

Practical File

Name : Ganesh Agrhari

Subject : Predictive Analytics (BCADSN15301)

Class : BCA DS&AI; 33

Roll No. : 1230258126

Submitted To : Mr. Robin Tyagi

Practical-5

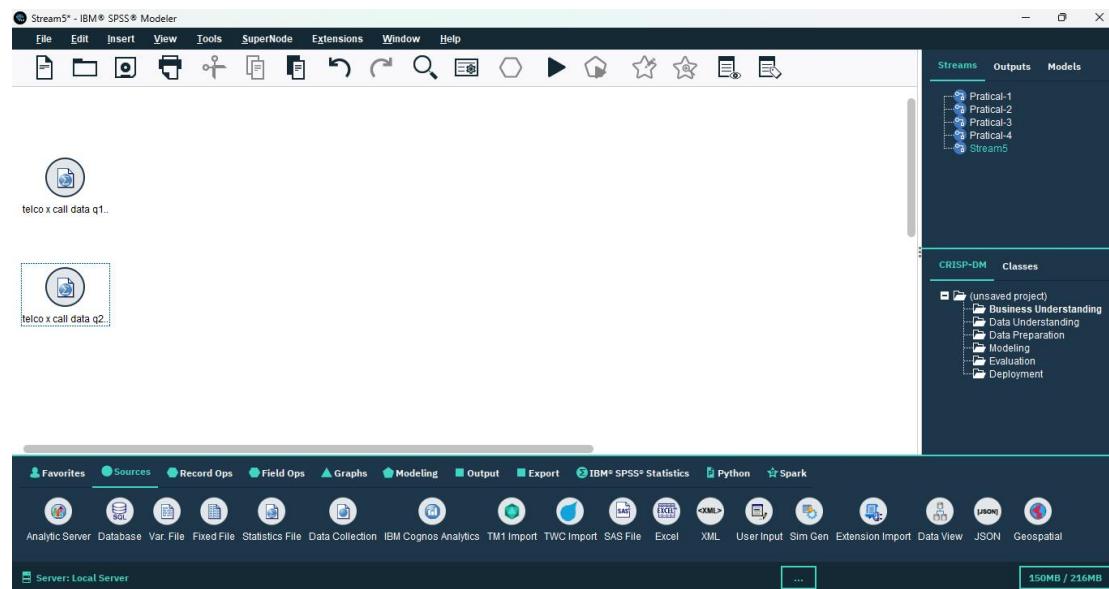
Definition: In SPSS Modeler, data integration is used to learn how to combine two datasets using methods such as merge, append, sampling, etc.

Outcomes/Learning: We learned to join one dataset with another using specific operations to integrate data effectively.

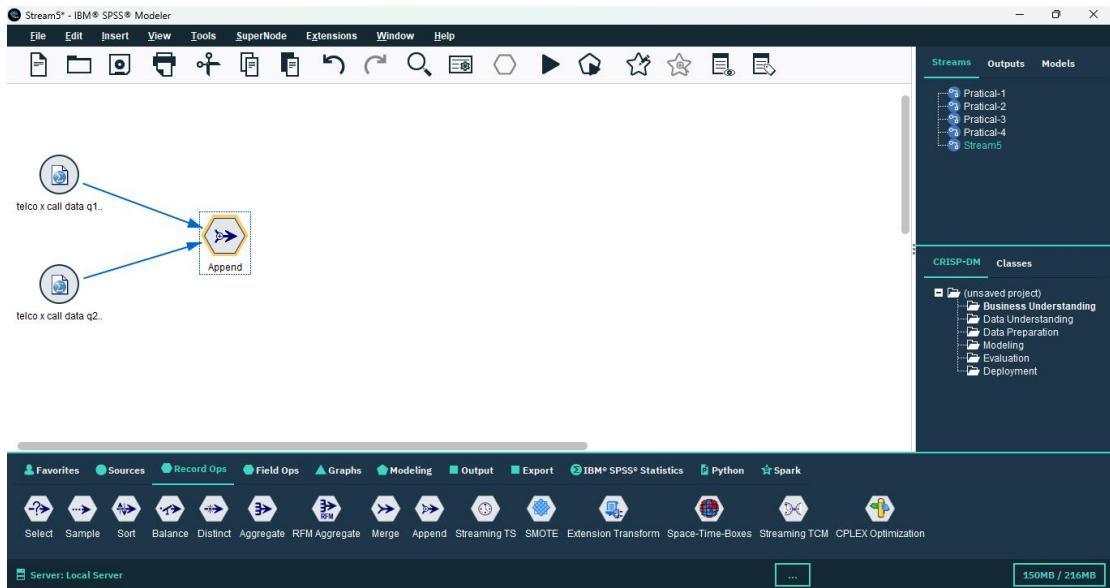
Required Tool: IBM SPSS Modeler tool.

Working: We used multiple nodes to combine datasets based on their fields or records.

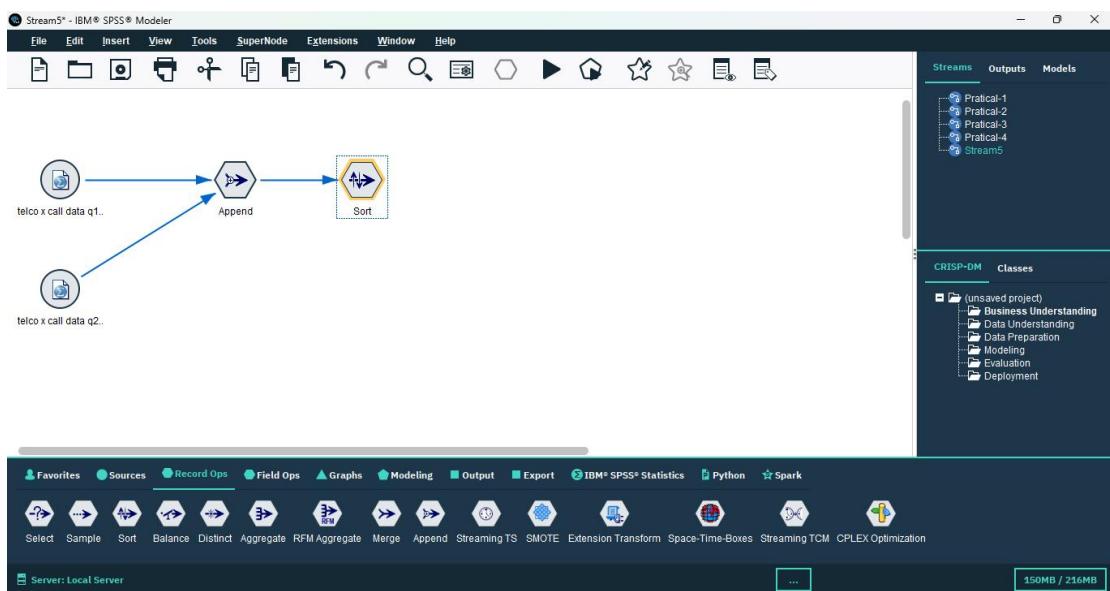
Step 1: First, we added two Statistics File nodes and imported the Telco X Call Data Q1 and Q2 datasets.



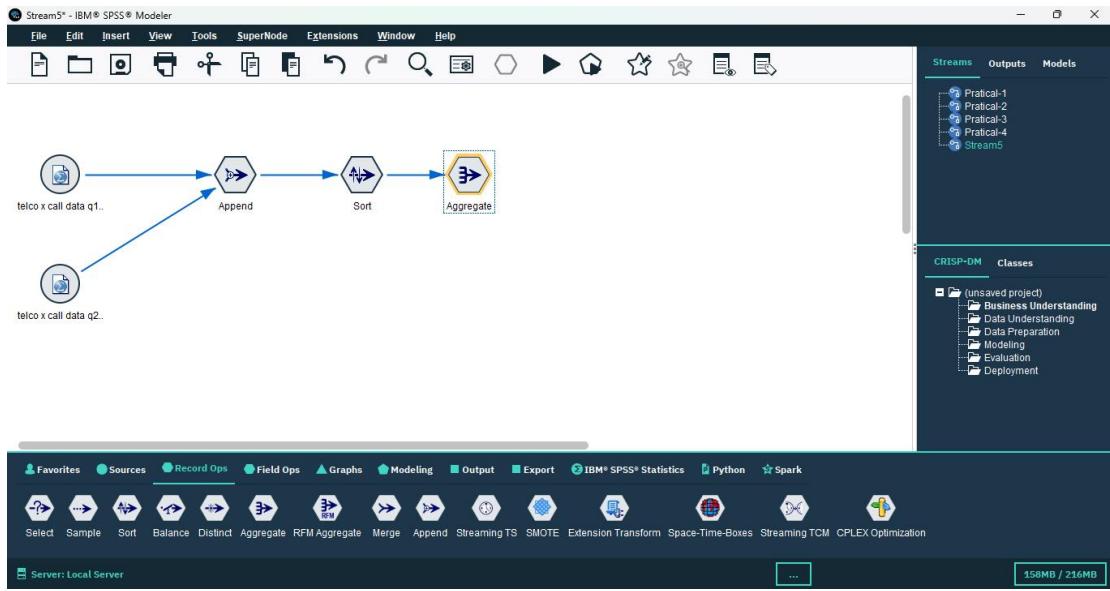
Step 2: We added the Append node from the Record Ops category and connected both datasets to it. After configuring the node, we applied the settings and clicked OK.



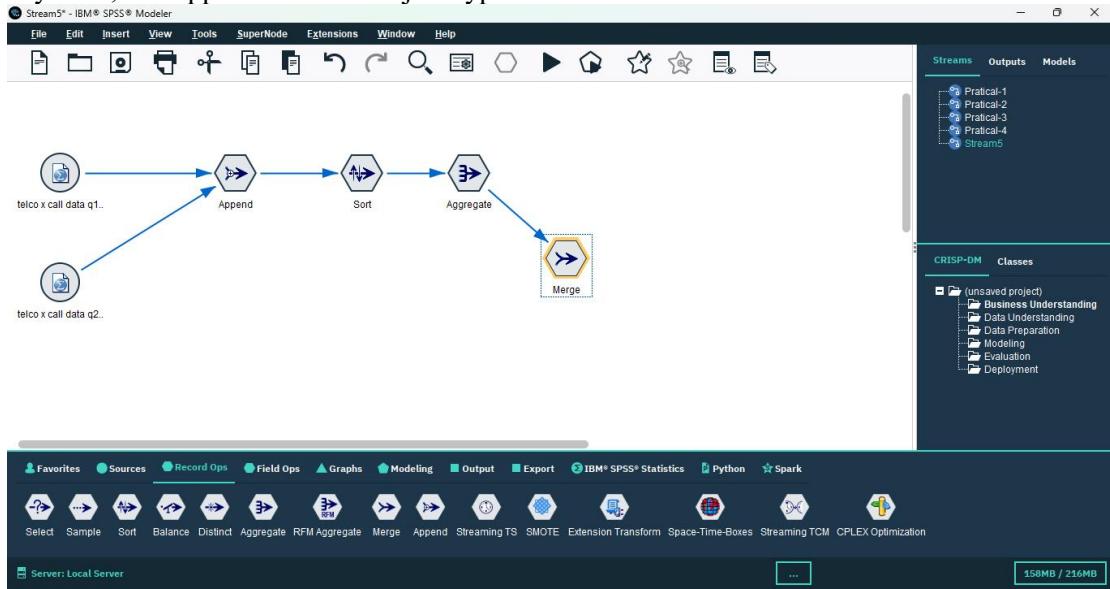
Step 3: We used the Sort node from the Record Ops category to sort the data by Customer ID and Month in ascending order.



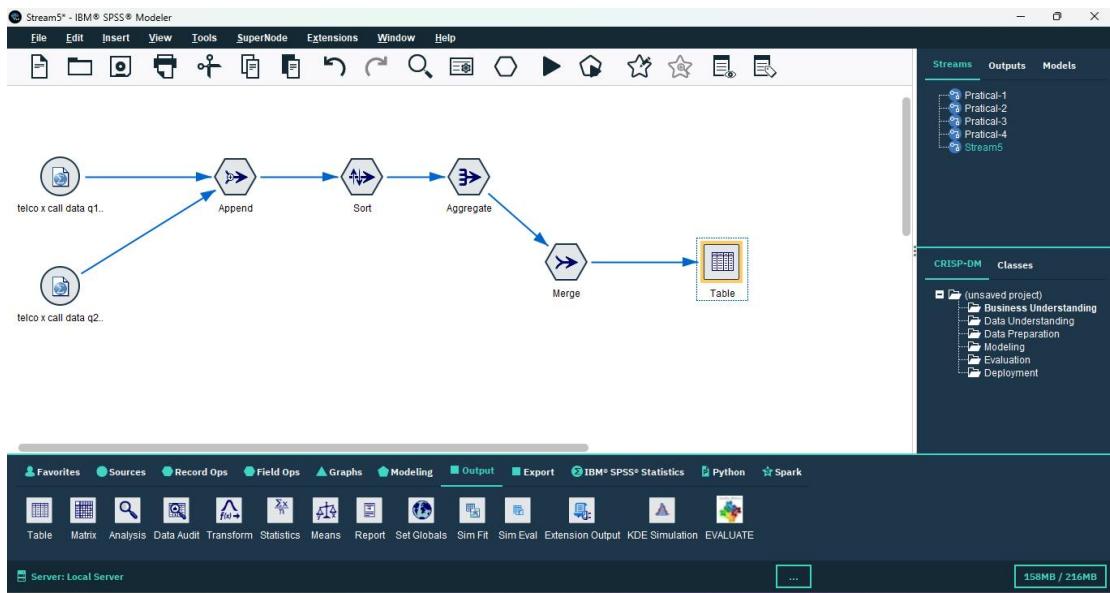
Step 4: Next, we connected the Aggregate node and used Customer ID as the key field to perform summation or aggregation of the data.



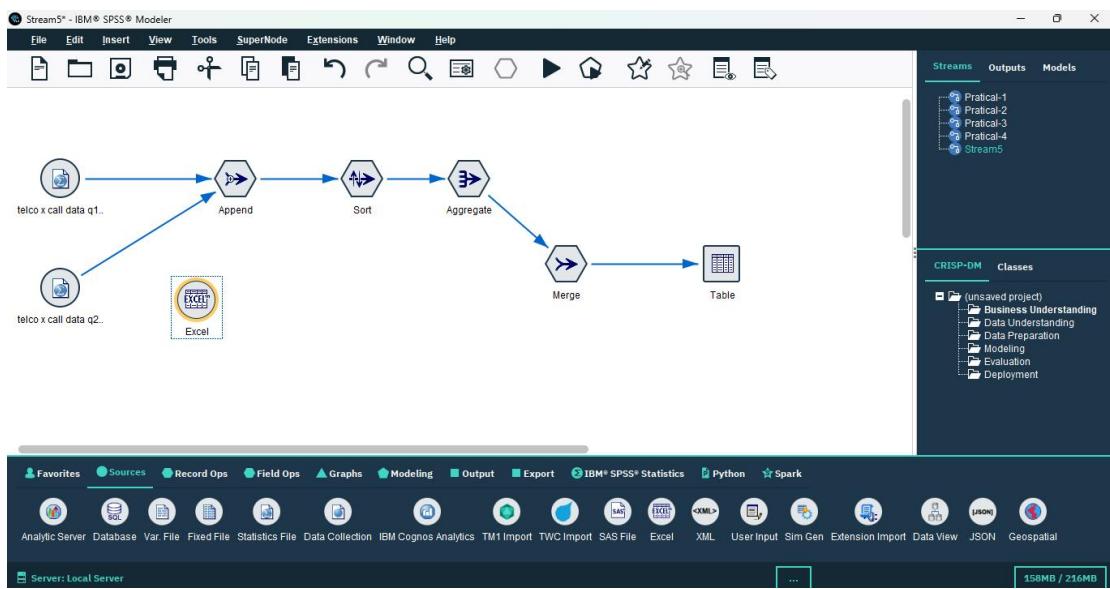
Step 5: We added the Merge node from the Record Ops category, used Customer ID as the key field, and applied the desired join type to combine the datasets.



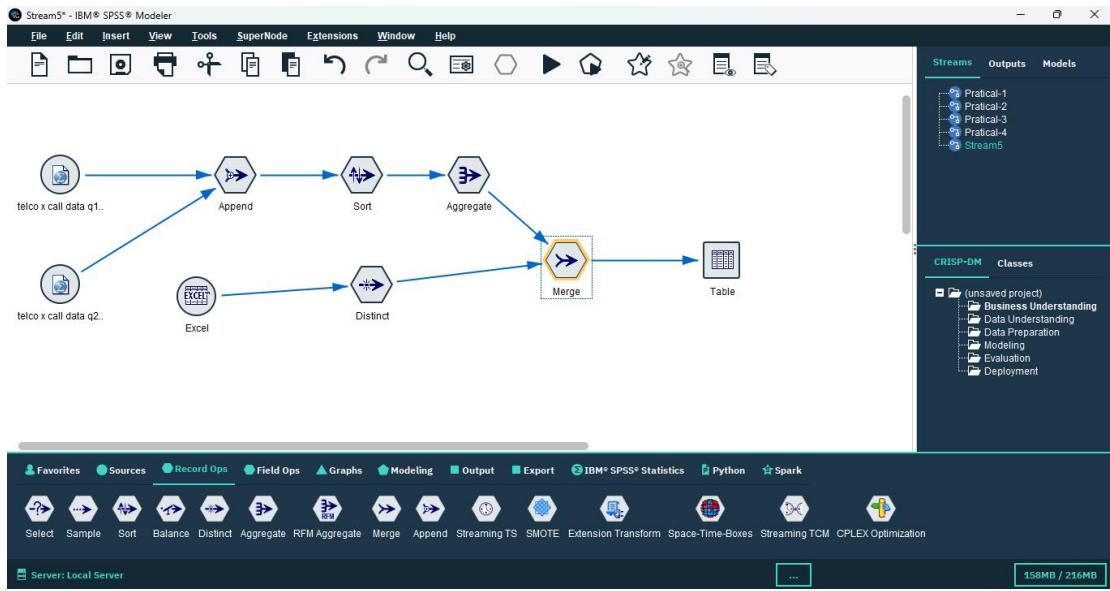
Step-6: We apply the table node to see the changing of dataset or used to run it.



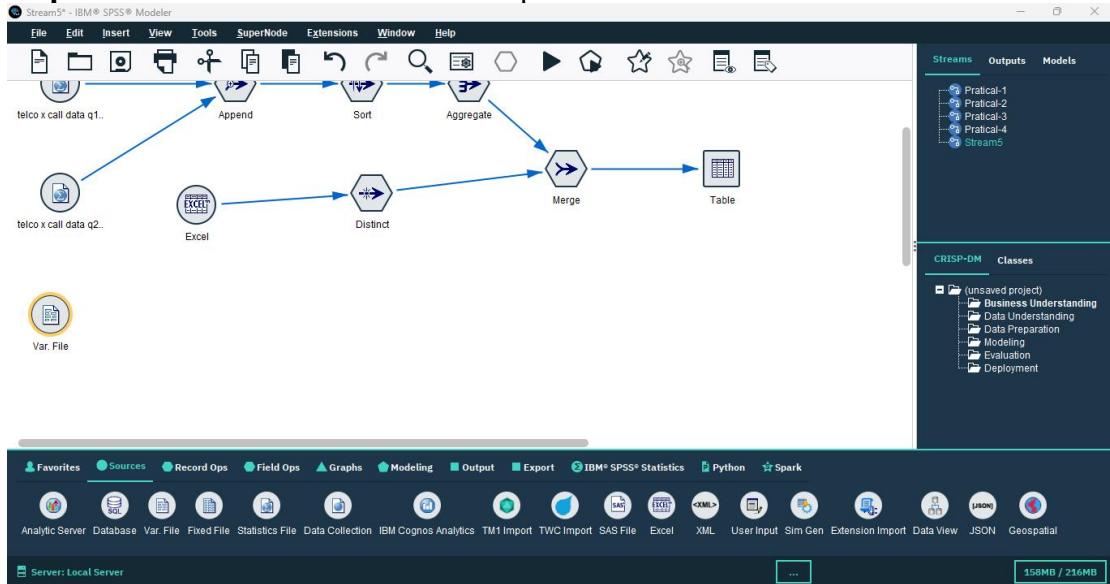
Step 7: Next, we added an Excel File node to import the Telco X Customer Data.



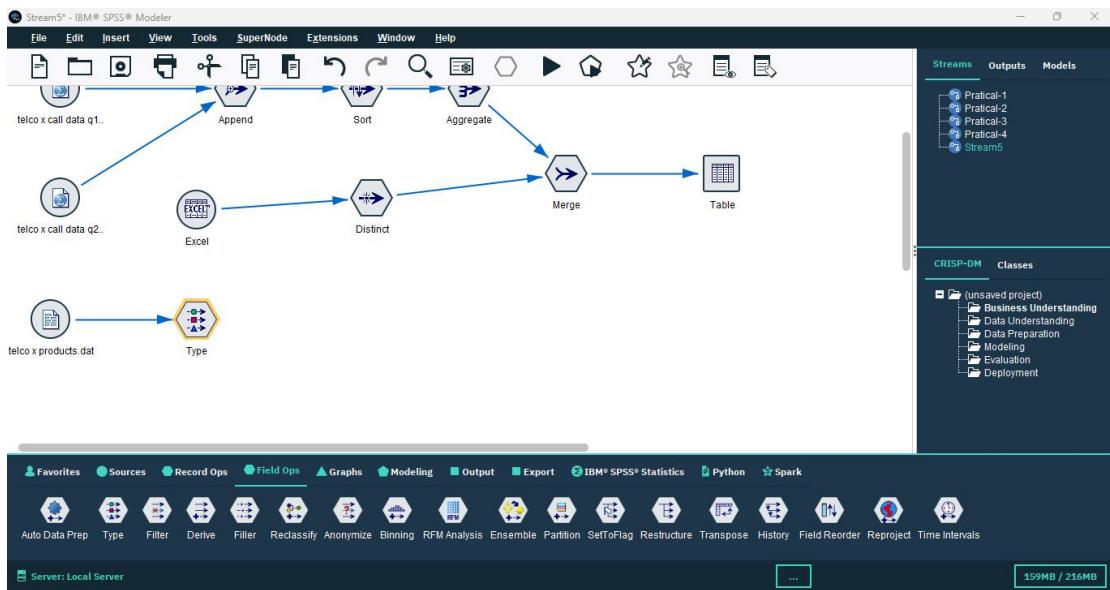
Step 8: We added a Distinct node, selecting all fields as key values to remove duplicate records from the dataset. After configuring the Distinct node, we connected it to the Merge node.



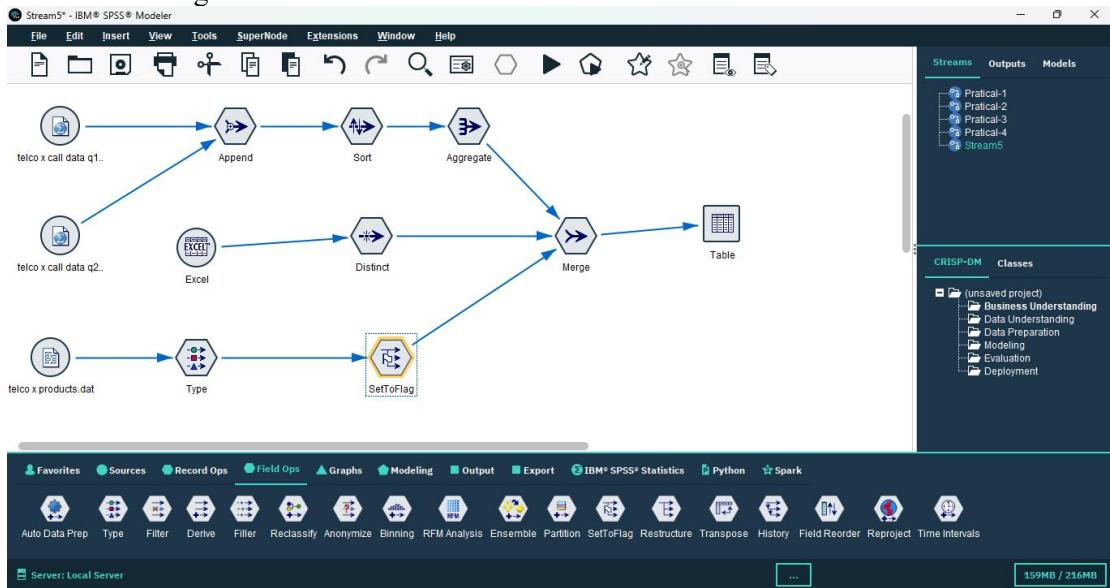
Step 9: We added a Var File node to import the Telco X Product Data.



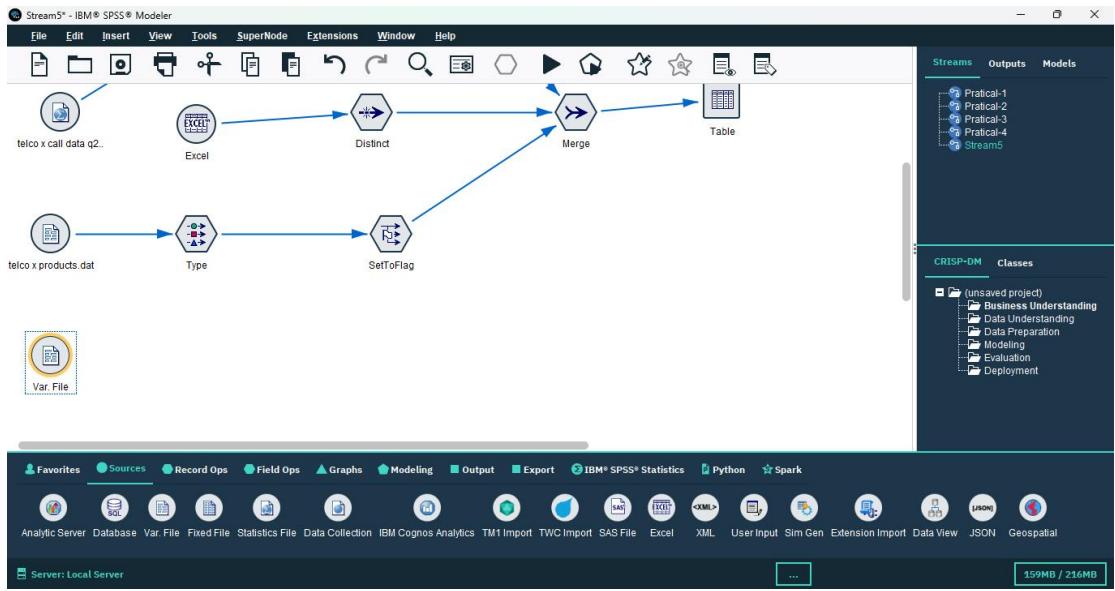
Step 10: Next, we connected a Type node to examine the measurement level of the fields. Initially, the data consisted of two types: continuous and categorical. After processing, the data was classified into flag, typeless, nominal, and ordinal categories.



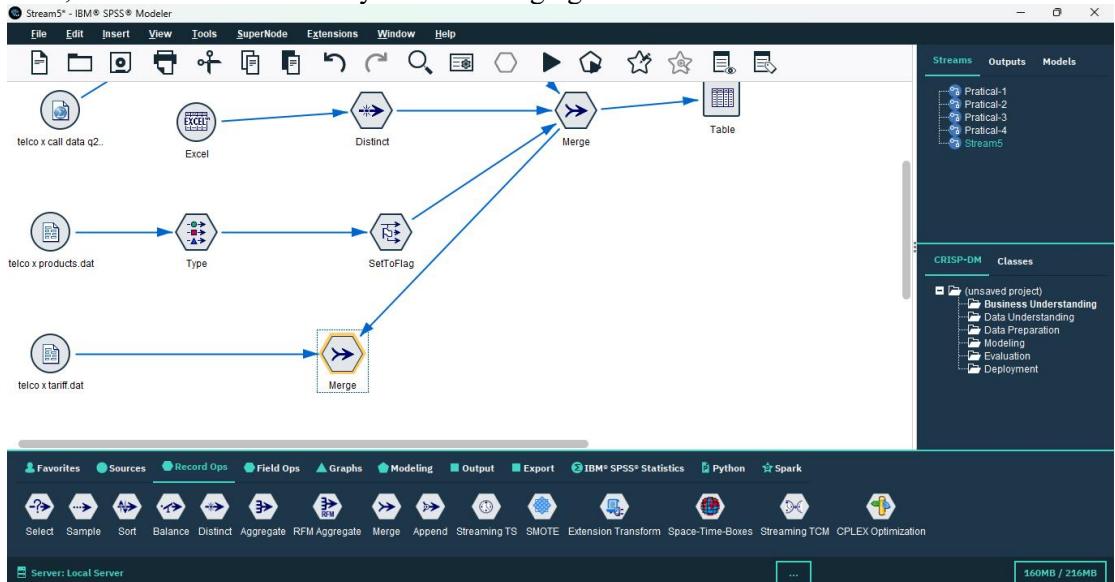
Step 11: In the Type node, we connected a Set to Flag node, which converts all field values into flag format. For example, we set all values of Gadgets as a flag and then connected this node to the Merge node.



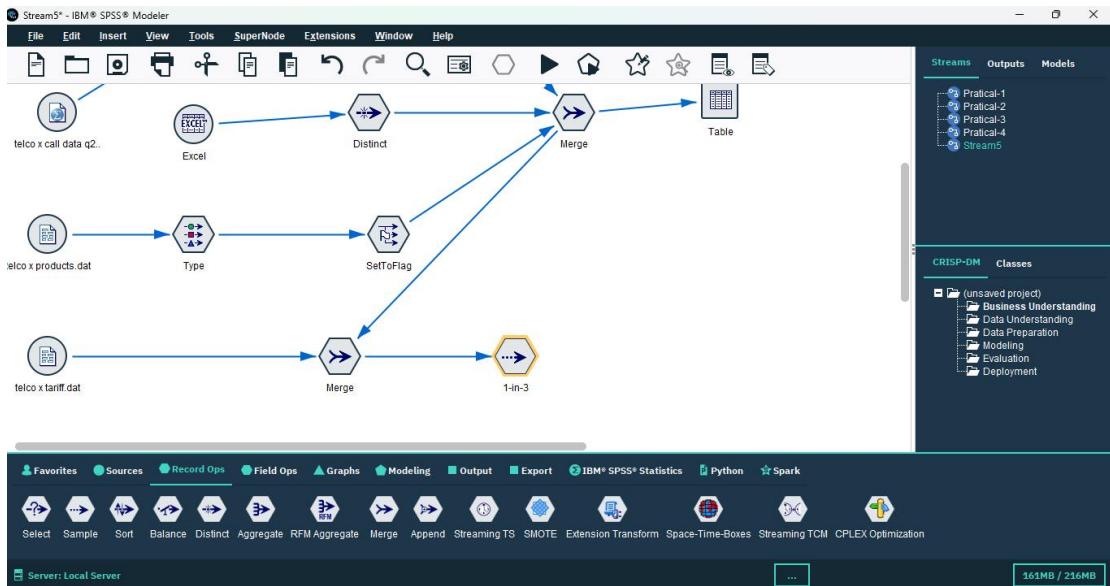
Step-12 : Again we take a var file and import the telco x tariff data in this.



Step 13: We added another Merge node and connected it to the previous Merge node. In this node, we used Tariff as the key field for merging.



Step 14: Next, we added a Sampling node from the Record Ops category. This node is used to select a sample from a large dataset. We applied the 1-in-n rule, taking every 3rd record (1 in 3) for the sample.



Step 15: Finally, we added a Table node from the Output category to display the final output and ran it.

