

CSE564 Visualization

Yinlong Su, #110461173

April 24, 2016

Row Clean Report

0 Version History

First built on April 24, 2016.

1 Country Disassemble

In the original dataset, there are three columns represents the countries of the food: countries, countries_tags and countries_en. Basically, they are just the same information represented in different forms. We choose the most understandable column 'countries_tags' as our country classify attribute.

This attribute is a multivalue attribute. So we need to disassemble the values and choose the ones we want to keep.

2 Column Report

Column 'countries_en' also has null values. The disassemble process includes: filter out the null values, group on the values of column (multivalue) and disassemble the multivalues into single countries. See table 1.

Table 1: Column Report Table

Item	Count
Invalid(Null)	211
Multivalue	369
Disassemble	149

3 Disassemble Report

Table 2 shows the report of disassembled column 'countries_en' (sorted on occurrence count).

Table 2: Disassemble Report Table

Country Name	Count	Percent
France	44069	67.3%
United Kingdom	3865	5.9%
Continued on next page		

Table 2 – continued from previous page

Country Name	Count	Percent
Spain	3534	5.4%
Germany	2934	4.48%
United States	2508	3.83%
Belgium	1715	2.62%
Switzerland	1464	2.24%
Australia	1299	1.98%
Italy	710	1.08%
Portugal	544	0.83%
Canada	482	0.74%
Denmark	363	0.55%
Brazil	309	0.47%
fr:Europe	262	0.4%
Netherlands	245	0.37%
Austria	211	0.32%
Guadeloupe	177	0.27%
Czech Republic	163	0.25%
R\xc3\xa9union	130	0.2%
China	119	0.18%
Japan	105	0.16%
Poland	103	0.16%
Tunisia	102	0.16%
French Guiana	100	0.15%
Luxembourg	90	0.14%
Morocco	78	0.12%
Russia	71	0.11%
Algeria	68	0.1%
New Zealand	66	0.1%
Romania	63	0.1%
Thailand	63	0.1%
Ireland	52	0.08%
Sweden	51	0.08%
Hong Kong	50	0.08%
Martinique	48	0.07%
New Caledonia	45	0.07%
Greece	43	0.07%
Saudi Arabia	41	0.06%
Saint Pierre and Miquelon	40	0.06%
Hungary	37	0.06%
Taiwan	34	0.05%
Cambodia	33	0.05%
Mexico	33	0.05%
Senegal	32	0.05%
United Arab Emirates	32	0.05%
Iceland	30	0.05%
fr:Quebec	29	0.04%
es:Europe	23	0.04%
India	23	0.04%
Argentina	22	0.03%

Continued on next page

Table 2 – continued from previous page

Country Name	Count	Percent
Lebanon	22	0.03%
Indonesia	21	0.03%
Israel	20	0.03%
Finland	20	0.03%
Bulgaria	20	0.03%
C\xc3\xb4te d'Ivoire	19	0.03%
South Africa	19	0.03%
Singapore	18	0.03%
Norway	17	0.03%
French Polynesia	17	0.03%
Turkey	17	0.03%
Qatar	15	0.02%
Chile	14	0.02%
Philippines	13	0.02%
Scotland	13	0.02%
Kuwait	12	0.02%
Slovenia	12	0.02%
Mauritius	11	0.02%
Kazakhstan	10	0.02%
Albania	10	0.02%
South Korea	9	0.01%
Iraq	8	0.01%
An	8	0.01%
European Union	8	0.01%
Mayotte	7	0.01%
Egypt	7	0.01%
Mongolia	7	0.01%
Latvia	7	0.01%
Cuba	7	0.01%
Malta	6	0.01%
\xe6\x97\xa5\xe6\x9c\xac	6	0.01%
Slovakia	6	0.01%
Aruba	6	0.01%
Vietnam	6	0.01%
Malaysia	5	0.01%
Cook Islands	5	0.01%
Burkina Faso	5	0.01%
Andorra	5	0.01%
Togo	5	0.01%
Venezuela	5	0.01%
Iran	4	0.01%
Croatia	4	0.01%
Serbia	4	0.01%
Ukraine	4	0.01%
Gabon	4	0.01%
Colombia	3	0.0%
Belarus	3	0.0%
Lithuania	3	0.0%

Continued on next page

Table 2 – continued from previous page

Country Name	Count	Percent
El Salvador	3	0.0%
Mozambique	3	0.0%
Peru	3	0.0%
Ecuador	3	0.0%
fr:Angleterre	2	0.0%
Estonia	2	0.0%
Dominican Republic	2	0.0%
Kenya	2	0.0%
Republic of the Congo	2	0.0%
Armenia	2	0.0%
Guinea	2	0.0%
Other-\xe6\x97\xa5\xe6\x9c\xac	2	0.0%
Azerbaijan	2	0.0%
Monaco	2	0.0%
Brunei	2	0.0%
Vanuatu	2	0.0%
Ghana	2	0.0%
Tanzania	2	0.0%
Moldova	2	0.0%
Costa Rica	2	0.0%
Mauritania	2	0.0%
Polska	1	0.0%
Other-\xd8\xa7\xd9\x84\xd8\xb9\xd8\xb1\xd8\xa7\xd9\x82	1	0.0%
Thailanda	1	0.0%
Belize	1	0.0%
Hawaii	1	0.0%
Other-other-\xe6\x97\xa5\xe6\x9c\xac	1	0.0%
Bangladesh	1	0.0%
Panama	1	0.0%
Yemen	1	0.0%
Syria	1	0.0%
Republique-de-chine	1	0.0%
Nederland	1	0.0%
Irlande	1	0.0%
Sri Lanka	1	0.0%
Georgia	1	0.0%
Cyprus	1	0.0%
Burundi	1	0.0%
Guatemala	1	0.0%
Other-turquie	1	0.0%
Macau	1	0.0%
Mali	1	0.0%
Pakistan	1	0.0%
Bahrain	1	0.0%
Sint Maarten	1	0.0%
fr:\xe6\x97\xa5\xe6\x9c\xac	1	0.0%
Other-japon	1	0.0%
Democratic Republic of the Congo	1	0.0%

Continued on next page

Table 2 – continued from previous page

Country Name	Count	Percent
Niger	1	0.0%
Maldives	1	0.0%
Jordan	1	0.0%

There are still some dirty data we need to clean.

4 Contact Information

If there is any problem please contact me.

Name: Yinlong Su

SBU ID: 110461173

Email: yinlsu@cs.stonybrook.edu