### Real

| | workclass | occupation | race | sex | `marital.status` | income | relationship | capital | count |
|---|---|---|---|---|---|---|---|---|---|
| 320 | Federal-gov | Exec-managerial | White | Female | Married-spouse-absent | >50K | Not-in-family | 10520 | 1 |
| 2147 | Private | Prof-specialty | Amer-Indian-Eskimo | Female | Divorced | <=50K | Unmarried | 0 | 1 |
| 3741 | State-gov | Protective-serv | Amer-Indian-Eskimo | Male | Married-civ-spouse | <=50K | Husband | 0 | 3 |
| 2935 | Self-emp-inc | Sales | White | Female | Widowed | <=50K | Not-in-family | 0 | 5 |
| 542 | Federal-gov | Transport-moving | Black | Male | Married-civ-spouse | >50K | Husband | 0 | 4 |

### Synthetic

| | workclass | occupation | race | sex | `marital.status` | income | relationship | capital | count |
|---|---|---|---|---|---|---|---|---|---|
| 320 | Federal-gov | Exec-managerial | White | Female | Married-spouse-absent | >50K | Not-in-family | nan | nan |
| 2147 | Private | Prof-specialty | Amer-Indian-Eskimo | Female | Divorced | <=50K | Unmarried | nan | nan |
| 3741 | State-gov | Protective-serv | Amer-Indian-Eskimo | Male | Married-civ-spouse | <=50K | Husband | 80.809213 | 1.000000 |
| 2935 | Self-emp-inc | Sales | White | Female | Widowed | <=50K | Not-in-family | -315.136364 | 8.000000 |
| 542 | Federal-gov | Transport-moving | Black | Male | Married-civ-spouse | >50K | Husband | nan | nan |

SQL for Real:
SELECT workclass,occupation,race,sex,`marital.status`,income,relationship,MIN(capital), COUNT(*) FROM C1 GROUP BY workclass,occupation,race,sex,`marital.status`,income,relationship

Resulted in 3839 records

SQL for Synthetic:
SELECT workclass,occupation,race,sex,`marital.status`,income,relationship,MIN(capital), COUNT(*) FROM C1_syn_06 GROUP BY workclass,occupation,race,sex,`marital.status`,income,relationship

Resulted in 3486 records

Normalized Euclidean distance for (capital): 47.72

Hellinger Distance: 0.237

==========================================

### Real

| | sex | race | relationship | workclass | occupation | `native.country` | income | `marital.status` | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 2922 | Female | White | Own-child | Private | Prof-specialty | Guatemala | >50K | Divorced | nan | nan |
| 255 | Female | Asian-Pac-Islander | Unmarried | State-gov | Other-service | United-States | <=50K | Never-married | nan | nan |
| 3493 | Female | White | Unmarried | Private | Machine-op-inspct | United-States | <=50K | Never-married | 45.000000 | 20.000000 |
| 3019 | Female | White | Own-child | Private | Transport-moving | Puerto-Rico | <=50K | Never-married | nan | nan |
| 2672 | Female | White | Own-child | Local-gov | Adm-clerical | United-States | <=50K | Widowed | nan | nan |

### Synthetic

| | sex | race | relationship | workclass | occupation | `native.country` | income | `marital.status` | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 2922 | Female | White | Own-child | Private | Prof-specialty | Guatemala | >50K | Divorced | 33 | 1 |
| 255 | Female | Asian-Pac-Islander | Unmarried | State-gov | Other-service | United-States | <=50K | Never-married | 41 | 1 |
| 3493 | Female | White | Unmarried | Private | Machine-op-inspct | United-States | <=50K | Never-married | 59 | 16 |
| 3019 | Female | White | Own-child | Private | Transport-moving | Puerto-Rico | <=50K | Never-married | 18 | 1 |
| 2672 | Female | White | Own-child | Local-gov | Adm-clerical | United-States | <=50K | Widowed | 56 | 2 |

SQL for Real:
SELECT sex,race,relationship,workclass,occupation,`native.country`,income,`marital.status`,MAX(age), COUNT(*) FROM C1 GROUP BY sex,race,relationship,workclass,occupation,`native.country`,income,`marital.status`

Resulted in 6492 records

SQL for Synthetic:
SELECT sex,race,relationship,workclass,occupation,`native.country`,income,`marital.status`,MAX(age), COUNT(*) FROM C1_syn_06 GROUP BY sex,race,relationship,workclass,occupation,`native.country`,income,`marital.status`

Resulted in 8900 records

Normalized Euclidean distance for (age): 49.69

Hellinger Distance: 0.502

==========================================

### Real

| | income | capital | count |
|---|---|---|---|
| 1 | >50K | -3683 | 11687 |
| 0 | <=50K | -4356 | 37155 |

### Synthetic

| | income | capital | count |
|---|---|---|---|
| 1 | >50K | -439.660050 | 4762.000000 |
| 0 | <=50K | -421.663504 | 44080.000000 |

SQL for Real:
SELECT income,MIN(capital), COUNT(*) FROM C1 GROUP BY income

Resulted in 2 records

SQL for Synthetic:
SELECT income,MIN(capital), COUNT(*) FROM C1_syn_06 GROUP BY income

Resulted in 2 records

Normalized Euclidean distance for (capital): 1.41

Hellinger Distance: 0.137

==========================================

|  | Real |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | race | education | workclass | occupation | `hours.per.week` | count |
| 544 | Black | 5th-6th | Private | Machine-op-inspct | 208 | 5 |
| 1319 | White | 9th | Self-emp-not-inc | Prof-specialty | 35 | 1 |
| 152 | Amer-Indian-Eskimo | Some-college | Local-gov | Exec-managerial | 202 | 5 |
| 690 | Black | Bachelors | Private | Tech-support | 647 | 16 |
| 1595 | White | HS-grad | Self-emp-inc | Craft-repair | 3226 | 69 |

|  | Synthetic |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | race | education | workclass | occupation | `hours.per.week` | count |
| 544 | Black | 5th-6th | Private | Machine-op-inspct | 400.012602 | 10.000000 |
| 1319 | White | 9th | Self-emp-not-inc | Prof-specialty | 39.076040 | 1.000000 |
| 152 | Amer-Indian-Eskimo | Some-college | Local-gov | Exec-managerial | 40.027960 | 1.000000 |
| 690 | Black | Bachelors | Private | Tech-support | 679.915977 | 17.000000 |
| 1595 | White | HS-grad | Self-emp-inc | Craft-repair | 2625.662088 | 65.000000 |

SQL for Real:
SELECT race,education,workclass,occupation,SUM(`hours.per.week`), COUNT(*) FROM C1 GROUP BY race,education,workclass,occupation

Resulted in 1814 records

SQL for Synthetic:
SELECT race,education,workclass,occupation,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY race,education,workclass,occupation

Resulted in 1746 records

Normalized Euclidean distance for (`hours.per.week`): 36.37

Hellinger Distance: 0.123

==========================================

|  | Real |  |  |
| --- | --- | --- | --- |
|  | sex | age | count |
| 1 | Male | 90 | 32650 |
| 0 | Female | 90 | 16192 |

|  | Synthetic |  |  |
| --- | --- | --- | --- |
|  | sex | age | count |
| 1 | Male | 90.000000 | 31728.000000 |
| 0 | Female | 90.000000 | 17114.000000 |

SQL for Real:
SELECT sex,MAX(age), COUNT(*) FROM C1 GROUP BY sex

Resulted in 2 records

SQL for Synthetic:
SELECT sex,MAX(age), COUNT(*) FROM C1_syn_06 GROUP BY sex

Resulted in 2 records

Normalized Euclidean distance for (age): 0.0

Hellinger Distance: 0.014

==========================================

|  | Real |  |  |  |
| --- | --- | --- | --- | --- |
|  | income | occupation | age | count |
| 21 | >50K | Handlers-cleaners | 70 | 138 |
| 10 | <=50K | Prof-specialty | 90 | 3388 |
| 7 | <=50K | Machine-op-inspct | 90 | 2650 |
| 2 | <=50K | Armed-Forces | 39 | 10 |
| 23 | >50K | Other-service | 77 | 204 |

|  | Synthetic |  |  |  |
| --- | --- | --- | --- | --- |
|  | income | occupation | age | count |
| 21 | >50K | Handlers-cleaners | 72.000000 | 56.000000 |
| 10 | <=50K | Prof-specialty | 90.000000 | 4931.000000 |
| 7 | <=50K | Machine-op-inspct | 90.000000 | 2944.000000 |
| 2 | <=50K | Armed-Forces | 52.000000 | 114.000000 |
| 23 | >50K | Other-service | 70.000000 | 64.000000 |

SQL for Real:
SELECT income,occupation,MAX(age), COUNT(*) FROM C1 GROUP BY income,occupation

Resulted in 30 records

SQL for Synthetic:
SELECT income,occupation,MAX(age), COUNT(*) FROM C1_syn_06 GROUP BY income,occupation

Resulted in 30 records

Normalized Euclidean distance for (age): 5.48

Hellinger Distance: 0.146

==========================================

### Real

| | occupation | income | education | `hours.per.week` | count |
|---|---|---|---|---|---|
| 281 | Prof-specialty | >50K | HS-grad | 42.397727 | 88 |
| 279 | Prof-specialty | >50K | Bachelors | 43.680540 | 889 |
| 168 | Handlers-cleaners | <=50K | HS-grad | 39.068886 | 871 |
| 1 | ? | <=50K | 11th | 30.374302 | 179 |
| 188 | Machine-op-inspct | <=50K | 9th | 39.826531 | 98 |

### Synthetic

| | occupation | income | education | `hours.per.week` | count |
|---|---|---|---|---|---|
| 281 | Prof-specialty | >50K | HS-grad | 40.097297 | 35.000000 |
| 279 | Prof-specialty | >50K | Bachelors | 40.106597 | 368.000000 |
| 168 | Handlers-cleaners | <=50K | HS-grad | 39.932715 | 1016.000000 |
| 1 | ? | <=50K | 11th | 39.583715 | 156.000000 |
| 188 | Machine-op-inspct | <=50K | 9th | 40.007418 | 84.000000 |

SQL for Real:
SELECT occupation,income,education,AVG(`hours.per.week`), COUNT(*) FROM C1 GROUP BY occupation,income,education

Resulted in 391 records

SQL for Synthetic:
SELECT occupation,income,education,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY occupation,income,education

Resulted in 379 records

Normalized Euclidean distance for (`hours.per.week`): 18.71

Hellinger Distance: 0.16

==========================================

### Real

| | occupation | workclass | race | education | `native.country` | sex | `marital.status` | capital | count |
|---|---|---|---|---|---|---|---|---|---|
| 573 | ? | ? | White | Masters | ? | Female | Divorced | 0.000000 | 1.000000 |
| 12428 | Transport-moving | Private | White | 9th | ? | Male | Married-civ-spouse | nan | nan |
| 8020 | Other-service | Self-emp-not-inc | White | 10th | United-States | Female | Divorced | 0.000000 | 1.000000 |
| 3029 | Craft-repair | Private | White | HS-grad | United-States | Female | Separated | 0.000000 | 7.000000 |
| 712 | ? | Private | White | Bachelors | Puerto-Rico | Male | Married-civ-spouse | nan | nan |

### Synthetic

| | occupation | workclass | race | education | `native.country` | sex | `marital.status` | capital | count |
|---|---|---|---|---|---|---|---|---|---|
| 573 | ? | ? | White | Masters | ? | Female | Divorced | -35.824637 | 1 |
| 12428 | Transport-moving | Private | White | 9th | ? | Male | Married-civ-spouse | -144.166221 | 1 |
| 8020 | Other-service | Self-emp-not-inc | White | 10th | United-States | Female | Divorced | 51.772781 | 1 |
| 3029 | Craft-repair | Private | White | HS-grad | United-States | Female | Separated | -281.802755 | 6 |
| 712 | ? | Private | White | Bachelors | Puerto-Rico | Male | Married-civ-spouse | -123.683775 | 1 |

SQL for Real:
SELECT occupation,workclass,race,education,`native.country`,sex,`marital.status`,MIN(capital), COUNT(*) FROM C1 GROUP BY occupation,workclass,race,education,`native.country`,sex,`marital.status`

Resulted in 8789 records

SQL for Synthetic:
SELECT occupation,workclass,race,education,`native.country`,sex,`marital.status`,MIN(capital), COUNT(*) FROM C1_syn_06 GROUP BY occupation,workclass,race,education,`native.country`,sex,`marital.status`

Resulted in 12726 records

Normalized Euclidean distance for (capital): 57.97

Hellinger Distance: 0.508

==========================================

### Real

| | education | sex | race | occupation | relationship | income | `native.country` | workclass | `marital.status` | age | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5978 | Bachelors | Female | White | Prof-specialty | Not-in-family | <=50K | United-States | State-gov | Divorced | 60.000000 | 7.000000 |
| 826 | 11th | Female | White | ? | Wife | <=50K | United-States | ? | Married-civ-spouse | 40.000000 | 6.000000 |
| 16413 | Some-college | Male | White | Handlers-cleaners | Husband | <=50K | Columbia | Local-gov | Married-civ-spouse | nan | nan |
| 10588 | HS-grad | Male | Black | Transport-moving | Not-in-family | <=50K | United-States | Private | Divorced | 54.000000 | 2.000000 |
| 13759 | Prof-school | Male | White | Exec-managerial | Husband | >50K | Puerto-Rico | Federal-gov | Married-civ-spouse | nan | nan |

### Synthetic

| | education | sex | race | occupation | relationship | income | `native.country` | workclass | `marital.status` | age | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5978 | Bachelors | Female | White | Prof-specialty | Not-in-family | <=50K | United-States | State-gov | Divorced | 51 | 3 |
| 826 | 11th | Female | White | ? | Wife | <=50K | United-States | ? | Married-civ-spouse | 23 | 2 |
| 16413 | Some-college | Male | White | Handlers-cleaners | Husband | <=50K | Columbia | Local-gov | Married-civ-spouse | 42 | 1 |
| 10588 | HS-grad | Male | Black | Transport-moving | Not-in-family | <=50K | United-States | Private | Divorced | 54 | 2 |
| 13759 | Prof-school | Male | White | Exec-managerial | Husband | >50K | Puerto-Rico | Federal-gov | Married-civ-spouse | 43 | 1 |

SQL for Real:
SELECT education,sex,race,occupation,relationship,income,`native.country`,workclass,`marital.status`,MAX(age), COUNT(*) FROM C1 GROUP BY education,sex,race,occupation,relationship,income,`native.country`,workclass,`marital.status`

Resulted in 12664 records

SQL for Synthetic:
SELECT education,sex,race,occupation,relationship,income,`native.country`,workclass,`marital.status`,MAX(age), COUNT(*) FROM C1_syn_06 GROUP BY education,sex,race,occupation,relationship,income,`native.country`,workclass,`marital.status`

Resulted in 17127 records

Normalized Euclidean distance for (age): 64.61

Hellinger Distance: 0.549

================================================

Real

| | `native.country` | age | count |
|---|---|---|---|
| 37 | Thailand | 1130 | 30 |
| 32 | Portugal | 2763 | 67 |
| 23 | Jamaica | 3937 | 106 |
| 40 | Vietnam | 2977 | 86 |
| 24 | Japan | 3437 | 92 |

Synthetic

| | `native.country` | age | count |
|---|---|---|---|
| 37 | Thailand | 4821.000000 | 211.000000 |
| 32 | Portugal | 117.000000 | 2.000000 |
| 23 | Jamaica | 3109.000000 | 59.000000 |
| 40 | Vietnam | 957.000000 | 39.000000 |
| 24 | Japan | 49167.000000 | 1156.000000 |

SQL for Real:
SELECT `native.country`,SUM(age), COUNT(*) FROM C1 GROUP BY `native.country`

Resulted in 42 records

SQL for Synthetic:
SELECT `native.country`,SUM(age), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`

Resulted in 41 records

Normalized Euclidean distance for (age): 6.4

Hellinger Distance: 0.396

================================================

Real

| | sex | education | `native.country` | race | `marital.status` | relationship | workclass | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|
| 1982 | Female | Bachelors | United-States | White | Never-married | Other-relative | State-gov | 208483.000000 | 2.000000 |
| 5441 | Male | Assoc-acdm | Peru | White | Never-married | Not-in-family | Private | 304082.000000 | 1.000000 |
| 8389 | Male | Some-college | Yugoslavia | Other | Never-married | Unmarried | Private | nan | nan |
| 8132 | Male | Some-college | Puerto-Rico | White | Married-spouse-absent | Own-child | Federal-gov | nan | nan |
| 3359 | Female | Masters | United-States | White | Married-spouse-absent | Unmarried | Self-emp-not-inc | nan | nan |

Synthetic

| | sex | education | `native.country` | race | `marital.status` | relationship | workclass | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|
| 1982 | Female | Bachelors | United-States | White | Never-married | Other-relative | State-gov | 196613.966348 | 1 |
| 5441 | Male | Assoc-acdm | Peru | White | Never-married | Not-in-family | Private | 204875.435507 | 1 |
| 8389 | Male | Some-college | Yugoslavia | Other | Never-married | Unmarried | Private | 180950.415632 | 1 |
| 8132 | Male | Some-college | Puerto-Rico | White | Married-spouse-absent | Own-child | Federal-gov | 176176.139524 | 1 |
| 3359 | Female | Masters | United-States | White | Married-spouse-absent | Unmarried | Self-emp-not-inc | 173062.821206 | 1 |

SQL for Real:
SELECT sex,education,`native.country`,race,`marital.status`,relationship,workclass,AVG(fnlwgt), COUNT(*) FROM C1 GROUP BY sex,education,`native.country`,race,`marital.status`,relationship,workclass

Resulted in 6002 records

SQL for Synthetic:
SELECT sex,education,`native.country`,race,`marital.status`,relationship,workclass,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY sex,education,`native.country`,race,`marital.status`,relationship,workclass

Resulted in 8391 records

Normalized Euclidean distance for (fnlwgt): 48.46

Hellinger Distance: 0.482

================================================

Real

| | workclass | fnlwgt | count |
|---|---|---|---|
| 5 | Self-emp-not-inc | 678086119 | 3862 |
| 2 | Local-gov | 596345319 | 3136 |
| 6 | State-gov | 360410194 | 1981 |
| 4 | Self-emp-inc | 303388390 | 1695 |
| 0 | ? | 529802400 | 2830 |

Synthetic

| | workclass | fnlwgt | count |
|---|---|---|---|
| 5 | Self-emp-not-inc | 658445746.320339 | 3818.000000 |
| 2 | Local-gov | 567199972.519647 | 3105.000000 |
| 6 | State-gov | 338630858.672139 | 1982.000000 |
| 4 | Self-emp-inc | 284019318.551430 | 1645.000000 |
| 0 | ? | 506616985.296546 | 2844.000000 |

SQL for Real:
SELECT workclass,SUM(fnlwgt), COUNT(*) FROM C1 GROUP BY workclass

Resulted in 7 records

SQL for Synthetic:
SELECT workclass,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY workclass

Resulted in 7 records

Normalized Euclidean distance for (fnlwgt): 2.65

Hellinger Distance: 0.003

==========================================

Real

| | workclass | `hours.per.week` | count |
|---|---|---|---|
| 2 | Local-gov | 2 | 3136 |
| 0 | ? | 1 | 2830 |
| 4 | Self-emp-inc | 1 | 1695 |
| 5 | Self-emp-not-inc | 1 | 3862 |
| 1 | Federal-gov | 4 | 1432 |

Synthetic

| | workclass | `hours.per.week` | count |
|---|---|---|---|
| 2 | Local-gov | 38.913042 | 3105.000000 |
| 0 | ? | 38.865705 | 2844.000000 |
| 4 | Self-emp-inc | 38.915161 | 1645.000000 |
| 5 | Self-emp-not-inc | 38.938152 | 3818.000000 |
| 1 | Federal-gov | 38.935892 | 1426.000000 |

SQL for Real:
SELECT workclass,MIN(`hours.per.week`), COUNT(*) FROM C1 GROUP BY workclass

Resulted in 7 records

SQL for Synthetic:
SELECT workclass,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY workclass

Resulted in 7 records

Normalized Euclidean distance for (`hours.per.week`): 2.65

Hellinger Distance: 0.003

==========================================

Real

| | relationship | workclass | education | `native.country` | age | count |
|---|---|---|---|---|---|---|
| 524 | Husband | Private | Prof-school | Germany | 50.000000 | 2.000000 |
| 798 | Husband | State-gov | Doctorate | United-States | 49.818182 | 55.000000 |
| 1499 | Not-in-family | Self-emp-not-inc | Prof-school | ? | 66.000000 | 1.000000 |
| 813 | Husband | State-gov | Masters | Peru | nan | nan |
| 377 | Husband | Private | 9th | Ireland | nan | nan |

Synthetic

| | relationship | workclass | education | `native.country` | age | count |
|---|---|---|---|---|---|---|
| 524 | Husband | Private | Prof-school | Germany | 34.000000 | 1 |
| 798 | Husband | State-gov | Doctorate | United-States | 50.121212 | 33 |
| 1499 | Not-in-family | Self-emp-not-inc | Prof-school | ? | 48.000000 | 1 |
| 813 | Husband | State-gov | Masters | Peru | 42.000000 | 1 |
| 377 | Husband | Private | 9th | Ireland | 26.000000 | 1 |

SQL for Real:
SELECT relationship,workclass,education,`native.country`,AVG(age), COUNT(*) FROM C1 GROUP BY relationship,workclass,education,`native.country`

Resulted in 2547 records

SQL for Synthetic:
SELECT relationship,workclass,education,`native.country`,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY relationship,workclass,education,`native.country`

Resulted in 3287 records

Normalized Euclidean distance for (age): 35.03

Hellinger Distance: 0.437

==========================================

Real

| | `native.country` | income | workclass | occupation | `marital.status` | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|
| 3058 | United-States | <=50K | State-gov | Craft-repair | Never-married | 745.000000 | 19.000000 |
| 2768 | United-States | <=50K | Federal-gov | Handlers-cleaners | Never-married | 437.000000 | 11.000000 |
| 2061 | Puerto-Rico | <=50K | Local-gov | Machine-op-inspct | Divorced | nan | nan |
| 1156 | Ireland | <=50K | Local-gov | Transport-moving | Divorced | nan | nan |
| 858 | Holand-Netherlands | <=50K | Private | Handlers-cleaners | Married-civ-spouse | nan | nan |

Synthetic

| | `native.country` | income | workclass | occupation | `marital.status` | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|
| 3058 | United-States | <=50K | State-gov | Craft-repair | Never-married | 239.957072 | 6 |
| 2768 | United-States | <=50K | Federal-gov | Handlers-cleaners | Never-married | 240.012326 | 6 |
| 2061 | Puerto-Rico | <=50K | Local-gov | Machine-op-inspct | Divorced | 40.000596 | 1 |
| 1156 | Ireland | <=50K | Local-gov | Transport-moving | Divorced | 39.947221 | 1 |
| 858 | Holand-Netherlands | <=50K | Private | Handlers-cleaners | Married-civ-spouse | 39.968699 | 1 |

SQL for Real:
SELECT `native.country`,income,workclass,occupation,`marital.status`,SUM(`hours.per.week`), COUNT(*) FROM C1 GROUP BY

`native.country`,income,workclass,occupation,`marital.status`

Resulted in 2676 records

SQL for Synthetic:
SELECT `native.country`,income,workclass,occupation,`marital.status`,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`,income,workclass,occupation,`marital.status`

Resulted in 3329 records

Normalized Euclidean distance for (`hours.per.week`): 35.85

Hellinger Distance: 0.457

==========================================

Real

| | education | workclass | occupation | relationship | `marital.status` | `native.country` | race | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|
| 7497 | HS-grad | Local-gov | Other-service | Unmarried | Widowed | United-States | White | 679570.000000 | 4.000000 |
| 467 | 10th | Private | Sales | Own-child | Never-married | United-States | Black | 393035.000000 | 2.000000 |
| 13785 | Some-college | Self-emp-inc | Prof-specialty | Husband | Married-civ-spouse | Columbia | White | nan | nan |
| 5590 | Bachelors | Private | Other-service | Not-in-family | Divorced | Puerto-Rico | White | nan | nan |
| 11183 | Masters | Self-emp-not-inc | Prof-specialty | Husband | Married-civ-spouse | Puerto-Rico | Black | nan | nan |

Synthetic

| | education | workclass | occupation | relationship | `marital.status` | `native.country` | race | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|
| 7497 | HS-grad | Local-gov | Other-service | Unmarried | Widowed | United-States | White | 693964.333869 | 4 |
| 467 | 10th | Private | Sales | Own-child | Never-married | United-States | Black | 559718.786063 | 3 |
| 13785 | Some-college | Self-emp-inc | Prof-specialty | Husband | Married-civ-spouse | Columbia | White | 355178.737856 | 2 |
| 5590 | Bachelors | Private | Other-service | Not-in-family | Divorced | Puerto-Rico | White | 343647.242145 | 2 |
| 11183 | Masters | Self-emp-not-inc | Prof-specialty | Husband | Married-civ-spouse | Puerto-Rico | Black | 169208.282081 | 1 |

SQL for Real:
SELECT education,workclass,occupation,relationship,`marital.status`,`native.country`,race,SUM(fnlwgt), COUNT(*) FROM C1 GROUP BY education,workclass,occupation,relationship,`marital.status`,`native.country`,race

Resulted in 10061 records

SQL for Synthetic:
SELECT education,workclass,occupation,relationship,`marital.status`,`native.country`,race,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY education,workclass,occupation,relationship,`marital.status`,`native.country`,race

Resulted in 14266 records

Normalized Euclidean distance for (fnlwgt): 60.38

Hellinger Distance: 0.519

==========================================

Real

| | sex | `hours.per.week` | count |
|---|---|---|---|
| 0 | Female | 99 | 16192 |
| 1 | Male | 99 | 32650 |

Synthetic

| | sex | `hours.per.week` | count |
|---|---|---|---|
| 0 | Female | 40.150468 | 17114.000000 |
| 1 | Male | 40.592070 | 31728.000000 |

SQL for Real:
SELECT sex,MAX(`hours.per.week`), COUNT(*) FROM C1 GROUP BY sex

Resulted in 2 records

SQL for Synthetic:
SELECT sex,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY sex

Resulted in 2 records

Normalized Euclidean distance for (`hours.per.week`): 1.41

Hellinger Distance: 0.014

==========================================

Real

| | `marital.status` | occupation | income | sex | education | race | workclass | relationship | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 2491 | Married-civ-spouse | Craft-repair | <=50K | Male | 9th | White | Private | Husband | 68 | 50 |
| 7537 | Never-married | Prof-specialty | >50K | Female | Bachelors | White | Private | Not-in-family | 45 | 13 |
| 7470 | Never-married | Prof-specialty | <=50K | Male | HS-grad | White | Federal-gov | Not-in-family | 23 | 1 |
| 9472 | Widowed | Handlers-cleaners | <=50K | Female | 11th | Black | Private | Not-in-family | 62 | 1 |
| 3246 | Married-civ-spouse | Handlers-cleaners | <=50K | Male | Some-college | White | Self-emp-not-inc | Husband | 45 | 1 |

Synthetic

| | `marital.status` | occupation | income | sex | education | race | workclass | relationship | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 2491 | Married-civ-spouse | Craft-repair | <=50K | Male | 9th | White | Private | Husband | 81.000000 | 59.000000 |
| 7537 | Never-married | Prof-specialty | >50K | Female | Bachelors | White | Private | Not-in-family | 42.000000 | 4.000000 |
| 7470 | Never-married | Prof-specialty | <=50K | Male | HS-grad | White | Federal-gov | Not-in-family | 18.000000 | 1.000000 |
| 9472 | Widowed | Handlers-cleaners | <=50K | Female | 11th | Black | Private | Not-in-family | nan | nan |
| 3246 | Married-civ-spouse | Handlers-cleaners | <=50K | Male | Some-college | White | Self-emp-not-inc | Husband | 34.000000 | 3.000000 |

SQL for Real:
SELECT `marital.status`,occupation,income,sex,education,race,workclass,relationship,MAX(age), COUNT(*) FROM C1 GROUP BY `marital.status`,occupation,income,sex,education,race,workclass,relationship

Resulted in 9897 records

SQL for Synthetic:
SELECT `marital.status`,occupation,income,sex,education,race,workclass,relationship,MAX(age), COUNT(*) FROM C1_syn_06 GROUP BY `marital.status`,occupation,income,sex,education,race,workclass,relationship

Resulted in 9113 records

Normalized Euclidean distance for (age): 69.99

Hellinger Distance: 0.328

==========================================

### Real

|  | `native.country` | race | sex | `marital.status` | workclass | income | capital | count |
|---|---|---|---|---|---|---|---|---|
| 2091 | United-States | White | Male | Married-spouse-absent | Self-emp-not-inc | >50K | 0.000000 | 4.000000 |
| 757 | Ireland | White | Female | Divorced | Local-gov | <=50K | nan | nan |
| 639 | Hungary | White | Male | Divorced | Private | >50K | nan | nan |
| 250 | Columbia | Black | Female | Married-civ-spouse | Private | <=50K | nan | nan |
| 1229 | Philippines | White | Female | Married-civ-spouse | Local-gov | >50K | nan | nan |

### Synthetic

|  | `native.country` | race | sex | `marital.status` | workclass | income | capital | count |
|---|---|---|---|---|---|---|---|---|
| 2091 | United-States | White | Male | Married-spouse-absent | Self-emp-not-inc | >50K | -11.114547 | 1 |
| 757 | Ireland | White | Female | Divorced | Local-gov | <=50K | 198.609019 | 7 |
| 639 | Hungary | White | Male | Divorced | Private | >50K | -125.971624 | 2 |
| 250 | Columbia | Black | Female | Married-civ-spouse | Private | <=50K | 62.614144 | 6 |
| 1229 | Philippines | White | Female | Married-civ-spouse | Local-gov | >50K | -60.287934 | 2 |

SQL for Real:
SELECT `native.country`,race,sex,`marital.status`,workclass,income,MAX(capital), COUNT(*) FROM C1 GROUP BY `native.country`,race,sex,`marital.status`,workclass,income

Resulted in 1867 records

SQL for Synthetic:
SELECT `native.country`,race,sex,`marital.status`,workclass,income,MAX(capital), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`,race,sex,`marital.status`,workclass,income

Resulted in 2171 records

Normalized Euclidean distance for (capital): 28.65

Hellinger Distance: 0.455

==========================================

### Real

|  | race | workclass | sex | income | occupation | relationship | `native.country` | `marital.status` | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 5900 | White | Private | Male | <=50K | Adm-clerical | Unmarried | Columbia | Married-spouse-absent | nan | nan |
| 7206 | White | Private | Male | >50K | Craft-repair | Unmarried | Puerto-Rico | Divorced | nan | nan |
| 8160 | White | Self-emp-not-inc | Male | <=50K | Farming-fishing | Own-child | Japan | Never-married | nan | nan |
| 6970 | White | Private | Male | <=50K | Tech-support | Husband | Ecuador | Married-civ-spouse | nan | nan |
| 4474 | White | Private | Female | <=50K | Craft-repair | Not-in-family | United-States | Divorced | 43.680000 | 25.000000 |

### Synthetic

|  | race | workclass | sex | income | occupation | relationship | `native.country` | `marital.status` | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 5900 | White | Private | Male | <=50K | Adm-clerical | Unmarried | Columbia | Married-spouse-absent | 56.000000 | 1 |
| 7206 | White | Private | Male | >50K | Craft-repair | Unmarried | Puerto-Rico | Divorced | 46.000000 | 1 |
| 8160 | White | Self-emp-not-inc | Male | <=50K | Farming-fishing | Own-child | Japan | Never-married | 38.000000 | 1 |
| 6970 | White | Private | Male | <=50K | Tech-support | Husband | Ecuador | Married-civ-spouse | 29.000000 | 2 |
| 4474 | White | Private | Female | <=50K | Craft-repair | Not-in-family | United-States | Divorced | 41.333333 | 12 |

SQL for Real:
SELECT race,workclass,sex,income,occupation,relationship,`native.country`,`marital.status`,AVG(age), COUNT(*) FROM C1 GROUP BY race,workclass,sex,income,occupation,relationship,`native.country`,`marital.status`

Resulted in 6492 records

SQL for Synthetic:
SELECT race,workclass,sex,income,occupation,relationship,`native.country`,`marital.status`,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY race,workclass,sex,income,occupation,relationship,`native.country`,`marital.status`

Resulted in 8900 records

Normalized Euclidean distance for (age): 49.69

Hellinger Distance: 0.502

==========================================

### Real

|  | race | education | age | count |
|---|---|---|---|---|
| 79 | White | Some-college | 17 | 9289 |
| 23 | Asian-Pac-Islander | Assoc-acdm | 22 | 49 |
| 31 | Asian-Pac-Islander | Some-college | 17 | 302 |
| 72 | White | Assoc-voc | 19 | 1803 |
| 42 | Black | Doctorate | 26 | 16 |

|     | race | Synthetic education | age | count |
|-----|------|-----------|-----|-------|
| 79  | White | Some-college | 17.000000 | 9391.000000 |
| 23  | Asian-Pac-Islander | Assoc-acdm | 20.000000 | 32.000000 |
| 31  | Asian-Pac-Islander | Some-college | 18.000000 | 148.000000 |
| 72  | White | Assoc-voc | 17.000000 | 1764.000000 |
| 42  | Black | Doctorate | 31.000000 | 15.000000 |

SQL for Real:
SELECT race,education,MIN(age), COUNT(*) FROM C1 GROUP BY race,education

Resulted in 80 records

SQL for Synthetic:
SELECT race,education,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY race,education

Resulted in 77 records

Normalized Euclidean distance for (age): 8.77

Hellinger Distance: 0.053

==========================================

|     | sex | race | workclass | income | Real capital | count |
|-----|-----|------|-----------|--------|---------|-------|
| 84  | Male | Asian-Pac-Islander | Self-emp-inc | >50K | 7602.757576 | 33 |
| 11  | Female | Amer-Indian-Eskimo | State-gov | >50K | 0.000000 | 3 |
| 117 | Male | White | ? | <=50K | 78.526042 | 1152 |
| 18  | Female | Asian-Pac-Islander | Private | >50K | 1225.106383 | 47 |
| 75  | Male | Asian-Pac-Islander | ? | <=50K | -49.704545 | 44 |

|     | sex | race | workclass | income | Synthetic capital | count |
|-----|-----|------|-----------|--------|---------|-------|
| 84  | Male | Asian-Pac-Islander | Self-emp-inc | >50K | 55.721717 | 8.000000 |
| 11  | Female | Amer-Indian-Eskimo | State-gov | >50K | nan | nan |
| 117 | Male | White | ? | <=50K | 5.804332 | 1247.000000 |
| 18  | Female | Asian-Pac-Islander | Private | >50K | -115.704524 | 4.000000 |
| 75  | Male | Asian-Pac-Islander | ? | <=50K | 24.960637 | 39.000000 |

SQL for Real:
SELECT sex,race,workclass,income,AVG(capital), COUNT(*) FROM C1 GROUP BY sex,race,workclass,income
Resulted in 131 records

SQL for Synthetic:
SELECT sex,race,workclass,income,AVG(capital), COUNT(*) FROM C1_syn_06 GROUP BY sex,race,workclass,income
Resulted in 122 records

Normalized Euclidean distance for (capital): 10.95

Hellinger Distance: 0.15

==========================================

|      | sex | `marital.status` | workclass | occupation | Real `native.country` | age | count |
|------|-----|-----------------|-----------|------------|-----------------|-----|-------|
| 634  | Female | Married-civ-spouse | Self-emp-inc | Exec-managerial | ? | 42.666667 | 3.000000 |
| 943  | Female | Never-married | Local-gov | Adm-clerical | Yugoslavia | nan | nan |
| 3241 | Male | Never-married | Private | Machine-op-inspct | Philippines | 28.000000 | 7.000000 |
| 3498 | Male | Never-married | State-gov | Other-service | Columbia | nan | nan |
| 1345 | Female | Never-married | Self-emp-not-inc | Prof-specialty | United-States | 35.837838 | 37.000000 |

|      | sex | `marital.status` | workclass | occupation | Synthetic `native.country` | age | count |
|------|-----|-----------------|-----------|------------|-----------------|-----|-------|
| 634  | Female | Married-civ-spouse | Self-emp-inc | Exec-managerial | ? | 28.000000 | 2 |
| 943  | Female | Never-married | Local-gov | Adm-clerical | Yugoslavia | 35.000000 | 1 |
| 3241 | Male | Never-married | Private | Machine-op-inspct | Philippines | 45.000000 | 1 |
| 3498 | Male | Never-married | State-gov | Other-service | Columbia | 32.000000 | 1 |
| 1345 | Female | Never-married | Self-emp-not-inc | Prof-specialty | United-States | 37.142857 | 21 |

SQL for Real:
SELECT sex,`marital.status`,workclass,occupation,`native.country`,AVG(age), COUNT(*) FROM C1 GROUP BY sex,`marital.status`,workclass,occupation,`native.country`
Resulted in 2817 records

SQL for Synthetic:
SELECT sex,`marital.status`,workclass,occupation,`native.country`,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY
sex,`marital.status`,workclass,occupation,`native.country`

Resulted in 3807 records

Normalized Euclidean distance for (age): 37.13

Hellinger Distance: 0.446

==========================================

Real

| | `native.country` | occupation | sex | income | `marital.status` | workclass | relationship | education | race | capital | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 15243 | United-States | Prof-specialty | Male | <=50K | Married-civ-spouse | Federal-gov | Husband | Prof-school | White | 0.000000 | 2.000000 |
| 7350 | Puerto-Rico | Exec-managerial | Female | <=50K | Married-civ-spouse | Self-emp-inc | Wife | Bachelors | White | nan | nan |
| 2435 | Guatemala | Sales | Male | <=50K | Married-civ-spouse | Private | Husband | HS-grad | White | 0.000000 | 1.000000 |
| 4698 | Laos | Exec-managerial | Female | <=50K | Divorced | Private | Not-in-family | Bachelors | White | nan | nan |
| 11146 | United-States | Adm-clerical | Female | <=50K | Never-married | Private | Not-in-family | HS-grad | Asian-Pac-Islander | 0.000000 | 1.000000 |

Synthetic

| | `native.country` | occupation | sex | income | `marital.status` | workclass | relationship | education | race | capital | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 15243 | United-States | Prof-specialty | Male | <=50K | Married-civ-spouse | Federal-gov | Husband | Prof-school | White | 98.457523 | 7 |
| 7350 | Puerto-Rico | Exec-managerial | Female | <=50K | Married-civ-spouse | Self-emp-inc | Wife | Bachelors | White | 138.526103 | 1 |
| 2435 | Guatemala | Sales | Male | <=50K | Married-civ-spouse | Private | Husband | HS-grad | White | -113.292638 | 3 |
| 4698 | Laos | Exec-managerial | Female | <=50K | Divorced | Private | Not-in-family | Bachelors | White | -9.636421 | 1 |
| 11146 | United-States | Adm-clerical | Female | <=50K | Never-married | Private | Not-in-family | HS-grad | Asian-Pac-Islander | 57.569155 | 1 |

SQL for Real:
SELECT `native.country`,occupation,sex,income,`marital.status`,workclass,relationship,education,race,MIN(capital), COUNT(*) FROM C1 GROUP BY `native.country`,occupation,sex,income,`marital.status`,workclass,relationship,education,race

Resulted in 12664 records

SQL for Synthetic:
SELECT `native.country`,occupation,sex,income,`marital.status`,workclass,relationship,education,race,MIN(capital), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`,occupation,sex,income,`marital.status`,workclass,relationship,education,race

Resulted in 17127 records

Normalized Euclidean distance for (capital): 64.61

Hellinger Distance: 0.549

================================================

Real

| | occupation | `marital.status` | workclass | sex | race | fnlwgt | count |
|---|---|---|---|---|---|---|---|
| 1302 | Protective-serv | Married-spouse-absent | Local-gov | Male | White | 202207 | 1 |
| 1577 | Tech-support | Separated | Private | Male | Asian-Pac-Islander | 341672 | 1 |
| 61 | Adm-clerical | Divorced | Federal-gov | Male | White | 3741548 | 20 |
| 125 | Adm-clerical | Married-civ-spouse | Self-emp-not-inc | Female | White | 4883377 | 23 |
| 1094 | Prof-specialty | Divorced | State-gov | Female | Asian-Pac-Islander | 700765 | 4 |

Synthetic

| | occupation | `marital.status` | workclass | sex | race | fnlwgt | count |
|---|---|---|---|---|---|---|---|
| 1302 | Protective-serv | Married-spouse-absent | Local-gov | Male | White | 194883.976133 | 1.000000 |
| 1577 | Tech-support | Separated | Private | Male | Asian-Pac-Islander | 176284.846677 | 1.000000 |
| 61 | Adm-clerical | Divorced | Federal-gov | Male | White | 3043524.538681 | 18.000000 |
| 125 | Adm-clerical | Married-civ-spouse | Self-emp-not-inc | Female | White | 3741210.846971 | 22.000000 |
| 1094 | Prof-specialty | Divorced | State-gov | Female | Asian-Pac-Islander | 162009.465884 | 1.000000 |

SQL for Real:
SELECT occupation,`marital.status`,workclass,sex,race,SUM(fnlwgt), COUNT(*) FROM C1 GROUP BY occupation,`marital.status`,workclass,sex,race
Resulted in 1695 records

SQL for Synthetic:
SELECT occupation,`marital.status`,workclass,sex,race,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY occupation,`marital.status`,workclass,sex,race
Resulted in 1584 records

Normalized Euclidean distance for (fnlwgt): 35.06

Hellinger Distance: 0.119

================================================

Real

| | income | relationship | sex | fnlwgt | count |
|---|---|---|---|---|---|
| 23 | >50K | Wife | Male | 350162.000000 | 1.000000 |
| 9 | <=50K | Unmarried | Male | 195616.585714 | 1050.000000 |
| 8 | <=50K | Unmarried | Female | 190393.823686 | 3766.000000 |
| 14 | >50K | Not-in-family | Female | 182632.548889 | 450.000000 |
| 11 | <=50K | Wife | Male | 113481.500000 | 2.000000 |

Synthetic

| | income | relationship | sex | fnlwgt | count |
|---|---|---|---|---|---|
| 23 | >50K | Wife | Male | 179441.058854 | 3 |
| 9 | <=50K | Unmarried | Male | 183636.319012 | 1155 |
| 8 | <=50K | Unmarried | Female | 179919.004125 | 3760 |
| 14 | >50K | Not-in-family | Female | 175831.078344 | 134 |
| 11 | <=50K | Wife | Male | 177956.939889 | 7 |

SQL for Real:
SELECT income,relationship,sex,AVG(fnlwgt), COUNT(*) FROM C1 GROUP BY income,relationship,sex
Resulted in 23 records

SQL for Synthetic:
SELECT income,relationship,sex,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY income,relationship,sex
Resulted in 24 records

Normalized Euclidean distance for (fnlwgt): 4.8

Hellinger Distance: 0.149

========================================

Real

| | `native.country` | relationship | income | `hours.per.week` | count |
|---|---|---|---|---|---|
| 310 | Scotland | Wife | >50K | 40 | 1 |
| 105 | Germany | Husband | >50K | 15 | 44 |
| 52 | Cuba | Husband | >50K | 24 | 27 |
| 290 | Portugal | Unmarried | >50K | 50 | 1 |
| 22 | Canada | Husband | >50K | 20 | 42 |

Synthetic

| | `native.country` | relationship | income | `hours.per.week` | count |
|---|---|---|---|---|---|
| 310 | Scotland | Wife | >50K | 39.931315 | 11.000000 |
| 105 | Germany | Husband | >50K | 40.338012 | 2.000000 |
| 52 | Cuba | Husband | >50K | 40.010078 | 3.000000 |
| 290 | Portugal | Unmarried | >50K | nan | nan |
| 22 | Canada | Husband | >50K | 39.976995 | 3.000000 |

SQL for Real:
SELECT `native.country`,relationship,income,MIN(`hours.per.week`), COUNT(*) FROM C1 GROUP BY `native.country`,relationship,income

Resulted in 372 records

SQL for Synthetic:
SELECT `native.country`,relationship,income,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`,relationship,income

Resulted in 299 records

Normalized Euclidean distance for (`hours.per.week`): 16.16

Hellinger Distance: 0.421

========================================

Real

| | relationship | `native.country` | workclass | race | occupation | sex | education | income | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 10408 | Own-child | United-States | Private | White | Other-service | Male | Some-college | <=50K | 22.370629 | 143.000000 |
| 1852 | Husband | Puerto-Rico | Federal-gov | White | Craft-repair | Male | Some-college | >50K | nan | nan |
| 6372 | Not-in-family | Scotland | Local-gov | White | Adm-clerical | Female | Some-college | <=50K | nan | nan |
| 6083 | Not-in-family | Puerto-Rico | Private | White | Prof-specialty | Male | Some-college | <=50K | nan | nan |
| 11398 | Unmarried | Peru | Local-gov | White | Other-service | Female | HS-grad | <=50K | nan | nan |

Synthetic

| | relationship | `native.country` | workclass | race | occupation | sex | education | income | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 10408 | Own-child | United-States | Private | White | Other-service | Male | Some-college | <=50K | 21.932584 | 89 |
| 1852 | Husband | Puerto-Rico | Federal-gov | White | Craft-repair | Male | Some-college | >50K | 53.000000 | 1 |
| 6372 | Not-in-family | Scotland | Local-gov | White | Adm-clerical | Female | Some-college | <=50K | 65.000000 | 1 |
| 6083 | Not-in-family | Puerto-Rico | Private | White | Prof-specialty | Male | Some-college | <=50K | 43.000000 | 3 |
| 11398 | Unmarried | Peru | Local-gov | White | Other-service | Female | HS-grad | <=50K | 52.000000 | 1 |

SQL for Real:
SELECT relationship,`native.country`,workclass,race,occupation,sex,education,income,AVG(age), COUNT(*) FROM C1 GROUP BY relationship,`native.country`,workclass,race,occupation,sex,education,income

Resulted in 9992 records

SQL for Synthetic:
SELECT relationship,`native.country`,workclass,race,occupation,sex,education,income,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY relationship,`native.country`,workclass,race,occupation,sex,education,income

Resulted in 13906 records

Normalized Euclidean distance for (age): 60.46

Hellinger Distance: 0.528

========================================

Real

| | relationship | race | sex | workclass | capital | count |
|---|---|---|---|---|---|---|
| 6 | Husband | Amer-Indian-Eskimo | Male | State-gov | 0 | 10 |
| 31 | Husband | White | Male | Local-gov | 99999 | 1131 |
| 289 | Wife | Black | Female | State-gov | 3103 | 19 |
| 182 | Own-child | Black | Female | State-gov | 0 | 16 |
| 138 | Other-relative | Other | Male | Private | 0 | 20 |

Synthetic

| | relationship | race | sex | workclass | capital | count |
|---|---|---|---|---|---|---|
| 6 | Husband | Amer-Indian-Eskimo | Male | State-gov | 234.669398 | 7.000000 |
| 31 | Husband | White | Male | Local-gov | 448.600229 | 1095.000000 |
| 289 | Wife | Black | Female | State-gov | 207.718738 | 14.000000 |
| 182 | Own-child | Black | Female | State-gov | 192.543029 | 24.000000 |
| 138 | Other-relative | Other | Male | Private | 298.366398 | 16.000000 |

SQL for Real:
SELECT relationship,race,sex,workclass,MAX(capital), COUNT(*) FROM C1 GROUP BY relationship,race,sex,workclass

Resulted in 302 records

SQL for Synthetic:

SELECT relationship,race,sex,workclass,MAX(capital), COUNT(*) FROM C1_syn_06 GROUP BY relationship,race,sex,workclass

Resulted in 289 records

Normalized Euclidean distance for (capital): 16.31

Hellinger Distance: 0.065

==============================================

### Real

| | race | workclass | relationship | `hours.per.week` | count |
|---|---|---|---|---|---|
| 25 | Amer-Indian-Eskimo | Self-emp-inc | Wife | 40 | 1 |
| 183 | White | Self-emp-not-inc | Own-child | 3 | 223 |
| 105 | Black | Self-emp-inc | Other-relative | 40 | 1 |
| 125 | Other | ? | Unmarried | 40 | 2 |
| 175 | White | Self-emp-inc | Not-in-family | 5 | 238 |

### Synthetic

| | race | workclass | relationship | `hours.per.week` | count |
|---|---|---|---|---|---|
| 25 | Amer-Indian-Eskimo | Self-emp-inc | Wife | nan | nan |
| 183 | White | Self-emp-not-inc | Own-child | 38.938152 | 268.000000 |
| 105 | Black | Self-emp-inc | Other-relative | 39.955479 | 1.000000 |
| 125 | Other | ? | Unmarried | 39.521403 | 2.000000 |
| 175 | White | Self-emp-inc | Not-in-family | 39.040644 | 248.000000 |

SQL for Real:
SELECT race,workclass,relationship,MIN(`hours.per.week`), COUNT(*) FROM C1 GROUP BY race,workclass,relationship

Resulted in 192 records

SQL for Synthetic:
SELECT race,workclass,relationship,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY race,workclass,relationship

Resulted in 181 records

Normalized Euclidean distance for (`hours.per.week`): 13.27

Hellinger Distance: 0.057

==============================================

### Real

| | education | sex | `marital.status` | race | workclass | occupation | `native.country` | income | relationship | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 13344 | Masters | Male | Married-spouse-absent | Asian-Pac-Islander | Private | Prof-specialty | United-States | <=50K | Not-in-family | nan | nan |
| 13813 | Prof-school | Male | Married-civ-spouse | White | Self-emp-inc | Exec-managerial | Peru | <=50K | Husband | nan | nan |
| 16436 | Some-college | Male | Never-married | Asian-Pac-Islander | Private | Adm-clerical | Columbia | <=50K | Own-child | nan | nan |
| 2799 | 9th | Female | Divorced | White | Private | Other-service | Puerto-Rico | <=50K | Unmarried | nan | nan |
| 4994 | Assoc-voc | Male | Never-married | Asian-Pac-Islander | Self-emp-inc | Sales | United-States | <=50K | Own-child | nan | nan |

### Synthetic

| | education | sex | `marital.status` | race | workclass | occupation | `native.country` | income | relationship | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 13344 | Masters | Male | Married-spouse-absent | Asian-Pac-Islander | Private | Prof-specialty | United-States | <=50K | Not-in-family | 193613.583186 | 1 |
| 13813 | Prof-school | Male | Married-civ-spouse | White | Self-emp-inc | Exec-managerial | Peru | <=50K | Husband | 178789.821231 | 1 |
| 16436 | Some-college | Male | Never-married | Asian-Pac-Islander | Private | Adm-clerical | Columbia | <=50K | Own-child | 177590.512608 | 1 |
| 2799 | 9th | Female | Divorced | White | Private | Other-service | Puerto-Rico | <=50K | Unmarried | 173157.046856 | 2 |
| 4994 | Assoc-voc | Male | Never-married | Asian-Pac-Islander | Self-emp-inc | Sales | United-States | <=50K | Own-child | 186483.684155 | 1 |

SQL for Real:
SELECT education,sex,`marital.status`,race,workclass,occupation,`native.country`,income,relationship,AVG(fnlwgt), COUNT(*) FROM C1 GROUP BY education,sex,`marital.status`,race,workclass,occupation,`native.country`,income,relationship

Resulted in 12664 records

SQL for Synthetic:
SELECT education,sex,`marital.status`,race,workclass,occupation,`native.country`,income,relationship,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY education,sex,`marital.status`,race,workclass,occupation,`native.country`,income,relationship

Resulted in 17127 records

Normalized Euclidean distance for (fnlwgt): 64.61

Hellinger Distance: 0.549

==============================================

### Real

| | education | income | sex | race | `marital.status` | occupation | workclass | `native.country` | relationship | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8499 | HS-grad | <=50K | Female | Black | Separated | Machine-op-inspct | Private | Puerto-Rico | Unmarried | nan | nan |
| 616 | 10th | <=50K | Male | White | Never-married | Other-service | Private | United-States | Other-relative | 26.250000 | 4.000000 |
| 12939 | Masters | <=50K | Male | White | Divorced | ? | ? | United-States | Other-relative | nan | nan |
| 4339 | Assoc-voc | <=50K | Female | White | Divorced | Machine-op-inspct | Private | United-States | Unmarried | nan | nan |
| 6401 | Bachelors | <=50K | Male | Black | Married-civ-spouse | Sales | Private | Puerto-Rico | Husband | nan | nan |

### Synthetic

| | education | income | sex | race | `marital.status` | occupation | workclass | `native.country` | relationship | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8499 | HS-grad | <=50K | Female | Black | Separated | Machine-op-inspct | Private | Puerto-Rico | Unmarried | 39.954001 | 1 |
| 616 | 10th | <=50K | Male | White | Never-married | Other-service | Private | United-States | Other-relative | 39.360393 | 3 |
| 12939 | Masters | <=50K | Male | White | Divorced | ? | ? | United-States | Other-relative | 39.990727 | 1 |
| 4339 | Assoc-voc | <=50K | Female | White | Divorced | Machine-op-inspct | Private | United-States | Unmarried | 39.999587 | 3 |
| 6401 | Bachelors | <=50K | Male | Black | Married-civ-spouse | Sales | Private | Puerto-Rico | Husband | 40.393387 | 2 |

SQL for Real:

SELECT education,income,sex,race,`marital.status`,occupation,workclass,`native.country`,relationship,AVG(`hours.per.week`), COUNT(*) FROM C1 GROUP BY education,income,sex,race,`marital.status`,occupation,workclass,`native.country`,relationship

Resulted in 12664 records

SQL for Synthetic:
SELECT education,income,sex,race,`marital.status`,occupation,workclass,`native.country`,relationship,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY education,income,sex,race,`marital.status`,occupation,workclass,`native.country`,relationship

Resulted in 17127 records

Normalized Euclidean distance for (`hours.per.week`): 64.61

Hellinger Distance: 0.549

==========================================

### Real

| | sex | `marital.status` | race | relationship | income | age | count |
|---|---|---|---|---|---|---|---|
| 184 | Male | Divorced | Other | Own-child | <=50K | 67 | 2 |
| 9 | Female | Divorced | Asian-Pac-Islander | Own-child | <=50K | 224 | 7 |
| 217 | Male | Married-civ-spouse | Other | Other-relative | >50K | 27 | 1 |
| 74 | Female | Married-spouse-absent | Black | Not-in-family | <=50K | 462 | 11 |
| 318 | Male | Widowed | Asian-Pac-Islander | Other-relative | >50K | 75 | 1 |

### Synthetic

| | sex | `marital.status` | race | relationship | income | age | count |
|---|---|---|---|---|---|---|---|
| 184 | Male | Divorced | Other | Own-child | <=50K | nan | nan |
| 9 | Female | Divorced | Asian-Pac-Islander | Own-child | <=50K | 102.000000 | 3.000000 |
| 217 | Male | Married-civ-spouse | Other | Other-relative | >50K | nan | nan |
| 74 | Female | Married-spouse-absent | Black | Not-in-family | <=50K | 456.000000 | 11.000000 |
| 318 | Male | Widowed | Asian-Pac-Islander | Other-relative | >50K | nan | nan |

SQL for Real:
SELECT sex,`marital.status`,race,relationship,income,SUM(age), COUNT(*) FROM C1 GROUP BY sex,`marital.status`,race,relationship,income

Resulted in 334 records

SQL for Synthetic:
SELECT sex,`marital.status`,race,relationship,income,SUM(age), COUNT(*) FROM C1_syn_06 GROUP BY sex,`marital.status`,race,relationship,income

Resulted in 306 records

Normalized Euclidean distance for (age): 16.0

Hellinger Distance: 0.164

==========================================

### Real

| | income | workclass | `native.country` | occupation | `marital.status` | sex | education | age | count |
|---|---|---|---|---|---|---|---|---|---|
| 3042 | <=50K | Private | Ecuador | Craft-repair | Married-civ-spouse | Male | Assoc-voc | nan | nan |
| 4252 | <=50K | Private | Japan | Other-service | Married-civ-spouse | Male | Assoc-acdm | nan | nan |
| 2004 | <=50K | Local-gov | United-States | Other-service | Divorced | Female | Prof-school | nan | nan |
| 1368 | <=50K | Local-gov | Japan | Exec-managerial | Divorced | Female | 9th | nan | nan |
| 4976 | <=50K | Private | Peru | Machine-op-inspct | Married-civ-spouse | Male | Assoc-acdm | nan | nan |

### Synthetic

| | income | workclass | `native.country` | occupation | `marital.status` | sex | education | age | count |
|---|---|---|---|---|---|---|---|---|---|
| 3042 | <=50K | Private | Ecuador | Craft-repair | Married-civ-spouse | Male | Assoc-voc | 29 | 1 |
| 4252 | <=50K | Private | Japan | Other-service | Married-civ-spouse | Male | Assoc-acdm | 32 | 1 |
| 2004 | <=50K | Local-gov | United-States | Other-service | Divorced | Female | Prof-school | 44 | 1 |
| 1368 | <=50K | Local-gov | Japan | Exec-managerial | Divorced | Female | 9th | 57 | 1 |
| 4976 | <=50K | Private | Peru | Machine-op-inspct | Married-civ-spouse | Male | Assoc-acdm | 41 | 1 |

SQL for Real:
SELECT income,workclass,`native.country`,occupation,`marital.status`,sex,education,MIN(age), COUNT(*) FROM C1 GROUP BY income,workclass,`native.country`,occupation,`marital.status`,sex,education

Resulted in 7960 records

SQL for Synthetic:
SELECT income,workclass,`native.country`,occupation,`marital.status`,sex,education,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY income,workclass,`native.country`,occupation,`marital.status`,sex,education

Resulted in 11425 records

Normalized Euclidean distance for (age): 56.2

Hellinger Distance: 0.509

==========================================

### Real

| | `native.country` | relationship | income | fnlwgt | count |
|---|---|---|---|---|---|
| 30 | Canada | Wife | >50K | 1840042 | 11 |
| 28 | Canada | Unmarried | >50K | 118139 | 2 |
| 193 | Italy | Not-in-family | >50K | 635318 | 4 |
| 245 | Nicaragua | Wife | >50K | 376302 | 1 |
| 254 | Peru | Husband | >50K | 1284191 | 3 |

| | `native.country` | relationship | income | fnlwgt | count |
|---|---|---|---|---|---|
| 30 | Canada | Wife | >50K | nan | nan |
| 28 | Canada | Unmarried | >50K | nan | nan |
| 193 | Italy | Not-in-family | >50K | nan | nan |
| 245 | Nicaragua | Wife | >50K | nan | nan |
| 254 | Peru | Husband | >50K | 11819181.656750 | 66.000000 |

SQL for Real:
SELECT `native.country`,relationship,income,SUM(fnlwgt), COUNT(*) FROM C1 GROUP BY `native.country`,relationship,income

Resulted in 372 records

SQL for Synthetic:
SELECT `native.country`,relationship,income,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`,relationship,income

Resulted in 299 records

Normalized Euclidean distance for (fnlwgt): 16.16

Hellinger Distance: 0.421

===========================================

Real

| | race | sex | fnlwgt | count |
|---|---|---|---|---|
| 9 | White | Male | 13769 | 28735 |
| 1 | Amer-Indian-Eskimo | Male | 13492 | 285 |
| 6 | Other | Female | 24562 | 155 |
| 5 | Black | Male | 21856 | 2377 |
| 0 | Amer-Indian-Eskimo | Female | 12285 | 185 |

Synthetic

| | race | sex | fnlwgt | count |
|---|---|---|---|---|
| 9 | White | Male | 117274.397609 | 28906.000000 |
| 1 | Amer-Indian-Eskimo | Male | 146906.905111 | 133.000000 |
| 6 | Other | Female | 134958.723599 | 159.000000 |
| 5 | Black | Male | 132147.182998 | 1830.000000 |
| 0 | Amer-Indian-Eskimo | Female | 160187.303733 | 69.000000 |

SQL for Real:
SELECT race,sex,MIN(fnlwgt), COUNT(*) FROM C1 GROUP BY race,sex

Resulted in 10 records

SQL for Synthetic:
SELECT race,sex,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY race,sex

Resulted in 10 records

Normalized Euclidean distance for (fnlwgt): 3.16

Hellinger Distance: 0.045

===========================================

Real

| | workclass | `marital.status` | occupation | `native.country` | age | count |
|---|---|---|---|---|---|---|
| 2565 | State-gov | Divorced | Other-service | Puerto-Rico | nan | nan |
| 1698 | Private | Never-married | Sales | Columbia | 21.000000 | 2.000000 |
| 2540 | State-gov | Divorced | Adm-clerical | ? | 45.000000 | 1.000000 |
| 611 | Local-gov | Married-spouse-absent | Transport-moving | Puerto-Rico | nan | nan |
| 2758 | State-gov | Separated | Exec-managerial | Puerto-Rico | nan | nan |

Synthetic

| | workclass | `marital.status` | occupation | `native.country` | age | count |
|---|---|---|---|---|---|---|
| 2565 | State-gov | Divorced | Other-service | Puerto-Rico | 29 | 11 |
| 1698 | Private | Never-married | Sales | Columbia | 26 | 27 |
| 2540 | State-gov | Divorced | Adm-clerical | ? | 40 | 6 |
| 611 | Local-gov | Married-spouse-absent | Transport-moving | Puerto-Rico | 61 | 1 |
| 2758 | State-gov | Separated | Exec-managerial | Puerto-Rico | 33 | 3 |

SQL for Real:
SELECT workclass,`marital.status`,occupation,`native.country`,MIN(age), COUNT(*) FROM C1 GROUP BY workclass,`marital.status`,occupation,`native.country`

Resulted in 2171 records

SQL for Synthetic:
SELECT workclass,`marital.status`,occupation,`native.country`,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY workclass,`marital.status`,occupation,`native.country`

Resulted in 2790 records

Normalized Euclidean distance for (age): 33.36

Hellinger Distance: 0.436

===========================================

### Real

| | occupation | `marital.status` | race | fnlwgt | count |
|---|---|---|---|---|---|
| 267 | Prof-specialty | Never-married | Black | 217159.601399 | 143 |
| 94 | Exec-managerial | Divorced | Black | 212223.720588 | 68 |
| 74 | Craft-repair | Married-civ-spouse | Other | 250221.628571 | 35 |
| 278 | Prof-specialty | Widowed | Other | 142909.000000 | 2 |
| 164 | Handlers-cleaners | Separated | Black | 265429.705882 | 17 |

### Synthetic

| | occupation | `marital.status` | race | fnlwgt | count |
|---|---|---|---|---|---|
| 267 | Prof-specialty | Never-married | Black | 182931.793926 | 128.000000 |
| 94 | Exec-managerial | Divorced | Black | 176166.072491 | 56.000000 |
| 74 | Craft-repair | Married-civ-spouse | Other | 181289.161511 | 25.000000 |
| 278 | Prof-specialty | Widowed | Other | 170709.366902 | 1.000000 |
| 164 | Handlers-cleaners | Separated | Black | 181554.390088 | 13.000000 |

SQL for Real:
SELECT occupation,`marital.status`,race,AVG(fnlwgt), COUNT(*) FROM C1 GROUP BY occupation,`marital.status`,race

Resulted in 383 records

SQL for Synthetic:
SELECT occupation,`marital.status`,race,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY occupation,`marital.status`,race

Resulted in 359 records

Normalized Euclidean distance for (fnlwgt): 18.25

Hellinger Distance: 0.07

==========================================

### Real

| | income | `native.country` | relationship | `hours.per.week` | count |
|---|---|---|---|---|---|
| 113 | <=50K | Ireland | Husband | 25 | 5 |
| 193 | <=50K | Scotland | Own-child | 20 | 1 |
| 356 | >50K | Taiwan | Not-in-family | 40 | 2 |
| 324 | >50K | Mexico | Unmarried | 35 | 3 |
| 327 | >50K | Nicaragua | Wife | 35 | 1 |

### Synthetic

| | income | `native.country` | relationship | `hours.per.week` | count |
|---|---|---|---|---|---|
| 113 | <=50K | Ireland | Husband | 39.912408 | 109.000000 |
| 193 | <=50K | Scotland | Own-child | 39.450670 | 34.000000 |
| 356 | >50K | Taiwan | Not-in-family | nan | nan |
| 324 | >50K | Mexico | Unmarried | nan | nan |
| 327 | >50K | Nicaragua | Wife | nan | nan |

SQL for Real:
SELECT income,`native.country`,relationship,MIN(`hours.per.week`), COUNT(*) FROM C1 GROUP BY income,`native.country`,relationship

Resulted in 372 records

SQL for Synthetic:
SELECT income,`native.country`,relationship,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY income,`native.country`,relationship

Resulted in 299 records

Normalized Euclidean distance for (`hours.per.week`): 16.16

Hellinger Distance: 0.421

==========================================

### Real

| | race | workclass | income | occupation | `native.country` | relationship | sex | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|
| 2329 | White | ? | <=50K | ? | Peru | Not-in-family | Female | nan | nan |
| 3061 | White | Local-gov | <=50K | Prof-specialty | Peru | Not-in-family | Male | nan | nan |
| 4531 | White | Private | <=50K | Protective-serv | Peru | Other-relative | Male | nan | nan |
| 1771 | Black | Private | >50K | Handlers-cleaners | Hungary | Husband | Male | nan | nan |
| 3148 | White | Local-gov | <=50K | Protective-serv | United-States | Not-in-family | Male | 201112.609756 | 41.000000 |

### Synthetic

| | race | workclass | income | occupation | `native.country` | relationship | sex | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|
| 2329 | White | ? | <=50K | ? | Peru | Not-in-family | Female | 178962.961657 | 6 |
| 3061 | White | Local-gov | <=50K | Prof-specialty | Peru | Not-in-family | Male | 179199.925369 | 4 |
| 4531 | White | Private | <=50K | Protective-serv | Peru | Other-relative | Male | 179458.917653 | 1 |
| 1771 | Black | Private | >50K | Handlers-cleaners | Hungary | Husband | Male | 182057.167035 | 2 |
| 3148 | White | Local-gov | <=50K | Protective-serv | United-States | Not-in-family | Male | 186226.256904 | 33 |

SQL for Real:
SELECT race,workclass,income,occupation,`native.country`,relationship,sex,AVG(fnlwgt), COUNT(*) FROM C1 GROUP BY race,workclass,income,occupation,`native.country`,relationship,sex

Resulted in 4642 records

SQL for Synthetic:
SELECT race,workclass,income,occupation,`native.country`,relationship,sex,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY race,workclass,income,occupation,`native.country`,relationship,sex

Resulted in 6316 records

Normalized Euclidean distance for (fnlwgt): 43.95

Hellinger Distance: 0.484

============================================

Real

| | relationship | education | income | race | `marital.status` | `native.country` | sex | workclass | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 4274 | Not-in-family | Some-college | <=50K | White | Divorced | Puerto-Rico | Male | ? | 65.000000 | 1.000000 |
| 7829 | Unmarried | HS-grad | <=50K | White | Never-married | Puerto-Rico | Male | Local-gov | nan | nan |
| 2153 | Not-in-family | 7th-8th | <=50K | White | Never-married | Peru | Male | Private | nan | nan |
| 3193 | Not-in-family | HS-grad | <=50K | Asian-Pac-Islander | Never-married | Thailand | Female | Self-emp-not-inc | nan | nan |
| 8099 | Unmarried | Prof-school | <=50K | White | Separated | United-States | Male | Self-emp-not-inc | nan | nan |

Synthetic

| | relationship | education | income | race | `marital.status` | `native.country` | sex | workclass | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 4274 | Not-in-family | Some-college | <=50K | White | Divorced | Puerto-Rico | Male | ? | 39.840864 | 6 |
| 7829 | Unmarried | HS-grad | <=50K | White | Never-married | Puerto-Rico | Male | Local-gov | 40.014015 | 1 |
| 2153 | Not-in-family | 7th-8th | <=50K | White | Never-married | Peru | Male | Private | 39.532139 | 1 |
| 3193 | Not-in-family | HS-grad | <=50K | Asian-Pac-Islander | Never-married | Thailand | Female | Self-emp-not-inc | 40.023884 | 1 |
| 8099 | Unmarried | Prof-school | <=50K | White | Separated | United-States | Male | Self-emp-not-inc | 40.020488 | 1 |

SQL for Real:
SELECT relationship,education,income,race,`marital.status`,`native.country`,sex,workclass,AVG(`hours.per.week`), COUNT(*) FROM C1 GROUP BY relationship,education,income,race,`marital.status`,`native.country`,sex,workclass

Resulted in 6863 records

SQL for Synthetic:
SELECT relationship,education,income,race,`marital.status`,`native.country`,sex,workclass,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY relationship,education,income,race,`marital.status`,`native.country`,sex,workclass

Resulted in 9165 records

Normalized Euclidean distance for (`hours.per.week`): 50.11

Hellinger Distance: 0.502

============================================

Real

| | occupation | sex | income | `marital.status` | race | education | workclass | age | count |
|---|---|---|---|---|---|---|---|---|---|
| 1012 | Adm-clerical | Male | <=50K | Separated | White | Some-college | State-gov | 37 | 1 |
| 4014 | Other-service | Female | <=50K | Widowed | White | HS-grad | Private | 3496 | 59 |
| 5456 | Prof-specialty | Male | >50K | Separated | White | Masters | Local-gov | 51 | 1 |
| 1788 | Exec-managerial | Female | <=50K | Divorced | White | Doctorate | State-gov | 51 | 1 |
| 4077 | Other-service | Male | <=50K | Divorced | Black | 11th | Local-gov | 47 | 1 |

Synthetic

| | occupation | sex | income | `marital.status` | race | education | workclass | age | count |
|---|---|---|---|---|---|---|---|---|---|
| 1012 | Adm-clerical | Male | <=50K | Separated | White | Some-college | State-gov | 67.000000 | 1.000000 |
| 4014 | Other-service | Female | <=50K | Widowed | White | HS-grad | Private | 2790.000000 | 49.000000 |
| 5456 | Prof-specialty | Male | >50K | Separated | White | Masters | Local-gov | 47.000000 | 1.000000 |
| 1788 | Exec-managerial | Female | <=50K | Divorced | White | Doctorate | State-gov | 62.000000 | 1.000000 |
| 4077 | Other-service | Male | <=50K | Divorced | Black | 11th | Local-gov | nan | nan |

SQL for Real:
SELECT occupation,sex,income,`marital.status`,race,education,workclass,SUM(age), COUNT(*) FROM C1 GROUP BY occupation,sex,income,`marital.status`,race,education,workclass

Resulted in 7145 records

SQL for Synthetic:
SELECT occupation,sex,income,`marital.status`,race,education,workclass,SUM(age), COUNT(*) FROM C1_syn_06 GROUP BY occupation,sex,income,`marital.status`,race,education,workclass

Resulted in 6386 records

Normalized Euclidean distance for (age): 62.59

Hellinger Distance: 0.289

============================================

Real

| | `native.country` | capital | count |
|---|---|---|---|
| 32 | Portugal | 5178 | 67 |
| 3 | China | 99999 | 122 |
| 27 | Nicaragua | 3887 | 49 |
| 22 | Italy | 20051 | 105 |
| 17 | Hong | 15024 | 30 |

Synthetic

| | `native.country` | capital | count |
|---|---|---|---|
| 32 | Portugal | 122.067401 | 2.000000 |
| 3 | China | 262.346894 | 94.000000 |
| 27 | Nicaragua | 294.797976 | 184.000000 |
| 22 | Italy | 124.148738 | 9.000000 |
| 17 | Hong | nan | nan |

SQL for Real:
SELECT `native.country`,MAX(capital), COUNT(*) FROM C1 GROUP BY `native.country`

Resulted in 42 records

Resulted in 41 records

Normalized Euclidean distance for (capital): 6.4

Hellinger Distance: 0.396

==========================================

### Real

| | workclass | relationship | `native.country` | sex | income | `marital.status` | occupation | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|
| 4401 | Private | Unmarried | Hungary | Female | <=50K | Never-married | Transport-moving | nan | nan |
| 1970 | Private | Husband | Puerto-Rico | Male | <=50K | Married-spouse-absent | Prof-specialty | nan | nan |
| 6590 | State-gov | Own-child | Laos | Male | <=50K | Never-married | Adm-clerical | nan | nan |
| 1391 | Local-gov | Own-child | United-States | Male | <=50K | Never-married | Farming-fishing | 332249.000000 | 10.000000 |
| 259 | ? | Unmarried | Japan | Female | <=50K | Divorced | Machine-op-inspct | nan | nan |

### Synthetic

| | workclass | relationship | `native.country` | sex | income | `marital.status` | occupation | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|
| 4401 | Private | Unmarried | Hungary | Female | <=50K | Never-married | Transport-moving | 183116.218862 | 1 |
| 1970 | Private | Husband | Puerto-Rico | Male | <=50K | Married-spouse-absent | Prof-specialty | 170348.457060 | 1 |
| 6590 | State-gov | Own-child | Laos | Male | <=50K | Never-married | Adm-clerical | 179514.759314 | 1 |
| 1391 | Local-gov | Own-child | United-States | Male | <=50K | Never-married | Farming-fishing | 189447.483094 | 8 |
| 259 | ? | Unmarried | Japan | Female | <=50K | Divorced | Machine-op-inspct | 178105.482200 | 1 |

Resulted in 4878 records

Resulted in 6808 records

Normalized Euclidean distance for (fnlwgt): 45.73

Hellinger Distance: 0.482

==========================================

### Real

| | income | workclass | race | occupation | `native.country` | `marital.status` | capital | count |
|---|---|---|---|---|---|---|---|---|
| 1312 | <=50K | Private | Asian-Pac-Islander | Other-service | Ireland | Married-civ-spouse | nan | nan |
| 1341 | <=50K | Private | Asian-Pac-Islander | Prof-specialty | Columbia | Divorced | nan | nan |
| 2898 | <=50K | Private | White | Sales | Ecuador | Never-married | nan | nan |
| 3195 | <=50K | Self-emp-inc | White | Craft-repair | Columbia | Married-civ-spouse | nan | nan |
| 4136 | <=50K | State-gov | White | Transport-moving | Columbia | Married-civ-spouse | nan | nan |

### Synthetic

| | income | workclass | race | occupation | `native.country` | `marital.status` | capital | count |
|---|---|---|---|---|---|---|---|---|
| 1312 | <=50K | Private | Asian-Pac-Islander | Other-service | Ireland | Married-civ-spouse | -188.623797 | 1 |
| 1341 | <=50K | Private | Asian-Pac-Islander | Prof-specialty | Columbia | Divorced | 206.041992 | 1 |
| 2898 | <=50K | Private | White | Sales | Ecuador | Never-married | 278.022809 | 12 |
| 3195 | <=50K | Self-emp-inc | White | Craft-repair | Columbia | Married-civ-spouse | 19.109056 | 2 |
| 4136 | <=50K | State-gov | White | Transport-moving | Columbia | Married-civ-spouse | 81.476716 | 1 |

Resulted in 3734 records

Resulted in 4911 records

Normalized Euclidean distance for (capital): 39.97

Hellinger Distance: 0.476

==========================================

### Real

| | race | `native.country` | education | sex | capital | count |
|---|---|---|---|---|---|---|
| 1048 | White | Jamaica | 9th | Female | nan | nan |
| 1077 | White | Japan | HS-grad | Male | 9472.000000 | 5.000000 |
| 163 | Asian-Pac-Islander | Laos | Assoc-voc | Female | nan | nan |
| 1031 | White | Ireland | HS-grad | Female | 0.000000 | 3.000000 |
| 671 | Other | Puerto-Rico | Prof-school | Male | nan | nan |

| | race | `native.country` | education | sex | capital | count |
|---|---|---|---|---|---|---|
| | | | Synthetic | | | |
| 1048 | White | Jamaica | 9th | Female | -35.617099 | 1 |
| 1077 | White | Japan | HS-grad | Male | -3709.139058 | 275 |
| 163 | Asian-Pac-Islander | Laos | Assoc-voc | Female | -212.294102 | 1 |
| 1031 | White | Ireland | HS-grad | Female | 408.690922 | 37 |
| 671 | Other | Puerto-Rico | Prof-school | Male | 21.705955 | 1 |

SQL for Real:
SELECT race,`native.country`,education,sex,SUM(capital), COUNT(*) FROM C1 GROUP BY race,`native.country`,education,sex

Resulted in 1171 records

SQL for Synthetic:
SELECT race,`native.country`,education,sex,SUM(capital), COUNT(*) FROM C1_syn_06 GROUP BY race,`native.country`,education,sex

Resulted in 1368 records

Normalized Euclidean distance for (capital): 24.76

Hellinger Distance: 0.428

========================================

Real
| | education | fnlwgt | count |
|---|---|---|---|
| 2 | 12th | 197538.184170 | 657 |
| 12 | Masters | 181480.074896 | 2657 |
| 0 | 10th | 196532.558675 | 1389 |
| 4 | 5th-6th | 229858.791749 | 509 |
| 7 | Assoc-acdm | 193700.075578 | 1601 |

Synthetic
| | education | fnlwgt | count |
|---|---|---|---|
| 2 | 12th | 187614.998110 | 626.000000 |
| 12 | Masters | 177106.686122 | 2810.000000 |
| 0 | 10th | 177291.796309 | 1430.000000 |
| 4 | 5th-6th | 178682.695677 | 342.000000 |
| 7 | Assoc-acdm | 181913.434789 | 1755.000000 |

SQL for Real:
SELECT education,AVG(fnlwgt), COUNT(*) FROM C1 GROUP BY education

Resulted in 16 records

SQL for Synthetic:
SELECT education,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY education

Resulted in 16 records

Normalized Euclidean distance for (fnlwgt): 4.0

Hellinger Distance: 0.026

========================================

Real
| | income | `native.country` | workclass | education | race | occupation | `marital.status` | relationship | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 12493 | <=50K | United-States | Self-emp-inc | HS-grad | White | Sales | Married-civ-spouse | Other-relative | nan | nan |
| 12028 | <=50K | United-States | Private | Some-college | Black | Handlers-cleaners | Never-married | Other-relative | nan | nan |
| 4196 | <=50K | Laos | Private | Masters | White | Exec-managerial | Never-married | Not-in-family | nan | nan |
| 11579 | <=50K | United-States | Private | HS-grad | White | Craft-repair | Separated | Not-in-family | 38.375000 | 24.000000 |
| 14111 | >50K | Japan | Private | HS-grad | Black | Machine-op-inspct | Married-civ-spouse | Husband | nan | nan |

Synthetic
| | income | `native.country` | workclass | education | race | occupation | `marital.status` | relationship | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 12493 | <=50K | United-States | Self-emp-inc | HS-grad | White | Sales | Married-civ-spouse | Other-relative | 28.000000 | 1 |
| 12028 | <=50K | United-States | Private | Some-college | Black | Handlers-cleaners | Never-married | Other-relative | 25.333333 | 3 |
| 4196 | <=50K | Laos | Private | Masters | White | Exec-managerial | Never-married | Not-in-family | 28.000000 | 2 |
| 11579 | <=50K | United-States | Private | HS-grad | White | Craft-repair | Separated | Not-in-family | 39.076923 | 13 |
| 14111 | >50K | Japan | Private | HS-grad | Black | Machine-op-inspct | Married-civ-spouse | Husband | 40.000000 | 1 |

SQL for Real:
SELECT income,`native.country`,workclass,education,race,occupation,`marital.status`,relationship,AVG(age), COUNT(*) FROM C1 GROUP BY income,`native.country`,workclass,education,race,occupation,`marital.status`,relationship

Resulted in 11324 records

SQL for Synthetic:
SELECT income,`native.country`,workclass,education,race,occupation,`marital.status`,relationship,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY income,`native.country`,workclass,education,race,occupation,`marital.status`,relationship

Resulted in 15410 records

Normalized Euclidean distance for (age): 62.14

Hellinger Distance: 0.538

========================================

Real

| | relationship | income | race | occupation | age | count |
|---|---|---|---|---|---|---|
| 93 | Husband | >50K | Black | Handlers-cleaners | 27 | 13 |
| 512 | Wife | <=50K | Amer-Indian-Eskimo | Exec-managerial | 24 | 1 |
| 545 | Wife | <=50K | Other | Machine-op-inspct | 24 | 4 |
| 386 | Own-child | <=50K | White | Transport-moving | 17 | 194 |
| 521 | Wife | <=50K | Asian-Pac-Islander | Exec-managerial | 27 | 5 |

Synthetic

| | relationship | income | race | occupation | age | count |
|---|---|---|---|---|---|---|
| 93 | Husband | >50K | Black | Handlers-cleaners | 28.000000 | 5.000000 |
| 512 | Wife | <=50K | Amer-Indian-Eskimo | Exec-managerial | nan | nan |
| 545 | Wife | <=50K | Other | Machine-op-inspct | 32.000000 | 2.000000 |
| 386 | Own-child | <=50K | White | Transport-moving | 17.000000 | 220.000000 |
| 521 | Wife | <=50K | Asian-Pac-Islander | Exec-managerial | 25.000000 | 8.000000 |

SQL for Real:
SELECT relationship,income,race,occupation,MIN(age), COUNT(*) FROM C1 GROUP BY relationship,income,race,occupation

Resulted in 608 records

SQL for Synthetic:
SELECT relationship,income,race,occupation,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY relationship,income,race,occupation

Resulted in 511 records

Normalized Euclidean distance for (age): 22.02

Hellinger Distance: 0.176

==========================================

Real

| | education | `marital.status` | fnlwgt | count |
|---|---|---|---|---|
| 57 | Bachelors | Divorced | 183708.214709 | 843 |
| 92 | Prof-school | Married-civ-spouse | 189783.506711 | 596 |
| 89 | Preschool | Widowed | 244417.800000 | 5 |
| 59 | Bachelors | Married-civ-spouse | 187092.705754 | 4136 |
| 70 | Doctorate | Widowed | 178553.384615 | 13 |

Synthetic

| | education | `marital.status` | fnlwgt | count |
|---|---|---|---|---|
| 57 | Bachelors | Divorced | 178357.999267 | 882.000000 |
| 92 | Prof-school | Married-civ-spouse | 172431.853565 | 614.000000 |
| 89 | Preschool | Widowed | 166876.754434 | 4.000000 |
| 59 | Bachelors | Married-civ-spouse | 178859.244094 | 4102.000000 |
| 70 | Doctorate | Widowed | 171632.909668 | 7.000000 |

SQL for Real:
SELECT education,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1 GROUP BY education,`marital.status`
Resulted in 104 records

SQL for Synthetic:
SELECT education,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY education,`marital.status`
Resulted in 101 records

Normalized Euclidean distance for (fnlwgt): 10.0

Hellinger Distance: 0.035

==========================================

Real

| | occupation | `marital.status` | race | relationship | education | age | count |
|---|---|---|---|---|---|---|---|
| 2704 | Other-service | Never-married | Asian-Pac-Islander | Own-child | HS-grad | 18 | 8 |
| 445 | Adm-clerical | Divorced | White | Not-in-family | Masters | 31 | 7 |
| 3956 | Sales | Separated | White | Unmarried | 12th | 34 | 1 |
| 1204 | Exec-managerial | Divorced | Asian-Pac-Islander | Not-in-family | HS-grad | 37 | 1 |
| 3402 | Prof-specialty | Separated | White | Not-in-family | Masters | 32 | 11 |

Synthetic

| | occupation | `marital.status` | race | relationship | education | age | count |
|---|---|---|---|---|---|---|---|
| 2704 | Other-service | Never-married | Asian-Pac-Islander | Own-child | HS-grad | 34.000000 | 2.000000 |
| 445 | Adm-clerical | Divorced | White | Not-in-family | Masters | 25.000000 | 15.000000 |
| 3956 | Sales | Separated | White | Unmarried | 12th | nan | nan |
| 1204 | Exec-managerial | Divorced | Asian-Pac-Islander | Not-in-family | HS-grad | nan | nan |
| 3402 | Prof-specialty | Separated | White | Not-in-family | Masters | 33.000000 | 10.000000 |

SQL for Real:
SELECT occupation,`marital.status`,race,relationship,education,MIN(age), COUNT(*) FROM C1 GROUP BY occupation,`marital.status`,race,relationship,education
Resulted in 4464 records

SQL for Synthetic:
SELECT occupation,`marital.status`,race,relationship,education,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY occupation,`marital.status`,race,relationship,education
Resulted in 4151 records

Normalized Euclidean distance for (age): 52.97

Hellinger Distance: 0.185

```
================================================
```

|  | Real |  |  |
| --- | --- | --- | --- |
| | income | workclass | fnlwgt | count |
| 7 | >50K | ? | 584259 | 267 |
| 10 | >50K | Private | 1226583 | 7387 |
| 13 | >50K | State-gov | 764638 | 530 |
| 2 | <=50K | Local-gov | 1125613 | 2209 |
| 12 | >50K | Self-emp-not-inc | 607848 | 1077 |

|  | Synthetic |  |  |
| --- | --- | --- | --- |
| | income | workclass | fnlwgt | count |
| 7 | >50K | ? | 1330514.072832 | 95.000000 |
| 10 | >50K | Private | 1362953.001310 | 2930.000000 |
| 13 | >50K | State-gov | 1340619.957202 | 230.000000 |
| 2 | <=50K | Local-gov | 1351242.858809 | 2683.000000 |
| 12 | >50K | Self-emp-not-inc | 196233.581006 | 433.000000 |

SQL for Real:
SELECT income,workclass,MAX(fnlwgt), COUNT(*) FROM C1 GROUP BY income,workclass

Resulted in 14 records

SQL for Synthetic:
SELECT income,workclass,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY income,workclass

Resulted in 14 records

Normalized Euclidean distance for (fnlwgt): 3.74

Hellinger Distance: 0.138

```
================================================
```

|  | Real |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | education | occupation | relationship | sex | `hours.per.week` | count |
| 1078 | HS-grad | Adm-clerical | Not-in-family | Male | 6.000000 | 132.000000 |
| 1246 | Masters | Craft-repair | Not-in-family | Male | 40.000000 | 4.000000 |
| 1450 | Some-college | Farming-fishing | Unmarried | Female | 25.000000 | 7.000000 |
| 341 | 1st-4th | Handlers-cleaners | Not-in-family | Male | 37.000000 | 4.000000 |
| 1366 | Prof-school | Machine-op-inspct | Own-child | Female | nan | nan |

|  | Synthetic |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | education | occupation | relationship | sex | `hours.per.week` | count |
| 1078 | HS-grad | Adm-clerical | Not-in-family | Male | 38.933067 | 139 |
| 1246 | Masters | Craft-repair | Not-in-family | Male | 39.964770 | 11 |
| 1450 | Some-college | Farming-fishing | Unmarried | Female | 39.950302 | 6 |
| 341 | 1st-4th | Handlers-cleaners | Not-in-family | Male | 39.998806 | 2 |
| 1366 | Prof-school | Machine-op-inspct | Own-child | Female | 39.979876 | 1 |

SQL for Real:
SELECT education,occupation,relationship,sex,MIN(`hours.per.week`), COUNT(*) FROM C1 GROUP BY education,occupation,relationship,sex

Resulted in 1503 records

SQL for Synthetic:
SELECT education,occupation,relationship,sex,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY education,occupation,relationship,sex

Resulted in 1539 records

Normalized Euclidean distance for (`hours.per.week`): 35.87

Hellinger Distance: 0.102

```
================================================
```

|  | Real |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | occupation | `marital.status` | relationship | education | income | workclass | race | `native.country` | sex | `hours.per.week` | count |
| 9510 | Other-service | Divorced | Unmarried | HS-grad | <=50K | Self-emp-not-inc | Black | United-States | Female | nan | nan |
| 15293 | Sales | Never-married | Own-child | HS-grad | <=50K | Private | Black | Nicaragua | Female | nan | nan |
| 12767 | Prof-specialty | Never-married | Not-in-family | Doctorate | >50K | Self-emp-inc | White | Puerto-Rico | Male | nan | nan |
| 11253 | Priv-house-serv | Never-married | Unmarried | 5th-6th | <=50K | Private | White | United-States | Female | nan | nan |
| 15100 | Sales | Never-married | Not-in-family | Some-college | <=50K | Private | White | Ireland | Female | nan | nan |

|  | Synthetic |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | occupation | `marital.status` | relationship | education | income | workclass | race | `native.country` | sex | `hours.per.week` | count |
| 9510 | Other-service | Divorced | Unmarried | HS-grad | <=50K | Self-emp-not-inc | Black | United-States | Female | 39.966967 | 1 |
| 15293 | Sales | Never-married | Own-child | HS-grad | <=50K | Private | Black | Nicaragua | Female | 39.450803 | 3 |
| 12767 | Prof-specialty | Never-married | Not-in-family | Doctorate | >50K | Self-emp-inc | White | Puerto-Rico | Male | 40.212530 | 1 |
| 11253 | Priv-house-serv | Never-married | Unmarried | 5th-6th | <=50K | Private | White | United-States | Female | 40.014485 | 1 |
| 15100 | Sales | Never-married | Not-in-family | Some-college | <=50K | Private | White | Ireland | Female | 39.984600 | 1 |

SQL for Real:
SELECT occupation,`marital.status`,relationship,education,income,workclass,race,`native.country`,sex,MIN(`hours.per.week`), COUNT(*) FROM C1 GROUP BY occupation,`marital.status`,relationship,education,income,workclass,race,`native.country`,sex

Resulted in 12664 records

SQL for Synthetic:
SELECT occupation,`marital.status`,relationship,education,income,workclass,race,`native.country`,sex,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY occupation,`marital.status`,relationship,education,income,workclass,race,`native.country`,sex

Resulted in 17127 records

Normalized Euclidean distance for (`hours.per.week`): 64.61

Hellinger Distance: 0.549

============================================

### Real

|      | sex    | occupation        | relationship   | income | workclass | race      | capital | count |
|------|--------|-------------------|----------------|--------|-----------|-----------|---------|-------|
| 450  | Female | Other-service     | Not-in-family  | >50K   | Private   | White     | 105789  | 8     |
| 912  | Male   | ?                 | Husband        | <=50K  | ?         | Black     | 2410    | 33    |
| 372  | Female | Handlers-cleaners | Unmarried      | <=50K  | Local-gov | Black     | 0       | 1     |
| 1106 | Male   | Craft-repair      | Not-in-family  | <=50K  | Private   | White     | 40970   | 722   |
| 1655 | Male   | Priv-house-serv   | Other-relative | <=50K  | Private   | White     | 0       | 2     |

### Synthetic

|      | sex    | occupation        | relationship   | income | workclass | race      | capital      | count       |
|------|--------|-------------------|----------------|--------|-----------|-----------|--------------|-------------|
| 450  | Female | Other-service     | Not-in-family  | >50K   | Private   | White     | -364.324528  | 6.000000    |
| 912  | Male   | ?                 | Husband        | <=50K  | ?         | Black     | 312.797875   | 21.000000   |
| 372  | Female | Handlers-cleaners | Unmarried      | <=50K  | Local-gov | Black     | -51.790690   | 2.000000    |
| 1106 | Male   | Craft-repair      | Not-in-family  | <=50K  | Private   | White     | 2118.500694  | 729.000000  |
| 1655 | Male   | Priv-house-serv   | Other-relative | <=50K  | Private   | White     | nan          | nan         |

SQL for Real:
SELECT sex,occupation,relationship,income,workclass,race,SUM(capital), COUNT(*) FROM C1 GROUP BY sex,occupation,relationship,income,workclass,race

Resulted in 2132 records

SQL for Synthetic:
SELECT sex,occupation,relationship,income,workclass,race,SUM(capital), COUNT(*) FROM C1_syn_06 GROUP BY sex,occupation,relationship,income,workclass,race

Resulted in 1881 records

Normalized Euclidean distance for (capital): 38.0

Hellinger Distance: 0.206

============================================

### Real

|   | workclass       | `hours.per.week` | count |
|---|-----------------|------------------|-------|
| 4 | Self-emp-inc    | 82327            | 1695  |
| 5 | Self-emp-not-inc| 171454           | 3862  |
| 6 | State-gov       | 77439            | 1981  |
| 1 | Federal-gov     | 59447            | 1432  |
| 3 | Private         | 1365501          | 33906 |

### Synthetic

|   | workclass       | `hours.per.week` | count       |
|---|-----------------|------------------|-------------|
| 4 | Self-emp-inc    | 66101.036292     | 1645.000000 |
| 5 | Self-emp-not-inc| 153032.465823    | 3818.000000 |
| 6 | State-gov       | 79189.970518     | 1982.000000 |
| 1 | Federal-gov     | 57003.626201     | 1426.000000 |
| 3 | Private         | 1360644.696869   | 34022.000000|

SQL for Real:
SELECT workclass,SUM(`hours.per.week`), COUNT(*) FROM C1 GROUP BY workclass

Resulted in 7 records

SQL for Synthetic:
SELECT workclass,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY workclass

Resulted in 7 records

Normalized Euclidean distance for (`hours.per.week`): 2.65

Hellinger Distance: 0.003

============================================

### Real

|       | relationship  | `marital.status` | race               | occupation  | education    | sex    | workclass | `native.country` | capital | count |
|-------|---------------|------------------|--------------------|-------------|--------------|--------|-----------|------------------|---------|-------|
| 10070 | Own-child     | Never-married    | Asian-Pac-Islander | ?           | Some-college | Female | ?         | ?                | nan     | nan   |
| 10510 | Own-child     | Never-married    | Other              | Adm-clerical| HS-grad      | Female | Private   | United-States    | nan     | nan   |
| 7514  | Not-in-family | Never-married    | White              | Tech-support| HS-grad      | Male   | Private   | Puerto-Rico      | nan     | nan   |
| 14071 | Unmarried     | Never-married    | White              | Sales       | HS-grad      | Female | Private   | China            | nan     | nan   |
| 14954 | Unmarried     | Widowed          | White              | Tech-support| HS-grad      | Female | Private   | Puerto-Rico      | nan     | nan   |

### Synthetic

|       | relationship  | `marital.status` | race               | occupation  | education    | sex    | workclass | `native.country` | capital      | count |
|-------|---------------|------------------|--------------------|-------------|--------------|--------|-----------|------------------|--------------|-------|
| 10070 | Own-child     | Never-married    | Asian-Pac-Islander | ?           | Some-college | Female | ?         | ?                | -5.257136    | 1     |
| 10510 | Own-child     | Never-married    | Other              | Adm-clerical| HS-grad      | Female | Private   | United-States    | 61.306068    | 1     |
| 7514  | Not-in-family | Never-married    | White              | Tech-support| HS-grad      | Male   | Private   | Puerto-Rico      | -68.835725   | 1     |
| 14071 | Unmarried     | Never-married    | White              | Sales       | HS-grad      | Female | Private   | China            | -124.853495  | 1     |
| 14954 | Unmarried     | Widowed          | White              | Tech-support| HS-grad      | Female | Private   | Puerto-Rico      | -59.072154   | 2     |

SQL for Real:
SELECT relationship,`marital.status`,race,occupation,education,sex,workclass,`native.country`,SUM(capital), COUNT(*) FROM C1 GROUP BY relationship,`marital.status`,race,occupation,education,sex,workclass,`native.country`

Resulted in 11374 records

SQL for Synthetic:

SELECT relationship,`marital.status`,race,occupation,education,sex,workclass,`native.country`,SUM(capital), COUNT(*) FROM C1_syn_06 GROUP BY relationship,`marital.status`,race,occupation,education,sex,workclass,`native.country`

Resulted in 15994 records

Normalized Euclidean distance for (capital): 63.12

Hellinger Distance: 0.53

=============================================

### Real

| | `native.country` | age | count |
|---|---|---|---|
| 36 | Taiwan | 34.184615 | 65 |
| 29 | Peru | 36.434783 | 46 |
| 37 | Thailand | 37.666667 | 30 |
| 2 | Canada | 44.049451 | 182 |
| 39 | United-States | 38.698690 | 43832 |

### Synthetic

| | `native.country` | age | count |
|---|---|---|---|
| 36 | Taiwan | 47.344595 | 148.000000 |
| 29 | Peru | 39.286661 | 1207.000000 |
| 37 | Thailand | 22.848341 | 211.000000 |
| 2 | Canada | 26.833333 | 84.000000 |
| 39 | United-States | 36.402528 | 27931.000000 |

SQL for Real:
SELECT `native.country`,AVG(age), COUNT(*) FROM C1 GROUP BY `native.country`

Resulted in 42 records

SQL for Synthetic:
SELECT `native.country`,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`

Resulted in 41 records

Normalized Euclidean distance for (age): 6.4

Hellinger Distance: 0.396

=============================================

### Real

| | sex | occupation | race | education | income | `marital.status` | `native.country` | workclass | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 8562 | Male | Exec-managerial | White | HS-grad | <=50K | Separated | Puerto-Rico | Private | nan | nan |
| 3790 | Female | Other-service | White | Some-college | >50K | Separated | Hungary | Private | nan | nan |
| 11574 | Male | Prof-specialty | White | Masters | <=50K | Married-civ-spouse | ? | State-gov | 28.000000 | 1.000000 |
| 7877 | Male | Craft-repair | White | Some-college | >50K | Married-civ-spouse | Japan | Private | nan | nan |
| 4605 | Female | Prof-specialty | White | Masters | <=50K | Never-married | Puerto-Rico | Local-gov | nan | nan |

### Synthetic

| | sex | occupation | race | education | income | `marital.status` | `native.country` | workclass | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 8562 | Male | Exec-managerial | White | HS-grad | <=50K | Separated | Puerto-Rico | Private | 38 | 1 |
| 3790 | Female | Other-service | White | Some-college | >50K | Separated | Hungary | Private | 43 | 1 |
| 11574 | Male | Prof-specialty | White | Masters | <=50K | Married-civ-spouse | ? | State-gov | 43 | 2 |
| 7877 | Male | Craft-repair | White | Some-college | >50K | Married-civ-spouse | Japan | Private | 39 | 1 |
| 4605 | Female | Prof-specialty | White | Masters | <=50K | Never-married | Puerto-Rico | Local-gov | 33 | 6 |

SQL for Real:
SELECT sex,occupation,race,education,income,`marital.status`,`native.country`,workclass,MIN(age), COUNT(*) FROM C1 GROUP BY sex,occupation,race,education,income,`marital.status`,`native.country`,workclass

Resulted in 10025 records

SQL for Synthetic:
SELECT sex,occupation,race,education,income,`marital.status`,`native.country`,workclass,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY sex,occupation,race,education,income,`marital.status`,`native.country`,workclass

Resulted in 13875 records

Normalized Euclidean distance for (age): 59.76

Hellinger Distance: 0.527

=============================================

### Real

| | workclass | race | relationship | `native.country` | sex | occupation | capital | count |
|---|---|---|---|---|---|---|---|---|
| 5462 | State-gov | White | Own-child | Columbia | Female | Adm-clerical | nan | nan |
| 356 | Federal-gov | Black | Not-in-family | Jamaica | Female | Prof-specialty | nan | nan |
| 4835 | Self-emp-not-inc | White | Not-in-family | ? | Male | Farming-fishing | 0.000000 | 2.000000 |
| 673 | Federal-gov | White | Unmarried | United-States | Male | Exec-managerial | 2204.500000 | 6.000000 |
| 4535 | Self-emp-inc | White | Unmarried | ? | Female | Prof-specialty | nan | nan |

### Synthetic

| | workclass | race | relationship | `native.country` | sex | occupation | capital | count |
|---|---|---|---|---|---|---|---|---|
| 5462 | State-gov | White | Own-child | Columbia | Female | Adm-clerical | -24.634660 | 1 |
| 356 | Federal-gov | Black | Not-in-family | Jamaica | Female | Prof-specialty | 175.648484 | 1 |
| 4835 | Self-emp-not-inc | White | Not-in-family | ? | Male | Farming-fishing | 27.793069 | 2 |
| 673 | Federal-gov | White | Unmarried | United-States | Male | Exec-managerial | 173.957275 | 1 |
| 4535 | Self-emp-inc | White | Unmarried | ? | Female | Prof-specialty | 54.366888 | 1 |

SQL for Real:
SELECT workclass,race,relationship,`native.country`,sex,occupation,AVG(capital), COUNT(*) FROM C1 GROUP BY workclass,race,relationship,`native.country`,sex,occupation

Resulted in 3919 records

SQL for Synthetic:
SELECT workclass,race,relationship,`native.country`,sex,occupation,AVG(capital), COUNT(*) FROM C1_syn_06 GROUP BY workclass,race,relationship,`native.country`,sex,occupation

Resulted in 5595 records

Normalized Euclidean distance for (capital): 41.82

Hellinger Distance: 0.463

===============================================

**Real**

| | relationship | `native.country` | education | workclass | capital | count |
|---|---|---|---|---|---|---|
| **251** | Husband | Hungary | Assoc-voc | Local-gov | nan | nan |
| **1786** | Other-relative | United-States | 5th-6th | Self-emp-not-inc | nan | nan |
| **2645** | Unmarried | Peru | Masters | Local-gov | nan | nan |
| **1760** | Other-relative | Scotland | Some-college | Local-gov | nan | nan |
| **1362** | Not-in-family | Puerto-Rico | Bachelors | Private | 0.000000 | 2.000000 |

**Synthetic**

| | relationship | `native.country` | education | workclass | capital | count |
|---|---|---|---|---|---|---|
| **251** | Husband | Hungary | Assoc-voc | Local-gov | 82.918198 | 1 |
| **1786** | Other-relative | United-States | 5th-6th | Self-emp-not-inc | -56.972206 | 1 |
| **2645** | Unmarried | Peru | Masters | Local-gov | -16.168854 | 1 |
| **1760** | Other-relative | Scotland | Some-college | Local-gov | -6.715164 | 2 |
| **1362** | Not-in-family | Puerto-Rico | Bachelors | Private | 4.662578 | 219 |

SQL for Real:
SELECT relationship,`native.country`,education,workclass,AVG(capital), COUNT(*) FROM C1 GROUP BY relationship,`native.country`,education,workclass

Resulted in 2547 records

SQL for Synthetic:
SELECT relationship,`native.country`,education,workclass,AVG(capital), COUNT(*) FROM C1_syn_06 GROUP BY relationship,`native.country`,education,workclass

Resulted in 3287 records

Normalized Euclidean distance for (capital): 35.03

Hellinger Distance: 0.437

===============================================

**Real**

| | occupation | fnlwgt | count |
|---|---|---|---|
| **3** | Craft-repair | 19302 | 6112 |
| **4** | Exec-managerial | 13769 | 6086 |
| **9** | Priv-house-serv | 22406 | 242 |
| **2** | Armed-Forces | 76313 | 15 |
| **6** | Handlers-cleaners | 19214 | 2072 |

**Synthetic**

| | occupation | fnlwgt | count |
|---|---|---|---|
| **3** | Craft-repair | 132974.972969 | 5926.000000 |
| **4** | Exec-managerial | 121336.094168 | 6095.000000 |
| **9** | Priv-house-serv | 138109.939556 | 223.000000 |
| **2** | Armed-Forces | 162713.581466 | 121.000000 |
| **6** | Handlers-cleaners | 139621.315300 | 2129.000000 |

SQL for Real:
SELECT occupation,MIN(fnlwgt), COUNT(*) FROM C1 GROUP BY occupation

Resulted in 15 records

SQL for Synthetic:
SELECT occupation,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY occupation

Resulted in 15 records

Normalized Euclidean distance for (fnlwgt): 3.87

Hellinger Distance: 0.027

===============================================

**Real**

| | income | occupation | relationship | workclass | education | race | `native.country` | sex | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|---|---|---|
| **11933** | <=50K | Transport-moving | Not-in-family | Private | Bachelors | White | Laos | Male | nan | nan |
| **6465** | <=50K | Machine-op-inspct | Unmarried | Private | Some-college | Other | United-States | Female | nan | nan |
| **4410** | <=50K | Exec-managerial | Unmarried | Private | Bachelors | White | United-States | Female | 40.567568 | 37.000000 |
| **6097** | <=50K | Machine-op-inspct | Not-in-family | Private | HS-grad | White | ? | Male | nan | nan |
| **9483** | <=50K | Prof-specialty | Wife | Private | HS-grad | Black | Laos | Female | nan | nan |

| | income | occupation | relationship | workclass | Synthetic education | race | `native.country` | sex | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 11933 | <=50K | Transport-moving | Not-in-family | Private | Bachelors | White | Laos | Male | 40.224536 | 1 |
| 6465 | <=50K | Machine-op-inspct | Unmarried | Private | Some-college | Other | United-States | Female | 40.059249 | 1 |
| 4410 | <=50K | Exec-managerial | Unmarried | Private | Bachelors | White | United-States | Female | 39.983609 | 20 |
| 6097 | <=50K | Machine-op-inspct | Not-in-family | Private | HS-grad | White | ? | Male | 39.992386 | 4 |
| 9483 | <=50K | Prof-specialty | Wife | Private | HS-grad | Black | Laos | Female | 39.994116 | 1 |

SQL for Real:
SELECT income,occupation,relationship,workclass,education,race,`native.country`,sex,AVG(`hours.per.week`), COUNT(*) FROM C1 GROUP BY income,occupation,relationship,workclass,education,race,`native.country`,sex

Resulted in 9992 records

SQL for Synthetic:
SELECT income,occupation,relationship,workclass,education,race,`native.country`,sex,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY income,occupation,relationship,workclass,education,race,`native.country`,sex

Resulted in 13906 records

Normalized Euclidean distance for (`hours.per.week`): 60.46

Hellinger Distance: 0.528

===========================================

| | income | workclass | race | Real relationship | `marital.status` | fnlwgt | count |
|---|---|---|---|---|---|---|---|
| 497 | <=50K | State-gov | White | Husband | Married-civ-spouse | 21412 | 326 |
| 336 | <=50K | Private | White | Own-child | Widowed | 57827 | 13 |
| 667 | >50K | Private | White | Unmarried | Widowed | 34269 | 25 |
| 379 | <=50K | Self-emp-inc | White | Unmarried | Widowed | 46577 | 10 |
| 169 | <=50K | Local-gov | Asian-Pac-Islander | Unmarried | Divorced | 73199 | 4 |

| | income | workclass | race | Synthetic relationship | `marital.status` | fnlwgt | count |
|---|---|---|---|---|---|---|---|
| 497 | <=50K | State-gov | White | Husband | Married-civ-spouse | 123630.347178 | 509.000000 |
| 336 | <=50K | Private | White | Own-child | Widowed | 164093.336246 | 18.000000 |
| 667 | >50K | Private | White | Unmarried | Widowed | 152336.175644 | 10.000000 |
| 379 | <=50K | Self-emp-inc | White | Unmarried | Widowed | 143919.539628 | 11.000000 |
| 169 | <=50K | Local-gov | Asian-Pac-Islander | Unmarried | Divorced | 180914.746175 | 2.000000 |

SQL for Real:
SELECT income,workclass,race,relationship,`marital.status`,MIN(fnlwgt), COUNT(*) FROM C1 GROUP BY income,workclass,race,relationship,`marital.status`

Resulted in 767 records

SQL for Synthetic:
SELECT income,workclass,race,relationship,`marital.status`,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY income,workclass,race,relationship,`marital.status`

Resulted in 685 records

Normalized Euclidean distance for (fnlwgt): 23.41

Hellinger Distance: 0.173

===========================================

| | income | race | Real `marital.status` | `native.country` | `hours.per.week` | count |
|---|---|---|---|---|---|---|
| 390 | <=50K | White | Married-spouse-absent | Scotland | nan | nan |
| 134 | <=50K | Black | Married-civ-spouse | ? | 1108.000000 | 27.000000 |
| 601 | >50K | White | Separated | United-States | 3719.000000 | 80.000000 |
| 50 | <=50K | Asian-Pac-Islander | Divorced | Yugoslavia | nan | nan |
| 168 | <=50K | Black | Married-spouse-absent | Jamaica | 224.000000 | 6.000000 |

| | income | race | Synthetic `marital.status` | `native.country` | `hours.per.week` | count |
|---|---|---|---|---|---|---|
| 390 | <=50K | White | Married-spouse-absent | Scotland | 320.199928 | 8 |
| 134 | <=50K | Black | Married-civ-spouse | ? | 1280.309417 | 32 |
| 601 | >50K | White | Separated | United-States | 600.793411 | 15 |
| 50 | <=50K | Asian-Pac-Islander | Divorced | Yugoslavia | 39.961838 | 1 |
| 168 | <=50K | Black | Married-spouse-absent | Jamaica | 40.011006 | 1 |

SQL for Real:
SELECT income,race,`marital.status`,`native.country`,SUM(`hours.per.week`), COUNT(*) FROM C1 GROUP BY income,race,`marital.status`,`native.country`

Resulted in 568 records

SQL for Synthetic:
SELECT income,race,`marital.status`,`native.country`,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY income,race,`marital.status`,`native.country`

Resulted in 606 records

Normalized Euclidean distance for (`hours.per.week`): 18.03

Hellinger Distance: 0.44

===========================================

**Real**

| | `native.country` | occupation | fnlwgt | count |
|---|---|---|---|---|
| 347 | Philippines | Exec-managerial | 355856 | 18 |
| 346 | Philippines | Craft-repair | 375313 | 22 |
| 224 | India | Handlers-cleaners | 357954 | 3 |
| 373 | Portugal | Adm-clerical | 216934 | 6 |
| 254 | Italy | Craft-repair | 420054 | 19 |

**Synthetic**

| | `native.country` | occupation | fnlwgt | count |
|---|---|---|---|---|
| 347 | Philippines | Exec-managerial | 188383.229560 | 9.000000 |
| 346 | Philippines | Craft-repair | 182512.247765 | 1.000000 |
| 224 | India | Handlers-cleaners | 184771.970411 | 2.000000 |
| 373 | Portugal | Adm-clerical | nan | nan |
| 254 | Italy | Craft-repair | 182696.036515 | 3.000000 |

SQL for Real:
SELECT `native.country`,occupation,MAX(fnlwgt), COUNT(*) FROM C1 GROUP BY `native.country`,occupation

Resulted in 481 records

SQL for Synthetic:
SELECT `native.country`,occupation,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`,occupation

Resulted in 453 records

Normalized Euclidean distance for (fnlwgt): 19.34

Hellinger Distance: 0.396

==========================================

**Real**

| | sex | occupation | race | fnlwgt | count |
|---|---|---|---|---|---|
| 130 | Male | Tech-support | Asian-Pac-Islander | 8280669.000000 | 44.000000 |
| 115 | Male | Prof-specialty | Asian-Pac-Islander | 35560521.000000 | 211.000000 |
| 18 | Female | Exec-managerial | Asian-Pac-Islander | 9308021.000000 | 58.000000 |
| 57 | Female | Sales | Asian-Pac-Islander | 7973557.000000 | 60.000000 |
| 45 | Female | Priv-house-serv | White | 32820486.000000 | 167.000000 |

**Synthetic**

| | sex | occupation | race | fnlwgt | count |
|---|---|---|---|---|---|
| 130 | Male | Tech-support | Asian-Pac-Islander | 3809703.739566 | 21 |
| 115 | Male | Prof-specialty | Asian-Pac-Islander | 23738139.540976 | 135 |
| 18 | Female | Exec-managerial | Asian-Pac-Islander | 6196791.890354 | 35 |
| 57 | Female | Sales | Asian-Pac-Islander | 3932138.468065 | 22 |
| 45 | Female | Priv-house-serv | White | 28163752.216671 | 148 |

SQL for Real:
SELECT sex,occupation,race,SUM(fnlwgt), COUNT(*) FROM C1 GROUP BY sex,occupation,race

Resulted in 137 records

SQL for Synthetic:
SELECT sex,occupation,race,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY sex,occupation,race

Resulted in 139 records

Normalized Euclidean distance for (fnlwgt): 11.58

Hellinger Distance: 0.059

==========================================

**Real**

| | occupation | race | income | capital | count |
|---|---|---|---|---|---|
| 115 | Tech-support | Amer-Indian-Eskimo | <=50K | -282.857143 | 7 |
| 62 | Machine-op-inspct | Asian-Pac-Islander | >50K | 899.250000 | 12 |
| 134 | Transport-moving | White | >50K | 1461.718605 | 430 |
| 33 | Craft-repair | White | >50K | 2552.527316 | 1263 |
| 31 | Craft-repair | Other | >50K | 920.111111 | 9 |

**Synthetic**

| | occupation | race | income | capital | count |
|---|---|---|---|---|---|
| 115 | Tech-support | Amer-Indian-Eskimo | <=50K | -16.419541 | 13.000000 |
| 62 | Machine-op-inspct | Asian-Pac-Islander | >50K | -74.948094 | 5.000000 |
| 134 | Transport-moving | White | >50K | -57.986179 | 137.000000 |
| 33 | Craft-repair | White | >50K | -56.370328 | 463.000000 |
| 31 | Craft-repair | Other | >50K | 120.408838 | 3.000000 |

SQL for Real:
SELECT occupation,race,income,AVG(capital), COUNT(*) FROM C1 GROUP BY occupation,race,income

Resulted in 135 records

SQL for Synthetic:
SELECT occupation,race,income,AVG(capital), COUNT(*) FROM C1_syn_06 GROUP BY occupation,race,income

Resulted in 126 records

Normalized Euclidean distance for (capital): 11.05

Hellinger Distance: 0.156

==================================================

#### Real

| | `native.country` | fnlwgt | count |
|---|---|---|---|
| **21** | Ireland | 390746 | 37 |
| **12** | Greece | 288229 | 49 |
| **36** | Taiwan | 506329 | 65 |
| **30** | Philippines | 554206 | 295 |
| **8** | El-Salvador | 795830 | 155 |

#### Synthetic

| | `native.country` | fnlwgt | count |
|---|---|---|---|
| **21** | Ireland | 196496.155005 | 384.000000 |
| **12** | Greece | 208268.210536 | 4.000000 |
| **36** | Taiwan | 200456.608706 | 148.000000 |
| **30** | Philippines | 193607.294609 | 80.000000 |
| **8** | El-Salvador | 1353982.527147 | 30.000000 |

SQL for Real:
SELECT `native.country`,MAX(fnlwgt), COUNT(*) FROM C1 GROUP BY `native.country`

Resulted in 42 records

SQL for Synthetic:
SELECT `native.country`,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`

Resulted in 41 records

Normalized Euclidean distance for (fnlwgt): 6.4

Hellinger Distance: 0.396

==================================================

#### Real

| | sex | income | occupation | `marital.status` | race | workclass | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|---|
| **2178** | Male | >50K | Protective-serv | Married-civ-spouse | White | Private | 78 | 28 |
| **604** | Female | <=50K | Sales | Divorced | Other | Private | 40 | 3 |
| **937** | Female | >50K | Prof-specialty | Separated | Asian-Pac-Islander | Private | 40 | 1 |
| **1955** | Male | >50K | Exec-managerial | Married-civ-spouse | Asian-Pac-Islander | Federal-gov | 60 | 5 |
| **1178** | Male | <=50K | Craft-repair | Never-married | Other | Private | 58 | 11 |

#### Synthetic

| | sex | income | occupation | `marital.status` | race | workclass | `hours.per.week` | count |
|---|---|---|---|---|---|---|---|---|
| **2178** | Male | >50K | Protective-serv | Married-civ-spouse | White | Private | 40.058344 | 22.000000 |
| **604** | Female | <=50K | Sales | Divorced | Other | Private | nan | nan |
| **937** | Female | >50K | Prof-specialty | Separated | Asian-Pac-Islander | Private | nan | nan |
| **1955** | Male | >50K | Exec-managerial | Married-civ-spouse | Asian-Pac-Islander | Federal-gov | 40.023201 | 2.000000 |
| **1178** | Male | <=50K | Craft-repair | Never-married | Other | Private | 40.064599 | 12.000000 |

SQL for Real:
SELECT sex,income,occupation,`marital.status`,race,workclass,MAX(`hours.per.week`), COUNT(*) FROM C1 GROUP BY sex,income,occupation,`marital.status`,race,workclass

Resulted in 2280 records

SQL for Synthetic:
SELECT sex,income,occupation,`marital.status`,race,workclass,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY sex,income,occupation,`marital.status`,race,workclass

Resulted in 1937 records

Normalized Euclidean distance for (`hours.per.week`): 38.38

Hellinger Distance: 0.21

==================================================

#### Real

| | `native.country` | relationship | sex | `marital.status` | race | occupation | education | age | count |
|---|---|---|---|---|---|---|---|---|---|
| **11253** | United-States | Unmarried | Female | Never-married | White | Handlers-cleaners | Assoc-acdm | nan | nan |
| **857** | Columbia | Husband | Male | Married-civ-spouse | Asian-Pac-Islander | Exec-managerial | Bachelors | nan | nan |
| **5974** | Puerto-Rico | Not-in-family | Male | Divorced | Black | Transport-moving | Some-college | nan | nan |
| **3391** | Japan | Not-in-family | Female | Separated | White | ? | HS-grad | nan | nan |
| **7255** | Puerto-Rico | Unmarried | Male | Divorced | White | Handlers-cleaners | HS-grad | nan | nan |

#### Synthetic

| | `native.country` | relationship | sex | `marital.status` | race | occupation | education | age | count |
|---|---|---|---|---|---|---|---|---|---|
| **11253** | United-States | Unmarried | Female | Never-married | White | Handlers-cleaners | Assoc-acdm | 49 | 2 |
| **857** | Columbia | Husband | Male | Married-civ-spouse | Asian-Pac-Islander | Exec-managerial | Bachelors | 40 | 1 |
| **5974** | Puerto-Rico | Not-in-family | Male | Divorced | Black | Transport-moving | Some-college | 51 | 1 |
| **3391** | Japan | Not-in-family | Female | Separated | White | ? | HS-grad | 38 | 1 |
| **7255** | Puerto-Rico | Unmarried | Male | Divorced | White | Handlers-cleaners | HS-grad | 52 | 1 |

SQL for Real:
SELECT `native.country`,relationship,sex,`marital.status`,race,occupation,education,SUM(age), COUNT(*) FROM C1 GROUP BY `native.country`,relationship,sex,`marital.status`,race,occupation,education

Resulted in 8385 records

SQL for Synthetic:
SELECT `native.country`,relationship,sex,`marital.status`,race,occupation,education,SUM(age), COUNT(*) FROM C1_syn_06 GROUP BY

`native.country`,relationship,sex,`marital.status`,race,occupation,education

Resulted in 12141 records

Normalized Euclidean distance for (age): 56.07

Hellinger Distance: 0.505

===========================================

### Real

| | workclass | `marital.status` | `native.country` | fnlwgt | count |
|---|---|---|---|---|---|
| 204 | Local-gov | Married-civ-spouse | Puerto-Rico | 312427 | 6 |
| 498 | Self-emp-inc | Married-civ-spouse | Italy | 373593 | 6 |
| 597 | Self-emp-not-inc | Never-married | ? | 467936 | 16 |
| 513 | Self-emp-inc | Married-spouse-absent | India | 138370 | 1 |
| 123 | Federal-gov | Married-civ-spouse | India | 350387 | 4 |

### Synthetic

| | workclass | `marital.status` | `native.country` | fnlwgt | count |
|---|---|---|---|---|---|
| 204 | Local-gov | Married-civ-spouse | Puerto-Rico | 191359.137860 | 264.000000 |
| 498 | Self-emp-inc | Married-civ-spouse | Italy | nan | nan |
| 597 | Self-emp-not-inc | Never-married | ? | 188340.990142 | 26.000000 |
| 513 | Self-emp-inc | Married-spouse-absent | India | nan | nan |
| 123 | Federal-gov | Married-civ-spouse | India | nan | nan |

SQL for Real:
SELECT workclass,`marital.status`,`native.country`,MAX(fnlwgt), COUNT(*) FROM C1 GROUP BY workclass,`marital.status`,`native.country`

Resulted in 705 records

SQL for Synthetic:
SELECT workclass,`marital.status`,`native.country`,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY workclass,`marital.status`,`native.country`

Resulted in 657 records

Normalized Euclidean distance for (fnlwgt): 19.92

Hellinger Distance: 0.399

===========================================

### Real

| | workclass | race | relationship | `marital.status` | sex | occupation | education | income | capital | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 7484 | Self-emp-inc | Asian-Pac-Islander | Other-relative | Married-civ-spouse | Female | Other-service | Bachelors | <=50K | 0.000000 | 1 |
| 7338 | Private | White | Wife | Married-civ-spouse | Female | Exec-managerial | HS-grad | >50K | 1775.608696 | 23 |
| 2412 | Private | Amer-Indian-Eskimo | Not-in-family | Divorced | Male | Craft-repair | HS-grad | <=50K | 0.000000 | 2 |
| 2028 | Local-gov | White | Not-in-family | Widowed | Female | Transport-moving | HS-grad | <=50K | 0.000000 | 1 |
| 6942 | Private | White | Unmarried | Never-married | Female | Other-service | 11th | <=50K | 0.000000 | 1 |

### Synthetic

| | workclass | race | relationship | `marital.status` | sex | occupation | education | income | capital | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 7484 | Self-emp-inc | Asian-Pac-Islander | Other-relative | Married-civ-spouse | Female | Other-service | Bachelors | <=50K | nan | nan |
| 7338 | Private | White | Wife | Married-civ-spouse | Female | Exec-managerial | HS-grad | >50K | -76.384177 | 10.000000 |
| 2412 | Private | Amer-Indian-Eskimo | Not-in-family | Divorced | Male | Craft-repair | HS-grad | <=50K | nan | nan |
| 2028 | Local-gov | White | Not-in-family | Widowed | Female | Transport-moving | HS-grad | <=50K | 94.075249 | 1.000000 |
| 6942 | Private | White | Unmarried | Never-married | Female | Other-service | 11th | <=50K | 57.169051 | 6.000000 |

SQL for Real:
SELECT workclass,race,relationship,`marital.status`,sex,occupation,education,income,AVG(capital), COUNT(*) FROM C1 GROUP BY workclass,race,relationship,`marital.status`,sex,occupation,education,income

Resulted in 9897 records

SQL for Synthetic:
SELECT workclass,race,relationship,`marital.status`,sex,occupation,education,income,AVG(capital), COUNT(*) FROM C1_syn_06 GROUP BY workclass,race,relationship,`marital.status`,sex,occupation,education,income

Resulted in 9113 records

Normalized Euclidean distance for (capital): 69.99

Hellinger Distance: 0.328

===========================================

### Real

| | `marital.status` | age | count |
|---|---|---|---|
| 4 | Never-married | 90 | 16117 |
| 5 | Separated | 90 | 1530 |
| 3 | Married-spouse-absent | 87 | 628 |
| 2 | Married-civ-spouse | 90 | 22379 |
| 1 | Married-AF-spouse | 75 | 37 |

### Synthetic

| | `marital.status` | age | count |
|---|---|---|---|
| 4 | Never-married | 90.000000 | 16343.000000 |
| 5 | Separated | 90.000000 | 1504.000000 |
| 3 | Married-spouse-absent | 87.000000 | 540.000000 |
| 2 | Married-civ-spouse | 90.000000 | 22152.000000 |
| 1 | Married-AF-spouse | 50.000000 | 8.000000 |

SQL for Real:

SELECT `marital.status`,MAX(age), COUNT(*) FROM C1 GROUP BY `marital.status`

Resulted in 7 records

SQL for Synthetic:
SELECT `marital.status`,MAX(age), COUNT(*) FROM C1_syn_06 GROUP BY `marital.status`

Resulted in 7 records

Normalized Euclidean distance for (age): 2.65

Hellinger Distance: 0.013

================================================

Real

| | `native.country` | fnlwgt | count |
|---|---|---|---|
| 5 | Cuba | 48853 | 138 |
| 7 | Ecuador | 61298 | 45 |
| 20 | Iran | 35864 | 59 |
| 10 | France | 29235 | 38 |
| 18 | Hungary | 114263 | 19 |

Synthetic

| | `native.country` | fnlwgt | count |
|---|---|---|---|
| 5 | Cuba | 179577.687682 | 12.000000 |
| 7 | Ecuador | 177899.523835 | 200.000000 |
| 20 | Iran | 179545.975184 | 205.000000 |
| 10 | France | 182709.774219 | 1.000000 |
| 18 | Hungary | 179550.245039 | 1045.000000 |

SQL for Real:
SELECT `native.country`,MIN(fnlwgt), COUNT(*) FROM C1 GROUP BY `native.country`

Resulted in 42 records

SQL for Synthetic:
SELECT `native.country`,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`

Resulted in 41 records

Normalized Euclidean distance for (fnlwgt): 6.4

Hellinger Distance: 0.396

================================================

Real

| | race | income | `native.country` | sex | age | count |
|---|---|---|---|---|---|---|
| 333 | White | >50K | Portugal | Female | 24.000000 | 2.000000 |
| 149 | Black | <=50K | Vietnam | Female | nan | nan |
| 58 | Asian-Pac-Islander | <=50K | Philippines | Male | 18.000000 | 108.000000 |
| 302 | White | >50K | Cuba | Male | 30.000000 | 29.000000 |
| 254 | White | <=50K | Iran | Female | 22.000000 | 10.000000 |

Synthetic

| | race | income | `native.country` | sex | age | count |
|---|---|---|---|---|---|---|
| 333 | White | >50K | Portugal | Female | 83 | 1 |
| 149 | Black | <=50K | Vietnam | Female | 20 | 2 |
| 58 | Asian-Pac-Islander | <=50K | Philippines | Male | 31 | 4 |
| 302 | White | >50K | Cuba | Male | 35 | 3 |
| 254 | White | <=50K | Iran | Female | 23 | 46 |

SQL for Real:
SELECT race,income,`native.country`,sex,MIN(age), COUNT(*) FROM C1 GROUP BY race,income,`native.country`,sex

Resulted in 325 records

SQL for Synthetic:
SELECT race,income,`native.country`,sex,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY race,income,`native.country`,sex

Resulted in 344 records

Normalized Euclidean distance for (age): 14.56

Hellinger Distance: 0.434

================================================

Real

| | income | sex | `native.country` | `hours.per.week` | count |
|---|---|---|---|---|---|
| 71 | <=50K | Male | Philippines | 4403 | 115 |
| 24 | <=50K | Female | Japan | 931 | 25 |
| 50 | <=50K | Male | El-Salvador | 3382 | 91 |
| 29 | <=50K | Female | Peru | 645 | 17 |
| 23 | <=50K | Female | Jamaica | 2077 | 55 |

## Synthetic

|  | income | sex | `native.country` | `hours.per.week` | count |
|---|---|---|---|---|---|
| 71 | <=50K | Male | Philippines | 481.874109 | 12.000000 |
| 24 | <=50K | Female | Japan | 13115.709789 | 328.000000 |
| 50 | <=50K | Male | El-Salvador | 398.024997 | 10.000000 |
| 29 | <=50K | Female | Peru | 15772.970632 | 395.000000 |
| 23 | <=50K | Female | Jamaica | 2080.429446 | 52.000000 |

SQL for Real:
SELECT income,sex,`native.country`,SUM(`hours.per.week`), COUNT(*) FROM C1 GROUP BY income,sex,`native.country`

Resulted in 160 records

SQL for Synthetic:
SELECT income,sex,`native.country`,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY income,sex,`native.country`

Resulted in 130 records

Normalized Euclidean distance for (`hours.per.week`): 11.18

Hellinger Distance: 0.418

==========================================

## Real

|  | race | workclass | income | `native.country` | education | relationship | sex | occupation | `marital.status` | age | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 16279 | White | Self-emp-not-inc | >50K | United-States | Some-college | Own-child | Male | Craft-repair | Never-married | nan | nan |
| 10672 | White | Private | <=50K | Puerto-Rico | HS-grad | Not-in-family | Male | Farming-fishing | Never-married | nan | nan |
| 14926 | White | Self-emp-not-inc | <=50K | ? | HS-grad | Husband | Male | Farming-fishing | Married-civ-spouse | nan | nan |
| 10534 | White | Private | <=50K | Puerto-Rico | Bachelors | Own-child | Female | Prof-specialty | Separated | nan | nan |
| 4176 | White | ? | <=50K | Laos | HS-grad | Own-child | Male | ? | Never-married | nan | nan |

## Synthetic

|  | race | workclass | income | `native.country` | education | relationship | sex | occupation | `marital.status` | age | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 16279 | White | Self-emp-not-inc | >50K | United-States | Some-college | Own-child | Male | Craft-repair | Never-married | 30 | 1 |
| 10672 | White | Private | <=50K | Puerto-Rico | HS-grad | Not-in-family | Male | Farming-fishing | Never-married | 18 | 2 |
| 14926 | White | Self-emp-not-inc | <=50K | ? | HS-grad | Husband | Male | Farming-fishing | Married-civ-spouse | 50 | 9 |
| 10534 | White | Private | <=50K | Puerto-Rico | Bachelors | Own-child | Female | Prof-specialty | Separated | 61 | 1 |
| 4176 | White | ? | <=50K | Laos | HS-grad | Own-child | Male | ? | Never-married | 19 | 1 |

SQL for Real:
SELECT race,workclass,income,`native.country`,education,relationship,sex,occupation,`marital.status`,MIN(age), COUNT(*) FROM C1 GROUP BY race,workclass,income,`native.country`,education,relationship,sex,occupation,`marital.status`

Resulted in 12664 records

SQL for Synthetic:
SELECT race,workclass,income,`native.country`,education,relationship,sex,occupation,`marital.status`,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY race,workclass,income,`native.country`,education,relationship,sex,occupation,`marital.status`

Resulted in 17127 records

Normalized Euclidean distance for (age): 64.61

Hellinger Distance: 0.549

==========================================

## Real

|  | relationship | `native.country` | sex | workclass | education | race | income | capital | count |
|---|---|---|---|---|---|---|---|---|---|
| 3508 | Other-relative | Nicaragua | Female | Private | Some-college | White | <=50K | nan | nan |
| 784 | Husband | Peru | Male | Private | Bachelors | Asian-Pac-Islander | <=50K | nan | nan |
| 6607 | Wife | United-States | Female | Self-emp-not-inc | Assoc-acdm | White | <=50K | 0.000000 | 4.000000 |
| 5579 | Unmarried | Puerto-Rico | Male | Private | HS-grad | Other | <=50K | 0.000000 | 1.000000 |
| 2999 | Not-in-family | United-States | Female | Private | Bachelors | Asian-Pac-Islander | <=50K | 0.000000 | 8.000000 |

## Synthetic

|  | relationship | `native.country` | sex | workclass | education | race | income | capital | count |
|---|---|---|---|---|---|---|---|---|---|
| 3508 | Other-relative | Nicaragua | Female | Private | Some-college | White | <=50K | -19.387890 | 2 |
| 784 | Husband | Peru | Male | Private | Bachelors | Asian-Pac-Islander | <=50K | -28.371567 | 2 |
| 6607 | Wife | United-States | Female | Self-emp-not-inc | Assoc-acdm | White | <=50K | 0.451539 | 1 |
| 5579 | Unmarried | Puerto-Rico | Male | Private | HS-grad | Other | <=50K | 2.653915 | 1 |
| 2999 | Not-in-family | United-States | Female | Private | Bachelors | Asian-Pac-Islander | <=50K | -138.331132 | 14 |

SQL for Real:
SELECT relationship,`native.country`,sex,workclass,education,race,income,MIN(capital), COUNT(*) FROM C1 GROUP BY relationship,`native.country`,sex,workclass,education,race,income

Resulted in 5000 records

SQL for Synthetic:
SELECT relationship,`native.country`,sex,workclass,education,race,income,MIN(capital), COUNT(*) FROM C1_syn_06 GROUP BY relationship,`native.country`,sex,workclass,education,race,income

Resulted in 6659 records

Normalized Euclidean distance for (capital): 44.68

Hellinger Distance: 0.485

==========================================

Real

| | sex | race | `native.country` | workclass | `marital.status` | occupation | capital | count |
|---|---|---|---|---|---|---|---|---|
| 5329 | Male | White | United-States | Private | Married-civ-spouse | Protective-serv | 15024.000000 | 101.000000 |
| 1161 | Female | White | China | Private | Divorced | Craft-repair | nan | nan |
| 2281 | Female | White | Taiwan | Private | Married-civ-spouse | Machine-op-inspct | nan | nan |
| 1449 | Female | White | Hungary | Federal-gov | Never-married | Adm-clerical | nan | nan |
| 3376 | Male | Black | Puerto-Rico | State-gov | Married-civ-spouse | Other-service | nan | nan |

Synthetic

| | sex | race | `native.country` | workclass | `marital.status` | occupation | capital | count |
|---|---|---|---|---|---|---|---|---|
| 5329 | Male | White | United-States | Private | Married-civ-spouse | Protective-serv | 635.308276 | 76 |
| 1161 | Female | White | China | Private | Divorced | Craft-repair | -2.724966 | 1 |
| 2281 | Female | White | Taiwan | Private | Married-civ-spouse | Machine-op-inspct | -204.296613 | 1 |
| 1449 | Female | White | Hungary | Federal-gov | Never-married | Adm-clerical | -39.560578 | 1 |
| 3376 | Male | Black | Puerto-Rico | State-gov | Married-civ-spouse | Other-service | 62.647144 | 1 |

SQL for Real:
SELECT sex,race,`native.country`,workclass,`marital.status`,occupation,MAX(capital), COUNT(*) FROM C1 GROUP BY sex,race,`native.country`,workclass,`marital.status`,occupation

Resulted in 3937 records

SQL for Synthetic:
SELECT sex,race,`native.country`,workclass,`marital.status`,occupation,MAX(capital), COUNT(*) FROM C1_syn_06 GROUP BY sex,race,`native.country`,workclass,`marital.status`,occupation

Resulted in 5543 records

Normalized Euclidean distance for (capital): 41.57

Hellinger Distance: 0.465

========================================

Real

| | race | occupation | relationship | fnlwgt | count |
|---|---|---|---|---|---|
| 97 | Asian-Pac-Islander | Handlers-cleaners | Other-relative | 150707.500000 | 4 |
| 349 | White | Other-service | Unmarried | 182325.233522 | 531 |
| 196 | Black | Priv-house-serv | Wife | 190912.500000 | 6 |
| 212 | Black | Sales | Own-child | 242872.538462 | 104 |
| 50 | Amer-Indian-Eskimo | Prof-specialty | Unmarried | 69804.583333 | 12 |

Synthetic

| | race | occupation | relationship | fnlwgt | count |
|---|---|---|---|---|---|
| 97 | Asian-Pac-Islander | Handlers-cleaners | Other-relative | 181532.840752 | 2.000000 |
| 349 | White | Other-service | Unmarried | 181402.453491 | 567.000000 |
| 196 | Black | Priv-house-serv | Wife | nan | nan |
| 212 | Black | Sales | Own-child | 187340.929707 | 93.000000 |
| 50 | Amer-Indian-Eskimo | Prof-specialty | Unmarried | 181623.590065 | 3.000000 |

SQL for Real:
SELECT race,occupation,relationship,AVG(fnlwgt), COUNT(*) FROM C1 GROUP BY race,occupation,relationship
Resulted in 387 records

SQL for Synthetic:
SELECT race,occupation,relationship,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY race,occupation,relationship
Resulted in 376 records

Normalized Euclidean distance for (fnlwgt): 19.0

Hellinger Distance: 0.071

========================================

Real

| | education | capital | count |
|---|---|---|---|
| 3 | 1st-4th | 7688 | 247 |
| 7 | Assoc-acdm | 99999 | 1601 |
| 15 | Some-college | 99999 | 10878 |
| 9 | Bachelors | 99999 | 8025 |
| 11 | HS-grad | 99999 | 15784 |

Synthetic

| | education | capital | count |
|---|---|---|---|
| 3 | 1st-4th | 315.046646 | 154.000000 |
| 7 | Assoc-acdm | 364.880776 | 1755.000000 |
| 15 | Some-college | 414.355939 | 10540.000000 |
| 9 | Bachelors | 354.918062 | 8164.000000 |
| 11 | HS-grad | 374.086257 | 16455.000000 |

SQL for Real:
SELECT education,MAX(capital), COUNT(*) FROM C1 GROUP BY education
Resulted in 16 records

SQL for Synthetic:
SELECT education,MAX(capital), COUNT(*) FROM C1_syn_06 GROUP BY education
Resulted in 16 records

Normalized Euclidean distance for (capital): 4.0

Hellinger Distance: 0.026

========================================

Real

| | relationship | race | `native.country` | workclass | income | `marital.status` | occupation | sex | education | age | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3142 | Husband | White | Scotland | Private | <=50K | Married-civ-spouse | Craft-repair | Male | 11th | nan | nan |
| 12580 | Own-child | White | United-States | Private | <=50K | Married-spouse-absent | Farming-fishing | Male | 11th | nan | nan |
| 8259 | Not-in-family | White | United-States | Private | <=50K | Divorced | Adm-clerical | Female | 10th | 61.000000 | 1.000000 |
| 13826 | Unmarried | Black | United-States | State-gov | <=50K | Never-married | Adm-clerical | Female | Assoc-voc | nan | nan |
| 4919 | Not-in-family | Black | United-States | Private | <=50K | Divorced | Exec-managerial | Female | Bachelors | nan | nan |

Synthetic

| | relationship | race | `native.country` | workclass | income | `marital.status` | occupation | sex | education | age | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3142 | Husband | White | Scotland | Private | <=50K | Married-civ-spouse | Craft-repair | Male | 11th | 33 | 2 |
| 12580 | Own-child | White | United-States | Private | <=50K | Married-spouse-absent | Farming-fishing | Male | 11th | 27 | 1 |
| 8259 | Not-in-family | White | United-States | Private | <=50K | Divorced | Adm-clerical | Female | 10th | 57 | 1 |
| 13826 | Unmarried | Black | United-States | State-gov | <=50K | Never-married | Adm-clerical | Female | Assoc-voc | 23 | 1 |
| 4919 | Not-in-family | Black | United-States | Private | <=50K | Divorced | Exec-managerial | Female | Bachelors | 31 | 2 |

SQL for Real:
SELECT relationship,race,`native.country`,workclass,income,`marital.status`,occupation,sex,education,MIN(age), COUNT(*) FROM C1 GROUP BY relationship,race,`native.country`,workclass,income,`marital.status`,occupation,sex,education

Resulted in 12664 records

SQL for Synthetic:
SELECT relationship,race,`native.country`,workclass,income,`marital.status`,occupation,sex,education,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY relationship,race,`native.country`,workclass,income,`marital.status`,occupation,sex,education

Resulted in 17127 records

Normalized Euclidean distance for (age): 64.61

Hellinger Distance: 0.549

========================================

Real

| | workclass | `marital.status` | education | income | sex | race | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|
| 398 | Federal-gov | Divorced | Some-college | <=50K | Female | Amer-Indian-Eskimo | 21876 | 3 |
| 1527 | Private | Never-married | 9th | <=50K | Female | White | 53367 | 43 |
| 1990 | Self-emp-inc | Never-married | HS-grad | <=50K | Female | Asian-Pac-Islander | 119422 | 1 |
| 1906 | Self-emp-inc | Married-civ-spouse | Bachelors | <=50K | Female | Asian-Pac-Islander | 135643 | 4 |
| 296 | ? | Separated | HS-grad | <=50K | Female | Asian-Pac-Islander | 313445 | 1 |

Synthetic

| | workclass | `marital.status` | education | income | sex | race | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|
| 398 | Federal-gov | Divorced | Some-college | <=50K | Female | Amer-Indian-Eskimo | 170153.474415 | 1.000000 |
| 1527 | Private | Never-married | 9th | <=50K | Female | White | 161815.980784 | 47.000000 |
| 1990 | Self-emp-inc | Never-married | HS-grad | <=50K | Female | Asian-Pac-Islander | nan | nan |
| 1906 | Self-emp-inc | Married-civ-spouse | Bachelors | <=50K | Female | Asian-Pac-Islander | 166910.973650 | 1.000000 |
| 296 | ? | Separated | HS-grad | <=50K | Female | Asian-Pac-Islander | nan | nan |

SQL for Real:
SELECT workclass,`marital.status`,education,income,sex,race,MIN(fnlwgt), COUNT(*) FROM C1 GROUP BY workclass,`marital.status`,education,income,sex,race
Resulted in 2701 records

SQL for Synthetic:
SELECT workclass,`marital.status`,education,income,sex,race,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY workclass,`marital.status`,education,income,sex,race
Resulted in 2264 records

Normalized Euclidean distance for (fnlwgt): 41.5

Hellinger Distance: 0.219

========================================

Real

| | race | relationship | `native.country` | education | workclass | income | `marital.status` | sex | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 5117 | White | Not-in-family | Puerto-Rico | HS-grad | Self-emp-not-inc | <=50K | Widowed | Female | nan | nan |
| 6391 | White | Other-relative | United-States | Preschool | Private | <=50K | Never-married | Male | nan | nan |
| 1284 | Black | Not-in-family | Puerto-Rico | Masters | Private | <=50K | Divorced | Female | nan | nan |
| 7052 | White | Own-child | Puerto-Rico | HS-grad | State-gov | <=50K | Divorced | Male | nan | nan |
| 6668 | White | Own-child | Hungary | Some-college | Private | >50K | Married-civ-spouse | Female | nan | nan |

Synthetic

| | race | relationship | `native.country` | education | workclass | income | `marital.status` | sex | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 5117 | White | Not-in-family | Puerto-Rico | HS-grad | Self-emp-not-inc | <=50K | Widowed | Female | 64.428571 | 7 |
| 6391 | White | Other-relative | United-States | Preschool | Private | <=50K | Never-married | Male | 26.000000 | 1 |
| 1284 | Black | Not-in-family | Puerto-Rico | Masters | Private | <=50K | Divorced | Female | 45.000000 | 2 |
| 7052 | White | Own-child | Puerto-Rico | HS-grad | State-gov | <=50K | Divorced | Male | 33.000000 | 1 |
| 6668 | White | Own-child | Hungary | Some-college | Private | >50K | Married-civ-spouse | Female | 28.000000 | 1 |

SQL for Real:
SELECT race,relationship,`native.country`,education,workclass,income,`marital.status`,sex,AVG(age), COUNT(*) FROM C1 GROUP BY race,relationship,`native.country`,education,workclass,income,`marital.status`,sex

Resulted in 6863 records

SELECT race,relationship,`native.country`,education,workclass,income,`marital.status`,sex,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY race,relationship,`native.country`,education,workclass,income,`marital.status`,sex

Resulted in 9165 records

Normalized Euclidean distance for (age): 50.11

Hellinger Distance: 0.502

==========================================

### Real

| | sex | age | count |
|---|---|---|---|
| **0** | Female | 17 | 16192 |
| **1** | Male | 17 | 32650 |

### Synthetic

| | sex | age | count |
|---|---|---|---|
| **0** | Female | 17.000000 | 17114.000000 |
| **1** | Male | 17.000000 | 31728.000000 |

SQL for Real:
SELECT sex,MIN(age), COUNT(*) FROM C1 GROUP BY sex

Resulted in 2 records

SQL for Synthetic:
SELECT sex,MIN(age), COUNT(*) FROM C1_syn_06 GROUP BY sex

Resulted in 2 records

Normalized Euclidean distance for (age): 0.0

Hellinger Distance: 0.014

==========================================

### Real

| | `marital.status` | age | count |
|---|---|---|---|
| **6** | Widowed | 59.377470 | 1518 |
| **2** | Married-civ-spouse | 43.353724 | 22379 |
| **5** | Separated | 39.725490 | 1530 |
| **1** | Married-AF-spouse | 31.945946 | 37 |
| **3** | Married-spouse-absent | 40.613057 | 628 |

### Synthetic

| | `marital.status` | age | count |
|---|---|---|---|
| **6** | Widowed | 59.203927 | 1579.000000 |
| **2** | Married-civ-spouse | 43.436710 | 22152.000000 |
| **5** | Separated | 39.946809 | 1504.000000 |
| **1** | Married-AF-spouse | 30.125000 | 8.000000 |
| **3** | Married-spouse-absent | 40.679630 | 540.000000 |

SQL for Real:
SELECT `marital.status`,AVG(age), COUNT(*) FROM C1 GROUP BY `marital.status`

Resulted in 7 records

SQL for Synthetic:
SELECT `marital.status`,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY `marital.status`

Resulted in 7 records

Normalized Euclidean distance for (age): 2.65

Hellinger Distance: 0.013

==========================================

### Real

| | race | workclass | sex | `native.country` | occupation | education | relationship | `marital.status` | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|---|
| **9060** | White | Private | Female | United-States | Machine-op-inspct | Bachelors | Own-child | Never-married | 151910.000000 | 1.000000 |
| **4065** | White | ? | Female | Puerto-Rico | ? | Assoc-acdm | Own-child | Never-married | nan | nan |
| **15708** | White | State-gov | Male | Puerto-Rico | Farming-fishing | HS-grad | Husband | Married-civ-spouse | nan | nan |
| **6973** | White | Private | Female | Holand-Netherlands | Adm-clerical | Some-college | Own-child | Never-married | nan | nan |
| **7516** | White | Private | Female | Laos | Exec-managerial | Bachelors | Wife | Married-civ-spouse | nan | nan |

### Synthetic

| | race | workclass | sex | `native.country` | occupation | education | relationship | `marital.status` | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|---|
| **9060** | White | Private | Female | United-States | Machine-op-inspct | Bachelors | Own-child | Never-married | 189288.095565 | 2 |
| **4065** | White | ? | Female | Puerto-Rico | ? | Assoc-acdm | Own-child | Never-married | 170830.532865 | 1 |
| **15708** | White | State-gov | Male | Puerto-Rico | Farming-fishing | HS-grad | Husband | Married-civ-spouse | 175331.417896 | 1 |
| **6973** | White | Private | Female | Holand-Netherlands | Adm-clerical | Some-college | Own-child | Never-married | 204148.390461 | 5 |
| **7516** | White | Private | Female | Laos | Exec-managerial | Bachelors | Wife | Married-civ-spouse | 178932.458952 | 1 |

SQL for Real:
SELECT race,workclass,sex,`native.country`,occupation,education,relationship,`marital.status`,MAX(fnlwgt), COUNT(*) FROM C1 GROUP BY race,workclass,sex,`native.country`,occupation,education,relationship,`marital.status`

Resulted in 11374 records

SQL for Synthetic:
SELECT race,workclass,sex,`native.country`,occupation,education,relationship,`marital.status`,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY race,workclass,sex,`native.country`,occupation,education,relationship,`marital.status`

Resulted in 15994 records

Normalized Euclidean distance for (fnlwgt): 63.12

Hellinger Distance: 0.53

==============================================

### Real

|     | workclass    | race               | relationship   | fnlwgt    | count |
|-----|--------------|--------------------|----------------|-----------|-------|
| 98  | Private      | Black              | Husband        | 162557757 | 674   |
| 67  | Local-gov    | Asian-Pac-Islander | Own-child      | 1534167   | 8     |
| 177 | State-gov    | Black              | Not-in-family  | 18178926  | 76    |
| 172 | State-gov    | Asian-Pac-Islander | Other-relative | 242851    | 3     |
| 123 | Self-emp-inc | Asian-Pac-Islander | Wife           | 750383    | 6     |

### Synthetic

|     | workclass    | race               | relationship   | fnlwgt             | count      |
|-----|--------------|--------------------|----------------|--------------------|------------|
| 98  | Private      | Black              | Husband        | 100256109.355009   | 540.000000 |
| 67  | Local-gov    | Asian-Pac-Islander | Own-child      | 2086888.774294     | 11.000000  |
| 177 | State-gov    | Black              | Not-in-family  | 10007381.764456    | 59.000000  |
| 172 | State-gov    | Asian-Pac-Islander | Other-relative | 163164.612004      | 1.000000   |
| 123 | Self-emp-inc | Asian-Pac-Islander | Wife           | 1015372.858229     | 6.000000   |

SQL for Real:
SELECT workclass,race,relationship,SUM(fnlwgt), COUNT(*) FROM C1 GROUP BY workclass,race,relationship
Resulted in 192 records

SQL for Synthetic:
SELECT workclass,race,relationship,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY workclass,race,relationship
Resulted in 181 records

Normalized Euclidean distance for (fnlwgt): 13.27

Hellinger Distance: 0.057

==============================================

### Real

|   | race               | `hours.per.week` | count |
|---|--------------------|------------------|-------|
| 4 | White              | 1                | 41762 |
| 1 | Asian-Pac-Islander | 1                | 1519  |
| 3 | Other              | 5                | 406   |
| 0 | Amer-Indian-Eskimo | 3                | 470   |
| 2 | Black              | 1                | 4685  |

### Synthetic

|   | race               | `hours.per.week` | count        |
|---|--------------------|------------------|--------------|
| 4 | White              | 38.865705        | 43379.000000 |
| 1 | Asian-Pac-Islander | 38.929010        | 993.000000   |
| 3 | Other              | 38.904838        | 348.000000   |
| 0 | Amer-Indian-Eskimo | 38.913528        | 202.000000   |
| 2 | Black              | 38.885831        | 3920.000000  |

SQL for Real:
SELECT race,MIN(`hours.per.week`), COUNT(*) FROM C1 GROUP BY race
Resulted in 5 records

SQL for Synthetic:
SELECT race,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY race
Resulted in 5 records

Normalized Euclidean distance for (`hours.per.week`): 2.24

Hellinger Distance: 0.041

==============================================

### Real

|      | `marital.status`   | workclass   | relationship   | education    | income | fnlwgt | count |
|------|--------------------|-------------|----------------|--------------|--------|--------|-------|
| 1141 | Never-married      | Local-gov   | Unmarried      | Masters      | >50K   | 88638  | 1     |
| 1449 | Separated          | Local-gov   | Not-in-family  | Some-college | >50K   | 212600 | 1     |
| 1047 | Never-married      | Federal-gov | Not-in-family  | Doctorate    | <=50K  | 345259 | 1     |
| 463  | Married-civ-spouse | Federal-gov | Husband        | Some-college | >50K   | 914061 | 98    |
| 1112 | Never-married      | Local-gov   | Other-relative | Some-college | <=50K  | 315287 | 7     |

### Synthetic

|      | `marital.status`   | workclass   | relationship   | education    | income | fnlwgt          | count     |
|------|--------------------|-------------|----------------|--------------|--------|-----------------|-----------|
| 1141 | Never-married      | Local-gov   | Unmarried      | Masters      | >50K   | 189054.833184   | 2.000000  |
| 1449 | Separated          | Local-gov   | Not-in-family  | Some-college | >50K   | nan             | nan       |
| 1047 | Never-married      | Federal-gov | Not-in-family  | Doctorate    | <=50K  | 186827.789105   | 2.000000  |
| 463  | Married-civ-spouse | Federal-gov | Husband        | Some-college | >50K   | 196840.840373   | 41.000000 |
| 1112 | Never-married      | Local-gov   | Other-relative | Some-college | <=50K  | 204464.194409   | 7.000000  |

SQL for Real:
SELECT `marital.status`,workclass,relationship,education,income,MAX(fnlwgt), COUNT(*) FROM C1 GROUP BY `marital.status`,workclass,relationship,education,income
Resulted in 1842 records

SQL for Synthetic:
SELECT `marital.status`,workclass,relationship,education,income,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY `marital.status`,workclass,relationship,education,income

Resulted in 1748 records

Normalized Euclidean distance for (fnlwgt): 36.51

Hellinger Distance: 0.196

=============================================

Real

| | relationship | income | `native.country` | education | `marital.status` | race | workclass | sex | occupation | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9386 | Not-in-family | >50K | United-States | HS-grad | Never-married | White | Federal-gov | Female | Adm-clerical | nan | nan |
| 6361 | Not-in-family | <=50K | Puerto-Rico | Bachelors | Divorced | White | Private | Female | Sales | nan | nan |
| 5419 | Not-in-family | <=50K | Japan | HS-grad | Divorced | White | Private | Female | Transport-moving | nan | nan |
| 5408 | Not-in-family | <=50K | Japan | HS-grad | Divorced | White | Federal-gov | Male | Adm-clerical | nan | nan |
| 8920 | Not-in-family | <=50K | United-States | Some-college | Married-civ-spouse | White | Private | Male | Adm-clerical | nan | nan |

Synthetic

| | relationship | income | `native.country` | education | `marital.status` | race | workclass | sex | occupation | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9386 | Not-in-family | >50K | United-States | HS-grad | Never-married | White | Federal-gov | Female | Adm-clerical | 150560.706454 | 1 |
| 6361 | Not-in-family | <=50K | Puerto-Rico | Bachelors | Divorced | White | Private | Female | Sales | 176119.455004 | 3 |
| 5419 | Not-in-family | <=50K | Japan | HS-grad | Divorced | White | Private | Female | Transport-moving | 177937.440112 | 1 |
| 5408 | Not-in-family | <=50K | Japan | HS-grad | Divorced | White | Federal-gov | Male | Adm-clerical | 178313.894962 | 1 |
| 8920 | Not-in-family | <=50K | United-States | Some-college | Married-civ-spouse | White | Private | Male | Adm-clerical | 187076.947274 | 1 |

SQL for Real:
SELECT relationship,income,`native.country`,education,`marital.status`,race,workclass,sex,occupation,MAX(fnlwgt), COUNT(*) FROM C1 GROUP BY relationship,income,`native.country`,education,`marital.status`,race,workclass,sex,occupation

Resulted in 12664 records

SQL for Synthetic:
SELECT relationship,income,`native.country`,education,`marital.status`,race,workclass,sex,occupation,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY relationship,income,`native.country`,education,`marital.status`,race,workclass,sex,occupation

Resulted in 17127 records

Normalized Euclidean distance for (fnlwgt): 64.61

Hellinger Distance: 0.549

=============================================

Real

| | `marital.status` | sex | `native.country` | education | income | occupation | race | relationship | workclass | age | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8794 | Married-civ-spouse | Male | United-States | Some-college | >50K | Adm-clerical | Black | Husband | Federal-gov | 42.000000 | 4.000000 |
| 12502 | Never-married | Male | Holand-Netherlands | HS-grad | <=50K | Other-service | White | Unmarried | Private | nan | nan |
| 1828 | Divorced | Female | United-States | Some-college | <=50K | Tech-support | White | Not-in-family | State-gov | nan | nan |
| 5189 | Married-civ-spouse | Male | Columbia | Some-college | <=50K | Handlers-cleaners | White | Husband | Local-gov | nan | nan |
| 1572 | Divorced | Female | United-States | HS-grad | <=50K | Machine-op-inspct | Asian-Pac-Islander | Unmarried | Private | nan | nan |

Synthetic

| | `marital.status` | sex | `native.country` | education | income | occupation | race | relationship | workclass | age | count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8794 | Married-civ-spouse | Male | United-States | Some-college | >50K | Adm-clerical | Black | Husband | Federal-gov | 49.000000 | 1 |
| 12502 | Never-married | Male | Holand-Netherlands | HS-grad | <=50K | Other-service | White | Unmarried | Private | 19.000000 | 1 |
| 1828 | Divorced | Female | United-States | Some-college | <=50K | Tech-support | White | Not-in-family | State-gov | 39.000000 | 1 |
| 5189 | Married-civ-spouse | Male | Columbia | Some-college | <=50K | Handlers-cleaners | White | Husband | Local-gov | 42.000000 | 1 |
| 1572 | Divorced | Female | United-States | HS-grad | <=50K | Machine-op-inspct | Asian-Pac-Islander | Unmarried | Private | 34.000000 | 1 |

SQL for Real:
SELECT `marital.status`,sex,`native.country`,education,income,occupation,race,relationship,workclass,AVG(age), COUNT(*) FROM C1 GROUP BY `marital.status`,sex,`native.country`,education,income,occupation,race,relationship,workclass

Resulted in 12664 records

SQL for Synthetic:
SELECT `marital.status`,sex,`native.country`,education,income,occupation,race,relationship,workclass,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY `marital.status`,sex,`native.country`,education,income,occupation,race,relationship,workclass

Resulted in 17127 records

Normalized Euclidean distance for (age): 64.61

Hellinger Distance: 0.549

=============================================

Real

| | workclass | income | education | capital | count |
|---|---|---|---|---|---|
| 135 | Self-emp-inc | >50K | 12th | 0 | 2 |
| 182 | State-gov | <=50K | 5th-6th | 0 | 4 |
| 185 | State-gov | <=50K | Assoc-acdm | -1974 | 56 |
| 71 | Local-gov | <=50K | Some-college | -2377 | 448 |
| 132 | Self-emp-inc | <=50K | Some-college | -2339 | 167 |

|  | workclass | income | Synthetic education | capital | count |
|---|---|---|---|---|---|
| 135 | Self-emp-inc | >50K | 12th | -136.583767 | 1.000000 |
| 182 | State-gov | <=50K | 5th-6th | -104.209948 | 6.000000 |
| 185 | State-gov | <=50K | Assoc-acdm | -325.928398 | 60.000000 |
| 71 | Local-gov | <=50K | Some-college | -283.318818 | 540.000000 |
| 132 | Self-emp-inc | <=50K | Some-college | -225.915020 | 250.000000 |

SQL for Real:
SELECT workclass,income,education,MIN(capital), COUNT(*) FROM C1 GROUP BY workclass,income,education

Resulted in 205 records

SQL for Synthetic:
SELECT workclass,income,education,MIN(capital), COUNT(*) FROM C1_syn_06 GROUP BY workclass,income,education

Resulted in 192 records

Normalized Euclidean distance for (capital): 13.75

Hellinger Distance: 0.154

==========================================

Real

|  | occupation | sex | relationship | income | `native.country` | education | `marital.status` | workclass | capital | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 5248 | Exec-managerial | Male | Not-in-family | <=50K | ? | Masters | Never-married | Private | nan | nan |
| 12739 | Sales | Male | Husband | <=50K | Puerto-Rico | Masters | Married-civ-spouse | Self-emp-inc | nan | nan |
| 9431 | Other-service | Male | Unmarried | <=50K | United-States | HS-grad | Never-married | Self-emp-not-inc | 0.000000 | 1.000000 |
| 9204 | Other-service | Male | Not-in-family | <=50K | United-States | Some-college | Never-married | Private | -1602.000000 | 103.000000 |
| 4982 | Exec-managerial | Male | Husband | <=50K | Taiwan | Some-college | Married-civ-spouse | Federal-gov | nan | nan |

Synthetic

|  | occupation | sex | relationship | income | `native.country` | education | `marital.status` | workclass | capital | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 5248 | Exec-managerial | Male | Not-in-family | <=50K | ? | Masters | Never-married | Private | -152.915131 | 2 |
| 12739 | Sales | Male | Husband | <=50K | Puerto-Rico | Masters | Married-civ-spouse | Self-emp-inc | -119.718126 | 7 |
| 9431 | Other-service | Male | Unmarried | <=50K | United-States | HS-grad | Never-married | Self-emp-not-inc | -30.321700 | 2 |
| 9204 | Other-service | Male | Not-in-family | <=50K | United-States | Some-college | Never-married | Private | -219.135217 | 64 |
| 4982 | Exec-managerial | Male | Husband | <=50K | Taiwan | Some-college | Married-civ-spouse | Federal-gov | 42.035629 | 1 |

SQL for Real:
SELECT occupation,sex,relationship,income,`native.country`,education,`marital.status`,workclass,MIN(capital), COUNT(*) FROM C1 GROUP BY occupation,sex,relationship,income,`native.country`,education,`marital.status`,workclass

Resulted in 10400 records

SQL for Synthetic:
SELECT occupation,sex,relationship,income,`native.country`,education,`marital.status`,workclass,MIN(capital), COUNT(*) FROM C1_syn_06 GROUP BY occupation,sex,relationship,income,`native.country`,education,`marital.status`,workclass

Resulted in 14639 records

Normalized Euclidean distance for (capital): 62.27

Hellinger Distance: 0.529

==========================================

Real

|  | sex | relationship | workclass | `marital.status` | occupation | race | income | education | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 752 | Female | Not-in-family | Private | Never-married | Other-service | Black | <=50K | Assoc-acdm | 137296 | 1 |
| 2450 | Female | Own-child | State-gov | Never-married | Prof-specialty | White | <=50K | Prof-school | 105936 | 1 |
| 2568 | Female | Unmarried | Federal-gov | Divorced | Exec-managerial | White | >50K | Masters | 348059 | 1 |
| 2068 | Female | Own-child | Private | Married-civ-spouse | Transport-moving | White | <=50K | HS-grad | 211585 | 1 |
| 3525 | Female | Unmarried | Self-emp-inc | Divorced | Exec-managerial | White | <=50K | HS-grad | 257764 | 2 |

Synthetic

|  | sex | relationship | workclass | `marital.status` | occupation | race | income | education | fnlwgt | count |
|---|---|---|---|---|---|---|---|---|---|---|
| 752 | Female | Not-in-family | Private | Never-married | Other-service | Black | <=50K | Assoc-acdm | nan | nan |
| 2450 | Female | Own-child | State-gov | Never-married | Prof-specialty | White | <=50K | Prof-school | nan | nan |
| 2568 | Female | Unmarried | Federal-gov | Divorced | Exec-managerial | White | >50K | Masters | nan | nan |
| 2068 | Female | Own-child | Private | Married-civ-spouse | Transport-moving | White | <=50K | HS-grad | nan | nan |
| 3525 | Female | Unmarried | Self-emp-inc | Divorced | Exec-managerial | White | <=50K | HS-grad | nan | nan |

SQL for Real:
SELECT sex,relationship,workclass,`marital.status`,occupation,race,income,education,MAX(fnlwgt), COUNT(*) FROM C1 GROUP BY sex,relationship,workclass,`marital.status`,occupation,race,income,education

Resulted in 9897 records

SQL for Synthetic:
SELECT sex,relationship,workclass,`marital.status`,occupation,race,income,education,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 GROUP BY sex,relationship,workclass,`marital.status`,occupation,race,income,education

Resulted in 9113 records

Normalized Euclidean distance for (fnlwgt): 69.99

Hellinger Distance: 0.328

==========================================

### Real

| | occupation | race | `marital.status` | income | sex | education | capital | count |
|---|---|---|---|---|---|---|---|---|
| **1028** | Craft-repair | White | Never-married | <=50K | Male | 5th-6th | 0 | 10 |
| **1535** | Farming-fishing | White | Divorced | <=50K | Female | Some-college | -1590 | 4 |
| **2059** | Machine-op-inspct | Other | Married-spouse-absent | <=50K | Female | 7th-8th | 0 | 1 |
| **1446** | Exec-managerial | White | Widowed | <=50K | Male | 12th | 2009 | 1 |
| **1343** | Exec-managerial | White | Married-civ-spouse | >50K | Male | 12th | 0 | 3 |

### Synthetic

| | occupation | race | `marital.status` | income | sex | education | capital | count |
|---|---|---|---|---|---|---|---|---|
| **1028** | Craft-repair | White | Never-married | <=50K | Male | 5th-6th | 1.135600 | 5.000000 |
| **1535** | Farming-fishing | White | Divorced | <=50K | Female | Some-college | 47.265738 | 1.000000 |
| **2059** | Machine-op-inspct | Other | Married-spouse-absent | <=50K | Female | 7th-8th | nan | nan |
| **1446** | Exec-managerial | White | Widowed | <=50K | Male | 12th | nan | nan |
| **1343** | Exec-managerial | White | Married-civ-spouse | >50K | Male | 12th | -283.093856 | 2.000000 |

SQL for Real:

SELECT occupation,race,`marital.status`,income,sex,education,MIN(capital), COUNT(*) FROM C1 GROUP BY occupation,race,`marital.status`,income,sex,education

Resulted in 4178 records

SQL for Synthetic:

SELECT occupation,race,`marital.status`,income,sex,education,MIN(capital), COUNT(*) FROM C1_syn_06 GROUP BY occupation,race,`marital.status`,income,sex,education

Resulted in 3545 records

Normalized Euclidean distance for (capital): 50.4

Hellinger Distance: 0.244

==========================================

### Real

| | race | `native.country` | sex | education | `marital.status` | relationship | occupation | income | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| **2899** | Black | United-States | Male | Bachelors | Married-civ-spouse | Husband | Adm-clerical | <=50K | 33.000000 | 1.000000 |
| **1884** | Black | Puerto-Rico | Female | Prof-school | Married-civ-spouse | Wife | Exec-managerial | <=50K | nan | nan |
| **152** | Amer-Indian-Eskimo | United-States | Male | HS-grad | Never-married | Other-relative | Craft-repair | <=50K | 40.000000 | 2.000000 |
| **2619** | Black | United-States | Female | Some-college | Divorced | Unmarried | Sales | <=50K | 90.000000 | 4.000000 |
| **2164** | Black | Scotland | Female | Bachelors | Widowed | Unmarried | Prof-specialty | <=50K | nan | nan |

### Synthetic

| | race | `native.country` | sex | education | `marital.status` | relationship | occupation | income | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| **2899** | Black | United-States | Male | Bachelors | Married-civ-spouse | Husband | Adm-clerical | <=50K | 62 | 4 |
| **1884** | Black | Puerto-Rico | Female | Prof-school | Married-civ-spouse | Wife | Exec-managerial | <=50K | 40 | 1 |
| **152** | Amer-Indian-Eskimo | United-States | Male | HS-grad | Never-married | Other-relative | Craft-repair | <=50K | 37 | 1 |
| **2619** | Black | United-States | Female | Some-college | Divorced | Unmarried | Sales | <=50K | 47 | 2 |
| **2164** | Black | Scotland | Female | Bachelors | Widowed | Unmarried | Prof-specialty | <=50K | 52 | 1 |

SQL for Real:

SELECT race,`native.country`,sex,education,`marital.status`,relationship,occupation,income,MAX(age), COUNT(*) FROM C1 GROUP BY race,`native.country`,sex,education,`marital.status`,relationship,occupation,income

Resulted in 9390 records

SQL for Synthetic:

SELECT race,`native.country`,sex,education,`marital.status`,relationship,occupation,income,MAX(age), COUNT(*) FROM C1_syn_06 GROUP BY race,`native.country`,sex,education,`marital.status`,relationship,occupation,income

Resulted in 13048 records

Normalized Euclidean distance for (age): 57.52

Hellinger Distance: 0.524

==========================================

### Real

| | income | occupation | `native.country` | `marital.status` | education | race | relationship | sex | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| **5767** | <=50K | Machine-op-inspct | Japan | Divorced | Some-college | Black | Not-in-family | Female | nan | nan |
| **11937** | >50K | Adm-clerical | Scotland | Married-civ-spouse | Bachelors | White | Husband | Male | nan | nan |
| **4234** | <=50K | Exec-managerial | United-States | Married-civ-spouse | Assoc-acdm | White | Husband | Male | 45.068966 | 29.000000 |
| **8076** | <=50K | Priv-house-serv | United-States | Never-married | 1st-4th | Asian-Pac-Islander | Not-in-family | Female | nan | nan |
| **6896** | <=50K | Other-service | Japan | Married-civ-spouse | 12th | White | Wife | Female | nan | nan |

### Synthetic

| | income | occupation | `native.country` | `marital.status` | education | race | relationship | sex | age | count |
|---|---|---|---|---|---|---|---|---|---|---|
| **5767** | <=50K | Machine-op-inspct | Japan | Divorced | Some-college | Black | Not-in-family | Female | 44.000000 | 1 |
| **11937** | >50K | Adm-clerical | Scotland | Married-civ-spouse | Bachelors | White | Husband | Male | 58.000000 | 2 |
| **4234** | <=50K | Exec-managerial | United-States | Married-civ-spouse | Assoc-acdm | White | Husband | Male | 45.406250 | 32 |
| **8076** | <=50K | Priv-house-serv | United-States | Never-married | 1st-4th | Asian-Pac-Islander | Not-in-family | Female | 25.000000 | 1 |
| **6896** | <=50K | Other-service | Japan | Married-civ-spouse | 12th | White | Wife | Female | 36.000000 | 1 |

SQL for Real:

SELECT income,occupation,`native.country`,`marital.status`,education,race,relationship,sex,AVG(age), COUNT(*) FROM C1 GROUP BY income,occupation,`native.country`,`marital.status`,education,race,relationship,sex

Resulted in 9390 records

SQL for Synthetic:

SELECT income,occupation,`native.country`,`marital.status`,education,race,relationship,sex,AVG(age), COUNT(*) FROM C1_syn_06 GROUP BY income,occupation,`native.country`,`marital.status`,education,race,relationship,sex

Resulted in 13048 records

Normalized Euclidean distance for (age): 57.52

Hellinger Distance: 0.524

══════════════════════════════════════════

Real

| | `native.country` | `hours.per.week` | count |
|---|---|---|---|
| **33** | Puerto-Rico | 80 | 184 |
| **17** | Hong | 65 | 30 |
| **30** | Philippines | 99 | 295 |
| **29** | Peru | 60 | 46 |
| **20** | Iran | 70 | 59 |

Synthetic

| | `native.country` | `hours.per.week` | count |
|---|---|---|---|
| **33** | Puerto-Rico | 40.546981 | 10305.000000 |
| **17** | Hong | nan | nan |
| **30** | Philippines | 40.443251 | 80.000000 |
| **29** | Peru | 40.545033 | 1207.000000 |
| **20** | Iran | 40.513728 | 205.000000 |

<u>SQL for Real:</u>
SELECT `native.country`,MAX(`hours.per.week`), COUNT(*) FROM C1 GROUP BY `native.country`

Resulted in 42 records

<u>SQL for Synthetic:</u>
SELECT `native.country`,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 GROUP BY `native.country`

Resulted in 41 records

Normalized Euclidean distance for (`hours.per.week`): 6.4

Hellinger Distance: 0.396

══════════════════════════════════════════

Hellinger Distance Summary: {'mean': 0.3142675150677951, 'median': 0.3955457288889714, 'stddev': 0.1942676140300674}

Euclidean distance Summary : {'mean': 32.23007336935236, 'median': 35.028559776273994, 'stddev': 23.161441462183728}

===========================================