

دوستان، موضوعی که می‌خواهم به آن بپردازم این است که ما معمولاً نمی‌خواهیم نیازهای واقعی‌مان را بشناسیم. این جمله در تمام جنبه‌های زندگی ما صادق است. در اینجا منظور من دانش کار با انواع داده‌هاست

این دانش به شدت برای همه کسانی که در شرکت‌های دولتی و خصوصی با خواندن، تجمیع، تمیز کردن، بصری کردن (visualize) و تفسیر داده‌ها سروکار دارند، ضروری است. اما تعداد افرادی که برای کسب این مهارت تلاش می‌کنند، بسیار کم است. ما با طیفی از افراد روبرو هستیم. در یک سمت این طیف، کسانی هستند که به هیچ وجه تغییر را نمی‌پذیرند. راستش را بخواهید، تغییر ترسناک است؛ زیرا به معنای پذیرش این است که تا به حال اشتباه کرده‌ایم.

دسته‌ای دیگر تغییر را می‌پذیرند، اما مسئولیت آن را بر عهده نمی‌گیرند. این افراد منتظرند تا دیگران، مانند واحد IT، گزارشات مورد نیاز را تهیه کرده و در دسترس کاربران قرار دهند یا اینکه فقط به نوشتن چند فرمول ساده در اکسل بسنده کنند تا بتوانند چهار عمل اصلی را انجام دهند.

گروه سوم افرادی هستند که با هر انگیزه‌ای به دنبال کسب و توسعه این مهارت هستند. این افراد با دستکاری‌هایی که روی خروجی گزارشات سیستم انجام می‌دهند، سعی می‌کنند آن‌ها را قابل فهم‌تر و مفیدتر کنند و حتی کارکنان IT را تحت فشار قرار می‌دهند تا نیازهای کشف شده را برای بهبود سیستم برطرف کنند. گروه چهارم علاوه بر همه این‌ها، دلسوزانه - اگر نگویم پیغمبرانه - تلاش می‌کنند تا دانش خود را به دیگران منتقل کنند. این افراد گاهی در ذهنشان زمزمه می‌کنند: "کس ندارد ذوق مستی..."

حالا شاید صدای شما بلند شود و بگویید: "قرار بود درباره PDF صحبت کنی، چرا شعر تحویل من می‌دی؟" اما باید بگویم که پاور کوئری یک ابزار همه‌فن‌حریف است! یعنی چه؟ فرض کنید یک صفحه وب دارید که شامل جدولی است که برای شما مفید است. درست مانند همین جدول، شما یک جدول مشابه در اکسل دارید و همچنین جدولی دیگر در Access حالا انتظار دارید من همین جا تمامش کنم؟ نه! لطفاً کمی صبر کنید.

علاوه بر این‌ها، یک جدول دیگر به صورت CSV روی هارد دیسک شما ذخیره شده و از همه مهم‌تر، یک فایل PDF دارید! بله، درست شنیدید؛ یک فایل PDF که شامل جدولی مشابه سایر جداول است. پاور کوئری قابلیت خواندن همه این‌ها را دارد! شوخی نمی‌کنم؛ باور کنید!

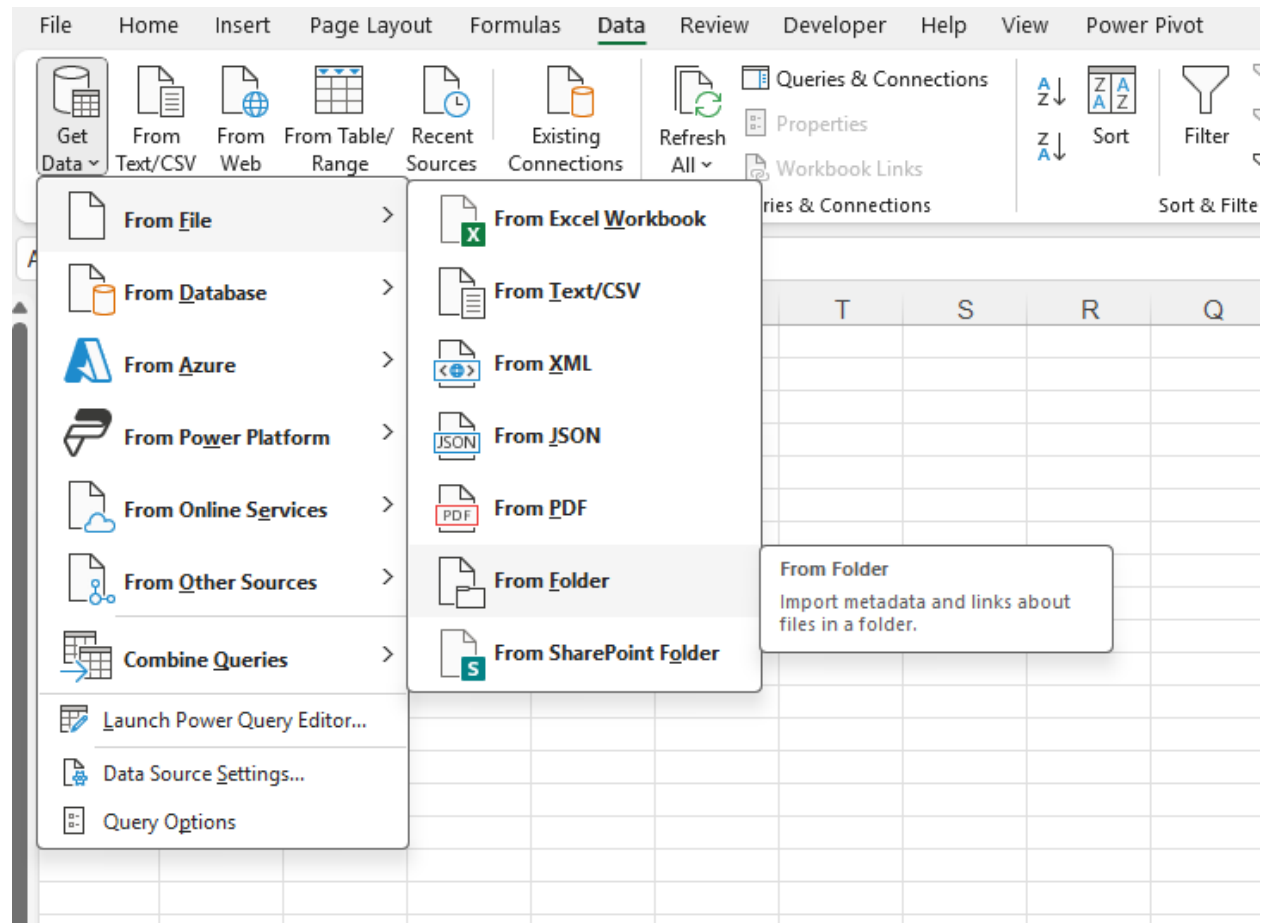
علاوه بر این، پاور کوئری در یک محیط کاربرپسند نه تنها همه داده‌ها را زیر هم قرار می‌دهد (Append)، بلکه نتیجه را به صورت جدول معمولی یا PivotTable یا نمودار در اکسل نمایش می‌دهد. اگر هر یک از فایل‌هایی که ذکر کردم آپدیت شوند (مثلاً فایلی که تحت وب است)، شما با یک Refresh می‌توانید همه آن‌ها را دوباره زیر هم بچینید.

شاید بپرسید: "چگونه می‌توانم پاور کوئری را تهیه کنم؟" خوشبختانه مایکروسافت از نسخه ۲۰۱۶ به بعد دسترسی به پاور کوئری را بسیار آسان کرده است. کافایت به سربرگ Data بالای صفحه نگاه کنید؛ گروه Get & Transform Data به وضوح قابل مشاهده است.

ما قصد داریم تمام جداول موجود در فایل‌های PDF را زیر هم در یک فایل اکسل بیاوریم. پیشنهاد می‌کنم قبل از هر چیز به ضمیمه مقاله مراجعه کنید و دو فایل PDF را با دقت بررسی کنید. توجه داشته باشید که اطلاعات مشابه هستند، اما ترتیب ستون‌ها ممکن است متفاوت باشد. بنابراین لطفاً با دقت به هر شش جدول نگاه

کنید و اختلافات آن‌ها را کشف کنید. سپس برای همراهی با من یک پوشه بسازید و هر دو فایل PDF را در آن کپی کنید.

در اکسل، برای دسترسی به پوشه‌ای که حاوی فایل‌های PDF است و ارسال آن‌ها به ادیتور پاور کوئری، مراحل زیر را دنبال کنید:



مسیر دسترسی:

از منوی Data به Get Data بروید.

سپس گزینه From File و بعد From Folder را انتخاب کنید.

توجه به توضیحات اکسل:

اکسل توضیحاتی درباره تب From Folder دارد. این توضیحات شامل نحوه انتقال داده‌های متا و لینک‌ها درباره فایل‌های موجود در پوشه است. بنابراین، انتظار نداشته باشید که پاور کوئری به طور خودکار فایل‌ها را باز کند.

انتقال به محیط پاور کوئری:

در پنجره بعدی، روی دکمه Transform Data کلیک کنید تا به محیط پاور کوئری بروید.

E:\مقالات\PDF_Folder\Pdfs

Content	Name	Extension	Date accessed	Date modified	Date created	Attributes	Folder Path
Binary	PDF_Data1.pdf	.pdf	1/16/2025 6:07:09 PM	1/14/2025 8:52:49 PM	1/14/2025 8:30:07 PM	Record	E:\مقالات\PDF_Folder\Pdfs\
Binary	PDF_Data2.pdf	.pdf	1/15/2025 11:09:14 PM	1/14/2025 9:08:00 PM	1/14/2025 9:08:00 PM	Record	E:\مقالات\PDF_Folder\Pdfs\

Combine

Load

Transform Data

Cancel

مشاهده ستون‌ها:

در سه ستون اول، نام فایل، پسوند و ستون Content را مشاهده می‌کنید. ستون Content شامل محتوای فایل‌هایی است که از پوشه انتخاب شده بارگذاری شده‌اند. این محتوا می‌تواند شامل متن، داده‌های جدولی یا هر نوع اطلاعات دیگری باشد. با استفاده از این ستون می‌توانید به راحتی به محتوای هر فایل دسترسی پیدا کرده و آن را پردازش کنید.

= Table.TransformColumns(Source , { "Content" , Pdf.Tables })

	Content	Name	Extension	Date accessed
1	Table	PDF_Data1.pdf	.pdf	1/14/2025 8:52:49 PM
2	Table	PDF_Data2.pdf	.pdf	1/14/2025 9:08:01 PM

< _____ >

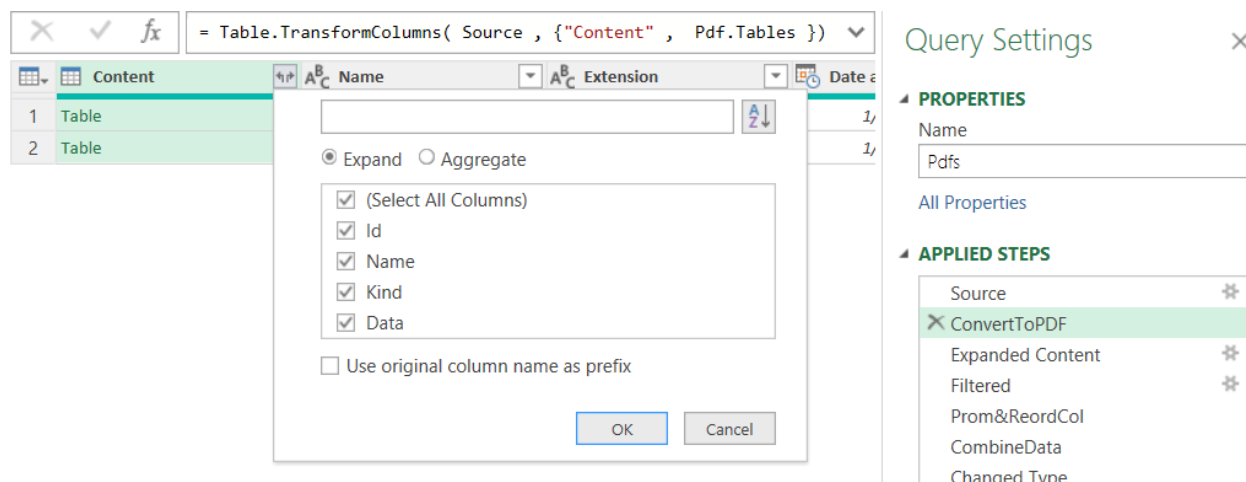
ID	Name	Kind	Data
Page001	Page001	Page	Table
Table001	Table001 (Page 1)	Table	Table
Page002	Page002	Page	Table
Table002	Table002 (Page 2)	Table	Table
Page003	Page003	Page	Table
Table003	Table003 (Page 3)	Table	Table

Query Settings

- PROPERTIES**
 - Name

PDFs
 - [All Properties](#)
- APPLIED STEPS**
 - Source
 - ConvertToPDF**
 - Expanded Content
 - Filtered
 - Prom&ReordCol
 - CombineData
 - Changed Type

سپس دکمه OK را فشار دهید.



کد تولید شده توسط پاور کوئری:

کد زیر توسط پاور کوئری تولید می‌شود:

```
Table.ExpandTableColumn(ConvertToPDF, "Content", {"Id", "Name", "Kind", "Data"}, {"Id", "Name.1", "Kind", "Data"})
```

با اجرای این کد، چهار ستون جدید جایگزین ستون Content می‌شوند.

بررسی داده‌ها:

بر روی بخش خالی هر سطر در ستون Data کلیک کنید و همزمان به آیتم‌های ستون Kind توجه کنید. اگر در ستون Kind با "Table" مواجه شدید، جداول موجود در فایل‌های PDF در ستون Data قابل مشاهده است.

اگر عبارت "Page" در ستون Kind وجود داشته باشد، علاوه بر جداول، قسمتی از متن فایل‌های PDF نیز قابل مشاهده خواهد بود.

مهم: پیشنهاد می‌شود ابتدا ردیف‌هایی که "Table" هستند را فیلتر کنید تا داده‌ها دبل نشوند.

`= Table.ExpandTableColumn(ConvertToPDF, "Content", {"Id", "Name", "Kind"},`

	A ^B C Id	A ^B C Name.1	A ^B C Kind	Data
1	Page001	Page001	Page	Table
2	Table001	Table001 (Page 1)	Table	Table
3	Page002	Page002	Page	Table
4	Table002	Table002 (Page 2)	Table	Table
5	Page003	Page003	Page	Table
6	Table003	Table003 (Page 3)	Table	Table
7	Page001	Page001	Page	Table
8	Table001	Table001 (Page 1)	Table	Table
9	Page002	Page002	Page	Table
10	Table002	Table002 (Page 2)	Table	Table
11	Page003	Page003	Page	Table
12	Table003	Table003 (Page 3)	Table	Table

Query Settings

PROPERTIES

Name
Pdfs

[All Properties](#)

APPLIED STEPS

- Source
- ConvertToPDF
- Expanded Content**
- Filtered
- Prom&ReordCol
- CombineData
- Changed Type

اهمیت مرحله: Expand:

اگر فقط تیک گزینه **Data** را نگه دارید و فیلتر نکنید، نمی‌توانید به داده‌های جداول دسترسی پیدا کنید. بعلاوه دست آخر با انبوهی از داده‌های تکراری روبرو هستید مرحله Expand بسیار مهم است.

✕ ✓ *f_x*

= Table.SelectRows("#Expanded Content", each ([Kind] = "Table"))

↗

	A ^B Id	A ^B Name.1	A ^B Kind	Data
1	Table001	Table001 (Page 1)	Table	Table
2	Table002	Table002 (Page 2)	Table	Table
3	Table003	Table003 (Page 3)	Table	Table
4	Table001	Table001 (Page 1)	Table	Table
5	Table002	Table002 (Page 2)	Table	Table
6	Table003	Table003 (Page 3)	Table	Table

⏪⏩

Column1	Column2	Column3	Column4	Column5	Column6
Date	SalesRep	Customer	Amt	Profit	Region
01-Jan-05	Jaspreet	VCC	10150	2200	North
01-Jan-05	Charley	Sharma & Co	12100	3640	East
06-Jan-05	Jaspreet	Good Fly	14750	3200	North
06-Jan-05	Ramesh	Data Tronics	15000	5280	West

✕

Query Settings

✕

➤ **PROPERTIES**

Name

Pdfs

All Properties

➤ **APPLIED STEPS**

Source

ConvertToPDF

Expanded Content

✕ Filtered

Prom&ReordCol

CombineData

Changed Type

تنظیم سرستون‌ها:

برای تنظیم سرستون‌ها ، باید یک مرحله (Step) اضافه کنید و از کد زیر استفاده کنید:

```
Table.TransformColumns(Filtered, {"Data", each Table.PromoteHeaders( )})
```

با این کار سرستون‌ها در جای خود قرار می‌گیرند.

تغییر ترتیب ستون‌ها:

برای تغییر ترتیب ستون‌ها، تابع زیر را استفاده کنید:

```
Table.TransformColumns(Filtered ,  
    {"Data", each Table.ReorderColumns (  
        Table.PromoteHeaders(_), ColumnList)  
    }  
)
```

با اجرای این کد، پاور کوئری سرستون‌ها را استخراج کرده و ترتیب آن‌ها را بر اساس لیست ColumnList تنظیم می‌کند. ColumnList لیستی از سر ستون‌هاست به ترتیبی که کاربر تصمیم می‌گیرد باشند آن را در فایل ضمیمه می‌توانید ببینید تابع Table.ReorderColumns در پاور کوئری

تابع **Table.ReorderColumns** یک تابع قدرتمند در پاور کوئری است که به شما اجازه می‌دهد ترتیب ستون‌های یک جدول را تغییر دهید.

ویژگی‌های اصلی

تغییر ترتیب ستون‌ها: می‌توانید ترتیب دلخواه ستون‌ها را مشخص کنید

حذف ستون‌های اضافی: ستون‌هایی که در لیست جدید نباشند، حذف می‌شوند

برخورد با ستون‌های وجود نداشته: با پارامتر MissingField می‌توانید تعیین کنید چگونه با ستون‌های موجود برخورد شود

$\times$ $\checkmark$ f_x

```
= Table.TransformColumns(Filterd ,
    {"Data" , each Table.ReorderColumns(
        Table.PromoteHeaders(_ ) , ColumnList)
    } )
```

Id	Name.1	Kind	Data
1	ble001	Table001 (Page 1)	Table
2	ble002	Table002 (Page 2)	Table
3	ble003	Table003 (Page 3)	Table

Date	SalesRep	Customer	Amt	Profit	Region
04-Feb-05	Charley	India Trotters	13350	2760	East
05-Feb-05	Anshika	Data Tronics	12550	4640	East
08-Feb-05	James	Boston Consultants	12800	2940	East
09-Feb-05	Charley	MNTL	10100	2780	West
13-Feb-05	Mark	Data Tronics	13750	3160	South
14-Feb-05	Varsha	Jindle Power Works	12500	5200	West
16-Feb-05	Bruce	MNTL	13500	2080	West
18-Feb-05	Mark	Sharma & Co	13400	4240	South
20-Feb-05	Ramesh	Shah Associates	14250	2980	South
20-Feb-05	James	Jindle Power Works	13050	5220	East
24-Feb-05	Charley	Shyam & Sharma Co	12950	3740	West

Query Settings

PROPERTIES

Name

Pdfs

[All Properties](#)

APPLIED STEPS

- Source
- ConvertToPDF
- Expanded Content
- Filtered
- Prom&ReordCol**
- CombineData
- Changed Type

حالا روی فضای خالی هر سلول در ستون Data کلیک کنید تا داده‌های صحیح و یکدست را مشاهده کنید.

حالا لیستی از Table ها دارید که قرار است این ها را با تابع Table.Combine زیر هم بچینید شبیه آنچه که در تصویر می بینید در فرمول بار آن را اضافه کنید .

در این مرحله اگر Changed Type پاور کوئری باب میلтан نبود خودتان آنگونه که می خواهید Type ستونها را تغییر دهید

نکاتی در باره باز کردن فایل هایی موجود در یک پوشه

فایل های CSV و Text

برای باز کردن و ترکیب فایل های CSV یا Text می توانید از روشی مشابه مقاله قبلی استفاده کنید.

این فایل ها معمولاً فقط شامل داده ها هستند و پاراگراف بین داده ها وجود ندارد.

به جای تابع Pdf.Table، از تابع Csv.Document استفاده کنید.

فایل های اکسل و Access

برای فایل های اکسل در یک پوشه: از تابع Excel.Workbook

برای فایل های Access در یک پوشه: از تابع Access.Database استفاده کنید

این توابع باید همراه با Table.TransformColumns به کار روند

نکته مهم درباره فایل های Access

در هر فایل Access معمولاً چند جدول وجود دارد

پاور کوئری امکان انتخاب و ترکیب فقط یک جدول خاص از هر فایل را دارد