

## 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\_without\_date\_normalization\_v1

CREATE TABLE IF NOT EXISTS synthetic.synthetic\_without\_date\_normalization\_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
)

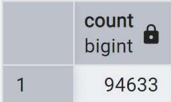


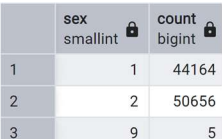
TABLESPACE pg_default;

ALTER TABLE IF EXISTS synthetic.synthetic_without_date_normalization_v1
OWNER to postgres;

## Synthetic Data Shape: 94633 rows, 10 columns

```

Comparison:

Table: synthetic.synthetic_without_date_normalization_v1	Table: long_to_wide.pirl_rentch_hdss_r29_rdlinkeageinputdatafinali_wide
<b>Total Rows</b>	
SELECT count(*) FROM synthetic.synthetic_without_date_normalization_v1  	SELECT count(*) FROM long_to_wide.pirl_rentch_hdss_r29_rdlinkeageinputdatafinali_wide  
<b>Sex Distribution</b>	
SELECT sex, count(*) FROM synthetic.synthetic_without_date_normalization_v1 GROUP BY sex ORDER BY sex  	SELECT sex, count(*) FROM long_to_wide.pirl_rentch_hdss_r29_rdlinkeageinputdatafinali_wide GROUP BY sex ORDER BY sex  
<b>Villagename Distribution</b>	
SELECT villagename, count(*)	SELECT villagename_1, count(*)

FROM  
 synthetic.synthetic\_without\_date\_normalization\_v1  
 GROUP BY villagename  
 ORDER BY villagename

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

FROM  
 long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkeinputdatafinali\_wide  
 GROUP BY villagename\_1  
 ORDER BY villagename\_1

	villagename_1 text	count 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\_normalization\_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
Iseni Bondeni	2708

SELECT subvillagename\_1, count(\*)  
 FROM  
 long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkeinputdatafinali\_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	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	
Ngĩ½wandulu	7195		Ngĩ½wandulu	4764	
Ngĩ½wangĩ½halanga	7168		Ngĩ½wangĩ½halanga	5881	
Njicha	2298		Njicha	1601	
Nkola	225		Nkola	1104	
None	24		Nyamikoma	744	
Nyamikoma	249		Nyang'hulukulu	2242	
Nyang'hulukulu	1042		Nyani½helela	1031	
Nyani½helela	2299		Welamasonga	1391	
Welamasonga	3061		Wita	6924	
Wita	2234				

#### Year of Birth Distribution

SELECT year\_of\_birth, count(\*)  
FROM  
synthetic.synthetic\_without\_date\_normalization\_v1  
GROUP BY year\_of\_birth  
ORDER BY year\_of\_birth

Year_of_Birth	Count
1895	1
1896	3
1898	3
1899	1
1900	3
1901	3
1902	1

SELECT EXTRACT(YEAR FROM dob), count(\*)  
FROM  
long\_to\_wide.pirl\_rentch\_hdss\_r29\_rdlinkeinputdatafinali\_wide  
GROUP BY EXTRACT(YEAR FROM dob)  
ORDER BY EXTRACT(YEAR FROM dob)

	1903	5			Year	Count	
	1904	4			1901	2	
	1905	5			1904	11	
	1906	5			1905	10	
	1907	5			1906	7	
	1908	11			1907	2	
	1909	5			1908	10	
	1910	18			1909	5	
	1911	9			1910	18	
	1912	7			1911	9	
	1913	17			1912	11	
	1914	18			1913	2	
	1915	21			1914	37	
	1916	19			1915	13	
	1917	23			1916	21	
	1918	29			1917	12	
	1919	36			1918	21	
	1920	37			1919	51	
	1921	53			1920	69	
	1922	44			1921	21	
	1923	45			1922	50	
	1924	67			1923	18	
	1925	59			1924	97	
	1926	85			1925	45	
	1927	64			1926	74	
	1928	88			1927	53	
	1929	89			1928	63	
	1930	117			1929	68	
	1931	92			1930	161	
	1932	120			1931	51	
	1933	110			1932	91	
	1934	111			1933	72	
	1935	129			1934	162	
	1936	125			1935	77	
	1937	142			1936	214	
	1938	145			1937	89	
	1939	192			1938	124	
	1940	179			1939	121	
	1941	166			1940	209	
	1942	152			1941	86	
	1943	188			1942	163	
	1944	165			1943	146	

	1945	180			1944	241	
	1946	194			1945	172	
	1947	177			1946	184	
	1948	203			1947	151	
	1949	188			1948	234	
	1950	158			1949	247	
	1951	209			1950	300	
	1952	184			1951	135	
	1953	170			1952	286	
	1954	181			1953	164	
	1955	196			1954	470	
	1956	233			1955	219	
	1957	225			1956	329	
	1958	265			1957	198	
	1959	275			1958	436	
	1960	336			1959	374	
	1961	315			1960	425	
	1962	344			1961	576	
	1963	395			1962	508	
	1964	413			1963	424	
	1965	455			1964	741	
	1966	520			1965	542	
	1967	601			1966	598	
	1968	791			1967	531	
	1969	911			1968	891	
	1970	1031			1969	725	
	1971	1162			1970	1090	
	1972	1302			1971	594	
	1973	1523			1972	1289	
	1974	1575			1973	954	
	1975	1706			1974	1811	
	1976	1758			1975	1225	
	1977	1750			1976	1494	
	1978	1774			1977	1359	
	1979	1736			1978	1630	
	1980	1669			1979	1724	
	1981	1609			1980	1904	
	1982	1566			1981	1279	
	1983	1598			1982	2065	
	1984	1468			1983	1521	
	1985	1528			1984	2093	
	1986	1631			1985	1573	

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

#### Month of Birth Distribution

```
SELECT month_of_birth, count(*)
FROM
synthetic.synthetic_without_date_normalization_v1
GROUP BY month_of_birth
ORDER BY month_of_birth
```

Month_of_Birth	Count
1	3402
2	2152
3	2434
4	2345

```
SELECT EXTRACT(MONTH FROM dob), count(*)
FROM
long_to_wide.pirl_rentch_hdss_r29_rdl linkageinputdatafinali_wide
GROUP BY EXTRACT(MONTH FROM dob)
ORDER BY EXTRACT(MONTH FROM dob)
```

Month	Count
1	3641
2	3498
3	3595
4	3838

5	1577	5	3406
6	5329	6	7053
7	56439	7	53256
8	5109	8	3512
9	3717	9	3333
10	3833	10	3095
11	3912	11	2829
12	4384	12	3646
		NULL	123

### Day of Birth Distribution

```
SELECT day_of_birth, count(*)
FROM
synthetic.synthetic_without_date_normalization_v1
GROUP BY day_of_birth
ORDER BY day_of_birth
```

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
26	700

```
SELECT EXTRACT(DAY FROM dob), count(*)
FROM
long_to_wide.pirl_rentch_hdss_r29_rdl linkageinputdatafinali_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	33		27	850	
	28	807		28	871	
	29	2725		29	620	
	30	180		30	740	
				31	282	
				NULL	123	