Nashville Housing Data SQL Queries

/\* Identify all null values in the owneraddress column. \*/

select \* from nashville\_housing\_data where owneraddress is null

/\* Because the owneraddress and propertyaddress columns have the same values,

update that null owneraddress's to the corresponding propertyaddress. \*/

update nashville\_housing\_data

set owneraddress = propertyaddress

where owneraddress is null

/\* There are some owneraddress fields that didnt get updated because the property address is null.

Reverse the query to fill out corresponding owner addresses. \*/

select \* from nashville\_housing\_data where propertyaddress is null

/\* Update that null propertyaddress fields to the corresponding owneraddress field. \*/

update nashville\_housing\_data

set propertyaddress = owneraddress

where propertyaddress is null

/\* 11 rows still empty because both property an owner address is null. Going to see if there are duplicates.

Wanted to make sure there are no duplicates of unique ids for the remaining 11 values that are still null.

No duplicates found. \*/

select

uniqueid,

count(\*) as count\_of\_duplicates

from

nashville\_housing\_data

group by

uniqueid

having

coumt(\*) > 1;

/\* Identify all unique landuse field. \*/

select

distinct landuse

from

nashville\_housing\_data

/\*1. Change the data type for the saleprice fiel. \*/ 4. self join for duplicates 5. group together\*/

alter table nashville\_housing\_data

alter column saleprice type bigint

using saleprice::bigint

/\* Find all distinct values for sold as vacant column. --4 different values yes, y, n, no \*/

select

distinct soldasvacant

from

nashville\_housing\_data

/\* Update both "Y" and "N" to reflect "Yes" and "No" to match the other fields. \*/

select

soldasvacant,

case

when soldasvacant = 'Y' then 'Yes'

when soldasvacant = 'N' then 'No'

else soldasvacant

end as soldasvacant

from

nashville\_housing\_data;

update nashville\_housing\_data

set soldasvacant =

case

when soldasvacant = 'Y' then 'Yes'

when soldasvacant = 'N' then 'No'

else soldasvacant

end

/\* Identify duplicates on parcel id. --7,225/56,477 are duplicates on parcelid. \*/

select

nh1.parcelid

, count(\*) AS duplicate\_count

from

nashville\_housing\_data nh1

join

nashville\_housing\_data nh2 on nh1.parcelid = nh2.parcelid

and nh1.UniqueID <> nh2.UniqueID

group by

nh1.parcelid

having

count(\*) > 1;

/\* Get rid of columns that are not needed. \*/

alter table nashville\_housing\_data

drop column owneraddress

alter table nashville\_housing\_data

drop column taxdistrict

alter table nashville\_housing\_data

drop column saledate