1)

#	id	~	ABC	full_name	~	RBC	first_name	~	ABC	last_name	~		email	~	*	gender	~	#	age	~
1 - 21			20 Car	tegories		20 Cate	gories		20 Cat	egories		19 Categ	gories		2 Cates	gories		-78 - 91	ılı.	d.
		1		el Finnigan		Mariel			Finni	gan		mfinni	gan0@usda.gov		Femal	Э				60
			,	on · Possek		Kenyon			Posse	k		kpossel	k1@ucoz.com		Male					12
		3	Lalo	Manifould		Lalo			Manif	ould		lmanif	ould2@pbs.org		Male					26
		4	Nicko	ola Carous		Nickol	a		Carou	S		ncarou	s3@phoca.cz		Male					4
		5	Norma	an Dubbin		Norman			Dubbi			ndubbi	n40wikipedia.or	g	Male					17
		6	Hasty	y · Perdue		Hasty			Perdu	е		hperdu	e5@qq.com							77
		7	Franz	z Castello		Franz			Caste	110		100000	llo6@1688.com		Male					25
				e-Tarney		Jorge			Tarne				y7@ft.com		Male					77
		9	Eunio	ce Blakebroug	h	Eunice			Blake	brough			brough8@sohu.co		Female	е				45
				topher Frankc	ombe	Kristo	pher		Frank	combe			combe9@slate.co	m	Male					old
		11	Palm	Domotor		Palm			Domot	or		pdomot	ora@github.io		Male					6
		12	Luz · l	Lansdowne		Luz			Lansd				owneb@theguardi	an.com	Femal					16
		13	Modes	stia Keble		Modest	ia		Keble			mkeble	c@cmu.edu		Femal	е				91
		14	Stace	ee-Bovis		Stacee			Bovis			sbovis	d@webeden.co.uk		Female					22
		15	Eden	Wace		Eden			Wace			ewacee	@marriott.com		Femal	9				16
		16	Eden	- Wace		Eden			Wace			ewacee	@marriott.com		Female	е				16
		17	Tobia	as Sherburn		Tobias			Sherb	urn		tsherb	urnf@facebook.c	om	Male					2
																				nul1
		19	Clai	r·Skillern		Clair			Skill	ern		cskill	erng@nih.gov		Male					-78
				ew-Addicott		Mathew			Addic	ott		maddic	otth@acquirethi	sname.com	Male					65
		21	Keria	anne Goacher		Kerian	ne		Goach	er					Female	е				45
			Mauri	its Shawl		Maurit	S		Shawl			mshawl	j@dmoz.org		Male					72
																				nul1

Abbildung 1: Rohdaten

#	id	~	RBC	full_name	~	ABC	first_name	~	ABC	last_name	~		email	~	*	gender	~	#	age	~
1-20			19 Cate	gories		19 Cated	gories		19 Cat	egories		19 Categor	ies		2 Cated	iories		-78 - 91	1	aL.
		1		Finnigan		Mariel			Finni				n0@usda.gov		Female					60
		2	Kenyor	· Possek		Kenyon			Posse	ek		kpossek1	@ucoz.com		Male					12
		3	Lalo N	Manifould		Lalo			Manif	ould		lmanifou	ld2@pbs.org		Male					26
		4	Nickol	la-Carous		Nickol	а		Carou	IS		ncarous3	@phoca.cz		Male					4
		5	Norman	-Dubbin		Norman			Dubbi	n		ndubbin4	@wikipedia.or	-g	Male					17
		6	Hasty	Perdue		Hasty			Perdu	ie		hperdue5	@qq.com							77
		7	Franz	Castello		Franz			Caste	110		fcastell	o6@1688.com		Male					25
		8	Jorge	Tarney		Jorge			Tarne	ey		jtarney7	@ft.com		Male					77
		9	Eunice	·Blakebrough		Eunice			Blake	ebrough		eblakebr	ough8@sohu.co	om	Female	е				45
				pher Frankco	mbe	Kristo	pher		Frank	combe			mbe9@slate.co	om	Male					old
			Palm·[Palm			Domot				a@github.io		Male					6
				insdowne		Luz			Lansd				neb@theguardi	ian.com	Female					16
		13	Modest	ia Keble		Modest	ia		Keble	•		mkeblec@			Female					91
			Stacee			Stacee			Bovis	3			webeden.co.uk	(Female					22
			Eden · V			Eden			Wace				arriott.com		Female					16
			Eden · V			Eden			Wace				arriott.com		Female	9				16
		17		Sherburn		Tobias			Sherb				nf@facebook.o	com	Male					2
		19		Skillern		Clair			Skill				ng@nih.gov		Male					-78
		20		· Addicott		Mathew			Addio				th@acquirethi	isname.com						65
			Maurit	s-Shawl		Maurit	S		Shawl			mshawlj@	dmoz.org		Male					72

Abbildung 2: filter type: missing missing: email action: Delete

#	id	~	ABC	full_name	~	RBC	first_name	~	RBC	last_name	~		email	v .	•	gender	~	#	age	~
			III			I														ы.
1 - 20			18 Cate			18 Categ				egories		18 Categor			2 Categ			-78 - 91		
				l Finnigan		Mariel			Finni	-		-	n0@usda.gov		emale	9				60
			-	n · Possek		Kenyon			Posse				Pucoz.com		Male					12
				Manifould		Lalo			Manif	ould			ld2@pbs.org		Male					26
				la-Carous		Nickol	а		Carou			ncarous30	0phoca.cz		Male					4
		5	Norman	n-Dubbin		Norman			Dubbi	.n		ndubbin4	@wikipedia.org	M	Male					17
		6	Hasty	Perdue		Hasty			Perdu			hperdue50								77
		7	Franz	Castello		Franz			Caste	110		fcastell	o6@1688.com	M	Male					25
		8	Jorge	Tarney		Jorge			Tarne	y		jtarney70	@ft.com	M	Male					77
		9	Eunice	e Blakebrough	ı	Eunice			Blake	brough		eblakebro	ough8@sohu.com	F	Female	9				45
		11	Palm·[Domotor		Palm			Domot	or		pdomotora	a@github.io	M	Male					6
		12	Luz · La	ansdowne		Luz			Lanso	lowne		llansdow	neb@theguardian.com	n F	emale	9				16
		13	Modest	tia Keble		Modest	ia		Keble			mkeblec@	cmu.edu	F	Female	•				91
		14	Stace	e-Bovis		Stacee			Bovis			sbovisd@v	webeden.co.uk	F	emale	9				22
		15	Eden · V	Nace		Eden			Wace			ewacee@ma	arriott.com	F	Female	•				16
		16	Eden · V	Wace		Eden			Wace			ewacee@ma	arriott.com	F	emale					16
		17	Tobias	s · Sherburn		Tobias			Sherb	urn		tsherbur	nf@facebook.com	M	Male					2
		19	Clair	Skillern		Clair			Skill	ern.		cskiller	ng@nih.gov	M	Male					-78
		20	Mathe	w · Addicott		Mathew			Addio	ott		maddicot	th@acquirethisname.	.com M	Male					65
			Maurit	ts-Shawl		Maurit	S		Shawl			mshawlj@d	dmoz.org	M	Male					72

Abbildung 3: filter type: mismatched col: age mismatched: Integer action: Delete

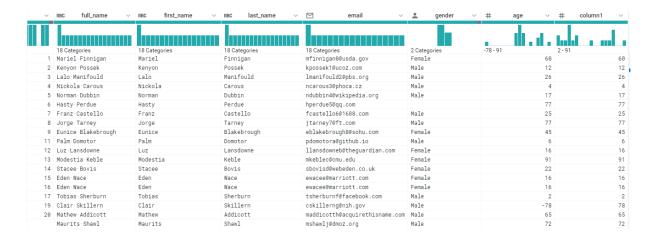


Abbildung 4: derive type: single value: if(age < 0, age * -1, age) as: 'column1'



Abbildung 5: drop col: age action: Drop



Abbildung 6: rename type: manual mapping: [column1,'age']

Aufgabe 7 Tool Supported Data Cleaning Exercise Carsten Brauer S8963350

#	id	~	RBC	full_name	~	RBC	first_name	~	RBC	last_name	~		email	~	<u>+</u>	gender	~	#	age	~
1-2	0		18 Cat	egories		18 Cated	ories		18 Cat	egories		18 Catego	pries		2 Cated	nories		2 - 91		
	-	1		l Finnigan		Mariel	,		Finni				an0@usda.gov		Femal					60
		2	Kenyo	n · Possek		Kenyon			Posse	k		kpossek	1@ucoz.com		Male					12
		3	Lalo	Manifould		Lalo			Manif	ould		lmanifo	uld2@pbs.org		Male					26
		4	Nicko	la Carous		Nickol	а		Carou	S		ncarous	3@phoca.cz		Male					4
		5	Norma	n Dubbin		Norman			Dubbi	n			4@wikipedia.org	l	Male					17
		6	Hasty	· Perdue		Hasty			Perdu				5@qq.com		nul1					77
				- Castello		Franz			Caste	110			lo6@1688.com		Male					25
			_	Tarney		Jorge			Tarne	-		, ,	7@ft.com		Male					77
				e Blakebroug	h	Eunice				brough			rough8@sohu.com	1	Femal	е				45
				Domotor		Palm			Domot				ra@github.io		Male					6
				ansdowne		Luz			Lansd				wneb@theguardia	in.com	Femal					16
				tia Keble		Modest	ia		Keble				@cmu.edu		Femal					91
				e Bovis		Stacee			Bovis				@webeden.co.uk		Femal					22
			Eden			Eden			Wace				marriott.com		Femal					16
			Eden			Eden			Wace				marriott.com		Femal	e				16
				s Sherburn		Tobias			Sherb				rnf@facebook.co	m	Male					2
				Skillern		Clair			Skill				rng@nih.gov		Male					78
		20		w Addicott		Mathew			Addic				tth@acquirethis	name.com	Male					65
			Mauri	ts Shawl		Maurit	S		Shawl			mshawlj	@dmoz.org		Male					72

Abbildung 7: set col: gender value: if(gender == ", null(), \$col)



Abbildung 8: set col: id value: ifmissing(\$col, max(id) + 1)

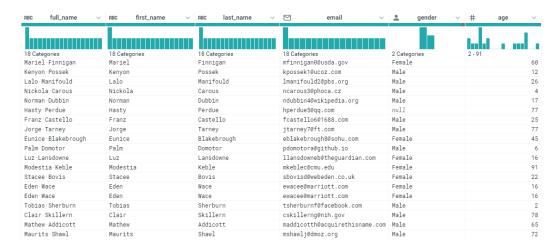


Abbildung 9: drop col: id action: Drop

Aufgabe 7 Tool Supported Data Cleaning Exercise Carsten Brauer S8963350

RBC full_name	∨ ABC firs	st_name v	RBC	last_name	~		email	~	*	gender	~	#	age	~
	ı I		Inn											
18 Categories	18 Categories	3	18 Catego			18 Categorie			2 Catego	ries		2 - 91		
Mariel Finnigan	Mariel		Finniga	n			0@usda.gov		Female					60
Kenyon · Possek	Kenyon		Possek			kpossek1@	ucoz.com		Male					12
Lalo Manifould	Lalo		Manifou	ld		lmanifoul	d2@pbs.org		Male					26
Nickola-Carous	Nickola		Carous			ncarous3@	phoca.cz		Male					4
Norman Dubbin	Norman		Dubbin			ndubbin4@	wikipedia.org		Male					17
Hasty Perdue	Hasty		Perdue			hperdue5@	qq.com		null					77
Franz · Castello	Franz		Castell	0		fcastello	6@1688.com		Male					25
Jorge Tarney	Jorge		Tarney			jtarney7@	ft.com		Male					77
Eunice Blakebrough	Eunice		Blakebr	ough		eblakebro	ugh8@sohu.com		Female					45
Palm Domotor	Palm		Domotor			pdomotora	@github.io		Male					6
Luz · Lansdowne	Luz		Lansdow	ne		llansdown	eb@theguardian.	com	Female					16
Modestia-Keble	Modestia		Keble			mkeblec@c	mu.edu		Female					91
Stacee Bovis	Stacee		Bovis			sbovisd@w	ebeden.co.uk		Female					22
Eden · Wace	Eden		Wace			ewacee@ma	rriott.com		Female					16
Tobias · Sherburn	Tobias		Sherbur	n		tsherburn	f@facebook.com		Male					2
Clair Skillern	Clair		Skiller	n		cskillern	g@nih.gov		Male					78
Mathew · Addicott	Mathew		Addicot	t		maddicott	h@acquirethisna	me.com	Male					65
Maurits Shawl	Maurits		Shawl			mshawlj@d	moz.org		Male					72

Abbildung 10: deduplicate



Abbildung 11: filter type: lessThan col: age lessThan: 10 action: Delete

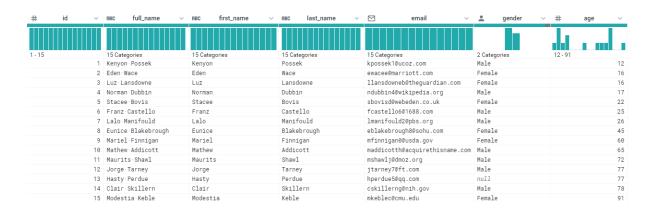


Abbildung 12: derive type: multiple value: rownumber() order: age as: 'id'

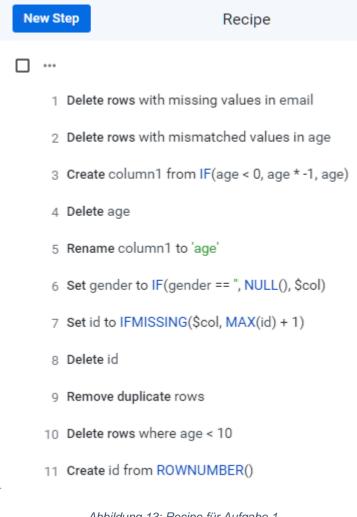


Abbildung 13: Recipe für Aufgabe 1

2)

- Schwerwiegende Fehler sind in den Spalten, die Zeit oder Datum enthalten.
 Diese Zeilen würde ich löschen, da es sich dabei nur um wenige Zeilen (67) handelt.
- 2. Dann lösche ich in allen Text-Spalten die Anführungszeichen.
- 3. Aus der Spalte "Number of Customers Affected" entferne ich alle Kommata, wandle den Datentyp in INT um und ersetze alle Missmatches mit NULL
- 4. Aus der Spalte "Demand Loss (MW)" entferne ich alle Buchstaben und Sonderzeichen und wandle sie in INT um, nachdem ich alle MISSMATCHED und MISSING in NULL umgewandelt habe.

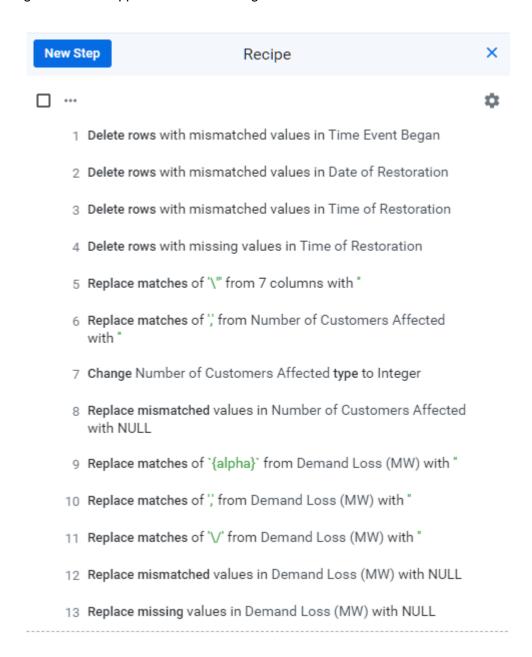


Abbildung 14: Recipe für Aufgabe 2