

Data Cleaning Data in SQL

Skills used : CREATE, UPDATE, SELECT, CTE, JOINS

Skills used : CREATE, UPDATE, SELECT, CTE, JOINS, ORDER BY, GROUP BY

Select * from Nashville housing

*/*Standardize Date Format*/*

```
select `Sale Date`  
from nashville_housing  
  
select `Sale Date`, convert(`Sale Date`, date)  
from nashville_housing  
  
update nashville_housing  
set `Sale Date` = convert(`Sale Date`, date)
```

*/*Populate Property Address Data*/*

```
select *  
from nashville_housing  
where `Property Address` is null  
order by `Parcel ID`  
  
select a.`Parcel ID`, b.`Property Address`, b.`Parcel ID`,  
b.`Property Address`,  
ifnull(a.`Property Address`, b.`Property Address`) as `Address To Be Filled`  
from nashville_housing a
```

```

join nashville_housing b

on a.`Parcel ID` = b.`Parcel ID`

and a.Column1 != b.Column1

where a.`Property Address` is null

update a

set a.`Property Address` = ifnull(a.`Property Address`, b.`Property Address`)

from nashville_housing a

join nashville_housing b

on a.`Parcel ID` = b.`Parcel ID`

and a.Column1 != b.Column1

where a.`Property Address` is null

```

```

/*Breaking out Address into Individual Columns (Address, City, State)*/

```

```

select `Property Address`

from nashville_housing

select

substring(`Property Address` , 1, locate(',', `Property Address`)) - 1 as Address ,

substring(`Property Address` , locate(',', `Property Address`) + 1 ,

length(`Property Address`)) as Address,

from nashville_housing

alter table nashville_housing

set `Property Split Address` varchar(255);

update nashville_housing

```

```

set `Property Split Address` = substring(`Property Address` , 1,
locate(',', `Property Address`)) - 1

alter table nashville_housing

set `Property City` varchar(255);

update nashville_housing

set `Property City` = substring(`Property Address` , locate(',', `Property Address`) + 1 ,
length(`Property Address`))

select `Owner Address`

from nashville_housing

```

```

/*CREATING A SPLIT STRING FUNCTION TO SPLIT THE OWNER ADDRESS*/

```

```

create function SPLIT_STR(
x varchar(255),
delim varchar(12),
pos int
)
return varchar(255)

return replace(substring(substring_index(x, delim, pos),
length(substring_index(x, delim, pos - 1)) + 1),
delim, "");

select

SPLIT_STR(`Owner Address`, ',', 1),

SPLIT_STR(`Owner Address`, ',', 2),

```

```
SPLIT_STR(`Owner Address`,',', 3)

from nashville_housing

alter table nashville_housing

set `Address` varchar(255);

update nashville_housing

set `Address` = SPLIT_STR(`Owner Address`,',', 1)

alter table nashville_housing

set `City` varchar(255);

update nashville_housing

set `City` = SPLIT_STR(`Owner Address`,',', 2)

alter table nashville_housing

set `State` varchar(255);

update nashville_housing

set `State` = SPLIT_STR(`Owner Address`,',', 3)
```

```
/*Change Y and N to Yes and No in `Sold as Vacant` */

select distinct (`Sold As Vacant`), count(`Sold As Vacant`)

from nashville_housing

group by `Sold As Vacant`

order by `Sold As Vacant`

select `Sold As Vacant` ,

case

when `Sold As Vacant` = 'Y' then 'Yes'
```

```
when `Sold As Vacant` = 'N' then 'No'

end as `Sold As Vacant`

from nashville_housing

update nashville_housing

set `Sold As Vacant` = case

when `Sold As Vacant` = 'Y' then 'Yes'

when `Sold As Vacant` = 'N' then 'No'

end as `Sold As Vacant`
```

```
/*Remove Duplicate*/
```

```
with RowNumCTE
```

```
as
```

```
(
```

```
select *,
```

```
row_number() OVER(
```

```
partition by
```

```
`Parcel ID`,
```

```
`Property Address`,
```

```
`Sale Price`,
```

```
`Sale Date`,
```

```
`Legal Reference`
```

```
order by
```

```
Column1
```

```
) row_num
```

```
from nashville_housing
```

```
)
```

```
delete
```

```
from RowNumCTE
```

```
where row_num > 1;
```

```
/*Delete Unused Column*/
```

```
alter table nashville_housing
```

```
drop table `Owner Address`, `Tax District`, `Property Address`, `Sale Date`;
```