89	230,000	29	Liverpool
S. Mané	89	270,000	29	Liverpool
M. Salah	89	270,000	29	Liverpool
Ederson	89	200,000	27	Manchester City
J. Kimmich	89	160,000	26	FC Bayern München

Agrupamento



Juntar registros de acordo com uma chave, podendo gerar sumarizações.

short_name	overall	wage_eur	age	club_name
L. Messi	93	320,000	34	Paris Saint-Germain
R. Lewandowski	92	270,000	32	FC Bayern München
Cristiano Ronaldo	91	270,000	36	Manchester United
Neymar Jr	91	270,000	29	Paris Saint-Germain
K. De Bruyne	91	350,000	30	Manchester City
J. Oblak	91	130,000	28	Atlético de Madrid
K. Mbappé	91	230,000	22	Paris Saint-Germain
M. Neuer	90	86,000	35	FC Bayern München
M. ter Stegen	90	250,000	29	FC Barcelona
H. Kane	90	240,000	27	Tottenham Hotspur
N. Kanté	90	230,000	30	Chelsea
K. Benzema	89	350,000	33	Real Madrid CF
T. Courtois	89	250,000	29	Real Madrid CF
H. Son	89	220,000	28	Tottenham Hotspur
Casemiro	89	310,000	29	Real Madrid CF
V. van Dijk	89	230,000	29	Liverpool
S. Mané	89	270,000	29	Liverpool
M. Salah	89	270,000	29	Liverpool
Ederson	89	200,000	27	Manchester City
J. Kimmich	89	160,000	26	FC Bayern München

club_name	#players	total_wage
Paris Saint-Germain	3	820,000
FC Bayern München	3	516,000
Manchester United	1	270,000
Manchester City	2	550,000
Atlético de Madrid	1	130,000
FC Barcelona	1	250,000
Tottenham Hotspur	2	460,000
Chelsea	1	230,000
Real Madrid CF	3	910,000
Liverpool	3	770,000

Pivoteamento



Semelhante ao agrupamento porém as colunas do resultado são baseadas nos valores de uma outra coluna.

Pivoteamento

Semelhante ao agrupamento porém as colunas do resultado são baseadas nos valores de uma outra coluna.

short_name	overall	wage_eur	age	club_name
L. Messi	93	320,000	34	Paris Saint-Germain
R. Lewandowski	92	270,000	32	FC Bayern München
Cristiano Ronaldo	91	270,000	36	Manchester United
Neymar Jr	91	270,000	29	Paris Saint-Germain
K. De Bruyne	91	350,000	30	Manchester City
J. Oblak	91	130,000	28	Atlético de Madrid
K. Mbappé	91	230,000	22	Paris Saint-Germain
M. Neuer	90	86,000	35	FC Bayern München
M. ter Stegen	90	250,000	29	FC Barcelona
H. Kane	90	240,000	27	Tottenham Hotspur
N. Kanté	90	230,000	30	Chelsea
K. Benzema	89	350,000	33	Real Madrid CF
T. Courtois	89	250,000	29	Real Madrid CF
H. Son	89	220,000	28	Tottenham Hotspur
Casemiro	89	310,000	29	Real Madrid CF
V. van Dijk	89	230,000	29	Liverpool
S. Mané	89	270,000	29	Liverpool
M. Salah	89	270,000	29	Liverpool
Ederson	89	200,000	27	Manchester City
J. Kimmich	89	160,000	26	FC Bayern München

club_name	age											Total
	22	26	27	28	29	30	32	33	34	35	36	
Atlético de Madrid				1								1
Chelsea						1						1
FC Barcelona					1							1
FC Bayern München		1					1			1		3
Liverpool					3							3
Manchester City						1						2
Manchester United			1								1	1
Paris Saint-Germain	1				1				1			3
Real Madrid CF					2			1				3
Tottenham Hotspur			1	1								2
Total	1	1	2	2	7	2	1	1	1	1	1	20

Pivoteamento



Figura 1: Salário médio de acordo com idade e clube (em milhares de euros)

club_name	age																											
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	38						
Atlético de Madrid		0.5	4	6		22	21	28	9	71	78	75	94	73	102	63	74		135									
Chelsea			2	6	31	43	107		76	108	62	115	187	105	163	125						105						
FC Barcelona	4	1	28	23	51	75	51	35	188		145	173		185	77	115	200	260	175									
FC Bayern München			19	13	26	17	70	40	49	94	126	110		53		140	125				86							
Liverpool			9		19		80		91	62	94	147	126	217	180	140			53	76								
Manchester City		0.7	8.6	14	10	86			138	178	153	178				252	170				15	100						
Manchester United		2	7.7	26		150	68	133	90	135	178	100	153	73	120		100	100	190	42	270	11						
Paris Saint-Germain		3	5	8	22		122		49	76		87	102	178	68	83			160	225	115							
Real Madrid CF		2	18	16	74	30	110	130	91	190	170	115		228	205	207		243		190								
Tottenham Hotspur	1	3	3.5	6	26	44	60	58	87	87	48	161	137	77											125			

Despivotamento (unpivot, melt)

club_name	age																					
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	38
Atlético de Madrid		0.5	4	6		22	21	28	9	71	78	75	94	73	102	63	74		135			
Chelsea			2	6	31	43	107		76	108	62	115	187	105	163	125					105	
FC Barcelona	4	1	28	23	51	75	51	35	188		145	173		185	77	115	200	260	175			
FC Bayern München			19	13	26	17	70	40	49	94	126	110		53		140	125			86		
Liverpool			9		19		80			91	62	94	147	126	217	180	140			53	76	
Manchester City		0.7	8.6	14	10	86				138	178	153	178		252	170				15	100	
Manchester United		2	7.7	26		150	68	133	90	135	178	100	153	73	120		100	100	190	42	270	11
Paris Saint-Germain		3	5	8	22		122		49	76		87	102	178	68	83		160	225	115		
Real Madrid CF		2	18	16	74	30	110	130	91	190	170	115		228	205	207		243		190		
Tottenham Hotspur	1	3	3.5	6	26	44	60	58	87	87	48	161	137	77						125		



club_name	age	value
Atlético de Madrid	17	0.50
Atlético de Madrid	18	4.00
Atlético de Madrid	19	6.00
Atlético de Madrid	21	22.33
Atlético de Madrid	22	21.33
Atlético de Madrid	23	28.00
Atlético de Madrid	24	9.00
Atlético de Madrid	25	71.00
Atlético de Madrid	26	77.75
Atlético de Madrid	27	74.50
Atlético de Madrid	28	94.00
Atlético de Madrid	29	73.00
Atlético de Madrid	30	102.25
Atlético de Madrid	31	63.00
Atlético de Madrid	32	74.00
Atlético de Madrid	34	135.00
Chelsea	18	2.00
Chelsea	19	6.00
Chelsea	20	30.67
Chelsea	21	43.33
Chelsea	22	106.50
Chelsea	24	75.50
Chelsea	25	108.00
Chelsea	26	61.67
Chelsea	27	115.00
Chelsea	28	186.67
Chelsea	29	105.00
Chelsea	30	162.50
Chelsea	31	125.00
Chelsea	36	105.00

Concatenar (union all)

União de tabelas ou conjuntos de dados

FIFA 22

short_name	club_name
L. Messi	Paris Saint-Germain
Neymar Jr	Paris Saint-Germain
K. Mbappé	Paris Saint-Germain
G. Donnarumm	Paris Saint-Germain
Sergio Ramos	Paris Saint-Germain
K. Navas	Paris Saint-Germain

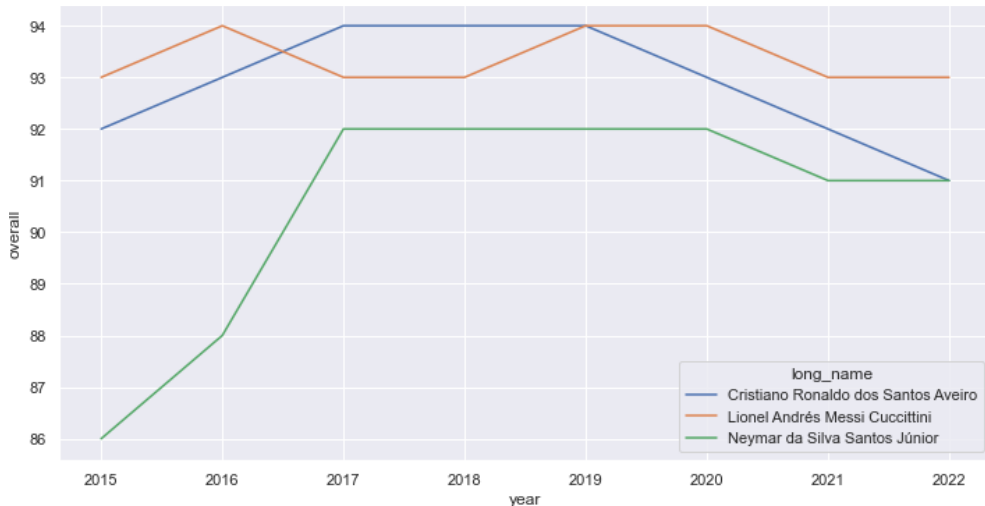
FIFA 21

short_name	club_name
Neymar Jr	Paris Saint-Germain
K. Mbappé	Paris Saint-Germain
Á. Di María	Paris Saint-Germain
K. Navas	Paris Saint-Germain
M. Verratti	Paris Saint-Germain
M. Icardi	Paris Saint-Germain



short_name	club_name
Neymar Jr	Paris Saint-Germain
K. Mbappé	Paris Saint-Germain
Á. Di María	Paris Saint-Germain
K. Navas	Paris Saint-Germain
M. Verratti	Paris Saint-Germain
M. Icardi	Paris Saint-Germain
L. Messi	Paris Saint-Germain
Neymar Jr	Paris Saint-Germain
K. Mbappé	Paris Saint-Germain
G. Donnarumm	Paris Saint-Germain
Sergio Ramos	Paris Saint-Germain
K. Navas	Paris Saint-Germain

Concatenar + pivoteamento



Mesclar (merge, join)

Produto cartesiano de duas tabelas onde a chave determinada coincide

sofifa_id	short_name	overall	age	club_team_id	club_team_id	club_name	league_name	level	#players	total_value	total_wage
158023	L. Messi	93	34	73	73	Paris Saint-Germain	French Ligue 1	1	33	1,227.40	3.00
188545	R. Lewandowski	92	32	21	21	FC Bayern München	German 1. Bundesliga	1	28	969.48	2.04
20801	Cristiano Ronaldo	91	36	11	11	Manchester United	English Premier League	1	33	1,008.73	3.41
190871	Neymar Jr	91	29	73	10	Manchester City	English Premier League	1	32	1,299.51	3.61
192985	K. De Bruyne	91	30	10	240	Atlético de Madrid	Spain Primera Division	1	32	952.48	1.78
200389	J. Oblak	91	28	240	241	FC Barcelona	Spain Primera Division	1	33	834.30	3.23
231747	K. Mbappé	91	22	73	18	Tottenham Hotspur	English Premier League	1	33	729.60	2.08
167495	M. Neuer	90	35	21	5	Chelsea	English Premier League	1	33	934.15	2.98
192448	M. ter Stegen	90	29	241	243	Real Madrid CF	Spain Primera Division	1	33	987.58	4.39
202126	H. Kane	90	27	18	9	Liverpool	English Premier League	1	33	1,046.38	3.21



sofifa_id	short_name	overall	age	club_team_id	club_team_id.1	club_name	league_name	level	#players	total_value	total_wage
158023	L. Messi	93	34	73	73	Paris Saint-Germain	French Ligue 1	1	33	1227.40	3.00
190871	Neymar Jr	91	29	73	73	Paris Saint-Germain	French Ligue 1	1	33	1227.40	3.00
231747	K. Mbappé	91	22	73	73	Paris Saint-Germain	French Ligue 1	1	33	1227.40	3.00
188545	R. Lewandowski	92	32	21	21	FC Bayern München	German 1. Bundesliga	1	28	969.48	2.04
167495	M. Neuer	90	35	21	21	FC Bayern München	German 1. Bundesliga	1	28	969.48	2.04
20801	Cristiano Ronaldo	91	36	11	11	Manchester United	English Premier League	1	33	1008.73	3.41
192985	K. De Bruyne	91	30	10	10	Manchester City	English Premier League	1	32	1299.51	3.61
200389	J. Oblak	91	28	240	240	Atlético de Madrid	Spain Primera Division	1	32	952.48	1.78
192448	M. ter Stegen	90	29	241	241	FC Barcelona	Spain Primera Division	1	33	834.30	3.23
202126	H. Kane	90	27	18	18	Tottenham Hotspur	English Premier League	1	33	729.60	2.08

Tipos de join



Figura 2: Fonte: <https://www.metabase.com/learn/sql-questions/sql-join-types>



Left outer join



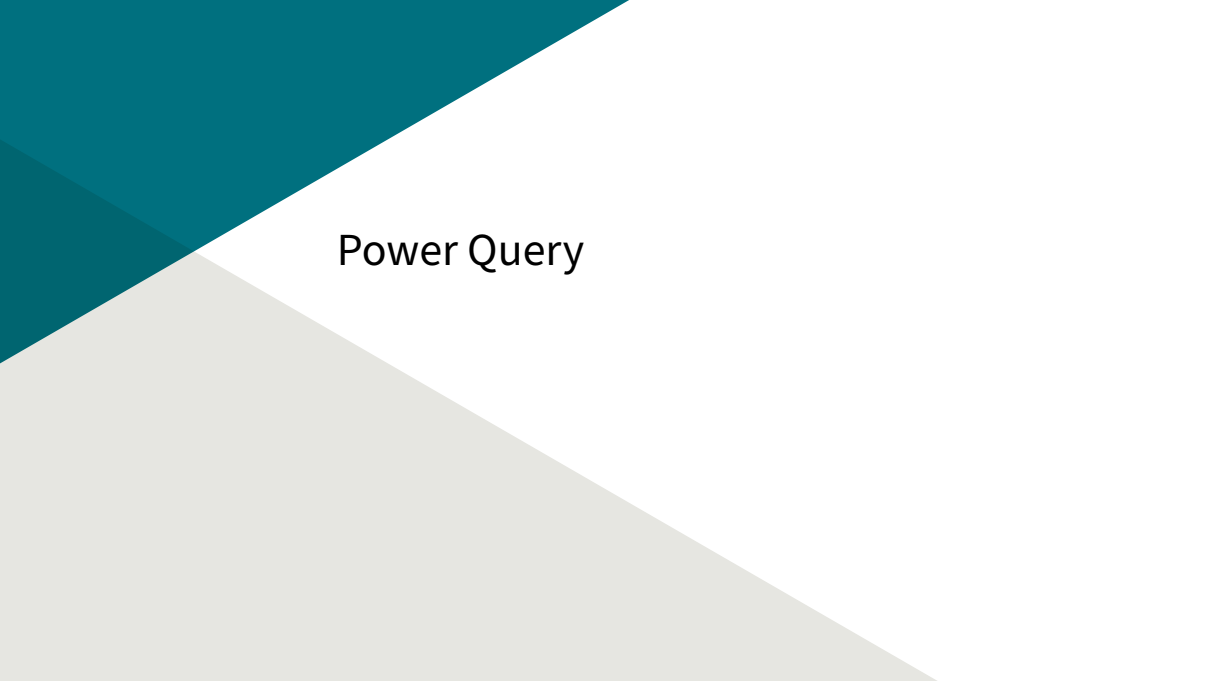
Inner join



Right outer join



Full outer join

The background of the slide is composed of two large, overlapping geometric shapes. A teal-colored shape occupies the top-left corner, while a light gray shape occupies the bottom-left corner. The rest of the slide is white. The text 'Power Query' is centered in the white area.

Power Query

Prós e contras



Pró	Contra
Completo	Ambiente Excel
Interface amigável	"Poucos" dados
Altamente personalizável	Gráficos limitados*
Pouquíssima infraestrutura	Execução lenta
Desenvolvimento ágil	



Python

Prós e contras



Pró

Contra

Linguagem de programação

Curva de aprendizado*

Fácil integração

Lento para muitos dados

Solução completa

Desenvolvimento ágil

Pouca infraestrutura

Exercício



Expandir as análises feitas. Sugestões:

- ▶ Use sua criatividade!
- ▶ Transformar para gráficos em interativos (sugestão: plotly)
- ▶ Explorar mais campos e relações entre campos
- ▶ Desenvolver método para otimizar formações dos times existentes
- ▶ Desenvolver método para otimizar links fortes (FIFA Ultimate Team)

The background of the slide is composed of two large, overlapping geometric shapes. A teal-colored shape occupies the top-left corner, while a light gray shape occupies the bottom-left corner. The rest of the slide is white.

Breves comentários sobre SQL

Prós e contras



Pró

Contra

Qualquer volume de dados

Curva de aprendizado*

Fácil integração

Infraestrutura grande

Eficiente

Fácil desenvolvimento