**1. Project Overview**

* **Title**: *Super Store Sales Data Analysis*
* **Objective**: The primary goal of this project is to analyze sales trends, customer segments, and product performance based on a historical dataset of store transactions.
* **Scope of Analysis**:
  + Analyze sales performance across various product categories, regions, and customer segments.
  + Understand delivery performance by analyzing shipping dates and shipping modes.
  + Provide actionable insights for business decision-making based on the dataset.

**2. Data Model Overview**

* **Data Source**: The dataset contains a single table with transactional data from a superstore.
* **Data Model**: A **flat file** model with each row representing a single transaction or order. All relevant details about the order, customer, and product are recorded in this table.

**3. Data Dictionary & Data Types**

The dataset has several columns, and each column’s data type plays a role in how it is used for analysis. Below is a detailed list of each column, its description, and its data type.

| **Column Name** | **Description** | **Data Type** |
| --- | --- | --- |
| Order ID | Unique identifier for each transaction or order. | **String** |
| Order Date | Date when the order was placed. | **Date** |
| Ship Date | Date when the order was shipped. | **Date** |
| Ship Mode | Mode of shipment used for the order (e.g., Standard, Second Class). | **String** |
| Customer ID | Unique identifier for the customer placing the order. | **String** |
| Customer Name | Full name of the customer. | **String** |
| Segment | Customer segment classification (e.g., Consumer, Corporate, Home Office). | **String** |
| Country | Country where the order was placed. Typically only one value (e.g., USA). | **String** |
| City | City associated with the customer’s shipping address. | **String** |
| State | State associated with the customer’s shipping address. | **String** |
| Postal Code | Postal or ZIP code of the shipping address. | **String** |
| Region | Geographic region (e.g., East, West, Central) where the customer is located. | **String** |
| Product ID | Unique identifier for the product purchased in the order. | **String** |
| Category | Product category (e.g., Furniture, Office Supplies). | **String** |
| Sub-Category | Sub-category within the product category (e.g., Chairs, Binders). | **String** |
| Product Name | Name of the product purchased in the order. | **String** |
| Sales | Total sales amount for the product in the order. | **Decimal** |

**4. Data Cleaning and Preparation**

**1. Handling Missing Data**

* **Issue:** The *postal code* (ZIP code) for Burlington, Vermont was missing.
* **Action Taken:** We filled in the missing postal code with 05401, which is the correct ZIP code for Burlington.

**2. Data Type Correction**

* **Issue:** The *Order Date* and *Ship Date* columns were not in the correct format.
* **Action Taken:** We converted both *Order Date* and *Ship Date* columns to Date format in Power Query to enable accurate date-based analysis.

**3. Duplicate Data**

* **Action Taken:** We checked for duplicates by examining all rows and the *Row ID* column. No duplicates were found in the dataset.

**4. Cleaning the Product Name Column**

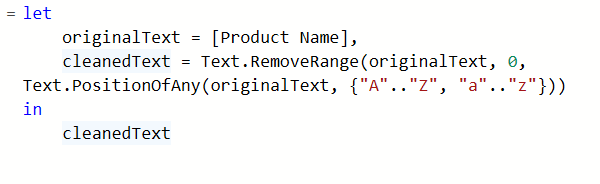
The product names contained unnecessary details, such as descriptions and numbers. We cleaned the column in the following steps:

**Step 1: Remove Unnecessary Descriptions**

* **Action:** After Trimming and Cleaning we split the Product Name column by the **comma** (,) delimiter, keeping only the main product name and removing the descriptive parts (e.g., size or dimensions). This was done by splitting at the **right-most** delimiter twice, which removed excess information after the product name.

**Step 2: Remove Numbers at the Beginning of Product Names**

* **Action:** We added a custom column to remove numbers and unnecessary characters from the beginning of the *product name*. The following M-code was used:



**Explanation:** This code removes any characters (such as numbers or symbols) before the first letter of the product name.

**Step 3: Remove Numbers at the End of Product Names**

* **Action:** We added another custom column to remove any numbers or unnecessary characters from the **end** of the product name. The following M-code was used:

A computer code with text

Description automatically generated with medium confidence

**Explanation:** We reversed the text, removed characters from the beginning (which were originally at the end), and then reversed it back to restore the original order without trailing numbers or symbols.

### ****5. Creating New Columns****

To enhance the dataset for further analysis, we added new calculated columns.

#### **Step 1: Shipping Time**

* **Action:** We calculated the shipping time by subtracting the Order Date from the Ship Date.



* **Explanation:** This calculates the number of days taken to ship the product, which can be used to analyze delivery performance.

**Step 2: Product Brand**

* **Action:** We extracted the first word by Splitting the cleaned Product Name to create a new column representing the product brand.