## Babu Banarasi Das University



## Predictive Analytics (BCADS15301)

LAB FILE

SUBMITTED TO: Mr. VIKASH **SUBMITTED BY:-**

Name: AYUSH YADAV

Roll No: 1230258136

**Class: BCADS31** 

## Practical Work: Data Integration and Analysis

**Definition:** As part of a telecommunications project, you are required to consolidate several datasets into one master dataset to facilitate data analysis and model development.

Outcomes/Learning: Learning how to join different data sets

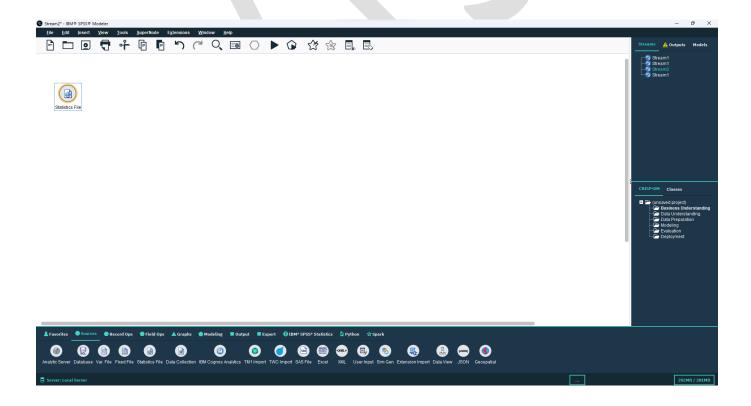
How to join records of different data sets

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.

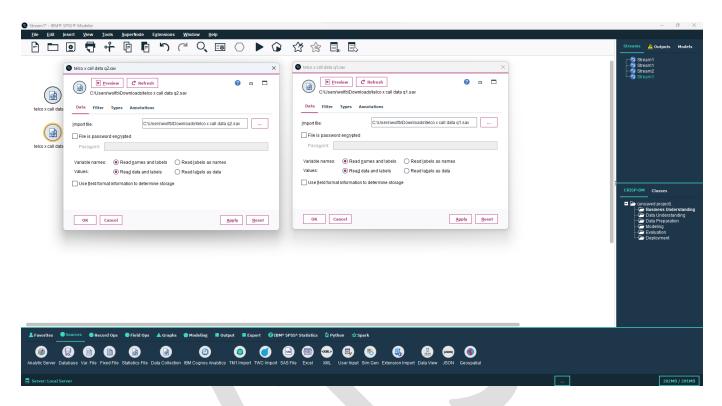
**Step 1:** Open SPSS Modeler tool then on Source category select statistics File node (we are selecting Statistics node because the data set we are using is an sav file.) Double clicking on statistics node will make it appear on canvas



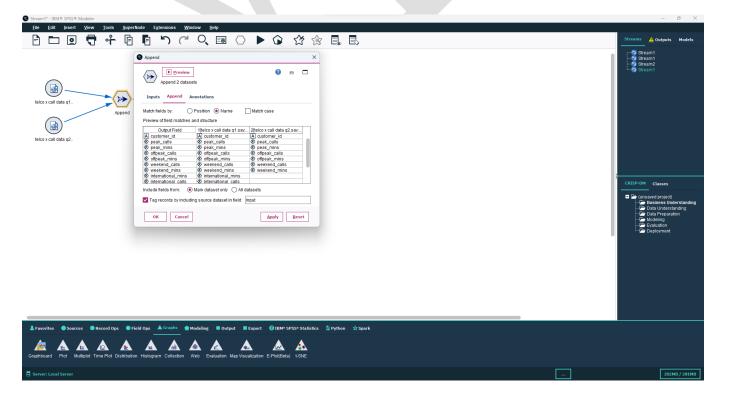
**Step 2:** Now we import a data set using the import option which can be accessed by double clicking on Statistics icon on the canvas.

We import a data set telco x call data q1.sav then import one more statistics file tele x call

data q2.sav.

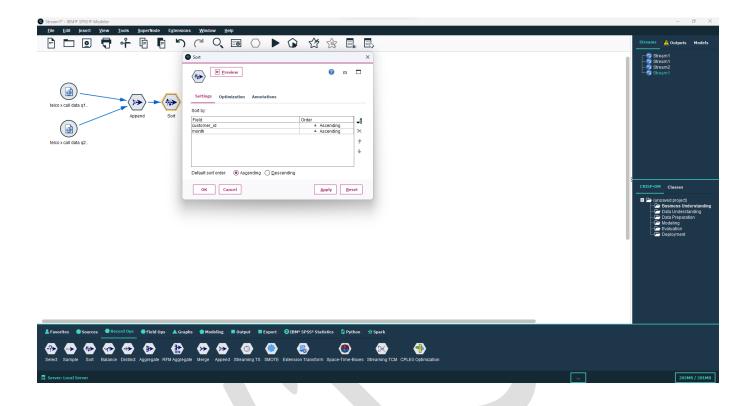


**Step 3:** We connect both of these files to Append node from Records Ops. click on apply and ok.

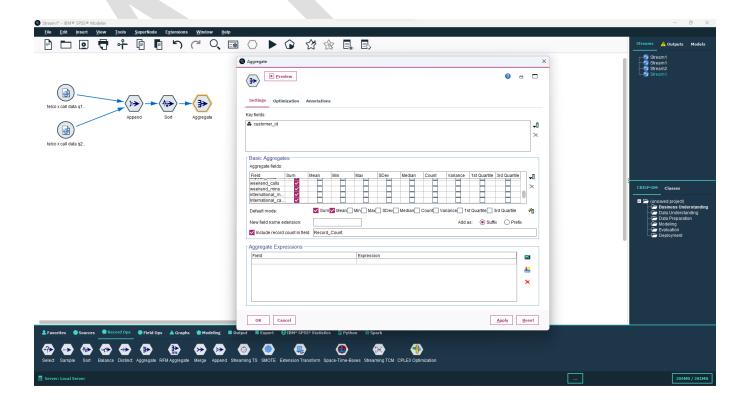


Step 4: Now we add sort node from Record Ops and connect it to Append node.

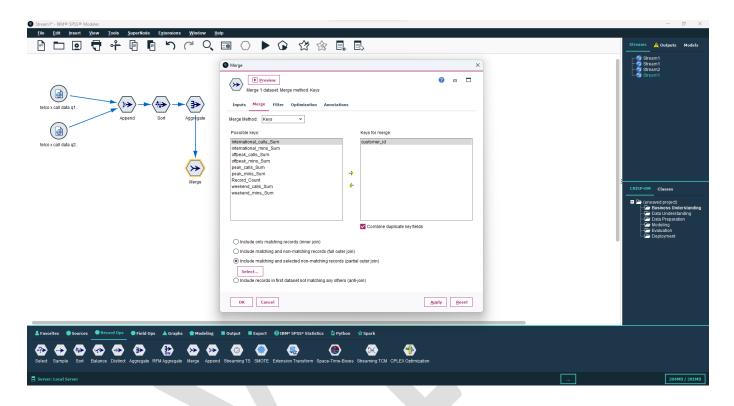
By double clicking on sort node a window opens in which under Sort by we select customer id and month fields and sort them in ascending order.



**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.



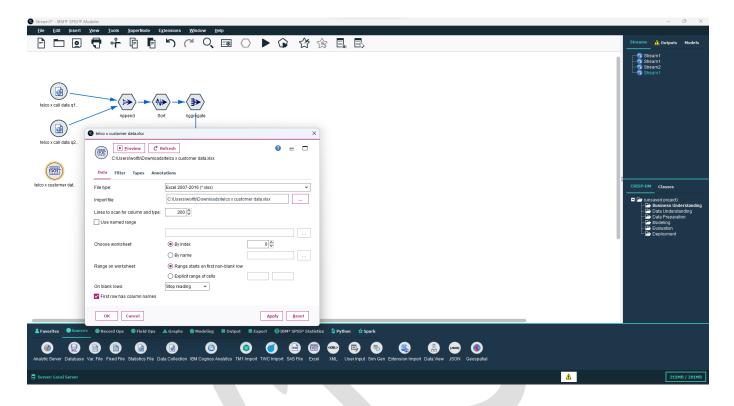
**Step 6:** Now add Merge node from Record ops and connect it to Aggregate node. Select customer\_id as the key for merge and partial outer join by double clicking on Merge node.



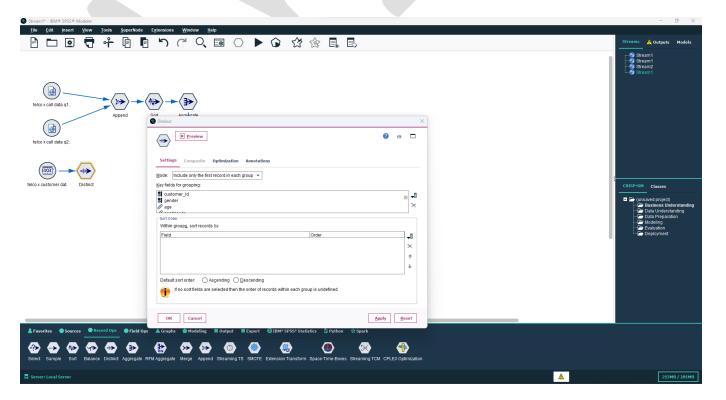
Step 7: We connect table node to see any change in data set.

Table	(10 fields, 31,76	9 records)							-	- 0	Χ
<u> </u>	<u></u> <u>E</u> dit	🖔 <u>G</u> enerate	<b>8 7</b> 4 Q							?	X
Table	Annotations										
	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_Cou	unt
1	K100010	14.000		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

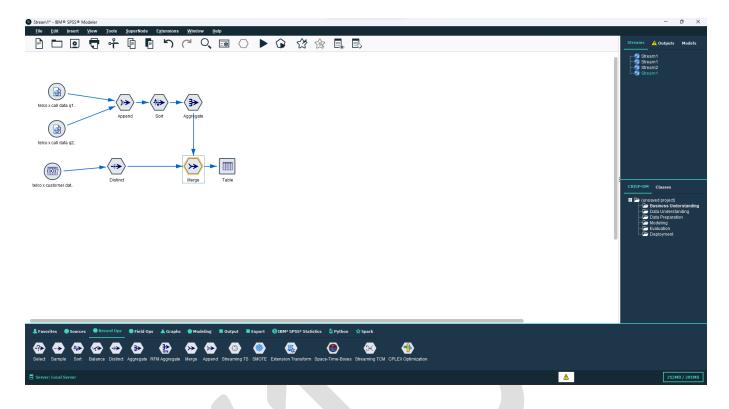
Step 8: Now we import a new data set telco x customer data.xlxs which is an excel file.



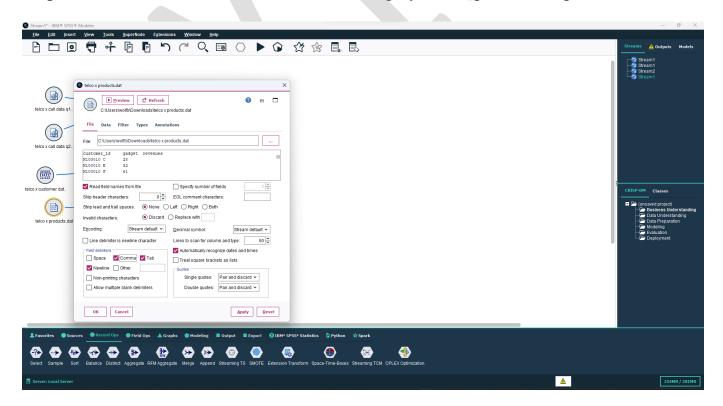
**Step 9:** Connect a Distinct node from Record ops and take all the fields of the data set as key fields in Distinct node by double clicking the Distinct node. And select Include only first record from each group.



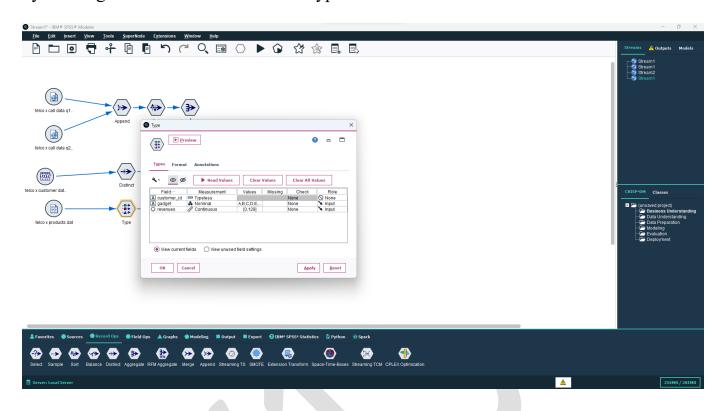
**Step 10:** Connect the Distinct node to Merge node.



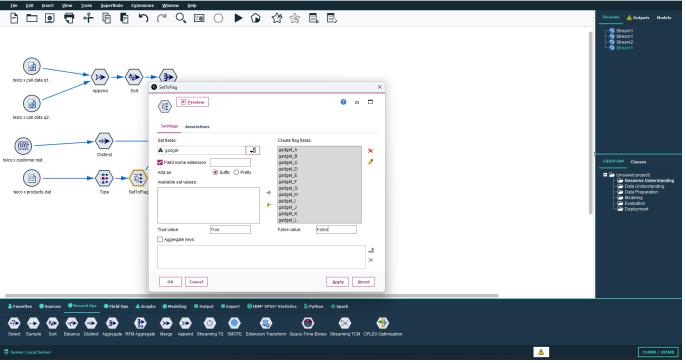
**Step 11:** Take a var file node from Sources Category and import telco x products.dat.



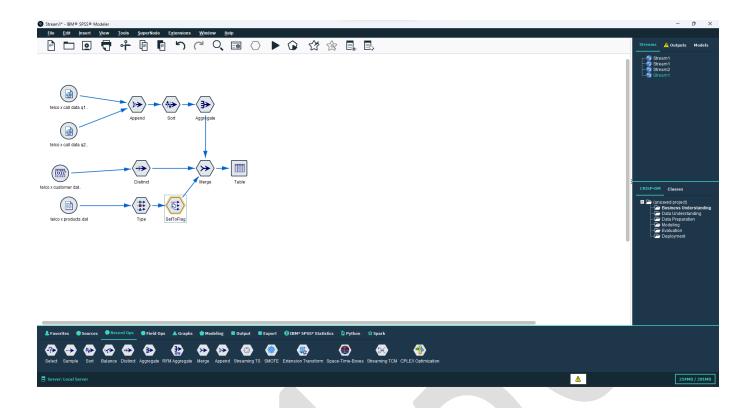
**Step 12:** Connect Type node to the new var file and get the specific category of all values by clicking on Read Values button in Type node window.



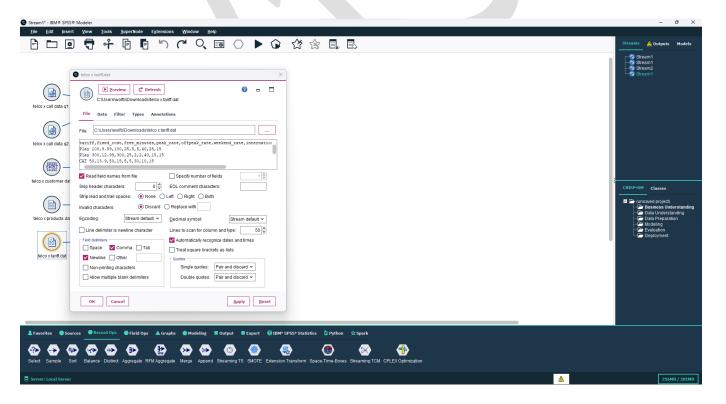
**Step 13:** Now we connect type node to Set to flag node and select gadget in Set fields section in Set to flag window and all the values should be selected and sent to Create Flag Fields.



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

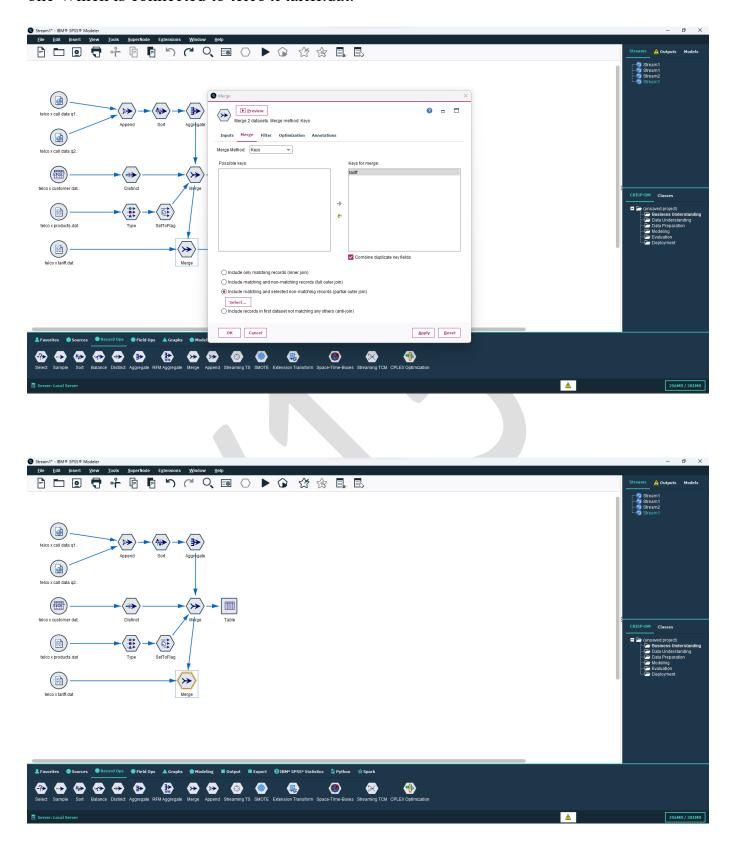


Step 15: Now we add another var file telco x tariff.dat



**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.

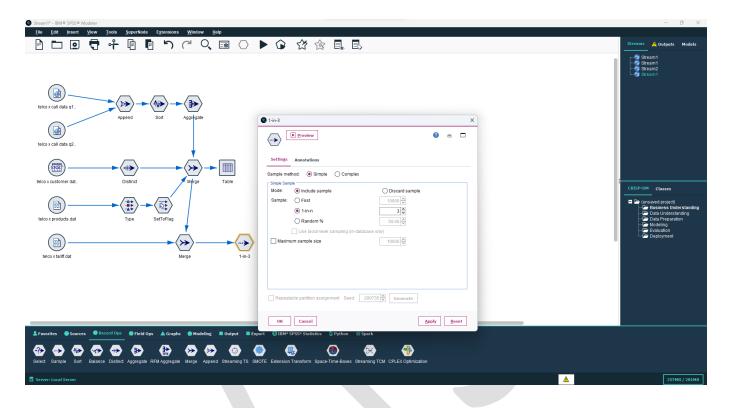


**Step 17:** Then connect the sample node from the Record ops to the Merge node connected to telco x tariff.dat.

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.

