BABU BANARASI DAS UNIVERSITY

CASE STUDY

ON

Predictive Analysis of Customer Purchase Behavior in Online Retail Using SPSS.

SUBMITTED TO:

Mr. Robin Tygai

SUBMITTED BY:

Name: Deeksha Singh

Roll No: 1230258150

Name: Janwi Kumari

Roll No: 1230258205

Name: Kirti Srivastava

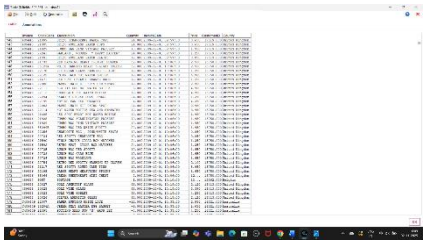
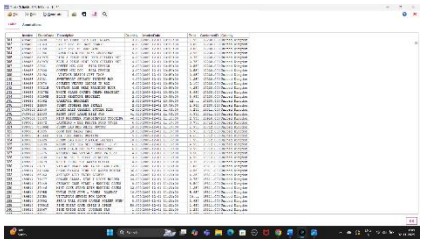
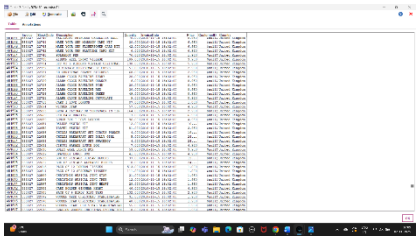
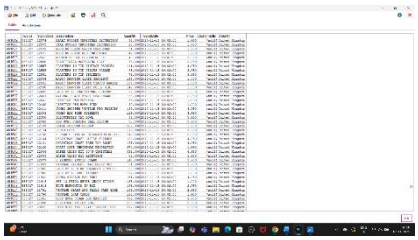
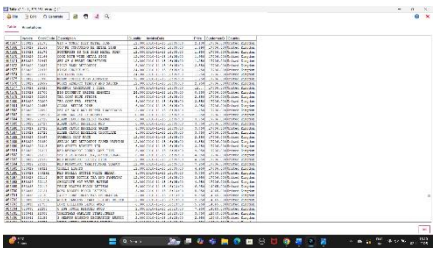
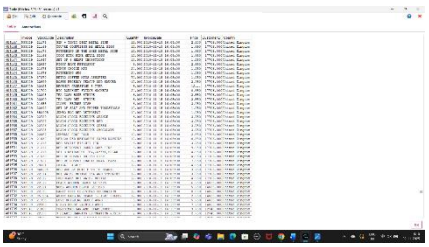
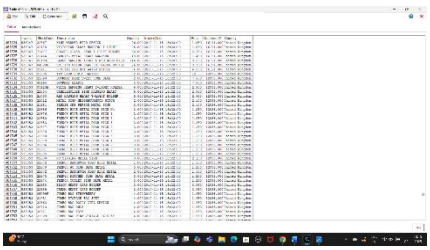
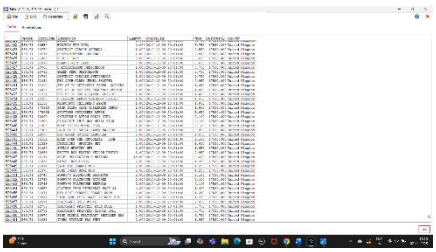
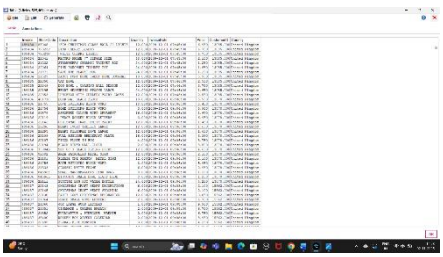
Roll No: 1230258220

Section: BCADS33

Dataset Used: Online Retail II Dataset (.xlsx)

Platform: Kaggle



CASE STUDY

Agenda/Definition: We work as a data miner for an online retail firm, where we have to clean, aggregate and transform the data to identify high-value customers and prepare the dataset for further modeling.

Learning: We will learn how to:

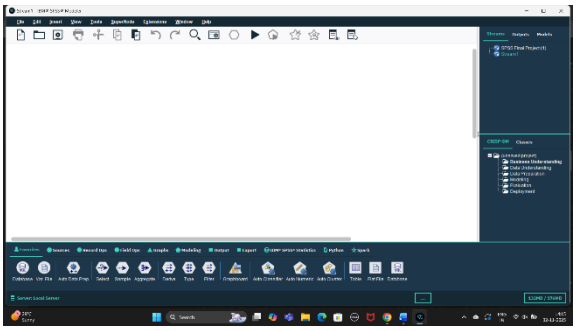
- Clean the dataset
- Aggregate the data
- Transform the data into relevant data

Required Tool: IBM SPSS Modeler 18.6

Dataset Used: Online Retail II Dataset (.xlsx)

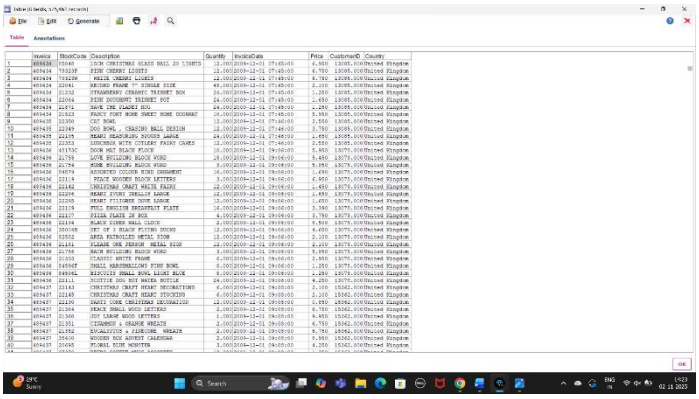
Working: This project analyzes customer purchasing patterns in an online retail store, where the aim is to identify high value customers and prepare the dataset for further modeling. To achieve this, follow the given steps:

STEP 1: Open IBM SPSS Modeler 18.6 in your PC.

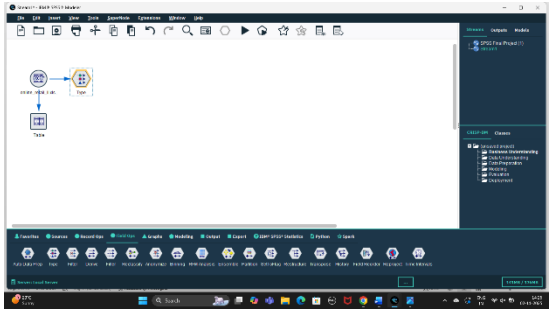


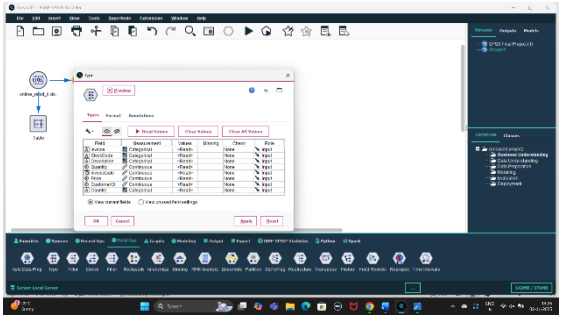
STEP 2: From the Sources Category, import the dataset named Online Retail II Dataset (.xlsx). This is an Excel dataset so, select the Excel Node.

STEP 3: Connect a Table from Output Category to Preview the dataset**.**

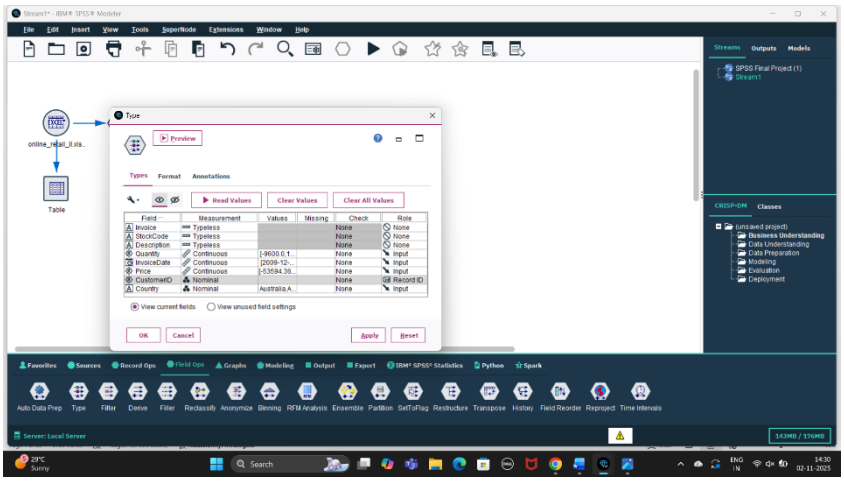


STEP 4: Connect the Type Node from Field Ops. This Node aids in understanding the different Fields Measurement Values.

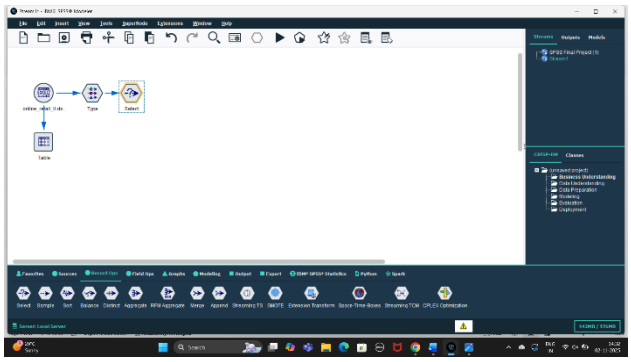


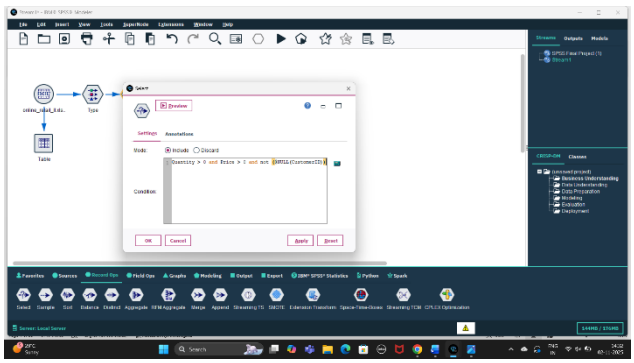


STEP 5: Rea the values to instantiate the different fields values. Now we will also change the role of some fields depending on our requirement. After the changes made, click on Apply and then on OK.

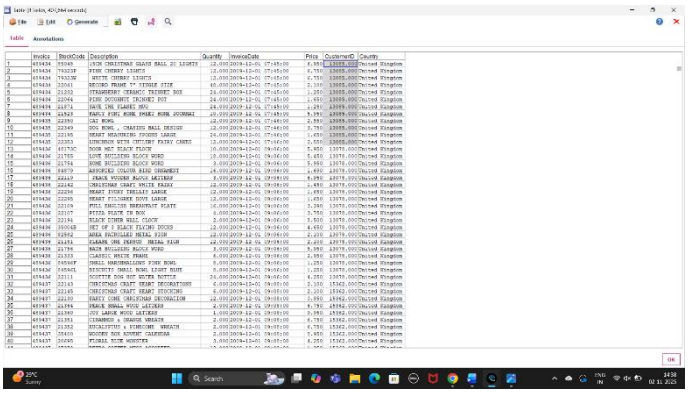


STEP 6: Connect A Select Node from Field Ops. This node allows us to include or discard the records. Type the following: Quantity > 0 and Price > 0 and not (@NULL(CustomerID)).

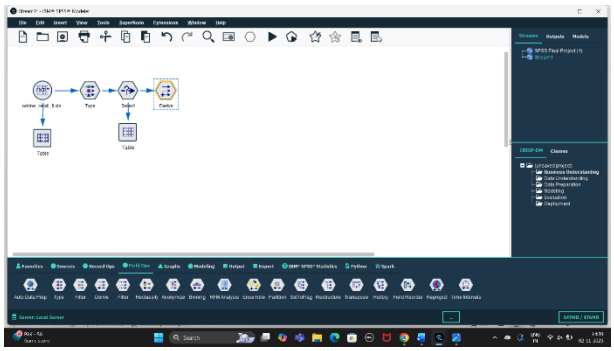


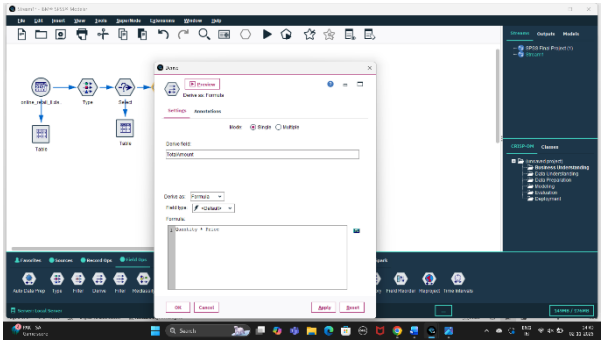


STEP 7: Connect a Table to see what changes your Select Node brings.

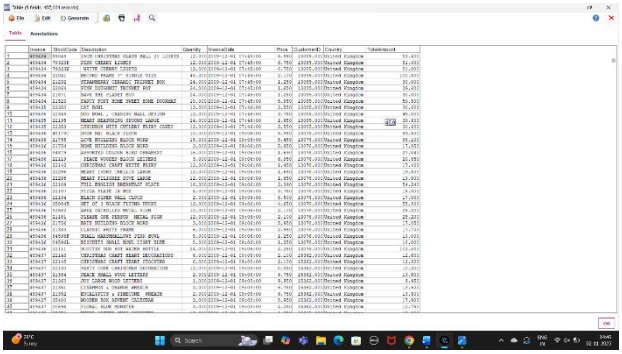


STEP 8: So now, we will create an additional filed named TotalAmount. To create a new field, we use Derive Node from Field Ops. Double click on the node and from the expression builder choose the formula: price*quantity. Click on Apply and OK.

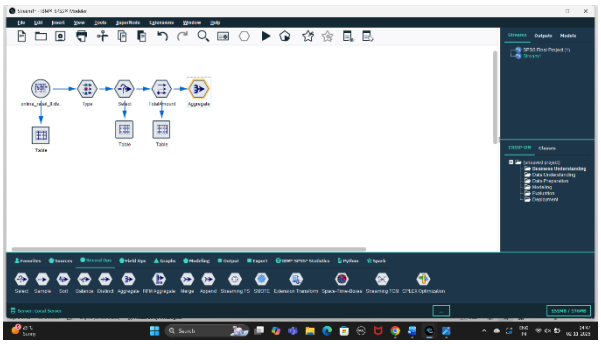


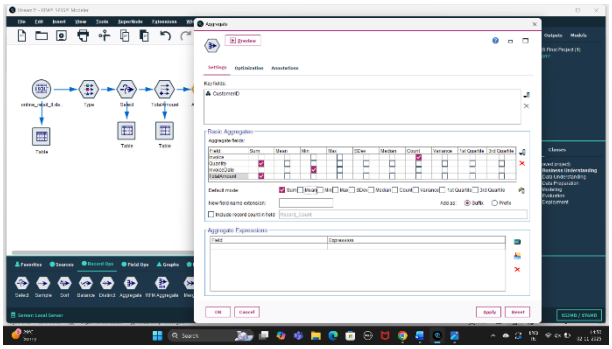


STEP 9: Connect a Table to see your new Field.

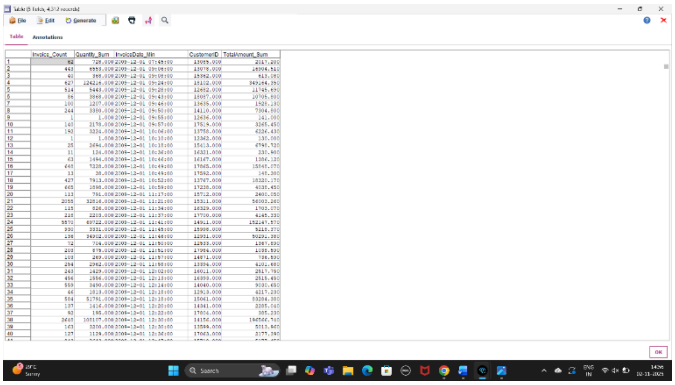


STEP 10: From the Record Ops, select Aggregate Node. This node combines multiple records into one row. Select the Key field as CUSTOMER_ID. In aggregate fileds choose Invoice, InvoiceDate, TotalAmount, Quantity.



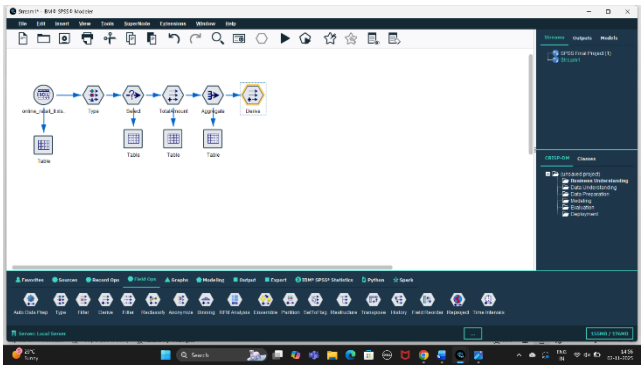


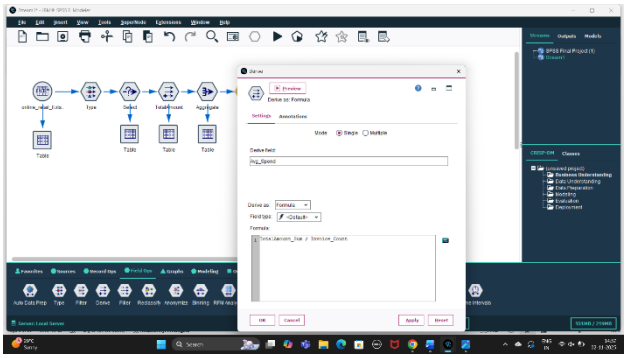
STEP 11: Connect a TABLE with this one as well.



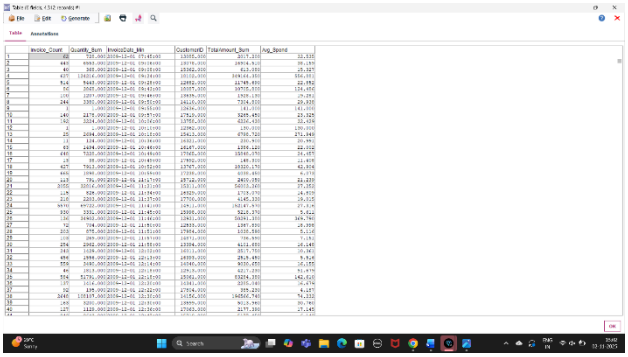
STEP 12: Choose one more Derive node. This time we will use it to derive a field called AVG_SPEND. From the Expression Builder choose the formula:

TotalAmount_Sum / Invoice_Count.

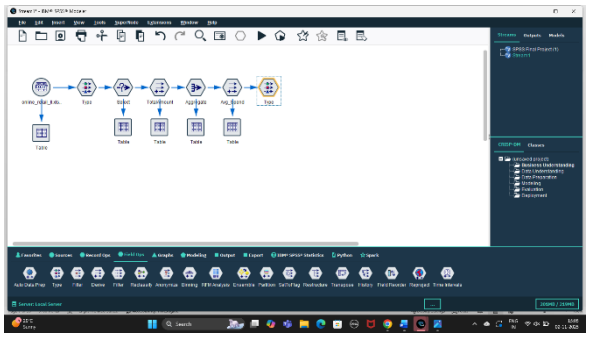


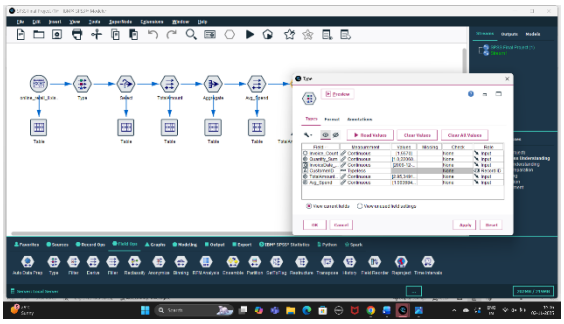


STEP 13: Connect a TABLE with this one also.

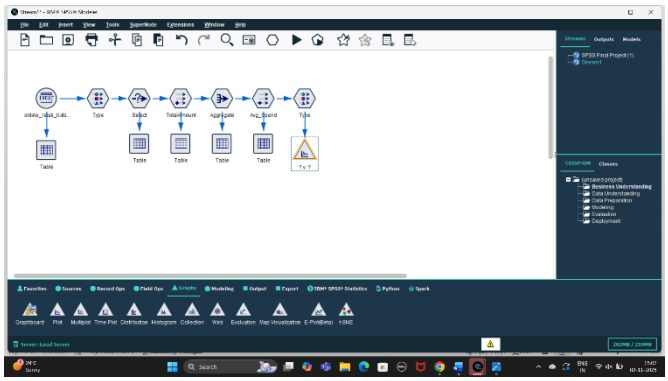


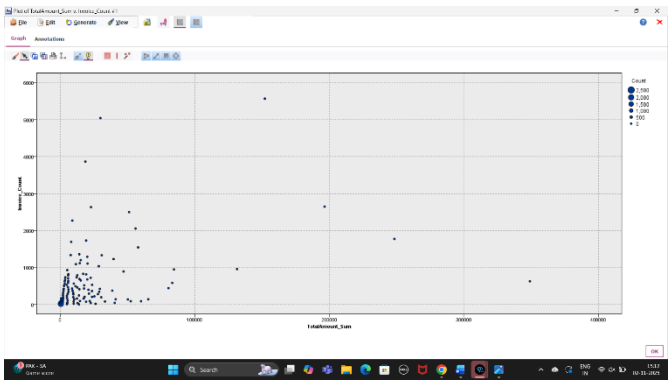
STEP 14: Connect the Type node. Read the values.



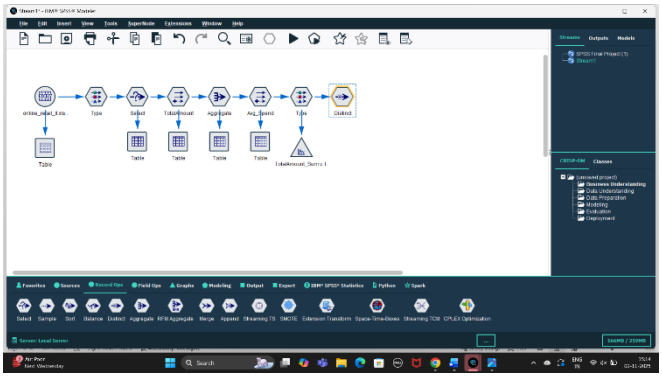


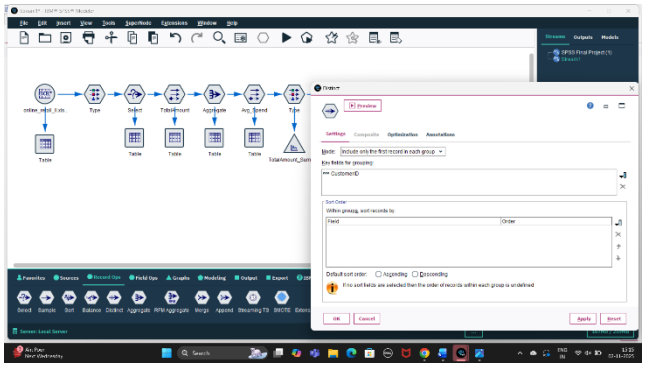
STEP 15: Connect a Plot graph from Graphs. Graphs aid in visualization and better understanding. Double click on it and give the X-axis and Y-axis TotalAmount_Sum and Invoice_Count respectively. Click on Apply and Run.



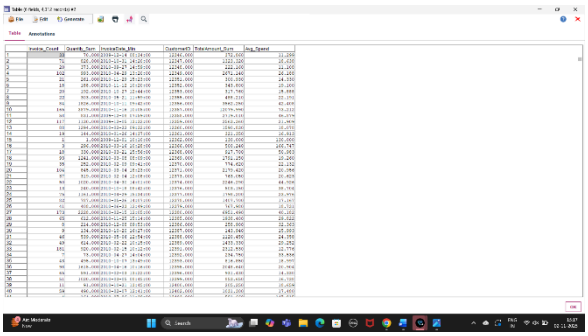


STEP 16: Select Distinct Node from Record Ops. Include only first record in each group should be chosen in Mode. In key fields for grouping choose CUSTOMER_ID. Click on Apply and OK.



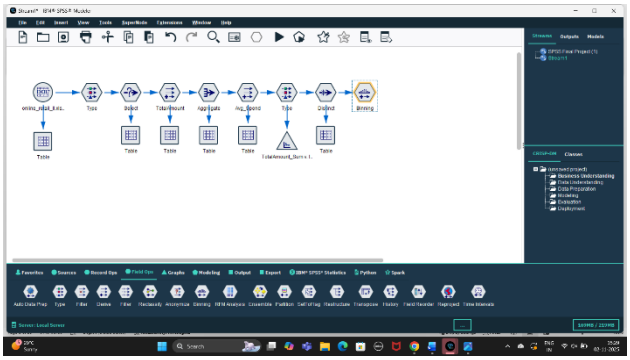


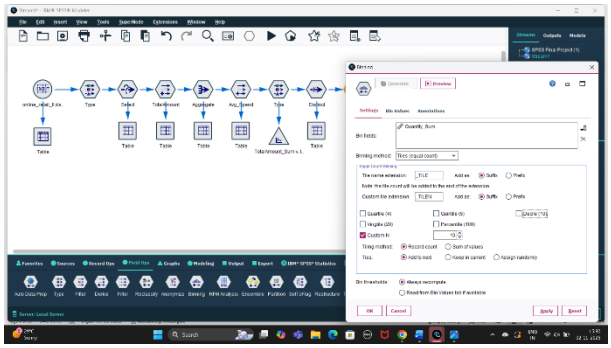
STEP 17: Connect a Table again to see how it works.



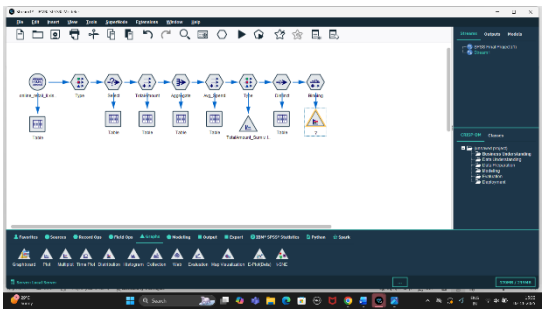
STEP 18: From Field Ops, choose Binning Node. This node bins the data into equal distributions.

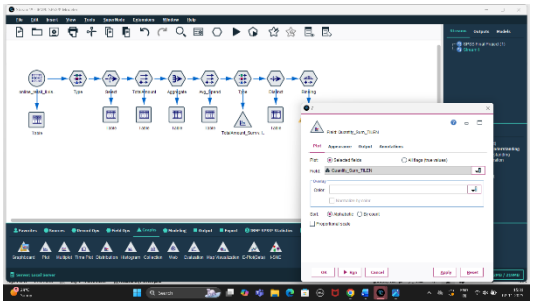
Double click on this node, select Binning method as Tiles and in binning fields choose Quantity_Sum.



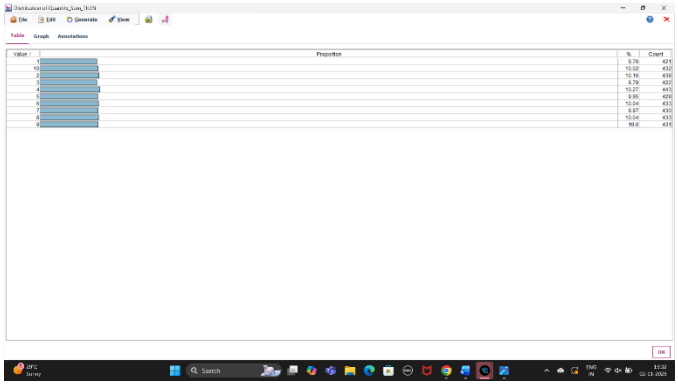


STEP 19: Choose Distribution node from Graphs. Choose Quantity_Sum_TILEN in fileds.

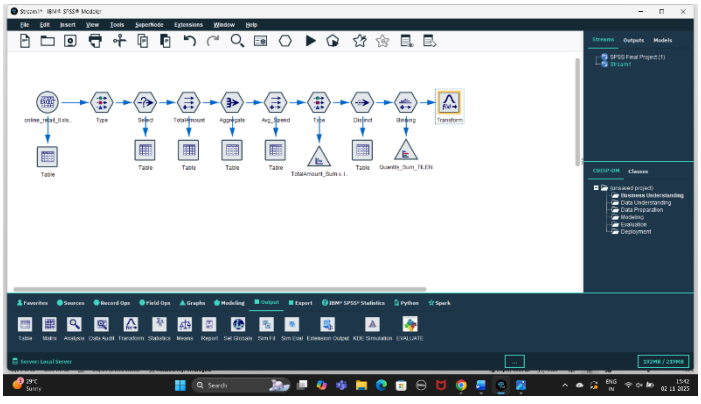


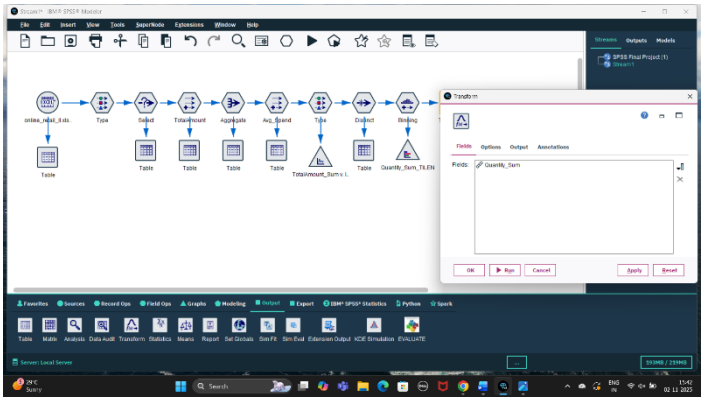


STEP 20: Click on Apply and Run.

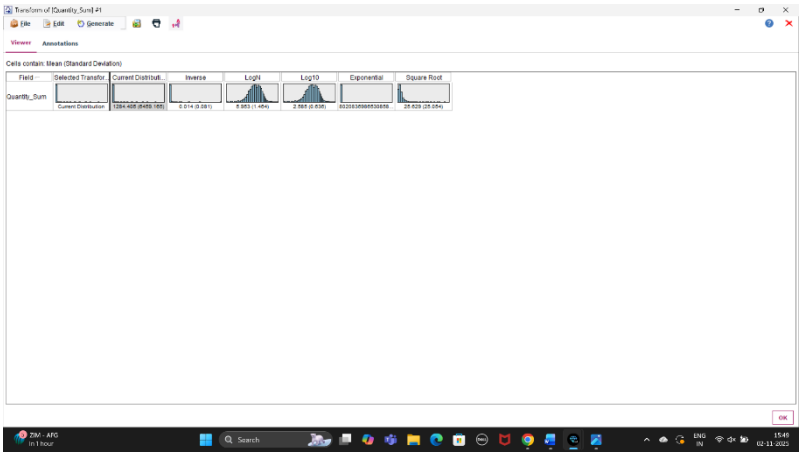


STEP 21: From Output Category choose Transform Node, this node will transform the data in a standard format. Double click on the node and choose Quantity_Sum in the fields.

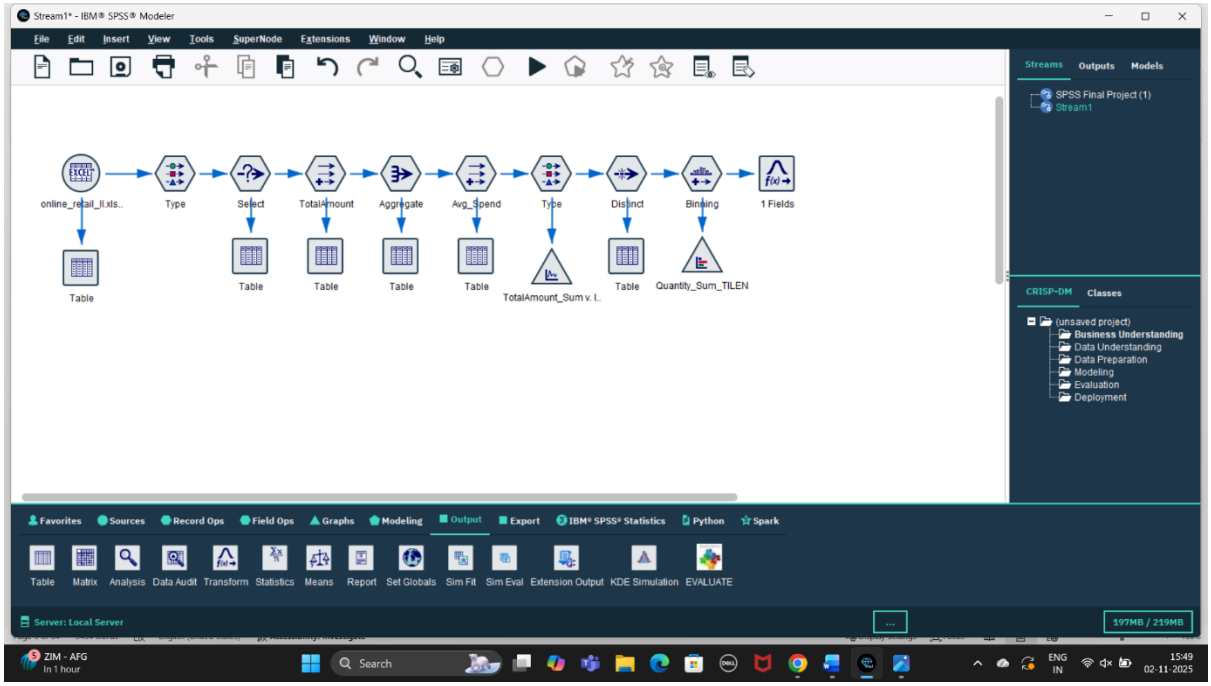




STEP 22: Click on Apply and Run.



STEP 23: The final flow of nodes looks like this:



This marks the end of our SPSS Case Study.

----------THANK YOU----------