Análises no dataset original

Analysis by age

age	income_per_year	number_kids
23	50560.900000	1.850000
24	47311.242424	1.909091
25	51602.483871	1.677419
26	54566.285714	2.238095
27	39126.045455	1.863636
28	41129.684211	2.052632
29	47792.968750	2.000000
30	42328.960000	2.440000
31	36774.961538	1.615385
32	50397.382353	1.882353
33	47980.000000	2.235294
34	56580.277778	1.555556
35	46414.680000	2.080000
36	41603.200000	1.680000
37	50578.333333	2.000000
38	48759.714286	2.357143
39	39590.578947	2.000000
40	55183.111111	1.388889
41	49715.333333	2.233333
42	46109.742857	2.114286
43	47043.000000	1.931034
44	43582.954545	2.227273
45	47463.421053	2.473684
46	46249.833333	1.933333
47	49076.400000	1.866667
48	45500.000000	1.433333
49	49735.880000	1.840000
50	45680.444444	2.037037
51	43346.448276	1.689655
52	43579.454545	2.136364
53	45500.038462	1.730769
54	51984.969697	1.787879
55	44640.920000	2.120000
56	40173.035714	2.107143
57	42264.935484	1.774194
58	47649.861111	1.805556
59	47712.346154	1.692308
60	46152.840000	2.600000

Analysis by gender

gender	income_per_year	number_kids
Female	46230.110220	2.02004
Male	46998.105788	1.89022

Analysis by city

city	income_per_year	number_kids
Aveiro	50495.992593	2.103704
Braga	45636.715517	1.965517
Castelo Branco	47734.728682	1.984496
Coimbra	48377.017241	1.965517
Faro	47572.396396	1.891892
Lisboa	46197.371901	1.983471
Porto	43783.720000	1.768000
Viseu	43477.285714	1.959184

Analysis by job

job	income_per_year	number_kids
Accountant	50019.400000	1.776923
Doctor	80684.625000	1.947368
Engineer	50370.781250	1.968750
Lawyer	53983.263889	2.131944
Manager	49782.144828	2.027586
Teacher	27477.147651	1.885906
Waiter	15228.519737	1.934211

Analysis by signs

signs	income_per_year	number_kids
Aquarius	48025.326087	1.793478
Aries	47358.603960	1.881188
Cancer	47310.526316	2.039474
Capricorn	45997.771429	2.014286
Gemini	46777.067568	2.162162
Leo	46765.565789	2.289474
Libra	43984.583333	1.802083
Pisces	45042.026667	2.013333
Sagittarius	48642.576471	1.941176
Scorpio	46077.930233	1.837209
Taurus	46456.712329	1.821918
Virgo	46800.666667	1.979167

Exercício 1

Atributos	**Tipo de Atributos**
Name	Identifying
email	Identifying
gender	Quasi-identifier
ip_address	Identifying
city	Quasi-identifier
phone	Identifying
job	Quasi-identifier
ssn	Identifying
credit_card	Identifying
income_per_year	Sensitive
age	Quasi-identifier
number_kids	Sensitive
signs	Quasi-identifier

Distinction and Separation - Original Dataset

Quasi-identifier	Distinction	Separation
gender	0.2%	50.04985%
job	0.7%	85.74034%
city	0.8%	87.49009%
signs	1.2%	91.63163%
age	3.8%	97.35816%
gender, job	1.4%	92.86927%
gender, city	1.6%	93.75716%
gender, signs	2.4%	95.84184%
job, city	5.6%	98.1952%
gender, age	7.6%	98.68529%
job, signs	8.4%	98.81642%
city, signs	9.6%	98.93734%
job, age	26,00%	99.62763%
age, city	29.3%	99.65926%
age, signs	41.3%	99.79139%
gender, job, city	11.2%	99.1017%
gender, job, signs	16.8%	99.40961%
gender, city, signs	19.2%	99.47287%
gender, job, age	45.1%	99.81602%
gender, age, city	48,00%	99.82823%
job, city, signs	51.4%	99.84104%
gender, age, signs	62.1%	99.9005%
job, age, city	79.3%	99.94955%
age, city, signs	86.4%	99.97077%
job, age, signs	87.5%	99.97337%
gender, job, city, signs	69.5%	99.92192%
gender, job, age, city	88.8%	99.97598%
gender, age, city, signs	92.5%	99.98478%
gender, job, age, signs	95,00%	99.98999%
job, age, city, signs	97.8%	99.9954%
gender, job, age, city, signs	99.4%	99.9988%

Privacy Risks - Original Dataset

Measure	Value [%]
Lowest prosecutor risk	50,00%
Records affected by lowest risk	1.2%
Average prosecutor risk	99.4%
Highest prosecutor risk	100,00%
Records affected by highest risk	98.8%
Estimated prosecutor risk	100,00%
Estimated journalist risk	100,00%
Estimated marketer risk	99.4%
Sample uniques	98.8%
Population uniques	97.6144%
Population model	ZAYATZ
Quasi-identifiers	age, city, gender, job, signs

Hierarquia usada para os quasi-Identifiers

- gender

  Aqui utilizamos sets com apenas 1 nível, ou seja, ou mostra o gênero, ou não mostra.

- city

  Aqui utilizamos set com 4 níveis como podemos observar na tabela embaixo.

Level-1	Level-2	Level-3	Level-4
{Aveiro}	{Av, Br}	{Av, Br, CasB, Cbr}	*
{Braga}	{Av, Br}	{Av, Br, CasB, Cbr}	*
{Castelo Branco}	{CasB, Cbr}	{Av, Br, CasB, Cbr}	*
{Coimbra}	{CasB, Cbr}	{Av, Br, CasB, Cbr}	*
{Faro}	{Fr, Lx}	{Fr, Lx, Pt, Vs}	*
{Lisboa}	{Fr, Lx}	{Fr, Lx, Pt, Vs}	*
{Porto}	{Pt, Vs}	{Fr, Lx, Pt, Vs}	*
{Viseu}	{Pt, Vs}	{Fr, Lx, Pt, Vs}	*

- job

Da mesma maneira que fizemos com a coluna City, aqui também utilizamos uma hierarquia de sets como podemos ver na seguinte tabela.

Level-0	Level-1	Level-2	Level-3
{Accountant}	{Acc,Doc}	{Acc,Doc,Eng,Law}	*
{Doctor}	{Acc,Doc}	{Acc,Doc,Eng,Law}	*
{Engineer}	{Eng,Law}	{Acc,Doc,Eng,Law}	*
{Lawyer}	{Eng,Law}	{Acc,Doc,Eng,Law}	*
{Manager}	{Man,Tch}	{Man,Tch,Waiter}	*
{Teacher}	{Man,Tch}	{Man,Tch,Waiter}	*
{Waiter}	{Waiter}	{Man,Tch,Waiter}	*

- age

  Na idade decidimos utilizar intervalos, como a sua variação varia entre 23 e 60 anos, utilizamos fazer 4 níveis em que o primeiro começa com 3 intervalos, convergindo num único intervalo no último nível como podemos observar abaixo. É de notar que o nível 0 são os valores contínuos, ou seja, sem estar em intervalos.

Level-0	Level-1	Level-2	Level-3
	[23,38[	[23,53[	[23,61[
	[38,53[	[23,53[	[23,61[
	[53,61[	[53,61[	[23,61[

- signs

  No que toca aos signos decidimos aplicar set como fizemos com os job e as city. Com um total de 4 níveis como podemos observar em baixo.

Level-0	Level-1	Level-2	Level-3
Aquarius	{Aq,Ar}	{Aq,Ar,Can,Cap}	*
Aries	{Aq,Ar}	{Aq,Ar,Can,Cap}	*
Cancer	{Can,Cap}	{Aq,Ar,Can,Cap}	*
Capricorn	{Can,Cap}	{Aq,Ar,Can,Cap}	*
Gemini	{Ge,Le}	{Li,Pi,Ge.Le}	*
Leo	{Ge,Le}	{Li,Pi,Ge.Le}	*
Libra	{Li,Pi}	{Li,Pi,Ge.Le}	*
Pisces	{Li,Pi}	{Li,Pi,Ge.Le}	*
Sagittarius	{Sa,Sco}	{Ta,Vi,Sa,Sco}	*
Scorpio	{Sa,Sco}	{Ta,Vi,Sa,Sco}	*
Taurus	{Ta,Vi}	{Ta,Vi,Sa,Sco}	*
Virgo	{Ta,Vi}	{Ta,Vi,Sa,Sco}	*

Nota: Os pesos dos atributos gender, birth, city e job estão todos a 0.5 e limitamos a supression a 10%.

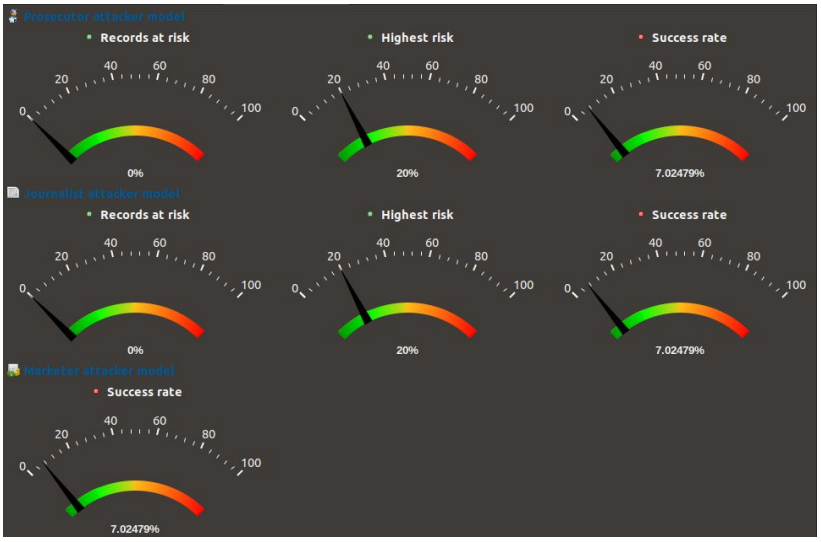
Usando o modelo de Privacidade L-Diversity, L=4 em todos os Sensitive

- Distinction and Separation - Resulting Dataset

Quasi-identifier	Distinction	Separation
job	0.20661%	49.49597%
city	0.20661%	49.96624%
gender	0.20661%	50.04636%
age	0.30992%	64.93735%
signs	0.30992%	66.70178%
job, city	0.41322%	74.70985%
gender, job	0.41322%	74.78634%
gender, city	0.41322%	75.00128%
job, age	0.61983%	82.1876%
gender, age	0.61983%	82.30362%
age, city	0.61983%	82.49058%
job, signs	0.61983%	83.14011%
gender, signs	0.61983%	83.36082%
city, signs	0.61983%	83.36852%
age, signs	0.92975%	88.22207%
gender, job, city	0.82645%	87.38366%
gender, job, age	1.23967%	91.03002%
job, age, city	1.23967%	91.07254%
gender, age, city	1.23967%	91.17083%
job, city, signs	1.23967%	91.57679%
gender, job, signs	1.23967%	91.58683%
gender, city, signs	1.23967%	91.6458%
job, age, signs	1.8595%	93.98305%
gender, age, signs	1.8595%	94.03732%
age, city, signs	1.8595%	94.11616%
gender, job, age, city	2.47934%	95.50753%
gender, job, city, signs	2.47934%	95.76713%
gender, job, age, signs	3.71901%	96.95189%
job, age, city, signs	3.71901%	96.96407%
gender, age, city, signs	3.71901%	97.02112%
gender, job, age, city, signs	7.02479%	98.44881%

- Privacy Risks - Original Dataset

Measure	Value [%]
Lowest prosecutor risk	3.84615%
Records affected by lowest risk	2.68595%
Average prosecutor risk	7.02479%
Highest prosecutor risk	20,00%
Records affected by highest risk	0.51653%
Estimated prosecutor risk	20,00%
Estimated journalist risk	20,00%
Estimated marketer risk	7.02479%
Sample uniques	0,00%
Population uniques	0,00%
Population model	DANKAR
Quasi-identifiers	age, city, gender, job, signs



Análises dos Datasets Transformados

Nota: Ocorreu uma supressão de 3.2% dos dados

Analysis by gender

gender	income_per_year	number_kids
*	51511.437500	2.093750
Female	45922.613497	1.997955
Male	46994.471816	1.901879

Analysis by city

city	income_per_year	number_kids
*	51511.437500	2.093750
{Aveiro, Braga, Castelo Branco, Coimbra}	47913.719828	2.002155
{Faro, Lisboa, Porto, Viseu}	45108.222222	1.902778

Analysis by age

age	income_per_year	number_kids
*	51511.437500	2.093750
[23, 38[	46651.083770	1.913613
[38, 53[	46998.856369	1.983740
[53, 61[	45176.119816	1.958525

Analysis by job

job	income_per_year	number_kids
*	51511.437500	2.093750
{Accountant, Doctor, Engineer, Lawyer}	59358.379439	1.940187
{Manager, Teacher, Waiter}	30507.568129	1.963048

Analysis by signs

signs	income_per_year	number_kids
*	51511.437500	2.093750
{Aquarius, Aries, Cancer, Capricorn}	47100.199377	1.925234
{Gemini, Leo, Libra, Pisces}	45027.945338	2.016077
{Sagittarius, Scorpio, Taurus, Virgo}	47153.735119	1.913690

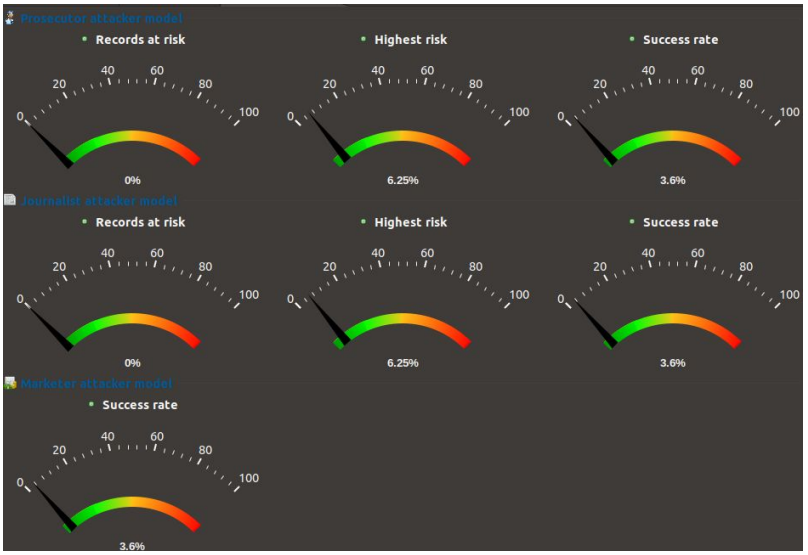
Usando o modelo de Privacidade K-Anonimity com K=10 e L-Diversity com L=8 no income_per_year e L=2 no number_kids

- Distinction and Separation - Resulting Dataset

Quasi-identifier	Distinction	Separation
city	0.1%	0,00%
job	0.2%	49.46627%
gender	0.2%	50.04985%
age	0.3%	65.12032%
signs	0.3%	66.71051%
job, city	0.2%	49.46627%
gender, city	0.2%	50.04985%
age, city	0.3%	65.12032%
city, signs	0.3%	66.71051%
gender, job	0.4%	74.78218%
job, age	0.6%	82.31471%
gender, age	0.6%	82.5039%
job, signs	0.6%	83.14454%
gender, signs	0.6%	83.3958%
age, signs	0.9%	88.36476%
gender, job, city	0.4%	74.78218%
job, age, city	0.6%	82.31471%
gender, age, city	0.6%	82.5039%
job, city, signs	0.6%	83.14454%
gender, city, signs	0.6%	83.3958%
age, city, signs	0.9%	88.36476%
gender, job, age	1.2%	91.15015%
gender, job, signs	1.2%	91.61321%
job, age, signs	1.8%	94.07628%
gender, age, signs	1.8%	94.15716%
gender, job, age, city	1.2%	91.15015%
gender, job, city, signs	1.2%	91.61321%
job, age, city, signs	1.8%	94.07628%
gender, age, city, signs	1.8%	94.15716%
gender, job, age, signs	3.6%	97.03904%
gender, job, age, city, signs	3.6%	97.03904%

- Privacy Risks - After Transformation

Measure	Value [%]
Lowest prosecutor risk	2,00%
Records affected by lowest risk	5,00%
Average prosecutor risk	3.6%
Highest prosecutor risk	6.25%
Records affected by highest risk	1.6%
Estimated prosecutor risk	6.25%
Estimated journalist risk	6.25%
Estimated marketer risk	3.6%
Sample uniques	0,00%
Population uniques	0,00%
Population model	DANKAR
Quasi-identifiers	age, city, gender, job, signs



Análises dos Datasets Transformados

Analysis by gender

gender	income_per_year	number_kids
Female	46230.110220	2.02004
Male	46998.105788	1.89022

Analysis by city

city	income_per_year	number_kids
{Aveiro, Braga, Castelo Branco, Coimbra, Faro, Lisboa, Porto, Viseu}	46614.876	1.955

Analysis by age

age	income_per_year	number_kids
[23, 38[	47023.931122	1.943878
[38, 53[	46619.502646	1.978836
[53, 61[	45910.100000	1.934783

Analysis by job

job	income_per_year	number_kids
{Accountant, Doctor, Engineer, Lawyer}	59544.467509	1.960289
{Manager, Teacher, Waiter}	30554.352018	1.948430

Analysis by signs

signs	income_per_year	number_kids
{Aquarius, Aries, Cancer, Capricorn}	47247.766962	1.920354
{Gemini, Leo, Libra, Pisces}	45533.825545	2.049844
{Sagittarius, Scorpio, Taurus, Virgo}	47004.485294	1.900000

Comparação da distinction entre o Dataset original e os 2 resultantes das transformações