

Real											
	education	sex	income	native.country	workclass	race	marital.status	relationship	occupation	SUM(capital)	COUNT(*)
16031	Some-college	Male	<=50K	Puerto-Rico	Private	White	Divorced	Own-child	Tech-support	nan	nan
3443	Assoc-acdm	Female	<=50K	Scotland	Private	White	Divorced	Not-in-family	Prof-specialty	nan	nan
958	11th	Female	<=50K	United-States	Private	White	Divorced	Unmarried	Sales	0.000000	6.000000
4633	Assoc-voc	Male	<=50K	Columbia	Local-gov	White	Divorced	Not-in-family	Machine-op-inspct	nan	nan
12351	Masters	Female	<=50K	?	Local-gov	White	Never-married	Unmarried	Prof-specialty	nan	nan

Synthetic											
	education	sex	income	native.country	workclass	race	marital.status	relationship	occupation	SUM(capital)	COUNT(*)
16031	Some-college	Male	<=50K	Puerto-Rico	Private	White	Divorced	Own-child	Tech-support	-90.607329	2
3443	Assoc-acdm	Female	<=50K	Scotland	Private	White	Divorced	Not-in-family	Prof-specialty	4.231951	1
958	11th	Female	<=50K	United-States	Private	White	Divorced	Unmarried	Sales	-250.396775	2
4633	Assoc-voc	Male	<=50K	Columbia	Local-gov	White	Divorced	Not-in-family	Machine-op-inspct	152.754300	1
12351	Masters	Female	<=50K	?	Local-gov	White	Never-married	Unmarried	Prof-specialty	-48.858825	1

SQL for Real:
SELECT education,sex,income,`native.country`,workclass,race,`marital.status`,relationship,occupation,SUM(capital), COUNT(*) FROM C1 WHERE (`native.country` <> 'United-States' AND fnlwgt = 167868) AND (`marital.status` <> 'Married-civ-spouse' AND `hours.per.week` < 49) OR (sex = 'Male' OR capital < -2377) AND (occupation <> 'Prof-specialty' AND education <> 'Bachelors') OR (relationship IN ('Not-in-family', 'Wife', 'Unmarried', 'Husband') OR age BETWEEN 31 AND 46) OR (race = 'White' OR workclass LIKE 'Private') GROUP BY education,sex,income,`native.country`,workclass,race,`marital.status`,relationship,occupation

Resulted in 12552 records

SQL for Synthetic:
SELECT education,sex,income,`native.country`,workclass,race,`marital.status`,relationship,occupation,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (`native.country` <> 'United-States' AND fnlwgt = 167868) AND (`marital.status` <> 'Married-civ-spouse' AND `hours.per.week` < 49) OR (sex = 'Male' OR capital < -2377) AND (occupation <> 'Prof-specialty' AND education <> 'Bachelors') OR (relationship IN ('Not-in-family', 'Wife', 'Unmarried', 'Husband') OR age BETWEEN 31 AND 46) OR (race = 'White' OR workclass LIKE 'Private') GROUP BY education,sex,income,`native.country`,workclass,race,`marital.status`,relationship,occupation

Resulted in 17007 records

Normalized Euclidean distance for (capital): 64.44

Hellinger Distance: 0.252

=====

Real					
	race	occupation	relationship	MAX(`hours.per.week`)	COUNT(*)
75	Black	?	Not-in-family	60.000000	36.000000
95	Black	Farming-fishing	Own-child	60.000000	12.000000
77	Black	?	Own-child	48.000000	88.000000
1	Amer-Indian-Eskimo	?	Unmarried	40.000000	2.000000
131	Black	Transport-moving	Not-in-family	80.000000	34.000000

Synthetic					
	race	occupation	relationship	MAX(`hours.per.week`)	COUNT(*)
75	Black	?	Not-in-family	40.049650	28
95	Black	Farming-fishing	Own-child	40.050697	11
77	Black	?	Own-child	40.058926	91
1	Amer-Indian-Eskimo	?	Unmarried	39.956267	1
131	Black	Transport-moving	Not-in-family	40.070279	18

SQL for Real:
SELECT race,occupation,relationship,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (`marital.status` = 'Never-married') GROUP BY race,occupation,relationship

Resulted in 253 records

SQL for Synthetic:
SELECT race,occupation,relationship,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` = 'Never-married') GROUP BY race,occupation,relationship

Resulted in 258 records

Normalized Euclidean distance for (`hours.per.week`): 14.8

Hellinger Distance: 0.078

=====

Real										
	workclass	native.country	occupation	race	income	marital.status	sex	relationship	SUM(fnlwgt)	COUNT(*)
8422	State-gov	Puerto-Rico	Adm-clerical	White	<=50K	Never-married	Female	Unmarried	nan	nan
262	?	Puerto-Rico	?	White	<=50K	Separated	Female	Unmarried	nan	nan
7283	Self-emp-inc	United-States	Handlers-cleaners	White	<=50K	Divorced	Male	Not-in-family	nan	nan
7158	Self-emp-inc	Puerto-Rico	Sales	Black	<=50K	Divorced	Female	Unmarried	nan	nan
8339	State-gov	Iran	Prof-specialty	White	<=50K	Never-married	Male	Not-in-family	nan	nan

Synthetic										
	workclass	native.country	occupation	race	income	marital.status	sex	relationship	SUM(fnlwgt)	COUNT(*)
8422	State-gov	Puerto-Rico	Adm-clerical	White	<=50K	Never-married	Female	Unmarried	690949.800936	4
262	?	Puerto-Rico	?	White	<=50K	Separated	Female	Unmarried	1384577.296412	8
7283	Self-emp-inc	United-States	Handlers-cleaners	White	<=50K	Divorced	Male	Not-in-family	140976.038166	1
7158	Self-emp-inc	Puerto-Rico	Sales	Black	<=50K	Divorced	Female	Unmarried	169416.911395	1
8339	State-gov	Iran	Prof-specialty	White	<=50K	Never-married	Male	Not-in-family	198805.723353	1

SQL for Real:
SELECT workclass,`native.country`,occupation,race,income,`marital.status`,sex,relationship,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (`marital.status` = 'Married-civ-spouse' AND sex = 'Male') OR (workclass IN ('Self-emp-inc', 'Local-gov', 'Self-emp-not-inc', 'Federal-gov', '?', 'State-gov', 'Private') OR fnlwgt BETWEEN 120471 AND 234460) OR (age <> 43 OR income = '>50K') OR (occupation <> 'Adm-clerical' OR `hours.per.week` BETWEEN 36 AND 70) GROUP BY workclass,`native.country`,occupation,race,income,`marital.status`,sex,relationship

Resulted in 6492 records

SQL for Synthetic:
SELECT workclass,`native.country`,occupation,`race`,`income`,`marital.status`,`sex`,`relationship`,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` = 'Married-civ-spouse' AND sex = 'Male') OR (workclass IN ('Self-emp-inc', 'Local-gov', 'Self-emp-not-inc', 'Federal-gov', '?', 'State-gov', 'Private') OR fnlwgt BETWEEN 120471 AND 234460) OR (age <> 43 OR income = '>50K') OR (occupation <> 'Adm-clerical' OR `hours.per.week` BETWEEN 36 AND 70) GROUP BY workclass,`native.country`,occupation,`race`,`income`,`marital.status`,`sex`,`relationship`

Resulted in 8900 records

Normalized Euclidean distance for (fnlwgt): 49.69

Hellinger Distance: 0.303

=====

Real									
	marital.status	income	education	workclass	occupation	relationship	native.country	MIN(capital)	COUNT(*)
11954	Separated	<=50K	Masters	Local-gov	Prof-specialty	Not-in-family	Scotland	nan	nan
12018	Separated	<=50K	Some-college	Local-gov	Other-service	Unmarried	Puerto-Rico	nan	nan
9887	Never-married	<=50K	HS-grad	Private	Sales	Unmarried	?	nan	nan
3283	Married-civ-spouse	<=50K	Assoc-acdm	Local-gov	Craft-repair	Husband	Columbia	nan	nan
1369	Divorced	<=50K	HS-grad	Private	Machine-op-inspct	Other-relative	Columbia	nan	nan

Synthetic									
	marital.status	income	education	workclass	occupation	relationship	native.country	MIN(capital)	COUNT(*)
11954	Separated	<=50K	Masters	Local-gov	Prof-specialty	Not-in-family	Scotland	-83.530398	1
12018	Separated	<=50K	Some-college	Local-gov	Other-service	Unmarried	Puerto-Rico	42.156971	1
9887	Never-married	<=50K	HS-grad	Private	Sales	Unmarried	?	-169.286342	1
3283	Married-civ-spouse	<=50K	Assoc-acdm	Local-gov	Craft-repair	Husband	Columbia	-193.545215	2
1369	Divorced	<=50K	HS-grad	Private	Machine-op-inspct	Other-relative	Columbia	60.661955	1

SQL for Real:
SELECT `marital.status`,`income`,`education`,`workclass`,`occupation`,`relationship`,`native.country`,`MIN(capital), COUNT(*) FROM C1 WHERE (race IN ('Black', 'Amer-Indian-Eskimo', 'Other', 'White', 'Asian-Pac-Islander') AND `hours.per.week` BETWEEN 30 AND 60) OR (`marital.status` <> 'Never-married' OR occupation <> '?') AND ((age BETWEEN 20 AND 31) AND capital BETWEEN 0 AND 0) OR (sex LIKE 'Male' OR workclass <> 'Private') AND (education LIKE 'HS-grad' AND `native.country` <> 'United-States') AND ((fnlwgt BETWEEN 132551 AND 68006) AND relationship LIKE 'Not-in-family') GROUP BY `marital.status`,`income`,`education`,`workclass`,`occupation`,`relationship`,`native.country`

Resulted in 8339 records

SQL for Synthetic:
SELECT `marital.status`,`income`,`education`,`workclass`,`occupation`,`relationship`,`native.country`,`MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (race IN ('Black', 'Amer-Indian-Eskimo', 'Other', 'White', 'Asian-Pac-Islander') AND `hours.per.week` BETWEEN 30 AND 60) OR (`marital.status` <> 'Never-married' OR occupation <> '?') AND ((age BETWEEN 20 AND 31) AND capital BETWEEN 0 AND 0) OR (sex LIKE 'Male' OR workclass <> 'Private') AND (education LIKE 'HS-grad' AND `native.country` <> 'United-States') AND ((fnlwgt BETWEEN 132551 AND 68006) AND relationship LIKE 'Not-in-family') GROUP BY `marital.status`,`income`,`education`,`workclass`,`occupation`,`relationship`,`native.country`

Resulted in 12979 records

Normalized Euclidean distance for (capital): 58.05

Hellinger Distance: 0.267

=====

Real						
	workclass	occupation	relationship	native.country	AVG(`hours.per.week`)	COUNT(*)
132	Local-gov	Tech-support	Husband	United-States	42.600000	5.000000
125	Local-gov	Protective-serv	Not-in-family	Philippines	nan	nan
113	Local-gov	Prof-specialty	Own-child	Scotland	nan	nan
454	Private	Tech-support	Wife	United-States	39.555556	9.000000
169	Private	Adm-clerical	Own-child	Nicaragua	nan	nan

Synthetic						
	workclass	occupation	relationship	native.country	AVG(`hours.per.week`)	COUNT(*)
132	Local-gov	Tech-support	Husband	United-States	40.026417	1
125	Local-gov	Protective-serv	Not-in-family	Philippines	40.064246	1
113	Local-gov	Prof-specialty	Own-child	Scotland	40.048953	1
454	Private	Tech-support	Wife	United-States	39.972136	5
169	Private	Adm-clerical	Own-child	Nicaragua	39.474984	1

SQL for Real:
SELECT workclass,occupation,relationship,`native.country`,`AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (education LIKE 'Assoc-acdm') OR (workclass = 'Private' AND race <> 'White') AND (`native.country` = 'United-States' AND age > 32) OR (income = '>50K' OR `hours.per.week` BETWEEN 32 AND 40) AND (fnlwgt = 67001 AND occupation IN ('?', 'Adm-clerical', 'Farming-fishing', 'Transport-moving', 'Tech-support', 'Other-service', 'Handlers-cleaners', 'Machine-op-inspct', 'Protective-serv', 'Priv-house-serv', 'Prof-specialty', 'Armed-Forces', 'Craft-repair', 'Exec-managerial', 'Sales')) AND ((capital BETWEEN 0 AND 0) AND relationship IN ('Husband', 'Unmarried', 'Other-relative')) GROUP BY workclass,occupation,relationship,`native.country`

Resulted in 343 records

SQL for Synthetic:
SELECT workclass,occupation,relationship,`native.country`,`AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (education LIKE 'Assoc-acdm') OR (workclass = 'Private' AND race <> 'White') AND (`native.country` = 'United-States' AND age > 32) OR (income = '>50K' OR `hours.per.week` BETWEEN 32 AND 40) AND (fnlwgt = 67001 AND occupation IN ('?', 'Adm-clerical', 'Farming-fishing', 'Transport-moving', 'Tech-support', 'Other-service', 'Handlers-cleaners', 'Machine-op-inspct', 'Protective-serv', 'Priv-house-serv', 'Prof-specialty', 'Armed-Forces', 'Craft-repair', 'Exec-managerial', 'Sales')) AND ((capital BETWEEN 0 AND 0) AND relationship IN ('Husband', 'Unmarried', 'Other-relative')) GROUP BY workclass,occupation,relationship,`native.country`

Resulted in 599 records

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

Hellinger Distance: 0.128

=====

Real											
	workclass	sex	relationship	race	education	marital.status	native.country	income	occupation	MIN(fnlwgt)	COUNT(*)
1550	Local-gov	Male	Husband	White	Assoc-voc	Married-civ-spouse	United-States	<=50K	Tech-support	242150	1
5473	Private	Male	Other-relative	Amer-Indian-Eskimo	Some-college	Never-married	United-States	<=50K	Craft-repair	21755	1
1290	Local-gov	Female	Unmarried	White	Assoc-acdm	Divorced	United-States	<=50K	Prof-specialty	119199	1
4072	Private	Male	Husband	Asian-Pac-Islander	Bachelors	Married-civ-spouse	Thailand	>50K	Sales	190319	1
1858	Local-gov	Male	Own-child	White	Masters	Never-married	United-States	<=50K	Prof-specialty	110965	3

Synthetic											
	workclass	sex	relationship	race	education	marital.status	native.country	income	occupation	MIN(fnlwgt)	COUNT(*)
1550	Local-gov	Male	Husband	White	Assoc-voc	Married-civ-spouse	United-States	<=50K	Tech-support	nan	nan
5473	Private	Male	Other-relative	Amer-Indian-Eskimo	Some-college	Never-married	United-States	<=50K	Craft-repair	nan	nan
1290	Local-gov	Female	Unmarried	White	Assoc-acdm	Divorced	United-States	<=50K	Prof-specialty	nan	nan
4072	Private	Male	Husband	Asian-Pac-Islander	Bachelors	Married-civ-spouse	Thailand	>50K	Sales	nan	nan
1858	Local-gov	Male	Own-child	White	Masters	Never-married	United-States	<=50K	Prof-specialty	nan	nan

SQL for Real:
SELECT workclass,sex,relationship,race,education,`marital.status`,`native.country`,income,occupation,MIN(fnlwgt), COUNT(*) FROM C1 WHERE (capital BETWEEN 0 AND 0) AND (fnlwgt > 188436 OR age > 33) AND (`native.country` LIKE 'United-States' AND relationship IN ('Other-relative', 'Unmarried', 'Not-in-family', 'Own-child', 'Wife', 'Husband')) OR (education LIKE 'Doctorate' OR race <> 'White') AND (workclass <> '?' AND `hours.per.week` BETWEEN 35 AND 40) GROUP BY workclass,sex,relationship,race,education,`marital.status`,`native.country`,income,occupation

Resulted in 8303 records

SQL for Synthetic:
SELECT workclass,sex,relationship,race,education,`marital.status`,`native.country`,income,occupation,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (capital BETWEEN 0 AND 0) AND (fnlwgt > 188436 OR age > 33) AND (`native.country` LIKE 'United-States' AND relationship IN ('Other-relative', 'Unmarried', 'Not-in-family', 'Own-child', 'Wife', 'Husband')) OR (education LIKE 'Doctorate' OR race <> 'White') AND (workclass <> '?' AND `hours.per.week` BETWEEN 35 AND 40) GROUP BY workclass,sex,relationship,race,education,`marital.status`,`native.country`,income,occupation

Resulted in 2136 records

Normalized Euclidean distance for (fnlwgt): 22.36

Hellinger Distance: 0.265

=====

Real							
	income	marital.status	education	native.country	race	sex	MIN(`hours.per.week`) COUNT(*)
164	<=50K	Widowed	5th-6th	United-States	White	Female	40 3
268	>50K	Never-married	Some-college	United-States	White	Female	12 16
131	<=50K	Separated	11th	United-States	White	Male	20 16
102	<=50K	Never-married	1st-4th	United-States	White	Female	24 1
265	>50K	Never-married	Masters	United-States	White	Male	32 42

Synthetic							
	income	marital.status	education	native.country	race	sex	MIN(`hours.per.week`) COUNT(*)
164	<=50K	Widowed	5th-6th	United-States	White	Female	nan nan
268	>50K	Never-married	Some-college	United-States	White	Female	nan nan
131	<=50K	Separated	11th	United-States	White	Male	nan nan
102	<=50K	Never-married	1st-4th	United-States	White	Female	nan nan
265	>50K	Never-married	Masters	United-States	White	Male	nan nan

SQL for Real:
SELECT income,`marital.status`,education,`native.country`,race,sex,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (`native.country` LIKE 'United-States') AND (race = 'White' AND capital BETWEEN 0 AND 0) GROUP BY income,`marital.status`,education,`native.country`,race,sex

Resulted in 298 records

SQL for Synthetic:
SELECT income,`marital.status`,education,`native.country`,race,sex,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`native.country` LIKE 'United-States') AND (race = 'White' AND capital BETWEEN 0 AND 0) GROUP BY income,`marital.status`,education,`native.country`,race,sex

Resulted in 0 records

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

Hellinger Distance: nan

=====

Real				
	workclass	marital.status	SUM(capital)	COUNT(*)
39	Self-emp-not-inc	Widowed	160168	127
5	?	Separated	9070	97
32	Self-emp-inc	Widowed	196084	41
40	State-gov	Divorced	126774	316
33	Self-emp-not-inc	Divorced	864447	432

Synthetic				
	workclass	marital.status	SUM(capital)	COUNT(*)
39	Self-emp-not-inc	Widowed	-2147.345970	144.000000
5	?	Separated	462.292496	85.000000
32	Self-emp-inc	Widowed	-95.740013	44.000000
40	State-gov	Divorced	215.444758	348.000000
33	Self-emp-not-inc	Divorced	-3529.007438	412.000000

SQL for Real:
SELECT workclass,`marital.status`,SUM(capital), COUNT(*) FROM C1 WHERE (sex IN ('Female', 'Male') OR race <> 'White') GROUP BY workclass,`marital.status`

Resulted in 47 records

SQL for Synthetic:
SELECT workclass,`marital.status`,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (sex IN ('Female', 'Male') OR race <> 'White') GROUP BY workclass,`marital.status`
Resulted in 47 records

Normalized Euclidean distance for (capital): 6.86

Hellinger Distance: 0.018

=====

Real							
	income	sex	relationship	education	marital.status	SUM(fnlwgt)	COUNT(*)
258	<=50K	Female	Unmarried	Bachelors	Married-spouse-absent	1875306	12
505	<=50K	Male	Own-child	Assoc-acdm	Divorced	1347024	7
491	<=50K	Male	Own-child	12th	Married-civ-spouse	814377	3
170	<=50K	Female	Own-child	7th-8th	Separated	446467	2
165	<=50K	Female	Own-child	5th-6th	Never-married	1861128	9

Synthetic							
	income	sex	relationship	education	marital.status	SUM(fnlwgt)	COUNT(*)
258	<=50K	Female	Unmarried	Bachelors	Married-spouse-absent	2113435.429239	12.000000
505	<=50K	Male	Own-child	Assoc-acdm	Divorced	2538802.443326	14.000000
491	<=50K	Male	Own-child	12th	Married-civ-spouse	746505.900045	4.000000
170	<=50K	Female	Own-child	7th-8th	Separated	nan	nan
165	<=50K	Female	Own-child	5th-6th	Never-married	721228.099998	4.000000

SQL for Real:
SELECT income,sex,relationship,education,`marital.status`,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (`hours.per.week` = 40) OR ((capital BETWEEN 4386 AND 0) OR relationship = 'Not-in-family') OR (education LIKE '7th-8th' OR `marital.status` LIKE 'Married-civ-spouse') AND (sex <> 'Male' AND occupation = 'Adm-clerical') OR (income = '<=50K' OR race = 'White') OR (`native.country` LIKE 'United-States' OR fnlwgt = 271466) GROUP BY income,sex,relationship,education, `marital.status`
Resulted in 870 records

SQL for Synthetic:
SELECT income,sex,relationship,education,`marital.status`,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` = 40) OR ((capital BETWEEN 4386 AND 0) OR relationship = 'Not-in-family') OR (education LIKE '7th-8th' OR `marital.status` LIKE 'Married-civ-spouse') AND (sex <> 'Male' AND occupation = 'Adm-clerical') OR (income = '<=50K' OR race = 'White') OR (`native.country` LIKE 'United-States' OR fnlwgt = 271466) GROUP BY income,sex,relationship,education,`marital.status`
Resulted in 836 records

Normalized Euclidean distance for (fnlwgt): 25.77

Hellinger Distance: 0.171

=====

Real										
	occupation	sex	native.country	marital.status	race	workclass	income	MAX(age)	COUNT(*)	
78	?	Female	Peru	Divorced	White	Local-gov	<=50K	nan	nan	
4305	Prof-specialty	Female	United-States	Widowed	Black	Local-gov	<=50K	66.000000	5.000000	
1295	Craft-repair	Male	Puerto-Rico	Married-civ-spouse	Black	Federal-gov	<=50K	nan	nan	
991	Craft-repair	Female	Puerto-Rico	Married-civ-spouse	White	Private	<=50K	nan	nan	
2231	Farming-fishing	Female	Puerto-Rico	Divorced	White	State-gov	<=50K	nan	nan	

Synthetic										
	occupation	sex	native.country	marital.status	race	workclass	income	MAX(age)	COUNT(*)	
78	?	Female	Peru	Divorced	White	Local-gov	<=50K	41	1	
4305	Prof-specialty	Female	United-States	Widowed	Black	Local-gov	<=50K	84	5	
1295	Craft-repair	Male	Puerto-Rico	Married-civ-spouse	Black	Federal-gov	<=50K	55	1	
991	Craft-repair	Female	Puerto-Rico	Married-civ-spouse	White	Private	<=50K	60	2	
2231	Farming-fishing	Female	Puerto-Rico	Divorced	White	State-gov	<=50K	51	1	

SQL for Real:
SELECT occupation,sex,`native.country`, `marital.status`,race,workclass,income,MAX(age), COUNT(*) FROM C1 WHERE (`native.country` <> 'United-States') OR (capital <> -2457 OR `marital.status` LIKE 'Never-married') AND (`hours.per.week` = 60 AND education = 'Some-college') OR (age = 41 OR occupation <> 'Adm-clerical') GROUP BY occupation,sex,`native.country`, `marital.status`,race,workclass,income
Resulted in 4526 records

SQL for Synthetic:
SELECT occupation,sex,`native.country`, `marital.status`,race,workclass,income,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (`native.country` <> 'United-States') OR (capital <> -2457 OR `marital.status` LIKE 'Never-married') AND (`hours.per.week` = 60 AND education = 'Some-college') OR (age = 41 OR occupation <> 'Adm-clerical') GROUP BY occupation,sex,`native.country`, `marital.status`,race,workclass,income
Resulted in 6148 records

Normalized Euclidean distance for (age): 42.41

Hellinger Distance: 0.328

=====

Real				
	race	income	native.country	MIN(age) COUNT(*)
198	White	>50K	Jamaica	31.000000 1.000000
163	White	<=50K	Jamaica	21.000000 4.000000
103	Black	>50K	Peru	nan nan
116	Other	<=50K	Guatemala	19.000000 8.000000
63	Black	<=50K	?	18.000000 113.000000

Synthetic				
	race	income	native.country	MIN(age) COUNT(*)
198	White	>50K	Jamaica	55 2
163	White	<=50K	Jamaica	31 42
103	Black	>50K	Peru	41 6
116	Other	<=50K	Guatemala	33 2
63	Black	<=50K	?	18 113

SQL for Real:
SELECT race,income,`native.country`,MIN(age), COUNT(*) FROM C1 WHERE (income = '<=50K') OR (workclass IN ('Self-emp-not-inc', 'Self-emp-inc', 'Federal-gov', '?', 'Private', 'Local-gov') OR `hours.per.week` BETWEEN 40 AND 40) OR (relationship IN ('Husband', 'Unmarried', 'Wife', 'Other-relative', 'Own-child') OR race IN ('White', 'Asian-Pac-Islander', 'Other')) GROUP BY race,income, `native.country`
Resulted in 199 records

SQL for Synthetic:
SELECT race,income,`native.country`,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (income = '<=50K') OR (workclass IN ('Self-emp-not-inc', 'Self-emp-inc', 'Federal-gov', '?', 'Private', 'Local-gov') OR `hours.per.week` BETWEEN 40 AND 40) OR (relationship IN ('Husband', 'Unmarried', 'Wife', 'Other-relative', 'Own-child') OR race IN ('White', 'Asian-Pac-Islander', 'Other')) GROUP BY race,income, `native.country`
Resulted in 211 records

Normalized Euclidean distance for (age): 11.92
Hellinger Distance: 0.416

=====

Real			
	income	SUM(fnlwgt)	COUNT(*)
1	>50K	1968158268	10436
0	<=50K	6383479598	33659

Synthetic			
	income	SUM(fnlwgt)	COUNT(*)
1	>50K	203087643.548069	1154.000000
0	<=50K	3978757941.296477	22069.000000

SQL for Real:
SELECT income,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (education IN ('Preschool', '5th-6th', '12th', '7th-8th', 'Some-college', 'Masters', 'Assoc-acdm', 'Assoc-voc', '1st-4th', 'Prof-school', '10th', 'Doctorate', 'Bachelors', 'HS-grad') AND capital >= 0) OR (race <> 'White' OR workclass IN ('Self-emp-inc', 'Federal-gov', 'Local-gov', 'Self-emp-not-inc', '?', 'State-gov')) AND (sex = 'Female' AND `hours.per.week` > 50) AND (occupation <> 'Other-service' AND `marital.status` IN ('Divorced', 'Married-civ-spouse', 'Never-married', 'Married-spouse-absent', 'Widowed', 'Separated', 'Married-AF-spouse')) GROUP BY income
Resulted in 2 records

SQL for Synthetic:
SELECT income,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (education IN ('Preschool', '5th-6th', '12th', '7th-8th', 'Some-college', 'Masters', 'Assoc-acdm', 'Assoc-voc', '1st-4th', 'Prof-school', '10th', 'Doctorate', 'Bachelors', 'HS-grad') AND capital >= 0) OR (race <> 'White' OR workclass IN ('Self-emp-inc', 'Federal-gov', 'Local-gov', 'Self-emp-not-inc', '?', 'State-gov')) AND (sex = 'Female' AND `hours.per.week` > 50) AND (occupation <> 'Other-service' AND `marital.status` IN ('Divorced', 'Married-civ-spouse', 'Never-married', 'Married-spouse-absent', 'Widowed', 'Separated', 'Married-AF-spouse')) GROUP BY income
Resulted in 2 records

Normalized Euclidean distance for (fnlwgt): 1.41
Hellinger Distance: 0.2

=====

Real								
	education	workclass	sex	occupation	marital.status	relationship	AVG(age)	COUNT(*)
5119	Some-college	Private	Female	Adm-clerical	Married-spouse-absent	Other-relative	18.000000	1.000000
5535	Some-college	Self-emp-inc	Male	Exec-managerial	Married-civ-spouse	Husband	47.375000	104.000000
3902	HS-grad	Self-emp-inc	Female	Sales	Never-married	Own-child	nan	nan
3778	HS-grad	Private	Male	Other-service	Married-spouse-absent	Not-in-family	52.000000	5.000000
1202	7th-8th	Local-gov	Female	Other-service	Divorced	Unmarried	51.000000	1.000000

Synthetic								
	education	workclass	sex	occupation	marital.status	relationship	AVG(age)	COUNT(*)
5119	Some-college	Private	Female	Adm-clerical	Married-spouse-absent	Other-relative	51.000000	1
5535	Some-college	Self-emp-inc	Male	Exec-managerial	Married-civ-spouse	Husband	46.690722	97
3902	HS-grad	Self-emp-inc	Female	Sales	Never-married	Own-child	22.000000	1
3778	HS-grad	Private	Male	Other-service	Married-spouse-absent	Not-in-family	37.428571	7
1202	7th-8th	Local-gov	Female	Other-service	Divorced	Unmarried	47.000000	3

SQL for Real:
SELECT education,workclass,sex,occupation,`marital.status`,relationship,AVG(age), COUNT(*) FROM C1 WHERE (education <> 'Bachelors' OR fnlwgt BETWEEN 180920 AND 401623) GROUP BY education,workclass,sex,occupation,`marital.status`,relationship
Resulted in 5745 records

SQL for Synthetic:
SELECT education,workclass,sex,occupation,`marital.status`,relationship,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (education <> 'Bachelors' OR fnlwgt BETWEEN 180920 AND 401623) GROUP BY education,workclass,sex,occupation,`marital.status`,relationship
Resulted in 5807 records

Normalized Euclidean distance for (age): 59.53
Hellinger Distance: 0.134

=====

Real						
	workclass	education	marital.status	native.country	MIN(age)	COUNT(*)
2098	Private	Some-college	Married-civ-spouse	Thailand	nan	nan
638	Local-gov	12th	Divorced	United-States	56.000000	1.000000
895	Local-gov	Masters	Divorced	Ireland	nan	nan
2404	Self-emp-not-inc	10th	Married-civ-spouse	?	33.000000	3.000000
893	Local-gov	Masters	Divorced	England	nan	nan

Synthetic						
	workclass	education	marital.status	native.country	MIN(age)	COUNT(*)
2098	Private	Some-college	Married-civ-spouse	Thailand	25	6
638	Local-gov	12th	Divorced	United-States	31	3
895	Local-gov	Masters	Divorced	Ireland	30	1
2404	Self-emp-not-inc	10th	Married-civ-spouse	?	48	3
893	Local-gov	Masters	Divorced	England	46	1

SQL for Real:
 (SELECT workclass,education,`marital.status`,`native.country`,MIN(age), COUNT(*) FROM C1 WHERE (workclass = 'Private' AND `native.country` LIKE 'United-States') OR ((capital BETWEEN 0 AND 0) OR age > 28) OR occupation IN ('Tech-support', 'Farming-fishing', 'Adm-clerical') OR `marital.status` IN ('Married-civ-spouse', 'Married-spouse-absent', 'Never-married', 'Married-AF-spouse')) OR (sex <> 'Male' OR `hours.per.week` BETWEEN 40 AND 40) GROUP BY workclass,education,`marital.status`,`native.country`

Resulted in 2389 records

SQL for Synthetic:
 SELECT workclass,education,`marital.status`,`native.country`,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (workclass = 'Private' AND `native.country` LIKE 'United-States') OR ((capital BETWEEN 0 AND 0) OR age > 28) OR occupation IN ('Tech-support', 'Farming-fishing', 'Adm-clerical') OR `marital.status` IN ('Married-civ-spouse', 'Married-spouse-absent', 'Never-married', 'Married-AF-spouse')) OR (sex <> 'Male' OR `hours.per.week` BETWEEN 40 AND 40) GROUP BY workclass,education,`marital.status`,`native.country`

Resulted in 2987 records

Normalized Euclidean distance for (age): 33.53

Hellinger Distance: 0.344

=====

Real			
	income	sex	AVG(`hours.per.week`) COUNT(*)
1	<=50K	Male	39.941628 7812
0	<=50K	Female	39.840917 5846
3	>50K	Male	39.933242 2187
2	>50K	Female	39.846575 730

Synthetic			
	income	sex	AVG(`hours.per.week`) COUNT(*)
1	<=50K	Male	39.821012 8376.000000
0	<=50K	Female	39.766481 8582.000000
3	>50K	Male	39.915380 654.000000
2	>50K	Female	39.928961 358.000000

SQL for Real:
 SELECT income,sex,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (fnlwtg < 176683 OR relationship <> 'Husband') AND (race IN ('Amer-Indian-Eskimo', 'Asian-Pac-Islander', 'Other', 'White') AND `hours.per.week` BETWEEN 37 AND 40) GROUP BY income,sex

Resulted in 4 records

SQL for Synthetic:
 SELECT income,sex,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt < 176683 OR relationship <> 'Husband') AND (race IN ('Amer-Indian-Eskimo', 'Asian-Pac-Islander', 'Other', 'White') AND `hours.per.week` BETWEEN 37 AND 40) GROUP BY income,sex

Resulted in 4 records

Normalized Euclidean distance for (`hours.per.week`): 2.0

Hellinger Distance: 0.148

=====

Real							
	native.country	education	marital.status	occupation	race	MIN(age)	COUNT(*)
5398	United-States	9th	Separated	Other-service	Black	46.000000	5.000000
2240	Japan	HS-grad	Divorced	Adm-clerical	Black	nan	nan
5281	United-States	7th-8th	Married-spouse-absent	Other-service	White	47.000000	2.000000
5547	United-States	Assoc-voc	Married-civ-spouse	Craft-repair	White	21.000000	211.000000
2527	Laos	Bachelors	Never-married	Sales	Asian-Pac-Islander	nan	nan

Synthetic							
	native.country	education	marital.status	occupation	race	MIN(age)	COUNT(*)
5398	United-States	9th	Separated	Other-service	Black	33	3
2240	Japan	HS-grad	Divorced	Adm-clerical	Black	37	2
5281	United-States	7th-8th	Married-spouse-absent	Other-service	White	36	1
5547	United-States	Assoc-voc	Married-civ-spouse	Craft-repair	White	21	118
2527	Laos	Bachelors	Never-married	Sales	Asian-Pac-Islander	27	2

SQL for Real:
 SELECT `native.country`,education,`marital.status`,occupation,race,MIN(age), COUNT(*) FROM C1 WHERE (`marital.status` LIKE 'Never-married' AND age = 19) OR (`native.country` = 'United-States' OR relationship IN ('Wife', 'Husband', 'Not-in-family', 'Unmarried')) GROUP BY `native.country`,education,`marital.status`,occupation,race

Resulted in 4528 records

SQL for Synthetic:
 SELECT `native.country`,education,`marital.status`,occupation,race,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` LIKE 'Never-married' AND age = 19)

OR (`native.country` = 'United-States' OR relationship IN ('Wife', 'Husband', 'Not-in-family', 'Unmarried')) GROUP BY `native.country`,education`,`marital.status`,occupation,race

Resulted in 6546 records

Normalized Euclidean distance for (age): 43.49

Hellinger Distance: 0.254

=====

Real										
	relationship	education	workclass	marital.status	occupation	race	income	native.country	MIN(capital)	COUNT(*)
2439	Husband	Prof-school	Self-emp-not-inc	Married-civ-spouse	Farming-fishing	White	<=50K	United-States	0	2
4594	Not-in-family	HS-grad	Federal-gov	Separated	Exec-managerial	Black	<=50K	United-States	0	1
3968	Not-in-family	Bachelors	Federal-gov	Divorced	Tech-support	Amer-Indian-Eskimo	<=50K	United-States	0	1
6057	Other-relative	10th	?	Separated	?	White	<=50K	Puerto-Rico	0	1
8617	Unmarried	12th	Private	Divorced	Other-service	White	<=50K	United-States	0	2

Synthetic										
	relationship	education	workclass	marital.status	occupation	race	income	native.country	MIN(capital)	COUNT(*)
2439	Husband	Prof-school	Self-emp-not-inc	Married-civ-spouse	Farming-fishing	White	<=50K	United-States	-191.937311	2.000000
4594	Not-in-family	HS-grad	Federal-gov	Separated	Exec-managerial	Black	<=50K	United-States	nan	nan
3968	Not-in-family	Bachelors	Federal-gov	Divorced	Tech-support	Amer-Indian-Eskimo	<=50K	United-States	17.205487	1.000000
6057	Other-relative	10th	?	Separated	?	White	<=50K	Puerto-Rico	nan	nan
8617	Unmarried	12th	Private	Divorced	Other-service	White	<=50K	United-States	nan	nan

SQL for Real:
SELECT relationship,education,workclass`,`marital.status`,`occupation,race,income`,`native.country`,`MIN(capital), COUNT(*) FROM C1 WHERE (education LIKE 'Bachelors') OR ((capital BETWEEN 0 AND 0) OR `hours.per.week` BETWEEN 40 AND 70) GROUP BY relationship,education,workclass`,`marital.status`,`occupation,race,income`,`native.country`

Resulted in 11159 records

SQL for Synthetic:
SELECT relationship,education,workclass`,`marital.status`,`occupation,race,income`,`native.country`,`MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (education LIKE 'Bachelors') OR ((capital BETWEEN 0 AND 0) OR `hours.per.week` BETWEEN 40 AND 70) GROUP BY relationship,education,workclass`,`marital.status`,`occupation,race,income`,`native.country`

Resulted in 10177 records

Normalized Euclidean distance for (capital): 53.77

Hellinger Distance: 0.302

=====

Real					
	marital.status	education	income	MIN(capital)	COUNT(*)
8	Never-married	Assoc-acdm	<=50K	0	1
4	Divorced	Some-college	<=50K	0	2
18	Widowed	Assoc-acdm	<=50K	0	1
19	Widowed	Bachelors	<=50K	0	1
10	Never-married	Bachelors	<=50K	0	11

Synthetic					
	marital.status	education	income	MIN(capital)	COUNT(*)
8	Never-married	Assoc-acdm	<=50K	nan	nan
4	Divorced	Some-college	<=50K	nan	nan
18	Widowed	Assoc-acdm	<=50K	nan	nan
19	Widowed	Bachelors	<=50K	nan	nan
10	Never-married	Bachelors	<=50K	nan	nan

SQL for Real:
SELECT `marital.status`,`education,income,MIN(capital), COUNT(*) FROM C1 WHERE (`hours.per.week` = 50) AND (education LIKE 'Some-college' OR workclass IN ('State-gov', 'Self-emp-not-inc', 'Private')) AND (race = 'White' AND sex = 'Female') AND (`native.country` IN ('China', 'Cambodia', 'Italy', 'Philippines', 'Japan', 'Trinidad&Tobago', 'Jamaica', 'Nicaragua', 'Taiwan', 'Iran', 'El-Salvador', 'Peru', 'Guatemala', 'Scotland', 'Puerto-Rico', 'Honduras', 'Haiti', 'Holand-Netherlands', 'South', 'United-States', 'Hungary', 'Mexico', 'France', 'Canada', 'Germany', 'Columbia', 'Laos', 'Greece', 'Portugal', 'India', 'Hong', 'England', 'Ireland', 'Vietnam') AND capital = 0) AND (occupation = 'Prof-specialty' AND income <> '>50K') GROUP BY `marital.status`,`education,income

Resulted in 20 records

SQL for Synthetic:
SELECT `marital.status`,`education,income,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` = 50) AND (education LIKE 'Some-college' OR workclass IN ('State-gov', 'Self-emp-not-inc', 'Private')) AND (race = 'White' AND sex = 'Female') AND (`native.country` IN ('China', 'Cambodia', 'Italy', 'Philippines', 'Japan', 'Trinidad&Tobago', 'Jamaica', 'Nicaragua', 'Taiwan', 'Iran', 'El-Salvador', 'Peru', 'Guatemala', 'Scotland', 'Puerto-Rico', 'Honduras', 'Haiti', 'Holand-Netherlands', 'South', 'United-States', 'Hungary', 'Mexico', 'France', 'Canada', 'Germany', 'Columbia', 'Laos', 'Greece', 'Portugal', 'India', 'Hong', 'England', 'Ireland', 'Vietnam') AND capital = 0) AND (occupation = 'Prof-specialty' AND income <> '>50K') GROUP BY `marital.status`,`education,income

Resulted in 0 records

Normalized Euclidean distance for (capital): nan

Hellinger Distance: nan

=====

Real		
	sex	SUM(capital) COUNT(*)
0	Female	7806422 16063
1	Male	2887029 22770

Synthetic			
	sex	SUM(capital)	COUNT(*)
0	Female	20424.064231	17063.000000
1	Male	202630.857775	27731.000000

SOL for Real:
SELECT sex,SUM(capital), COUNT(*) FROM C1 WHERE (income = '<=50K') OR ((fnlwgt BETWEEN 308365 AND 221581) OR age < 59) AND (sex <> 'Male' AND race IN ('Other', 'Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'Black', 'White')) OR (` marital.status` <> 'Married-civ-spouse' OR `hours.per.week` BETWEEN 60 AND 40) AND (workclass LIKE 'Private' AND education LIKE 'Prof-school') GROUP BY sex

Resulted in 2 records

SOL for Synthetic:
SELECT sex,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (income = '<=50K') OR ((fnlwgt BETWEEN 308365 AND 221581) OR age < 59) AND (sex <> 'Male' AND race IN ('Other', 'Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'Black', 'White')) OR (` marital.status` <> 'Married-civ-spouse' OR `hours.per.week` BETWEEN 60 AND 40) AND (workclass LIKE 'Private' AND education LIKE 'Prof-school') GROUP BY sex

Resulted in 2 records

Normalized Euclidean distance for (capital): 1.41

Hellinger Distance: 0.024

=====

Real						
	sex	income	relationship	native.country	AVG(age)	COUNT(*)
485	Male	>50K	Unmarried	France	47.000000	1
159	Female	<=50K	Wife	Portugal	32.000000	2
390	Male	<=50K	Unmarried	Mexico	31.275862	29
399	Male	<=50K	Unmarried	United-States	37.848907	503
218	Female	>50K	Wife	Iran	38.000000	2

Synthetic						
	sex	income	relationship	native.country	AVG(age)	COUNT(*)
485	Male	>50K	Unmarried	France	nan	nan
159	Female	<=50K	Wife	Portugal	nan	nan
390	Male	<=50K	Unmarried	Mexico	nan	nan
399	Male	<=50K	Unmarried	United-States	nan	nan
218	Female	>50K	Wife	Iran	30.666667	3.000000

SOL for Real:
SELECT sex,income,relationship,`native.country`,AVG(age), COUNT(*) FROM C1 WHERE (`hours.per.week` = 40) OR (education IN ('11th', 'HS-grad', '7th-8th', 'Bachelors', 'Doctorate', '5th-6th', '12th', 'Assoc-voc', 'Prof-school', '1st-4th', 'Some-college', 'Masters') AND `native.country` IN ('China', 'Yugoslavia', 'Greece', 'Poland', 'Trinidad&Tobago', 'England', 'Canada', 'Ireland', 'Dominican-Republic', 'Haiti', 'El-Salvador', 'Italy', 'Jamaica')) OR (occupation = 'Machine-op-inspct' OR workclass LIKE 'Private') AND (fnlwgt = 50149 AND sex = 'Male') AND (` marital.status` IN ('Widowed', 'Married-civ-spouse', 'Married-AF-spouse', 'Separated', 'Never-married', 'Married-spouse-absent') AND capital >= 0) OR ((age BETWEEN 37 AND 21) OR income LIKE '>50K') GROUP BY sex,income,relationship,`native.country`

Resulted in 493 records

SOL for Synthetic:
SELECT sex,income,relationship,`native.country`,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` = 40) OR (education IN ('11th', 'HS-grad', '7th-8th', 'Bachelors', 'Doctorate', '5th-6th', '12th', 'Assoc-voc', 'Prof-school', '1st-4th', 'Some-college', 'Masters') AND `native.country` IN ('China', 'Yugoslavia', 'Greece', 'Poland', 'Trinidad&Tobago', 'England', 'Canada', 'Ireland', 'Dominican-Republic', 'Haiti', 'El-Salvador', 'Italy', 'Jamaica')) OR (occupation = 'Machine-op-inspct' OR workclass LIKE 'Private') AND (fnlwgt = 50149 AND sex = 'Male') AND (` marital.status` IN ('Widowed', 'Married-civ-spouse', 'Married-AF-spouse', 'Separated', 'Never-married', 'Married-spouse-absent') AND capital >= 0) OR ((age BETWEEN 37 AND 21) OR income LIKE '>50K') GROUP BY sex,income,relationship,`native.country`

Resulted in 205 records

Normalized Euclidean distance for (age): 12.37

Hellinger Distance: 0.404

=====

Real								
	workclass	education	relationship	marital.status	occupation	income	sex	MAX(fnlwgt) COUNT(*)
34	Self-emp-not-inc	Bachelors	Own-child	Never-married	Craft-repair	<=50K	Male	216178 2
18	Private	Bachelors	Own-child	Never-married	Craft-repair	>50K	Male	31493 1
4	Local-gov	Bachelors	Not-in-family	Never-married	Craft-repair	>50K	Male	158291 1
13	Private	Bachelors	Not-in-family	Widowed	Craft-repair	>50K	Male	244605 1
1	Local-gov	Bachelors	Husband	Married-civ-spouse	Craft-repair	<=50K	Male	523910 6

Synthetic								
	workclass	education	relationship	marital.status	occupation	income	sex	MAX(fnlwgt) COUNT(*)
34	Self-emp-not-inc	Bachelors	Own-child	Never-married	Craft-repair	<=50K	Male	nan nan
18	Private	Bachelors	Own-child	Never-married	Craft-repair	>50K	Male	nan nan
4	Local-gov	Bachelors	Not-in-family	Never-married	Craft-repair	>50K	Male	nan nan
13	Private	Bachelors	Not-in-family	Widowed	Craft-repair	>50K	Male	nan nan
1	Local-gov	Bachelors	Husband	Married-civ-spouse	Craft-repair	<=50K	Male	193107.003786 4.000000

SOL for Real:
SELECT workclass,education,relationship,`marital.status`,occupation,income,sex,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (`native.country` LIKE 'United-States' OR relationship IN ('Other-relative', 'Wife')) AND (capital >= 0 AND sex = 'Male') AND (occupation LIKE 'Craft-repair' AND education LIKE 'Bachelors') GROUP BY workclass,education,relationship,`marital.status`,occupation,income,sex

Resulted in 41 records

SOL for Synthetic:
SELECT workclass,education,relationship,`marital.status`,occupation,income,sex,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`native.country` LIKE 'United-States' OR relationship IN ('Other-relative', 'Wife')) AND (capital >= 0 AND sex = 'Male') AND (occupation LIKE 'Craft-repair' AND education LIKE 'Bachelors') GROUP BY workclass,education,relationship,`marital.status`,occupation,income,sex

Resulted in 23 records

Normalized Euclidean distance for (fnlwgt): 4.47

Hellinger Distance: 0.327

=====

Real										
	native.country	race	relationship	sex	income	marital.status	workclass	occupation	education	MIN(`hours.per.week`) COUNT(*)
14825	United-States	White	Other-relative	Female	<=50K	Divorced	Private	Exec-managerial	HS-grad	40.000000 3.000000
1678	Columbia	White	Not-in-family	Male	<=50K	Widowed	Private	Transport-moving	HS-grad	nan nan
11124	United-States	Black	Husband	Male	<=50K	Married-civ-spouse	Private	Tech-support	Some-college	14.000000 4.000000
9302	Puerto-Rico	White	Unmarried	Female	<=50K	Married-spouse-absent	Private	Adm-clerical	10th	nan nan
9971	Scotland	White	Husband	Male	<=50K	Married-civ-spouse	Private	Other-service	HS-grad	nan nan

Synthetic										
	native.country	race	relationship	sex	income	marital.status	workclass	occupation	education	MIN(`hours.per.week`) COUNT(*)
14825	United-States	White	Other-relative	Female	<=50K	Divorced	Private	Exec-managerial	HS-grad	39.991997 1
1678	Columbia	White	Not-in-family	Male	<=50K	Widowed	Private	Transport-moving	HS-grad	39.994779 2
11124	United-States	Black	Husband	Male	<=50K	Married-civ-spouse	Private	Tech-support	Some-college	39.916384 2
9302	Puerto-Rico	White	Unmarried	Female	<=50K	Married-spouse-absent	Private	Adm-clerical	10th	40.032068 1
9971	Scotland	White	Husband	Male	<=50K	Married-civ-spouse	Private	Other-service	HS-grad	40.014116 1

SQL for Real:
SELECT `native.country`,race,relationship,sex,income,`marital.status`,workclass,occupation,education,MIN(` hours.per.week`), COUNT(*) FROM C1 WHERE (` hours.per.week` BETWEEN 45 AND 55) OR (occupation <> 'Priv-house-serv' AND income LIKE '>50K') OR (sex IN ('Female', 'Male') OR workclass = 'Private') OR (race IN ('White', 'Black', 'Amer-Indian-Eskimo', 'Asian-Pac-Islander', 'Other') OR education IN ('Some-college', 'Preschool', 'Bachelors', '1st-4th', '9th', 'Prof-school', 'HS-grad', '11th', '5th-6th', '10th', '12th', 'Masters', 'Assoc-acdm', '7th-8th', 'Doctorate')) OR (`native.country` IN ('Japan', 'China', 'England') OR capital = 0) GROUP BY `native.country`,race,relationship,sex,income,`marital.status`,workclass,occupation,education
Resulted in 12664 records

SQL for Synthetic:
SELECT `native.country`,race,relationship,sex,income,`marital.status`,workclass,occupation,education,MIN(` hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (` hours.per.week` BETWEEN 45 AND 55) OR (occupation <> 'Priv-house-serv' AND income LIKE '>50K') OR (sex IN ('Female', 'Male') OR workclass = 'Private') OR (race IN ('White', 'Black', 'Amer-Indian-Eskimo', 'Asian-Pac-Islander', 'Other') OR education IN ('Some-college', 'Preschool', 'Bachelors', '1st-4th', '9th', 'Prof-school', 'HS-grad', '11th', '5th-6th', '10th', '12th', 'Masters', 'Assoc-acdm', '7th-8th', 'Doctorate')) OR (`native.country` IN ('Japan', 'China', 'England') OR capital = 0) GROUP BY `native.country`,race,relationship,sex,income,`marital.status`,workclass,occupation,education
Resulted in 17127 records

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

Hellinger Distance: 0.252

=====

Real							
	sex	native.country	occupation	race	education	marital.status	AVG(fnlwgt) COUNT(*)
5605	Male	United-States	Prof-specialty	Amer-Indian-Eskimo	HS-grad	Never-married	nan nan
5533	Male	United-States	Other-service	Asian-Pac-Islander	Some-college	Separated	nan nan
4312	Male	Puerto-Rico	Craft-repair	Black	HS-grad	Never-married	nan nan
5599	Male	United-States	Other-service	White	Some-college	Separated	63602.500000 2.000000
5242	Male	United-States	Exec-managerial	Amer-Indian-Eskimo	Some-college	Married-civ-spouse	32587.000000 1.000000

Synthetic							
	sex	native.country	occupation	race	education	marital.status	AVG(fnlwgt) COUNT(*)
5605	Male	United-States	Prof-specialty	Amer-Indian-Eskimo	HS-grad	Never-married	185170.626323 1
5533	Male	United-States	Other-service	Asian-Pac-Islander	Some-college	Separated	173240.210958 1
4312	Male	Puerto-Rico	Craft-repair	Black	HS-grad	Never-married	176524.388731 2
5599	Male	United-States	Other-service	White	Some-college	Separated	175588.422500 2
5242	Male	United-States	Exec-managerial	Amer-Indian-Eskimo	Some-college	Married-civ-spouse	171862.046118 2

SQL for Real:
SELECT sex,`native.country`,occupation,race,education,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (capital < 0) GROUP BY sex,`native.country`,occupation,race,education,`marital.status`
Resulted in 756 records

SQL for Synthetic:
SELECT sex,`native.country`,occupation,race,education,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (capital < 0) GROUP BY sex,`native.country`,occupation,race,education,`marital.status`
Resulted in 5927 records

Normalized Euclidean distance for (fnlwgt): 22.87

Hellinger Distance: 0.268

=====

Real			
	workclass	MAX(` hours.per.week`)	COUNT(*)
3	Private	99	32689
5	Self-emp-not-inc	99	3862
0	?	99	2830
1	Federal-gov	99	1432
6	State-gov	99	1981

Synthetic			
	workclass	MAX(`hours.per.week`)	COUNT(*)
3	Private	40.592070	32654.000000
5	Self-emp-not-inc	40.527407	3818.000000
0	?	40.382698	2844.000000
1	Federal-gov	40.128197	1426.000000
6	State-gov	40.128433	1982.000000

SQL for Real:
SELECT workclass,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (capital > 0) AND (relationship = 'Husband' OR income LIKE '<=50K') OR (`native.country` IN ('Jamaica', 'Thailand', 'England', 'Nicaragua', 'Puerto-Rico', 'Japan', 'United-States', 'Laos') OR occupation LIKE '?') OR (age < 39 OR workclass <> 'Private') GROUP BY workclass

Resulted in 7 records

SQL for Synthetic:
SELECT workclass,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (capital > 0) AND (relationship = 'Husband' OR income LIKE '<=50K') OR (`native.country` IN ('Jamaica', 'Thailand', 'England', 'Nicaragua', 'Puerto-Rico', 'Japan', 'United-States', 'Laos') OR occupation LIKE '?') OR (age < 39 OR workclass <> 'Private') GROUP BY workclass

Resulted in 7 records

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

Hellinger Distance: 0.002

=====

Real				
	marital.status	race	MIN(age)	COUNT(*)
5	Married-civ-spouse	Asian-Pac-Islander	30	3
0	Divorced	Amer-Indian-Eskimo	30	1
9	Married-spouse-absent	White	30	1
3	Divorced	White	30	13
6	Married-civ-spouse	Black	31	1

Synthetic				
	marital.status	race	MIN(age)	COUNT(*)
5	Married-civ-spouse	Asian-Pac-Islander	nan	nan
0	Divorced	Amer-Indian-Eskimo	nan	nan
9	Married-spouse-absent	White	nan	nan
3	Divorced	White	nan	nan
6	Married-civ-spouse	Black	nan	nan

SQL for Real:
SELECT `marital.status`,race,MIN(age), COUNT(*) FROM C1 WHERE (sex <> 'Male' OR `marital.status` IN ('Married-AF-spouse', 'Married-spouse-absent')) AND (capital = 0 AND age BETWEEN 30 AND 33) AND (income = '<=50K' AND occupation LIKE 'Sales') AND (workclass = 'Private' AND relationship <> 'Not-in-family') GROUP BY `marital.status`,race

Resulted in 16 records

SQL for Synthetic:
SELECT `marital.status`,race,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (sex <> 'Male' OR `marital.status` IN ('Married-AF-spouse', 'Married-spouse-absent')) AND (capital = 0 AND age BETWEEN 30 AND 33) AND (income = '<=50K' AND occupation LIKE 'Sales') AND (workclass = 'Private' AND relationship <> 'Not-in-family') GROUP BY `marital.status`,race

Resulted in 0 records

Normalized Euclidean distance for (age): nan

Hellinger Distance: nan

=====

Real			
	income	MAX(age)	COUNT(*)
0	<=50K	90	35461
1	>50K	90	11557

Synthetic			
	income	MAX(age)	COUNT(*)
0	<=50K	90.000000	42168.000000
1	>50K	90.000000	4707.000000

SQL for Real:
SELECT income,MAX(age), COUNT(*) FROM C1 WHERE (occupation <> 'Handlers-cleaners') OR (sex <> 'Male' AND capital >= 0) GROUP BY income

Resulted in 2 records

SQL for Synthetic:
SELECT income,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (occupation <> 'Handlers-cleaners') OR (sex <> 'Male' AND capital >= 0) GROUP BY income

Resulted in 2 records

Normalized Euclidean distance for (age): 1.41

Hellinger Distance: 0.139

=====

Real								
	race	education	sex	workclass	marital.status	occupation	MIN(capital)	COUNT(*)
246	Black	Some-college	Female	Private	Never-married	Protective-serv	99999	1
115	Asian-Pac-Islander	Some-college	Male	Federal-gov	Married-civ-spouse	Exec-managerial	0	1
180	Black	HS-grad	Male	Local-gov	Married-civ-spouse	Protective-serv	0	4
191	Black	HS-grad	Male	Private	Married-civ-spouse	Transport-moving	-1887	11
228	Black	Masters	Male	Private	Never-married	Exec-managerial	-1564	3
Synthetic								
	race	education	sex	workclass	marital.status	occupation	MIN(capital)	COUNT(*)
246	Black	Some-college	Female	Private	Never-married	Protective-serv	nan	nan
115	Asian-Pac-Islander	Some-college	Male	Federal-gov	Married-civ-spouse	Exec-managerial	nan	nan
180	Black	HS-grad	Male	Local-gov	Married-civ-spouse	Protective-serv	-243.928161	2.000000
191	Black	HS-grad	Male	Private	Married-civ-spouse	Transport-moving	-270.066186	5.000000
228	Black	Masters	Male	Private	Never-married	Exec-managerial	nan	nan

SOL for Real:
SELECT race,education,sex,workclass,`marital.status`,occupation,MIN(capital), COUNT(*) FROM C1 WHERE (occupation <> 'Craft-repair' AND income LIKE '>50K') AND (relationship IN ('Own-child', 'Other-relative', 'Not-in-family', 'Wife', 'Unmarried', 'Husband') AND race <> 'White') AND (age <> 25 AND education IN ('Doctorate', '1st-4th', '9th', 'Some-college', '12th', '10th', 'Assoc-voc', 'Preschool', 'HS-grad', '5th-6th', 'Masters')) GROUP BY race,education,sex,workclass,`marital.status`,occupation

Resulted in 300 records

SOL for Synthetic:
SELECT race,education,sex,workclass,`marital.status`,occupation,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (occupation <> 'Craft-repair' AND income LIKE '>50K') AND (relationship IN ('Own-child', 'Other-relative', 'Not-in-family', 'Wife', 'Unmarried', 'Husband') AND race <> 'White') AND (age <> 25 AND education IN ('Doctorate', '1st-4th', '9th', 'Some-college', '12th', '10th', 'Assoc-voc', 'Preschool', 'HS-grad', '5th-6th', 'Masters')) GROUP BY race,education,sex,workclass,`marital.status`,occupation

Resulted in 125 records

Normalized Euclidean distance for (capital): 7.81

Hellinger Distance: 0.216

=====

Real											
	native.country	occupation	race	sex	marital.status	education	relationship	income	workclass	MAX(age)	COUNT(*)
7688	United-States	Machine-op-inspct	Asian-Pac-Islander	Male	Never-married	HS-grad	Own-child	<=50K	Private	nan	nan
4140	Puerto-Rico	Exec-managerial	White	Male	Married-civ-spouse	Assoc-acdm	Husband	>50K	Local-gov	nan	nan
208	?	Exec-managerial	White	Male	Married-civ-spouse	Some-college	Husband	>50K	Private	52.000000	1.000000
7194	United-States	Exec-managerial	White	Male	Widowed	Bachelors	Not-in-family	<=50K	State-gov	nan	nan
1297	Germany	Machine-op-inspct	White	Male	Married-civ-spouse	7th-8th	Husband	<=50K	Private	nan	nan
Synthetic											
	native.country	occupation	race	sex	marital.status	education	relationship	income	workclass	MAX(age)	COUNT(*)
7688	United-States	Machine-op-inspct	Asian-Pac-Islander	Male	Never-married	HS-grad	Own-child	<=50K	Private	23	2
4140	Puerto-Rico	Exec-managerial	White	Male	Married-civ-spouse	Assoc-acdm	Husband	>50K	Local-gov	60	1
208	?	Exec-managerial	White	Male	Married-civ-spouse	Some-college	Husband	>50K	Private	62	6
7194	United-States	Exec-managerial	White	Male	Widowed	Bachelors	Not-in-family	<=50K	State-gov	58	2
1297	Germany	Machine-op-inspct	White	Male	Married-civ-spouse	7th-8th	Husband	<=50K	Private	33	2

SOL for Real:
SELECT `native.country`,occupation,race,sex,`marital.status`,education,relationship,income,workclass,MAX(age), COUNT(*) FROM C1 WHERE (sex = 'Male') OR (fnlwgt <= 181130 AND income LIKE '>50K') AND ((age BETWEEN 43 AND 28) AND occupation = 'Sales') GROUP BY `native.country`,occupation,race,sex,`marital.status`,education,relationship,income,workclass

Resulted in 7386 records

SOL for Synthetic:
SELECT `native.country`,occupation,race,sex,`marital.status`,education,relationship,income,workclass,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (sex = 'Male') OR (fnlwgt <= 181130 AND income LIKE '>50K') AND ((age BETWEEN 43 AND 28) AND occupation = 'Sales') GROUP BY `native.country`,occupation,race,sex,`marital.status`,education,relationship,income,workclass

Resulted in 9548 records

Normalized Euclidean distance for (age): 49.57

Hellinger Distance: 0.274

=====

Real			
	relationship	native.country	SUM(`hours.per.week`) COUNT(*)
0	Unmarried	United-States	466 12
Synthetic			
	relationship	native.country	SUM(`hours.per.week`) COUNT(*)
0	Unmarried	United-States	239.608729 6.000000

SOL for Real:
SELECT relationship,`native.country`,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (sex LIKE 'Male' AND education IN ('Doctorate', 'Prof-school', 'Some-college', '5th-6th')) AND (occupation <> 'Adm-clerical' AND workclass IN ('?', 'Private', 'Local-gov')) AND (relationship LIKE 'Unmarried' AND capital <= 0) AND ((age BETWEEN 21 AND 48) AND `native.country` LIKE 'United-States') AND (race LIKE 'Black' AND income IN ('>50K', '<=50K')) GROUP BY relationship,`native.country`

Resulted in 1 records

SOL for Synthetic:
SELECT relationship,`native.country`,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (sex LIKE 'Male' AND education IN ('Doctorate', 'Prof-school', 'Some-college', '5th-6th')) AND (occupation <> 'Adm-clerical' AND workclass IN ('?', 'Private', 'Local-gov')) AND (relationship LIKE 'Unmarried' AND capital <= 0) AND ((age BETWEEN 21 AND 48) AND `native.country` LIKE 'United-States') AND (race LIKE 'Black' AND income IN ('>50K', '<=50K')) GROUP BY relationship,`native.country`

Resulted in 1 records

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

Hellinger Distance: 0.0

=====

Real							
income	native.country	race	relationship	sex	marital.status	MAX(capital)	COUNT(*)
909 <=50K	United-States	Asian-Pac-Islander	Wife	Female	Married-civ-spouse	3464.000000	3.000000
358 <=50K	Hungary	White	Own-child	Female	Never-married	nan	nan
731 <=50K	Puerto-Rico	Other	Wife	Female	Married-civ-spouse	nan	nan
200 <=50K	Ecuador	White	Own-child	Female	Never-married	0.000000	2.000000
603 <=50K	Outlying-US(Guam-USVI-etc)	White	Other-relative	Male	Divorced	nan	nan

Synthetic							
income	native.country	race	relationship	sex	marital.status	MAX(capital)	COUNT(*)
909 <=50K	United-States	Asian-Pac-Islander	Wife	Female	Married-civ-spouse	120.832636	3
358 <=50K	Hungary	White	Own-child	Female	Never-married	156.738520	18
731 <=50K	Puerto-Rico	Other	Wife	Female	Married-civ-spouse	-244.557949	1
200 <=50K	Ecuador	White	Own-child	Female	Never-married	278.022809	8
603 <=50K	Outlying-US(Guam-USVI-etc)	White	Other-relative	Male	Divorced	213.469716	1

SOL for Real:
SELECT income, `native.country`, race, relationship, sex, `marital.status`, MAX(capital), COUNT(*) FROM C1 WHERE (race = 'White') OR (education = 'HS-grad' AND `hours.per.week` < 40) GROUP BY income, `native.country`, race, relationship, sex, `marital.status`

Resulted in 886 records

SOL for Synthetic:
SELECT income, `native.country`, race, relationship, sex, `marital.status`, MAX(capital), COUNT(*) FROM C1_syn_06 WHERE (race = 'White') OR (education = 'HS-grad' AND `hours.per.week` < 40) GROUP BY income, `native.country`, race, relationship, sex, `marital.status`

Resulted in 1238 records

Normalized Euclidean distance for (capital): 21.79

Hellinger Distance: 0.408

=====

Real								
	occupation	sex	income	race	education	marital.status	native.country	SUM(fnlwgt) COUNT(*)
383	Transport-moving	Male	<=50K	White	10th	Separated	United-States	182237 1
372	Tech-support	Male	>50K	White	Some-college	Married-civ-spouse	Mexico	441949 1
56	Adm-clerical	Male	>50K	White	Assoc-acdm	Married-civ-spouse	United-States	114937 1
99	Craft-repair	Male	>50K	White	Bachelors	Married-civ-spouse	United-States	408700 2
337	Sales	Male	<=50K	White	HS-grad	Divorced	United-States	19491 1

Synthetic								
	occupation	sex	income	race	education	marital.status	native.country	SUM(fnlwgt) COUNT(*)
383	Transport-moving	Male	<=50K	White	10th	Separated	United-States	nan nan
372	Tech-support	Male	>50K	White	Some-college	Married-civ-spouse	Mexico	nan nan
56	Adm-clerical	Male	>50K	White	Assoc-acdm	Married-civ-spouse	United-States	nan nan
99	Craft-repair	Male	>50K	White	Bachelors	Married-civ-spouse	United-States	nan nan
337	Sales	Male	<=50K	White	HS-grad	Divorced	United-States	nan nan

SOL for Real:
SELECT occupation, sex, income, race, education, `marital.status`, `native.country`, SUM(fnlwgt), COUNT(*) FROM C1 WHERE (`marital.status` <> 'Never-married' OR education <> 'Bachelors') AND (age = 31 AND `hours.per.week` = 40) GROUP BY occupation, sex, income, race, education, `marital.status`, `native.country`

Resulted in 397 records

SOL for Synthetic:
SELECT occupation, sex, income, race, education, `marital.status`, `native.country`, SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` <> 'Never-married' OR education <> 'Bachelors') AND (age = 31 AND `hours.per.week` = 40) GROUP BY occupation, sex, income, race, education, `marital.status`, `native.country`

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

=====

Real				
	native.country	relationship	SUM(fnlwgt)	COUNT(*)
104	Iran	Not-in-family	2149124	12
158	Peru	Wife	727522	3
31	Cuba	Other-relative	1156985	5
48	El-Salvador	Other-relative	5847565	27
34	Cuba	Wife	1295401	6

Synthetic				
	native.country	relationship	SUM(fnlwgt)	COUNT(*)
104	Iran	Not-in-family	10324335.539470	55.000000
158	Peru	Wife	7404141.130616	41.000000
31	Cuba	Other-relative	nan	nan
48	El-Salvador	Other-relative	550881.436875	3.000000
34	Cuba	Wife	nan	nan

SOL for Real:
SELECT `native.country`, relationship, SUM(fnlwgt), COUNT(*) FROM C1 WHERE (race LIKE 'White') AND (sex <> 'Male' AND workclass IN ('Private', 'Federal-gov', 'Self-emp-not-inc', 'State-gov')) OR (`marital.status` <> 'Married-civ-spouse' OR occupation = 'Prof-specialty') OR (income = '<=50K' OR education <> 'HS-grad') AND (`native.country` LIKE 'United-States' AND capital = 0) GROUP BY `native.country`, relationship

Resulted in 225 records

```
SQL for Synthetic:
SELECT `native.country`,relationship,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (race LIKE 'White') AND (sex <> 'Male' AND workclass IN ('Private', 'Federal-gov', 'Self-emp-not-inc', 'State-gov')) OR (`marital.status` <> 'Married-civ-spouse' OR occupation = 'Prof-specialty') OR (income = '<=50K' OR education <> 'HS-grad') AND (`native.country` LIKE 'United-States' AND capital = 0) GROUP BY `native.country`,relationship
```

Resulted in 202 records

Normalized Euclidean distance for (fnlwgt): 13.6

Hellinger Distance: 0.436

=====

Real								
	sex	relationship	income	race	education	native.country	SUM(age)	COUNT(*)
1242	Female	Unmarried	<=50K	White	Some-college	?	190.000000	4.000000
2612	Male	Not-in-family	<=50K	White	Assoc-voc	Peru	nan	nan
2496	Male	Not-in-family	<=50K	Other	9th	United-States	31.000000	1.000000
498	Female	Other-relative	<=50K	Black	Bachelors	Hungary	nan	nan
2568	Male	Not-in-family	<=50K	White	7th-8th	Japan	nan	nan

Synthetic								
	sex	relationship	income	race	education	native.country	SUM(age)	COUNT(*)
1242	Female	Unmarried	<=50K	White	Some-college	?	1149	25
2612	Male	Not-in-family	<=50K	White	Assoc-voc	Peru	138	3
2496	Male	Not-in-family	<=50K	Other	9th	United-States	53	2
498	Female	Other-relative	<=50K	Black	Bachelors	Hungary	43	1
2568	Male	Not-in-family	<=50K	White	7th-8th	Japan	101	2

```
SQL for Real:
SELECT sex,relationship,income,race,education,`native.country`,SUM(age), COUNT(*) FROM C1 WHERE (`hours.per.week` > 40) OR (relationship <> 'Unmarried' AND sex <> 'Male') AND ((capital BETWEEN 0 AND 0) AND age = 38) OR (workclass LIKE 'Private' OR occupation IN ('Prof-specialty', 'Armed-Forces', 'Tech-support', 'Handlers-cleaners', 'Priv-house-serv', 'Craft-repair', 'Exec-managerial')) GROUP BY sex,relationship,income,race,education,`native.country`
```

Resulted in 2739 records

```
SQL for Synthetic:
SELECT sex,relationship,income,race,education,`native.country`,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` > 40) OR (relationship <> 'Unmarried' AND sex <> 'Male') AND ((capital BETWEEN 0 AND 0) AND age = 38) OR (workclass LIKE 'Private' OR occupation IN ('Prof-specialty', 'Armed-Forces', 'Tech-support', 'Handlers-cleaners', 'Priv-house-serv', 'Craft-repair', 'Exec-managerial')) GROUP BY sex,relationship,income,race,education,`native.country`
```

Resulted in 3399 records

Normalized Euclidean distance for (age): 33.9

Hellinger Distance: 0.357

=====

Real						
	relationship	occupation	sex	workclass	SUM(age)	COUNT(*)
237	Own-child	Farming-fishing	Female	Private	173	8
323	Own-child	Transport-moving	Male	Self-emp-inc	69	3
91	Not-in-family	Protective-serv	Male	Federal-gov	130	4
105	Not-in-family	Sales	Male	State-gov	38	1
349	Unmarried	Exec-managerial	Male	Private	432	13

Synthetic						
	relationship	occupation	sex	workclass	SUM(age)	COUNT(*)
237	Own-child	Farming-fishing	Female	Private	95.000000	4.000000
323	Own-child	Transport-moving	Male	Self-emp-inc	nan	nan
91	Not-in-family	Protective-serv	Male	Federal-gov	92.000000	3.000000
105	Not-in-family	Sales	Male	State-gov	154.000000	3.000000
349	Unmarried	Exec-managerial	Male	Private	356.000000	11.000000

```
SQL for Real:
SELECT relationship,occupation,sex,workclass,SUM(age), COUNT(*) FROM C1 WHERE (`marital.status` = 'Never-married' AND `native.country` = 'United-States') OR ((age BETWEEN 25 AND 26) OR occupation LIKE 'Sales') AND (fnlwgt = 204590 AND workclass IN ('Private', 'Local-gov', 'Self-emp-not-inc', '?', 'Self-emp-inc', 'State-gov', 'Federal-gov')) GROUP BY relationship,occupation,sex,workclass
```

Resulted in 410 records

```
SQL for Synthetic:
SELECT relationship,occupation,sex,workclass,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` = 'Never-married' AND `native.country` = 'United-States') OR ((age BETWEEN 25 AND 26) OR occupation LIKE 'Sales') AND (fnlwgt = 204590 AND workclass IN ('Private', 'Local-gov', 'Self-emp-not-inc', '?', 'Self-emp-inc', 'State-gov', 'Federal-gov')) GROUP BY relationship,occupation,sex,workclass
```

Resulted in 402 records

Normalized Euclidean distance for (age): 18.11

Hellinger Distance: 0.089

=====

Real					
	relationship	education	marital.status	MIN(age)	COUNT(*)
45	Not-in-family	5th-6th	Never-married	17	61
286	Unmarried	5th-6th	Married-spouse-absent	33	8
85	Not-in-family	HS-grad	Separated	21	225
210	Own-child	1st-4th	Never-married	24	8
205	Own-child	12th	Married-spouse-absent	19	1

	Synthetic				
	relationship	education	marital.status	MIN(age)	COUNT(*)
45	Not-in-family	5th-6th	Never-married	18.000000	41.000000
286	Unmarried	5th-6th	Married-spouse-absent	31.000000	3.000000
85	Not-in-family	HS-grad	Separated	22.000000	136.000000
210	Own-child	1st-4th	Never-married	20.000000	6.000000
205	Own-child	12th	Married-spouse-absent	nan	nan

SQL for Real:
SELECT relationship,education,`marital.status`,MIN(age), COUNT(*) FROM C1 WHERE (capital >= 0 OR age = 31) OR (`marital.status` = 'Never-married' OR income <> '<=50K') GROUP BY relationship,education,`marital.status`
Resulted in 369 records

SQL for Synthetic:
SELECT relationship,education,`marital.status`,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (capital >= 0 OR age = 31) OR (`marital.status` = 'Never-married' OR income <> '<=50K') GROUP BY relationship,education,`marital.status`
Resulted in 357 records

Normalized Euclidean distance for (age): 17.52

Hellinger Distance: 0.102

=====

	Real				
	education	income	relationship	SUM(fnlwgt)	COUNT(*)
54	7th-8th	<=50K	Own-child	12597723	63
35	1st-4th	<=50K	Husband	24204163	103
101	Bachelors	>50K	Not-in-family	81195440	410
24	12th	<=50K	Not-in-family	22067415	119
19	11th	>50K	Other-relative	121441	1

	Synthetic				
	education	income	relationship	SUM(fnlwgt)	COUNT(*)
54	7th-8th	<=50K	Own-child	11696985.026486	64.000000
35	1st-4th	<=50K	Husband	10290680.245706	60.000000
101	Bachelors	>50K	Not-in-family	14217814.913736	81.000000
24	12th	<=50K	Not-in-family	23253258.761162	115.000000
19	11th	>50K	Other-relative	nan	nan

SQL for Real:
SELECT education,income,relationship,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (fnlwgt = 184682 OR education LIKE 'Bachelors') OR (`native.country` = 'United-States' OR sex = 'Male') OR (income = '<=50K' OR workclass <> 'Private') AND (race = 'White' AND `hours.per.week` BETWEEN 60 AND 40) OR (age >= 37 OR `marital.status` = 'Widowed') GROUP BY education,income,relationship
Resulted in 172 records

SQL for Synthetic:
SELECT education,income,relationship,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt = 184682 OR education LIKE 'Bachelors') OR (`native.country` = 'United-States' OR sex = 'Male') OR (income = '<=50K' OR workclass <> 'Private') AND (race = 'White' AND `hours.per.week` BETWEEN 60 AND 40) OR (age >= 37 OR `marital.status` = 'Widowed') GROUP BY education,income,relationship
Resulted in 164 records

Normalized Euclidean distance for (fnlwgt): 12.61

Hellinger Distance: 0.165

=====

	Real									
	native.country	education	relationship	marital.status	income	race	occupation	workclass	AVG(fnlwgt)	COUNT(*)
8895	United-States	Masters	Not-in-family	Divorced	<=50K	White	Prof-specialty	Private	147786.000000	24
4879	United-States	7th-8th	Husband	Married-civ-spouse	<=50K	White	Farming-fishing	Local-gov	125927.000000	1
10286	United-States	Some-college	Unmarried	Divorced	<=50K	White	Craft-repair	State-gov	106466.000000	1
10112	United-States	Some-college	Own-child	Never-married	<=50K	Asian-Pac-Islander	Adm-clerical	State-gov	96483.000000	1
9561	United-States	Some-college	Husband	Married-civ-spouse	>50K	White	Tech-support	Self-emp-inc	106103.000000	1

	Synthetic									
	native.country	education	relationship	marital.status	income	race	occupation	workclass	AVG(fnlwgt)	COUNT(*)
8895	United-States	Masters	Not-in-family	Divorced	<=50K	White	Prof-specialty	Private	178838.234166	13.000000
4879	United-States	7th-8th	Husband	Married-civ-spouse	<=50K	White	Farming-fishing	Local-gov	nan	nan
10286	United-States	Some-college	Unmarried	Divorced	<=50K	White	Craft-repair	State-gov	171851.020297	1.000000
10112	United-States	Some-college	Own-child	Never-married	<=50K	Asian-Pac-Islander	Adm-clerical	State-gov	nan	nan
9561	United-States	Some-college	Husband	Married-civ-spouse	>50K	White	Tech-support	Self-emp-inc	149647.783640	1.000000

SQL for Real:
SELECT `native.country`,education,relationship,`marital.status`,income,race,occupation,workclass,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (capital <= 0) GROUP BY `native.country`,education,relationship,`marital.status`,income,race,occupation,workclass
Resulted in 10686 records

SQL for Synthetic:
SELECT `native.country`,education,relationship,`marital.status`,income,race,occupation,workclass,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (capital <= 0) GROUP BY `native.country`,education,relationship,`marital.status`,income,race,occupation,workclass
Resulted in 9656 records

Normalized Euclidean distance for (fnlwgt): 53.36

Hellinger Distance: 0.233

=====

Real							
	sex	race	income	occupation	native.country	education	AVG(age) COUNT(*)
4781	Male	White	<=50K	Tech-support	United-States	Some-college	34.302326 129.000000
3228	Male	Other	<=50K	Prof-specialty	Ecuador	Bachelors	32.000000 1.000000
1434	Female	White	<=50K	Handlers-cleaners	Puerto-Rico	11th	nan nan
531	Female	Black	<=50K	Other-service	England	HS-grad	nan nan
3196	Male	Other	<=50K	Handlers-cleaners	United-States	10th	23.000000 1.000000
Synthetic							
	sex	race	income	occupation	native.country	education	AVG(age) COUNT(*)
4781	Male	White	<=50K	Tech-support	United-States	Some-college	32.173469 98
3228	Male	Other	<=50K	Prof-specialty	Ecuador	Bachelors	29.000000 1
1434	Female	White	<=50K	Handlers-cleaners	Puerto-Rico	11th	62.000000 1
531	Female	Black	<=50K	Other-service	England	HS-grad	40.000000 1
3196	Male	Other	<=50K	Handlers-cleaners	United-States	10th	23.000000 1

SQL for Real:
SELECT sex,race,income,occupation,`native.country`,education,AVG(age), COUNT(*) FROM C1 WHERE (education LIKE 'Bachelors' AND `marital.status` = 'Married-civ-spouse') OR (sex IN ('Female', 'Male') OR `hours.per.week` BETWEEN 48 AND 40) AND ((capital BETWEEN 0 AND 7688) AND workclass IN ('Self-emp-inc', 'Federal-gov', '?', 'State-gov')) AND (race LIKE 'White' AND occupation IN ('Prof-specialty', 'Handlers-cleaners', '?', 'Machine-op-inspct', 'Tech-support', 'Adm-clerical', 'Sales', 'Exec-managerial', 'Protective-serv', 'Other-service', 'Craft-repair', 'Priv-house-serv', 'Transport-moving', 'Farming-fishing')) OR (income = '<=50K' OR fnlwtg <= 109532) GROUP BY sex,race,income,occupation,`native.country`,education
Resulted in 3805 records

SQL for Synthetic:
SELECT sex,race,income,occupation,`native.country`,education,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (education LIKE 'Bachelors' AND `marital.status` = 'Married-civ-spouse') OR (sex IN ('Female', 'Male') OR `hours.per.week` BETWEEN 48 AND 40) AND ((capital BETWEEN 0 AND 7688) AND workclass IN ('Self-emp-inc', 'Federal-gov', '?', 'State-gov')) AND (race LIKE 'White' AND occupation IN ('Prof-specialty', 'Handlers-cleaners', '?', 'Machine-op-inspct', 'Tech-support', 'Adm-clerical', 'Sales', 'Exec-managerial', 'Protective-serv', 'Other-service', 'Craft-repair', 'Priv-house-serv', 'Transport-moving', 'Farming-fishing')) OR (income = '<=50K' OR fnlwtg <= 109532) GROUP BY sex,race,income,occupation,`native.country`,education
Resulted in 5060 records

Normalized Euclidean distance for (age): 38.83
Hellinger Distance: 0.328

=====

Real					
	occupation	native.country	sex	income	SUM(capital) COUNT(*)
262	Craft-repair	Thailand	Male	<=50K	0 2
3	?	?	Male	>50K	0 3
647	Other-service	Philippines	Male	<=50K	10137 26
775	Prof-specialty	Jamaica	Female	<=50K	4101 6
45	?	Mexico	Female	<=50K	0 27
Synthetic					
	occupation	native.country	sex	income	SUM(capital) COUNT(*)
262	Craft-repair	Thailand	Male	<=50K	26.486360 14.000000
3	?	?	Male	>50K	-90.980955 2.000000
647	Other-service	Philippines	Male	<=50K	-81.814196 1.000000
775	Prof-specialty	Jamaica	Female	<=50K	364.986734 2.000000
45	?	Mexico	Female	<=50K	nan nan

SQL for Real:
SELECT occupation,`native.country`,sex,income,SUM(capital), COUNT(*) FROM C1 WHERE (fnlwtg <> 126614) OR (relationship = 'Not-in-family' AND age BETWEEN 40 AND 22) OR (`native.country` <> 'United-States' OR `marital.status` = 'Never-married') GROUP BY occupation,`native.country`,sex,income
Resulted in 1074 records

SQL for Synthetic:
SELECT occupation,`native.country`,sex,income,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (fnlwtg <> 126614) OR (relationship = 'Not-in-family' AND age BETWEEN 40 AND 22) OR (`native.country` <> 'United-States' OR `marital.status` = 'Never-married') GROUP BY occupation,`native.country`,sex,income
Resulted in 964 records

Normalized Euclidean distance for (capital): 25.0
Hellinger Distance: 0.398

=====

Real										
	education	sex	marital.status	race	relationship	income	workclass	native.country	occupation	SUM(capital) COUNT(*)
0	Assoc-voc	Female	Married-civ-spouse	White	Wife	>50K	Private	Japan	Tech-support	nan nan
2	HS-grad	Female	Married-civ-spouse	White	Wife	>50K	Local-gov	Puerto-Rico	Adm-clerical	nan nan
3	HS-grad	Female	Married-civ-spouse	White	Wife	>50K	Private	Puerto-Rico	Adm-clerical	nan nan
1	Bachelors	Female	Married-civ-spouse	White	Wife	>50K	Private	Taiwan	Prof-specialty	nan nan
5	Some-college	Female	Married-civ-spouse	White	Wife	>50K	Self-emp-not-inc	?	Prof-specialty	nan nan
Synthetic										
	education	sex	marital.status	race	relationship	income	workclass	native.country	occupation	SUM(capital) COUNT(*)
0	Assoc-voc	Female	Married-civ-spouse	White	Wife	>50K	Private	Japan	Tech-support	20.155352 1
2	HS-grad	Female	Married-civ-spouse	White	Wife	>50K	Local-gov	Puerto-Rico	Adm-clerical	-99.613852 1
3	HS-grad	Female	Married-civ-spouse	White	Wife	>50K	Private	Puerto-Rico	Adm-clerical	-173.661439 1
1	Bachelors	Female	Married-civ-spouse	White	Wife	>50K	Private	Taiwan	Prof-specialty	-14.039921 1
5	Some-college	Female	Married-civ-spouse	White	Wife	>50K	Self-emp-not-inc	?	Prof-specialty	-88.470403 1

SQL for Real:
SELECT education,sex,`marital.status`,race,relationship,income,workclass,`native.country`,occupation,SUM(capital), COUNT(*) FROM C1 WHERE (race LIKE 'White') AND ((fnlwgt BETWEEN 338320 AND 106552) OR workclass <> 'State-gov') AND (`marital.status` LIKE 'Married-civ-spouse' AND income <> '<=50K') AND (age = 41 AND relationship = 'Wife') AND (sex IN ('Female', 'Male') AND `native.country` <> 'United-States') GROUP BY education,sex,`marital.status`,race,relationship,income,workclass,`native.country`,occupation
Resulted in 3 records

SQL for Synthetic:
SELECT education,sex,`marital.status`,race,relationship,income,workclass,`native.country`,occupation,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (race LIKE 'White') AND ((fnlwgt BETWEEN 338320 AND 106552) OR workclass <> 'State-gov') AND (`marital.status` LIKE 'Married-civ-spouse' AND income <> '<=50K') AND (age = 41 AND relationship = 'Wife') AND (sex IN ('Female', 'Male') AND `native.country` <> 'United-States') GROUP BY education,sex,`marital.status`,race,relationship,income,workclass,`native.country`,occupation
Resulted in 6 records

Normalized Euclidean distance for (capital): nan
Hellinger Distance: nan

=====

Real								
	workclass	native.country	income	sex	occupation	relationship	MAX(`hours.per.week`)	COUNT(*)
1801	Private	Mexico	<=50K	Male	Transport-moving	Husband	55.000000	14.000000
1747	Private	Laos	<=50K	Male	Other-service	Own-child	nan	nan
2891	Self-emp-not-inc	Puerto-Rico	<=50K	Female	Sales	Other-relative	nan	nan
1470	Private	India	<=50K	Female	Sales	Unmarried	nan	nan
502	Local-gov	Holand-Netherlands	<=50K	Male	Handlers-cleaners	Own-child	nan	nan

Synthetic								
	workclass	native.country	income	sex	occupation	relationship	MAX(`hours.per.week`)	COUNT(*)
1801	Private	Mexico	<=50K	Male	Transport-moving	Husband	40.261340	3
1747	Private	Laos	<=50K	Male	Other-service	Own-child	40.105219	9
2891	Self-emp-not-inc	Puerto-Rico	<=50K	Female	Sales	Other-relative	39.010632	1
1470	Private	India	<=50K	Female	Sales	Unmarried	40.088326	1
502	Local-gov	Holand-Netherlands	<=50K	Male	Handlers-cleaners	Own-child	40.001186	1

SQL for Real:
SELECT workclass,`native.country`,income,sex,occupation,relationship,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (relationship <> 'Not-in-family') GROUP BY workclass,`native.country`,income,sex,occupation,relationship
Resulted in 2571 records

SQL for Synthetic:
SELECT workclass,`native.country`,income,sex,occupation,relationship,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (relationship <> 'Not-in-family') GROUP BY workclass,`native.country`,income,sex,occupation,relationship
Resulted in 3331 records

Normalized Euclidean distance for (`hours.per.week`): 34.04
Hellinger Distance: 0.343

=====

Real								
	native.country	occupation	sex	race	education	income	AVG(fnlwgt)	COUNT(*)
4561	United-States	Farming-fishing	Male	White	Assoc-acdm	<=50K	208004.230769	13.000000
4082	Thailand	Exec-managerial	Female	White	Some-college	<=50K	nan	nan
3513	Puerto-Rico	Prof-specialty	Female	Black	Assoc-voc	<=50K	nan	nan
2321	Laos	Machine-op-inspct	Female	White	10th	<=50K	nan	nan
1038	Germany	Prof-specialty	Female	Asian-Pac-Islander	Assoc-voc	<=50K	nan	nan

Synthetic								
	native.country	occupation	sex	race	education	income	AVG(fnlwgt)	COUNT(*)
4561	United-States	Farming-fishing	Male	White	Assoc-acdm	<=50K	172474.395027	5
4082	Thailand	Exec-managerial	Female	White	Some-college	<=50K	177340.672565	1
3513	Puerto-Rico	Prof-specialty	Female	Black	Assoc-voc	<=50K	170462.094245	1
2321	Laos	Machine-op-inspct	Female	White	10th	<=50K	178949.079041	1
1038	Germany	Prof-specialty	Female	Asian-Pac-Islander	Assoc-voc	<=50K	184620.456806	1

SQL for Real:
SELECT `native.country`,occupation,sex,race,education,income,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (race IN ('Asian-Pac-Islander', 'Black', 'White', 'Amer-Indian-Eskimo') AND capital <= 0) OR (education = 'HS-grad' OR workclass LIKE 'Private') GROUP BY `native.country`,occupation,sex,race,education,income
Resulted in 4191 records

SQL for Synthetic:
SELECT `native.country`,occupation,sex,race,education,income,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (race IN ('Asian-Pac-Islander', 'Black', 'White', 'Amer-Indian-Eskimo') AND capital <= 0) OR (education = 'HS-grad' OR workclass LIKE 'Private') GROUP BY `native.country`,occupation,sex,race,education,income
Resulted in 5315 records

Normalized Euclidean distance for (fnlwgt): 40.26
Hellinger Distance: 0.326

=====

Real

	marital.status	education	workclass	MIN(fnlwgt)	COUNT(*)
203	Married-civ-spouse	Preschool	Self-emp-not-inc	227253	3
321	Never-married	9th	?	30140	21
477	Widowed	11th	State-gov	159908	1
34	Divorced	9th	Private	22418	75
397	Separated	11th	Self-emp-not-inc	210926	3

Synthetic

	marital.status	education	workclass	MIN(fnlwgt)	COUNT(*)
203	Married-civ-spouse	Preschool	Self-emp-not-inc	157059.527311	2.000000
321	Never-married	9th	?	148353.803745	15.000000
477	Widowed	11th	State-gov	149395.201919	1.000000
34	Divorced	9th	Private	156179.028282	80.000000
397	Separated	11th	Self-emp-not-inc	159134.242918	2.000000

SQL for Real:
SELECT `marital.status`,education,workclass,MIN(fnlwgt), COUNT(*) FROM C1 WHERE (race IN ('Asian-Pac-Islander', 'White', 'Black')) GROUP BY `marital.status`,education,workclass

Resulted in 558 records

SQL for Synthetic:
SELECT `marital.status`,education,workclass,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (race IN ('Asian-Pac-Islander', 'White', 'Black')) GROUP BY `marital.status`,education,workclass

Resulted in 546 records

Normalized Euclidean distance for (fnlwgt): 22.29

Hellinger Distance: 0.054

=====

Real

	relationship	occupation	sex	education	workclass	race	AVG(`hours.per.week`)	COUNT(*)
3867	Own-child	Prof-specialty	Female	Bachelors	Local-gov	White	36.500000	28
2309	Not-in-family	Prof-specialty	Male	Some-college	Self-emp-not-inc	Black	40.000000	1
2178	Not-in-family	Prof-specialty	Female	Doctorate	Federal-gov	White	84.500000	2
2402	Not-in-family	Sales	Female	Assoc-acdm	Self-emp-inc	White	25.000000	1
433	Husband	Farming-fishing	Male	7th-8th	?	White	20.000000	1

Synthetic

	relationship	occupation	sex	education	workclass	race	AVG(`hours.per.week`)	COUNT(*)
3867	Own-child	Prof-specialty	Female	Bachelors	Local-gov	White	39.956706	31.000000
2309	Not-in-family	Prof-specialty	Male	Some-college	Self-emp-not-inc	Black	nan	nan
2178	Not-in-family	Prof-specialty	Female	Doctorate	Federal-gov	White	40.022744	1.000000
2402	Not-in-family	Sales	Female	Assoc-acdm	Self-emp-inc	White	40.041420	1.000000
433	Husband	Farming-fishing	Male	7th-8th	?	White	nan	nan

SQL for Real:
SELECT relationship,occupation,sex,education,workclass,race,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (`hours.per.week` <> 50) OR (relationship = 'Husband' OR age BETWEEN 58 AND 37) AND (workclass = 'Private' AND capital < 0) AND (education IN ('12th', 'HS-grad', '5th-6th', 'Doctorate', 'Prof-school', 'Assoc-voc', 'Some-college', '1st-4th', 'Preschool', 'Assoc-acdm', '11th', 'Bachelors', '10th', '9th', 'Masters') AND income LIKE '<=50K') AND (sex <> 'Male' AND occupation IN ('Tech-support', 'Machine-op-inspct')) GROUP BY relationship,occupation,sex,education,workclass,race

Resulted in 5668 records

SQL for Synthetic:
SELECT relationship,occupation,sex,education,workclass,race,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` <> 50) OR (relationship = 'Husband' OR age BETWEEN 58 AND 37) AND (workclass = 'Private' AND capital < 0) AND (education IN ('12th', 'HS-grad', '5th-6th', 'Doctorate', 'Prof-school', 'Assoc-voc', 'Some-college', '1st-4th', 'Preschool', 'Assoc-acdm', '11th', 'Bachelors', '10th', '9th', 'Masters') AND income LIKE '<=50K') AND (sex <> 'Male' AND occupation IN ('Tech-support', 'Machine-op-inspct')) GROUP BY relationship,occupation,sex,education,workclass,race

Resulted in 5511 records

Normalized Euclidean distance for (`hours.per.week`): 59.16

Hellinger Distance: 0.127

=====

Real

	native.country	workclass	race	marital.status	occupation	education	AVG(`hours.per.week`)	COUNT(*)
1086	Ireland	Private	White	Married-civ-spouse	Tech-support	Assoc-voc	nan	nan
234	?	Self-emp-not-inc	White	Married-civ-spouse	Exec-managerial	Bachelors	47.500000	2.000000
1183	Japan	Private	Black	Married-civ-spouse	Sales	Masters	nan	nan
2847	United-States	Federal-gov	White	Married-civ-spouse	Armed-Forces	Bachelors	nan	nan
3171	United-States	Private	Black	Married-civ-spouse	Prof-specialty	HS-grad	45.000000	4.000000

Synthetic

	native.country	workclass	race	marital.status	occupation	education	AVG(`hours.per.week`)	COUNT(*)
1086	Ireland	Private	White	Married-civ-spouse	Tech-support	Assoc-voc	40.000842	1
234	?	Self-emp-not-inc	White	Married-civ-spouse	Exec-managerial	Bachelors	40.389015	9
1183	Japan	Private	Black	Married-civ-spouse	Sales	Masters	40.447918	1
2847	United-States	Federal-gov	White	Married-civ-spouse	Armed-Forces	Bachelors	39.962412	1
3171	United-States	Private	Black	Married-civ-spouse	Prof-specialty	HS-grad	40.015199	3

SQL for Real:
SELECT `native.country`,workclass,race,`marital.status`,occupation,education,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (`marital.status` LIKE 'Married-civ-spouse') GROUP BY `native.country`,workclass,race,`marital.status`,occupation,education

Resulted in 2638 records

SQL for Synthetic:
SELECT `native.country`,workclass,race,`marital.status`,occupation,education,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`marital.status` LIKE 'Married-civ-spouse') GROUP BY `native.country`,workclass,race,`marital.status`,occupation,education

Resulted in 3777 records

Normalized Euclidean distance for (`hours.per.week`): 32.25

Hellinger Distance: 0.272

=====

Real								
	marital.status	workclass	education	relationship	native.country	race	SUM(fnlwgt)	COUNT(*)
2730	Married-civ-spouse	Self-emp-inc	Assoc-voc	Husband	Columbia	White	nan	nan
2766	Married-civ-spouse	Self-emp-inc	HS-grad	Husband	Guatemala	White	nan	nan
5766	Separated	Private	11th	Not-in-family	Puerto-Rico	White	319378.000000	2.000000
3162	Married-civ-spouse	State-gov	Masters	Husband	Japan	Black	nan	nan
1671	Married-civ-spouse	Local-gov	Assoc-acdm	Husband	Puerto-Rico	Black	nan	nan

Synthetic								
	marital.status	workclass	education	relationship	native.country	race	SUM(fnlwgt)	COUNT(*)
2730	Married-civ-spouse	Self-emp-inc	Assoc-voc	Husband	Columbia	White	355041.439223	2
2766	Married-civ-spouse	Self-emp-inc	HS-grad	Husband	Guatemala	White	187969.829613	1
5766	Separated	Private	11th	Not-in-family	Puerto-Rico	White	850749.111337	5
3162	Married-civ-spouse	State-gov	Masters	Husband	Japan	Black	178175.722695	1
1671	Married-civ-spouse	Local-gov	Assoc-acdm	Husband	Puerto-Rico	Black	171313.754768	1

SQL for Real:
SELECT `marital.status`,workclass,education,relationship,`native.country`,race,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (occupation LIKE 'Adm-clerical') OR (race IN ('Black', 'White', 'Amer-Indian-Eskimo', 'Other', 'Asian-Pac-Islander') AND capital <= 0) OR (workclass LIKE 'Private' OR `native.country` <> 'United-States') OR (age >= 65 OR sex LIKE 'Male') AND (`marital.status` LIKE 'Married-civ-spouse' AND education LIKE 'HS-grad') AND (relationship <> 'Husband' AND income LIKE '<=50K') GROUP BY `marital.status`,workclass,education,relationship,`native.country`,race

Resulted in 4960 records

SQL for Synthetic:
SELECT `marital.status`,workclass,education,relationship,`native.country`,race,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (occupation LIKE 'Adm-clerical') OR (race IN ('Black', 'White', 'Amer-Indian-Eskimo', 'Other', 'Asian-Pac-Islander') AND capital <= 0) OR (workclass LIKE 'Private' OR `native.country` <> 'United-States') OR (age >= 65 OR sex LIKE 'Male') AND (`marital.status` LIKE 'Married-civ-spouse' AND education LIKE 'HS-grad') AND (relationship <> 'Husband' AND income LIKE '<=50K') GROUP BY `marital.status`,workclass,education,relationship,`native.country`,race

Resulted in 6619 records

Normalized Euclidean distance for (fnlwgt): 42.45

Hellinger Distance: 0.34

=====

Real			
	race	AVG(age)	COUNT(*)
3	Other	33.682927	205
4	White	38.315777	38201
1	Asian-Pac-Islander	37.974901	757
0	Amer-Indian-Eskimo	36.918367	245
2	Black	38.181535	2567

Synthetic			
	race	AVG(age)	COUNT(*)
3	Other	nan	nan
4	White	38.805845	21694.000000
1	Asian-Pac-Islander	65.000000	2.000000
0	Amer-Indian-Eskimo	nan	nan
2	Black	47.500000	2.000000

SQL for Real:
SELECT race,AVG(age), COUNT(*) FROM C1 WHERE (race = 'White' OR `hours.per.week` = 40) AND (sex IN ('Female', 'Male') AND capital <= 0) OR (fnlwgt <= 35603 OR age > 45) AND (relationship <> 'Husband' AND education <> 'HS-grad') AND (income LIKE '>50K' AND workclass = 'Self-emp-inc') GROUP BY race

Resulted in 5 records

SQL for Synthetic:
SELECT race,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (race = 'White' OR `hours.per.week` = 40) AND (sex IN ('Female', 'Male') AND capital <= 0) OR (fnlwgt <= 35603 OR age > 45) AND (relationship <> 'Husband' AND education <> 'HS-grad') AND (income LIKE '>50K' AND workclass = 'Self-emp-inc') GROUP BY race

Resulted in 3 records

Normalized Euclidean distance for (age): 1.73

Hellinger Distance: 0.193

=====

Real							
	race	occupation	marital.status	income	workclass	AVG(age)	COUNT(*)
727	White	Transport-moving	Never-married	>50K	State-gov	41.000000	1
522	White	Craft-repair	Never-married	>50K	Local-gov	60.500000	2
227	Asian-Pac-Islander	Transport-moving	Never-married	<=50K	Private	28.125000	16
354	Black	Protective-serv	Married-civ-spouse	<=50K	Private	41.071429	14
34	Amer-Indian-Eskimo	Handlers-cleaners	Married-spouse-absent	<=50K	Private	29.500000	2

Synthetic							
	race	occupation	marital.status	income	workclass	AVG(age)	COUNT(*)
727	White	Transport-moving	Never-married	>50K	State-gov	nan	nan
522	White	Craft-repair	Never-married	>50K	Local-gov	nan	nan
227	Asian-Pac-Islander	Transport-moving	Never-married	<=50K	Private	25.625000	8.000000
354	Black	Protective-serv	Married-civ-spouse	<=50K	Private	40.466667	15.000000
34	Amer-Indian-Eskimo	Handlers-cleaners	Married-spouse-absent	<=50K	Private	nan	nan

SQL for Real:
SELECT race,occupation,`marital.status`,income,workclass,AVG(age), COUNT(*) FROM C1 WHERE (occupation = 'Other-service' AND fnlwgt BETWEEN 117789 AND 33811) OR (`marital.status` = 'Never-married' OR workclass = 'Private') GROUP BY race,occupation,`marital.status`,income,workclass
Resulted in 732 records

SQL for Synthetic:
SELECT race,occupation,`marital.status`,income,workclass,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (occupation = 'Other-service' AND fnlwgt BETWEEN 117789 AND 33811) OR (`marital.status` = 'Never-married' OR workclass = 'Private') GROUP BY race,occupation,`marital.status`,income,workclass
Resulted in 604 records

Normalized Euclidean distance for (age): 23.02

Hellinger Distance: 0.16

=====

Real										
	income	workclass	sex	race	marital.status	occupation	native.country	relationship	education	MAX(age) COUNT(*)
0	>50K	?	Female	White	Married-civ-spouse	?	United-States	Other-relative	12th	19 1

Synthetic										
	income	workclass	sex	race	marital.status	occupation	native.country	relationship	education	MAX(age) COUNT(*)
0	>50K	?	Female	White	Married-civ-spouse	?	United-States	Other-relative	12th	nan nan

SQL for Real:
SELECT income,workclass,sex,race,`marital.status`,occupation,`native.country`,relationship,education,MAX(age), COUNT(*) FROM C1 WHERE (`native.country` = 'United-States' AND age = 19) AND (`marital.status` <> 'Never-married' AND income <> '<=50K') AND (workclass IN ('Self-emp-inc', '?', 'Self-emp-not-inc', 'State-gov', 'Federal-gov') AND fnlwgt BETWEEN 182227 AND 239130) GROUP BY income,workclass,sex,race,`marital.status`,occupation,`native.country`,relationship,education
Resulted in 1 records

SQL for Synthetic:
SELECT income,workclass,sex,race,`marital.status`,occupation,`native.country`,relationship,education,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (`native.country` = 'United-States' AND age = 19) AND (`marital.status` <> 'Never-married' AND income <> '<=50K') AND (workclass IN ('Self-emp-inc', '?', 'Self-emp-not-inc', 'State-gov', 'Federal-gov') AND fnlwgt BETWEEN 182227 AND 239130) GROUP BY income,workclass,sex,race,`marital.status`,occupation,`native.country`,relationship,education
Resulted in 0 records

Normalized Euclidean distance for (age): nan

Hellinger Distance: nan

=====

Real									
	marital.status	race	income	education	occupation	relationship	workclass	AVG(`hours.per.week`)	COUNT(*)
654	Divorced	White	<=50K	Bachelors	Prof-specialty	Not-in-family	Private	40.439024	41
4028	Never-married	Asian-Pac-Islander	<=50K	Bachelors	Prof-specialty	Own-child	Self-emp-not-inc	20.000000	1
129	Divorced	Black	<=50K	11th	?	Unmarried	?	19.500000	4
2608	Married-civ-spouse	White	<=50K	HS-grad	Craft-repair	Husband	State-gov	40.086957	23
730	Divorced	White	<=50K	HS-grad	Other-service	Own-child	Self-emp-not-inc	32.000000	2

Synthetic									
	marital.status	race	income	education	occupation	relationship	workclass	AVG(`hours.per.week`)	COUNT(*)
654	Divorced	White	<=50K	Bachelors	Prof-specialty	Not-in-family	Private	40.080801	66.000000
4028	Never-married	Asian-Pac-Islander	<=50K	Bachelors	Prof-specialty	Own-child	Self-emp-not-inc	40.064651	1.000000
129	Divorced	Black	<=50K	11th	?	Unmarried	?	nan	nan
2608	Married-civ-spouse	White	<=50K	HS-grad	Craft-repair	Husband	State-gov	39.993623	24.000000
730	Divorced	White	<=50K	HS-grad	Other-service	Own-child	Self-emp-not-inc	nan	nan

SQL for Real:
SELECT `marital.status`,race,income,education,occupation,relationship,workclass,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (age <= 23 AND capital BETWEEN 0 AND 0) OR (relationship = 'Husband' OR workclass IN ('?', 'Self-emp-not-inc', 'Private')) OR (income = '<=50K' OR `native.country` = 'United-States') AND (education LIKE 'Bachelors' AND sex <> 'Male') AND (race <> 'White' AND occupation IN ('Adm-clerical', '?', 'Priv-house-serv', 'Machine-op-inspct', 'Craft-repair', 'Sales', 'Transport-moving', 'Farming-fishing', 'Protective-serv', 'Prof-specialty', 'Other-service', 'Handlers-cleaners')) AND (`marital.status` LIKE 'Married-civ-spouse' AND `hours.per.week` = 40) GROUP BY `marital.status`,race,income,education,occupation,relationship,workclass
Resulted in 6713 records

SQL for Synthetic:
SELECT `marital.status`,race,income,education,occupation,relationship,workclass,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (age <= 23 AND capital BETWEEN 0 AND 0) OR (relationship = 'Husband' OR workclass IN ('?', 'Self-emp-not-inc', 'Private')) OR (income = '<=50K' OR `native.country` = 'United-States') AND (education LIKE 'Bachelors' AND sex <> 'Male') AND (race <> 'White' AND occupation IN ('Adm-clerical', '?', 'Priv-house-serv', 'Machine-op-inspct', 'Craft-repair', 'Sales', 'Transport-moving', 'Farming-fishing', 'Protective-serv', 'Prof-specialty', 'Other-service', 'Handlers-cleaners')) AND (`marital.status` LIKE 'Married-civ-spouse' AND `hours.per.week` = 40) GROUP BY `marital.status`,race,income,education,occupation,relationship,workclass
Resulted in 5701 records

Normalized Euclidean distance for (`hours.per.week`): 59.87

Hellinger Distance: 0.203

=====

Real						
	marital.status	occupation	sex	relationship	income	native.country AVG(age) COUNT(*)
0	Never-married	Other-service	Female	Not-in-family	<=50K	United-States 20.000000 1

Synthetic

marital.status occupation sex relationship income native.country AVG(age) COUNT(*)

0 Never-married Other-service Female Not-in-family <=50K United-States nan nan

SQL for Real:

SELECT `marital.status`,occupation,sex,relationship,income,`native.country`,AVG(age), COUNT(*) FROM C1 WHERE (fnlwgt = 122649) GROUP BY `marital.status`,occupation,sex,relationship,income,`native.country`

Resulted in 1 records

SQL for Synthetic:

SELECT `marital.status`,occupation,sex,relationship,income,`native.country`,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt = 122649) GROUP BY `marital.status`,occupation,sex,relationship,income,`native.country`

Resulted in 0 records

Normalized Euclidean distance for (age): nan

Hellinger Distance: nan

=====

Real

occupation relationship native.country MIN(age) COUNT(*)

1236 Protective-serv Own-child Ireland nan nan

44 ? Other-relative Hungary nan nan

1317 Sales Other-relative Canada nan nan

944 Other-service Own-child Thailand nan nan

656 Handlers-cleaners Other-relative Japan nan nan

Synthetic

occupation relationship native.country MIN(age) COUNT(*)

1236 Protective-serv Own-child Ireland 30 5

44 ? Other-relative Hungary 49 1

1317 Sales Other-relative Canada 23 2

944 Other-service Own-child Thailand 18 26

656 Handlers-cleaners Other-relative Japan 45 1

SQL for Real:

SELECT occupation,relationship,`native.country`,MIN(age), COUNT(*) FROM C1 WHERE (capital > 0 OR `native.country` LIKE 'United-States') OR (race IN ('Other', 'Black', 'Asian-Pac-Islander', 'White') OR income IN ('>50K', '<=50K')) AND (workclass IN ('Private', 'Local-gov', 'Federal-gov', 'Self-emp-inc', '?', 'Self-emp-not-inc', 'State-gov') AND occupation <> 'Other-service') AND (relationship = 'Husband' AND sex <> 'Male') OR (`hours.per.week` <> 16 OR fnlwgt BETWEEN 136687 AND 74775) GROUP BY occupation,relationship,`native.country`

Resulted in 1480 records

SQL for Synthetic:

SELECT occupation,relationship,`native.country`,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (capital > 0 OR `native.country` LIKE 'United-States') OR (race IN ('Other', 'Black', 'Asian-Pac-Islander', 'White') OR income IN ('>50K', '<=50K')) AND (workclass IN ('Private', 'Local-gov', 'Federal-gov', 'Self-emp-inc', '?', 'Self-emp-not-inc', 'State-gov') AND occupation <> 'Other-service') AND (relationship = 'Husband' AND sex <> 'Male') OR (`hours.per.week` <> 16 OR fnlwgt BETWEEN 136687 AND 74775) GROUP BY occupation,relationship,`native.country`

Resulted in 1585 records

Normalized Euclidean distance for (age): 28.97

Hellinger Distance: 0.369

=====

Real

occupation workclass SUM(capital) COUNT(*)

88 Transport-moving ? 2414.000000 1.000000

6 Adm-clerical ? 0.000000 2.000000

64 Prof-specialty Private 8035962.000000 3397.000000

13 Armed-Forces Federal-gov 5411.000000 14.000000

35 Handlers-cleaners ? 4416.000000 2.000000

Synthetic

occupation workclass SUM(capital) COUNT(*)

88 Transport-moving ? -10.865275 2

6 Adm-clerical ? 39.801273 5

64 Prof-specialty Private 51297.453965 3461

13 Armed-Forces Federal-gov -1319.293983 117

35 Handlers-cleaners ? 13.734085 2

SQL for Real:

SELECT occupation,workclass,SUM(capital), COUNT(*) FROM C1 WHERE (education <> 'HS-grad' OR income = '>50K') OR (race LIKE 'White' OR `marital.status` LIKE 'Separated') OR (`native.country` LIKE 'United-States' OR capital < 0) AND (workclass = 'Private' AND sex <> 'Male') GROUP BY occupation,workclass

Resulted in 83 records

SQL for Synthetic:

SELECT occupation,workclass,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (education <> 'HS-grad' OR income = '>50K') OR (race LIKE 'White' OR `marital.status` LIKE 'Separated') OR (`native.country` LIKE 'United-States' OR capital < 0) AND (workclass = 'Private' AND sex <> 'Male') GROUP BY occupation,workclass

Resulted in 95 records

Normalized Euclidean distance for (capital): 9.11

Hellinger Distance: 0.036

=====

Real						
	marital.status	native.country	occupation	income	MIN(capital)	COUNT(*)
1587	Widowed	Japan	Adm-clerical	<=50K	nan	nan
1521	Separated	United-States	Other-service	<=50K	-1876.000000	212.000000
1164	Never-married	Mexico	?	<=50K	0.000000	17.000000
759	Married-civ-spouse	United-States	Armed-Forces	<=50K	0.000000	2.000000
224	Divorced	Puerto-Rico	Transport-moving	<=50K	0.000000	1.000000

Synthetic						
	marital.status	native.country	occupation	income	MIN(capital)	COUNT(*)
1587	Widowed	Japan	Adm-clerical	<=50K	-259.120491	9
1521	Separated	United-States	Other-service	<=50K	-273.925045	141
1164	Never-married	Mexico	?	<=50K	158.783815	1
759	Married-civ-spouse	United-States	Armed-Forces	<=50K	-135.748898	9
224	Divorced	Puerto-Rico	Transport-moving	<=50K	-226.296837	48

SQL for Real:
SELECT `marital.status`,`native.country`,occupation,income,MIN(capital), COUNT(*) FROM C1 WHERE (workclass <> 'Self-emp-not-inc') AND (education <> 'Some-college' OR `hours.per.week` > 11) GROUP BY `marital.status`,`native.country`,occupation,income

Resulted in 1608 records

SQL for Synthetic:
SELECT `marital.status`,`native.country`,occupation,income,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (workclass <> 'Self-emp-not-inc') AND (education <> 'Some-college' OR `hours.per.week` > 11) GROUP BY `marital.status`,`native.country`,occupation,income

Resulted in 1677 records

Normalized Euclidean distance for (capital): 28.79

Hellinger Distance: 0.373

=====

Real			
	education	AVG(capital)	COUNT(*)
7	Assoc-acdm	546.376015	1601
6	9th	277.482804	756
9	Bachelors	1637.793645	8025
2	12th	179.727549	657
0	10th	269.432997	1388

Synthetic			
	education	AVG(capital)	COUNT(*)
7	Assoc-acdm	-6.168588	1755.000000
6	9th	4.983828	737.000000
9	Bachelors	-0.440649	8164.000000
2	12th	0.125910	626.000000
0	10th	-0.458255	1430.000000

SQL for Real:
SELECT education,AVG(capital), COUNT(*) FROM C1 WHERE (age >= 26 OR education = 'Prof-school') OR (`hours.per.week` >= 35 OR relationship <> 'Husband') OR (income = '>50K' OR `native.country` = 'United-States') AND (sex <> 'Male' AND race IN ('Other', 'Asian-Pac-Islander', 'Black')) AND (`marital.status` IN ('Widowed', 'Married-civ-spouse', 'Separated', 'Divorced', 'Married-AF-spouse', 'Married-spouse-absent', 'Never-married') AND fnlwgt <> 160120) OR (capital >= 0 OR occupation = 'Exec-managerial') GROUP BY education

Resulted in 16 records

SQL for Synthetic:
SELECT education,AVG(capital), COUNT(*) FROM C1_syn_06 WHERE (age >= 26 OR education = 'Prof-school') OR (`hours.per.week` >= 35 OR relationship <> 'Husband') OR (income = '>50K' OR `native.country` = 'United-States') AND (sex <> 'Male' AND race IN ('Other', 'Asian-Pac-Islander', 'Black')) AND (`marital.status` IN ('Widowed', 'Married-civ-spouse', 'Separated', 'Divorced', 'Married-AF-spouse', 'Married-spouse-absent', 'Never-married') AND fnlwgt <> 160120) OR (capital >= 0 OR occupation = 'Exec-managerial') GROUP BY education

Resulted in 16 records

Normalized Euclidean distance for (capital): 4.0

Hellinger Distance: 0.026

=====

Real							
workclass	relationship	race	occupation	education	native.country	income	sex SUM(capital) COUNT(*)
Synthetic							
workclass	relationship	race	occupation	education	native.country	income	sex SUM(capital) COUNT(*)

SQL for Real:
SELECT workclass,relationship,race,occupation,education,`native.country`,income,sex,SUM(capital), COUNT(*) FROM C1 WHERE (race <> 'Black') AND (income <> '<=50K' AND `marital.status` LIKE 'Married-civ-spouse') AND (`hours.per.week` > 20 AND relationship = 'Unmarried') GROUP BY workclass,relationship,race,occupation,education,`native.country`,income,sex

Resulted in 0 records

SQL for Synthetic:
SELECT workclass,relationship,race,occupation,education,`native.country`,income,sex,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (race <> 'Black') AND (income <> '<=50K' AND `marital.status` LIKE 'Married-civ-spouse') AND (`hours.per.week` > 20 AND relationship = 'Unmarried') GROUP BY workclass,relationship,race,occupation,education,`native.country`,income,sex

Resulted in 0 records

Normalized Euclidean distance for (capital): nan

Hellinger Distance: nan

=====

Real					
	native.country	marital.status	relationship	AVG(`hours.per.week`)	COUNT(*)
21	?	Never-married	Not-in-family	40.217949	156.000000
319	Laos	Divorced	Own-child	nan	nan
64	Columbia	Divorced	Husband	nan	nan
225	Hungary	Widowed	Not-in-family	11.500000	2.000000
544	United-States	Married-spouse-absent	Own-child	36.795455	44.000000

Synthetic					
	native.country	marital.status	relationship	AVG(`hours.per.week`)	COUNT(*)
21	?	Never-married	Not-in-family	40.023821	164
319	Laos	Divorced	Own-child	40.014150	12
64	Columbia	Divorced	Husband	40.116834	2
225	Hungary	Widowed	Not-in-family	40.008019	4
544	United-States	Married-spouse-absent	Own-child	39.913841	39

SQL for Real:
SELECT `native.country`,`marital.status`,relationship,AVG(`hours.per.week`), COUNT(*) FROM C1 WHERE (`native.country` LIKE 'United-States' OR income <> '>50K') AND (fnlwgt >= 268832 AND workclass = 'Private') OR (`marital.status` <> 'Married-civ-spouse' OR age <= 30) OR (relationship IN ('Own-child', 'Wife', 'Husband', 'Unmarried', 'Other-relative') OR `hours.per.week` BETWEEN 20 AND 40) GROUP BY `native.country`,`marital.status`,relationship

Resulted in 575 records

SQL for Synthetic:
SELECT `native.country`,`marital.status`,relationship,AVG(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`native.country` LIKE 'United-States' OR income <> '>50K') AND (fnlwgt >= 268832 AND workclass = 'Private') OR (`marital.status` <> 'Married-civ-spouse' OR age <= 30) OR (relationship IN ('Own-child', 'Wife', 'Husband', 'Unmarried', 'Other-relative') OR `hours.per.week` BETWEEN 20 AND 40) GROUP BY `native.country`,`marital.status`,relationship

Resulted in 588 records

Normalized Euclidean distance for (`hours.per.week`): 19.13

Hellinger Distance: 0.397

=====

Real			
	education	MAX(capital)	COUNT(*)

Synthetic			
	education	MAX(capital)	COUNT(*)

SQL for Real:
SELECT education,MAX(capital), COUNT(*) FROM C1 WHERE (occupation <> '?') AND (`native.country` LIKE '?' OR `marital.status` <> 'Never-married') AND (education LIKE 'Bachelors' AND fnlwgt = 248445) AND (age >= 42 AND capital = 0) GROUP BY education

Resulted in 0 records

SQL for Synthetic:
SELECT education,MAX(capital), COUNT(*) FROM C1_syn_06 WHERE (occupation <> '?') AND (`native.country` LIKE '?' OR `marital.status` <> 'Never-married') AND (education LIKE 'Bachelors' AND fnlwgt = 248445) AND (age >= 42 AND capital = 0) GROUP BY education

Resulted in 0 records

Normalized Euclidean distance for (capital): nan

Hellinger Distance: nan

=====

Real				
	workclass	income	MIN(fnlwgt)	COUNT(*)
7	Private	>50K	13769	1330
1	?	>50K	109912	35
12	State-gov	<=50K	26880	29
4	Local-gov	<=50K	31251	56
5	Local-gov	>50K	24982	212

Synthetic				
	workclass	income	MIN(fnlwgt)	COUNT(*)
7	Private	>50K	167662.404018	98.000000
1	?	>50K	178959.896672	1.000000
12	State-gov	<=50K	154372.541348	45.000000
4	Local-gov	<=50K	170514.500354	80.000000
5	Local-gov	>50K	171963.230511	17.000000

SQL for Real:
SELECT workclass,income,MIN(fnlwgt), COUNT(*) FROM C1 WHERE (age = 44) OR (education IN ('Some-college', '12th', 'Assoc-voc', '5th-6th', '11th', 'Prof-school', 'Assoc-acdm', '7th-8th', '1st-4th', 'Preschool', 'HS-grad', '10th', 'Masters', 'Bachelors') OR capital > 0) AND ((fnlwgt BETWEEN 200381 AND 250804) AND income <> '<=50K') GROUP BY workclass,income

Resulted in 14 records

SQL for Synthetic:
SELECT workclass,income,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (age = 44) OR (education IN ('Some-college', '12th', 'Assoc-voc', '5th-6th', '11th', 'Prof-school', 'Assoc-acdm', '7th-8th', '1st-4th', 'Preschool', 'HS-grad', '10th', 'Masters', 'Bachelors') OR capital > 0) AND ((fnlwgt BETWEEN 200381 AND 250804) AND income <> '<=50K') GROUP BY workclass,income

Resulted in 14 records

Normalized Euclidean distance for (fnlwgt): 3.74

Hellinger Distance: 0.462

=====

Real			
	marital.status	MIN(`hours.per.week`)	COUNT(*)
0	Divorced	1	3301
6	Widowed	3	271
1	Married-AF-spouse	10	18
3	Married-spouse-absent	5	289
2	Married-civ-spouse	1	16591

SQL for Real:
SELECT `marital.status`,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (relationship IN ('Not-in-family', 'Wife', 'Husband', 'Other-relative', 'Own-child') AND age BETWEEN 30 AND 57) GROUP BY `marital.status`

Resulted in 7 records

SQL for Synthetic:
SELECT `marital.status`,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (relationship IN ('Not-in-family', 'Wife', 'Husband', 'Other-relative', 'Own-child') AND age BETWEEN 30 AND 57) GROUP BY `marital.status`

Resulted in 6 records

Normalized Euclidean distance for (`hours.per.week`): 2.45

Hellinger Distance: 0.014

=====

Real						
	marital.status	relationship	occupation	sex	MIN(capital)	COUNT(*)
512	Separated	Not-in-family	Transport-moving	Male	0.000000	27.000000
613	Widowed	Not-in-family	Prof-specialty	Male	-2282.000000	12.000000
590	Separated	Wife	Machine-op-inspct	Female	nan	nan
606	Widowed	Not-in-family	Machine-op-inspct	Female	-1870.000000	35.000000
281	Married-spouse-absent	Other-relative	Craft-repair	Male	0.000000	9.000000

Synthetic						
	marital.status	relationship	occupation	sex	MIN(capital)	COUNT(*)
512	Separated	Not-in-family	Transport-moving	Male	-202.109058	15
613	Widowed	Not-in-family	Prof-specialty	Male	-209.797813	13
590	Separated	Wife	Machine-op-inspct	Female	23.271080	1
606	Widowed	Not-in-family	Machine-op-inspct	Female	-267.752085	37
281	Married-spouse-absent	Other-relative	Craft-repair	Male	-57.081414	2

SQL for Real:
SELECT `marital.status`,relationship,occupation,sex,MIN(capital), COUNT(*) FROM C1 WHERE (capital = 15024 OR income = '<=50K') AND (relationship = 'Unmarried' AND `hours.per.week` = 40) AND (occupation LIKE 'Prof-specialty' AND age > 31) AND (workclass = 'Private' AND sex = 'Male') OR (`marital.status` IN ('Separated', 'Married-AF-spouse', 'Married-spouse-absent', 'Married-civ-spouse', 'Divorced', 'Widowed', 'Never-married') OR education IN ('HS-grad', '1st-4th', '12th', '10th', 'Preschool', 'Assoc-acdm', 'Assoc-voc')) GROUP BY `marital.status`,relationship,occupation,sex

Resulted in 611 records

SQL for Synthetic:
SELECT `marital.status`,relationship,occupation,sex,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (capital = 15024 OR income = '<=50K') AND (relationship = 'Unmarried' AND `hours.per.week` = 40) AND (occupation LIKE 'Prof-specialty' AND age > 31) AND (workclass = 'Private' AND sex = 'Male') OR (`marital.status` IN ('Separated', 'Married-AF-spouse', 'Married-spouse-absent', 'Married-civ-spouse', 'Divorced', 'Widowed', 'Never-married') OR education IN ('HS-grad', '1st-4th', '12th', '10th', 'Preschool', 'Assoc-acdm', 'Assoc-voc')) GROUP BY `marital.status`,relationship,occupation,sex

Resulted in 688 records

Normalized Euclidean distance for (capital): 23.49

Hellinger Distance: 0.059

=====

Real					
	workclass	relationship	education	SUM(fnlwgt)	COUNT(*)
54	State-gov	Unmarried	Assoc-voc	225160	2
1	?	Unmarried	11th	231016	2
22	Local-gov	Unmarried	Masters	468000	4
24	Local-gov	Unmarried	Some-college	2061205	16
0	?	Unmarried	10th	339787	3

Synthetic					
	workclass	relationship	education	SUM(fnlwgt)	COUNT(*)
54	State-gov	Unmarried	Assoc-voc	nan	nan
1	?	Unmarried	11th	nan	nan
22	Local-gov	Unmarried	Masters	nan	nan
24	Local-gov	Unmarried	Some-college	nan	nan
0	?	Unmarried	10th	nan	nan

SQL for Real:
SELECT workclass,relationship,education,SUM(fnlwgt), COUNT(*) FROM C1 WHERE ((`hours.per.week` BETWEEN 20 AND 40) AND relationship LIKE 'Unmarried') AND

(capital = 0 AND fnlwgt BETWEEN 99665 AND 156606) GROUP BY workclass,relationship,education

Resulted in 59 records

SQL for Synthetic:
SELECT workclass,relationship,education,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE ((` hours.per.week` BETWEEN 20 AND 40) AND relationship LIKE 'Unmarried') AND (capital = 0 AND fnlwgt BETWEEN 99665 AND 156606) GROUP BY workclass,relationship,education

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

=====

Real									
	race	native.country	sex	relationship	income	occupation	marital.status	workclass	SUM(` hours.per.week`) COUNT(*)
1660	Black	Puerto-Rico	Female	Wife	<=50K	Exec-managerial	Married-civ-spouse	Private	nan nan
5591	White	Peru	Male	Husband	>50K	Exec-managerial	Married-civ-spouse	Self-emp-inc	nan nan
3896	White	Germany	Male	Not-in-family	<=50K	Machine-op-inspct	Never-married	Private	nan nan
7139	White	Taiwan	Male	Husband	<=50K	Machine-op-inspct	Married-civ-spouse	Local-gov	nan nan
8406	White	United-States	Male	Not-in-family	>50K	Prof-specialty	Never-married	Private	3237.000000 71.000000

Synthetic									
	race	native.country	sex	relationship	income	occupation	marital.status	workclass	SUM(` hours.per.week`) COUNT(*)
1660	Black	Puerto-Rico	Female	Wife	<=50K	Exec-managerial	Married-civ-spouse	Private	200.046887 5
5591	White	Peru	Male	Husband	>50K	Exec-managerial	Married-civ-spouse	Self-emp-inc	40.353524 1
3896	White	Germany	Male	Not-in-family	<=50K	Machine-op-inspct	Never-married	Private	40.013104 1
7139	White	Taiwan	Male	Husband	<=50K	Machine-op-inspct	Married-civ-spouse	Local-gov	40.007008 1
8406	White	United-States	Male	Not-in-family	>50K	Prof-specialty	Never-married	Private	442.511919 11

SQL for Real:
SELECT race,` native.country`,sex,relationship,income,occupation,` marital.status`,workclass,SUM(` hours.per.week`), COUNT(*) FROM C1 WHERE (` hours.per.week` >= 40) OR ((age BETWEEN 45 AND 64) OR education IN ('1st-4th', 'Preschool', '5th-6th', 'Some-college', '11th', '10th')) OR (` marital.status` = 'Never-married' OR capital BETWEEN 0 AND 0) OR (sex IN ('Male', 'Female') OR race <> 'White') OR (workclass = 'Private' OR income LIKE '<=50K') OR (` native.country` = 'United-States' OR fnlwgt > 272950) GROUP BY race,` native.country`,sex,relationship,income,occupation,` marital.status`,workclass

Resulted in 6492 records

SQL for Synthetic:
SELECT race,` native.country`,sex,relationship,income,occupation,` marital.status`,workclass,SUM(` hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (` hours.per.week` >= 40) OR ((age BETWEEN 45 AND 64) OR education IN ('1st-4th', 'Preschool', '5th-6th', 'Some-college', '11th', '10th')) OR (` marital.status` = 'Never-married' OR capital BETWEEN 0 AND 0) OR (sex IN ('Male', 'Female') OR race <> 'White') OR (workclass = 'Private' OR income LIKE '<=50K') OR (` native.country` = 'United-States' OR fnlwgt > 272950) GROUP BY race,` native.country`,sex,relationship,income,occupation,` marital.status`,workclass

Resulted in 8900 records

Normalized Euclidean distance for (` hours.per.week`): 49.69

Hellinger Distance: 0.303

=====

Real			
	education	SUM(fnlwgt)	COUNT(*)
8	Assoc-voc	369786888	2061
12	Masters	482192559	2657
7	Assoc-acdm	310113821	1601
10	Doctorate	109349744	594
9	Bachelors	1511601409	8025

Synthetic			
	education	SUM(fnlwgt)	COUNT(*)
8	Assoc-voc	348829865.375829	1935.000000
12	Masters	497669788.004103	2810.000000
7	Assoc-acdm	319258078.054042	1755.000000
10	Doctorate	90887686.639099	526.000000
9	Bachelors	1479193991.288778	8164.000000

SQL for Real:
SELECT education,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (fnlwgt < 300812) AND (capital <> 0 AND age BETWEEN 49 AND 34) OR (occupation = 'Exec-managerial' OR race = 'White') OR (workclass <> 'Private' OR income = '<=50K') OR (` hours.per.week` >= 40 OR education <> '10th') GROUP BY education

Resulted in 16 records

SQL for Synthetic:
SELECT education,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt < 300812) AND (capital <> 0 AND age BETWEEN 49 AND 34) OR (occupation = 'Exec-managerial' OR race = 'White') OR (workclass <> 'Private' OR income = '<=50K') OR (` hours.per.week` >= 40 OR education <> '10th') GROUP BY education

Resulted in 16 records

Normalized Euclidean distance for (fnlwgt): 4.0

Hellinger Distance: 0.026

=====

Real									
	workclass	relationship	occupation	sex	income	marital.status	education	native.country	SUM(fnlwgt) COUNT(*)
24	Private	Unmarried	Handlers-cleaners	Female	<=50K	Separated	10th	United-States	269784 1
11	Private	Not-in-family	Machine-op-inspct	Female	<=50K	Separated	HS-grad	United-States	281138 1
19	Private	Unmarried	Adm-clerical	Female	<=50K	Separated	10th	United-States	280278 1
16	Private	Own-child	Machine-op-inspct	Female	<=50K	Separated	HS-grad	United-States	257500 1
13	Private	Not-in-family	Other-service	Female	<=50K	Separated	HS-grad	United-States	800168 3

Synthetic									
	workclass	relationship	occupation	sex	income	marital.status	education	native.country	SUM(fnlwgt) COUNT(*)
24	Private	Unmarried	Handlers-cleaners	Female	<=50K	Separated	10th	United-States	nan nan
11	Private	Not-in-family	Machine-op-inspct	Female	<=50K	Separated	HS-grad	United-States	nan nan
19	Private	Unmarried	Adm-clerical	Female	<=50K	Separated	10th	United-States	nan nan
16	Private	Own-child	Machine-op-inspct	Female	<=50K	Separated	HS-grad	United-States	nan nan
13	Private	Not-in-family	Other-service	Female	<=50K	Separated	HS-grad	United-States	nan nan

SOL for Real:
SELECT workclass,relationship,occupation,sex,income,`marital.status`,education,`native.country`,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (fnlwgt BETWEEN 254478 AND 281792) AND (`marital.status` LIKE 'Separated' AND sex = 'Female') GROUP BY workclass,relationship,occupation,sex,income,`marital.status`,education,`native.country`
Resulted in 38 records

SOL for Synthetic:
SELECT workclass,relationship,occupation,sex,income,`marital.status`,education,`native.country`,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt BETWEEN 254478 AND 281792) AND (`marital.status` LIKE 'Separated' AND sex = 'Female') GROUP BY workclass,relationship,occupation,sex,income,`marital.status`,education,`native.country`
Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

=====

Real					
	relationship	native.country	occupation	race	AVG(capital) COUNT(*)
1405	Other-relative	Puerto-Rico	Transport-moving	White	nan nan
1910	Own-child	Thailand	Farming-fishing	White	nan nan
769	Not-in-family	Hungary	Handlers-cleaners	White	nan nan
2645	Wife	Peru	Prof-specialty	White	0.000000 1.000000
2379	Unmarried	Scotland	Prof-specialty	White	nan nan

Synthetic					
	relationship	native.country	occupation	race	AVG(capital) COUNT(*)
1405	Other-relative	Puerto-Rico	Transport-moving	White	29.181273 3
1910	Own-child	Thailand	Farming-fishing	White	-30.281394 5
769	Not-in-family	Hungary	Handlers-cleaners	White	28.332768 9
2645	Wife	Peru	Prof-specialty	White	17.230924 7
2379	Unmarried	Scotland	Prof-specialty	White	-29.324200 7

SOL for Real:
SELECT relationship,`native.country`,occupation,race,AVG(capital), COUNT(*) FROM C1 WHERE (race IN ('Amer-Indian-Eskimo', 'Black', 'White') OR capital BETWEEN 0 AND 0) OR (sex LIKE 'Male' OR workclass LIKE 'Private') OR (income IN ('<=50K', '>50K') OR `marital.status` LIKE 'Married-civ-spouse') GROUP BY relationship,`native.country`,occupation,race
Resulted in 2127 records

SOL for Synthetic:
SELECT relationship,`native.country`,occupation,race,AVG(capital), COUNT(*) FROM C1_syn_06 WHERE (race IN ('Amer-Indian-Eskimo', 'Black', 'White') OR capital BETWEEN 0 AND 0) OR (sex LIKE 'Male' OR workclass LIKE 'Private') OR (income IN ('<=50K', '>50K') OR `marital.status` LIKE 'Married-civ-spouse') GROUP BY relationship,`native.country`,occupation,race
Resulted in 2763 records

Normalized Euclidean distance for (capital): 32.37

Hellinger Distance: 0.357

=====

Real					
	education	marital.status	race	relationship	SUM(`hours.per.week`) COUNT(*)
813	Masters	Married-civ-spouse	Other	Husband	175 4
21	10th	Married-civ-spouse	Other	Wife	48 1
914	Prof-school	Separated	White	Unmarried	128 3
709	HS-grad	Married-civ-spouse	Asian-Pac-Islander	Other-relative	424 10
109	11th	Never-married	Asian-Pac-Islander	Not-in-family	135 3

Synthetic					
	education	marital.status	race	relationship	SUM(`hours.per.week`) COUNT(*)
813	Masters	Married-civ-spouse	Other	Husband	40.046017 1.000000
21	10th	Married-civ-spouse	Other	Wife	nan nan
914	Prof-school	Separated	White	Unmarried	80.029681 2.000000
709	HS-grad	Married-civ-spouse	Asian-Pac-Islander	Other-relative	119.942627 3.000000
109	11th	Never-married	Asian-Pac-Islander	Not-in-family	119.953567 3.000000

SOL for Real:
SELECT education,`marital.status`,race,relationship,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (fnlwgt > 341368) AND (capital = 0 AND workclass LIKE 'Private') AND (`native.country` = 'United-States' AND relationship <> 'Husband') OR ((age BETWEEN 28 AND 46) OR race IN ('Amer-Indian-Eskimo', 'Asian-Pac-Islander')) OR (`marital.status` IN ('Separated', 'Divorced', 'Married-civ-spouse') OR income <> '>50K') GROUP BY education,`marital.status`,race,relationship
Resulted in 1023 records

SOL for Synthetic:
SELECT education,`marital.status`,race,relationship,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt > 341368) AND (capital = 0 AND workclass LIKE 'Private') AND (`native.country` = 'United-States' AND relationship <> 'Husband') OR ((age BETWEEN 28 AND 46) OR race IN ('Amer-Indian-Eskimo', 'Asian-Pac-Islander')) OR (`marital.status` IN ('Separated', 'Divorced', 'Married-civ-spouse') OR income <> '>50K') GROUP BY education,`marital.status`,race,relationship
Resulted in 964 records

Normalized Euclidean distance for (`hours.per.week`): 27.5

Hellinger Distance: 0.074

=====

Real			
	relationship	SUM(`hours.per.week`)	COUNT(*)
5	Wife	74073.000000	2023.000000
2	Other-relative	4665.000000	123.000000
0	Husband	784707.000000	17721.000000
1	Not-in-family	601.000000	12.000000
3	Own-child	4628.000000	124.000000

Synthetic			
	relationship	SUM(`hours.per.week`)	COUNT(*)
5	Wife	50886.984470	1273
2	Other-relative	3431.743160	86
0	Husband	409396.068011	10213
1	Not-in-family	1279.524913	32
3	Own-child	4550.749863	114

SQL for Real:
SELECT relationship,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (`native.country` LIKE 'United-States' AND `marital.status` = 'Married-civ-spouse') GROUP BY relationship

Resulted in 5 records

SQL for Synthetic:
SELECT relationship,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`native.country` LIKE 'United-States' AND `marital.status` = 'Married-civ-spouse') GROUP BY relationship

Resulted in 6 records

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

Hellinger Distance: 0.027

=====

Real								
	race	income	education	native.country	occupation	sex	SUM(`hours.per.week`)	COUNT(*)
249	Asian-Pac-Islander	<=50K	Assoc-acdm	United-States	Machine-op-inspct	Male	40	1
13	Amer-Indian-Eskimo	<=50K	11th	United-States	Craft-repair	Male	40	1
2750	White	<=50K	HS-grad	United-States	Prof-specialty	Male	3940	97
2897	White	<=50K	Some-college	?	Sales	Female	80	2
1790	White	<=50K	12th	?	Transport-moving	Male	10	1

Synthetic								
	race	income	education	native.country	occupation	sex	SUM(`hours.per.week`)	COUNT(*)
249	Asian-Pac-Islander	<=50K	Assoc-acdm	United-States	Machine-op-inspct	Male	nan	nan
13	Amer-Indian-Eskimo	<=50K	11th	United-States	Craft-repair	Male	nan	nan
2750	White	<=50K	HS-grad	United-States	Prof-specialty	Male	120.073569	3.000000
2897	White	<=50K	Some-college	?	Sales	Female	nan	nan
1790	White	<=50K	12th	?	Transport-moving	Male	nan	nan

SQL for Real:
SELECT race,income,education,`native.country`,occupation,sex,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (workclass = 'State-gov' AND `native.country` IN ('Scotland', 'Iran', 'Honduras', 'Hungary', 'India', 'England', 'Mexico', 'Taiwan', 'Poland', 'United-States', 'Peru', 'China', 'Cuba', 'Holand-Netherlands', 'Dominican-Republic', 'Hong', 'South', 'Portugal', 'France', 'Trinidad&Tobago', 'Cambodia', 'Nicaragua', 'Japan', 'Greece', '?', 'Ecuador', 'Puerto-Rico', 'Haiti', 'Vietnam', 'Thailand', 'Canada', 'Columbia', 'Laos', 'El-Salvador')) OR (age > 42 OR occupation <> 'Exec-managerial') AND (`marital.status` <> 'Divorced' AND capital BETWEEN 0 AND 0) GROUP BY race,income,education,`native.country`,occupation,sex

Resulted in 3583 records

SQL for Synthetic:
SELECT race,income,education,`native.country`,occupation,sex,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (workclass = 'State-gov' AND `native.country` IN ('Scotland', 'Iran', 'Honduras', 'Hungary', 'India', 'England', 'Mexico', 'Taiwan', 'Poland', 'United-States', 'Peru', 'China', 'Cuba', 'Holand-Netherlands', 'Dominican-Republic', 'Hong', 'South', 'Portugal', 'France', 'Trinidad&Tobago', 'Cambodia', 'Nicaragua', 'Japan', 'Greece', '?', 'Ecuador', 'Puerto-Rico', 'Haiti', 'Vietnam', 'Thailand', 'Canada', 'Columbia', 'Laos', 'El-Salvador')) OR (age > 42 OR occupation <> 'Exec-managerial') AND (`marital.status` <> 'Divorced' AND capital BETWEEN 0 AND 0) GROUP BY race,income,education,`native.country`,occupation,sex

Resulted in 672 records

Normalized Euclidean distance for (`hours.per.week`): 19.1

Hellinger Distance: 0.437

=====

Real				
	race	native.country	MAX(age)	COUNT(*)
55	Black	Holand-Netherlands	nan	nan
14	Amer-Indian-Eskimo	Scotland	nan	nan
83	Other	Guatemala	40.000000	8.000000
59	Black	Ireland	nan	nan
72	Black	Thailand	25.000000	1.000000

		Synthetic		
	race	native.country	MAX(age)	COUNT(*)
55	Black	Holand-Netherlands	26	23
14	Amer-Indian-Eskimo	Scotland	34	1
83	Other	Guatemala	34	2
59	Black	Ireland	31	41
72	Black	Thailand	20	20

SQL for Real:
SELECT race,`native.country`,MAX(age), COUNT(*) FROM C1 WHERE (capital = 0 AND race <> 'White') AND (occupation <> 'Sales' AND relationship IN ('Wife', 'Other-relative', 'Own-child', 'Husband')) AND (age < 19 AND education LIKE 'HS-grad') OR (income IN ('>50K', '<=50K') OR `marital.status` = 'Married-civ-spouse') AND ('native.country' <> 'United-States' AND sex IN ('Female', 'Male')) GROUP BY race,`native.country`

Resulted in 127 records

SQL for Synthetic:
SELECT race,`native.country`,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (capital = 0 AND race <> 'White') AND (occupation <> 'Sales' AND relationship IN ('Wife', 'Other-relative', 'Own-child', 'Husband')) AND (age < 19 AND education LIKE 'HS-grad') OR (income IN ('>50K', '<=50K') OR `marital.status` = 'Married-civ-spouse') AND ('native.country' <> 'United-States' AND sex IN ('Female', 'Male')) GROUP BY race,`native.country`

Resulted in 139 records

Normalized Euclidean distance for (age): 9.8

Hellinger Distance: 0.664

=====

			Real				
	income	native.country	workclass	marital.status	race	MIN(fnlwgt)	COUNT(*)
1199	>50K	United-States	Private	Divorced	Asian-Pac-Islander	85815	1
489	<=50K	Mexico	?	Married-civ-spouse	Other	291547	2
1075	>50K	Jamaica	Local-gov	Married-civ-spouse	Black	145178	1
421	<=50K	Ireland	Private	Never-married	Asian-Pac-Islander	194951	1
943	>50K	?	Private	Never-married	Black	143776	1

			Synthetic				
	income	native.country	workclass	marital.status	race	MIN(fnlwgt)	COUNT(*)
1199	>50K	United-States	Private	Divorced	Asian-Pac-Islander	nan	nan
489	<=50K	Mexico	?	Married-civ-spouse	Other	nan	nan
1075	>50K	Jamaica	Local-gov	Married-civ-spouse	Black	nan	nan
421	<=50K	Ireland	Private	Never-married	Asian-Pac-Islander	nan	nan
943	>50K	?	Private	Never-married	Black	nan	nan

SQL for Real:
SELECT income,`native.country`,workclass,`marital.status`,race,MIN(fnlwgt), COUNT(*) FROM C1 WHERE (capital BETWEEN 0 AND 0) OR (age <> 41 OR income IN ('>50K', '<=50K')) AND (`marital.status` IN ('Married-spouse-absent', 'Separated', 'Married-civ-spouse', 'Married-AF-spouse', 'Widowed') AND sex = 'Female') AND (race <> 'White' AND relationship <> 'Husband') AND (fnlwgt = 274398 AND `native.country` IN ('England', 'Peru')) AND (`hours.per.week` = 24 AND education = 'HS-grad') GROUP BY income,`native.country`,workclass,`marital.status`,race

Resulted in 1264 records

SQL for Synthetic:
SELECT income,`native.country`,workclass,`marital.status`,race,MIN(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (capital BETWEEN 0 AND 0) OR (age <> 41 OR income IN ('>50K', '<=50K')) AND (`marital.status` IN ('Married-spouse-absent', 'Separated', 'Married-civ-spouse', 'Married-AF-spouse', 'Widowed') AND sex = 'Female') AND (race <> 'White' AND relationship <> 'Husband') AND (fnlwgt = 274398 AND `native.country` IN ('England', 'Peru')) AND (`hours.per.week` = 24 AND education = 'HS-grad') GROUP BY income,`native.country`,workclass,`marital.status`,race

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

=====

	Real
education	AVG(age) COUNT(*)

	Synthetic
education	AVG(age) COUNT(*)

SQL for Real:
SELECT education,AVG(age), COUNT(*) FROM C1 WHERE ((age BETWEEN 54 AND 34) AND `native.country` <> 'United-States') AND (`marital.status` LIKE 'Married-civ-spouse' AND `hours.per.week` > 40) AND (relationship = 'Husband' AND race IN ('Asian-Pac-Islander', 'White', 'Black', 'Other', 'Amer-Indian-Eskimo')) GROUP BY education

Resulted in 0 records

SQL for Synthetic:
SELECT education,AVG(age), COUNT(*) FROM C1_syn_06 WHERE ((age BETWEEN 54 AND 34) AND `native.country` <> 'United-States') AND (`marital.status` LIKE 'Married-civ-spouse' AND `hours.per.week` > 40) AND (relationship = 'Husband' AND race IN ('Asian-Pac-Islander', 'White', 'Black', 'Other', 'Amer-Indian-Eskimo')) GROUP BY education

Resulted in 0 records

Normalized Euclidean distance for (age): nan

Hellinger Distance: nan

=====

Real								
	marital.status	education	occupation	relationship	race	income	AVG(fnlwgt)	COUNT(*)
2227	Separated	HS-grad	Protective-serv	Other-relative	Black	<=50K	171328.000000	1
1285	Never-married	10th	?	Own-child	Black	<=50K	217778.000000	9
1828	Never-married	HS-grad	Protective-serv	Other-relative	Black	<=50K	191376.500000	2
124	Divorced	Assoc-voc	Exec-managerial	Own-child	Amer-Indian-Eskimo	<=50K	134220.000000	1
74	Divorced	7th-8th	Other-service	Unmarried	Black	<=50K	334141.000000	1

Synthetic								
	marital.status	education	occupation	relationship	race	income	AVG(fnlwgt)	COUNT(*)
2227	Separated	HS-grad	Protective-serv	Other-relative	Black	<=50K	193690.626820	2.000000
1285	Never-married	10th	?	Own-child	Black	<=50K	169161.656152	8.000000
1828	Never-married	HS-grad	Protective-serv	Other-relative	Black	<=50K	182703.491313	1.000000
124	Divorced	Assoc-voc	Exec-managerial	Own-child	Amer-Indian-Eskimo	<=50K	nan	nan
74	Divorced	7th-8th	Other-service	Unmarried	Black	<=50K	185619.474668	1.000000

SQL for Real:
SELECT `marital.status`,education,occupation,relationship,race,income,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (race <> 'White' AND age <> 25) GROUP BY `marital.status`,education,occupation,relationship,race,income

Resulted in 2470 records

SQL for Synthetic:
SELECT `marital.status`,education,occupation,relationship,race,income,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (race <> 'White' AND age <> 25) GROUP BY `marital.status`,education,occupation,relationship,race,income

Resulted in 1896 records

Normalized Euclidean distance for (fnlwgt): 32.97

Hellinger Distance: 0.22

=====

Real								
	native.country	race	relationship	education	income	sex	AVG(age)	COUNT(*)
2912	United-States	Black	Not-in-family	Bachelors	<=50K	Male	35.333333	48.000000
1428	Japan	White	Not-in-family	Doctorate	>50K	Male	nan	nan
1362	Japan	Black	Own-child	12th	<=50K	Male	nan	nan
2171	Puerto-Rico	Black	Own-child	11th	<=50K	Female	nan	nan
3081	United-States	Other	Unmarried	HS-grad	<=50K	Male	22.000000	1.000000

Synthetic								
	native.country	race	relationship	education	income	sex	AVG(age)	COUNT(*)
2912	United-States	Black	Not-in-family	Bachelors	<=50K	Male	35.454545	22
1428	Japan	White	Not-in-family	Doctorate	>50K	Male	34.000000	1
1362	Japan	Black	Own-child	12th	<=50K	Male	42.000000	1
2171	Puerto-Rico	Black	Own-child	11th	<=50K	Female	18.000000	4
3081	United-States	Other	Unmarried	HS-grad	<=50K	Male	27.500000	4

SQL for Real:
SELECT `native.country`,race,relationship,education,income,sex,AVG(age), COUNT(*) FROM C1 WHERE (relationship <> 'Not-in-family') OR (occupation IN ('Protective-serv', 'Prof-specialty', 'Farming-fishing', '?', 'Machine-op-inspct', 'Other-service') AND race = 'White') OR (age = 32 OR capital <= 0) GROUP BY `native.country`,race,relationship,education,income,sex

Resulted in 2887 records

SQL for Synthetic:
SELECT `native.country`,race,relationship,education,income,sex,AVG(age), COUNT(*) FROM C1_syn_06 WHERE (relationship <> 'Not-in-family') OR (occupation IN ('Protective-serv', 'Prof-specialty', 'Farming-fishing', '?', 'Machine-op-inspct', 'Other-service') AND race = 'White') OR (age = 32 OR capital <= 0) GROUP BY `native.country`,race,relationship,education,income,sex

Resulted in 3396 records

Normalized Euclidean distance for (age): 34.25

Hellinger Distance: 0.364

=====

Real							
	education	occupation	race	marital.status	workclass	SUM(fnlwgt)	COUNT(*)
2529	Prof-school	Craft-repair	White	Married-civ-spouse	Self-emp-not-inc	422660	2
1852	HS-grad	Exec-managerial	Other	Married-civ-spouse	Local-gov	34640	1
1235	Bachelors	Adm-clerical	White	Married-civ-spouse	Local-gov	3320782	12
2210	HS-grad	Tech-support	White	Married-civ-spouse	State-gov	1247867	8
2165	HS-grad	Sales	Black	Divorced	Private	649321	2

Synthetic							
	education	occupation	race	marital.status	workclass	SUM(fnlwgt)	COUNT(*)
2529	Prof-school	Craft-repair	White	Married-civ-spouse	Self-emp-not-inc	1018890.224140	6.000000
1852	HS-grad	Exec-managerial	Other	Married-civ-spouse	Local-gov	nan	nan
1235	Bachelors	Adm-clerical	White	Married-civ-spouse	Local-gov	1267166.532916	7.000000
2210	HS-grad	Tech-support	White	Married-civ-spouse	State-gov	1342552.186451	8.000000
2165	HS-grad	Sales	Black	Divorced	Private	715468.095330	4.000000

SQL for Real:
SELECT education,occupation,race,`marital.status`,workclass,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (sex <> 'Female' AND education <> '11th') GROUP BY education,occupation,race,`marital.status`,workclass

Resulted in 3129 records

SQL for Synthetic:

SELECT education,occupation,race,`marital.status`,workclass,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (sex <> 'Female' AND education <> '11th') GROUP BY education,occupation,race,`marital.status`,workclass

Resulted in 2927 records

Normalized Euclidean distance for (fnlwgt): 43.69

Hellinger Distance: 0.113

=====

Real										
	marital.status	occupation	relationship	native.country	education	sex	race	workclass	MIN(capital)	COUNT(*)
1884	Widowed	Other-service	Unmarried	Puerto-Rico	HS-grad	Female	Black	State-gov	nan	nan
455	Married-civ-spouse	Other-service	Husband	Peru	9th	Male	White	Local-gov	nan	nan
55	Divorced	Other-service	Not-in-family	Puerto-Rico	5th-6th	Female	White	Private	nan	nan
606	Married-civ-spouse	Other-service	Husband	United-States	Masters	Male	White	Private	0.000000	7.000000
10	Divorced	Other-service	Not-in-family	?	Some-college	Female	White	Self-emp-not-inc	nan	nan

Synthetic										
	marital.status	occupation	relationship	native.country	education	sex	race	workclass	MIN(capital)	COUNT(*)
1884	Widowed	Other-service	Unmarried	Puerto-Rico	HS-grad	Female	Black	State-gov	-53.357951	1
455	Married-civ-spouse	Other-service	Husband	Peru	9th	Male	White	Local-gov	-60.426615	2
55	Divorced	Other-service	Not-in-family	Puerto-Rico	5th-6th	Female	White	Private	-66.990580	1
606	Married-civ-spouse	Other-service	Husband	United-States	Masters	Male	White	Private	-54.065722	2
10	Divorced	Other-service	Not-in-family	?	Some-college	Female	White	Self-emp-not-inc	-167.328720	1

SQL for Real:
SELECT `marital.status`,occupation,relationship,`native.country`,education,sex,race,workclass,MIN(capital), COUNT(*) FROM C1 WHERE (occupation LIKE 'Other-service') AND (`native.country` IN ('Outlying-US(Guam-USVI-etc)', 'Scotland', 'Germany', 'Guatemala', 'Hungary', 'Italy') OR workclass IN ('Self-emp-not-inc', 'Local-gov', 'State-gov', 'Self-emp-inc', 'Private', '?', 'Federal-gov')) GROUP BY `marital.status`,occupation,relationship,`native.country`,education,sex,race,workclass

Resulted in 1469 records

SQL for Synthetic:
SELECT `marital.status`,occupation,relationship,`native.country`,education,sex,race,workclass,MIN(capital), COUNT(*) FROM C1_syn_06 WHERE (occupation LIKE 'Other-service') AND (`native.country` IN ('Outlying-US(Guam-USVI-etc)', 'Scotland', 'Germany', 'Guatemala', 'Hungary', 'Italy') OR workclass IN ('Self-emp-not-inc', 'Local-gov', 'State-gov', 'Self-emp-inc', 'Private', '?', 'Federal-gov')) GROUP BY `marital.status`,occupation,relationship,`native.country`,education,sex,race,workclass

Resulted in 1917 records

Normalized Euclidean distance for (capital): 21.84

Hellinger Distance: 0.213

=====

Real												
	relationship	education	race	occupation	income	sex	marital.status	native.country	workclass	MIN(`hours.per.week`)	COUNT(*)	
2072	Other-relative	HS-grad	White	Sales	<=50K	Female	Married-spouse-absent	Hungary	Private	nan	nan	
2230	Own-child	1st-4th	White	Prof-specialty	<=50K	Female	Separated	India	Private	nan	nan	
808	Not-in-family	Bachelors	White	Prof-specialty	>50K	Male	Married-civ-spouse	United-States	Private	nan	nan	
2175	Own-child	10th	White	Handlers-cleaners	<=50K	Female	Separated	United-States	Private	nan	nan	
2288	Own-child	Assoc-voc	White	Adm-clerical	<=50K	Male	Separated	?	Private	nan	nan	

Synthetic												
	relationship	education	race	occupation	income	sex	marital.status	native.country	workclass	MIN(`hours.per.week`)	COUNT(*)	
2072	Other-relative	HS-grad	White	Sales	<=50K	Female	Married-spouse-absent	Hungary	Private	39.973932	1	
2230	Own-child	1st-4th	White	Prof-specialty	<=50K	Female	Separated	India	Private	39.961680	1	
808	Not-in-family	Bachelors	White	Prof-specialty	>50K	Male	Married-civ-spouse	United-States	Private	40.257286	1	
2175	Own-child	10th	White	Handlers-cleaners	<=50K	Female	Separated	United-States	Private	40.047494	1	
2288	Own-child	Assoc-voc	White	Adm-clerical	<=50K	Male	Separated	?	Private	40.063989	1	

SQL for Real:
SELECT relationship,education,race,occupation,income,sex,`marital.status`,`native.country`,workclass,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (relationship IN ('Own-child', 'Not-in-family', 'Other-relative')) AND (workclass LIKE 'Private' OR sex = 'Male') AND (race LIKE 'White' AND `marital.status` <> 'Never-married') GROUP BY relationship,education,race,occupation,income,sex,`marital.status`,`native.country`,workclass

Resulted in 1814 records

SQL for Synthetic:
SELECT relationship,education,race,occupation,income,sex,`marital.status`,`native.country`,workclass,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (relationship IN ('Own-child', 'Not-in-family', 'Other-relative')) AND (workclass LIKE 'Private' OR sex = 'Male') AND (race LIKE 'White' AND `marital.status` <> 'Never-married') GROUP BY relationship,education,race,occupation,income,sex,`marital.status`,`native.country`,workclass

Resulted in 2670 records

Normalized Euclidean distance for (`hours.per.week`): 24.86

Hellinger Distance: 0.203

=====

Real							
	native.country	sex	race	relationship	workclass	AVG(capital)	COUNT(*)
44	?	Female	White	Wife	State-gov	7688.000000	1.000000
1164	United-States	Male	Other	Husband	Local-gov	-943.500000	2.000000
932	Scotland	Male	White	Husband	Self-emp-inc	nan	nan
381	Hungary	Male	White	Wife	Private	nan	nan
1218	Vietnam	Male	White	Own-child	Private	nan	nan

Synthetic							
	native.country	sex	race	relationship	workclass	AVG(capital)	COUNT(*)
44	?	Female	White	Wife	State-gov	34.522002	4
1164	United-States	Male	Other	Husband	Local-gov	77.434212	1
932	Scotland	Male	White	Husband	Self-emp-inc	200.559628	4
381	Hungary	Male	White	Wife	Private	-39.870357	1
1218	Vietnam	Male	White	Own-child	Private	103.008154	1

SQL for Real:
SELECT `native.country`,sex,race,relationship,workclass,AVG(capital), COUNT(*) FROM C1 WHERE (fnlwgt BETWEEN 20956 AND 114032) OR (race IN ('Black', 'Other', 'Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'White') AND `marital.status` = 'Married-civ-spouse') AND (occupation = 'Sales' AND workclass <> 'Private') OR (education IN ('5th-6th', 'Prof-school', '10th', '1st-4th', 'HS-grad', 'Preschool') OR `hours.per.week` >= 56) GROUP BY `native.country`,sex,race,relationship,workclass

Resulted in 1068 records

SQL for Synthetic:
SELECT `native.country`,sex,race,relationship,workclass,AVG(capital), COUNT(*) FROM C1_syn_06 WHERE (fnlwgt BETWEEN 20956 AND 114032) OR (race IN ('Black', 'Other', 'Asian-Pac-Islander', 'Amer-Indian-Eskimo', 'White') AND `marital.status` = 'Married-civ-spouse') AND (occupation = 'Sales' AND workclass <> 'Private') OR (education IN ('5th-6th', 'Prof-school', '10th', '1st-4th', 'HS-grad', 'Preschool') OR `hours.per.week` >= 56) GROUP BY `native.country`,sex,race,relationship,workclass

Resulted in 1237 records

Normalized Euclidean distance for (capital): 22.09

Hellinger Distance: 0.379

=====

Real											
	relationship	sex	occupation	native.country	marital.status	income	workclass	race	education	MIN(`hours.per.week`)	COUNT(*)
14556	Unmarried	Female	Sales	Japan	Never-married	<=50K	Private	White	HS-grad	nan	nan
3274	Husband	Male	Sales	Hungary	Married-civ-spouse	>50K	Private	White	HS-grad	nan	nan
12719	Own-child	Male	Transport-moving	Puerto-Rico	Never-married	<=50K	State-gov	White	HS-grad	nan	nan
2678	Husband	Male	Prof-specialty	Peru	Married-civ-spouse	<=50K	Federal-gov	White	Bachelors	nan	nan
2427	Husband	Male	Other-service	United-States	Married-civ-spouse	<=50K	Self-emp-inc	White	HS-grad	30.000000	4.000000

Synthetic											
	relationship	sex	occupation	native.country	marital.status	income	workclass	race	education	MIN(`hours.per.week`)	COUNT(*)
14556	Unmarried	Female	Sales	Japan	Never-married	<=50K	Private	White	HS-grad	40.070797	1
3274	Husband	Male	Sales	Hungary	Married-civ-spouse	>50K	Private	White	HS-grad	40.331280	3
12719	Own-child	Male	Transport-moving	Puerto-Rico	Never-married	<=50K	State-gov	White	HS-grad	40.029514	1
2678	Husband	Male	Prof-specialty	Peru	Married-civ-spouse	<=50K	Federal-gov	White	Bachelors	39.998705	1
2427	Husband	Male	Other-service	United-States	Married-civ-spouse	<=50K	Self-emp-inc	White	HS-grad	39.003745	4

SQL for Real:
SELECT relationship,sex,occupation,`native.country`,`marital.status`,income,workclass,race,education,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (income <> '>50K' OR age BETWEEN 35 AND 29) OR (fnlwgt <= 165695 OR `native.country` <> 'United-States') GROUP BY relationship,sex,occupation,`native.country`,`marital.status`,income,workclass,race,education

Resulted in 11964 records

SQL for Synthetic:
SELECT relationship,sex,occupation,`native.country`,`marital.status`,income,workclass,race,education,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (income <> '>50K' OR age BETWEEN 35 AND 29) OR (fnlwgt <= 165695 OR `native.country` <> 'United-States') GROUP BY relationship,sex,occupation,`native.country`,`marital.status`,income,workclass,race,education

Resulted in 16719 records

Normalized Euclidean distance for (`hours.per.week`): 62.51

Hellinger Distance: 0.263

=====

Real			
	education	SUM(age)	COUNT(*)
3	1st-4th	11468	246
6	9th	30804	754
2	12th	21179	657
8	Assoc-voc	78895	2042
1	11th	58023	1808

Synthetic			
	education	SUM(age)	COUNT(*)
3	1st-4th	7495.000000	154.000000
6	9th	30820.000000	737.000000
2	12th	20073.000000	625.000000
8	Assoc-voc	73863.000000	1935.000000
1	11th	55332.000000	1692.000000

SQL for Real:
SELECT education,SUM(age), COUNT(*) FROM C1 WHERE (income = '<=50K') OR (occupation <> 'Farming-fishing' OR `marital.status` = 'Divorced') OR (fnlwgt >= 249392 OR relationship LIKE 'Husband') AND (race <> 'White' AND `hours.per.week` BETWEEN 40 AND 40) AND (`native.country` IN ('Holand-Netherlands', 'Thailand', 'South', 'Columbia', 'Scotland', 'Hungary', 'Canada', 'Germany', 'El-Salvador', 'Cuba', 'Trinidad&Tobago', 'Portugal', 'Greece', 'Cambodia', 'Vietnam', 'Puerto-Rico', 'India', 'Taiwan', 'Philippines', 'Peru', 'Mexico', 'Japan', 'Guatemala', 'Jamaica', 'Nicaragua', 'Haiti', '?') AND age BETWEEN 55 AND 18) GROUP BY education

Resulted in 16 records

SQL for Synthetic:
SELECT education,SUM(age), COUNT(*) FROM C1_syn_06 WHERE (income = '<=50K') OR (occupation <> 'Farming-fishing' OR `marital.status` = 'Divorced') OR (fnlwgt >= 249392 OR relationship LIKE 'Husband') AND (race <> 'White' AND `hours.per.week` BETWEEN 40 AND 40) AND (`native.country` IN ('Holand-Netherlands', 'Thailand', 'South', 'Columbia', 'Scotland', 'Hungary', 'Canada', 'Germany', 'El-Salvador', 'Cuba', 'Trinidad&Tobago', 'Portugal', 'Greece', 'Cambodia', 'Vietnam', 'Puerto-Rico', 'India', 'Taiwan', 'Philippines', 'Peru', 'Mexico', 'Japan', 'Guatemala', 'Jamaica', 'Nicaragua', 'Haiti', '?') AND age BETWEEN 55 AND 18) GROUP BY education

Resulted in 16 records

Normalized Euclidean distance for (age): 4.0

Hellinger Distance: 0.026

=====

Real					
	relationship	workclass	sex	marital.status	SUM(fnlwgt) COUNT(*)
7	Own-child	Self-emp-not-inc	Male	Married-civ-spouse	nan nan
1	Husband	Local-gov	Male	Married-civ-spouse	4046346.000000 15.000000
3	Husband	Self-emp-inc	Male	Married-civ-spouse	nan nan
4	Husband	Self-emp-not-inc	Male	Married-civ-spouse	4318141.000000 20.000000
6	Own-child	Private	Male	Married-civ-spouse	467540.000000 3.000000

Synthetic					
	relationship	workclass	sex	marital.status	SUM(fnlwgt) COUNT(*)
7	Own-child	Self-emp-not-inc	Male	Married-civ-spouse	174533.395671 1
1	Husband	Local-gov	Male	Married-civ-spouse	2281328.676901 13
3	Husband	Self-emp-inc	Male	Married-civ-spouse	517915.415016 3
4	Husband	Self-emp-not-inc	Male	Married-civ-spouse	6343638.060560 31
6	Own-child	Private	Male	Married-civ-spouse	177290.046958 1

SQL for Real:
SELECT relationship,workclass,sex,`marital.status`,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (occupation = 'Craft-repair') AND (race <> 'White' OR age = 55) AND (relationship IN ('Wife', 'Unmarried', 'Not-in-family', 'Husband', 'Own-child') AND sex <> 'Female') AND (`marital.status` LIKE 'Married-civ-spouse' AND `hours.per.week` < 42) GROUP BY relationship,workclass,sex,`marital.status`

Resulted in 7 records

SQL for Synthetic:
SELECT relationship,workclass,sex,`marital.status`,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (occupation = 'Craft-repair') AND (race <> 'White' OR age = 55) AND (relationship IN ('Wife', 'Unmarried', 'Not-in-family', 'Husband', 'Own-child') AND sex <> 'Female') AND (`marital.status` LIKE 'Married-civ-spouse' AND `hours.per.week` < 42) GROUP BY relationship,workclass,sex,`marital.status`

Resulted in 8 records

Normalized Euclidean distance for (fnlwgt): 2.45

Hellinger Distance: 0.067

=====

Real					
	income	education	workclass	race	MAX(fnlwgt) COUNT(*)
40	<=50K	Assoc-voc	Self-emp-not-inc	Amer-Indian-Eskimo	108506 1
42	<=50K	Bachelors	?	Other	123983 1
38	<=50K	Assoc-voc	Private	Black	257017 3
2	<=50K	10th	Private	Amer-Indian-Eskimo	214134 3
22	<=50K	5th-6th	Private	Other	82061 1

Synthetic					
	income	education	workclass	race	MAX(fnlwgt) COUNT(*)
40	<=50K	Assoc-voc	Self-emp-not-inc	Amer-Indian-Eskimo	nan nan
42	<=50K	Bachelors	?	Other	nan nan
38	<=50K	Assoc-voc	Private	Black	194787.835228 1.000000
2	<=50K	10th	Private	Amer-Indian-Eskimo	202213.887474 2.000000
22	<=50K	5th-6th	Private	Other	nan nan

SQL for Real:
SELECT income,education,workclass,race,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (sex LIKE 'Male' AND education <> 'Some-college') AND (race <> 'White' AND age <= 23) GROUP BY income,education,workclass,race

Resulted in 64 records

SQL for Synthetic:
SELECT income,education,workclass,race,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (sex LIKE 'Male' AND education <> 'Some-college') AND (race <> 'White' AND age <= 23) GROUP BY income,education,workclass,race

Resulted in 56 records

Normalized Euclidean distance for (fnlwgt): 6.24

Hellinger Distance: 0.146

=====

Real		
relationship	SUM(fnlwgt)	COUNT(*)

Synthetic		
relationship	SUM(fnlwgt)	COUNT(*)

SQL for Real:
SELECT relationship,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (sex = 'Male' AND race IN ('Other', 'Asian-Pac-Islander', 'White', 'Amer-Indian-Eskimo')) AND (relationship = 'Own-child' AND age BETWEEN 36 AND 34) GROUP BY relationship

Resulted in 0 records

SQL for Synthetic:
SELECT relationship,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (sex = 'Male' AND race IN ('Other', 'Asian-Pac-Islander', 'White', 'Amer-Indian-Eskimo')) AND (relationship = 'Own-child' AND age BETWEEN 36 AND 34) GROUP BY relationship

Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

=====

Real				
	income	sex	MAX(fnlwgt)	COUNT(*)
0	<=50K	Male	206008	1

Synthetic				
	income	sex	MAX(fnlwgt)	COUNT(*)
0	<=50K	Male	nan	nan

SQL for Real:
SELECT income,sex,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (`hours.per.week` = 40 AND relationship LIKE 'Unmarried') AND (age = 18 AND workclass = 'Private') AND (education = 'Some-college' AND race IN ('Other', 'Amer-Indian-Eskimo', 'Black', 'White', 'Asian-Pac-Islander')) GROUP BY income,sex
Resulted in 1 records

SQL for Synthetic:
SELECT income,sex,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (`hours.per.week` = 40 AND relationship LIKE 'Unmarried') AND (age = 18 AND workclass = 'Private') AND (education = 'Some-college' AND race IN ('Other', 'Amer-Indian-Eskimo', 'Black', 'White', 'Asian-Pac-Islander')) GROUP BY income,sex
Resulted in 0 records

Normalized Euclidean distance for (fnlwgt): nan

Hellinger Distance: nan

=====

Real				
	education	sex	AVG(fnlwgt)	COUNT(*)
5	12th	Male	204798.275785	446
16	Assoc-voc	Female	183048.084469	734
24	Masters	Female	175166.269822	845
19	Bachelors	Male	191192.134643	5548
29	Prof-school	Male	189146.198006	702

Synthetic				
	education	sex	AVG(fnlwgt)	COUNT(*)
5	12th	Male	185693.362432	423.000000
16	Assoc-voc	Female	179631.047107	738.000000
24	Masters	Female	179204.437299	954.000000
19	Bachelors	Male	180664.782286	5396.000000
29	Prof-school	Male	172999.442744	695.000000

SQL for Real:
SELECT education,sex,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (race = 'White') OR (age > 23 AND relationship = 'Wife') AND (workclass <> 'Private' AND `native.country` IN ('Puerto-Rico', 'Iran', 'Ireland', 'China', 'Peru', 'Philippines', 'India', 'Nicaragua', 'Dominican-Republic', 'Honduras', 'England', 'United-States', 'Taiwan', 'Thailand', 'Vietnam', 'South', 'Poland', 'Yugoslavia', 'Mexico', 'France', 'Hungary', 'Portugal', 'Outlying-US(Guam-USVI-etc)', 'Trinidad&Tobago', 'Canada', 'Scotland', 'Italy', 'Cambodia', 'El-Salvador', 'Haiti', 'Cuba', 'Germany', 'Greece')) AND (`hours.per.week` = 50 AND fnlwgt BETWEEN 98076 AND 134808) AND (income LIKE '<=50K' AND education = 'Some-college') OR (`marital.status` = 'Never-married' OR sex IN ('Male', 'Female')) GROUP BY education,sex
Resulted in 32 records

SQL for Synthetic:
SELECT education,sex,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (race = 'White') OR (age > 23 AND relationship = 'Wife') AND (workclass <> 'Private' AND `native.country` IN ('Puerto-Rico', 'Iran', 'Ireland', 'China', 'Peru', 'Philippines', 'India', 'Nicaragua', 'Dominican-Republic', 'Honduras', 'England', 'United-States', 'Taiwan', 'Thailand', 'Vietnam', 'South', 'Poland', 'Yugoslavia', 'Mexico', 'France', 'Hungary', 'Portugal', 'Outlying-US(Guam-USVI-etc)', 'Trinidad&Tobago', 'Canada', 'Scotland', 'Italy', 'Cambodia', 'El-Salvador', 'Haiti', 'Cuba', 'Germany', 'Greece')) AND (`hours.per.week` = 50 AND fnlwgt BETWEEN 98076 AND 134808) AND (income LIKE '<=50K' AND education = 'Some-college') OR (`marital.status` = 'Never-married' OR sex IN ('Male', 'Female')) GROUP BY education,sex
Resulted in 32 records

Normalized Euclidean distance for (fnlwgt): 5.66

Hellinger Distance: 0.031

=====

Real						
	relationship	education	sex	income	SUM(`hours.per.week`)	COUNT(*)
63	Wife	10th	Female	<=50K	255	7
84	Wife	Some-college	Female	>50K	310	11
19	Not-in-family	Some-college	Female	<=50K	2282	83
8	Not-in-family	Assoc-acdm	Female	>50K	40	1
69	Wife	9th	Female	<=50K	147	4

Synthetic						
	relationship	education	sex	income	SUM(`hours.per.week`)	COUNT(*)
63	Wife	10th	Female	<=50K	239.931979	6.000000
84	Wife	Some-college	Female	>50K	39.058910	1.000000
19	Not-in-family	Some-college	Female	<=50K	3449.134586	87.000000
8	Not-in-family	Assoc-acdm	Female	>50K	nan	nan
69	Wife	9th	Female	<=50K	200.169854	5.000000

SQL for Real:
SELECT relationship,education,sex,income,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (occupation = '?' AND sex <> 'Male') GROUP BY relationship,education,sex,income
Resulted in 85 records

SQL for Synthetic:
SELECT relationship,education,sex,income,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (occupation = '?' AND sex <> 'Male') GROUP BY relationship,education,sex,income

Resulted in 76 records

Normalized Euclidean distance for (`hours.per.week`): 8.19

Hellinger Distance: 0.12

=====

Real											
	education	marital.status	race	income	relationship	workclass	native.country	occupation	sex	MAX(fnlwgt)	COUNT(*)
1975	Bachelors	Married-civ-spouse	White	<=50K	Husband	Private	Ecuador	Prof-specialty	Male	nan	nan
580	9th	Married-civ-spouse	White	>50K	Wife	Local-gov	Puerto-Rico	Other-service	Female	nan	nan
5012	Masters	Married-civ-spouse	White	>50K	Husband	Private	Scotland	Prof-specialty	Male	nan	nan
1388	Bachelors	Divorced	White	<=50K	Not-in-family	Self-emp-not-inc	Puerto-Rico	Craft-repair	Male	nan	nan
3959	HS-grad	Divorced	White	>50K	Unmarried	Private	Japan	Exec-managerial	Male	nan	nan

Synthetic											
	education	marital.status	race	income	relationship	workclass	native.country	occupation	sex	MAX(fnlwgt)	COUNT(*)
1975	Bachelors	Married-civ-spouse	White	<=50K	Husband	Private	Ecuador	Prof-specialty	Male	182895.681429	2
580	9th	Married-civ-spouse	White	>50K	Wife	Local-gov	Puerto-Rico	Other-service	Female	169156.838262	1
5012	Masters	Married-civ-spouse	White	>50K	Husband	Private	Scotland	Prof-specialty	Male	180775.197816	1
1388	Bachelors	Divorced	White	<=50K	Not-in-family	Self-emp-not-inc	Puerto-Rico	Craft-repair	Male	170458.898725	3
3959	HS-grad	Divorced	White	>50K	Unmarried	Private	Japan	Exec-managerial	Male	178621.793945	1

SQL for Real:
SELECT education, `marital.status`,race,income,relationship,workclass, `native.country`,occupation,sex,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (income <> '<=50K') OR (education = 'Bachelors' OR relationship LIKE 'Own-child') GROUP BY education, `marital.status`,race,income,relationship,workclass, `native.country`,occupation,sex

Resulted in 5402 records

SQL for Synthetic:
SELECT education, `marital.status`,race,income,relationship,workclass, `native.country`,occupation,sex,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (income <> '<=50K') OR (education = 'Bachelors' OR relationship LIKE 'Own-child') GROUP BY education, `marital.status`,race,income,relationship,workclass, `native.country`,occupation,sex

Resulted in 6215 records

Normalized Euclidean distance for (fnlwgt): 40.27

Hellinger Distance: 0.245

=====

Real						
	native.country	income	relationship	race	SUM(`hours.per.week`)	COUNT(*)
510	Puerto-Rico	<=50K	Husband	Black	145.000000	4.000000
190	Guatemala	<=50K	Husband	Black	nan	nan
594	Taiwan	<=50K	Wife	White	nan	nan
229	Holand-Netherlands	<=50K	Own-child	White	nan	nan
157	El-Salvador	>50K	Wife	White	40.000000	1.000000

Synthetic						
	native.country	income	relationship	race	SUM(`hours.per.week`)	COUNT(*)
510	Puerto-Rico	<=50K	Husband	Black	7603.683892	190
190	Guatemala	<=50K	Husband	Black	79.927204	2
594	Taiwan	<=50K	Wife	White	279.816536	7
229	Holand-Netherlands	<=50K	Own-child	White	2777.464021	70
157	El-Salvador	>50K	Wife	White	40.056577	1

SQL for Real:
SELECT `native.country`,income,relationship,race,SUM(`hours.per.week`), COUNT(*) FROM C1 WHERE (age <= 56) OR (`hours.per.week` = 30 AND `native.country` LIKE 'United-States') OR (relationship = 'Own-child' OR sex IN ('Male', 'Female')) AND (`marital.status` <> 'Separated' AND income IN ('>50K', '<=50K')) OR (workclass LIKE 'Private' OR occupation <> 'Prof-specialty') GROUP BY `native.country`,income,relationship,race

Resulted in 635 records

SQL for Synthetic:
SELECT `native.country`,income,relationship,race,SUM(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (age <= 56) OR (`hours.per.week` = 30 AND `native.country` LIKE 'United-States') OR (relationship = 'Own-child' OR sex IN ('Male', 'Female')) AND (`marital.status` <> 'Separated' AND income IN ('>50K', '<=50K')) OR (workclass LIKE 'Private' OR occupation <> 'Prof-specialty') GROUP BY `native.country`,income,relationship,race

Resulted in 697 records

Normalized Euclidean distance for (`hours.per.week`): 19.31

Hellinger Distance: 0.412

=====

Real						
	marital.status	sex	relationship	education	MAX(`hours.per.week`)	COUNT(*)
489	Separated	Male	Not-in-family	5th-6th	50	8
346	Never-married	Female	Own-child	Assoc-acdm	80	84
369	Never-married	Female	Unmarried	Some-college	99	204
392	Never-married	Male	Other-relative	9th	60	23
531	Separated	Male	Unmarried	HS-grad	80	45

Synthetic						
	marital.status	sex	relationship	education	MAX(`hours.per.week`)	COUNT(*)
489	Separated	Male	Not-in-family	5th-6th	40.028519	5.000000
346	Never-married	Female	Own-child	Assoc-acdm	40.048147	66.000000
369	Never-married	Female	Unmarried	Some-college	40.101259	83.000000
392	Never-married	Male	Other-relative	9th	40.029833	7.000000
531	Separated	Male	Unmarried	HS-grad	40.058127	53.000000

SQL for Real:
SELECT `marital.status`,`sex`,relationship,education,MAX(`hours.per.week`), COUNT(*) FROM C1 WHERE (education LIKE 'Some-college') AND (race <> 'White' AND sex IN ('Male', 'Female')) AND ((`hours.per.week` BETWEEN 40 AND 46) AND workclass LIKE 'Self-emp-not-inc') OR (income = '<=50K' OR fnlwgt <> 99065) AND ((age BETWEEN 62 AND 32) AND relationship = 'Own-child') OR (`marital.status` IN ('Separated', 'Widowed') OR capital <= 0) GROUP BY `marital.status`,`sex`,relationship,education

Resulted in 624 records

SQL for Synthetic:
SELECT `marital.status`,`sex`,relationship,education,MAX(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (education LIKE 'Some-college') AND (race <> 'White' AND sex IN ('Male', 'Female')) AND ((`hours.per.week` BETWEEN 40 AND 46) AND workclass LIKE 'Self-emp-not-inc') OR (income = '<=50K' OR fnlwgt <> 99065) AND ((age BETWEEN 62 AND 32) AND relationship = 'Own-child') OR (`marital.status` IN ('Separated', 'Widowed') OR capital <= 0) GROUP BY `marital.status`,`sex`,relationship,education

Resulted in 596 records

Normalized Euclidean distance for (`hours.per.week`): 22.29

Hellinger Distance: 0.094

=====

Real			
	occupation	relationship	SUM(capital) COUNT(*)
82	Transport-moving	Unmarried	31016.000000 140.000000
52	Priv-house-serv	Other-relative	0.000000 3.000000
29	Farming-fishing	Not-in-family	144818.000000 277.000000
49	Other-service	Wife	0.000000 1.000000
47	Other-service	Own-child	44426.000000 1305.000000

Synthetic			
	occupation	relationship	SUM(capital) COUNT(*)
82	Transport-moving	Unmarried	1530.608179 132
52	Priv-house-serv	Other-relative	443.862186 6
29	Farming-fishing	Not-in-family	-1020.493997 223
49	Other-service	Wife	-27.598692 1
47	Other-service	Own-child	-403.568725 1384

SQL for Real:
SELECT occupation,relationship,SUM(capital), COUNT(*) FROM C1 WHERE (relationship IN ('Not-in-family', 'Husband', 'Own-child', 'Unmarried')) OR (workclass LIKE 'Private' OR education IN ('7th-8th', 'Preschool', '12th')) AND (capital <> 0 AND fnlwgt BETWEEN 229225 AND 212588) OR (occupation IN ('Armed-Forces', 'Farming-fishing') OR `hours.per.week` < 40) AND (income IN ('>50K', '<=50K') AND race LIKE 'White') AND (`marital.status` <> 'Married-civ-spouse' AND age > 26) GROUP BY occupation,relationship

Resulted in 76 records

SQL for Synthetic:
SELECT occupation,relationship,SUM(capital), COUNT(*) FROM C1_syn_06 WHERE (relationship IN ('Not-in-family', 'Husband', 'Own-child', 'Unmarried')) OR (workclass LIKE 'Private' OR education IN ('7th-8th', 'Preschool', '12th')) AND (capital <> 0 AND fnlwgt BETWEEN 229225 AND 212588) OR (occupation IN ('Armed-Forces', 'Farming-fishing') OR `hours.per.week` < 40) AND (income IN ('>50K', '<=50K') AND race LIKE 'White') AND (`marital.status` <> 'Married-civ-spouse' AND age > 26) GROUP BY occupation,relationship

Resulted in 83 records

Normalized Euclidean distance for (capital): 8.72

Hellinger Distance: 0.038

=====

Real							
	race	occupation	education	relationship	sex	MAX(fnlwgt)	COUNT(*)
2360	White	Priv-house-serv	9th	Own-child	Female	nan	nan
1990	White	Farming-fishing	Assoc-voc	Unmarried	Male	nan	nan
2623	White	Sales	Assoc-acdm	Husband	Male	289349.000000	2.000000
706	Black	Exec-managerial	Assoc-acdm	Not-in-family	Female	190591.000000	1.000000
226	Asian-Pac-Islander	Craft-repair	Doctorate	Husband	Male	nan	nan

Synthetic							
	race	occupation	education	relationship	sex	MAX(fnlwgt)	COUNT(*)
2360	White	Priv-house-serv	9th	Own-child	Female	196332.055962	4
1990	White	Farming-fishing	Assoc-voc	Unmarried	Male	182948.797109	5
2623	White	Sales	Assoc-acdm	Husband	Male	196827.159459	78
706	Black	Exec-managerial	Assoc-acdm	Not-in-family	Female	189281.752204	1
226	Asian-Pac-Islander	Craft-repair	Doctorate	Husband	Male	167902.818436	1

SQL for Real:
SELECT race,occupation,education,relationship,sex,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (capital <> 0) AND (fnlwgt < 99199 OR occupation <> 'Machine-op-inspct') AND (relationship IN ('Unmarried', 'Other-relative', 'Wife', 'Not-in-family', 'Own-child', 'Husband') AND income = '<=50K') OR (age > 35 OR `marital.status` IN ('Married-AF-spouse', 'Separated', 'Divorced')) AND (`native.country` <> 'United-States' AND education IN ('Assoc-acdm', 'Doctorate', '1st-4th', '12th', '11th', 'Some-college', 'Assoc-voc', 'Masters', 'Prof-school', '10th', 'Preschool', 'HS-grad', '5th-6th')) GROUP BY race,occupation,education,relationship,sex

Resulted in 1303 records

SQL for Synthetic:
SELECT race,occupation,education,relationship,sex,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (capital <> 0) AND (fnlwgt < 99199 OR occupation <> 'Machine-op-inspct') AND (relationship IN ('Unmarried', 'Other-relative', 'Wife', 'Not-in-family', 'Own-child', 'Husband') AND income = '<=50K') OR (age > 35 OR `marital.status` IN ('Married-AF-spouse', 'Separated', 'Divorced')) AND (`native.country` <> 'United-States' AND education IN ('Assoc-acdm', 'Doctorate', '1st-4th', '12th', '11th', 'Some-college',

'Assoc-voc', 'Masters', 'Prof-school', '10th', 'Preschool', 'HS-grad', '5th-6th')) GROUP BY race,occupation,education,relationship,sex

Resulted in 2858 records

Normalized Euclidean distance for (fnlwgt): 32.03

Hellinger Distance: 0.277

=====

Real					
	relationship	marital.status	education	MAX(age)	COUNT(*)
32	Not-in-family	Married-spouse-absent	7th-8th	77	10
2	Not-in-family	Divorced	12th	77	36
18	Not-in-family	Married-civ-spouse	1st-4th	33	1
66	Not-in-family	Separated	Assoc-acdm	64	22
12	Not-in-family	Divorced	Masters	90	229

Synthetic					
	relationship	marital.status	education	MAX(age)	COUNT(*)
32	Not-in-family	Married-spouse-absent	7th-8th	74.000000	10.000000
2	Not-in-family	Divorced	12th	65.000000	33.000000
18	Not-in-family	Married-civ-spouse	1st-4th	nan	nan
66	Not-in-family	Separated	Assoc-acdm	49.000000	16.000000
12	Not-in-family	Divorced	Masters	68.000000	258.000000

SQL for Real:
SELECT relationship,`marital.status`,education,MAX(age), COUNT(*) FROM C1 WHERE (sex IN ('Female', 'Male') AND relationship LIKE 'Not-in-family') GROUP BY relationship,`marital.status`,education

Resulted in 91 records

SQL for Synthetic:
SELECT relationship,`marital.status`,education,MAX(age), COUNT(*) FROM C1_syn_06 WHERE (sex IN ('Female', 'Male') AND relationship LIKE 'Not-in-family') GROUP BY relationship,`marital.status`,education

Resulted in 91 records

Normalized Euclidean distance for (age): 9.33

Hellinger Distance: 0.053

=====

Real											
	race	native.country	relationship	education	sex	occupation	workclass	income	marital.status	AVG(fnlwgt)	COUNT(*)
9986	White	Puerto-Rico	Not-in-family	Masters	Male	Prof-specialty	Private	<=50K	Divorced	nan	nan
8318	White	Peru	Not-in-family	HS-grad	Female	Exec-managerial	Private	<=50K	Divorced	nan	nan
3071	Black	United-States	Own-child	HS-grad	Male	Other-service	Local-gov	<=50K	Never-married	241022.750000	4.000000
15794	White	United-States	Unmarried	Masters	Female	Prof-specialty	Local-gov	<=50K	Divorced	167440.185185	27.000000
5757	White	Haiti	Own-child	Some-college	Male	Prof-specialty	Private	<=50K	Never-married	nan	nan

Synthetic											
	race	native.country	relationship	education	sex	occupation	workclass	income	marital.status	AVG(fnlwgt)	COUNT(*)
9986	White	Puerto-Rico	Not-in-family	Masters	Male	Prof-specialty	Private	<=50K	Divorced	172931.175301	5
8318	White	Peru	Not-in-family	HS-grad	Female	Exec-managerial	Private	<=50K	Divorced	179202.973396	3
3071	Black	United-States	Own-child	HS-grad	Male	Other-service	Local-gov	<=50K	Never-married	204409.121762	2
15794	White	United-States	Unmarried	Masters	Female	Prof-specialty	Local-gov	<=50K	Divorced	180238.235699	14
5757	White	Haiti	Own-child	Some-college	Male	Prof-specialty	Private	<=50K	Never-married	205827.053049	1

SQL for Real:
SELECT race,`native.country`,relationship,education,sex,occupation,workclass,income,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (occupation <> 'Transport-moving') OR (education = 'Some-college' AND age <= 21) AND (race LIKE 'White' AND `hours.per.week` BETWEEN 12 AND 40) AND (`native.country` = 'United-States' AND capital <= 0) AND (income <> '<=50K' AND relationship LIKE 'Wife') GROUP BY race,`native.country`,relationship,education,sex,occupation,workclass,income,`marital.status`

Resulted in 12048 records

SQL for Synthetic:
SELECT race,`native.country`,relationship,education,sex,occupation,workclass,income,`marital.status`,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (occupation <> 'Transport-moving') OR (education = 'Some-college' AND age <= 21) AND (race LIKE 'White' AND `hours.per.week` BETWEEN 12 AND 40) AND (`native.country` = 'United-States' AND capital <= 0) AND (income <> '<=50K' AND relationship LIKE 'Wife') GROUP BY race,`native.country`,relationship,education,sex,occupation,workclass,income,`marital.status`

Resulted in 16375 records

Normalized Euclidean distance for (fnlwgt): 63.13

Hellinger Distance: 0.249

=====

Real											
	relationship	sex	income	workclass	marital.status	occupation	native.country	race	MAX(fnlwgt)	COUNT(*)	
689	Wife	Female	>50K	Private	Married-civ-spouse	Exec-managerial	United-States	White	585203.000000	18.000000	
96	Husband	Male	<=50K	State-gov	Married-civ-spouse	Protective-serv	Hungary	White	nan	nan	
429	Own-child	Female	<=50K	Private	Never-married	Exec-managerial	Laos	Black	nan	nan	
568	Own-child	Male	<=50K	State-gov	Never-married	Craft-repair	United-States	Black	nan	nan	
153	Not-in-family	Female	<=50K	Private	Never-married	Craft-repair	Dominican-Republic	Black	nan	nan	

Synthetic										
	relationship	sex	income	workclass	marital.status	occupation	native.country	race	MAX(fnlwgt)	COUNT(*)
689	Wife	Female	>50K	Private	Married-civ-spouse	Exec-managerial	United-States	White	190244.956467	4
96	Husband	Male	<=50K	State-gov	Married-civ-spouse	Protective-serv	Hungary	White	183020.374251	1
429	Own-child	Female	<=50K	Private	Never-married	Exec-managerial	Laos	Black	179492.885253	1
568	Own-child	Male	<=50K	State-gov	Never-married	Craft-repair	United-States	Black	187699.640929	1
153	Not-in-family	Female	<=50K	Private	Never-married	Craft-repair	Dominican-Republic	Black	188239.026530	1

SQL for Real:
SELECT relationship,sex,income,workclass,`marital.status`,occupation,`native.country`,race,MAX(fnlwgt), COUNT(*) FROM C1 WHERE (occupation IN ('Craft-repair', 'Protective-serv', 'Exec-managerial') AND age < 30) GROUP BY relationship,sex,income,workclass,`marital.status`,occupation,`native.country`,race

Resulted in 512 records

SQL for Synthetic:
SELECT relationship,sex,income,workclass,`marital.status`,occupation,`native.country`,race,MAX(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (occupation IN ('Craft-repair', 'Protective-serv', 'Exec-managerial') AND age < 30) GROUP BY relationship,sex,income,workclass,`marital.status`,occupation,`native.country`,race

Resulted in 693 records

Normalized Euclidean distance for (fnlwgt): 13.75

Hellinger Distance: 0.166

=====

Real							
	marital.status	relationship	race	workclass	education	MIN(`hours.per.week`)	COUNT(*)
1481	Never-married	Not-in-family	Other	Local-gov	Bachelors	45	1
1333	Married-spouse-absent	Unmarried	White	Private	Bachelors	20	13
2139	Separated	Not-in-family	White	Private	5th-6th	12	7
351	Divorced	Unmarried	Black	?	11th	10	4
509	Married-civ-spouse	Husband	Amer-Indian-Eskimo	Federal-gov	Bachelors	40	1

Synthetic							
	marital.status	relationship	race	workclass	education	MIN(`hours.per.week`)	COUNT(*)
1481	Never-married	Not-in-family	Other	Local-gov	Bachelors	39.990876	2.000000
1333	Married-spouse-absent	Unmarried	White	Private	Bachelors	39.916537	10.000000
2139	Separated	Not-in-family	White	Private	5th-6th	39.924088	6.000000
351	Divorced	Unmarried	Black	?	11th	nan	nan
509	Married-civ-spouse	Husband	Amer-Indian-Eskimo	Federal-gov	Bachelors	39.965063	1.000000

SQL for Real:
SELECT `marital.status`,relationship,race,workclass,education,MIN(`hours.per.week`), COUNT(*) FROM C1 WHERE (`native.country` <> 'United-States') OR (education = 'Some-college' AND `marital.status` = 'Never-married') OR (workclass <> 'State-gov' OR income LIKE '<=50K') OR (`hours.per.week` > 40 OR relationship = 'Unmarried') AND (race IN ('White', 'Other', 'Amer-Indian-Eskimo') AND fnlwgt BETWEEN 233571 AND 178319) GROUP BY `marital.status`,relationship,race,workclass,education

Resulted in 2653 records

SQL for Synthetic:
SELECT `marital.status`,relationship,race,workclass,education,MIN(`hours.per.week`), COUNT(*) FROM C1_syn_06 WHERE (`native.country` <> 'United-States') OR (education = 'Some-college' AND `marital.status` = 'Never-married') OR (workclass <> 'State-gov' OR income LIKE '<=50K') OR (`hours.per.week` > 40 OR relationship = 'Unmarried') AND (race IN ('White', 'Other', 'Amer-Indian-Eskimo') AND fnlwgt BETWEEN 233571 AND 178319) GROUP BY `marital.status`,relationship,race,workclass,education

Resulted in 2519 records

Normalized Euclidean distance for (`hours.per.week`): 41.88

Hellinger Distance: 0.097

=====

Real											
	native.country	income	workclass	relationship	occupation	marital.status	education	race	sex	SUM(fnlwgt)	COUNT(*)
41	United-States	<=50K	Private	Unmarried	Other-service	Separated	HS-grad	Black	Female	313546.000000	1.000000
36	United-States	<=50K	Private	Not-in-family	Sales	Separated	11th	Black	Male	nan	nan
3	Peru	<=50K	Private	Not-in-family	Prof-specialty	Never-married	HS-grad	Black	Female	nan	nan
45	United-States	<=50K	Private	Wife	Other-service	Married-civ-spouse	HS-grad	Black	Female	nan	nan
13	Puerto-Rico	<=50K	Private	Not-in-family	Machine-op-inspct	Divorced	9th	Black	Male	nan	nan

Synthetic											
	native.country	income	workclass	relationship	occupation	marital.status	education	race	sex	SUM(fnlwgt)	COUNT(*)
41	United-States	<=50K	Private	Unmarried	Other-service	Separated	HS-grad	Black	Female	157632.588051	1
36	United-States	<=50K	Private	Not-in-family	Sales	Separated	11th	Black	Male	173383.007423	1
3	Peru	<=50K	Private	Not-in-family	Prof-specialty	Never-married	HS-grad	Black	Female	178737.780669	1
45	United-States	<=50K	Private	Wife	Other-service	Married-civ-spouse	HS-grad	Black	Female	172623.230190	1
13	Puerto-Rico	<=50K	Private	Not-in-family	Machine-op-inspct	Divorced	9th	Black	Male	170714.697749	1

SQL for Real:
SELECT `native.country`,income,workclass,relationship,occupation,`marital.status`,education,race,sex,SUM(fnlwgt), COUNT(*) FROM C1 WHERE ((`hours.per.week` BETWEEN 52 AND 40) AND occupation LIKE 'Craft-repair') AND (`native.country` LIKE 'United-States' AND education = 'Some-college') OR (workclass = 'Private' OR income <> '<=50K') AND (age = 50 AND race = 'Black') GROUP BY `native.country`,income,workclass,relationship,occupation,`marital.status`,education,race,sex

Resulted in 42 records

SQL for Synthetic:
SELECT `native.country`,income,workclass,relationship,occupation,`marital.status`,education,race,sex,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE ((`hours.per.week` BETWEEN 52 AND 40) AND occupation LIKE 'Craft-repair') AND (`native.country` LIKE 'United-States' AND education = 'Some-college') OR (workclass = 'Private' OR income <> '<=50K') AND (age = 50 AND race = 'Black') GROUP BY `native.country`,income,workclass,relationship,occupation,`marital.status`,education,race,sex

Resulted in 51 records

Normalized Euclidean distance for (fnlwgt): 1.41

Hellinger Distance: 0.12

=====

Real					
	sex	race	native.country	SUM(fnlwgt)	COUNT(*)
178	Male	Black	Philippines	606167.000000	2.000000
35	Female	Black	England	389674.000000	2.000000
65	Female	Other	England	nan	nan
49	Female	Black	Nicaragua	200734.000000	1.000000
56	Female	Black	Taiwan	nan	nan
Synthetic					
	sex	race	native.country	SUM(fnlwgt)	COUNT(*)
178	Male	Black	Philippines	189202.605907	1
35	Female	Black	England	184498.374745	1
65	Female	Other	England	181415.486752	1
49	Female	Black	Nicaragua	2813470.522182	15
56	Female	Black	Taiwan	725988.763380	4

SQL for Real:
SELECT sex,race,`native.country`,SUM(fnlwgt), COUNT(*) FROM C1 WHERE (sex LIKE 'Male' OR occupation <> 'Machine-op-inspct') OR (relationship LIKE 'Unmarried' OR `native.country` LIKE 'United-States') AND (capital >= 0 AND income LIKE '<=50K') AND ((age BETWEEN 33 AND 44) AND fnlwgt = 226311) AND (workclass <> 'Self-emp-not-inc' AND `marital.status` IN ('Widowed', 'Married-AF-spouse', 'Divorced', 'Separated')) GROUP BY sex,race,`native.country`
Resulted in 210 records

SQL for Synthetic:
SELECT sex,race,`native.country`,SUM(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (sex LIKE 'Male' OR occupation <> 'Machine-op-inspct') OR (relationship LIKE 'Unmarried' OR `native.country` LIKE 'United-States') AND (capital >= 0 AND income LIKE '<=50K') AND ((age BETWEEN 33 AND 44) AND fnlwgt = 226311) AND (workclass <> 'Self-emp-not-inc' AND `marital.status` IN ('Widowed', 'Married-AF-spouse', 'Divorced', 'Separated')) GROUP BY sex,race,`native.country`
Resulted in 242 records

Normalized Euclidean distance for (fnlwgt): 12.37
Hellinger Distance: 0.398

=====

Real						
	workclass	occupation	sex	native.country	MAX(age)	COUNT(*)
615	Private	Craft-repair	Male	Ireland	32.000000	6.000000
606	Private	Craft-repair	Male	Ecuador	90.000000	8.000000
1101	Private	Tech-support	Male	Thailand	nan	nan
1113	Private	Transport-moving	Female	Puerto-Rico	61.000000	3.000000
1177	Self-emp-inc	Exec-managerial	Male	Guatemala	nan	nan
Synthetic						
	workclass	occupation	sex	native.country	MAX(age)	COUNT(*)
615	Private	Craft-repair	Male	Ireland	31	36
606	Private	Craft-repair	Male	Ecuador	29	21
1101	Private	Tech-support	Male	Thailand	20	2
1113	Private	Transport-moving	Female	Puerto-Rico	65	14
1177	Self-emp-inc	Exec-managerial	Male	Guatemala	35	1

SQL for Real:
SELECT workclass,occupation,sex,`native.country`,MAX(age), COUNT(*) FROM C1 WHERE ((capital BETWEEN 0 AND 0) OR fnlwgt >= 126754) OR (`hours.per.week` <= 40 OR sex IN ('Female', 'Male')) GROUP BY workclass,occupation,sex,`native.country`
Resulted in 1407 records

SQL for Synthetic:
SELECT workclass,occupation,sex,`native.country`,MAX(age), COUNT(*) FROM C1_syn_06 WHERE ((capital BETWEEN 0 AND 0) OR fnlwgt >= 126754) OR (`hours.per.week` <= 40 OR sex IN ('Female', 'Male')) GROUP BY workclass,occupation,sex,`native.country`
Resulted in 1636 records
Normalized Euclidean distance for (age): 27.93
Hellinger Distance: 0.364

=====

Real			
	workclass	MIN(age)	COUNT(*)
4	Self-emp-not-inc	19	4
2	Private	17	301
3	Self-emp-inc	22	3
1	Local-gov	17	92
0	Federal-gov	17	105
Synthetic			
	workclass	MIN(age)	COUNT(*)
4	Self-emp-not-inc	29.000000	3.000000
2	Private	17.000000	219.000000
3	Self-emp-inc	41.000000	4.000000
1	Local-gov	19.000000	66.000000
0	Federal-gov	18.000000	58.000000

SQL for Real:
SELECT workclass,MIN(age), COUNT(*) FROM C1 WHERE (occupation LIKE 'Adm-clerical') AND (workclass <> 'Private' OR education IN ('11th', 'Some-college', '10th', 'Assoc-acdm', '12th')) AND (relationship <> 'Husband' AND race <> 'White') GROUP BY workclass

Resulted in 6 records

SQL for Synthetic:
SELECT workclass,MIN(age), COUNT(*) FROM C1_syn_06 WHERE (occupation LIKE 'Adm-clerical') AND (workclass <> 'Private' OR education IN ('11th', 'Some-college', '10th', 'Assoc-acdm', '12th')) AND (relationship <> 'Husband' AND race <> 'White') GROUP BY workclass

Resulted in 6 records

Normalized Euclidean distance for (age): 2.45

Hellinger Distance: 0.046

=====

Real								
	workclass	native.country	race	relationship	occupation	income	AVG(fnlwgt)	COUNT(*)
2602	Private	Vietnam	White	Own-child	Prof-specialty	<=50K	nan	nan
18	?	Columbia	White	Own-child	?	<=50K	nan	nan
2811	Self-emp-inc	United-States	White	Husband	Handlers-cleaners	<=50K	nan	nan
503	Local-gov	Hungary	White	Husband	Exec-managerial	<=50K	nan	nan
1457	Private	Hungary	White	Own-child	Prof-specialty	<=50K	nan	nan

Synthetic								
	workclass	native.country	race	relationship	occupation	income	AVG(fnlwgt)	COUNT(*)
2602	Private	Vietnam	White	Own-child	Prof-specialty	<=50K	188560.934934	2
18	?	Columbia	White	Own-child	?	<=50K	177361.522217	2
2811	Self-emp-inc	United-States	White	Husband	Handlers-cleaners	<=50K	154673.715025	2
503	Local-gov	Hungary	White	Husband	Exec-managerial	<=50K	180468.654787	1
1457	Private	Hungary	White	Own-child	Prof-specialty	<=50K	181971.521270	2

SQL for Real:
SELECT workclass,`native.country`,race,relationship,occupation,income,AVG(fnlwgt), COUNT(*) FROM C1 WHERE (relationship = 'Husband' OR race LIKE 'White') AND (sex IN ('Female', 'Male') AND education IN ('Assoc-acdm', 'Some-college', '12th', '10th', '5th-6th', 'Assoc-voc', 'Bachelors', 'Prof-school', '9th', '1st-4th', '11th', 'Doctorate')) OR (workclass <> 'State-gov' OR occupation <> 'Adm-clerical') AND (`marital.status` IN ('Married-AF-spouse', 'Married-spouse-absent', 'Widowed', 'Never-married') AND age > 37) GROUP BY workclass,`native.country`,race,relationship,occupation,income

Resulted in 2493 records

SQL for Synthetic:
SELECT workclass,`native.country`,race,relationship,occupation,income,AVG(fnlwgt), COUNT(*) FROM C1_syn_06 WHERE (relationship = 'Husband' OR race LIKE 'White') AND (sex IN ('Female', 'Male') AND education IN ('Assoc-acdm', 'Some-college', '12th', '10th', '5th-6th', 'Assoc-voc', 'Bachelors', 'Prof-school', '9th', '1st-4th', '11th', 'Doctorate')) OR (workclass <> 'State-gov' OR occupation <> 'Adm-clerical') AND (`marital.status` IN ('Married-AF-spouse', 'Married-spouse-absent', 'Widowed', 'Never-married') AND age > 37) GROUP BY workclass,`native.country`,race,relationship,occupation,income

Resulted in 3537 records

Normalized Euclidean distance for (fnlwgt): 33.17

Hellinger Distance: 0.314

=====

Hellinger Distance Summary: {'mean': 0.2198806828617489, 'median': 0.2203291864871219, 'stddev': 0.14051246619427304}

Euclidean distance Summary : {'mean': 24.487519116718367, 'median': 22.293496809607948, 'stddev': 18.68428708778959}

=====