

Babu Banarasi Das University



Case Study

On

Predictive Analysis of Medical Insurance Charges Using IBM SPSS Modeler

SUBMITTED TO:

Mr. ROBIN TYAGI

SUBMITTED BY:

Name: Jai

Roll No: 1230258201

Name: Devesh Prakash

Roll No: 1230258201

Class: BCADS3

CASE STUDY

Definition: As a Data Miner in the insurance company we conduct a predictive analysis of medical insurance charges using IBM SPSS Modeler to support data-driven decision-making within the company. The primary goal is to understand how various customer attributes — such as age, BMI, smoking status, and region — influence insurance costs.

Learning:

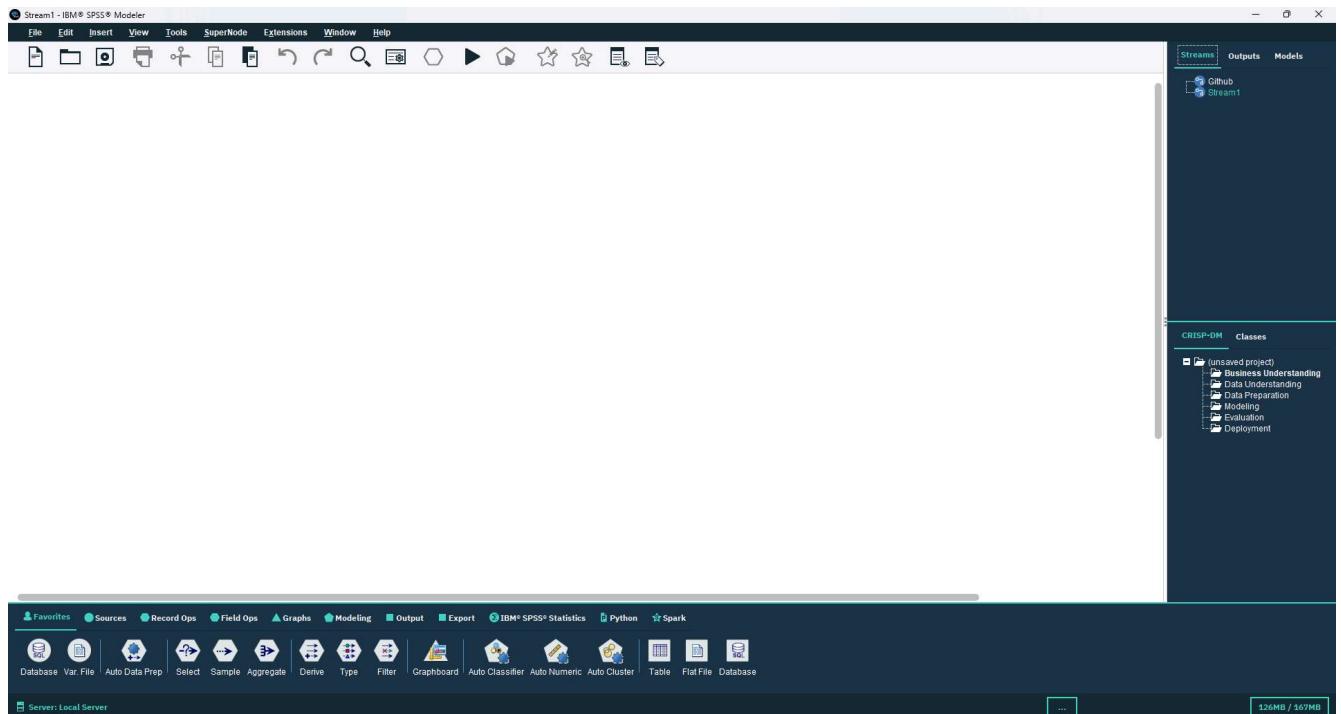
- Using the Derive node to create new fields (e.g., BMI Category, Age Group) from existing data.
- Using the Reclassify node to simplify or merge categorical values
- Identifying relationships between numeric and categorical variables

Required Tool: IBM SPSS Modeler 18.6

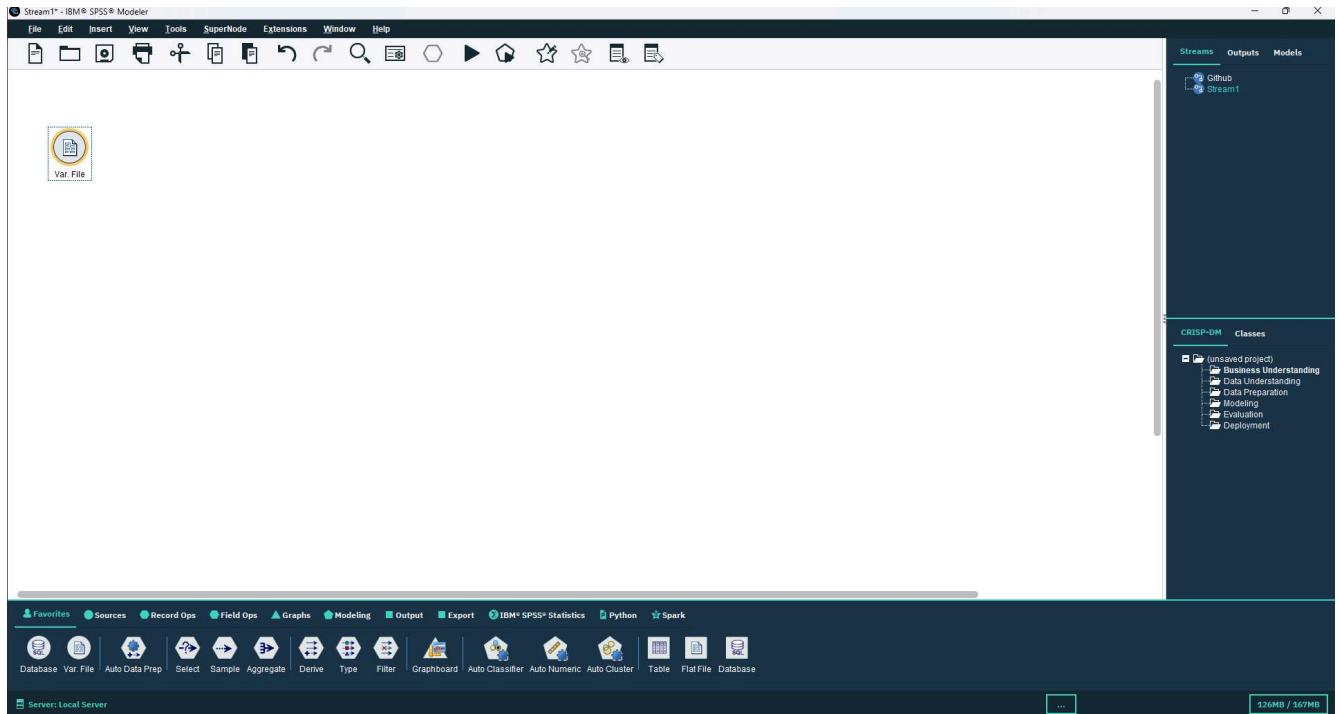
Working: This project focuses on the predictive analysis of medical insurance charges using the Insurance dataset in IBM SPSS Modeler. The aim is to understand how various demographic and lifestyle factors — such as age, BMI, smoking habits, and region — influence the cost of medical insurance.

Dataset Used: insurance.csv

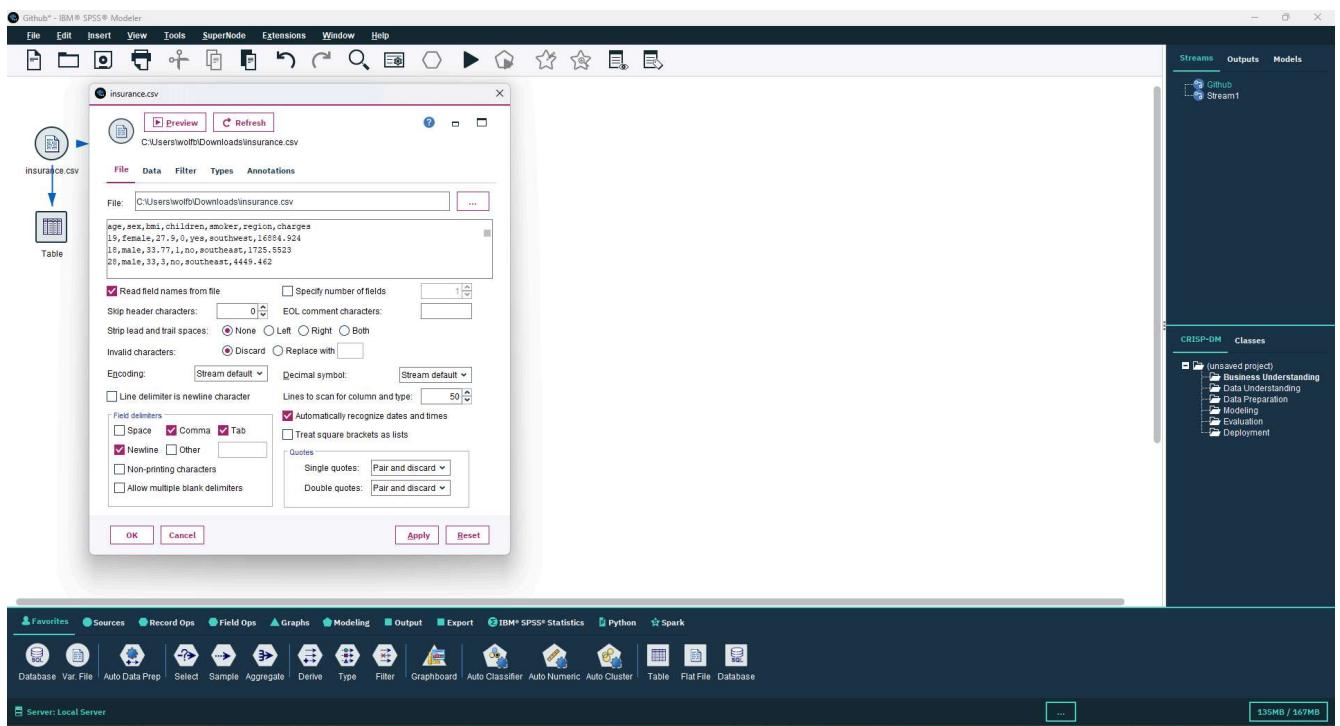
Step 1: Open IBM SPSS Modeler 18.6 on your pc



Step 2: Open SPSS Modeler tool then on Source category select Var File node (we are selecting Var file node because the data set we are using is an .csv file.)
 Double clicking on the var node will make it appear on canvas.



Step 3: Now we import a data set using the import option which can be accessed by double clicking on the Var icon on the canvas. We import a dataset “insurance.csv”.



Step 4: Connect the table node from Output Category to see the dataset.

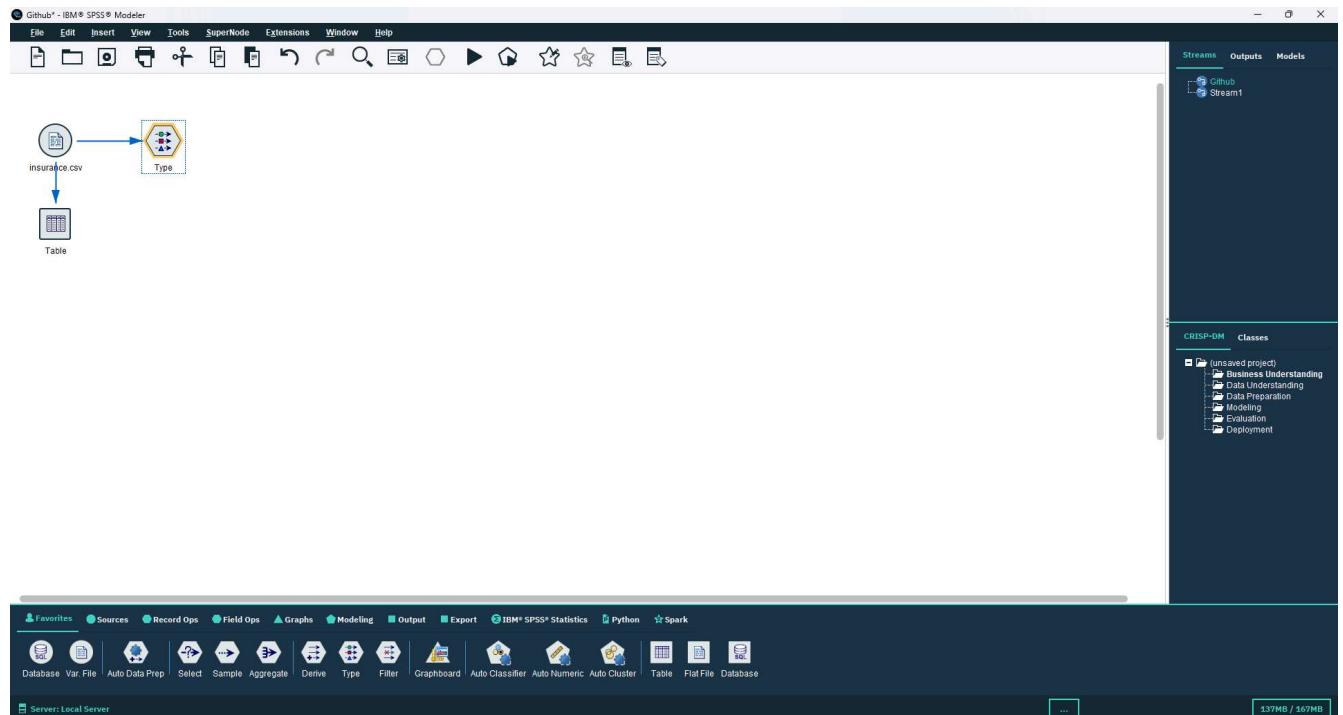
Table (7 fields, 1,338 records)

File Edit Generate Table Annotations

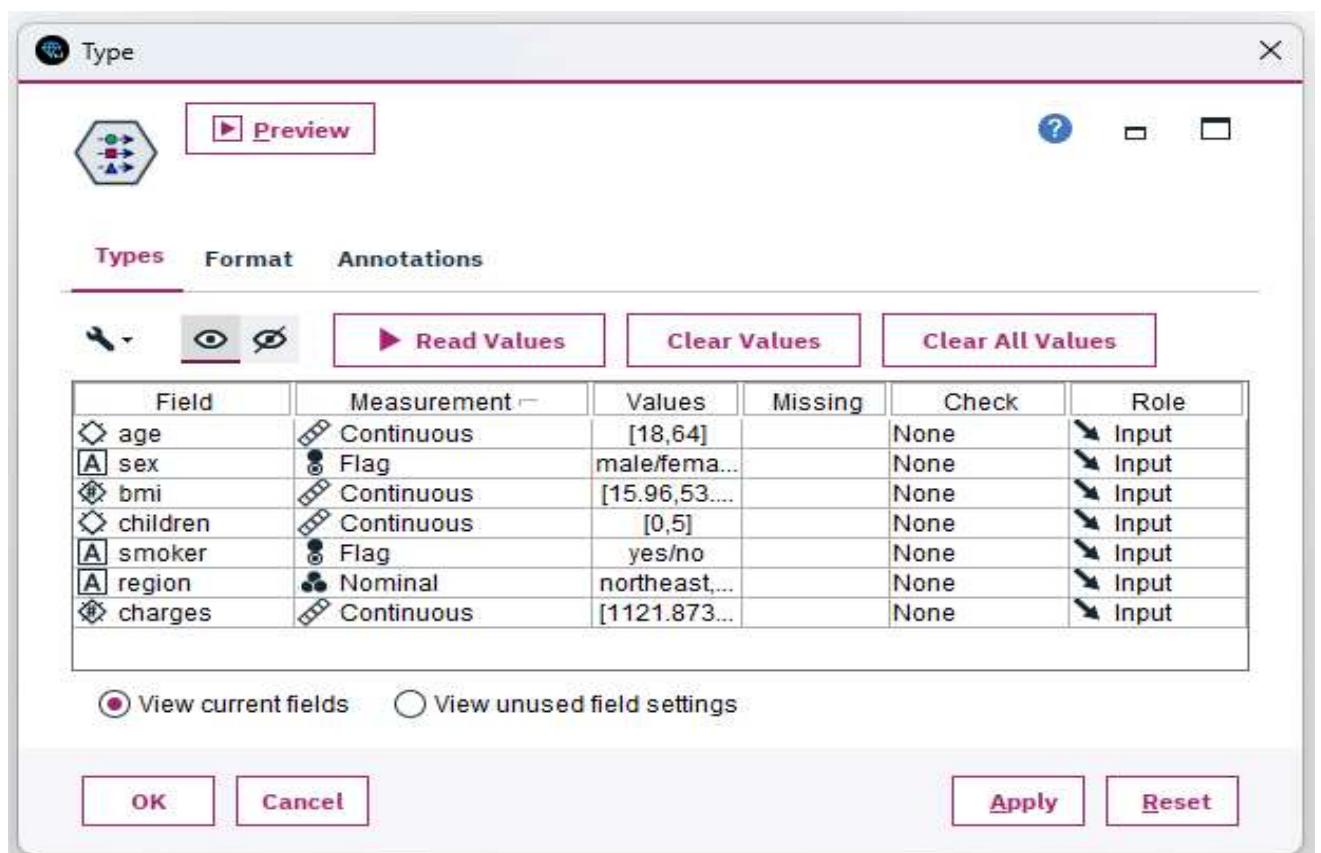
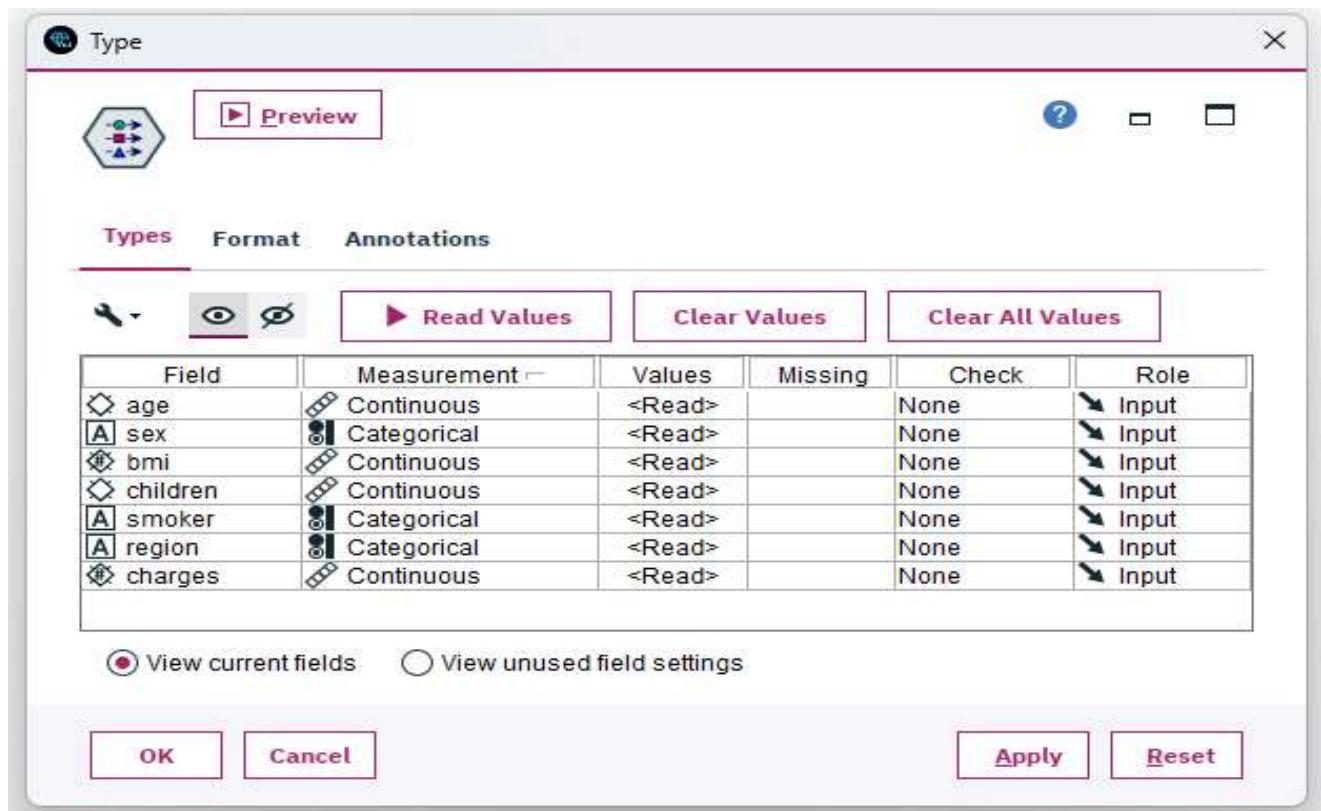
	age	sex	bmi	children	smoker	region	charges
1	19	female	27.900	1	yes	southeast	1604.524
2	19	male	27.770	1	no	southeast	1725.552
3	20	male	33.000	3	no	southeast	4445.462
4	35	male	22.705	0	no	northwest	21984.471
5	32	male	28.850	0	no	northwest	3866.055
6	31	male	29.900	0	no	southeast	1604.522
7	41	female	33.440	1	no	southeast	6240.590
8	37	female	27.740	3	no	northwest	7281.506
9	37	male	29.830	2	no	northwest	6406.411
10	60	female	25.840	0	no	northwest	29523.137
11	21	male	26.250	0	no	southeast	1604.521
12	42	female	26.250	0	yes	southeast	27805.325
13	23	male	34.400	0	no	southwest	1824.843
14	50	female	39.820	0	no	southeast	11090.718
15	27	male	42.150	0	yes	southeast	39611.758
16	18	male	24.100	0	no	southeast	1604.527
17	52	female	30.750	1	no	northeast	10797.336
18	23	male	23.845	0	no	northeast	2395.172
19	50	male	40.300	0	no	southwest	10602.395
20	30	male	32.300	0	yes	southeast	36537.457
21	60	female	32.008	0	no	northeast	11090.847
22	30	female	32.400	1	no	southwest	4145.736
23	19	male	34.100	0	no	southeast	1137.011
24	34	female	31.920	1	yes	northwest	37701.877
25	37	male	32.000	0	no	southeast	1604.528
26	50	female	37.720	3	no	northeast	14001.134
27	60	female	33.085	0	no	northeast	14451.835
28	55	female	32.775	2	no	northwest	12265.632
29	25	male	17.385	1	no	northwest	2775.192
30	31	male	34.200	2	yes	southeast	11090.800
31	22	male	35.600	1	yes	southwest	35555.576
32	10	female	26.315	0	no	northwest	21981.190
33	19	female	28.600	5	no	southwest	4687.797
34	63	male	28.310	0	no	northwest	13290.098
35	30	male	34.190	1	yes	southeast	51194.159
36	19	male	20.425	0	no	northwest	51625.434
37	62	female	32.945	3	no	northwest	15812.193
38	24	male	20.800	0	no	southwest	3302.300
39	39	female	36.670	1	yes	northwest	3393.176
40	60	male	35.590	0	no	southeast	43175.151
41	24	female	26.600	0	no	northeast	10404.062
42	31	female	36.630	2	no	southeast	6945.759
43	45	male	21.780	1	no	southeast	6272.477
44	37	male	20.200	0	no	southeast	1604.519
45	30	male	37.050	1	no	northeast	6079.472
46	50	male	37.300	0	no	southwest	20430.284
47	10	female	30.665	2	no	northwest	3393.356
48	20	female	34.770	0	no	northwest	3556.922
49	60	male	35.050	0	no	southeast	12265.477
50	34	male	35.200	1	yes	southeast	35705.176
51	10	female	35.625	0	no	northeast	2211.131
52	21	female	33.630	2	no	northwest	3575.829
53	40	male	28.000	1	yes	southwest	23566.272
54	36	male	34.430	0	yes	southeast	37742.576

OK

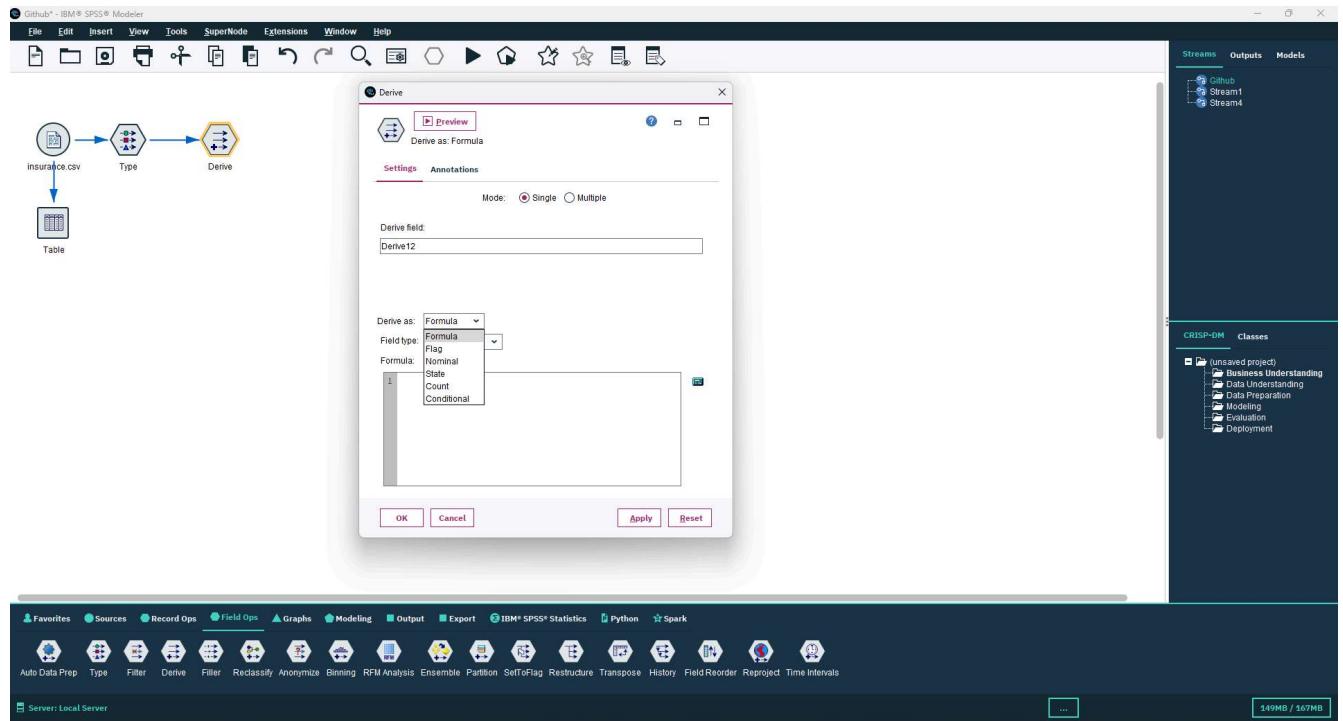
Step 5: Connect the Type Node from Field Ops. This Node aids in understanding the different Fields Measurement Values.



Step 6: Double click on the Type node on the canvas and then click on Read Values option to get the different Measurement of the fields.



Step 7: Now connect the Derive Node from Field Ops. Derive Node helps us make new fields using the existing fields in the Dataset it can be through Formula, Conditional, Nominal, Flag, State and Count.



Step 8: Name the Derive field as “BMI_Category”. In the “Derive as” section select “Nominal” which will let us divide our data into different categories based on the condition we set and let the “Field type” and “Default value” to be set as Default.

Now we set the condition and decide the value which will be set if the condition is true.

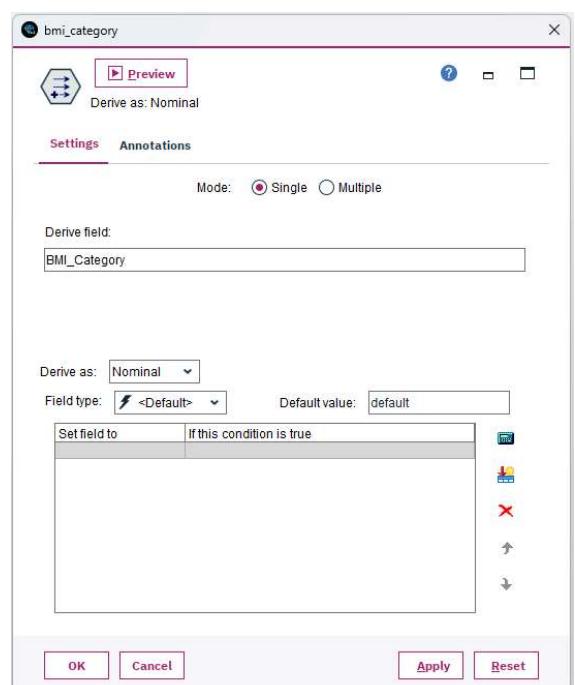
The condition is as follows:

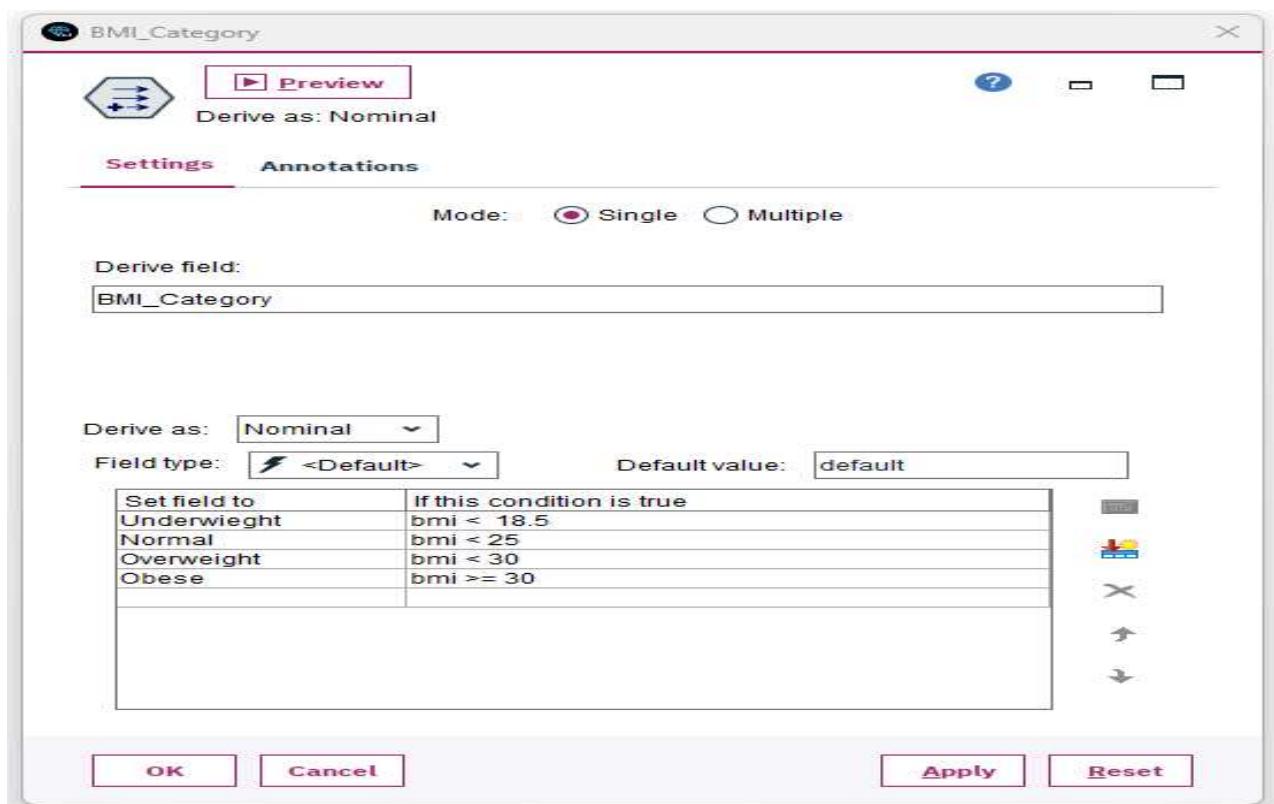
If ($bmi < 18.5$) Then the value will be 'Underweight'.

If ($bmi < 25$) Then the value will be 'Normal'.

If ($bmi < 30$) Then 'Overweight'.

And if ($bmi \geq 30$) Then the value will be 'Obese'.





Step 9: Connect the table node from Output Category to see the new “BMI_Category” Field in the Dataset.

	age	sex	bmi	children	smoker	region	charges	BMI_Category
1	19	female	27.500	0yes	southwest	1683.524	Overweight	
2	19	female	26.770	1no	southwest	1725.566	Obese	
3	20	male	33.000	3no	southeast	4445.462	Obese	
4	33	male	22.705	0no	northwest	21964.471	Normal	
5	32	male	28.880	0no	northwest	3866.855	Overweight	
6	31	female	25.740	0no	southeast	3756.622	Overweight	
7	46	female	25.440	1no	southeast	6240.390	Obese	
8	31	male	27.350	2no	southeast	5186.558	Overweight	
9	37	male	29.830	2no	northwest	6406.411	Overweight	
10	60	female	25.840	0no	northwest	29323.137	Overweight	
11	25	male	26.220	0no	northwest	2721.321	Overweight	
12	62	female	26.250	0yes	southeast	27808.725	Overweight	
13	23	male	34.400	0no	southwest	1826.843	Obese	
14	56	female	29.020	0no	southeast	11861.178	Obese	
15	31	male	42.130	0yes	southeast	51161.178	Obese	
16	19	male	24.600	1no	southeast	1387.237	Normal	
17	52	female	30.780	1no	northwest	10797.336	Obese	
18	23	male	23.845	0no	northwest	2395.172	Normal	
19	56	male	40.300	0no	southwest	10602.393	Obese	
20	30	male	35.300	0yes	southwest	36837.467	Obese	
21	39	female	26.000	0no	southeast	15301.194	Obese	
22	39	female	32.400	1no	southeast	1145.734	Obese	
23	19	male	34.100	0no	southeast	1137.011	Obese	
24	34	female	31.920	1yes	northwest	37701.877	Obese	
25	37	male	28.025	2no	northwest	6203.902	Overweight	
26	59	female	27.720	3no	southeast	14205.934	Overweight	
27	67	female	30.095	0no	southeast	14851.833	Obese	
28	51	female	32.775	2no	northwest	11246.432	Obese	
29	23	male	17.315	1no	northwest	2775.192	Underweight	
30	31	male	36.300	2yes	southwest	35711.500	Obese	
31	22	male	35.600	0yes	southwest	35555.576	Obese	
32	32	female	26.315	0no	northwest	2195.190	Overweight	
33	19	female	29.300	0no	southeast	11861.178	Overweight	
34	60	male	33.310	0no	northwest	15770.089	Overweight	
35	28	male	36.400	1yes	southwest	51164.559	Obese	
36	19	male	20.428	0no	northwest	1625.434	Normal	
37	62	female	32.965	3no	northwest	15412.193	Obese	
38	24	male	20.800	0no	southwest	2302.300	Normal	
39	35	male	36.670	1yes	northwest	35512.176	Obese	
40	49	female	31.000	0no	southeast	49173.444	Obese	
41	24	female	26.600	1no	northwest	1044.062	Overweight	
42	31	female	36.630	2no	southeast	4545.759	Obese	
43	41	male	21.780	1no	southeast	6272.477	Normal	
44	37	female	30.800	2no	southeast	6313.759	Obese	
45	38	male	37.050	1no	northwest	6145.672	Obese	
46	50	female	37.000	0no	southeast	20430.166	Obese	
47	19	female	38.665	2no	northwest	3391.356	Obese	
48	20	female	34.770	0no	northwest	3556.922	Obese	
49	60	female	24.530	0no	southeast	12625.897	Normal	
50	36	male	35.200	1yes	southeast	35709.176	Obese	
51	19	female	35.625	0no	northwest	2211.131	Obese	
52	23	female	33.630	2no	northwest	3571.839	Obese	
53	40	male	20.000	1yes	southeast	23580.272	Overweight	
54	38	male	34.450	0yes	southeast	37742.176	Obese	

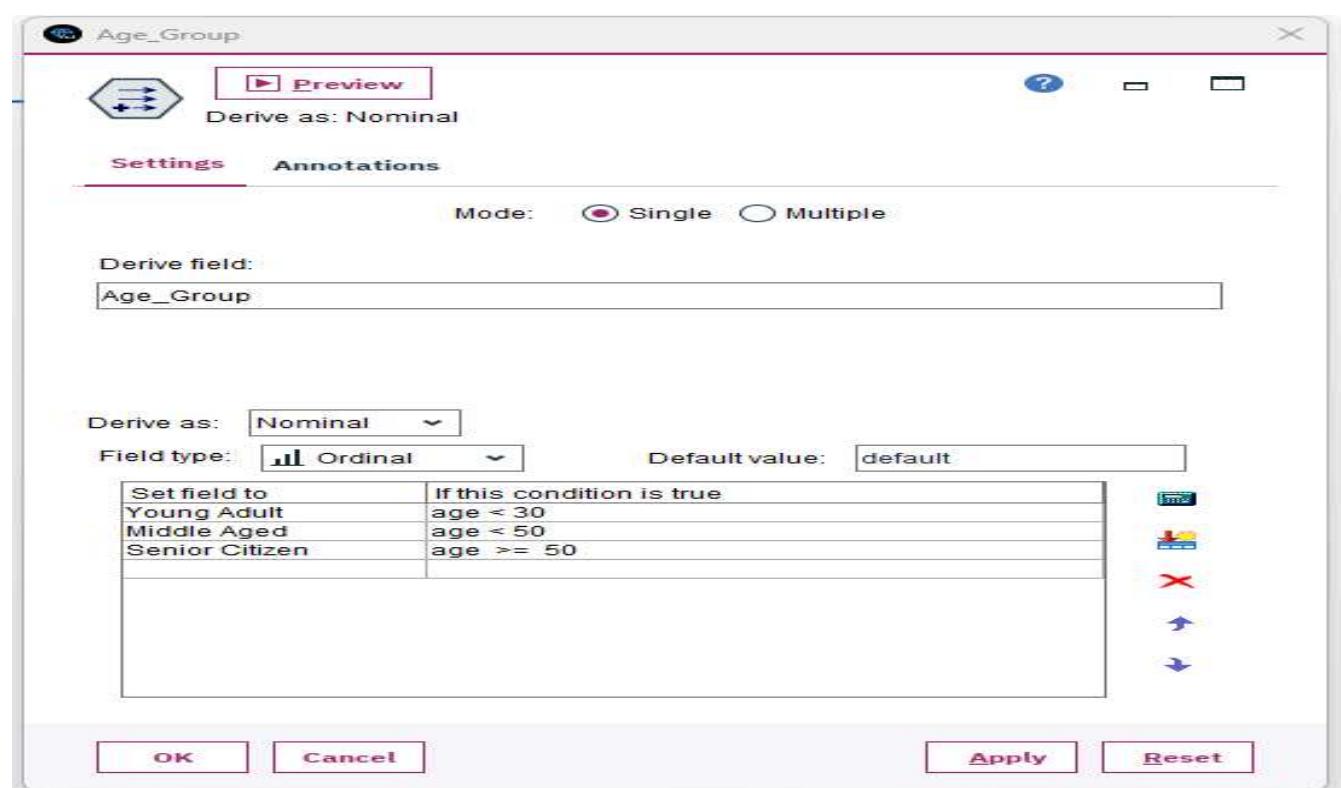
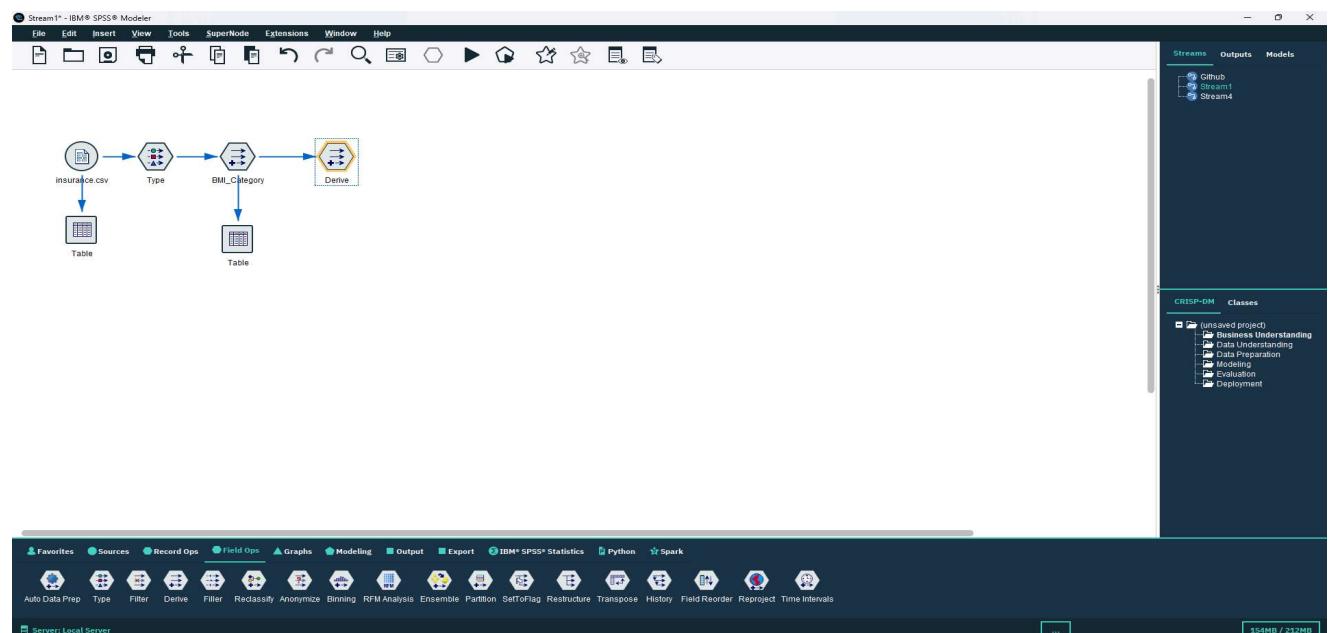
Step 10: Connect a new Derive Node and name it as “Age_Group”. We are dividing our customers by their age. As done in above Step 8 we select “Nominal” in the “Derive as” section and let the other options set as Default.

Now we set condition as follows:

If (age < 30) Then value will be 'Young Adult'

If (age < 50) Then value will be 'Middle Aged'

And if (age >= 50) Then value will be 'Senior Citizen'



Step 11: Connect the table node from Output Category to see the new “Age_Group” Field in the Dataset.

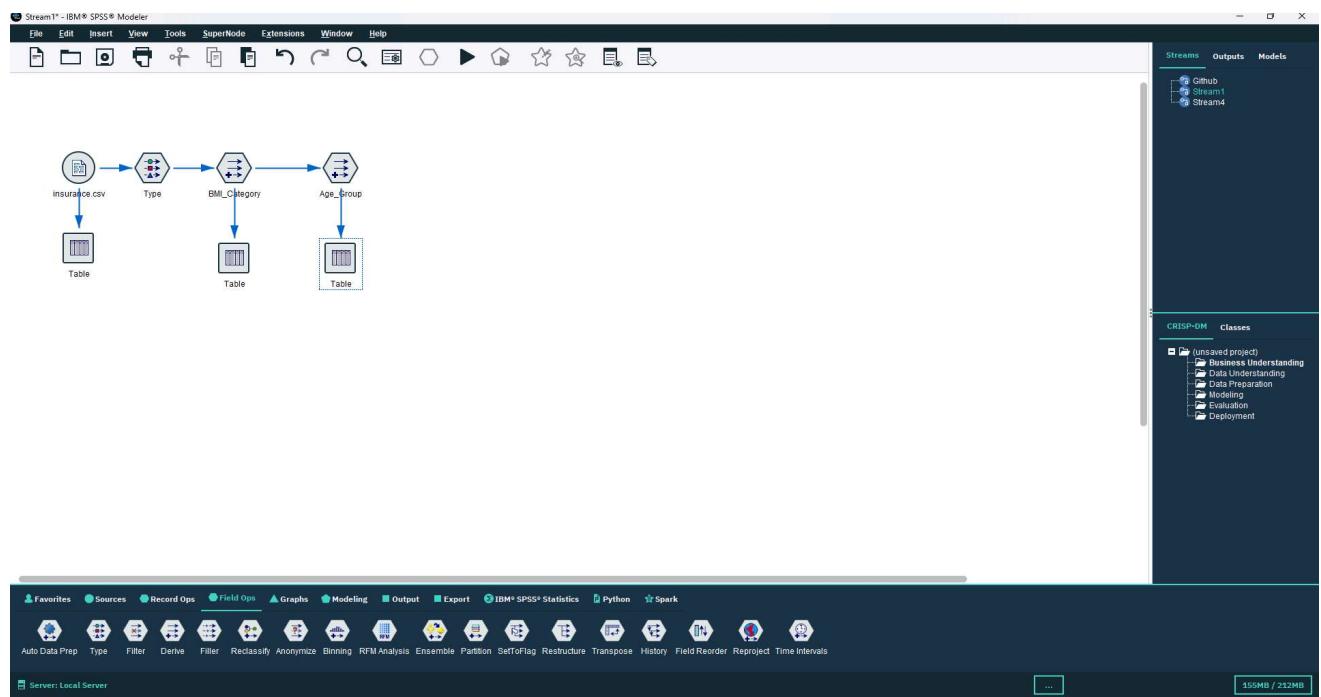


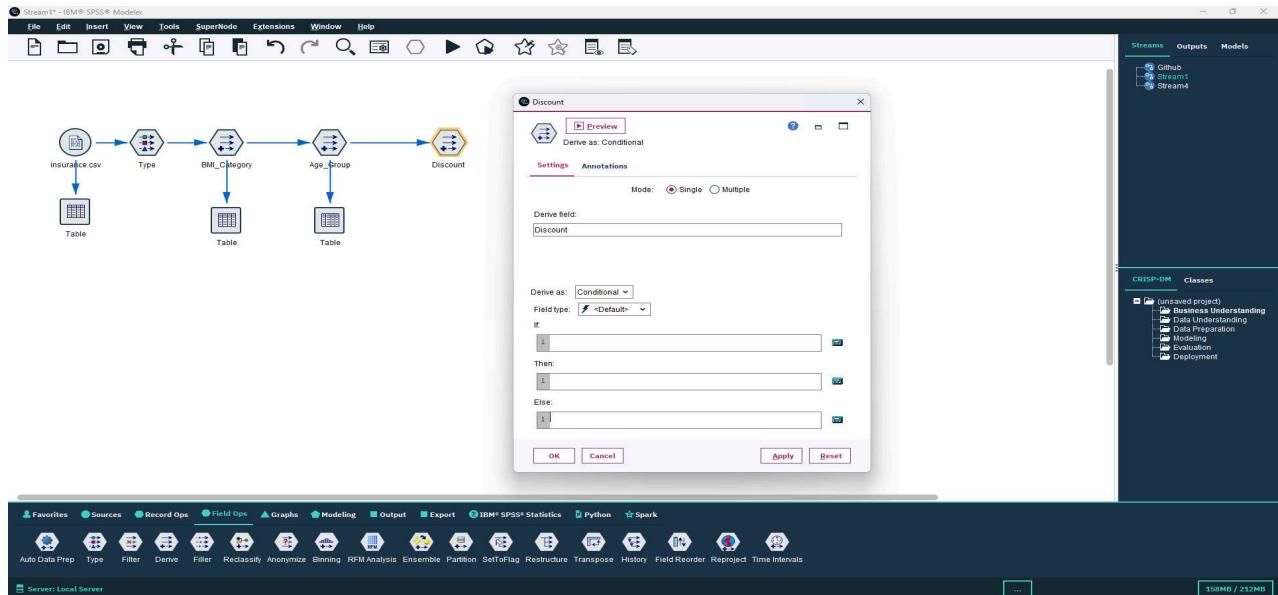
Table (9 fields, 1338 records) #2

	age	sex	bmi	children	smoker	region	charges	BMI_Catagory	Age_Group
1	18	female	27.900	0yes	southwest	16884.924	Overweight	Young Adult	
2	18	male	33.770	1no	southeast	1725.552	Obese	Young Adult	
3	28	male	33.000	3no	southeast	4445.462	Obese	Young Adult	
4	35	male	22.705	0no	northwest	21584.471	Normal	Middle Aged	
5	32	female	28.880	0no	northwest	3866.355	Overweight	Middle Aged	
6	31	female	25.740	0no	southeast	3756.622	Normal	Middle Aged	
7	46	female	33.440	1no	southeast	9240.590	Obese	Middle Aged	
8	37	female	27.740	3no	northwest	7281.506	Overweight	Middle Aged	
9	37	male	29.830	2no	northeast	4406.411	Overweight	Middle Aged	
10	60	female	25.840	0no	northwest	29523.137	Overweight	Senior Citizen	
11	25	male	26.220	0no	northeast	2721.321	Overweight	Young Adult	
12	62	female	26.250	0yes	southeast	27085.725	Overweight	Senior Citizen	
13	23	male	34.400	0no	southwest	1326.843	Obese	Young Adult	
14	54	female	35.820	0no	southeast	11090.718	Obese	Senior Citizen	
15	27	male	42.130	0yes	southeast	39611.159	Obese	Young Adult	
16	19	male	24.600	1no	southeast	1337.237	Normal	Young Adult	
17	52	female	30.760	1no	northeast	10197.336	Obese	Senior Citizen	
18	22	male	23.845	0no	northeast	2395.172	Normal	Young Adult	
19	56	male	40.300	0no	southwest	16012.355	Obese	Senior Citizen	
20	36	male	35.300	0yes	southwest	36837.467	Obese	Middle Aged	
21	60	female	36.005	0no	northeast	13228.847	Obese	Senior Citizen	
22	30	female	32.400	1no	southwest	4145.736	Obese	Middle Aged	
23	18	male	34.100	0no	southeast	1137.011	Obese	Young Adult	
24	34	female	31.920	1yes	northeast	37701.877	Obese	Middle Aged	
25	37	male	28.025	2no	northeast	6203.902	Overweight	Middle Aged	
26	59	female	27.720	3no	southeast	14001.134	Overweight	Senior Citizen	
27	63	female	23.085	0no	northeast	14451.835	Normal	Senior Citizen	
28	55	female	32.775	2no	northwest	12265.632	Obese	Senior Citizen	
29	23	male	17.385	1no	northwest	2775.192	Underweight	Young Adult	
30	31	male	36.300	2yes	southwest	35711.000	Obese	Middle Aged	
31	22	male	35.600	0yes	southwest	35585.576	Obese	Young Adult	
32	19	female	26.315	0no	northeast	2198.190	Overweight	Young Adult	
33	19	female	28.600	5no	southwest	4687.797	Overweight	Young Adult	
34	65	male	28.310	0no	northwest	13770.098	Overweight	Senior Citizen	
35	28	male	36.400	1yes	southwest	51164.559	Obese	Young Adult	
36	19	male	20.425	0no	northwest	1625.434	Normal	Young Adult	
37	62	female	32.945	3no	northwest	15612.193	Obese	Senior Citizen	
38	24	male	20.800	0no	southwest	2302.300	Normal	Young Adult	
39	35	male	36.670	1yes	northeast	39774.276	Obese	Middle Aged	
40	60	male	35.900	0yes	southwest	49173.361	Obese	Senior Citizen	
41	24	female	26.600	1no	northeast	3044.062	Overweight	Young Adult	
42	32	female	36.630	2no	southeast	4445.759	Obese	Middle Aged	
43	41	male	21.780	1no	southeast	6272.477	Normal	Middle Aged	
44	37	female	30.800	2no	southeast	6313.759	Obese	Middle Aged	
45	38	male	37.050	1no	northeast	4075.472	Obese	Middle Aged	
46	55	male	37.300	0no	southwest	20630.284	Obese	Senior Citizen	
47	18	female	35.665	2no	northeast	3393.356	Obese	Young Adult	
48	28	female	34.770	1no	northwest	3356.922	Obese	Young Adult	
49	60	female	24.530	0no	southeast	12625.897	Normal	Senior Citizen	
50	36	male	35.200	1yes	southeast	35705.176	Obese	Middle Aged	
51	18	female	35.625	0no	northeast	2211.131	Obese	Young Adult	
52	21	female	33.630	2no	northwest	3575.829	Obese	Young Adult	
53	49	male	28.000	1yes	southwest	23565.272	Overweight	Middle Aged	
54	36	male	34.430	0yes	southeast	37742.576	Obese	Middle Aged	

OK

- Now we can analyze charges through the new fields “Age_Group” and “BMI_Category”.

Step 12: Now we connect a new Derive node to derive a new field with the help of the “Age_Group” field that we created with the help of “age” field. We name this Derive field as “Discount”. In this new field we are using the “Age_Group” field to give Senior Citizens a discount on charges.



Step 13: We use Conditional as the Derive method here and Set the condition as follows:
If

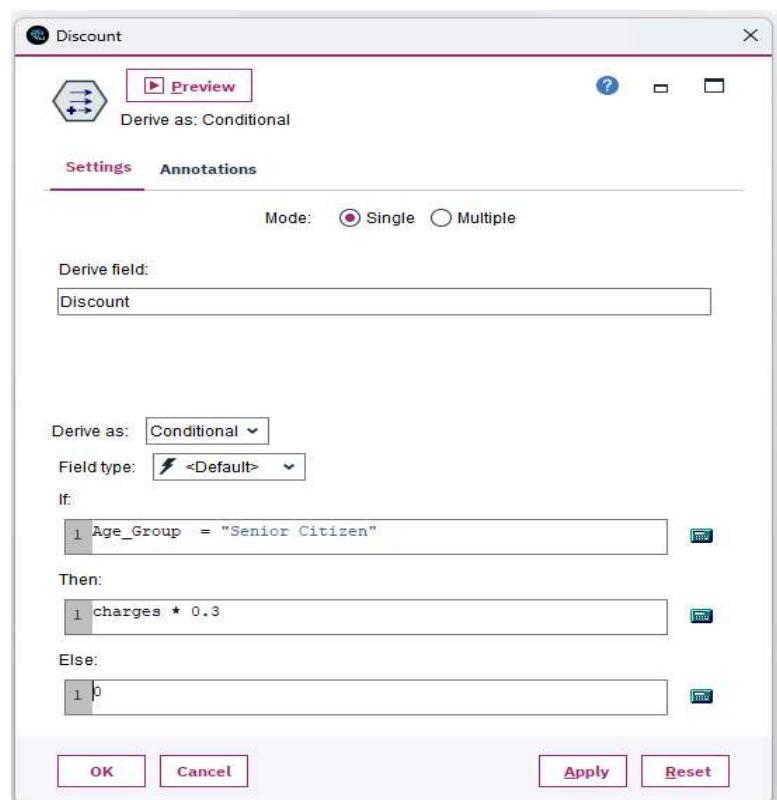
Age_Group = “Senior Citizen”

Then

Charges * 0.3

Else

0



Step 14: Connect the table node from Output Category to see the new “Discount” Field in the Dataset.

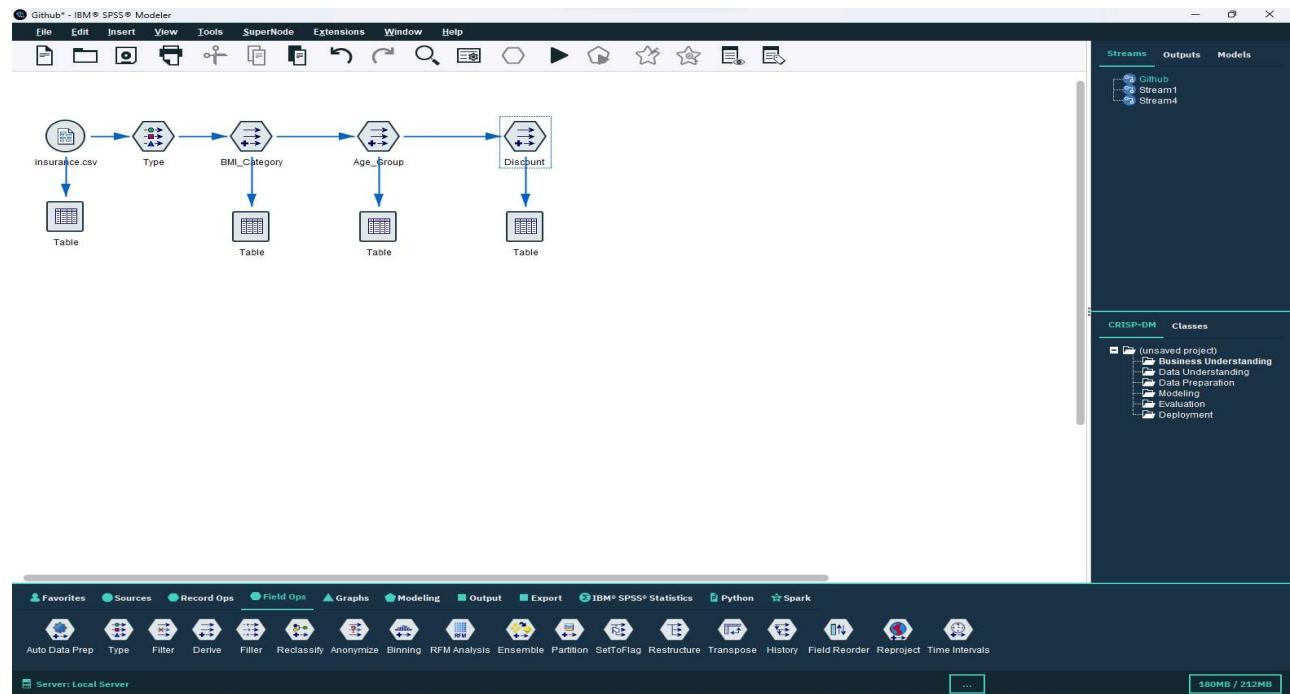


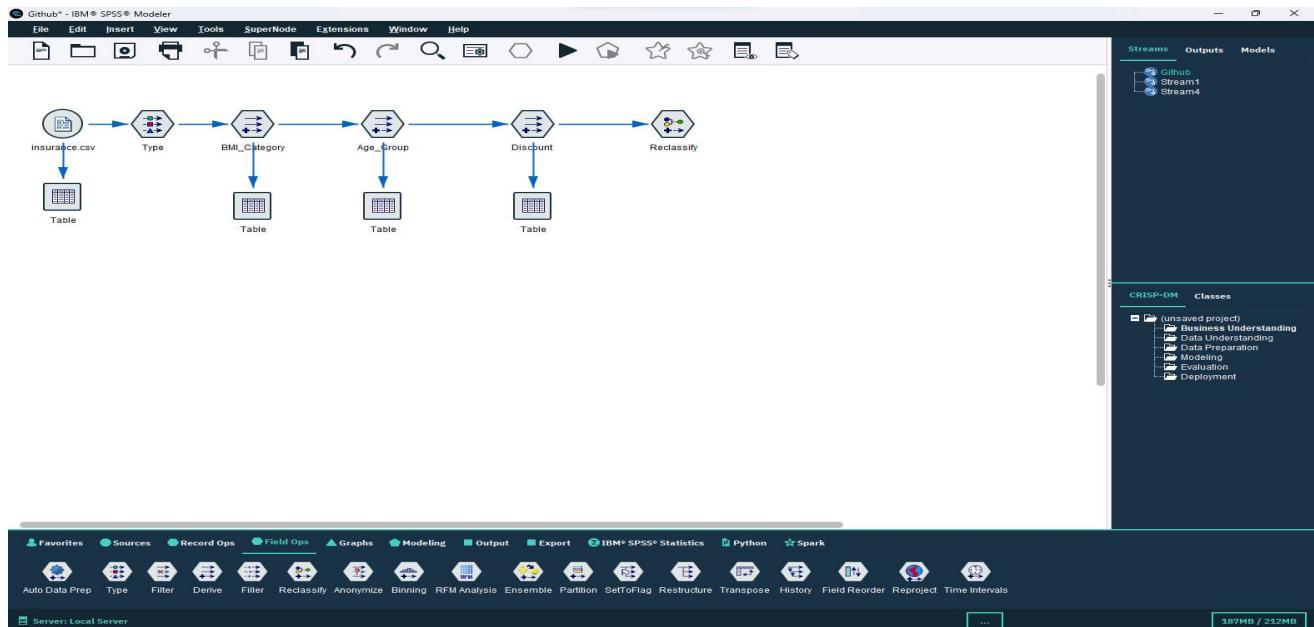
Table (10 fields, 1,338 records) #2

Table Annotations

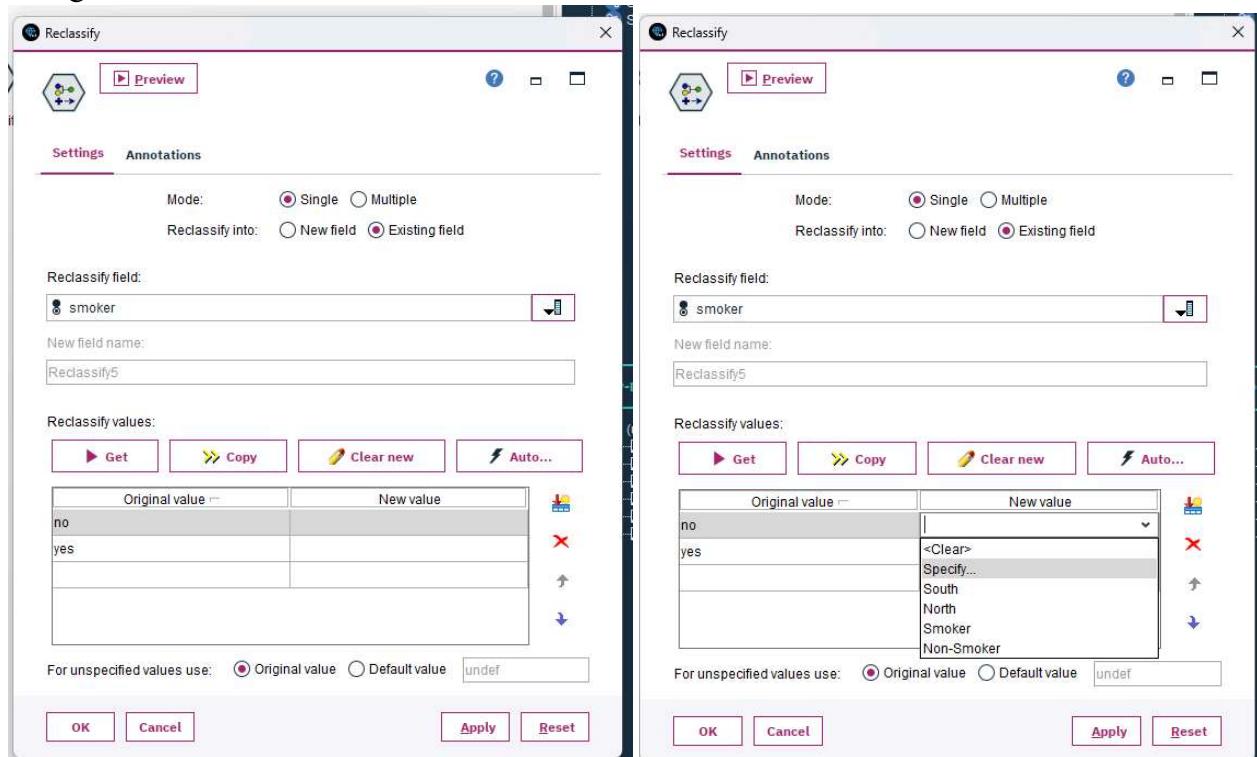
	age	sex	bmi	children	smoker	region	charges	BMI_Catagory	Age_Group	Discount
1	19	female	27.900	0	yes	southwest	16584.592	Overweight	Young Adult	0.000
2	18	male	31.770	1	no	southeast	12050.562	Obese	Young Adult	0.000
3	35	male	33.000	0	no	southeast	4446.162	Obese	Young Adult	0.000
4	44	male	22.705	0	no	northwest	21884.471	Normal	Middle Aged	0.000
5	32	male	28.880	0	no	northwest	3864.655	Overweight	Middle Aged	0.000
6	31	female	25.740	0	no	southeast	3756.622	Overweight	Middle Aged	0.000
7	46	female	33.440	1	no	southeast	8240.590	Obese	Middle Aged	0.000
8	37	female	27.740	3	no	northwest	7281.506	Overweight	Middle Aged	0.000
9	37	male	28.830	2	no	northwest	6406.411	Overweight	Middle Aged	0.000
10	60	female	25.840	0	no	northwest	28923.137	Overweight	Senior Citizen	8676.941
11	25	male	26.220	0	no	northwest	2721.321	Overweight	Young Adult	0.000
12	42	female	26.290	0	yes	southeast	27808.725	Overweight	Senior Citizen	8342.618
13	23	male	34.400	0	no	southwest	1826.643	Obese	Young Adult	0.000
14	56	female	35.920	0	no	southeast	11090.718	Obese	Senior Citizen	3327.215
15	27	male	42.130	0	yes	southeast	36611.758	Obese	Young Adult	0.000
16	19	male	24.600	1	no	southwest	1837.237	Normal	Young Adult	0.000
17	52	female	30.780	1	no	northwest	10797.336	Obese	Senior Citizen	3239.201
18	23	male	23.845	0	no	northwest	2395.172	Normal	Young Adult	0.000
19	56	male	40.300	0	no	southwest	10602.385	Obese	Senior Citizen	3180.715
20	30	male	35.300	0	yes	southwest	36837.467	Obese	Middle Aged	0.000
21	60	female	36.005	0	no	northwest	13228.847	Obese	Senior Citizen	3968.654
22	30	female	32.400	1	no	southwest	4149.736	Obese	Middle Aged	0.000
23	18	male	34.100	0	no	southeast	1137.011	Obese	Young Adult	0.000
24	34	female	31.920	1	yes	northwest	37701.877	Obese	Middle Aged	0.000
25	37	male	28.025	2	no	northwest	6203.902	Overweight	Middle Aged	0.000
26	59	female	27.720	3	no	southeast	14001.134	Overweight	Senior Citizen	4200.340
27	63	male	33.085	0	no	northwest	14451.835	Normal	Senior Citizen	4335.551
28	55	female	32.075	2	no	northwest	12268.632	Obese	Senior Citizen	3680.590
29	23	male	17.385	1	no	northwest	2775.192	Underweight	Young Adult	0.000
30	31	male	36.300	2	yes	southwest	38711.000	Obese	Middle Aged	0.000
31	22	male	35.600	0	yes	southeast	35585.576	Obese	Young Adult	0.000
32	18	female	36.315	0	no	northwest	2198.150	Overweight	Young Adult	0.000
33	19	female	28.600	5	no	southwest	4687.797	Overweight	Young Adult	0.000
34	63	male	28.310	0	no	northwest	13770.098	Overweight	Senior Citizen	4131.029
35	28	male	36.400	1	yes	southeast	51194.549	Obese	Young Adult	0.000
36	19	male	20.425	0	no	northwest	14166.434	Normal	Young Adult	0.000
37	62	female	32.465	3	no	northwest	15612.593	Obese	Senior Citizen	4650.658
38	30	male	20.500	0	no	southeast	23052.300	Normal	Young Adult	0.000
39	35	male	36.470	1	yes	northwest	39774.276	Obese	Middle Aged	0.000
40	60	male	39.500	0	yes	southeast	48173.361	Obese	Senior Citizen	14452.008
41	24	female	28.600	0	no	northwest	3046.062	Overweight	Young Adult	0.000
42	31	female	36.820	2	no	southeast	4945.759	Obese	Middle Aged	0.000
43	41	male	21.780	1	no	southeast	6272.477	Normal	Middle Aged	0.000
44	37	male	30.800	2	no	southeast	6313.759	Obese	Middle Aged	0.000
45	38	male	37.050	1	no	northwest	6079.672	Obese	Middle Aged	0.000
46	55	male	37.300	0	no	southwest	20630.284	Obese	Senior Citizen	6189.085
47	18	female	38.665	2	no	northwest	3393.356	Obese	Young Adult	0.000
48	28	female	34.770	0	no	northwest	3556.922	Obese	Young Adult	0.000
49	60	female	24.530	0	no	southeast	12625.897	Normal	Senior Citizen	3788.969
50	36	male	35.200	1	yes	southeast	38709.176	Obese	Middle Aged	0.000
51	18	female	35.625	0	no	northeast	2211.131	Obese	Young Adult	0.000
52	21	female	33.630	2	no	northwest	3579.829	Obese	Young Adult	0.000
53	48	male	28.000	1	yes	southwest	23568.272	Overweight	Middle Aged	0.000
54	36	male	34.430	0	yes	southeast	37742.576	Obese	Middle Aged	0.000

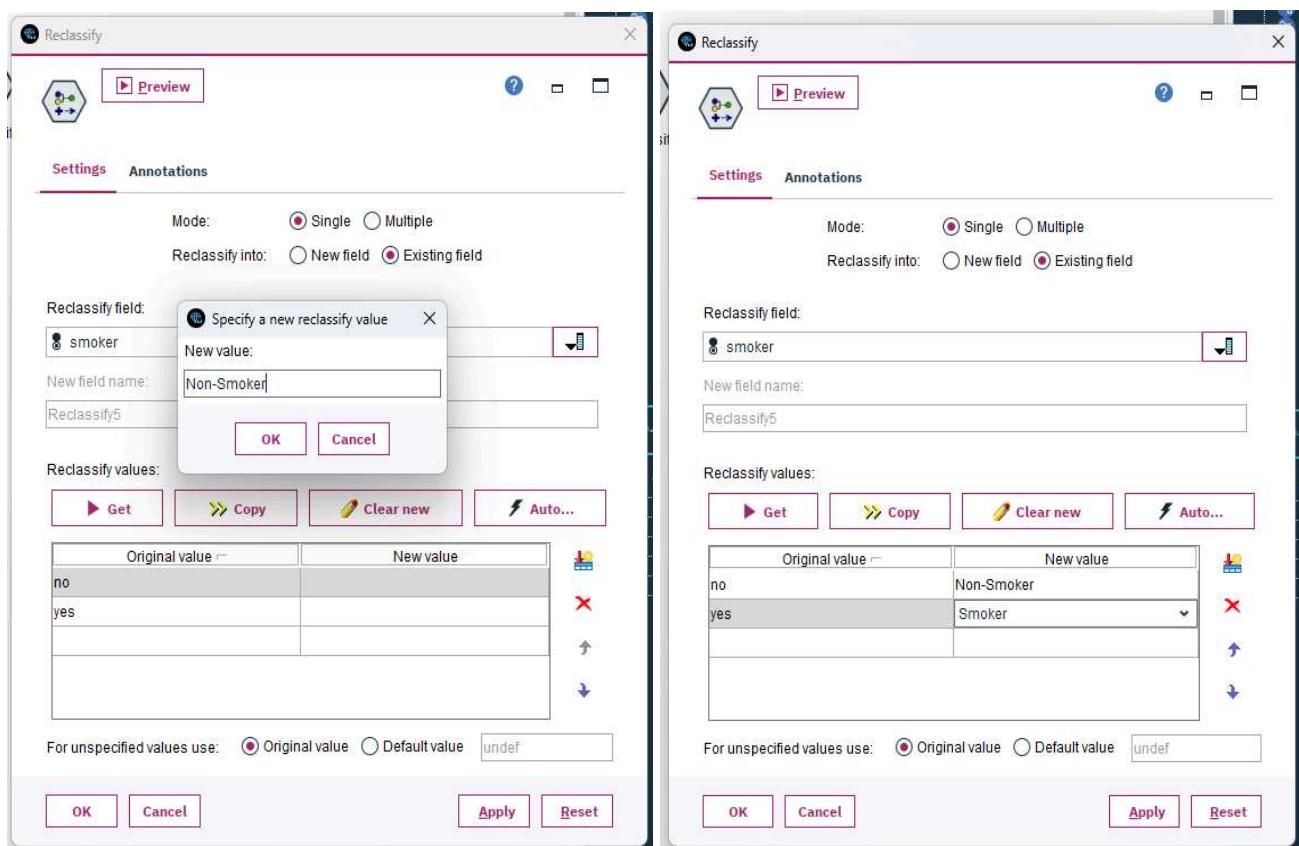
OK

Step 15: Connect a new node Reclassify from Field ops. We use Reclassify to change values in the existing field without making a new Field. This helps to make the values more specific and understandable.



Step 16: Now double click on the Reclassify node on the canvas and select “smoker” in Reclassify Field and then click on the “Get Values” option to get all the different values in the smoker field i.e ‘yes’ and ‘no’. Now in the New value column select the box and click on specify and we specify the new value as ‘Non-Smoker’ in the place of ‘no’ and do the same for ‘yes’ and change it to ‘Smoker’.





Step 17: Connect the table node from Output Category to see the changes in values in “smoker” Field in the Dataset.

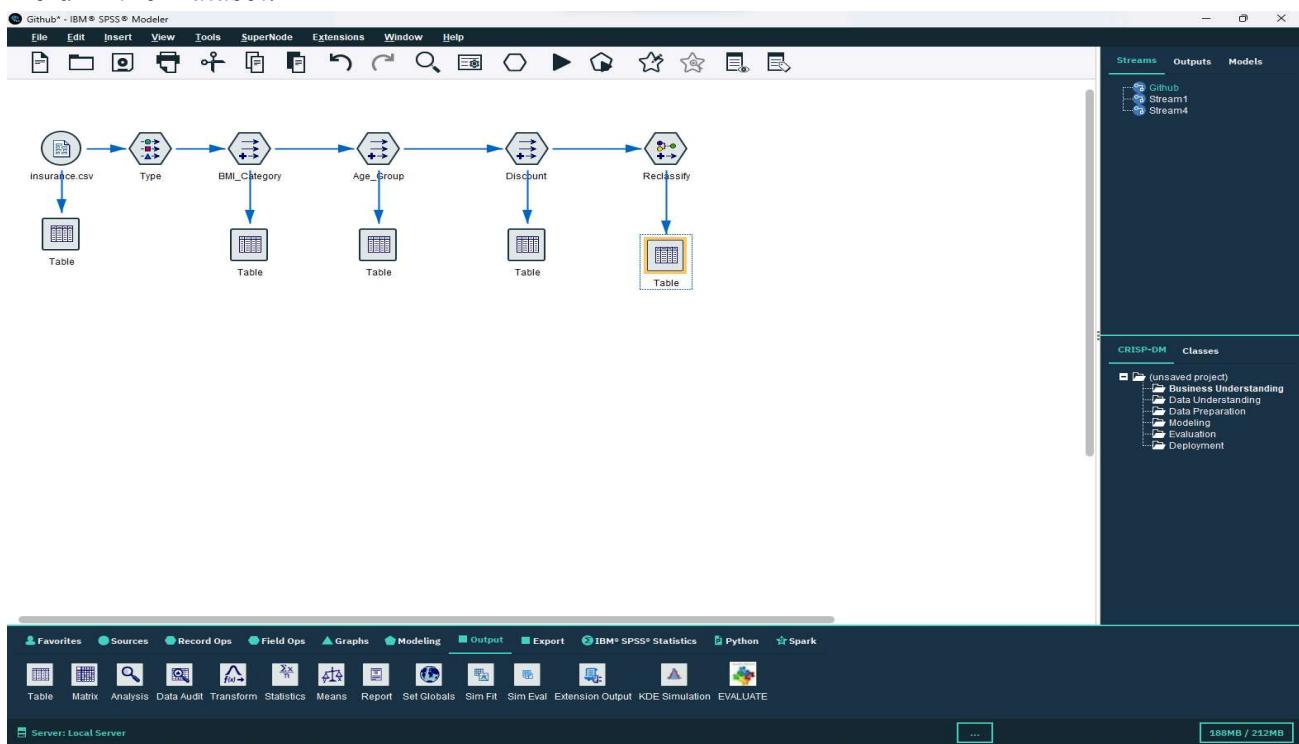


Table (10 fields, 1,338 records) #3

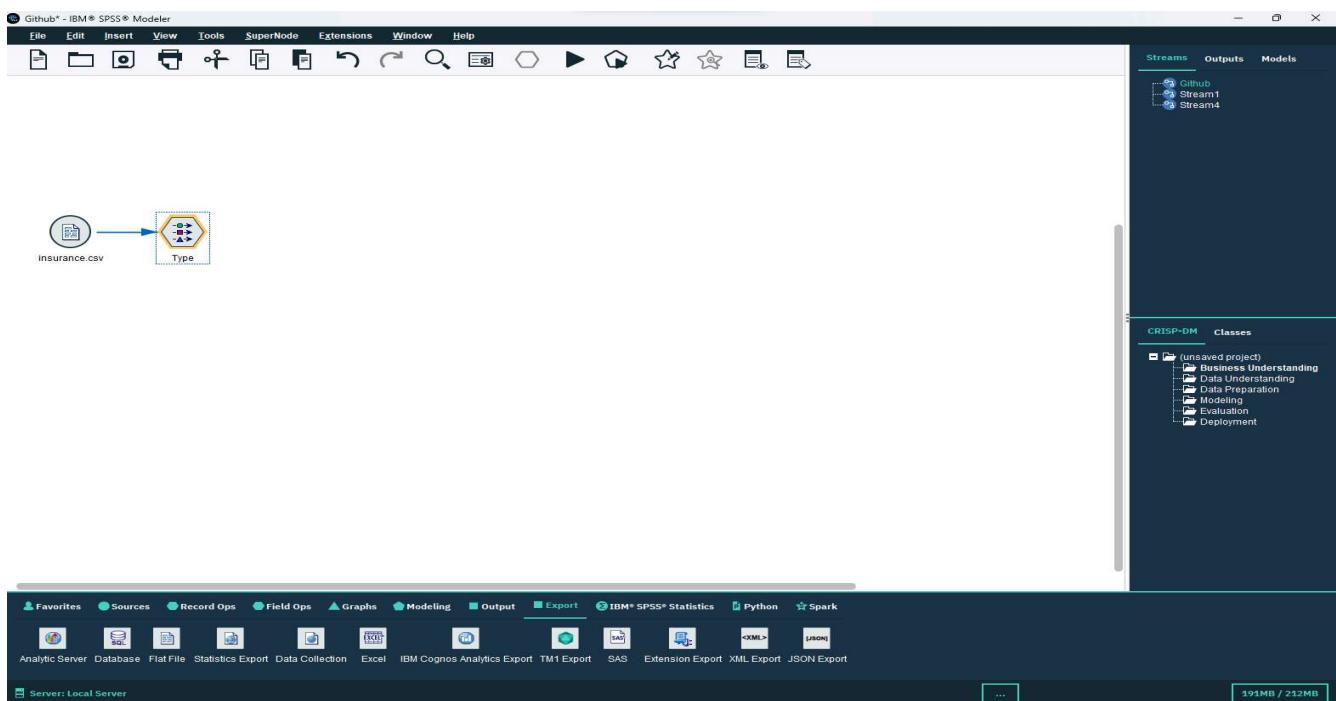
File Edit Generate

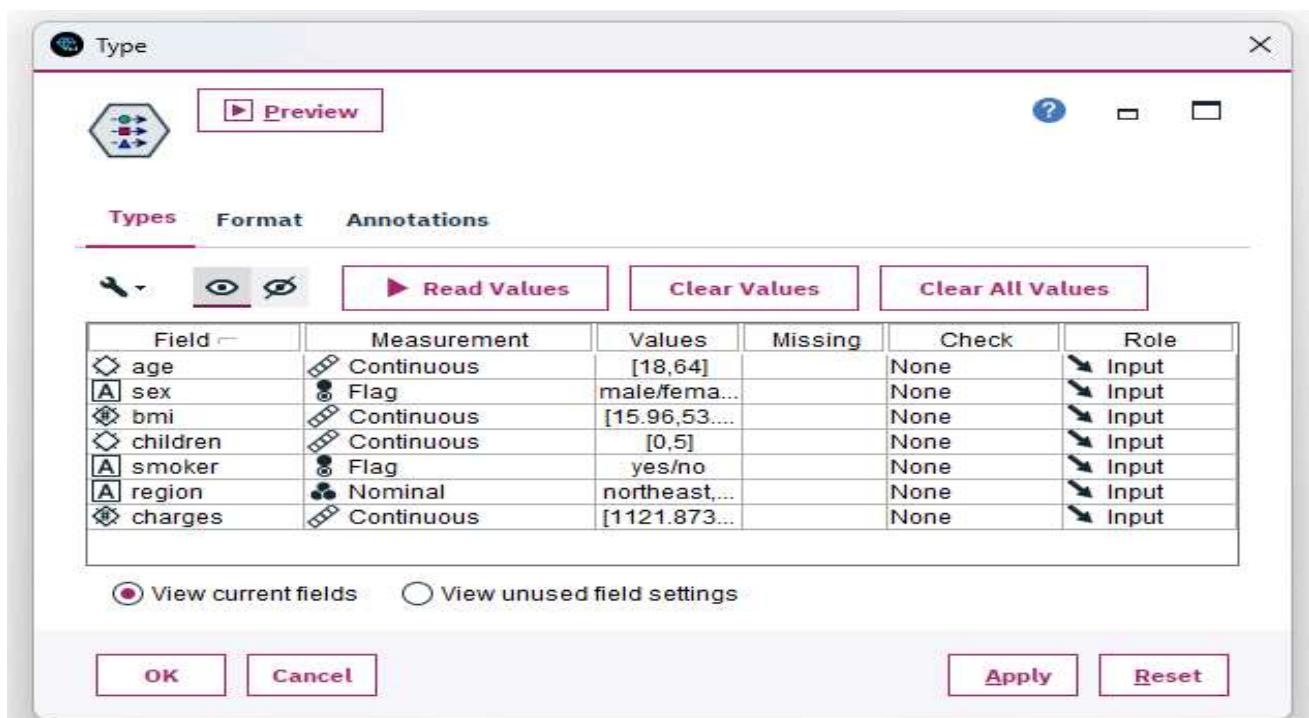
Table Annotations

	age	sex	bmi	children	smoker	region	charges	BMI_Category	Age_Group	Discount
1	18	female	27.900	0	Smoker	southeast	16884.924	Overweight	Young Adult	0.000
2	18	male	33.770	1	Non-Smoker	southeast	1725.552	Obese	Young Adult	0.000
3	28	male	33.000	3	Non-Smoker	southeast	4449.462	Obese	Young Adult	0.000
4	33	male	22.705	0	Non-Smoker	northwest	21984.471	Normal	Middle Aged	0.000
5	32	male	28.880	0	Non-Smoker	northwest	3866.855	Overweight	Middle Aged	0.000
6	31	female	25.740	0	Non-Smoker	Non-Smoked	3756.620	Overweight	Middle Aged	0.000
7	46	female	33.440	1	Non-Smoker	southeast	8240.590	Obese	Middle Aged	0.000
8	37	male	27.740	3	Non-Smoker	northwest	7281.506	Overweight	Middle Aged	0.000
9	37	male	29.830	2	Non-Smoker	northwest	6406.411	Overweight	Middle Aged	0.000
10	60	female	25.840	0	Non-Smoker	northwest	28923.137	Overweight	Senior Citizen	8676.941
11	25	male	26.220	0	Non-Smoker	northwest	2721.321	Overweight	Young Adult	0.000
12	62	female	26.290	0	Smoker	southeast	27808.725	Overweight	Senior Citizen	8342.618
13	23	male	34.400	0	Non-Smoker	southwest	1826.843	Obese	Young Adult	0.000
14	56	female	39.820	0	Non-Smoker	southeast	11090.718	Obese	Senior Citizen	3327.215
15	27	male	42.130	0	Smoker	southeast	39611.758	Obese	Young Adult	0.000
16	19	male	24.600	1	Non-Smoker	southeast	1837.237	Normal	Young Adult	0.000
17	52	female	30.780	1	Non-Smoker	northwest	10797.336	Obese	Senior Citizen	3239.201
18	23	male	23.845	0	Non-Smoker	northwest	2395.172	Normal	Young Adult	0.000
19	56	male	40.300	0	Non-Smoker	southwest	10602.385	Obese	Senior Citizen	3180.715
20	30	male	35.300	0	Smoker	southwest	36937.467	Obese	Middle Aged	0.000
21	60	female	36.005	0	Non-Smoker	northwest	13228.847	Obese	Senior Citizen	3968.654
22	30	female	32.400	1	Non-Smoker	southeast	4149.736	Obese	Middle Aged	0.000
23	18	male	34.000	0	Non-Smoker	southeast	11090.718	Obese	Young Adult	0.000
24	34	female	31.920	1	Smoker	northwest	37701.877	Obese	Middle Aged	0.000
25	34	male	20.025	2	Non-Smoker	northwest	6203.902	Overweight	Middle Aged	0.000
26	59	female	27.720	3	Non-Smoker	southeast	14001.134	Overweight	Senior Citizen	4200.340
27	63	female	23.085	0	Non-Smoker	northwest	14451.835	Normal	Senior Citizen	4335.551
28	55	female	32.775	2	Non-Smoker	northwest	12268.632	Obese	Senior Citizen	3880.590
29	23	male	17.385	1	Non-Smoker	northwest	2775.192	Underweight	Young Adult	0.000
30	31	male	36.300	2	Smoker	southeast	38711.000	Obese	Middle Aged	0.000
31	22	male	35.600	0	Smoker	southeast	35585.576	Obese	Young Adult	0.000
32	18	female	26.315	0	Non-Smoker	northwest	2198.190	Overweight	Young Adult	0.000
33	19	female	28.600	5	Non-Smoker	southeast	4687.797	Overweight	Young Adult	0.000
34	63	male	28.310	0	Non-Smoker	northwest	13770.098	Overweight	Senior Citizen	4131.029
35	28	male	36.400	1	Smoker	southeast	51194.559	Obese	Young Adult	0.000
36	19	male	20.425	0	Non-Smoker	northwest	1625.434	Normal	Young Adult	0.000
37	62	female	32.965	3	Non-Smoker	northwest	15612.193	Obese	Senior Citizen	4683.658
38	26	male	20.800	0	Non-Smoker	southwest	2302.300	Normal	Young Adult	0.000
39	35	male	36.670	1	Smoker	northwest	39774.276	Obese	Middle Aged	0.000
40	60	male	39.900	0	Smoker	southwest	48173.361	Obese	Senior Citizen	14452.008
41	24	female	26.600	0	Non-Smoker	northwest	3046.062	Overweight	Young Adult	0.000
42	31	female	36.630	2	Non-Smoker	southeast	4949.759	Obese	Middle Aged	0.000
43	41	male	21.780	1	Non-Smoker	southeast	6272.477	Normal	Middle Aged	0.000
44	37	female	30.800	2	Non-Smoker	southeast	6313.759	Obese	Middle Aged	0.000
45	38	male	37.050	1	Non-Smoker	northwest	6079.672	Obese	Middle Aged	0.000
46	55	male	37.300	0	Non-Smoker	southwest	20630.284	Obese	Senior Citizen	6189.085
47	18	female	38.665	2	Non-Smoker	northwest	3393.356	Obese	Young Adult	0.000
48	28	female	34.770	0	Non-Smoker	northwest	3556.922	Obese	Young Adult	0.000
49	60	male	24.530	0	Non-Smoker	southeast	12629.897	Normal	Senior Citizen	3788.969
50	36	male	35.200	1	Smoker	southeast	38709.176	Obese	Middle Aged	0.000
51	18	female	35.600	0	Non-Smoker	northwest	2211.131	Obese	Young Adult	0.000
52	21	female	33.630	2	Non-Smoker	northwest	3579.829	Obese	Young Adult	0.000
53	48	male	28.000	1	Smoker	southeast	23568.272	Overweight	Middle Aged	0.000
54	36	male	34.430	0	Smoker	southeast	37742.576	Obese	Middle Aged	0.000

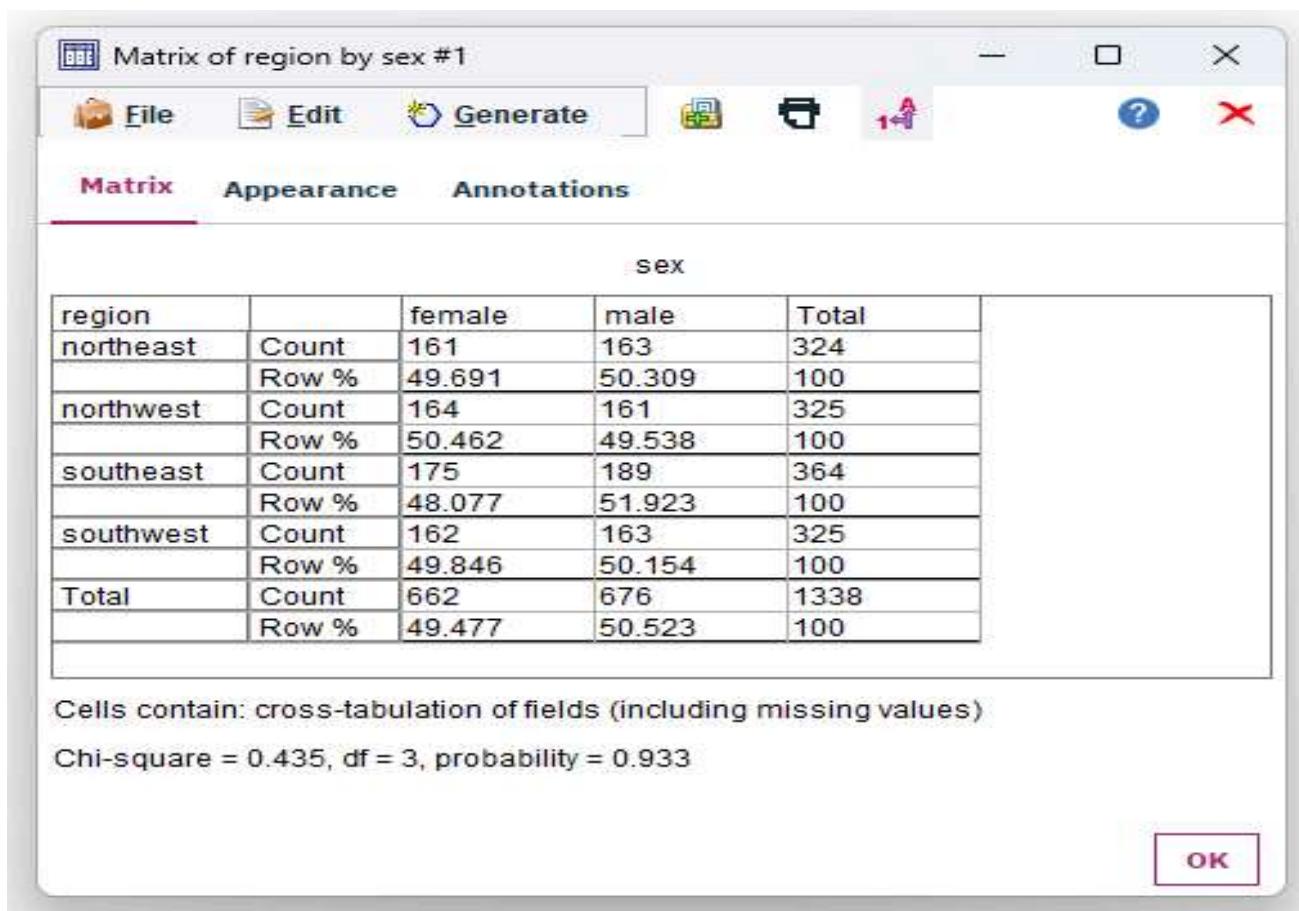
OK

Step 18: Now we find relationships between different types of Measurement values within this dataset. Import the same dataset through Var File node in the Source Category. And Type node to read the values and get measurement of the fields.

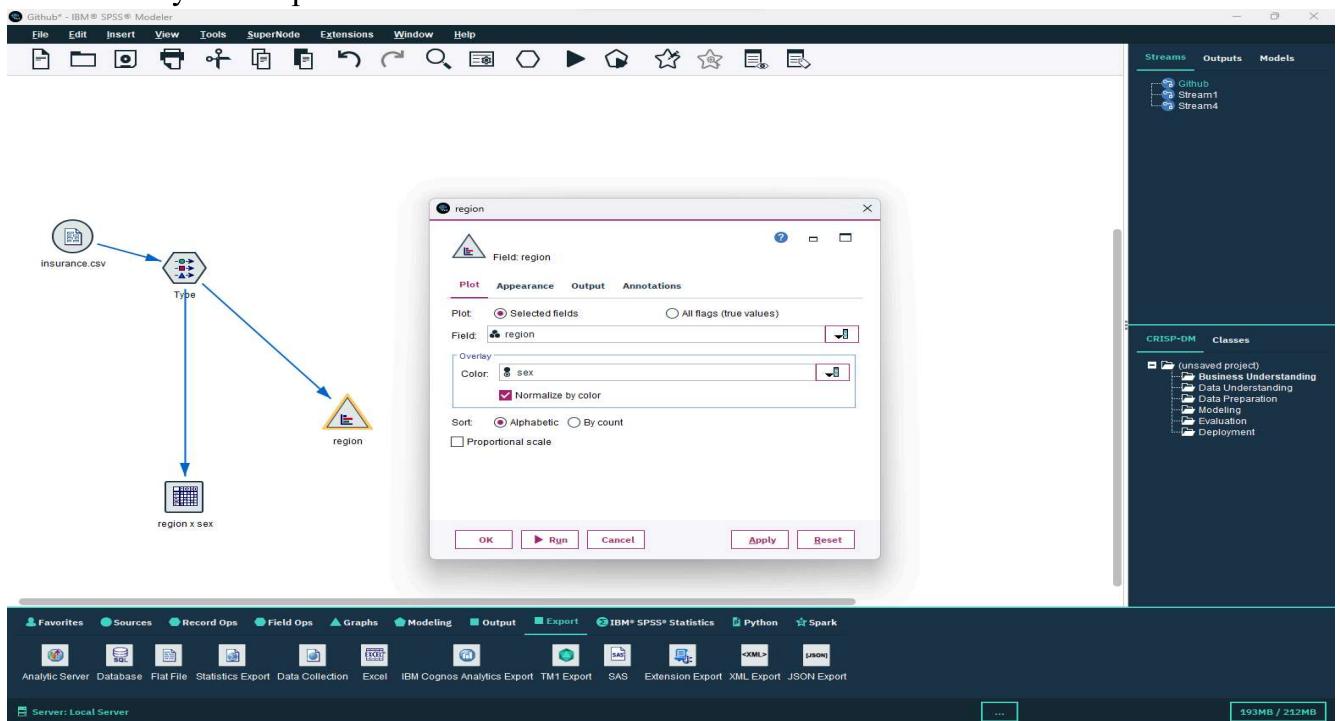


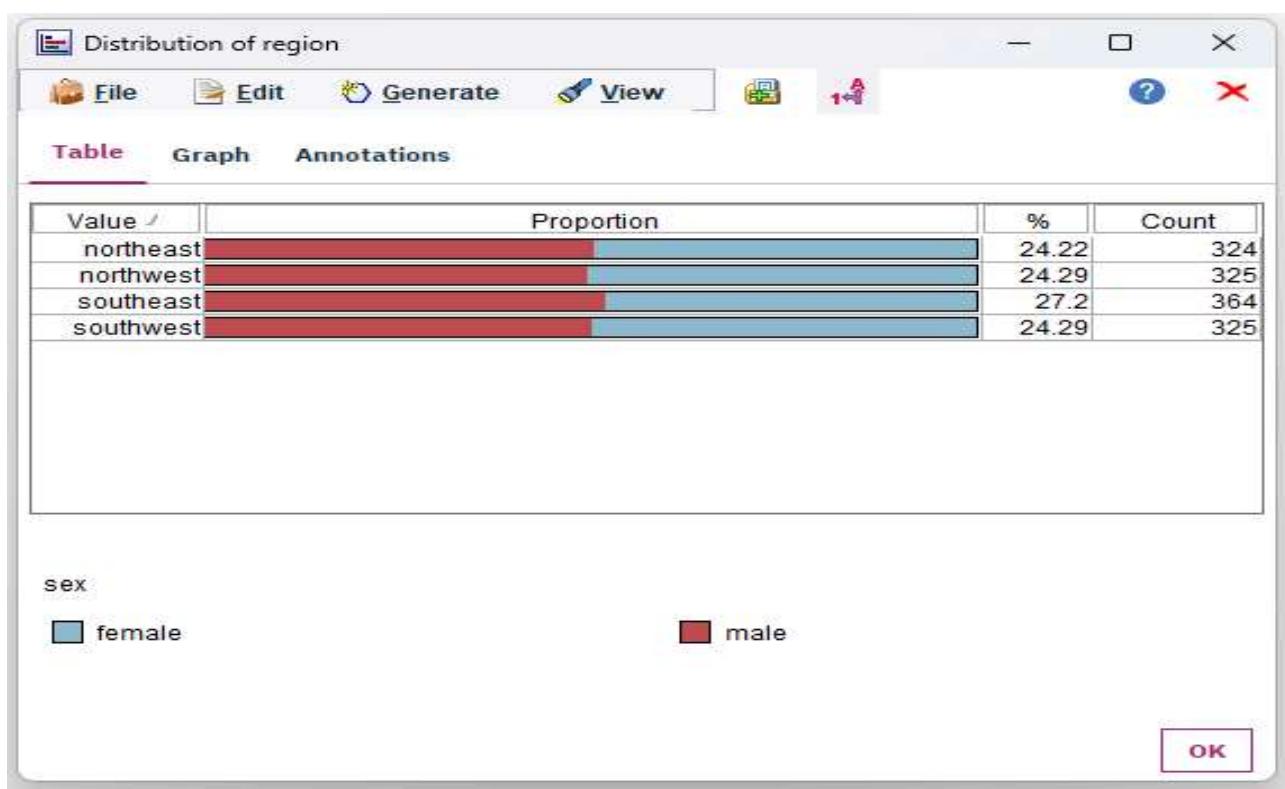


Step 19: We add the matrix node from the Output category to see our output or result in matrix form .When we used to make the relationship between two categorical field then used matrix for tabular output and distribution node from Graph category for graph output.
Take ‘region’ as row and ‘sex’ as column.

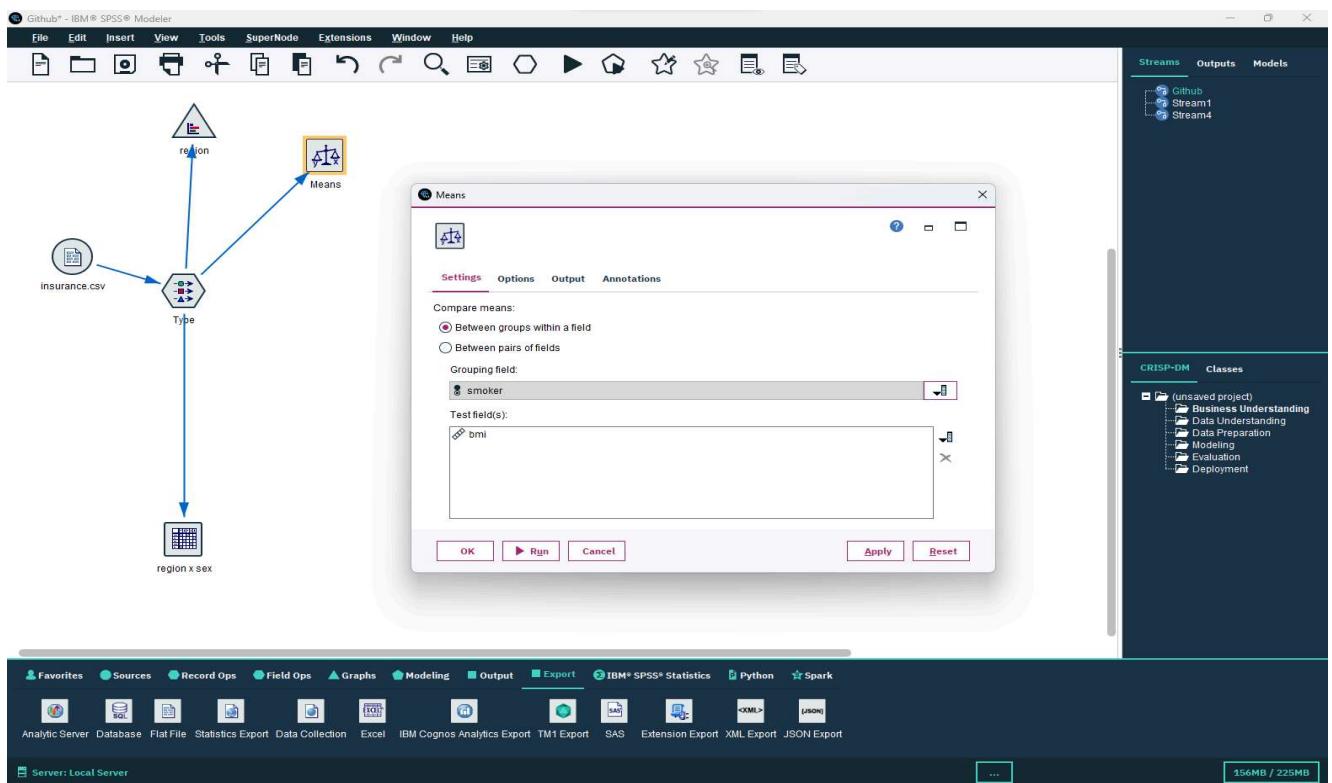


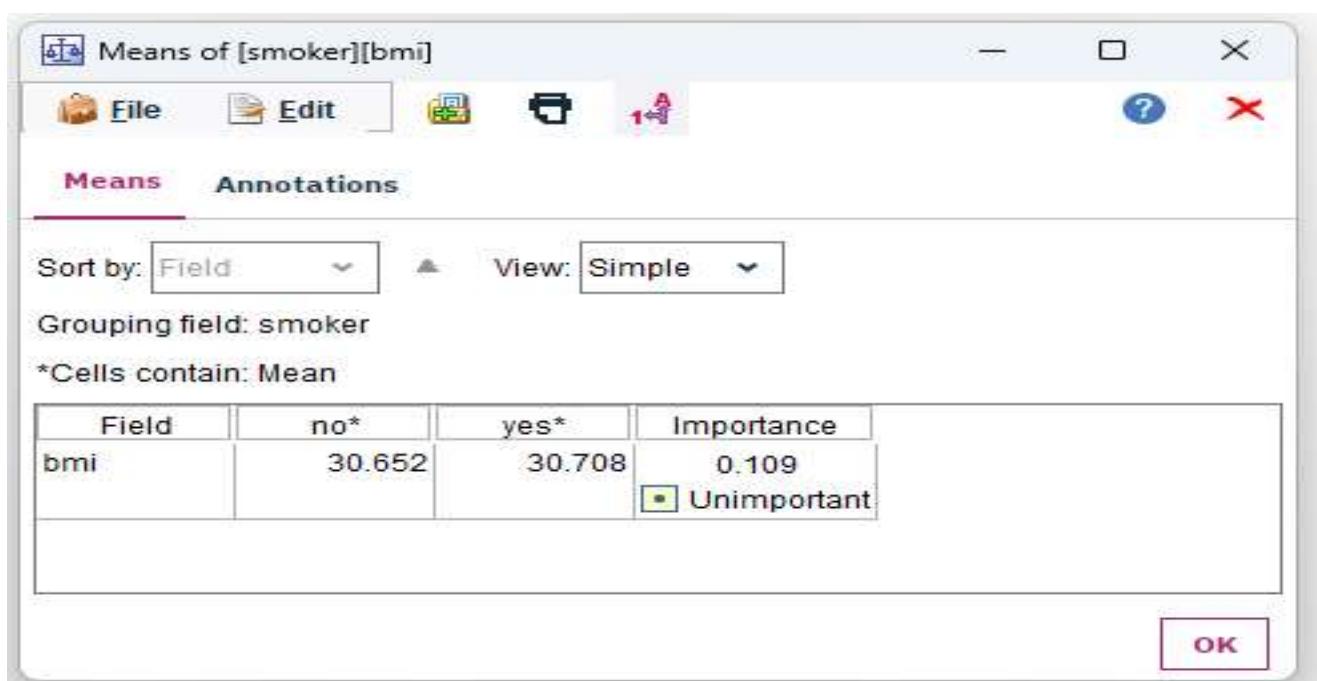
Step 20: For Distribution Graph take ‘region’ as Field and ‘sex’ as color and check the Normalize by color option.



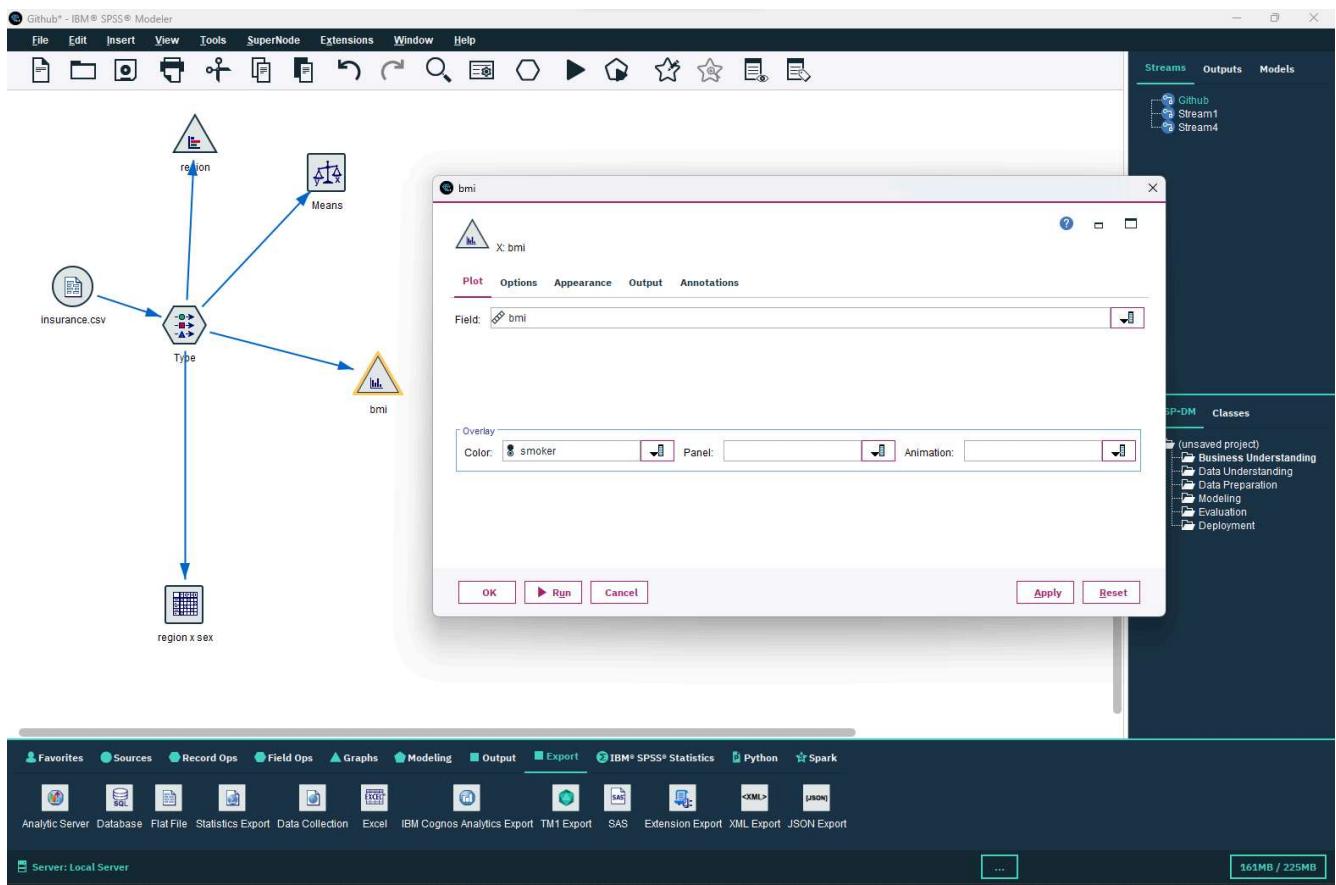


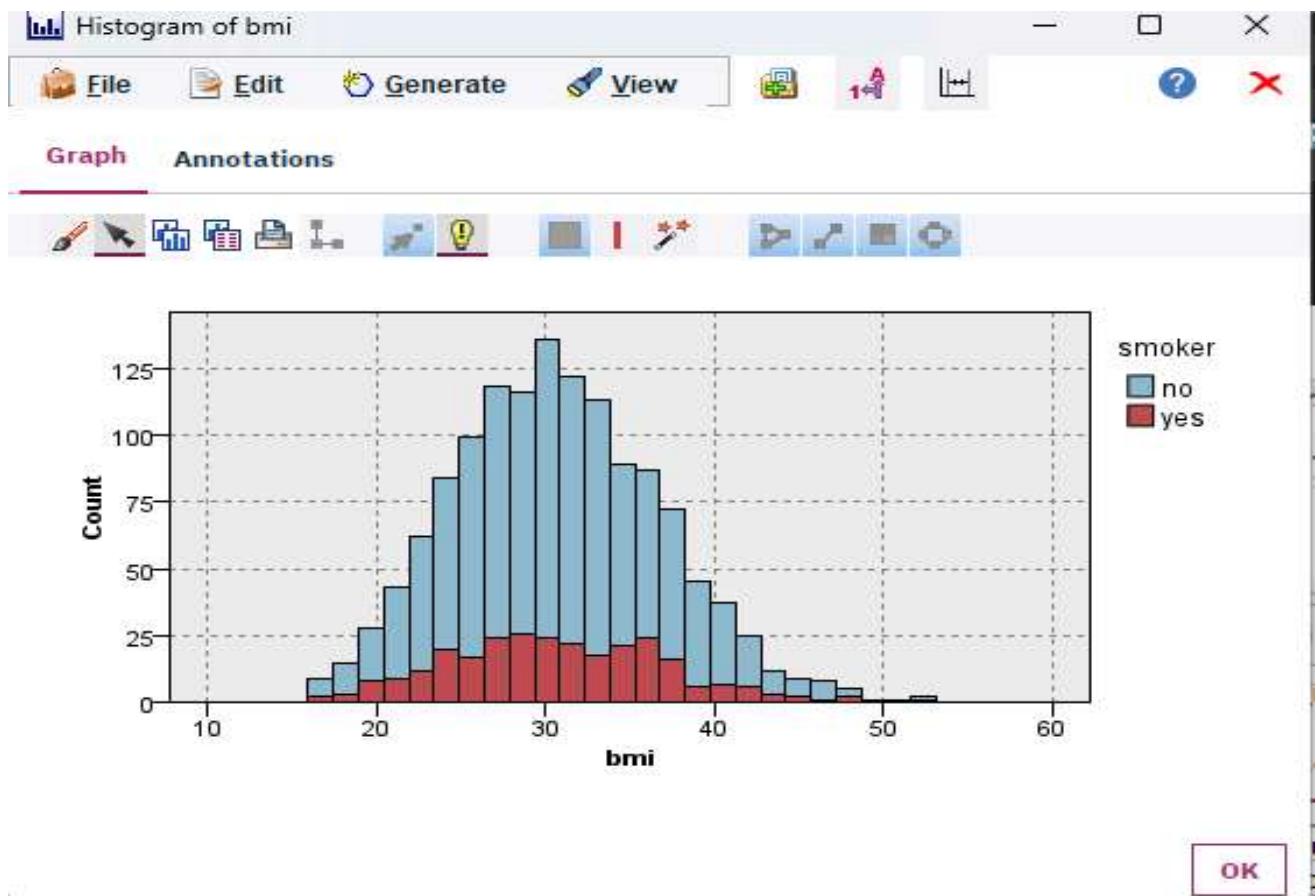
Step 21: We take the mean node from the output category because we want to find the relationship between one categorical and one continuous field. In this node we take the 'smoker' which is a categorical field and 'bmi' which is a continuous field .





Step 22: We connect the histogram graph to the type to see the output in the graph . In this we find the relationship of one categorical and one continuous field. In this we select bmi as field and smoker as color and run it.





Step 23: After this we connect the statistics node from the output category to find the relationship between two continuous fields in the tabular form. In this statistics node we take two continuous fields to find the relationship. In this we select the children field as the examine node and charges field as the correlate. In this we add the multi statistics operation like min, max, mean count, etc.

Github* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

The screenshot shows the IBM SPSS Modeler interface. On the left, a data flow diagram is displayed with nodes: 'insurance.csv' (represented by a document icon), 'Type' (represented by a hexagon icon), 'relation' (represented by a triangle icon), 'Means' (represented by a bar chart icon), 'bmi' (represented by a triangle icon), and 'Statistics' (represented by a summation symbol icon). Arrows indicate connections between these nodes. A context menu is open over the 'Type' node. On the right, a 'Statistics' dialog box is open, showing settings for examining 'children' and correlating with 'charges'. The dialog includes options for Count, Mean, Sum, Min, Max, Range, Variance, Std Dev, Std Error of Mean, Median, and Mode. Below the dialog, there are OK, Run, Cancel, Apply, and Reset buttons. The top right corner shows a sidebar with 'Streams', 'Outputs', and 'Models' tabs, and a list of streams: 'Github', 'Stream1', and 'Stream4'. The bottom left shows a navigation bar with various icons and tabs like Favorites, Sources, Record Ops, Field Ops, Modeling, Output, Export, IBM SPSS Statistics, Python, and Spark. The bottom right shows memory usage: 163MB / 225MB.

Statistics of [children][charges] #1

File Edit Generate

Statistics Annotations

children

Statistics

Count	1338
Mean	1.095
Min	0
Max	5
Range	5
Variance	1.453
Standard Deviation	1.205
Standard Error of Mean	0.033

Pearson Correlations

	charges
children	0.068/Strong

OK

This screenshot shows the results of a statistical analysis for the 'children' variable against 'charges'. It displays basic statistics like count, mean, and standard deviation, along with a Pearson correlation coefficient of 0.068, described as 'Strong'.

Step 24: We connect plot graph node from Graph Category to see the relationship between two continuous fields. We take the charges field as X field and bmi field as y field and both are continuous fields and then run to see the output.

