

Adatok átalakítása Power Queryvel

4

Ebben a fejezetben:

- megismerjük a Power Query szerkesztőt,
- megértjük az adatminőség és az adattisztítás szerepét,
- meghatározzuk az adattípusokat,
- oszlopokat alakítunk át,
- táblázatszerkezeteket módosítunk,
- táblákat egyesítünk és adunk hozzá a modellhez.

Adatokkal való munkavégzés során ideális esetben az adatok elemzésre optimális állapotban állnak rendelkezésre. Általában azonban ennek az ellenkezője igaz, vagyis gyakran szükség van bizonyos változtatásokra vagy átalakításokra, hogy az adatok egységesek, rendezettek és értelmezhetők legyenek. Más esetekben az elvégzendő elemzés követelményeinek megfelelően új oszlopokkal kell kiegészíteni a táblákat, kijavítani a hibákat, kicserélni a rekordokat stb.

A Power Queryben megtalálható egy sor olyan funkció, amely a felhasználói felületen keresztül, *programozási ismeretek nélkül* lehetővé teszi az adatok átalakítását, néhány kattintással történő módosítását. Ez a látszólagos egyszerűség hatalmas rejtett előny a Power BI-ban, mivel lehetővé teszi az adatminőség és a pontosság követelményeinek való megfelelést anélkül, hogy módosítani kellene az eredeti forrás adatait.

Amikor adatátalakításról beszélünk, azokra a műveletekre vagy tevékenységekre utalunk, amelyeket az adatok minőségének biztosítása érdekében hajtanak végre. A hibák általában problémákhoz vezetnek, mert befolyásolják a számítások pontosságát, következetlenségeket eredményeznek, amelyek később hátráltatják a jelentés kialakítását és kárt okoznak – ez a bemutatott adatokba vetett bizalom elvesztésével járhat.

Az adatok szemantikai modellbe történő betöltése előtt az elsődleges feladat annak megállapítása, hogy az adatok megfelelnek-e a számítások elindításához szükséges követelményeknek. Ha nem, akkor több lépésben kell azokat kiigazítani. Erről lesz szó ebben a fejezetben.

ISMERJÜK MEG A POWER QUERYT

A Power Query az az adat-előkészítő és -átalakító motor, amely minden Power BI-ban létrehozott kapcsolatban megtalálható. Ez az eszköz egy M nevű, több mint 700 függvényből álló, hatékony nyelvet használ.

A Power Query számos előre definiált átalakítást kínál a munka felgyorsításához, például rekordok szűrését, oszlopok kiválasztását, értékek rendezését, adattípusok meghatározását, új oszlopok hozzáadását, lekérdezések egyesítését és hozzáfűzését. Az egyik legjellegzetesebb tulajdonsága azonban az, hogy a felhasználói felület segítségével anélkül érhetjük el a kívánt eredményt, hogy pontosan tudnánk, melyik a megfelelő függvény.

Ezzel az eszközzel egyensúlyt teremthetünk a néhány kattintással elérhető funkciók és a feladatok megoldásában nagyobb rugalmasságot biztosító speciális funkciók közötti használat között.

AZ ADATMINŐSÉG ÉS ADATTISZTÍTÁS SZEREPE

A szükséges minőségű, tiszta adatokkal való munka kulcsfontosságú, mivel ez később, a jelentés készítésekor is megmutatkozik. Az inkonzisztenciák általában az eredeti adatforrásból vagy az azokat generáló közbenső folyamatokból származnak. Bár nem lehetetlen, de nagy valószínűséggel nem fogsz olyan helyzetbe kerülni, ahol az adatok 100%-ban optimálisak lennének a szemantikai modellbe való azonnali betöltéshez.

A hibás, rendezetlen adatok betöltésének következményei sokféle formát ölthetnek, a számítások pontosságával kapcsolatos problémáktól kezdve az ismétlődő értékeket tükröző inkonzisztenciákig, vagy akár a tényadatok teljességét torzító rekordvesztésig. Mivel ez a Power BI-ban végzett munka egyik első lépése, ezáltal lehetővé válik, hogy döntéseket hozz az adatok állapotáról, hogy azokat az igényeidhez igazítsd és alakítsd.

ADATTÍPUSOK MEGHATÁROZÁSA

Az adattípusok hozzárendelése fontos lépés a minőség biztosításában. Először is, mert meghatározza az adatok jellegét, osztályozza azokat, másodsor pedig azért, mert meghatározza a számunkra elérhető átalakítási lehetőségeket. Nem minden átalakítás működik általánosan a különböző adattípusokon; éppen ellenkezőleg, egyesek műveletek csak bizonyos típusok esetében engedélyezettek.

Alapértelmezés szerint a Power BI az automatikus felismerés révén határozza meg az egyes oszlopok adattípusait. Ezeket az adattípusokat az **Átalakítás > Adattípus** alatt legördülő panelen módosíthatjuk. Ezenkívül, miután kiválasztottunk egy vagy több oszlopot, az **Átalakítás > Adattípus észlelése** gombra kattintással gyorsan hozzárendelhetjük az oszlopokhoz a megfelelő adattípusokat. (Több oszlop kijelöléséhez a Ctrl, illetve Shift billentyűket használhatjuk.)



TIPP Általánosságban elmondható, hogy az adattípusok meghatározása az egyik első lépés, amelyre figyelmet kell fordítani a Power Query használata során.

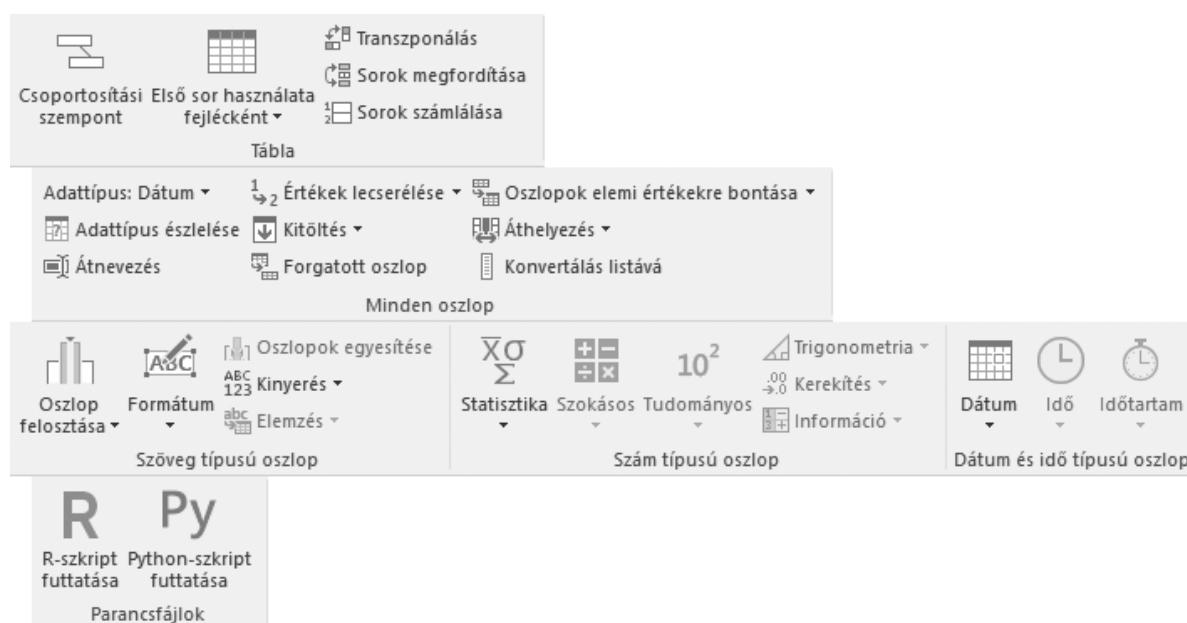
Tizedes tört
Fixpontos tizedes tört
Egész szám
Százalék
Dátum/idő
Dátum
Idő
Dátum/idő/időzóna
Időtartam
Szöveg
Igaz/Hamis
Bináris

Az oszlopokban elérhető adattípusok

OSZLOPALAPÚ ÁTALAKÍTÁSOK

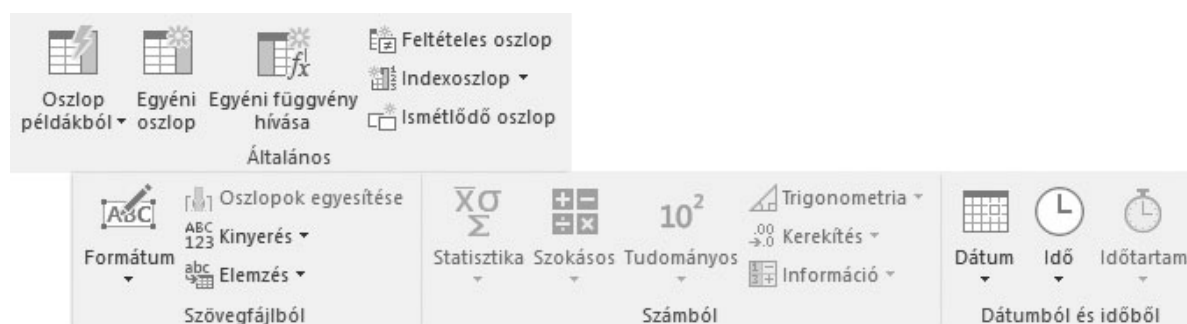
Mielőtt megismerkednél az oszlopalapú átalakításokkal, gondold át: az átalakítások a Power Querybe betöltött táblából származó oszlopokon történnek. Ugyanakkor olyan átalakításokat is végrehajthatsz, amelyek eredménye új oszlopokban jelenik meg, amelyek szintén a szemantikai modell részét fogják képezni.

Az **Átalakítás** menü olyan opciókat tartalmaz, amelyek közvetlenül egy vagy több kiválasztott oszlopra vonatkoznak.



*Átalakítási lehetőségek az **Átalakítás** menüben*

Másrészt az **Oszlop hozzáadása** menüben új oszlopok hozzáadásával is átalakíthatjuk az adatokat. Hogy melyik opciót válassza, az elsősorban a feladattól függ, így nincs egyetlen helyes megoldás erre a feladatra.



*Átalakítási lehetőségek az **Oszlop hozzáadása** menüben*

Hiányzó tartalom kitöltése

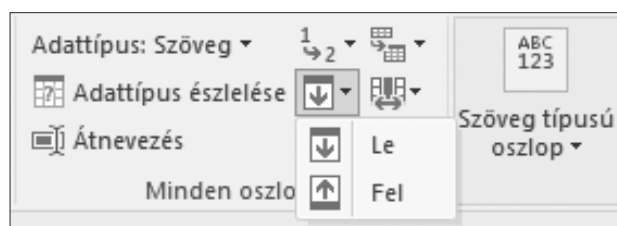
Előfordulhat, hogy a rekordokban hiányosságok vannak. Ezek kitöltéséhez használhatod a *felfelé*, illetve *lefelé kitöltés* funkciót. A két lehetőség közötti különbség a kitöltés irányában rejlik. A lefelé kitöltés közvetlenül az egymást követő null értékek előtt álló rekordot, míg a felfelé kitöltés közvetlenül az egymást követő null értékek után álló rekordot veszi figyelembe.

	A ^B _C Város	A ^B _C Állam	A ^B _C ÁllamTeljes
1	Corona	CA	California
2	Los Angeles	CA	null
3	Salinas	CA	null
4	Santa Fe Springs	CA	null
5	New Leake	Boston	Boston
6	Hoffleet Stow	Boston	Boston
7	Wrangle	Boston	null
8	Little Rock	AR	Arkansas
9	Elaine	AR	Arkansas
10	Lake Village	AR	Arkansas

Az **ÁllamTeljes** oszlop kitöltésre váró üres mezőket tartalmaz

Kövesd az alábbi lépéseket a lefelé kitöltés végrehajtásához:

1. Válaszd ki az **ÁllamTeljes** oszlopot az oszlopnévre való kattintással.
2. Lépj az **Átalakítás** fülre.
3. Kattints a **Kitöltés > Le** ikonra.



A nyíl iránya jelzi a kitöltés típusát

A lefelé kitöltés alkalmazásával a hiányos rekordok null értékeinek a helyére az előző sorban lévő érték kerül, anélkül hogy a már adatokat tartalmazó mezők megváltoznának.

	A ^B _C Város	A ^B _C Állam	A ^B _C ÁllamTeljes
1	Corona	CA	California
2	Los Angeles	CA	California
3	Salinas	CA	California
4	Santa Fe Springs	CA	California
5	New Leake	Boston	Boston
6	Hoffleet Stow	Boston	Boston
7	Wrangle	Boston	Boston
8	Little Rock	AR	Arkansas
9	Elaine	AR	Arkansas
10	Lake Village	AR	Arkansas

Most az **ÁllamTeljes** oszlop minden sorában található adatok



TIPP Az Excel-munkafüzet beolvasása után előfordul, hogy a Power Queryben az oszlopok fejléceiben a **Column1**, **Column2** stb. szavak szerepelnek, a valódi oszlopfejlécek pedig a tábla első sorába kerülnek. A probléma helyrehozásához kattints az **Átalakítás** menüben az **Első sor használata fejlécként** gombra.

Duplikátumok eltávolítása

A duplikált értékek eltávolítása egy másik gyakori művelet az adatok átalakításakor. Ezúttal nem az **Átalakítás** menüből érheted el a műveletet, hanem jobb gombbal kattintva.

	A ^B _C Keresztnév	A ^B _C Vezetéknév	A ^B _C Cím	A ^B _C Város
1	Julian	McGuigan	25 Railway Street	Canning Creek
2	Julian	McGuigan	25 Railway Street	Canning Creek
3	Annabelle	Townsend	83 Webb Road	Mayfield North
4	Jamie	Hetherington	3 Link Road	Reekara
5	Jamie	Hetherington	3 Link Road	Reekara
6	Gabrielle	Castella	74 Old Tenterfield Road	Mallanganee
7	Hayley	Jull	61 McKillop Street	Rocky Point
8	Natalie	Hilder	63 Davenport Street	Paddys Flat
9	Hunter	Hutchins	81 Bayley Street	Kinglake
10	Maya	Atherton	53 Pipeclay Road	Kindee

Az oszlopok duplikált értékeket tartalmaznak Julian és Jamie esetében