Министр науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

Факультет информационных технологий и программирования

Лабораторная работа №4

Маскировка и анонимизация данных

Выполнил студент группы № M34041
Титов Даниил Ярославович

Подпись

Проверила:
Демина Лилия Станиславовна

Санкт-Петербург

2023

1 Задачи

1. Замаскировать поля с конфиденциальными данными.
2. Провести анонимизацию данных.

2 Предметная область

Поиск работы (https://hh.ru/)

3 Установка

3.1 Создание docker контейнера с расширением PostgreSQL Anonymizer

```
docker pull registry.gitlab.com/dalibo/postgresql_anonymizer:stable
```

4 Маскировка и анонимизация данных

4.1 Создание роли для маскировки данных

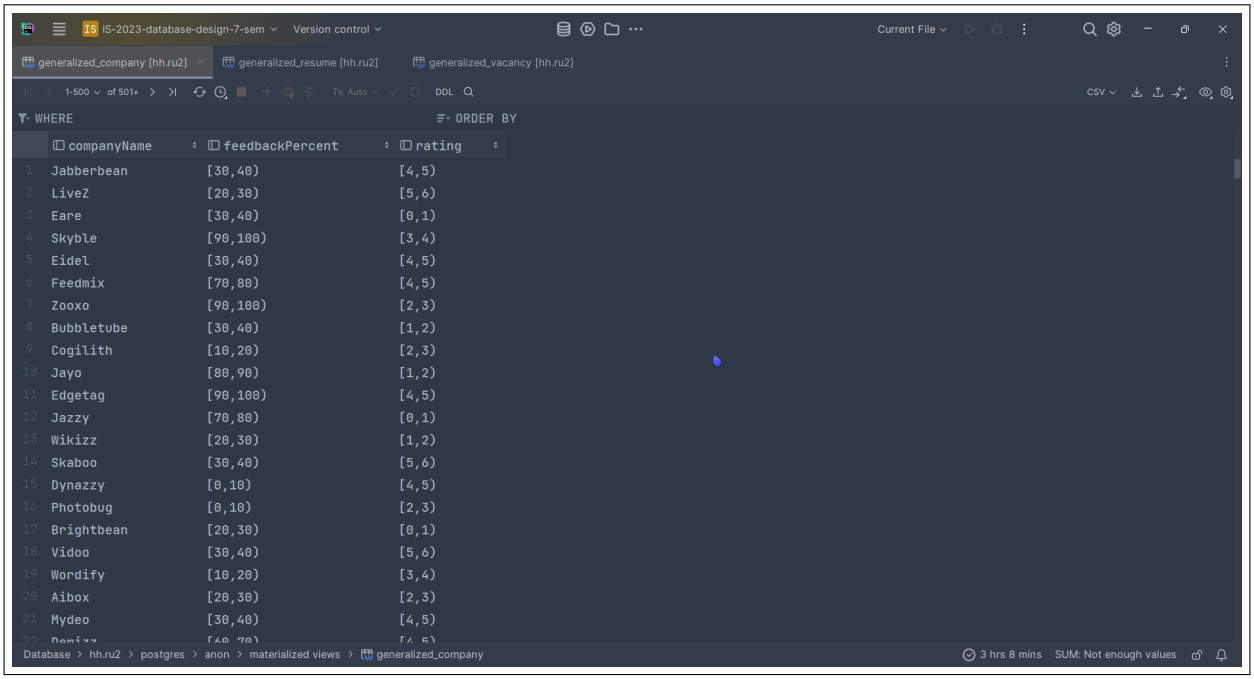
```
CREATE ROLE "custom2" WITH LOGIN PASSWORD '1753';
GRANT USAGE ON SCHEMA "public" TO "custom2";
GRANT SELECT ON ALL TABLES IN SCHEMA "public" TO "custom2";
CREATE EXTENSION IF NOT EXISTS anon CASCADE;

SELECT anon.start_dynamic_masking();
SECURITY LABEL FOR anon ON ROLE "custom2" IS 'MASKED';
```

5 Generalization

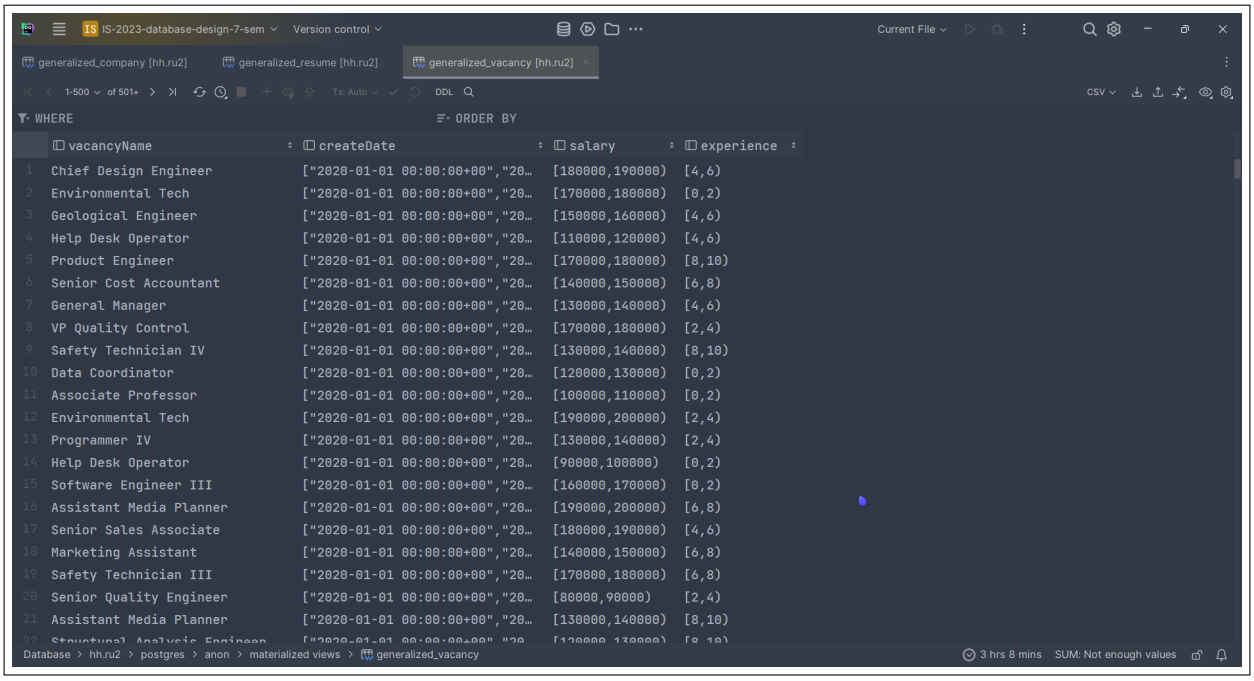
5.1

```
CREATE MATERIALIZED VIEW generalized_company AS
SELECT "companyName",
       anon.generalize_numrange("feedbackPercent", 10) AS "feedbackPercent",
       anon.generalize_numrange("rating", 1)           AS "rating"
FROM public."company";
```



5.2

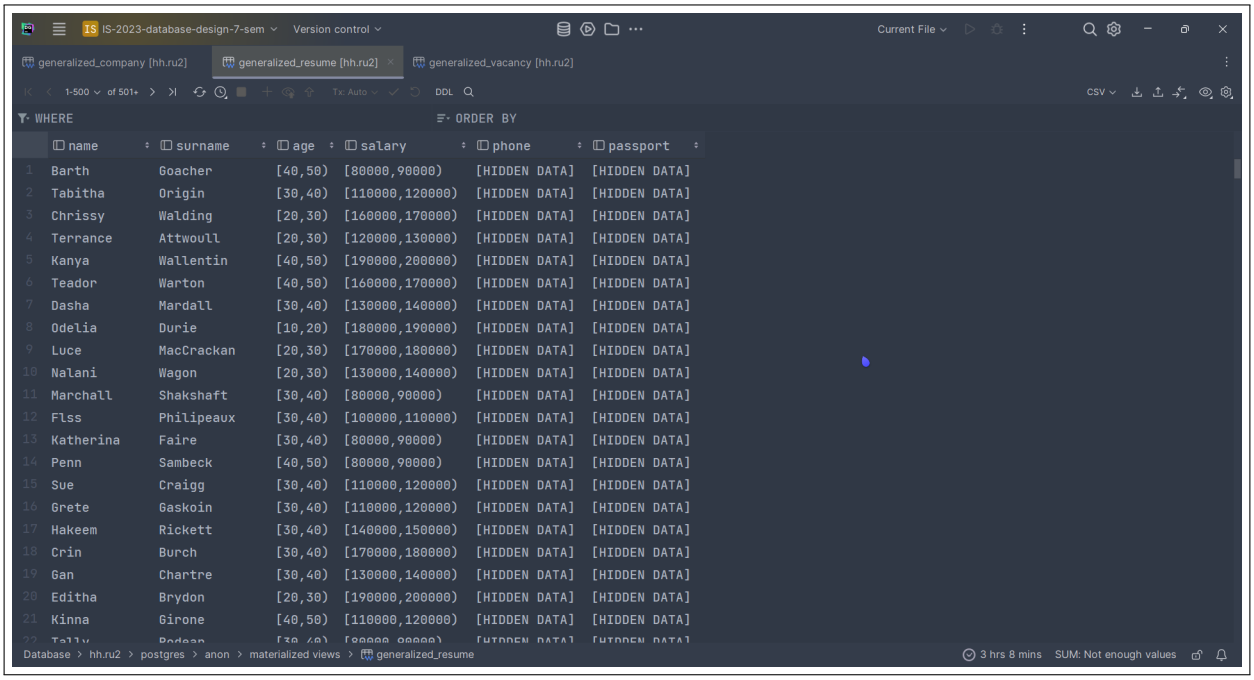
```
CREATE MATERIALIZED VIEW generalized_vacancy AS
SELECT "vacancyName",
       anon.generalize_tstzrange("createDate", 'decade') AS "createDate",
       anon.generalize_int4range("salary", 10000)        AS "salary",
       anon.generalize_int4range("experience", 2)        AS "experience"
FROM "vacancy";
```



	vacancyName	createDate	salary	experience
1	Chief Design Engineer	["2020-01-01 00:00:00+00","20…	[180000,190000)	[4,6)
2	Environmental Tech	["2020-01-01 00:00:00+00","20…	[170000,180000)	[0,2)
3	Geological Engineer	["2020-01-01 00:00:00+00","20…	[150000,160000)	[4,6)
4	Help Desk Operator	["2020-01-01 00:00:00+00","20…	[110000,120000)	[4,6)
5	Product Engineer	["2020-01-01 00:00:00+00","20…	[170000,180000)	[8,10)
6	Senior Cost Accountant	["2020-01-01 00:00:00+00","20…	[140000,150000)	[6,8)
7	General Manager	["2020-01-01 00:00:00+00","20…	[130000,140000)	[4,6)
8	VP Quality Control	["2020-01-01 00:00:00+00","20…	[170000,180000)	[2,4)
9	Safety Technician IV	["2020-01-01 00:00:00+00","20…	[130000,140000)	[8,10)
10	Data Coordinator	["2020-01-01 00:00:00+00","20…	[120000,130000)	[0,2)
11	Associate Professor	["2020-01-01 00:00:00+00","20…	[100000,110000)	[0,2)
12	Environmental Tech	["2020-01-01 00:00:00+00","20…	[190000,200000)	[2,4)
13	Programmer IV	["2020-01-01 00:00:00+00","20…	[130000,140000)	[2,4)
14	Help Desk Operator	["2020-01-01 00:00:00+00","20…	[90000,100000)	[0,2)
15	Software Engineer III	["2020-01-01 00:00:00+00","20…	[160000,170000)	[0,2)
16	Assistant Media Planner	["2020-01-01 00:00:00+00","20…	[190000,200000)	[6,8)
17	Senior Sales Associate	["2020-01-01 00:00:00+00","20…	[180000,190000)	[4,6)
18	Marketing Assistant	["2020-01-01 00:00:00+00","20…	[140000,150000)	[6,8)
19	Safety Technician III	["2020-01-01 00:00:00+00","20…	[170000,180000)	[6,8)
20	Senior Quality Engineer	["2020-01-01 00:00:00+00","20…	[80000,90000)	[2,4)
21	Assistant Media Planner	["2020-01-01 00:00:00+00","20…	[130000,140000)	[8,10)
22	Structural Analysis Engineer	["2020-01-01 00:00:00+00","20…	[120000,130000)	[8,10)

5.3

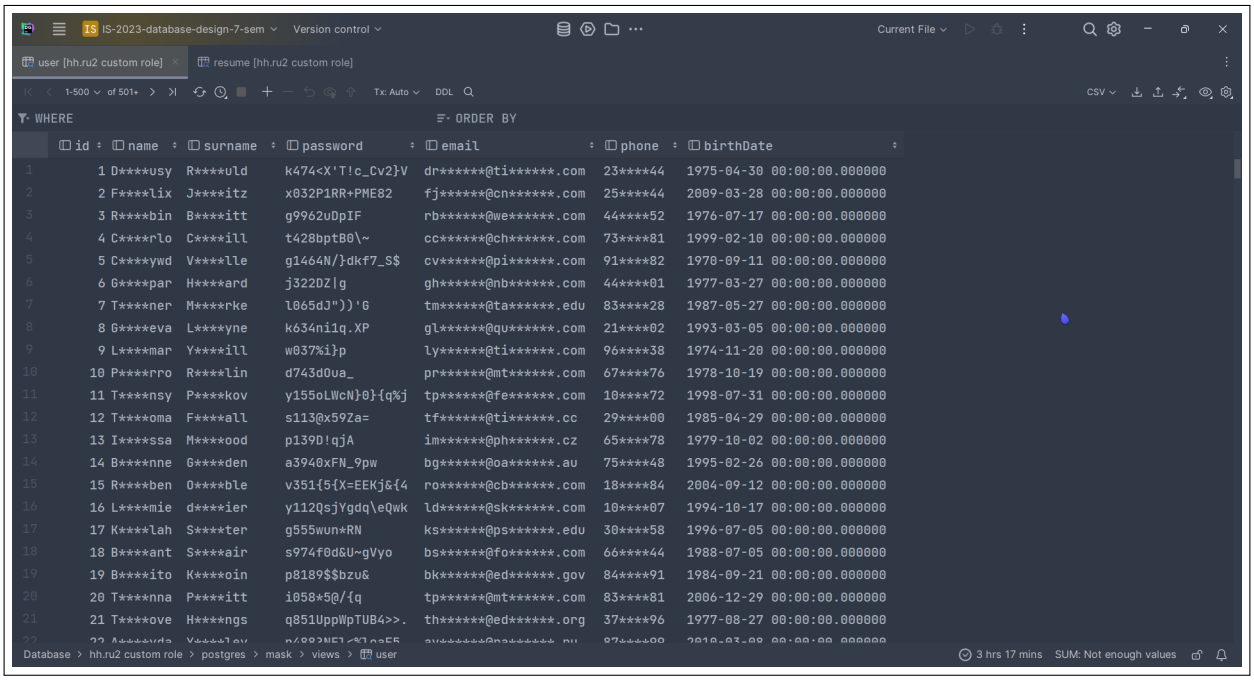
```
CREATE MATERIALIZED VIEW generalized_resume AS
SELECT "name",
       "surname",
       anon.generalize_int4range("age", 10)        AS "age",
       anon.generalize_int4range("salary", 10000) AS "salary",
       '[HIDDEN_DATA]'::TEXT                       AS "phone",
       '[HIDDEN_DATA]'::TEXT                       AS "passport"
FROM public."resume";
```



6 Стратегии анонимизации

6.1 Partial scrambling

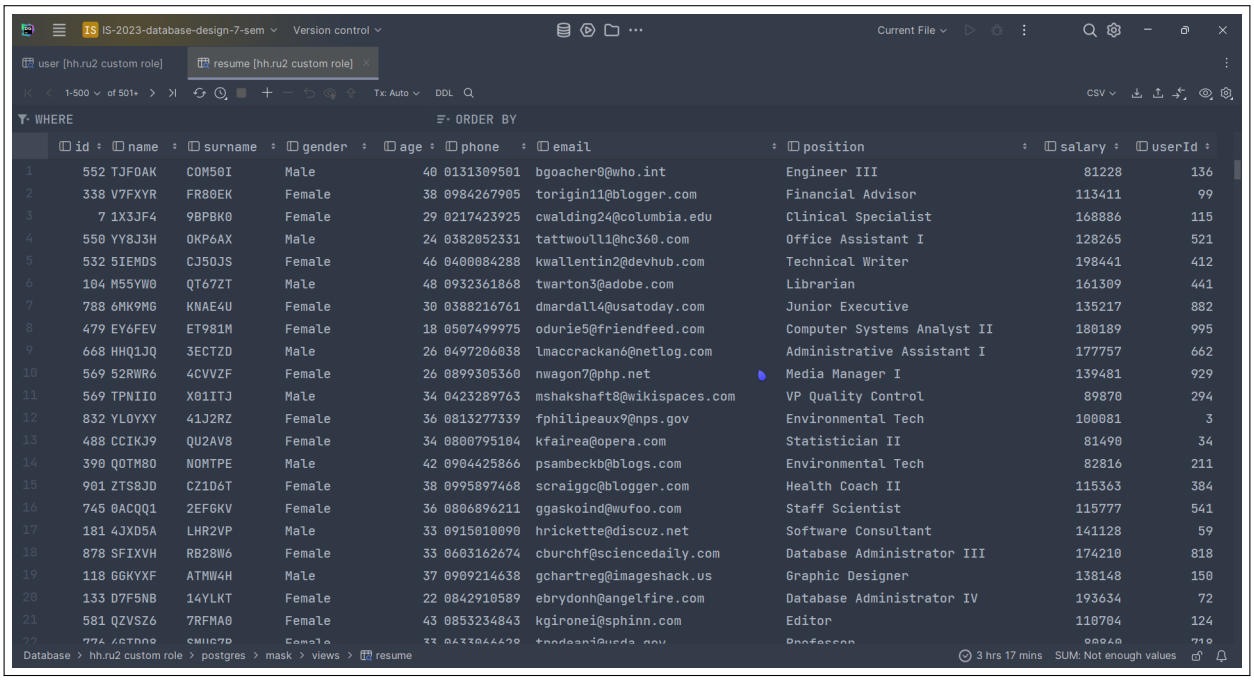
```
SECURITY LABEL FOR anon ON COLUMN "user"."name"
    IS 'MASKED_WITH_FUNCTION_anon.partial("name",1,''****'',3)';
SECURITY LABEL FOR anon ON COLUMN "user"."surname"
    IS 'MASKED_WITH_FUNCTION_anon.partial("surname",1,''****'',3)';
SECURITY LABEL FOR anon ON COLUMN "user".email
    IS 'MASKED_WITH_FUNCTION_anon.partial_email(email)';
SECURITY LABEL FOR anon ON COLUMN "user".phone
    IS 'MASKED_WITH_FUNCTION_anon.partial("phone",2,''****'',2)';
```



id	name	surname	password	email	phone	birthDate
1	D****usy	R****uld	k474<X'T!c_Cv2}V	dr******@ti******.com	23****44	1975-04-30 00:00:00.000000
2	F****lix	J****itz	x032P1RR+PME82	fj******@cn******.com	25****44	2009-03-28 00:00:00.000000
3	R****bin	B****itt	g9962uDpIF	rb******@we******.com	44****52	1976-07-17 00:00:00.000000
4	C****rlo	C****ill	t428bptB0\~	cc******@ch******.com	73****81	1999-02-10 00:00:00.000000
5	C****ywd	V****lle	g1464N/}dkf7_S$	cv******@pi******.com	91****82	1970-09-11 00:00:00.000000
6	G****par	H****ard	j322DZ|g	gh******@nb******.com	44****01	1977-03-27 00:00:00.000000
7	T****ner	M****rke	l065dJ"))'G	tm******@ta******.edu	83****28	1987-05-27 00:00:00.000000
8	G****eva	L****yne	k634ni1q.XP	gl******@qu******.com	21****02	1993-03-05 00:00:00.000000
9	L****mar	Y****ill	w037%i}p	ly******@ti******.com	96****38	1974-11-20 00:00:00.000000
10	P****rro	R****lin	d743dOua_	pr******@mt******.com	67****76	1978-10-19 00:00:00.000000
11	T****nsy	P****kov	y155oLWcN}0}{q%j	tp******@fe******.com	10****72	1998-07-31 00:00:00.000000
12	T****oma	F****all	s113@x59Za=	tf******@ti******.cc	29****00	1985-04-29 00:00:00.000000
13	I****ssa	M****ood	p139D!qjA	im******@ph******.cz	65****78	1979-10-02 00:00:00.000000
14	B****nne	G****den	a3940xFN_9pw	bg******@oa******.au	75****48	1995-02-26 00:00:00.000000
15	R****ben	O****ble	v351{5{X=EEKj&{4	ro******@cb******.com	18****84	2004-09-12 00:00:00.000000
16	L****mie	d****ier	y112QsjYgdq\eQwk	ld******@sk******.com	10****07	1994-10-17 00:00:00.000000
17	K****lah	S****ter	g555wun*RN	ks******@ps******.edu	30****58	1996-07-05 00:00:00.000000
18	B****ant	S****air	s974f0d&U~gVyo	bs******@fo******.com	66****44	1988-07-05 00:00:00.000000
19	B****ito	K****oin	p8189$$bzu&	bk******@ed******.gov	84****91	1984-09-21 00:00:00.000000
20	T****nna	P****itt	i058*5@/{q	tp******@mt******.com	83****81	2006-12-29 00:00:00.000000
21	T****ove	H****ngs	q851UppWpTUB4>>.	th******@ed******.org	37****96	1977-08-27 00:00:00.000000

6.2 Randomization

```
SECURITY LABEL FOR anon ON COLUMN "resume"."id"
    IS 'MASKED_WITH_FUNCTION_anon.random_int_between(0,999)';
SECURITY LABEL FOR anon ON COLUMN "resume"."userId"
    IS 'MASKED_WITH_FUNCTION_anon.random_int_between(0,999)';
SECURITY LABEL FOR anon ON COLUMN "resume"."name"
    IS 'MASKED_WITH_FUNCTION_anon.random_string(6)';
SECURITY LABEL FOR anon ON COLUMN "resume"."surname"
    IS 'MASKED_WITH_FUNCTION_anon.random_string(6)';
SECURITY LABEL FOR anon ON COLUMN "resume"."phone"
    IS 'MASKED_WITH_FUNCTION_anon.random_phone()';
```



WHERE ORDER BY

	id	name	surname	gender	age	phone	email	position	salary	userId
1	552	TJFOAK	COM50I	Male	40	0131309501	bgoacher0@who.int	Engineer III	81228	136
2	338	V7FXYR	FR80EK	Female	38	0984267905	torigin11@blogger.com	Financial Advisor	113411	99
3	7	1X3JF4	9BPBK0	Female	29	0217423925	cwalding24@columbia.edu	Clinical Specialist	168886	115
4	550	YY8J3H	OKP6AX	Male	24	0382052331	tattwoull1@hc360.com	Office Assistant I	128265	521
5	532	5IEMDS	CJ5OJS	Female	46	0400084288	kwallentin2@devhub.com	Technical Writer	198441	412
6	104	M55YW0	QT67ZT	Male	48	0932361868	twarton3@adobe.com	Librarian	161309	441
7	788	6MK9MG	KNAE4U	Female	30	0388216761	dmardall4@usatoday.com	Junior Executive	135217	882
8	479	EY6FEV	ET981M	Female	18	0507499975	odurie5@friendfeed.com	Computer Systems Analyst II	180189	995
9	668	HHQ1JQ	3ECTZD	Male	26	0497206038	lmaccrackan6@netlog.com	Administrative Assistant I	177757	662
10	569	52RWR6	4CVVZF	Female	26	0899305360	nwagon7@php.net	Media Manager I	139481	929
11	569	TPNIIO	X01ITJ	Male	34	0423289763	mshakshaft8@wikispaces.com	VP Quality Control	89870	294
12	832	YLOYXY	41J2RZ	Female	36	0813277339	fphilipeaux9@nps.gov	Environmental Tech	100081	3
13	488	CCIKJ9	QU2AV8	Female	34	0800795104	kfairea@opera.com	Statistician II	81490	34
14	390	QOTM8O	NOMTPE	Male	42	0904425866	psambeckb@blogs.com	Environmental Tech	82816	211
15	901	ZTS8JD	CZ1D6T	Female	38	0995897468	scraiggc@blogger.com	Health Coach II	115363	384
16	745	0ACQQ1	2EFGKV	Female	36	0806896211	ggaskoind@wufoo.com	Staff Scientist	115777	541
17	181	4JXD5A	LHR2VP	Male	33	0915010090	hrickette@discuz.net	Software Consultant	141128	59
18	878	SFIXVH	RB28W6	Female	33	0603162674	cburchf@sciencedaily.com	Database Administrator III	174210	818
19	118	GGKYXF	ATMW4H	Male	37	0909214638	gchartreg@imageshack.us	Graphic Designer	138148	150
20	133	D7F5NB	14YLKT	Female	22	0842910589	ebrydonh@angelfire.com	Database Administrator IV	193634	72
21	581	QZVSZ6	7RFMA0	Female	43	0853234843	kgironei@sphinn.com	Editor	110704	124