**Project Recap: Data Cleaning in Microsoft SQL Server (Nashville Housing Dataset)**

In this project, I performed a series of data cleaning and transformation steps on the **Nashville Housing** dataset using Microsoft SQL Server. Below is a summary of the key operations and transformations:

**1. Standardizing Date Format**

* **Issue**: The SaleDate column was not properly formatted and included unnecessary time data.
* **Solution**:
  + Added a new column SaleDateConverted to store the date in the correct format without time.
  + Used the CONVERT() function to update the new column with only the date from the SaleDate column.

**2. Populating Missing Property Address Data**

* **Issue**: Many records had ParcelID duplicates with some missing PropertyAddress data.
* **Solution**:
  + Identified rows where ParcelID was the same but PropertyAddress was missing.
  + Used a JOIN on ParcelID to fill in the missing addresses from matching rows with the same ParcelID.

**3. Breaking Out Property Address into Separate Columns**

* **Issue**: The PropertyAddress column contained both street and city data in a single field.
* **Solution**:
  + Used string functions (SUBSTRING() and CHARINDEX()) to split PropertyAddress into two columns: PropertySplitAddress (Street) and PropertySplitCity (City).

**4. Breaking Out Owner Address into Separate Columns**

* **Issue**: Similar to PropertyAddress, the OwnerAddress field contained street, city, and state in a single field.
* **Solution**:
  + Used the PARSENAME() function to split OwnerAddress into three separate columns: OwnerSplitAddress1, OwnerSplitCity2, and OwnerSplitState3.

**5. Standardizing the “Sold as Vacant” Field**

* **Issue**: The SoldAsVacant field had inconsistent values like 'Y', 'N', 'Yes', and 'No'.
* **Solution**:
  + Used a CASE statement to standardize the values, replacing 'Y' with 'Yes' and 'N' with 'No'.

**6. Deleting Unused Columns**

* **Issue**: Several columns, such as OwnerAddress, TaxDistrict, and PropertyAddress, were no longer necessary after the transformations.
* **Solution**: Dropped the unused columns to streamline the dataset.

**Conclusion:**

These transformations improved the overall quality of the Nashville Housing dataset by standardizing dates, filling missing addresses, splitting complex fields into individual columns, and cleaning up inconsistent data. These steps are crucial for enhancing the dataset’s usability for future analysis and reporting.