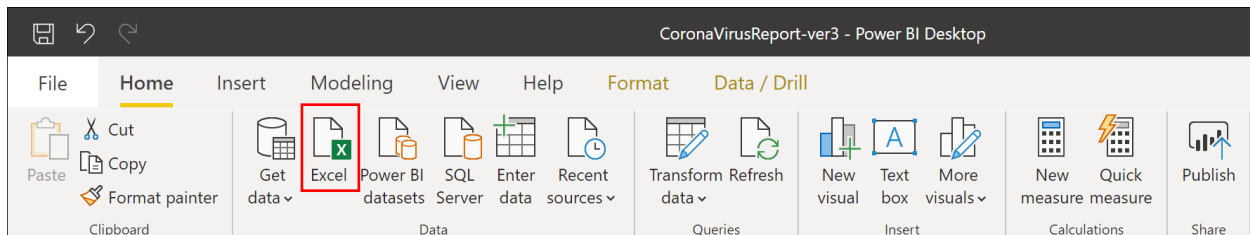


ETL

Hands on Labs

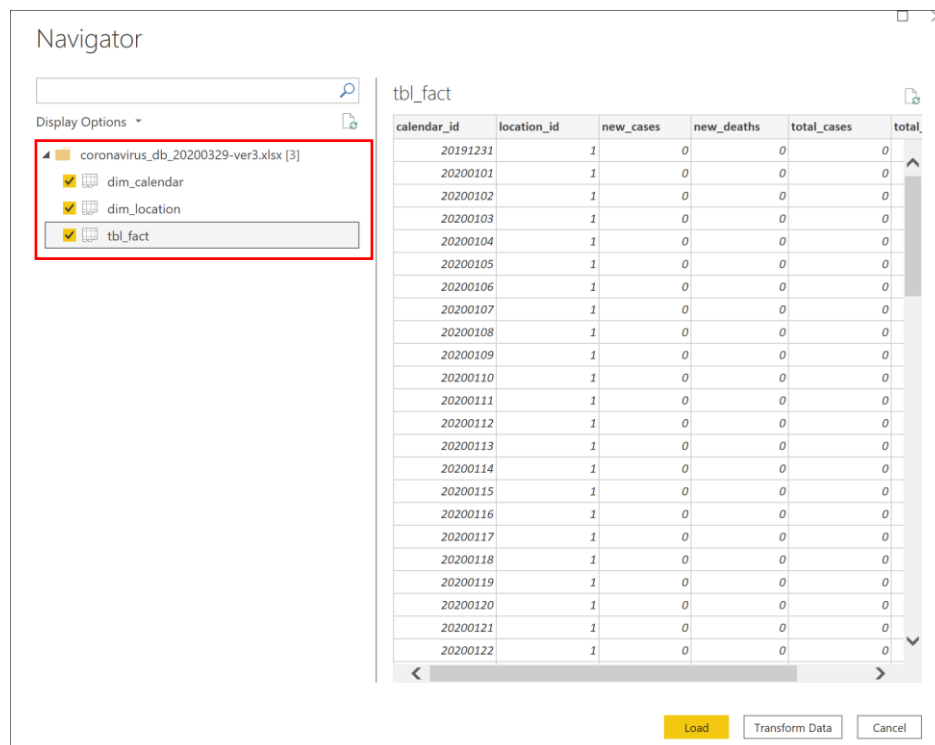
1. Data dari File Excel

Pada workshop ini format data yang digunakan adalah MS Excel. Jika telah mengerjakan tugas pada workshop sebelumnya. Maka bersihkan atau hapus Text box yang telah dibuat. Selanjutnya pada menu Home pilih tombol Excel pada group Data seperti yang terlihat pada Gambar 1.



Gambar 1. Memuat file Excel.

Pilih file coronavirus_db_20200329-ver3.xlsx yang telah diberikan. Selanjutnya akan ditampilkan window Navigator seperti pada Gambar 2.



Gambar 2. Window Navigator.

Secara otomatis window ini akan membaca dan menampilkan sheet yang memiliki data. Pada contoh di atas ditampilkan 3 sheet yaitu:

- a. dim_calendar.
- b. dim_location.
- c. tbl_fact.

Centang semua sheet tersebut kemudian klik tombol Load.

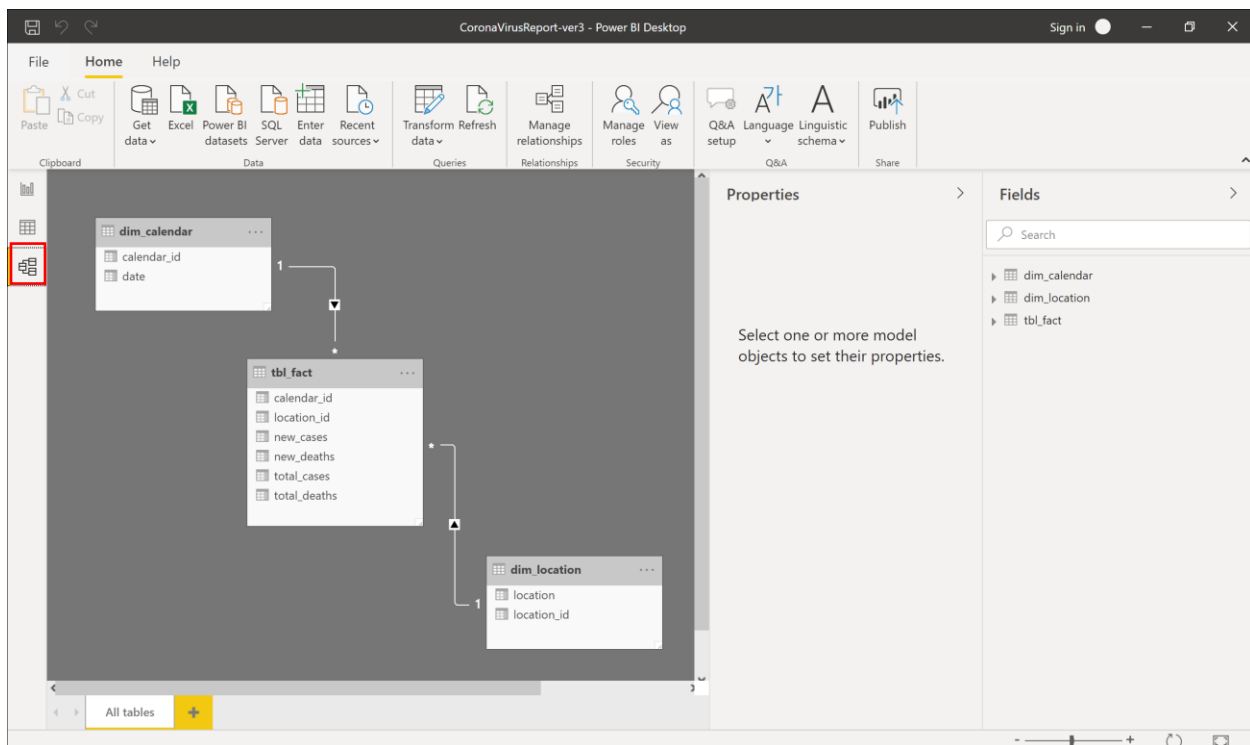
2. Side Bar

Power BI memiliki side bar yang berisi 3 tombol, yaitu:

- a. Report.
- b. Data.
- c. Model.

Model

Fitur ini menampilkan visualisasi struktur data. Fitur ini dapat diaktifkan dengan mengklik tombol pada kotak merah (Gambar 3). Jika struktur data memiliki relasi dimana pada tabel memiliki foreign key ke tabel yang lain maka secara otomatis Power BI akan menggambarkan relasi. Pada Gambar 3 diperlihatkan sheet pada file Excel dianggap sebagai tabel dan relasi antar tabel berdasarkan foreign key calendar_id dan location_id.

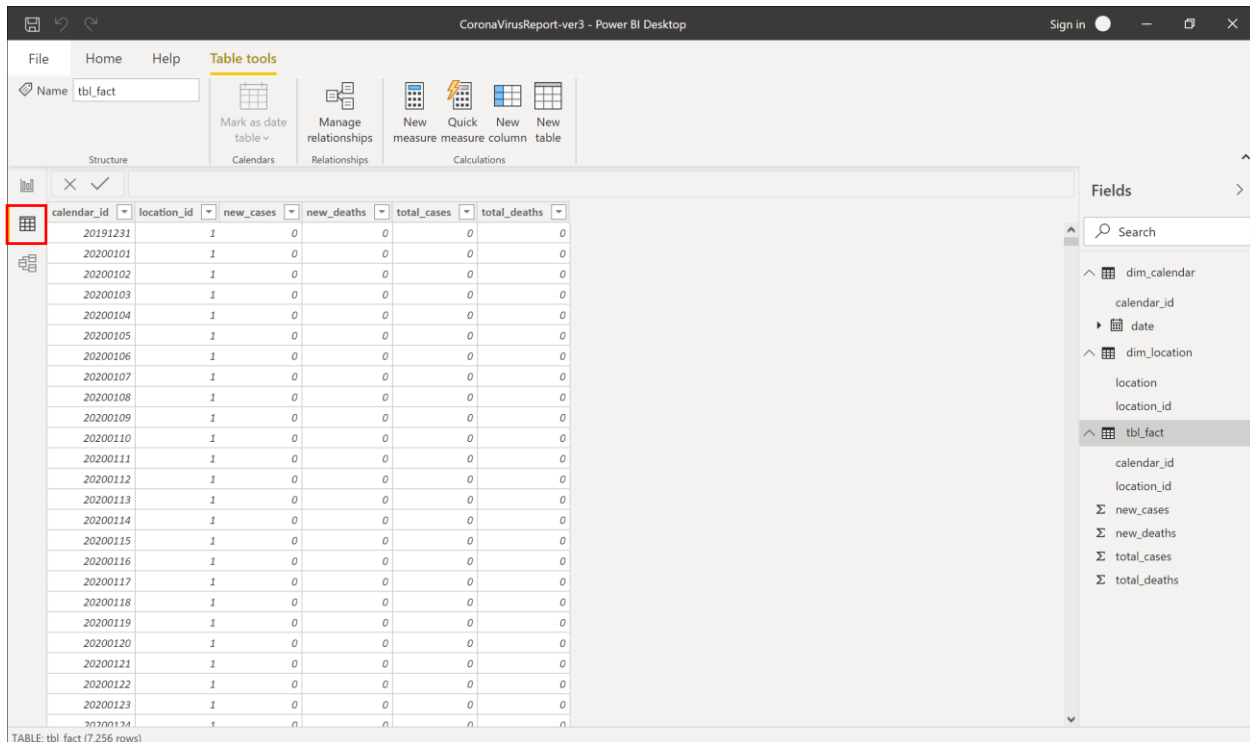


Gambar 3. Model.

Data

Fitur ini berfungsi untuk melihat data dari masing-masing tabel tersebut. Fitur ini dapat diaktifkan dengan mengklik tombol pada kotak merah pada side bar (Gambar 4). Pada sisi kanan terdapat area Fields yang

menampilkan daftar tabel beserta field yang dimilikinya. Jika memilih tbl_fact, maka data pada tabel ini ditampilkan.



CoronaVirusReport-ver3 - Power BI Desktop

File Home Help Table tools

Name: tbl_fact

Structure: Mark as date table, Manage relationships, New measure, Quick measure column, New table

calendar_id	location_id	new_cases	new_deaths	total_cases	total_deaths
20191231	1	0	0	0	0
20200101	1	0	0	0	0
20200102	1	0	0	0	0
20200103	1	0	0	0	0
20200104	1	0	0	0	0
20200105	1	0	0	0	0
20200106	1	0	0	0	0
20200107	1	0	0	0	0
20200108	1	0	0	0	0
20200109	1	0	0	0	0
20200110	1	0	0	0	0
20200111	1	0	0	0	0
20200112	1	0	0	0	0
20200113	1	0	0	0	0
20200114	1	0	0	0	0
20200115	1	0	0	0	0
20200116	1	0	0	0	0
20200117	1	0	0	0	0
20200118	1	0	0	0	0
20200119	1	0	0	0	0
20200120	1	0	0	0	0
20200121	1	0	0	0	0
20200122	1	0	0	0	0
20200123	1	0	0	0	0
20200124	1	0	0	0	0

TABLE: tbl_fact (7,256 rows)

Fields

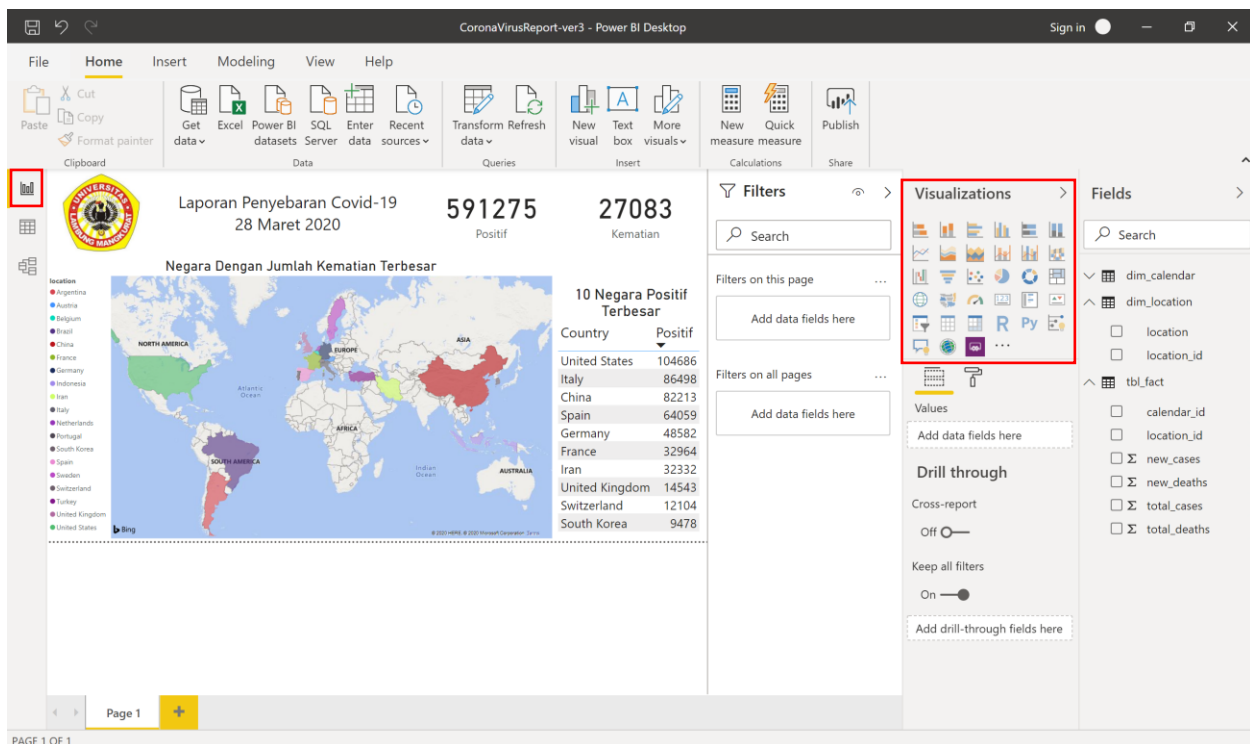
- dim_calendar
 - calendar_id
 - date
- dim_location
 - location
 - location_id
- tbl_fact
 - calendar_id
 - location_id
 - new_cases
 - new_deaths
 - total_cases
 - total_deaths

Gambar 4. Data.

Report

Setelah model dan data siap, maka selanjutnya adalah membuat report atau dashboard. Report dan dashboard ini adalah representasi informasi dari data yang dapat disampaikan dalam bentuk tabel atau visualisasi grafik.

Fitur ini dapat diaktifkan dengan mengklik tombol Report pada kota merah (Gambar 5). Pada halaman fitur ini, dapat dilihat tombol-tombol visualisasi yang dapat dipilih dan ditambahkan ke dalam report seperti yang terlihat pada kotak merah di area Visualizations.



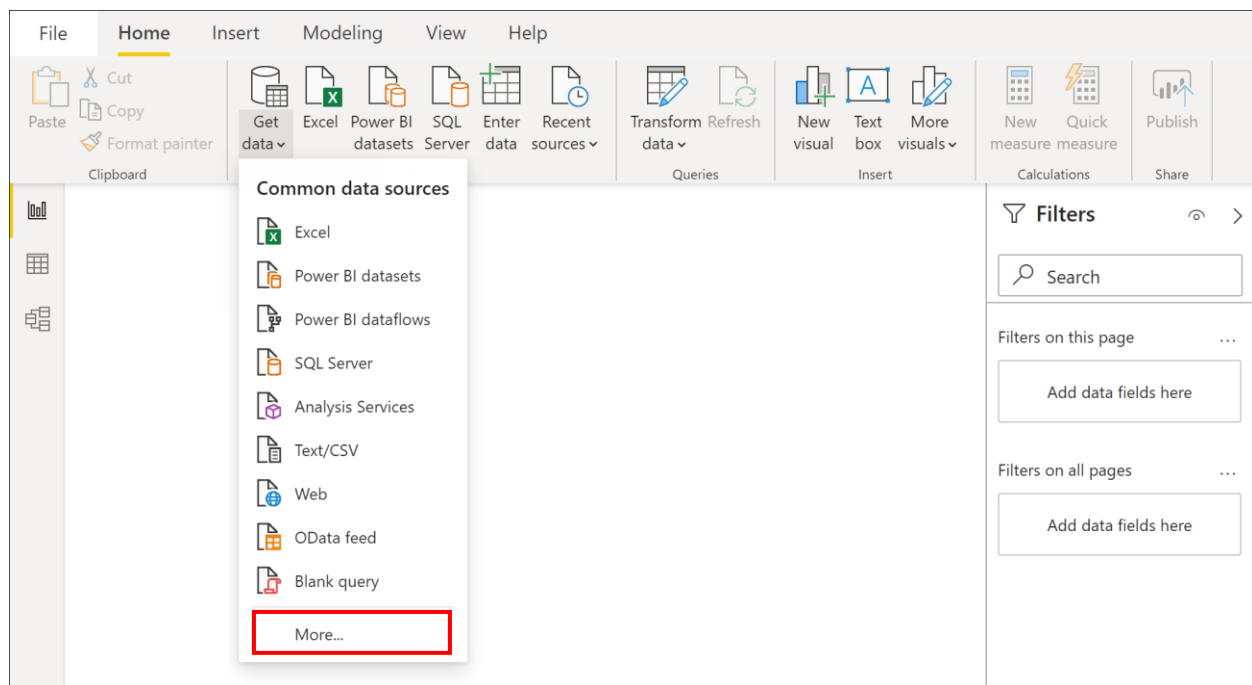
Gambar 5. Report.

3. Sumber Data

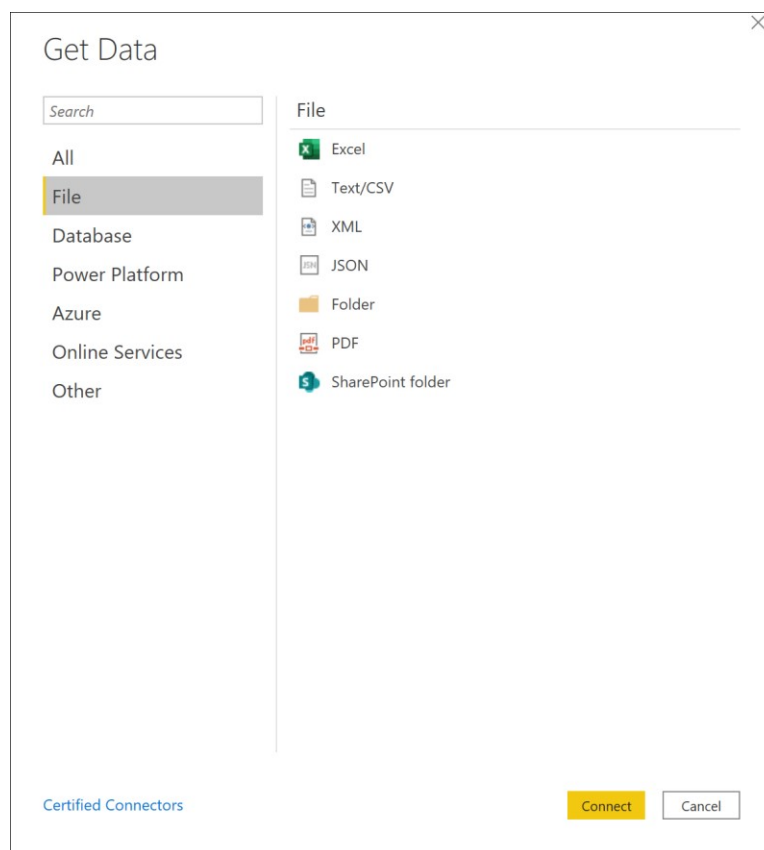
Pada paparan di atas diberikan contoh memuat data dari sumber berupa format MS Excel. Power BI memiliki kemampuan untuk melakukan proses extract dari berbagai sumber data. Untuk melihat sumber data yang didukung oleh Power BI dapat memilih menu Home > Get data kemudian pilih More... (Gambar 6).

Maka akan didapat dilihat daftar sumber data (Gambar 7) yang dikelompokkan sebagai:

- File.
- Database.
- Power Platform.
- Azure.
- Online Services.
- Other.

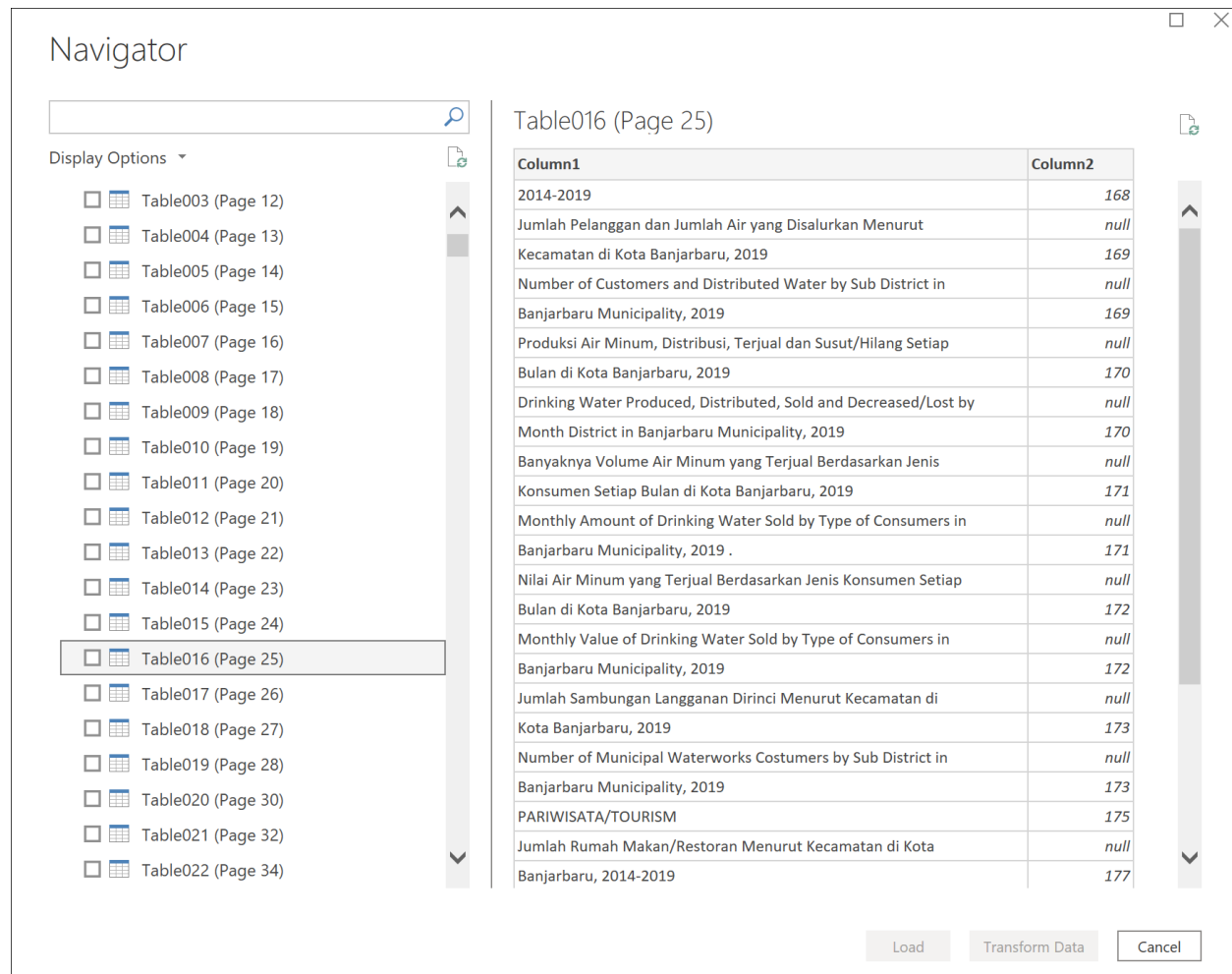


Gambar 6. Get more data.



Gambar 7. Daftar kelompok sumber data.

Misal dipilih File > PDF kemudian klik tombol Connect. Kemudian pilih file PDF yang telah disediakan pada workshop ini yaitu Kota Banjar Baru Dalam Angka 2020.pdf. Selanjutnya engine Power BI akan melakukan proses extract dengan mencari table yang ada pada dokumen ini. Hal ini karena umumnya data ditulis di dalam tabel. Setelah semua table di-extract maka ditampilkan daftar data pada window Navigator (Gambar 8).



Gambar 8. Daftar data dari table pada file PDF.

Klik salah satu table kemudian klik tombol Load jika ingin memuat data table tersebut untuk diolah pada Power BI.

Pada Gambar 7 dapat dilihat juga kemampuan untuk melakukan koneksi ke berbagai database seperti:

- SQL Server database
- MySQL database.
- Oracle database.
- PostgreSQL database.
- Sybase databse.
- Dan lain-lain.

Hal ini membuktikan jika kita memiliki akses ke database berupa user dan password maka proses extract ke database tersebut dapat dilakukan. Hal ini menunjukkan proses extract pada database yang berbeda-beda dapat langsung dilakukan tanpa susah payah membuat perantara seperti API dengan format Rest API.

4. Transform

Pada paparan ini dan sebelumnya sebenarnya telah dilakukan proses extract, transform dan load. Namun yang terlihat seperti hanya melakukan proses extract dengan memilih sumber data kemudian mengklik tombol load. Hal ini dikarenakan data yang didapat dari hasil extract sudah sesuai dengan keinginan. Namun jika data belum sesuai dengan yang diinginkan maka dapat dilakukan proses transform. Berikut langkah-langkah singkat untuk melakukan transform.

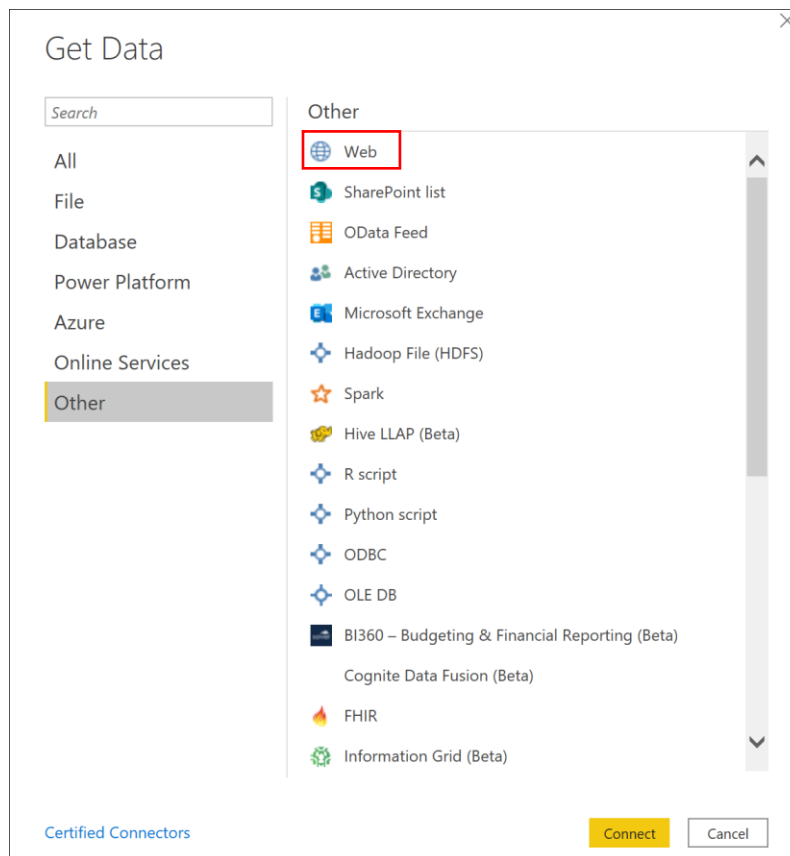
Sumber data yang digunakan adalah data pada web dengan link berikut ini <https://www.worldometers.info/coronavirus/#countries>. Pada contoh di bagian 3. Sumber Data diketahui bahwa Power BI dapat meng-extract table di dalam file PDF. Artinya dapat juga meng-extract table pada halaman web. Berikut adalah sebagai dari isi halaman web dari link di atas. Pada Gambar 9 terlihat dua table yang dapat diakses dengan mengklik tombol Now dan Yesterday.

Report coronavirus cases

Now		Yesterday		Search:						
Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Reported 1 st case
World	885,344	+27,025	44,214	+1,912	185,215	655,915	33,778	113.6	5.7	Jan 10
USA	189,754	+1,224	4,099	+46	7,274	178,381	4,576	573	12	Jan 20
Italy	105,792		12,428		15,729	77,635	4,023	1,750	206	Jan 29
Spain	102,136	+6,213	9,053	+589	22,647	70,436	5,872	2,185	194	Jan 30
China	81,554	+36	3,312	+7	76,238	2,004	466	57	2	Jan 10
Germany	74,508	+2,700	821	+46	16,100	57,587	2,675	889	10	Jan 26
France	52,128		3,523		9,444	39,161	5,565	799	54	Jan 23
Iran	47,593	+2,988	3,036	+138	15,473	29,084	3,871	567	36	Feb 18
UK	29,474	+4,324	2,352	+563	135	26,987	163	434	35	Jan 30
Switzerland	17,137	+532	461	+28	2,967	13,709	348	1,980	53	Feb 24
Belgium	13,964	+1,189	828	+123	2,132	11,004	1,088	1,205	71	Feb 03
Netherlands	13,614	+1,019	1,173	+134	250	12,191	1,053	795	68	Feb 26

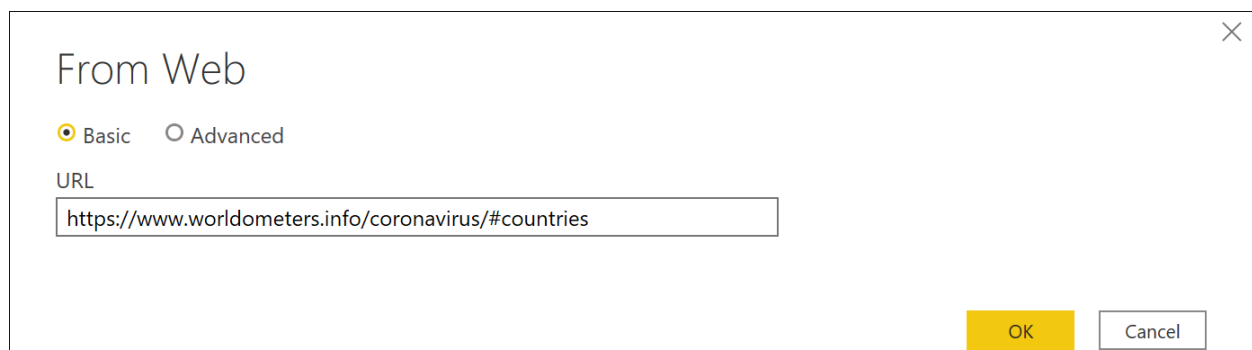
Gambar 9. Halaman web <https://www.worldometers.info/coronavirus/#countries>.

Langkah pertama adalah pilih menu Home > Get data > More ..., kemudian ditampilkan window Get Data seperti pada Gambar 10.



Gambar 10. Window Get Data > Web.

Pilih Other kemudian Web setelah itu klik tombol Connect. Kemudian pada window From Web (Gambar 11) ketikkan alamat berikut ini <https://www.worldometers.info/coronavirus/#countries>.



Gambar 11. Window From Web.

Kemudian klik tombol OK. Pastikan komputer telah tersambung dengan internet. Selanjutnya dapat dilihat window Navigator (Gambar 12).

The screenshot shows the Power BI Navigator window. On the left, the 'Display Options' section shows a list of tables: 'Document', 'Table 0' (selected with a checkmark), and 'Table 1'. The main area displays 'Table 0' with a preview of data downloaded on Wednesday. The table has five columns: 'Country, Other', 'Total Cases', 'New Cases', 'Total Deaths', and 'New Deaths'. The data lists various countries and the world total, with values for cases and deaths. At the bottom, there are buttons for 'Add table using examples', 'Load', 'Transform Data', and 'Cancel'.

Country, Other	Total Cases	New Cases	Total Deaths	New Deaths
World	854307	69569	42016	
USA	185270	21482	3780	
Italy	105792	4053	12428	
Spain	95923	7967	8464	
China	81518	null	3305	
Germany	71808	4923	775	
France	52128	7578	3523	
Iran	44605	3110	2898	
UK	25150	3009	1789	
Switzerland	16605	683	433	
Turkey	13531	2704	214	
Belgium	12775	876	705	
Netherlands	12595	845	1039	
Austria	10180	562	128	
S. Korea	9786	125	162	
Canada	8505	1057	101	
Portugal	7443	1035	160	
Brazil	5717	1087	201	
Israel	5358	663	20	
Australia	4711	251	20	

Gambar 12. Window Navigator data web.

Pada daftar data yang ada di atas terdapat:

- Table 0.
- Table 1.

Jika dilihat pada web di alamat <https://www.worldometers.info/coronavirus/#countries> maka dapat dilihat terdapat 2 table seperti pada Gambar 9 maka diketahui bahwa Table 0 adalah data table Now dan Table 1 adalah data table Yesterday. Jika yang ingin diambil hanya data Now saja maka centang Table 0 saja kemudian klik tombol Load.

Kemudian klik tombol Data (kotak merah pada bagian kiri pada Gambar 13) pada side bar untuk melihat data yang berhasil kita extract dan load.

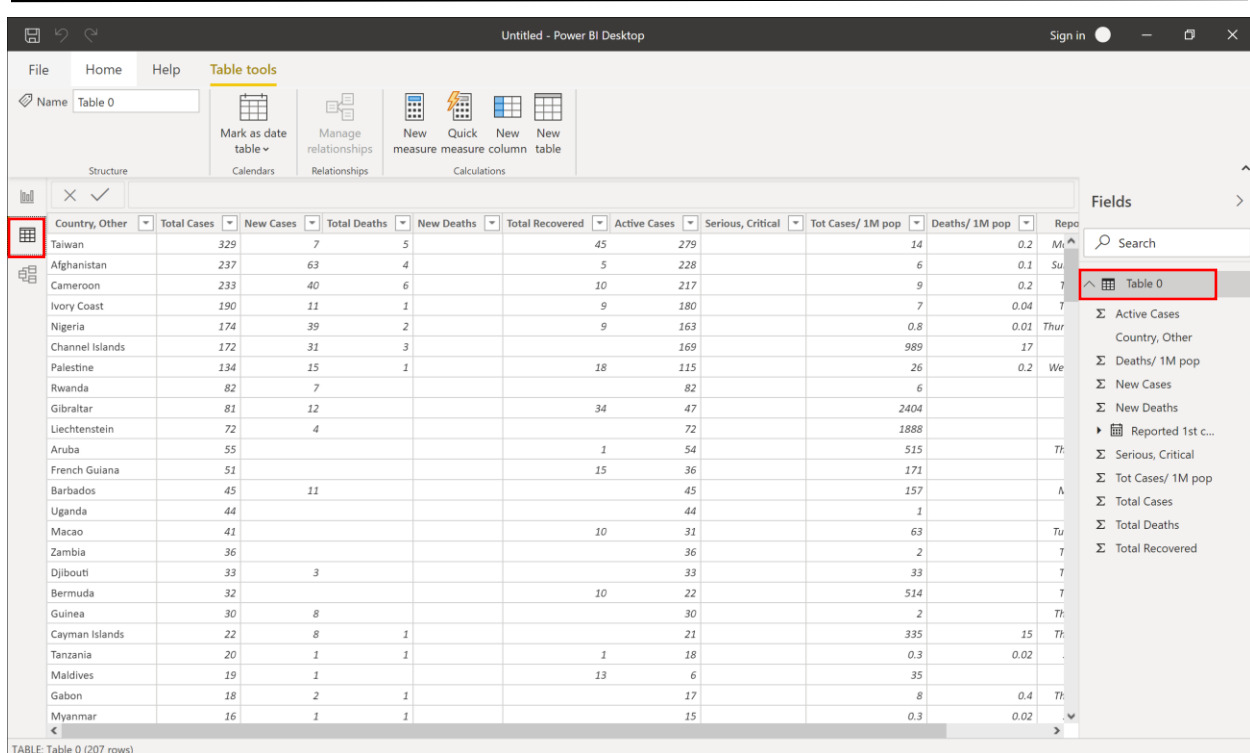


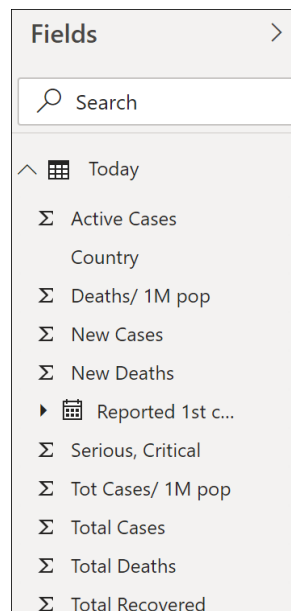
Table 0

Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Reported 1st c...
Taiwan	329	7	5		45	279		14	0.2	M...
Afghanistan	237	63	4		5	228		6	0.1	Su...
Cameroon	233	40	6		10	217		9	0.2	T...
Ivory Coast	190	11	1		9	180		7	0.04	T...
Nigeria	174	39	2		9	163		0.8	0.01	Thur...
Channel Islands	172	31	3			169		989	17	We...
Palestine	134	15	1		18	115		26	0.2	We...
Rwanda	82	7				82		6		
Gibraltar	81	12			34	47		2404		
Liechtenstein	72	4				72		1888		
Aruba	55				1	54		515		Th...
French Guiana	51				15	36		171		
Barbados	45	11				45		157		A...
Uganda	44					44		1		
Macao	41				10	31		63		Tu...
Zambia	36					36		2		T...
Djibouti	33	3				33		33		T...
Bermuda	32				10	22		514		T...
Guinea	30	8				30		2		Th...
Cayman Islands	22	8	1			21		335	15	Th...
Tanzania	20	1	1		1	18		0.3	0.02	
Maldives	19	1			13	6		35		Th...
Gabon	18	2	1			17		8	0.4	Th...
Myanmar	16	1	1			15		0.3	0.02	

TABLE: Table 0 (207 rows)

Gambar 13. Data table Today.

Table 0 dapat diberi nama baru dengan cara klik kanan pada Table 0 yang ada pada bagian Fields (kotak merah pada bagian kanan gambar) kemudian pilih Rename pada context menu. Ganti nama Table 0 menjadi Today. Kemudian field "Country, Other" diganti menjadi Country. Maka hasilnya dapat dilihat nama table dan field sebagai berikut.



Fields

Search

Today

- Active Cases
- Country
- Deaths/ 1M pop
- New Cases
- New Deaths
- Reported 1st c...
- Serious, Critical
- Tot Cases/ 1M pop
- Total Cases
- Total Deaths
- Total Recovered

Gambar 14. Table Today.

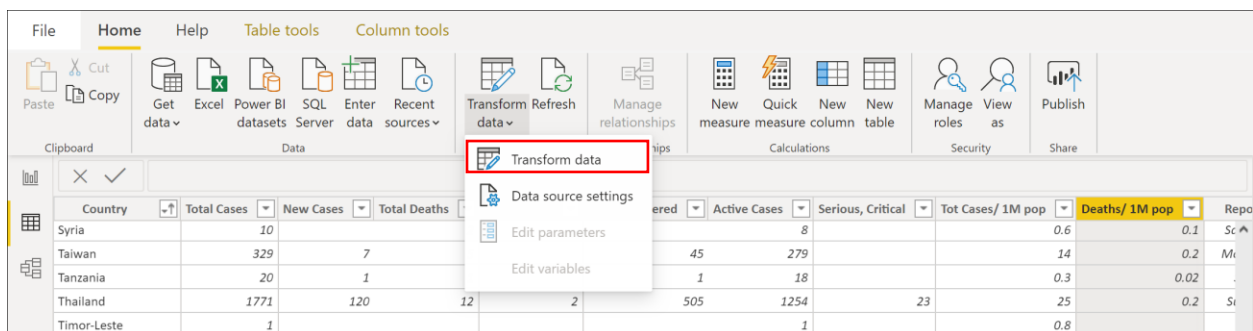
Selanjutnya perhatikan data (Gambar 15) terdapat dua record yang tidak diperlukan yaitu Total: dan World karena keduanya dapat dicari dengan menghitung jumlah nilai yang dimiliki seluruh negara.

Country	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Repo
Syria	10		2			8		0.6	0.1	Sc
Taiwan	329	7	5		45	279		14	0.2	Mi
Tanzania	20	1	1		1	18		0.3	0.02	
Thailand	1771	120	12	2	505	1254	23	25	0.2	Si
Timor-Leste	1					1		0.8		
Togo	36	2	2	1	10	24		4	0.2	7
Total:	935190	76835	47192	4883	193989	694009	35478	120	6.1	
Trinidad and Tobago	90	3	5	2	1	84		64	4	Wed
Tunisia	423	29	12	2	5	406	10	36	1	
Turkey	15679	2148	277	63	333	15069	847	186	3	
Turks and Caicos	6	1				6		155		
UAE	814	150	8	2	61	745	2	82	0.8	Tu
Uganda	44					44		1		
UK	29474	4324	2352	563	135	26987	163	434	35	Thu
Ukraine	794	149	20	3	13	761		18	0.5	
Uruguay	350	12	2	1	62	286	15	101	0.6	Th
USA	215003	26473	5102	1049	8878	201023	5005	650	15	Mi
Uzbekistan	181	9	2		12	167	8	5	0.06	Sc
Vatican City	6					6		7491		7
Venezuela	144	9	3		43	98	6	5	0.1	Th
Vietnam	218	6			63	155	3	2		Wed
World	935190	76835	47192	4883	193989	694009	35478	120	6.1	
Zambia	36					36		2		7
Zimbabwe	8		1			7		0.5	0.07	Th

Gambar 15. Data Today.

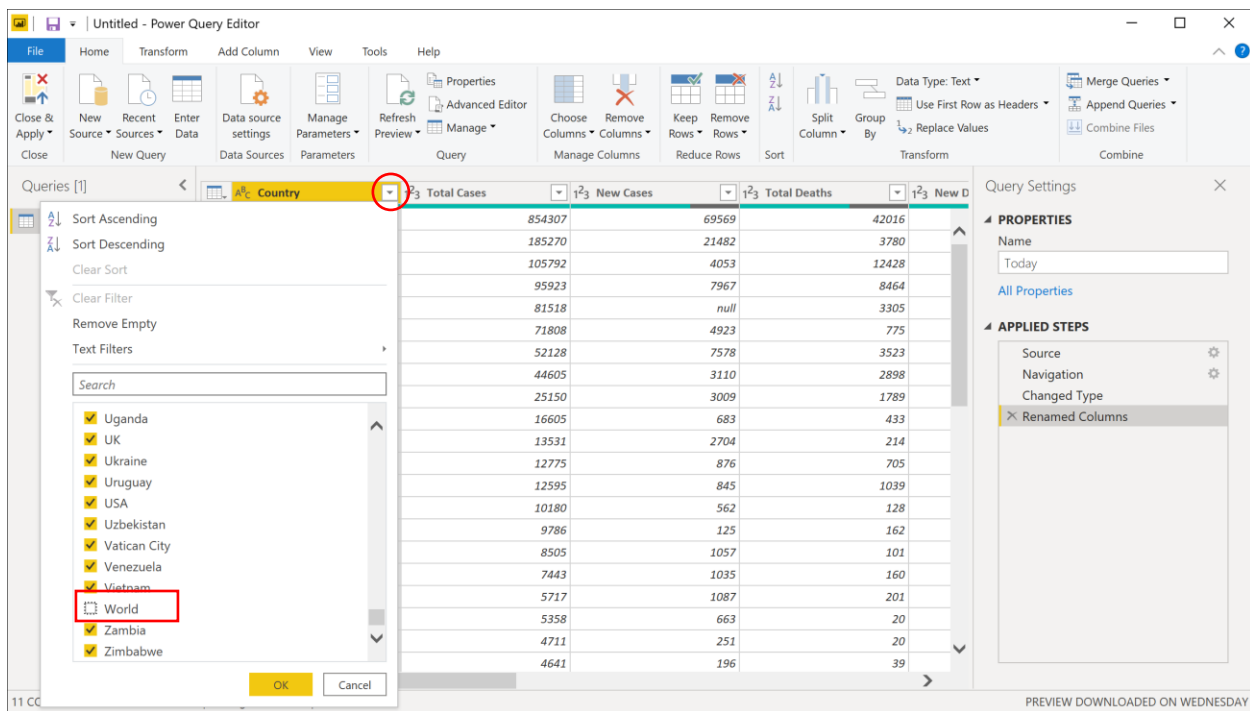
Selain itu terdapat nilai-nilai yang kosong atau missing value. Oleh karena itu perlu dilakukan proses transform data yang meliputi penghapusan data yang tidak diperlukan, pengantian nilai dan hal-hal yang diperlukan.

Untuk melakukan transform klik menu Home > Transform data



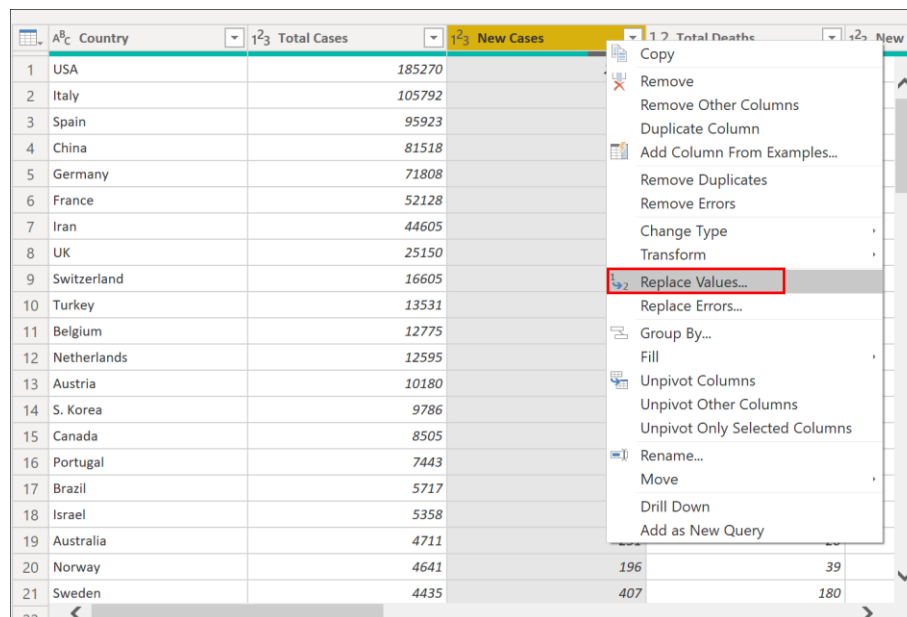
Gambar 16. Transform data.

Selanjutnya akan ditampilkan data pada window Power Query Editor (Gambar 17). Pada kolom Country klik tanda panah ke bawah (lingkaran merah), maka akan ditampilkan context menu kemudian hilangkan centang "Total:" dan "World" kemudian klik tombol OK.



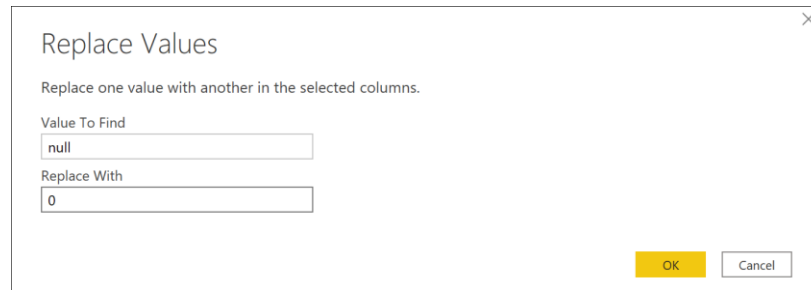
Gambar 17. Power Query Editor.

Kemudian pada beberapa kolom dapat dilihat ada nilai kosong (*null*), nilai ini dapat diganti dengan 0. Caranya adalah dengan memilih kolom dengan cara klik pada header (misal kolom New Cases). Kemudian klik kanan pada header kolom dan pilih Replace Values (Gambar 18).



Gambar 18. Context menu Replace Values.

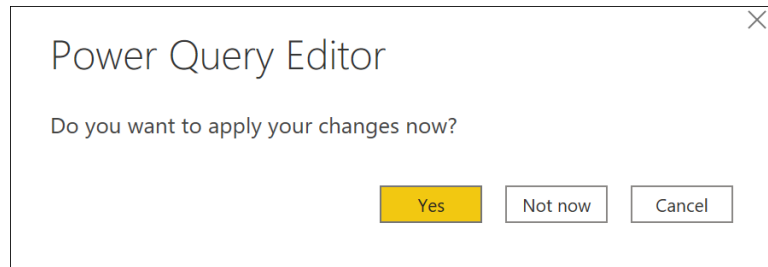
Kemudian akan ditampilkan window Replace Values (Gambar 19) masukkan null pada kolom input Value To Find, dan nilai 0 pada kolom input Replace With. Kemudian klik tombol OK.



Gambar 19. Window replace values.

Pada window ini tidak ada tombol Undo, tetapi aksi yang dilakukan dapat dilihat pada area APPLIED STEPS pada sisi sebelah kanan window Power Query Editor. Jika ingin membatalkan suatu aksi, maka tinggal hapus saja aksi tersebut dari daftar APPLIED STEPS.

Setelah proses transform selesai, klik tombol [x] pada ujung kanan atas untuk menutup window Power Query Editor ini. Kemudian ditampilkan window konfirmasi (Gambar 20), klik tombol Yes.



Gambar 20. Window konfirmasi.

Sekarang perhatikan bahwa record “Total:” dan “World” telah hilang dari data.

Karena data diambil dari halaman web di internet, maka pastikan saat melakukan hal ini komputer terkoneksi dengan internet.

Kesimpulan

Dari paparan di atas dapat diambil beberapa kesimpulan:

- Power BI dapat melakukan proses Extract-Transform-Load, tetapi Power BI bukan Data Warehouse karena data tidak disimpan ke suatu media penyimpanan seperti halnya Data Warehouse sebenarnya.
- Pada paparan dan contoh di atas, dapat dilihat bahwa proses extract dapat dilakukan langsung dengan mengakses database atau data tanpa harus membuat media perantara seperti Rest API. Hal dapat menjadi ide solusi untuk membuat program yang dapat mengakses data dari bermacam-macam database untuk melakukan proses extract, kemudian dilanjutkan dengan proses transform dan di-load ke data warehouse.

Tugas

- a. Lakukan langkah-langkah yang diberikan pada bagian **1. Data dari File Excel**. Kemudian simpan pekerjaan dengan menu File > Save. Simpan ke dalam file excel_NIM.pbix. Kumpulkan file pada <https://elearning.ulm.ac.id/>.
- b. Lakukan langkah-langkah yang diberikan pada bagian **4. Transform**, kemudian simpan ke dalam file web_NIM.pbix. kumpulkan file pada <https://elearning.ulm.ac.id/>.
- c. Download script MySQL pada link berikut ini <https://github.com/dalers/mywind>. Buat database dan table-tablenya dengan menggunakan northwind.sql dan northwind-data.sql. Kemudian gunakan Power BI untuk mengakses database tersebut dan table-tablenya. Tuliskan:
 - Langkah-langkah membuat database dan table pada MySQL dengan uraikan kalimat yang singkat dan gambar screenshot jika diperlukan.
 - Langkah-langkah Power Bi untuk melakukan koneksi ke database MySQL dengan uraikan kalimat singkat dan gambar screenshot (harus ada).
 - Format file adalah PDF, dan dikumpulkan pada <https://elearning.ulm.ac.id/>.