

انتخاب، ساده سازی و متناسب سازی داده ها

۱-انتخاب چند کشور که داده های کاملی در مجموع دادگان دارند (حداقل ۵ کشور) ۲-فیلتر و ساده سازی داده های موجود متناسب هدف پژوهشی ۳- تطبیق مجموعه دادگان با یکدیگر

برای انتخاب ۵ کشور که دارای بیشترین اطلاعات ثبتی و زمانی باشند، ابتدا دیتاست اول (کودها) را فیلتر و مورد تجزیه و تحلیل قرار داریم. به عنوان مثال برای کشور افغانستان به اطلاعات زیر دست پیدا کردیم :

1- Afghanistan 322 All

نام کشور

از همه سال ها (۲۰۰۲-۲۰۱۷) برای این کشور اطلاعات داریم

تعداد رکوردهای ثبت شده برای این کشور

با همین منطق، کل کشورها را فیلتر می کنیم تا بعد از آن به انتخاب ۵ کشور حداکثری اقدام کنیم :

1- Afghanistan 322 All
2- Albania 854 All
3- Algeria 972 All
4- Angola 781 All
5- Antigua and Barbuda 387 All
6- Argentina 1408 All *
7- Armenia 583 All
8- Australia 1461 All *
9- Austria 1240 All
10- Azerbaijan 822 All
11- Bahamas 544 2002-2015
12- Bahrain 706 All
13- Bangladesh 862 All
14- Barbados 731 All
15- Belarus 1170 All
16- Belgium 1306 All
17- Belize 595 All
18- Benin 311 All
19- Bermuda 440 All
20- Bhutan 410 2002-2016
21- Bolivia (Plurinational State of) 828 All
22- Bosnia and Herzegovina 952 All
23- Botswana 890 All
24- Brazil 1779 All
25- Brunei Darussalam 678 All
26- Bulgaria 1327 All
27- Burkina Faso 647 All
28- Burundi 572 All
29- Cambodia 444 2002-2016
30- Cameroon 1086 All
31- Canada 1512 All
32- Central African Republic 164 All
33- Chile 1393 All
34- China, Hong Kong SAR 880 All
35- China, Macao SAR 281 All
36- China, mainland 1303 All
37- China, Taiwan Province of 1329 All
38- Colombia 1337 All
39- Congo 288 2004-2017
40- Cook Islands 449 All

41- Costa Rica 1207 All
42- Côte d'Ivoire 1060 All
43- Croatia 1472 All
44- Cuba 809
45- Cyprus 1085 All
46- Czechia 1507 All
47- Democratic Republic of the Congo 738 All
48- Denmark 1621 All
49- Dominica 339 All
50- Dominican Republic 1098 All
51- Ecuador 1158 All
52- Egypt 1274 All
53- El Salvador 1083 All
54- Eritrea 184 All
55- Estonia 1154 All
56- Ethiopia 669 All
57- Fiji 766 All
58- Finland 1334 All
59- France 1466 All
60- French Polynesia 527 2002-2015
61- Gabon 581 All
62- Gambia 288 All
63- Georgia 747 All
64- Germany 1327 All
65- Ghana 853 All
66- Greece 1477 All
67- Guatemala 1241 All
68- Guinea 495 All
69- Guyana 642 All
70- Honduras 909 All
71- Hungary 1380 All
72- Iceland 837 All
73- India 1514 All
74- Indonesia 1285 All
75- Iran (Islamic Republic of) 1300 All
76- Iraq 677 All
77- Ireland 1409 All
78- Israel 1291
79- Italy 1478 All
80- Jamaica 782 All

81- Japan 1382 All
82- Jordan 1107
83- Kazakhstan 1136 All
84- Kenya 1364 All
85- Kuwait 569
86- Kyrgyzstan 668
87- Latvia 1407
88- Lebanon 1039 All
89- Libya 683 All
90- Lithuania 1483 All
91- Luxembourg 1455 All
92- Madagascar 685 All
93- Malawi 978 All
94- Malaysia 1409 All
95- Maldives 619 All
96- Mali 546 All
97- Malta 704
98- Marshall Islands 58 All
99- Mauritius 1255 All
100- Mexico 1462 All
101- Mongolia 454 All
102- Montenegro 457 2006-2017
103- Morocco 1287 All
104- Mozambique 868 All
105- Myanmar 656 All
106- Namibia 1114 All
107- Nauru 266 2003-2017
108- Nepal 550 2003-2017
109- Netherlands 1380 All
110- New Caledonia 722 2002-2016
111- New Zealand 1371 All
112- Nicaragua 1102 All
113- Niger 548 2002-2016
114- Nigeria 606 All
115- North Macedonia 1107 All
116- Norway 1414 All
117- Oman 906 All
118- Pakistan 1119 2003-2017
119- Panama 1077 All
120- Papua New Guinea 694 All

121- Paraguay 745 All
122- Peru 1370 All
123- Philippines 1159 All
124- Poland 1520 All
125- Portugal 1508 All
126- Qatar 683 All
127- Republic of Korea 1432 All
128- Republic of Moldova 790 All
129- Romania 1264 All
130- Russian Federation 1386 All
131- Rwanda 611 2002-2016
132- Saint Kitts and Nevis 533 All
133- Saint Lucia 472 All
134- Samoa 363 All
135- Saudi Arabia 1238 All
136- Senegal 996 All
137- Serbia 1048 All
138- Serbia and Montenegro 106 2002-2005
139- Seychelles 602 All
140- Singapore 1169 All
141- Slovakia 1313 All
142- Slovenia 1626 All
143- South Africa 1359 All
144- Spain 1516 All
145- Sri Lanka 1154 All
146- Sudan 289 2012-2017
147- Sudan (former) 315 2002-2011
148- Suriname 484 All
149- Sweden 1575 All
150- Switzerland 1472 All
151- Syrian Arab Republic 748 All
152- Tajikistan 359 All
153- Thailand 1296 All
154- Togo 474 All
155- Tonga 238 2003-2017
156- Trinidad and Tobago 877 All
157- Tunisia 1287 All
158- Turkey 1566 All
159- Uganda 949 All
160- Ukraine 1380 All

161- United Arab Emirates 1406 All
162- United Kingdom 1388 All
163- United Republic of Tanzania 1148 All
164- United States of America 1592 All
165- Uruguay 1186 All
166- Uzbekistan 919 All
167- Venezuela (Bolivarian Republic of) 988 All
168- Viet Nam 1198 All
169- Yemen 562 All
170- Zambia 1057 All
171- Zimbabwe 937 All

بعد از مشخص کردن تعداد رکوردهای ثبت شده برای هر کشور و در نظر گرفتن رزولوشن زمانی، کشورها را سورت کردیم:

```
24- Brazil 1779 All
142- Slovenia 1626 All
48- Denmark 1621 All
164- United States of America 1592 All
149- Sweden 1575 All
158- Turkey 1566 All
124- Poland 1520 All
144- Spain 1516 All
73- India 1514 All
46- Czechia 1507 All
90- Lithuania 1483 All
79- Italy 1478 All
66- Greece 1477 All
150- Switzerland 1472 All
43- Croatia 1472 All
59- France 1466 All
100- Mexico 1462 All
```

اکنون کشورهای سورت شده را با دیتاست دوم (میزان تولید محصولات زراعی) مطابقت دادیم تا ببینیم برای کدام یک از این کشورها، در مجموعه دوم اطلاعات داریم که به لیست زیر رسیدیم:

```
24- Brazil 1779 All + 444
142- Slovenia 1626 All + No Data
48- Denmark 1621 All + No Data
164- United States of America 1592 All + 444
149- Sweden 1575 All + No Data
158- Turkey 1566 All + 444
124- Poland 1520 All + No Data
144- Spain 1516 All + No Data
73- India 1514 All + 444
46- Czechia 1507 All + No Data
90- Lithuania 1483 All + No Data
79- Italy 1478 All + No Data
66- Greece 1477 All + No Data
150- Switzerland 1472 All + No Data
43- Croatia 1472 All + No Data
59- France 1466 All + No Data
100- Mexico 1462 All + 444
```

همان طور که در تصویر بالا پیداست برای بعضی کشورها در مجموعه دوم داده ای نداشتیم و برای برخی از کشورها هم ۴۴۴ دیتا داشتیم. و نهایتاً با در نظر گرفتن همه این موارد ۵ کشور زیر را انتخاب کردیم:

```
24- Brazil 1779 All + 444 (BRA)
164- United States of America 1592 All + 444 (USA)
158- Turkey 1566 All + 444 (TUR)
73- India 1514 All + 444 (IND)
100- Mexico 1462 All + 444 (MEX)
```

در گام بعد ستون های غیر ضرور را از دیتاست اول حذف کردیم تا داده ها ساده تر شوند:

1	Area Code	Area	Item Code	Item	Element Code	Element	Year Code	Year	Unit	Value	Flag

در گام آخر برای تطبیق دیتاست ها و سینک کردن آن ها به هم از اشتراک آن ها استفاده کردیم (سال ثبت و نام کشور)

1770	21	Brazil	4006	Urea and	2012	2012	1000 US\$	356.41			
1771	21	Brazil	4006	Urea and	2013	2013	1000 US\$	713.26			
1772	21	Brazil	4006	Urea and	2014	2014	1000 US\$	711.13			
1773	21	Brazil	4006	Urea and	2015	2015	1000 US\$	1090.17			
1774	21	Brazil	4006	Urea and	2016	2016	1000 US\$	1920.66			
1775	21	Brazil	4006	Urea and	2017	2017	1000 US\$	1922.86			
1776	21	Brazil	4006	Urea and	2013	2013	tonnes	0			
1777	21	Brazil	4006	Urea and	2014	2014	tonnes	0			
1778	21	Brazil	4006	Urea and	2015	2015	tonnes	0			
1779	21	Brazil	4006	Urea and	2016	2016	tonnes	0			
1780	21	Brazil	4006	Urea and	2017	2017	tonnes	0			
1781	21	Brazil			2002	2002	RICE	TONNE_H	2.2443		
1782	21	Brazil			2002	2002	WHEAT	TONNE_H	1.519569		
1783	21	Brazil			2002	2002	RICE	THND_TOI	7225.748		
1784	21	Brazil			2002	2002	RICE	THND_HA	3219.6		
1785	21	Brazil			2002	2002	WHEAT	THND_HA	2044		
1786	21	Brazil			2002	2002	WHEAT	THND_TOI	3106		
1787	21	Brazil			2002	2002	SOYBEAN	THND_HA	18524.77		
1788	21	Brazil			2002	2002	SOYBEAN	THND_TOI	51919.44		
1789	21	Brazil			2002	2002	MAIZE	TONNE_H	3.048088		
1790	21	Brazil			2002	2002	MAIZE	THND_HA	11791		

ستون مشترک دیتاست ها

ستون مشترک دیتاست ها

*** فایل نهایی به نام FinalDataset ارسال گردید.**