

NASHVILLE HOUSING DATA (DATA CLEANING)

SQL PROJECT

The dataset contained information on housing data in the Nashville, TN area. I used SQL Server to clean the data to make it easier to use. For example, I converted some dates to remove unnecessary timestamps; I populated data for null values; I changed address columns from containing all of the address, city and state into separate columns; I changed a column that had different representations of the same data into consistent usage; I removed duplicate rows; and I deleted unused columns.

This dataset contains information about housing sales in Nashville, TN such as property, owner, sales, and tax information.

vpn_keyUniqueIDSort

Unique ID

text_formatParcelIDSort

Parcel ID

text_formatLandUseSort

Use of Land

text_formatPropertyAddressSort

Property Address

calendar_todaySaleDateSort

Sold Date

text_formatSalePriceSort

Purchase Price

text_formatLegalReferenceSort

Legal Doc Number

checkSoldAsVacantSort

Sold as Vacant or Not

text_formatOwnerNameSort

Owner

text_formatOwnerAddressSort

Owner Address

```
/* Cleaning Data in SQL Queries*/
```

```
/* Creating new schema and using the same*/
```

```
use newshema;
```

```
show tables;
```

```
/* Rename the data or table for our convenience */
```

```
rename table `nashville housing data for data cleaning` to nashville_housing_data;
```

```
/* Checking for the missing value*/
```

```
UPDATE nashville_housing_data
```

```
SET PropertyAddress = CASE WHEN PropertyAddress = '' THEN NULL ELSE PropertyAddress END,
```

```
ParcelID = CASE WHEN ParcelID = '' THEN NULL ELSE ParcelID END;
```

```
/* Standardize Date Format in SaleDate */
```

```
ALTER TABLE nashville_housing_data ADD COLUMN date DATE;
```

```
/*SELECT STR_TO_DATE(saleDate, '%M %e, %Y') AS date
```

```
FROM nashville_housing_data;*/
```

```
UPDATE nashville_housing_data
```

```
SET date = STR_TO_DATE(saleDate, '%M %e, %Y');
```

```
/* To delete the previous date column as it is of no use now after modification*/
```

```
alter table nashville_housing_data drop column date;
```

```
select * from nashville_housing_data;
```

```
select PropertyAddress, count(*) as times from nashville_housing_data group by PropertyAddress  
having count(*)>1;
```

```
/* To check the duplicate values*/
```

```
select * from nashville_housing_data
```

```
where PropertyAddress in
```



```
else soldAsVacant
```

```
end;
```

```
SELECT * FROM (SELECT *, row_number() over (
```

```
partition by parcelID,
```

```
propertyAddress,
```

```
salePrice,
```

```
saleDate,
```

```
legalReference
```

```
ORDER BY uniqueID
```

```
) as row_num
```

```
from nashville_housing_data
```

```
) t
```

```
where row_num > 1;
```

```
/* Delete unused columns. */
```

```
alter table nashville_housing_data
```

```
drop column OwnerAddress, drop column TaxDistrict, drop column PropertyAddress;
```