**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_data_v1**

```
CREATE TABLE IF NOT EXISTS synthetic.synthetic_CTGAN_data_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 integer,
    day_of_birth integer,
    date_of_birth timestamp without time zone
)

TABLESPACE pg_default;

ALTER TABLE IF EXISTS synthetic."synthetic_CTGAN_data_v1"
    OWNER to postgres;
```
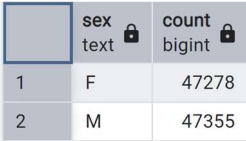
## Synthetic Data Shape: 94633 rows, 10 columns

Comparison:

| Table: synthetic.synthetic_CTGAN_data_v1 | Table: long_to_wide.pirl_rentch_hdss_r29_rdlinkageinputdatafinali_wide |
|---|---|
| **Total Rows** | |
| SELECT count(*) FROM synthetic.synthetic_without_date_normalizatio n_v1 | SELECT count(*) FROM long_to_wide.pirl_rentch_hdss_r29_rdlinkageinputdatafinali_ wide |
| count bigint: 1 → 94633 | count bigint: 1 → 94825 |
| **Sex Distribution** | |
| SELECT sex, count(*) FROM synthetic.synthetic_without_date_normalizatio n_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 |
| sex (text), count (bigint): 1 → F, 47278; 2 → M, 47355 | sex (smallint), count (bigint): 1 → 1, 44164; 2 → 2, 50656; 3 → 9, 5 |

## Villagename Distribution

SELECT villagename, count(*)
FROM
synthetic.synthetic_without_date_normalizatio
n_v1
GROUP BY villagename
ORDER BY villagename

| | villagename<br>text 🔒 | count 🔒<br>bigint |
|---|---|---|
| 1 | Igekemaja | 8842 |
| 2 | Ihayabuyaga | 10860 |
| 3 | Isangijo | 8272 |
| 4 | Kanyama | 14301 |
| 5 | Kisesa | 26221 |
| 6 | Kitumba | 15545 |
| 7 | Welamasonga | 10592 |

SELECT villagename_1, count(*)
FROM
long_to_wide.pirl_rentch_hdss_r29_rdlinkageinputdatafinali_
wide
GROUP BY villagename_1
ORDER BY villagename_1

| | villagename_1<br>text 🔒 | count 🔒<br>bigint |
|---|---|---|
| 1 | Igekemaja | 8619 |
| 2 | Ihayabuyaga | 10583 |
| 3 | Isangijo | 8175 |
| 4 | Kanyama | 14969 |
| 5 | Kisesa | 27087 |
| 6 | Kitumba | 14897 |
| 7 | Welamasonga | 10495 |

## Sub Villagename Distribution

SELECT subvillagename, count(*)
FROM
synthetic.synthetic_without_date_normalizatio
n_v1
GROUP BY subvillagename
ORDER BY subvillagename

| subvillagename | count |
|---|---|
| | 992 |
| Bukala | 832 |
| Bukandwe | 1942 |
| Bukelebe | 1676 |
| Changabe | 569 |
| Igadya | 342 |
| Igekemaja | 796 |
| Igeye | 1639 |
| Igudija A | 2626 |
| Igudija B | 2906 |
| Igumo | 2176 |
| Igunga | 1515 |
| Ihala | 1675 |
| Ihayabuyaga | 2462 |
| Ikangabuta | 1940 |
| Ikengele | 843 |
| Ikulicha | 610 |
| Ilagaja | 1396 |
| Ilangale | 2351 |
| Ilendeja | 3000 |

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 | 2708 | | Iseni Bondeni | 4516 |
| Iseni Mlimani | 1749 | | Iseni Mlimani | 1968 |
| Kanami | 1144 | | Kanami | 1402 |
| Kanyama | 2305 | | Kanyama | 3152 |
| Kigungumuli | 3043 | | Kigungumuli | 1339 |
| Kimaga | 1073 | | Kimaga | 1527 |
| Kisesa Kati | 7497 | | Kisesa Kati | 8178 |
| Kisha | 568 | | Kisha | 913 |
| Kitumba A | 1165 | | Kitumba A | 1320 |
| Kitumba B | 1624 | | Kitumba B | 2350 |
| Lumve | 1868 | | Lumve | 1337 |
| Mahilinga | 1580 | | Mahilinga | 1622 |
| Mondo | 1190 | | Mondo | 1264 |
| Mwabuki | 1047 | | Mwabuki | 2073 |
| Mwadubi | 1076 | | Mwadubi | 716 |
| Mwamanyili | 1213 | | Mwamanyili | 1254 |
| Mwaneneka | 1217 | | Mwaneneka | 916 |
| Ng'wabongoso | 1257 | | Ng'wabongoso | 1457 |
| Ng'wahuli | 3226 | | Ng'wahuli | 1630 |
| Ngï¿½wandulu | 7195 | | Ngï¿½wandulu | 4764 |
| Ngï¿½wangï¿½halanga | 7168 | | Ngï¿½wangï¿½halanga | 5881 |
| Njicha | 2298 | | Njicha | 1601 |
| Nkola | 225 | | Nkola | 1104 |
| None | 24 | | Nyamikoma | 744 |
| Nyamikoma | 249 | | Nyang'hulukulu | 2242 |
| Nyang'hulukulu | 1042 | | Nyanï¿½helela | 1031 |
| Nyanï¿½helela | 2299 | | Welamasonga | 1391 |
| Welamasonga | 3061 | | Wita | 6924 |
| Wita | 2234 | | | |

## Year of Birth Distribution

| | |
|---|---|
| SELECT year_of_birth, count(*) FROM synthetic.synthetic_without_date_normalizatio n_v1 GROUP BY year_of_birth ORDER BY year_of_birth | 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_of_Birth | Count |
|---|---|
| 1895 | 1 |
| 1896 | 3 |
| 1898 | 3 |
| 1899 | 1 |
| 1900 | 3 |
| 1901 | 3 |

| Year | Count |
| --- | --- |
| 1902 | 1 |
| 1903 | 5 |
| 1904 | 4 |
| 1905 | 5 |
| 1906 | 5 |
| 1907 | 5 |
| 1908 | 11 |
| 1909 | 5 |
| 1910 | 18 |
| 1911 | 9 |
| 1912 | 7 |
| 1913 | 17 |
| 1914 | 18 |
| 1915 | 21 |
| 1916 | 19 |
| 1917 | 23 |
| 1918 | 29 |
| 1919 | 36 |
| 1920 | 37 |
| 1921 | 53 |
| 1922 | 44 |
| 1923 | 45 |
| 1924 | 67 |
| 1925 | 59 |
| 1926 | 85 |
| 1927 | 64 |
| 1928 | 88 |
| 1929 | 89 |
| 1930 | 117 |
| 1931 | 92 |
| 1932 | 120 |
| 1933 | 110 |
| 1934 | 111 |
| 1935 | 129 |
| 1936 | 125 |
| 1937 | 142 |
| 1938 | 145 |
| 1939 | 192 |
| 1940 | 179 |
| 1941 | 166 |
| 1942 | 152 |
| 1943 | 188 |

| 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 | 165 | | | 1944 | 241 |
| | 1945 | 180 | | | 1945 | 172 |
| | 1946 | 194 | | | 1946 | 184 |
| | 1947 | 177 | | | 1947 | 151 |
| | 1948 | 203 | | | 1948 | 234 |
| | 1949 | 188 | | | 1949 | 247 |
| | 1950 | 158 | | | 1950 | 300 |
| | 1951 | 209 | | | 1951 | 135 |
| | 1952 | 184 | | | 1952 | 286 |
| | 1953 | 170 | | | 1953 | 164 |
| | 1954 | 181 | | | 1954 | 470 |
| | 1955 | 196 | | | 1955 | 219 |
| | 1956 | 233 | | | 1956 | 329 |
| | 1957 | 225 | | | 1957 | 198 |
| | 1958 | 265 | | | 1958 | 436 |
| | 1959 | 275 | | | 1959 | 374 |
| | 1960 | 336 | | | 1960 | 425 |
| | 1961 | 315 | | | 1961 | 576 |
| | 1962 | 344 | | | 1962 | 508 |
| | 1963 | 395 | | | 1963 | 424 |
| | 1964 | 413 | | | 1964 | 741 |
| | 1965 | 455 | | | 1965 | 542 |
| | 1966 | 520 | | | 1966 | 598 |
| | 1967 | 601 | | | 1967 | 531 |
| | 1968 | 791 | | | 1968 | 891 |
| | 1969 | 911 | | | 1969 | 725 |
| | 1970 | 1031 | | | 1970 | 1090 |
| | 1971 | 1162 | | | 1971 | 594 |
| | 1972 | 1302 | | | 1972 | 1289 |
| | 1973 | 1523 | | | 1973 | 954 |
| | 1974 | 1575 | | | 1974 | 1811 |
| | 1975 | 1706 | | | 1975 | 1225 |
| | 1976 | 1758 | | | 1976 | 1494 |
| | 1977 | 1750 | | | 1977 | 1359 |
| | 1978 | 1774 | | | 1978 | 1630 |
| | 1979 | 1736 | | | 1979 | 1724 |
| | 1980 | 1669 | | | 1980 | 1904 |
| | 1981 | 1609 | | | 1981 | 1279 |
| | 1982 | 1566 | | | 1982 | 2065 |
| | 1983 | 1598 | | | 1983 | 1521 |
| | 1984 | 1468 | | | 1984 | 2093 |
| | 1985 | 1528 | | | 1985 | 1573 |

| 1986 | 1631 | | 1986 | 2083 |
|------|------|--|------|------|
| 1987 | 1676 | | 1987 | 1653 |
| 1988 | 1898 | | 1988 | 2044 |
| 1989 | 1976 | | 1989 | 1697 |
| 1990 | 2089 | | 1990 | 1934 |
| 1991 | 2127 | | 1991 | 1673 |
| 1992 | 2253 | | 1992 | 2090 |
| 1993 | 2227 | | 1993 | 1924 |
| 1994 | 2197 | | 1994 | 2084 |
| 1995 | 2062 | | 1995 | 2145 |
| 1996 | 2036 | | 1996 | 2138 |
| 1997 | 2052 | | 1997 | 2031 |
| 1998 | 2035 | | 1998 | 2057 |
| 1999 | 2029 | | 1999 | 2224 |
| 2000 | 2203 | | 2000 | 2329 |
| 2001 | 2270 | | 2001 | 2175 |
| 2002 | 2248 | | 2002 | 2468 |
| 2003 | 2206 | | 2003 | 2480 |
| 2004 | 2422 | | 2004 | 2453 |
| 2005 | 2269 | | 2005 | 2440 |
| 2006 | 2220 | | 2006 | 2392 |
| 2007 | 2096 | | 2007 | 2387 |
| 2008 | 2005 | | 2008 | 2300 |
| 2009 | 1914 | | 2009 | 2148 |
| 2010 | 2031 | | 2010 | 1905 |
| 2011 | 1969 | | 2011 | 1864 |
| 2012 | 1813 | | 2012 | 1529 |
| 2013 | 1386 | | 2013 | 1168 |
| 2014 | 911  | | 2014 | 603  |
| 2015 | 376  | | 9999 | 12   |
| 2016 | 106  | | NULL | 123  |
| 2017 | 10   | |      |      |

## Month of Birth Distribution

```
SELECT month_of_birth, count(*)
FROM
synthetic.synthetic_without_date_normalizatio
n_v1
GROUP BY month_of_birth
ORDER BY month_of_birth
```

```
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_of_Birth | Count | | Month | Count |
|----------------|-------|--|-------|-------|
| 1 | 3402 | | 1 | 3641 |
| 2 | 2152 | | 2 | 3498 |
| 3 | 2434 | | 3 | 3595 |

| 4 | 2345 |
|---|---|
| 5 | 1577 |
| 6 | 5329 |
| 7 | 56439 |
| 8 | 5109 |
| 9 | 3717 |
| 10 | 3833 |
| 11 | 3912 |
| 12 | 4384 |

| 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_without_date_normalizatio
n_v1
GROUP BY day_of_birth
ORDER BY day_of_birth
```

```
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_of_Birth | Count |
|---|---|
| 1 | 11414 |
| 2 | 26 |
| 3 | 1426 |
| 4 | 2293 |
| 5 | 101 |
| 6 | 23 |
| 7 | 528 |
| 8 | 1742 |
| 9 | 1356 |
| 10 | 211 |
| 11 | 5 |
| 12 | 795 |
| 13 | 2619 |
| 14 | 48 |
| 15 | 57636 |
| 16 | 785 |
| 17 | 147 |
| 18 | 1470 |
| 19 | 2990 |
| 20 | 355 |
| 21 | 389 |
| 22 | 893 |
| 23 | 670 |
| 24 | 582 |
| 25 | 1684 |

| 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 | 700 | | 26 | 819 |
| 27 | 33 | | 27 | 850 |
| 28 | 807 | | 28 | 871 |
| 29 | 2725 | | 29 | 620 |
| 30 | 180 | | 30 | 740 |
| | | | 31 | 282 |
| | | | NULL | 123 |