

Comparison between tables

Table 1: pirl_rentch_hdss_r29_rdlinkeinputdatafinali_wide

CREATE TABLE IF NOT EXISTS

long_to_wide.pirl_rentch_hdss_r29_rdlinkeinputdatafinali_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_rdlinkeinputdatafinali_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_rdlinkeinputdatafinali _wide
Total Rows	
SELECT count(*) FROM synthetic.synthetic_CTGAN_distributed_dob_v1	SELECT count(*) FROM long_to_wide.pirl_rentch_hdss_r29_rdlinkeinputdatafinali _wide

<table><tr><td></td><td>count bigint</td></tr><tr><td>1</td><td>94690</td></tr></table>			count bigint	1	94690	<table><tr><td></td><td>count bigint</td></tr><tr><td>1</td><td>94825</td></tr></table>			count bigint	1	94825																																								
	count bigint																																																		
1	94690																																																		
	count bigint																																																		
1	94825																																																		
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																																																	
<table><tr><td></td><td>sex text</td><td>count bigint</td></tr><tr><td>1</td><td>F</td><td>47241</td></tr><tr><td>2</td><td>M</td><td>47449</td></tr></table>			sex text	count bigint	1	F	47241	2	M	47449	<table><tr><td></td><td>sex smallint</td><td>count bigint</td></tr><tr><td>1</td><td>1</td><td>44164</td></tr><tr><td>2</td><td>2</td><td>50656</td></tr><tr><td>3</td><td>9</td><td>5</td></tr></table>			sex smallint	count bigint	1	1	44164	2	2	50656	3	9	5																											
	sex text	count bigint																																																	
1	F	47241																																																	
2	M	47449																																																	
	sex smallint	count bigint																																																	
1	1	44164																																																	
2	2	50656																																																	
3	9	5																																																	
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																																																	
<table><tr><td></td><td>villagename text</td><td>count bigint</td></tr><tr><td>1</td><td>Igekemaja</td><td>8726</td></tr><tr><td>2</td><td>Ihayabuyaga</td><td>10849</td></tr><tr><td>3</td><td>Isangijo</td><td>8749</td></tr><tr><td>4</td><td>Kanyama</td><td>15085</td></tr><tr><td>5</td><td>Kisesa</td><td>26319</td></tr><tr><td>6</td><td>Kitumba</td><td>14173</td></tr><tr><td>7</td><td>Welamasonga</td><td>10789</td></tr></table>			villagename text	count bigint	1	Igekemaja	8726	2	Ihayabuyaga	10849	3	Isangijo	8749	4	Kanyama	15085	5	Kisesa	26319	6	Kitumba	14173	7	Welamasonga	10789	<table><tr><td></td><td>villagename_1 text</td><td>count bigint</td></tr><tr><td>1</td><td>Igekemaja</td><td>8619</td></tr><tr><td>2</td><td>Ihayabuyaga</td><td>10583</td></tr><tr><td>3</td><td>Isangijo</td><td>8175</td></tr><tr><td>4</td><td>Kanyama</td><td>14969</td></tr><tr><td>5</td><td>Kisesa</td><td>27087</td></tr><tr><td>6</td><td>Kitumba</td><td>14897</td></tr><tr><td>7</td><td>Welamasonga</td><td>10495</td></tr></table>			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
	villagename text	count bigint																																																	
1	Igekemaja	8726																																																	
2	Ihayabuyaga	10849																																																	
3	Isangijo	8749																																																	
4	Kanyama	15085																																																	
5	Kisesa	26319																																																	
6	Kitumba	14173																																																	
7	Welamasonga	10789																																																	
	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_CTGAN_distributed_dob_v1 GROUP BY subvillagename ORDER BY subvillagename		SELECT subvillagename_1, count(*) FROM long_to_wide.pirl_rentch_hdss_r29_rdlinkageinputdatafinali_wide GROUP BY subvillagename_1 ORDER BY subvillagename_1																																																	
<table><tr><td>subvillagename</td><td>count</td></tr><tr><td></td><td>4081</td></tr><tr><td>Bukala</td><td>231</td></tr><tr><td>Bukandwe</td><td>674</td></tr><tr><td>Bukelebe</td><td>933</td></tr><tr><td>Changabe</td><td>762</td></tr><tr><td>Igadya</td><td>586</td></tr><tr><td>Igekemaja</td><td>1019</td></tr><tr><td>Igeye</td><td>1731</td></tr></table>		subvillagename	count		4081	Bukala	231	Bukandwe	674	Bukelebe	933	Changabe	762	Igadya	586	Igekemaja	1019	Igeye	1731	<table><tr><td>subvillagename_1</td><td>count</td></tr><tr><td>NULL</td><td>16</td></tr><tr><td>Bukala</td><td>810</td></tr><tr><td>Bukandwe</td><td>1093</td></tr><tr><td>Bukelebe</td><td>1769</td></tr><tr><td>Changabe</td><td>621</td></tr><tr><td>Igadya</td><td>577</td></tr><tr><td>Igekemaja</td><td>1380</td></tr></table>		subvillagename_1	count	NULL	16	Bukala	810	Bukandwe	1093	Bukelebe	1769	Changabe	621	Igadya	577	Igekemaja	1380														
subvillagename	count																																																		
	4081																																																		
Bukala	231																																																		
Bukandwe	674																																																		
Bukelebe	933																																																		
Changabe	762																																																		
Igadya	586																																																		
Igekemaja	1019																																																		
Igeye	1731																																																		
subvillagename_1	count																																																		
NULL	16																																																		
Bukala	810																																																		
Bukandwe	1093																																																		
Bukelebe	1769																																																		
Changabe	621																																																		
Igadya	577																																																		
Igekemaja	1380																																																		

Igudija A	2193	Igeye	865
Igudija B	1958	Igudija A	2240
Igumo	1176	Igudija B	3364
Igunga	812	Igumo	1228
Ihala	1591	Igunga	1832
Ihayabuyaga	2927	Ihala	1446
Ikangabuta	2812	Ihayabuyaga	4822
Ikengele	1271	Ikangabuta	983
Ikulicha	828	Ikengele	1706
Ilagaja	1465	Ikulicha	799
Ilangale	2415	Ilagaja	1717
Ilendeja	3355	Ilangale	1105
Iseni Bondeni	3849	Ilendeja	1836
Iseni Mlimani	3668	Iseni Bondeni	4516
Kanami	3009	Iseni Mlimani	1968
Kanyama	2489	Kanami	1402
Kigungumuli	1788	Kanyama	3152
Kimaga	789	Kigungumuli	1339
Kisesa Kati	6073	Kimaga	1527
Kisha	1972	Kisesa Kati	8178
Kitumba A	2612	Kisha	913
Kitumba B	2172	Kitumba A	1320
Lumve	1420	Kitumba B	2350
Mahilinga	1021	Lumve	1337
Mondo	968	Mahilinga	1622
Mwabuki	986	Mondo	1264
Mwadubi	890	Mwabuki	2073
Mwamanyili	835	Mwadubi	716
Mwaneneka	559	Mwamanyili	1254
Ng'wabongoso	605	Mwaneneka	916
Ng'wahuli	1447	Ng'wabongoso	1457
Ngĩ½wandulu	3548	Ng'wahuli	1630
Ngĩ½wangĩ½hala nga	4852	Ngĩ½wandulu	4764
Njicha	3229	Ngĩ½wangĩ½halanga	5881
Nkola	629	Njicha	1601
None	200	Nkola	1104
Nyamikoma	649	Nyamikoma	744
Nyang'hulukulu	1365	Nyang'hulukulu	2242
Nyani½helela	2214	Nyani½helela	1031
Welamasonga	3570	Welamasonga	1391
Wita	4462	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

```
SELECT EXTRACT(YEAR FROM dob), count(*)
FROM
long_to_wide.pirl_rentch_hdss_r29_rdl linkageinputdatafinali
_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

1937	158		1938	124
1938	138		1939	121
1939	129		1940	209
1940	124		1941	86
1941	147		1942	163
1942	136		1943	146
1943	109		1944	241
1944	124		1945	172
1945	150		1946	184
1946	140		1947	151
1947	159		1948	234
1948	185		1949	247
1949	222		1950	300
1950	220		1951	135
1951	276		1952	286
1952	278		1953	164
1953	328		1954	470
1954	347		1955	219
1955	376		1956	329
1956	438		1957	198
1957	493		1958	436
1958	531		1959	374
1959	606		1960	425
1960	607		1961	576
1961	643		1962	508
1962	652		1963	424
1963	691		1964	741
1964	608		1965	542
1965	632		1966	598
1966	642		1967	531
1967	615		1968	891
1968	712		1969	725
1969	850		1970	1090
1970	958		1971	594
1971	1154		1972	1289
1972	1330		1973	954
1973	1335		1974	1811
1974	1284		1975	1225
1975	1383		1976	1494
1976	1434		1977	1359
1977	1435		1978	1630
1978	1420		1979	1724

1979	1369		1980	1904	
1980	1283		1981	1279	
1981	1293		1982	2065	
1982	1409		1983	1521	
1983	1524		1984	2093	
1984	1709		1985	1573	
1985	1887		1986	2083	
1986	1837		1987	1653	
1987	1850		1988	2044	
1988	1827		1989	1697	
1989	1922		1990	1934	
1990	1963		1991	1673	
1991	2066		1992	2090	
1992	2147		1993	1924	
1993	2244		1994	2084	
1994	2159		1995	2145	
1995	2160		1996	2138	
1996	2072		1997	2031	
1997	2217		1998	2057	
1998	2172		1999	2224	
1999	2247		2000	2329	
2000	2193		2001	2175	
2001	2200		2002	2468	
2002	2261		2003	2480	
2003	2219		2004	2453	
2004	2124		2005	2440	
2005	1979		2006	2392	
2006	1927		2007	2387	
2007	1852		2008	2300	
2008	2050		2009	2148	
2009	2193		2010	1905	
2010	2151		2011	1864	
2011	2083		2012	1529	
2012	1568		2013	1168	
2013	972		2014	603	
2014	532		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			SELECT EXTRACT(MONTH FROM dob), count(*) FROM long_to_wide.pirl_rentch_hdss_r29_rdlinkeinputdatafinali_wide GROUP BY EXTRACT(MONTH FROM dob)		

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

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

SELECT EXTRACT(DAY FROM dob), count(*)
FROM
long_to_wide.pirl_rentch_hdss_r29_rdlinkeinputdatafinali
_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

21	3337			20	756	
22	3447			21	582	
23	3390			22	689	
24	3363			23	675	
25	3347			24	770	
26	3343			25	1145	
27	3474			26	819	
28	3321			27	850	
				28	871	
				29	620	
				30	740	
				31	282	
				NULL	123	