

Data Cleaning menggunakan SQL

Introduction

Dalam project ini saya melakukan data cleaning untuk raw data housing di Amerika dan mentransformasi data tersebut dengan Microsoft SQL Server agar lebih mudah untuk dianalisis.

Step 1. Load data ke SQL Server

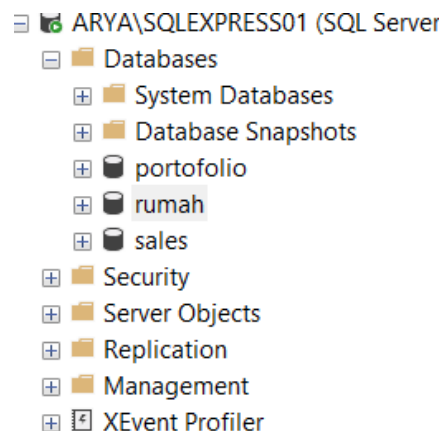
UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant	OwnerName
2045	007 00 0 125.00	SINGLE FAMILY	1808 FOX CHASE DR, GOODLETTSVILLE	April 9, 2013	240000	20130412-0036474	No	FRAZIER, CYRENTHA LYNETTE
16918	007 00 0 130.00	SINGLE FAMILY	1832 FOX CHASE DR, GOODLETTSVILLE	June 10, 2014	366000	20140619-0053768	No	BONER, CHARLES & LESLIE
54582	007 00 0 138.00	SINGLE FAMILY	1864 FOX CHASE DR, GOODLETTSVILLE	September 26, 2016	435000	20160927-0101718	No	WILSON, JAMES E. & JOANNE
43070	007 00 0 143.00	SINGLE FAMILY	1853 FOX CHASE DR, GOODLETTSVILLE	January 29, 2016	255000	20160129-0008913	No	BAKER, JAY K. & SUSAN E.
22714	007 00 0 149.00	SINGLE FAMILY	1829 FOX CHASE DR, GOODLETTSVILLE	October 10, 2014	278000	20141015-0095255	No	POST, CHRISTOPHER M. & SAMANTHA C
18367	007 00 0 151.00	SINGLE FAMILY	1821 FOX CHASE DR, GOODLETTSVILLE	July 16, 2014	267000	20140718-0063802	No	FIELDS, KAREN L. & BRENT A.
19804	007 14 0 002.00	SINGLE FAMILY	2005 SADIÉ LN, GOODLETTSVILLE	August 28, 2014	171000	20140903-0080214	No	HINTON, MICHAEL R. & CYNTHIA M. MC
54583	007 14 0 024.00	SINGLE FAMILY	1917 GRACELAND DR, GOODLETTSVILLE	September 27, 2016	262000	20161005-0105441	No	BAILOR, DARRELL & TAMMY
36500	007 14 0 026.00	SINGLE FAMILY	1428 SPRINGFIELD HWY, GOODLETTSVILLE	August 14, 2015	285000	20150819-0083440	No	ROBERTS, MISTY L. & ROBERT M.
19805	007 14 0 034.00	SINGLE FAMILY	1420 SPRINGFIELD HWY, GOODLETTSVILLE	August 29, 2014	340000	20140909-0082348	No	LEE, JEFFREY & NANCY
29467	007 14 0A 024.00	SINGLE FAMILY	2209 KAYLA DR, GOODLETTSVILLE	April 14, 2015	425000	20150415-0033442	No	
10754	007 14 0A 027.00	SINGLE FAMILY	109 BAILEY VIEW CT, GOODLETTSVILLE	December 12, 2013	585000	20131227-0130352	No	
34751	007 14 0B 010.00	RESIDENTIAL CONDO	1900 TININN RD, GOODLETTSVILLE	July 13, 2015	190000	20150717-0069947	No	
4512	007 15 0 002.00	SINGLE FAMILY	629 GAYLEMORE DR, GOODLETTSVILLE	June 7, 2013	189900	20130612-0059715	No	URRUTIA, CARLOS MIGUEL & REBECCA
16919	007 15 0 003.00	SINGLE FAMILY	633 GAYLEMORE DR, GOODLETTSVILLE	June 30, 2014	157500	20140702-0058050	No	SALDANA, ALMA J. & EDWARD A., II
16920	007 15 0 004.00	SINGLE FAMILY	637 GAYLEMORE DR, GOODLETTSVILLE	June 30, 2014	247400	20140630-0057267	No	MCKINNEY, ROBERT K. & BETH E.
51967	007 15 0 008.00	SINGLE FAMILY	1976 SADIÉ LN, GOODLETTSVILLE	July 15, 2016	211500	20160720-0074793	No	MILLER, JAMES L. & EILEEN T.
28155	007 15 0 014.00	SINGLE FAMILY	644 GAYLEMORE DR, GOODLETTSVILLE	March 31, 2015	185900	20150402-0029022	No	SANDAGE, LEAH S.
8899	007 15 0 020.00	SINGLE FAMILY	1921 NORMERLE DR, GOODLETTSVILLE	October 11, 2013	349900	20131018-0109102	No	VAUGHN, JOHN
4513	007 15 0 031.00	SINGLE FAMILY	1916 NORMERLE DR, GOODLETTSVILLE	June 28, 2013	192500	20130711-0071698	No	PARKS, KRISTEN LEIGH & WILLIAMS, PAI
27161	007 15 0 044.00	SINGLE FAMILY	2050 GRACELAND DR, GOODLETTSVILLE	February 10, 2015	279900	20150212-0012993	No	ANGELL, MARGO G. & CHARLES D.
46859	007 15 0 048.00	SINGLE FAMILY	2034 GRACELAND DR, GOODLETTSVILLE	April 14, 2016	379900	20160418-0036715	No	DOSS, GLEN KEVIN & ANGELA DAWN
5802	007 15 0 052.00	SINGLE FAMILY	811 BENTON CT, GOODLETTSVILLE	July 8, 2013	192500	20130712-0072376	No	BRAKE, MICHAEL E. & ERIN L.
36501	010 00 0 045.00	SINGLE FAMILY	331 VIEW RIDGE DR, GOODLETTSVILLE	August 31, 2015	193500	20150903-0090036	No	BENJAMIN, DAVID AND LINDA REVOCAL
8900	010 00 0 052.00	SINGLE FAMILY	361 VIEW RIDGE DR, GOODLETTSVILLE	October 21, 2013	172400	20131030-0112723	No	WILLIAMS, JOSEPH D. & AUDREY
48716	011 15 0A 004.00	SINGLE FAMILY	119 LAKESIDE DR, GOODLETTSVILLE	May 31, 2016	435000	20160602-0055827	No	

data di atas merupakan raw data housing yang masih dalam bentuk excel, masih kotor, dan berantakan.

a. Buat database pada SQL Server untuk meletakkan data raw tersebut ke dalamnya

```
SQLQuery2.sql - AR...S01.rumah (sa (58)) *
create database rumah
use rumah

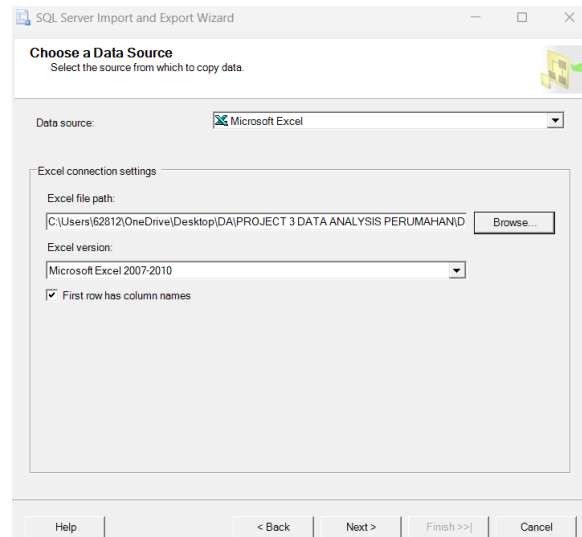
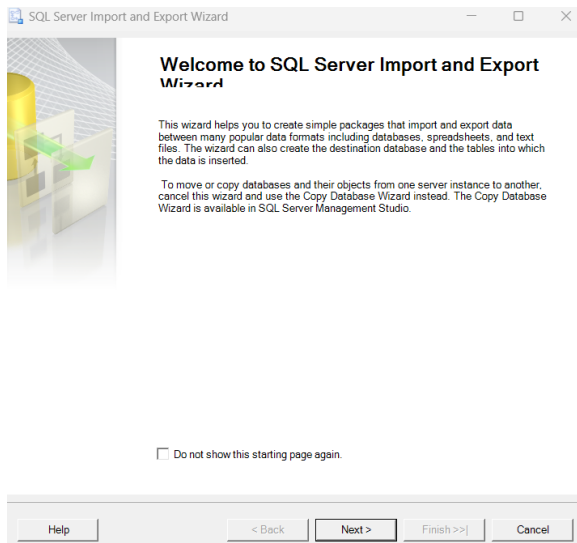
Messages
Commands completed successfully.
Completion time: 2023-08-10T11:08:59.8576250+07:00
```



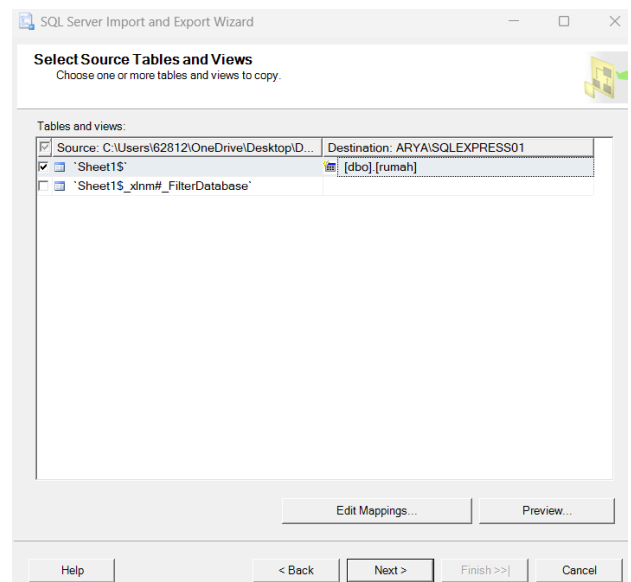
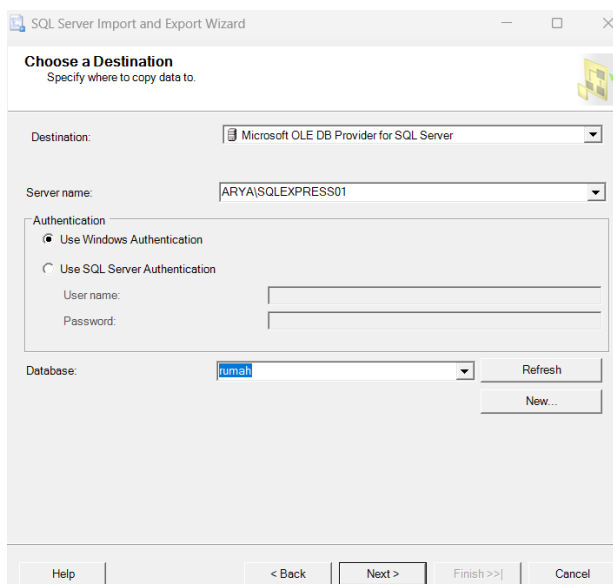
b. Import raw data housing ke dalam database

Bisa dilakukan dengan klik kanan database -> tasks -> import data..

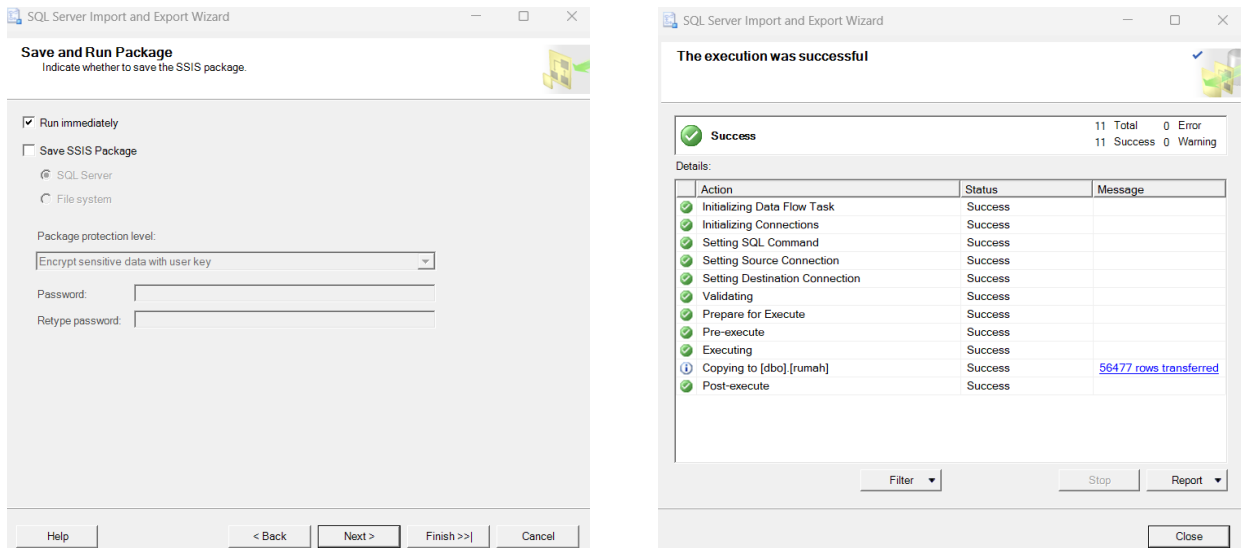
Maka akan muncul menu, kemudian klik next -> pilih data source yaitu 'Ms Excel' dan file excelnya -> klik next



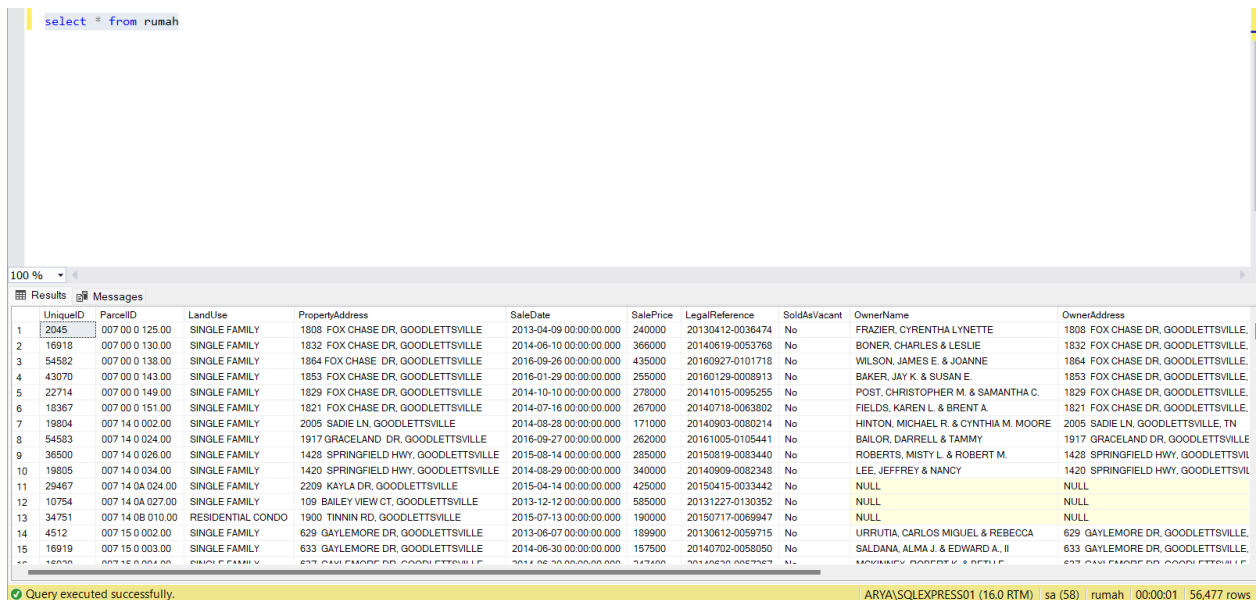
Kemudian pilih destinasinya yaitu Microsoft OLE DB Provider for SQL Server dan Databasenya sesuai yang sudah kita buat yaitu 'rumah' -> klik next dan kita rename nama tablenya yaitu 'rumah' -> klik next



dan klik finish, maka akan muncul tampilan seperti berikut apabila data berhasil di import ke dalam database



kemudian kita cek apakah data sudah berhasil masuk ke database, dan ternyata berhasil.



Setelah data raw berhasil masuk ke dalam database, baru kita dapat melakukan cleaning data tersebut dengan tools SQL Server

Step 2. Cleaning Data Raw

a. Standarisasi format tanggal

SaleDate
2013-04-09 00:00:00.000
2014-06-10 00:00:00.000
2016-09-26 00:00:00.000
2016-01-29 00:00:00.000
2014-10-10 00:00:00.000
2014-07-16 00:00:00.000
2014-08-28 00:00:00.000
2016-09-27 00:00:00.000
2015-08-14 00:00:00.000
2014-08-29 00:00:00.000
2015-04-14 00:00:00.000
2013-12-12 00:00:00.000
2015-07-13 00:00:00.000
2013-06-07 00:00:00.000

Terlihat format tanggal tidak standar karna masih dalam format datetime, maka kita harus ubah ke dalam format date dengan perintah convert.

```
select saledate, convert(date, saledate) as tanggal
from rumah
```

100 %

Results Messages

	saledate	tanggal
1	2013-04-09 00:00:00.000	2013-04-09
2	2014-06-10 00:00:00.000	2014-06-10
3	2016-09-26 00:00:00.000	2016-09-26
4	2016-01-29 00:00:00.000	2016-01-29
5	2014-10-10 00:00:00.000	2014-10-10
6	2014-07-16 00:00:00.000	2014-07-16
7	2014-08-28 00:00:00.000	2014-08-28

setelah berhasil, kita buat colom baru untuk memasukkan saledate yang telah diconvert, kemudian kita lakukan update data tersebut.

```
alter table rumah
add saledatebaru date

update rumah
set saledatebaru = convert(date, saledate)
```

kemudian kita cek dengan melakukan perintah select, dan format tanggal berhasil berubah

```
select saledatebaru
from rumah
```

100 %

Results Messages

	saledatebaru
1	2013-04-09
2	2014-06-10
3	2016-09-26
4	2016-01-29
5	2014-10-10

b. Mengisi data kolom propertyaddress yang kosong (NULL)

```
select * from rumah
where propertyaddress is null
```

100 %

Results Messages

	UniquelD	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant
1	43076	025 07 0 031.00	SINGLE FAMILY	NULL	2016-01-15 00:00:00.000	179900	20160120-0005776	No
2	39432	026 01 0 069.00	VACANT RESIDENTIAL LAND	NULL	2015-10-23 00:00:00.000	153000	20151028-0109602	No
3	45290	026 05 0 017.00	SINGLE FAMILY	NULL	2016-03-29 00:00:00.000	155000	20160330-0029941	No
4	53147	026 06 0A 038.00	RESIDENTIAL CONDO	NULL	2016-08-25 00:00:00.000	144900	20160831-0091567	No
5	43080	033 06 0 041.00	SINGLE FAMILY	NULL	2016-01-04 00:00:00.000	170000	20160107-0001526	No
6	45295	033 06 0A 002.00	SINGLE FAMILY	NULL	2016-03-29 00:00:00.000	210000	20160331-0030709	No
7	48731	033 15 0 123.00	SINGLE FAMILY	NULL	2016-05-05 00:00:00.000	199900	20160506-0045368	No

Kita bisa lihat bahwa kolom propertyaddress banyak yang NULL, namun jika diperhatikan kembali, apabila kita urutkan berdasarkan parcelID, jika terdapat parcelID yang sama, maka propertyaddressnya pun juga sama, itu artinya kita bisa mencari propertyaddress yang bernilai NULL dengan melihat parcelIDnya, karena parcelID tidak akan kosong

```
select * from rumah
--where propertyaddress is null
order by ParcelID
```

100 %

Results Messages

	UniquelD	ParcelID	LandUse	PropertyAddress
85	34756	018 07 0 109.00	SINGLE FAMILY	105 BRYAN HOUSE DR, GOODLETTSVILLE
86	22716	018 07 0 142.00	SINGLE FAMILY	301 MYSTIC HILL DR, GOODLETTSVILLE
87	50584	018 07 0 142.00	SINGLE FAMILY	301 MYSTIC HILL DR, GOODLETTSVILLE

Untuk mencari propertyaddress berdasarkan parcelID, kita dapat melakukan join dengan table itu sendiri

```

select a.ParcelID, a.PropertyAddress, b.ParcelID, b.PropertyAddress, isnull(a.propertyaddress, b.PropertyAddress) as propertyaddressnew
from rumah a
join rumah b
on a.ParcelID = b.ParcelID
and a.[UniqueID ] <> b.[UniqueID ]
where a.propertyaddress is null

```

100 %

Results Messages

	ParcelID	PropertyAddress	ParcelID	PropertyAddress	propertyaddressnew
1	025 07 0 031.00	NULL	025 07 0 031.00	410 ROSEHILL CT, GOODLETTSVILLE	410 ROSEHILL CT, GOODLETTSVILLE
2	026 01 0 069.00	NULL	026 01 0 069.00	141 TWO MILE PIKE, GOODLETTSVILLE	141 TWO MILE PIKE, GOODLETTSVILLE
3	026 05 0 017.00	NULL	026 05 0 017.00	208 EAST AVE, GOODLETTSVILLE	208 EAST AVE, GOODLETTSVILLE
4	026 06 0A 038.00	NULL	026 06 0A 038.00	109 CANTON CT, GOODLETTSVILLE	109 CANTON CT, GOODLETTSVILLE
5	033 06 0 041.00	NULL	033 06 0 041.00	1129 CAMPBELL RD, GOODLETTSVILLE	1129 CAMPBELL RD, GOODLETTSVILLE
6	033 06 0A 002.00	NULL	033 06 0A 002.00	1116 CAMPBELL RD, GOODLETTSVILLE	1116 CAMPBELL RD, GOODLETTSVILLE

kemudian kita update kolom propertyaddress

```

update a
set PropertyAddress = isnull(a.propertyaddress, b.PropertyAddress)
from rumah a
join rumah b
on a.ParcelID = b.ParcelID
and a.[UniqueID ] <> b.[UniqueID ]
where a.propertyaddress is null

```

100 %

Messages

(0 rows affected)

Completion time: 2023-08-10T12:25:46.7987632+07:00

dan null value berhasil hilang dari kolom propertyaddress

```

select a.ParcelID, a.PropertyAddress, b.ParcelID, b.PropertyAddress, isnull(a.propertyaddress, b.PropertyAddress)
from rumah a
join rumah b
on a.ParcelID = b.ParcelID
and a.[UniqueID ] <> b.[UniqueID ]
where a.propertyaddress is null

```

100 %

Results Messages

ParcelID	PropertyAddress	ParcelID	PropertyAddress	propertyaddressnew
----------	-----------------	----------	-----------------	--------------------

c. memecah alamat menjadi individual column (address, city, state)

```
select PropertyAddress from rumah
```

	PropertyAddress
1	1808 FOX CHASE DR, GOODLETTSVILLE
2	1832 FOX CHASE DR, GOODLETTSVILLE
3	1864 FOX CHASE DR, GOODLETTSVILLE
4	1853 FOX CHASE DR, GOODLETTSVILLE
5	1829 FOX CHASE DR, GOODLETTSVILLE

kita dapat memecah menjadi address dan city dengan perintah SQL substring dan charIndex, hasilnya seperti ini

```
select
    substring(propertyaddress, 1, CHARINDEX(',',propertyaddress)-1) as address
    , substring(propertyaddress, CHARINDEX(',',propertyaddress)+1, len(propertyaddress)) as city
from rumah
```

	address	city
1	1808 FOX CHASE DR	GOODLETTSVILLE
2	1832 FOX CHASE DR	GOODLETTSVILLE
3	1864 FOX CHASE DR	GOODLETTSVILLE

kemudian kita buat kolom baru untuk mengisi address dan city yang telah kita split

```
alter table rumah
add propertysplitaddress varchar(255)

alter table rumah
add propertysplitcity varchar(255)
```

kemudian kita update table untuk mengisi kolom tersebut

```
update rumah
set propertysplitaddress = substring(propertyaddress, 1, CHARINDEX(',',propertyaddress)-1)

update rumah
set propertysplitcity = substring(propertyaddress, CHARINDEX(',',propertyaddress)+1, len(propertyaddress))
```

(56477 rows affected)

(56477 rows affected)

Completion time: 2023-08-10T12:37:06.1599920+07:00

dan hasilnya seperti ini

propertysplitaddress	propertysplitcity
1808 FOX CHASE DR	GOODLETTSVILLE
1832 FOX CHASE DR	GOODLETTSVILLE
1864 FOX CHASE DR	GOODLETTSVILLE
1853 FOX CHASE DR	GOODLETTSVILLE
1829 FOX CHASE DR	GOODLETTSVILLE
1821 FOX CHASE DR	GOODLETTSVILLE
2005 SADIE LN	GOODLETTSVILLE
1917 GRACELAND DR	GOODLETTSVILLE
1428 SPRINGFIELD HWY	GOODLETTSVILLE
1420 SPRINGFIELD HWY	GOODLETTSVILLE
2209 KAYLA DR	GOODLETTSVILLE
109 BAILEY VIEW CT	GOODLETTSVILLE

d. Memecah Alamat menjadi individual kolom dengan cara yang beda

Kolom owneraddress ternyata juga masih menyatu dengan city dan state, jadi kita perlu memisahkan menjadi 3 komponen yakni address, city, dan state

```
select owneraddress from rumah
```

owneraddress
1808 FOX CHASE DR, GOODLETTSVILLE, TN
1832 FOX CHASE DR, GOODLETTSVILLE, TN
1864 FOX CHASE DR, GOODLETTSVILLE, TN
1853 FOX CHASE DR, GOODLETTSVILLE, TN

Agar lebih praktis, kita gunakan cara dengan perintah SQL parselname dan replace, Kemudian kita tinggal membuat kolom baru untuk meletakkan data yang telah dipecah tersebut

```
select  
    parselname(replace(owneraddress, ',', '.'),3) as address  
    , parselname(replace(owneraddress, ',', '.'),2) as city  
    , parselname(replace(owneraddress, ',', '.'),1) as state  
from rumah
```

address	city	state
1808 FOX CHASE DR	GOODLETTSVILLE	TN
1832 FOX CHASE DR	GOODLETTSVILLE	TN
1864 FOX CHASE DR	GOODLETTSVILLE	TN
1853 FOX CHASE DR	GOODLETTSVILLE	TN
1829 FOX CHASE DR	GOODLETTSVILLE	TN

```
alter table rumah  
    add ownersplitaddress varchar(255)  
  
alter table rumah  
    add ownersplitcity varchar(255)  
  
alter table rumah  
    add ownersplitstate varchar(255)
```


Dan kita lakukan update seperti biasa

```
update rumah
set ownersplitaddress = parsename(replace(owneraddress, ',', '.'), 3)

update rumah
set ownersplitcity = parsename(replace(owneraddress, ',', '.'), 2)

update rumah
set ownersplitstate = parsename(replace(owneraddress, ',', '.'), 1)
```

Hasilnya seperti ini

ownersplitaddress	ownersplitcity	ownersplitstate
1808 FOX CHASE DR	GOODLETTSVILLE	TN
1832 FOX CHASE DR	GOODLETTSVILLE	TN
1864 FOX CHASE DR	GOODLETTSVILLE	TN
1853 FOX CHASE DR	GOODLETTSVILLE	TN
1829 FOX CHASE DR	GOODLETTSVILLE	TN
1821 FOX CHASE DR	GOODLETTSVILLE	TN

e. Mengubah Y and N to Yes and No in "Sold as Vacant" Field.

Kita bisa lihat bawa pada kolom sold as vacant data di dalamnya tidak sama

```
select distinct(soldasvacant), count(soldasvacant)
from rumah
group by SoldAsVacant
order by 2
```

.00 %

Results Messages

	soldasvacant	(No column name)
1	Y	52
2	N	399
3	Yes	4623
4	No	51403

kita akan mentransformasi agar data yang di dalamnya semua sama yakni menggunakan Yes dan No saja.

Kita akan menggunakan perintah case, saat terdapat data 'Y' maka akan diubah menjadi Yes dan 'N' menjadi No

```
select soldasvacant, case when soldasvacant = 'Y' then 'Yes'
                        when soldasvacant = 'N' then 'No'
                        else soldasvacant
                        end
from rumah
```

100 %

Results Messages

	soldasvacant	(No column name)
1	No	No
2	No	No
3	No	No
4	No	No

Kemudian kita lakukan update

```
update rumah
set SoldAsVacant = case when soldasvacant = 'Y' then 'Yes'
                        when soldasvacant = 'N' then 'No'
                        else soldasvacant
                        end
from rumah
```

Dan hasilnya akan seperti ini, semua data akan menjadi Yes dan No

```
select distinct(soldasvacant), count(soldasvacant)
from rumah
group by SoldAsVacant
order by 2
```

100 %

Results Messages

	soldasvacant	(No column name)
1	Yes	4675
2	No	51802

f. Menghilangkan data duplikat

data duplikat sudah biasa di dalam sebuah table, maka kita harus bisa menghilangkannya agar tidak menjadi bias.

Kita dapat melakukan identifikasi data duplikat dengan perintah SQL `row_number`, perintah ini akan menghitung data yang duplikat yang akan muncul di kolom `row_num`, apabila lebih dari 1, artinya data tersebut memiliki duplikat.

```
select *, row_number() over (
  partition by parcelid,
    propertyaddress,
    saleprice,
    saledate,
    legalreference
  order by
    uniqueid) as row_num
from rumah
```

Kemudian untuk mencari data yang duplikat, kita hanya perlu melakukan perintah dengan filter `row_num > 1`, hasil yang muncul adalah data duplikat

```
with rownumCTE AS(
  select *, row_number() over (
    partition by parcelid,
      propertyaddress,
      saleprice,
      saledate,
      legalreference
    order by
      uniqueid) as row_num
  from rumah
  --order by parcelid
)
select * from rownumCTE
where row_num > 1
order by propertyaddress
```

Results Messages

Acreage	TaxDistrict	LandValue	BuildingValue	TotalValue	YearBuilt	Bedrooms	FullBath	HalfBath	saledatebaru	propertysplitaddress	propertysplitcity	ownersplitaddress	ownersplitcity	ownersplitstate	row_num
0.17	URBAN SERVICES DISTRICT	20000	41600	61600	1940	3	2	0	2015-02-23	1011 43RD AVE N	NASHVILLE	1011 43RD AVE N	NASHVILLE	TN	2
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2015-02-06	1023 CAPITAL FUNDS CT	NASHVILLE	NULL	NULL	NULL	2
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2015-02-06	1025 CAPITAL FUNDS CT	NASHVILLE	NULL	NULL	NULL	2

Dalam query ini kita menggunakan CTE (common table expression) karena filter `row_num > 1` tidak akan bisa dijalankan apabila hanya di dalam select biasa karena `row_num` berada di dalam fungsi `row_number`, sehingga perlu di jalankan melalui CTE.

Kemudian, kita perlu menghapus semua data yang muncul karena itu semua adalah data duplikat

Caranya mudah, kita hanya menggunakan perintah delete

```
with rownumCTE AS(  
    select *, row_number() over (  
        partition by parcelid,  
            propertyaddress,  
            saleprice,  
            saledate,  
            legalreference  
        order by  
            uniqueid) as row_num  
    from rumah  
    --order by parcelid  
)  
delete from rownumCTE  
where row_num > 1  
--order by propertyaddress
```

.00 %

Messages

(104 rows affected)

Completion time: 2023-08-10T13:22:44.9935030+07:00

Dan hasilnya data duplikat telah hilang

```
with rownumCTE AS(  
    select *, row_number() over (  
        partition by parcelid,  
            propertyaddress,  
            saleprice,  
            saledate,  
            legalreference  
        order by  
            uniqueid) as row_num  
    from rumah  
    --order by parcelid  
)  
select * from rownumCTE  
where row_num > 1  
--order by propertyaddress
```

10 %

Results Messages

UniqueID	ParcelID	LandUse	PropertyAddress	SaleDate	SalePrice	LegalReference	SoldAsVacant	OwnerName	OwnerAddress	Acreage	TaxDistrict	LandValue	BuildingValue	TotalValue	YearBuilt	Bedrooms	FullBath	HalfBath	saledate
----------	----------	---------	-----------------	----------	-----------	----------------	--------------	-----------	--------------	---------	-------------	-----------	---------------	------------	-----------	----------	----------	----------	----------

g. Menghapus kolom yang tidak terpakai

setelah kita melakukan transformasi di atas, kita banyak melakukan penambahan kolom yang berakibat tidak terpakainya kolom yang lama, sehingga kita perlu menghapus semua kolom-kolom yang tidak terpakai tersebut

```
alter table rumah
drop column owneraddress, taxdistrict, propertyaddress, saledate
```

100 %

Messages

Commands completed successfully.

Completion time: 2023-08-10T13:29:45.3360402+07:00

akhirnya kita telah selesai melakukan cleaning data pada raw data housing ini, dengan hasil akhir seperti ini

```
select *
from rumah
```

100 %

Results Messages

	UniqueID	ParcelID	LandUse	SalePrice	LegalReference	SoldAs/Vacant	OwnerName	Acreage	LandValue	BuildingValue	TotalValue	YearBuilt	Bedrooms	FullBath	HalfBath	saledatebaru
1	2045	007 00 0 125.00	SINGLE FAMILY	240000	20130412-0036474	No	FRAZIER, CYRENTHA LYNETTE	2.3	50000	168200	235700	1986	3	3	0	2013-04-09
2	16918	007 00 0 130.00	SINGLE FAMILY	366000	20140619-0053768	No	BONER, CHARLES & LESLIE	3.5	50000	264100	319000	1998	3	3	2	2014-06-10
3	54582	007 00 0 138.00	SINGLE FAMILY	435000	20160927-0101718	No	WILSON, JAMES E. & JOANNE	2.9	50000	216200	298000	1987	4	3	0	2016-09-26
4	43070	007 00 0 143.00	SINGLE FAMILY	255000	20160129-0008913	No	BAKER, JAY K. & SUSAN E.	2.6	50000	147300	197300	1985	3	3	0	2016-01-29
5	22714	007 00 0 149.00	SINGLE FAMILY	278000	20141015-0095255	No	POST, CHRISTOPHER M. & SAMANTHA C.	2	50000	152300	202300	1984	4	3	0	2014-10-10

terlihat data sudah bersih dan siap untuk dilakukan analisis lebih lanjut