

# **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.

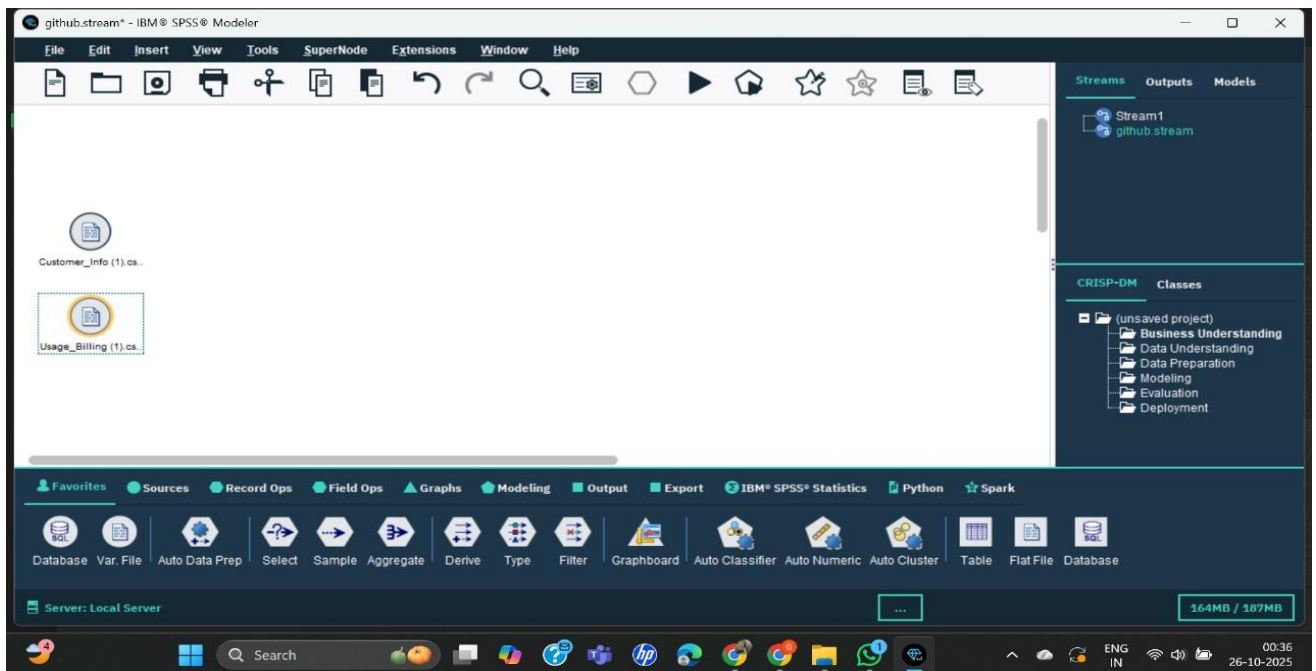
**Datasets:** 1. Customer\_Info.csv: 2. Usage\_Billing.csv

**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

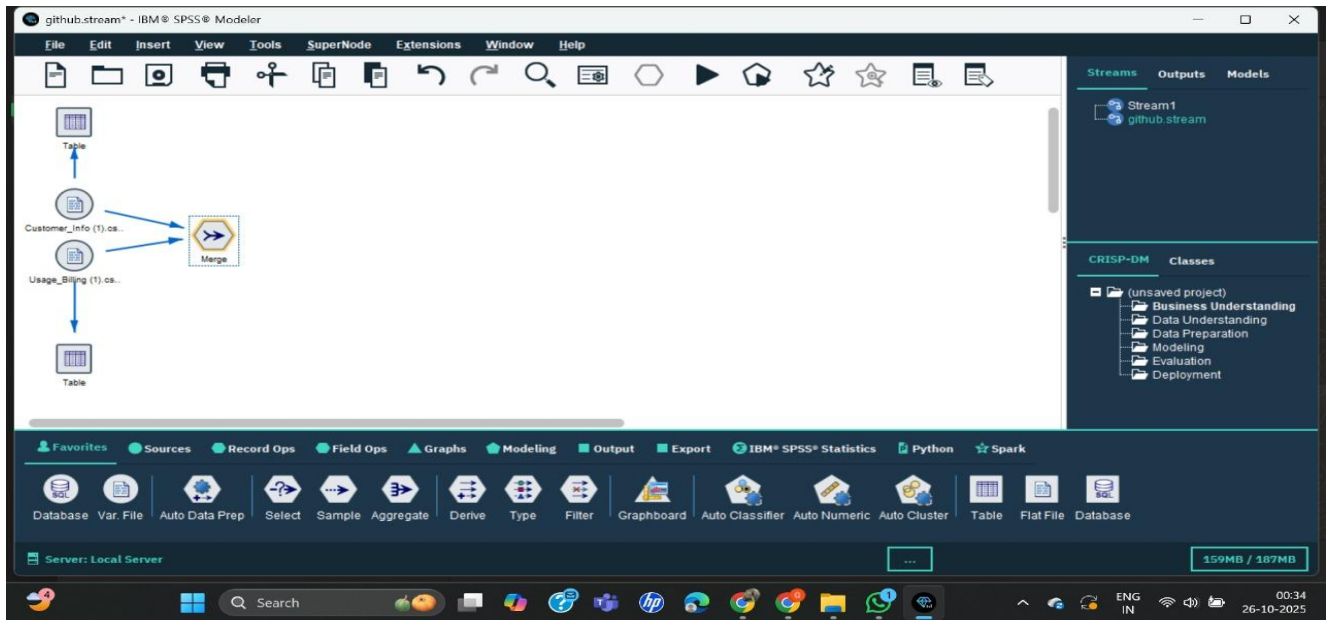
The screenshot shows the IBM SPSS Modeler interface. A 'Table' node output window is open, displaying a table with 8 fields and 200 records. The table columns are: CustomerID, Gender, SeniorCitizen, Partner, Dependents, Tenure, Contract, and Churn. The data is listed for 20 customers (CUST0001 to CUST0020).

	CustomerID	Gender	SeniorCitizen	Partner	Dependents	Tenure	Contract	Churn
1	CUST0001	Male	0	Yes	No	1	Month-to-month	Yes
2	CUST0002	Female	1	No	Yes	2	One year	No
3	CUST0003	Male	0	Yes	No	3	Two year	No
4	CUST0004	Female	1	No	Yes	4	Month-to-month	Yes
5	CUST0005	Male	0	Yes	No	5	One year	No
6	CUST0006	Female	1	No	Yes	6	Two year	No
7	CUST0007	Male	0	Yes	No	7	Month-to-month	Yes
8	CUST0008	Female	1	No	Yes	8	One year	No
9	CUST0009	Male	0	Yes	No	9	Two year	No
10	CUST0010	Female	1	No	Yes	10	Month-to-month	Yes
11	CUST0011	Male	0	Yes	No	11	One year	No
12	CUST0012	Female	1	No	Yes	12	Two year	No
13	CUST0013	Male	0	Yes	No	13	Month-to-month	Yes
14	CUST0014	Female	1	No	Yes	14	One year	No
15	CUST0015	Male	0	Yes	No	15	Two year	No
16	CUST0016	Female	1	No	Yes	16	Month-to-month	Yes
17	CUST0017	Male	0	Yes	No	17	One year	No
18	CUST0018	Female	1	No	Yes	18	Two year	No
19	CUST0019	Male	0	Yes	No	19	Month-to-month	Yes
20	CUST0020	Female	1	No	Yes	20	One year	No

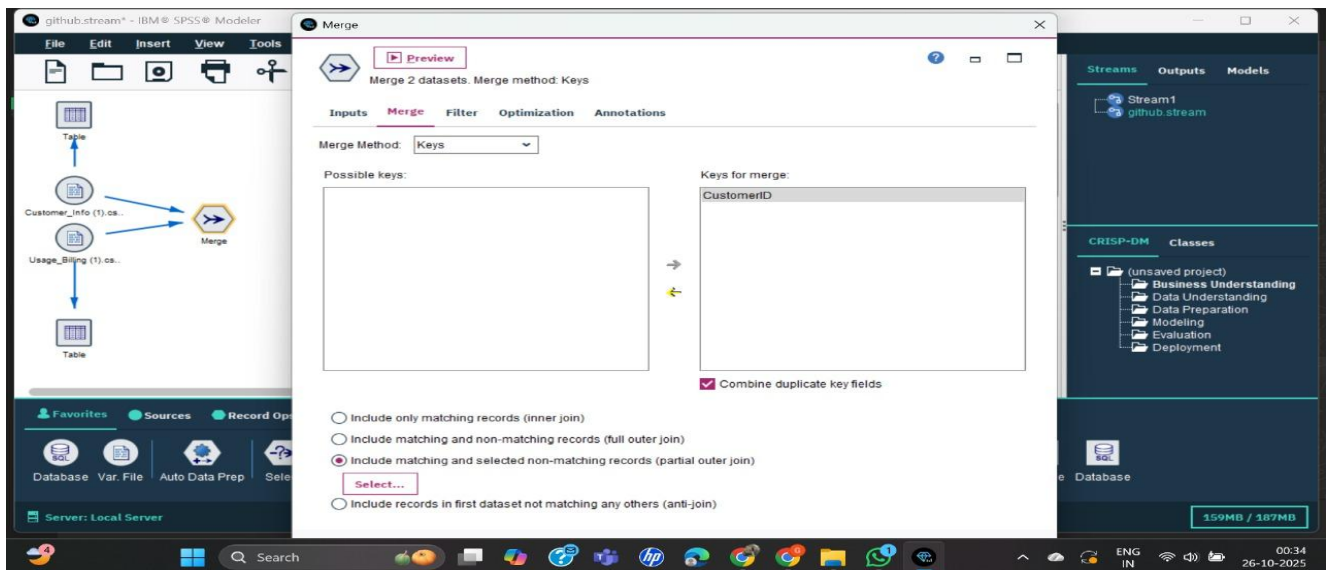
The screenshot shows the IBM SPSS Modeler interface. A 'Table' node output window is open, displaying a table with 6 fields and 200 records. The table columns are: CustomerID, PhoneService, InternetService, PaymentMethod, MonthlyCharges, and TotalCharges. The data is listed for 20 customers (CUST0001 to CUST0020).

	CustomerID	PhoneService	InternetService	PaymentMethod	MonthlyCharges	TotalCharges
1	CUST0001	Yes	DSL	Electronic check	30	100.000
2	CUST0002	No	Fiber optic	Mailed check	31	150.750
3	CUST0003	Yes	No	Credit card	32	201.500
4	CUST0004	No	DSL	Bank transfer	33	252.250
5	CUST0005	Yes	Fiber optic	Electronic check	34	303.000
6	CUST0006	No	No	Mailed check	35	353.750
7	CUST0007	Yes	DSL	Credit card	36	404.500
8	CUST0008	No	Fiber optic	Bank transfer	37	455.250
9	CUST0009	Yes	No	Electronic check	38	506.000
10	CUST0010	No	DSL	Mailed check	39	556.750
11	CUST0011	Yes	Fiber optic	Credit card	40	607.500
12	CUST0012	No	No	Bank transfer	41	658.250
13	CUST0013	Yes	DSL	Electronic check	42	709.000
14	CUST0014	No	Fiber optic	Mailed check	43	759.750
15	CUST0015	Yes	No	Credit card	44	810.500
16	CUST0016	No	DSL	Bank transfer	45	861.250
17	CUST0017	Yes	Fiber optic	Electronic check	46	912.000
18	CUST0018	No	No	Mailed check	47	962.750
19	CUST0019	Yes	DSL	Credit card	48	1013.500
20	CUST0020	No	Fiber optic	Bank transfer	49	1064.250

**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 records(**partial outer join**).





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

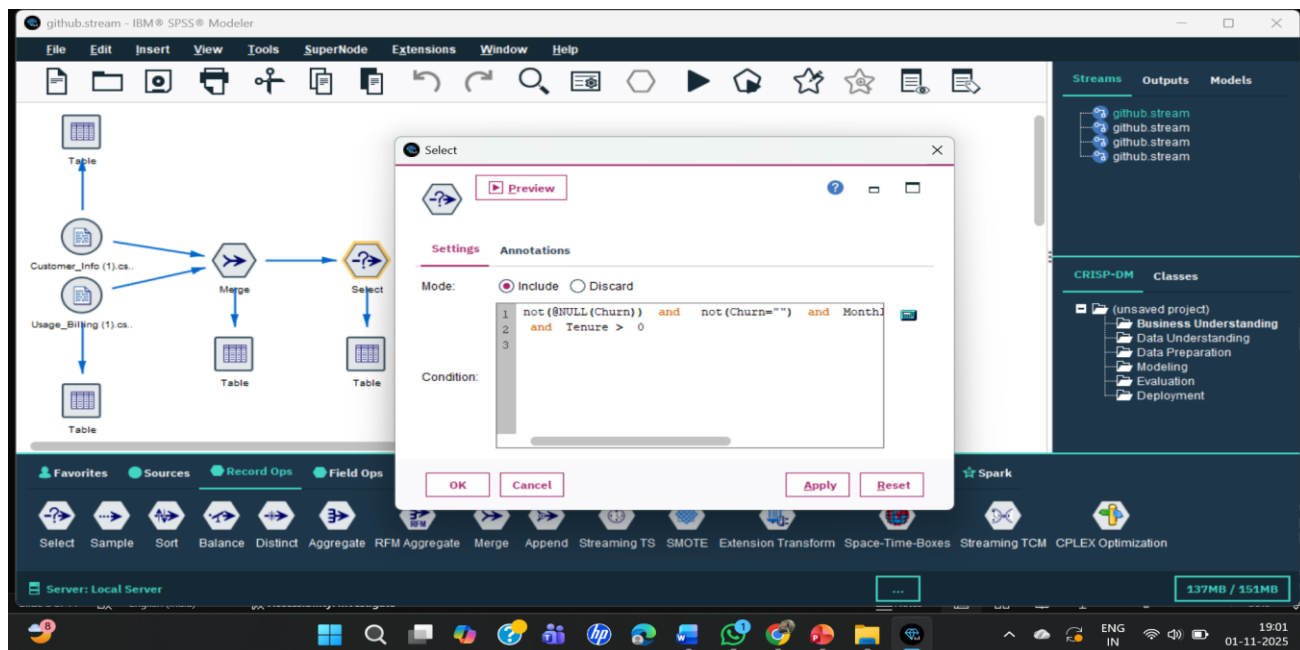
The screenshot shows the IBM SPSS Modeler interface. A 'Table' node is selected, displaying a dataset with 13 fields and 200 records. The dataset is a customer information table with the following columns: CustomerID, Gender, SeniorCitizen, Partner, Dependents, Tenure, Contract, Churn, and PhoneSe. The data is presented in a table format with 20 rows visible.

	CustomerID	Gender	SeniorCitizen	Partner	Dependents	Tenure	Contract	Churn	PhoneSe
1	CUST0001	Male	0	Yes	No	1	Month-to-month	Yes	Yes
2	CUST0002	Female	1	No	Yes	2	One year	No	No
3	CUST0003	Male	0	Yes	No	3	Two year	No	Yes
4	CUST0004	Female	1	No	Yes	4	Month-to-month	Yes	No
5	CUST0005	Male	0	Yes	No	5	One year	No	Yes
6	CUST0006	Female	1	No	Yes	6	Two year	No	No
7	CUST0007	Male	0	Yes	No	7	Month-to-month	Yes	Yes
8	CUST0008	Female	1	No	Yes	8	One year	No	No
9	CUST0009	Male	0	Yes	No	9	Two year	No	Yes
10	CUST0010	Female	1	No	Yes	10	Month-to-month	Yes	No
11	CUST0011	Male	0	Yes	No	11	One year	No	Yes
12	CUST0012	Female	1	No	Yes	12	Two year	No	No
13	CUST0013	Male	0	Yes	No	13	Month-to-month	Yes	Yes
14	CUST0014	Female	1	No	Yes	14	One year	No	No
15	CUST0015	Male	0	Yes	No	15	Two year	No	Yes
16	CUST0016	Female	1	No	Yes	16	Month-to-month	Yes	No
17	CUST0017	Male	0	Yes	No	17	One year	No	Yes
18	CUST0018	Female	1	No	Yes	18	Two year	No	No
19	CUST0019	Male	0	Yes	No	19	Month-to-month	Yes	Yes
20	CUST0020	Female	1	No	Yes	20	One year	No	No

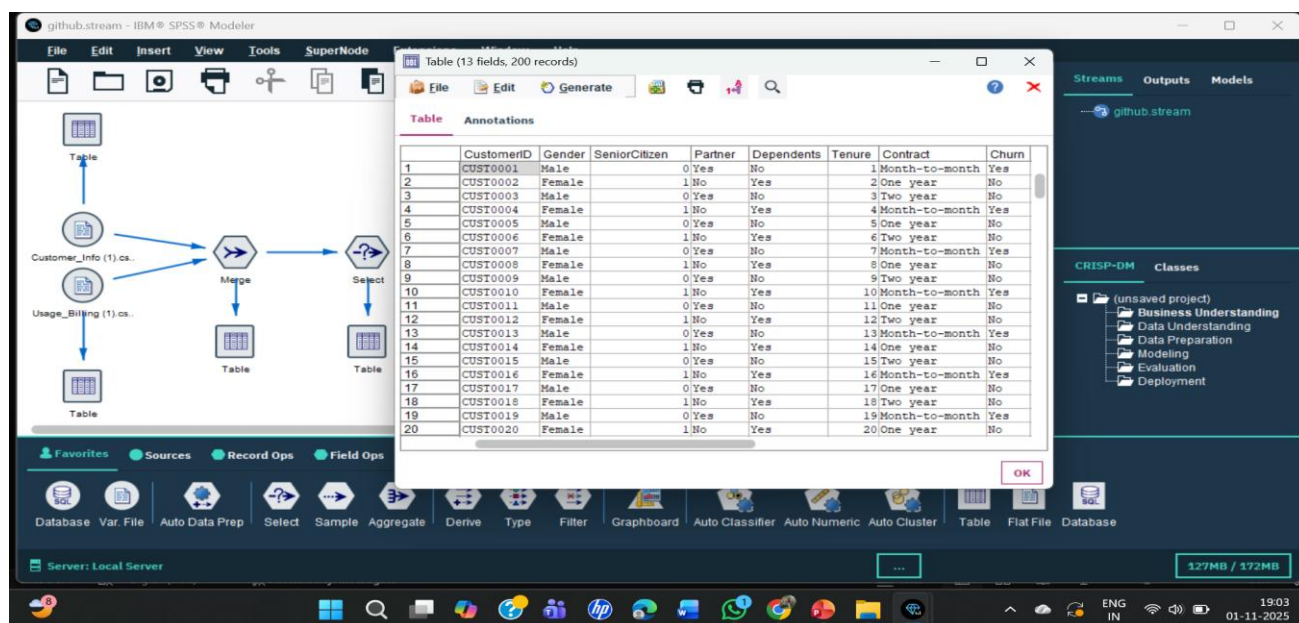
**Step 6:-** After that add Select node from the record option and connect with the Merge node

The screenshot shows the IBM SPSS Modeler interface with a workflow diagram. The workflow starts with two input nodes: 'Customer\_Info (1).cs.' and 'Usage\_Billing (1).cs.'. These are connected to a 'Merge' node. A 'Select' node is added to the workflow, connected to the 'Merge' node. The 'Select' node is highlighted with a dashed border. The right sidebar shows the 'Streams' tab with 'Stream1' and 'github.stream'. The bottom status bar shows 'Server: Local Server' and '153MB / 187MB'.

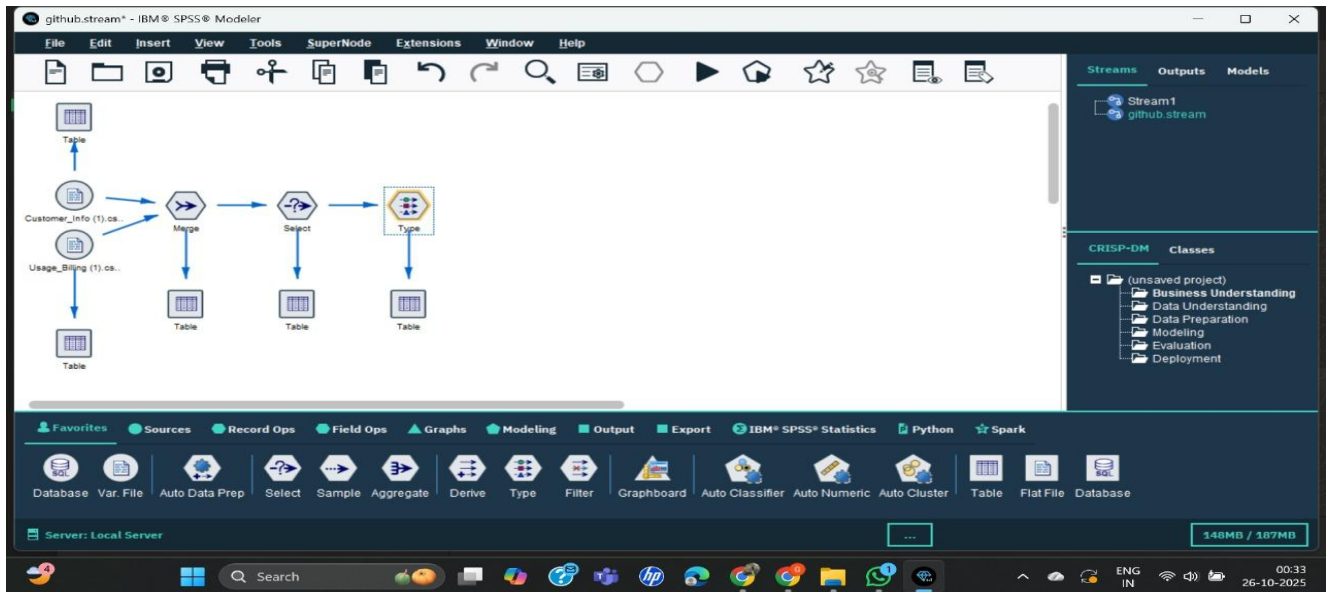
**Step 7:-** Now Double click on select node and filter the dataset by applying the condition **not(@NULL(Churn))** and **not(Churn="")** and **MonthlyCharges > 0** and **TotalCharges > 0** and **Tenure > 0** " to keep only valid records.



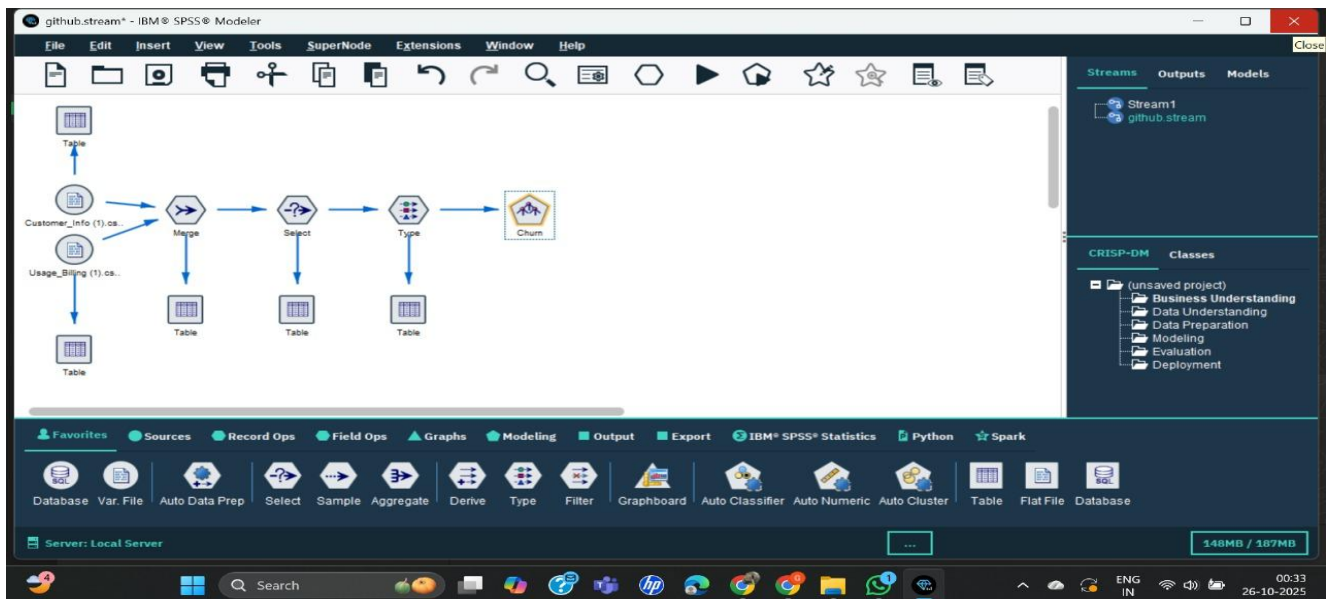
**Step 8:-** Now connect the tabel to see the data is running or not.



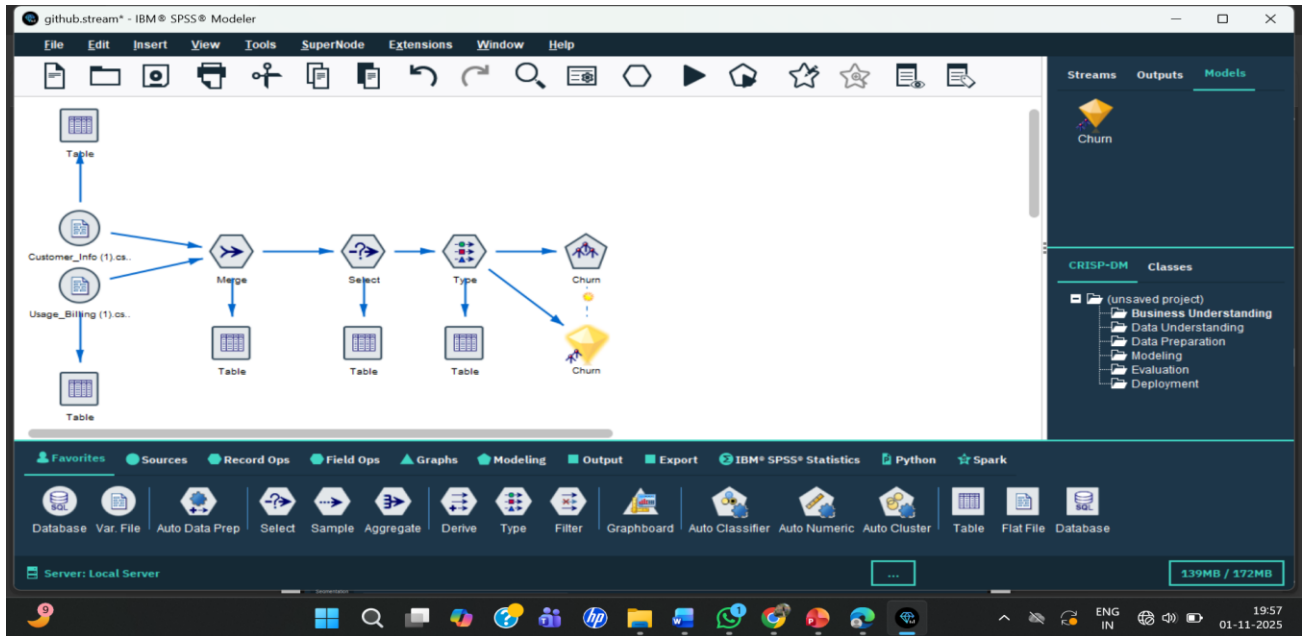
**Step 9:-** From the Field tab, drag the Type node and connect it to the select node. In the Type node, choose all the field as input such as gender, senior citizen, partner, etc.. except customer id and set churn as the target variable , and add the table node to see the output



**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



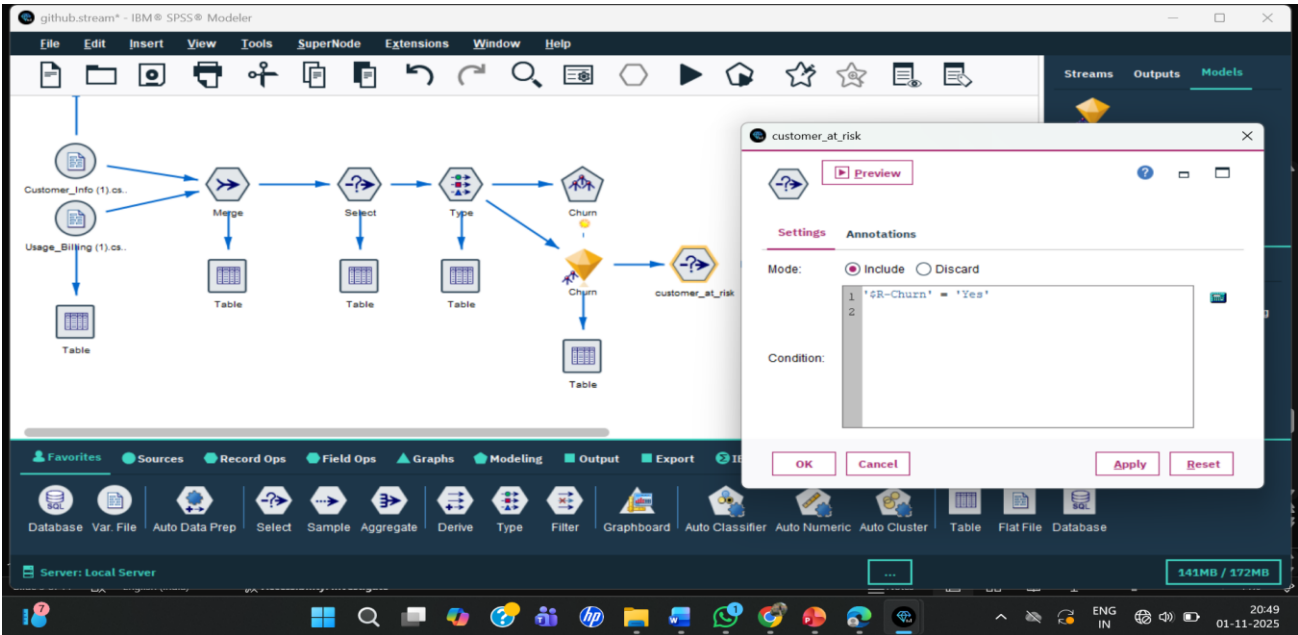
**Step 12:-** Now add table to check the churn we created is showing a churn fields or not like \$R-churn, \$RC-churn, \$RI-churn.

The screenshot displays the IBM SPSS Modeler interface with a 'Table' output window open. The window shows a table with 16 fields and 200 records. The table has columns: 'MonthlyCharges', 'TotalCharges', '\$R-Churn', '\$RC-Churn', and '\$RI-Churn'. The data shows various values for these fields, including 'ck' for 'MonthlyCharges' and 'TotalCharges', and 'Yes'/'No' for the churn-related fields. The bottom status bar indicates 'Server: Local Server' and '140MB / 172MB'.

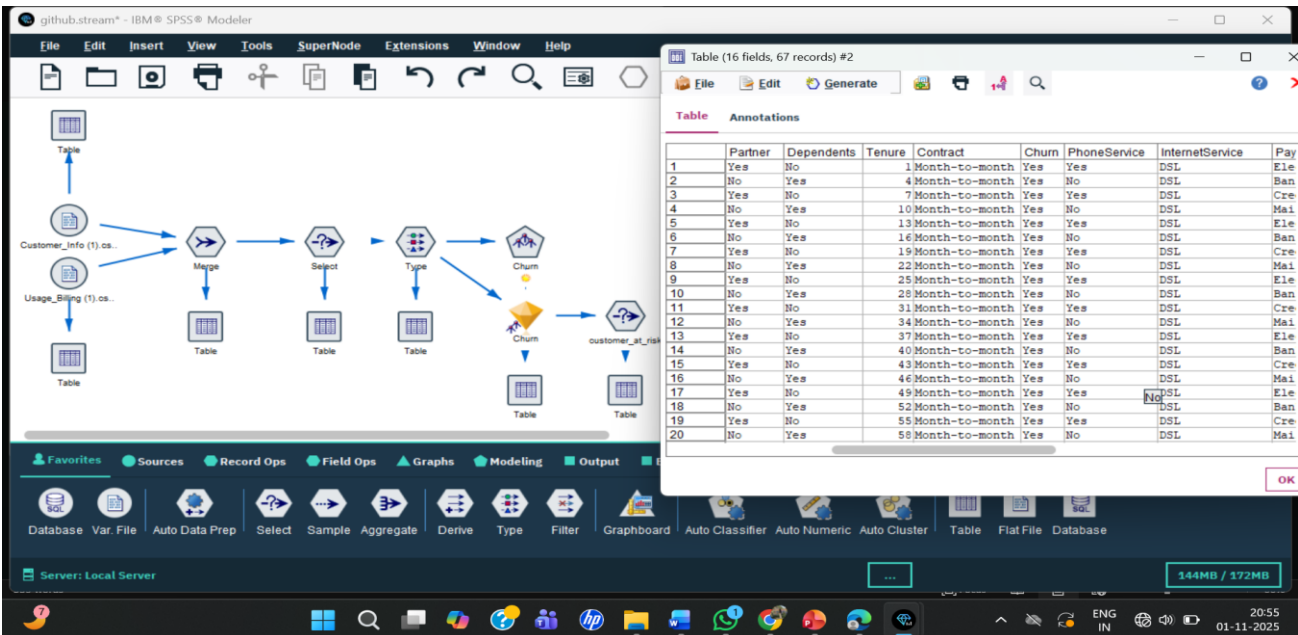
	MonthlyCharges	TotalCharges	\$R-Churn	\$RC-Churn	\$RI-Churn
1		30	100.000	Yes	0.986
2	ck	31	150.750	No	0.993
3		32	201.500	No	0.993
4		33	252.250	Yes	0.986
5	ck	34	303.000	No	0.993
6		35	353.750	No	0.993
7		36	404.500	Yes	0.986
8		37	455.250	No	0.993
9		38	506.000	No	0.993
10		39	556.750	Yes	0.986
11		40	607.500	No	0.993
12		41	658.250	No	0.993
13	ck	42	709.000	Yes	0.986
14		43	759.750	No	0.993
15		44	810.500	No	0.993
16		45	861.250	Yes	0.986
17	ck	46	912.000	No	0.993
18		47	962.750	No	0.993
19		48	1013.500	Yes	0.986
20		49	1064.250	No	0.993



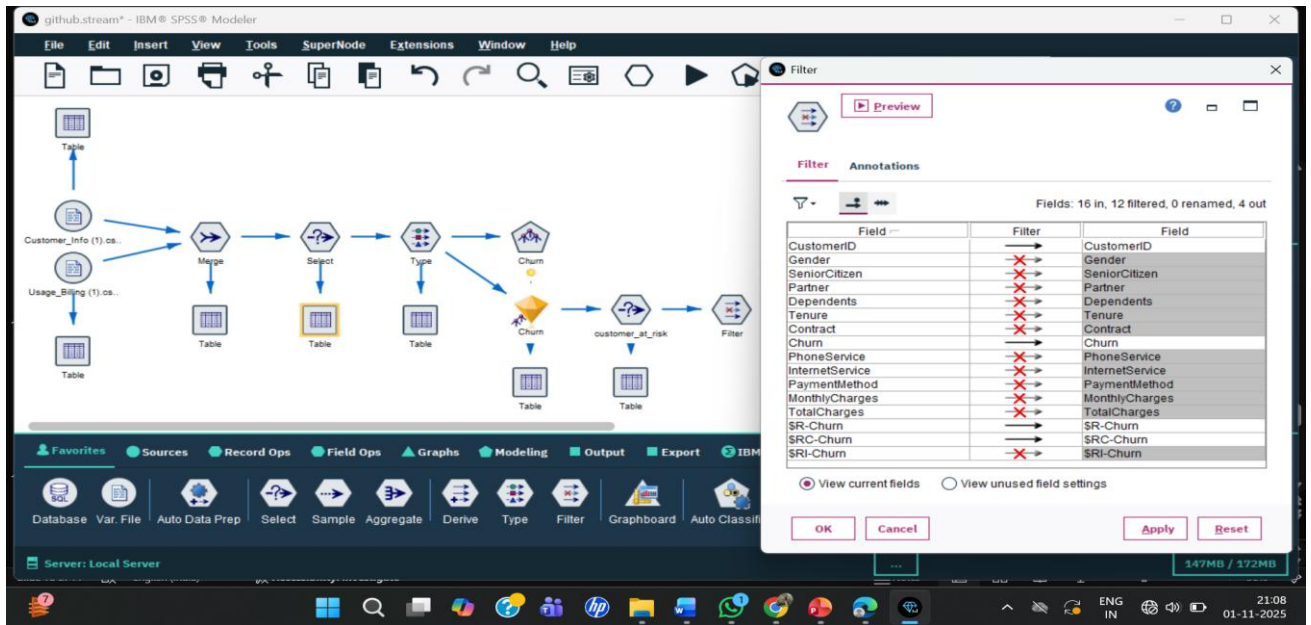
**Step13:-** Now we go to the Record Ops and click on select node rename the node as **Customer\_at\_Risk** and write query ‘\$R-Churn’ = ‘Yes’



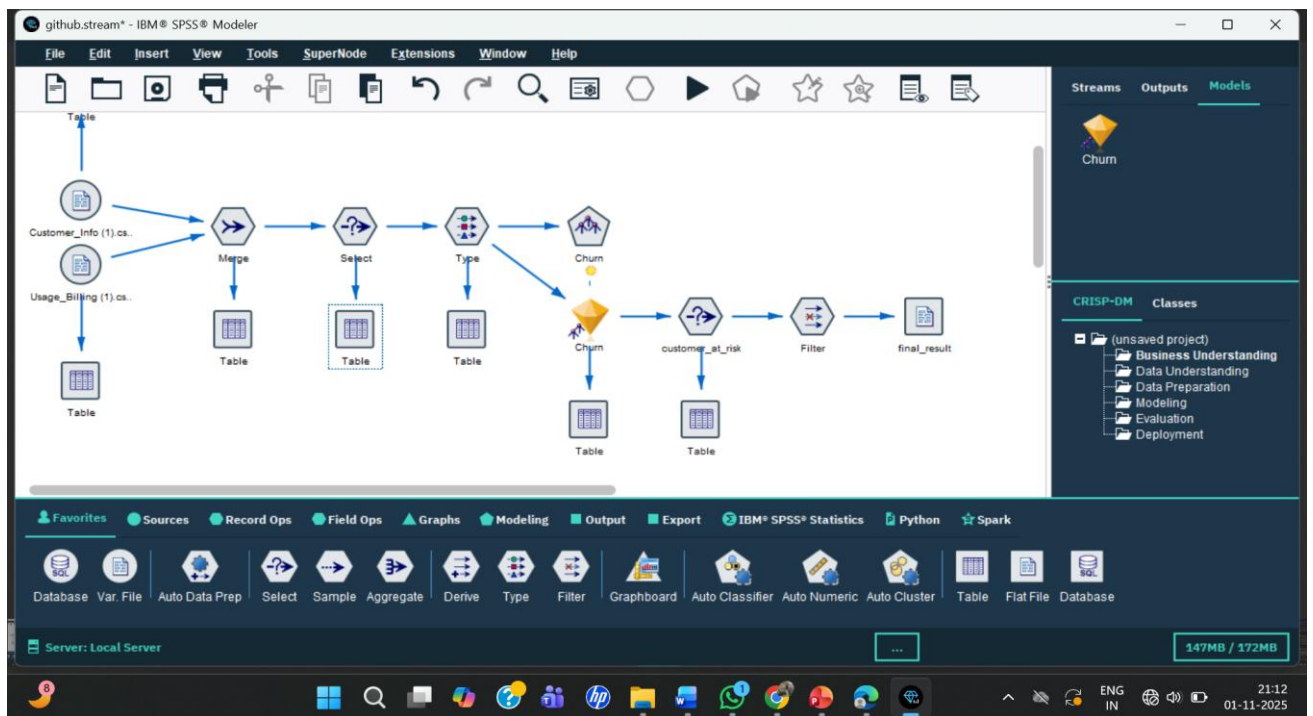
**Step14:-** Now connect a tabel node and see the customers ones who are most likely to Churn... high-risk(Churn=” YES”) customers



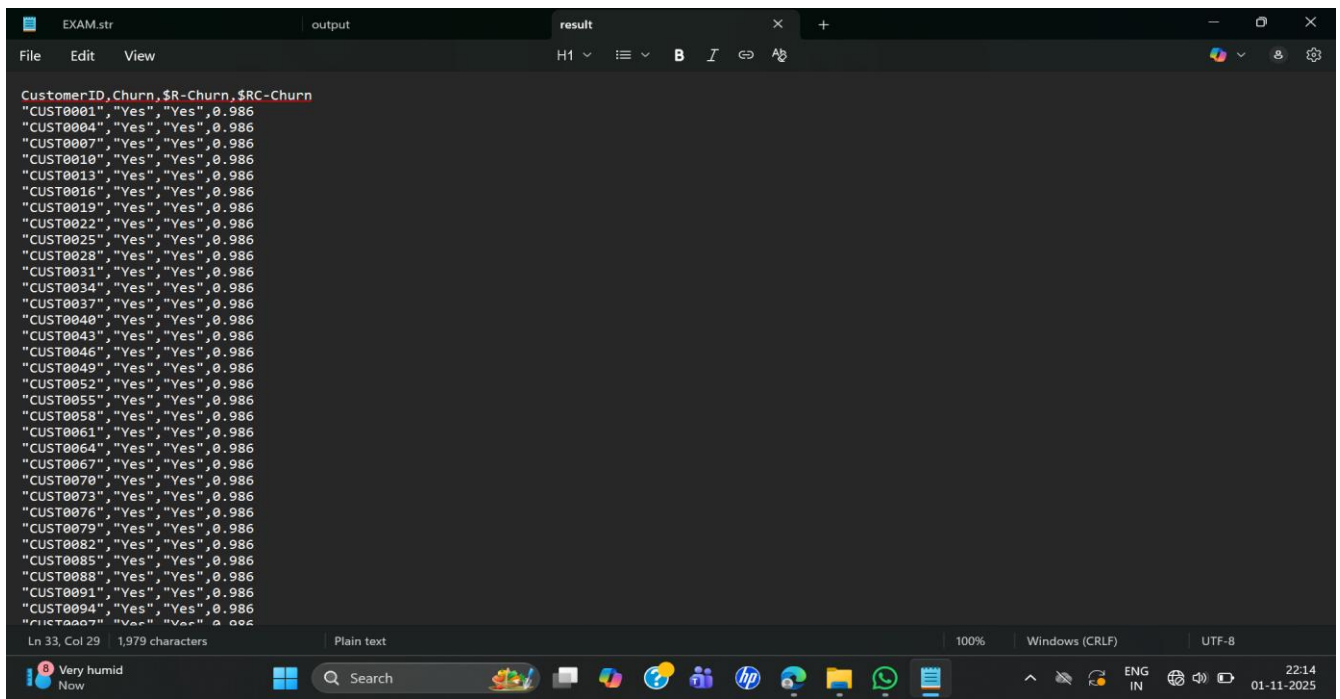
**Step15:-** We go to field category to add the filter node to filter the data means which field we want to see as final output



**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



The screenshot shows a Notepad window with a dark theme. The title bar indicates three open files: 'EXAM.str', 'output', and 'result'. The 'result' file is the active document. The text content is a CSV dataset with four columns: 'CustomerID', 'Churn', '\$R-Churn', and '\$RC-Churn'. The first row is the header, and the subsequent rows contain 32 data entries, each with a unique CustomerID, a 'Yes' churn status, and a churn rate of 0.986. The status bar at the bottom shows the cursor is at line 33, column 29, with a total of 1,979 characters. The Windows taskbar is visible at the very bottom, showing the system clock as 22:14 on 01-11-2025.

```
CustomerID,Churn,$R-Churn,$RC-Churn
"CUST0001","Yes","Yes",0.986
"CUST0004","Yes","Yes",0.986
"CUST0007","Yes","Yes",0.986
"CUST0010","Yes","Yes",0.986
"CUST0013","Yes","Yes",0.986
"CUST0016","Yes","Yes",0.986
"CUST0019","Yes","Yes",0.986
"CUST0022","Yes","Yes",0.986
"CUST0025","Yes","Yes",0.986
"CUST0028","Yes","Yes",0.986
"CUST0031","Yes","Yes",0.986
"CUST0034","Yes","Yes",0.986
"CUST0037","Yes","Yes",0.986
"CUST0040","Yes","Yes",0.986
"CUST0043","Yes","Yes",0.986
"CUST0046","Yes","Yes",0.986
"CUST0049","Yes","Yes",0.986
"CUST0052","Yes","Yes",0.986
"CUST0055","Yes","Yes",0.986
"CUST0058","Yes","Yes",0.986
"CUST0061","Yes","Yes",0.986
"CUST0064","Yes","Yes",0.986
"CUST0067","Yes","Yes",0.986
"CUST0070","Yes","Yes",0.986
"CUST0073","Yes","Yes",0.986
"CUST0076","Yes","Yes",0.986
"CUST0079","Yes","Yes",0.986
"CUST0082","Yes","Yes",0.986
"CUST0085","Yes","Yes",0.986
"CUST0088","Yes","Yes",0.986
"CUST0091","Yes","Yes",0.986
"CUST0094","Yes","Yes",0.986
"CUST0097","Yes","Yes",0.986
```