Babu Banarasi Das University

Case Study
On
Data Mining And Integration Of Telecommunication dataset for Predicting Customer Churn Using SPSS

SUBMITTED TO:-
MR. ROBIN TYAGI

SUBMITTED BY:-
Gayatri Singh Manral
Garima Singh
Jyoti Rani

Telecom Customer Churn Analysis

Agenda/Definition: You work for a telecom company and need to combine, clean, and process datasets to create a solid base for building churn prediction models using SPSS.

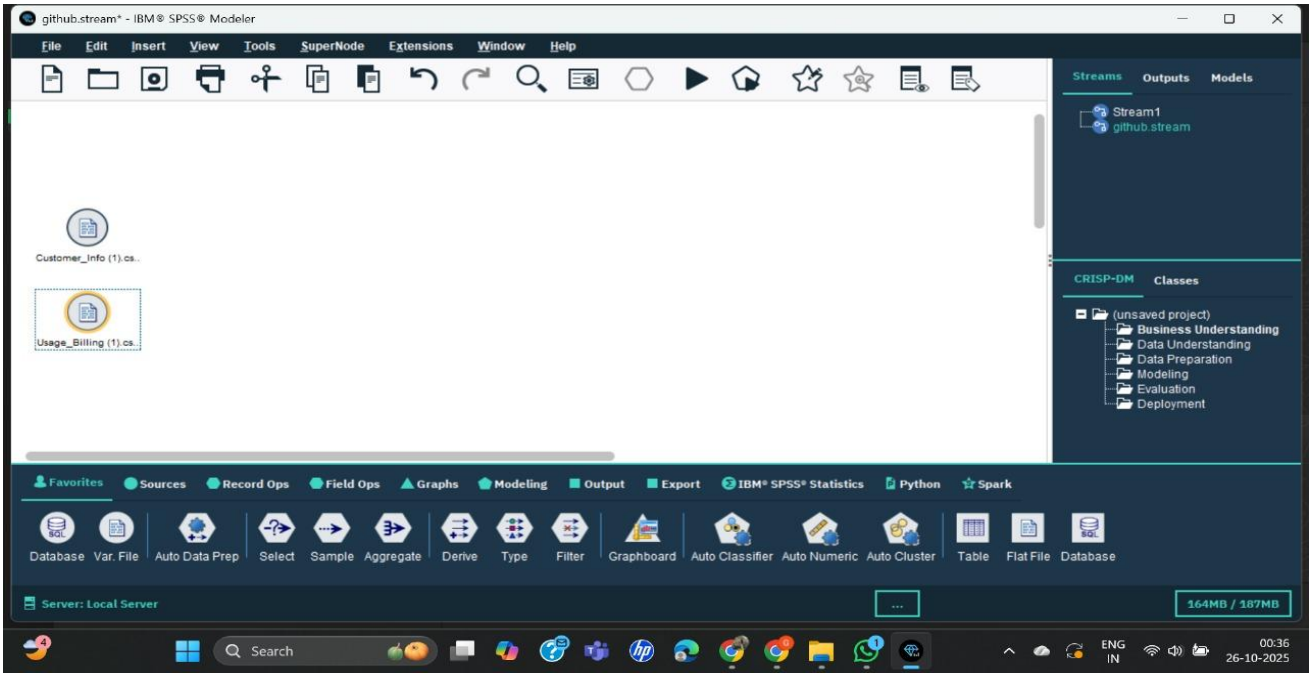
Datasets: 1. Customer_Info.csv: 2. Usage_Billing.csv

Outcomes/Learning: This practical teach how to clean, merge, and prepare telecom data for accurate churn prediction in SPSS.

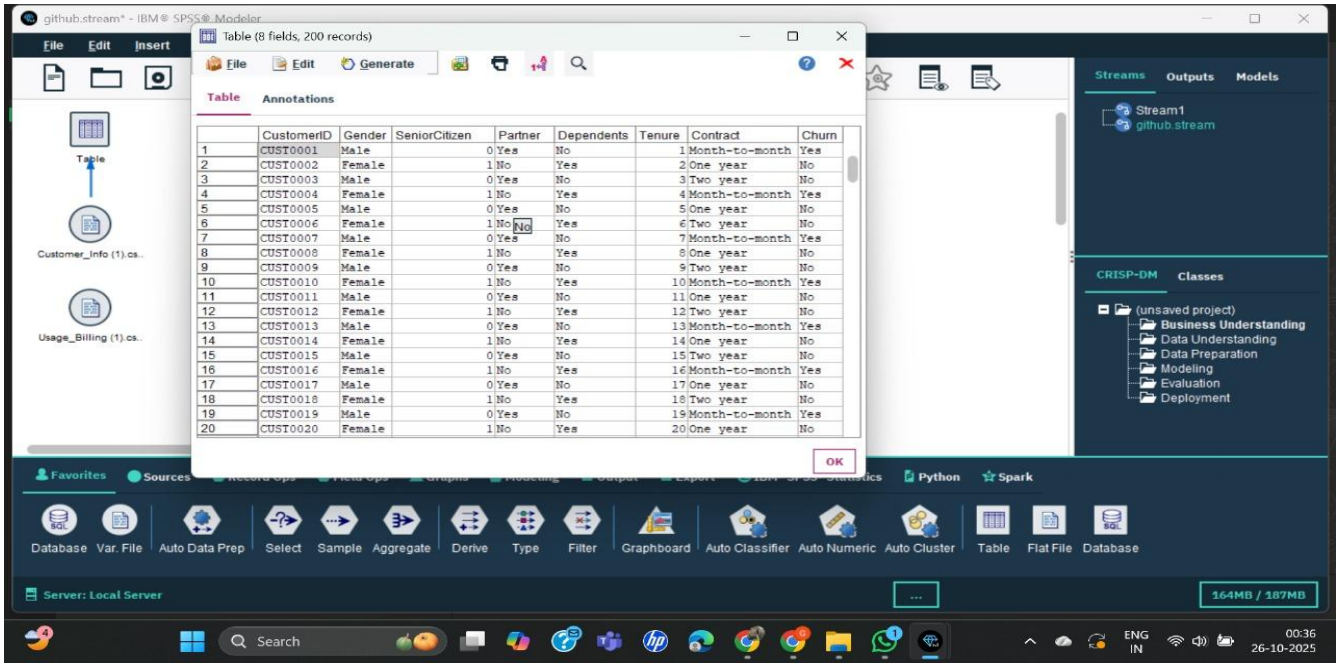
Required Tool: IBM SPSS Modeler tool.

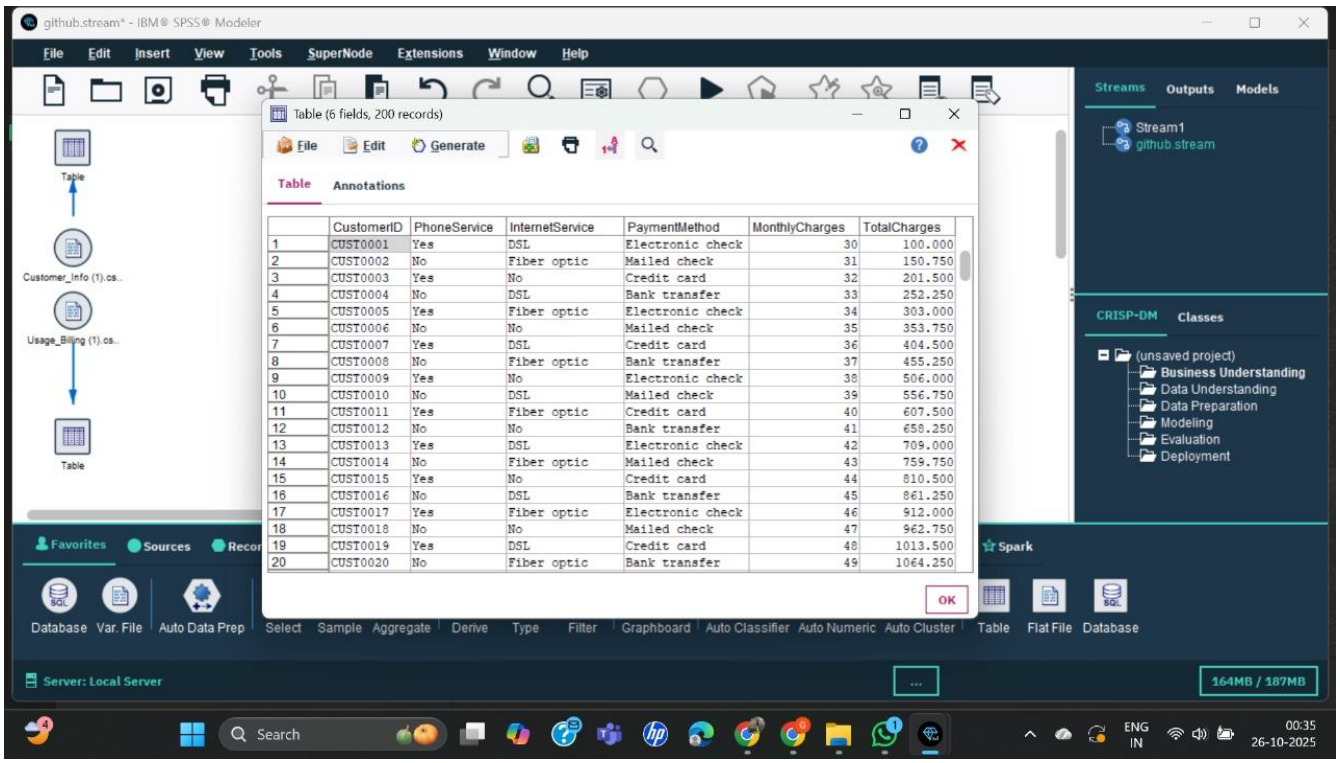
Working: You remove errors, combine datasets, and organize data for model building

Step1: Open IBM SPSS Modular, go to the Source tab and add two Excel node to the canvas and import datasets. **(Customer_Info.csv and Usage_Billing.csv)**

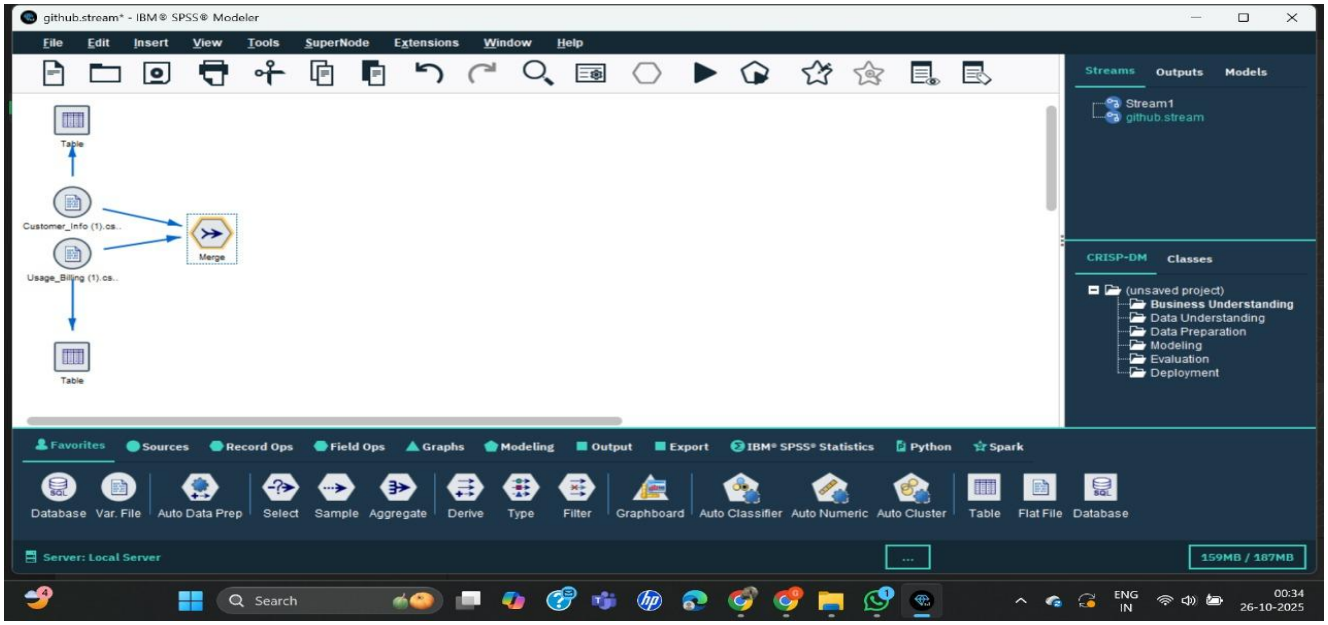


Step2 : From the Output tab, drag the Table node and connect it to see the results in table form

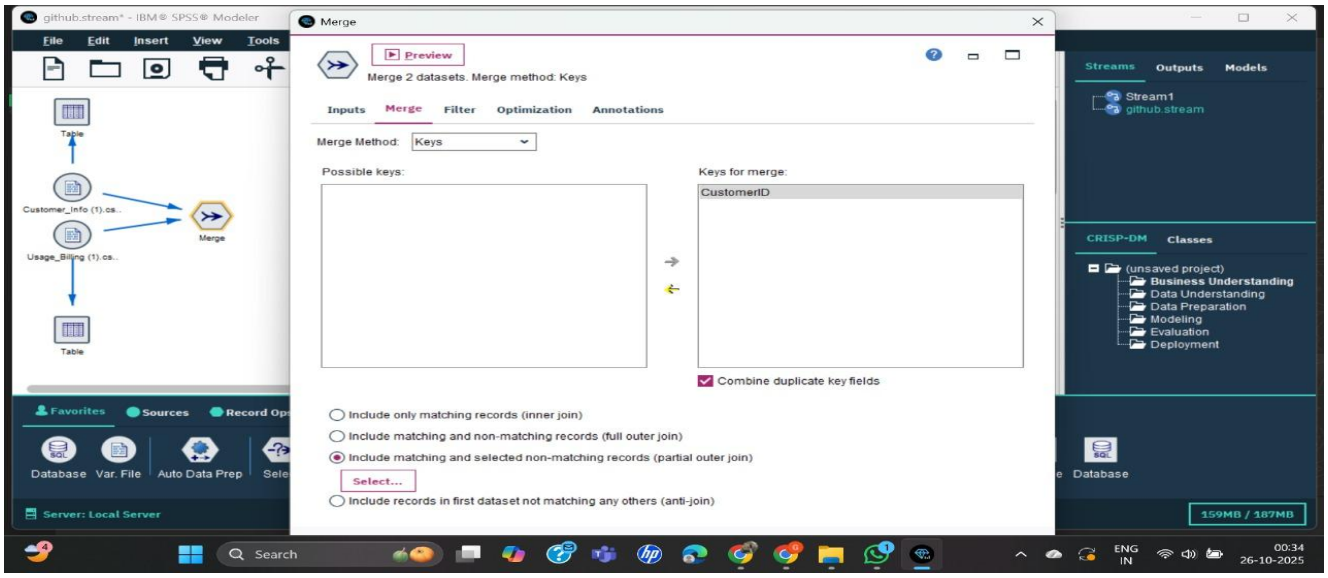




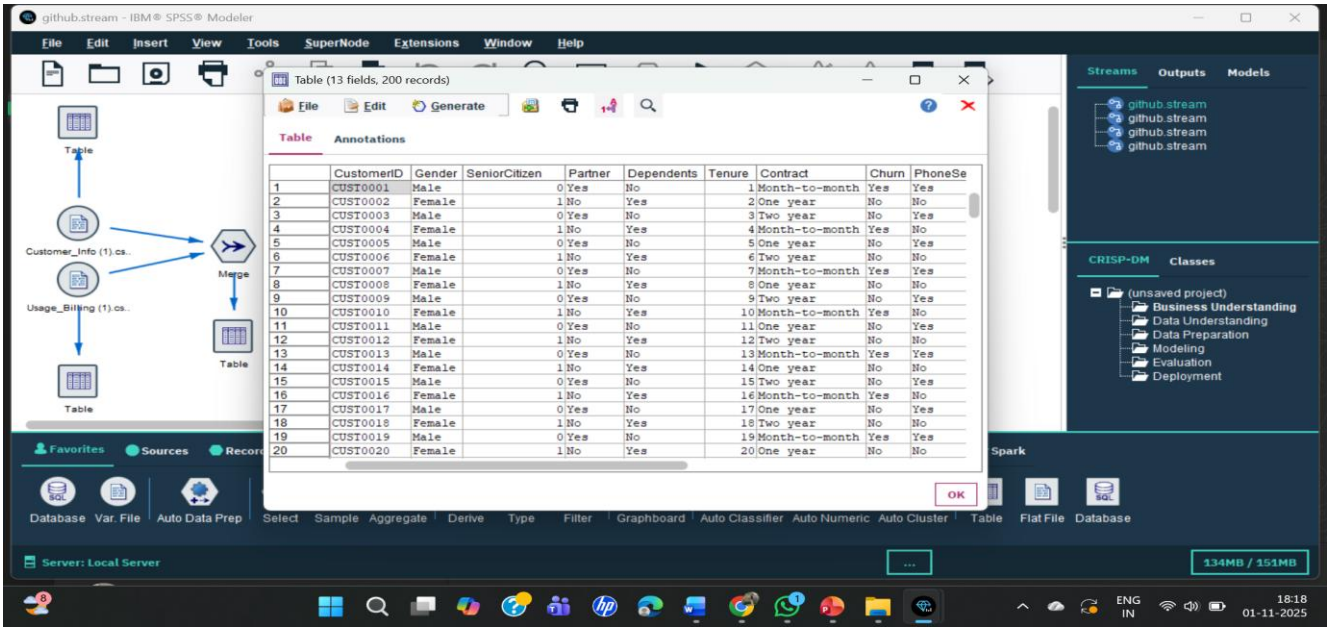
Step 3:- Now we go to the Record Ops and select Merge node then connect the Merge node with the both datasets. These helps to combines both dataset using the **common key field**(Customer ID)



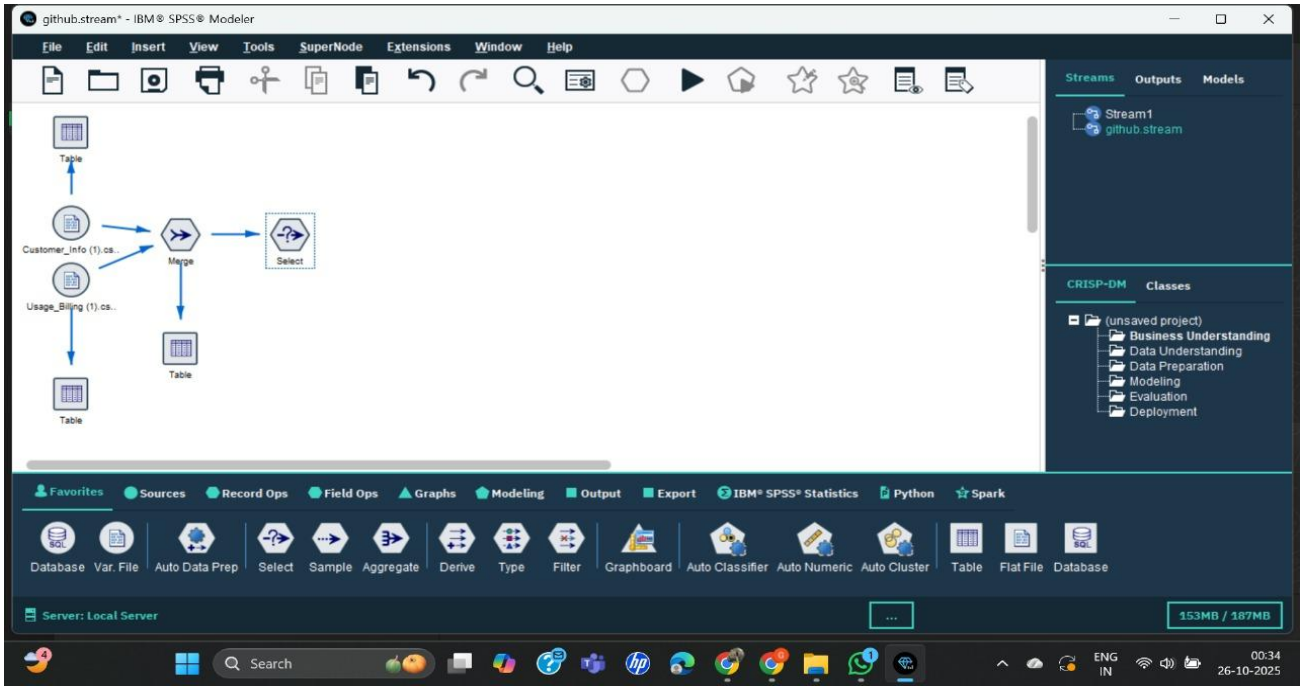
Step 4:- After connecting the Merge node double click on merge node and select merge method from the merge method dropdown, choose **keys.** Then from the **"Possible keys"** list, select **CustomerID** and move it to **"Keys for Merge".** After that choose Merge type where you select include matching and selected non-matching recoreds(**partial outer join**).



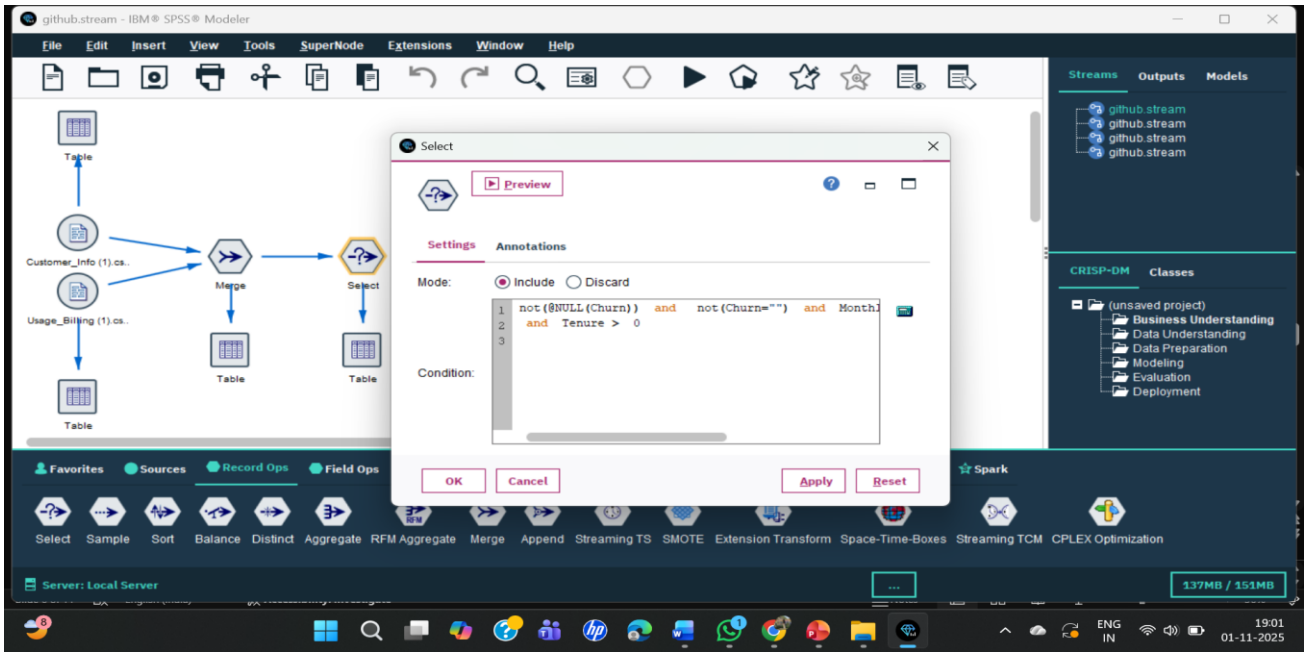
Step 5:- Now go to the output tab and select table node for see the dataset.



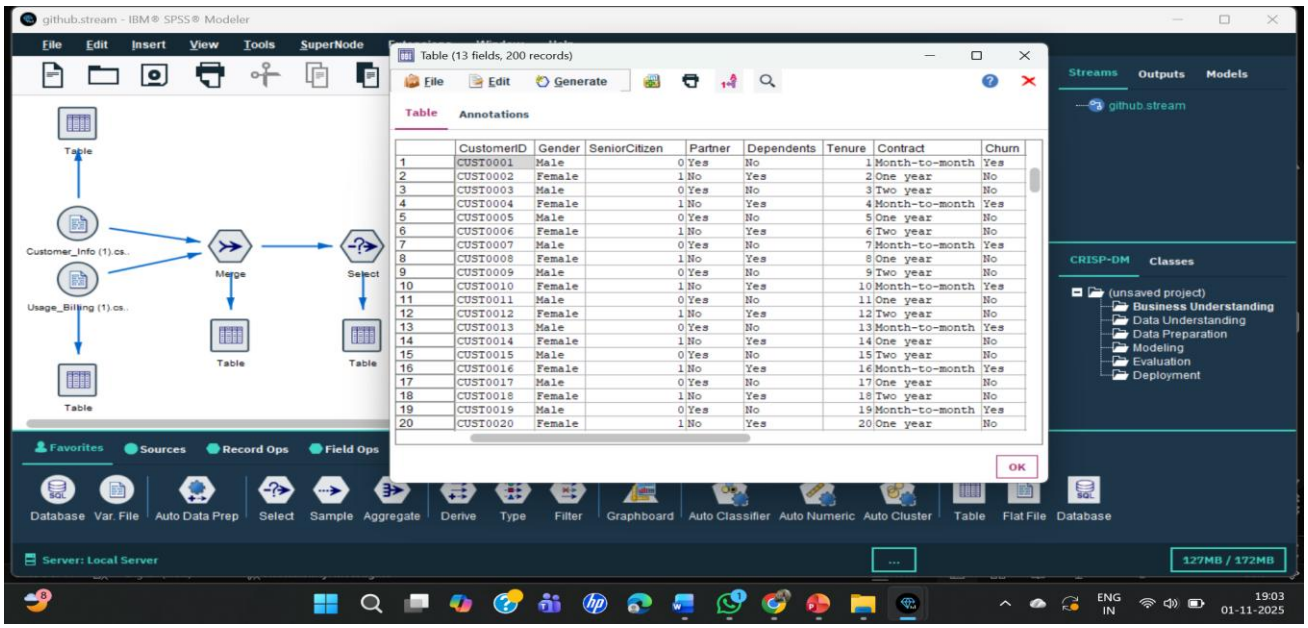
Step 6:- After that add Select node from the record option and connect with the Merge node



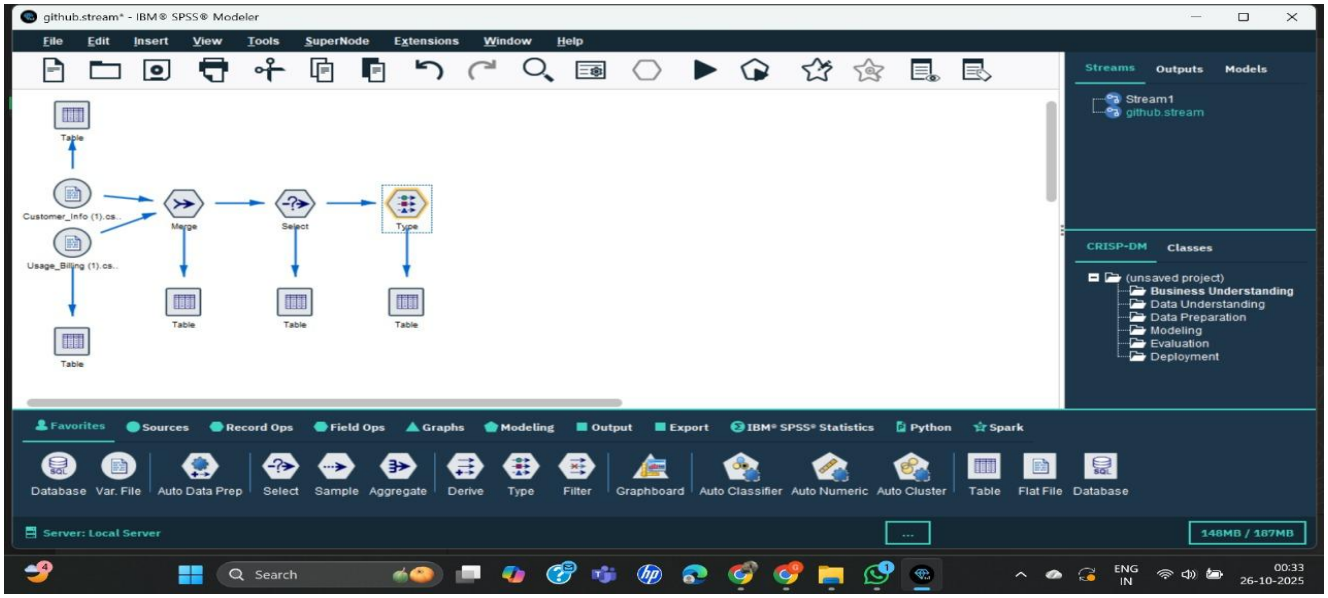
Step 7:- Now Double click on select node and filter the dataset by applying the condition **not(@NULL(Churn)) and not(Churn="") and MonthlyCharges > 0 and TotalCharges > 0 and Tenure > 0** " to keep only valid records.



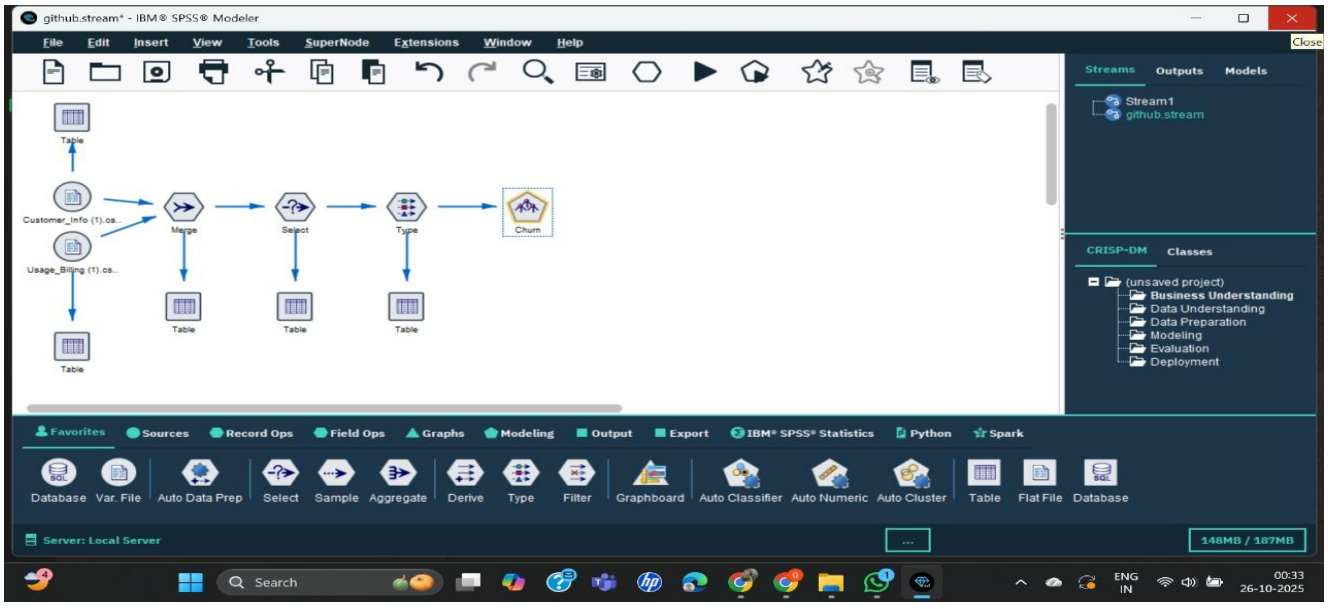
Step 8:- Now connect the tabel to see the data is running or not.



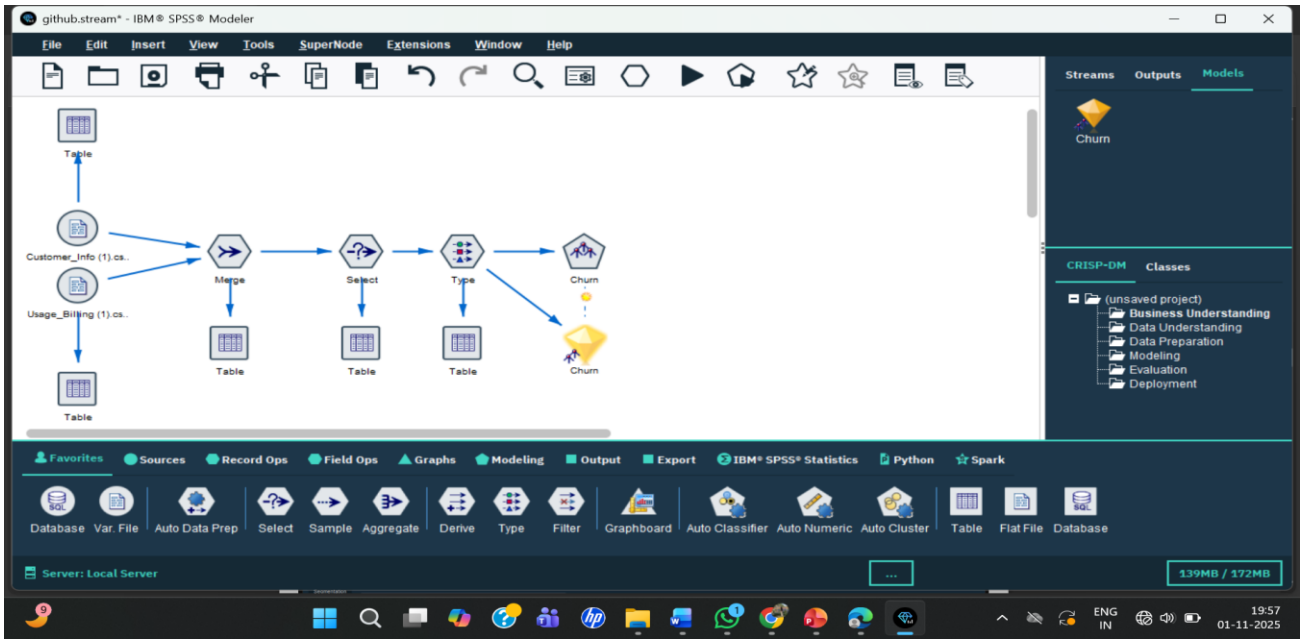
Step 9:-From the Field tab, drag the Type node and connect it to the select node. In the Type node, choose all the field as input such as gender, senior citizen, partner, etc.. except customer id and set churn as the target variable , and add the table node to see the output



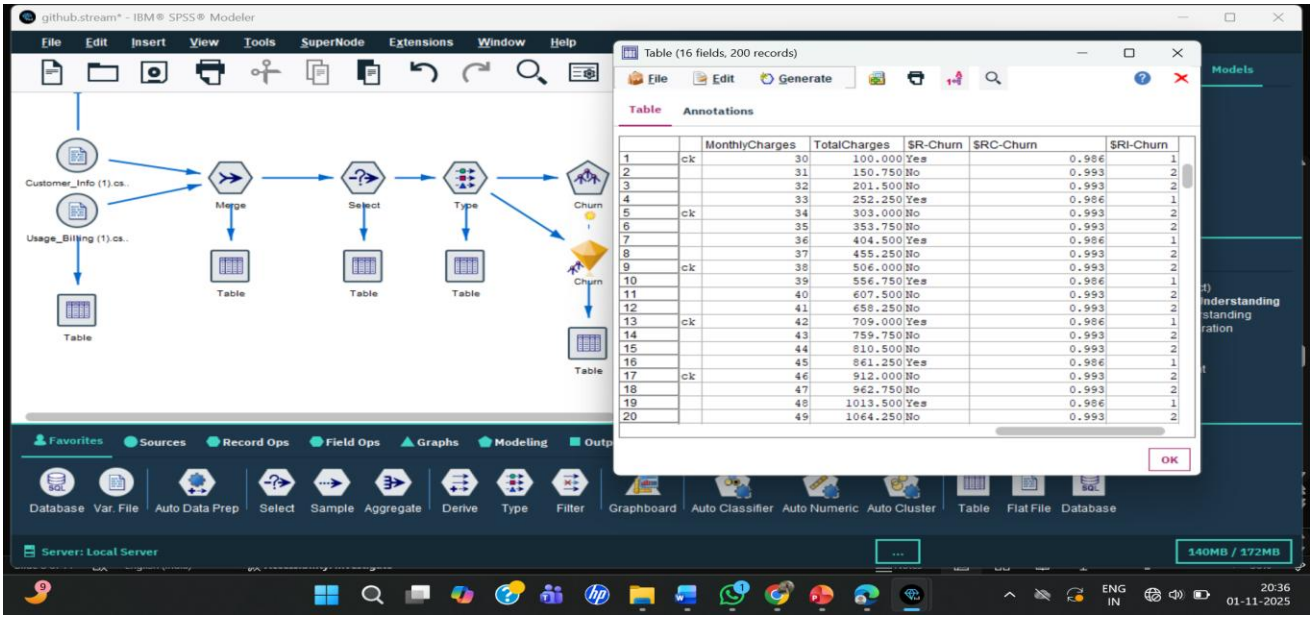
Step10:- now from the modelling category add churn node We used this to apply that value which we select as input in type node and target value.



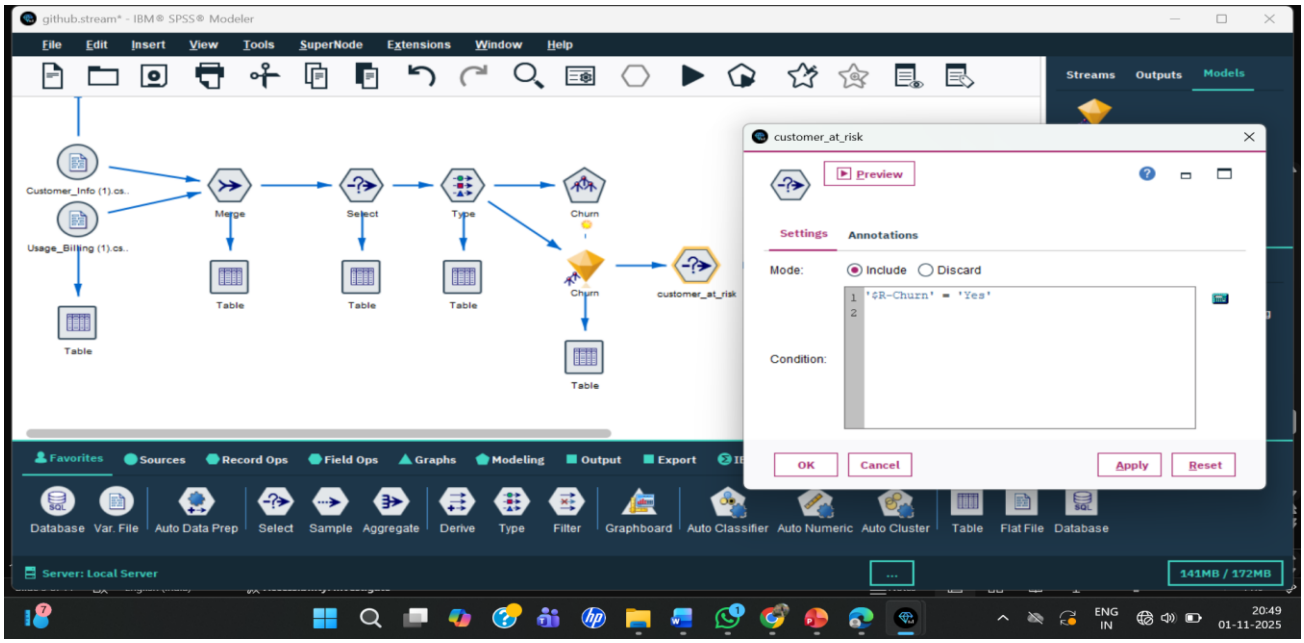
Step11:- After run the Chaid you can see the nugget on your canvas and linked to the Chaid itself



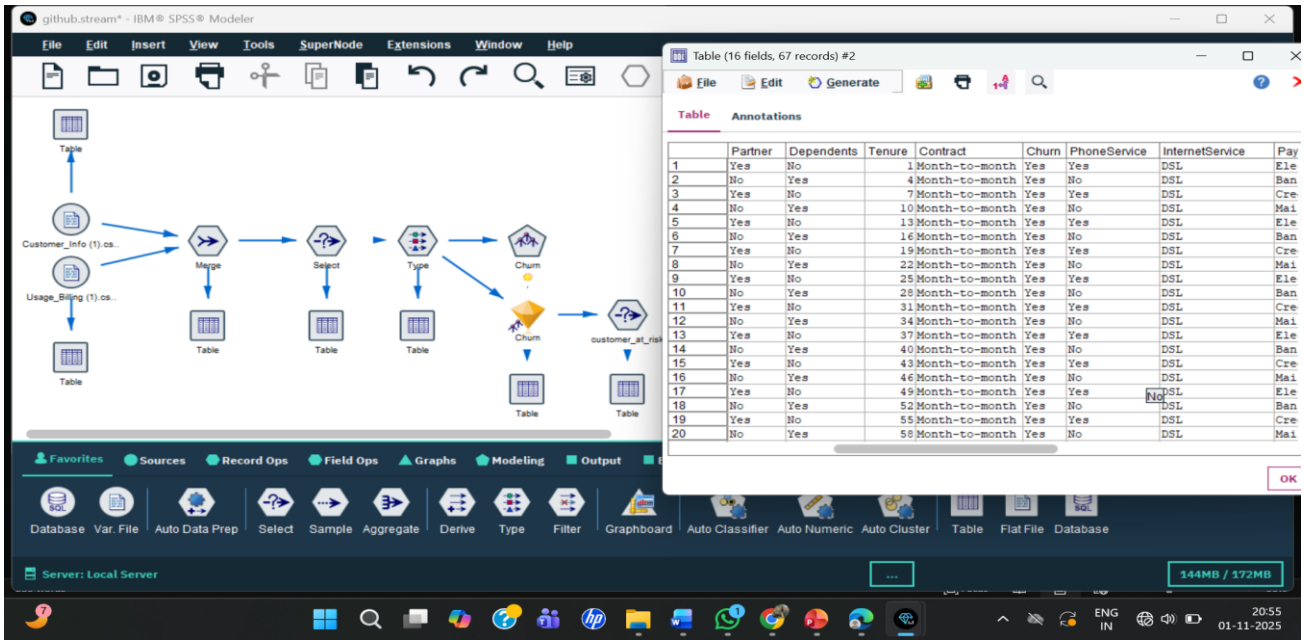
Step 12:- Now add table to check the churn we created is showing a churn fields or not like $R-churn, $RC-churn, $RI-churn.



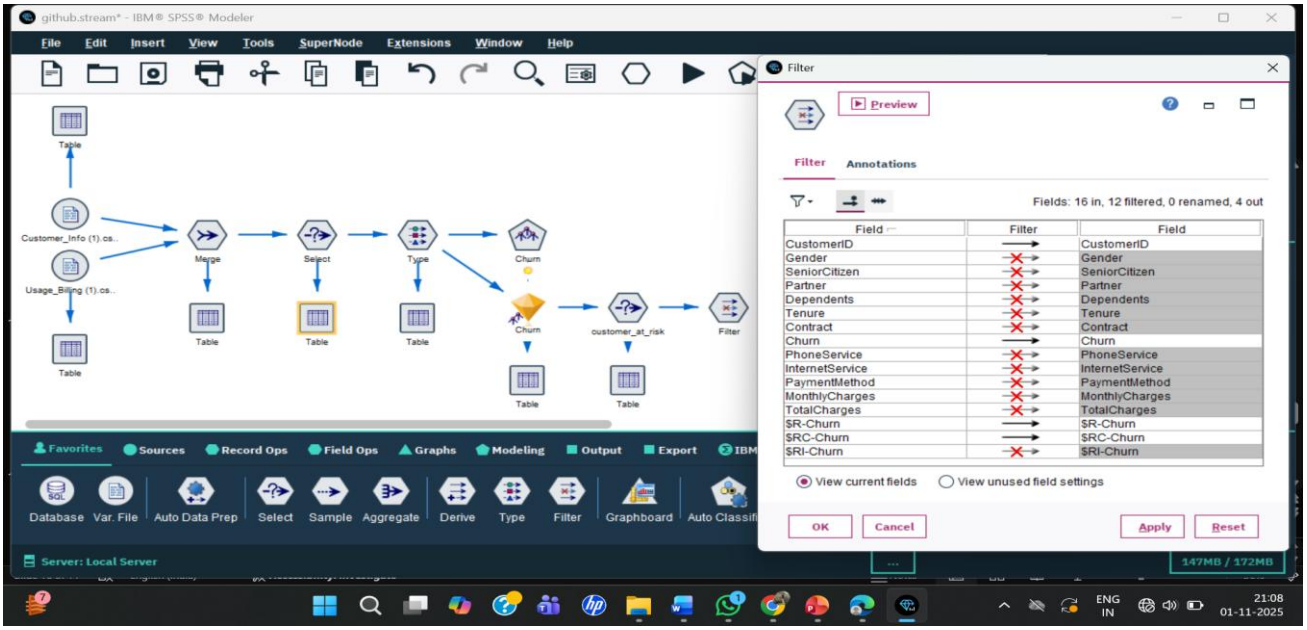
Step13:- Now we go to the Record Ops and click on select node rename the node as **Customer_at_Risk** and write query **'$R-Churn' = 'Yes'**



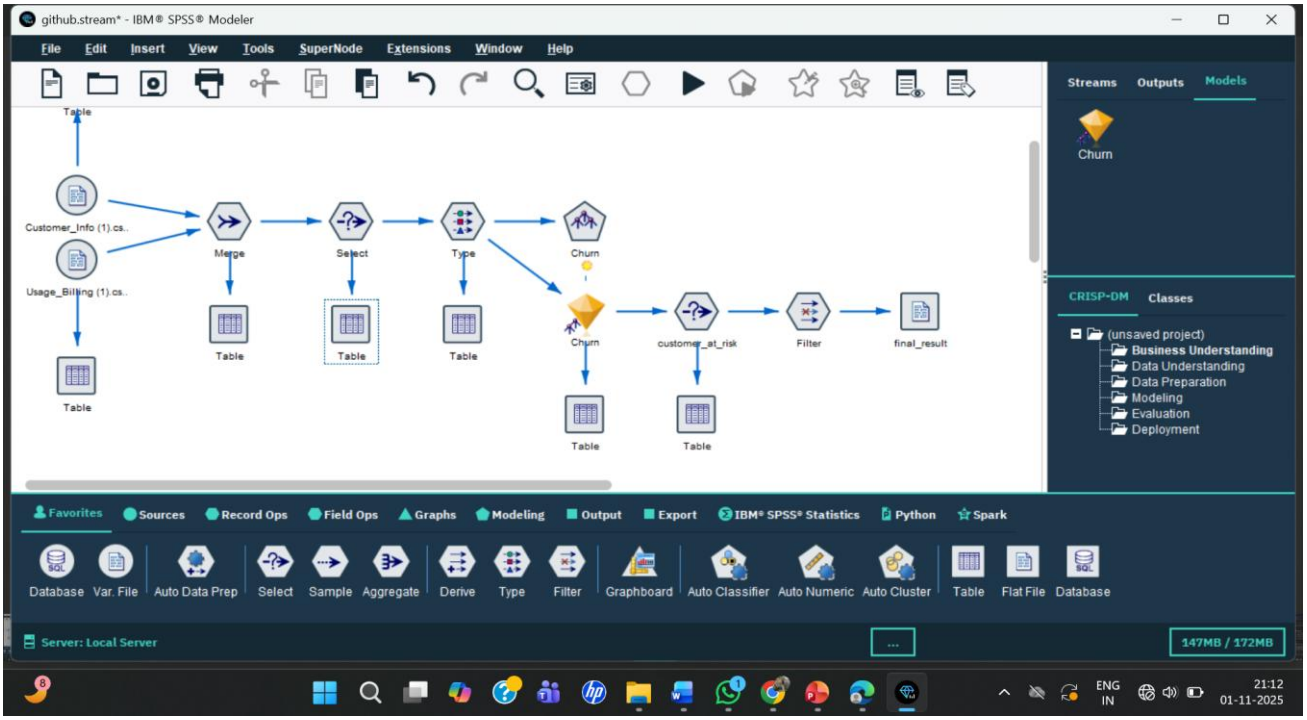
Step14:- Now connect a tabel node and see the customers ones who are most likely to Churn… high-risk(Churn=" YES") customers



Step15:- We go to field category to add the filter node to filter the data means which field we want to see as final output



Step 16:- Now connect a flat file from the export to see the final result . In final result we give the path of data and apply and save it, by this we can see and read the data in notepad , word , etc



Step 17:- Here our final output of our data set in notepad