**Comparison between tables**

Table 1: pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide

CREATE TABLE IF NOT EXISTS long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide

(

idlong bigint,

firstname character varying(15) COLLATE pg\_catalog."default",

lastname character varying(15) COLLATE pg\_catalog."default",

dob timestamp without time zone,

sex smallint,

villagename\_1 text COLLATE pg\_catalog."default",

villagename\_2 text COLLATE pg\_catalog."default",

villagename\_3 text COLLATE pg\_catalog."default",

villagename\_4 text COLLATE pg\_catalog."default",

villagename\_5 text COLLATE pg\_catalog."default",

subvillagename\_1 text COLLATE pg\_catalog."default",

subvillagename\_2 text COLLATE pg\_catalog."default",

subvillagename\_3 text COLLATE pg\_catalog."default",

subvillagename\_4 text COLLATE pg\_catalog."default",

subvillagename\_5 text COLLATE pg\_catalog."default",

hhfname character varying(15) COLLATE pg\_catalog."default",

hhsname character varying(15) COLLATE pg\_catalog."default",

hhnumber bigint

)

TABLESPACE pg\_default;

ALTER TABLE IF EXISTS long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide

OWNER to postgres;

## Original Data Shape: 94825 rows, 7 columns

**Table 2: synthetic\_CTGAN\_distributed\_dob\_v1**

CREATE TABLE IF NOT EXISTS synthetic."synthetic\_CTGAN\_distributed\_dob\_v1"

(

idlong bigint,

firstname text COLLATE pg\_catalog."default",

lastname text COLLATE pg\_catalog."default",

sex text COLLATE pg\_catalog."default",

villagename text COLLATE pg\_catalog."default",

subvillagename text COLLATE pg\_catalog."default",

year\_of\_birth integer,

month\_of\_birth bigint,

day\_of\_birth bigint,

date\_of\_birth timestamp without time zone

)

TABLESPACE pg\_default;

ALTER TABLE IF EXISTS synthetic."synthetic\_CTGAN\_distributed\_dob\_v1"

OWNER to postgres;

## Synthetic Data Shape: 94690 rows, 10 columns

Comparison:

|  |  |
| --- | --- |
| Table: synthetic\_CTGAN\_distributed\_dob\_v1 | Table: long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide |
| **Total Rows** | |
| SELECT count(\*)  FROM synthetic.synthetic\_CTGAN\_distributed\_dob\_v1 | SELECT count(\*)  FROM long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide |
| **Sex Distribution** | |
| SELECT sex, count(\*)  FROM synthetic.synthetic\_CTGAN\_distributed\_dob\_v1  GROUP BY sex  ORDER BY sex | SELECT sex, count(\*)  FROM long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide  GROUP BY sex  ORDER BY sex |
| **Villagename Distribution** | |
| SELECT villagename, count(\*)  FROM synthetic.synthetic\_CTGAN\_distributed\_dob\_v1  GROUP BY villagename  ORDER BY villagename | SELECT villagename\_1, count(\*)  FROM long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide  GROUP BY villagename\_1  ORDER BY villagename\_1 |
| **Sub Villagename Distribution** | |
| SELECT subvillagename, count(\*)  FROM synthetic.synthetic\_CTGAN\_distributed\_dob\_v1  GROUP BY subvillagename  ORDER BY subvillagename   |  |  | | --- | --- | | subvillagename | count | |  | 4081 | | Bukala | 231 | | Bukandwe | 674 | | Bukelebe | 933 | | Changabe | 762 | | Igadya | 586 | | Igekemaja | 1019 | | Igeye | 1731 | | Igudija A | 2193 | | Igudija B | 1958 | | Igumo | 1176 | | Igunga | 812 | | Ihala | 1591 | | Ihayabuyaga | 2927 | | Ikangabuta | 2812 | | Ikengele | 1271 | | Ikulicha | 828 | | Ilagaja | 1465 | | Ilangale | 2415 | | Ilendeja | 3355 | | Iseni Bondeni | 3849 | | Iseni Mlimani | 3668 | | Kanami | 3009 | | Kanyama | 2489 | | Kigungumuli | 1788 | | Kimaga | 789 | | Kisesa Kati | 6073 | | Kisha | 1972 | | Kitumba A | 2612 | | Kitumba B | 2172 | | Lumve | 1420 | | Mahilinga | 1021 | | Mondo | 968 | | Mwabuki | 986 | | Mwadubi | 890 | | Mwamanyili | 835 | | Mwaneneka | 559 | | Ng'wabongoso | 605 | | Ng'wahuli | 1447 | | Ngï¿½wandulu | 3548 | | Ngï¿½wangï¿½halanga | 4852 | | Njicha | 3229 | | Nkola | 629 | | None | 200 | | Nyamikoma | 649 | | Nyang'hulukulu | 1365 | | Nyanï¿½helela | 2214 | | Welamasonga | 3570 | | Wita | 4462 | | SELECT subvillagename\_1, count(\*)  FROM long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide  GROUP BY subvillagename\_1  ORDER BY subvillagename\_1   |  |  | | --- | --- | | subvillagename\_1 | count | | NULL | 16 | | Bukala | 810 | | Bukandwe | 1093 | | Bukelebe | 1769 | | Changabe | 621 | | Igadya | 577 | | Igekemaja | 1380 | | Igeye | 865 | | Igudija A | 2240 | | Igudija B | 3364 | | Igumo | 1228 | | Igunga | 1832 | | Ihala | 1446 | | Ihayabuyaga | 4822 | | Ikangabuta | 983 | | Ikengele | 1706 | | Ikulicha | 799 | | Ilagaja | 1717 | | Ilangale | 1105 | | Ilendeja | 1836 | | Iseni Bondeni | 4516 | | Iseni Mlimani | 1968 | | Kanami | 1402 | | Kanyama | 3152 | | Kigungumuli | 1339 | | Kimaga | 1527 | | Kisesa Kati | 8178 | | Kisha | 913 | | Kitumba A | 1320 | | Kitumba B | 2350 | | Lumve | 1337 | | Mahilinga | 1622 | | Mondo | 1264 | | Mwabuki | 2073 | | Mwadubi | 716 | | Mwamanyili | 1254 | | Mwaneneka | 916 | | Ng'wabongoso | 1457 | | Ng'wahuli | 1630 | | Ngï¿½wandulu | 4764 | | Ngï¿½wangï¿½halanga | 5881 | | Njicha | 1601 | | Nkola | 1104 | | Nyamikoma | 744 | | Nyang'hulukulu | 2242 | | Nyanï¿½helela | 1031 | | Welamasonga | 1391 | | Wita | 6924 | |
| **Year of Birth Distribution** | |
| SELECT year\_of\_birth, count(\*)  FROM synthetic."synthetic\_CTGAN\_distributed\_dob\_v1"  GROUP BY year\_of\_birth  ORDER BY year\_of\_birth   |  |  | | --- | --- | | year\_of\_birth | count | | 1901 | 13 | | 1902 | 4 | | 1903 | 8 | | 1904 | 15 | | 1905 | 14 | | 1906 | 21 | | 1907 | 30 | | 1908 | 32 | | 1909 | 35 | | 1910 | 44 | | 1911 | 50 | | 1912 | 52 | | 1913 | 68 | | 1914 | 68 | | 1915 | 76 | | 1916 | 72 | | 1917 | 79 | | 1918 | 105 | | 1919 | 87 | | 1920 | 103 | | 1921 | 102 | | 1922 | 90 | | 1923 | 85 | | 1924 | 80 | | 1925 | 106 | | 1926 | 100 | | 1927 | 113 | | 1928 | 114 | | 1929 | 117 | | 1930 | 100 | | 1931 | 127 | | 1932 | 129 | | 1933 | 145 | | 1934 | 147 | | 1935 | 149 | | 1936 | 150 | | 1937 | 158 | | 1938 | 138 | | 1939 | 129 | | 1940 | 124 | | 1941 | 147 | | 1942 | 136 | | 1943 | 109 | | 1944 | 124 | | 1945 | 150 | | 1946 | 140 | | 1947 | 159 | | 1948 | 185 | | 1949 | 222 | | 1950 | 220 | | 1951 | 276 | | 1952 | 278 | | 1953 | 328 | | 1954 | 347 | | 1955 | 376 | | 1956 | 438 | | 1957 | 493 | | 1958 | 531 | | 1959 | 606 | | 1960 | 607 | | 1961 | 643 | | 1962 | 652 | | 1963 | 691 | | 1964 | 608 | | 1965 | 632 | | 1966 | 642 | | 1967 | 615 | | 1968 | 712 | | 1969 | 850 | | 1970 | 958 | | 1971 | 1154 | | 1972 | 1330 | | 1973 | 1335 | | 1974 | 1284 | | 1975 | 1383 | | 1976 | 1434 | | 1977 | 1435 | | 1978 | 1420 | | 1979 | 1369 | | 1980 | 1283 | | 1981 | 1293 | | 1982 | 1409 | | 1983 | 1524 | | 1984 | 1709 | | 1985 | 1887 | | 1986 | 1837 | | 1987 | 1850 | | 1988 | 1827 | | 1989 | 1922 | | 1990 | 1963 | | 1991 | 2066 | | 1992 | 2147 | | 1993 | 2244 | | 1994 | 2159 | | 1995 | 2160 | | 1996 | 2072 | | 1997 | 2217 | | 1998 | 2172 | | 1999 | 2247 | | 2000 | 2193 | | 2001 | 2200 | | 2002 | 2261 | | 2003 | 2219 | | 2004 | 2124 | | 2005 | 1979 | | 2006 | 1927 | | 2007 | 1852 | | 2008 | 2050 | | 2009 | 2193 | | 2010 | 2151 | | 2011 | 2083 | | 2012 | 1568 | | 2013 | 972 | | 2014 | 532 | | SELECT EXTRACT(YEAR FROM dob), count(\*)  FROM long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide  GROUP BY EXTRACT(YEAR FROM dob)  ORDER BY EXTRACT(YEAR FROM dob   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  | | --- | --- | | Year | Count | | 1901 | 2 | | 1904 | 11 | | 1905 | 10 | | 1906 | 7 | | 1907 | 2 | | 1908 | 10 | | 1909 | 5 | | 1910 | 18 | | 1911 | 9 | | 1912 | 11 | | 1913 | 2 | | 1914 | 37 | | 1915 | 13 | | 1916 | 21 | | 1917 | 12 | | 1918 | 21 | | 1919 | 51 | | 1920 | 69 | | 1921 | 21 | | 1922 | 50 | | 1923 | 18 | | 1924 | 97 | | 1925 | 45 | | 1926 | 74 | | 1927 | 53 | | 1928 | 63 | | 1929 | 68 | | 1930 | 161 | | 1931 | 51 | | 1932 | 91 | | 1933 | 72 | | 1934 | 162 | | 1935 | 77 | | 1936 | 214 | | 1937 | 89 | | 1938 | 124 | | 1939 | 121 | | 1940 | 209 | | 1941 | 86 | | 1942 | 163 | | 1943 | 146 | | 1944 | 241 | | 1945 | 172 | | 1946 | 184 | | 1947 | 151 | | 1948 | 234 | | 1949 | 247 | | 1950 | 300 | | 1951 | 135 | | 1952 | 286 | | 1953 | 164 | | 1954 | 470 | | 1955 | 219 | | 1956 | 329 | | 1957 | 198 | | 1958 | 436 | | 1959 | 374 | | 1960 | 425 | | 1961 | 576 | | 1962 | 508 | | 1963 | 424 | | 1964 | 741 | | 1965 | 542 | | 1966 | 598 | | 1967 | 531 | | 1968 | 891 | | 1969 | 725 | | 1970 | 1090 | | 1971 | 594 | | 1972 | 1289 | | 1973 | 954 | | 1974 | 1811 | | 1975 | 1225 | | 1976 | 1494 | | 1977 | 1359 | | 1978 | 1630 | | 1979 | 1724 | | 1980 | 1904 | | 1981 | 1279 | | 1982 | 2065 | | 1983 | 1521 | | 1984 | 2093 | | 1985 | 1573 | | 1986 | 2083 | | 1987 | 1653 | | 1988 | 2044 | | 1989 | 1697 | | 1990 | 1934 | | 1991 | 1673 | | 1992 | 2090 | | 1993 | 1924 | | 1994 | 2084 | | 1995 | 2145 | | 1996 | 2138 | | 1997 | 2031 | | 1998 | 2057 | | 1999 | 2224 | | 2000 | 2329 | | 2001 | 2175 | | 2002 | 2468 | | 2003 | 2480 | | 2004 | 2453 | | 2005 | 2440 | | 2006 | 2392 | | 2007 | 2387 | | 2008 | 2300 | | 2009 | 2148 | | 2010 | 1905 | | 2011 | 1864 | | 2012 | 1529 | | 2013 | 1168 | | 2014 | 603 | | 9999 | 12 | | NULL | 123 | | |
| **Month of Birth Distribution** | |
| SELECT month\_of\_birth, count(\*)  FROM synthetic.synthetic\_CTGAN\_distributed\_dob\_v1  GROUP BY month\_of\_birth  ORDER BY month\_of\_birth   |  |  | | --- | --- | | month\_of\_birth | count | | 1 | 7914 | | 2 | 8037 | | 3 | 7722 | | 4 | 7916 | | 5 | 7710 | | 6 | 7796 | | 7 | 7905 | | 8 | 7838 | | 9 | 7848 | | 10 | 8024 | | 11 | 8009 | | 12 | 7971 | | SELECT EXTRACT(MONTH FROM dob), count(\*)  FROM long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide  GROUP BY EXTRACT(MONTH FROM dob)  ORDER BY EXTRACT(MONTH FROM dob)   |  |  | | --- | --- | | Month | Count | | 1 | 3641 | | 2 | 3498 | | 3 | 3595 | | 4 | 3838 | | 5 | 3406 | | 6 | 7053 | | 7 | 53256 | | 8 | 3512 | | 9 | 3333 | | 10 | 3095 | | 11 | 2829 | | 12 | 3646 | | NULL | 123 | |
| **Day of Birth Distribution** | |
| SELECT day\_of\_birth, count(\*)  FROM synthetic.synthetic\_CTGAN\_distributed\_dob\_v1  GROUP BY day\_of\_birth  ORDER BY day\_of\_birth   |  |  | | --- | --- | | day\_of\_birth | count | | 1 | 3546 | | 2 | 3295 | | 3 | 3404 | | 4 | 3422 | | 5 | 3295 | | 6 | 3419 | | 7 | 3366 | | 8 | 3351 | | 9 | 3367 | | 10 | 3526 | | 11 | 3365 | | 12 | 3407 | | 13 | 3373 | | 14 | 3397 | | 15 | 3275 | | 16 | 3436 | | 17 | 3392 | | 18 | 3299 | | 19 | 3403 | | 20 | 3330 | | 21 | 3337 | | 22 | 3447 | | 23 | 3390 | | 24 | 3363 | | 25 | 3347 | | 26 | 3343 | | 27 | 3474 | | 28 | 3321 | | SELECT EXTRACT(DAY FROM dob), count(\*)  FROM long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkageinputdatafinali\_wide  GROUP BY EXTRACT(DAY FROM dob)  ORDER BY EXTRACT(DAY FROM dob)   |  |  | | --- | --- | | Day | Count | | 1 | 8277 | | 2 | 798 | | 3 | 919 | | 4 | 827 | | 5 | 1133 | | 6 | 873 | | 7 | 830 | | 8 | 823 | | 9 | 777 | | 10 | 937 | | 11 | 743 | | 12 | 970 | | 13 | 661 | | 14 | 785 | | 15 | 63802 | | 16 | 786 | | 17 | 680 | | 18 | 723 | | 19 | 559 | | 20 | 756 | | 21 | 582 | | 22 | 689 | | 23 | 675 | | 24 | 770 | | 25 | 1145 | | 26 | 819 | | 27 | 850 | | 28 | 871 | | 29 | 620 | | 30 | 740 | | 31 | 282 | | NULL | 123 | |
|  |  |
|  |  |