

--cleaning data in SQL

```
Select *  
from Housing
```

-----  
----

```
--Standardize Date Format  
Select SaleDate, CONVERT(date, SaleDate)  
from Housing
```

```
update housing  
set SaleDate = CONVERT(date, SaleDate)
```

-----  
-----

```
--Populate Property Address Data  
--By looking at the data we determine that if parcel id is same then property  
address is also the same i.e for parcel id 1 property adress is ABC
```

```
Select PropertyAddress  
from Housing  
ORDER BY ParcelID
```

```
select a.ParcelID, a.PropertyAddress, b.ParcelID, b.PropertyAddress, ISNULL  
      (a.PropertyAddress, b.PropertyAddress)  
from Housing as a  
JOIN Housing as b  
on a.ParcelID=b.ParcelID  
AND a.UniqueID <> b.UniqueID --make row pairs where parcel id is same by unique id  
is different  
WHERE a.PropertyAddress is NULL
```

```
update a  
SET PropertyAddress=ISNULL(a.PropertyAddress, b.PropertyAddress)  
from Housing as a  
JOIN Housing as b  
on a.ParcelID=b.ParcelID  
AND a.UniqueID <> b.UniqueID --make row pairs where parcel id is same by unique id  
is different  
WHERE a.PropertyAddress is NULL
```

-----  
-----

```
--Breaking out address into individual columns(address,city,state)
```

```
--for property address
select
SUBSTRING(PropertyAddress,1,CHARINDEX(',',PropertyAddress)-1),
SUBSTRING(PropertyAddress,CHARINDEX(',',PropertyAddress)+1,len(propertyaddress))
from Housing
```

```
alter table housing
add propertysplitAddress Nvarchar(255)
```

```
alter table housing
add propertysplitCity Nvarchar(255)
```

```
update Housing
set propertysplitAddress = SUBSTRING(PropertyAddress,1,CHARINDEX
(',',PropertyAddress)-1)
```



```
update Housing
set propertysplitCity =SUBSTRING(PropertyAddress,CHARINDEX(',',PropertyAddress)
+1,len(propertyaddress))
```



```
--for owner address
select OwnerAddress
from Housing
```

```
select
PARSENAME(REPLACE(ownerAddress,',','.'),1),
PARSENAME(REPLACE(ownerAddress,',','.'),2),
PARSENAME(REPLACE(ownerAddress,',','.'),3)
from Housing
```

```
alter table housing
add ownersplitAddress Nvarchar(255)
```

```
alter table housing
add ownersplitCity Nvarchar(255)
```

```
alter table housing
add ownersplitstate Nvarchar(255)
```

```
update Housing
set ownersplitAddress =PARSENAME(REPLACE(ownerAddress,',','.'),3)
```

```
update Housing
set ownersplitcity=PARSENAME(REPLACE(ownerAddress,',','.'),2)
```

```
update Housing
set ownersplitstate=PARSENAME(REPLACE(ownerAddress,',','.'),1)
```

----- ↗  
-----  
--change y to yes and n to no in sold as vacant field

```
Select Distinct(SoldAsVacant), Count(SoldAsVacant)
From Housing
Group by SoldAsVacant
order by 2
```

```
Select SoldAsVacant
, CASE When SoldAsVacant = 'Y' THEN 'Yes'
      When SoldAsVacant = 'N' THEN 'No'
      ELSE SoldAsVacant
      END
From Housing
```

```
Update Housing
SET SoldAsVacant = CASE When SoldAsVacant = 'Y' THEN 'Yes'
                      When SoldAsVacant = 'N' THEN 'No'
                      ELSE SoldAsVacant
                      END
```

----- ↗  
-----  
--delete unused columns  
Select \*  
From Housing

```
ALTER TABLE Housing
DROP COLUMN OwnerAddress, TaxDistrict, PropertyAddress, SaleDate
```