

Felipe Guth - 14210231

P1

Missing values are identified with an "?".

Percentage of missing values: 379/10722. 3.53%

The percentage is not very important because it represents a small fraction of the data.

The best way would be filling the age as 17 if the column juvenile contain the value yes and fill with the mean value in case of not juvenile or in case of missing this second information as well.

RM technique, replace missing values with average of attribute.

The screenshot shows the RapidMiner Studio interface. The main canvas displays a workflow with two processes: 'Retrieve practice data' and 'Replace Missing Values'. The 'Replace Missing Values' process is configured with the following parameters:

- attribute filter type: single
- attribute: Age at time of check
- invert selection: ☐
- include special attributes: ☐
- default: average

The screenshot shows the 'Result Overview' of the 'Replace Missing Values' process. The table displays 10722 examples with 16 regular attributes. The attributes are: Row No., Age at time of check, Reference No., Police Area, Date of Check, Time of Check, To/From, Self Defined, Officer Defined, Search Authority, and Juvenile.

Row No.	Age at time of check	Reference No.	Police Area	Date of Check	Time of Check	To/From	Self Defined	Officer Defined	Search Authority	Juvenile
8643	7	8643	JQ	Oct 18, 2003	Dec 30, 189	ESHFORT	White - Britis	1 - WHITE E!	A. Police anc	Yes
842	9	842	AQ	Nov 8, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE E!	A. Police anc	Yes
2869	9	2869	CQ	Jan 16, 2004	Dec 30, 189	MEOTSTONI	White - Britis	1 - WHITE E!	A. Police anc	?
4393	9	4393	EQ	Jul 28, 2003	Dec 30, 189	SHEERNES	White - Britis	1 - WHITE E!	A. Police anc	?
4988	9	4988	FQ	Sep 2, 2003	Dec 30, 189	CENTERBU	White - Britis	1 - WHITE E!	A. Police anc	Yes
10305	9	10305	LQ	Jul 24, 2003	Dec 30, 189	FRENCE	White - Britis	1 - WHITE E!	F. Terrorism	Yes
1087	10	1087	AQ	Nov 8, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE E!	A. Police anc	Yes
1176	10	1176	AQ	Dec 1, 2003	Dec 30, 189	TREVESENTI	N - Ethnic Cr	1 - WHITE E!	A. Police anc	Yes
4478	10	4478	EQ	Sep 14, 200	Dec 30, 189	FEVERSHAM	White - Britis	1 - WHITE E!	A. Police anc	Yes
5045	10	5045	FQ	Aug 23, 200	Dec 30, 189	THOTSTEBL	White - Britis	1 - WHITE E!	A. Police anc	Yes
6916	10	6916	HQ	Jul 27, 2003	Dec 30, 189	HYTHE	White - Britis	1 - WHITE E!	C. Firearms	Yes
7500	10	7500	HQ	Apr 17, 2003	Dec 30, 189	TOWCHEST	White - Britis	1 - WHITE E!	A. Police anc	Yes
8393	10	8393	JQ	Aug 26, 200	Dec 30, 189	ESHFORT	White - Britis	1 - WHITE E!	A. Police anc	Yes
412	11	412	AQ	Jun 12, 200	Dec 30, 189	TERTFORT	White - Britis	1 - WHITE E!	A. Police anc	Yes
605	11	605	AQ	Jul 19, 2003	Dec 30, 189	TREENHOTI	Mixed - Any c	1 - WHITE E!	A. Police anc	Yes
949	11	949	AQ	Oct 17, 2003	Dec 30, 189	TREVESENTI	N - Ethnic Cr	0 - UNKNOW	D. Criminal -	Yes
2195	11	2195	CQ	Jan 26, 2004	Dec 30, 189	MEOTSTONI	White - Britis	1 - WHITE E!	A. Police anc	Yes
2375	11	2375	CQ	Aug 1, 2003	Dec 30, 189	MEOTSTONI	White - Britis	1 - WHITE E!	A. Police anc	Yes
2557	11	2557	CQ	Oct 24, 2003	Dec 30, 189	MEOTSTONI	White - Britis	1 - WHITE E!	A. Police anc	Yes
2558	11	2558	CQ	Oct 23, 2003	Dec 30, 189	MEOTSTONI	White - Britis	1 - WHITE E!	A. Police anc	Yes
2592	11	2592	CQ	Nov 2, 2003	Dec 30, 189	MEOTSTONI	White - Britis	1 - WHITE E!	A. Police anc	?
2870	11	2870	CQ	Jan 16, 2004	Dec 30, 189	MEOTSTONI	White - Britis	1 - WHITE E!	A. Police anc	?
3711	11	3711	DQ	Apr 10, 2003	Dec 30, 189	CHETHEM	White - Britis	1 - WHITE E!	A. Police anc	Yes
4720	11	4720	EQ	Nov 18, 200	Dec 30, 189	SOTTONTB	White - Britis	1 - WHITE E!	A. Police anc	Yes
5837	11	5837	GQ	Jul 10, 2003	Dec 30, 189	REMSSTETE	White - Britis	1 - WHITE E!	A. Police anc	Yes
7186	11	7186	HQ	Sep 28, 200	Dec 30, 189	TOVER	White - Britis	1 - WHITE E!	A. Police anc	Yes

P2

The noise in the case is the variance error due to data entry problems. The officers in the moment of inputting the data may have had problems with the equipment or performed a wrong operation leading to erroneous inserting of data.

Binning method to categorize the data into 5 classes.

The screenshot shows the RapidMiner Studio Community 6.5.002 interface. The main process canvas displays a workflow with three nodes: 'Retrieve practice...', 'Replace Missing Values', and 'Discretize'. The 'Discretize' node is selected, and its parameters are shown in the right-hand pane. The parameters are as follows:

- create view**: ☐
- attribute filter type**: single
- attribute**: Age at time of check
- invert selection**: ☐
- include special attributes**: ☐
- number of bins**: 5
- define boundaries**: ☒
- min value**: 6.0
- max value**: 74.0
- range name type**: long

Nominal values			
Index	Nominal value	Absolute count	Fraction
1	range2 [6 - 19.600]	4907	0.458
2	range3 [19.600 - 33.200]	4009	0.374
3	range4 [33.200 - 46.800]	1422	0.133
4	range5 [46.800 - 60.400]	345	0.032
5	range6 [60.400 - 74]	39	0.004
6	range1 [-∞ - 6]	0	0
7	range7 [74 - ∞]	0	0

<new process> - RapidMiner Studio Community 6.5.002 @ FG-lapt

File Edit Process Tools View Help

Result Overview ExampleSet (Discretize) ExampleSet (/Local Repository/data/practicalRM1) ExampleSet (/Local Repository/data/practicalRM1)

ExampleSet (10722 examples, 0 special attributes, 16 regular attributes) Filter (10,722 / 10,722 examples): all

Row No.	Age at time ...	Reference ...	Police Area	Date of Che...	Time of Ch...	ToTn	Self Define...	Officer Defi...	Search Auth...	Juvenile	Search Type	Search Rea...	Arrested
1	range2 [6 - 1 1	1	AQ	Dec 8, 2003	Dec 30, 189	TREENHOTI	White - Britis	1 - WHITE EI	A. Police anc	Yes	Person	Stolen Prop	No
2	range2 [6 - 1 2		AQ	Oct 19, 2003	Dec 30, 189	TERTFORT	White - Britis	1 - WHITE EI	B. Misuse of	Yes	Person	Drugs	No
3	range2 [6 - 1 3		AQ	Apr 1, 2003	Dec 30, 189	TREVESEN1	White - Britis	1 - WHITE EI	B. Misuse of	Yes	Person	Drugs	No
4	range2 [6 - 1 4		AQ	Dec 8, 2003	Dec 30, 189	TREENHOTI	White - Britis	1 - WHITE EI	A. Police anc	Yes	Person	Stolen Prop	No
5	range2 [6 - 1 5		AQ	Apr 2, 2003	Dec 30, 189	TEST DRIVE	White - Britis	1 - WHITE EI	B. Misuse of	Yes	Person	Drugs	No
6	range2 [6 - 1 6		AQ	Apr 2, 2003	Dec 30, 189	TEST DRIVE	White - Britis	1 - WHITE EI	B. Misuse of	No	Driver & Veh	Drugs	No
7	range2 [6 - 1 7		AQ	Apr 7, 2003	Dec 30, 189	TEST DRIVE	White - Britis	1 - WHITE EI	B. Misuse of	No	Driver & Veh	Drugs	Yes
8	range3 [19.6 8		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
9	range2 [6 - 1 9		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	Yes	Person	Drugs	No
10	range3 [19.6 10		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
11	range2 [6 - 1 11		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
12	range3 [19.6 12		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
13	range3 [19.6 13		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
14	range2 [6 - 1 14		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
15	range2 [6 - 1 15		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	Mixed - Whitr	1 - WHITE EI	B. Misuse of	Yes	Person	Drugs	No
16	range2 [6 - 1 16		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
17	range3 [19.6 17		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
18	range2 [6 - 1 18		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
19	range3 [19.6 19		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No
20	range3 [19.6 20		AQ	Apr 13, 2003	Dec 30, 189	STENLEY	White - Britis	1 - WHITE EI	B. Misuse of	No	Person	Drugs	No

P3

It is not possible to normalize the attributes Juvenile, Search Reason and Police area, once they are string, non-numerical values.

The only attributes possible of normalization are age at time of check and reference number (which do not make sense).

