Real

| | education | workclass | marital.status | relationship | native.country | race | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **196** | 10th | State-gov | Never-married | Not-in-family | United-States | White | nan | nan |
| **341** | 11th | Private | Never-married | Other-relative | Puerto-Rico | White | 40.000000 | 1.000000 |
| **2168** | HS-grad | ? | Never-married | Not-in-family | Peru | Black | nan | nan |
| **2132** | HS-grad | ? | Married-civ-spouse | Other-relative | Columbia | Black | nan | nan |
| **944** | Assoc-acdm | Federal-gov | Separated | Unmarried | United-States | White | 40.000000 | 1.000000 |

Synthetic

| | education | workclass | marital.status | relationship | native.country | race | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **196** | 10th | State-gov | Never-married | Not-in-family | United-States | White | 39.987552 | 1 |
| **341** | 11th | Private | Never-married | Other-relative | Puerto-Rico | White | 79.068085 | 2 |
| **2168** | HS-grad | ? | Never-married | Not-in-family | Peru | Black | 40.020854 | 1 |
| **2132** | HS-grad | ? | Married-civ-spouse | Other-relative | Columbia | Black | 40.020158 | 1 |
| **944** | Assoc-acdm | Federal-gov | Separated | Unmarried | United-States | White | 40.024314 | 1 |

SQL for Real:
SELECT education,workclass,`marital.status`,relationship,`native.country`,race,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (sex <> 'Male') GROUP BY education,workclass,`marital.status`,relationship,`native.country`,race

Resulted in 2769 records

SQL for Synthetic:
SELECT education,workclass,`marital.status`,relationship,`native.country`,race,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (sex <> 'Male') GROUP BY education,workclass,`marital.status`,relationship,`native.country`,race

Resulted in 4276 records

Normalized Euclidean distance for (`hours.per.week`): 33.0

Hellinger Distance: 0.237

==========

Real

| | marital.status | education | relationship | workclass | income | race | sex | native.country | occupation | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **2548** | Divorced | Masters | Unmarried | Private | <=50K | White | Male | United-States | Adm-clerical | 113323.000000 | 1.000000 |
| **11502** | Never-married | Bachelors | Not-in-family | State-gov | <=50K | Black | Female | Nicaragua | Adm-clerical | nan | nan |
| **12169** | Never-married | HS-grad | Not-in-family | Private | <=50K | Black | Male | United-States | Adm-clerical | 216187.500000 | 6.000000 |
| **6094** | Married-civ-spouse | HS-grad | Husband | ? | <=50K | White | Male | Hungary | ? | nan | nan |
| **9781** | Never-married | 11th | Own-child | Private | <=50K | White | Female | United-States | Protective-serv | 154097.000000 | 2.000000 |

Synthetic

| | marital.status | education | relationship | workclass | income | race | sex | native.country | occupation | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **2548** | Divorced | Masters | Unmarried | Private | <=50K | White | Male | United-States | Adm-clerical | 179784.544501 | 1 |
| **11502** | Never-married | Bachelors | Not-in-family | State-gov | <=50K | Black | Female | Nicaragua | Adm-clerical | 179668.147338 | 1 |
| **12169** | Never-married | HS-grad | Not-in-family | Private | <=50K | Black | Male | United-States | Adm-clerical | 186389.722403 | 4 |
| **6094** | Married-civ-spouse | HS-grad | Husband | ? | <=50K | White | Male | Hungary | ? | 180281.934249 | 2 |
| **9781** | Never-married | 11th | Own-child | Private | <=50K | White | Female | United-States | Protective-serv | 198448.023388 | 1 |

SQL for Real:
SELECT `marital.status`,education,relationship,workclass,income,race,sex,`native.country`,occupation,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (education IN ('7th-8th', 'Some-college', 'Assoc-acdm', '1st-4th', 'Masters', 'HS-grad', 'Bachelors', '10th', 'Assoc-voc', '12th', 'Doctorate', '9th', 'Prof-school', '11th', 'Preschool', '5th-6th')) OR (income LIKE '<=50K' OR age = 34) OR (`marital.status` = 'Married-civ-spouse' OR fnlwgt <= 325374) OR (race <> 'White' OR sex <> 'Male') OR (occupation IN ('Adm-clerical', 'Other-service', 'Tech-support', 'Armed-Forces', 'Prof-specialty', 'Exec-managerial', '?') OR relationship = 'Not-in-family') AND (`hours.per.week` = 2 AND capital >= 0) GROUP BY `marital.status`,education,relationship,workclass,income,race,sex,`native.country`,occupation

Resulted in 12664 records

SQL for Synthetic:
SELECT `marital.status`,education,relationship,workclass,income,race,sex,`native.country`,occupation,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (education IN ('7th-8th', 'Some-college', 'Assoc-acdm', '1st-4th', 'Masters', 'HS-grad', 'Bachelors', '10th', 'Assoc-voc', '12th', 'Doctorate', '9th', 'Prof-school', '11th', 'Preschool', '5th-6th')) OR (income LIKE '<=50K' OR age = 34) OR (`marital.status` = 'Married-civ-spouse' OR fnlwgt <= 325374) OR (race <> 'White' OR sex <> 'Male') OR (occupation IN ('Adm-clerical', 'Other-service', 'Tech-support', 'Armed-Forces', 'Prof-specialty', 'Exec-managerial', '?') OR relationship = 'Not-in-family') AND (`hours.per.week` = 2 AND capital >= 0) GROUP BY `marital.status`,education,relationship,workclass,income,race,sex,`native.country`,occupation

Resulted in 17127 records

Normalized Euclidean distance for (fnlwgt): 64.61

Hellinger Distance: 0.252

==========

Real

| occupation | sex | race | marital.status | income | relationship | education | native.country | workclass | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|

Synthetic

| occupation | sex | race | marital.status | income | relationship | education | native.country | workclass | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|

SQL for Real:
SELECT occupation,sex,race,`marital.status`,income,relationship,education,`native.country`,workclass,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (`marital.status` IN ('Divorced', 'Married-AF-spouse', 'Widowed', 'Never-married', 'Married-spouse-absent', 'Married-civ-spouse')) AND (race = 'White' AND fnlwgt BETWEEN 180052 AND 177487) GROUP BY occupation,sex,race,`marital.status`,income,relationship,education,`native.country`,workclass

Resulted in 0 records

SQL for Synthetic:
SELECT occupation,sex,race,`marital.status`,income,relationship,education,`native.country`,workclass,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` IN ('Divorced', 'Married-AF-spouse', 'Widowed', 'Never-married', 'Married-spouse-absent', 'Married-civ-spouse')) AND (race = 'White' AND fnlwgt BETWEEN 180052 AND 177487) GROUP BY occupation,sex,race,`marital.status`,income,relationship,education,`native.country`,workclass

Resulted in 0 records

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

Hellinger Distance: nan

==========

Real

| | sex | education | workclass | marital.status | race | relationship | income | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **1824** | Female | Some-college | ? | Never-married | Amer-Indian-Eskimo | Not-in-family | <=50K | 45 | 2 |
| **301** | Female | 1st-4th | ? | Widowed | Black | Not-in-family | <=50K | 90 | 1 |
| **3771** | Male | Masters | Local-gov | Divorced | Black | Not-in-family | <=50K | 122 | 3 |
| **2960** | Male | Assoc-voc | Private | Never-married | Asian-Pac-Islander | Own-child | <=50K | 153 | 6 |
| **1989** | Female | Some-college | Private | Married-spouse-absent | White | Own-child | <=50K | 104 | 4 |

Synthetic

| | sex | education | workclass | marital.status | race | relationship | income | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **1824** | Female | Some-college | ? | Never-married | Amer-Indian-Eskimo | Not-in-family | <=50K | nan | nan |
| **301** | Female | 1st-4th | ? | Widowed | Black | Not-in-family | <=50K | nan | nan |
| **3771** | Male | Masters | Local-gov | Divorced | Black | Not-in-family | <=50K | 36.000000 | 1.000000 |
| **2960** | Male | Assoc-voc | Private | Never-married | Asian-Pac-Islander | Own-child | <=50K | nan | nan |
| **1989** | Female | Some-college | Private | Married-spouse-absent | White | Own-child | <=50K | 55.000000 | 2.000000 |

SQL for Real:
SELECT sex,education,workclass,`marital.status`,race,relationship,income,SUM(age), COUNT(*) FROM C1 WHERE (capital = 0 OR occupation IN ('Other-service', 'Craft-repair', 'Tech-support', 'Farming-fishing', 'Transport-moving', 'Sales', 'Handlers-cleaners', 'Machine-op-inspct')) OR (`marital.status` = 'Never-married' OR sex <> 'Female') OR (relationship <> 'Own-child' OR race IN ('Other', 'White')) GROUP BY sex,education,workclass,`marital.status`,race,relationship,income

Resulted in 4330 records

SQL for Synthetic:
SELECT sex,education,workclass,`marital.status`,race,relationship,income,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (capital = 0 OR occupation IN ('Other-service', 'Craft-repair', 'Tech-support', 'Farming-fishing', 'Transport-moving', 'Sales', 'Handlers-cleaners', 'Machine-op-inspct')) OR (`marital.status` = 'Never-married' OR sex <> 'Female') OR (relationship <> 'Own-child' OR race IN ('Other', 'White')) GROUP BY sex,education,workclass,`marital.status`,race,relationship,income

Resulted in 3809 records

Normalized Euclidean distance for (age): 49.57

Hellinger Distance: 0.188

==========

Real

| | race | relationship | marital.status | occupation | sex | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **1062** | White | Not-in-family | Widowed | Exec-managerial | Female | 58.291667 | 48 |
| **188** | Asian-Pac-Islander | Not-in-family | Married-spouse-absent | Prof-specialty | Male | 36.200000 | 5 |
| **330** | Asian-Pac-Islander | Unmarried | Divorced | Other-service | Female | 42.333333 | 3 |
| **368** | Asian-Pac-Islander | Unmarried | Separated | Other-service | Female | 34.666667 | 3 |
| **341** | Asian-Pac-Islander | Unmarried | Married-spouse-absent | Other-service | Female | 39.000000 | 1 |

Synthetic

| | race | relationship | marital.status | occupation | sex | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **1062** | White | Not-in-family | Widowed | Exec-managerial | Female | 58.553846 | 65.000000 |
| **188** | Asian-Pac-Islander | Not-in-family | Married-spouse-absent | Prof-specialty | Male | 24.000000 | 1.000000 |
| **330** | Asian-Pac-Islander | Unmarried | Divorced | Other-service | Female | 54.500000 | 2.000000 |
| **368** | Asian-Pac-Islander | Unmarried | Separated | Other-service | Female | 41.000000 | 1.000000 |
| **341** | Asian-Pac-Islander | Unmarried | Married-spouse-absent | Other-service | Female | 33.000000 | 1.000000 |

SQL for Real:
SELECT race,relationship,`marital.status`,occupation,sex,AVG(age), COUNT(*) FROM C1 WHERE (education <> 'Some-college') GROUP BY race,relationship,`marital.status`,occupation,sex

Resulted in 1486 records

SQL for Synthetic:
SELECT race,relationship,`marital.status`,occupation,sex,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (education <> 'Some-college') GROUP BY race,relationship,`marital.status`,occupation,sex

Resulted in 1397 records

Normalized Euclidean distance for (age): 32.53

Hellinger Distance: 0.087

==========

Real

| | education | relationship | income | sex | workclass | native.country | race | marital.status | occupation | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | Some-college | Unmarried | <=50K | Female | Private | United-States | White | Married-civ-spouse | Craft-repair | nan | nan |
| **3** | HS-grad | Unmarried | <=50K | Female | ? | United-States | White | Married-civ-spouse | ? | nan | nan |
| **5** | HS-grad | Unmarried | <=50K | Female | State-gov | United-States | White | Married-civ-spouse | Other-service | nan | nan |
| **0** | 11th | Unmarried | <=50K | Male | Private | United-States | White | Married-civ-spouse | Craft-repair | nan | nan |
| **4** | HS-grad | Unmarried | <=50K | Female | Private | United-States | White | Married-civ-spouse | Machine-op-inspct | nan | nan |

Synthetic

| | education | relationship | income | sex | workclass | native.country | race | marital.status | occupation | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | Some-college | Unmarried | <=50K | Female | Private | United-States | White | Married-civ-spouse | Craft-repair | 51 | 1 |
| **3** | HS-grad | Unmarried | <=50K | Female | ? | United-States | White | Married-civ-spouse | ? | 42 | 1 |
| **5** | HS-grad | Unmarried | <=50K | Female | State-gov | United-States | White | Married-civ-spouse | Other-service | 37 | 1 |
| **0** | 11th | Unmarried | <=50K | Male | Private | United-States | White | Married-civ-spouse | Craft-repair | 27 | 1 |
| **4** | HS-grad | Unmarried | <=50K | Female | Private | United-States | White | Married-civ-spouse | Machine-op-inspct | 49 | 1 |

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

40 AND 20) AND (fnlwgt = 32280 OR occupation LIKE 'Other-service') OR (capital = 0 OR age <> 38) AND (income = '<=50K' AND relationship LIKE 'Unmarried') AND (`marital.status` = 'Married-civ-spouse' AND `native.country` = 'United-States') GROUP BY education,relationship,income,sex,workclass,`native.country`,race,`marital.status`,occupation

Resulted in 0 records

SQL for Synthetic:
SELECT education,relationship,income,sex,workclass,`native.country`,race,`marital.status`,occupation,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` BETWEEN 40 AND 20) AND (fnlwgt = 32280 OR occupation LIKE 'Other-service') OR (capital = 0 OR age <> 38) AND (income = '<=50K' AND relationship LIKE 'Unmarried') AND (`marital.status` = 'Married-civ-spouse' AND `native.country` = 'United-States') GROUP BY education,relationship,income,sex,workclass,`native.country`,race,`marital.status`,occupation

Resulted in 8 records

Normalized Euclidean distance for (age): nan

Hellinger Distance: nan

==========

Real

| | sex | native.country | MIN(fnlwgt) | COUNT(*) |
|---|---|---|---|---|
| **65** | Male | Japan | 22201 | 61 |
| **18** | Female | Hungary | 114263 | 7 |
| **46** | Male | Columbia | 107732 | 53 |
| **5** | Female | Cuba | 48853 | 50 |
| **70** | Male | Peru | 149366 | 28 |

Synthetic

| | sex | native.country | MIN(fnlwgt) | COUNT(*) |
|---|---|---|---|---|
| **65** | Male | Japan | 177902.403884 | 801.000000 |
| **18** | Female | Hungary | 179577.131609 | 307.000000 |
| **46** | Male | Columbia | 177158.044047 | 929.000000 |
| **5** | Female | Cuba | 181386.001124 | 3.000000 |
| **70** | Male | Peru | 178625.998440 | 794.000000 |

SQL for Real:
SELECT sex,`native.country`,MIN(fnlwgt), COUNT(*) FROM C1 WHERE (capital >= 0 OR workclass <> 'Private') OR (fnlwgt = 344351 OR `hours.per.week` BETWEEN 60 AND 30) OR (income = '<=50K' OR education LIKE 'Masters') OR (`marital.status` = 'Married-civ-spouse' OR occupation IN ('Other-service', 'Adm-clerical', 'Craft-repair', 'Armed-Forces', 'Tech-support', 'Priv-house-serv', 'Exec-managerial', '?', 'Prof-specialty', 'Sales', 'Protective-serv')) AND (race LIKE 'White' AND sex <> 'Female') GROUP BY sex,`native.country`

Resulted in 83 records

SQL for Synthetic:
SELECT sex,`native.country`,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (capital >= 0 OR workclass <> 'Private') OR (fnlwgt = 344351 OR `hours.per.week` BETWEEN 60 AND 30) OR (income = '<=50K' OR education LIKE 'Masters') OR (`marital.status` = 'Married-civ-spouse' OR occupation IN ('Other-service', 'Adm-clerical', 'Craft-repair', 'Armed-Forces', 'Tech-support', 'Priv-house-serv', 'Exec-managerial', '?', 'Prof-specialty', 'Sales', 'Protective-serv')) AND (race LIKE 'White' AND sex <> 'Female') GROUP BY sex,`native.country`

Resulted in 78 records

Normalized Euclidean distance for (fnlwgt): 8.77

Hellinger Distance: 0.399

==========

Real

| | sex | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|
| **0** | Female | 36.996553 | 13053 |
| **1** | Male | 43.197277 | 27104 |

Synthetic

| | sex | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|
| **0** | Female | 39.888529 | 13179.000000 |
| **1** | Male | 40.052696 | 23780.000000 |

SQL for Real:
SELECT sex,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE ((`hours.per.week` BETWEEN 60 AND 72) OR income LIKE '>50K') OR (sex = 'Male' OR `native.country` = 'United-States') AND (capital < 0 AND `marital.status` LIKE 'Never-married') AND (fnlwgt >= 34066 AND relationship LIKE 'Husband') AND (occupation LIKE 'Adm-clerical' AND education <> 'Bachelors') OR (workclass = 'Private' OR race = 'Black') GROUP BY sex

Resulted in 2 records

SQL for Synthetic:
SELECT sex,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE ((`hours.per.week` BETWEEN 60 AND 72) OR income LIKE '>50K') OR (sex = 'Male' OR `native.country` = 'United-States') AND (capital < 0 AND `marital.status` LIKE 'Never-married') AND (fnlwgt >= 34066 AND relationship LIKE 'Husband') AND (occupation LIKE 'Adm-clerical' AND education <> 'Bachelors') OR (workclass = 'Private' OR race = 'Black') GROUP BY sex

Resulted in 2 records

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

Hellinger Distance: 0.024

==========

Real

| | workclass | occupation | relationship | marital.status | native.country | education | race | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **4499** | Private | Craft-repair | Not-in-family | Never-married | Hungary | 1st-4th | White | nan | nan |
| **13502** | State-gov | Craft-repair | Husband | Married-civ-spouse | Laos | HS-grad | White | nan | nan |
| **9192** | Private | Prof-specialty | Unmarried | Divorced | Laos | Some-college | White | nan | nan |
| **2767** | Local-gov | Protective-serv | Husband | Married-civ-spouse | Hungary | Some-college | White | nan | nan |
| **5669** | Private | Exec-managerial | Unmarried | Separated | Puerto-Rico | Some-college | White | nan | nan |

Synthetic

| | workclass | occupation | relationship | marital.status | native.country | education | race | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **4499** | Private | Craft-repair | Not-in-family | Never-married | Hungary | 1st-4th | White | 40.040170 | 1 |
| **13502** | State-gov | Craft-repair | Husband | Married-civ-spouse | Laos | HS-grad | White | 40.026051 | 1 |
| **9192** | Private | Prof-specialty | Unmarried | Divorced | Laos | Some-college | White | 40.002672 | 1 |
| **2767** | Local-gov | Protective-serv | Husband | Married-civ-spouse | Hungary | Some-college | White | 39.952612 | 1 |
| **5669** | Private | Exec-managerial | Unmarried | Separated | Puerto-Rico | Some-college | White | 39.992909 | 4 |

SQL for Real:
SELECT workclass,occupation,relationship,`marital.status`,`native.country`,education,race,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (sex IN ('Male', 'Female') OR `marital.status` = 'Married-civ-spouse') GROUP BY workclass,occupation,relationship,`marital.status`,`native.country`,education,race

Resulted in 10061 records

SQL for Synthetic:
SELECT workclass,occupation,relationship,`marital.status`,`native.country`,education,race,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (sex IN ('Male', 'Female') OR `marital.status` = 'Married-civ-spouse') GROUP BY workclass,occupation,relationship,`marital.status`,`native.country`,education,race

Resulted in 14266 records

Normalized Euclidean distance for (`hours.per.week`): 60.38

Hellinger Distance: 0.229

==========

Real

| | income | education | native.country | workclass | sex | marital.status | race | occupation | relationship | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **6967** | <=50K | Some-college | China | Private | Female | Never-married | White | Machine-op-inspct | Own-child | nan | nan |
| **4270** | <=50K | HS-grad | Ecuador | Private | Male | Never-married | White | Handlers-cleaners | Own-child | nan | nan |
| **7950** | <=50K | Some-college | United-States | Local-gov | Female | Married-civ-spouse | White | Adm-clerical | Wife | 60.000000 | 9.000000 |
| **7034** | <=50K | Some-college | Columbia | Self-emp-not-inc | Female | Never-married | White | Prof-specialty | Not-in-family | nan | nan |
| **6042** | <=50K | HS-grad | United-States | Self-emp-not-inc | Male | Married-spouse-absent | White | Transport-moving | Unmarried | nan | nan |

Synthetic

| | income | education | native.country | workclass | sex | marital.status | race | occupation | relationship | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **6967** | <=50K | Some-college | China | Private | Female | Never-married | White | Machine-op-inspct | Own-child | 28 | 1 |
| **4270** | <=50K | HS-grad | Ecuador | Private | Male | Never-married | White | Handlers-cleaners | Own-child | 29 | 1 |
| **7950** | <=50K | Some-college | United-States | Local-gov | Female | Married-civ-spouse | White | Adm-clerical | Wife | 38 | 3 |
| **7034** | <=50K | Some-college | Columbia | Self-emp-not-inc | Female | Never-married | White | Prof-specialty | Not-in-family | 25 | 1 |
| **6042** | <=50K | HS-grad | United-States | Self-emp-not-inc | Male | Married-spouse-absent | White | Transport-moving | Unmarried | 60 | 1 |

SQL for Real:
SELECT income,education,`native.country`,workclass,sex,`marital.status`,race,occupation,relationship,MAX(age), COUNT(*) FROM C1 WHERE (occupation IN ('?', 'Machine-op-inspct', 'Prof-specialty', 'Adm-clerical', 'Handlers-cleaners', 'Transport-moving', 'Priv-house-serv', 'Tech-support', 'Farming-fishing')) GROUP BY income,education,`native.country`,workclass,sex,`marital.status`,race,occupation,relationship

Resulted in 6962 records

SQL for Synthetic:
SELECT income,education,`native.country`,workclass,sex,`marital.status`,race,occupation,relationship,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (occupation IN ('?', 'Machine-op-inspct', 'Prof-specialty', 'Adm-clerical', 'Handlers-cleaners', 'Transport-moving', 'Priv-house-serv', 'Tech-support', 'Farming-fishing')) GROUP BY income,education,`native.country`,workclass,sex,`marital.status`,race,occupation,relationship

Resulted in 9334 records

Normalized Euclidean distance for (age): 47.77

Hellinger Distance: 0.238

==========

Real

| | occupation | sex | education | income | native.country | relationship | workclass | race | marital.status | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **5012** | Handlers-cleaners | Male | Assoc-voc | <=50K | Laos | Not-in-family | Private | White | Divorced | nan | nan |
| **10796** | Sales | Male | Some-college | <=50K | Canada | Not-in-family | Private | White | Never-married | nan | nan |
| **3060** | Exec-managerial | Female | 9th | <=50K | Ireland | Not-in-family | Private | White | Widowed | nan | nan |
| **1018** | Adm-clerical | Female | HS-grad | <=50K | United-States | Unmarried | State-gov | White | Married-spouse-absent | nan | nan |
| **5340** | Machine-op-inspct | Female | 10th | <=50K | United-States | Unmarried | Private | White | Divorced | -1594.000000 | 3.000000 |

Synthetic

| | occupation | sex | education | income | native.country | relationship | workclass | race | marital.status | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **5012** | Handlers-cleaners | Male | Assoc-voc | <=50K | Laos | Not-in-family | Private | White | Divorced | -21.072608 | 1 |
| **10796** | Sales | Male | Some-college | <=50K | Canada | Not-in-family | Private | White | Never-married | -32.573375 | 2 |
| **3060** | Exec-managerial | Female | 9th | <=50K | Ireland | Not-in-family | Private | White | Widowed | 11.097125 | 1 |
| **1018** | Adm-clerical | Female | HS-grad | <=50K | United-States | Unmarried | State-gov | White | Married-spouse-absent | 13.761959 | 1 |
| **5340** | Machine-op-inspct | Female | 10th | <=50K | United-States | Unmarried | Private | White | Divorced | -57.501421 | 2 |

SQL for Real:
SELECT occupation,sex,education,income,`native.country`,relationship,workclass,race,`marital.status`,MIN(capital), COUNT(*) FROM C1 WHERE (income = '<=50K') AND (sex <> 'Male' AND occupation IN ('Armed-Forces', 'Sales', '?')) AND (race <> 'Black' AND education IN ('9th', '12th', '7th-8th', 'Bachelors', 'Prof-school')) OR ((`hours.per.week` BETWEEN 48 AND 40) OR workclass LIKE 'Private') OR (capital >= 0 OR relationship IN ('Own-child', 'Unmarried')) AND (age < 54 AND `native.country` = 'United-States') GROUP BY occupation,sex,education,income,`native.country`,relationship,workclass,race,`marital.status`

Resulted in 10783 records

SQL for Synthetic:
SELECT occupation,sex,education,income,`native.country`,relationship,workclass,race,`marital.status`,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (income = '<=50K') AND (sex <> 'Male' AND occupation IN ('Armed-Forces', 'Sales', '?')) AND (race <> 'Black' AND education IN ('9th', '12th', '7th-8th', 'Bachelors', 'Prof-school')) OR ((`hours.per.week` BETWEEN 48 AND 40) OR workclass LIKE 'Private') OR (capital >= 0 OR relationship IN ('Own-child', 'Unmarried')) AND (age < 54 AND `native.country` = 'United-States') GROUP BY occupation,sex,education,income,`native.country`,relationship,workclass,race,`marital.status`

Resulted in 12065 records

Normalized Euclidean distance for (capital): 58.86

Hellinger Distance: 0.266

==========

Real

| | marital.status | sex | relationship | race | occupation | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **1500** | Widowed | Female | Unmarried | Black | Sales | 223 | 4 |
| **384** | Married-civ-spouse | Female | Wife | Amer-Indian-Eskimo | Farming-fishing | 29 | 1 |
| **399** | Married-civ-spouse | Female | Wife | Asian-Pac-Islander | Transport-moving | 63 | 1 |
| **792** | Never-married | Female | Not-in-family | White | ? | 2858 | 102 |
| **705** | Married-spouse-absent | Male | Other-relative | White | Farming-fishing | 178 | 5 |

Synthetic

| | marital.status | sex | relationship | race | occupation | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **1500** | Widowed | Female | Unmarried | Black | Sales | 70.000000 | 1.000000 |
| **384** | Married-civ-spouse | Female | Wife | Amer-Indian-Eskimo | Farming-fishing | 36.000000 | 1.000000 |
| **399** | Married-civ-spouse | Female | Wife | Asian-Pac-Islander | Transport-moving | nan | nan |
| **792** | Never-married | Female | Not-in-family | White | ? | 3788.000000 | 113.000000 |
| **705** | Married-spouse-absent | Male | Other-relative | White | Farming-fishing | 106.000000 | 2.000000 |

SQL for Real:
SELECT `marital.status`,sex,relationship,race,occupation,SUM(age), COUNT(*) FROM C1 WHERE (workclass IN ('Self-emp-inc', 'Private', '?', 'State-gov', 'Self-emp-not-inc', 'Federal-gov', 'Local-gov')) OR (`native.country` = 'United-States' AND `hours.per.week` BETWEEN 55 AND 36) AND (age > 45 AND sex LIKE 'Male') AND (race <> 'White' AND income <> '<=50K') AND (occupation = 'Sales' AND relationship <> 'Own-child') AND (`marital.status` LIKE 'Married-spouse-absent' AND capital BETWEEN 0 AND 0) GROUP BY `marital.status`,sex,relationship,race,occupation

Resulted in 1585 records

SQL for Synthetic:
SELECT `marital.status`,sex,relationship,race,occupation,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (workclass IN ('Self-emp-inc', 'Private', '?', 'State-gov', 'Self-emp-not-inc', 'Federal-gov', 'Local-gov')) OR (`native.country` = 'United-States' AND `hours.per.week` BETWEEN 55 AND 36) AND (age > 45 AND sex LIKE 'Male') AND (race <> 'White' AND income <> '<=50K') AND (occupation = 'Sales' AND relationship <> 'Own-child') AND (`marital.status` LIKE 'Married-spouse-absent' AND capital BETWEEN 0 AND 0) GROUP BY `marital.status`,sex,relationship,race,occupation

Resulted in 1503 records

Normalized Euclidean distance for (age): 33.82

Hellinger Distance: 0.082

==========

Real

| | workclass | SUM(age) | COUNT(*) |
|---|---|---|---|
| **5** | Self-emp-not-inc | 133798 | 2912 |
| **2** | Local-gov | 95620 | 2285 |
| **1** | Federal-gov | 44527 | 1038 |
| **0** | ? | 79356 | 2016 |
| **4** | Self-emp-inc | 63233 | 1363 |

Synthetic

| | workclass | SUM(age) | COUNT(*) |
|---|---|---|---|
| **5** | Self-emp-not-inc | 116489.000000 | 2521.000000 |
| **2** | Local-gov | 88785.000000 | 2123.000000 |
| **1** | Federal-gov | 40574.000000 | 936.000000 |
| **0** | ? | 68289.000000 | 1813.000000 |
| **4** | Self-emp-inc | 53760.000000 | 1181.000000 |

SQL for Real:
SELECT workclass,SUM(age), COUNT(*) FROM C1 WHERE (race <> 'White') OR (relationship <> 'Not-in-family' AND education <> 'HS-grad') OR (fnlwgt > 337064 OR `native.country` = 'United-States') AND (sex = 'Male' AND age >= 24) AND (`marital.status` = 'Married-civ-spouse' AND workclass <> 'Private') GROUP BY workclass

Resulted in 7 records

SQL for Synthetic:
SELECT workclass,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (race <> 'White') OR (relationship <> 'Not-in-family' AND education <> 'HS-grad') OR (fnlwgt > 337064 OR `native.country` = 'United-States') AND (sex = 'Male' AND age >= 24) AND (`marital.status` = 'Married-civ-spouse' AND workclass <> 'Private') GROUP BY workclass

Resulted in 7 records

Normalized Euclidean distance for (age): 2.65

Hellinger Distance: 0.013

==========

Real

| | relationship | education | native.country | race | sex | workclass | occupation | marital.status | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **6127** | Not-in-family | HS-grad | Puerto-Rico | White | Male | State-gov | Other-service | Never-married | nan | nan |
| **11330** | Unmarried | HS-grad | Japan | White | Female | State-gov | Adm-clerical | Divorced | nan | nan |
| **7034** | Not-in-family | Prof-school | India | Asian-Pac-Islander | Male | Private | Prof-specialty | Never-married | 236481.000000 | 2.000000 |
| **3892** | Not-in-family | Assoc-acdm | Peru | Other | Female | Local-gov | Adm-clerical | Divorced | nan | nan |
| **4723** | Not-in-family | Bachelors | Peru | Asian-Pac-Islander | Female | State-gov | Adm-clerical | Widowed | nan | nan |

Synthetic

| | relationship | education | native.country | race | sex | workclass | occupation | marital.status | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **6127** | Not-in-family | HS-grad | Puerto-Rico | White | Male | State-gov | Other-service | Never-married | 169934.478193 | 1 |
| **11330** | Unmarried | HS-grad | Japan | White | Female | State-gov | Adm-clerical | Divorced | 178281.490826 | 1 |
| **7034** | Not-in-family | Prof-school | India | Asian-Pac-Islander | Male | Private | Prof-specialty | Never-married | 181019.620018 | 1 |
| **3892** | Not-in-family | Assoc-acdm | Peru | Other | Female | Local-gov | Adm-clerical | Divorced | 178630.487495 | 1 |
| **4723** | Not-in-family | Bachelors | Peru | Asian-Pac-Islander | Female | State-gov | Adm-clerical | Widowed | 179378.354720 | 1 |

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

Resulted in 9335 records

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

Resulted in 12998 records

Normalized Euclidean distance for (fnlwgt): 56.56

Hellinger Distance: 0.228

==========

Real

| | marital.status | education | sex | income | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|
| **166** | Never-married | 10th | Female | <=50K | 634226 | 184 |
| **189** | Never-married | Assoc-voc | Female | <=50K | 557805 | 233 |
| **8** | Divorced | 12th | Male | <=50K | 849067 | 30 |
| **100** | Married-civ-spouse | Assoc-voc | Male | >50K | 673764 | 362 |
| **33** | Divorced | Doctorate | Female | >50K | 295566 | 15 |

Synthetic

| | marital.status | education | sex | income | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|
| **166** | Never-married | 10th | Female | <=50K | 199960.552073 | 146.000000 |
| **189** | Never-married | Assoc-voc | Female | <=50K | 199578.642449 | 175.000000 |
| **8** | Divorced | 12th | Male | <=50K | 199821.814539 | 13.000000 |
| **100** | Married-civ-spouse | Assoc-voc | Male | >50K | 196489.936888 | 106.000000 |
| **33** | Divorced | Doctorate | Female | >50K | 203890.233745 | 4.000000 |

SQL for Real:
SELECT `marital.status`,education,sex,income,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (`native.country` IN ('El-Salvador', 'Italy', 'Taiwan', 'United-States', 'Ireland', 'India', 'Trinadad&Tobago', 'Nicaragua')) GROUP BY `marital.status`,education,sex,income

Resulted in 320 records

SQL for Synthetic:
SELECT `marital.status`,education,sex,income,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`native.country` IN ('El-Salvador', 'Italy', 'Taiwan', 'United-States', 'Ireland', 'India', 'Trinadad&Tobago', 'Nicaragua')) GROUP BY `marital.status`,education,sex,income

Resulted in 272 records

Normalized Euclidean distance for (fnlwgt): 16.22

Hellinger Distance: 0.157

==========

Real

| | marital.status | sex | native.country | occupation | workclass | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **88** | Married-spouse-absent | Male | United-States | Farming-fishing | Local-gov | 43.000000 | 2.000000 |
| **8** | Divorced | Male | ? | Farming-fishing | Self-emp-not-inc | nan | nan |
| **133** | Never-married | Male | Puerto-Rico | Farming-fishing | State-gov | nan | nan |
| **62** | Married-civ-spouse | Male | Mexico | Farming-fishing | State-gov | nan | nan |
| **154** | Widowed | Female | Mexico | Farming-fishing | Local-gov | nan | nan |

Synthetic

| | marital.status | sex | native.country | occupation | workclass | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **88** | Married-spouse-absent | Male | United-States | Farming-fishing | Local-gov | 34 | 1 |
| **8** | Divorced | Male | ? | Farming-fishing | Self-emp-not-inc | 49 | 4 |
| **133** | Never-married | Male | Puerto-Rico | Farming-fishing | State-gov | 51 | 2 |
| **62** | Married-civ-spouse | Male | Mexico | Farming-fishing | State-gov | 64 | 1 |
| **154** | Widowed | Female | Mexico | Farming-fishing | Local-gov | 68 | 1 |

SQL for Real:
SELECT `marital.status`,sex,`native.country`,occupation,workclass,MAX(age), COUNT(*) FROM C1 WHERE (occupation LIKE 'Farming-fishing') GROUP BY `marital.status`,sex,`native.country`,occupation,workclass

Resulted in 107 records

SQL for Synthetic:
SELECT `marital.status`,sex,`native.country`,occupation,workclass,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (occupation LIKE 'Farming-fishing') GROUP BY `marital.status`,sex,`native.country`,occupation,workclass

Resulted in 167 records

Normalized Euclidean distance for (age): 7.21

Hellinger Distance: 0.182

==========

Real

| | sex | workclass | SUM(capital) | COUNT(*) |
|---|---|---|---|---|
| **8** | Male | Federal-gov | 767742 | 661 |
| **7** | Male | ? | 641229 | 725 |
| **5** | Female | Self-emp-not-inc | 385562 | 195 |
| **11** | Male | Self-emp-inc | 7089058 | 1192 |
| **3** | Female | Private | 2259788 | 1595 |

Synthetic

| | sex | workclass | SUM(capital) | COUNT(*) |
|---|---|---|---|---|
| **8** | Male | Federal-gov | 3807.038206 | 698.000000 |
| **7** | Male | ? | 1146.063085 | 698.000000 |
| **5** | Female | Self-emp-not-inc | 2833.023414 | 204.000000 |
| **11** | Male | Self-emp-inc | 12627.611293 | 1083.000000 |
| **3** | Female | Private | -1582.709730 | 1693.000000 |

SQL for Real:
SELECT sex,workclass,SUM(capital), COUNT(*) FROM C1 WHERE (`marital.status` LIKE 'Married-civ-spouse') GROUP BY sex,workclass

Resulted in 14 records

SQL for Synthetic:
SELECT sex,workclass,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` LIKE 'Married-civ-spouse') GROUP BY sex,workclass

Resulted in 14 records

Normalized Euclidean distance for (capital): 3.74

Hellinger Distance: 0.016

==========

Real

| | race | marital.status | income | relationship | education | sex | occupation | native.country | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **8933** | White | Never-married | <=50K | Not-in-family | HS-grad | Female | Adm-clerical | Taiwan | nan | nan |
| **3912** | White | Divorced | <=50K | Not-in-family | Prof-school | Male | Prof-specialty | Puerto-Rico | nan | nan |
| **4522** | White | Divorced | <=50K | Unmarried | 9th | Female | Sales | Jamaica | nan | nan |
| **6766** | White | Married-civ-spouse | <=50K | Wife | 12th | Female | Craft-repair | United-States | 0.000000 | 1.000000 |
| **556** | Asian-Pac-Islander | Married-civ-spouse | >50K | Husband | Bachelors | Male | Machine-op-inspct | Puerto-Rico | nan | nan |

Synthetic

| | race | marital.status | income | relationship | education | sex | occupation | native.country | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **8933** | White | Never-married | <=50K | Not-in-family | HS-grad | Female | Adm-clerical | Taiwan | 116.399689 | 1 |
| **3912** | White | Divorced | <=50K | Not-in-family | Prof-school | Male | Prof-specialty | Puerto-Rico | 18.614639 | 2 |
| **4522** | White | Divorced | <=50K | Unmarried | 9th | Female | Sales | Jamaica | -35.617099 | 1 |
| **6766** | White | Married-civ-spouse | <=50K | Wife | 12th | Female | Craft-repair | United-States | 137.294560 | 1 |
| **556** | Asian-Pac-Islander | Married-civ-spouse | >50K | Husband | Bachelors | Male | Machine-op-inspct | Puerto-Rico | -107.069280 | 1 |

SQL for Real:
SELECT race,`marital.status`,income,relationship,education,sex,occupation,`native.country`,MIN(capital), COUNT(*) FROM C1 WHERE (race <> 'Other') OR (capital <> 0 OR `marital.status` IN ('Separated', 'Never-married', 'Married-AF-spouse', 'Married-civ-spouse')) AND ((age BETWEEN 45 AND 34) AND `native.country` = 'United-States') GROUP BY race,`marital.status`,income,relationship,education,sex,occupation,`native.country`

Resulted in 9029 records

SQL for Synthetic:
SELECT race,`marital.status`,income,relationship,education,sex,occupation,`native.country`,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (race <> 'Other') OR (capital <> 0 OR `marital.status` IN ('Separated', 'Never-married', 'Married-AF-spouse', 'Married-civ-spouse')) AND ((age BETWEEN 45 AND 34) AND `native.country` = 'United-States') GROUP BY race,`marital.status`,income,relationship,education,sex,occupation,`native.country`

Resulted in 12738 records

Normalized Euclidean distance for (capital): 57.09

Hellinger Distance: 0.264

==========

Real

| | workclass | native.country | sex | MIN(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|
| **14** | Federal-gov | Poland | Male | 190160 | 1 |
| **60** | Self-emp-not-inc | Iran | Male | 127295 | 1 |
| **43** | Private | Philippines | Male | 137192 | 1 |
| **29** | Private | England | Female | 77271 | 2 |
| **3** | ? | Cuba | Female | 178013 | 1 |

Synthetic

| | workclass | native.country | sex | MIN(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|
| **14** | Federal-gov | Poland | Male | nan | nan |
| **60** | Self-emp-not-inc | Iran | Male | nan | nan |
| **43** | Private | Philippines | Male | nan | nan |
| **29** | Private | England | Female | nan | nan |
| **3** | ? | Cuba | Female | nan | nan |

SQL for Real:
SELECT workclass,`native.country`,sex,MIN(fnlwgt), COUNT(*) FROM C1 WHERE (sex IN ('Male', 'Female')) AND (race IN ('White', 'Other', 'Amer-Indian-Eskimo') OR education <> 'Some-college') AND (age > 26 AND `hours.per.week` = 20) GROUP BY workclass,`native.country`,sex

Resulted in 75 records

SQL for Synthetic:
SELECT workclass,`native.country`,sex,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (sex IN ('Male', 'Female')) AND (race IN ('White', 'Other', 'Amer-Indian-Eskimo') OR education <> 'Some-college') AND (age > 26 AND `hours.per.week` = 20) GROUP BY workclass,`native.country`,sex

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

==========

Real

| | education | income | native.country | occupation | relationship | race | sex | SUM(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **734** | 11th | <=50K | United-States | ? | Own-child | White | Male | -1602.000000 | 34.000000 |
| **2776** | Assoc-voc | <=50K | Ecuador | Tech-support | Husband | White | Male | nan | nan |
| **1078** | 12th | <=50K | United-States | ? | Own-child | Black | Male | 0.000000 | 5.000000 |
| **8788** | Some-college | <=50K | United-States | Craft-repair | Husband | White | Male | 55998.000000 | 398.000000 |
| **7482** | Prof-school | <=50K | Peru | ? | Husband | White | Male | nan | nan |

Synthetic

| | education | income | native.country | occupation | relationship | race | sex | SUM(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **734** | 11th | <=50K | United-States | ? | Own-child | White | Male | 275.081699 | 11 |
| **2776** | Assoc-voc | <=50K | Ecuador | Tech-support | Husband | White | Male | 34.433163 | 2 |
| **1078** | 12th | <=50K | United-States | ? | Own-child | Black | Male | 83.159748 | 1 |
| **8788** | Some-college | <=50K | United-States | Craft-repair | Husband | White | Male | 6883.559039 | 319 |
| **7482** | Prof-school | <=50K | Peru | ? | Husband | White | Male | 153.534672 | 1 |

SQL for Real:
SELECT education,income,`native.country`,occupation,relationship,race,sex,SUM(capital), COUNT(*) FROM C1 WHERE (relationship = 'Husband') OR (`native.country` LIKE 'United-States' OR `hours.per.week` BETWEEN 40 AND 45) OR (capital > 0 OR race IN ('Black', 'Other', 'Asian-Pac-Islander', 'Amer-Indian-Eskimo')) GROUP BY education,income,`native.country`,occupation,relationship,race,sex

Resulted in 6482 records

SQL for Synthetic:
SELECT education,income,`native.country`,occupation,relationship,race,sex,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (relationship = 'Husband') OR (`native.country` LIKE 'United-States' OR `hours.per.week` BETWEEN 40 AND 45) OR (capital > 0 OR race IN ('Black', 'Other', 'Asian-Pac-Islander', 'Amer-Indian-Eskimo')) GROUP BY education,income,`native.country`,occupation,relationship,race,sex

Resulted in 9249 records

Normalized Euclidean distance for (capital): 50.21

Hellinger Distance: 0.263

==========

Real

| native.country | marital.status | race | occupation | income | workclass | relationship | education | sex | AVG(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|

Synthetic

| native.country | marital.status | race | occupation | income | workclass | relationship | education | sex | AVG(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|

SQL for Real:
SELECT `native.country`,`marital.status`,race,occupation,income,workclass,relationship,education,sex,AVG(capital), COUNT(*) FROM C1 WHERE (`marital.status` LIKE 'Married-civ-spouse' OR sex <> 'Male') AND (race LIKE 'White' AND fnlwgt BETWEEN 136331 AND 48093) AND (`hours.per.week` <= 40 AND capital BETWEEN 0 AND 0) GROUP BY `native.country`,`marital.status`,race,occupation,income,workclass,relationship,education,sex

Resulted in 0 records

SQL for Synthetic:
SELECT `native.country`,`marital.status`,race,occupation,income,workclass,relationship,education,sex,AVG(capital), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` LIKE 'Married-civ-spouse' OR sex <> 'Male') AND (race LIKE 'White' AND fnlwgt BETWEEN 136331 AND 48093) AND (`hours.per.week` <= 40 AND capital BETWEEN 0 AND 0) GROUP BY `native.country`,`marital.status`,race,occupation,income,workclass,relationship,education,sex

Resulted in 0 records

Normalized Euclidean distance for (capital): nan

Hellinger Distance: nan

==========

Real

| | relationship | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|
| **0** | Other-relative | 1 | 689 |

Synthetic

| | relationship | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|
| **0** | Other-relative | 38.918517 | 682.000000 |

SQL for Real:
SELECT relationship,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (sex LIKE 'Female' AND relationship = 'Other-relative') GROUP BY relationship

Resulted in 1 records

SQL for Synthetic:
SELECT relationship,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (sex LIKE 'Female' AND relationship = 'Other-relative') GROUP BY relationship

Resulted in 1 records

Normalized Euclidean distance for (`hours.per.week`): 0.0

Hellinger Distance: 0.0

==========

Real

| | sex | race | marital.status | workclass | native.country | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **182** | Female | White | Divorced | Self-emp-not-inc | China | nan | nan |
| **995** | Male | White | Never-married | Local-gov | Guatemala | 1024369.000000 | 3.000000 |
| **473** | Female | White | Never-married | State-gov | Ireland | nan | nan |
| **584** | Female | White | Widowed | Self-emp-inc | Puerto-Rico | nan | nan |
| **950** | Male | White | Married-spouse-absent | Private | Scotland | nan | nan |

Synthetic

| | sex | race | marital.status | workclass | native.country | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **182** | Female | White | Divorced | Self-emp-not-inc | China | 178622.729328 | 1 |
| **995** | Male | White | Never-married | Local-gov | Guatemala | 187784.919759 | 1 |
| **473** | Female | White | Never-married | State-gov | Ireland | 356246.130158 | 2 |
| **584** | Female | White | Widowed | Self-emp-inc | Puerto-Rico | 1368991.688997 | 8 |
| **950** | Male | White | Married-spouse-absent | Private | Scotland | 553294.103922 | 3 |

SQL for Real:
SELECT sex,race,`marital.status`,workclass,`native.country`,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (race LIKE 'White' OR `hours.per.week` BETWEEN 40 AND 34) OR (income LIKE '>50K' OR `native.country` IN ('Trinadad&Tobago', 'Vietnam', 'Dominican-Republic', 'Cuba', 'Hong', 'Ireland', 'El-Salvador', 'Outlying-US(Guam-USVI-etc)', 'Germany', 'Guatemala', 'Mexico')) OR ((fnlwgt BETWEEN 146325 AND 191202) OR relationship LIKE 'Own-child') AND (education = 'Bachelors' AND workclass IN ('Private', '?', 'Federal-gov', 'Self-emp-not-inc')) AND (age <= 28 AND occupation IN ('Other-service', 'Exec-managerial', 'Handlers-cleaners')) GROUP BY sex,race,`marital.status`,workclass,`native.country`

Resulted in 1067 records

SQL for Synthetic:
SELECT sex,race,`marital.status`,workclass,`native.country`,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (race LIKE 'White' OR `hours.per.week` BETWEEN 40 AND 34) OR (income LIKE '>50K' OR `native.country` IN ('Trinadad&Tobago', 'Vietnam', 'Dominican-Republic', 'Cuba', 'Hong', 'Ireland', 'El-Salvador', 'Outlying-US(Guam-USVI-etc)', 'Germany', 'Guatemala', 'Mexico')) OR ((fnlwgt BETWEEN 146325 AND 191202) OR relationship LIKE 'Own-child') AND (education = 'Bachelors' AND workclass IN ('Private', '?', 'Federal-gov', 'Self-emp-not-inc')) AND (age <= 28 AND occupation IN ('Other-service', 'Exec-managerial', 'Handlers-cleaners')) GROUP BY sex,race,`marital.status`,workclass,`native.country`

Resulted in 1157 records

Normalized Euclidean distance for (fnlwgt): 21.66

Hellinger Distance: 0.383

==========

Real

| | education | marital.status | workclass | occupation | sex | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **2963** | Some-college | Married-civ-spouse | Self-emp-inc | Other-service | Male | nan | nan |
| **515** | 12th | Widowed | Private | Machine-op-inspct | Female | 40.000000 | 1.000000 |
| **1121** | Assoc-acdm | Never-married | Local-gov | Other-service | Male | nan | nan |
| **1539** | Bachelors | Married-civ-spouse | Federal-gov | Other-service | Male | nan | nan |
| **1213** | Assoc-voc | Divorced | Local-gov | Craft-repair | Male | 40.000000 | 1.000000 |

Synthetic

| | education | marital.status | workclass | occupation | sex | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **2963** | Some-college | Married-civ-spouse | Self-emp-inc | Other-service | Male | 160.860361 | 4 |
| **515** | 12th | Widowed | Private | Machine-op-inspct | Female | 39.994496 | 1 |
| **1121** | Assoc-acdm | Never-married | Local-gov | Other-service | Male | 40.075299 | 1 |
| **1539** | Bachelors | Married-civ-spouse | Federal-gov | Other-service | Male | 80.036479 | 2 |
| **1213** | Assoc-voc | Divorced | Local-gov | Craft-repair | Male | 40.004403 | 1 |

SQL for Real:
SELECT education,`marital.status`,workclass,occupation,sex,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (age BETWEEN 46 AND 47) AND (income IN ('>50K', '<=50K') AND capital BETWEEN 0 AND 0) AND (race <> 'White' AND fnlwgt >= 98515) OR (`native.country` <> 'United-States' OR sex = 'Male') GROUP BY education,`marital.status`,workclass,occupation,sex

Resulted in 2672 records

SQL for Synthetic:
SELECT education,`marital.status`,workclass,occupation,sex,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (age BETWEEN 46 AND 47) AND (income IN ('>50K', '<=50K') AND capital BETWEEN 0 AND 0) AND (race <> 'White' AND fnlwgt >= 98515) OR (`native.country` <> 'United-States' OR sex = 'Male') GROUP BY education,`marital.status`,workclass,occupation,sex

Resulted in 3260 records

Normalized Euclidean distance for (`hours.per.week`): 43.95

Hellinger Distance: 0.189

==========

Real

| | workclass | education | native.country | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|
| **576** | Private | 5th-6th | Ireland | nan | nan |
| **948** | Self-emp-inc | Masters | Japan | 43.000000 | 2.000000 |
| **637** | Private | Assoc-acdm | Ecuador | nan | nan |
| **278** | Local-gov | 10th | Poland | nan | nan |
| **772** | Private | HS-grad | Philippines | 73.000000 | 44.000000 |

Synthetic

| | workclass | education | native.country | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|
| **576** | Private | 5th-6th | Ireland | 26 | 1 |
| **948** | Self-emp-inc | Masters | Japan | 49 | 1 |
| **637** | Private | Assoc-acdm | Ecuador | 29 | 5 |
| **278** | Local-gov | 10th | Poland | 17 | 5 |
| **772** | Private | HS-grad | Philippines | 47 | 19 |

SQL for Real:
SELECT workclass,education,`native.country`,MAX(age), COUNT(*) FROM C1 WHERE (`native.country` LIKE 'United-States') OR (income <> '<=50K' OR race LIKE 'White') OR (fnlwgt = 35626 OR sex IN ('Male', 'Female')) OR (`hours.per.week` <= 6 OR age = 28) AND (workclass <> 'Self-emp-not-inc' AND capital BETWEEN 0 AND 0) AND (occupation <> '?' AND `marital.status` IN ('Separated', 'Never-married', 'Married-spouse-absent', 'Married-AF-spouse')) GROUP BY workclass,education,`native.country`

Resulted in 1170 records

SQL for Synthetic:
SELECT workclass,education,`native.country`,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (`native.country` LIKE 'United-States') OR (income <> '<=50K' OR race LIKE 'White') OR (fnlwgt = 35626 OR sex IN ('Male', 'Female')) OR (`hours.per.week` <= 6 OR age = 28) AND (workclass <> 'Self-emp-not-inc' AND capital BETWEEN 0 AND 0) AND (occupation <> '?' AND `marital.status` IN ('Separated', 'Never-married', 'Married-spouse-absent', 'Married-AF-spouse')) GROUP BY

workclass,education,`native.country`

Resulted in 1259 records

Normalized Euclidean distance for (age): 24.66

Hellinger Distance: 0.368

==========

Real

| | marital.status | education | workclass | relationship | income | sex | native.country | occupation | SUM(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **2459** | Married-civ-spouse | Assoc-acdm | Self-emp-not-inc | Husband | >50K | Male | United-States | Craft-repair | -1887.000000 | 3.000000 |
| **6666** | Never-married | HS-grad | Private | Not-in-family | <=50K | Male | Puerto-Rico | Craft-repair | nan | nan |
| **1556** | Divorced | Some-college | Local-gov | Not-in-family | <=50K | Female | Scotland | Adm-clerical | nan | nan |
| **1479** | Divorced | Preschool | Private | Not-in-family | <=50K | Male | United-States | Farming-fishing | nan | nan |
| **5031** | Never-married | 11th | Private | Own-child | <=50K | Male | Puerto-Rico | Farming-fishing | nan | nan |

Synthetic

| | marital.status | education | workclass | relationship | income | sex | native.country | occupation | SUM(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **2459** | Married-civ-spouse | Assoc-acdm | Self-emp-not-inc | Husband | >50K | Male | United-States | Craft-repair | -82.967266 | 3 |
| **6666** | Never-married | HS-grad | Private | Not-in-family | <=50K | Male | Puerto-Rico | Craft-repair | -709.013774 | 11 |
| **1556** | Divorced | Some-college | Local-gov | Not-in-family | <=50K | Female | Scotland | Adm-clerical | -136.674603 | 1 |
| **1479** | Divorced | Preschool | Private | Not-in-family | <=50K | Male | United-States | Farming-fishing | 128.790123 | 1 |
| **5031** | Never-married | 11th | Private | Own-child | <=50K | Male | Puerto-Rico | Farming-fishing | -264.152052 | 3 |

SQL for Real:
SELECT `marital.status`,education,workclass,relationship,income,sex,`native.country`,occupation,SUM(capital), COUNT(*) FROM C1 WHERE (`native.country` IN ('El-Salvador', 'Iran', 'Peru', 'United-States', 'Taiwan', 'Thailand', 'Jamaica', 'Yugoslavia', 'China', 'France', '?', 'Guatemala', 'Honduras', 'Scotland') OR relationship LIKE 'Own-child') OR (workclass LIKE 'Private' OR `hours.per.week` > 52) AND (education LIKE 'HS-grad' AND capital < 0) GROUP BY `marital.status`,education,workclass,relationship,income,sex,`native.country`,occupation

Resulted in 8036 records

SQL for Synthetic:
SELECT `marital.status`,education,workclass,relationship,income,sex,`native.country`,occupation,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (`native.country` IN ('El-Salvador', 'Iran', 'Peru', 'United-States', 'Taiwan', 'Thailand', 'Jamaica', 'Yugoslavia', 'China', 'France', '?', 'Guatemala', 'Honduras', 'Scotland') OR relationship LIKE 'Own-child') OR (workclass LIKE 'Private' OR `hours.per.week` > 52) AND (education LIKE 'HS-grad' AND capital < 0) GROUP BY `marital.status`,education,workclass,relationship,income,sex,`native.country`,occupation

Resulted in 9408 records

Normalized Euclidean distance for (capital): 59.97

Hellinger Distance: 0.207

==========

Real

| | occupation | sex | native.country | income | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|
| **690** | Tech-support | Female | Puerto-Rico | <=50K | nan | nan |
| **598** | Sales | Female | Canada | <=50K | 37.000000 | 2.000000 |
| **756** | Transport-moving | Male | Japan | <=50K | nan | nan |
| **461** | Priv-house-serv | Female | Hungary | <=50K | 26.000000 | 1.000000 |
| **524** | Prof-specialty | Male | Cambodia | <=50K | 36.000000 | 1.000000 |

Synthetic

| | occupation | sex | native.country | income | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|
| **690** | Tech-support | Female | Puerto-Rico | <=50K | 43 | 32 |
| **598** | Sales | Female | Canada | <=50K | 23 | 4 |
| **756** | Transport-moving | Male | Japan | <=50K | 42 | 22 |
| **461** | Priv-house-serv | Female | Hungary | <=50K | 28 | 1 |
| **524** | Prof-specialty | Male | Cambodia | <=50K | 37 | 1 |

SQL for Real:
SELECT occupation,sex,`native.country`,income,MAX(age), COUNT(*) FROM C1 WHERE (age <= 43 AND workclass LIKE 'Private') AND (income IN ('>50K', '<=50K') AND sex IN ('Male', 'Female')) GROUP BY occupation,sex,`native.country`,income

Resulted in 747 records

SQL for Synthetic:
SELECT occupation,sex,`native.country`,income,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (age <= 43 AND workclass LIKE 'Private') AND (income IN ('>50K', '<=50K') AND sex IN ('Male', 'Female')) GROUP BY occupation,sex,`native.country`,income

Resulted in 773 records

Normalized Euclidean distance for (age): 20.49

Hellinger Distance: 0.3

==========

Real

| | income | education | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|
| **14** | <=50K | Prof-school | 197769.756757 | 148 |
| **18** | >50K | 12th | 210650.977273 | 44 |
| **9** | <=50K | Bachelors | 190191.201089 | 2755 |
| **27** | >50K | HS-grad | 183933.766682 | 2173 |
| **16** | >50K | 10th | 197666.855422 | 83 |

Synthetic

| | income | education | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|
| **14** | <=50K | Prof-school | 173767.954946 | 460.000000 |
| **18** | >50K | 12th | 177037.539026 | 17.000000 |
| **9** | <=50K | Bachelors | 181270.549337 | 4312.000000 |
| **27** | >50K | HS-grad | 178210.560484 | 858.000000 |
| **16** | >50K | 10th | 173937.873068 | 29.000000 |

SQL for Real:
SELECT income,education,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (relationship IN ('Own-child', 'Other-relative', 'Not-in-family', 'Husband', 'Unmarried') AND sex = 'Male') GROUP BY income,education

Resulted in 32 records

SQL for Synthetic:
SELECT income,education,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (relationship IN ('Own-child', 'Other-relative', 'Not-in-family', 'Husband', 'Unmarried') AND sex = 'Male') GROUP BY income,education

Resulted in 31 records

Normalized Euclidean distance for (fnlwgt): 5.57

Hellinger Distance: 0.171

==========

Real

| | marital.status | relationship | occupation | education | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|---|
| **1667** | Widowed | Own-child | Craft-repair | HS-grad | 0.000000 | 2.000000 |
| **441** | Married-civ-spouse | Husband | Other-service | 7th-8th | 0.000000 | 1.000000 |
| **18** | Divorced | Not-in-family | Craft-repair | 11th | -1876.000000 | 1.000000 |
| **701** | Married-civ-spouse | Wife | Tech-support | 7th-8th | nan | nan |
| **299** | Divorced | Unmarried | Priv-house-serv | 9th | nan | nan |

Synthetic

| | marital.status | relationship | occupation | education | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|---|
| **1667** | Widowed | Own-child | Craft-repair | HS-grad | 142.123879 | 1 |
| **441** | Married-civ-spouse | Husband | Other-service | 7th-8th | -10.611902 | 8 |
| **18** | Divorced | Not-in-family | Craft-repair | 11th | -22.105065 | 7 |
| **701** | Married-civ-spouse | Wife | Tech-support | 7th-8th | -46.897516 | 1 |
| **299** | Divorced | Unmarried | Priv-house-serv | 9th | 141.085768 | 2 |

SQL for Real:
SELECT `marital.status`,relationship,occupation,education,MAX(capital), COUNT(*) FROM C1 WHERE (`marital.status` = 'Widowed') OR (workclass LIKE 'Private' AND race LIKE 'White') AND (capital < 0 AND income = '<=50K') OR (`hours.per.week` < 12 OR occupation IN ('Priv-house-serv', 'Prof-specialty', 'Protective-serv', 'Handlers-cleaners')) GROUP BY `marital.status`,relationship,occupation,education

Resulted in 1191 records

SQL for Synthetic:
SELECT `marital.status`,relationship,occupation,education,MAX(capital), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` = 'Widowed') OR (workclass LIKE 'Private' AND race LIKE 'White') AND (capital < 0 AND income = '<=50K') OR (`hours.per.week` < 12 OR occupation IN ('Priv-house-serv', 'Prof-specialty', 'Protective-serv', 'Handlers-cleaners')) GROUP BY `marital.status`,relationship,occupation,education

Resulted in 1796 records

Normalized Euclidean distance for (capital): 29.97

Hellinger Distance: 0.331

==========

Real

| | workclass | MAX(age) | COUNT(*) |
|---|---|---|---|
| **1** | Federal-gov | 90 | 1432 |
| **5** | Self-emp-not-inc | 90 | 3862 |
| **4** | Self-emp-inc | 85 | 1695 |
| **0** | ? | 90 | 2830 |
| **3** | Private | 90 | 33906 |

Synthetic

| | workclass | MAX(age) | COUNT(*) |
|---|---|---|---|
| **1** | Federal-gov | 79.000000 | 1426.000000 |
| **5** | Self-emp-not-inc | 90.000000 | 3818.000000 |
| **4** | Self-emp-inc | 84.000000 | 1645.000000 |
| **0** | ? | 90.000000 | 2844.000000 |
| **3** | Private | 90.000000 | 34022.000000 |

SQL for Real:
SELECT workclass,MAX(age), COUNT(*) FROM C1 WHERE (sex IN ('Male', 'Female') OR fnlwgt BETWEEN 103435 AND 123031) OR (workclass <> 'Private' OR `hours.per.week` <= 17) AND (`native.country` IN ('Peru', 'United-States', 'Hong', 'Trinadad&Tobago', 'Italy', 'Nicaragua', 'Cambodia', 'Canada', 'Dominican-Republic', 'Vietnam', 'Portugal') AND `marital.status` = 'Never-married') OR (occupation <> 'Other-service' OR age <= 24) OR (capital < 3137 OR race <> 'White') AND (education = '10th' AND income <> '<=50K') GROUP BY workclass

Resulted in 7 records

SQL for Synthetic:
SELECT workclass,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (sex IN ('Male', 'Female') OR fnlwgt BETWEEN 103435 AND 123031) OR (workclass <> 'Private' OR `hours.per.week` <= 17) AND (`native.country` IN ('Peru', 'United-States', 'Hong', 'Trinadad&Tobago', 'Italy', 'Nicaragua', 'Cambodia', 'Canada', 'Dominican-Republic', 'Vietnam', 'Portugal') AND `marital.status` = 'Never-married') OR (occupation <> 'Other-service' OR age <= 24) OR (capital < 3137 OR race <> 'White') AND (education = '10th' AND income <> '<=50K') GROUP BY workclass

Resulted in 7 records

Normalized Euclidean distance for (age): 2.65

Hellinger Distance: 0.003

==========

Real

| | native.country | education | race | workclass | relationship | occupation | marital.status | income | sex | AVG(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **8769** | Puerto-Rico | Masters | Black | State-gov | Husband | Prof-specialty | Married-civ-spouse | <=50K | Male | nan | nan |
| **2539** | Holand-Netherlands | HS-grad | White | Private | Own-child | Transport-moving | Never-married | <=50K | Male | nan | nan |
| **8123** | Puerto-Rico | HS-grad | White | ? | Unmarried | ? | Separated | <=50K | Male | nan | nan |
| **3107** | Hungary | Some-college | Black | Private | Own-child | Other-service | Never-married | <=50K | Female | nan | nan |
| **6255** | Puerto-Rico | 10th | White | Self-emp-not-inc | Husband | Other-service | Married-civ-spouse | <=50K | Male | nan | nan |

Synthetic

| | native.country | education | race | workclass | relationship | occupation | marital.status | income | sex | AVG(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **8769** | Puerto-Rico | Masters | Black | State-gov | Husband | Prof-specialty | Married-civ-spouse | <=50K | Male | -64.801767 | 1 |
| **2539** | Holand-Netherlands | HS-grad | White | Private | Own-child | Transport-moving | Never-married | <=50K | Male | -100.929840 | 1 |
| **8123** | Puerto-Rico | HS-grad | White | ? | Unmarried | ? | Separated | <=50K | Male | -81.202327 | 1 |
| **3107** | Hungary | Some-college | Black | Private | Own-child | Other-service | Never-married | <=50K | Female | 159.982969 | 1 |
| **6255** | Puerto-Rico | 10th | White | Self-emp-not-inc | Husband | Other-service | Married-civ-spouse | <=50K | Male | 77.691078 | 1 |

SQL for Real:
SELECT `native.country`,education,race,workclass,relationship,occupation,`marital.status`,income,sex,AVG(capital), COUNT(*) FROM C1 WHERE (race IN ('Amer-Indian-Eskimo', 'Black', 'White') AND capital BETWEEN 0 AND 4386) AND (income IN ('>50K', '<=50K') AND `native.country` <> 'United-States') OR (occupation <> 'Sales' OR `hours.per.week` <= 40) OR (relationship LIKE 'Unmarried' OR sex <> 'Male') GROUP BY
`native.country`,education,race,workclass,relationship,occupation,`marital.status`,income,sex

Resulted in 12533 records

SQL for Synthetic:
SELECT `native.country`,education,race,workclass,relationship,occupation,`marital.status`,income,sex,AVG(capital), COUNT(*) FROM C1_syn_06 WHERE (race IN ('Amer-Indian-Eskimo', 'Black', 'White') AND capital BETWEEN 0 AND 4386) AND (income IN ('>50K', '<=50K') AND `native.country` <> 'United-States') OR (occupation <> 'Sales' OR `hours.per.week` <= 40) OR (relationship LIKE 'Unmarried' OR sex <> 'Male') GROUP BY
`native.country`,education,race,workclass,relationship,occupation,`marital.status`,income,sex

Resulted in 16816 records

Normalized Euclidean distance for (capital): 63.96

Hellinger Distance: 0.256

==========

Real

| | marital.status | workclass | sex | race | relationship | native.country | occupation | income | education | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **4110** | Married-civ-spouse | Local-gov | Male | White | Husband | Puerto-Rico | Prof-specialty | <=50K | Bachelors | nan | nan |
| **13993** | Separated | Local-gov | Female | Black | Unmarried | Puerto-Rico | Other-service | <=50K | Some-college | nan | nan |
| **13503** | Never-married | State-gov | Female | Asian-Pac-Islander | Not-in-family | United-States | Prof-specialty | <=50K | Bachelors | 0.000000 | 1.000000 |
| **5410** | Married-civ-spouse | Private | Male | Black | Husband | United-States | Other-service | <=50K | Some-college | 0.000000 | 4.000000 |
| **1446** | Divorced | Private | Female | White | Unmarried | ? | Priv-house-serv | <=50K | HS-grad | nan | nan |

Synthetic

| | marital.status | workclass | sex | race | relationship | native.country | occupation | income | education | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **4110** | Married-civ-spouse | Local-gov | Male | White | Husband | Puerto-Rico | Prof-specialty | <=50K | Bachelors | 67.890254 | 8 |
| **13993** | Separated | Local-gov | Female | Black | Unmarried | Puerto-Rico | Other-service | <=50K | Some-college | 42.156971 | 1 |
| **13503** | Never-married | State-gov | Female | Asian-Pac-Islander | Not-in-family | United-States | Prof-specialty | <=50K | Bachelors | -33.525904 | 2 |
| **5410** | Married-civ-spouse | Private | Male | Black | Husband | United-States | Other-service | <=50K | Some-college | 122.215307 | 2 |
| **1446** | Divorced | Private | Female | White | Unmarried | ? | Priv-house-serv | <=50K | HS-grad | -178.581472 | 1 |

SQL for Real:
SELECT `marital.status`,workclass,sex,race,relationship,`native.country`,occupation,income,education,MAX(capital), COUNT(*) FROM C1 WHERE (fnlwgt > 182401) AND (education <> 'HS-grad' AND race <> 'Asian-Pac-Islander') OR (sex <> 'Male' OR occupation IN ('Adm-clerical', 'Sales', 'Craft-repair', 'Prof-specialty', 'Protective-serv', 'Tech-support', 'Exec-managerial', 'Other-service', '?', 'Armed-Forces', 'Transport-moving', 'Priv-house-serv')) GROUP BY
`marital.status`,workclass,sex,race,relationship,`native.country`,occupation,income,education

Resulted in 11954 records

SQL for Synthetic:
SELECT `marital.status`,workclass,sex,race,relationship,`native.country`,occupation,income,education,MAX(capital), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt > 182401) AND (education <> 'HS-grad' AND race <> 'Asian-Pac-Islander') OR (sex <> 'Male' OR occupation IN ('Adm-clerical', 'Sales', 'Craft-repair', 'Prof-specialty', 'Protective-serv', 'Tech-support', 'Exec-managerial', 'Other-service', '?', 'Armed-Forces', 'Transport-moving', 'Priv-house-serv')) GROUP BY
`marital.status`,workclass,sex,race,relationship,`native.country`,occupation,income,education

Resulted in 15860 records

Normalized Euclidean distance for (capital): 62.59

Hellinger Distance: 0.256

==========

Real

| | sex | occupation | marital.status | workclass | income | relationship | race | native.country | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **138** | Female | Adm-clerical | Separated | Local-gov | <=50K | Not-in-family | Black | United-States | 48 | 1 |
| **679** | Male | ? | Married-spouse-absent | ? | <=50K | Not-in-family | Black | United-States | 71 | 1 |
| **934** | Male | Exec-managerial | Divorced | State-gov | <=50K | Not-in-family | White | United-States | 53 | 1 |
| **1416** | Male | Prof-specialty | Separated | Private | <=50K | Not-in-family | White | United-States | 33 | 1 |
| **1020** | Male | Exec-managerial | Widowed | Private | <=50K | Not-in-family | White | United-States | 120 | 2 |

Synthetic

| | sex | occupation | marital.status | workclass | income | relationship | race | native.country | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **138** | Female | Adm-clerical | Separated | Local-gov | <=50K | Not-in-family | Black | United-States | nan | nan |
| **679** | Male | ? | Married-spouse-absent | ? | <=50K | Not-in-family | Black | United-States | nan | nan |
| **934** | Male | Exec-managerial | Divorced | State-gov | <=50K | Not-in-family | White | United-States | nan | nan |
| **1416** | Male | Prof-specialty | Separated | Private | <=50K | Not-in-family | White | United-States | nan | nan |
| **1020** | Male | Exec-managerial | Widowed | Private | <=50K | Not-in-family | White | United-States | nan | nan |

SQL for Real:
SELECT sex,occupation,`marital.status`,workclass,income,relationship,race,`native.country`,SUM(age), COUNT(*) FROM C1 WHERE (fnlwgt BETWEEN 244147 AND 308686) OR (occupation IN ('Farming-fishing', 'Other-service', 'Armed-Forces', 'Transport-moving', 'Exec-managerial', 'Craft-repair', 'Priv-house-serv', 'Prof-specialty', 'Sales', 'Tech-support', 'Protective-serv', '?', 'Adm-clerical', 'Machine-op-inspct', 'Handlers-cleaners') AND capital = 0) AND (`marital.status` = 'Never-married' AND `native.country` = 'United-States') AND (workclass <> 'Private' AND income <> '<=50K') GROUP BY sex,occupation,`marital.status`,workclass,income,relationship,race,`native.country`

Resulted in 1673 records

SQL for Synthetic:
SELECT sex,occupation,`marital.status`,workclass,income,relationship,race,`native.country`,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt BETWEEN 244147 AND 308686) OR (occupation IN ('Farming-fishing', 'Other-service', 'Armed-Forces', 'Transport-moving', 'Exec-managerial', 'Craft-repair', 'Priv-house-serv', 'Prof-specialty', 'Sales', 'Tech-support', 'Protective-serv', '?', 'Adm-clerical', 'Machine-op-inspct', 'Handlers-cleaners') AND capital = 0) AND (`marital.status` = 'Never-married' AND `native.country` = 'United-States') AND (workclass <> 'Private' AND income <> '<=50K') GROUP BY sex,occupation,`marital.status`,workclass,income,relationship,race,`native.country`

Resulted in 0 records

Normalized Euclidean distance for (age): nan

Hellinger Distance: nan

==========

Real

| | education | native.country | income | marital.status | race | relationship | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **2394** | Some-college | ? | <=50K | Divorced | White | Not-in-family | nan | nan |
| **2452** | Some-college | Columbia | <=50K | Separated | Amer-Indian-Eskimo | Unmarried | nan | nan |
| **639** | 9th | El-Salvador | <=50K | Divorced | White | Unmarried | nan | nan |
| **2426** | Some-college | China | <=50K | Married-civ-spouse | White | Wife | nan | nan |
| **87** | 10th | Puerto-Rico | <=50K | Separated | White | Own-child | nan | nan |

Synthetic

| | education | native.country | income | marital.status | race | relationship | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **2394** | Some-college | ? | <=50K | Divorced | White | Not-in-family | 40.076548 | 17 |
| **2452** | Some-college | Columbia | <=50K | Separated | Amer-Indian-Eskimo | Unmarried | 40.021791 | 1 |
| **639** | 9th | El-Salvador | <=50K | Divorced | White | Unmarried | 39.999984 | 2 |
| **2426** | Some-college | China | <=50K | Married-civ-spouse | White | Wife | 40.063405 | 1 |
| **87** | 10th | Puerto-Rico | <=50K | Separated | White | Own-child | 40.033509 | 1 |

SQL for Real:
SELECT education,`native.country`,income,`marital.status`,race,relationship,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (age BETWEEN 29 AND 53) AND (fnlwgt > 187649 OR income LIKE '<=50K') GROUP BY education,`native.country`,income,`marital.status`,race,relationship

Resulted in 2280 records

SQL for Synthetic:
SELECT education,`native.country`,income,`marital.status`,race,relationship,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (age BETWEEN 29 AND 53) AND (fnlwgt > 187649 OR income LIKE '<=50K') GROUP BY education,`native.country`,income,`marital.status`,race,relationship

Resulted in 2814 records

Normalized Euclidean distance for (`hours.per.week`): 28.71

Hellinger Distance: 0.41

==========

Real

| | education | occupation | native.country | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|
| **662** | 9th | Adm-clerical | Puerto-Rico | nan | nan |
| **600** | 7th-8th | Machine-op-inspct | Puerto-Rico | 189336.500000 | 2.000000 |
| **901** | Assoc-acdm | Prof-specialty | Thailand | nan | nan |
| **536** | 7th-8th | ? | Columbia | nan | nan |
| **1161** | Bachelors | Adm-clerical | Jamaica | nan | nan |

Synthetic

| | education | occupation | native.country | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|
| **662** | 9th | Adm-clerical | Puerto-Rico | 171279.267775 | 7 |
| **600** | 7th-8th | Machine-op-inspct | Puerto-Rico | 171178.620321 | 45 |
| **901** | Assoc-acdm | Prof-specialty | Thailand | 177462.778514 | 1 |
| **536** | 7th-8th | ? | Columbia | 177526.773129 | 3 |
| **1161** | Bachelors | Adm-clerical | Jamaica | 188521.767296 | 1 |

SQL for Real:
SELECT education,occupation,`native.country`,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (relationship IN ('Unmarried', 'Own-child') OR fnlwgt < 419895) GROUP BY education,occupation,`native.country`

Resulted in 1994 records

SQL for Synthetic:
SELECT education,occupation,`native.country`,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (relationship IN ('Unmarried', 'Own-child') OR fnlwgt < 419895) GROUP BY education,occupation,`native.country`

Resulted in 2374 records

Normalized Euclidean distance for (fnlwgt): 32.39

Hellinger Distance: 0.342

==========

Real

| | relationship | workclass | marital.status | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|
| **19** | Husband | Self-emp-not-inc | Widowed | nan | nan |
| **169** | Unmarried | Self-emp-inc | Divorced | 176674.239130 | 46.000000 |
| **68** | Other-relative | ? | Never-married | 194720.289855 | 69.000000 |
| **66** | Other-relative | ? | Married-civ-spouse | 249134.727273 | 11.000000 |
| **41** | Not-in-family | Private | Married-AF-spouse | nan | nan |

Synthetic

| | relationship | workclass | marital.status | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|
| **19** | Husband | Self-emp-not-inc | Widowed | 161040.240767 | 1 |
| **169** | Unmarried | Self-emp-inc | Divorced | 174183.955784 | 34 |
| **68** | Other-relative | ? | Never-married | 182904.442904 | 74 |
| **66** | Other-relative | ? | Married-civ-spouse | 178818.102039 | 16 |
| **41** | Not-in-family | Private | Married-AF-spouse | 188527.416317 | 2 |

SQL for Real:
SELECT relationship,workclass,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (race = 'White') AND (relationship = 'Own-child' AND income <> '>50K') OR (fnlwgt <> 202191 OR capital = 0) OR (occupation <> 'Adm-clerical' OR `marital.status` <> 'Never-married') OR (sex = 'Female' OR education IN ('9th', 'Masters', 'Preschool', 'Assoc-voc', '11th', '1st-4th', 'Doctorate', '7th-8th', '12th', '10th', 'Prof-school', 'Bachelors')) OR (workclass LIKE 'Private' OR `hours.per.week` BETWEEN 60 AND 40) GROUP BY relationship,workclass,`marital.status`

Resulted in 171 records

SQL for Synthetic:
SELECT relationship,workclass,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (race = 'White') AND (relationship = 'Own-child' AND income <> '>50K') OR (fnlwgt <> 202191 OR capital = 0) OR (occupation <> 'Adm-clerical' OR `marital.status` <> 'Never-married') OR (sex = 'Female' OR education IN ('9th', 'Masters', 'Preschool', 'Assoc-voc', '11th', '1st-4th', 'Doctorate', '7th-8th', '12th', '10th', 'Prof-school', 'Bachelors')) OR (workclass LIKE 'Private' OR `hours.per.week` BETWEEN 60 AND 40) GROUP BY relationship,workclass,`marital.status`

Resulted in 209 records

Normalized Euclidean distance for (fnlwgt): 12.57

Hellinger Distance: 0.032

==========

Real

| | income | relationship | SUM(age) | COUNT(*) |
|---|---|---|---|---|
| **6** | >50K | Husband | 389664 | 8686 |
| **1** | <=50K | Not-in-family | 428238 | 11307 |
| **10** | >50K | Unmarried | 13607 | 296 |
| **11** | >50K | Wife | 43776 | 1068 |
| **7** | >50K | Not-in-family | 54154 | 1250 |

Synthetic

| | income | relationship | SUM(age) | COUNT(*) |
|---|---|---|---|---|
| **6** | >50K | Husband | 164056.000000 | 3652.000000 |
| **1** | <=50K | Not-in-family | 467606.000000 | 12106.000000 |
| **10** | >50K | Unmarried | 3842.000000 | 82.000000 |
| **11** | >50K | Wife | 21604.000000 | 533.000000 |
| **7** | >50K | Not-in-family | 14365.000000 | 318.000000 |

SQL for Real:
SELECT income,relationship,SUM(age), COUNT(*) FROM C1 WHERE (education LIKE 'Some-college' OR workclass IN ('Local-gov', '?', 'Private')) OR (`native.country` LIKE 'United-States' OR sex = 'Male') AND ((capital BETWEEN 0 AND -2042) AND age = 37) OR (race = 'White' OR income LIKE '<=50K') GROUP BY income,relationship

Resulted in 12 records

SQL for Synthetic:
SELECT income,relationship,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (education LIKE 'Some-college' OR workclass IN ('Local-gov', '?', 'Private')) OR (`native.country` LIKE 'United-States' OR sex = 'Male') AND ((capital BETWEEN 0 AND -2042) AND age = 37) OR (race = 'White' OR income LIKE '<=50K') GROUP BY income,relationship

Resulted in 12 records

Normalized Euclidean distance for (age): 3.46

Hellinger Distance: 0.147

==========

Real

| | occupation | workclass | sex | marital.status | education | income | race | relationship | native.country | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **104** | Transport-moving | Self-emp-not-inc | Male | Married-civ-spouse | Some-college | <=50K | White | Husband | United-States | 1988773.000000 | 11.000000 |
| **40** | Transport-moving | Private | Male | Married-civ-spouse | Some-college | <=50K | Amer-Indian-Eskimo | Husband | United-States | nan | nan |
| **68** | Transport-moving | Private | Male | Married-spouse-absent | Some-college | <=50K | White | Not-in-family | United-States | 147989.000000 | 1.000000 |
| **112** | Transport-moving | Self-emp-not-inc | Male | Separated | Some-college | <=50K | White | Own-child | Puerto-Rico | nan | nan |
| **77** | Transport-moving | Private | Male | Never-married | Some-college | <=50K | White | Not-in-family | United-States | 5182219.000000 | 30.000000 |

Synthetic

| | occupation | workclass | sex | marital.status | education | income | race | relationship | native.country | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **104** | Transport-moving | Self-emp-not-inc | Male | Married-civ-spouse | Some-college | <=50K | White | Husband | United-States | 870995.964396 | 5 |
| **40** | Transport-moving | Private | Male | Married-civ-spouse | Some-college | <=50K | Amer-Indian-Eskimo | Husband | United-States | 164152.927823 | 1 |
| **68** | Transport-moving | Private | Male | Married-spouse-absent | Some-college | <=50K | White | Not-in-family | United-States | 181085.518171 | 1 |
| **112** | Transport-moving | Self-emp-not-inc | Male | Separated | Some-college | <=50K | White | Own-child | Puerto-Rico | 168157.046218 | 1 |
| **77** | Transport-moving | Private | Male | Never-married | Some-college | <=50K | White | Not-in-family | United-States | 2807629.042519 | 15 |

SQL for Real:
SELECT occupation,workclass,sex,`marital.status`,education,income,race,relationship,`native.country`,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (education LIKE 'Some-college') AND ((capital BETWEEN 15024 AND 0) OR occupation LIKE 'Transport-moving') GROUP BY occupation,workclass,sex,`marital.status`,education,income,race,relationship,`native.country`

Resulted in 94 records

SQL for Synthetic:
SELECT occupation,workclass,sex,`marital.status`,education,income,race,relationship,`native.country`,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (education LIKE 'Some-college') AND ((capital BETWEEN 15024 AND 0) OR occupation LIKE 'Transport-moving') GROUP BY occupation,workclass,sex,`marital.status`,education,income,race,relationship,`native.country`

Resulted in 123 records

Normalized Euclidean distance for (fnlwgt): 5.83

Hellinger Distance: 0.215

==========

Real

| | occupation | sex | native.country | education | income | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **1411** | Exec-managerial | Male | Thailand | Masters | >50K | nan | nan |
| **1367** | Exec-managerial | Male | Puerto-Rico | 1st-4th | <=50K | nan | nan |
| **776** | Craft-repair | Male | ? | Assoc-voc | >50K | 140.000000 | 3.000000 |
| **3576** | Sales | Male | United-States | 5th-6th | <=50K | 98.000000 | 3.000000 |
| **2202** | Other-service | Female | Holand-Netherlands | HS-grad | <=50K | nan | nan |

Synthetic

| | occupation | sex | native.country | education | income | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **1411** | Exec-managerial | Male | Thailand | Masters | >50K | 40.387970 | 1 |
| **1367** | Exec-managerial | Male | Puerto-Rico | 1st-4th | <=50K | 40.025847 | 1 |
| **776** | Craft-repair | Male | ? | Assoc-voc | >50K | 119.902595 | 3 |
| **3576** | Sales | Male | United-States | 5th-6th | <=50K | 119.813716 | 3 |
| **2202** | Other-service | Female | Holand-Netherlands | HS-grad | <=50K | 119.461778 | 3 |

SQL for Real:
SELECT occupation,sex,`native.country`,education,income,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (race = 'White' OR `marital.status` = 'Never-married') AND (`native.country` LIKE 'United-States' AND age BETWEEN 48 AND 50) AND ((capital BETWEEN 0 AND 0) AND workclass IN ('Self-emp-inc', '?', 'Self-emp-not-inc', 'Private')) AND ((fnlwgt BETWEEN 190628 AND 184471) AND occupation LIKE 'Prof-specialty') OR (relationship = 'Husband' OR education <> 'HS-grad') OR (income LIKE '<=50K' OR sex <> 'Male') GROUP BY occupation,sex,`native.country`,education,income

Resulted in 3172 records

SQL for Synthetic:
SELECT occupation,sex,`native.country`,education,income,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (race = 'White' OR `marital.status` = 'Never-married') AND (`native.country` LIKE 'United-States' AND age BETWEEN 48 AND 50) AND ((capital BETWEEN 0 AND 0) AND workclass IN ('Self-emp-inc', '?', 'Self-emp-not-inc', 'Private')) AND ((fnlwgt BETWEEN 190628 AND 184471) AND occupation LIKE 'Prof-specialty') OR (relationship = 'Husband' OR education <> 'HS-grad') OR (income LIKE '<=50K' OR sex <> 'Male') GROUP BY occupation,sex,`native.country`,education,income

Resulted in 4017 records

Normalized Euclidean distance for (`hours.per.week`): 37.26

Hellinger Distance: 0.335

==========

Real

| | marital.status | native.country | education | relationship | race | workclass | occupation | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **987** | Divorced | Puerto-Rico | 11th | Not-in-family | White | ? | Craft-repair | nan | nan |
| **4672** | Married-civ-spouse | Puerto-Rico | Bachelors | Wife | White | Self-emp-not-inc | Prof-specialty | nan | nan |
| **1484** | Divorced | Puerto-Rico | Some-college | Unmarried | Amer-Indian-Eskimo | Private | Exec-managerial | nan | nan |
| **2698** | Married-civ-spouse | ? | HS-grad | Husband | White | State-gov | Craft-repair | nan | nan |
| **8017** | Never-married | Laos | Bachelors | Own-child | White | Self-emp-not-inc | Craft-repair | nan | nan |

Synthetic

| | marital.status | native.country | education | relationship | race | workclass | occupation | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **987** | Divorced | Puerto-Rico | 11th | Not-in-family | White | ? | Craft-repair | 62 | 1 |
| **4672** | Married-civ-spouse | Puerto-Rico | Bachelors | Wife | White | Self-emp-not-inc | Prof-specialty | 45 | 4 |
| **1484** | Divorced | Puerto-Rico | Some-college | Unmarried | Amer-Indian-Eskimo | Private | Exec-managerial | 44 | 1 |
| **2698** | Married-civ-spouse | ? | HS-grad | Husband | White | State-gov | Craft-repair | 36 | 1 |
| **8017** | Never-married | Laos | Bachelors | Own-child | White | Self-emp-not-inc | Craft-repair | 34 | 1 |

SQL for Real:
SELECT `marital.status`,`native.country`,education,relationship,race,workclass,occupation,MIN(age), COUNT(*) FROM C1 WHERE (sex <> 'Male' OR race LIKE 'White') GROUP BY `marital.status`,`native.country`,education,relationship,race,workclass,occupation

Resulted in 8130 records

SQL for Synthetic:
SELECT `marital.status`,`native.country`,education,relationship,race,workclass,occupation,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (sex <> 'Male' OR race LIKE 'White') GROUP BY `marital.status`,`native.country`,education,relationship,race,workclass,occupation

Resulted in 12535 records

Normalized Euclidean distance for (age): 56.58

Hellinger Distance: 0.228

==========

Real

| | marital.status | workclass | income | relationship | race | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **109** | Married-civ-spouse | Self-emp-inc | <=50K | Wife | White | 40 | 4 |
| **23** | Divorced | Local-gov | <=50K | Unmarried | White | 40 | 29 |
| **0** | Divorced | ? | <=50K | Not-in-family | Asian-Pac-Islander | 40 | 1 |
| **99** | Married-civ-spouse | Private | <=50K | Wife | Asian-Pac-Islander | 40 | 9 |
| **54** | Divorced | State-gov | <=50K | Other-relative | Black | 32 | 1 |

Synthetic

| | marital.status | workclass | income | relationship | race | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **109** | Married-civ-spouse | Self-emp-inc | <=50K | Wife | White | nan | nan |
| **23** | Divorced | Local-gov | <=50K | Unmarried | White | nan | nan |
| **0** | Divorced | ? | <=50K | Not-in-family | Asian-Pac-Islander | nan | nan |
| **99** | Married-civ-spouse | Private | <=50K | Wife | Asian-Pac-Islander | nan | nan |
| **54** | Divorced | State-gov | <=50K | Other-relative | Black | nan | nan |

SQL for Real:
SELECT `marital.status`,workclass,income,relationship,race,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (race IN ('Other', 'Asian-Pac-Islander', 'White') OR sex IN ('Male', 'Female')) AND (income LIKE '<=50K' AND capital BETWEEN 0 AND 0) AND ((`hours.per.week` BETWEEN 20 AND 40) AND education IN ('1st-4th', 'Bachelors', '11th', '9th', 'Assoc-acdm', '7th-8th')) OR (relationship = 'Not-in-family' OR age = 43) AND (`marital.status` LIKE 'Married-civ-spouse' AND `native.country` IN ('Dominican-Republic', 'England', 'Ireland', 'Poland')) GROUP BY `marital.status`,workclass,income,relationship,race

Resulted in 327 records

SQL for Synthetic:
SELECT `marital.status`,workclass,income,relationship,race,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (race IN ('Other', 'Asian-Pac-Islander', 'White') OR sex IN ('Male', 'Female')) AND (income LIKE '<=50K' AND capital BETWEEN 0 AND 0) AND ((`hours.per.week` BETWEEN 20 AND 40) AND education IN ('1st-4th', 'Bachelors', '11th', '9th', 'Assoc-acdm', '7th-8th')) OR (relationship = 'Not-in-family' OR age = 43) AND (`marital.status` LIKE 'Married-civ-spouse' AND `native.country` IN ('Dominican-Republic', 'England', 'Ireland', 'Poland')) GROUP BY `marital.status`,workclass,income,relationship,race

Resulted in 1 records

Normalized Euclidean distance for (`hours.per.week`): 0.0

Hellinger Distance: 0.0

==========

Real

| | income | sex | marital.status | relationship | education | occupation | race | native.country | workclass | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **494** | <=50K | Female | Divorced | Own-child | Assoc-acdm | Adm-clerical | White | United-States | Private | 30 | 3 |
| **8778** | <=50K | Male | Never-married | Own-child | HS-grad | Adm-clerical | White | United-States | Local-gov | 19 | 5 |
| **4655** | <=50K | Male | Divorced | Not-in-family | Assoc-voc | Exec-managerial | Amer-Indian-Eskimo | United-States | Local-gov | 53 | 1 |
| **8657** | <=50K | Male | Never-married | Own-child | Assoc-voc | Farming-fishing | White | United-States | Private | 20 | 4 |
| **7724** | <=50K | Male | Never-married | Not-in-family | Doctorate | Prof-specialty | White | United-States | Self-emp-not-inc | 41 | 1 |

Synthetic

| | income | sex | marital.status | relationship | education | occupation | race | native.country | workclass | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **494** | <=50K | Female | Divorced | Own-child | Assoc-acdm | Adm-clerical | White | United-States | Private | nan | nan |
| **8778** | <=50K | Male | Never-married | Own-child | HS-grad | Adm-clerical | White | United-States | Local-gov | nan | nan |
| **4655** | <=50K | Male | Divorced | Not-in-family | Assoc-voc | Exec-managerial | Amer-Indian-Eskimo | United-States | Local-gov | nan | nan |
| **8657** | <=50K | Male | Never-married | Own-child | Assoc-voc | Farming-fishing | White | United-States | Private | nan | nan |
| **7724** | <=50K | Male | Never-married | Not-in-family | Doctorate | Prof-specialty | White | United-States | Self-emp-not-inc | nan | nan |

SQL for Real:
SELECT income,sex,`marital.status`,relationship,education,occupation,race,`native.country`,workclass,MIN(age), COUNT(*) FROM C1 WHERE (sex LIKE 'Female') OR ((age BETWEEN 55 AND 25) OR capital = 0) GROUP BY income,sex,`marital.status`,relationship,education,occupation,race,`native.country`,workclass

Resulted in 11974 records

SQL for Synthetic:
SELECT income,sex,`marital.status`,relationship,education,occupation,race,`native.country`,workclass,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (sex LIKE 'Female') OR ((age BETWEEN 55 AND 25) OR capital = 0) GROUP BY income,sex,`marital.status`,relationship,education,occupation,race,`native.country`,workclass

Resulted in 7579 records

Normalized Euclidean distance for (age): 41.44

Hellinger Distance: 0.196

==========

Real

| | marital.status | native.country | race | workclass | income | relationship | sex | occupation | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **2171** | Never-married | United-States | White | Private | <=50K | Other-relative | Female | Adm-clerical | 0 | 41 |
| **1410** | Married-spouse-absent | United-States | Black | Private | <=50K | Not-in-family | Male | Other-service | 0 | 1 |
| **2238** | Never-married | United-States | White | Private | <=50K | Unmarried | Male | Farming-fishing | 0 | 15 |
| **360** | Divorced | United-States | White | Local-gov | <=50K | Not-in-family | Female | Transport-moving | 0 | 3 |
| **1828** | Never-married | United-States | Black | Private | <=50K | Other-relative | Female | Other-service | 0 | 11 |

Synthetic

| | marital.status | native.country | race | workclass | income | relationship | sex | occupation | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **2171** | Never-married | United-States | White | Private | <=50K | Other-relative | Female | Adm-clerical | nan | nan |
| **1410** | Married-spouse-absent | United-States | Black | Private | <=50K | Not-in-family | Male | Other-service | nan | nan |
| **2238** | Never-married | United-States | White | Private | <=50K | Unmarried | Male | Farming-fishing | nan | nan |
| **360** | Divorced | United-States | White | Local-gov | <=50K | Not-in-family | Female | Transport-moving | nan | nan |
| **1828** | Never-married | United-States | Black | Private | <=50K | Other-relative | Female | Other-service | nan | nan |

SQL for Real:
SELECT `marital.status`,`native.country`,race,workclass,income,relationship,sex,occupation,MIN(capital), COUNT(*) FROM C1 WHERE (capital = 0 AND `native.country` LIKE 'United-States') GROUP BY `marital.status`,`native.country`,race,workclass,income,relationship,sex,occupation

Resulted in 3105 records

SQL for Synthetic:
SELECT `marital.status`,`native.country`,race,workclass,income,relationship,sex,occupation,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (capital = 0 AND `native.country` LIKE 'United-States') GROUP BY `marital.status`,`native.country`,race,workclass,income,relationship,sex,occupation

Resulted in 0 records

Normalized Euclidean distance for (capital): nan

Hellinger Distance: nan

==========

Real

| | native.country | race | sex | workclass | relationship | education | marital.status | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **1953** | Japan | Black | Male | Private | Husband | 9th | Married-civ-spouse | nan | nan |
| **5360** | United-States | Black | Female | State-gov | Unmarried | Some-college | Divorced | 0.000000 | 7.000000 |
| **3864** | Puerto-Rico | White | Female | Self-emp-not-inc | Not-in-family | HS-grad | Never-married | nan | nan |
| **4372** | Puerto-Rico | White | Male | Self-emp-not-inc | Own-child | Assoc-acdm | Married-spouse-absent | nan | nan |
| **6149** | United-States | White | Female | Private | Unmarried | Assoc-acdm | Divorced | -2238.000000 | 47.000000 |

Synthetic

| | native.country | race | sex | workclass | relationship | education | marital.status | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **1953** | Japan | Black | Male | Private | Husband | 9th | Married-civ-spouse | 30.655048 | 1 |
| **5360** | United-States | Black | Female | State-gov | Unmarried | Some-college | Divorced | -108.514460 | 4 |
| **3864** | Puerto-Rico | White | Female | Self-emp-not-inc | Not-in-family | HS-grad | Never-married | -84.214585 | 5 |
| **4372** | Puerto-Rico | White | Male | Self-emp-not-inc | Own-child | Assoc-acdm | Married-spouse-absent | -114.909826 | 1 |
| **6149** | United-States | White | Female | Private | Unmarried | Assoc-acdm | Divorced | -236.962343 | 29 |

SQL for Real:
SELECT `native.country`,race,sex,workclass,relationship,education,`marital.status`,MIN(capital), COUNT(*) FROM C1 WHERE (education <> 'Bachelors') AND (occupation <> 'Sales' AND race IN ('Asian-Pac-Islander', 'Other', 'Black', 'White', 'Amer-Indian-Eskimo')) OR (capital = 0 OR age BETWEEN 27 AND 62) AND (income LIKE '>50K' AND fnlwgt BETWEEN 135388 AND 236091) OR (workclass IN ('Local-gov', 'Self-emp-not-inc', '?', 'State-gov', 'Self-emp-inc', 'Private', 'Federal-gov') OR `marital.status` <> 'Married-civ-spouse') AND (relationship LIKE 'Husband' AND `native.country` = 'United-States') GROUP BY `native.country`,race,sex,workclass,relationship,education,`marital.status`

Resulted in 5167 records

SQL for Synthetic:
SELECT `native.country`,race,sex,workclass,relationship,education,`marital.status`,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (education <> 'Bachelors') AND (occupation <> 'Sales' AND race IN ('Asian-Pac-Islander', 'Other', 'Black', 'White', 'Amer-Indian-Eskimo')) OR (capital = 0 OR age BETWEEN 27 AND 62) AND (income LIKE '>50K' AND fnlwgt BETWEEN 135388 AND 236091) OR (workclass IN ('Local-gov', 'Self-emp-not-inc', '?', 'State-gov', 'Self-emp-inc', 'Private', 'Federal-gov') OR `marital.status` <> 'Married-civ-spouse') AND (relationship LIKE 'Husband' AND `native.country` = 'United-States') GROUP BY `native.country`,race,sex,workclass,relationship,education,`marital.status`

Resulted in 7159 records

Normalized Euclidean distance for (capital): 44.78

Hellinger Distance: 0.29

==========

Real

| | occupation | income | relationship | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|
| **50** | Other-service | >50K | Unmarried | 0 | 1 |
| **33** | Exec-managerial | >50K | Wife | 15024 | 1 |
| **0** | ? | <=50K | Husband | 6767 | 534 |
| **75** | Transport-moving | <=50K | Unmarried | 0 | 1 |
| **3** | ? | <=50K | Own-child | 34095 | 872 |

Synthetic

| | occupation | income | relationship | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|
| **50** | Other-service | >50K | Unmarried | nan | nan |
| **33** | Exec-managerial | >50K | Wife | -3.395264 | 3.000000 |
| **0** | ? | <=50K | Husband | 483.749628 | 629.000000 |
| **75** | Transport-moving | <=50K | Unmarried | 10.560411 | 1.000000 |
| **3** | ? | <=50K | Own-child | 354.918062 | 883.000000 |

SQL for Real:
SELECT occupation,income,relationship,MAX(capital), COUNT(*) FROM C1 WHERE (relationship IN ('Not-in-family', 'Wife', 'Other-relative', 'Unmarried', 'Husband', 'Own-child')) AND (education <> 'HS-grad' AND fnlwgt >= 183884) AND (sex = 'Male' AND age BETWEEN 38 AND 26) OR (`native.country` IN ('Iran', 'Trinadad&Tobago', 'South', 'Outlying-US(Guam-USVI-etc)') OR occupation = '?') GROUP BY occupation,income,relationship

Resulted in 76 records

SQL for Synthetic:
SELECT occupation,income,relationship,MAX(capital), COUNT(*) FROM C1_syn_06 WHERE (relationship IN ('Not-in-family', 'Wife', 'Other-relative', 'Unmarried', 'Husband', 'Own-child')) AND (education <> 'HS-grad' AND fnlwgt >= 183884) AND (sex = 'Male' AND age BETWEEN 38 AND 26) OR (`native.country` IN ('Iran', 'Trinadad&Tobago', 'South', 'Outlying-US(Guam-USVI-etc)') OR occupation = '?') GROUP BY occupation,income,relationship

Resulted in 72 records

Normalized Euclidean distance for (capital): 7.55

Hellinger Distance: 0.136

==========

Real

| | income | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|
| **1** | >50K | 531168 | 11686 |
| **0** | <=50K | 1443102 | 37155 |

Synthetic

| | income | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|
| **1** | >50K | 191049.236539 | 4762.000000 |
| **0** | <=50K | 1761886.259794 | 44080.000000 |

SQL for Real:
SELECT income,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (age >= 57) AND (income = '>50K' AND race = 'White') AND (relationship IN ('Wife', 'Unmarried', 'Other-relative', 'Husband', 'Own-child', 'Not-in-family') AND `hours.per.week` BETWEEN 38 AND 40) OR (education LIKE 'HS-grad' OR capital < 0) OR (workclass LIKE 'Private' OR fnlwgt <> 510072) GROUP BY income

Resulted in 2 records

SQL for Synthetic:
SELECT income,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (age >= 57) AND (income = '>50K' AND race = 'White') AND (relationship IN ('Wife', 'Unmarried', 'Other-relative', 'Husband', 'Own-child', 'Not-in-family') AND `hours.per.week` BETWEEN 38 AND 40) OR (education LIKE 'HS-grad' OR capital < 0) OR (workclass LIKE 'Private' OR fnlwgt <> 510072) GROUP BY income

Resulted in 2 records

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

Hellinger Distance: 0.137

==========

Real

| education | relationship | income | marital.status | native.country | workclass | sex | occupation | race | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|

Synthetic

| education | relationship | income | marital.status | native.country | workclass | sex | occupation | race | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|

SQL for Real:
SELECT education,relationship,income,`marital.status`,`native.country`,workclass,sex,occupation,race,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (`native.country` <> 'United-States') AND ((fnlwgt BETWEEN 200560 AND 108438) AND age > 34) OR (income LIKE '<=50K' OR sex LIKE 'Male') AND (capital > 0 AND `hours.per.week` BETWEEN 40 AND 20) AND (education LIKE '9th' AND `marital.status` LIKE 'Never-married') AND (occupation LIKE 'Tech-support' AND relationship LIKE 'Not-in-family') GROUP BY education,relationship,income,`marital.status`,`native.country`,workclass,sex,occupation,race

Resulted in 0 records

SQL for Synthetic:
SELECT education,relationship,income,`marital.status`,`native.country`,workclass,sex,occupation,race,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`native.country` <> 'United-States') AND ((fnlwgt BETWEEN 200560 AND 108438) AND age > 34) OR (income LIKE '<=50K' OR sex LIKE 'Male') AND (capital > 0 AND `hours.per.week` BETWEEN 40 AND 20) AND (education LIKE '9th' AND `marital.status` LIKE 'Never-married') AND (occupation LIKE 'Tech-support' AND relationship LIKE 'Not-in-family') GROUP BY education,relationship,income,`marital.status`,`native.country`,workclass,sex,occupation,race

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

==========

Real

| | relationship | marital.status | workclass | occupation | native.country | race | income | education | sex | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **12767** | Own-child | Never-married | Self-emp-not-inc | Sales | United-States | White | <=50K | Some-college | Female | 64.000000 | 2.000000 |
| **6691** | Not-in-family | Never-married | Private | Adm-clerical | United-States | White | <=50K | Bachelors | Male | 1337.000000 | 44.000000 |
| **4400** | Not-in-family | Divorced | Local-gov | Exec-managerial | Japan | White | <=50K | Masters | Male | nan | nan |
| **6572** | Not-in-family | Never-married | Private | Adm-clerical | India | Black | <=50K | HS-grad | Female | nan | nan |
| **15765** | Unmarried | Widowed | Private | Machine-op-inspct | United-States | Black | <=50K | HS-grad | Female | 236.000000 | 5.000000 |

Synthetic

| | relationship | marital.status | workclass | occupation | native.country | race | income | education | sex | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **12767** | Own-child | Never-married | Self-emp-not-inc | Sales | United-States | White | <=50K | Some-college | Female | 43 | 2 |
| **6691** | Not-in-family | Never-married | Private | Adm-clerical | United-States | White | <=50K | Bachelors | Male | 1037 | 34 |
| **4400** | Not-in-family | Divorced | Local-gov | Exec-managerial | Japan | White | <=50K | Masters | Male | 40 | 1 |
| **6572** | Not-in-family | Never-married | Private | Adm-clerical | India | Black | <=50K | HS-grad | Female | 28 | 1 |
| **15765** | Unmarried | Widowed | Private | Machine-op-inspct | United-States | Black | <=50K | HS-grad | Female | 119 | 2 |

SQL for Real:
SELECT relationship,`marital.status`,workclass,occupation,`native.country`,race,income,education,sex,SUM(age), COUNT(*) FROM C1 WHERE (`marital.status` LIKE 'Divorced' OR workclass = 'Private') AND (relationship LIKE 'Husband' AND occupation LIKE 'Machine-op-inspct') OR (sex LIKE 'Female' OR `hours.per.week` BETWEEN 10 AND 45) OR (capital = 0 OR fnlwgt = 241895) GROUP BY relationship,`marital.status`,workclass,occupation,`native.country`,race,income,education,sex

Resulted in 12433 records

SQL for Synthetic:
SELECT relationship,`marital.status`,workclass,occupation,`native.country`,race,income,education,sex,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` LIKE 'Divorced' OR workclass = 'Private') AND (relationship LIKE 'Husband' AND occupation LIKE 'Machine-op-inspct') OR (sex LIKE 'Female' OR `hours.per.week` BETWEEN 10 AND 45) OR (capital = 0 OR fnlwgt = 241895) GROUP BY relationship,`marital.status`,workclass,occupation,`native.country`,race,income,education,sex

Resulted in 17127 records

Normalized Euclidean distance for (age): 64.44

Hellinger Distance: 0.245

==========

Real

| income | race | workclass | relationship | sex | education | marital.status | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|

Synthetic

**income race workclass relationship sex education marital.status SUM(`hours.per.week`) COUNT(*)**

SQL for Real:
SELECT income,race,workclass,relationship,sex,education,`marital.status`,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (relationship IN ('Unmarried', 'Not-in-family', 'Husband', 'Own-child')) AND (`native.country` IN ('Germany', 'Columbia', 'Ecuador', 'Italy', 'Guatemala', 'Jamaica', 'Hong', 'China', 'Honduras', 'Outlying-US(Guam-USVI-etc)', 'Philippines', 'Puerto-Rico', 'Poland', 'Canada', '?', 'Cambodia', 'Thailand', 'Taiwan', 'Holand-Netherlands', 'Cuba', 'Haiti', 'India') AND `marital.status` LIKE 'Never-married') AND (fnlwgt = 113843 AND occupation IN ('Farming-fishing', 'Handlers-cleaners', 'Tech-support', 'Priv-house-serv', 'Transport-moving', 'Exec-managerial', 'Sales', 'Prof-specialty', '?', 'Adm-clerical', 'Craft-repair')) AND (education IN ('Some-college', '5th-6th', 'Bachelors', 'Prof-school', '9th', '11th', 'Assoc-voc', 'Assoc-acdm', 'HS-grad', 'Preschool', '12th', 'Doctorate', 'Masters', '10th') AND capital BETWEEN 0 AND 0) GROUP BY income,race,workclass,relationship,sex,education,`marital.status`

Resulted in 0 records

SQL for Synthetic:
SELECT income,race,workclass,relationship,sex,education,`marital.status`,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (relationship IN ('Unmarried', 'Not-in-family', 'Husband', 'Own-child')) AND (`native.country` IN ('Germany', 'Columbia', 'Ecuador', 'Italy', 'Guatemala', 'Jamaica', 'Hong', 'China', 'Honduras', 'Outlying-US(Guam-USVI-etc)', 'Philippines', 'Puerto-Rico', 'Poland', 'Canada', '?', 'Cambodia', 'Thailand', 'Taiwan', 'Holand-Netherlands', 'Cuba', 'Haiti', 'India') AND `marital.status` LIKE 'Never-married') AND (fnlwgt = 113843 AND occupation IN ('Farming-fishing', 'Handlers-cleaners', 'Tech-support', 'Priv-house-serv', 'Transport-moving', 'Exec-managerial', 'Sales', 'Prof-specialty', '?', 'Adm-clerical', 'Craft-repair')) AND (education IN ('Some-college', '5th-6th', 'Bachelors', 'Prof-school', '9th', '11th', 'Assoc-voc', 'Assoc-acdm', 'HS-grad', 'Preschool', '12th', 'Doctorate', 'Masters', '10th') AND capital BETWEEN 0 AND 0) GROUP BY income,race,workclass,relationship,sex,education,`marital.status`

Resulted in 0 records

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

Hellinger Distance: nan

==========

Real

| | income | marital.status | native.country | occupation | relationship | workclass | education | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **8331** | >50K | Married-civ-spouse | Japan | Exec-managerial | Other-relative | Private | Bachelors | nan | nan |
| **242** | <=50K | Divorced | Hungary | Exec-managerial | Unmarried | Private | Some-college | nan | nan |
| **7952** | <=50K | Widowed | United-States | Sales | Unmarried | Private | 12th | 40.000000 | 1.000000 |
| **3147** | <=50K | Married-civ-spouse | Puerto-Rico | Sales | Wife | Private | HS-grad | nan | nan |
| **7080** | <=50K | Separated | Puerto-Rico | Tech-support | Not-in-family | Private | Some-college | nan | nan |

Synthetic

| | income | marital.status | native.country | occupation | relationship | workclass | education | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **8331** | >50K | Married-civ-spouse | Japan | Exec-managerial | Other-relative | Private | Bachelors | 40.078594 | 1 |
| **242** | <=50K | Divorced | Hungary | Exec-managerial | Unmarried | Private | Some-college | 40.013028 | 2 |
| **7952** | <=50K | Widowed | United-States | Sales | Unmarried | Private | 12th | 39.968278 | 1 |
| **3147** | <=50K | Married-civ-spouse | Puerto-Rico | Sales | Wife | Private | HS-grad | 40.005109 | 5 |
| **7080** | <=50K | Separated | Puerto-Rico | Tech-support | Not-in-family | Private | Some-college | 40.033564 | 1 |

SQL for Real:
SELECT income,`marital.status`,`native.country`,occupation,relationship,workclass,education,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (sex = 'Female') AND (`marital.status` IN ('Married-AF-spouse', 'Never-married') AND relationship LIKE 'Husband') OR (`native.country` = 'United-States' OR fnlwgt >= 357596) OR (workclass <> 'Private' OR race = 'White') AND (`hours.per.week` > 40 AND income IN ('<=50K', '>50K')) GROUP BY
income,`marital.status`,`native.country`,occupation,relationship,workclass,education

Resulted in 6399 records

SQL for Synthetic:
SELECT income,`marital.status`,`native.country`,occupation,relationship,workclass,education,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (sex = 'Female') AND (`marital.status` IN ('Married-AF-spouse', 'Never-married') AND relationship LIKE 'Husband') OR (`native.country` = 'United-States' OR fnlwgt >= 357596) OR (workclass <> 'Private' OR race = 'White') AND (`hours.per.week` > 40 AND income IN ('<=50K', '>50K')) GROUP BY
income,`marital.status`,`native.country`,occupation,relationship,workclass,education

Resulted in 9203 records

Normalized Euclidean distance for (`hours.per.week`): 54.71

Hellinger Distance: 0.201

==========

Real

| | native.country | marital.status | race | occupation | relationship | income | workclass | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **4516** | Puerto-Rico | Never-married | White | Adm-clerical | Own-child | <=50K | Self-emp-inc | nan | nan |
| **1015** | Columbia | Never-married | White | Protective-serv | Not-in-family | <=50K | State-gov | nan | nan |
| **3814** | Puerto-Rico | Divorced | White | Machine-op-inspct | Other-relative | <=50K | Private | nan | nan |
| **2984** | Mexico | Married-civ-spouse | White | Exec-managerial | Husband | <=50K | Private | 29.000000 | 7.000000 |
| **3829** | Puerto-Rico | Divorced | White | Other-service | Unmarried | <=50K | Private | 41.000000 | 2.000000 |

Synthetic

| | native.country | marital.status | race | occupation | relationship | income | workclass | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **4516** | Puerto-Rico | Never-married | White | Adm-clerical | Own-child | <=50K | Self-emp-inc | 20 | 2 |
| **1015** | Columbia | Never-married | White | Protective-serv | Not-in-family | <=50K | State-gov | 32 | 1 |
| **3814** | Puerto-Rico | Divorced | White | Machine-op-inspct | Other-relative | <=50K | Private | 46 | 2 |
| **2984** | Mexico | Married-civ-spouse | White | Exec-managerial | Husband | <=50K | Private | 34 | 2 |
| **3829** | Puerto-Rico | Divorced | White | Other-service | Unmarried | <=50K | Private | 28 | 54 |

SQL for Real:
SELECT `native.country`,`marital.status`,race,occupation,relationship,income,workclass,MIN(age), COUNT(*) FROM C1 WHERE (education IN ('Some-college', 'Masters', '5th-6th', 'Preschool', '10th') OR `native.country` <> 'United-States') OR (`hours.per.week` < 12 OR workclass = 'Private') GROUP BY
`native.country`,`marital.status`,race,occupation,relationship,income,workclass

Resulted in 4888 records

SQL for Synthetic:
SELECT `native.country`,`marital.status`,race,occupation,relationship,income,workclass,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (education IN ('Some-college', 'Masters', '5th-6th', 'Preschool', '10th') OR `native.country` <> 'United-States') OR (`hours.per.week` < 12 OR workclass = 'Private') GROUP BY
`native.country`,`marital.status`,race,occupation,relationship,income,workclass

Resulted in 7015 records

Normalized Euclidean distance for (age): 42.01

Hellinger Distance: 0.327

==========

Real

| | relationship | income | native.country | education | occupation | sex | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **2158** | Not-in-family | <=50K | Columbia | Masters | Sales | Female | nan | nan |
| **5494** | Own-child | <=50K | United-States | 11th | Transport-moving | Female | 475582.000000 | 2.000000 |
| **1207** | Husband | <=50K | United-States | 11th | Sales | Male | 2825278.000000 | 19.000000 |
| **5501** | Own-child | <=50K | United-States | 12th | Craft-repair | Male | 2549041.000000 | 11.000000 |
| **7260** | Wife | <=50K | Puerto-Rico | Bachelors | Adm-clerical | Female | 26668.000000 | 1.000000 |

Synthetic

| | relationship | income | native.country | education | occupation | sex | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **2158** | Not-in-family | <=50K | Columbia | Masters | Sales | Female | 355358.739936 | 2 |
| **5494** | Own-child | <=50K | United-States | 11th | Transport-moving | Female | 159604.372382 | 1 |
| **1207** | Husband | <=50K | United-States | 11th | Sales | Male | 3897825.998871 | 22 |
| **5501** | Own-child | <=50K | United-States | 12th | Craft-repair | Male | 1874941.563083 | 10 |
| **7260** | Wife | <=50K | Puerto-Rico | Bachelors | Adm-clerical | Female | 2216541.075038 | 13 |

SQL for Real:
SELECT relationship,income,`native.country`,education,occupation,sex,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (education IN ('Doctorate', 'Some-college', '7th-8th', '9th', 'Masters', '5th-6th', 'HS-grad', 'Prof-school', 'Preschool', '11th') AND age = 45) OR (`native.country` LIKE 'United-States' OR income IN ('>50K', '<=50K')) GROUP BY relationship,income,`native.country`,education,occupation,sex

Resulted in 5296 records

SQL for Synthetic:
SELECT relationship,income,`native.country`,education,occupation,sex,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (education IN ('Doctorate', 'Some-college', '7th-8th', '9th', 'Masters', '5th-6th', 'HS-grad', 'Prof-school', 'Preschool', '11th') AND age = 45) OR (`native.country` LIKE 'United-States' OR income IN ('>50K', '<=50K')) GROUP BY relationship,income,`native.country`,education,occupation,sex

Resulted in 7620 records

Normalized Euclidean distance for (fnlwgt): 46.97

Hellinger Distance: 0.286

==========

Real

| | native.country | sex | income | workclass | marital.status | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **184** | England | Male | >50K | Federal-gov | Married-civ-spouse | 32528.000000 | 1 |
| **627** | United-States | Female | <=50K | State-gov | Widowed | 146221.846154 | 26 |
| **746** | United-States | Male | >50K | Self-emp-not-inc | Widowed | 169330.250000 | 4 |
| **719** | United-States | Male | >50K | Federal-gov | Never-married | 166191.727273 | 11 |
| **589** | United-States | Female | <=50K | ? | Widowed | 162419.707317 | 123 |

Synthetic

| | native.country | sex | income | workclass | marital.status | AVG(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **184** | England | Male | >50K | Federal-gov | Married-civ-spouse | nan | nan |
| **627** | United-States | Female | <=50K | State-gov | Widowed | 159898.082099 | 34.000000 |
| **746** | United-States | Male | >50K | Self-emp-not-inc | Widowed | nan | nan |
| **719** | United-States | Male | >50K | Federal-gov | Never-married | 179078.889857 | 7.000000 |
| **589** | United-States | Female | <=50K | ? | Widowed | 166630.970127 | 100.000000 |

SQL for Real:
SELECT `native.country`,sex,income,workclass,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (fnlwgt BETWEEN 83827 AND 31327) OR (relationship <> 'Not-in-family' AND age = 30) OR (`native.country` LIKE 'United-States' OR sex LIKE 'Male') AND (income IN ('>50K', '<=50K') AND `hours.per.week` <> 40) OR (`marital.status` LIKE 'Never-married' OR education IN ('11th', '5th-6th', 'Doctorate', '9th', 'Some-college', '10th', '7th-8th', 'Assoc-voc', 'Preschool', 'HS-grad', '1st-4th', '12th', 'Prof-school', 'Bachelors', 'Assoc-acdm', 'Masters')) AND (occupation LIKE 'Adm-clerical' AND capital BETWEEN 0 AND 0) GROUP BY `native.country`,sex,income,workclass,`marital.status`

Resulted in 769 records

SQL for Synthetic:
SELECT `native.country`,sex,income,workclass,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt BETWEEN 83827 AND 31327) OR (relationship <> 'Not-in-family' AND age = 30) OR (`native.country` LIKE 'United-States' OR sex LIKE 'Male') AND (income IN ('>50K', '<=50K') AND `hours.per.week` <> 40) OR (`marital.status` LIKE 'Never-married' OR education IN ('11th', '5th-6th', 'Doctorate', '9th', 'Some-college', '10th', '7th-8th', 'Assoc-voc', 'Preschool', 'HS-grad', '1st-4th', '12th', 'Prof-school', 'Bachelors', 'Assoc-acdm', 'Masters')) AND (occupation LIKE 'Adm-clerical' AND capital BETWEEN 0 AND 0) GROUP BY `native.country`,sex,income,workclass,`marital.status`

Resulted in 755 records

Normalized Euclidean distance for (fnlwgt): 18.84

Hellinger Distance: 0.353

==========

Real

| | marital.status | workclass | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|
| **46** | Widowed | State-gov | 420526 | 42 |
| **19** | Married-spouse-absent | ? | 558183 | 45 |
| **18** | Married-civ-spouse | State-gov | 1033222 | 464 |
| **33** | Separated | ? | 672412 | 97 |
| **37** | Separated | Self-emp-inc | 505365 | 25 |

Synthetic

| | marital.status | workclass | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|
| **46** | Widowed | State-gov | 203170.265382 | 58.000000 |
| **19** | Married-spouse-absent | ? | 192738.629280 | 33.000000 |
| **18** | Married-civ-spouse | State-gov | 1340619.957202 | 664.000000 |
| **33** | Separated | ? | 1338996.446401 | 85.000000 |
| **37** | Separated | Self-emp-inc | 188480.770664 | 20.000000 |

SQL for Real:
SELECT `marital.status`,workclass,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (sex = 'Male' AND education <> 'Some-college') AND (race <> 'White' AND `hours.per.week` <= 40) OR (`marital.status` = 'Separated' OR income = '<=50K') GROUP BY `marital.status`,workclass

Resulted in 47 records

SQL for Synthetic:
SELECT `marital.status`,workclass,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (sex = 'Male' AND education <> 'Some-college') AND (race <> 'White' AND `hours.per.week` <= 40) OR (`marital.status` = 'Separated' OR income = '<=50K') GROUP BY `marital.status`,workclass

Resulted in 47 records

Normalized Euclidean distance for (fnlwgt): 6.86

Hellinger Distance: 0.056

==========

Real

| | relationship | marital.status | race | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|
| **63** | Own-child | Separated | White | 43.833333 | 12.000000 |
| **37** | Other-relative | Married-spouse-absent | Black | nan | nan |
| **90** | Wife | Never-married | Other | nan | nan |
| **82** | Unmarried | Widowed | Asian-Pac-Islander | nan | nan |
| **79** | Unmarried | Never-married | White | 38.838588 | 793.000000 |

Synthetic

| | relationship | marital.status | race | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|
| **63** | Own-child | Separated | White | 39.975196 | 15 |
| **37** | Other-relative | Married-spouse-absent | Black | 40.027019 | 1 |
| **90** | Wife | Never-married | Other | 40.007400 | 1 |
| **82** | Unmarried | Widowed | Asian-Pac-Islander | 39.003188 | 1 |
| **79** | Unmarried | Never-married | White | 39.939168 | 890 |

SQL for Real:
SELECT relationship,`marital.status`,race,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (`marital.status` IN ('Married-civ-spouse', 'Never-married')) OR (occupation <> 'Other-service' AND education = 'Some-college') AND (workclass IN ('Self-emp-not-inc', 'Federal-gov', 'Local-gov', 'State-gov', '?', 'Private') AND income IN ('>50K', '<=50K')) AND (relationship IN ('Husband', 'Unmarried', 'Other-relative', 'Own-child', 'Not-in-family') AND sex LIKE 'Male') GROUP BY relationship,`marital.status`,race

Resulted in 89 records

SQL for Synthetic:
SELECT relationship,`marital.status`,race,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` IN ('Married-civ-spouse', 'Never-married')) OR (occupation <> 'Other-service' AND education = 'Some-college') AND (workclass IN ('Self-emp-not-inc', 'Federal-gov', 'Local-gov', 'State-gov', '?', 'Private') AND income IN ('>50K', '<=50K')) AND (relationship IN ('Husband', 'Unmarried', 'Other-relative', 'Own-child', 'Not-in-family') AND sex LIKE 'Male') GROUP BY relationship,`marital.status`,race

Resulted in 92 records

Normalized Euclidean distance for (`hours.per.week`): 8.6

Hellinger Distance: 0.052

==========

Real

| | education | relationship | income | workclass | occupation | sex | race | marital.status | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **186** | 12th | Husband | >50K | Self-emp-not-inc | Craft-repair | Male | White | Married-civ-spouse | 70 | 2 |
| **291** | 7th-8th | Husband | <=50K | Self-emp-not-inc | Farming-fishing | Male | White | Married-civ-spouse | 80 | 33 |
| **399** | Assoc-acdm | Husband | <=50K | Self-emp-inc | Prof-specialty | Male | White | Married-civ-spouse | 70 | 1 |
| **646** | Assoc-voc | Wife | >50K | Private | Other-service | Female | White | Married-civ-spouse | 20 | 1 |
| **57** | 10th | Husband | >50K | State-gov | Exec-managerial | Male | White | Married-civ-spouse | 40 | 1 |

Synthetic

| | education | relationship | income | workclass | occupation | sex | race | marital.status | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **186** | 12th | Husband | >50K | Self-emp-not-inc | Craft-repair | Male | White | Married-civ-spouse | nan | nan |
| **291** | 7th-8th | Husband | <=50K | Self-emp-not-inc | Farming-fishing | Male | White | Married-civ-spouse | 40.063029 | 25.000000 |
| **399** | Assoc-acdm | Husband | <=50K | Self-emp-inc | Prof-specialty | Male | White | Married-civ-spouse | 39.952743 | 1.000000 |
| **646** | Assoc-voc | Wife | >50K | Private | Other-service | Female | White | Married-civ-spouse | nan | nan |
| **57** | 10th | Husband | >50K | State-gov | Exec-managerial | Male | White | Married-civ-spouse | nan | nan |

SQL for Real:
SELECT education,relationship,income,workclass,occupation,sex,race,`marital.status`,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (workclass = 'Self-emp-inc' AND `hours.per.week` < 40) OR (`marital.status` LIKE 'Married-civ-spouse' OR occupation LIKE 'Craft-repair') AND (education IN ('Bachelors', '10th', 'Masters', 'Preschool', 'Assoc-acdm', 'Doctorate', 'Prof-school', '11th', 'HS-grad', '7th-8th', 'Some-college', 'Assoc-voc', '12th', '1st-4th') AND income LIKE '>50K') OR (sex = 'Male' OR capital = 0) AND (race = 'White' AND relationship = 'Husband') GROUP BY education,relationship,income,workclass,occupation,sex,race,`marital.status`

Resulted in 1813 records

SQL for Synthetic:
SELECT education,relationship,income,workclass,occupation,sex,race,`marital.status`,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (workclass = 'Self-emp-inc' AND `hours.per.week` < 40) OR (`marital.status` LIKE 'Married-civ-spouse' OR occupation LIKE 'Craft-repair') AND (education IN ('Bachelors', '10th', 'Masters', 'Preschool', 'Assoc-acdm', 'Doctorate', 'Prof-school', '11th', 'HS-grad', '7th-8th', 'Some-college', 'Assoc-voc', '12th', '1st-4th') AND income LIKE '>50K') OR (sex = 'Male' OR capital = 0) AND (race = 'White' AND relationship = 'Husband') GROUP BY education,relationship,income,workclass,occupation,sex,race,`marital.status`

Resulted in 1624 records

Normalized Euclidean distance for (`hours.per.week`): 32.19

Hellinger Distance: 0.243

==========

Real

| | sex | workclass | native.country | income | relationship | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **75** | Female | Local-gov | Vietnam | <=50K | Not-in-family | nan | nan |
| **145** | Female | Private | Puerto-Rico | >50K | Not-in-family | nan | nan |
| **288** | Male | Local-gov | United-States | <=50K | Husband | 19820009.000000 | 106.000000 |
| **245** | Male | Federal-gov | Puerto-Rico | <=50K | Not-in-family | nan | nan |
| **381** | Male | Self-emp-inc | ? | >50K | Husband | 99185.000000 | 1.000000 |

Synthetic

| | sex | workclass | native.country | income | relationship | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **75** | Female | Local-gov | Vietnam | <=50K | Not-in-family | 198931.530547 | 1 |
| **145** | Female | Private | Puerto-Rico | >50K | Not-in-family | 704621.054045 | 4 |
| **288** | Male | Local-gov | United-States | <=50K | Husband | 7658754.049169 | 42 |
| **245** | Male | Federal-gov | Puerto-Rico | <=50K | Not-in-family | 694968.112070 | 4 |
| **381** | Male | Self-emp-inc | ? | >50K | Husband | 336802.291206 | 2 |

SQL for Real:
SELECT sex,workclass,`native.country`,income,relationship,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (workclass = 'Private' AND `native.country` LIKE 'United-States') AND (`hours.per.week` > 38 AND `marital.status` LIKE 'Never-married') AND (race <> 'White' AND education LIKE 'Some-college') OR (income IN ('<=50K', '>50K') OR fnlwgt = 143766) AND (occupation = 'Prof-specialty' AND capital <= 0) GROUP BY sex,workclass,`native.country`,income,relationship

Resulted in 420 records

SQL for Synthetic:
SELECT sex,workclass,`native.country`,income,relationship,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (workclass = 'Private' AND `native.country` LIKE 'United-States') AND (`hours.per.week` > 38 AND `marital.status` LIKE 'Never-married') AND (race <> 'White' AND education LIKE 'Some-college') OR (income IN ('<=50K', '>50K') OR fnlwgt = 143766) AND (occupation = 'Prof-specialty' AND capital <= 0) GROUP BY sex,workclass,`native.country`,income,relationship

Resulted in 454 records

Normalized Euclidean distance for (fnlwgt): 12.73

Hellinger Distance: 0.205

==========

Real

| | race | occupation | workclass | education | native.country | relationship | income | sex | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **11067** | White | Prof-specialty | Private | Prof-school | United-States | Unmarried | <=50K | Male | 43.500000 | 2.000000 |
| **8211** | White | Handlers-cleaners | Private | Assoc-acdm | Outlying-US(Guam-USVI-etc) | Husband | <=50K | Male | nan | nan |
| **9563** | White | Other-service | Private | Bachelors | United-States | Husband | >50K | Male | 40.285714 | 7.000000 |
| **8415** | White | Handlers-cleaners | Private | Some-college | Thailand | Own-child | <=50K | Male | nan | nan |
| **5393** | White | Craft-repair | Local-gov | HS-grad | United-States | Not-in-family | <=50K | Male | 43.800000 | 10.000000 |

Synthetic

| | race | occupation | workclass | education | native.country | relationship | income | sex | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **11067** | White | Prof-specialty | Private | Prof-school | United-States | Unmarried | <=50K | Male | 28.000000 | 1 |
| **8211** | White | Handlers-cleaners | Private | Assoc-acdm | Outlying-US(Guam-USVI-etc) | Husband | <=50K | Male | 24.000000 | 1 |
| **9563** | White | Other-service | Private | Bachelors | United-States | Husband | >50K | Male | 37.000000 | 3 |
| **8415** | White | Handlers-cleaners | Private | Some-college | Thailand | Own-child | <=50K | Male | 18.000000 | 2 |
| **5393** | White | Craft-repair | Local-gov | HS-grad | United-States | Not-in-family | <=50K | Male | 35.000000 | 7 |

SQL for Real:
SELECT race,occupation,workclass,education,`native.country`,relationship,income,sex,AVG(age), COUNT(*) FROM C1 WHERE (`native.country` LIKE 'United-States') OR (sex = 'Male' AND `hours.per.week` <= 40) OR (race LIKE 'White' OR age BETWEEN 56 AND 41) OR (relationship <> 'Not-in-family' OR occupation <> 'Craft-repair') OR (`marital.status` = 'Divorced' OR education = 'Assoc-voc') OR (income IN ('<=50K', '>50K') OR fnlwgt > 160431) GROUP BY race,occupation,workclass,education,`native.country`,relationship,income,sex

Resulted in 9992 records

SQL for Synthetic:
SELECT race,occupation,workclass,education,`native.country`,relationship,income,sex,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (`native.country` LIKE 'United-States') OR (sex = 'Male' AND `hours.per.week` <= 40) OR (race LIKE 'White' OR age BETWEEN 56 AND 41) OR (relationship <> 'Not-in-family' OR occupation <> 'Craft-repair') OR (`marital.status` = 'Divorced' OR education = 'Assoc-voc') OR (income IN ('<=50K', '>50K') OR fnlwgt > 160431) GROUP BY race,occupation,workclass,education,`native.country`,relationship,income,sex

Resulted in 13906 records

Normalized Euclidean distance for (age): 60.46

Hellinger Distance: 0.257

==========

Real

| | race | education | native.country | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|
| **143** | Asian-Pac-Islander | Prof-school | India | 59.000000 | 17.000000 |
| **247** | Black | 9th | United-States | 81.000000 | 84.000000 |
| **135** | Asian-Pac-Islander | Masters | Philippines | 59.000000 | 10.000000 |
| **295** | Black | Bachelors | Yugoslavia | nan | nan |
| **701** | White | Bachelors | Nicaragua | 46.000000 | 3.000000 |

Synthetic

| | race | education | native.country | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|
| **143** | Asian-Pac-Islander | Prof-school | India | 45 | 3 |
| **247** | Black | 9th | United-States | 79 | 34 |
| **135** | Asian-Pac-Islander | Masters | Philippines | 44 | 2 |
| **295** | Black | Bachelors | Yugoslavia | 40 | 1 |
| **701** | White | Bachelors | Nicaragua | 22 | 6 |

SQL for Real:
SELECT race,education,`native.country`,MAX(age), COUNT(*) FROM C1 WHERE (sex <> 'Male') OR (race IN ('White', 'Black', 'Other', 'Asian-Pac-Islander') AND workclass IN ('Private', 'State-gov', 'Self-emp-inc', 'Federal-gov')) OR (fnlwgt >= 201292 OR `hours.per.week` BETWEEN 40 AND 48) AND (age <= 31 AND education = '11th') AND (income = '>50K' AND capital = 0) GROUP BY race,education,`native.country`

Resulted in 736 records

SQL for Synthetic:
SELECT race,education,`native.country`,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (sex <> 'Male') OR (race IN ('White', 'Black', 'Other', 'Asian-Pac-Islander') AND workclass IN ('Private', 'State-gov', 'Self-emp-inc', 'Federal-gov')) OR (fnlwgt >= 201292 OR `hours.per.week` BETWEEN 40 AND 48) AND (age <= 31 AND education = '11th') AND (income = '>50K' AND capital = 0) GROUP BY race,education,`native.country`

Resulted in 859 records

Normalized Euclidean distance for (age): 20.83

Hellinger Distance: 0.383

==========

Real

| occupation | marital.status | native.country | workclass | race | sex | education | income | AVG(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|

Synthetic

| occupation | marital.status | native.country | workclass | race | sex | education | income | AVG(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|

SQL for Real:
SELECT occupation,`marital.status`,`native.country`,workclass,race,sex,education,income,AVG(capital), COUNT(*) FROM C1 WHERE (`hours.per.week` < 60 OR age BETWEEN 17 AND 35) AND ((fnlwgt BETWEEN 184176 AND 129513) AND workclass IN ('Self-emp-inc', '?', 'State-gov', 'Private')) AND (race IN ('Other', 'Black') AND relationship <> 'Husband') AND (occupation IN ('Other-service', 'Machine-op-inspct', 'Priv-house-serv', 'Protective-serv', 'Craft-repair', 'Adm-clerical', 'Handlers-cleaners', 'Transport-moving') AND `native.country` LIKE 'United-States') AND (income LIKE '<=50K' AND sex IN ('Male', 'Female')) GROUP BY occupation,`marital.status`,`native.country`,workclass,race,sex,education,income

Resulted in 0 records

SQL for Synthetic:
SELECT occupation,`marital.status`,`native.country`,workclass,race,sex,education,income,AVG(capital), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` < 60 OR age BETWEEN 17 AND 35) AND ((fnlwgt BETWEEN 184176 AND 129513) AND workclass IN ('Self-emp-inc', '?', 'State-gov', 'Private')) AND (race IN ('Other', 'Black') AND relationship <> 'Husband') AND (occupation IN ('Other-service', 'Machine-op-inspct', 'Priv-house-serv', 'Protective-serv', 'Craft-repair', 'Adm-clerical', 'Handlers-cleaners', 'Transport-moving') AND `native.country` LIKE 'United-States') AND (income LIKE '<=50K' AND sex IN ('Male', 'Female')) GROUP BY occupation,`marital.status`,`native.country`,workclass,race,sex,education,income

Resulted in 0 records

Normalized Euclidean distance for (capital): nan

Hellinger Distance: nan

==========

Real

| | marital.status | occupation | SUM(capital) | COUNT(*) |
|---|---|---|---|---|
| **26** | Married-civ-spouse | Other-service | 324186.000000 | 174.000000 |
| **83** | Widowed | Priv-house-serv | 25236.000000 | 1.000000 |
| **37** | Married-spouse-absent | Farming-fishing | 34426.000000 | 3.000000 |
| **67** | Separated | Machine-op-inspct | 6356.000000 | 2.000000 |
| **33** | Married-spouse-absent | ? | 0.000000 | 1.000000 |

Synthetic

| | marital.status | occupation | SUM(capital) | COUNT(*) |
|---|---|---|---|---|
| **26** | Married-civ-spouse | Other-service | 11189.046326 | 221 |
| **83** | Widowed | Priv-house-serv | -12.830576 | 1 |
| **37** | Married-spouse-absent | Farming-fishing | 440.615619 | 7 |
| **67** | Separated | Machine-op-inspct | -110.996995 | 4 |
| **33** | Married-spouse-absent | ? | 854.122179 | 13 |

SQL for Real:
SELECT `marital.status`,occupation,SUM(capital), COUNT(*) FROM C1 WHERE (workclass <> 'Private' AND capital > 0) OR (fnlwgt = 88781 OR age BETWEEN 20 AND 23) AND (education LIKE 'Masters' AND sex = 'Female') AND (`hours.per.week` = 40 AND `native.country` = 'United-States') OR (`marital.status` LIKE 'Never-married' OR income LIKE '>50K') GROUP BY `marital.status`,occupation

Resulted in 88 records

SQL for Synthetic:
SELECT `marital.status`,occupation,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (workclass <> 'Private' AND capital > 0) OR (fnlwgt = 88781 OR age BETWEEN 20 AND 23) AND (education LIKE 'Masters' AND sex = 'Female') AND (`hours.per.week` = 40 AND `native.country` = 'United-States') OR (`marital.status` LIKE 'Never-married' OR income LIKE '>50K') GROUP BY `marital.status`,occupation

Resulted in 89 records

Normalized Euclidean distance for (capital): 9.11

Hellinger Distance: 0.105

==========

Real

| | education | income | MIN(capital) | COUNT(*) |
|---|---|---|---|---|
| **28** | Prof-school | <=50K | -2339 | 216 |
| **30** | Some-college | <=50K | -4356 | 8249 |
| **8** | 5th-6th | <=50K | -2603 | 462 |
| **18** | Bachelors | <=50K | -3770 | 4580 |
| **15** | Assoc-acdm | >50K | -2824 | 391 |

Synthetic

| | education | income | MIN(capital) | COUNT(*) |
|---|---|---|---|---|
| **28** | Prof-school | <=50K | -313.886797 | 566.000000 |
| **30** | Some-college | <=50K | -391.476155 | 8639.000000 |
| **8** | 5th-6th | <=50K | -316.679482 | 288.000000 |
| **18** | Bachelors | <=50K | -370.806040 | 6560.000000 |
| **15** | Assoc-acdm | >50K | -244.110975 | 175.000000 |

SQL for Real:
SELECT education,income,MIN(capital), COUNT(*) FROM C1 WHERE (age > 45) AND (`native.country` <> 'United-States' OR `marital.status` IN ('Married-spouse-absent', 'Separated', 'Divorced', 'Married-civ-spouse', 'Widowed', 'Married-AF-spouse')) AND ((`hours.per.week` BETWEEN 40 AND 40) AND relationship = 'Husband') OR (occupation <> 'Craft-repair' OR fnlwgt >= 216858) GROUP BY education,income

Resulted in 32 records

SQL for Synthetic:
SELECT education,income,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (age > 45) AND (`native.country` <> 'United-States' OR `marital.status` IN ('Married-spouse-absent', 'Separated', 'Divorced', 'Married-civ-spouse', 'Widowed', 'Married-AF-spouse')) AND ((`hours.per.week` BETWEEN 40 AND 40) AND relationship = 'Husband') OR (occupation <> 'Craft-repair' OR fnlwgt >= 216858) GROUP BY education,income

Resulted in 31 records

Normalized Euclidean distance for (capital): 5.57

Hellinger Distance: 0.153

==========

Real

| | education | income | workclass | sex | occupation | race | relationship | native.country | marital.status | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **8163** | HS-grad | <=50K | Local-gov | Male | Craft-repair | White | Husband | United-States | Married-civ-spouse | 5013.000000 | 48.000000 |
| **8883** | HS-grad | <=50K | Private | Female | Other-service | Amer-Indian-Eskimo | Unmarried | Puerto-Rico | Married-civ-spouse | nan | nan |
| **9257** | HS-grad | <=50K | Private | Female | Sales | White | Not-in-family | Puerto-Rico | Married-spouse-absent | nan | nan |
| **4848** | Assoc-voc | >50K | Private | Female | Prof-specialty | White | Husband | Columbia | Married-civ-spouse | nan | nan |
| **10287** | HS-grad | <=50K | Private | Male | Sales | White | Not-in-family | Holand-Netherlands | Never-married | nan | nan |

Synthetic

| | education | income | workclass | sex | occupation | race | relationship | native.country | marital.status | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **8163** | HS-grad | <=50K | Local-gov | Male | Craft-repair | White | Husband | United-States | Married-civ-spouse | 174.915746 | 27 |
| **8883** | HS-grad | <=50K | Private | Female | Other-service | Amer-Indian-Eskimo | Unmarried | Puerto-Rico | Married-civ-spouse | -84.613600 | 1 |
| **9257** | HS-grad | <=50K | Private | Female | Sales | White | Not-in-family | Puerto-Rico | Married-spouse-absent | 46.518664 | 2 |
| **4848** | Assoc-voc | >50K | Private | Female | Prof-specialty | White | Husband | Columbia | Married-civ-spouse | -31.772576 | 1 |
| **10287** | HS-grad | <=50K | Private | Male | Sales | White | Not-in-family | Holand-Netherlands | Never-married | -101.943232 | 1 |

SQL for Real:
SELECT education,income,workclass,sex,occupation,race,relationship,`native.country`,`marital.status`,MAX(capital), COUNT(*) FROM C1 WHERE (workclass <> 'Private') AND (age <= 49 OR `hours.per.week` BETWEEN 75 AND 40) AND ((fnlwgt BETWEEN 422013 AND 194537) AND sex <> 'Female') OR (`native.country` IN ('?', 'Vietnam', 'Portugal', 'Cambodia', 'Italy', 'Hong', 'France', 'Peru', 'South', 'Thailand', 'Laos', 'Outlying-US(Guam-USVI-etc)', 'Ecuador', 'Honduras', 'Jamaica', 'Columbia', 'Japan', 'Cuba', 'Puerto-Rico', 'Mexico', 'Hungary', 'Trinadad&Tobago', 'United-States', 'Taiwan', 'Philippines', 'Nicaragua', 'Holand-Netherlands', 'Scotland') OR education = '10th') GROUP BY education,income,workclass,sex,occupation,race,relationship,`native.country`,`marital.status`

Resulted in 11409 records

SQL for Synthetic:
SELECT education,income,workclass,sex,occupation,race,relationship,`native.country`,`marital.status`,MAX(capital), COUNT(*) FROM C1_syn_06 WHERE (workclass <> 'Private') AND (age <= 49 OR `hours.per.week` BETWEEN 75 AND 40) AND ((fnlwgt BETWEEN 422013 AND 194537) AND sex <> 'Female') OR (`native.country` IN ('?', 'Vietnam', 'Portugal', 'Cambodia', 'Italy', 'Hong', 'France', 'Peru', 'South', 'Thailand', 'Laos', 'Outlying-US(Guam-USVI-etc)', 'Ecuador', 'Honduras', 'Jamaica', 'Columbia', 'Japan', 'Cuba', 'Puerto-Rico', 'Mexico', 'Hungary', 'Trinadad&Tobago', 'United-States', 'Taiwan', 'Philippines', 'Nicaragua', 'Holand-Netherlands', 'Scotland') OR education = '10th') GROUP BY education,income,workclass,sex,occupation,race,relationship,`native.country`,`marital.status`

Resulted in 15947 records

Normalized Euclidean distance for (capital): 63.98

Hellinger Distance: 0.252

==========

Real

| | workclass | race | relationship | sex | native.country | occupation | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **1066** | Self-emp-not-inc | White | Own-child | Male | Puerto-Rico | Sales | nan | nan |
| **1022** | Self-emp-not-inc | White | Own-child | Female | Puerto-Rico | Farming-fishing | nan | nan |
| **57** | Local-gov | Black | Husband | Male | Scotland | Transport-moving | nan | nan |
| **129** | Local-gov | Black | Other-relative | Male | United-States | Handlers-cleaners | 661358.000000 | 2.000000 |
| **736** | Self-emp-not-inc | Black | Husband | Male | Puerto-Rico | Transport-moving | nan | nan |

Synthetic

| | workclass | race | relationship | sex | native.country | occupation | SUM(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **1066** | Self-emp-not-inc | White | Own-child | Male | Puerto-Rico | Sales | 692268.079644 | 4 |
| **1022** | Self-emp-not-inc | White | Own-child | Female | Puerto-Rico | Farming-fishing | 676299.984980 | 4 |
| **57** | Local-gov | Black | Husband | Male | Scotland | Transport-moving | 368579.302095 | 2 |
| **129** | Local-gov | Black | Other-relative | Male | United-States | Handlers-cleaners | 185753.004498 | 1 |
| **736** | Self-emp-not-inc | Black | Husband | Male | Puerto-Rico | Transport-moving | 175736.887313 | 1 |

SQL for Real:
SELECT workclass,race,relationship,sex,`native.country`,occupation,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (workclass IN ('Self-emp-not-inc', 'Local-gov')) AND (race IN ('Black', 'Asian-Pac-Islander', 'White') OR age < 21) AND (income LIKE '<=50K' AND relationship IN ('Other-relative', 'Unmarried', 'Own-child', 'Husband', 'Not-in-family', 'Wife')) OR (capital = 0 OR `marital.status` IN ('Married-AF-spouse', 'Never-married', 'Married-civ-spouse', 'Separated', 'Divorced', 'Married-spouse-absent', 'Widowed')) AND (occupation = 'Prof-specialty' AND fnlwgt = 211695) AND (`hours.per.week` = 40 AND education <> 'HS-grad') GROUP BY workclass,race,relationship,sex,`native.country`,occupation

Resulted in 623 records

SQL for Synthetic:
SELECT workclass,race,relationship,sex,`native.country`,occupation,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (workclass IN ('Self-emp-not-inc', 'Local-gov')) AND (race IN ('Black', 'Asian-Pac-Islander', 'White') OR age < 21) AND (income LIKE '<=50K' AND relationship IN ('Other-relative', 'Unmarried', 'Own-child', 'Husband', 'Not-in-family', 'Wife')) OR (capital = 0 OR `marital.status` IN ('Married-AF-spouse', 'Never-married', 'Married-civ-spouse', 'Separated', 'Divorced', 'Married-spouse-absent', 'Widowed')) AND (occupation = 'Prof-specialty' AND fnlwgt = 211695) AND (`hours.per.week` = 40 AND education <> 'HS-grad') GROUP BY workclass,race,relationship,sex,`native.country`,occupation

Resulted in 1177 records

Normalized Euclidean distance for (fnlwgt): 17.83

Hellinger Distance: 0.214

==========

Real

| | workclass | race | income | marital.status | education | occupation | relationship | sex | native.country | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **13570** | Self-emp-not-inc | White | <=50K | Married-civ-spouse | Bachelors | Exec-managerial | Husband | Male | Mexico | nan | nan |
| **1549** | Federal-gov | White | <=50K | Separated | Some-college | Adm-clerical | Not-in-family | Female | United-States | nan | nan |
| **4175** | Private | Black | <=50K | Married-civ-spouse | 9th | Machine-op-inspct | Wife | Female | Hungary | nan | nan |
| **3786** | Private | Asian-Pac-Islander | <=50K | Widowed | Assoc-voc | Exec-managerial | Unmarried | Female | United-States | nan | nan |
| **7495** | Private | White | <=50K | Married-civ-spouse | Bachelors | Prof-specialty | Not-in-family | Male | United-States | 60.000000 | 1.000000 |

Synthetic

| | workclass | race | income | marital.status | education | occupation | relationship | sex | native.country | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **13570** | Self-emp-not-inc | White | <=50K | Married-civ-spouse | Bachelors | Exec-managerial | Husband | Male | Mexico | 40.366009 | 1 |
| **1549** | Federal-gov | White | <=50K | Separated | Some-college | Adm-clerical | Not-in-family | Female | United-States | 40.012611 | 1 |
| **4175** | Private | Black | <=50K | Married-civ-spouse | 9th | Machine-op-inspct | Wife | Female | Hungary | 39.998754 | 1 |
| **3786** | Private | Asian-Pac-Islander | <=50K | Widowed | Assoc-voc | Exec-managerial | Unmarried | Female | United-States | 39.956478 | 1 |
| **7495** | Private | White | <=50K | Married-civ-spouse | Bachelors | Prof-specialty | Not-in-family | Male | United-States | 40.180498 | 1 |

SQL for Real:
SELECT workclass,race,income,`marital.status`,education,occupation,relationship,sex,`native.country`,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (income = '<=50K') AND (`hours.per.week` > 40 AND age = 38) OR (`marital.status` = 'Married-civ-spouse' OR fnlwgt > 247515) OR (sex = 'Male' OR education <> 'HS-grad') OR (occupation = 'Transport-moving' OR capital BETWEEN 0 AND 0) GROUP BY workclass,race,income,`marital.status`,education,occupation,relationship,sex,`native.country`

Resulted in 12632 records

SQL for Synthetic:
SELECT workclass,race,income,`marital.status`,education,occupation,relationship,sex,`native.country`,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (income = '<=50K') AND (`hours.per.week` > 40 AND age = 38) OR (`marital.status` = 'Married-civ-spouse' OR fnlwgt > 247515) OR (sex = 'Male' OR education <> 'HS-grad') OR (occupation = 'Transport-moving' OR capital BETWEEN 0 AND 0) GROUP BY workclass,race,income,`marital.status`,education,occupation,relationship,sex,`native.country`

Resulted in 15508 records

Normalized Euclidean distance for (`hours.per.week`): 61.82

Hellinger Distance: 0.269

==========

Real

| | education | relationship | income | workclass | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|
| **544** | HS-grad | Own-child | >50K | Local-gov | 40 | 1 |
| **287** | Assoc-acdm | Own-child | <=50K | Federal-gov | 40 | 2 |
| **47** | 11th | Husband | <=50K | Private | 15 | 317 |
| **756** | Some-college | Unmarried | <=50K | ? | 4 | 49 |
| **764** | Some-college | Unmarried | >50K | Federal-gov | 40 | 2 |

Synthetic

| | education | relationship | income | workclass | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|
| **544** | HS-grad | Own-child | >50K | Local-gov | nan | nan |
| **287** | Assoc-acdm | Own-child | <=50K | Federal-gov | 39.554523 | 2.000000 |
| **47** | 11th | Husband | <=50K | Private | 39.199172 | 331.000000 |
| **756** | Some-college | Unmarried | <=50K | ? | 39.005285 | 62.000000 |
| **764** | Some-college | Unmarried | >50K | Federal-gov | nan | nan |

SQL for Real:
SELECT education,relationship,income,workclass,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (sex = 'Female') OR (occupation <> 'Craft-repair' OR relationship IN ('Husband', 'Wife', 'Other-relative', 'Not-in-family', 'Unmarried', 'Own-child')) AND (workclass <> 'Private' AND capital < 0) AND (age < 42 AND fnlwgt <> 231482) AND (`hours.per.week` <= 20 AND race <> 'White') OR (education LIKE 'Bachelors' OR `marital.status` LIKE 'Married-civ-spouse') GROUP BY education,relationship,income,workclass

Resulted in 782 records

SQL for Synthetic:
SELECT education,relationship,income,workclass,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (sex = 'Female') OR (occupation <> 'Craft-repair' OR relationship IN ('Husband', 'Wife', 'Other-relative', 'Not-in-family', 'Unmarried', 'Own-child')) AND (workclass <> 'Private' AND capital < 0) AND (age < 42 AND fnlwgt <>

231482) AND (`hours.per.week` <= 20 AND race <> 'White') OR (education LIKE 'Bachelors' OR `marital.status` LIKE 'Married-civ-spouse') GROUP BY education,relationship,income,workclass

Resulted in 725 records

Normalized Euclidean distance for (`hours.per.week`): 25.38

Hellinger Distance: 0.183

==========

Real

| | education | sex | workclass | relationship | native.country | race | marital.status | occupation | income | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **96** | Doctorate | Male | Private | Husband | Puerto-Rico | Asian-Pac-Islander | Married-civ-spouse | Craft-repair | <=50K | nan | nan |
| **62** | Bachelors | Male | Private | Husband | Puerto-Rico | Asian-Pac-Islander | Married-civ-spouse | Adm-clerical | >50K | nan | nan |
| **87** | Bachelors | Male | Self-emp-not-inc | Husband | Puerto-Rico | Asian-Pac-Islander | Married-civ-spouse | Exec-managerial | <=50K | nan | nan |
| **69** | Bachelors | Male | Private | Husband | Puerto-Rico | Asian-Pac-Islander | Married-civ-spouse | Prof-specialty | <=50K | nan | nan |
| **39** | Bachelors | Female | Private | Not-in-family | Puerto-Rico | Asian-Pac-Islander | Married-spouse-absent | Sales | <=50K | nan | nan |

Synthetic

| | education | sex | workclass | relationship | native.country | race | marital.status | occupation | income | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **96** | Doctorate | Male | Private | Husband | Puerto-Rico | Asian-Pac-Islander | Married-civ-spouse | Craft-repair | <=50K | 40.053984 | 1 |
| **62** | Bachelors | Male | Private | Husband | Puerto-Rico | Asian-Pac-Islander | Married-civ-spouse | Adm-clerical | >50K | 39.971058 | 1 |
| **87** | Bachelors | Male | Self-emp-not-inc | Husband | Puerto-Rico | Asian-Pac-Islander | Married-civ-spouse | Exec-managerial | <=50K | 40.359326 | 1 |
| **69** | Bachelors | Male | Private | Husband | Puerto-Rico | Asian-Pac-Islander | Married-civ-spouse | Prof-specialty | <=50K | 39.965902 | 1 |
| **39** | Bachelors | Female | Private | Not-in-family | Puerto-Rico | Asian-Pac-Islander | Married-spouse-absent | Sales | <=50K | 40.018025 | 1 |

SQL for Real:
SELECT education,sex,workclass,relationship,`native.country`,race,`marital.status`,occupation,income,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (race IN ('Asian-Pac-Islander', 'Amer-Indian-Eskimo') AND `hours.per.week` <= 50) AND (income IN ('<=50K', '>50K') AND `native.country` = 'Puerto-Rico') GROUP BY education,sex,workclass,relationship,`native.country`,race,`marital.status`,occupation,income

Resulted in 1 records

SQL for Synthetic:
SELECT education,sex,workclass,relationship,`native.country`,race,`marital.status`,occupation,income,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (race IN ('Asian-Pac-Islander', 'Amer-Indian-Eskimo') AND `hours.per.week` <= 50) AND (income IN ('<=50K', '>50K') AND `native.country` = 'Puerto-Rico') GROUP BY education,sex,workclass,relationship,`native.country`,race,`marital.status`,occupation,income

Resulted in 208 records

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

Hellinger Distance: nan

==========

Real

| | income | marital.status | race | sex | education | workclass | relationship | occupation | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **28** | <=50K | Divorced | Amer-Indian-Eskimo | Male | HS-grad | Local-gov | Not-in-family | Transport-moving | 40.000000 | 1 |
| **2366** | <=50K | Never-married | White | Female | Assoc-voc | ? | Own-child | ? | 37.333333 | 3 |
| **1285** | <=50K | Married-civ-spouse | White | Male | 11th | Self-emp-not-inc | Husband | Other-service | 46.666667 | 3 |
| **2511** | <=50K | Never-married | White | Female | HS-grad | State-gov | Unmarried | Adm-clerical | 40.000000 | 1 |
| **2046** | <=50K | Never-married | Black | Female | HS-grad | Federal-gov | Own-child | Craft-repair | 40.000000 | 1 |

Synthetic

| | income | marital.status | race | sex | education | workclass | relationship | occupation | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **28** | <=50K | Divorced | Amer-Indian-Eskimo | Male | HS-grad | Local-gov | Not-in-family | Transport-moving | nan | nan |
| **2366** | <=50K | Never-married | White | Female | Assoc-voc | ? | Own-child | ? | 40.012631 | 3.000000 |
| **1285** | <=50K | Married-civ-spouse | White | Male | 11th | Self-emp-not-inc | Husband | Other-service | 39.613293 | 2.000000 |
| **2511** | <=50K | Never-married | White | Female | HS-grad | State-gov | Unmarried | Adm-clerical | 39.981678 | 3.000000 |
| **2046** | <=50K | Never-married | Black | Female | HS-grad | Federal-gov | Own-child | Craft-repair | nan | nan |

SQL for Real:
SELECT income,`marital.status`,race,sex,education,workclass,relationship,occupation,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (workclass <> 'Private') GROUP BY income,`marital.status`,race,sex,education,workclass,relationship,occupation

Resulted in 4796 records

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

Resulted in 4581 records

Normalized Euclidean distance for (`hours.per.week`): 46.25

Hellinger Distance: 0.229

==========

Real

| | sex | relationship | race | occupation | native.country | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **2957** | Male | Own-child | Black | Craft-repair | Hungary | nan | nan |
| **731** | Female | Own-child | Black | Handlers-cleaners | United-States | 427541.000000 | 12.000000 |
| **2623** | Male | Not-in-family | White | Other-service | Thailand | nan | nan |
| **2999** | Male | Own-child | Black | Protective-serv | Hungary | nan | nan |
| **856** | Female | Own-child | White | Armed-Forces | United-States | nan | nan |

Synthetic

| | sex | relationship | race | occupation | native.country | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **2957** | Male | Own-child | Black | Craft-repair | Hungary | 180580.328757 | 1 |
| **731** | Female | Own-child | Black | Handlers-cleaners | United-States | 1367980.872487 | 13 |
| **2623** | Male | Not-in-family | White | Other-service | Thailand | 178325.603589 | 2 |
| **2999** | Male | Own-child | Black | Protective-serv | Hungary | 181751.586615 | 1 |
| **856** | Female | Own-child | White | Armed-Forces | United-States | 197247.474259 | 6 |

SQL for Real:
SELECT sex,relationship,race,occupation,`native.country`,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (occupation <> 'Craft-repair') OR (`hours.per.week` < 40 OR fnlwgt >= 164526) AND (sex = 'Male' AND relationship <> 'Own-child') OR (`marital.status` IN ('Separated', 'Divorced', 'Married-AF-spouse', 'Married-civ-spouse', 'Married-spouse-absent', 'Never-married') OR capital > 0) GROUP BY sex,relationship,race,occupation,`native.country`

Resulted in 2564 records

SQL for Synthetic:
SELECT sex,relationship,race,occupation,`native.country`,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (occupation <> 'Craft-repair') OR (`hours.per.week` < 40 OR fnlwgt >= 164526) AND (sex = 'Male' AND relationship <> 'Own-child') OR (`marital.status` IN ('Separated', 'Divorced', 'Married-AF-spouse', 'Married-civ-spouse', 'Married-spouse-absent', 'Never-married') OR capital > 0) GROUP BY sex,relationship,race,occupation,`native.country`

Resulted in 3455 records

Normalized Euclidean distance for (fnlwgt): 34.53

Hellinger Distance: 0.345

==========

Real

| | workclass | marital.status | sex | income | native.country | education | relationship | occupation | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **9700** | Private | Separated | Male | <=50K | Puerto-Rico | HS-grad | Own-child | Sales | nan | nan |
| **12161** | Self-emp-not-inc | Never-married | Male | <=50K | Puerto-Rico | Some-college | Own-child | Other-service | nan | nan |
| **8695** | Private | Never-married | Male | <=50K | Puerto-Rico | Masters | Not-in-family | Prof-specialty | nan | nan |
| **460** | ? | Never-married | Female | <=50K | United-States | Assoc-acdm | Own-child | ? | 242736.000000 | 7.000000 |
| **13310** | State-gov | Never-married | Male | <=50K | United-States | Some-college | Own-child | Sales | nan | nan |

Synthetic

| | workclass | marital.status | sex | income | native.country | education | relationship | occupation | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **9700** | Private | Separated | Male | <=50K | Puerto-Rico | HS-grad | Own-child | Sales | 177005.043649 | 1 |
| **12161** | Self-emp-not-inc | Never-married | Male | <=50K | Puerto-Rico | Some-college | Own-child | Other-service | 169028.998155 | 1 |
| **8695** | Private | Never-married | Male | <=50K | Puerto-Rico | Masters | Not-in-family | Prof-specialty | 177079.533338 | 12 |
| **460** | ? | Never-married | Female | <=50K | United-States | Assoc-acdm | Own-child | ? | 196119.959717 | 11 |
| **13310** | State-gov | Never-married | Male | <=50K | United-States | Some-college | Own-child | Sales | 150659.388183 | 1 |

SQL for Real:
SELECT workclass,`marital.status`,sex,income,`native.country`,education,relationship,occupation,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (`native.country` IN ('Trinadad&Tobago', 'Columbia', 'France', 'Yugoslavia', 'Vietnam', 'Guatemala', 'Greece', 'Outlying-US(Guam-USVI-etc)', 'Iran', 'Germany', 'Puerto-Rico', 'England', 'Japan', 'Cambodia', 'Canada', 'India', 'Hong', 'Laos', 'Portugal', 'Hungary', 'Nicaragua', 'Ireland', 'Jamaica', 'Mexico', 'Honduras', 'Peru', 'Poland', '?', 'United-States', 'El-Salvador')) GROUP BY workclass,`marital.status`,sex,income,`native.country`,education,relationship,occupation

Resulted in 9427 records

SQL for Synthetic:
SELECT workclass,`marital.status`,sex,income,`native.country`,education,relationship,occupation,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`native.country` IN ('Trinadad&Tobago', 'Columbia', 'France', 'Yugoslavia', 'Vietnam', 'Guatemala', 'Greece', 'Outlying-US(Guam-USVI-etc)', 'Iran', 'Germany', 'Puerto-Rico', 'England', 'Japan', 'Cambodia', 'Canada', 'India', 'Hong', 'Laos', 'Portugal', 'Hungary', 'Nicaragua', 'Ireland', 'Jamaica', 'Mexico', 'Honduras', 'Peru', 'Poland', '?', 'United-States', 'El-Salvador')) GROUP BY workclass,`marital.status`,sex,income,`native.country`,education,relationship,occupation

Resulted in 13412 records

Normalized Euclidean distance for (fnlwgt): 61.92

Hellinger Distance: 0.26

==========

Real

| | relationship | native.country | education | occupation | workclass | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **6656** | Unmarried | China | Some-college | Tech-support | Private | nan | nan |
| **1282** | Husband | Peru | Masters | Craft-repair | Federal-gov | nan | nan |
| **5211** | Own-child | ? | 11th | Protective-serv | Private | nan | nan |
| **7410** | Unmarried | United-States | 10th | Transport-moving | Private | 28.000000 | 9.000000 |
| **5497** | Own-child | Hungary | 11th | Prof-specialty | Private | nan | nan |

Synthetic

| | relationship | native.country | education | occupation | workclass | MIN(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **6656** | Unmarried | China | Some-college | Tech-support | Private | 28 | 1 |
| **1282** | Husband | Peru | Masters | Craft-repair | Federal-gov | 39 | 1 |
| **5211** | Own-child | ? | 11th | Protective-serv | Private | 48 | 1 |
| **7410** | Unmarried | United-States | 10th | Transport-moving | Private | 30 | 4 |
| **5497** | Own-child | Hungary | 11th | Prof-specialty | Private | 28 | 1 |

SQL for Real:
SELECT relationship,`native.country`,education,occupation,workclass,MIN(age), COUNT(*) FROM C1 WHERE (race = 'White' AND income = '<=50K') OR (`native.country`

IN ('El-Salvador', '?', 'Puerto-Rico', 'Columbia', 'Haiti') OR sex = 'Male') GROUP BY relationship,`native.country`,education,occupation,workclass

Resulted in 5291 records

SQL for Synthetic:
SELECT relationship,`native.country`,education,occupation,workclass,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (race = 'White' AND income = '<=50K') OR (`native.country` IN ('El-Salvador', '?', 'Puerto-Rico', 'Columbia', 'Haiti') OR sex = 'Male') GROUP BY relationship,`native.country`,education,occupation,workclass

Resulted in 8429 records

Normalized Euclidean distance for (age): 49.11

Hellinger Distance: 0.263

==========

Real

| | workclass | occupation | relationship | education | native.country | marital.status | income | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **2825** | Self-emp-not-inc | Craft-repair | Husband | Doctorate | Puerto-Rico | Married-civ-spouse | <=50K | nan | nan |
| **3031** | Self-emp-not-inc | Exec-managerial | Husband | Bachelors | Puerto-Rico | Married-civ-spouse | >50K | 37.000000 | 1.000000 |
| **3460** | Self-emp-not-inc | Other-service | Not-in-family | HS-grad | Puerto-Rico | Never-married | <=50K | nan | nan |
| **2765** | Self-emp-not-inc | Adm-clerical | Unmarried | 11th | Puerto-Rico | Divorced | <=50K | nan | nan |
| **112** | ? | ? | Husband | Some-college | Laos | Married-civ-spouse | >50K | nan | nan |

Synthetic

| | workclass | occupation | relationship | education | native.country | marital.status | income | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **2825** | Self-emp-not-inc | Craft-repair | Husband | Doctorate | Puerto-Rico | Married-civ-spouse | <=50K | 42.000000 | 1 |
| **3031** | Self-emp-not-inc | Exec-managerial | Husband | Bachelors | Puerto-Rico | Married-civ-spouse | >50K | 45.625000 | 8 |
| **3460** | Self-emp-not-inc | Other-service | Not-in-family | HS-grad | Puerto-Rico | Never-married | <=50K | 39.000000 | 5 |
| **2765** | Self-emp-not-inc | Adm-clerical | Unmarried | 11th | Puerto-Rico | Divorced | <=50K | 49.000000 | 1 |
| **112** | ? | ? | Husband | Some-college | Laos | Married-civ-spouse | >50K | 31.000000 | 1 |

SQL for Real:
SELECT workclass,occupation,relationship,education,`native.country`,`marital.status`,income,AVG(age), COUNT(*) FROM C1 WHERE (`hours.per.week` = 40 AND `native.country` LIKE 'Mexico') AND (age = 61 AND income LIKE '<=50K') OR (relationship LIKE 'Husband' OR workclass LIKE 'Self-emp-not-inc') GROUP BY workclass,occupation,relationship,education,`native.country`,`marital.status`,income

Resulted in 3088 records

SQL for Synthetic:
SELECT workclass,occupation,relationship,education,`native.country`,`marital.status`,income,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` = 40 AND `native.country` LIKE 'Mexico') AND (age = 61 AND income LIKE '<=50K') OR (relationship LIKE 'Husband' OR workclass LIKE 'Self-emp-not-inc') GROUP BY workclass,occupation,relationship,education,`native.country`,`marital.status`,income

Resulted in 4314 records

Normalized Euclidean distance for (age): 35.19

Hellinger Distance: 0.324

==========

Real

| | occupation | race | marital.status | relationship | income | workclass | native.country | sex | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **316** | Protective-serv | Black | Never-married | Not-in-family | >50K | Private | United-States | Male | 36.000000 | 1 |
| **148** | Farming-fishing | White | Married-AF-spouse | Husband | >50K | Private | United-States | Male | 29.000000 | 1 |
| **54** | Craft-repair | Other | Married-civ-spouse | Husband | >50K | Private | United-States | Male | 43.666667 | 3 |
| **217** | Other-service | White | Divorced | Not-in-family | >50K | Private | ? | Male | 45.000000 | 1 |
| **156** | Handlers-cleaners | Asian-Pac-Islander | Married-civ-spouse | Husband | >50K | Private | Philippines | Male | 62.000000 | 2 |

Synthetic

| | occupation | race | marital.status | relationship | income | workclass | native.country | sex | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **316** | Protective-serv | Black | Never-married | Not-in-family | >50K | Private | United-States | Male | nan | nan |
| **148** | Farming-fishing | White | Married-AF-spouse | Husband | >50K | Private | United-States | Male | nan | nan |
| **54** | Craft-repair | Other | Married-civ-spouse | Husband | >50K | Private | United-States | Male | nan | nan |
| **217** | Other-service | White | Divorced | Not-in-family | >50K | Private | ? | Male | nan | nan |
| **156** | Handlers-cleaners | Asian-Pac-Islander | Married-civ-spouse | Husband | >50K | Private | Philippines | Male | nan | nan |

SQL for Real:
SELECT occupation,race,`marital.status`,relationship,income,workclass,`native.country`,sex,AVG(age), COUNT(*) FROM C1 WHERE (income = '>50K' AND workclass = 'Private') AND (relationship <> 'Own-child' AND sex LIKE 'Male') GROUP BY occupation,race,`marital.status`,relationship,income,workclass,`native.country`,sex

Resulted in 416 records

SQL for Synthetic:
SELECT occupation,race,`marital.status`,relationship,income,workclass,`native.country`,sex,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (income = '>50K' AND workclass = 'Private') AND (relationship <> 'Own-child' AND sex LIKE 'Male') GROUP BY occupation,race,`marital.status`,relationship,income,workclass,`native.country`,sex

Resulted in 296 records

Normalized Euclidean distance for (age): 10.3

Hellinger Distance: 0.244

==========

Real

| | workclass | marital.status | relationship | SUM(capital) | COUNT(*) |
|---|---|---|---|---|---|
| **146** | Self-emp-not-inc | Married-spouse-absent | Other-relative | -1564.000000 | 1.000000 |
| **15** | ? | Married-spouse-absent | Unmarried | 0.000000 | 12.000000 |
| **121** | Self-emp-inc | Married-civ-spouse | Wife | -12625.000000 | 58.000000 |
| **143** | Self-emp-not-inc | Married-civ-spouse | Wife | -27781.000000 | 169.000000 |
| **32** | Federal-gov | Divorced | Own-child | 0.000000 | 8.000000 |

Synthetic

| | workclass | marital.status | relationship | SUM(capital) | COUNT(*) |
|---|---|---|---|---|---|
| **146** | Self-emp-not-inc | Married-spouse-absent | Other-relative | -57.081414 | 1 |
| **15** | ? | Married-spouse-absent | Unmarried | -461.194586 | 5 |
| **121** | Self-emp-inc | Married-civ-spouse | Wife | -2792.338970 | 38 |
| **143** | Self-emp-not-inc | Married-civ-spouse | Wife | -6869.066081 | 85 |
| **32** | Federal-gov | Divorced | Own-child | -450.855735 | 5 |

SQL for Real:
SELECT workclass,`marital.status`,relationship,SUM(capital), COUNT(*) FROM C1 WHERE (capital <= 0) GROUP BY workclass,`marital.status`,relationship

Resulted in 170 records

SQL for Synthetic:
SELECT workclass,`marital.status`,relationship,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (capital <= 0) GROUP BY workclass,`marital.status`,relationship

Resulted in 179 records

Normalized Euclidean distance for (capital): 12.21

Hellinger Distance: 0.039

==========

Real

| | native.country | education | marital.status | race | relationship | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **1180** | India | 12th | Divorced | White | Unmarried | nan | nan |
| **2394** | Puerto-Rico | 7th-8th | Separated | Asian-Pac-Islander | Not-in-family | nan | nan |
| **1294** | Ireland | 1st-4th | Separated | White | Unmarried | nan | nan |
| **1665** | Laos | 11th | Separated | White | Unmarried | nan | nan |
| **1257** | Iran | HS-grad | Married-civ-spouse | White | Husband | 55.000000 | 1.000000 |

Synthetic

| | native.country | education | marital.status | race | relationship | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **1180** | India | 12th | Divorced | White | Unmarried | 40.052175 | 1 |
| **2394** | Puerto-Rico | 7th-8th | Separated | Asian-Pac-Islander | Not-in-family | 40.025162 | 1 |
| **1294** | Ireland | 1st-4th | Separated | White | Unmarried | 39.992412 | 1 |
| **1665** | Laos | 11th | Separated | White | Unmarried | 79.957776 | 2 |
| **1257** | Iran | HS-grad | Married-civ-spouse | White | Husband | 1162.534391 | 29 |

SQL for Real:
SELECT `native.country`,education,`marital.status`,race,relationship,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE ((age BETWEEN 20 AND 35) OR relationship IN ('Wife', 'Not-in-family', 'Other-relative', 'Own-child', 'Husband', 'Unmarried')) OR (`marital.status` = 'Married-spouse-absent' OR sex IN ('Female', 'Male')) OR (occupation LIKE 'Other-service' OR race IN ('Other', 'White', 'Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'Black')) GROUP BY `native.country`,education,`marital.status`,race,relationship

Resulted in 2950 records

SQL for Synthetic:
SELECT `native.country`,education,`marital.status`,race,relationship,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE ((age BETWEEN 20 AND 35) OR relationship IN ('Wife', 'Not-in-family', 'Other-relative', 'Own-child', 'Husband', 'Unmarried')) OR (`marital.status` = 'Married-spouse-absent' OR sex IN ('Female', 'Male')) OR (occupation LIKE 'Other-service' OR race IN ('Other', 'White', 'Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'Black')) GROUP BY `native.country`,education,`marital.status`,race,relationship

Resulted in 3922 records

Normalized Euclidean distance for (`hours.per.week`): 35.04

Hellinger Distance: 0.35

==========

Real

| | relationship | workclass | income | education | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|
| **184** | Other-relative | Private | <=50K | Some-college | 34.375000 | 16 |
| **270** | Unmarried | Private | <=50K | 9th | 37.727273 | 11 |
| **22** | Husband | Local-gov | <=50K | 10th | 20.000000 | 1 |
| **167** | Other-relative | ? | <=50K | HS-grad | 32.800000 | 5 |
| **315** | Wife | Private | <=50K | 10th | 36.400000 | 5 |

Synthetic

| | relationship | workclass | income | education | AVG(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|
| **184** | Other-relative | Private | <=50K | Some-college | 39.817394 | 9.000000 |
| **270** | Unmarried | Private | <=50K | 9th | nan | nan |
| **22** | Husband | Local-gov | <=50K | 10th | nan | nan |
| **167** | Other-relative | ? | <=50K | HS-grad | 40.016429 | 1.000000 |
| **315** | Wife | Private | <=50K | 10th | nan | nan |

SQL for Real:
SELECT relationship,workclass,income,education,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (race <> 'White') AND (age <= 39 OR relationship <> 'Own-child') AND (`native.country` LIKE 'United-States' AND fnlwgt >= 194869) GROUP BY relationship,workclass,income,education

Resulted in 340 records

SQL for Synthetic:
SELECT relationship,workclass,income,education,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (race <> 'White') AND (age <= 39 OR relationship <> 'Own-child') AND (`native.country` LIKE 'United-States' AND fnlwgt >= 194869) GROUP BY relationship,workclass,income,education

Resulted in 68 records

Normalized Euclidean distance for (`hours.per.week`): 8.06

Hellinger Distance: 0.252

==========

Real

| | workclass | education | occupation | race | marital.status | sex | relationship | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **3704** | Private | HS-grad | Priv-house-serv | White | Widowed | Female | Other-relative | 226084 | 1 |
| **156** | ? | Bachelors | ? | Black | Married-civ-spouse | Male | Husband | 230165 | 1 |
| **702** | Federal-gov | Some-college | Adm-clerical | White | Never-married | Male | Unmarried | 113398 | 1 |
| **4851** | Self-emp-not-inc | 12th | Craft-repair | White | Married-civ-spouse | Male | Not-in-family | 107236 | 1 |
| **5100** | Self-emp-not-inc | Masters | Exec-managerial | White | Divorced | Female | Unmarried | 109684 | 1 |

Synthetic

| | workclass | education | occupation | race | marital.status | sex | relationship | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **3704** | Private | HS-grad | Priv-house-serv | White | Widowed | Female | Other-relative | nan | nan |
| **156** | ? | Bachelors | ? | Black | Married-civ-spouse | Male | Husband | nan | nan |
| **702** | Federal-gov | Some-college | Adm-clerical | White | Never-married | Male | Unmarried | nan | nan |
| **4851** | Self-emp-not-inc | 12th | Craft-repair | White | Married-civ-spouse | Male | Not-in-family | nan | nan |
| **5100** | Self-emp-not-inc | Masters | Exec-managerial | White | Divorced | Female | Unmarried | nan | nan |

SQL for Real:
SELECT workclass,education,occupation,race,`marital.status`,sex,relationship,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (`hours.per.week` = 40) GROUP BY workclass,education,occupation,race,`marital.status`,sex,relationship

Resulted in 5715 records

SQL for Synthetic:
SELECT workclass,education,occupation,race,`marital.status`,sex,relationship,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` = 40) GROUP BY workclass,education,occupation,race,`marital.status`,sex,relationship

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

==========

Real

| | income | native.country | occupation | marital.status | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|
| **695** | <=50K | Mexico | Transport-moving | Divorced | nan | nan |
| **1160** | <=50K | Yugoslavia | Other-service | Married-spouse-absent | nan | nan |
| **1170** | >50K | ? | Craft-repair | Married-spouse-absent | 82.000000 | 2.000000 |
| **593** | <=50K | Japan | Priv-house-serv | Separated | nan | nan |
| **945** | <=50K | Scotland | Farming-fishing | Never-married | nan | nan |

Synthetic

| | income | native.country | occupation | marital.status | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|
| **695** | <=50K | Mexico | Transport-moving | Divorced | 65 | 2 |
| **1160** | <=50K | Yugoslavia | Other-service | Married-spouse-absent | 81 | 2 |
| **1170** | >50K | ? | Craft-repair | Married-spouse-absent | 47 | 1 |
| **593** | <=50K | Japan | Priv-house-serv | Separated | 55 | 1 |
| **945** | <=50K | Scotland | Farming-fishing | Never-married | 65 | 2 |

SQL for Real:
SELECT income,`native.country`,occupation,`marital.status`,SUM(age), COUNT(*) FROM C1 WHERE (relationship = 'Unmarried' AND education LIKE 'Bachelors') AND (income <> '>50K' AND capital <= 0) AND (race IN ('Amer-Indian-Eskimo', 'Other', 'Black', 'Asian-Pac-Islander', 'White') AND `native.country` = 'United-States') OR (`marital.status` <> 'Married-civ-spouse' OR occupation IN ('?', 'Machine-op-inspct', 'Sales', 'Handlers-cleaners', 'Protective-serv')) GROUP BY income,`native.country`,occupation,`marital.status`

Resulted in 1293 records

SQL for Synthetic:
SELECT income,`native.country`,occupation,`marital.status`,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (relationship = 'Unmarried' AND education LIKE 'Bachelors') AND (income <> '>50K' AND capital <= 0) AND (race IN ('Amer-Indian-Eskimo', 'Other', 'Black', 'Asian-Pac-Islander', 'White') AND `native.country` = 'United-States') OR (`marital.status` <> 'Married-civ-spouse' OR occupation IN ('?', 'Machine-op-inspct', 'Sales', 'Handlers-cleaners', 'Protective-serv')) GROUP BY income,`native.country`,occupation,`marital.status`

Resulted in 1367 records

Normalized Euclidean distance for (age): 25.65

Hellinger Distance: 0.349

==========

Real

| | marital.status | occupation | sex | relationship | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|---|
| **26** | Never-married | ? | Male | Unmarried | 0 | 9 |
| **30** | Separated | ? | Female | Unmarried | 0 | 8 |
| **18** | Married-spouse-absent | ? | Male | Other-relative | 0 | 1 |
| **5** | Divorced | ? | Male | Own-child | -1602 | 1 |
| **3** | Divorced | ? | Female | Unmarried | 0 | 6 |

Synthetic

| | marital.status | occupation | sex | relationship | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|---|
| **26** | Never-married | ? | Male | Unmarried | 240.649821 | 14.000000 |
| **30** | Separated | ? | Female | Unmarried | 40.831991 | 4.000000 |
| **18** | Married-spouse-absent | ? | Male | Other-relative | nan | nan |
| **5** | Divorced | ? | Male | Own-child | 164.723799 | 3.000000 |
| **3** | Divorced | ? | Female | Unmarried | 165.702610 | 4.000000 |

SQL for Real:
SELECT `marital.status`,occupation,sex,relationship,MAX(capital), COUNT(*) FROM C1 WHERE (occupation LIKE '?') AND (age <= 26 AND race IN ('White', 'Black')) GROUP BY `marital.status`,occupation,sex,relationship

Resulted in 34 records

SQL for Synthetic:
SELECT `marital.status`,occupation,sex,relationship,MAX(capital), COUNT(*) FROM C1_syn_06 WHERE (occupation LIKE '?') AND (age <= 26 AND race IN ('White', 'Black')) GROUP BY `marital.status`,occupation,sex,relationship

Resulted in 29 records

Normalized Euclidean distance for (capital): 5.1

Hellinger Distance: 0.081

==========

Real

| | marital.status | education | workclass | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|
| **48** | Never-married | 1st-4th | Private | 49.000000 | 32.000000 |
| **72** | Separated | Some-college | Private | 90.000000 | 7.000000 |
| **53** | Never-married | Assoc-voc | Private | 45.000000 | 12.000000 |
| **4** | Divorced | 5th-6th | Private | 60.000000 | 9.000000 |
| **68** | Separated | Bachelors | Private | 50.000000 | 9.000000 |

Synthetic

| | marital.status | education | workclass | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|
| **48** | Never-married | 1st-4th | Private | 54 | 5 |
| **72** | Separated | Some-college | Private | 60 | 30 |
| **53** | Never-married | Assoc-voc | Private | 58 | 21 |
| **4** | Divorced | 5th-6th | Private | 56 | 6 |
| **68** | Separated | Bachelors | Private | 61 | 13 |

SQL for Real:
SELECT `marital.status`,education,workclass,MAX(age), COUNT(*) FROM C1 WHERE (capital BETWEEN 0 AND 4650) AND (workclass = 'Private' AND `native.country` IN ('Thailand', 'Philippines', 'Vietnam', 'Puerto-Rico', 'Greece', 'France', 'Mexico', 'Holand-Netherlands', 'Cambodia', 'El-Salvador', 'Scotland', 'Italy', 'England', 'Jamaica', 'Portugal')) GROUP BY `marital.status`,education,workclass

Resulted in 80 records

SQL for Synthetic:
SELECT `marital.status`,education,workclass,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (capital BETWEEN 0 AND 4650) AND (workclass = 'Private' AND `native.country` IN ('Thailand', 'Philippines', 'Vietnam', 'Puerto-Rico', 'Greece', 'France', 'Mexico', 'Holand-Netherlands', 'Cambodia', 'El-Salvador', 'Scotland', 'Italy', 'England', 'Jamaica', 'Portugal')) GROUP BY `marital.status`,education,workclass

Resulted in 88 records

Normalized Euclidean distance for (age): 8.54

Hellinger Distance: 0.341

==========

Real

| | relationship | education | occupation | marital.status | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|
| **655** | Not-in-family | Bachelors | Craft-repair | Never-married | 34.285714 | 42.000000 |
| **836** | Not-in-family | Masters | Prof-specialty | Married-spouse-absent | 54.500000 | 6.000000 |
| **1171** | Other-relative | HS-grad | Protective-serv | Separated | 28.500000 | 2.000000 |
| **754** | Not-in-family | HS-grad | Exec-managerial | Widowed | 61.441176 | 34.000000 |
| **1267** | Own-child | 10th | Farming-fishing | Divorced | nan | nan |

Synthetic

| | relationship | education | occupation | marital.status | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|
| **655** | Not-in-family | Bachelors | Craft-repair | Never-married | 30.862745 | 51 |
| **836** | Not-in-family | Masters | Prof-specialty | Married-spouse-absent | 45.500000 | 2 |
| **1171** | Other-relative | HS-grad | Protective-serv | Separated | 49.000000 | 2 |
| **754** | Not-in-family | HS-grad | Exec-managerial | Widowed | 60.981132 | 53 |
| **1267** | Own-child | 10th | Farming-fishing | Divorced | 50.000000 | 2 |

SQL for Real:
SELECT relationship,education,occupation,`marital.status`,AVG(age), COUNT(*) FROM C1 WHERE (`marital.status` = 'Married-civ-spouse' OR race LIKE 'White') GROUP BY relationship,education,occupation,`marital.status`

Resulted in 2276 records

SQL for Synthetic:
SELECT relationship,education,occupation,`marital.status`,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` = 'Married-civ-spouse' OR race LIKE 'White') GROUP BY relationship,education,occupation,`marital.status`

Resulted in 2367 records

Normalized Euclidean distance for (age): 41.86

Hellinger Distance: 0.091

==========

Real

| | income | relationship | sex | education | race | marital.status | workclass | occupation | native.country | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **2072** | <=50K | Other-relative | Male | 11th | White | Never-married | Private | Sales | United-States | nan | nan |
| **1157** | <=50K | Not-in-family | Female | Masters | White | Divorced | Private | Prof-specialty | United-States | 52.000000 | 2.000000 |
| **1994** | <=50K | Other-relative | Female | HS-grad | White | Divorced | Private | Machine-op-inspct | United-States | nan | nan |
| **945** | <=50K | Not-in-family | Female | Bachelors | White | Never-married | Private | Exec-managerial | Laos | nan | nan |
| **2823** | <=50K | Own-child | Male | HS-grad | White | Never-married | Private | Tech-support | Laos | nan | nan |

Synthetic

| | income | relationship | sex | education | race | marital.status | workclass | occupation | native.country | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **2072** | <=50K | Other-relative | Male | 11th | White | Never-married | Private | Sales | United-States | 17 | 1 |
| **1157** | <=50K | Not-in-family | Female | Masters | White | Divorced | Private | Prof-specialty | United-States | 62 | 3 |
| **1994** | <=50K | Other-relative | Female | HS-grad | White | Divorced | Private | Machine-op-inspct | United-States | 42 | 2 |
| **945** | <=50K | Not-in-family | Female | Bachelors | White | Never-married | Private | Exec-managerial | Laos | 36 | 2 |
| **2823** | <=50K | Own-child | Male | HS-grad | White | Never-married | Private | Tech-support | Laos | 26 | 1 |

SQL for Real:
SELECT income,relationship,sex,education,race,`marital.status`,workclass,occupation,`native.country`,MAX(age), COUNT(*) FROM C1 WHERE (education IN ('10th', '12th', 'Preschool', '9th', '1st-4th', 'Doctorate', 'Assoc-voc') OR workclass = 'Private') AND (`native.country` IN ('Germany', 'Hong', 'Outlying-US(Guam-USVI-etc)', 'Greece', 'Cambodia', 'France', 'Portugal', 'Iran', 'Scotland', 'Yugoslavia', 'Laos', 'Honduras', 'Peru', 'Italy', 'Taiwan', 'United-States', 'Nicaragua', 'Jamaica', 'Japan', 'Dominican-Republic', 'China', 'Cuba', 'Ecuador', 'Thailand', 'Trinadad&Tobago', 'Philippines') AND capital > 0) GROUP BY income,relationship,sex,education,race,`marital.status`,workclass,occupation,`native.country`

Resulted in 1053 records

SQL for Synthetic:
SELECT income,relationship,sex,education,race,`marital.status`,workclass,occupation,`native.country`,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (education IN ('10th', '12th', 'Preschool', '9th', '1st-4th', 'Doctorate', 'Assoc-voc') OR workclass = 'Private') AND (`native.country` IN ('Germany', 'Hong', 'Outlying-US(Guam-USVI-etc)', 'Greece', 'Cambodia', 'France', 'Portugal', 'Iran', 'Scotland', 'Yugoslavia', 'Laos', 'Honduras', 'Peru', 'Italy', 'Taiwan', 'United-States', 'Nicaragua', 'Jamaica', 'Japan', 'Dominican-Republic', 'China', 'Cuba', 'Ecuador', 'Thailand', 'Trinadad&Tobago', 'Philippines') AND capital > 0) GROUP BY income,relationship,sex,education,race,`marital.status`,workclass,occupation,`native.country`

Resulted in 4082 records

Normalized Euclidean distance for (age): 21.95

Hellinger Distance: 0.469

==========

Real

| | workclass | relationship | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|
| **35** | Self-emp-not-inc | Wife | 462832 | 191 |
| **25** | Self-emp-inc | Not-in-family | 593246 | 253 |
| **1** | ? | Not-in-family | 913447 | 667 |
| **20** | Private | Other-relative | 874728 | 1158 |
| **11** | Federal-gov | Wife | 436341 | 61 |

Synthetic

| | workclass | relationship | MAX(fnlwgt) | COUNT(*) |
|---|---|---|---|---|
| **35** | Self-emp-not-inc | Wife | 191686.848684 | 198.000000 |
| **25** | Self-emp-inc | Not-in-family | 204293.928722 | 272.000000 |
| **1** | ? | Not-in-family | 1338996.446401 | 658.000000 |
| **20** | Private | Other-relative | 209208.041466 | 594.000000 |
| **11** | Federal-gov | Wife | 194985.521430 | 71.000000 |

SQL for Real:
SELECT workclass,relationship,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (race IN ('Other', 'Amer-Indian-Eskimo')) AND (`hours.per.week` > 40 OR `marital.status` LIKE 'Never-married') OR (education LIKE 'Some-college' OR capital BETWEEN 0 AND 0) AND (fnlwgt <> 172415 AND sex IN ('Male', 'Female')) OR (workclass <> 'Private' OR `native.country` <> 'United-States') GROUP BY workclass,relationship

Resulted in 42 records

SQL for Synthetic:
SELECT workclass,relationship,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (race IN ('Other', 'Amer-Indian-Eskimo')) AND (`hours.per.week` > 40 OR `marital.status` LIKE 'Never-married') OR (education LIKE 'Some-college' OR capital BETWEEN 0 AND 0) AND (fnlwgt <> 172415 AND sex IN ('Male', 'Female')) OR (workclass <> 'Private' OR `native.country` <> 'United-States') GROUP BY workclass,relationship

Resulted in 42 records

Normalized Euclidean distance for (fnlwgt): 6.48

Hellinger Distance: 0.085

==========

Real

| | education | relationship | workclass | MIN(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|
| **0** | 11th | Unmarried | Private | 159109 | 1 |
| **3** | Some-college | Unmarried | Private | 277342 | 1 |
| **2** | HS-grad | Unmarried | Self-emp-not-inc | 229842 | 1 |
| **1** | HS-grad | Unmarried | Private | 189346 | 2 |

Synthetic

| | education | relationship | workclass | MIN(fnlwgt) | COUNT(*) |
|---|---|---|---|---|---|
| **0** | 11th | Unmarried | Private | nan | nan |
| **3** | Some-college | Unmarried | Private | nan | nan |
| **2** | HS-grad | Unmarried | Self-emp-not-inc | nan | nan |
| **1** | HS-grad | Unmarried | Private | nan | nan |

SQL for Real:
SELECT education,relationship,workclass,MIN(fnlwgt), COUNT(*) FROM C1 WHERE (fnlwgt <> 343061 OR `hours.per.week` <= 40) AND (age = 29 AND workclass <> 'Local-gov') AND (income IN ('>50K', '<=50K') AND occupation LIKE 'Transport-moving') AND (relationship LIKE 'Unmarried' AND sex = 'Male') AND ((capital BETWEEN 0 AND 0) AND education <> '10th') GROUP BY education,relationship,workclass

Resulted in 4 records

SQL for Synthetic:
SELECT education,relationship,workclass,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt <> 343061 OR `hours.per.week` <= 40) AND (age = 29 AND workclass <> 'Local-gov') AND (income IN ('>50K', '<=50K') AND occupation LIKE 'Transport-moving') AND (relationship LIKE 'Unmarried' AND sex = 'Male') AND ((capital BETWEEN 0 AND 0) AND education <> '10th') GROUP BY education,relationship,workclass

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

==========

Real

| | relationship | workclass | sex | education | native.country | occupation | race | marital.status | AVG(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **15690** | Wife | Private | Female | Some-college | Ireland | Exec-managerial | White | Married-civ-spouse | nan | nan |
| **10871** | Own-child | Private | Male | 11th | Laos | Farming-fishing | White | Never-married | nan | nan |
| **9540** | Own-child | ? | Female | HS-grad | Ireland | ? | Black | Never-married | nan | nan |
| **7938** | Not-in-family | Self-emp-not-inc | Male | Assoc-acdm | United-States | Other-service | White | Divorced | nan | nan |
| **4657** | Not-in-family | Private | Female | 5th-6th | United-States | Priv-house-serv | White | Never-married | nan | nan |

Synthetic

| | relationship | workclass | sex | education | native.country | occupation | race | marital.status | AVG(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **15690** | Wife | Private | Female | Some-college | Ireland | Exec-managerial | White | Married-civ-spouse | 68.042294 | 1 |
| **10871** | Own-child | Private | Male | 11th | Laos | Farming-fishing | White | Never-married | -45.444730 | 2 |
| **9540** | Own-child | ? | Female | HS-grad | Ireland | ? | Black | Never-married | 44.174087 | 1 |
| **7938** | Not-in-family | Self-emp-not-inc | Male | Assoc-acdm | United-States | Other-service | White | Divorced | -208.488082 | 1 |
| **4657** | Not-in-family | Private | Female | 5th-6th | United-States | Priv-house-serv | White | Never-married | -218.716237 | 1 |

SQL for Real:
SELECT relationship,workclass,sex,education,`native.country`,occupation,race,`marital.status`,AVG(capital), COUNT(*) FROM C1 WHERE (education = 'HS-grad') AND (`hours.per.week` < 16 AND relationship <> 'Own-child') OR (income = '<=50K' OR occupation = 'Sales') OR (race IN ('Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'Other') OR capital <> 0) GROUP BY relationship,workclass,sex,education,`native.country`,occupation,race,`marital.status`

Resulted in 10822 records

SQL for Synthetic:
SELECT relationship,workclass,sex,education,`native.country`,occupation,race,`marital.status`,AVG(capital), COUNT(*) FROM C1_syn_06 WHERE (education = 'HS-grad') AND (`hours.per.week` < 16 AND relationship <> 'Own-child') OR (income = '<=50K' OR occupation = 'Sales') OR (race IN ('Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'Other') OR capital <> 0) GROUP BY relationship,workclass,sex,education,`native.country`,occupation,race,`marital.status`

Resulted in 15994 records

Normalized Euclidean distance for (capital): 62.06

Hellinger Distance: 0.23

==========

Real

| | education | marital.status | native.country | race | relationship | income | workclass | occupation | sex | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **16138** | Some-college | Never-married | Puerto-Rico | White | Not-in-family | <=50K | Private | ? | Male | nan | nan |
| **13542** | Preschool | Married-civ-spouse | Puerto-Rico | White | Husband | <=50K | Private | Craft-repair | Male | nan | nan |
| **11983** | HS-grad | Separated | United-States | White | Not-in-family | <=50K | Private | Craft-repair | Female | 40.000000 | 1.000000 |
| **6926** | Bachelors | Never-married | Laos | White | Not-in-family | <=50K | Private | Adm-clerical | Male | nan | nan |
| **2109** | 5th-6th | Married-civ-spouse | Mexico | White | Husband | <=50K | Private | Transport-moving | Male | 37.000000 | 4.000000 |

Synthetic

| | education | marital.status | native.country | race | relationship | income | workclass | occupation | sex | MIN(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **16138** | Some-college | Never-married | Puerto-Rico | White | Not-in-family | <=50K | Private | ? | Male | 40.053874 | 1 |
| **13542** | Preschool | Married-civ-spouse | Puerto-Rico | White | Husband | <=50K | Private | Craft-repair | Male | 39.974042 | 3 |
| **11983** | HS-grad | Separated | United-States | White | Not-in-family | <=50K | Private | Craft-repair | Female | 39.943371 | 1 |
| **6926** | Bachelors | Never-married | Laos | White | Not-in-family | <=50K | Private | Adm-clerical | Male | 40.011355 | 2 |
| **2109** | 5th-6th | Married-civ-spouse | Mexico | White | Husband | <=50K | Private | Transport-moving | Male | 40.215503 | 1 |

SQL for Real:
SELECT education,`marital.status`,`native.country`,race,relationship,income,workclass,occupation,sex,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (income IN ('>50K', '<=50K') AND education = 'Some-college') OR (age < 21 OR sex IN ('Male', 'Female')) AND (race <> 'White' AND `native.country` IN ('Peru', 'China', 'Yugoslavia', 'Germany', 'Italy', 'Mexico', 'Jamaica', 'Columbia', 'Hong', 'Cambodia', '?', 'Poland', 'Scotland', 'South', 'Vietnam', 'Puerto-Rico', 'Laos', 'Taiwan', 'India', 'Ecuador')) OR (`hours.per.week` > 20 OR `marital.status` IN ('Married-AF-spouse', 'Never-married', 'Married-spouse-absent', 'Divorced')) GROUP BY education,`marital.status`,`native.country`,race,relationship,income,workclass,occupation,sex

Resulted in 12377 records

SQL for Synthetic:
SELECT education,`marital.status`,`native.country`,race,relationship,income,workclass,occupation,sex,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (income IN ('>50K', '<=50K') AND education = 'Some-college') OR (age < 21 OR sex IN ('Male', 'Female')) AND (race <> 'White' AND `native.country` IN ('Peru', 'China', 'Yugoslavia', 'Germany', 'Italy', 'Mexico', 'Jamaica', 'Columbia', 'Hong', 'Cambodia', '?', 'Poland', 'Scotland', 'South', 'Vietnam', 'Puerto-Rico', 'Laos', 'Taiwan', 'India', 'Ecuador')) OR (`hours.per.week` > 20 OR `marital.status` IN ('Married-AF-spouse', 'Never-married', 'Married-spouse-absent', 'Divorced')) GROUP BY education,`marital.status`,`native.country`,race,relationship,income,workclass,occupation,sex

Resulted in 17127 records

Normalized Euclidean distance for (`hours.per.week`): 64.22

Hellinger Distance: 0.253

==========

Real

| | relationship | marital.status | SUM(age) | COUNT(*) |
|---|---|---|---|---|
| **0** | Unmarried | Divorced | 99037.000000 | 2369.000000 |
| **5** | Unmarried | Separated | 26328.000000 | 668.000000 |
| **2** | Unmarried | Married-civ-spouse | nan | nan |
| **4** | Unmarried | Never-married | 42301.000000 | 1333.000000 |
| **3** | Unmarried | Married-spouse-absent | 7554.000000 | 183.000000 |

Synthetic

| | relationship | marital.status | SUM(age) | COUNT(*) |
|---|---|---|---|---|
| **0** | Unmarried | Divorced | 97032 | 2307 |
| **5** | Unmarried | Separated | 25405 | 638 |
| **2** | Unmarried | Married-civ-spouse | 879 | 21 |
| **4** | Unmarried | Never-married | 41576 | 1302 |
| **3** | Unmarried | Married-spouse-absent | 6290 | 155 |

SQL for Real:
SELECT relationship,`marital.status`,SUM(age), COUNT(*) FROM C1 WHERE (relationship LIKE 'Unmarried') GROUP BY relationship,`marital.status`

Resulted in 5 records

SQL for Synthetic:
SELECT relationship,`marital.status`,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (relationship LIKE 'Unmarried') GROUP BY relationship,`marital.status`

Resulted in 7 records

Normalized Euclidean distance for (age): 2.24

Hellinger Distance: 0.01

==========

Real

| | occupation | income | native.country | marital.status | sex | relationship | race | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **341** | ? | <=50K | United-States | Never-married | Male | Husband | White | nan | nan |
| **117** | ? | <=50K | Japan | Widowed | Female | Unmarried | Black | nan | nan |
| **5043** | Tech-support | <=50K | Puerto-Rico | Widowed | Female | Unmarried | White | nan | nan |
| **250** | ? | <=50K | Scotland | Married-civ-spouse | Male | Husband | White | nan | nan |
| **1707** | Exec-managerial | <=50K | Peru | Never-married | Female | Not-in-family | Asian-Pac-Islander | nan | nan |

Synthetic

| | occupation | income | native.country | marital.status | sex | relationship | race | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **341** | ? | <=50K | United-States | Never-married | Male | Husband | White | 60 | 3 |
| **117** | ? | <=50K | Japan | Widowed | Female | Unmarried | Black | 66 | 1 |
| **5043** | Tech-support | <=50K | Puerto-Rico | Widowed | Female | Unmarried | White | 272 | 5 |
| **250** | ? | <=50K | Scotland | Married-civ-spouse | Male | Husband | White | 1039 | 18 |
| **1707** | Exec-managerial | <=50K | Peru | Never-married | Female | Not-in-family | Asian-Pac-Islander | 38 | 1 |

SQL for Real:
SELECT occupation,income,`native.country`,`marital.status`,sex,relationship,race,SUM(age), COUNT(*) FROM C1 WHERE (sex = 'Female' OR occupation = 'Other-service') OR ((fnlwgt BETWEEN 325802 AND 331395) OR income <> '<=50K') OR (race = 'White' OR age BETWEEN 50 AND 55) OR (workclass LIKE '?' OR `native.country` IN ('Outlying-US(Guam-USVI-etc)', 'Honduras', 'Vietnam', 'El-Salvador', 'Holand-Netherlands', 'Trinadad&Tobago', 'Dominican-Republic', 'Cuba', 'Columbia')) AND ((capital BETWEEN -1902 AND 0) AND education <> 'Assoc-voc') GROUP BY occupation,income,`native.country`,`marital.status`,sex,relationship,race

Resulted in 3897 records

SQL for Synthetic:
SELECT occupation,income,`native.country`,`marital.status`,sex,relationship,race,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (sex = 'Female' OR occupation = 'Other-service') OR ((fnlwgt BETWEEN 325802 AND 331395) OR income <> '<=50K') OR (race = 'White' OR age BETWEEN 50 AND 55) OR (workclass LIKE '?' OR `native.country` IN ('Outlying-US(Guam-USVI-etc)', 'Honduras', 'Vietnam', 'El-Salvador', 'Holand-Netherlands', 'Trinadad&Tobago', 'Dominican-Republic', 'Cuba', 'Columbia')) AND ((capital BETWEEN -1902 AND 0) AND education <> 'Assoc-voc') GROUP BY occupation,income,`native.country`,`marital.status`,sex,relationship,race

Resulted in 5370 records

Normalized Euclidean distance for (age): 39.31

Hellinger Distance: 0.336

==========

Real

| | marital.status | native.country | education | income | occupation | workclass | relationship | race | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **8294** | Never-married | United-States | Bachelors | <=50K | Other-service | Private | Own-child | Black | 54 | 4 |
| **4878** | Married-civ-spouse | United-States | HS-grad | <=50K | Tech-support | Federal-gov | Husband | White | 58 | 5 |
| **6243** | Never-married | ? | Bachelors | <=50K | Sales | Private | Not-in-family | Black | 37 | 1 |
| **1874** | Married-civ-spouse | ? | 7th-8th | <=50K | Exec-managerial | Self-emp-inc | Husband | White | 45 | 1 |
| **7914** | Never-married | United-States | Assoc-acdm | <=50K | Craft-repair | Private | Not-in-family | White | 44 | 13 |

Synthetic

| | marital.status | native.country | education | income | occupation | workclass | relationship | race | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **8294** | Never-married | United-States | Bachelors | <=50K | Other-service | Private | Own-child | Black | nan | nan |
| **4878** | Married-civ-spouse | United-States | HS-grad | <=50K | Tech-support | Federal-gov | Husband | White | nan | nan |
| **6243** | Never-married | ? | Bachelors | <=50K | Sales | Private | Not-in-family | Black | nan | nan |
| **1874** | Married-civ-spouse | ? | 7th-8th | <=50K | Exec-managerial | Self-emp-inc | Husband | White | nan | nan |
| **7914** | Never-married | United-States | Assoc-acdm | <=50K | Craft-repair | Private | Not-in-family | White | nan | nan |

SQL for Real:
SELECT `marital.status`,`native.country`,education,income,occupation,workclass,relationship,race,MAX(age), COUNT(*) FROM C1 WHERE (capital BETWEEN 0 AND 0) OR (income = '>50K' OR `hours.per.week` = 35) GROUP BY `marital.status`,`native.country`,education,income,occupation,workclass,relationship,race

Resulted in 10936 records

SQL for Synthetic:
SELECT `marital.status`,`native.country`,education,income,occupation,workclass,relationship,race,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (capital BETWEEN 0 AND 0) OR (income = '>50K' OR `hours.per.week` = 35) GROUP BY `marital.status`,`native.country`,education,income,occupation,workclass,relationship,race

Resulted in 1720 records

Normalized Euclidean distance for (age): 23.66

Hellinger Distance: 0.191

==========

Real

| | marital.status | native.country | education | relationship | workclass | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **2596** | Never-married | United-States | Assoc-acdm | Not-in-family | Self-emp-inc | 25.000000 | 1.000000 |
| **413** | Divorced | Puerto-Rico | 7th-8th | Unmarried | Private | nan | nan |
| **1484** | Married-spouse-absent | Hungary | Bachelors | Unmarried | Local-gov | nan | nan |
| **1** | Divorced | ? | 11th | Not-in-family | Self-emp-not-inc | nan | nan |
| **1991** | Never-married | India | Some-college | Other-relative | Private | nan | nan |

Synthetic

| | marital.status | native.country | education | relationship | workclass | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **2596** | Never-married | United-States | Assoc-acdm | Not-in-family | Self-emp-inc | 39.981229 | 1 |
| **413** | Divorced | Puerto-Rico | 7th-8th | Unmarried | Private | 40.031691 | 7 |
| **1484** | Married-spouse-absent | Hungary | Bachelors | Unmarried | Local-gov | 40.031471 | 1 |
| **1** | Divorced | ? | 11th | Not-in-family | Self-emp-not-inc | 40.025133 | 1 |
| **1991** | Never-married | India | Some-college | Other-relative | Private | 39.994124 | 2 |

SQL for Real:
SELECT `marital.status`,`native.country`,education,relationship,workclass,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (sex <> 'Male' OR education LIKE 'Bachelors') GROUP BY `marital.status`,`native.country`,education,relationship,workclass

Resulted in 2286 records

SQL for Synthetic:
SELECT `marital.status`,`native.country`,education,relationship,workclass,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (sex <> 'Male' OR education LIKE 'Bachelors') GROUP BY `marital.status`,`native.country`,education,relationship,workclass

Resulted in 3495 records

Normalized Euclidean distance for (`hours.per.week`): 32.28

Hellinger Distance: 0.29

==========

Real

| | income | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|
| **0** | <=50K | 99 | 36334 |
| **1** | >50K | 99 | 11676 |

Synthetic

| | income | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|
| **0** | <=50K | 40.592070 | 43463.000000 |
| **1** | >50K | 40.563159 | 4759.000000 |

SQL for Real:
SELECT income,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (`marital.status` <> 'Never-married' OR race = 'White') OR (relationship LIKE 'Not-in-family' OR sex <> 'Male') GROUP BY income

Resulted in 2 records

SQL for Synthetic:
SELECT income,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` <> 'Never-married' OR race = 'White') OR (relationship LIKE 'Not-in-family' OR sex <> 'Male') GROUP BY income

Resulted in 2 records

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

Hellinger Distance: 0.138

==========

Real

| | income | race | marital.status | native.country | workclass | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **585** | <=50K | White | Married-civ-spouse | Thailand | ? | nan | nan |
| **1012** | >50K | White | Widowed | United-States | Federal-gov | 45.000000 | 2.000000 |
| **792** | <=50K | White | Separated | Puerto-Rico | Local-gov | 51.000000 | 1.000000 |
| **554** | <=50K | White | Married-civ-spouse | Japan | State-gov | nan | nan |
| **980** | >50K | White | Married-civ-spouse | Yugoslavia | Private | nan | nan |

Synthetic

| | income | race | marital.status | native.country | workclass | MAX(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|
| **585** | <=50K | White | Married-civ-spouse | Thailand | ? | 24 | 3 |
| **1012** | >50K | White | Widowed | United-States | Federal-gov | 59 | 1 |
| **792** | <=50K | White | Separated | Puerto-Rico | Local-gov | 60 | 8 |
| **554** | <=50K | White | Married-civ-spouse | Japan | State-gov | 49 | 2 |
| **980** | >50K | White | Married-civ-spouse | Yugoslavia | Private | 42 | 2 |

SQL for Real:
SELECT income,race,`marital.status`,`native.country`,workclass,MAX(age), COUNT(*) FROM C1 WHERE (sex = 'Female' AND workclass <> 'Self-emp-not-inc') GROUP BY income,race,`marital.status`,`native.country`,workclass

Resulted in 763 records

SQL for Synthetic:
SELECT income,race,`marital.status`,`native.country`,workclass,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (sex = 'Female' AND workclass <> 'Self-emp-not-inc') GROUP BY income,race,`marital.status`,`native.country`,workclass

Resulted in 1016 records

Normalized Euclidean distance for (age): 18.79

Hellinger Distance: 0.355

==========

Real

| | native.country | occupation | income | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|
| **239** | Greece | Sales | <=50K | 140 | 3 |
| **77** | China | Transport-moving | <=50K | 40 | 1 |
| **393** | Jamaica | Prof-specialty | <=50K | 226 | 6 |
| **405** | Japan | Craft-repair | <=50K | 230 | 5 |
| **71** | China | Prof-specialty | <=50K | 668 | 19 |

Synthetic

| | native.country | occupation | income | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|
| **239** | Greece | Sales | <=50K | nan | nan |
| **77** | China | Transport-moving | <=50K | nan | nan |
| **393** | Jamaica | Prof-specialty | <=50K | 80.114613 | 2.000000 |
| **405** | Japan | Craft-repair | <=50K | 5960.462707 | 149.000000 |
| **71** | China | Prof-specialty | <=50K | 600.096475 | 15.000000 |

SQL for Real:
SELECT `native.country`,occupation,income,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (`marital.status` <> 'Never-married') OR (education LIKE 'HS-grad' OR `native.country` = 'United-States') OR (relationship = 'Not-in-family' OR `hours.per.week` BETWEEN 41 AND 40) GROUP BY `native.country`,occupation,income

Resulted in 699 records

SQL for Synthetic:
SELECT `native.country`,occupation,income,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` <> 'Never-married') OR (education LIKE 'HS-grad' OR `native.country` = 'United-States') OR (relationship = 'Not-in-family' OR `hours.per.week` BETWEEN 41 AND 40) GROUP BY `native.country`,occupation,income

Resulted in 608 records

Normalized Euclidean distance for (`hours.per.week`): 20.9

Hellinger Distance: 0.396

==========

Real

| | occupation | education | relationship | income | race | sex | workclass | native.country | marital.status | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **97** | Handlers-cleaners | 10th | Unmarried | <=50K | White | Male | Private | United-States | Never-married | 0.000000 | 3.000000 |
| **64** | Exec-managerial | 10th | Unmarried | <=50K | White | Female | Self-emp-not-inc | United-States | Never-married | nan | nan |
| **19** | ? | 10th | Own-child | <=50K | White | Male | ? | United-States | Never-married | -1721.000000 | 30.000000 |
| **120** | Machine-op-inspct | 10th | Wife | <=50K | White | Female | Private | United-States | Never-married | nan | nan |
| **66** | Farming-fishing | 10th | Not-in-family | <=50K | White | Male | Private | United-States | Never-married | 0.000000 | 2.000000 |

Synthetic

| | occupation | education | relationship | income | race | sex | workclass | native.country | marital.status | MIN(capital) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **97** | Handlers-cleaners | 10th | Unmarried | <=50K | White | Male | Private | United-States | Never-married | -90.573985 | 2 |
| **64** | Exec-managerial | 10th | Unmarried | <=50K | White | Female | Self-emp-not-inc | United-States | Never-married | -35.471121 | 1 |
| **19** | ? | 10th | Own-child | <=50K | White | Male | ? | United-States | Never-married | -242.652523 | 10 |
| **120** | Machine-op-inspct | 10th | Wife | <=50K | White | Female | Private | United-States | Never-married | -26.645626 | 1 |
| **66** | Farming-fishing | 10th | Not-in-family | <=50K | White | Male | Private | United-States | Never-married | -82.138422 | 4 |

SQL for Real:
SELECT occupation,education,relationship,income,race,sex,workclass,`native.country`,`marital.status`,MIN(capital), COUNT(*) FROM C1 WHERE (relationship IN ('Own-child', 'Wife', 'Unmarried') OR race IN ('Amer-Indian-Eskimo', 'Other', 'Asian-Pac-Islander', 'White', 'Black')) AND (`marital.status` LIKE 'Never-married' AND education = '10th') GROUP BY occupation,education,relationship,income,race,sex,workclass,`native.country`,`marital.status`

Resulted in 188 records

SQL for Synthetic:
SELECT occupation,education,relationship,income,race,sex,workclass,`native.country`,`marital.status`,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (relationship IN ('Own-child', 'Wife', 'Unmarried') OR race IN ('Amer-Indian-Eskimo', 'Other', 'Asian-Pac-Islander', 'White', 'Black')) AND (`marital.status` LIKE 'Never-married' AND education = '10th') GROUP BY occupation,education,relationship,income,race,sex,workclass,`native.country`,`marital.status`

Resulted in 223 records

Normalized Euclidean distance for (capital): 8.19

Hellinger Distance: 0.198

==========

Real

| | native.country | occupation | workclass | marital.status | education | sex | relationship | income | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **6919** | Puerto-Rico | Exec-managerial | Self-emp-not-inc | Married-civ-spouse | Assoc-acdm | Female | Wife | <=50K | nan | nan |
| **8839** | Scotland | Adm-clerical | Federal-gov | Divorced | HS-grad | Female | Unmarried | <=50K | nan | nan |
| **4246** | Laos | ? | ? | Never-married | 10th | Male | Own-child | <=50K | nan | nan |
| **7824** | Puerto-Rico | Prof-specialty | Local-gov | Married-civ-spouse | Some-college | Female | Wife | <=50K | nan | nan |
| **9667** | United-States | ? | ? | Separated | 7th-8th | Male | Unmarried | <=50K | nan | nan |

Synthetic

| | native.country | occupation | workclass | marital.status | education | sex | relationship | income | SUM(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **6919** | Puerto-Rico | Exec-managerial | Self-emp-not-inc | Married-civ-spouse | Assoc-acdm | Female | Wife | <=50K | 39.980840 | 1 |
| **8839** | Scotland | Adm-clerical | Federal-gov | Divorced | HS-grad | Female | Unmarried | <=50K | 40.051883 | 1 |
| **4246** | Laos | ? | ? | Never-married | 10th | Male | Own-child | <=50K | 39.546903 | 1 |
| **7824** | Puerto-Rico | Prof-specialty | Local-gov | Married-civ-spouse | Some-college | Female | Wife | <=50K | 79.998304 | 2 |
| **9667** | United-States | ? | ? | Separated | 7th-8th | Male | Unmarried | <=50K | 40.019019 | 1 |

SQL for Real:
SELECT `native.country`,occupation,workclass,`marital.status`,education,sex,relationship,income,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (capital >= 15024 OR workclass IN ('State-gov', 'Federal-gov', 'Private', 'Self-emp-inc', 'Local-gov', 'Self-emp-not-inc', '?')) OR (age <= 56 OR education = 'Bachelors') GROUP BY `native.country`,occupation,workclass,`marital.status`,education,sex,relationship,income

Resulted in 10400 records

SQL for Synthetic:
SELECT `native.country`,occupation,workclass,`marital.status`,education,sex,relationship,income,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (capital >= 15024 OR workclass IN ('State-gov', 'Federal-gov', 'Private', 'Self-emp-inc', 'Local-gov', 'Self-emp-not-inc', '?')) OR (age <= 56 OR education = 'Bachelors') GROUP BY `native.country`,occupation,workclass,`marital.status`,education,sex,relationship,income

Resulted in 14639 records

Normalized Euclidean distance for (`hours.per.week`): 62.27

Hellinger Distance: 0.26

==========

Real

| | occupation | relationship | income | race | sex | marital.status | native.country | workclass | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **4567** | Other-service | Own-child | <=50K | White | Female | Married-civ-spouse | United-States | Private | 177.000000 | 6.000000 |
| **1640** | Craft-repair | Not-in-family | <=50K | White | Male | Widowed | Guatemala | Private | nan | nan |
| **2178** | Exec-managerial | Not-in-family | <=50K | White | Female | Divorced | ? | State-gov | nan | nan |
| **5068** | Prof-specialty | Husband | <=50K | Asian-Pac-Islander | Male | Married-civ-spouse | India | Private | 266.000000 | 7.000000 |
| **546** | Adm-clerical | Not-in-family | <=50K | White | Female | Never-married | Puerto-Rico | Federal-gov | nan | nan |

Synthetic

| | occupation | relationship | income | race | sex | marital.status | native.country | workclass | SUM(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|---|
| **4567** | Other-service | Own-child | <=50K | White | Female | Married-civ-spouse | United-States | Private | 304 | 11 |
| **1640** | Craft-repair | Not-in-family | <=50K | White | Male | Widowed | Guatemala | Private | 34 | 1 |
| **2178** | Exec-managerial | Not-in-family | <=50K | White | Female | Divorced | ? | State-gov | 38 | 1 |
| **5068** | Prof-specialty | Husband | <=50K | Asian-Pac-Islander | Male | Married-civ-spouse | India | Private | 76 | 2 |
| **546** | Adm-clerical | Not-in-family | <=50K | White | Female | Never-married | Puerto-Rico | Federal-gov | 268 | 7 |

SQL for Real:
SELECT occupation,relationship,income,race,sex,`marital.status`,`native.country`,workclass,SUM(age), COUNT(*) FROM C1 WHERE (race LIKE 'White' OR workclass = 'Private') OR (sex <> 'Male' OR occupation IN ('Craft-repair', 'Exec-managerial', 'Prof-specialty', 'Sales', 'Transport-moving', 'Adm-clerical', 'Other-service', '?', 'Priv-house-serv', 'Handlers-cleaners', 'Armed-Forces', 'Tech-support')) AND (`native.country` <> 'United-States' AND capital BETWEEN 0 AND 0) AND (fnlwgt = 52537 AND `marital.status` = 'Divorced') GROUP BY occupation,relationship,income,race,sex,`marital.status`,`native.country`,workclass

Resulted in 5318 records

SQL for Synthetic:
SELECT occupation,relationship,income,race,sex,`marital.status`,`native.country`,workclass,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (race LIKE 'White' OR workclass = 'Private') OR (sex <> 'Male' OR occupation IN ('Craft-repair', 'Exec-managerial', 'Prof-specialty', 'Sales', 'Transport-moving', 'Adm-clerical', 'Other-service', '?', 'Priv-house-serv', 'Handlers-cleaners', 'Armed-Forces', 'Tech-support')) AND (`native.country` <> 'United-States' AND capital BETWEEN 0 AND 0) AND (fnlwgt = 52537 AND `marital.status` = 'Divorced') GROUP BY occupation,relationship,income,race,sex,`marital.status`,`native.country`,workclass

Resulted in 7754 records

Normalized Euclidean distance for (age): 46.82

Hellinger Distance: 0.304

==========

Real

| education | occupation | workclass | MIN(fnlwgt) | COUNT(*) |
|---|---|---|---|---|

Synthetic

| education | occupation | workclass | MIN(fnlwgt) | COUNT(*) |
|---|---|---|---|---|

SQL for Real:
SELECT education,occupation,workclass,MIN(fnlwgt), COUNT(*) FROM C1 WHERE (education IN ('5th-6th', 'HS-grad', 'Bachelors', '12th', 'Assoc-acdm', 'Masters', 'Prof-school', '7th-8th', '11th', 'Preschool', '9th', 'Some-college', 'Doctorate', 'Assoc-voc', '1st-4th') OR `native.country` <> 'United-States') AND (`hours.per.week` < 40 AND fnlwgt = 47707) GROUP BY education,occupation,workclass

Resulted in 0 records

SQL for Synthetic:
SELECT education,occupation,workclass,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (education IN ('5th-6th', 'HS-grad', 'Bachelors', '12th', 'Assoc-acdm', 'Masters', 'Prof-school', '7th-8th', '11th', 'Preschool', '9th', 'Some-college', 'Doctorate', 'Assoc-voc', '1st-4th') OR `native.country` <> 'United-States') AND (`hours.per.week` < 40 AND fnlwgt = 47707) GROUP BY education,occupation,workclass

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

==========

Real

| | sex | marital.status | education | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|
| **14** | Male | Married-civ-spouse | Preschool | 0 | 18 |
| **9** | Male | Married-civ-spouse | Assoc-voc | 0 | 395 |
| **15** | Male | Married-civ-spouse | Prof-school | 0 | 205 |
| **0** | Male | Married-AF-spouse | HS-grad | 0 | 1 |
| **2** | Male | Married-civ-spouse | 11th | 0 | 256 |

Synthetic

| | sex | marital.status | education | MAX(capital) | COUNT(*) |
|---|---|---|---|---|---|
| **14** | Male | Married-civ-spouse | Preschool | nan | nan |
| **9** | Male | Married-civ-spouse | Assoc-voc | nan | nan |
| **15** | Male | Married-civ-spouse | Prof-school | nan | nan |
| **0** | Male | Married-AF-spouse | HS-grad | nan | nan |
| **2** | Male | Married-civ-spouse | 11th | nan | nan |

SQL for Real:
SELECT sex,`marital.status`,education,MAX(capital), COUNT(*) FROM C1 WHERE (`hours.per.week` = 40 OR `marital.status` LIKE 'Married-civ-spouse') AND (sex = 'Male'

AND fnlwgt >= 158451) AND ((capital BETWEEN 0 AND 0) AND relationship LIKE 'Husband') GROUP BY sex,`marital.status`,education

Resulted in 17 records

SQL for Synthetic:
SELECT sex,`marital.status`,education,MAX(capital), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` = 40 OR `marital.status` LIKE 'Married-civ-spouse') AND (sex = 'Male' AND fnlwgt >= 158451) AND ((capital BETWEEN 0 AND 0) AND relationship LIKE 'Husband') GROUP BY sex,`marital.status`,education

Resulted in 0 records

Normalized Euclidean distance for (capital): nan

Hellinger Distance: nan

==========

Real

| | sex | occupation | marital.status | workclass | relationship | income | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **809** | Female | Sales | Married-civ-spouse | Private | Other-relative | >50K | 24 | 1 |
| **1036** | Male | Adm-clerical | Divorced | Private | Unmarried | >50K | 45 | 1 |
| **431** | Female | Machine-op-inspct | Married-civ-spouse | Private | Not-in-family | <=50K | 40 | 2 |
| **1204** | Male | Craft-repair | Never-married | Private | Other-relative | <=50K | 60 | 73 |
| **2055** | Male | Tech-support | Separated | Federal-gov | Unmarried | <=50K | 40 | 1 |

Synthetic

| | sex | occupation | marital.status | workclass | relationship | income | MAX(`hours.per.week`) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|
| **809** | Female | Sales | Married-civ-spouse | Private | Other-relative | >50K | nan | nan |
| **1036** | Male | Adm-clerical | Divorced | Private | Unmarried | >50K | nan | nan |
| **431** | Female | Machine-op-inspct | Married-civ-spouse | Private | Not-in-family | <=50K | 40.056874 | 4.000000 |
| **1204** | Male | Craft-repair | Never-married | Private | Other-relative | <=50K | 40.081915 | 65.000000 |
| **2055** | Male | Tech-support | Separated | Federal-gov | Unmarried | <=50K | nan | nan |

SQL for Real:
SELECT sex,occupation,`marital.status`,workclass,relationship,income,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (`hours.per.week` > 60) AND (age <> 31 AND income LIKE '>50K') OR (race IN ('Black', 'White') OR sex IN ('Male', 'Female')) OR (occupation LIKE 'Transport-moving' OR education <> 'Some-college') AND (relationship <> 'Not-in-family' AND capital BETWEEN 0 AND 0) OR (fnlwgt <= 356934 OR `native.country` <> 'United-States') GROUP BY sex,occupation,`marital.status`,workclass,relationship,income

Resulted in 2148 records

SQL for Synthetic:
SELECT sex,occupation,`marital.status`,workclass,relationship,income,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` > 60) AND (age <> 31 AND income LIKE '>50K') OR (race IN ('Black', 'White') OR sex IN ('Male', 'Female')) OR (occupation LIKE 'Transport-moving' OR education <> 'Some-college') AND (relationship <> 'Not-in-family' AND capital BETWEEN 0 AND 0) OR (fnlwgt <= 356934 OR `native.country` <> 'United-States') GROUP BY sex,occupation,`marital.status`,workclass,relationship,income

Resulted in 2126 records

Normalized Euclidean distance for (`hours.per.week`): 38.99

Hellinger Distance: 0.175

==========

Real

| | occupation | sex | income | native.country | relationship | marital.status | workclass | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **597** | Adm-clerical | Female | <=50K | Philippines | Wife | Married-civ-spouse | Private | 36.000000 | 1.000000 |
| **2177** | Exec-managerial | Male | <=50K | Columbia | Not-in-family | Married-spouse-absent | Self-emp-not-inc | nan | nan |
| **1475** | Craft-repair | Male | <=50K | Mexico | Not-in-family | Divorced | Private | 41.000000 | 2.000000 |
| **2608** | Farming-fishing | Male | <=50K | Columbia | Not-in-family | Divorced | Self-emp-not-inc | nan | nan |
| **4685** | Prof-specialty | Female | <=50K | United-States | Not-in-family | Divorced | Self-emp-not-inc | 48.190476 | 21.000000 |

Synthetic

| | occupation | sex | income | native.country | relationship | marital.status | workclass | AVG(age) | COUNT(*) |
|---|---|---|---|---|---|---|---|---|---|
| **597** | Adm-clerical | Female | <=50K | Philippines | Wife | Married-civ-spouse | Private | 39.000000 | 3 |
| **2177** | Exec-managerial | Male | <=50K | Columbia | Not-in-family | Married-spouse-absent | Self-emp-not-inc | 33.000000 | 1 |
| **1475** | Craft-repair | Male | <=50K | Mexico | Not-in-family | Divorced | Private | 50.000000 | 1 |
| **2608** | Farming-fishing | Male | <=50K | Columbia | Not-in-family | Divorced | Self-emp-not-inc | 46.000000 | 1 |
| **4685** | Prof-specialty | Female | <=50K | United-States | Not-in-family | Divorced | Self-emp-not-inc | 48.500000 | 12 |

SQL for Real:
SELECT occupation,sex,income,`native.country`,relationship,`marital.status`,workclass,AVG(age), COUNT(*) FROM C1 WHERE (relationship LIKE 'Own-child' OR `hours.per.week` BETWEEN 40 AND 40) AND (sex LIKE 'Female' AND capital BETWEEN 0 AND 0) AND (race IN ('Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'Black') AND `native.country` LIKE 'United-States') OR (education = 'Some-college' OR fnlwgt <> 236396) OR (`marital.status` = 'Married-civ-spouse' OR occupation LIKE 'Adm-clerical') GROUP BY occupation,sex,income,`native.country`,relationship,`marital.status`,workclass

Resulted in 4878 records

SQL for Synthetic:
SELECT occupation,sex,income,`native.country`,relationship,`marital.status`,workclass,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (relationship LIKE 'Own-child' OR `hours.per.week` BETWEEN 40 AND 40) AND (sex LIKE 'Female' AND capital BETWEEN 0 AND 0) AND (race IN ('Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'Black') AND `native.country` LIKE 'United-States') OR (education = 'Some-college' OR fnlwgt <> 236396) OR (`marital.status` = 'Married-civ-spouse' OR occupation LIKE 'Adm-clerical') GROUP BY occupation,sex,income,`native.country`,relationship,`marital.status`,workclass

Resulted in 6808 records

Normalized Euclidean distance for (age): 45.73

Hellinger Distance: 0.311

==========

Hellinger Distance Summary: {'mean': 0.2210146331272294, 'median': 0.23770563481546514, 'stddev': 0.11083381530412842}

Euclidean distance Summary : {'mean': 30.38649259011477, 'median': 31.07680100320278, 'stddev': 21.201374421796952}

==========