

Deforestation

Data Science for Business Analytics Project

Mathieu Maréchal & Martina Kofler

14 novembre, 2021

Introduction

- Overview and Motivation
- Related Work
- Research questions

Data

Dataset index and URLs:

- AgriLandData : <https://data.worldbank.org/indicator/AG.LND.AGRI.ZS>
- ForestAreaData : <https://data.worldbank.org/indicator/AG.LND.FRST.K2>
- PopGrowth : <https://data.worldbank.org/indicator/SP.POP.GROWL>
- AgriVarData : <https://data.worldbank.org/topic/trade>
- SoybeansData : <https://www.fao.org/faostat/en/#search/soy>
- TradeOpenessData : <https://ourworldindata.org/grapher/trade-openness>
- ResourcesExtractionData : <https://www.eia.gov/international/data/world>
- LandAreaData : <https://data.worldbank.org/indicator/AG.LND.TOTL.K2>

Agricultural Land

The AgriLandData dataset shows the total area of agricultural land in square km per country/region for the years between 1960 and 2020. In this project we are just interested in the years starting from 2000.

Country	Year	AgriLand
Aruba	2000	11.1
Aruba	2001	11.1
Aruba	2002	11.1
Aruba	2003	11.1
Aruba	2004	11.1
Aruba	2005	11.1
Aruba	2006	11.1

To know how the area of agricultural land changed over time we created two new columns that contain information about the change to the year before, one in square km and one in percentage.

Country	Year	AgriLand	Change agricultural land	Change agricultural land %
Aruba	2019	NA	NA	NA
Aruba	2020	NA	NA	NA
Africa Eastern and Southern	2000	42.5	NA	NA
Africa Eastern and Southern	2001	42.6	-0.126	-0.295
Africa Eastern and Southern	2002	42.8	-0.122	-0.286
Africa Eastern and Southern	2003	43.0	-0.233	-0.545
Africa Eastern and Southern	2004	42.8	0.163	0.380
Africa Eastern and Southern	2005	43.0	-0.126	-0.293

Forest Area

The AgriLandData dataset shows the total forest area in square km per country/region for the years between 1960 and 2020. In this project we are just interested in the years starting from 2000.

Country	Year	ForestArea
Africa Western and Central	2016	20.2
Africa Western and Central	2017	20.1
Africa Western and Central	2018	20.0
Africa Western and Central	2019	19.9
Africa Western and Central	2020	19.8
Angola	2000	62.3
Angola	2001	61.9
Angola	2002	61.4

Also for this dataset we created two new columns that contain information about the change to the year before, one in square km and one in percentage.

Country	Year	ForestArea	Deforestation	Deforestation %
Africa Western and Central	2016	20.2	0.082	0.403
Africa Western and Central	2017	20.1	0.081	0.403
Africa Western and Central	2018	20.0	0.085	0.425
Africa Western and Central	2019	19.9	0.085	0.425
Africa Western and Central	2020	19.8	0.084	0.424
Angola	2000	62.3	NA	NA
Angola	2001	61.9	0.445	0.714
Angola	2002	61.4	0.445	0.719

Agricultural raw material exports, Merchandise Exports, Food exports

The dataset we took includes information about 149 indicators for 237 countries from 1960 to 2020. To get an impression about the data we show the 7 rows and 7 columns from the dataset.

Country	Indicator Name	1960	1961	1962	1963	1964
Aruba	Merchandise exports by the reporting economy, residual (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise exports to low- and middle-income economies in Sub-Saharan Africa (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise exports to low- and middle-income economies in South Asia (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise exports to low- and middle-income economies in Middle East & North Africa (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise exports to low- and middle-income economies in Latin America & the Caribbean (% of total merchandise exports)	NA	NA	NA	NA	NA
Aruba	Merchandise	NA	NA	NA	NA	NA

Before selecting the variables we need we changed the structure of the dataset so that it has the year and the different indicator names as column names and we also just selected data from 2000 and above.

Country	Year	Travel services (% of commercial service exports)	Transport services (% of commercial service exports)	High-technology exports (% of manufactured exports)
Aruba	2000	52.1	5.32	NA
Aruba	2001	45.3	2.86	NA
Aruba	2002	69.1	2.96	NA
Aruba	2003	53.9	2.44	NA
Aruba	2004	85.4	4.29	NA
Aruba	2005	85.0	4.39	NA
Aruba	2006	82.5	4.15	NA

Form all this variables just two are of interest for our project. Therefore we selected just these indicators and created a new dataset for each of them. The result was:

A dataset that shows the percentage of food exports from total merchandise exports in US\$ per country and year

Country	Year	Food exports (% of merchandise exports)
Aruba	2000	49.96
Aruba	2001	45.08
Aruba	2002	44.09
Aruba	2003	46.87
Aruba	2004	35.86
Aruba	2005	1.18
Aruba	2006	34.38

and a dataset that shows the percentage agricultural raw material exports from total merchandise exports in US\$ per country and year.

Country	Year	Agricultural raw materials exports (% of merchandise exports)
Aruba	2000	0.737
Aruba	2001	0.549
Aruba	2002	0.993
Aruba	2003	1.172
Aruba	2004	1.249
Aruba	2005	0.317
Aruba	2006	1.096

LandArea

This dataset shows how much squared kilometer every country has, this makes it possible identify how much of percent of this area is forest and how much of percent of the land is agricultural land.

Country	Year	Land Area in sq km
Aruba	2000	180
Aruba	2001	180
Aruba	2002	180
Aruba	2003	180
Aruba	2004	180
Aruba	2005	180
Aruba	2006	180

Extraction of Minerals Data

This data set contains the different amount of extraction of different sources of energy taken worldwide. Although it was a difficult data set to find for the whole world, it was also a hard one to tidy and here is why.

After running the four first line of code in order to delete the unnecessary information lines and name the columns to be able to use it, we end up to this stage:

	Extr	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
Algeria		168087682	1828384085	868788989	90999293	949596	9799	9920	1010	1020	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050	1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064	1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092	1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106	1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190	1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302	1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316	1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330	1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358	1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	1426	1427	1428	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439	1440	1441	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498	1499	1500	1501	1502	1503	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515	1516	1517	1518	1519	1520	1521	1522	1523	1524	1525	1526	1527	1528	1529	1530	1531	1532	1533	1534	1535	1536	1537	1538	1539	1540	1541	1542	1543	1544	1545	1546	1547	1548	1549	1550	1551	1552	1553	1554	1555	1556	1557	1558	1559	1560	1561	1562	1563	1564	1565	1566	1567	1568	1569	1570	1571	1572	1573	1574	1575	1576	1577	1578	1579	1580	1581	1582	1583	1584	1585	1586	1587	1588	1589	1590	1591	1592	1593	1594	1595	1596	1597	1598	1599	1600	1601	1602	1603	1604	1605	1606	1607	1608	1609	1610	1611	1612	1613	1614	1615	1616	1617	1618	1619	1620	1621	1622	1623	1624	1625	1626	1627	1628	1629	1630	1631	1632	1633	1634	1635	1636	1637	1638	1639	1640	1641	1642	1643	1644	1645	1646	1647	1648	1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661	1662	1663	1664	1665	1666	1667	1668	1669	1670	1671	1672	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682	1683	1684	1685	1686	1687	1688	1689	1690	1691	1692	1693	1694	1695	1696	1697	1698	1699	1700	1701	1702	1703	1704	1705	1706	1707	1708	1709	1710	1711	1712	1713	1714	1715	1716	1717	1718	1719	1720	1721	1722	1723	1724	1725	1726	1727	1728	1729	1730	1731	1732	1733	1734	1735	1736	1737	1738	1739	1740	1741	1742	1743	1744	1745	1746	1747	1748	1749	1750	1751	1752	1753	1754	1755	1756	1757	1758	1759	1760	1761	1762	1763	1764	1765	1766	1767	1768	1769	1770	1771	1772	1773	1774	1775	1776	1777	1778	1779	1780	1781	1782	1783	1784	1785	1786	1787	1788	1789	1790	1791	1792	1793	1794	1795	1796	1797	1798	1799	1800	1801	1802	1803	1804	1805	1806	1807	1808	1809	1810	1811	1812	1813	1814	1815	1816	1817	1818	1819	1820	1821	1822	1823	1824	1825	1826	1827	1828	1829	1830	1831	1832	1833	1834	1835	1836	1837	1838	1839	1840	1841	1842	1843	1844	1845	1846	1847	1848	1849	1850	1851	1852	1853	1854	1855	1856	1857	1858	1859	1860	1861	1862	1863	1864	1865	1866	1867	1868	1869	1870	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342

Now we can see our problem: The name of the country appear as a title that is in the same column as the

variables names we need. We are therefore at this stage unable to pivot our tibble. In order to solve this problem we created a new column “Country” where we take only one row every eight of the first column and copy it for all the variable. Then we deleted the row corresponding at the country name that is now store in the column “Country” and obtain something like this:

Now we are finally able to pivot our tibble to have a tidy dataset usable for the rest of our analysis:

Country	Year	Production Coal (quad Btu)	Natural gas (quad Btu)	Petroleum and other liquids (quad Btu)	Nuclear, renew- ables, and other (quad Btu)	Nuclear (quad Btu)	Renewables and other (quad Btu)
World	1980	296.21435257	102.05404252	99.77610455	9133.11110888	28.03702555	56.58558046
World	1981	291.26914348	102.14027573	99.57353674	125.43894942	20.81389983	52.92153469
World	1982	290.16459398	102.33134119	99.55226611	119.76280547	20.68922428	50.67686427
World	1983	293.08843019	102.25167485	99.55951591	119.26919638	20.97914475	50.82484409
World	1984	308.88419728	106.32244006	108.77259753	222.55382666	20.98229077	52.99150787
World	1985	316.38674879	106.02187139	109.50052282	211.13478963	19.32524919	57.29561514
World	1986	326.69952978	108.83261653	109.48286482	126.54884517	18.83053793	56.28769514

Population Growth Data

Country	NA	Year	PopulationGrowth
Aruba	NA	2000	2.065
Aruba	NA	2001	2.205
Aruba	NA	2002	2.236
Aruba	NA	2003	2.108
Aruba	NA	2004	1.765
Aruba	NA	2005	1.292
Aruba	NA	2006	0.799

Openness to Trade Data

Country	Year	Ratio.of.exports.and.imports.to.GDP.....PWT.9.1..
Afghanistan	2000	109.2
Afghanistan	2001	109.4
Afghanistan	2002	89.4
Afghanistan	2003	128.7
Afghanistan	2004	110.0
Afghanistan	2005	99.6
Afghanistan	2006	94.4

Soybeans dataset

Country	Year	Gross Production Value Soybeans
Albania	2000	292
Albania	2001	389
Albania	2002	200
Albania	2003	287
Albania	2004	272
Albania	2005	491
Albania	2006	285

Merged Data

Country	Year	Agricultural land	Change agri-land	Grazing cul- land	Roads ma- ri-	Exports ports of mer- chan- dise (% ex- of ports)	Dates res	Deforestation % Area in sq km	Land NA	Population	Gross Pro- duction Value Soy- beans	National Roth (quad Btu)	Capital Btu) (quad Btu)	Natural gas and re- Btu) liq- uids and other (quad Btu)	Pulp and Kraft (quad Btu)	New and other (quad Btu)	Net F	9.1..201		
Albania	2000	41.8	NA	NA	5.97	6.63	28.1	NA	NA	27400	NA	-	292	63.5	0.0608	3.3082	270880	1890587	237107	463194148
Albania	2001	41.6	0.182	0.437	5.54	5.79	28.1	-	-	27400	NA	-	389	66.5	0.0513	3.6007	356800	1065822	308504	136381827
Albania	2002	41.6	-	-	6.52	3.56	28.2	-	-	27400	NA	-	200	68.5	0.0512	5.6013	308002	1065822	308504	1357711521
Albania	2003	40.9	0.693	1.667	5.14	5.63	28.2	-	-	27400	NA	-	287	67.0	0.0689	0.0653	470002	2063557	1087960	30322879625
Albania	2004	40.9	-	-	4.49	5.80	28.3	-	-	27400	NA	-	272	67.0	0.0712	8846	101857	2063557	1087960	30322879625
Albania	2005	39.3	1.642	4.011	113.94	9.77	28.3	-	-	27400	NA	-	491	70.9	0.0708	10058	110800	2063557	1087960	30322879625
Albania	2006	40.9	-	-	13.127	9.4	28.4	-	-	27400	NA	-	285	74.3	0.0689	5249	108203	2063557	1087960	30322879625

- Sources
- Description
- Wrangling/cleaning
- Spotting mistakes and missing data (could be part of EDA too)
- Listing anomalies and outliers (could be part of EDA too)

Analysis

- Answers to the research questions

- Different methods considered
- Competing approaches
- Justifications

Conclusion

- Take home message
- Limitations
- Future work?
, ‘eda.Rmd’