Practical File

Name : Ganesh Agrhari

Subject : Predictive Analytics (BCADSN15301)

Class : BCA DS&AI; 33

Roll No. : 1230258126

Submitted To : Mr. Robin Tyagi

Practical-4

Definition:-The Unit of Analysis refers to the level at which data is grouped and analyzed. It is used to:

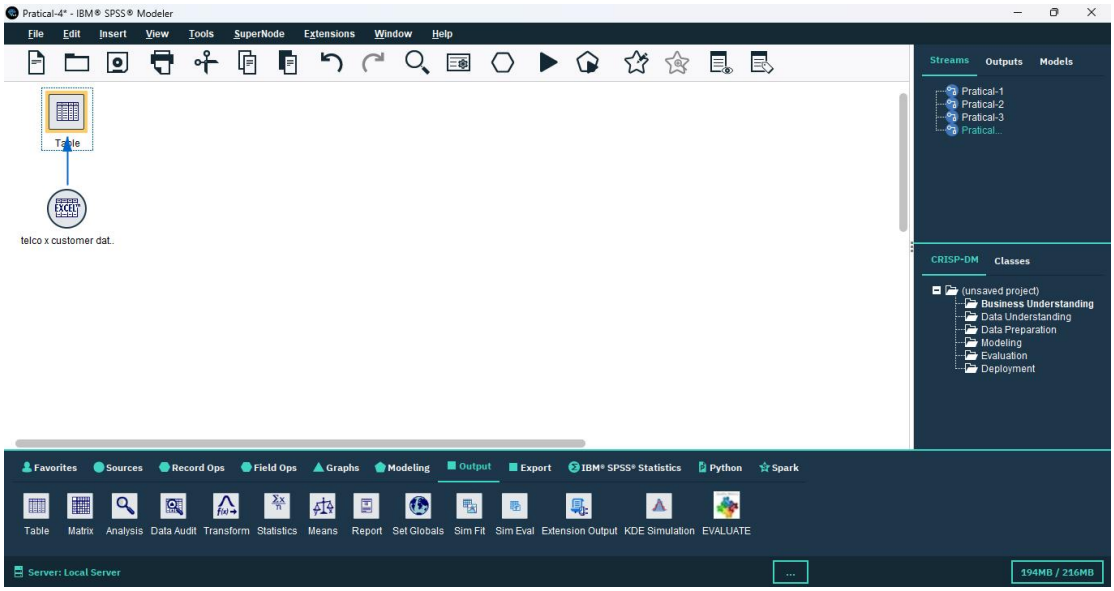
- Remove duplicate rows, ensuring each record is unique based on selected fields.
- Summarize multiple rows into one per group, thereby changing the unit of analysis.
- Create a binary (0/1) field based on specific conditions to label or filter data at a chosen level.

Outcomes/Learning: Learned to remove duplicate data, merge revenue using aggregation, and assign values using flag fields.

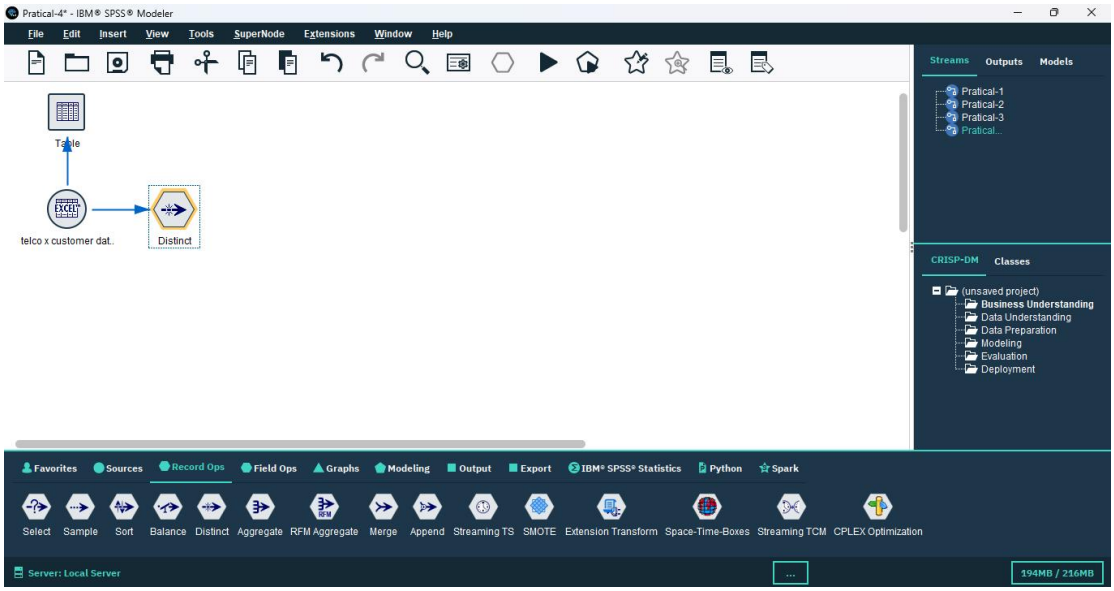
Required Tool: IBM SPSS Modeler tool.

Working: We used this data to identify and retain valid customer records, keeping only the known data and discarding the rest.

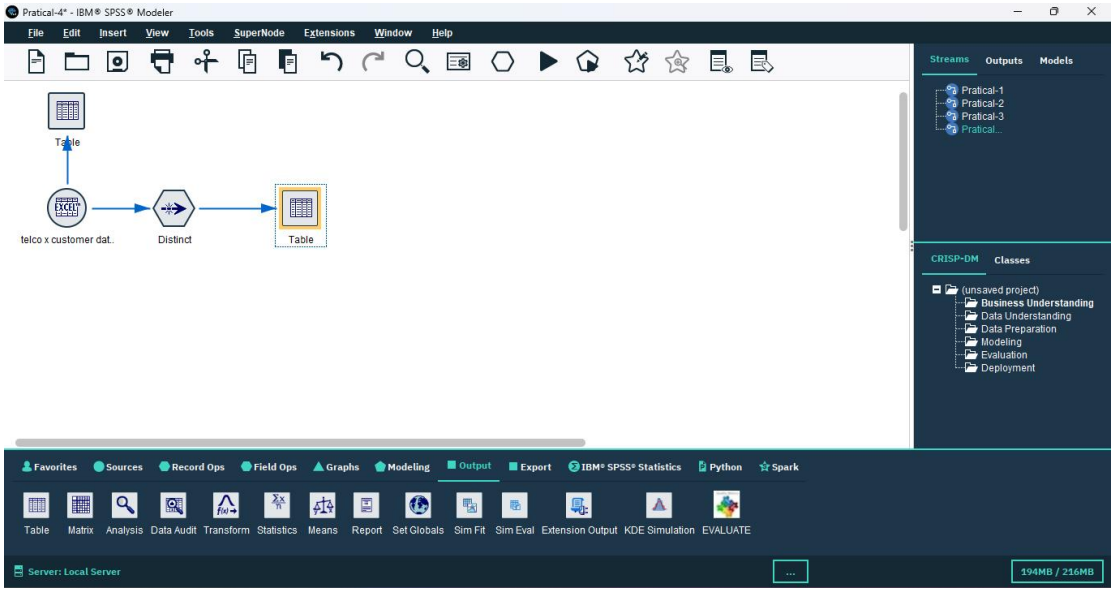
Step 1: Open the SPSS Modeler tool, navigate to the Source category, and double-click the Excel node to place it on the canvas. Then, import the Excel data into this node.



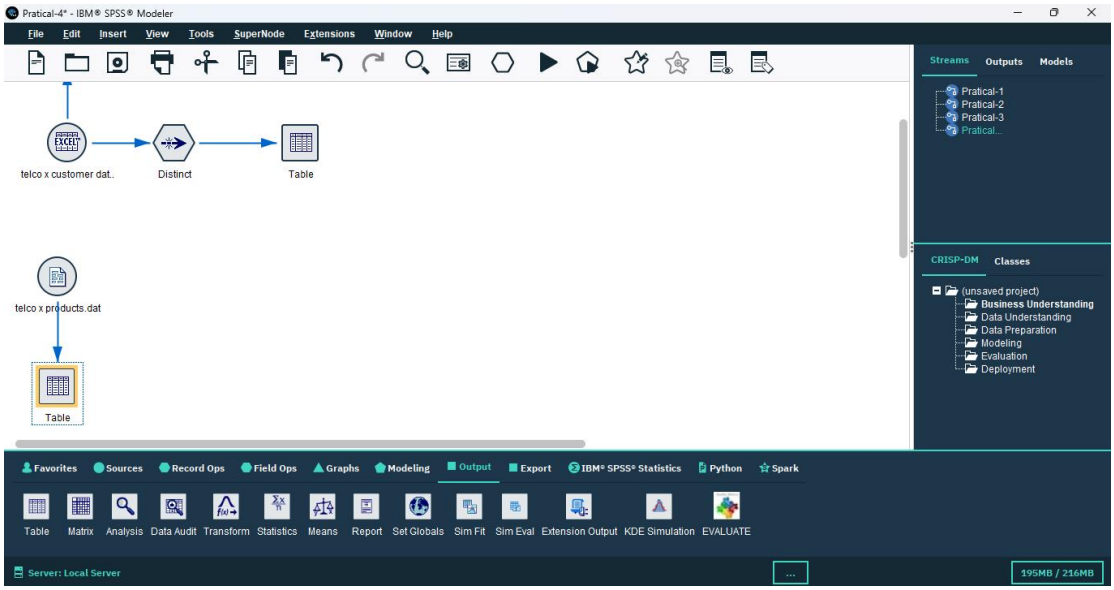
Step 2: We used the Record Ops node to select distinct data, which removes duplicate records from the dataset. In this process, we set all fields as key values to ensure that only unique.



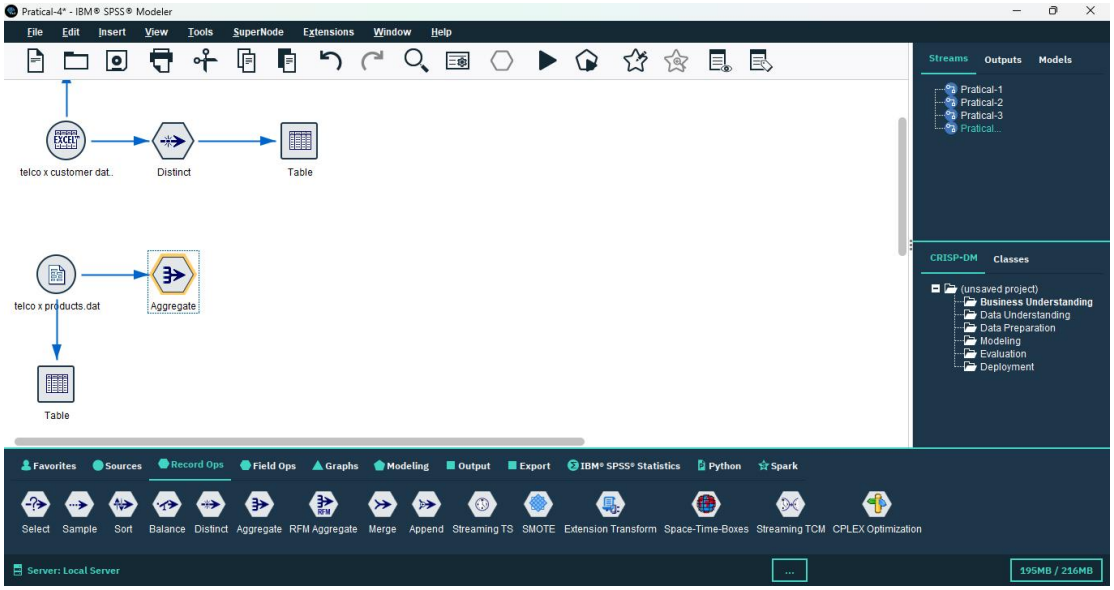
Step 3: We connected a Table node to check whether the duplicate records were removed successfully or not.



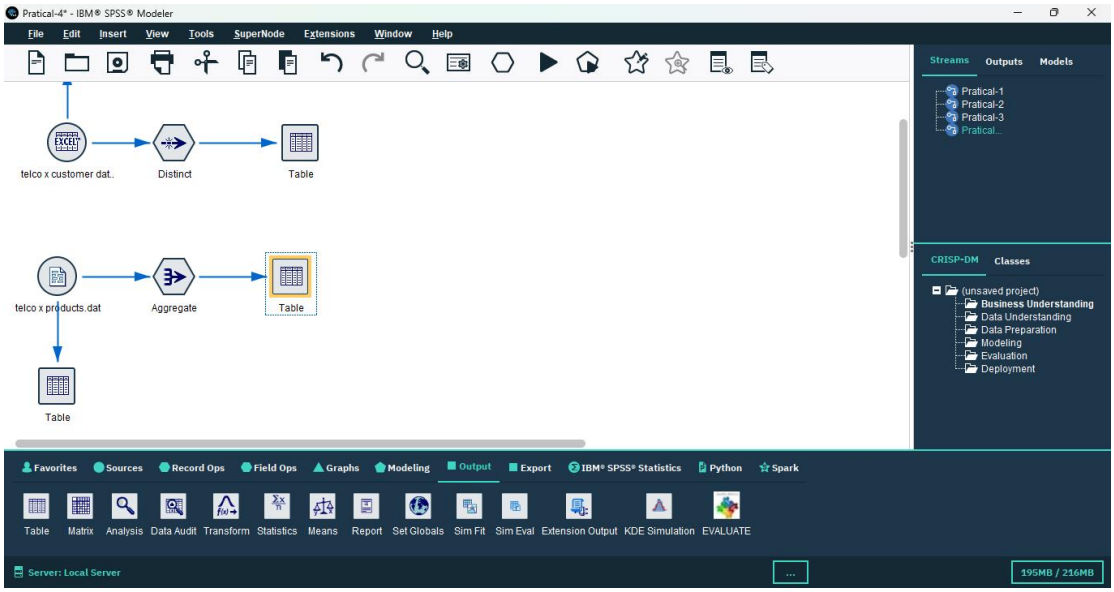
Step 4: We used a Var File node to import the flat data, and then connected a Table node to display the output in table format.



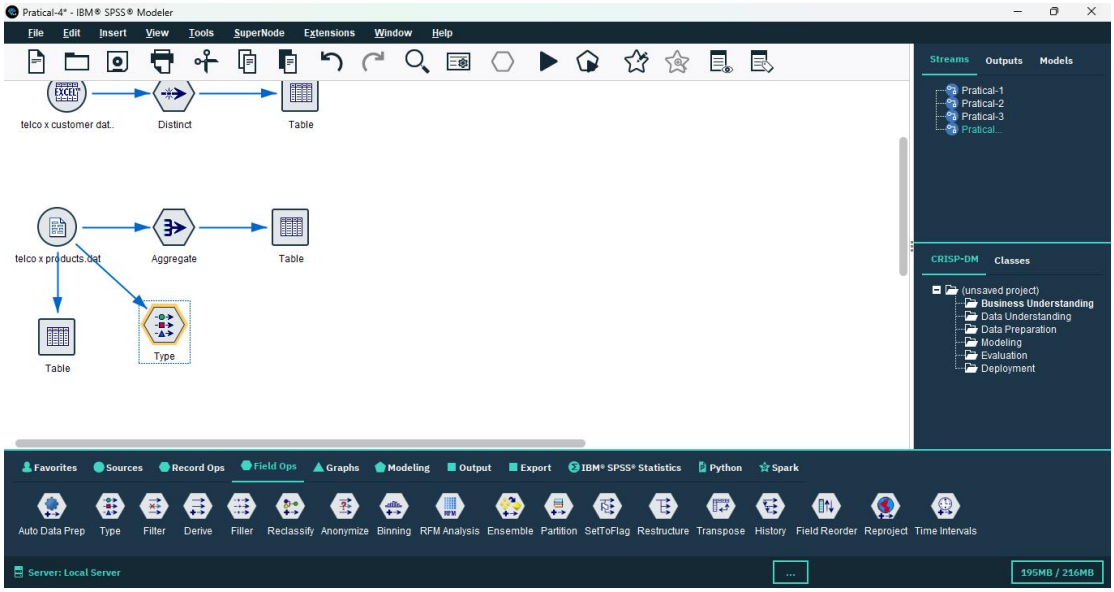
Step-5: We used the record ops to connect a aggregate node. In this we use to sum the revenue field values. Then we take key value as revenue and apply sum and mean .



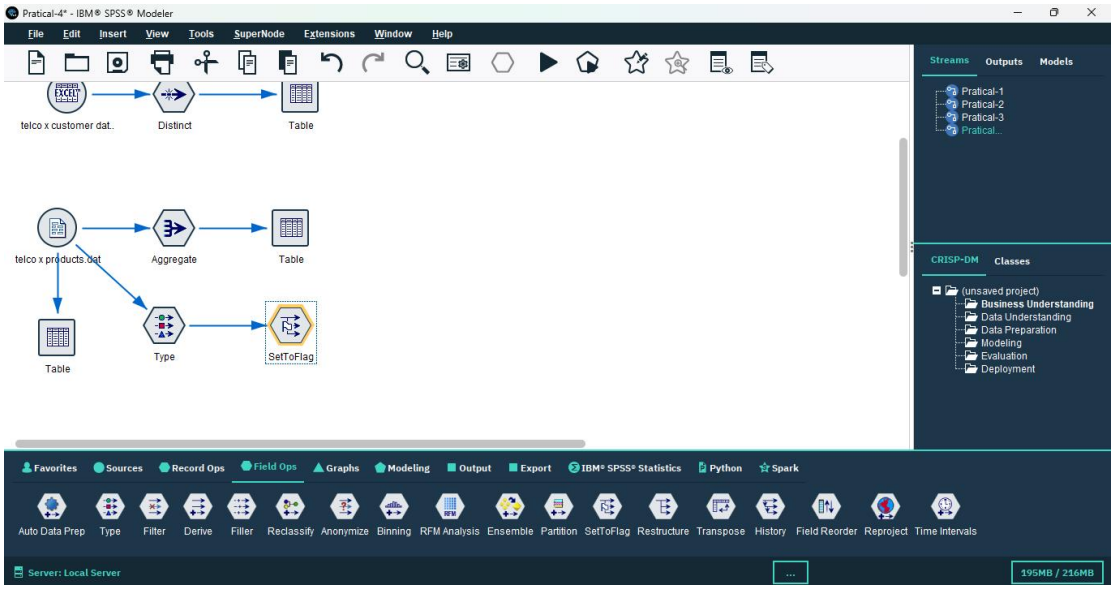
Step 6: We connected a Table node to view our output in a table format.



Step-7: We connect a type node to read the data from the field ops category.



Step 8: We used the Field Ops node to add the Set to Flag node. This converted all product values into a flag format, representing them as Yes/No or 0/1.



Step 9: We added a Table node from the Output category to view the final results of the Var File data.