

SCHOOL OF COMPUTER APPLICATION



BBD UNIVERSITY

Case Study On

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.

Submitted To :

Mr Ayushman Bhadauriya

Submitted By

Sharandhar pandey

Babu Banarasi Das University, Lucknow

BBD city, Faizabad road, Lucknow Uttarpradesh- 226 028 India

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:

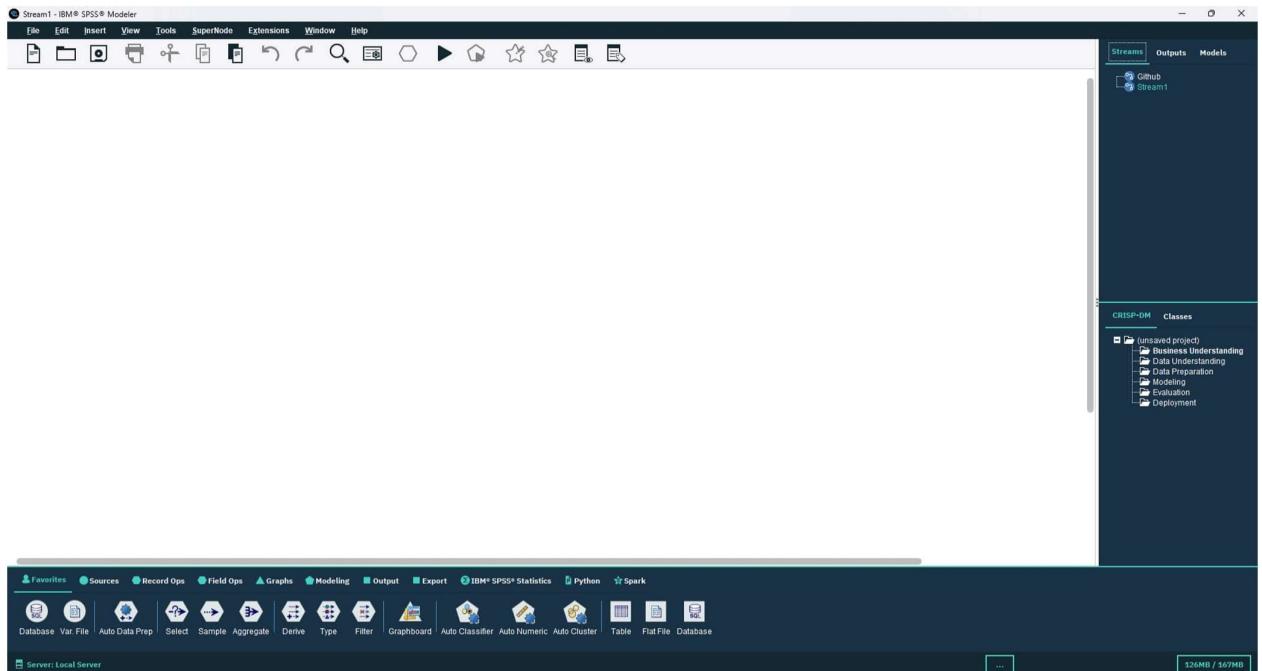
- Using the Derive node to create new fields (e.g., BMI Category, Age Group) from existing data.
- Using the Reclassify node to simplify or merge categorical values
- Identifying relationships between numeric and categorical variables

Required Tool: IBM SPSS Modeler 18.6

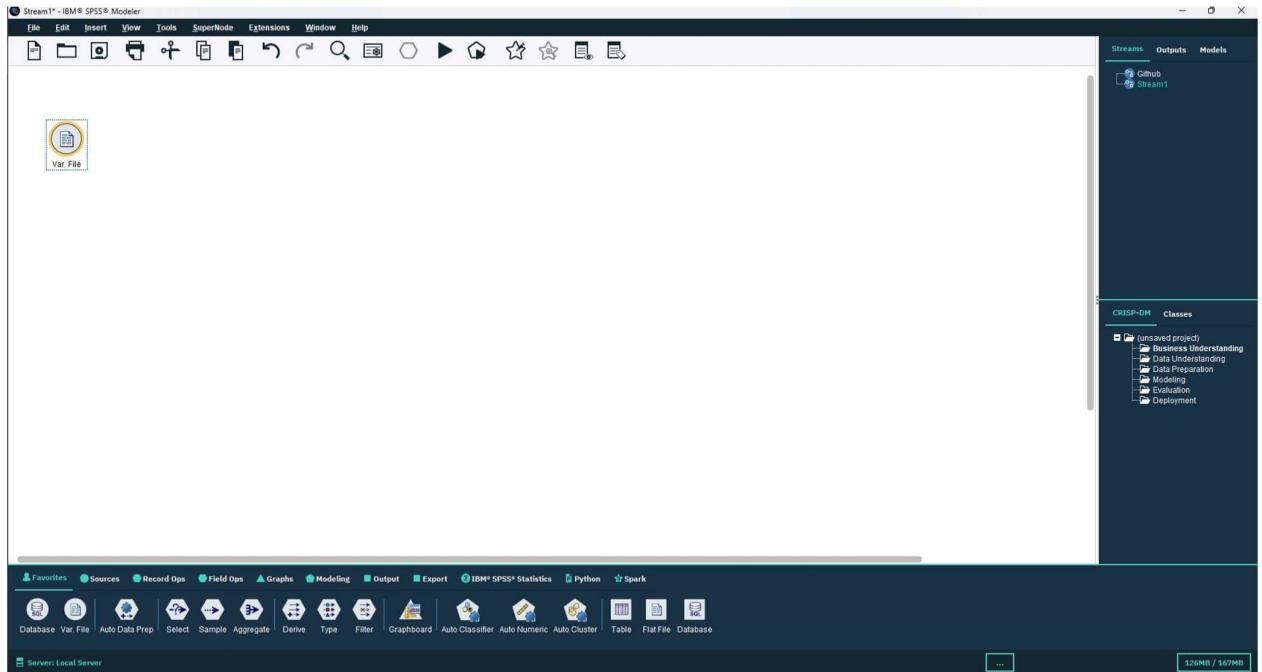
Working: This project focuses on the predictive analysis of medical insurance charges using the Insurance dataset in IBM SPSS Modeler. The aim is to understand how various demographic and lifestyle factors — such as age, BMI, smoking habits, and region — influence the cost of medical insurance.

Dataset Used: insurance.csv

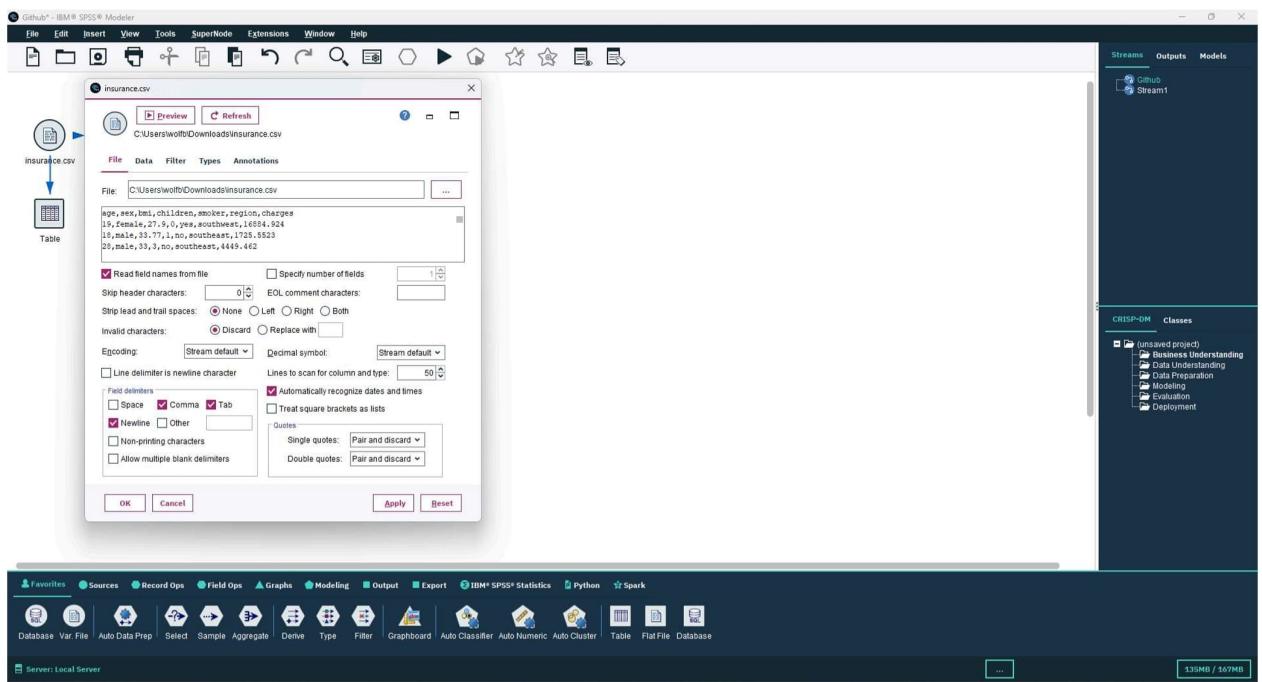
Step 1: Open IBM SPSS Modeler 18.6 on your pc



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)

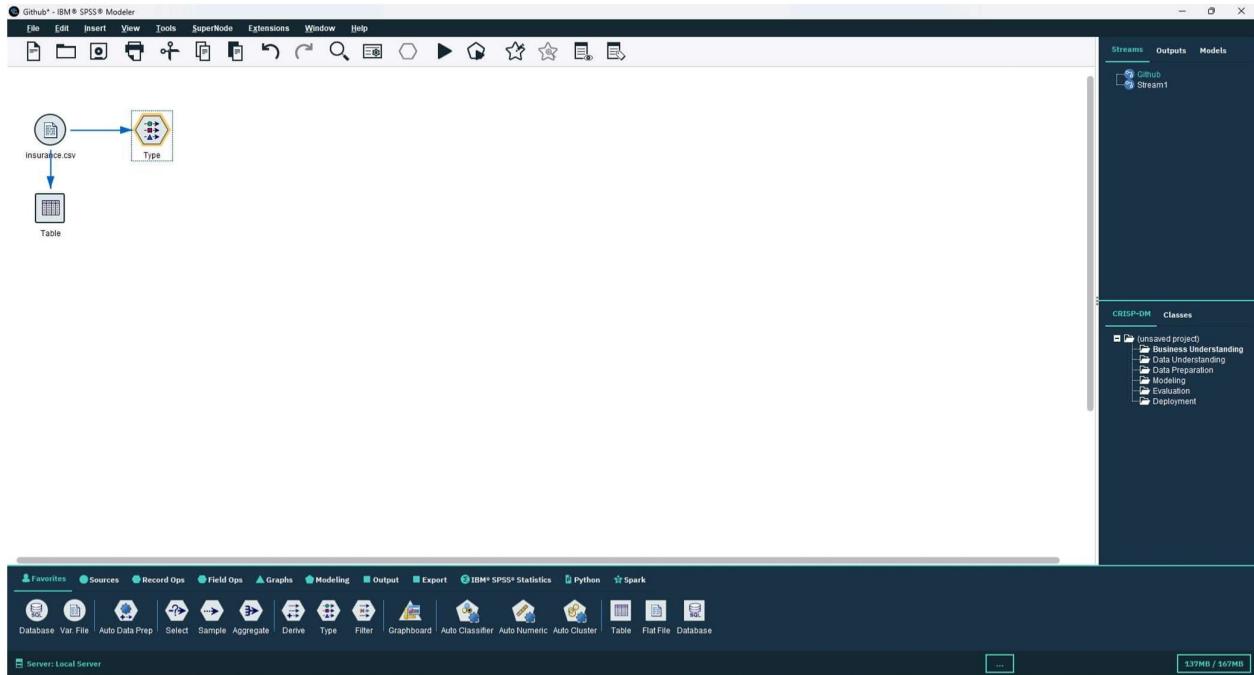
File Edit Generate

Table Annotations

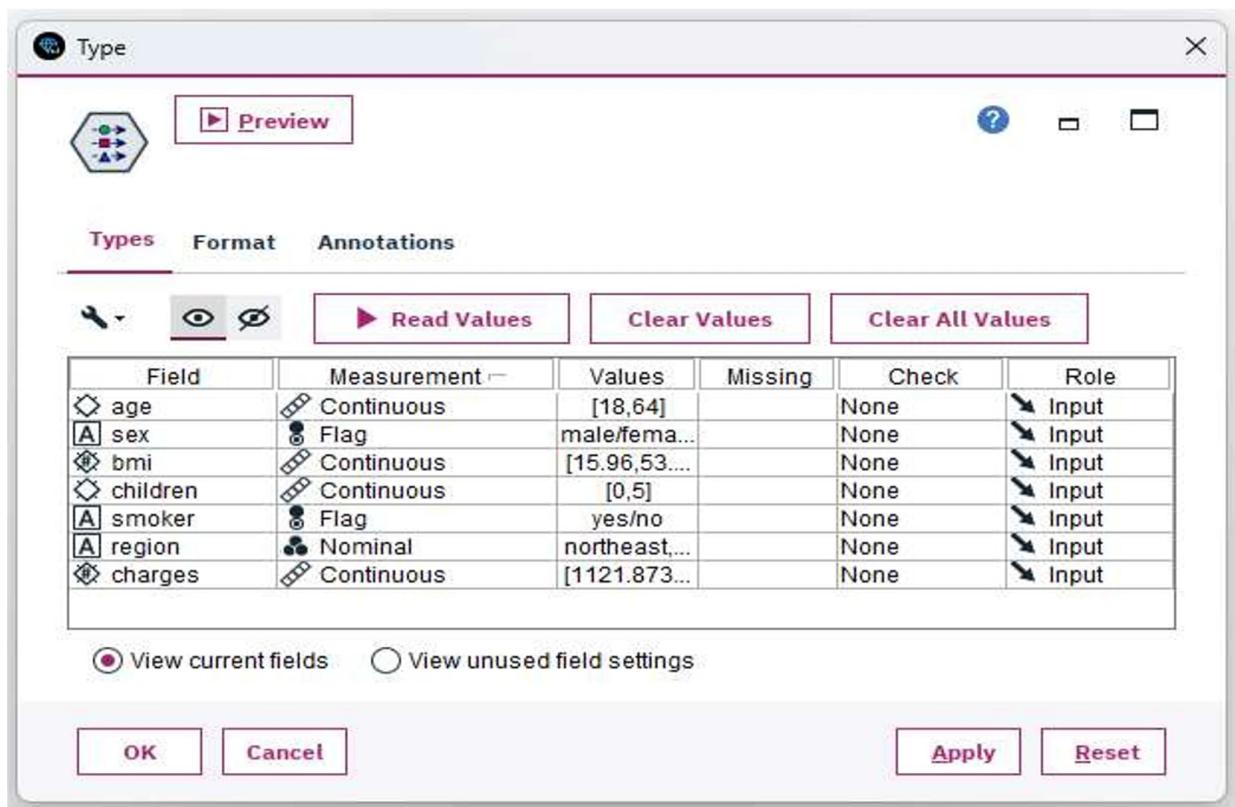
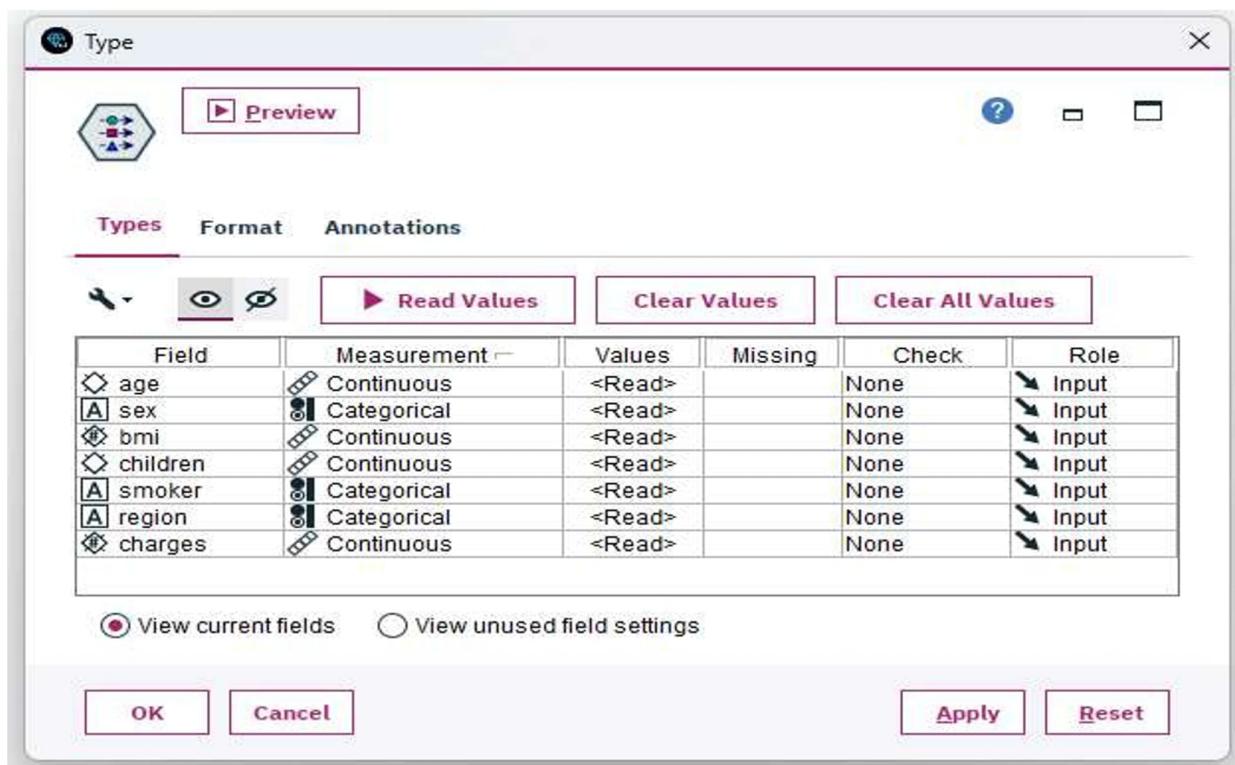
	age	sex	bmi	children	smoker	region	charges
1	19	female	27.800	0	yes	southeast	16984.564
2	19	male	27.770	1	no	southeast	12350.552
3	20	male	33.000	3	no	southeast	4449.460
4	33	male	22.705	0	no	northwest	21964.471
5	35	male	28.880	0	no	northwest	3864.855
6	31	male	28.740	0	no	southeast	3710.265
7	44	female	33.440	1	no	southeast	8240.590
8	37	female	27.740	3	no	northwest	7281.504
9	37	male	29.530	2	no	northwest	1611.461
10	60	female	31.540	0	no	southeast	28923.137
11	25	male	26.220	0	no	northwest	2721.321
12	62	female	26.290	0	yes	southeast	27608.726
13	35	male	33.000	0	no	southeast	4819.478
14	54	female	35.420	0	no	northwest	11050.710
15	27	male	42.130	0	yes	southeast	39611.758
16	19	male	24.400	1	no	southeast	1837.237
17	53	female	29.100	1	no	northwest	10700.865
18	23	male	23.945	0	no	northwest	2395.172
19	56	male	40.300	0	no	southeast	10602.385
20	30	male	35.300	0	yes	southeast	36337.447
21	60	male	30.000	0	no	northwest	13700.177
22	30	female	32.400	1	no	southeast	4149.736
23	19	male	34.100	0	no	southeast	1137.011
24	34	female	31.920	1	yes	northwest	37701.877
25	37	male	28.120	2	no	southeast	6900.565
26	59	female	27.720	3	no	southeast	14001.134
27	63	female	23.095	0	no	northwest	14451.836
28	55	female	32.775	2	no	northwest	12265.632
29	33	male	27.155	1	no	northwest	2710.182
30	31	male	36.300	2	yes	southeast	38711.000
31	22	male	35.600	0	yes	southeast	35585.576
32	18	female	26.215	0	no	northwest	2198.190
33	19	male	30.100	5	no	southeast	4410.177
34	43	male	28.310	0	no	northwest	13770.098
35	20	male	36.400	1	yes	southeast	51194.559
36	19	male	28.425	0	no	northwest	12416.144
37	6	female	32.945	1	no	northwest	15412.193
38	24	male	20.800	0	no	southeast	2302.300
39	35	male	36.470	1	yes	northwest	39774.276
40	40	male	36.400	0	yes	northwest	4811.141
41	24	female	26.400	0	no	northwest	30446.043
42	31	female	36.430	2	no	southeast	49495.759
43	41	male	21.780	1	no	southeast	6372.477
44	37	male	35.600	2	no	southeast	6931.795
45	38	male	37.080	1	no	northwest	6079.476
46	55	male	37.300	0	no	southeast	20630.284
47	18	female	38.465	2	no	northwest	3393.356
48	21	male	28.770	0	no	northwest	2872.522
49	60	female	24.530	0	no	southeast	12429.097
50	36	male	35.200	1	yes	southeast	38709.176
51	18	female	35.425	0	no	northwest	2211.131
52	31	male	28.120	2	no	southeast	3710.265
53	45	male	28.000	1	yes	southeast	23365.272
54	36	male	34.430	0	yes	southeast	37742.576

OK

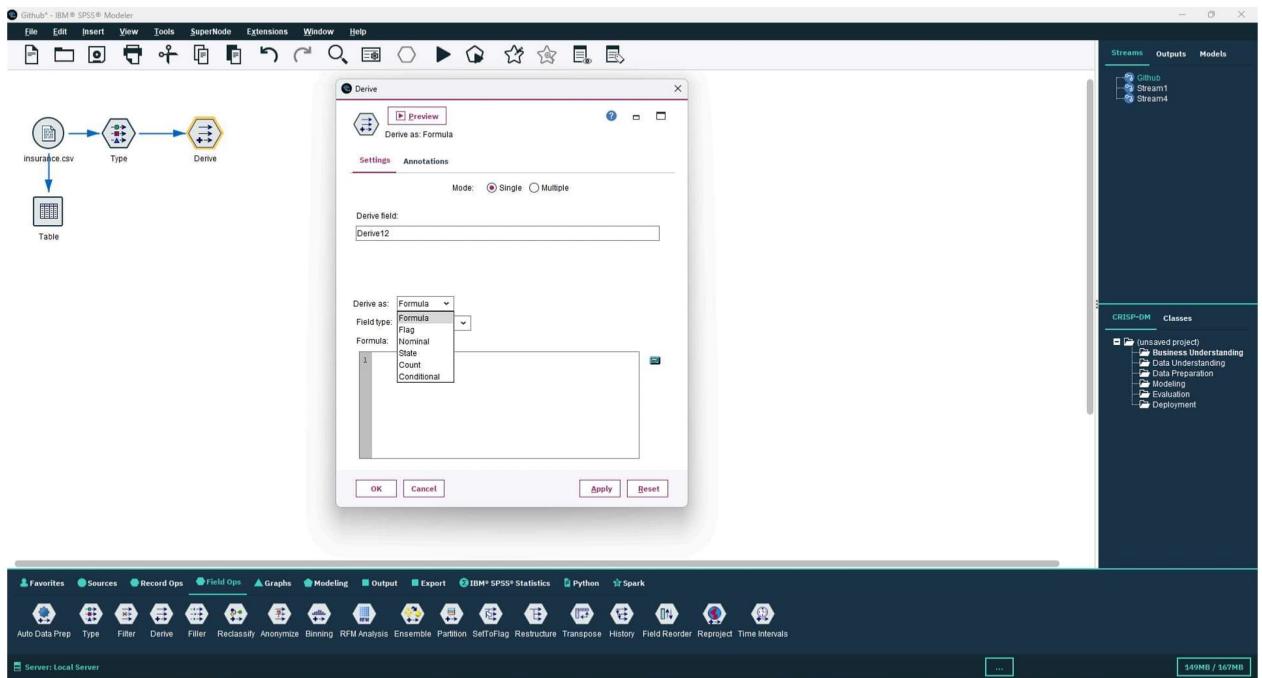
Step 5: Connect the Type Node from Field Ops. This Node aids in understanding the different Fields Measurement Values.



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.

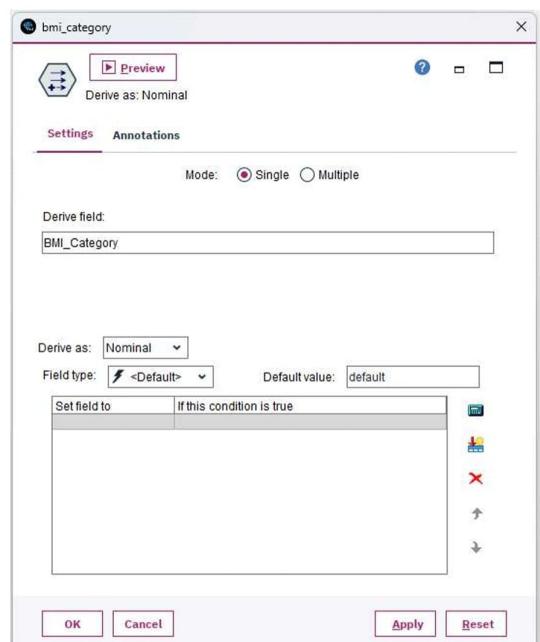
The condition is as follows:

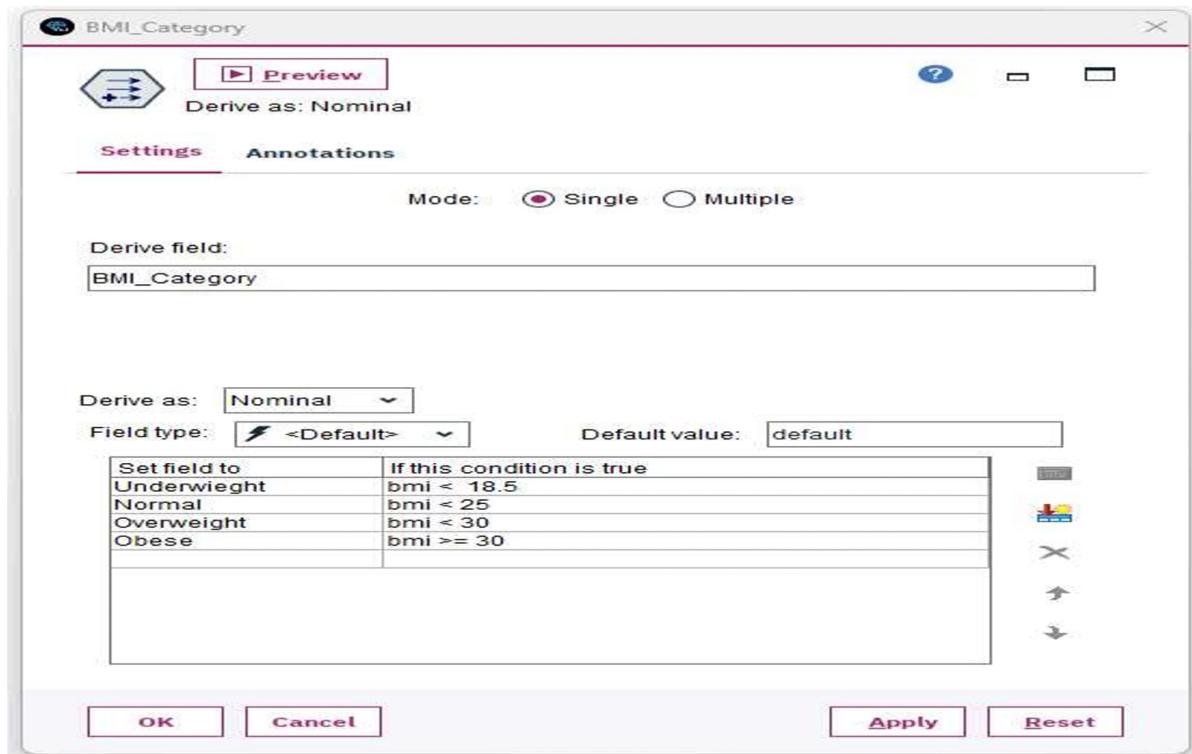
If $(\text{bmi} < 18.5)$ Then the value will be 'Underweight'.

If $(\text{bmi} < 25)$ Then the value will be 'Normal'.

If $(\text{bmi} < 30)$ Then 'Overweight'.

And if $(\text{bmi} \geq 30)$ 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.

Table (8 fields, 1,338 records) #3								
	age	sex	bmi	children	smoker	region	charge	BMI_Category
1	19	female	27.900	0	yes	southeast	16834.463	Overweight
2	18	male	27.700	1	no	southeast	12000.558	Obese
3	20	female	33.000	3	no	southeast	4449.463	Obese
4	33	male	22.705	0	no	northwest	21584.473	Normal
5	32	male	28.880	0	no	northwest	3866.855	Overweight
6	31	female	23.740	0	no	southeast	3784.622	Overweight
7	44	male	24.140	1	no	southeast	8280.443	Normal
8	37	female	27.740	3	no	northwest	7281.504	Overweight
9	37	male	29.830	2	no	northwest	6406.411	Overweight
10	60	female	25.940	0	no	northwest	28923.137	Overweight
11	25	male	26.220	0	no	northwest	2721.321	Overweight
12	52	female	24.150	0	no	southeast	2700.443	Overweight
13	23	male	34.400	0	no	southwest	1826.840	Obese
14	56	female	39.820	0	no	southeast	11090.710	Obese
15	27	male	42.130	0	yes	southeast	39611.795	Obese
16	19	female	24.200	1	no	southeast	1493.943	Normal
17	52	female	30.780	1	no	northwest	10797.330	Obese
18	23	male	23.845	0	no	northwest	2395.170	Normal
19	56	male	40.800	0	no	southwest	10602.385	Obese
20	30	male	35.300	0	yes	southeast	36337.497	Obese
21	61	female	24.040	0	no	northwest	13000.443	Overweight
22	30	female	32.400	1	no	southwest	4149.736	Obese
23	19	male	34.100	0	no	southeast	1137.011	Obese
24	34	female	31.920	1	yes	northwest	37701.877	Obese
25	37	male	28.025	2	no	northwest	6203.962	Overweight
26	51	female	30.120	0	no	northwest	14487.890	Overweight
27	63	female	23.085	0	no	northwest	14451.631	Normal
28	55	female	32.775	2	no	northwest	12268.632	Obese
29	23	male	17.385	1	no	northwest	2775.192	Underweight
30	31	male	24.000	2	yes	southeast	31300.571	Obese
31	23	female	35.600	0	yes	southeast	35855.571	Obese
32	10	female	26.315	0	no	northwest	2198.190	Overweight
33	19	female	28.600	5	no	southeast	4687.797	Overweight
34	63	male	28.310	0	no	northwest	13770.090	Overweight
35	36	female	24.100	1	yes	southeast	31300.571	Obese
36	19	male	20.425	0	no	northwest	1625.434	Normal
37	62	female	32.945	3	no	northwest	15612.190	Obese
38	26	male	20.800	0	no	southwest	2302.300	Normal
39	35	male	34.470	1	yes	northwest	39774.276	Obese
40	60	female	24.400	0	no	northwest	47100.000	Obese
41	24	female	26.400	0	no	northwest	30446.043	Overweight
42	31	female	36.630	2	no	southeast	4949.750	Obese
43	41	male	21.780	1	no	southeast	6272.477	Normal
44	37	female	30.600	2	no	southeast	6313.750	Obese
45	39	female	24.050	1	yes	southeast	6000.000	Obese
46	55	male	37.300	0	no	southeast	20630.284	Obese
47	18	female	38.665	2	no	northwest	3393.354	Obese
48	28	female	34.770	0	no	northwest	3556.925	Obese
49	40	male	24.530	0	no	southeast	12800.000	Normal
50	34	female	24.100	1	yes	southeast	38709.170	Obese
51	18	female	35.625	0	no	northwest	2211.131	Obese
52	21	female	33.630	2	no	northwest	3579.629	Obese
53	49	male	28.000	1	yes	southeast	23560.270	Overweight
54	36	male	34.430	0	yes	southeast	37742.576	Obese

OK

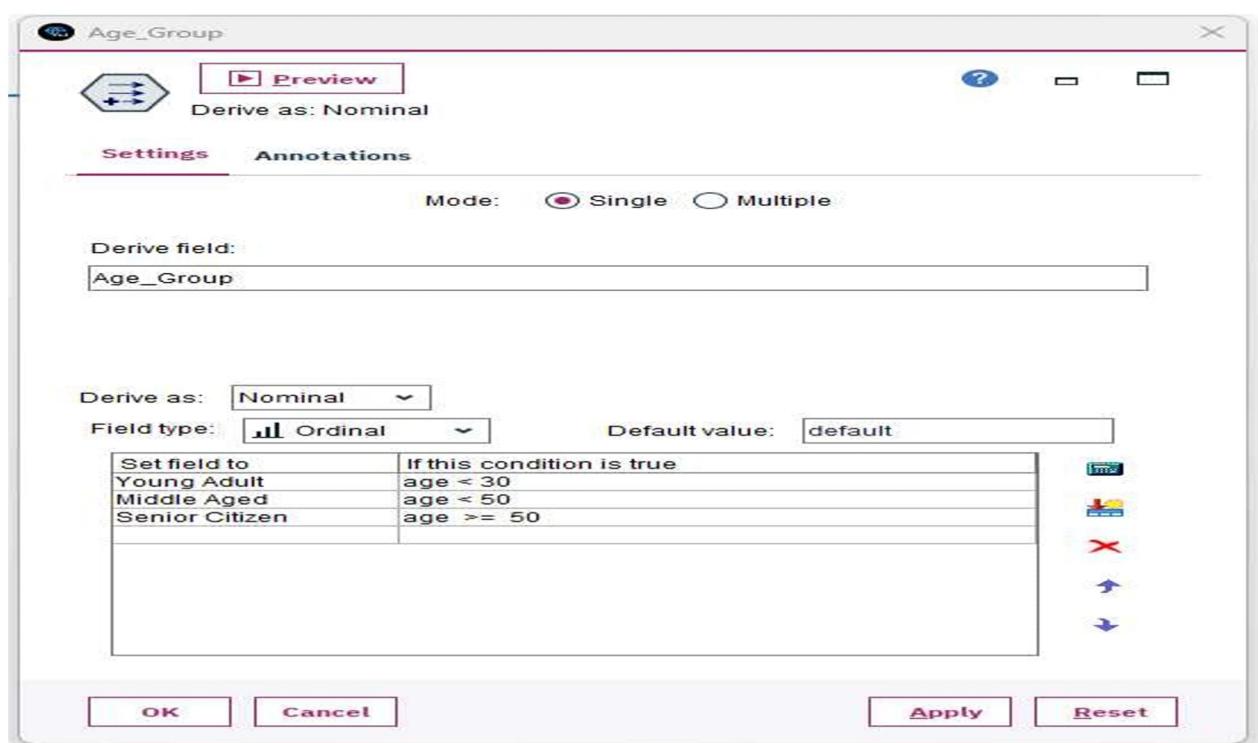
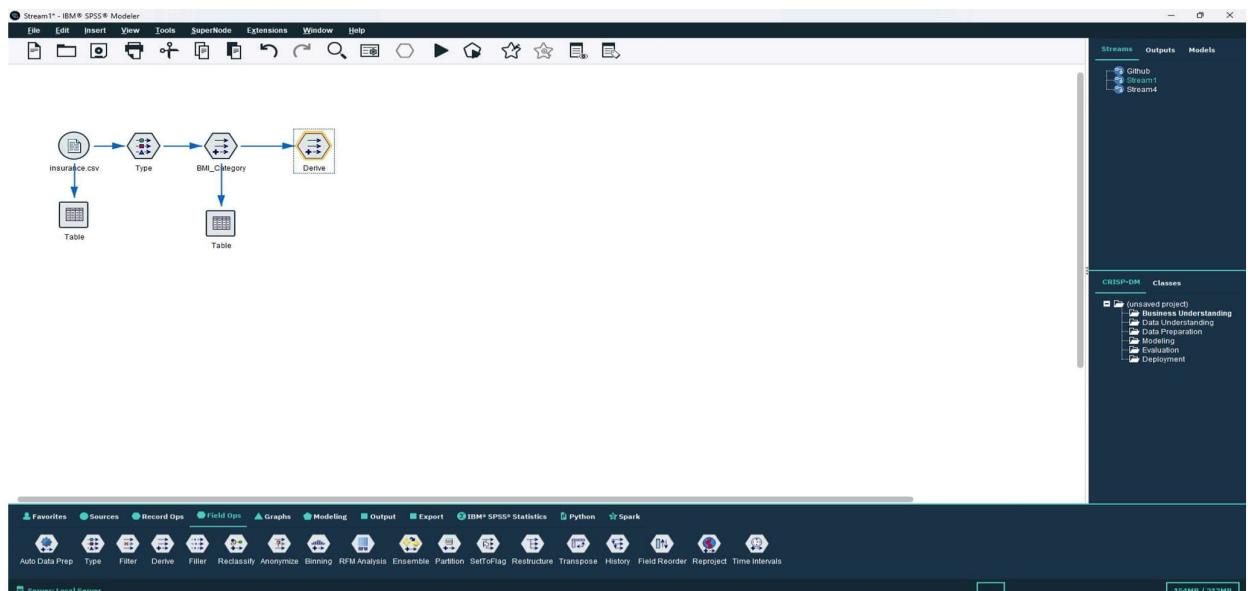
Step 10: Connect a new Derive Node and name it as “Age_Group”. We are dividing our customers by their age. As done in above Step 8 we select “Nominal” in the “Derive as” section and let the other options set as Default.

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 \geq 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.

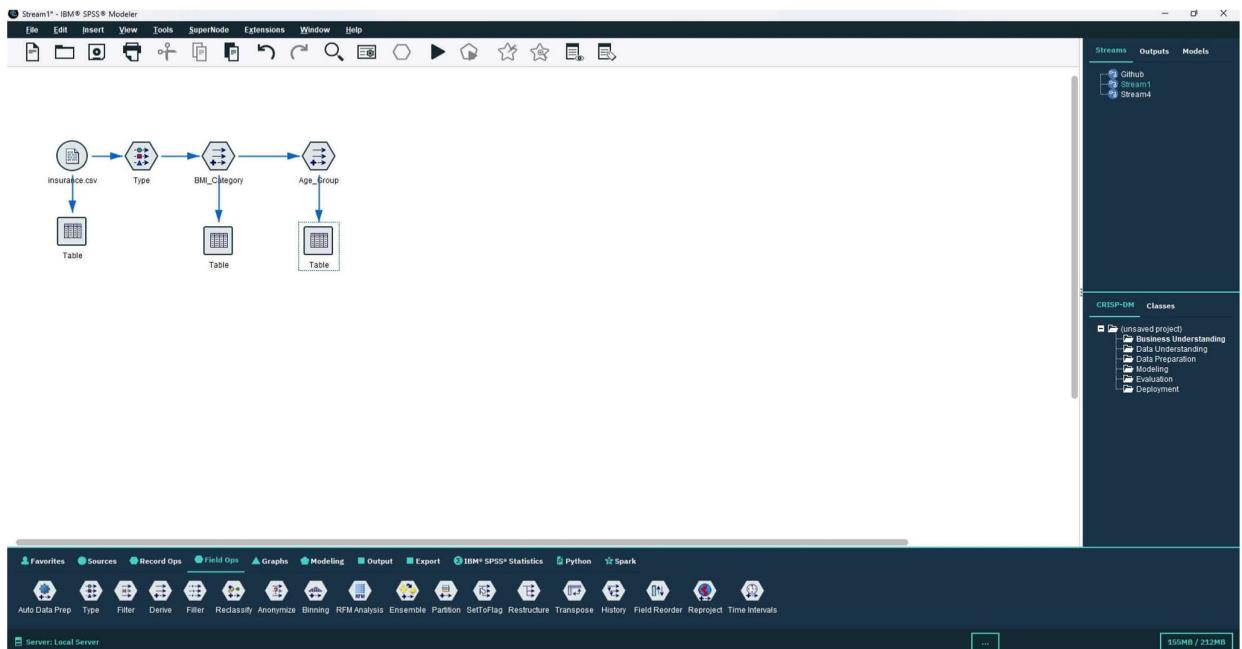


Table (9 fields, 1,338 records) #2

File Edit Generate

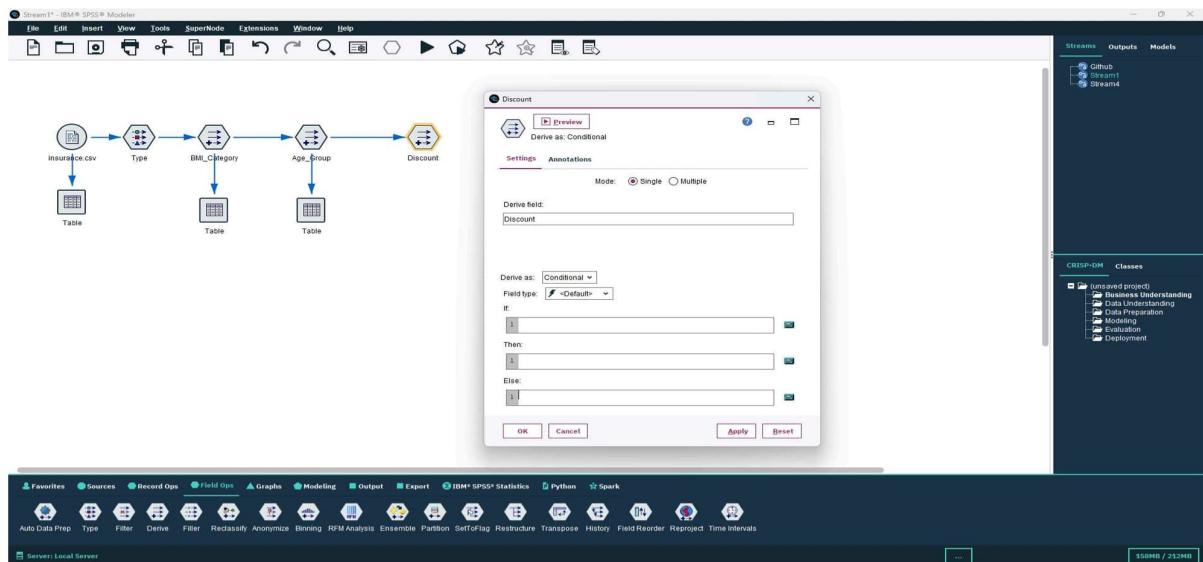
Table Annotations

age	sex	bmi	children	smoker	region	charges	BMI_CCategory	Age_Group
1	18 female	27.900	0 yes	southwest	16884.92 Overweight	Young Adult		
2	18 male	33.770	1 no	southeast	1725.55 Obese	Young Adult		
3	28 male	33.000	3 no	southeast	4449.46 Obese	Young Adult		
4	33 male	22.705	0 no	northwest	21894.47 Normal	Middle Aged		
5	32 male	28.880	0 no	northwest	3866.85 Overweight	Middle Aged		
6	31 female	25.740	0 no	southeast	3756.62 Normal	Middle Aged		
7	46 female	33.440	1 no	southeast	8240.59 Obese	Middle Aged		
8	37 female	27.740	3 no	northwest	7281.50 Overweight	Middle Aged		
9	37 male	29.830	2 no	northeast	6406.41 Overweight	Middle Aged		
10	60 female	25.840	0 no	northwest	28923.13 Overweight	Senior Citizen		
11	25 male	26.220	0 no	northeast	2721.32 Overweight	Young Adult		
12	62 female	26.290	0 yes	southeast	27808.72 Overweight	Senior Citizen		
13	23 male	34.400	0 no	southwest	1826.84 Obese	Young Adult		
14	56 female	35.820	0 no	southeast	11090.71 Obese	Senior Citizen		
15	27 male	42.130	0 yes	southeast	36411.75 Obese	Young Adult		
16	19 male	24.600	1 no	southwest	1837.23 Normal	Young Adult		
17	52 female	30.780	1 no	northeast	10797.33 Obese	Senior Citizen		
18	23 male	23.845	0 no	northeast	2395.17 Normal	Young Adult		
19	56 female	40.300	0 no	southwest	10602.38 Obese	Senior Citizen		
20	30 male	35.300	0 yes	southwest	36837.46 Obese	Middle Aged		
21	60 female	36.005	0 no	northeast	13228.84 Obese	Senior Citizen		
22	30 female	32.400	1 no	southwest	4149.73 Obese	Middle Aged		
23	18 male	34.100	0 no	southeast	1137.01 Obese	Young Adult		
24	34 female	31.920	1 yes	northwest	37701.87 Obese	Middle Aged		
25	37 male	28.025	2 no	northeast	6203.90 Overweight	Middle Aged		
26	59 female	27.720	3 no	southeast	14040.10 Overweight	Senior Citizen		
27	63 female	23.685	0 no	northeast	14451.80 Normal	Senior Citizen		
28	55 female	37.775	2 no	northeast	12285.63 Overweight	Senior Citizen		
29	23 male	17.155	1 no	northeast	2775.19 Underweight	Young Adult		
30	31 female	34.300	2 yes	southeast	39711.00 Obese	Middle Aged		
31	20 male	35.600	0 yes	southwest	36595.57 Obese	Young Adult		
32	18 female	26.315	0 no	northeast	2199.19 Overweight	Young Adult		
33	19 male	28.600	5 no	southeast	4657.79 Overweight	Young Adult		
34	63 male	28.310	0 no	northeast	13770.09 Overweight	Senior Citizen		
35	28 male	36.400	1 yes	southeast	51194.55 Obese	Young Adult		
36	19 male	20.425	0 no	northeast	1625.43 Normal	Young Adult		
37	62 female	32.945	3 no	northeast	15612.19 Obese	Senior Citizen		
38	24 male	20.800	0 no	southeast	2302.30 Normal	Young Adult		
39	35 male	36.470	1 yes	northeast	39774.27 Obese	Middle Aged		
40	60 male	39.900	0 yes	southeast	48173.36 Obese	Senior Citizen		
41	24 female	26.400	0 no	northeast	3046.06 Overweight	Young Adult		
42	31 female	36.430	2 no	southeast	4949.75 Obese	Middle Aged		
43	41 male	21.780	1 no	southeast	6372.47 Normal	Middle Aged		
44	37 female	30.800	2 no	southeast	6313.75 Obese	Middle Aged		
45	38 male	37.050	1 no	northeast	6079.67 Obese	Middle Aged		
46	55 male	37.300	0 no	southwest	20630.28 Obese	Senior Citizen		
47	18 female	38.665	2 no	northeast	3393.35 Obese	Young Adult		
48	28 female	34.770	0 no	northeast	3556.92 Obese	Young Adult		
49	60 female	24.530	0 no	southeast	12629.89 Normal	Senior Citizen		
50	36 male	35.200	1 yes	southeast	38709.17 Obese	Middle Aged		
51	18 female	35.625	0 no	northeast	2211.13 Obese	Young Adult		
52	21 female	33.630	2 no	northeast	3579.82 Obese	Young Adult		
53	49 male	28.000	1 yes	southeast	23568.27 Overweight	Middle Aged		
54	36 male	34.430	0 yes	southeast	37742.57 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

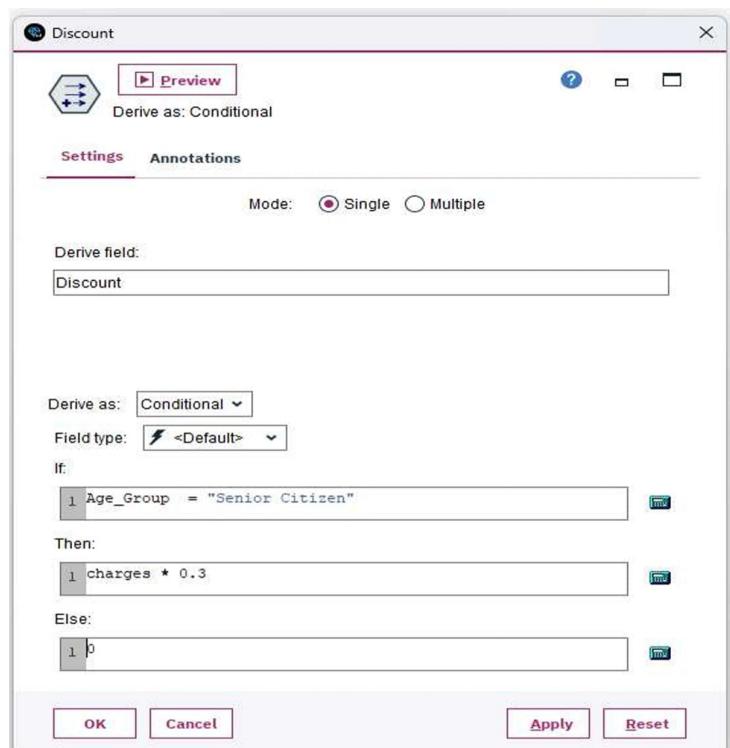
Age_Group = “Senior Citizen”

Then

Charges * 0.3

Else

0



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

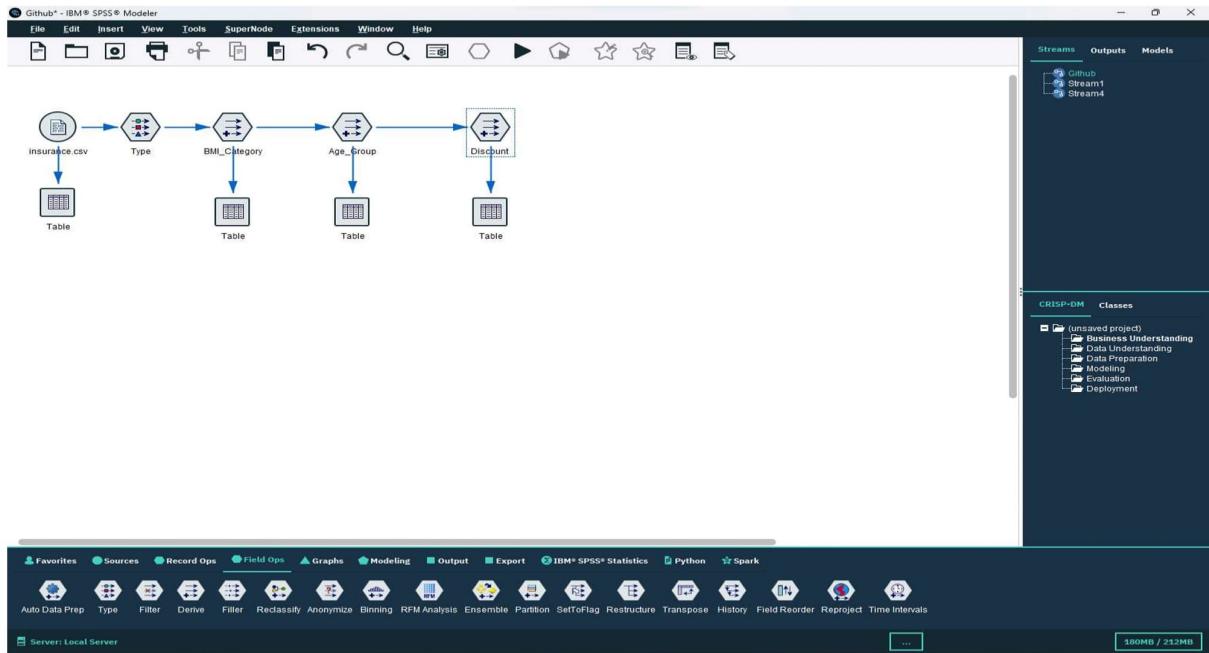


Table (10 fields, 1,338 records) #2

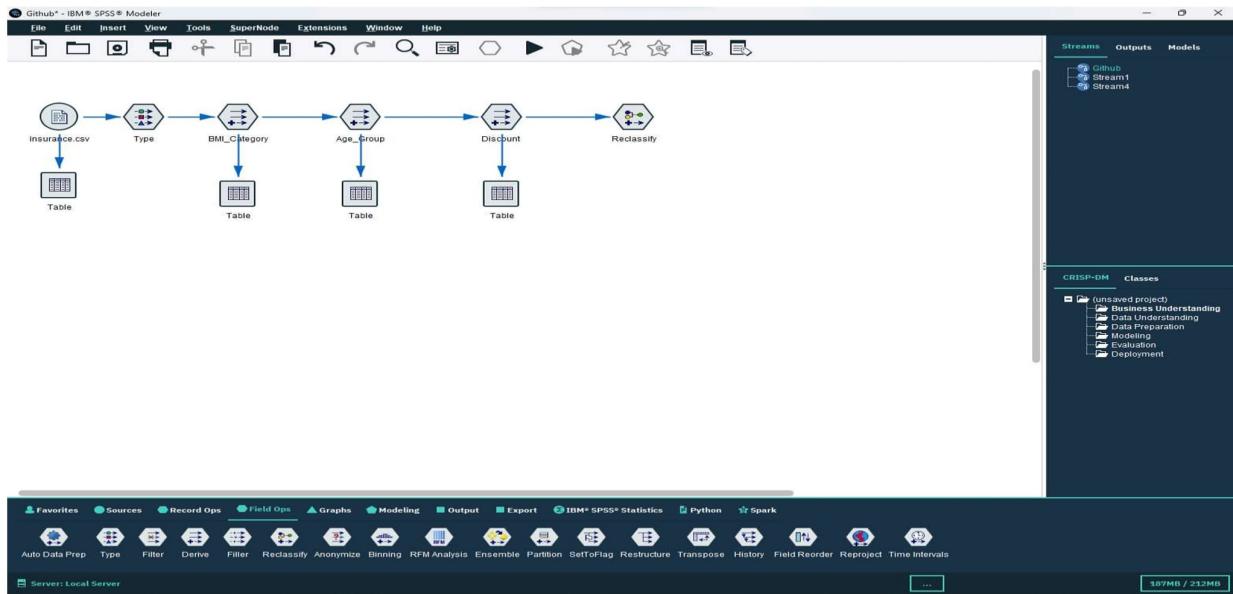
File Edit Generate

Table Annotations

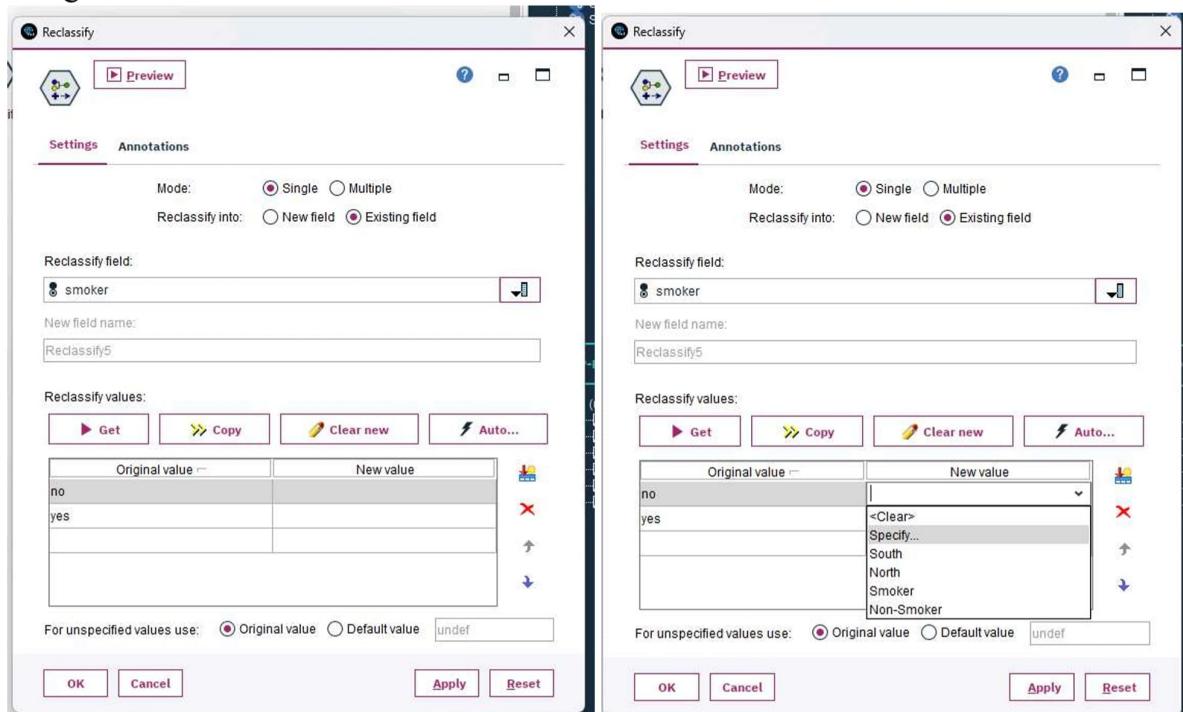
	age	sex	bmi	children	smoker	region	charges	BMI Category	Age Group	Discount
1	19	female	27.900	0	yes	southwest	16894.924	Overweight	Young Adult	0.000
2	18	male	33.770	1	no	southeast	1725.552	Obese	Young Adult	0.000
3	28	male	33.000	3	no	southeast	4449.462	Obese	Young Adult	0.000
4	33	male	22.705	0	no	northwest	21984.471	Normal	Middle Aged	0.000
5	32	male	28.880	0	no	northwest	3866.855	Overweight	Middle Aged	0.000
6	31	female	25.740	0	no	southeast	3756.622	Overweight	Middle Aged	0.000
7	46	male	31.000	1	no	southeast	2420.100	Normal	Young Adult	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	4406.411	Overweight	Middle Aged	0.000
10	60	female	25.840	0	no	northwest	29523.137	Overweight	Senior Citizen	8676.941
11	25	male	26.220	0	no	northwest	2721.321	Overweight	Young Adult	0.000
12	62	female	26.290	0	yes	southeast	27808.725	Overweight	Senior Citizen	8342.618
13	23	male	34.400	0	no	southwest	1826.843	Obese	Young Adult	0.000
14	56	female	39.820	0	no	southeast	11090.718	Obese	Senior Citizen	3327.215
15	27	male	42.130	0	yes	southeast	35611.758	Obese	Young Adult	0.000
16	19	male	24.600	1	no	southwest	1837.237	Normal	Young Adult	0.000
17	52	female	30.750	1	no	northwest	10787.170	Obese	Senior Citizen	3239.201
18	23	male	23.145	0	no	northwest	2384.172	Normal	Young Adult	0.000
19	56	male	40.300	0	no	southeast	10602.385	Obese	Senior Citizen	3180.715
20	39	male	35.300	0	yes	southeast	34837.467	Obese	Middle Aged	0.000
21	60	female	36.005	0	no	northwest	13228.847	Obese	Senior Citizen	3965.454
22	39	female	32.400	1	no	southeast	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	female	33.085	0	no	northwest	14451.835	Normal	Senior Citizen	4335.551
28	55	female	32.778	2	no	northwest	1488.150	Obese	Senior Citizen	3680.590
29	32	male	26.205	0	no	northwest	2774.152	Overweight	Young Adult	0.000
30	31	male	36.300	2	yes	southeast	38711.000	Obese	Middle Aged	0.000
31	23	male	35.600	0	yes	southeast	35585.576	Obese	Young Adult	0.000
32	18	female	26.315	0	no	northwest	2198.190	Overweight	Young Adult	0.000
33	19	female	28.600	5	no	southeast	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	51194.559	Obese	Young Adult	0.000
36	19	male	20.425	0	no	northwest	1625.434	Normal	Young Adult	0.000
37	62	female	32.965	3	no	northwest	15612.193	Obese	Senior Citizen	4683.658
38	26	male	30.800	0	no	southeast	2302.300	Normal	Young Adult	0.000
39	35	male	36.670	1	yes	northwest	3774.276	Obese	Middle Aged	0.000
40	60	male	39.400	0	yes	southeast	4515.436	Obese	Senior Citizen	14452.038
41	24	male	36.600	0	no	northwest	3044.062	Overweight	Young Adult	0.000
42	31	female	36.630	2	no	southeast	4949.785	Obese	Middle Aged	0.000
43	41	male	21.780	1	no	southeast	4272.477	Normal	Middle Aged	0.000
44	37	female	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	southeast	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	3566.522	Obese	Young Adult	0.000
49	60	male	24.530	0	no	southeast	12629.697	Normal	Senior Citizen	3785.969
50	36	male	35.200	1	yes	southeast	38709.176	Obese	Middle Aged	0.000
51	18	female	35.625	0	no	northwest	10711.285	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	southeast	23568.272	Overweight	Middle Aged	0.000
54	36	male	34.430	0	yes	southeast	37742.576	Obese	Middle Aged	0.000

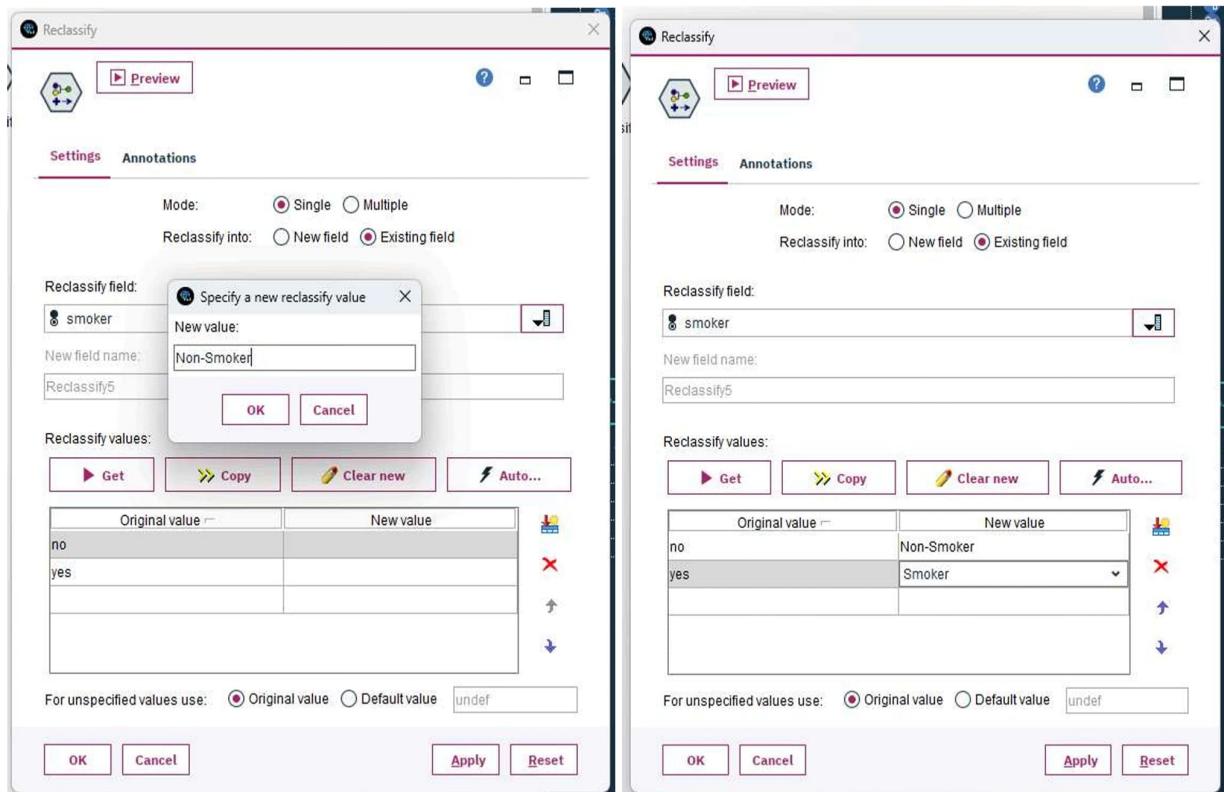
OK

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.

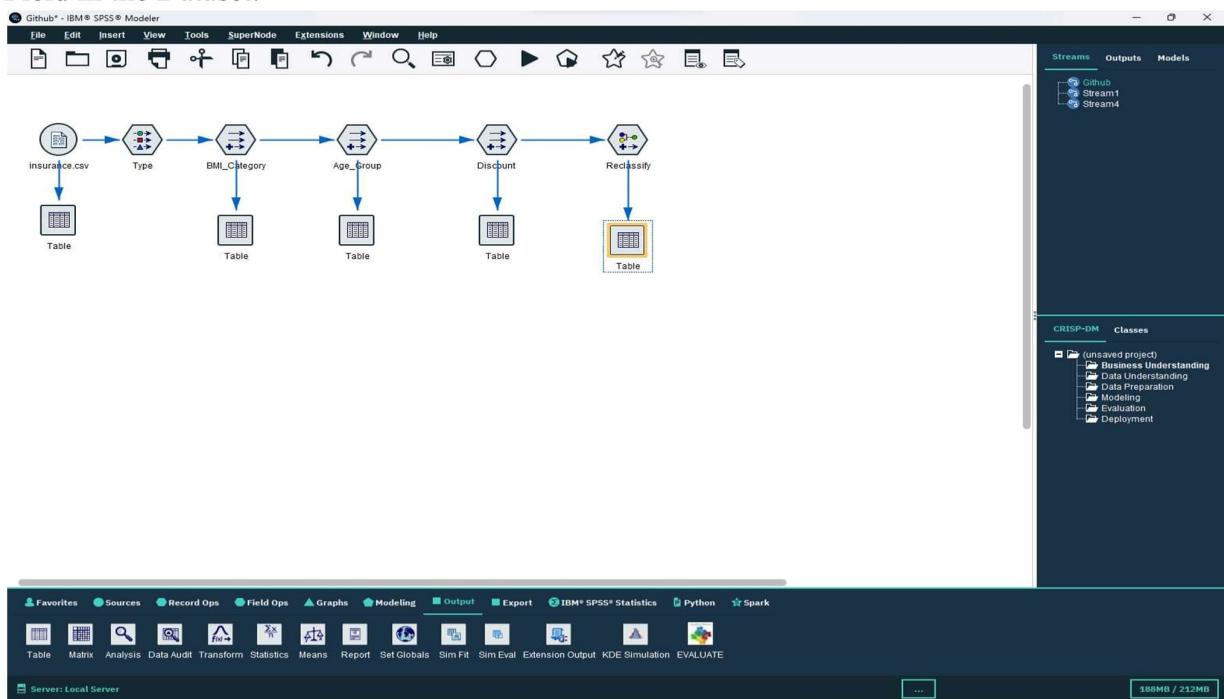


Table (10 fields, 1,338 records) #3

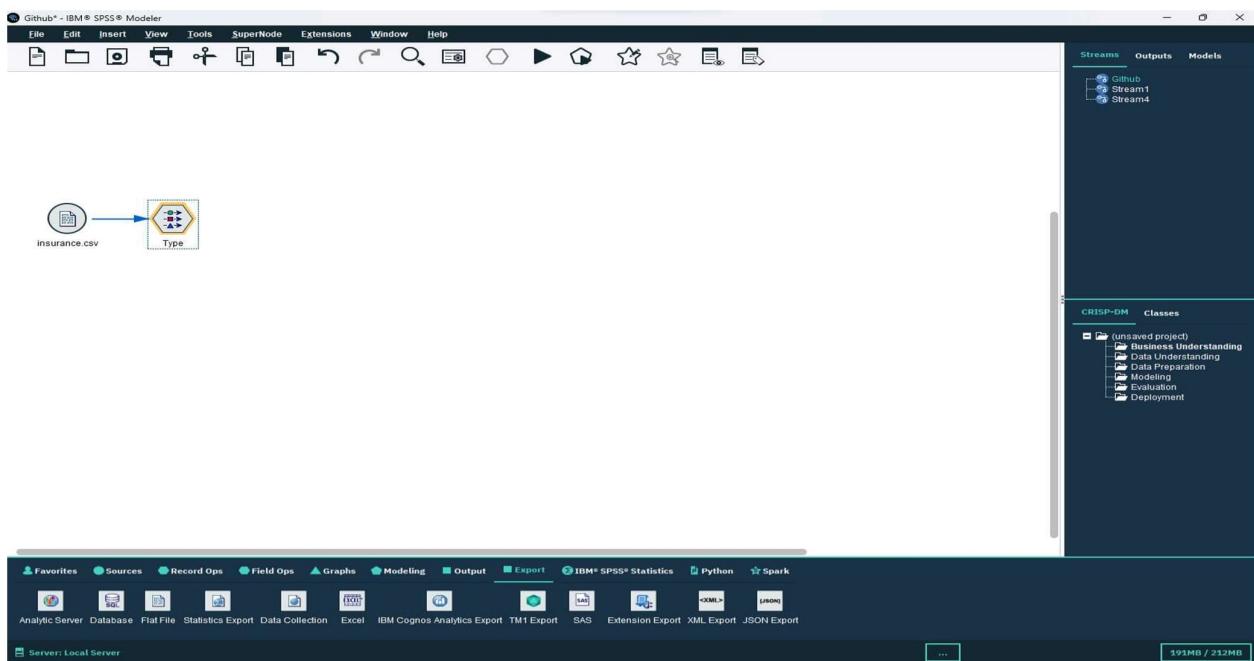
File Edit Generate ? X

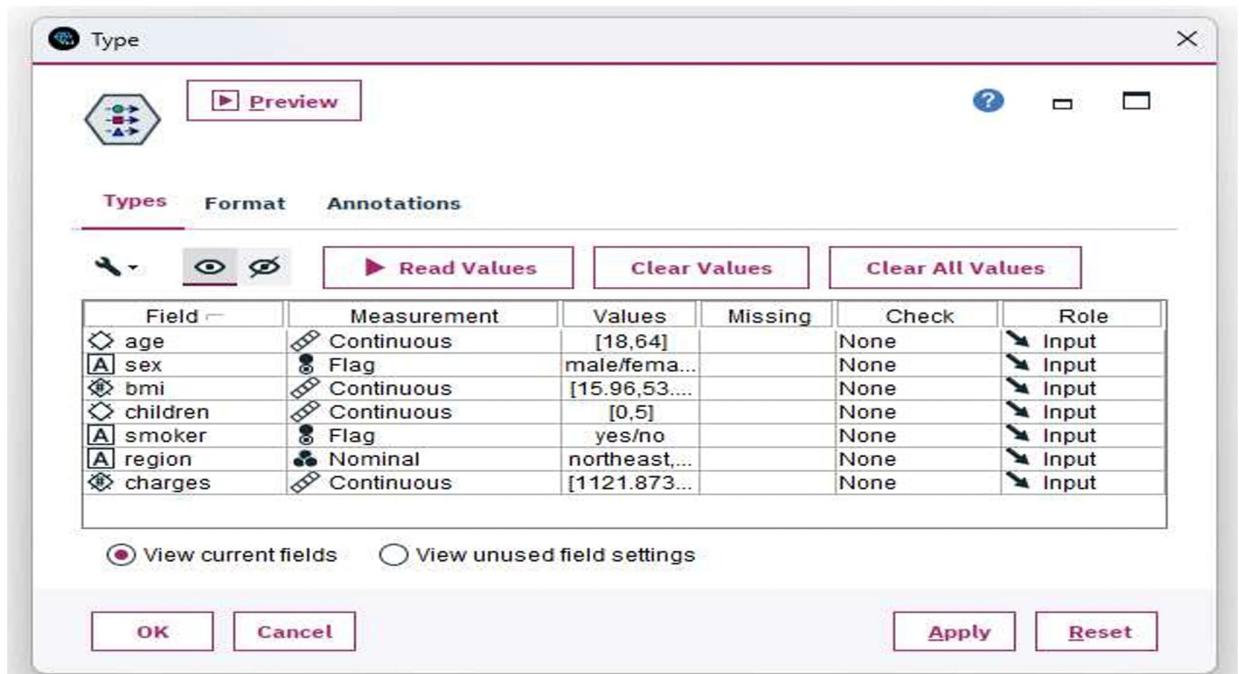
Table Annotations

	age	sex	bmi	children	smoker	region	charges	BMI_Catagory	Age_Group	Discount
1	19	female	27.900	0	1Non-Smoker	southwest	1784.324	Overweight	Young Adult	0.000
2	18	male	31.770	0	1Non-Smoker	northwest	1784.324	Overweight	Young Adult	0.000
3	20	male	33.000	0	0Non-Smoker	southeast	4449.462	Obese	Young Adult	0.000
4	33	male	22.705	0	0Non-Smoker	northwest	21884.471	Normal	Middle Aged	0.000
5	32	male	28.880	0	0Non-Smoker	northwest	3866.855	Overweight	Middle Aged	0.000
6	31	female	25.740	0	0Non-Smoker	Non-Smoked	3756.622	Overweight	Middle Aged	0.000
7	46	female	33.440	0	1Non-Smoker	southeast	8240.590	Obese	Middle Aged	0.000
8	37	female	27.740	3	0Non-Smoker	northwest	7281.506	Overweight	Middle Aged	0.000
9	37	male	29.830	2	0Non-Smoker	northwest	6406.411	Overweight	Middle Aged	0.000
10	60	female	25.840	0	0Non-Smoker	northwest	28923.137	Overweight	Senior Citizen	8676.941
11	25	male	26.220	0	0Non-Smoker	northwest	2721.321	Overweight	Young Adult	0.000
12	62	female	24.850	0	0Smoker	northwest	4704.321	Overweight	Senior Citizen	8342.18
13	23	male	34.140	0	0Non-Smoker	southeast	1824.843	Obese	Young Adult	0.000
14	56	female	39.830	0	0Non-Smoker	southeast	11990.718	Obese	Senior Citizen	3327.215
15	27	male	42.130	0	0Smoker	southeast	39611.758	Obese	Young Adult	0.000
16	19	male	24.600	1	1Non-Smoker	northwest	1837.237	Normal	Young Adult	0.000
17	52	female	30.780	1	1Non-Smoker	northwest	10797.336	Obese	Senior Citizen	3239.201
18	23	male	23.845	0	0Non-Smoker	northwest	2395.172	Normal	Young Adult	0.000
19	56	female	40.300	0	0Non-Smoker	southwest	10602.385	Obese	Senior Citizen	3180.715
20	30	male	35.300	0	0Smoker	southwest	36837.467	Obese	Middle Aged	0.000
21	60	female	36.005	0	0Non-Smoker	northwest	31228.847	Obese	Senior Citizen	3965.654
22	30	female	32.400	0	1Non-Smoker	southwest	4169.324	Obese	Middle Aged	0.000
23	18	male	30.000	0	0Non-Smoker	northwest	11990.718	Obese	Young Adult	0.000
24	34	female	31.920	1	0Smoker	northwest	37701.877	Obese	Middle Aged	0.000
25	37	male	28.025	2	0Non-Smoker	northwest	3203.902	Overweight	Middle Aged	0.000
26	59	female	27.720	0	0Non-Smoker	southeast	14001.134	Overweight	Senior Citizen	4200.340
27	63	female	23.085	0	0Non-Smoker	northwest	14451.835	Normal	Senior Citizen	4335.551
28	55	female	32.775	0	0Non-Smoker	northwest	12268.632	Obese	Senior Citizen	3680.590
29	23	male	17.385	1	0Non-Smoker	northwest	2775.192	Underweight	Young Adult	0.000
30	31	male	36.300	2	0Smoker	southwest	38711.000	Obese	Middle Aged	0.000
31	22	male	35.600	0	0Smoker	southwest	35585.576	Obese	Young Adult	0.000
32	18	female	26.315	0	0Non-Smoker	northwest	2198.190	Overweight	Young Adult	0.000
33	33	male	25.600	0	0Non-Smoker	southwest	4687.97	Overweight	Young Adult	0.000
34	63	male	31.110	0	0Non-Smoker	northwest	13781.059	Overweight	Senior Citizen	4131.059
35	26	male	36.400	1	1Smoker	southwest	51194.559	Obese	Young Adult	0.000
36	19	male	20.425	0	0Non-Smoker	northwest	1625.434	Normal	Young Adult	0.000
37	62	female	32.965	3	0Non-Smoker	northwest	15612.193	Obese	Senior Citizen	4483.658
38	26	male	20.800	0	0Non-Smoker	southwest	2302.300	Normal	Young Adult	0.000
39	35	male	36.670	1	0Smoker	northwest	39774.276	Obese	Middle Aged	0.000
40	60	male	39.900	0	0Smoker	southwest	40173.361	Obese	Senior Citizen	14452.008
41	24	female	26.600	0	0Non-Smoker	northwest	3046.062	Overweight	Young Adult	0.000
42	31	female	36.630	2	0Non-Smoker	southwest	4949.759	Obese	Middle Aged	0.000
43	41	male	21.780	1	0Non-Smoker	northwest	6272.477	Normal	Middle Aged	0.000
44	37	female	30.800	0	0Non-Smoker	northwest	4304.595	Obese	Middle Aged	0.000
45	36	male	37.150	1	0Non-Smoker	northwest	6074.679	Obese	Middle Aged	0.000
46	55	male	37.300	0	0Non-Smoker	northwest	20630.284	Obese	Senior Citizen	6189.085
47	18	female	35.665	2	0Non-Smoker	northwest	3393.356	Obese	Young Adult	0.000
48	28	female	34.770	0	0Non-Smoker	northwest	3556.922	Obese	Young Adult	0.000
49	60	female	24.530	0	0Non-Smoker	southeast	12625.897	Normal	Senior Citizen	3788.969
50	36	male	35.200	1	0Smoker	southeast	38709.176	Obese	Middle Aged	0.000
51	18	female	35.625	0	0Non-Smoker	northwest	2211.131	Obese	Young Adult	0.000
52	21	female	33.630	2	0Non-Smoker	northwest	3579.829	Obese	Young Adult	0.000
53	48	male	28.000	1	0Smoker	southwest	23568.272	Overweight	Middle Aged	0.000
54	36	male	34.430	0	0Smoker	southeast	37742.576	Obese	Middle Aged	0.000

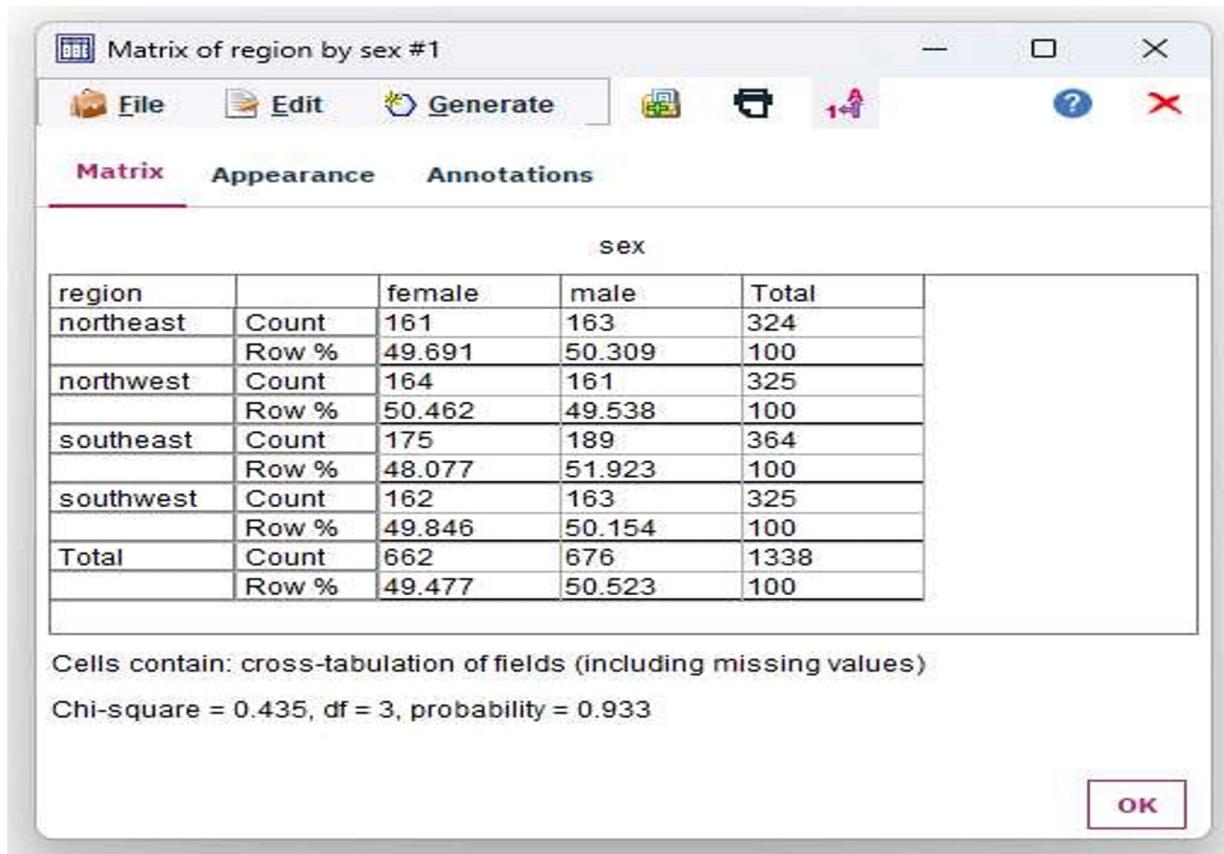
OK

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.

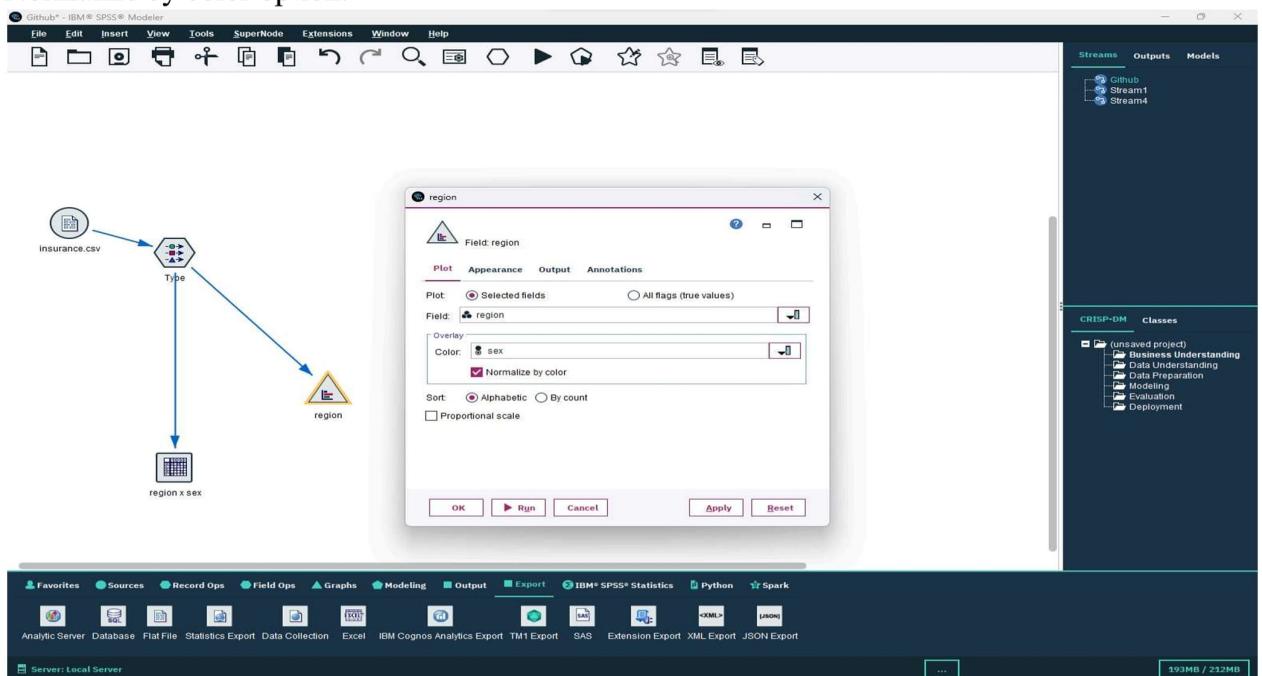


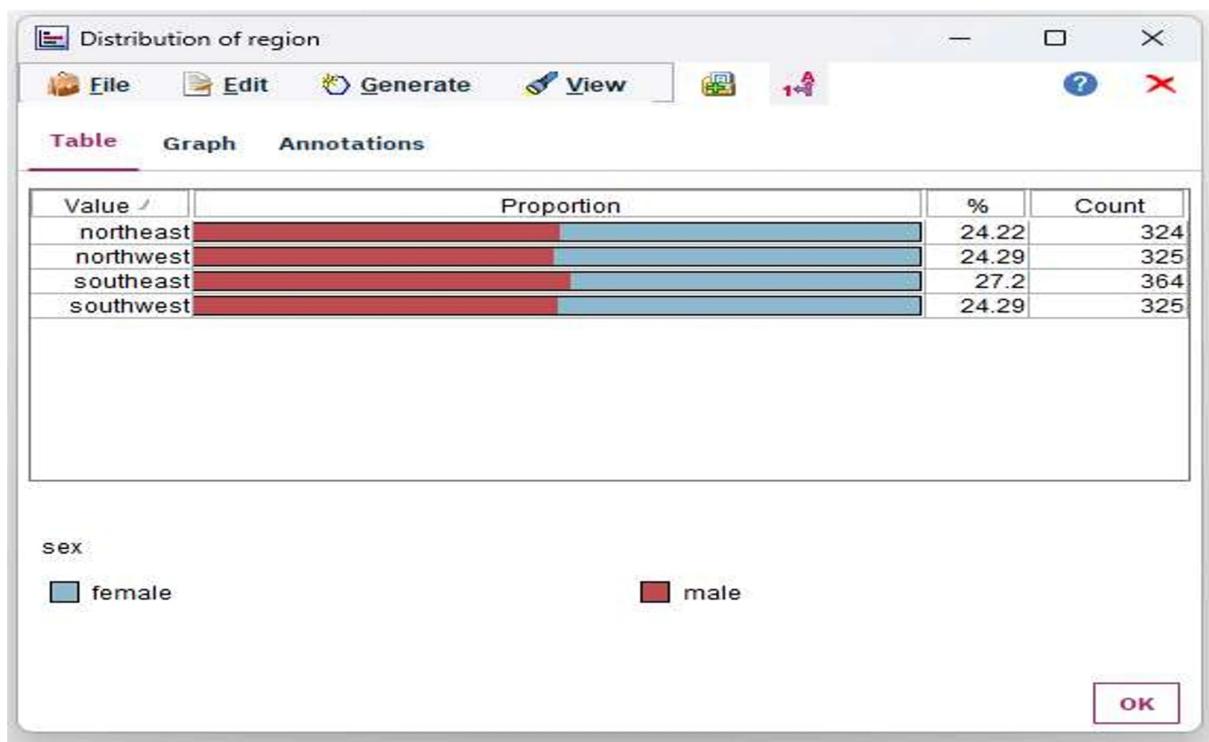


Step 19: We add the matrix node from the Output category to see our output or result in matrix form .When we used to make the relationship between two categorical field then used matrix for tabular output and distribution node from Graph category for graph output.
Take ‘region’ as row and ‘sex’ as column.

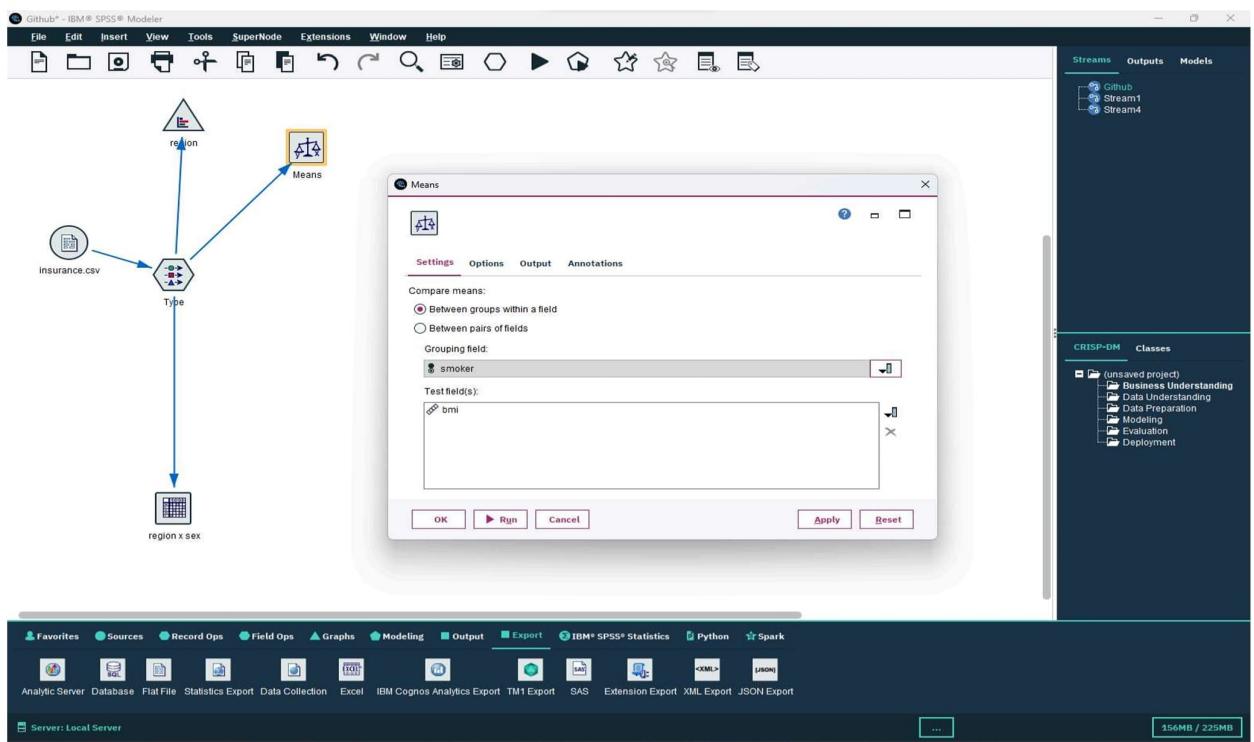


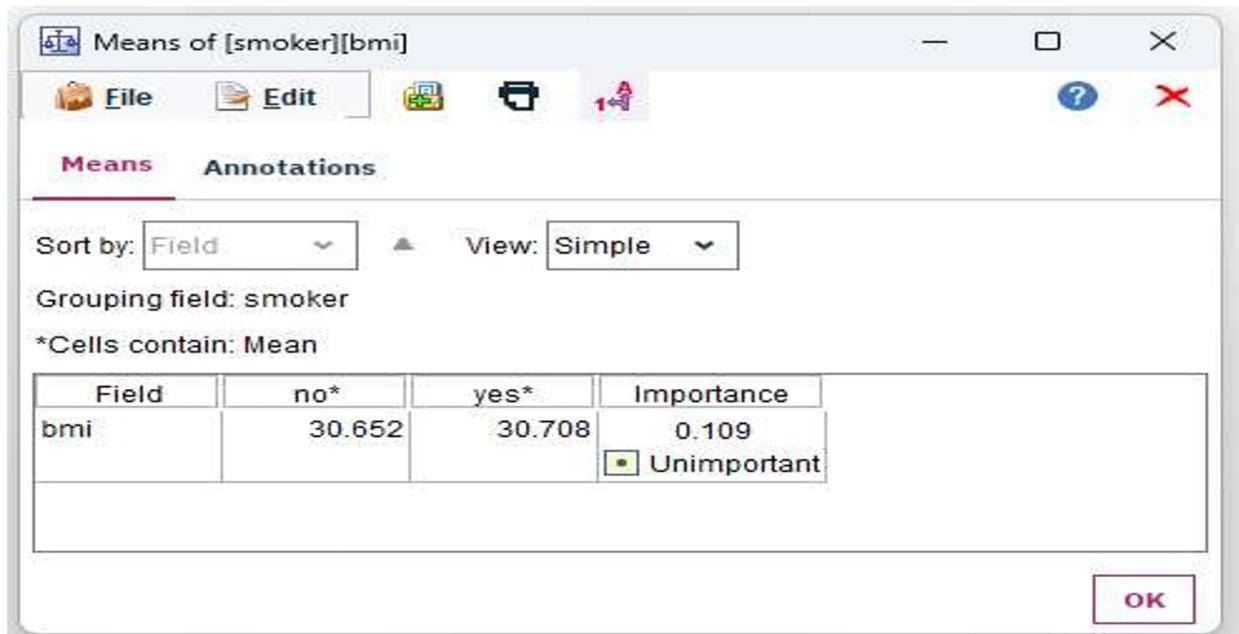
Step 20: For Distribution Graph take ‘region’ as Field and ‘sex’ as color and check the Normalize by color option.



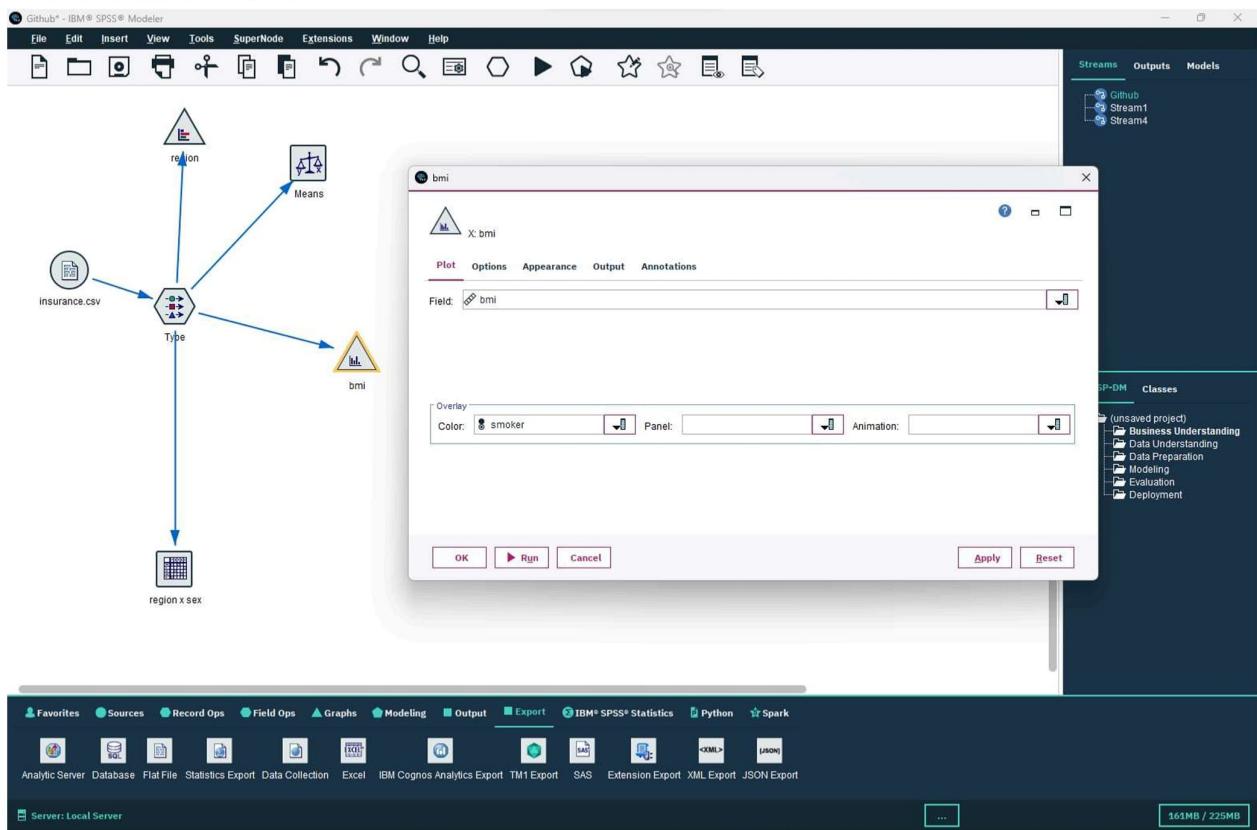


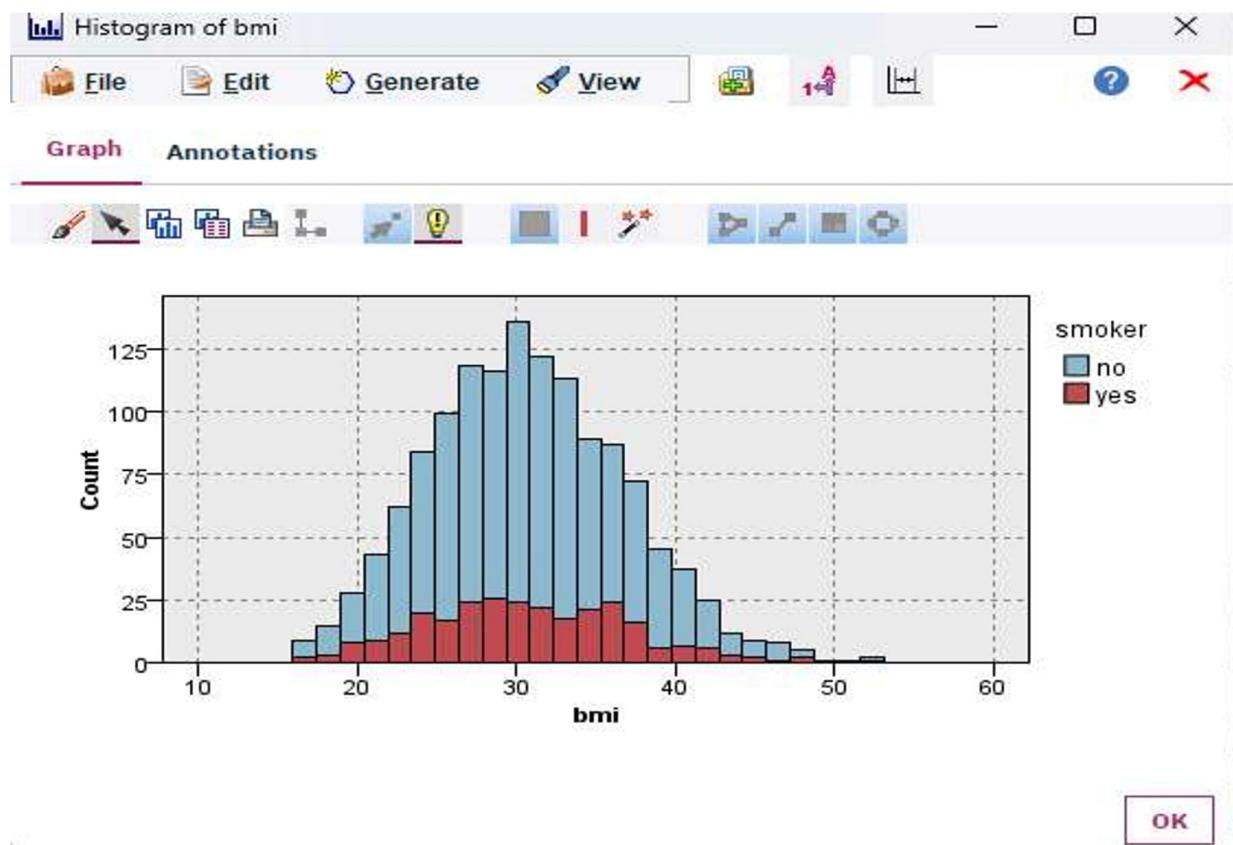
Step 21: We take the mean node from the output category because we want to find the relationship between one categorical and one continuous field. In this node we take the 'smoker' which is a categorical field and 'bmi' which is a continuous field .





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 Stream diagram shows the following flow:

- An input node "insurance.csv" feeds into a "Type" node.
- The "Type" node has three outgoing paths:
 - To "region": A triangle node labeled "region".
 - To "Means": A square node labeled "Means".
 - To "Statistics": A square node labeled "Statistics".
- The "region" path leads to a triangle node labeled "bmi".
- The "Means" path leads to a triangle node labeled "bmi".
- The "Statistics" path leads to a rectangle node labeled "region x sex".

A Statistics dialog box is open, showing settings for examining "children" and correlating "charges".

Streams Outputs Models

CRISP-DM Classes

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark

Analytic Server Database Flat File Statistics Export Data Collection Excel IBM Cognos Analytics Export TM1 Export SAS Extension Export XML Export JSON Export

Server: Local Server 163MB / 225MB

Statistics of [children][charges] #1

File Edit Generate

Statistics Annotations

Collapse All Expand All

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

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.

