BABU BANARASI DAS UNIVERSITY

SCHOOL OF COMPUTER APPLICATION

CASE STUDY

ON

Integration of Telecommunications Datasets using SPSS Modeler

SUBMITTED TO:

MR. AYUSHMAN BHADURIYA

SUBMITTED BY:

NAME: BHUMI YADAV

CLASS: BCADS32

ROLL NO: 1230258145

ASSIGNMENT

Integration of Telecommunications Datasets using SPSS Modeler

DEFINITION:

This case study demonstrates the process of combining different datasets related to telecommunications, including call data, customer data, product details, and tariff information. Using IBM SPSS Modeler, these datasets are merged, cleansed, and transformed to create a single dataset suitable for analytical tasks and predictive modeling.

OUTCOME:

- Integrating multiple telecommunications datasets into one unified file using IBM SPSS Modeler.
- Ensuring consistency and accuracy across customer, call, product, and tariff data.
- How dataset integration simplifies data preparation for advanced analysis and modeling.

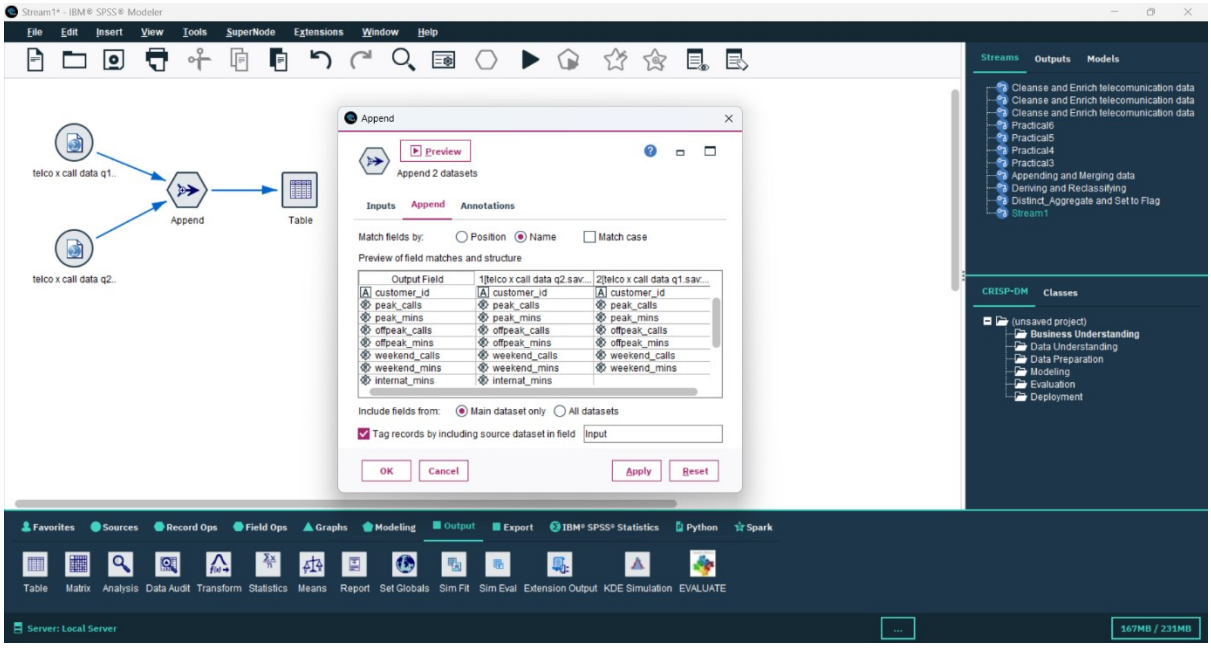
REQUIRED TOOL: IBM SPSS Modeler Tool

WORKING: Using Merge, Append and Sample nodes to merge two datasets, to append records from two or more data sets and to sample data respectively.

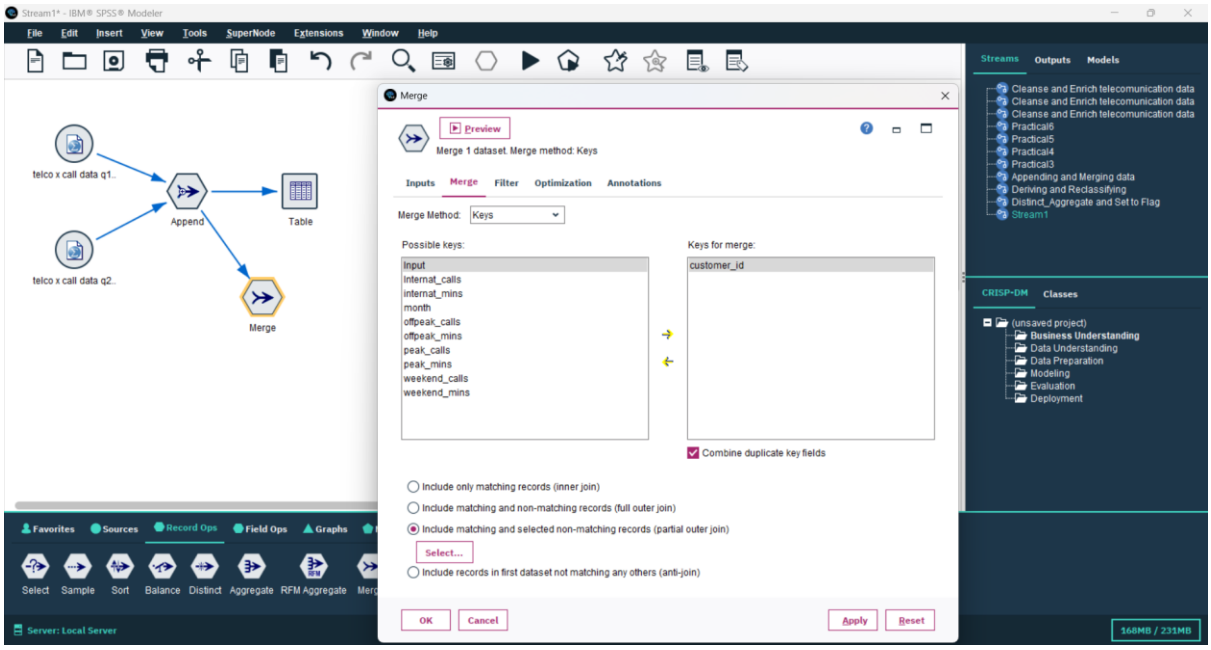
Steps to perform the task :-

Step 1: Open the SPSS Modeler and from the Source palette take two Statistics File node and import two datasets (telco x call data q1.sav and telco x call data q2.sav)

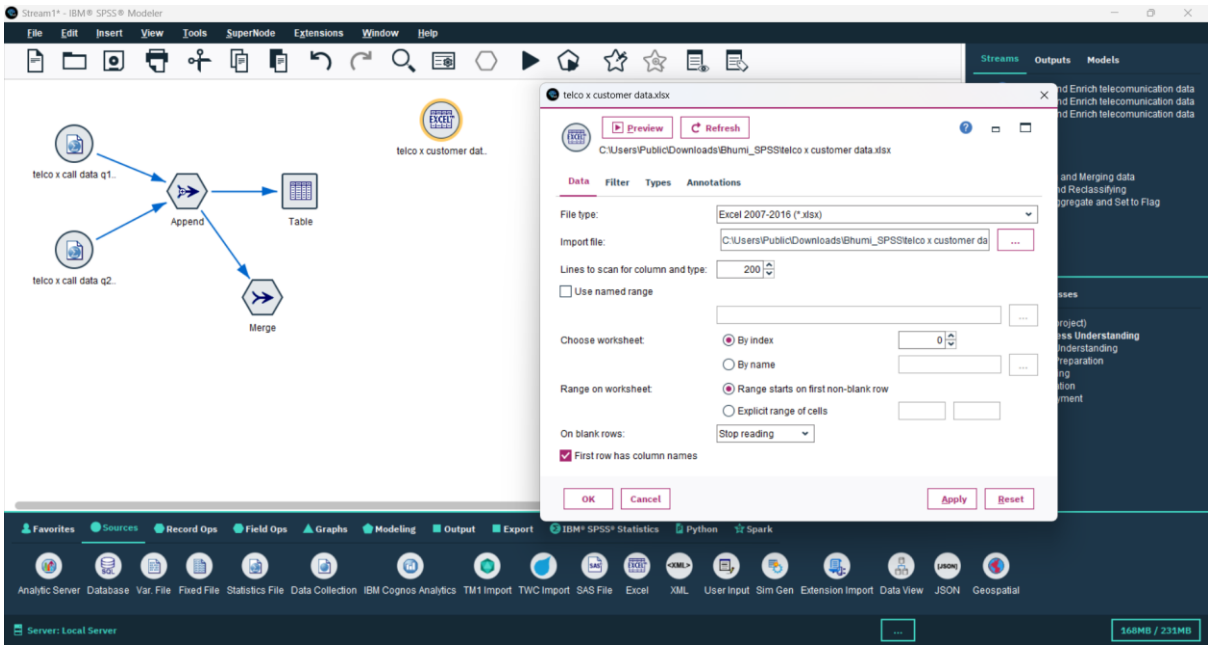
Step 2: Now from the Record Ops palette, take the Append node and connect both datasets to this. This will merge the two datasets together.



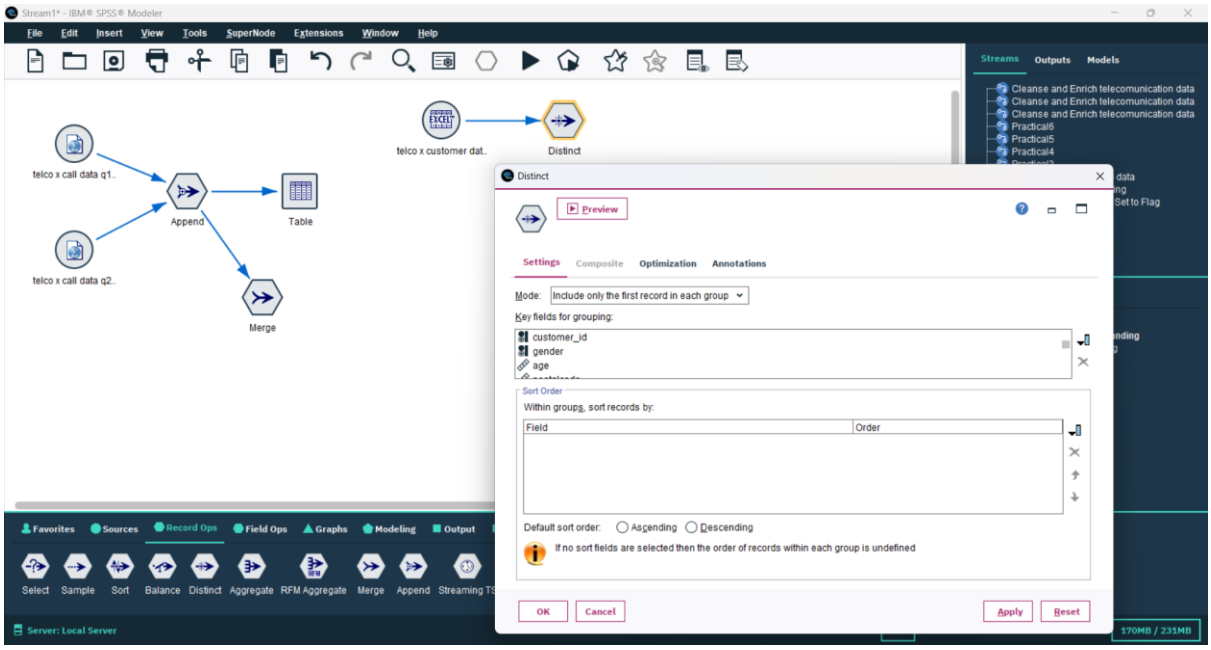
Step 3: Now connect Append node to Merge node from Record Ops palette. Select customer_id as the key for merge and partial outer join by double clicking on merge node.



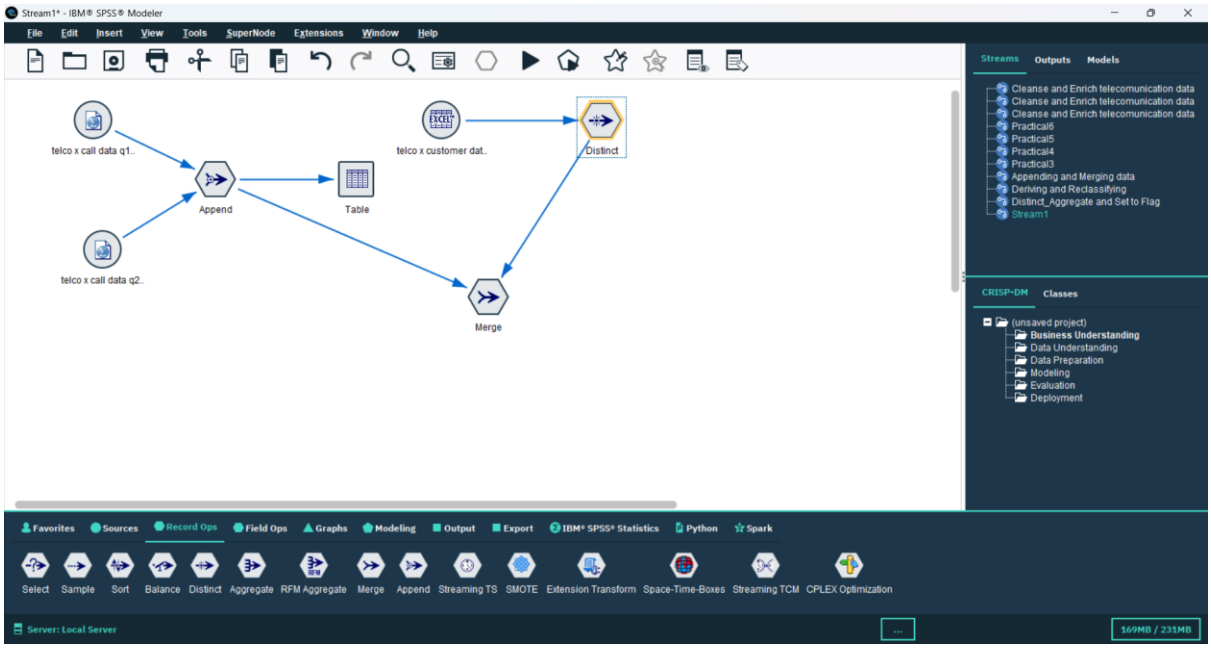
Step 4: Now we will import a new dataset (telco x customer data.xlsx) in the excel node from Sources.



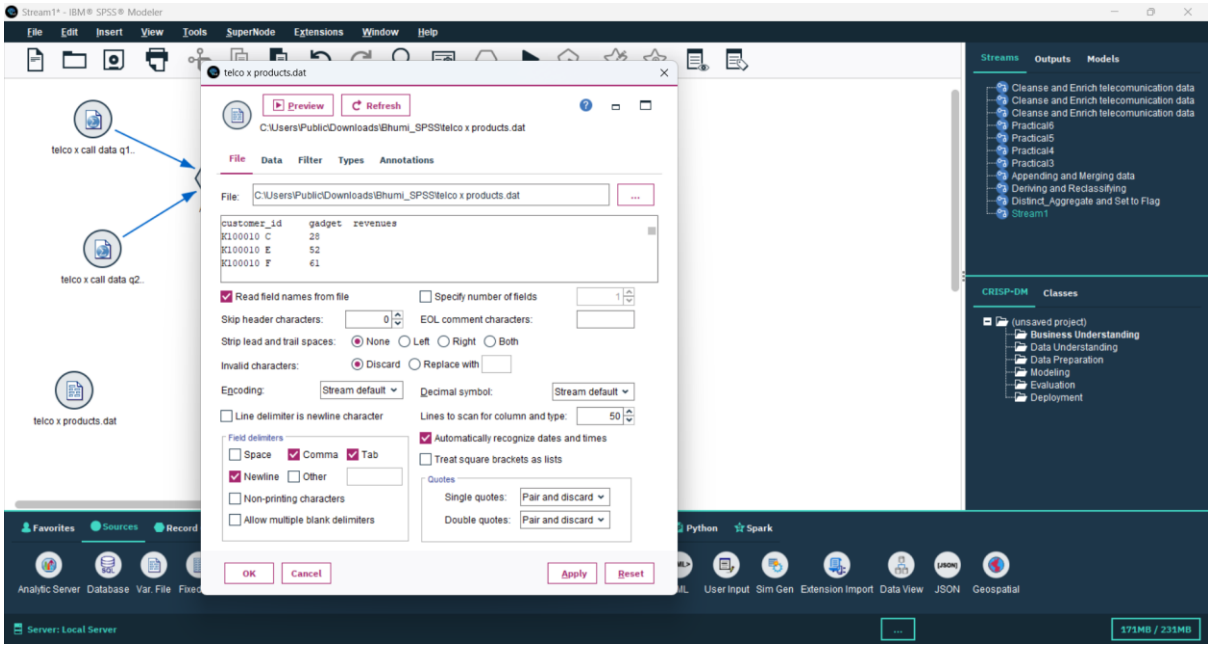
Step 5: Connect Distinct node to the excel dataset and take all the columns as keys. Now choose “Include only the first record in each group” and click on Apply and OK.



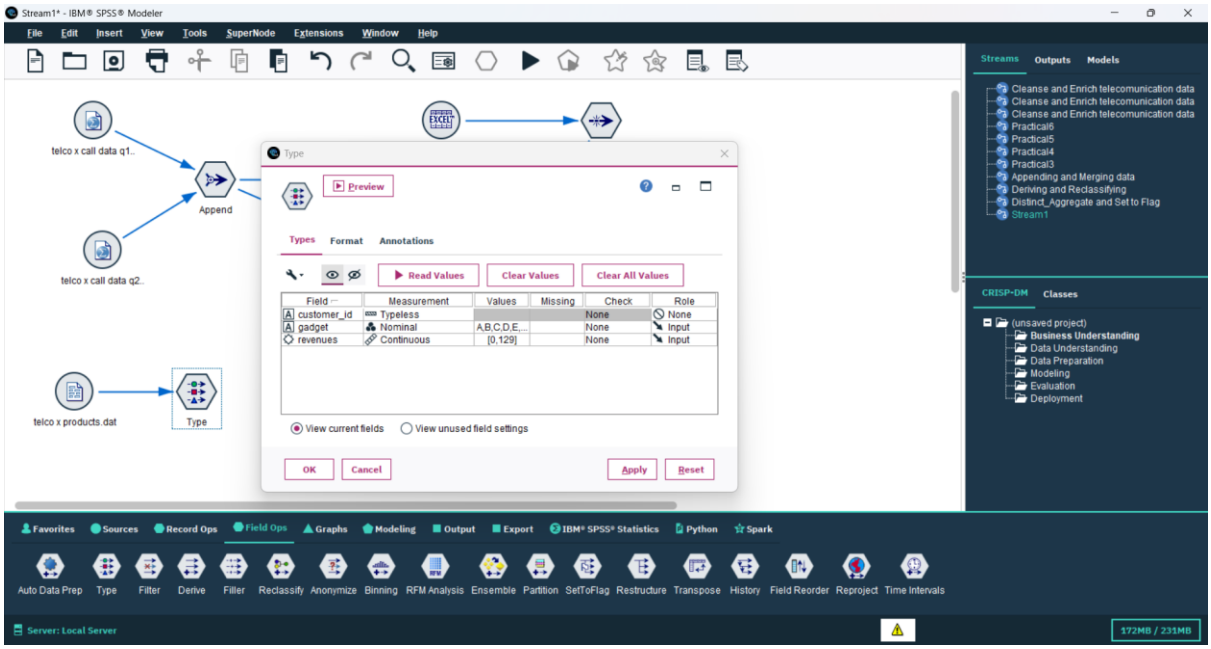
Step 6: Connect the Distinct node to Merge node.



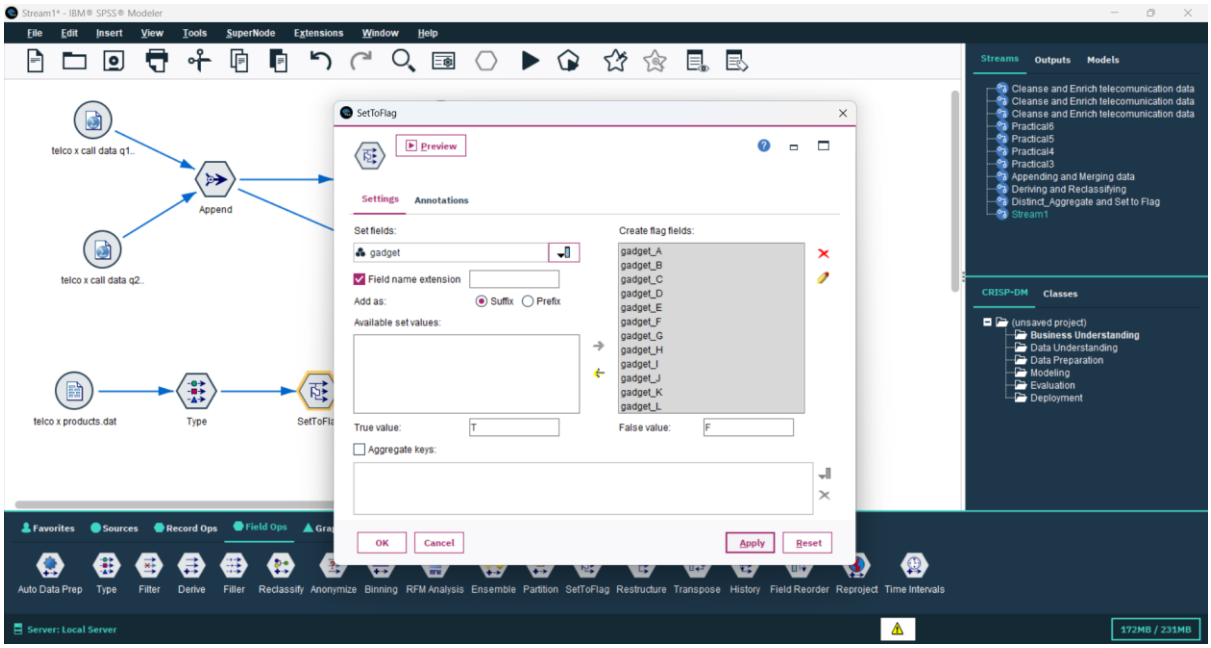
Step 7: Next take a Var. File from Sources and import telco x products.dat



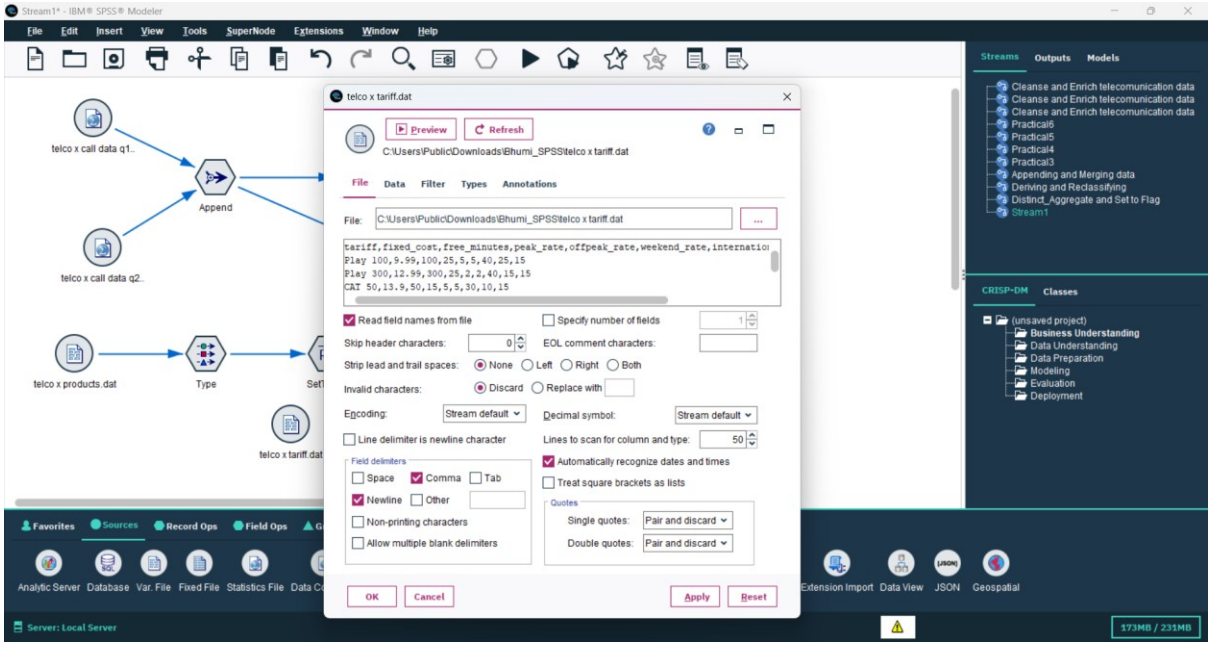
Step 8: Take Type node from Field Ops and connect to our var file and get specific category of all values by clicking on Read Values in Type window.



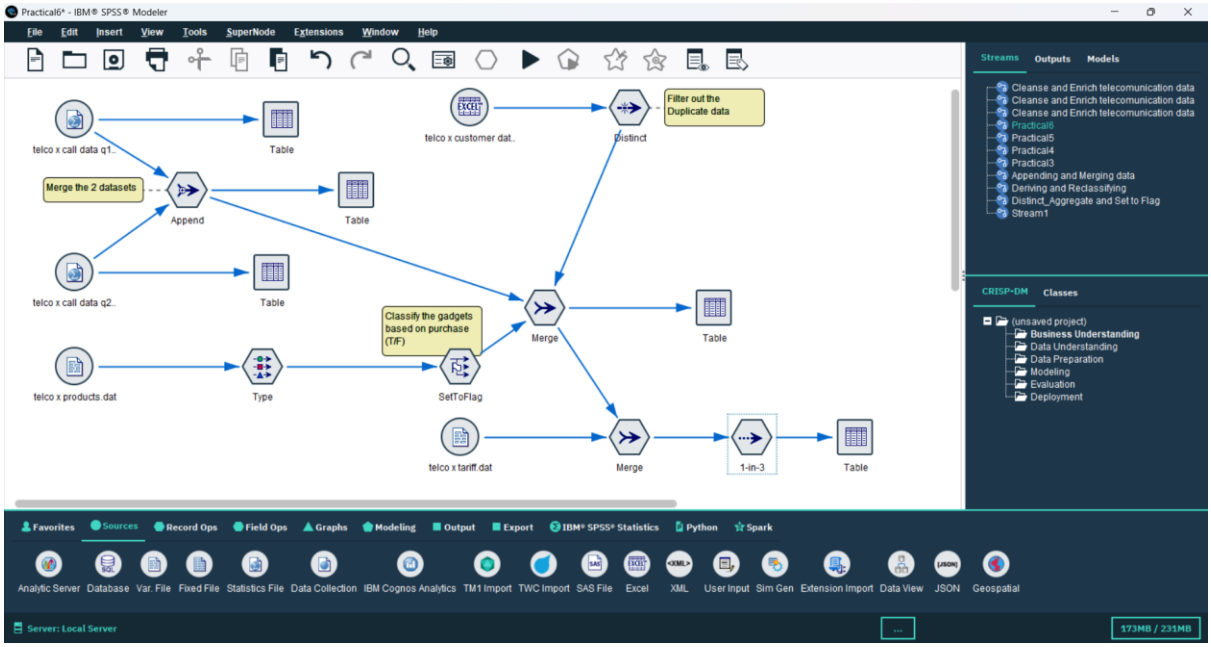
Step 9: Connect SetToFlag node to Type node and select gadget in Set fields section and send all selected values to Create flag fields. Now connect the SetToFlag node to Merge node.



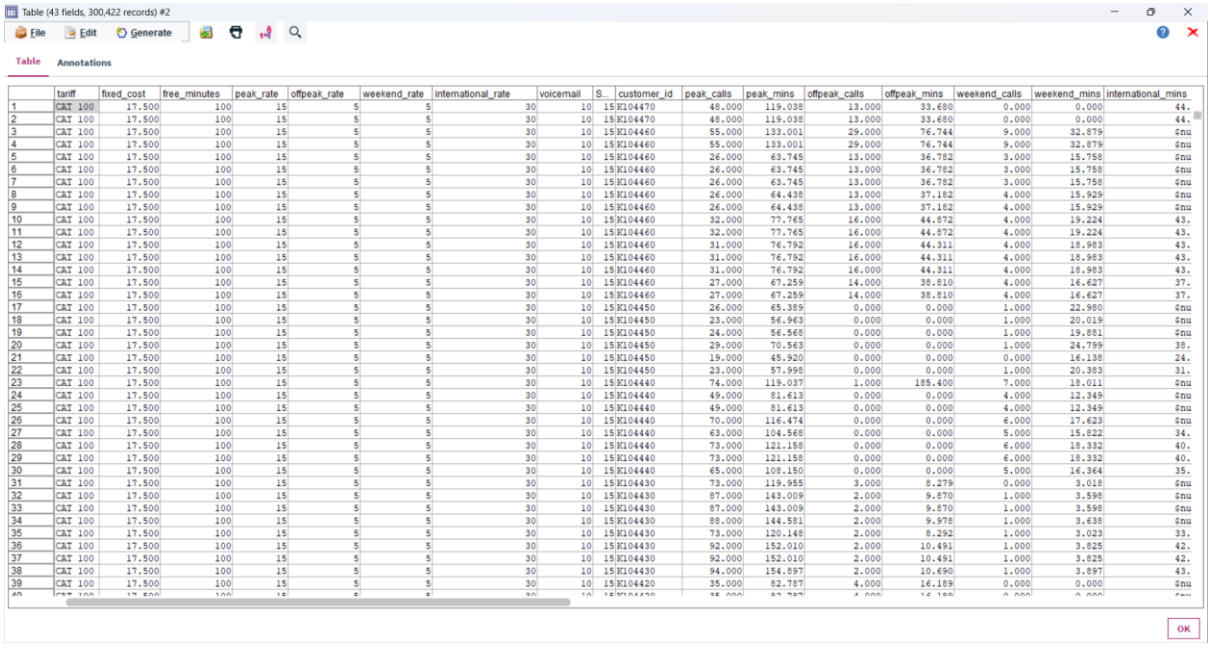
Step 10: Now we add another Var. File (telco x tariff.dat). Connect Merge node to the new var file and select tariff as the key and partial outer join in the Merge window. Then connect our other Merge node to this one.



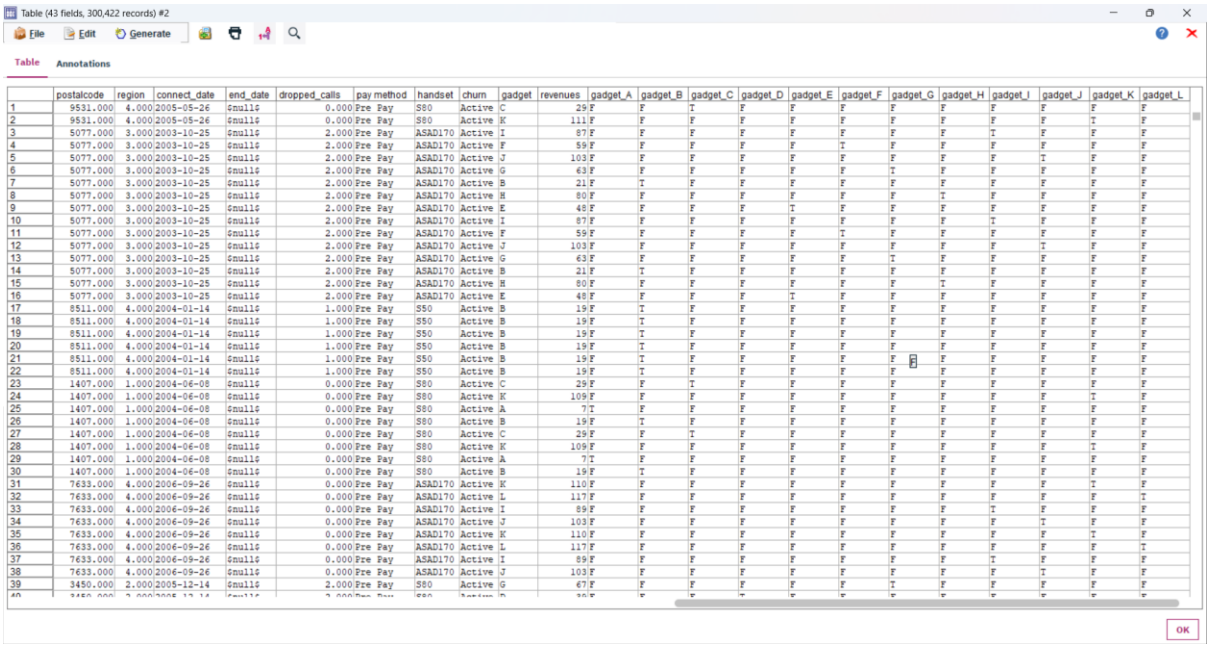
<u>Step 11:</u> Then connect the sample node from Record Ops palette to the new merge node. Open sample node window and select 1-in-n option. Choosing 3 in place of n will show the first record and then the thirds record skipping the second one.



Finally connect the Table node to the Sample node to view the result which shows 4 joint datasets.



	tariff	fixed_cost	free_minutes	peak_rate	offpeak_rate	weekend_rate	international_rate	voicemail	S...	customer_id	peak_calls	peak_mins	offpeak_calls	offpeak_mins	weekend_calls	weekend_mins	international_mins
1	CAT 100	17.500	100	15	5	5	30	10	15	K104470	48.000	119.038	13.000	33.680	0.000	0.000	44.
2	CAT 100	17.500	100	15	5	5	30	10	15	K104470	48.000	119.038	13.000	33.680	0.000	0.000	44.
3	CAT 100	17.500	100	15	5	5	30	10	15	K104460	55.000	133.001	29.000	76.744	9.000	32.879	$nu
4	CAT 100	17.500	100	15	5	5	30	10	15	K104460	55.000	133.001	29.000	76.744	9.000	32.879	$nu
5	CAT 100	17.500	100	15	5	5	30	10	15	K104460	26.000	63.745	13.000	36.782	3.000	15.758	$nu
6	CAT 100	17.500	100	15	5	5	30	10	15	K104460	26.000	63.745	13.000	36.782	3.000	15.758	$nu
7	CAT 100	17.500	100	15	5	5	30	10	15	K104460	26.000	63.745	13.000	36.782	3.000	15.758	$nu
8	CAT 100	17.500	100	15	5	5	30	10	15	K104460	26.000	64.438	13.000	37.182	4.000	15.929	$nu
9	CAT 100	17.500	100	15	5	5	30	10	15	K104460	26.000	64.438	13.000	37.182	4.000	15.929	$nu
10	CAT 100	17.500	100	15	5	5	30	10	15	K104460	32.000	77.765	16.000	44.872	4.000	19.224	43.
11	CAT 100	17.500	100	15	5	5	30	10	15	K104460	32.000	77.765	16.000	44.872	4.000	19.224	43.
12	CAT 100	17.500	100	15	5	5	30	10	15	K104460	31.000	76.792	16.000	44.311	4.000	18.983	43.
13	CAT 100	17.500	100	15	5	5	30	10	15	K104460	31.000	76.792	16.000	44.311	4.000	18.983	43.
14	CAT 100	17.500	100	15	5	5	30	10	15	K104460	31.000	76.792	16.000	44.311	4.000	18.983	43.
15	CAT 100	17.500	100	15	5	5	30	10	15	K104460	27.000	67.259	14.000	38.810	4.000	16.627	37.
16	CAT 100	17.500	100	15	5	5	30	10	15	K104460	27.000	67.259	14.000	38.810	4.000	16.627	37.
17	CAT 100	17.500	100	15	5	5	30	10	15	K104450	26.000	65.389	0.000	0.000	1.000	22.980	$nu
18	CAT 100	17.500	100	15	5	5	30	10	15	K104450	23.000	56.963	0.000	0.000	1.000	20.019	$nu
19	CAT 100	17.500	100	15	5	5	30	10	15	K104450	24.000	56.568	0.000	0.000	1.000	19.881	$nu
20	CAT 100	17.500	100	15	5	5	30	10	15	K104450	29.000	70.563	0.000	0.000	1.000	24.799	38.
21	CAT 100	17.500	100	15	5	5	30	10	15	K104450	19.000	45.920	0.000	0.000	0.000	16.138	24.
22	CAT 100	17.500	100	15	5	5	30	10	15	K104450	23.000	57.998	0.000	0.000	1.000	20.383	31.
23	CAT 100	17.500	100	15	5	5	30	10	15	K104440	74.000	119.037	1.000	185.400	7.000	18.011	$nu
24	CAT 100	17.500	100	15	5	5	30	10	15	K104440	49.000	81.613	0.000	0.000	4.000	12.349	$nu
25	CAT 100	17.500	100	15	5	5	30	10	15	K104440	49.000	81.613	0.000	0.000	4.000	12.349	$nu
26	CAT 100	17.500	100	15	5	5	30	10	15	K104440	70.000	116.474	0.000	0.000	6.000	17.623	$nu
27	CAT 100	17.500	100	15	5	5	30	10	15	K104440	63.000	104.569	0.000	0.000	5.000	15.822	34.
28	CAT 100	17.500	100	15	5	5	30	10	15	K104440	73.000	121.158	0.000	0.000	6.000	18.332	40.
29	CAT 100	17.500	100	15	5	5	30	10	15	K104440	73.000	121.158	0.000	0.000	6.000	18.332	40.
30	CAT 100	17.500	100	15	5	5	30	10	15	K104440	65.000	108.150	0.000	0.000	5.000	16.364	35.
31	CAT 100	17.500	100	15	5	5	30	10	15	K104430	73.000	119.955	3.000	8.279	0.000	3.018	$nu
32	CAT 100	17.500	100	15	5	5	30	10	15	K104430	87.000	143.009	2.000	9.870	1.000	3.598	$nu
33	CAT 100	17.500	100	15	5	5	30	10	15	K104430	87.000	143.009	2.000	9.870	1.000	3.598	$nu
34	CAT 100	17.500	100	15	5	5	30	10	15	K104430	88.000	144.581	2.000	9.978	1.000	3.638	$nu
35	CAT 100	17.500	100	15	5	5	30	10	15	K104430	73.000	120.148	2.000	8.292	1.000	3.023	33.
36	CAT 100	17.500	100	15	5	5	30	10	15	K104430	92.000	152.010	2.000	10.491	1.000	3.825	42.
37	CAT 100	17.500	100	15	5	5	30	10	15	K104430	92.000	152.010	2.000	10.491	1.000	3.825	42.
38	CAT 100	17.500	100	15	5	5	30	10	15	K104430	94.000	154.897	2.000	10.690	1.000	3.897	43.
39	CAT 100	17.500	100	15	5	5	30	10	15	K104420	35.000	82.787	4.000	16.189	0.000	0.000	$nu



	postalcode	region	connect_date	end_date	dropped_calls	pay method	handset	churn	gadget	revenues	gadget_A	gadget_B	gadget_C	gadget_D	gadget_E	gadget_F	gadget_G	gadget_H	gadget_I	gadget_J	gadget_K	gadget_L
1	9531.000	4.000	2005-05-26	$null$	0.000	Pre Pay	S80	Active	C	29	F	F	T	F	F	F	F	F	F	F	F	F
2	9531.000	4.000	2005-05-26	$null$	0.000	Pre Pay	S80	Active	K	111	F	F	F	F	F	F	F	F	F	F	T	F
3	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	I	87	F	F	F	F	F	F	F	F	T	F	F	F
4	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	F	59	F	F	F	F	F	T	F	F	F	F	F	F
5	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	J	103	F	F	F	F	F	F	F	F	F	T	F	F
6	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	G	63	F	F	F	F	F	F	T	F	F	F	F	F
7	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	B	21	F	T	F	F	F	F	F	F	F	F	F	F
8	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	H	80	F	F	F	F	F	F	F	T	F	F	F	F
9	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	E	48	F	F	F	F	T	F	F	F	F	F	F	F
10	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	I	87	F	F	F	F	F	F	F	F	T	F	F	F
11	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	F	59	F	F	F	F	T	F	F	F	F	F	F	F
12	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	J	103	F	F	F	F	F	F	F	F	F	T	F	F
13	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	G	63	F	F	F	F	F	F	T	F	F	F	F	F
14	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	B	21	F	T	F	F	F	F	F	F	F	F	F	F
15	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	H	80	F	F	F	F	F	F	F	T	F	F	F	F
16	5077.000	3.000	2003-10-25	$null$	2.000	Pre Pay	ASAD170	Active	E	48	F	F	F	F	T	F	F	F	F	F	F	F
17	8511.000	4.000	2004-01-14	$null$	1.000	Pre Pay	S50	Active	B	19	F	T	F	F	F	F	F	F	F	F	F	F
18	8511.000	4.000	2004-01-14	$null$	1.000	Pre Pay	S50	Active	B	19	F	T	F	F	F	F	F	F	F	F	F	F
19	8511.000	4.000	2004-01-14	$null$	1.000	Pre Pay	S50	Active	B	19	F	T	F	F	F	F	F	F	F	F	F	F
20	8511.000	4.000	2004-01-14	$null$	1.000	Pre Pay	S50	Active	B	19	F	T	F	F	F	F	F	F	F	F	F	F
21	8511.000	4.000	2004-01-14	$null$	1.000	Pre Pay	S50	Active	B	19	F	T	F	F	F	F	F	F	F	F	F	F
22	8511.000	4.000	2004-01-14	$null$	1.000	Pre Pay	S50	Active	B	19	F	T	F	F	F	F	F	F	F	F	F	F
23	1407.000	1.000	2004-06-08	$null$	0.000	Pre Pay	S80	Active	C	29	F	F	T	F	F	F	F	F	F	F	F	F
24	1407.000	1.000	2004-06-08	$null$	0.000	Pre Pay	S80	Active	K	109	F	F	F	F	F	F	F	F	F	F	T	F
25	1407.000	1.000	2004-06-08	$null$	0.000	Pre Pay	S80	Active	A	7	T	F	F	F	F	F	F	F	F	F	F	F
26	1407.000	1.000	2004-06-08	$null$	0.000	Pre Pay	S80	Active	B	19	F	T	F	F	F	F	F	F	F	F	F	F
27	1407.000	1.000	2004-06-08	$null$	0.000	Pre Pay	S80	Active	C	29	F	F	T	F	F	F	F	F	F	F	F	F
28	1407.000	1.000	2004-06-08	$null$	0.000	Pre Pay	S80	Active	K	109	F	F	F	F	F	F	F	F	F	F	T	F
29	1407.000	1.000	2004-06-08	$null$	0.000	Pre Pay	S80	Active	A	7	T	F	F	F	F	F	F	F	F	F	F	F
30	1407.000	1.000	2004-06-08	$null$	0.000	Pre Pay	S80	Active	B	19	F	T	F	F	F	F	F	F	F	F	F	F
31	7633.000	4.000	2006-09-26	$null$	0.000	Pre Pay	ASAD170	Active	K	110	F	F	F	F	F	F	F	F	F	F	T	F
32	7633.000	4.000	2006-09-26	$null$	0.000	Pre Pay	ASAD170	Active	L	117	F	F	F	F	F	F	F	F	F	F	F	T
33	7633.000	4.000	2006-09-26	$null$	0.000	Pre Pay	ASAD170	Active	I	89	F	F	F	F	F	F	F	F	T	F	F	F
34	7633.000	4.000	2006-09-26	$null$	0.000	Pre Pay	ASAD170	Active	J	103	F	F	F	F	F	F	F	F	F	T	F	F
35	7633.000	4.000	2006-09-26	$null$	0.000	Pre Pay	ASAD170	Active	K	110	F	F	F	F	F	F	F	F	F	F	T	F
36	7633.000	4.000	2006-09-26	$null$	0.000	Pre Pay	ASAD170	Active	L	117	F	F	F	F	F	F	F	F	F	F	F	T
37	7633.000	4.000	2006-09-26	$null$	0.000	Pre Pay	ASAD170	Active	I	89	F	F	F	F	F	F	F	F	T	F	F	F
38	7633.000	4.000	2006-09-26	$null$	0.000	Pre Pay	ASAD170	Active	J	103	F	F	F	F	F	F	F	F	F	T	F	F
39	3450.000	2.000	2005-12-14	$null$	2.000	Pre Pay	S80	Active	G	67	F	F	F	F	F	F	T	F	F	F	F	F