

## Data Preparation using SPSS Modeler

Some useful data exploration considerations:

- ✓ Prepare a data audit for each source file and examine each field individually
- ✓ Find the total number of transactions for each customer
- ✓ Find the number of different products each customer purchased
- ✓ Find the total dollars spent by each customer
- ✓ Find the average dollars spent per transaction by each customer
- ✓ Calculate the following correlations:
  - Income and total dollars spent
  - Income and number of transactions
- ✓ Create a matrix to visualize the relationship between a customer's experience score and whether or not they are in a loyalty group
  - Note: Experience score will need to be casted to a string before creating the matrix

# SPSS Modeler Data Preparation Stream

## Step 1 –

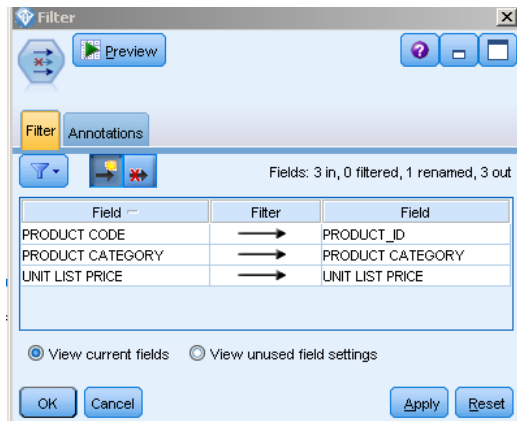
Read the 3 data files for “Products”, “Transactions”, and “Customers” from the .csv inputs by choosing “Fixed File” from SPSS Modeler “Sources” section



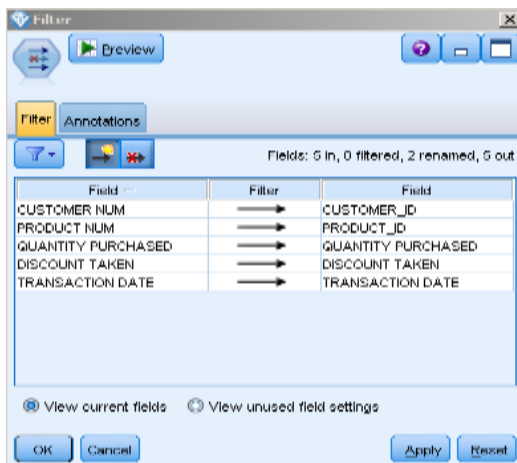
Once reading is done, for all three files, use “Filter node” from “Field Ops” section of SPSS Modeler to name the columns of the 3 files in a uniform way so that joins can be performed.



## Products File –



## Transactions File –



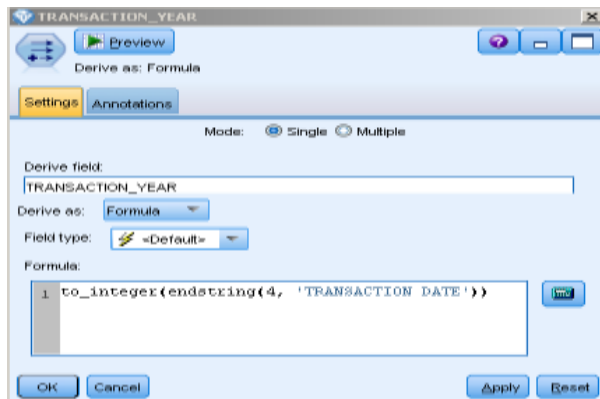
## Customers File -



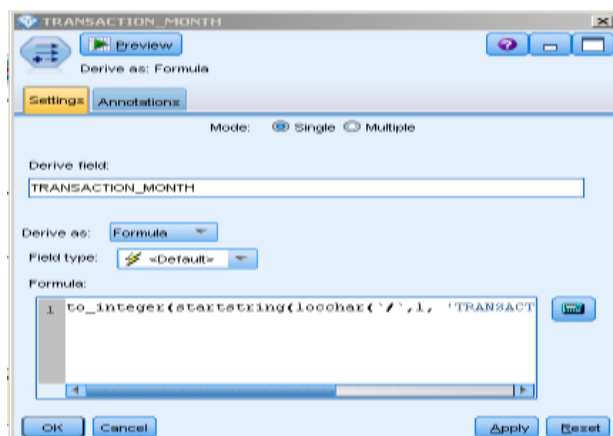
## Step 2 –

From the “Transactions” file, 3 fields have to be derived to facilitate calculations at a later stage. These fields are Transaction\_Year, Transaction\_Month, and Transaction\_Day.

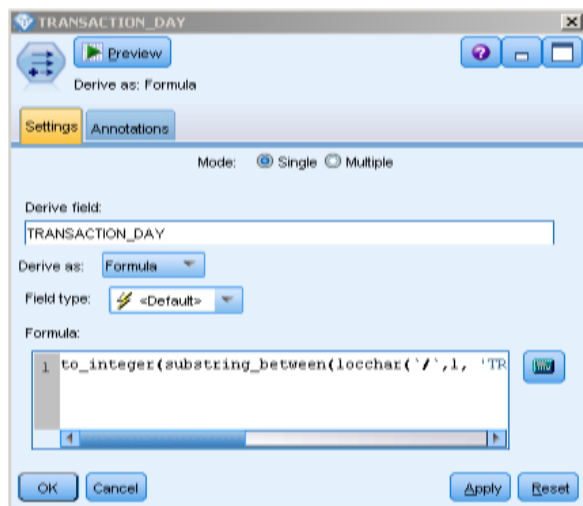
## Transaction\_Year:



## Transaction\_Month:

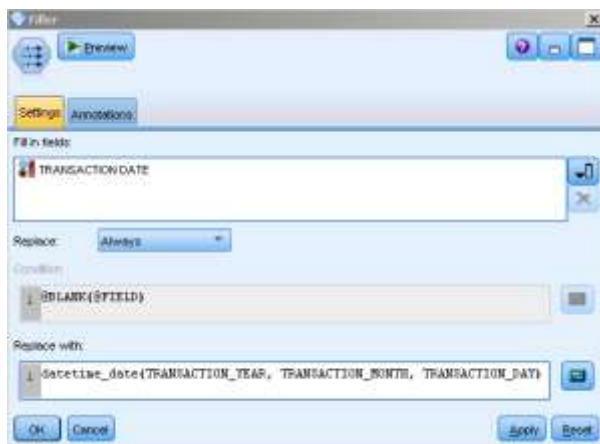


Transaction\_Day:



Step 3 –

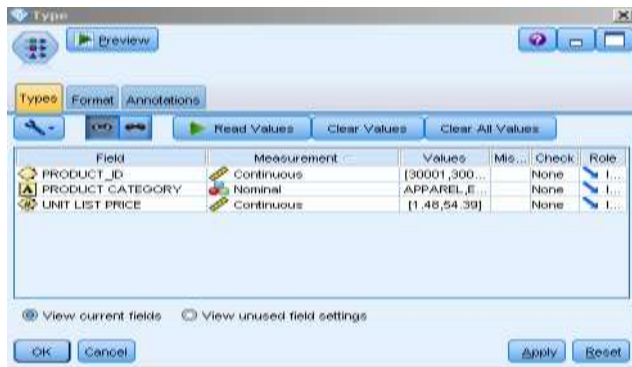
For Transactions file, the date format needs to be changed for SPSS input.



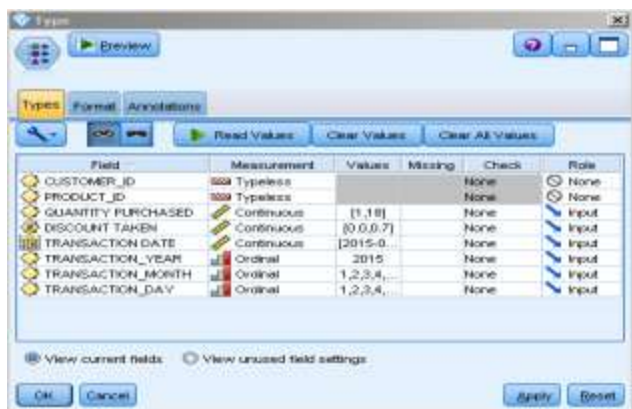
#### Step 4 –

Before performing “join” operation, a “Type” node needs to be used to arrange the data types in different categories (such as: Continuous, Nominal etc.)

Modify Products file with “Type” node –



Modify Transactions file with “Type” node –



#### Step 5 –

Merge all three files. First, merge Customers and Transactions file on CUSTOMER\_ID.

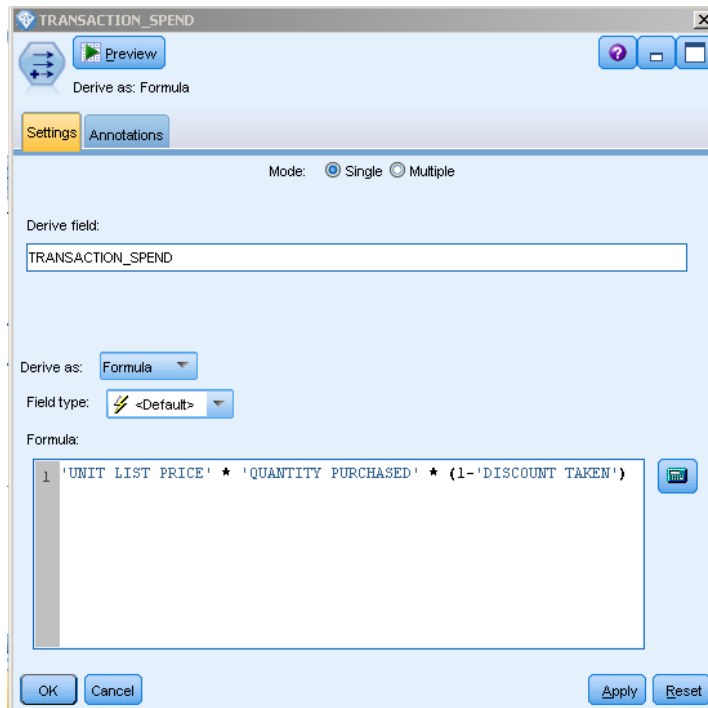


Then merge the modified file with Products file on PRODUCT\_ID.

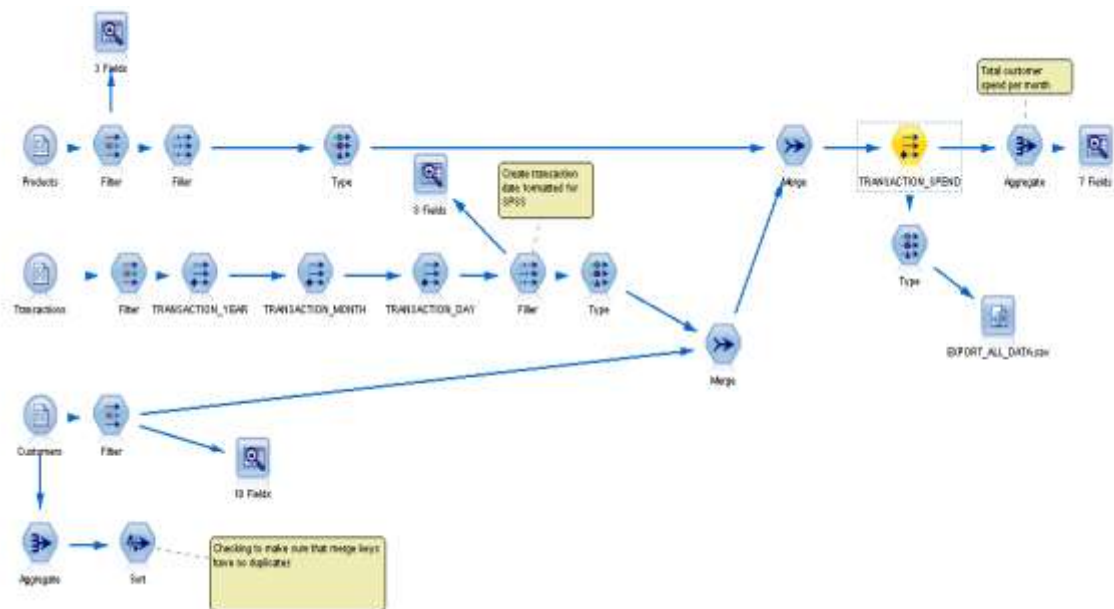


Step 6–

Finally per Transaction Spend is calculated on the derived file as an additional column.



End to End Data creation diagram –



The data is extracted as a Statistics File –

