

# Maimoona Khilji

## BS-DS

### Semester 6

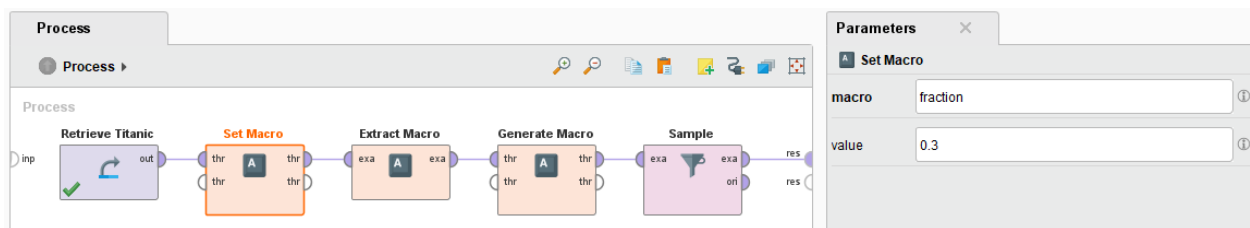
#### Lab Submission 08

Instructor: Basit Ali

1. Try to build samples of 30% and 80% of the original size. What do you need to change? What are the resulting example set sizes?

Size= 30%

Number of rows: 393



Open in Turbo Prep Auto Model

Filter (393 / 393 examples): all

Row No.	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Numb...	Passenger F...	Cabin	Port of Emb...	Life Boat	S
1	First	Allison, Mast...	Male	0.917	1	2	113781	151.550	C22 C26	Southampton	11	Y
2	First	Allison, Miss. ...	Female	2	1	2	113781	151.550	C22 C26	Southampton	?	N
3	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550	C22 C26	Southampton	?	N
4	First	Andrews, Mis...	Female	63	1	0	13502	77.958	D7	Southampton	10	Y
5	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300	B35	Cherbourg	9	Y
6	First	Baxter, Mr. Qu...	Male	24	0	1	PC 17558	247.521	B58 B60	Cherbourg	?	N
7	First	Bazzani, Miss...	Female	32	0	0	11813	76.292	D15	Cherbourg	8	Y
8	First	Beattie, Mr. T...	Male	36	0	0	13050	75.242	C6	Cherbourg	A	N
9	First	Beckwith, Mrs...	Female	47	1	1	11751	52.554	D35	Southampton	5	Y
10	First	Bidois, Miss. ...	Female	42	0	0	PC 17757	227.525	?	Cherbourg	4	Y
11	First	Bird, Miss. Eli...	Female	29	0	0	PC 17483	221.779	C97	Southampton	8	Y
12	First	Bjornstrom-S...	Male	28	0	0	110564	26.550	C52	Southampton	D	Y
13	First	Cardeza, Mr. ...	Male	36	0	1	PC 17755	512.329	B51 B53 B55	Cherbourg	3	Y
14	First	Carlsson, Mr....	Male	33	0	0	695	5	B51 B53 B55	Southampton	?	N

ExampleSet (393 examples, 0 special attributes, 12 regular attributes)

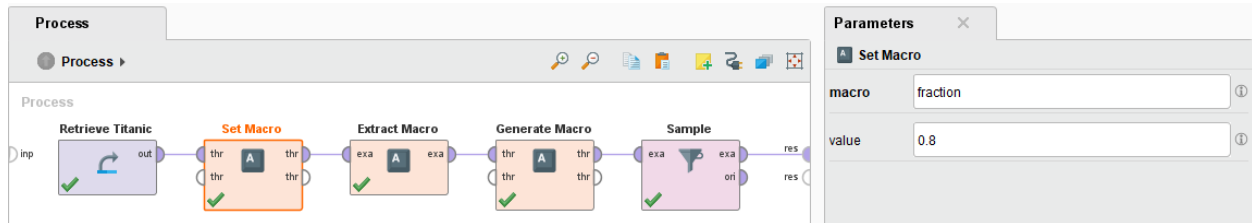
# Maimoona Khilji

## BS-DS

### Semester 6

Size= 80%

Number of rows: 1047



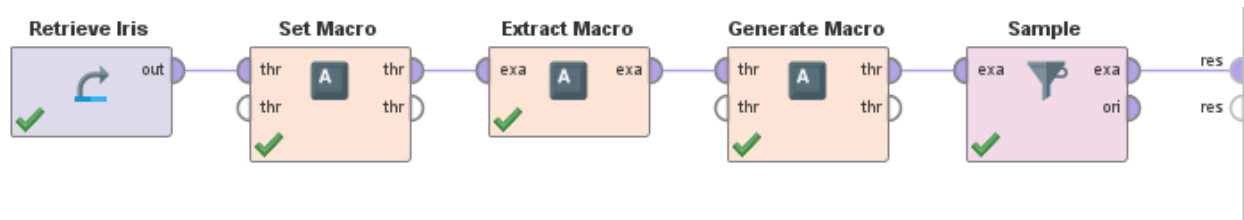
Open in [Turbo Prep](#) [Auto Model](#) Filter (1,047 / 1,047 examples): all

Row No.	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Num...	Passenger F...	Cabin	Port of Emb...	Life Boat	S
1	First	Allison, Mast...	Male	0.917	1	2	113781	151.550	C22 C26	Southampton	11	Y
2	First	Allison, Miss. ...	Female	2	1	2	113781	151.550	C22 C26	Southampton	?	N
3	First	Allison, Mr. H...	Male	30	1	2	113781	151.550	C22 C26	Southampton	?	N
4	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550	C22 C26	Southampton	?	N
5	First	Andrews, Mis...	Female	63	1	0	13502	77.958	D7	Southampton	10	Y
6	First	Appleton, Mrs...	Female	53	2	0	11769	51.479	C101	Southampton	D	Y
7	First	Artagaveytia, ...	Male	71	0	0	PC 17609	49.504	?	Cherbourg	?	N
8	First	Astor, Mrs. Jo...	Female	18	1	0	PC 17757	227.525	C62 C64	Cherbourg	4	Y
9	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300	B35	Cherbourg	9	Y
10	First	Barber, Miss. ...	Female	26	0	0	19877	78.850	?	Southampton	6	Y
11	First	Barkworth, Mr...	Male	80	0	0	27042	30	A23	Southampton	B	Y
12	First	Baumann, Mr...	Male	?	0	0	PC 17318	25.925	?	Southampton	?	N
13	First	Baxter, Mr. Qu...	Male	24	0	1	PC 17558	247.521	B58 B60	Cherbourg	?	N
14	First	Baxter, Mrs. J...	Female	50	0	1	PC 17558	247.521	B58 B60	Cherbourg	6	Y

ExampleSet (1,047 examples, 0 special attributes, 12 regular attributes)

2. Replace the input example set (Titanic) with Iris from the Samples repository. Do you need to change anything else or will the process just execute fine? Try it!

After replacing example set with Iris, it works correctly.



100% data = 150 rows

# Maimoona Khilji

## BS-DS

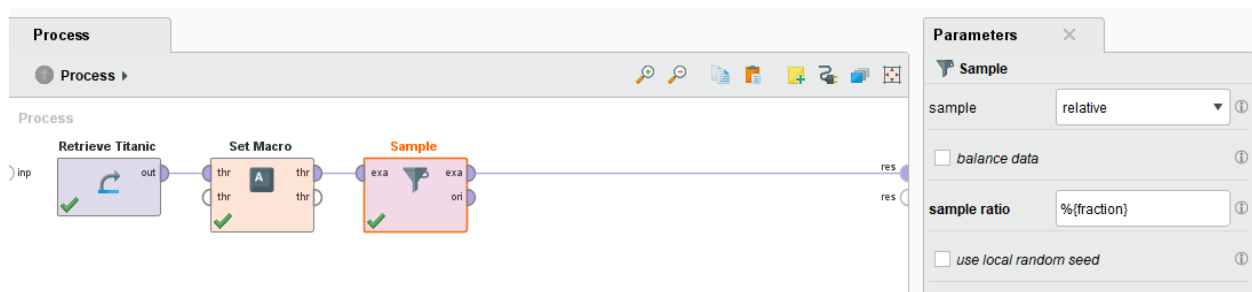
### Semester 6

Open in [Turbo Prep](#) [Auto Model](#)

Row No.	id	label	a1	a2	a3	a4
1	id_1	Iris-setosa	5.100	3.500	1.400	0.200
2	id_2	Iris-setosa	4.900	3	1.400	0.200
3	id_3	Iris-setosa	4.700	3.200	1.300	0.200
4	id_4	Iris-setosa	4.600	3.100	1.500	0.200
5	id_5	Iris-setosa	5	3.600	1.400	0.200
6	id_6	Iris-setosa	5.400	3.900	1.700	0.400
7	id_7	Iris-setosa	4.600	3.400	1.400	0.300
8	id_8	Iris-setosa	5	3.400	1.500	0.200
9	id_9	Iris-setosa	4.400	2.900	1.400	0.200
10	id_10	Iris-setosa	4.900	3.100	1.500	0.100
11	id_11	Iris-setosa	5.400	3.700	1.500	0.200
12	id_12	Iris-setosa	4.800	3.400	1.600	0.200
13	id_13	Iris-setosa	4.800	3	1.400	0.100
14	id_14	Iris-setosa	4.300	3	1.100	0.100
15	id_15	Iris-setosa	5.800	4	1.200	0.200

ExampleSet (150 examples, 2 special attributes, 4 regular attributes)

- Have a look into the Parameters of Sample. Do you find a setting which would allow you to create a 50% sample of the data without calculating the size macro? Try to change the process but keep the fraction macro and use it directly for setting the fraction in the Sample operator



Total number of rows of 50 % data = 654.

Maimoona Khilji

BS-DS

Semester 6

Open in

Turbo Prep

Auto Model

Filter (654 / 654 examples):

all

Row No.	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Numb...	Passenger F...	Cabin	Port of Emb...	Life Boat	S
1	First	Allison, Mast...	Male	0.917	1	2	113781	151.550	C22 C26	Southampton	11	Y
2	First	Allison, Miss. ...	Female	2	1	2	113781	151.550	C22 C26	Southampton	?	N
3	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550	C22 C26	Southampton	?	N
4	First	Andrews, Mis...	Female	63	1	0	13502	77.958	D7	Southampton	10	Y
5	First	Artagaveytia, ...	Male	71	0	0	PC 17609	49.504	?	Cherbourg	?	N
6	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300	B35	Cherbourg	9	Y
7	First	Baxter, Mr. Qu...	Male	24	0	1	PC 17558	247.521	B58 B60	Cherbourg	?	N
8	First	Bazzani, Miss...	Female	32	0	0	11813	76.292	D15	Cherbourg	8	Y
9	First	Beattie, Mr. T...	Male	36	0	0	13050	75.242	C6	Cherbourg	A	N
10	First	Beckwith, Mr. ...	Male	37	1	1	11751	52.554	D35	Southampton	5	Y
11	First	Beckwith, Mrs...	Female	47	1	1	11751	52.554	D35	Southampton	5	Y
12	First	Bidois, Miss. ...	Female	42	0	0	PC 17757	227.525	?	Cherbourg	4	Y
13	First	Bird, Miss. Eli...	Female	29	0	0	PC 17483	221.779	C97	Southampton	8	Y
14	First	Bishop, Mr. D...	Male	25	1	0	11967	91.079	B49	Cherbourg	7	Y

ExampleSet (654 examples, 0 special attributes, 12 regular attributes)