

Power BI: Web Scraping e Tratamento de Dados

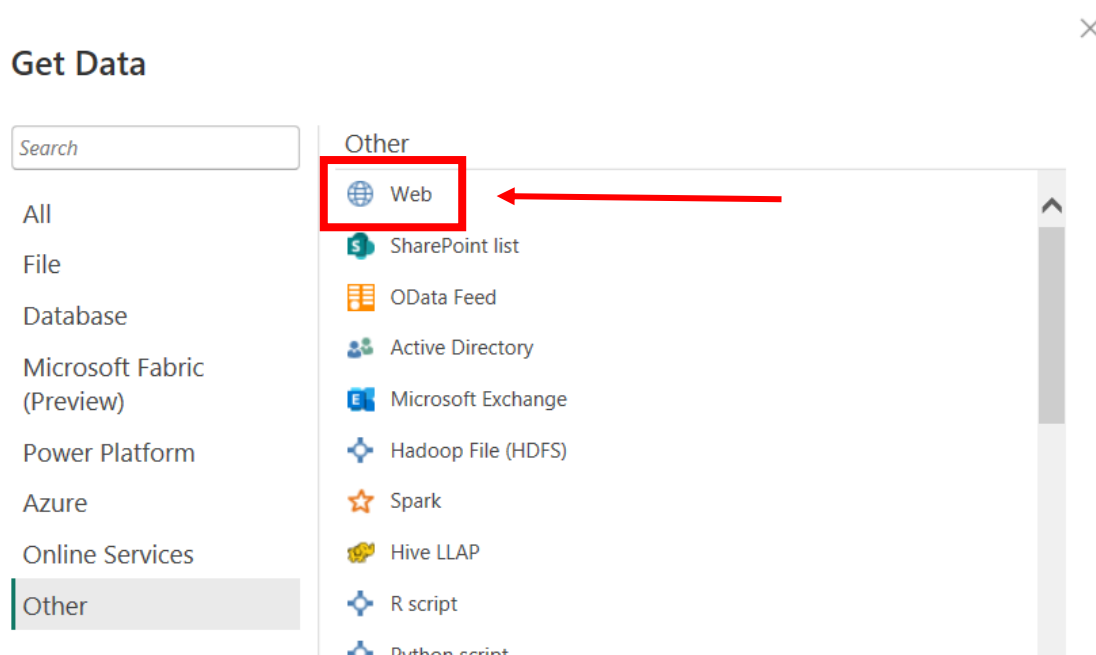
Web Scraping?

Podemos usar o Power BI para fazer uma coisa chamada Web Scraping, mas o que isso significa?

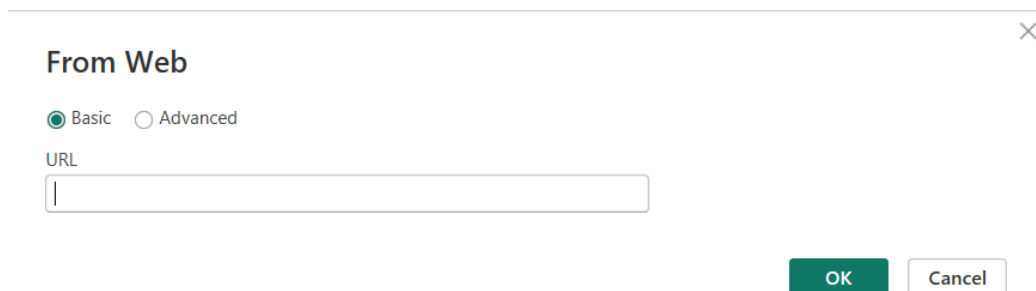
Web Scraping é a coleta de dados da web. Podemos extrair dados de um website, tratar esses dados e com base nos mesmos, podemos criar nossos relatórios.

Como eu extraio os dados?

Vá em **Get Data** e procure a opção **Web**.



Em seguida cole a URL do site que você deseja.



Navigator

The screenshot shows the Power BI Navigator interface. On the left, there is a search bar and a list of tables under 'Suggested Tables [7]'. 'Table 34' is selected. On the right, the 'Table View' tab is active, displaying a preview of 'Table 34' with 5 columns and 15 rows of data. Below the table, there are buttons for 'Add Table Using Examples', 'Load', 'Transform Data', and 'Cancel'.

Column1	Column2	Column3	Column4	Column5
€180.00m	5	2	Kylian Mbappé	Centre-Forward
€120.00m	8	4	Bukayo Saka	Right Winger
€110.00m	9	6	Phil Foden	Left Winger
€90.00m	6	8	Gavi	Central Midfield
€90.00m	8	10	Rodri	Defensive Midfield
€90.00m	8	12	Martin Ødegaard	Attacking Midfield
€85.00m	7	14	Eduardo Camavinga	Central Midfield
€85.00m	4	16	Khvicha Kvaratskhelia	Left Winger
€80.00m	7	18	Randal Kolo Muani	Centre-Forward
€80.00m	7	20	Marcus Rashford	Left Winger
€80.00m	6	22	Bernardo Silva	Attacking Midfield
€75.00m	6	24	Frenkie de Jong	Central Midfield

Depois o Power BI mostrará opções para você, você pode selecionar o que achar melhor e começar o tratamento, caso você queira uma tabela mais personalizada ao seu gosto, clique em **Add Table Using Examples**, que é uma opção que abre o site dentro do Power BI e você pode criar sua tabela de uma maneira mais personalizada.

No meu caso eu extraí os dados do site do Transfermarkt, para ser mais específico, extraí os 25 jogadores mais valiosos do mundo.

Depois, utilize o Power Query para tratar os seus dados e depois de pronto, você pode começar a criar o seu relatório.

Meu exemplo

Para colorar em prática esses ensinamentos eu fiz a extração de dados do site Transfermarkt onde eu peguei os 25 jogadores mais valiosos do mundo.

Depois da extração, fiz o tratamento usando o Power Query, segue abaixo a imagem com os tratamentos feitos.

Aç Player Name	Aç Position	Aç Nationality	Yç Age	Aç Club	Aç League	\$ Market Value
Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
Erling Haaland	Centre-Forward	Norway		23 Manchester City	Premier League	180.000.000,00
Kylian Mbappé	Centre-Forward	France/Cameroon		24 Paris Saint Germain	Ligue 1	180.000.000,00
Jude Bellingham	Central Midfield	English/Ireland		20 Real Madrid	La Liga	120.000.000,00
Bukayo Saka	Right Winger	English/Nigeria		22 Arsenal	Premier League	120.000.000,00
Jamal Musiala	Attacking Midfield	Germany/English		20 Bayern München	Bundesliga	110.000.000,00
Phil Foden	Left Winger	English		23 Manchester City	Premier League	110.000.000,00
Pedri	Central Midfield	Spain		20 FC Barcelona	La Liga	100.000.000,00
Gavi	Central Midfield	Spain		19 FC Barcelona	La Liga	90.000.000,00
Declan Rice	Defensive Midfield	English/Ireland		24 Arsenal	Premier League	90.000.000,00
Rodri	Defensive Midfield	Spain		27 Manchester City	Premier League	90.000.000,00
Rafael Leão	Left Winger	Portugal/Angola		24 AC Milan	Serie A	90.000.000,00
Martin Ødegaard	Attacking Midfield	Norway		24 Arsenal	Premier League	90.000.000,00
Harry Kane	Centre-Forward	English		30 Bayern München	Bundesliga	90.000.000,00
Eduardo Camavinga	Central Midfield	France/Congo		20 Real Madrid	La Liga	85.000.000,00
Florian Wirtz	Attacking Midfield	Germany		20 Bayer Leverkusen	Bundesliga	85.000.000,00
Khvicha Kvaratskhelia	Left Winger	Georgia		22 Napoli	Serie A	85.000.000,00
Aurélien Tchouaméni	Defensive Midfield	France/Cameroon		23 Real Madrid	La Liga	85.000.000,00
Randal Kolo Muani	Centre-Forward	France/DR Congo		24 Paris Saint Germain	Ligue 1	80.000.000,00
Christopher Nkunku	Second Striker	France/DR Congo		25 Chelsea	Premier League	80.000.000,00
Marcus Rashford	Left Winger	English/ST.Kitts & Nevis		25 Manchester United	Premier League	80.000.000,00
Rúben Dias	Centre-Back	Portugal		26 Manchester City	Premier League	80.000.000,00
Bernardo Silva	Attacking Midfield	Portugal		29 Manchester City	Premier League	80.000.000,00
Josko Gvardiol	Centre-Back	Croatia		21 Manchester City	Premier League	75.000.000,00
Freddie de Jong	Central Midfield	Netherlands		26 FC Barcelona	La Liga	75.000.000,00
Matthijs de Ligt	Centre-Back	Netherlands		24 Bayern München	Bundesliga	75.000.000,00

Falta uma coluna que é o **Id Player**. Tirando isso são os dados extraídos do Transfermarkt mais os tratamentos que fiz.



Em seguida resolvi criar alguns insights, como o valor total do valor de mercado de todas as ligas, de todos os 25 jogadores que é 2,43 bilhões de euros.

A média de idade dos 25 jogadores que é 23,40.

A quantidade de jogadores por liga, como você pode ver 11 são da Premier League da Inglaterra, 6 são de La Liga (Espanha), 4 da Bundesliga (Alemanha), 2 da Série A (Itália) e Ligue 1 (França).

Por último um gráfico que mostra a soma do valor de mercado baseado na liga, para exemplificas, a soma do valor de mercados dos 11 jogadores da Premier League em euros é 1,08bn.

Dos 6 jogadores de La Liga é de 550M de euros.



Na segunda página eu criei uma tabela onde mostram os jogadores cujo o valor de mercado é igual ou superior a 100M de euros. Embaixo tem a 25 média dos jogadores que jogam nas ligas registradas nos dados que eu extraí.

Dos 11 que jogam na Premier League, a média de idade é 24,45. Dos 6 de La Liga é 24 e da Bundesliga é 23,50.

Vou disponibilizar esse estudo no meu repositório no GitHub.