**Cleaning Tasks in ETL**

The ETL process involved several data cleaning tasks to ensure the accuracy and consistency of the data:

* **Duplicate Removal**: I identified unique transactions by combining the fields ReceiptNbr, ProductNbr, Transaction Date, and Quantity, leading to the removal of 23,466 transactions. The final table with no duplicate transactions contains 74,111 records.
* **Invalid Values Correction**: I rectified invalid values like '0' for some products by replacing them with 'NA' in MenuName and ProductDesc variables.
* **Whitespace Correction**: Extra spaces in MenuName were corrected and replaced with a single space.
* **Special Characters Removal**: Special characters in ProductDesc were corrected, with “?” and extra spaces removed.
* **Consistency Maintenance**: I replaced values in the MenuType field by substituting “Soups & Salads” with “Soup & Salad”.
* **NULL Value Handling**: I replaced all NULL values in the Building Type field with 'NA'.

**Assumptions**

**Dimension Tables Creation**:

* For the CustomerInfo dimension table, I assumed that each receipt number corresponds to a unique customer. This could represent a family or an individual making multiple purchases in a single transaction, particularly when analyzing the impact of the number of drive-through lanes on sales.
* I did not truncate string formats (varchar) to low character lengths to avoid affecting future lookups through new dimension tables. Since the source was a database table and not a flat file, truncating variables to limit memory usage was unnecessary.
* The ETL package was designed to only load facts with valid dimensions. If a fact did not have an associated surrogate key in a dimension table, it was not loaded into the final fact table as the lookup returned a NULL value for the dimension primary ID.

**View Statements for Executive Questions**:

* For Questions 1 and 2, I considered a specific date to identify the top 3 locations or the number of customers served by each location, assuming the analyst would be interested in a specific period or time of analysis.
* For Question 3, I assumed that the analyst would be interested in a specific store for understanding the top 10 products sold during a certain period or on a certain day. Hence, I included a WHERE statement to specify the store number.
* For Question 4, I considered Gross Amount as the key business metric to identify sales and calculate total sales. However, Quantity could also be a metric of interest. I assumed that executives are not interested in a specific location or duration when conducting this analysis.
* For Question 5, I considered the receipt number/customer as a unit of analysis, assuming that customers usually come in their vehicles and purchase multiple products on a single receipt. I assumed that executives are interested in the effects of drive-thru lanes on both net profits and sales, and thus, I created two view statements.