**Case Study1: Retail Sales Data Analysis**

**Scenario:**

You’ve been hired by a retail chain to analyse their sales and product data. The data comes from different systems and formats. Your task is to clean, prepare, and transform the data to generate useful insights for the operations team.



**In/Out Tools + Data Types**

* Import all 3 datasets using Input Data tool.
* Use **Browse** and **Auto Field** tools to check and adjust data types.
* Validate that SaleDate is Date, UnitPrice and Quantity are Numeric.

**Filter, Select, Formula, Sample, Unique, Sort**

* Filter: Sales from “South” region.
* Formula: Add a new field TotalSale = Quantity \* UnitPrice.
* Select: Reorder columns and remove unwanted fields.
* Sample: Take first 3 records and last 2 records from sales table.
* Unique: Get unique list of CustomerIDs.
* Sort: Sort sales by SaleDate descending.

**Preparation Tools**

* Auto Field: Automatically assign optimal data types.
* Data Cleansing: Remove null values and unwanted characters from customer names.
* Create Samples: Split sales data into 70% training and 30% testing.

**Join Tools**

* Join Sales with Product Data using ProductID.
* Join Sales with Customer Data using CustomerID.
* Use Union Tool to merge data from multiple regions (simulate with region-wise files).
* Use Join Multiple to join all 3 datasets in one go.

**Record-Level Tools**

* Generate Rows: Create a running date range for February 2024.
* Record ID: Add a sequential record number to each sale.
* Append Fields: Add region manager details (create a dummy manager file).
* Find Replace: Standardize the details, Replace the category with the description below

|  |  |
| --- | --- |
| Category | Description |
| Home | Home & Living |
| Clothing | Fashion & Lifestyle |
| Books | Books & Magazines |
| Toys | Outdoor & Indoor Play |
| Electronics | Electronics & Gadgets |

* **Transform Tools**
* Count Records: Total number of sales transactions.
* Cross Tab: Pivot data to show TotalSale by Region and Category.
* Running Total: Running total of TotalSale by Customer.
* Summarize: Total quantity and sale per product.
* Transpose: Convert wide format to long for product analysis.

**Parse Tools**

* DateTime: Extract Month and Year from SaleDate.
* Regex: Extract username from Email (^(.+)@).
* Text To Columns: Split PhoneNumber by digits if formatted.

**Documentation Tools**

* Add **Comment** tools to document step.
* Use **Tool Container** to group "Data Cleansing" tasks.

**Sample Tasks:**

1. Identify the top-selling product in each region.
2. Create a sales trend report using DateTime parsing.
3. Identify unique customers and sort by highest total spend.
4. Use containers and comments to make your workflow readable.