

Babu Banarasi Das University



Predictive Analytics (BCADS15301)

LAB FILE

**SUBMITTED TO:
Mr. AYUSHMAN BHADURIA**

**SUBMITTED BY:-
Name: Vishal chaudhary
Roll No: 1230258472
Class: BCADS36**

SPSS Modeler Practical Workbook: Data Integration and Analysis

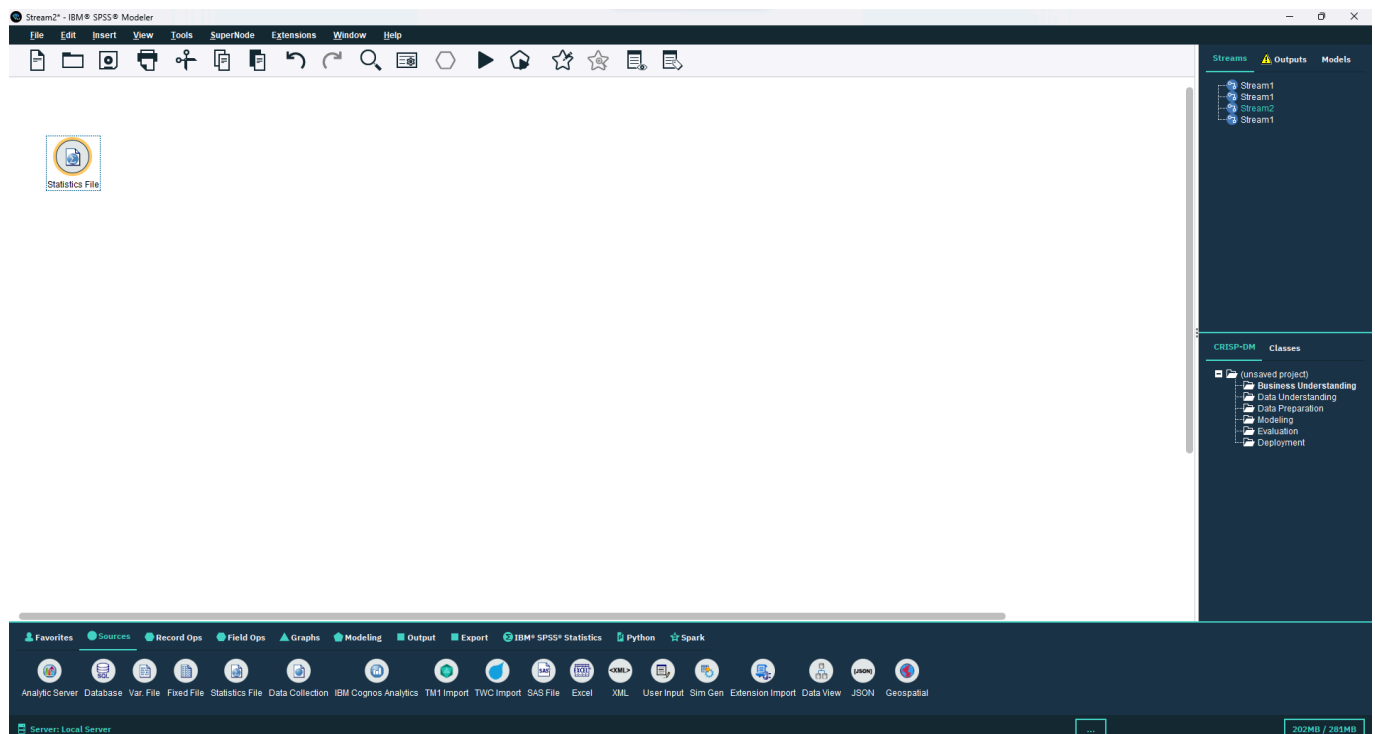
Definition: You work for a telecommunications firm and have to combine a number of datasets into a single dataset for analyses and modeling later.

Outcomes/Learning: Learning how to join different data sets
How to join records of different data sets
How to sample data from data sets

Required Tool: IBM SPSS Modeler Tool

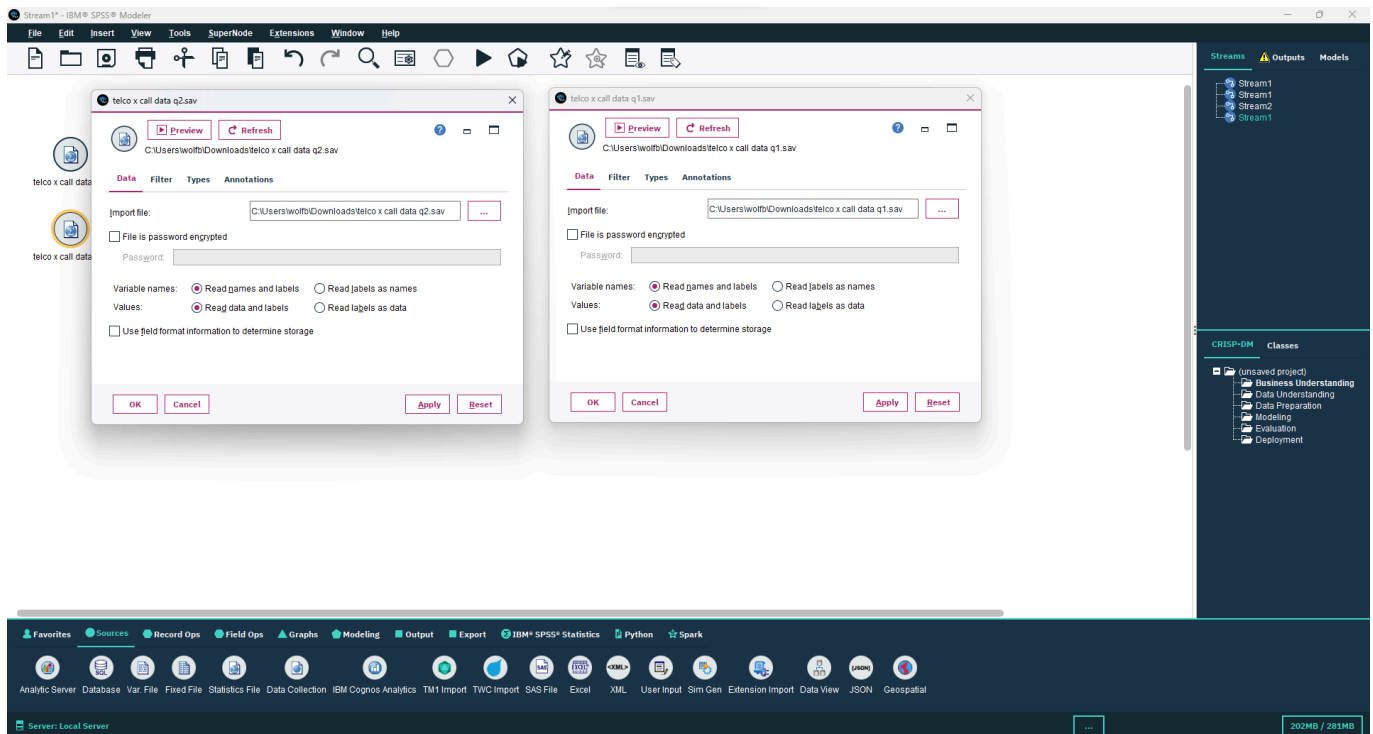
Working: Using Merge, Append, Sample nodes to merge two data sets, to append records from two or more data sets, To sample data from whole data sets.

Step 1: Open SPSS Modeler tool then on Source category select statistics File node (we are selecting Statistics node because the data set we are using is an sav file.)
Double clicking on statistics node will make it appear on canvas

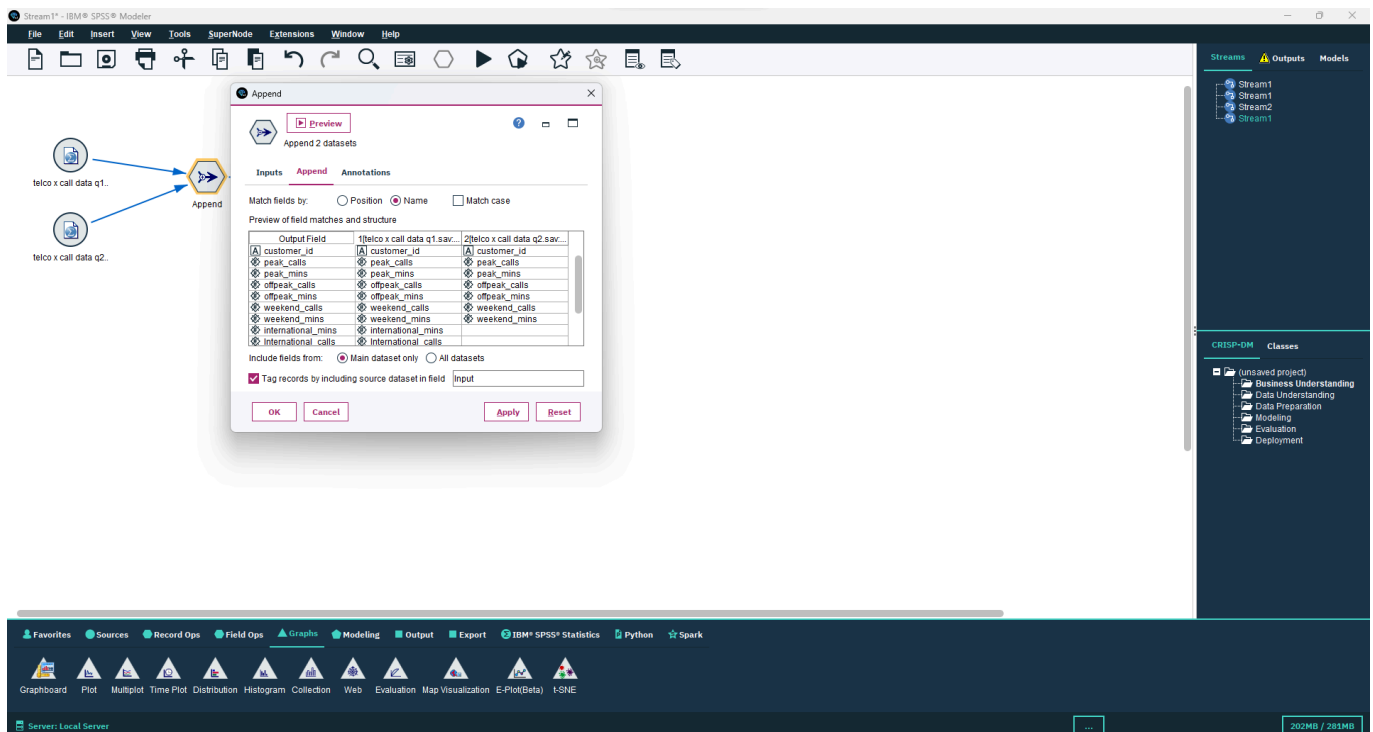


Step 2: Now we import a data set using the import option which can be accessed by double clicking on Statistics icon on the canvas.
We import a data set telco x call data q1.sav then import one more statistics file tele x call

data q2.sav.

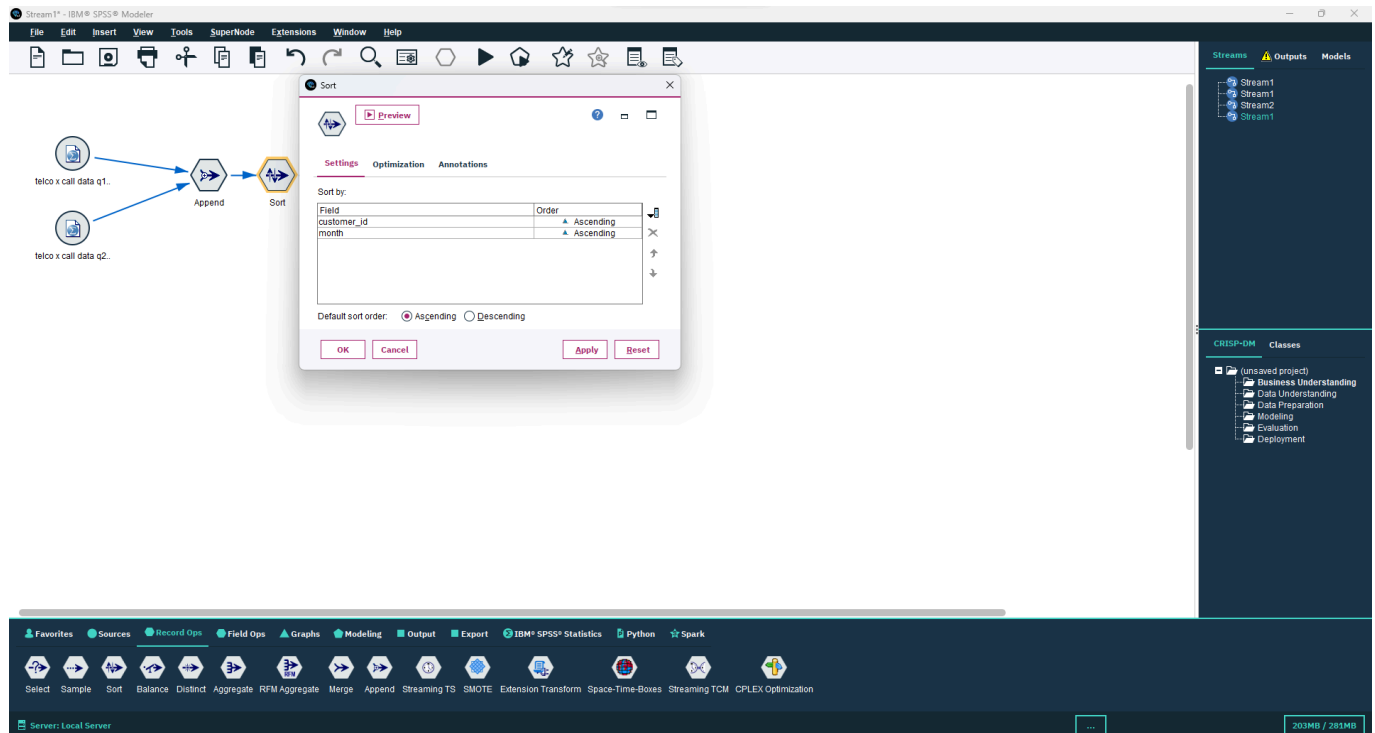


Step 3: We connect both of these files to Append node from Records Ops. click on apply and ok.

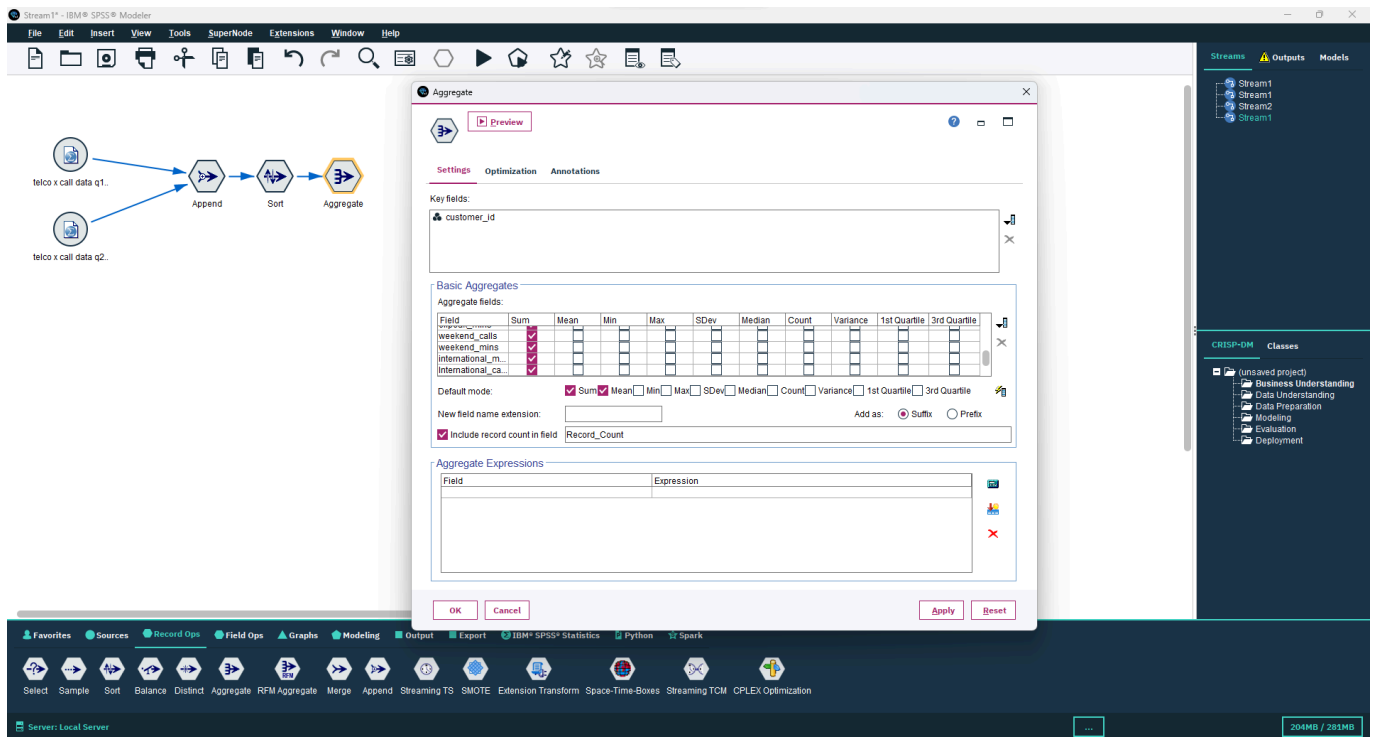


Step 4: Now we add sort node from Record Ops and connect it to Append node.

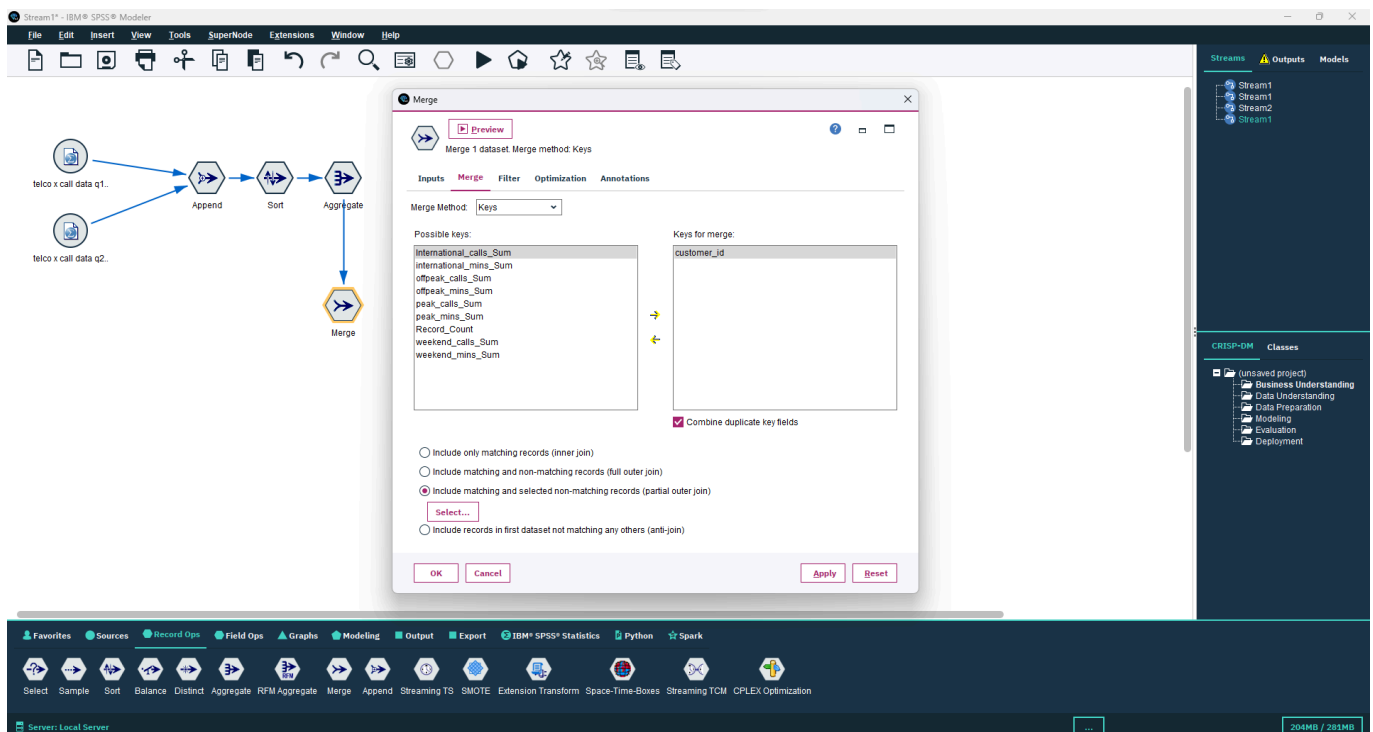
By double clicking on sort node a window opens in which under Sort by we select customer_id and month fields and sort them in ascending order.



Step 5: We add Aggregate node from Record ops and connect it to sort node. We customer_id as Key Field and all other fields should be selected in Aggregate fields and all should have only sum box as checked.



Step 6: Now add Merge node from Record ops and connect it to Aggregate node. Select customer_id as the key for merge and partial outer join by double clicking on Merge node.



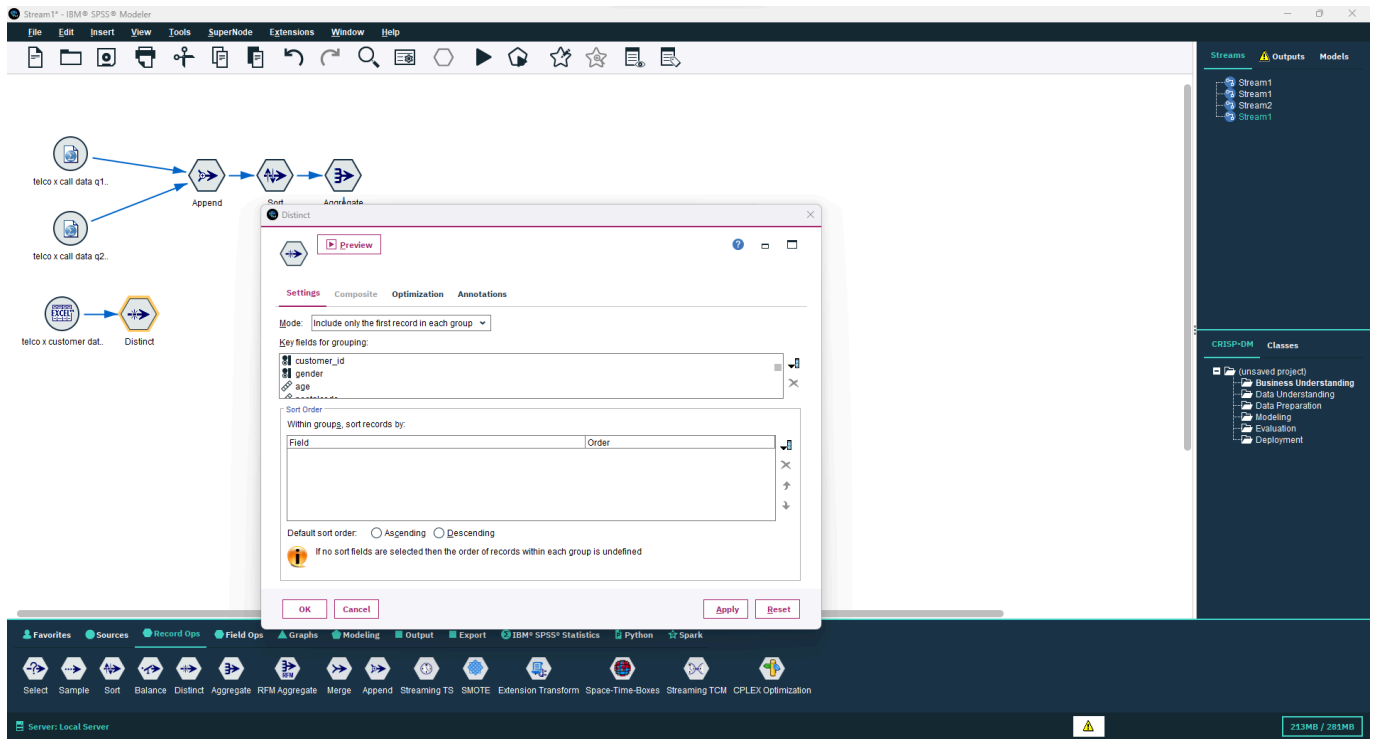
Step 7: We connect table node to see any change in data set.

	customer_id	peak_calls_Sum	peak_mins_Sum	offpeak_calls_Sum	offpeak_mins_Sum	weekend_calls_Sum	weekend_mins_Sum	international_mins_Sum	International_calls_Sum	Record_Count
1	K100010	14.000	36.131	10.000	7.973	24.000	14.533	0.705	7.000	6
2	K100020	54.000	39.437	34.000	21.153	0.000	0.000	4.609	0.000	6
3	K100030	44.000	72.600	1.000	27.600	22.000	37.200	7.700	0.000	6
4	K100040	44.000	72.600	1.000	27.600	22.000	37.200	8.621	1.000	6
5	K100050	32.000	40.608	14.000	18.824	1.000	1.234	2.494	0.000	6
6	K100060	56.000	46.260	6.000	11.085	8.000	10.233	0.858	0.000	6
7	K100070	68.000	56.370	6.000	4.012	1.000	4.012	3.189	0.000	6
8	K100080	40.000	51.043	13.000	34.802	5.000	14.230	6.734	0.000	6
9	K100090	54.000	99.000	34.000	53.100	0.000	0.000	12.404	0.000	6
10	K100100	37.000	65.400	17.000	29.400	0.000	0.000	2.074	0.000	6
11	K100110	37.000	65.400	17.000	29.400	0.000	0.000	2.062	0.000	6
12	K100120	61.000	68.123	34.000	33.880	8.000	8.334	6.031	0.000	6
13	K100130	73.000	63.352	31.000	26.803	1.000	3.046	12.062	0.000	6
14	K100140	40.000	41.447	23.000	16.133	16.000	14.209	8.364	4.000	6
15	K100150	54.000	96.000	9.000	34.800	14.000	22.800	25.436	2.000	6
16	K100160	15.000	41.342	17.000	23.841	6.000	13.189	3.675	0.000	6
17	K100170	62.000	100.200	17.000	31.500	5.000	9.000	2.547	1.000	6
18	K100180	57.000	41.918	32.000	21.827	6.000	11.988	10.687	0.000	6
19	K100190	64.000	58.782	12.000	28.749	2.000	3.533	9.348	0.000	6
20	K100200	50.000	84.600	15.000	27.900	21.000	42.000	6.155	1.000	6
21	K100210	50.000	84.600	15.000	27.900	21.000	42.000	6.355	0.000	6

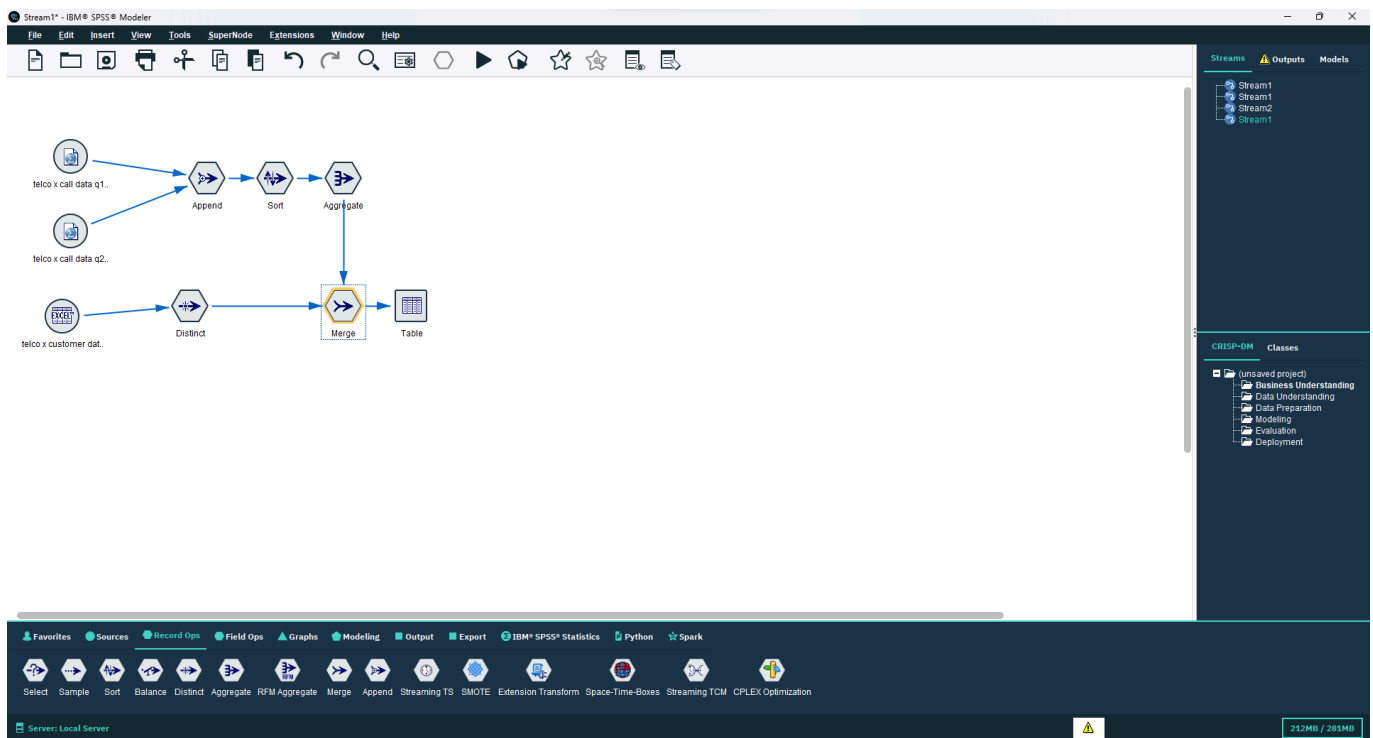
Step 8: Now we import a new data set telco x customer data.xlsx which is an excel file.

The screenshot shows the IBM SPSS Modeler interface. A workflow is visible with three nodes: 'Append', 'Sort', and 'Aggregate'. Two data sources, 'telco x call data q1' and 'telco x call data q2', are connected to the 'Append' node. A third data source, 'telco x customer data', is connected to the 'Sort' node. A dialog box titled 'telco x customer data.xlsx' is open, showing the 'Data' tab. The 'File type' is set to 'Excel 2007-2016 (*.xlsx)'. The 'Import file' is 'C:\Users\wohlf\Downloads\telco x customer data.xlsx'. The 'Lines to scan for column and type' is set to '200'. The 'Choose worksheet' section has 'By index' selected with a value of '0'. The 'Range on worksheet' section has 'Range starts on first non-blank row' selected. The 'On blank rows' section has 'Stop reading' selected. The 'First row has column names' checkbox is checked. The 'OK' button is highlighted.

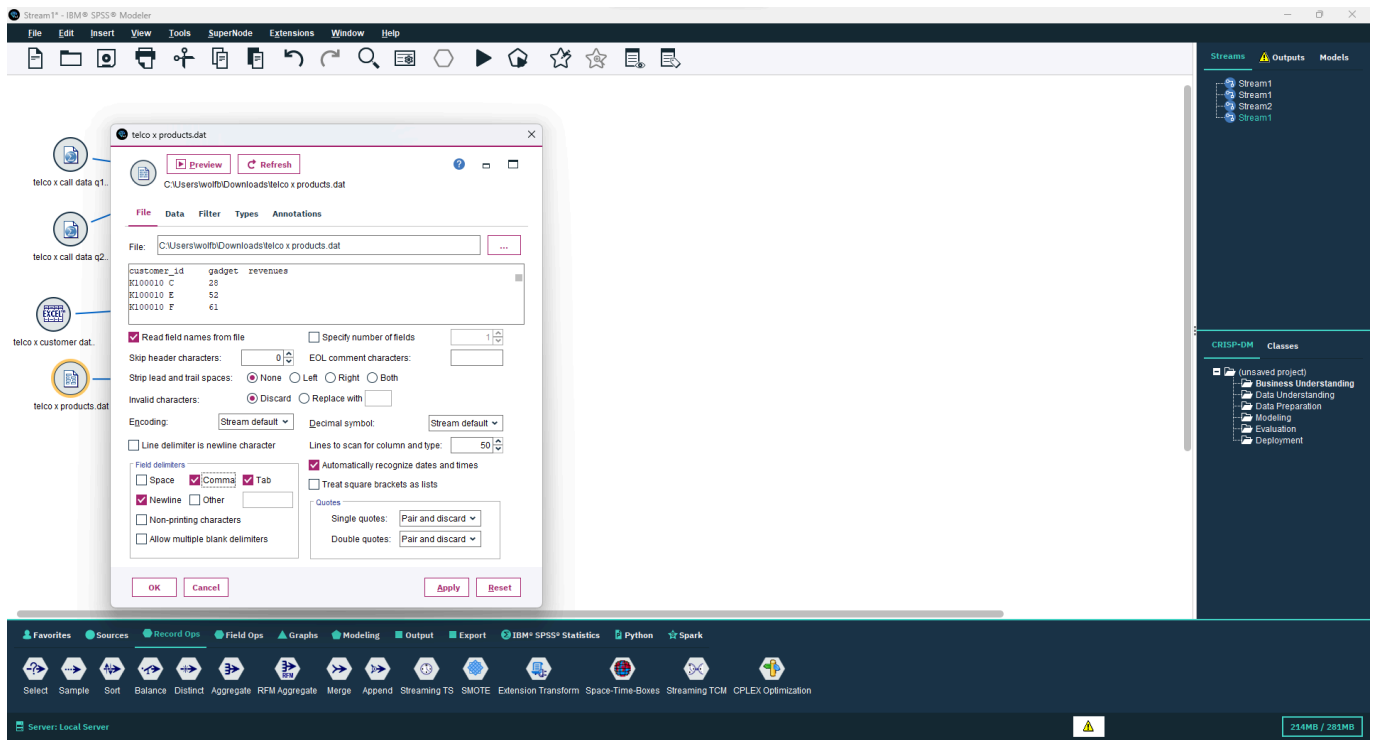
Step 9: Connect a Distinct node from Record ops and take all the fields of the data set as key fields in Distinct node by double clicking the Distinct node. And select Include only first record from each group.



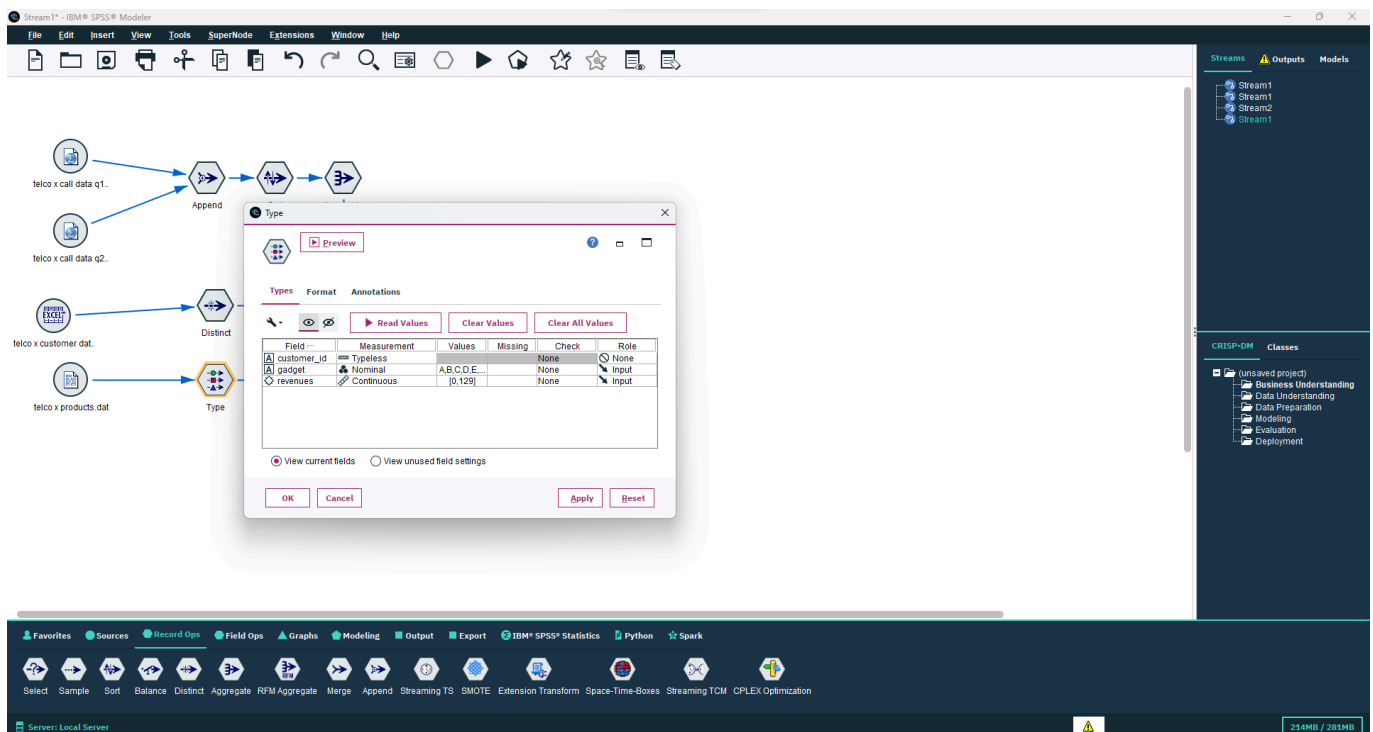
Step 10: Connect the Distinct node to Merge node.



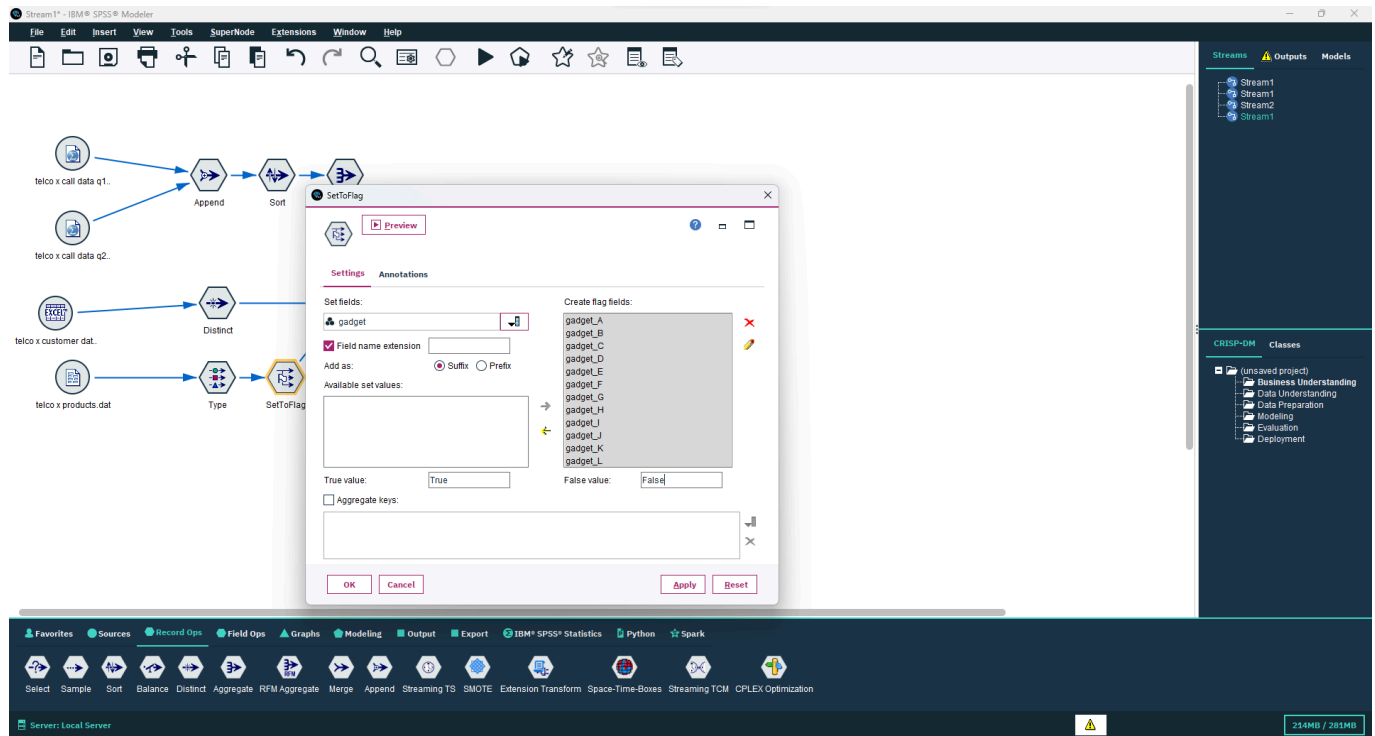
Step 11: Take a var file node from Sources Category and import telco x products.dat.



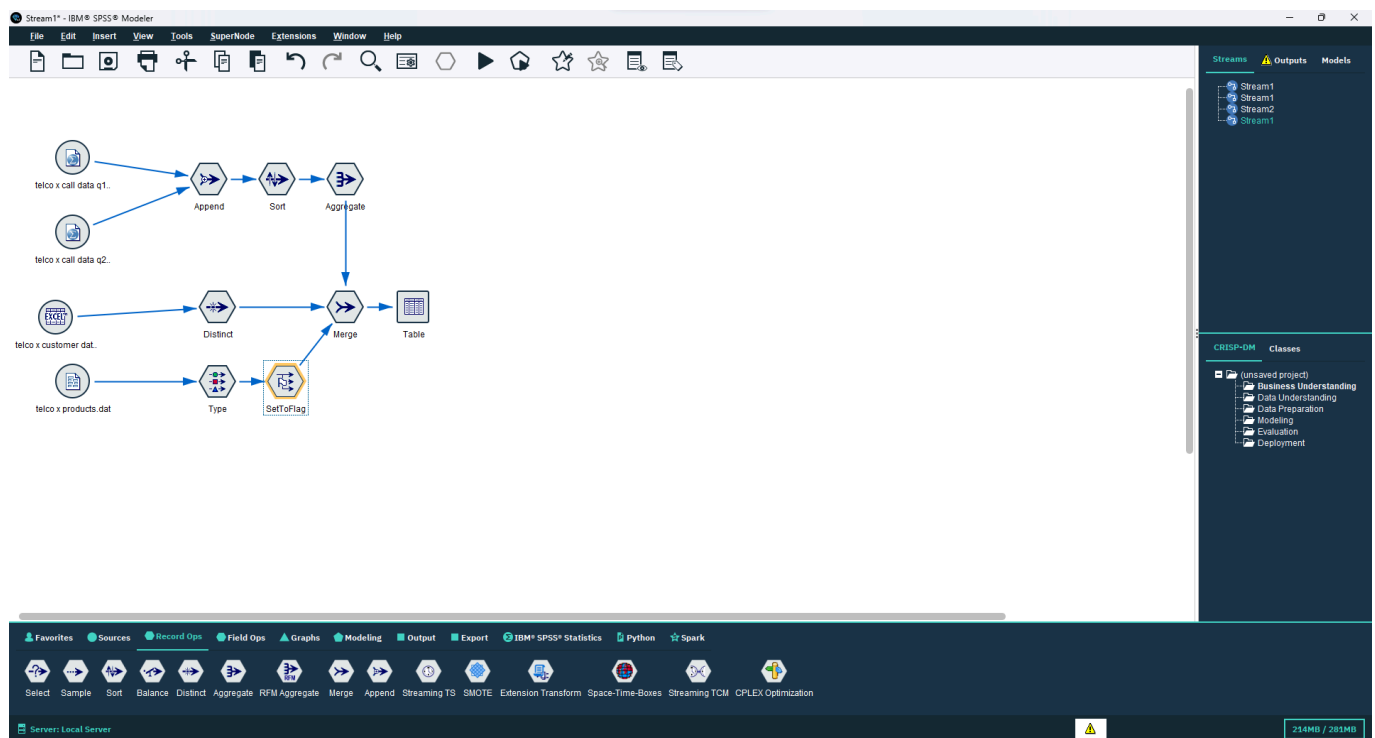
Step 12: Connect Type node to the new var file and get the specific category of all values by clicking on Read Values button in Type node window.



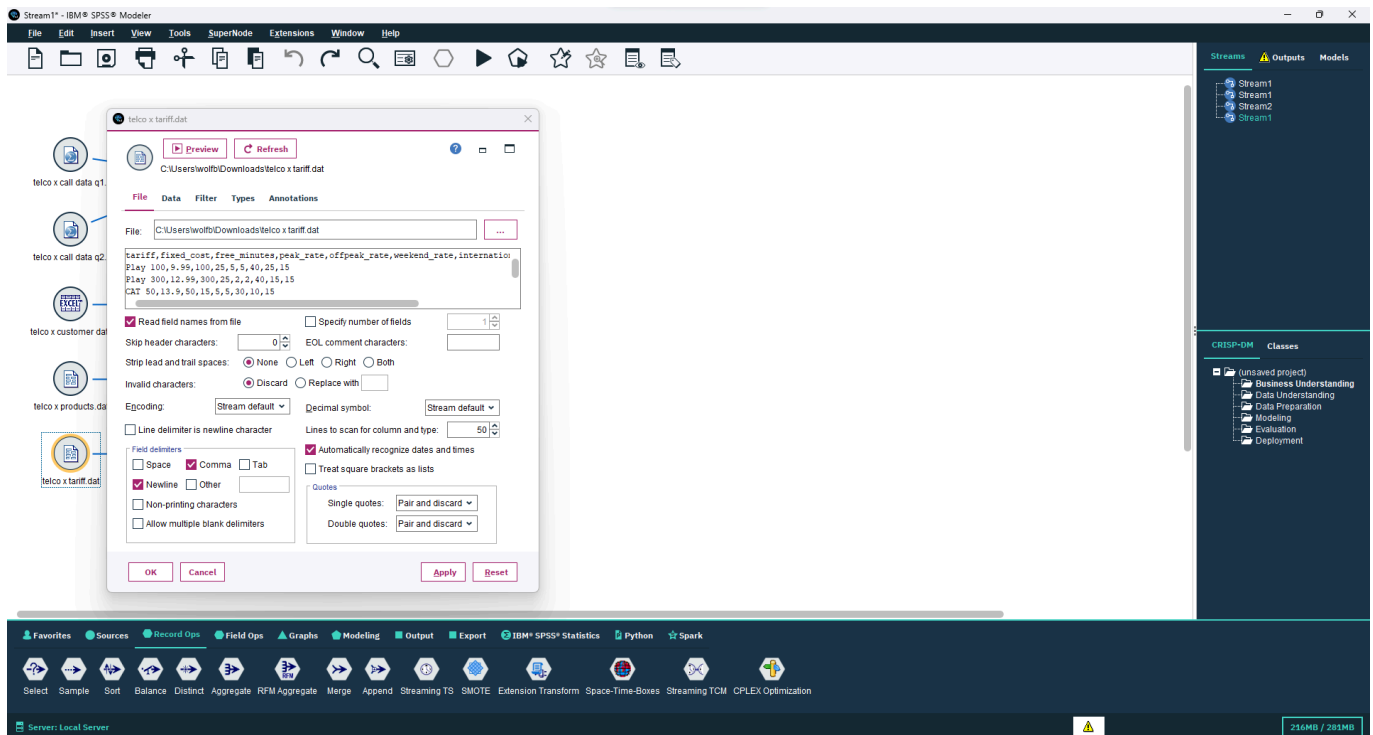
Step 13: Now we connect type node to Set to flag node and select gadget in Set fields section in Set to flag window and all the values should be selected and sent to Create Flag Fields.



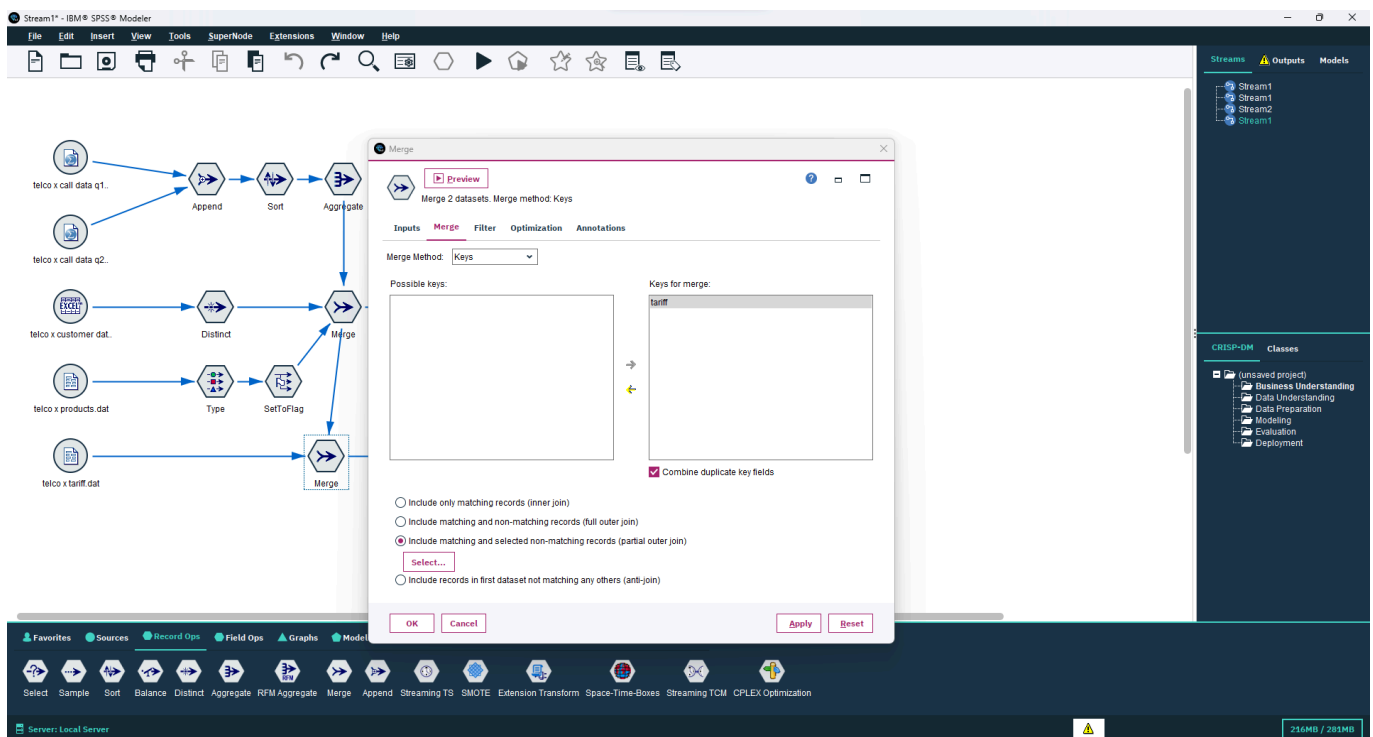
Step 14: Connect the Set to Flag node to Merge node.

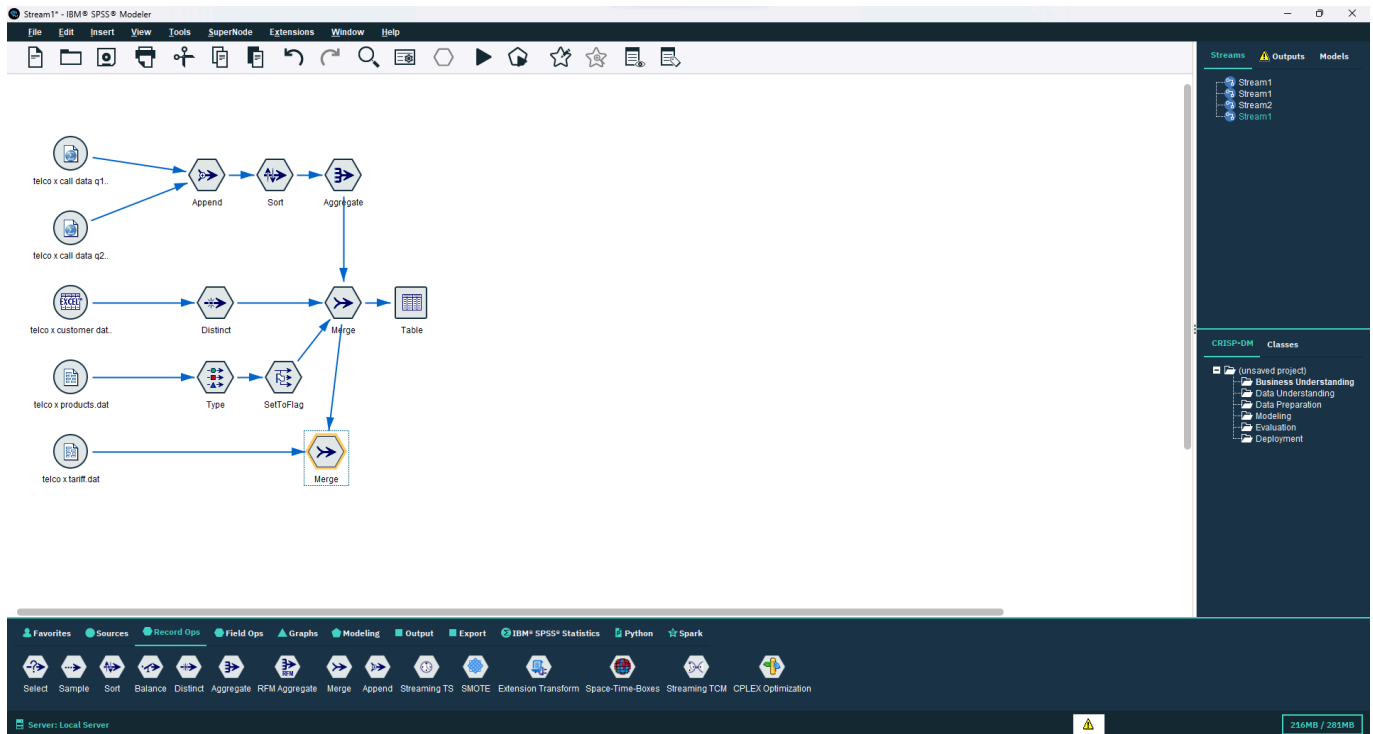


Step 15: Now we add another var file telco x tariff.dat



Step 16: Connect Merge node to the new var file and take tariff as key field and partial outer join in the Merge node window then connect the previous Merge Node to the new one Which is connected to telco x tariff.dat.



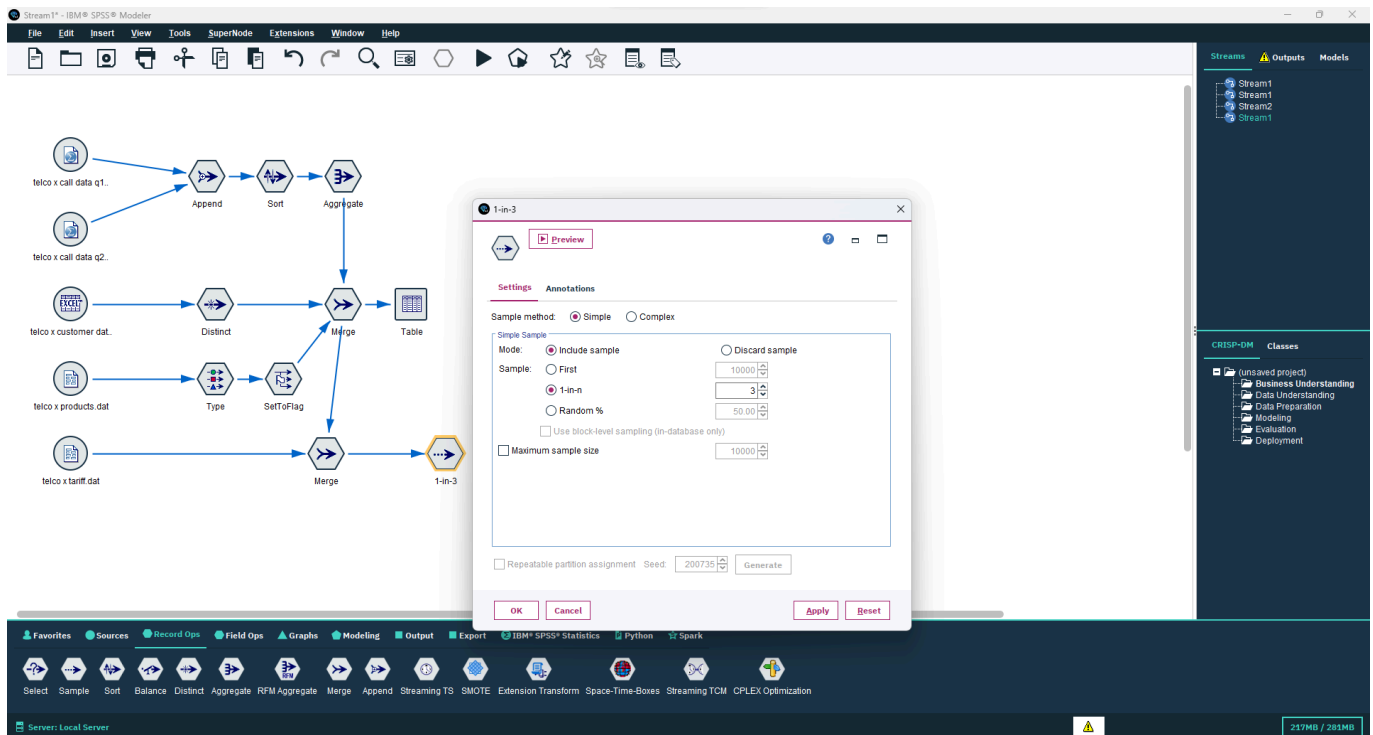


Step 17: Then connect the sample node from the Record ops to the Merge node connected to telco x tariff.dat.

Sample node is used to take a part of data from a data set to test the data set.

Open the Sample node window by double clicking on it and now select 1 in n rule now we choose a integer in place of n.

If we choose 3 in place of n the result will show the first record and then the third record skipping the record on the second place.



Step 18: Connect Table node to Sample node to view the final result which will show sample data of the four joint data sets.

Table (43 fields, 50,463 records)																										
File		Edit		Generate																						
Table		Annotations																								
		tariff	fixed_cost	free_minutes	peak_rate	offpeak_rate	weekend_rate	international_rate	voicemail	S	customer_id	gender	age	postalcode	region	connect_date	end_date	dropped_calls	paymethod	handset	churn	peak_calls_Sum	peak_mins_Sum	offpeak_calls_Sum	peak_calls_Sum	
1	CAT 100	17.500	100	15	5	5	5	30	10	159110780	Male	2...	4548.000	2.000	2004-09-25	2004-09-25	1.000	Free Pay	850	Active	13.000	664.200	72.00			
2	CAT 100	17.500	100	15	5	5	5	30	10	159110770	Male	3...	1112.000	1.000	2003-09-25	2003-09-25	6.000	Free Pay	850	Active	13.000	668.400	124.00			
3	CAT 100	17.500	100	15	5	5	5	30	10	159110770	Male	3...	1112.000	1.000	2003-09-25	2003-09-25	6.000	Free Pay	850	Active	13.000	668.400	124.00			
4	CAT 100	17.500	100	15	5	5	5	30	10	159110760	Female	4...	8287.000	4.000	2003-01-22	2004-03-19	2.000	Free Pay	ASAD90	Churned	8.000	438.600	157.00			
5	CAT 100	17.500	100	15	5	5	5	30	10	159110760	Female	4...	8287.000	4.000	2003-01-22	2004-03-19	2.000	Free Pay	ASAD90	Churned	8.000	438.600	157.00			
6	CAT 100	17.500	100	15	5	5	5	30	10	159110750	Male	3...	8960.000	4.000	2003-08-03	2003-08-03	2.000	Free Pay	850	Active	2.000	826.200	4.00			
7	CAT 100	17.500	100	15	5	5	5	30	10	159110740	Male	3...	3914.000	2.000	2004-05-27	2004-05-27	1.000	Free Pay	850	Active	1.000	461.400	2.00			
8	CAT 100	17.500	100	15	5	5	5	30	10	159110740	Male	3...	3914.000	2.000	2004-05-27	2004-05-27	1.000	Free Pay	850	Active	1.000	461.400	2.00			
9	CAT 100	17.500	100	15	5	5	5	30	10	159110730	Female	4...	6245.000	3.000	2003-04-09	2004-10-04	2.000	Free Pay	ASAD90	Churned	9.000	464.400	85.00			
10	CAT 100	17.500	100	15	5	5	5	30	10	159110720	Female	2...	4191.000	2.000	2005-06-27	2005-06-27	2.000	Free Pay	850	Active	2.000	473.400	39.00			
11	CAT 100	17.500	100	15	5	5	5	30	10	159110720	Female	2...	4191.000	2.000	2005-06-27	2005-06-27	2.000	Free Pay	850	Active	2.000	473.400	39.00			
12	CAT 100	17.500	100	15	5	5	5	30	10	159110720	Female	2...	4191.000	2.000	2005-06-27	2005-06-27	2.000	Free Pay	850	Active	2.000	473.400	39.00			
13	CAT 100	17.500	100	15	5	5	5	30	10	159110710	Male	3...	2167.000	1.000	2003-04-20	2003-04-20	0.000	Free Pay	ASAD10	Active	2.000	417.600	147.00			
14	CAT 100	17.500	100	15	5	5	5	30	10	159110700	Male	4...	9763.000	4.000	2005-05-19	2005-05-19	0.000	Free Pay	850	Active	3.000	719.400	90.00			
15	CAT 100	17.500	100	15	5	5	5	30	10	159110690	Male	4...	9919.000	4.000	2004-06-01	2004-06-01	2.000	Free Pay	850	Active	3.000	644.400	59.00			
16	CAT 100	17.500	100	15	5	5	5	30	10	159110690	Male	4...	9919.000	4.000	2004-06-01	2004-06-01	2.000	Free Pay	850	Active	3.000	644.400	59.00			
17	CAT 100	17.500	100	15	5	5	5	30	10	159110680	Female	4...	5068.000	3.000	2005-12-14	2008-05-21	1.000	Free Pay	ASAD90	Churned	8.000	522.000	63.00			
18	CAT 100	17.500	100	15	5	5	5	30	10	159110680	Female	4...	5068.000	3.000	2005-12-14	2008-05-21	1.000	Free Pay	ASAD90	Churned	8.000	522.000	63.00			
19	CAT 100	17.500	100	15	5	5	5	30	10	159110680	Female	4...	5068.000	3.000	2005-12-14	2008-05-21	1.000	Free Pay	ASAD90	Churned	8.000	522.000	63.00			
20	CAT 100	17.500	100	15	5	5	5	30	10	159110670	Female	5...	1054.000	1.000	2004-08-26	2007-08-25	1.000	Free Pay	ASAD90	Churned	10.000	459.600	79.00			
21	CAT 100	17.500	100	15	5	5	5	30	10	159110670	Female	5...	1054.000	1.000	2004-08-26	2007-08-25	1.000	Free Pay	ASAD90	Churned	10.000	459.600	79.00			
22	CAT 100	17.500	100	15	5	5	5	30	10	159110660	Female	3...	8800.000	4.000	2004-01-03	2006-01-21	11.000	Free Pay	ASAD90	Churned	3.000	637.200	20.00			
23	CAT 100	17.500	100	15	5	5	5	30	10	159110650	Female	5...	3225.000	2.000	2003-01-01	2003-01-01	1.000	Free Pay	ASAD10	Active	2.000	436.200	163.00			
24	CAT 100	17.500	100	15	5	5	5	30	10	159110650	Female	5...	3225.000	2.000	2003-01-01	2003-01-01	1.000	Free Pay	ASAD10	Active	2.000	436.200	163.00			
25	CAT 100	17.500	100	15	5	5	5	30	10	159110640	Male	3...	9303.000	4.000	2005-10-02	2005-10-02	8.000	Free Pay	ASAD10	Active	3.000	533.400	22.00			
26	CAT 100	17.500	100	15	5	5	5	30	10	159110640	Male	3...	9303.000	4.000	2005-10-02	2005-10-02	8.000	Free Pay	ASAD10	Active	3.000	533.400	22.00			
27	CAT 100	17.500	100	15	5	5	5	30	10	159110630	Female	3...	9087.000	4.000	2003-08-09	2003-08-09	2.000	Free Pay	850	Active	12.000	579.600	20.00			
28	CAT 100	17.500	100	15	5	5	5	30	10	159110620	Female	3...	4644.000	2.000	2006-12-26	2006-12-26	1.000	Free Pay	ASAD10	Active	4.000	603.000	34.00			
29	CAT 100	17.500	100	15	5	5	5	30	10	159110610	Female	2...	2218.000	1.000	2005-09-06	2009-09-13	0.000	Free Pay	ASAD90	Churned	9.000	549.000	19.00			
30	CAT 100	17.500	100	15	5	5	5	30	10	159110610	Female	2...	2218.000	1.000	2005-09-06	2009-09-13	0.000	Free Pay	ASAD90	Churned	9.000	549.000	19.00			
31	CAT 100	17.500	100	15	5	5	5	30	10	159110610	Female	2...	2218.000	1.000	2005-09-06	2009-09-13	0.000	Free Pay	ASAD90	Churned	9.000	549.000	19.00			
32	CAT 100	17.500	100	15	5	5	5	30	10	159110600	Male	3...	4097.000	2.000	2003-06-06	2004-04-12	2.000	Free Pay	ASAD90	Churned	12.000	535.200	55.00			
33	CAT 100	17.500	100	15	5	5	5	30	10	159110600	Male	3...	4097.000	2.000	2003-06-06	2004-04-12	2.000	Free Pay	ASAD90	Churned	12.000	535.200	55.00			
34	CAT 100	17.500	100	15	5	5	5	30	10	159110590	Male	5...	5627.000	3.000	2003-01-15	2004-11-22	2.000	Free Pay	ASAD90	Churned	14.000	682.000	19.00			
35	CAT 100	17.500	100	15	5	5	5	30	10	159110580	Male	---	2247.000	1.000	2005-10-15	2008-07-14	4.000	Free Pay	850	Churned	12.000	678.600	78.00			
36	CAT 100	17.500	100	15	5	5	5	30	10	159110580	Male	---	2247.000	1.000	2005-10-15	2008-07-14	4.000	Free Pay	850	Churned	12.000	678.600	78.00			
37	CAT 100	17.500	100	15	5	5	5	30	10	159110570	Male	4...	7611.000	4.000	2005-06-11	2009-09-16	2.000	Free Pay	ASAD90	Churned	11.000	540.000	99.00			
38	CAT 100	17.500	100	15	5	5	5	30	10	159110560	Female	4...	1580.000	1.000	2005-08-22	2005-08-22	1.000	Free Pay	850	Active	5.000	506.400	38.00			
39	CAT 100	17.500	100	15	5	5	5	30	10	159110550	Male	4...	5675.000	3.000	2005-12-18	2007-01-22	11.000	Free Pay	ASAD90	Churned	7.000	587.400	39.00			
40	CAT 100	17.500	100	15	5	5	5	30	10	159110550	Male	4...	5675.000	3.000	2005-12-18	2007-01-22	11.000	Free Pay	ASAD90	Churned	7.000	587.400	39.00			
41	CAT 100	17.500	100	15	5	5	5	30	10	159110540	Female	2...	6053.000	3.000	2005-04-22	2009-07-01	3.000	Free Pay	850	Churned	12.000	560.400	44.00			
42	CAT 100	17.500	100	15	5	5	5	30	10	159110540	Female	2...	6053.000	3.000	2005-04-22	2009-07-01	3.000	Free Pay	850	Churned	12.000	560.400	44.00			
43	CAT 100	17.500	100	15	5	5	5	30	10	159110530	Male	5...	2015.000	1.000	2003-04-15	2003-04-15	2.000	Free Pay	850	Active	11.000	672.000	0.00			
44	CAT 100	17.500	100	15	5	5	5	30	10	159110530	Male	5...	2015.000	1.000	2003-04-15	2003-04-15	2.000	Free Pay	850	Active	11.000	672.000	0.00			
45	CAT 100	17.500	100	15	5	5	5	30	10	159110520	Female	2...	4795.000	2.000	2003-05-19	2009-02-14	0.000	Free Pay	ASAD10	Active	8.000	676.800	178.00			
46	CAT 100	17.500	100	15	5	5	5	30	10	159110510	Male	4...	8705.000	4.000	2003-07-18	2003-07-18	10.000	Free Pay	850	Active	8.000	765.600	78.00			
47	CAT 100	17.500	100	15	5	5	5	30	10	159110500	Male	2...	5686.000	3.000	2003-01-19	2007-02-14	11.000	Free Pay	ASAD90	Churned	189.000	729.000	5.00			
48	CAT 100	17.500	100	15	5	5	5	30	10	159110490	Male	5...	4255.000	2.000	2006-06-25	2006-06-25	1.000	Free Pay	850	Active	206.000	630.000	106.00			
49	CAT 100	17.500	100	15	5	5	5	30	10	159110490	Male	5...	4255.000	2.000	2006-06-25	2006-06-25	1.000	Free Pay	850	Active	206.000	630.000	106.00			
50	CAT 100	17.500	100	15	5	5	5	30	10	159110480	Male	3...	8861.000	4.000	2003-11-27	2003-11-27	1.000	Free Pay	850	Active	103.					

