Lab Record

of

PREDICTIVE ANALYTICS MODELER (IB411)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Session 2022-23

Submitted to: -

Mrs. Rohini Matta
Assistant Professor,
School of Computing

Submitted by: -

Devesh Tripathi
190111022
1000012472
B.Tech CSE-CSF

SCHOOL OF COMPUTING
DIT UNIVERSITY, DEHRADUN
(State Private University through State Legislature Act No. 10 of 2013 of Uttarakhand and approved by UGC)
Mussoorie Diversion Road, Dehradun, Uttarakhand - 248009, India.

Aug - Dec.2022

INDEX

S.NO.	NAME OF THE EXPERIMENTS	DATE OF CONDUCTION	DATE OF SUBMISSION	REMARKS
1	Create a data-mining project to predict churn in telecommunications.	24/08/2022	24/08/2022	
2	You work as a data miner for a telecommunications firm. You have to import data from various sources and examine the unit of analysis and the measurement levels of the fields.	31/08/2022	31/08/2022	
3	You work as a data miner for a telecommunications firm. You have imported data, and you need to assess the quality of the data and define blanks where appropriate.	07/09/2022	07/09/2022	
4	You work for a telecommunications firm and have to combine a number of datasets into a single dataset for analyses and modelling.	14/09/2022	14/09/2022	
5	You work for a telecommunications firm and you need to cleanse and enrich a dataset in order to build models	21/09/2022	21/09/2022	
6	Identify relationships in the telecommunications data.	28/09/2022	28/09/2022	
7	Predict churn in telecommunications and cluster customers into segments	19/10/2022	19/10/2022	
8	Use CLEM functions to cleanse and enrich telecommunications data.	02/11/2022	02/11/2022	
9	Use additional field transformations to prepare the telecommunications data for modelling.	09/11/2022	09/11/2022	
10	Prepare sequence data of a bank for analysis.	16/11/2022	16/11/2022	

Experiment - 1

Objective : Create a data-mining project to predict churn in telecommunications.

PROCEDURE:

1. Drag and drop excel node from sources palette, and import telco x modelling dataset inside it.
2. Generate a select node and set expression as data_known = "yes".
3. Add a type node and change the measurements of the fields as required.
4. Add CHAID node from model palette and choose target -> churn, and as required.
5. Add another excel node with telco x deployment dataset, and generate same data known node as we did above.
6. Connect this generated node with one of the nuggets of chaid node.
7. Generate another node with expression:
   '$R-churn' = "Active" and '$RC-churn' > 0.9 and connect with that nugget.
8. Add a filter node from field ops and select customer_id, $R-churn,$RC-churn only.
9. Add a report node from output palette which will show list of fields that we selected previously.

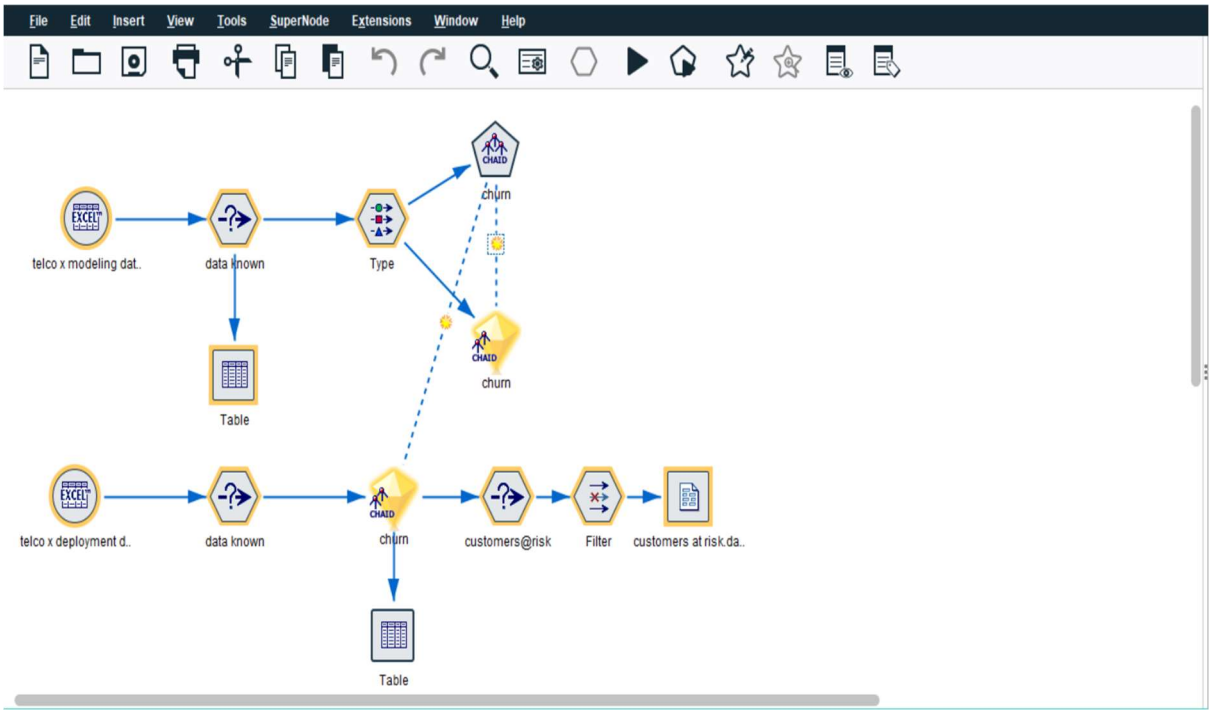


Table (23 fields, 9,994 records)

File Edit Generate

Table Annotations

	customer_id	data_known	gender	age	tariff	dropped_calls	handset	peak_mins
1	N247880	yes	Female	31....	CAT 50	0.000	BS110	393.600
2	N376450	yes	Female	42....	CAT 50	6.000	CAS30	136.200
3	N109880	yes	Female	62....	CAT 100	9.000	ASAD1...	657.600
4	N188070	yes	Male	24....	CAT 200	0.000	S50	1335.600
5	N201040	yes	Female	20....	CAT 200	2.000	ASAD1...	844.800
6	N164070	yes	Female	31....	Play 300	10.000	ASAD1...	336.600
7	N201620	yes	Male	15....	CAT 200	1.000	S50	1243.800
8	N408820	yes	Female	18....	CAT 200	12.000	S50	2346.000
9	N178490	yes	Male	25....	CAT 200	9.000	S50	976.800
10	N348730	yes	Male	31....	CAT 100	1.000	CAS30	481.200
11	N173280	yes	Female	17....	CAT 200	2.000	S50	927.600
12	N262370	yes	Male	49....	Play 100	0.000	BS210	64.200
13	N207500	yes	Female	14....	Play 300	2.000	S50	486.600
14	N204780	yes	Male	23....	Play 100	0.000	ASAD90	122.400
15	N286390	yes	Male	24....	CAT 100	0.000	BS110	537.000
16	N392950	yes	Male	57....	Play 300	9.000	CAS60	204.600
17	N273340	yes	Male	56....	Play 300	1.000	S80	304.800
18	N310150	yes	Female	13....	CAT 200	0.000	WC95	718.800
19	N313300	yes	Male	41....	CAT 200	2.000	S50	637.200
20	N271610	yes	Male	51....	Play 100	1.000	S80	36.600

OK

Table (23 fields, 31,769 records)

File Edit Generate

Table Annotations

	customer_id	data_known	gender	age	tariff	dropped_calls	handset	peak_mins
1	K100010	yes	Male	46....	CA...	1.000	SOP10	36.131
2	K100020	yes	Male	27....	CA...	0.000	SOP10	39.437
3	K100030	yes	Male	39....	CA...	2.000	SOP20	72.600
4	K100040	yes	Male	28....	CA...	2.000	SOP10	72.600
5	K100050	yes	Male	47....	CA...	0.000	SOP10	40.608
6	K100060	yes	Male	29....	CA...	1.000	SOP10	46.260
7	K100070	yes	Male	38....	CA...	1.000	SOP20	56.370
8	K100080	yes	Male	27....	CA...	0.000	SOP20	51.043
9	K100090	yes	Male	38....	CA...	1.000	SOP20	99.000
10	K100100	yes	Male	48....	CA...	0.000	SOP20	65.400
11	K100110	yes	Male	34....	CA...	2.000	SOP20	65.400
12	K100120	yes	Male	32....	CA...	0.000	SOP20	68.123
13	K100130	yes	Male	45....	CA...	2.000	SOP20	63.352
14	K100140	yes	Male	29....	CA...	1.000	SOP20	41.447
15	K100150	yes	Male	30....	CA...	2.000	SOP20	96.000
16	K100160	yes	Male	28....	CA...	9.000	SOP20	41.342
17	K100170	yes	Male	26....	CA...	1.000	SOP10	100.200
18	K100180	yes	Male	37....	CA...	1.000	SOP10	41.918
19	K100190	yes	Male	38....	CA...	2.000	SOP10	58.782
20	K100200	yes	Male	46....	CA...	0.000	SOP10	84.600

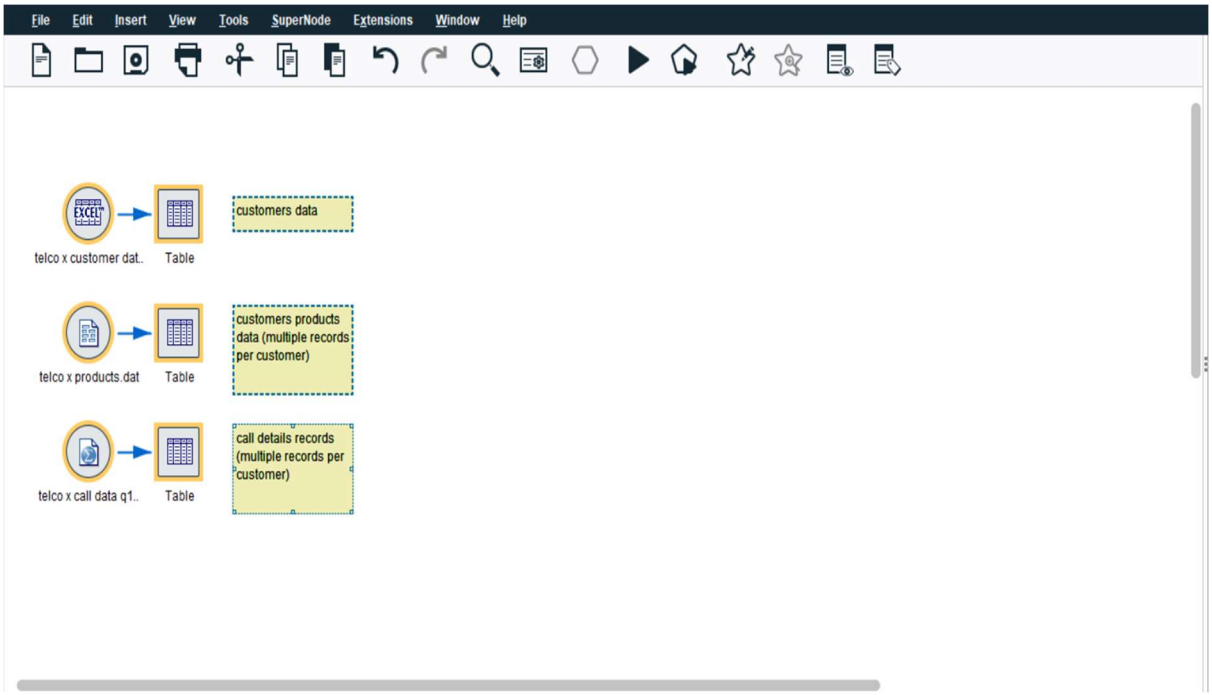
OK

Experiment - 2

Objective : You work as a data miner for a telecommunications firm. You have to import data from various sources and examine the unit of analysis and the measurement levels of the fields.

PROCEDURE:

1. Drag and drop 2 excel nodes from sources palette.
2. Import telco x customers dataset inside it.
3. Add and connect a table from output palette.
4. Import telco x products dataset in 2nd excel node.
5. Add and connect a table from output palette.
6. Add a statistics file from source and import telco x data q1 in it.
7. Add and connect a table from output palette.



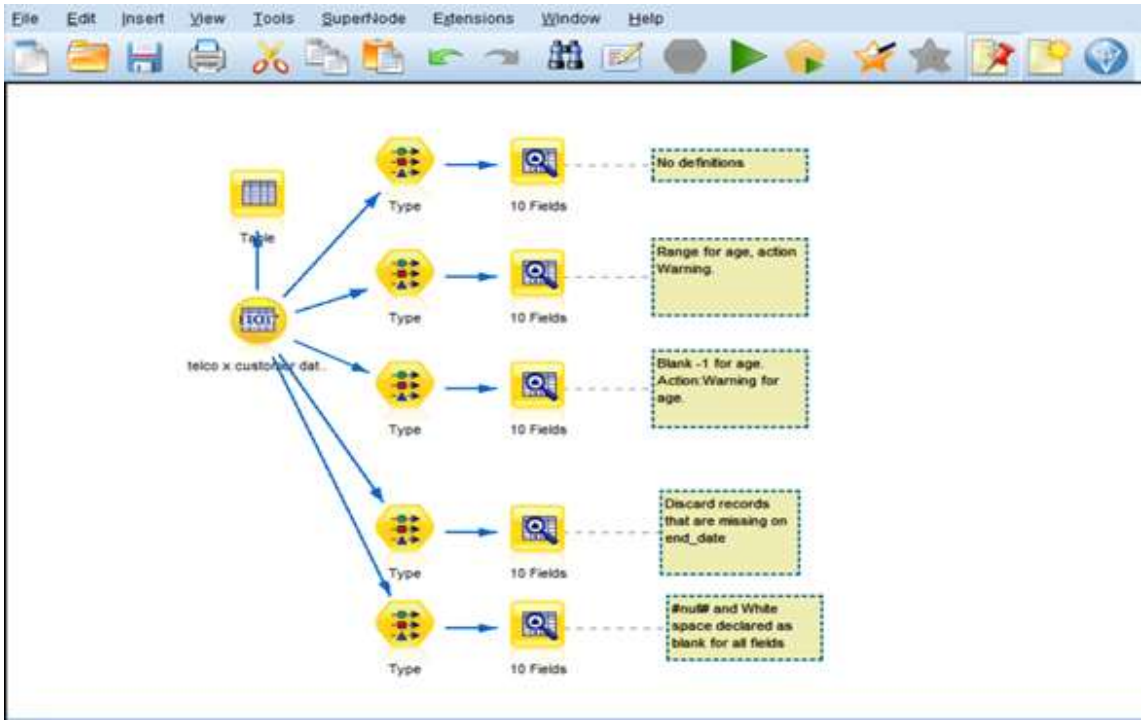


Table (12 fields, 31,781 records) #1

File Edit Generate

Table Annotations

	customer_id	gender	age	postalcode	region	connect_date	end_date	dropped_calls
1	K100010	Male	46....	6253.000	3.000	2006-07-18	2010-05-10	1.000
2	K100020	Male	27....	4121.000	2.000	2004-09-18	2007-02-07	0.000
3	K100030	Male	39....	3870.000	2.000	2003-08-23	2006-08-13	2.000
4	K100040	Male	28....	8322.000	4.000	2005-08-18	2006-06-04	2.000
5	K100050	Male	47....	2614.000	2.000	2005-08-11	2006-08-13	0.000
6	K100060	Male	29....	1891.000	1.000	2004-07-06	2005-03-13	1.000
7	K100070	Male	38....	1741.000	1.000	2006-10-14	2007-01-01	1.000
8	K100080	Male	27....	9770.000	4.000	2005-01-20	2009-08-27	0.000
9	K100090	Male	38....	8967.000	4.000	2003-05-31	2004-05-19	1.000
10	K100100	Male	48....	3438.000	2.000	2003-07-28	2007-07-05	0.000
11	K100110	Male	34....	1071.000	1.000	2005-12-11	2007-09-25	2.000
12	K100120	Male	32....	2875.000	2.000	2004-11-06	2007-10-06	0.000
13	K100130	Male	45....	8772.000	4.000	2003-12-15	2006-08-16	2.000
14	K100140	Male	29....	9016.000	4.000	2006-03-24	2008-03-13	1.000
15	K100150	Male	30....	5034.000	3.000	2005-07-28	2007-06-11	2.000
16	K100160	Male	28....	1709.000	1.000	2006-07-18	2009-09-04	9.000
17	K100170	Male	26....	1363.000	1.000	2004-06-28	2005-08-04	1.000
18	K100180	Male	37....	6455.000	3.000	2005-06-26	2006-08-06	1.000
19	K100190	Male	38....	4199.000	2.000	2004-08-06	2005-04-01	2.000
20	K100200	Male	46....	7532.000	4.000	2003-10-09	2005-08-10	0.000

OK

Table (10 fields, 95,307 records)

File Edit Generate

Table Annotations

	customer_id	peak_calls	peak_mins	offpeak_calls	offpeak_mins	weekend_calls	weekend_mins	international_mins
1	K100010	2	6.086	1	1.343	4	2.448	0.243
2	K100010	2	6.060	1	1.337	4	2.437	0.242
3	K100010	2	5.494	1	1.212	3	2.210	0.220
4	K100020	7	5.538	5	2.970	0	0.000	1.373
5	K100020	9	6.875	6	3.688	0	0.000	1.705
6	K100020	8	6.172	5	3.310	0	0.000	1.531
7	K100030	6	9.471	0	0.000	3	4.853	2.166
8	K100030	8	13.963	0	0.000	4	7.154	3.193
9	K100030	6	10.239	0	0.000	3	5.247	2.341
10	K100040	8	13.417	0	0.000	4	6.875	3.068
11	K100040	8	13.516	0	0.000	4	6.925	3.091
12	K100040	6	10.766	0	0.000	3	5.517	2.462
13	K100050	6	7.534	2	3.492	0	0.000	0.830
14	K100050	5	7.062	2	3.274	0	0.000	0.778
15	K100050	6	8.046	3	3.730	0	0.000	0.886
16	K100060	8	6.364	1	1.525	1	1.408	0.256
17	K100060	10	8.751	1	2.097	1	1.936	0.352
18	K100060	7	6.234	1	1.494	1	1.379	0.250
19	K100070	9	7.524	1	0.535	0	0.000	0.891
20	K100070	12	10.188	1	0.725	0	0.000	1.206

OK

Data Audit of [10 fields] #2

File Edit Generate

Audit Quality Annotations

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
gender		Nominal	--	--	--	--	--	6	31781
age		Continuous	-1.000	82.000	30.313	12.873	0.822	--	31781
region		Nominal	1.000	4.000	--	--	--	4	31781
connect_date		Continuous	2003-01-01	2006-12-31	--	--	--	--	31781
end_date		Continuous	2004-01-01	2010-12-29	--	--	--	--	14686
dropped_ca...		Continuous	0.000	45.000	3.190	4.195	2.152	--	31781
pay method		Flag	--	--	--	--	--	1	31781
tariff		Nominal	--	--	--	--	--	5	31781

[1] Indicates a multimode result [2] Indicates a sampled result

OK

Data Audit of [10 fields] #1

File Edit Generate

Audit Quality Annotations

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
gender		Nominal	--	--	--	--	--	6	31781
age		Continuous	12.000	82.000	30.327	12.858	0.829	--	31766
region		Nominal	1.000	4.000	--	--	--	4	31781
connect_date		Continuous	2003-01-01	2006-12-31	--	--	--	--	31781
end_date		Continuous	2004-01-01	2010-12-29	--	--	--	--	14686
dropped_ca...		Continuous	0.000	45.000	3.190	4.195	2.152	--	31781
pay method		Flag	--	--	--	--	--	1	31781
tariff		Nominal	--	--	--	--	--	5	31781

[1] Indicates a multimode result [2] Indicates a sampled result

OK

Data Audit of [10 fields]

File Edit Generate

Audit Quality Annotations

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
gender		Nominal	--	--	--	--	--	6	14686
age		Continuous	12.000	82.000	28.609	12.655	0.974	--	14674
region		Nominal	1.000	4.000	--	--	--	4	14686
connect_date		Continuous	2003-01-01	2006-12-31	--	--	--	--	14686
end_date		Continuous	2004-01-01	2010-12-29	--	--	--	--	14686
dropped_ca...		Continuous	0.000	41.000	3.909	4.700	1.221	--	14686
pay method		Flag	--	--	--	--	--	1	14686
tariff		Nominal	--	--	--	--	--	5	14686

[1] Indicates a multimode result [2] Indicates a sampled result

OK

Experiment - 3

Objective : You work as a data miner for a telecommunications firm. You have imported data, and you need to assess the quality of the data and define blanks where appropriate.

PROCEDURE:

1. Drag and drop an excel node from sources palette and import telco x customer info dataset in it.
2. Connect the excel node to a table.
3. Add a distinct node from record ops and connect it to excel node.
4. In distinct node select all the fields for grouping so that when all values in all fields are same then it will consider it only once.
5. Connect it to a table to display the results.
6. Add a var file node from sources and import telco x products dataset init.
7. Add an aggregate node and select customer_id field. In basic aggregates we select sum and connect it to a table. In output we will get sum of revenues and count/sum of gadgets a customer is using at once.
8. Add a type node and connect it to the setToFlag node. In that select gadgets field which will thus produce different-different fields for each gadget having values as T and F.

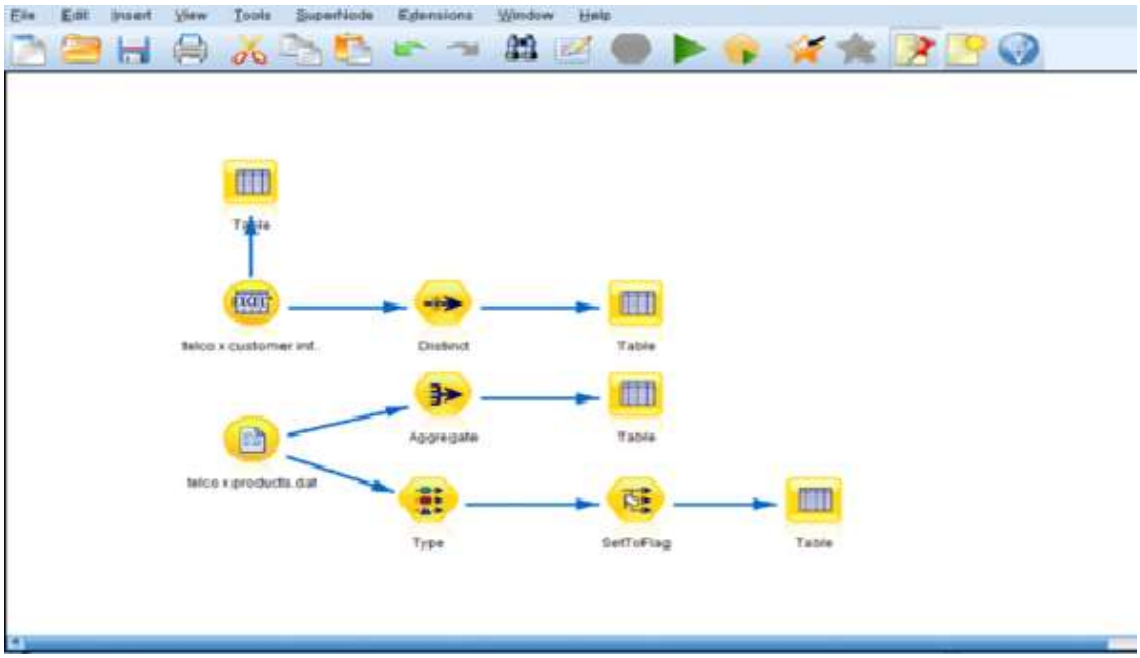


Table (10 fields, 31,769 records) #1

File Edit Generate

Table Annotations

	customer_id	gender	age	connect_date	end_date	dropped_calls	pay method	tariff
1	K100010	Male	46....	2006-07-18	2010-05-10	1.000	Pre Pay	CA...
2	K100020	Male	27....	2004-09-18	2007-02-07	0.000	Pre Pay	CA...
3	K100030	Male	39....	2003-08-23	2006-08-13	2.000	Pre Pay	CA...
4	K100040	Male	28....	2005-08-18	2006-06-04	2.000	Pre Pay	CA...
5	K100050	Male	47....	2005-08-11	2006-08-13	0.000	Pre Pay	CA...
6	K100060	Male	29....	2004-07-06	2005-03-13	1.000	Pre Pay	CA...
7	K100070	Male	38....	2006-10-14	2007-01-01	1.000	Pre Pay	CA...
8	K100080	Male	27....	2005-01-20	2009-08-27	0.000	Pre Pay	CA...
9	K100090	Male	38....	2003-05-31	2004-05-19	1.000	Pre Pay	CA...
10	K100100	Male	48....	2003-07-28	2007-07-05	0.000	Pre Pay	CA...
11	K100110	Male	34....	2005-12-11	2007-09-25	2.000	Pre Pay	CA...
12	K100120	Male	32....	2004-11-06	2007-10-06	0.000	Pre Pay	CA...
13	K100130	Male	45....	2003-12-15	2006-08-16	2.000	Pre Pay	CA...
14	K100140	Male	29....	2006-03-24	2008-03-13	1.000	Pre Pay	CA...
15	K100150	Male	30....	2005-07-28	2007-06-11	2.000	Pre Pay	CA...
16	K100160	Male	28....	2006-07-18	2009-09-04	9.000	Pre Pay	CA...
17	K100170	Male	26....	2004-06-28	2005-08-04	1.000	Pre Pay	CA...
18	K100180	Male	37....	2005-06-26	2006-08-06	1.000	Pre Pay	CA...
19	K100190	Male	38....	2004-08-06	2005-04-01	2.000	Pre Pay	CA...
20	K100200	Male	46....	2003-10-09	2005-08-10	0.000	Pre Pay	CA...

OK

Table (3 fields, 30,590 records) #1

File Edit Generate

Table Annotations

	customer_id	revenues_Sum	number_of_gadgets
1	K100010	250	4
2	K100020	141	3
3	K100030	455	7
4	K100040	72	3
5	K100070	94	2
6	K100080	86	1
7	K100090	171	2
8	K100100	71	3
9	K100110	39	1
10	K100120	117	1
11	K100130	197	4
12	K100140	198	2
13	K100150	160	4
14	K100160	109	1
15	K100180	245	3
16	K100200	125	2
17	K100220	50	1
18	K100230	108	1
19	K100240	92	1
20	K100260	284	3

OK

Table (13 fields, 30,590 records) #1

File Edit Generate

Table Annotations

	customer_id	gadget_A	gadget_B	gadget_C	gadget_D	gadget_E	gadget_F	gadget_G
1	K100010	F	F	T	F	T	T	F
2	K100020	T	F	F	F	F	T	T
3	K100030	T	T	F	T	F	F	F
4	K100040	T	T	F	T	F	F	F
5	K100070	F	F	T	F	F	T	F
6	K100080	F	F	F	F	F	F	F
7	K100090	F	F	F	F	F	F	F
8	K100100	T	T	F	T	F	F	F
9	K100110	F	F	F	T	F	F	F
10	K100120	F	F	F	F	F	F	F
11	K100130	F	T	T	F	T	F	F
12	K100140	F	F	F	F	F	F	F
13	K100150	F	T	T	T	F	F	T
14	K100160	F	F	F	F	F	F	F
15	K100180	F	F	F	F	F	T	F
16	K100200	F	F	F	F	F	T	T
17	K100220	F	F	F	F	T	F	F
18	K100230	F	F	F	F	F	F	F
19	K100240	F	F	F	F	F	F	F
20	K100260	F	F	F	F	F	F	F

OK

Experiment - 4

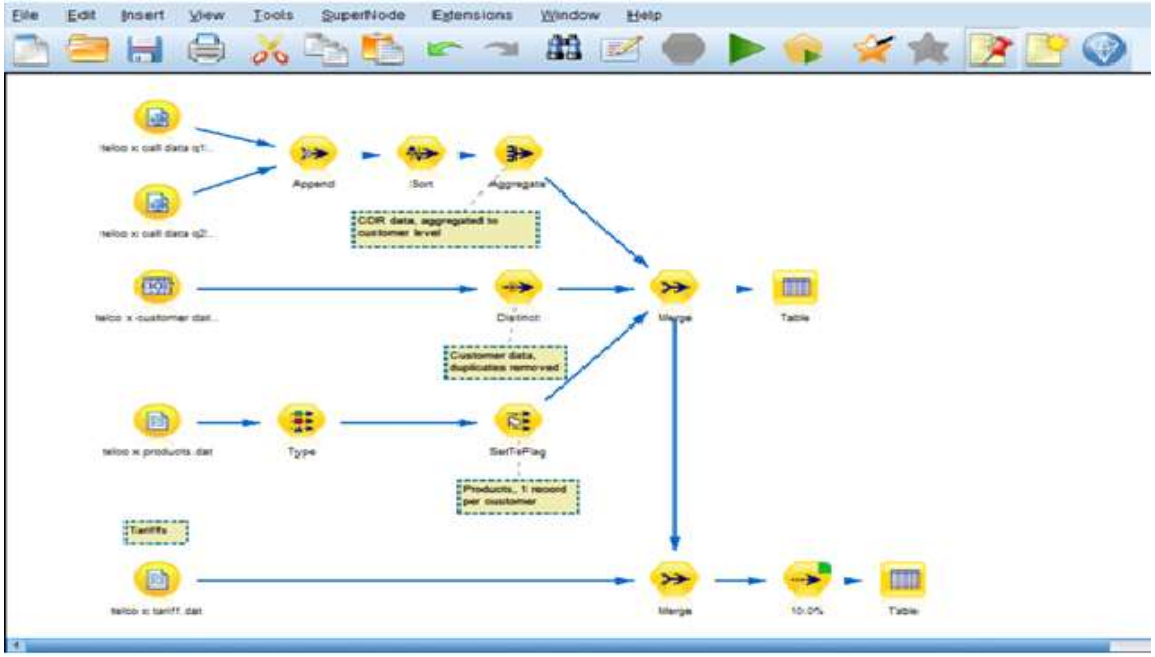
Objective : You work for a telecommunications firm and have to combine a number of datasets into a single dataset for analyses and modelling.

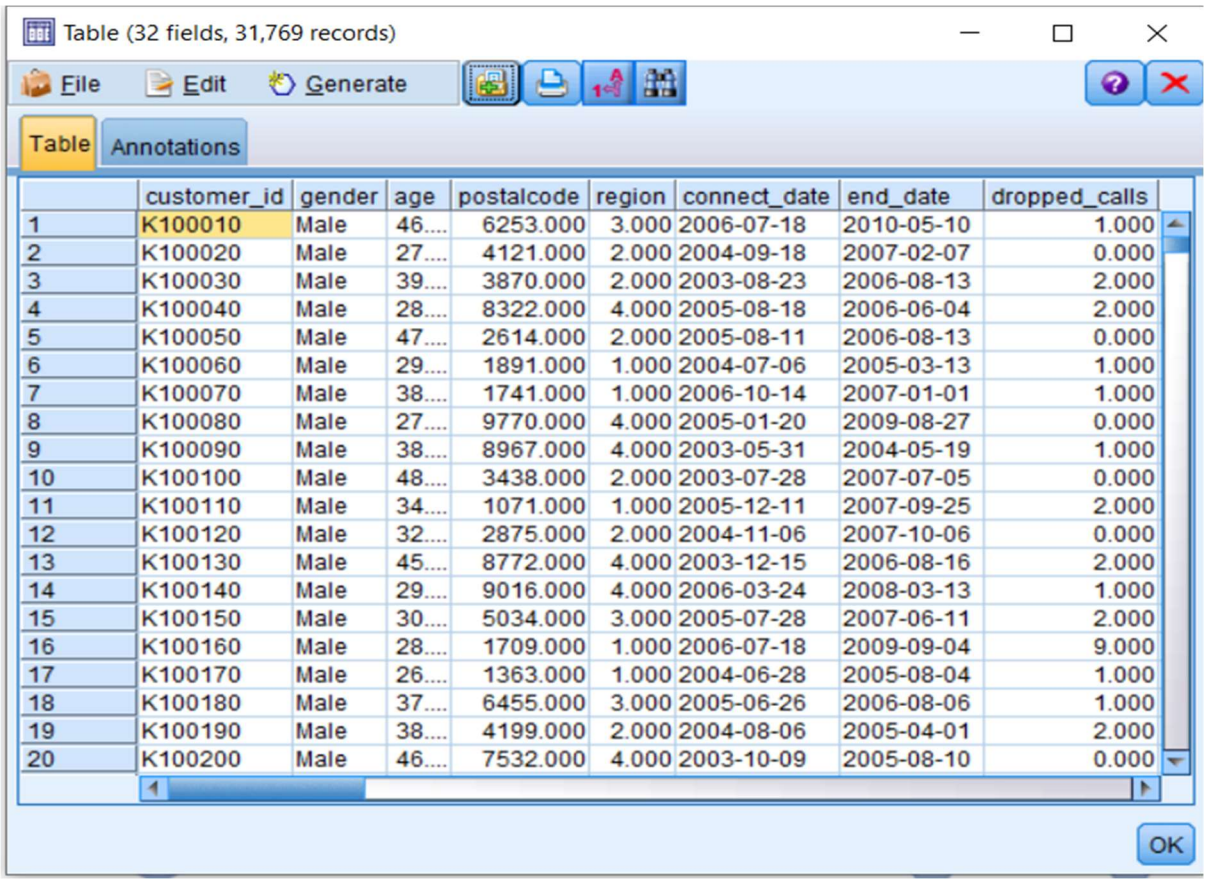
PROCEDURE:

1. From the Record Ops palette, place an Append node on the stream canvas to the right of the two Statistics File sources nodes named telco x call data q1.sav and telco x call data q2.sav.
2. Connect the Statistics Files node named telco x call data q1.sav and telco x call q2.sav to the Append node.
3. Edit the Append node, and beside Match fields by, select Position, enable the Tag records by including source dataset in field option and rename the Input field to quarter.
4. From the Record Ops palette, add a Sort node downstream from the Append node.
5. Edit the Sort node, and under Sort by, Ctrl-click to select customer_id and month.
6. Edit the Aggregate node, and under Key fields, select customer_id, under Aggregate fields, click peak_calls, and Shift-click international_calls, and then click. Then under Aggregate fields, click peak_calls and drag to international_calls, and then click Mean for any field, to deselect Mean for all selected fields. Disable the Include record count in field option.
7. From the Record Ops palette, drag a Merge node to the right of the Distinct node
8. Connect the Distinct, Aggregate, Set to flag node to the Merge node.
9. For Merge Method, select Keys. Under Possible keys, click customer_id, and then click the Add keys button to move it into the Keys for merge box.
10. Enable the Include matching and selected non-matching records (partial outer join) option, ensure that telco x customer data.xlsx is the leading dataset.
11. From the Record Ops palette, place a second Merge node on the stream canvas, to the right of the data source telco x tariff.dat.
12. Connect the first Merge node to the second Merge node, Also Connect the telco x tariff.dat Var File node to the second Merge node.
13. Edit the second Merge node. For Merge Method, select Keys. Under Possible keys, click tariff, and then move it into the Keys for merge box. Enable the option Include matching and selected non-matching records (partial outer join).
14. From the Record Ops palette, add a Sample node downstream from the second Merge node. Right-click the Sample node, click Cache, and then click Enable from the context menu.

15. Under Sample, select Random %, and change the associated value to check the Repeatable partition assignment box. Set the Seed value to 1.

16. Produce the desired output.





	customer_id	gender	age	postalcode	region	connect_date	end_date	dropped_calls
1	K100010	Male	46....	6253.000	3.000	2006-07-18	2010-05-10	1.000
2	K100020	Male	27....	4121.000	2.000	2004-09-18	2007-02-07	0.000
3	K100030	Male	39....	3870.000	2.000	2003-08-23	2006-08-13	2.000
4	K100040	Male	28....	8322.000	4.000	2005-08-18	2006-06-04	2.000
5	K100050	Male	47....	2614.000	2.000	2005-08-11	2006-08-13	0.000
6	K100060	Male	29....	1891.000	1.000	2004-07-06	2005-03-13	1.000
7	K100070	Male	38....	1741.000	1.000	2006-10-14	2007-01-01	1.000
8	K100080	Male	27....	9770.000	4.000	2005-01-20	2009-08-27	0.000
9	K100090	Male	38....	8967.000	4.000	2003-05-31	2004-05-19	1.000
10	K100100	Male	48....	3438.000	2.000	2003-07-28	2007-07-05	0.000
11	K100110	Male	34....	1071.000	1.000	2005-12-11	2007-09-25	2.000
12	K100120	Male	32....	2875.000	2.000	2004-11-06	2007-10-06	0.000
13	K100130	Male	45....	8772.000	4.000	2003-12-15	2006-08-16	2.000
14	K100140	Male	29....	9016.000	4.000	2006-03-24	2008-03-13	1.000
15	K100150	Male	30....	5034.000	3.000	2005-07-28	2007-06-11	2.000
16	K100160	Male	28....	1709.000	1.000	2006-07-18	2009-09-04	9.000
17	K100170	Male	26....	1363.000	1.000	2004-06-28	2005-08-04	1.000
18	K100180	Male	37....	6455.000	3.000	2005-06-26	2006-08-06	1.000
19	K100190	Male	38....	4199.000	2.000	2004-08-06	2005-04-01	2.000
20	K100200	Male	46....	7532.000	4.000	2003-10-09	2005-08-10	0.000

Table (40 fields, 3,252 records)

File Edit Generate

Table Annotations

	tariff	customer_id	gender	age	postalcode	region	connect_date	end_date
1	CAT 100	K342620	Female	24....	9143.000	4.000	2005-06-26	$null$
2	CAT 100	K342450	Female	56....	7261.000	3.000	2005-12-07	$null$
3	CAT 100	K342350	Male	31....	6331.000	3.000	2005-12-02	$null$
4	CAT 100	K342240	Male	12....	3956.000	2.000	2003-04-21	$null$
5	CAT 100	K341870	Male	14....	8702.000	4.000	2006-07-31	2008-10-29
6	CAT 100	K341730	Female	22....	5834.000	3.000	2003-10-31	2006-02-20
7	CAT 100	K341710	Female	13....	3485.000	2.000	2005-02-12	$null$
8	CAT 100	K341690	Female	30....	1663.000	1.000	2006-06-26	$null$
9	CAT 100	K341680	Male	53....	5693.000	3.000	2003-07-23	$null$
10	CAT 100	K341530	Male	41....	5012.000	3.000	2005-05-31	2008-02-18
11	CAT 100	K341490	Male	30....	7463.000	3.000	2004-09-23	$null$
12	CAT 100	K341450	Female	35....	7723.000	4.000	2006-02-20	$null$
13	CAT 100	K341340	Male	19....	4538.000	2.000	2003-12-21	$null$
14	CAT 100	K341140	Female	36....	8059.000	4.000	2004-01-14	2008-12-04
15	CAT 100	K341080	Female	22....	4066.000	2.000	2006-05-06	2010-12-28
16	CAT 100	K340960	Female	24....	6174.000	3.000	2003-12-02	$null$
17	CAT 100	K340690	Female	56....	5905.000	3.000	2005-03-12	$null$
18	CAT 100	K340590	Male	51....	3278.000	2.000	2005-06-07	$null$
19	CAT 100	K340570	Female	40....	1260.000	1.000	2005-02-09	$null$
20	CAT 100	K340540	Female	30....	3465.000	2.000	2005-02-08	$null$

OK

Experiment - 5

Objective : You work for a telecommunications firm and you need to cleanse and enrich a dataset in order to build models.

PROCEDURE:

1. Import the telco x data.txt.
2. Run the Distribution node for gender, downstream from the Type node.
3. From the Field Ops palette, add a Derive node downstream from the Type node.
4. In the Derive field text box, type gender_OK.
5. Click the Launch expression builder button to launch the Expression Builder. On the left side, click General Functions, and select the String group of functions from the list.
6. From the Graphs palette, add a Distribution node, downstream from the gender_OK Derive node. Edit the Distribution node, and then beside Field select gender_OK from the list.
7. From the Field Ops palette, add a Derive node downstream from the gender_OK Derive node
8. Edit the Derive node as follows:

   • Derive field: bill_peak.

   • Derive as: Formula

   • Expression Builder: peak_mins * peak_rate/100.

9. Add a Derive node downstream from the bill_peak node, and edit it as follows:

   • Derive field: bill_offpeak

   • Derive as: Formula (the default)

   • Expression: offpeak_mins * offpeak_rate / 100.

10. Add a Derive node downstream from the bill_offpeak node, and edit it as follows:

   • Derive field: bill_total.

   • Derive as: Formula (the default)

   • Expression bill_peak + bill_offpeak.

11. Add a Derive node, downstream from the bill_total node, and edit it as follows:

   • Derive field: bill_total_greater_than_0

   • Derive as: Flag

   • Field type: Flag (default value when the Derive type is Flag).

   • True when: bill_total > 0

12. Add a Derive node, downstream from the bill_total_greater_than_0 node, and edit it

as follows:

- Derive field: segment
- Derive as: Nominal
- Field type: Ordinal

13. Under Set field to, enter: 1, and under If this condition is true, enter: bill_total <= 100.

14. Add a Derive node, downstream from segment node, and define it as follows:

- Derive field: discount.
- Derive as: Conditional.
- Field type: Continuous
- If: segment = 3.
- Then: bill_total * 0.1
- Under Else, type 0.

15. In the Table output window, find K237920, and then scroll to the right.

16. From the Field Ops palette, add a Reclassify node, downstream from the discount node.

17. Beside Reclassify into, select Existing field. Under Reclassify field, select gender from the list.

18. From the Field Ops palette, add a Reclassify node, downstream from the first Reclassify node, and edit it as follows:

- Reclassify field: handset
- New field name: brand.

19. For the first new brand name, type ASAD in the New value column.

20. For the second new brand name, type ASAD, or pick ASAD from the drop-down list.

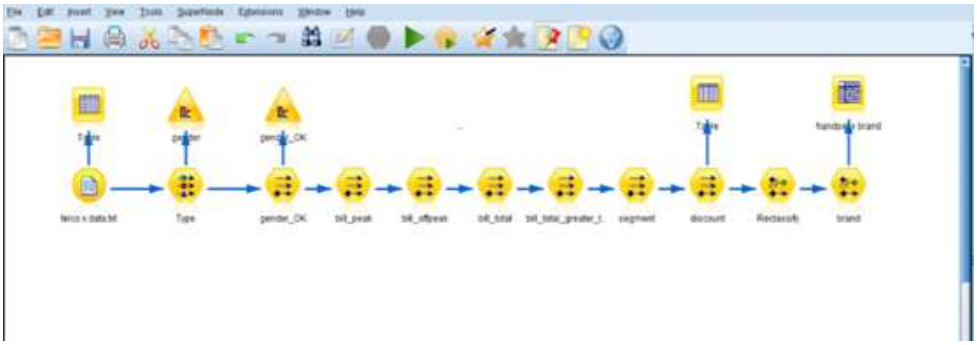
21. Repeat for the other brands, using only the letters from the original value as the new value.

22. From the Output palette, add a Matrix node, downstream from the brand Reclassify node, and define it as follows:

Rows: handset

Columns: brand

23. Get the desired result.



Matrix of handset by brand

File Edit Generate

Matrix Appearance Annotations

brand

handset	ASAD	BS	CAS	S	SOP
ASAD170	2680	0	0	0	0
ASAD90	4355	0	0	0	0
BS110	0	5340	0	0	0
BS210	0	1391	0	0	0
CAS01	0	0	8	0	0
CAS30	0	0	2843	0	0
CAS60	0	0	506	0	0
S50	0	0	0	7076	0
S80	0	0	0	3442	0
SOP10	0	0	0	0	632

Cells contain: cross-tabulation of fields (including missing values)

Chi-square = 158,845, df = 55, probability = 0

OK

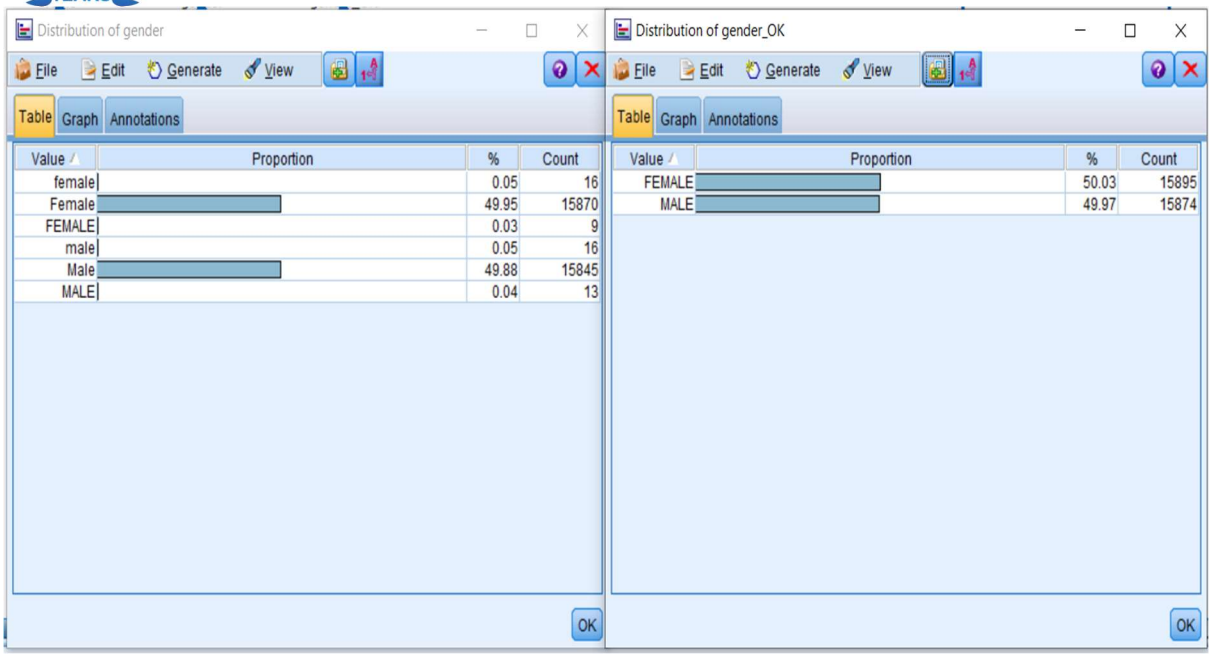
Table (61 fields, 31,769 records)

File Edit Generate

Table Annotations

	customer_id	gender	age	postalcode	region	tariff	connect_date	end_date
1	K338270	Female	20....	6599.000	3.000	CAT 100	2006-09-08	$null$
2	K342660	Male	21....	1635.000	1.000	CAT 100	2006-04-09	$null$
3	K342650	Male	17....	2149.000	1.000	CAT 100	2003-10-11	$null$
4	K342640	Male	20....	7788.000	4.000	CAT 100	2004-11-24	$null$
5	K342630	Male	46....	5223.000	3.000	CAT 100	2004-01-16	$null$
6	K342620	Female	24....	9143.000	4.000	CAT 100	2005-06-26	$null$
7	K342610	Male	32....	4223.000	2.000	CAT 100	2003-07-26	$null$
8	K342600	Female	51....	6165.000	3.000	CAT 100	2005-09-26	2007-06-15
9	K342590	Male	23....	7210.000	3.000	CAT 100	2003-10-24	$null$
10	K342580	Female	15....	1630.000	1.000	CAT 100	2006-03-02	$null$
11	K342570	Male	15....	9467.000	4.000	CAT 100	2004-11-09	$null$
12	K342560	Male	20....	9287.000	4.000	CAT 100	2005-10-24	2008-11-26
13	K342550	Female	43....	3498.000	2.000	CAT 100	2006-01-02	$null$
14	K342540	Female	19....	7618.000	4.000	CAT 100	2006-04-24	2008-06-07
15	K342530	Female	53....	9556.000	4.000	CAT 100	2003-09-09	2004-09-09
16	K342520	Male	44....	9179.000	4.000	CAT 100	2004-10-10	$null$
17	K342510	Male	21....	2428.000	1.000	CAT 100	2006-04-06	2009-03-06
18	K342500	Female	24....	5867.000	3.000	CAT 100	2006-05-06	2009-07-17
19	K342490	Male	22....	5627.000	3.000	CAT 100	2003-09-14	2004-03-15
20	K342480	Male	16....	5207.000	3.000	CAT 100	2004-10-21	2008-01-23

OK



Value	Proportion	%	Count
female		0.05	16
Female		49.95	15870
FEMALE		0.03	9
male		0.05	16
Male		49.88	15845
MALE		0.04	13

OK

Value	Proportion	%	Count
FEMALE		50.03	15895
MALE		49.97	15874

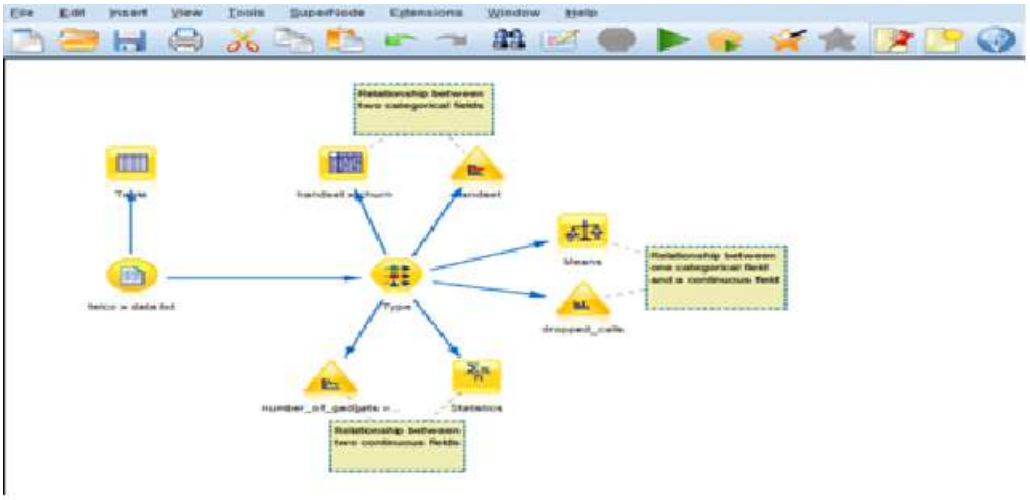
OK

Experiment - 6

Objective : Identify relationships in the telecommunications data.

PROCEDURE:

1. On the IBM SPSS modeler, From the Output palette, add a Matrix node downstream from the Type node.
2. Edit the Matrix node. Beside Rows, select handset from the list. Beside Columns, select churn from the list.
3. Click the Appearance tab. Select the Percentage of row option. Select the Include row and column totals option.
4. From the Graphs palette, add a Distribution node downstream from the Type node. Edit the Distribution node. Beside Field, select handset, Beside Colour, select churn, Enable the Normalize by colour check box.
5. From the Output palette, add a Means node downstream from the Type node.
6. Edit the Means node. For Grouping field, select churn. For Test field(s), select dropped_calls.
7. From the Graphs palette, add a Histogram node downstream from the Type node.
8. Edit the Histogram node. Beside Field, select dropped_calls. Beside Color, select churn.
9. From the Output palette, add a Statistics node downstream from the Type node.
10. Edit the Statistics node. Beside Examine, select revenues from the list. Beside Correlate, select number_of_gadgets from the list.
11. From the Graphs palette, add a Plot node downstream from the Type node.
12. Edit the Plot node. For X field, select number_of_gadgets. For Y field, select revenues.
13. Get the desired Output.



Means of [churn][dropped_calls]

File Edit

Means Annotations

Sort by: Field View: Simple

Grouping field: churn

*Cells contain: Mean

Field	Active*	Churned*	Importance
dropped_calls	2.573	3.906	1.000 Important

OK

Matrix of handset by churn

File Edit Generate

Matrix Appearance Annotations

churn

handset		Active	Churned	Total
ASAD170	Count	2556	124	2680
	Row %	95.373	4.627	100
ASAD90	Count	224	4131	4355
	Row %	5.144	94.856	100
BS110	Count	2991	2349	5340
	Row %	56.011	43.989	100
BS210	Count	989	402	1391
	Row %	71.100	28.900	100
CAS01	Count	3	5	8
	Row %	37.500	62.500	100
CAS30	Count	153	2690	2843
	Row %	5.382	94.618	100
CAS60	Count	476	30	506
	Row %	94.071	5.929	100

Cells contain: cross-tabulation of fields (including missing values)

Chi-square = 13,941.022, df = 11, probability = 0

OK

Table (54 fields, 31,769 records)

File Edit Generate

Table Annotations

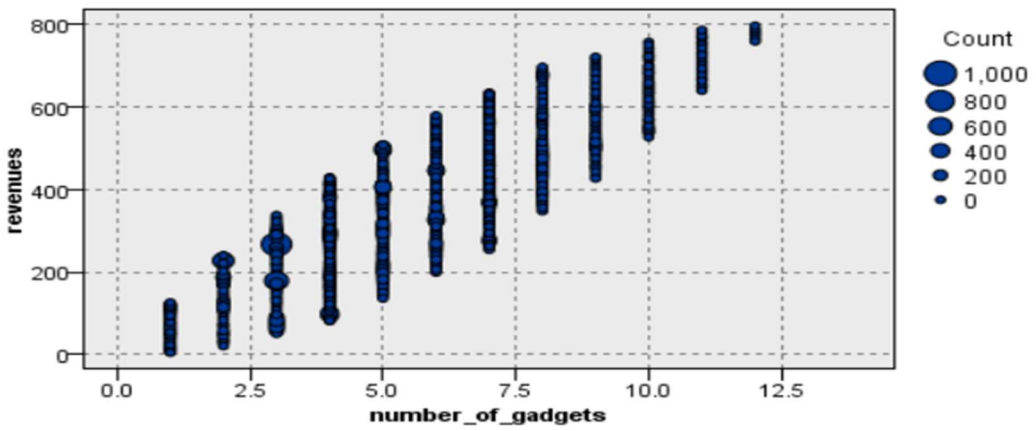
	customer_id	gender	age	postalcode	region	tariff	connect_date	ei
1	K338270	Female	20....	6599.000	3.000	CAT 100	2006-09-08	$r
2	K342660	Male	21....	1635.000	1.000	CAT 100	2006-04-09	$r
3	K342650	Male	17....	2149.000	1.000	CAT 100	2003-10-11	$r
4	K342640	Male	20....	7788.000	4.000	CAT 100	2004-11-24	$r
5	K342630	Male	46....	5223.000	3.000	CAT 100	2004-01-16	$r
6	K342620	Female	24....	9143.000	4.000	CAT 100	2005-06-26	$r
7	K342610	Male	32....	4223.000	2.000	CAT 100	2003-07-26	$r
8	K342600	Female	51....	6165.000	3.000	CAT 100	2005-09-26	20
9	K342590	Male	23....	7210.000	3.000	CAT 100	2003-10-24	$r
10	K342580	Female	15....	1630.000	1.000	CAT 100	2006-03-02	$r
11	K342570	Male	15....	9467.000	4.000	CAT 100	2004-11-09	$r
12	K342560	Male	20....	9287.000	4.000	CAT 100	2005-10-24	20
13	K342550	Female	43....	3498.000	2.000	CAT 100	2006-01-02	$r
14	K342540	Female	19....	7618.000	4.000	CAT 100	2006-04-24	20
15	K342530	Female	53....	9556.000	4.000	CAT 100	2003-09-09	20
16	K342520	Male	44....	9179.000	4.000	CAT 100	2004-10-10	$r
17	K342510	Male	21....	2428.000	1.000	CAT 100	2006-04-06	20
18	K342500	Female	24....	5867.000	3.000	CAT 100	2006-05-06	20
19	K342490	Male	22....	5627.000	3.000	CAT 100	2003-09-14	20
20	K342480	Male	16....	5207.000	3.000	CAT 100	2004-10-21	20

OK

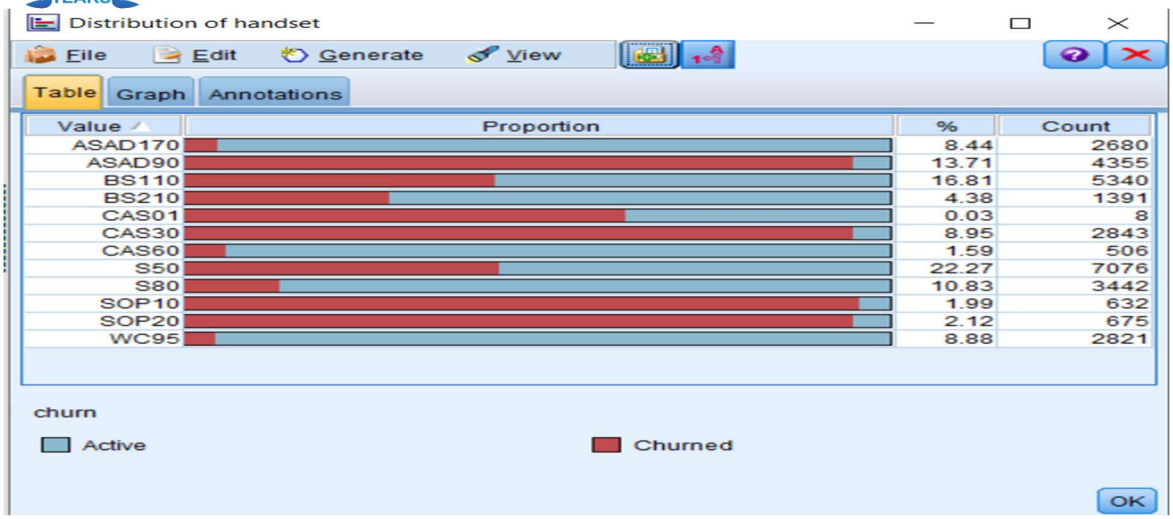
Plot of number_of_gadgets v. revenues

File Edit Generate View

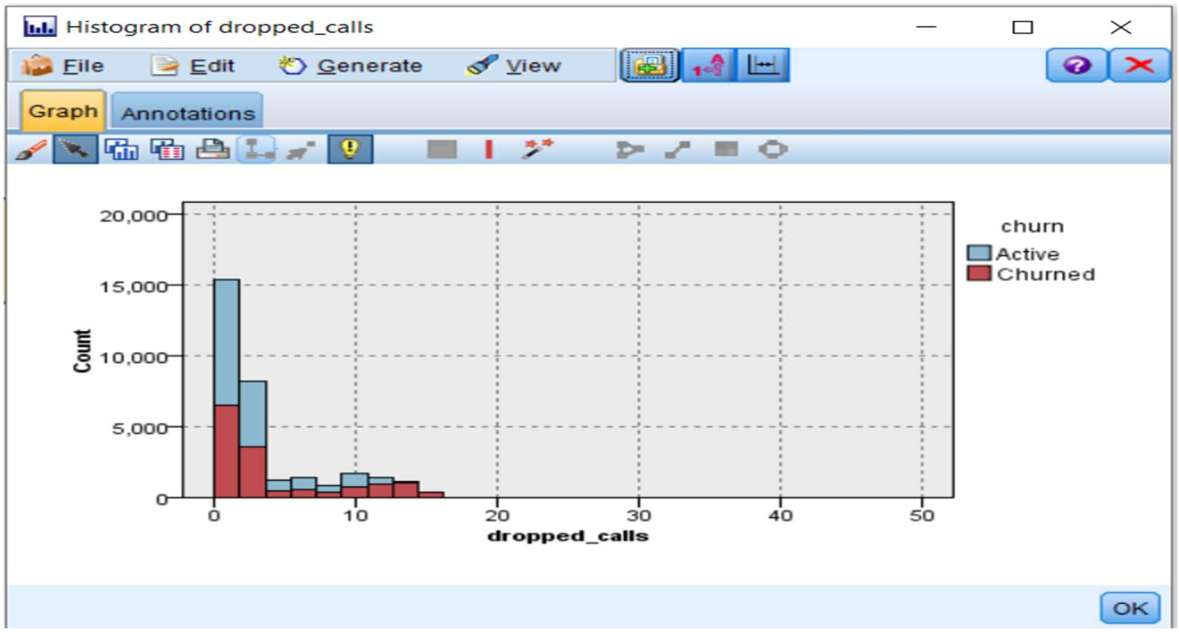
Graph Annotations



OK



Value	Proportion	%	Count
ASAD170		8.44	2680
ASAD90		13.71	4355
BS110		16.81	5340
BS210		4.38	1391
CAS01		0.03	8
CAS30		8.95	2843
CAS60		1.59	506
S50		22.27	7076
S80		10.83	3442
SOP10		1.99	632
SOP20		2.12	675
WC95		8.88	2821

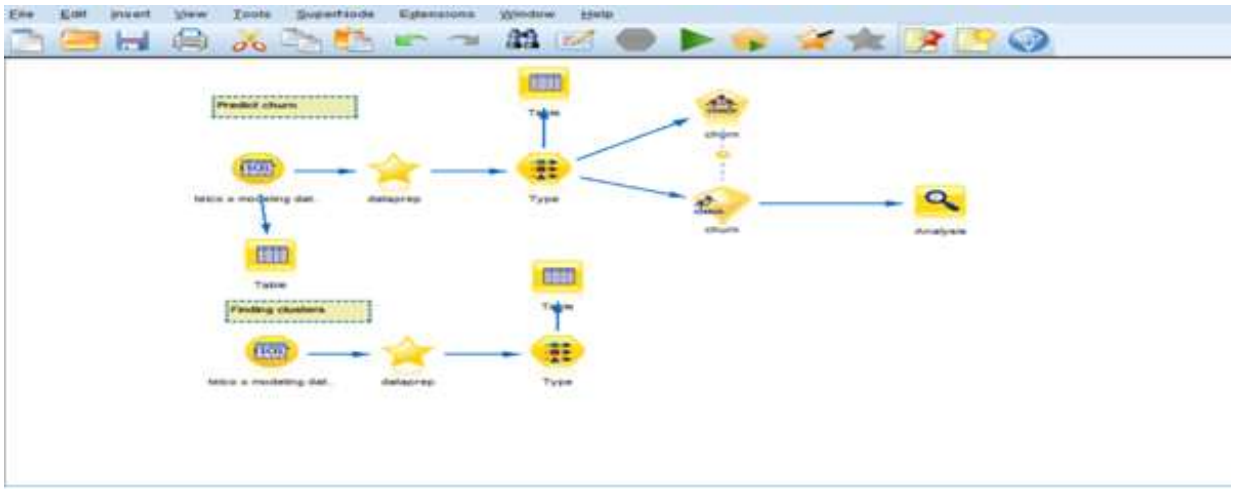


Experiment - 7

Objective : Predict churn in telecommunications and cluster customers into segments.

PROCEDURE:

1. Drag and drop an excel node from sources palette and import telco x modelling dataset in it.
2. Add a filler node from field ops and select gender field. We replace all the values of gender field with their respective uppercase values using expression lowertoupper(@FIELD).
3. Connect this filler node to a select node in which we use expression data_known = "yes" to select only those records of which data is known.
4. Above nodes were of data preparation so we combine them to form a super node.
5. Add a type node and select correct measurements as required for each field.
6. From modelling palette select CHAID node and select target -> churn and predictors -> gender, age, tariff, dropped_calls,handset.
7. A nugget will be formed. Connect it to an analysis node from output palette.
8. In this select/check coincidences matrices. This will help us to analyse the probability of correct or wrong predictions.



Analysis of [churn]

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field churn

Comparing $R-churn with churn

Correct	27,810	87.54%
Wrong	3,959	12.46%
Total	31,769	

Coincidence Matrix for $R-churn (rows show actuals)

	Active	Churned
Active	15,200	1,856
Churned	2,103	12,610

OK

Table (23 fields, 31,769 records) #2

File Edit Generate

Table Annotations

	customer_id	data_known	gender	age	tariff	dropped_calls	handset	peak_mins
1	K100010	yes	MALE	46....	CA...	1.000	SOP10	36.131
2	K100020	yes	MALE	27....	CA...	0.000	SOP10	39.437
3	K100030	yes	MALE	39....	CA...	2.000	SOP20	72.600
4	K100040	yes	MALE	28....	CA...	2.000	SOP10	72.600
5	K100050	yes	MALE	47....	CA...	0.000	SOP10	40.608
6	K100060	yes	MALE	29....	CA...	1.000	SOP10	46.260
7	K100070	yes	MALE	38....	CA...	1.000	SOP20	56.370
8	K100080	yes	MALE	27....	CA...	0.000	SOP20	51.043
9	K100090	yes	MALE	38....	CA...	1.000	SOP20	99.000
10	K100100	yes	MALE	48....	CA...	0.000	SOP20	65.400
11	K100110	yes	MALE	34....	CA...	2.000	SOP20	65.400
12	K100120	yes	MALE	32....	CA...	0.000	SOP20	68.123
13	K100130	yes	MALE	45....	CA...	2.000	SOP20	63.352
14	K100140	yes	MALE	29....	CA...	1.000	SOP20	41.447
15	K100150	yes	MALE	30....	CA...	2.000	SOP20	96.000
16	K100160	yes	MALE	28....	CA...	9.000	SOP20	41.342
17	K100170	yes	MALE	26....	CA...	1.000	SOP10	100.200
18	K100180	yes	MALE	37....	CA...	1.000	SOP10	41.918
19	K100190	yes	MALE	38....	CA...	2.000	SOP10	58.782
20	K100200	yes	MALE	46....	CA...	0.000	SOP10	84.600

OK

Experiment - 8

Objective : Use CLEM functions to cleanse and enrich telecommunications data.

PROCEDURE:

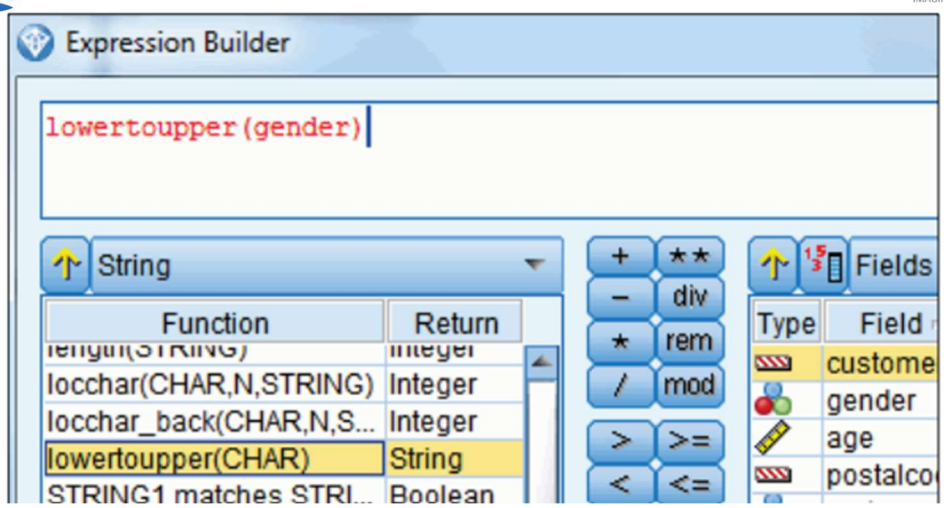
1) Run the Distribution node for gender, downstream from the Type node.
2) Close the Distribution output window.
3) From the Field Ops palette, add a Derive node downstream from the Type Node
4) Edit the Derive node as formula.
5) In the Derive field text box, type gender_OK
6) Click the Launch expression builder button to launch the Expression Builder to write the CLEM expression.
7) On the left side, click General Functions, and select the String group of functions from the list.
8) Under Function, select lowertoupper, and then click the Insert button to paste it into the expression area.
9) In the right pane, under Field, double-click gender, to add it to the expression area. This should create the CLEM expression as :

   Lowertoupper(gender)

10) Save the node and from the Graphs palette, add a Distribution node, downstream from the gender_OK Derive node. Edit the Distribution node, and then beside Field select gender_OK from the list.
11) Run stream.

Table | Graph | Annotations

Value	Proportion	%	Count
female		0.05	16
Female		49.95	15870
FEMALE		0.03	9
male		0.05	16
Male		49.88	15845
MALE		0.04	13



	rnational_rate	voicemail	SMS	gender_OK
1	30	10	15	FEMALE
2	30	10	15	MALE
3	30	10	15	MALE
4	30	10	15	MALE
5	30	10	15	MALE
6	30	10	15	FEMALE
7	30	10	15	MALE

Table | Graph | Annotations

Value	Proportion	%	Count
FEMALE		50.03	15895
MALE		49.97	15874

Experiment - 9

Objective: Use additional field transformations to prepare the telecommunications data for modelling.

PROCEDURE:

1. From the Sources palette, double-click the Var. File node to add it to the stream canvas. Select telco x churn data.dat.

2. From the Field Ops palette, add a Type node downstream from the Var. File node. Edit the Type node. Click Read Values to instantiate the data.

3. From the Field Ops palette, add a Filler node downstream from the Type node. Under Fill in fields, select AGE, POSTALCODE, REGION, and DROPPED_CALLS. Beside Replace, click Always from the list. Under Replace with, enter to_integer (@FIELD).

4. From the Field Ops palette, add a Filler node downstream from to the first Filler node. Under Fill in fields, select A_REVENUES to L_ REVENUES, and then click OK. Beside Replace, click Null values from the list. Under Replace with, ensure that the value is 0.

5. Add a Filler node downstream from the second Filler node. Under Fill in fields, select GENDER, and then click OK. Beside Replace, click Always from the list. Under Replace with, enter lowertoupper (@FIELD).

6. From the Field Ops palette, add a Binning node downstream from the third Filler node. Edit the Binning node. Beside Bin fields, select BILL_TOTAL, and then click OK. In the Binning method list, select Tiles (equal count). Clear the Decile (10) check box. Click the Custom N check box to select it, and then replace the current value with 3. Click the Bin Values tab, and then click the Read Values button.

7. From the Graphs palette, add a Distribution node downstream from the Binning node. Edit the Distribution node. Beside Field, click BILL_TOTAL_TILEN from the list. Click Run.

8. From the Field Ops palette, add a Binning node downstream from the first Binning node. Edit the second Binning node. Next to Bin fields, select BILL_TOTAL, and then click OK. Beside Binning method, click Optimal from the list. Beside Supervisor field, select CHURN. Click the Bin Values tab, and then click Read Values.

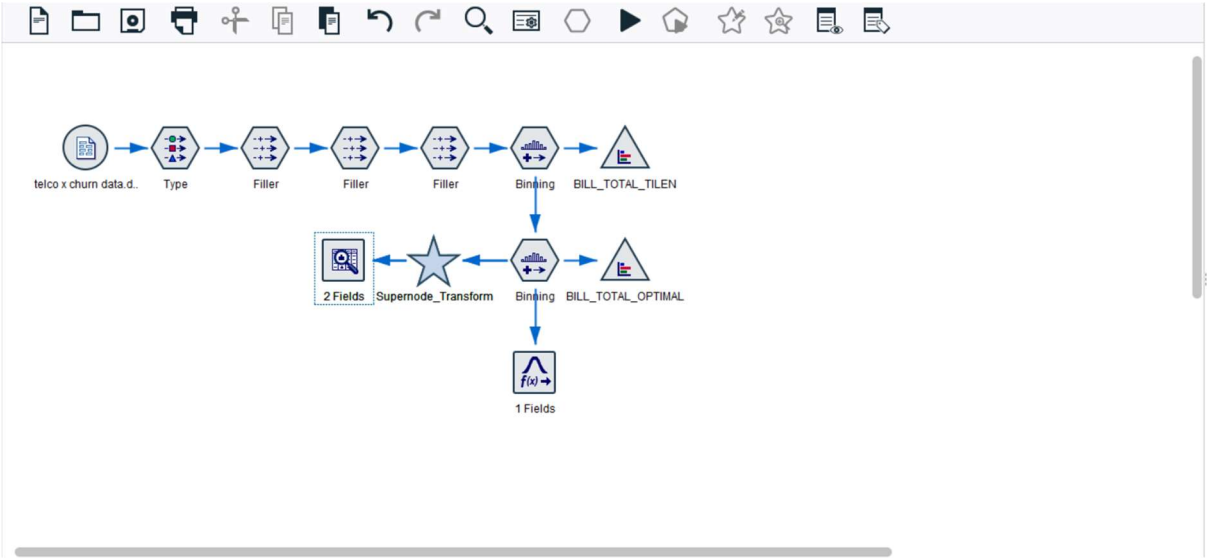
9. From the Graphs palette, add a Distribution node downstream from the second Binning node. Edit the Distribution node. Beside Field, select BILL_TOTAL_OPTIMAL. Beside Color, select CHURN. Click the Normalize by color check box to select it. Click Run.

10. From the Output palette, add a Transform node downstream from the second Binning node. Edit the Transform node. Fields are selected on the Fields tab. Beside Fields, click BILL_TOTAL from the list, and then click OK. Click the Options tab. Click Run.

11. Click the Square Root thumbnail so that it is selected. From the Generate menu, click Derive Node, and then click OK in the dialog box that displays (you will compute non-standardized scores). Click OK to close the Transform output window. Right click the second Binning node, click Connect, and then click the generated SuperNode, located in the upper left corner on the stream canvas, to add downstream from the second Binning node. You can drag the SuperNode downstream of the Binning node. Edit the SuperNode, click Zoom in, and then edit the node named BILL_TOTAL_Square Root.

12. From the Output palette, add a Data Audit node downstream from the SuperNode. Edit the Data Audit node. On the Settings tab, click the Use custom fields radio button. Beside Fields, select BILL_TOTAL and BILL_TOTAL_Square Root, and then click OK. Click Run

Output:



Preview from Type Node (45 fields, 10 records)

File Edit Generate

Table Annotations

	CUSTOMER_ID	GENDER	AGE	POSTALCODE	REGION	HANDSET	CONNECT_DATE	END_DATE
1	K338270	Female	20....	6599.000	3.000	WC95	2006-09-08	$null$
2	K342660	Male	21....	1635.000	1.000	S50	2006-04-09	$null$
3	K342650	Male	17....	2149.000	1.000	S50	2003-10-11	$null$
4	K342640	Male	20....	7788.000	4.000	S50	2004-11-24	$null$
5	K342630	Male	46....	5223.000	3.000	CAS30	2004-01-16	$null$
6	K342620	Female	24....	9143.000	4.000	BS110	2005-06-26	$null$
7	K342610	Male	32....	4223.000	2.000	S50	2003-07-26	$null$
8	K342600	Female	51....	6165.000	3.000	CAS30	2005-09-26	2007-06-15
9	K342590	Male	23....	7210.000	3.000	BS110	2003-10-24	$null$
10	K342580	Female	15....	1630.000	1.000	WC95	2006-03-02	$null$

OK

Preview from Filler Node (45 fields, 10 records)

File Edit Generate

Table Annotations

	CUSTOMER_ID	GENDER	AGE	POSTALCODE	REGION	HANDSET	CONNECT_DATE	END_DATE
1	K338270	Female	20	6599	3	WC95	2006-09-08	$null$
2	K342660	Male	21	1635	1	S50	2006-04-09	$null$
3	K342650	Male	17	2149	1	S50	2003-10-11	$null$
4	K342640	Male	20	7788	4	S50	2004-11-24	$null$
5	K342630	Male	46	5223	3	CAS30	2004-01-16	$null$
6	K342620	Female	24	9143	4	BS110	2005-06-26	$null$
7	K342610	Male	32	4223	2	S50	2003-07-26	$null$
8	K342600	Female	51	6165	3	CAS30	2005-09-26	2007-06-15
9	K342590	Male	23	7210	3	BS110	2003-10-24	$null$
10	K342580	Female	15	1630	1	WC95	2006-03-02	$null$

OK

Preview from Filler Node (45 fields, 10 records) #1

File Edit Generate

Table Annotations

	REVENUES	A_REVENUES	B_REVENUES	C_REVENUES	D_REVENUES	E_REVENUES	F_REVE
1	495	0	0	0	0	0	
2	276	0	0	0	0	0	
3	624	0	0	0	0	0	
4	498	0	0	0	0	0	
5	572	13	20	35	0	0	
6	367	0	23	29	42	0	
7	339	9	19	0	40	0	
8	270	0	0	0	0	0	
9	451	0	0	0	0	52	
10	351	0	0	0	0	49	

OK

Preview from Filler Node (45 fields, 10 records) #2

File Edit Generate

Table Annotations

	CUSTOMER_ID	GENDER	AGE	POSTALCODE	REGION	HANDSET	CONNECT_DATE	END_DATE
1	K338270	FEMALE	20	6599	3	WC95	2006-09-08	$null$
2	K342660	MALE	21	1635	1	S50	2006-04-09	$null$
3	K342650	MALE	17	2149	1	S50	2003-10-11	$null$
4	K342640	MALE	20	7788	4	S50	2004-11-24	$null$
5	K342630	MALE	46	5223	3	CAS30	2004-01-16	$null$
6	K342620	FEMALE	24	9143	4	BS110	2005-06-26	$null$
7	K342610	MALE	32	4223	2	S50	2003-07-26	$null$
8	K342600	FEMALE	51	6165	3	CAS30	2005-09-26	2007-06-15
9	K342590	MALE	23	7210	3	BS110	2003-10-24	$null$
10	K342580	FEMALE	15	1630	1	WC95	2006-03-02	$null$

OK

Distribution of BILL_TOTAL_TILEN #1

File Edit Generate View

Table Graph Annotations

Value	Proportion	%	Count
1		33.33	10590
2		33.33	10590
3		33.33	10589

OK

Distribution of BILL_TOTAL_OPTIMAL #1

File Edit Generate View

Table Graph Annotations

Value	Proportion	%	Count
1		7.1	2255
2		4.1	1303
3		7.1	2255
4		41.0	13026
5		40.7	12930

CHURN

Active Churned

OK

Data Audit of [BILL_TOTAL BILL_TOTAL_Square Root] #1

File Edit Generate

Audit Quality Annotations

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
BILL_TOTAL		Continuous	0.000	582.141	155.757	82.286	0.917	--	31769
BILL_TOTAL_Square R...		Continuous	0.000	24.128	12.022	3.351	-0.003	--	31769

[1] Indicates a multimode result [2] Indicates a sampled result

OK

Experiment - 10

Objective: Prepare sequence data of a bank for analysis.

PROCEDURE:

1. From the Sources palette, add a Var. File node to the stream canvas, edit then node, and import data from year_balances.txt.

2. From the Field Ops palette, add a Derive node downstream from the Var. File node. Edit the Derive node. Under Derive field, type RECORD_ID. Under Formula, enter @INDEX.

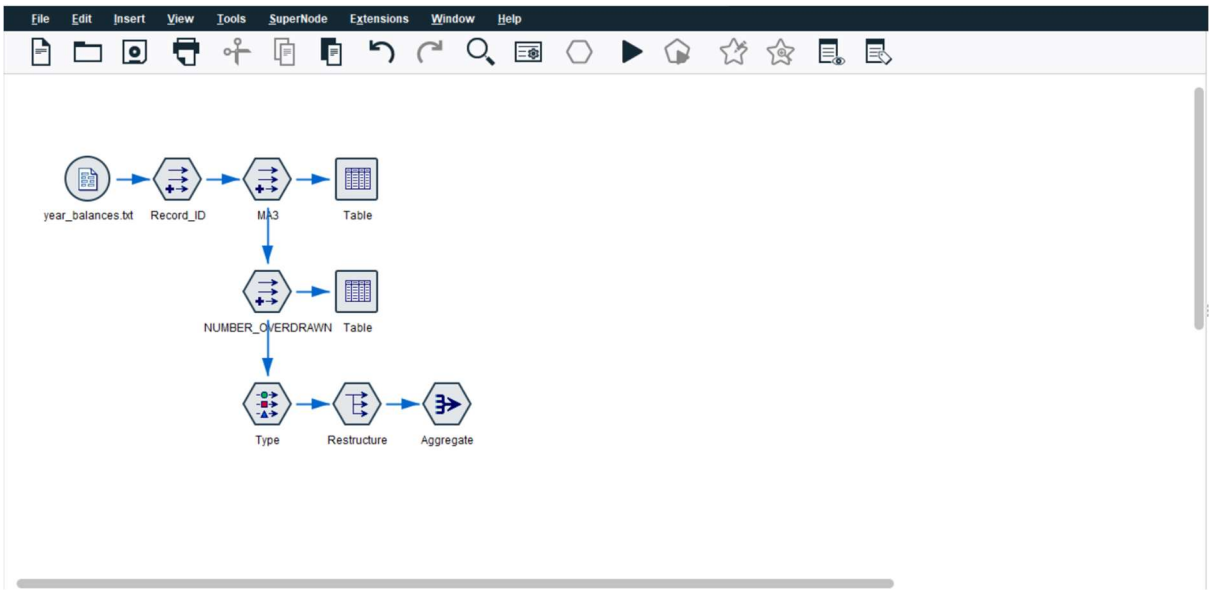
3. From the Field Ops palette, add a Derive node downstream from the Derive node named RECORD_ID. Edit the Derive node. Under Derive field, type MA3. Beside Derive as, select Conditional from the drop-down list. Under If, enter ACCTNO = @OFFSET(ACCTNO, 2). Under Then, enter @MEAN(BALANCE, 3). Under Else, type undef. From the Output palette, add a Table node downstream from the Derive node, and then run the Table node.

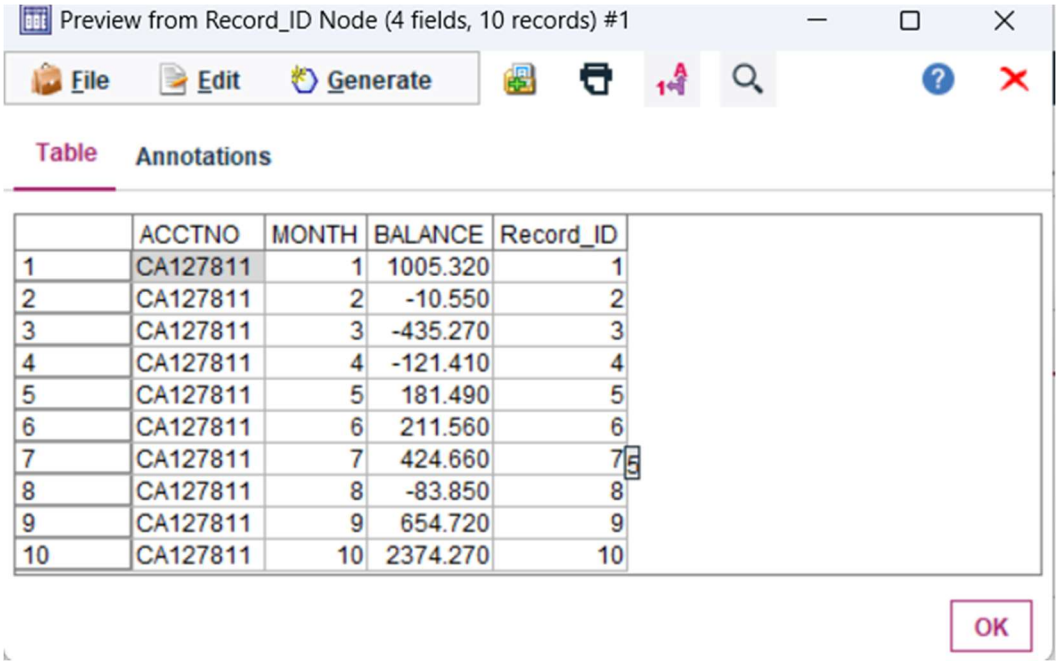
4. From the Field Ops palette, add a Derive node downstream from the Derive node named MA3. Edit the Derive node. Under Derive field, enter NUMBER_OVERDRAWN. Beside Derive as, select Count from the drop-down list. Beside Initial value, ensure that the value is 0. Under Increment when, enter BALANCE < 0. Under Increment by, ensure that the value is 1. Under Reset when, enter ACCTNO /= @OFFSET(ACCTNO,1). From the Output palette, add a Table node downstream from the Derive node, and then run the Table node.

5. From the Field Ops palette, add a Type node downstream from the Derive node named NUMBER_OVERDRAWN. Edit the Type node. Set the measurement level for MONTH to Categorical by clicking the Measurement column of the MONTH row and then clicking Categorical from the list. Click Read Values. From the Field Ops palette, add a Restructure node downstream from the Type node. Edit the Restructure node. Under Available fields, select MONTH. Select all values, and then click the Add selected fields arrow to add them into the Create restructured fields box. Clear the Include field names check box. Under Value field(s), select BALANCE from the list.

6. From the Record Ops palette, add an Aggregate node downstream from the Restructure node. Edit the Aggregate node. For Key fields, select ACCTNO. Under Basic Aggregates, Aggregate fields, select 1_BALANCE to 12_BALANCE. For Aggregate fields, ensure that the Sum statistic is set for 1_BALANCE to 12_BALANCE. Clear any other statistic check box and then click Apply default operations to all fields

Output:





	ACCTNO	MONTH	BALANCE	Record_ID
1	CA127811	1	1005.320	1
2	CA127811	2	-10.550	2
3	CA127811	3	-435.270	3
4	CA127811	4	-121.410	4
5	CA127811	5	181.490	5
6	CA127811	6	211.560	6
7	CA127811	7	424.660	7
8	CA127811	8	-83.850	8
9	CA127811	9	654.720	9
10	CA127811	10	2374.270	10

Preview from MA3 Node (5 fields, 10 records) #1

File Edit Generate

Table Annotations

	ACCTNO	MONTH	BALANCE	Record_ID	MA3
1	CA127811	1	1005.320	1	$null$
2	CA127811	2	-10.550	2	$null$
3	CA127811	3	-435.270	3	186.500
4	CA127811	4	-121.410	4	-189.077
5	CA127811	5	181.490	5	-125.063
6	CA127811	6	211.560	6	90.547
7	CA127811	7	424.660	7	272.570
8	CA127811	8	-83.850	8	184.123
9	CA127811	9	654.720	9	331.843
10	CA127811	10	2374.270	10	981.713

OK

Table (5 fields, 144 records) #1

File Edit Generate

Table Annotations

	ACCTNO	MONTH	BALANCE	Record_ID	MA3
1	CA127811	1	1005.320	1	$null$
2	CA127811	2	-10.550	2	$null$
3	CA127811	3	-435.270	3	186.500
4	CA127811	4	-121.410	4	-189.077
5	CA127811	5	181.490	5	-125.063
6	CA127811	6	211.560	6	90.547
7	CA127811	7	424.660	7	272.570
8	CA127811	8	-83.850	8	184.123
9	CA127811	9	654.720	9	331.843
10	CA127811	10	2374.270	10	981.713
11	CA127811	11	313.330	11	1114.107
12	CA127811	12	216.250	12	967.950
13	CA127812	1	144.510	13	$null$
14	CA127812	2	57.210	14	$null$
15	CA127812	3	514.130	15	238.617
16	CA127812	4	5095.030	16	1888.790
17	CA127812	5	-202.250	17	1802.303
18	CA127812	6	-305.590	18	1529.063
19	CA127812	7	241.860	19	-88.660
20	CA127812	8	1060.410	20	332.227

OK

Table (6 fields, 144 records) #1

File Edit Generate

Table Annotations

	ACCTNO	MONTH	BALANCE	Record_ID	MA3	NUMBER_OVERDRAWN
1	CA127811	1	1005.320	1	$null$	0
2	CA127811	2	-10.550	2	$null$	1
3	CA127811	3	-435.270	3	186.500	2
4	CA127811	4	-121.410	4	-189.077	3
5	CA127811	5	181.490	5	-125.063	3
6	CA127811	6	211.560	6	90.547	3
7	CA127811	7	424.660	7	272.570	3
8	CA127811	8	-83.850	8	184.123	4
9	CA127811	9	654.720	9	331.843	4
10	CA127811	10	2374.270	10	981.713	4
11	CA127811	11	313.330	11	1114.107	4
12	CA127811	12	216.250	12	967.950	4
13	CA127812	1	144.510	13	$null$	0
14	CA127812	2	57.210	14	$null$	0
15	CA127812	3	514.130	15	238.617	0
16	CA127812	4	5095.030	16	1888.790	0
17	CA127812	5	-202.250	17	1802.303	1
18	CA127812	6	-305.590	18	1529.063	2
19	CA127812	7	241.860	19	-88.660	2
20	CA127812	8	1060.410	20	332.227	2

OK

Preview from Aggregate Node (14 fields, 10 records) #1

File Edit Generate

Table Annotations

	ACCTNO	1_BALANCE_Sum	2_BALANCE_Sum	3_BALANCE_Sum	4_BALANCE_Sum	5_BALANCE_Sum	6_BALANCE_Sum	7_BALANCE_Sum
1	CA127811	1005.320	-10.550	-435.270	-121.410	181.490	211.560	424.660
2	CA127812	144.510	57.210	514.130	5095.030	-202.250	-305.590	241.860
3	CA127813	321.200	121.610	-684.590	-2320.500	-156.520	426.030	2092.180
4	CA127814	-204.510	1052.070	294.390	2218.450	-1.430	-141.580	-312.990
5	CA127815	2022.020	1082.270	73.700	4574.650	-81.980	-213.340	-2858.470
6	CA127816	287.800	1123.280	-47.230	-333.540	-53.870	2009.640	321.150
7	CA127817	2762.990	122.110	583.150	182.450	83.130	-140.450	-405.510
8	CA127818	191.090	35.920	288.850	-1200.200	-105.000	74.730	1484.720
9	CA127819	353.690	44.100	1060.630	211.890	306.340	393.080	-213.380
10	CA127820	1490.110	-89.100	1280.400	825.930	244.990	-253.770	-46.000

OK