



New Step

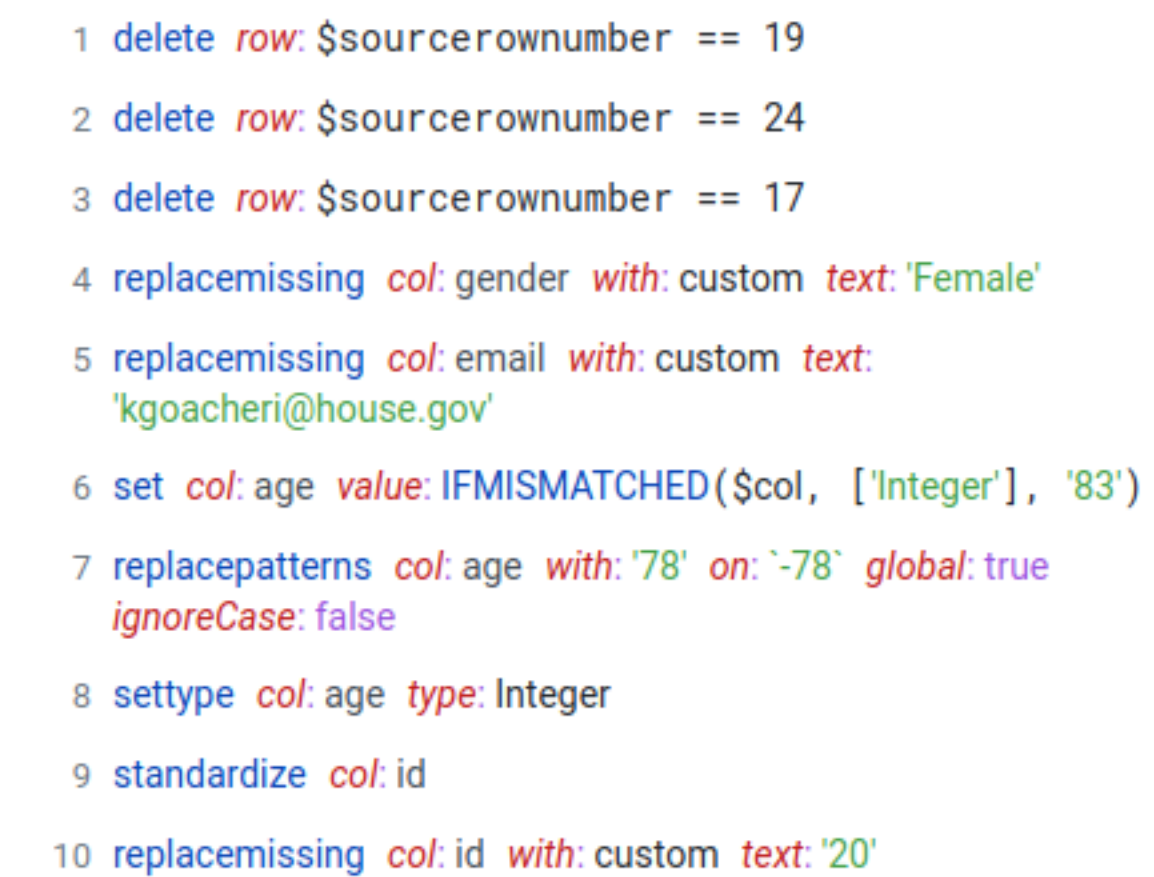
Recipe



- 1 Delete rows where \$sourcerownumber == 19
- 2 Delete rows where \$sourcerownumber == 24
- 3 Delete rows where \$sourcerownumber == 17
- 4 Replace missing values in gender with 'Female'
- 5 Replace missing values in email with 'kgoacheri@house.gov'
- 6 Set age to IFMISMATCHED(\$col, ['Integer'], '83')
- 7 Replace matches of '-78' from age with '78'
- 8 Change age type to Integer
- 9 Standardize id
- 10 Replace missing values in id with '20'



#	id	ABC	full_name	ABC	first_name	ABC	last_name	✉	email	👤	gender	#	age
1 - 20			20 Categories		20 Categories		20 Categories		20 Categories		2 Categories		2 - 91
	1		Mariel Finnigan		Mariel		Finnigan		mfinnigan0@usda.gov		Female		60
	2		Kenyon Possek		Kenyon		Possek		kpossek1@ucoz.com		Male		12
	3		Lalo Manifould		Lalo		Manifould		lmanifould2@pbs.org		Male		26
	4		Nickola Carous		Nickola		Carous		ncarous3@phoca.cz		Male		4
	5		Norman Dubbin		Norman		Dubbin		ndubbin4@wikipedia.org		Male		17
	6		Hasty Perdue		Hasty		Perdue		hperdue5@qq.com		Female		77
	7		Franz Castello		Franz		Castello		fcastello6@1688.com		Male		25
	8		Jorge Tarney		Jorge		Tarney		jtarney7@ft.com		Male		77
	9		Eunice Blakebrough		Eunice		Blakebrough		eblakebrough8@sohu.com		Female		45
	10		Kristopher Frankcombe		Kristopher		Frankcombe		kfrankcombe9@slate.com		Male		83
	11		Palm Domotor		Palm		Domotor		pdomotora@github.io		Male		6
	12		Luz Lansdowne		Luz		Lansdowne		llansdowneb@theguardian.com		Female		16
	13		Modestia Keble		Modestia		Keble		mkeblec@cmu.edu		Female		91
	14		Stacee Bovis		Stacee		Bovis		sbovisd@webeden.co.uk		Female		22
	15		Eden Wace		Eden		Wace		ewacee@marriott.com		Female		16
	16		Tobias Sherburn		Tobias		Sherburn		tsherburnf@facebook.com		Male		2
	17		Clair Skillern		Clair		Skillern		cskillerng@nih.gov		Male		78
	18		Mathew Addicott		Mathew		Addicott		maddicotth@acquirethisname.com		Male		65
	19		Kerianne Goacher		Kerianne		Goacher		kgoacheri@house.gov		Female		45
	20		Maurits Shawl		Maurits		Shawl		mshawlj@dmoz.org		Male		72



Download results

...

Details

×

Previewing 01\_dsm-beuth-edl-demo

#	id	ABC	full_name	ABC	first_name
1			Mariel Finnigan		Mariel
2			Kenyon Possek		Kenyon
3			Lalo Manifould		Lalo
4			Nickola Carous		Nickola
5			Norman Dubbin		Norman
6			Hasty Perdue		Hasty
7			Franz Castello		Franz
8			Jorge Tarney		Jorge
9			Eunice Blakebrough		Eunice
10			Kristopher Frankcombe		Kristopher
11			Palm Domotor		Palm
12			Luz Lansdowne		Luz
13			Modestia Keble		Modestia
14			Stacee Bovis		Stacee
15			Eden Wace		Eden
16			Tobias Sherburn		Tobias
17			Clair Skillern		Clair
18			Mathew Addicott		Mathew
19			Kerianne Goacher		Kerianne
20			Maurits Shawl		Maurits

Previewing 01\_dsm-beuth-edl-demodata-dirty - 2.csv.csv

×

ame	ABC	last_name	✉	email	👤	gender	#	age
		Finnigan		mfinnigan0@usda.gov		Female	60	
		Possek		kpossek1@ucoz.com		Male	12	
		Manifould		lmanifould2@pbs.org		Male	26	
		Carous		ncarous3@phoca.cz		Male	4	
		Dubbin		ndubbin4@wikipedia.org		Male	17	
		Perdue		hperdue5@qq.com		Female	77	
		Castello		fcastello6@1688.com		Male	25	
		Tarney		jtarney7@ft.com		Male	77	
		Blakebrough		eblakebrough8@sohu.com		Female	45	
		Frankcombe		kfrankcombe9@slate.com		Male	83	
		Domotor		pdomotora@github.io		Male	6	
		Lansdowne		llansdowneb@theguardian.com		Female	16	
		Keble		mkeblec@cmu.edu		Female	91	
		Bovis		sbovisd@webeden.co.uk		Female	22	
		Wace		ewacee@marriott.com		Female	16	
		Sherburn		tsherburnf@facebook.com		Male	2	
		Skillern		cskillerng@nih.gov		Male	78	
		Addicott		maddicotth@acquirethisname.com		Male	65	
		Goacher		kgoacheri@house.gov		Female	45	
		Shawl		mshawlj@dmoz.org		Male	72	

## Part 2:

Mögliche Lösungen sind in Orange hinterlegt.

### 1. Spalte

390 Kategorien bei 1652 Datensätzen sind zu viele.

Kategorien zusammenfassen.

### 2. Spalte Year + 3. Spalte Date Event Began

Aus den Jahren 2000-2003 sind relativ wenig Datensätze vorhanden.

Deshalb eventuell weniger aussagekräftig.

Rauswerfen oder mehr Datensätze suchen.

### 4. Spalte Time Event Began

Enthält NA Werte (23) – Rauswerfen.

Schreibweise von p.m. und a.m. nicht eindeutig, auch PM AM

Eine der Schreibweisen anpassen.

### 5. Spalte Date of Restoration

35 „mismatched values“

Datensatz Fehler zwei Spalten in einer: 09/23/04·12:00·p.m.

Per Hand aufteilen.

### 6. Spalte Time of Restoration

72 „mismatched values“

„ongoing“ Ereignisse, kritisch weil Zeitraum nicht konstant. – Rauswerfen.

### 7. Spalte Respondent

Es sind zwar nur die „Befragten“, aber 523 Kategorien bei 1652 Datensätzen sind nicht sinnvoll.

Eventuell Spalte dropen ?

### 8. Spalte Geographic Areas

Hier sind es auch eventuell zu viele Kategorien.

### 9. Spalte NERC Region

Wenige Datensätzen haben 2 Kategorien.

Rauswerfen, dann eindeutig, oder für eins entscheiden.

### 10. Spalte Demand Loss (MW)

Formate der Inhalte nicht eindeutig Strings wie „1,250“ und Integer wie 500.

Strings in Integer ändern.

24% N/A

1% NA

13% Unknown

Für mich ist Unknown eigentlich fast das gleiche wie wie N/A or. NA.

Die 3 Datensatzmöglichkeiten in N/A ändern.

11. Spalte Number of Customers Affected

12,35% N/A

8,78% Unknown

Anpassen.

Werte mit Abkürzungen dazwischen.

12. Spalte Tags

Die Werte der Spalte sind teilweise in „ “ oder ohne vorhanden.

Anpassen.