

# CSE564 Visualization

Yinlong Su, #110461173

April 10, 2016

## Row Clean Report

### 0 Version History

First built on April 10, 2016.

### 1 Row Cleaning

In the original dataset, there are some illegal rows have wrong data. For example, row 11,865 has wrong data on almost every column (shown in Figure 1). It has location information stored in time column, ingredient information stored in product name, etc.

Those illegal rows cause the warning from our csv parser pandas: “sys:1: DtypeWarning: Columns (0,3,5,27,36) have mixed types. Specify dtype option on import or set low\_memory=False.” We remove the rows that has a null value on “code” column and solve this problem.

The row number reduces from 65,503 to 65,486.

	A	B	C	D	E	F	G	H	I	J	K	L	M
11858	3.03E+12	http://wo	openfood	1.4E+09	2014-03-2	1.43E+09	2015-05-24T15:37:51Z			120 g	Frais	frais	Taillefine,
11859	3.03E+12	http://wo	openfood	1.43E+09	2015-04-0	1.43E+09	2015-04-0	Taille fine recette fr		120g			danone
11860	3.03E+12	http://wo	max1986	1.39E+09	2014-02-2	1.44E+09	2015-07-2	Taillefine Sp??cialit		850 g	Frais,Oper	frais,oper	Danone,T
11861	3.03E+12	http://wo	tacite	1.41E+09	2014-09-0	1.43E+09	2015-06-1	Danette (Cr??me de		500 g (4 x	Frais,Pot,	frais,pot,	Danone,D
11862	3.03E+12	http://wo	saiyanriku	1.4E+09	2014-05-0	1.43E+09	2015-06-1	Danette (Cr??me de		500 g (4 x	Frais,Pot,	frais,pot,	Danone,D
11863	3.03E+12	http://wo	openfood	1.36E+09	2013-03-2	1.44E+09	2015-07-2	Danette (Cr??me de		500 g (4 x	Frais,Pot,	frais,pot,	Danone,D
11864	3.03E+12	http://wo	penelope	1.33E+09	2012-04-0	1.43E+09	2015-06-1	Activia Fr Lait ferme		2 kg (16 x	Frais,Pot,	frais,pot,	Danone,A
11865			villecomt	France		Belgique,	len:belgiu	Belgium,F	Lait entier, fruits : fraise (20 %), mangue (10 %), abri				
11866	3.03E+12	http://wo	openfood	1.4E+09	2014-05-2	1.4E+09	2014-05-2	Activia 0%		8			Danone
11867	3.03E+12	http://wo	sighelm	1.39E+09	2014-03-0	1.44E+09	2015-08-1	Gervais Pe Fromage f		720 g [2 x	Frais,Pot,	frais,pot,	Danone,G
11868	3.03E+12	http://wo	date-limit	1.45E+09	2015-11-2	1.45E+09	2015-11-2	Petit frais			pots	pots	
11869	3.03E+12	http://wo	hommeda	1.36E+09	2012-12-0	1.45E+09	2016-01-0	Danette C Cr??me de		468 g (4 x	Frais,Pot,	frais,pot,	Danone,D
11870	3.03E+12	http://wo	delicesno	1.4E+09	2014-06-0	1.44E+09	2015-07-2	Jockey Av Dessert la		1 kg	Frais,Pot,	frais,pot,	Danone,Jc
11871	3.03E+12	http://wo	openfood	1.37E+09	2013-04-1	1.44E+09	2015-07-2	Jockey (3, Fromage k		1 kg	Frais,Pot,	frais,pot,	Danone,Jc
11872	3.03E+12	http://wo	jacob80	1.41E+09	2014-09-2	1.45E+09	2015-12-3	Activia Ecl Sp??cialit		500 g (4 x	Frais,Oper	frais,oper	Danone,A
11873	3.03E+12	http://wo	openfood	1.42E+09	2014-11-0	1.42E+09	2014-11-03T11:41:23Z						
11874	3.03E+12	http://wo	jeanbono	1.43E+09	2015-03-0	1.43E+09	2015-06-1	Activia (Se Lait ferme		2 kg [2 x	Frais,Pot,	frais,pot,	Danone,A
11875	3.03E+12	http://wo	jeanbono	1.42E+09	2015-01-1	1.44E+09	2015-07-2	Activia (N Lait ferme		2 kg [2 x	Frais,Oper	frais,oper	Danone,A
11876	3.03E+12	http://wo	openfood	1.39E+09	2014-02-2	1.39E+09	2014-02-2	Velout?? Yaourt bra		12x125g	Frais	frais	Danone
11877	3.03E+12	http://wo	openfood	1.37E+09	2013-04-2	1.37E+09	2013-05-2	Danette			1 sans surer	sans-surel	Danone
11878	3.03E+12	http://wo	beniben	1.44E+09	2015-10-1	1.44E+09	2015-10-1	Danette s Cr??me de		4x125 g	Pot,Plasti	pot,plasti	Danone,D
11879	3.03E+12	http://wo	leternnoz	1.39E+09	2013-12-2	1.39E+09	2013-12-2	ACTIVIA N Lait ferme		1kg	Frais	frais	Danone

Figure 1: Illegal Row in the dataset

## 2 Type Correction

The “code” column need to be TextType. Although it contains only number, but the prefix 0s cannot be omitted. And some code numbers are too long to store in an int variable. Indicate the dtype when parsing the csv file.

```
df = pandas.read_csv('row.clean.csv', dtype={'code': str})
```

## 3 Contact Information

If there is any problem please contact me.

Name: Yinlong Su

SBU ID: 110461173

Email: yinlsu@cs.stonybrook.edu