**Nashville Housing Data\_cleaning Report**

The Dataset was imported into SSMS software, and I used the ‘select \*’ to virtualize the dataset. The dataset contained 169,431 rows and 19 columns.

**A. Checking for Null and Missing values**

While studying the dataset for Null and missing values, I noticed that PropertyAddress has 87 rows of Null values. The rows contained useful information across the columns. I then decided to populate the PropertyAddress column by SELFJOIN using ParcelID and UniqueID.

The PropertyAddress was properly updated and the dataset was confirmed to still contain 169,431 rows and 19 columns.

**B. Clustered Information**

PropertyAddress and OwnerAddress contained varying information separated by delimiters such as ‘,’ and ‘.’. The details includes property address and city in one column; owner address, city and state in one column.

On propertyAddress, SUBSTRING and CHARINDEX function was used to split the columns into two named PropertySplitAddress and PropertySplitCity.

On OwnerAddress, PARSENAME and REPLACE function was used to separate the columns into three parts named OwnerSplitAddress, OwnerCity and OwnerState.

**C. Drop columns not needed for analysis**

The OwnerAddress, TaxDistrict, PropertyAddress, SaleDate, Bedrooms, FullBath, HalfBathcolumns were dropped using the DROP function.

**D. Remove duplicates**

I used the ROW\_NUMBER() OVER PARTITION BY and DELETE function to remove duplicates to avoid inaccuracy.

The final table now has 56,357 rows and 18 columns.