



Centrālā statistikas
pārvalde

2000. gada tautas skaitīšanas datu ģeokodēšana

Dāvis Kļaviņš, ĢIS speciālists | davis.klavins@csb.gov.lv

12.01.2018.



Centrālā statistikas
pārvalde

Lai veiktu telpiski detalizētus pētījumus laikā un telpā, piemēram, par migrāciju un iedzīvotāju skaita samazināšanos, nepieciešama ģeokodēta datu kopa par iedzīvotājiem, nodarbinātību un mājokļu apstākļiem. Šādus datus vispilnīgāk iespējams iegūt tautas skaitīšanās.

Atšķirībā no 2011. gada tautas skaitīšanas (TS) datiem, 2000. gada TS dati līdz šim nebija ģeokodēti, un to izmantošanas iespējas pētījumiem laikā un telpā bija ierobežotas



Centrālā statistikas
pārvalde

Eurostat atbalstīta granta projekta ietvaros veikta:

- **2000. gada TS datu kopas ģeokodēšana**
- Iedzīvotāju rādītāju publicēšana, tai skaitā 1×1 km režģa šūnās visai Latvijas teritorijai un 100×100 m režģa šūnās republikas pilsētām, kā arī attālinātas pieejas anonimizētas datu kopas veidā zinātniskiem mērķiem (līdzīgi kā publicēti 2011. gada TS dati)



Centrālā statistikas
pārvalde

Adrešu dati no anketām tika optiski atpazīti un satur kļūdas:

Mājas Nr.	Ielas nosaukums
/1	TARTU
11	
11	TART4
11	TARTU
2A	TA.RTU .
2A	TAR+4
2A	TART4
2A	TARTU
2A	TART-U
-2A	TARTU
5	
5	TART4
5	TARTU
5	TARTU -
5	TARTU.
ZA	TARTU



Centrālā statistikas
pārvalde

Savienojot adresu datus no 2000. gada TS un Iedzīvotāju reģistra (IR) 01.03.2000., tika aprēķināta varbūtība, ka adreses sakrīt

Adresu dati IR arī satur pieraksta kļūdas, bet būtiski mazāk nekā 2000. gada TS dati

Varbūtības tika aprēķinātas R programmā, izmantojot adreses pierakstu veidojošos laukus (ielas/mājas nosaukumu, mājas Nr. u.c.)



Centrālā statistikas
pārvalde

Katram laukam visiem ierakstiem tika aprēķināta Lēvenšteina distance – minimālais izmaiņu daudzums, kas nepieciešams, lai vienu vārdu pārveidotu par otru. Pēc tam tā tika izdalīta ar garākā vārda garumu un atņemta no 1, tādējādi aprēķinot varbūtību, ka lauki sakrīt.

DAUGAVPIL.5 → DAUGAVPILS (distance vienāda ar 2)

$$\text{Varbūtība} = 1 - \frac{2}{11} \approx 0.82$$



Centrālā statistikas
pārvalde

Specifiski gadījumi, kad vārds pilnībā iekļaujas citā. Lai arī lauki bieži vien apraksta vienu un to pašu ģeogrāfisko objektu, aprēķinātās varbūtības bija zemas, jo viens no laukiem bija saīsināts. Līdz ar to tika veiktas izmaiņas varbūtību aprēķināšanā šādos gadījumos.

KR BARONA → KRIŠJĀNA BARONA

Varbūtība_{pirms} = 0.6

Varbūtība_{pēc} ≈ 0.87



Centrālā statistikas
pārvalde

Varbūtības tika apvienotas ar svariem, lai aprēķinātu apvienoto varbūtību katram ierakstam

Vadoties no adreses pieraksta veida un izmantojot MS SQL Server, vairākās (~30) grupās adreses no 2000. gada TS datu kopas tika savienotas ar

1) Valsts adrešu reģistra (VAR) vēsturiskajiem datiem vai, ja ar tiem nesavienojās –

2) 2011. gada TS datu kopu

Pievienotās datu kopas satur 2000. gada TS datu kopai pievienojamās koordinātas



Centrālā statistikas
pārvalde

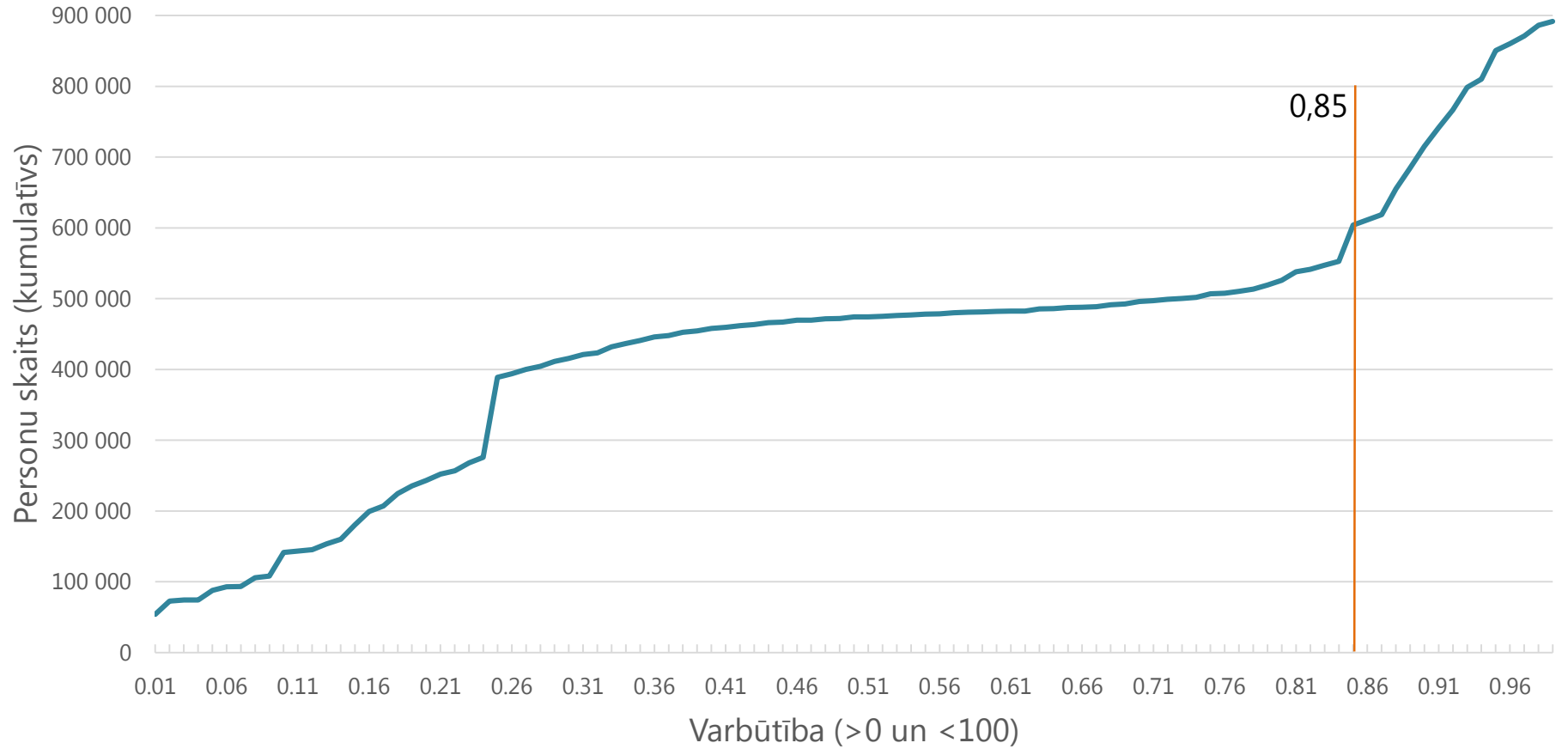
Atsevišķās grupās aprēķinātās varbūtības tika izmantotas, lai aizvietotu adresu datus no anketām ar datiem no IR

Piemēram, ja adresu līdzības varbūtība starp datiem no anketām un IR $>0,85$, savienošanai ar VAR izmantota adrese no IR (savienotie lauki – ciema nosaukums, ielas nosaukums un mājas Nr.)



Centrālā statistikas
pārvalde

Varbūtība, ka adrese datos no anketām un IR sakrīt





Centrālā statistikas
pārvalde

Atlikušajām adresēm tika aprēķināta varbūtība, ka tās sakrīt 2000. un 2011. gada TS datu kopās, un pie vairākiem nosacījumiem izmantota, lai pieņemtu, ka var tikt lietota adrese no 2011. gada TS

Piemēram, ja varbūtība līdzībai starp mājas nosaukumu anketā 2000. gada TS datu kopā un ciema nosaukumu 2011. gada TS datu kopā $\geq 0,7$ (2000. gada TS datu kopā mājas nosaukuma laukā norādīts tikai ciema nosaukums, un adrese no 2011. gada TS datu kopas var tik izmantota, ja vismaz viena persona mājstaimniecībā dzīvo tajā pašā ciemā saskaņā ar 2011. gada TS)



Centrālā statistikas
pārvalde

Adresēm, kuras nebija iespējams automātiski savienot un aptver ~245 000 personas (10%), tika aprēķinātas varbūtības sakritībai ar VAR adresēm. 10 lielākās varbūtības katrai adresei tika izmantotas manuālajai savienošanai (visas, ja lielākā varbūtība sakrita vairāk nekā 10 adresēm).

Ierakstu skaits manuālajai savienošanai ir būtiski mazāks nekā ~245 000 grupēšanas pēc adresei dēļ (sākotnēji ~37 000, bet atsevišķos gadījumos adrese nepieciešams sadalīt mājstāpniecībās vai personās)



Centrālā statistikas
pārvalde

Lai atvieglotu manuālo savienošanu, tika izveidota programmatūra, kurā bez VAR adresēm ar lielākajām sakritības varbūtībām bija dota pieeja arī šādiem avotiem:

- Skanētajām anketām
- Visām intervētāja apsekotajām adresēm
- Deklarētajām dzīvesvietām līdz 2016. gadam un adresei atbilstoši 2011. gada TS



Centrālā statistikas
pārvalde

Darbs ar programmu adrešu manuālajai savienošanai

Merge

Papildus informācija

ATK2000: 7642 MAJANOS: MAJASNUM: 9
ATVK_NAME: Aglonas pagasts KORPUSS:
NOV_NAME: Aglonas novads SKAITIT: 76057
CIEMS: RU30NA5 JAUNAĢLON SKAITIS: 1
IELA: RU30NA5 JAUNAĢLON SKAITIS: 1

Piev. veids: X
Adrešu kods: Y
Prezīmes: Lietotājs

Skaitlāja ATVK kodi

ATK	count
7642	419

Adrešu saraksts

ATVK	ciems	iele	majasnos	majasnum	korpus	skaitis	adr_merged
7642		RUŠONAS JAUNAĢLONA		9		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLONA
7642		RUŠONAS JAUNAĢLO...		5		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		13		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		7		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		9		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		9		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		9		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		5		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		15		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		22		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		4		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		5		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		7		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		9		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		-9		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...				76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUNAĢLO...		5		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUNAĢLO...
7642		RUŠONAS JAUN		9		76057	1-76057ATVK:7642IELA/MAJA RUŠONAS JAUN
7642		RUŠONAS				76057	1-76057ATVK:7642IELA/MAJA RUŠONAS
7642		SKUDRI GREBEŽI				76057	1-76057ATVK:7642IELA/MAJA SKUDRI GREBEŽI
7642		STRAZDI JAUNAĢLONA				76057	1-76057ATVK:7642IELA/MAJA STRAZDI JAUNAĢLONA
7642		TRUŠU JAUNAĢLONA				76057	1-76057ATVK:7642IELA/MAJA TRUŠU JAUNAĢLONA
7642		7AUDZEMU		2B		76057	1-76057ATVK:7642IELA/MAJA 7AUDZEMU
7642		7AUDZEMU		2B		76057	1-76057ATVK:7642IELA/MAJA 7AUDZEMU
7642		7AUDZEMU		5A		76057	1-76057ATVK:7642IELA/MAJA 7AUDZEMU
7642		7		6		76057	1-76057ATVK:7642IELA/MAJA 7
7642		A BROKA		3		76057	1-76057ATVK:7642IELA/MAJA A BROKA
7642		A BROKA		5		76057	1-76057ATVK:7642IELA/MAJA A BROKA
7642		A BROKA		7		76057	1-76057ATVK:7642IELA/MAJA A BROKA

Adrešu saraksts (VARIS)

ATVK	Ciems	Majas_nosaucums	iele	Majas_Nr	Korpus	Adrešu_kods
						Optiski atpazītā adrese no anketas
						Vislīdzīgākā adrese VAR
	Ciems					Jaunaglona
	Iela		RU30NA5 JAUNAĢLONA			RUŠONAS IELA
	Mājas Nr.	9				9



Centrālā statistikas
pārvalde

Darbs ar programmu adrešu manuālajai savienošanai

Papildus informācija

ATK2000: 7842
 ATKV_NAME: Aglonas pagasts
 NOV_NAME: Aglonas novads
 CIEMS: 7
 IELA: 7

MAJANOS: MAJASNUM: 6
 KORPUSS: SKAITIT: 76067
 SKAITIS: 2

Piev. veids: X
 Adreses kods: Y
 Piezīmes: Lietotājs

Skaitlāja ATVK kodi

ATK	count
7842	405

Adrešu saraksts

ATVK	ciems	iele	majasnos	majasnum	korpus	skaitit	adr_merged
7842		RUŠONAS JAUNAGLONA		9		76057	1-76057ATVK:7842IELA/MAJA:RUŠON
7842		RUŠONAS JAUNAG/O...		5		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		13		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		7		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		9		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		9		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAG/O...		5		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		15		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		22		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		4		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		5		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		7		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		9		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		-9		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...				76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		5		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUN		9		76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS				76057	1-76057ATVK:7842IELA/MAJA:RUŠOI
7842		SKIDRI GREBEŽI				76057	1-76057ATVK:7842IELA/MAJA:SKIDF
7842		STRAZDI JAUNAGLONA				76057	1-76057ATVK:7842IELA/MAJA:STRAZ
7842		TRUŠU JAUNAGLONA				76057	1-76057ATVK:7842IELA/MAJA:TRUŠI
7842		7AUDZEMU		2B		76067	1-76067ATVK:7842IELA/MAJA:7AUD2
7842		7AUDZEMU		2B		76067	1-76067ATVK:7842IELA/MAJA:7AUD2
7842		7AUDZEMU		5A		76067	1-76067ATVK:7842IELA/MAJA:7AUD2
7842		7		6		76067	1-76067ATVK:7842IELA/MAJA:7MNI
7842		A BROKA		3		76067	1-76067ATVK:7842IELA/MAJA:A BRO
7842		A BROKA		5		76067	1-76067ATVK:7842IELA/MAJA:A BRO
7842		A BROKA		7		76067	1-76067ATVK:7842IELA/MAJA:A BRO

Līdzīgas adreses

Ciems_VARIS	ID	Majas_nosaukums	iele_VARIS	Majas_Nr_VARIS	Korpus_VARIS	ATVK_VARIS	Pazime	Verbubbas_kart
Agļona	1-76067ATVK:76...		ALOTU IELA	6		0604342	0	1
Jaunagļona	1-76067ATVK:76...		DARZA IELA	6		0604342	0	1
Agļona	1-76067ATVK:76...		EZERA IELA	6		0604342	0	1
Agļona	1-76067ATVK:76...		KALNA IELA	6		0604342	0	1
Jaunagļona	1-76067ATVK:76...		PARKA IELA	6		0604342	0	1
Agļona	1-76067ATVK:76...		ALEJAS IELA	6		0604342	0	2
Agļona	1-76067ATVK:76...		ALOŽA BROKA ...	6		0604342	1	2
Agļona	1-76067ATVK:76...		CIRIŠA IELA	6		0604342	0	2
Agļona	1-76067ATVK:76...		CIRIŠA IELA	6	7	0604342	0	2
Agļona	1-76067ATVK:76...		JAUDZEMU IELA	6		0604342	1	2

Adrešu saraksts (VARIS)

ATVK	Ciems	Majas_nosaukums	iele	Majas_Nr	Korpus	Adreses_kods
						Optiski atpazītā adrese no anketas
						Vislīdzīgākā adrese VAR
	Ciems					Daudzas ar vienādu mājas Nr.
	Iela	7				
	Mājas Nr.	7				7



Centrālā statistikas
pārvalde

Darbs ar programmu adrešu manuālajai savienošanai

maj_num	atk_n	iel_maj	maj_num	corp_num	dzvokls	skatls	adr_merged	ATVK_piek	Ciems_piek	Majas_nosaukums	Iela_piek	Majas_nr_piek	Korpus_piek
1090116	7642	JAUDZEMU	6		8	76067	1-76067ATVK	0604342	Adona		JAUDZEMU I.	6	
1090117	7642	JAUDZEMU	7		11	76067	1-76067ATVK.76						
1090118	7642	JAUDZEMU	6		10	76067	1-76067ATVK	0604342	Adona		JAUDZEMU I.	1	

Starp visām intervētāja apsekotajām adresēm –
iepriekšējā un nākamā mājsaimniecība atrodas tajā pašā ielā



Centrālā statistikas
pārvalde

Darbs ar programmu adrešu manuālajai savienošanai

Merge

Papildus informācija

ATK2000: 7842 MAJANOS: MAJANOSUM: 6
ATVK_NAME: Aglonas pagasts KORPUSS: KORPUSS: 78067
NOV_NAME: Aglonas novads SKAITIT: SKAITIT: 78067
CIEMS: 7 SKAITIS: SKAITIS: 2
IELA: 7

Piev. veids: X:
Adreses kods: Y:
Prezīmes: Lietotājs

Skaitlāja ATVK kodi

ATK	count
7842	405

Adrešu saraksts

ATVK	ciems	iele	majasnos	majasnum	korpus	skaitl	adr_merged
7842		RUŠONAS JAUNAGLONA		9		78057	1-78057ATVK:7842IELA/MAJA:RUŠON
7842		RUŠONAS JAUNAG/O...		5		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		13		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		7		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		9		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		9		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAG/O...		5		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		15		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		22		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		4		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		5		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		7		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		9		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		-9		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...				78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUNAGLO...		5		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS JAUN		9		78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		RUŠONAS				78057	1-78057ATVK:7842IELA/MAJA:RUŠOI
7842		SKUDRI GREBEŽI				78057	1-78057ATVK:7842IELA/MAJA:SKUDF
7842		STRAZDI JAUNAGLONA				78057	1-78057ATVK:7842IELA/MAJA:STRAZ
7842		TRUŠU JAUNAGLONA				78057	1-78057ATVK:7842IELA/MAJA:TRUŠI
7842		7AUDZEMU		2B		78067	1-78067ATVK:7842IELA/MAJA:7AUD2
7842		7AUDZEMU		2B		78067	1-78067ATVK:7842IELA/MAJA:7AUD2
7842		7AUDZEMU		5A		78067	1-78067ATVK:7842IELA/MAJA:7AUD2
7842		7		6		78067	1-78067ATVK:7842IELA/MAJA:7MNI
7842		A BROKA		3		78067	1-78067ATVK:7842IELA/MAJA:A BRO
7842		A BROKA		5		78067	1-78067ATVK:7842IELA/MAJA:A BRO
7842		A BROKA		7		78067	1-78067ATVK:7842IELA/MAJA:A BRO

Uzdīgās adreses

Ciems_VARIS	ID	Majas_nosaukums	iele_VARIS	Majas_Nr_VARIS	Korpuss_VARIS	ATVK_VARIS	Pazīme	Verbubbas_kart
Agļona	1-78067ATVK:78...		ALOTU IELA	6		0604342	0	1
Jaunagļona	1-78067ATVK:78...		DARZA IELA	6		0604342	0	1
Agļona	1-78067ATVK:78...		EZERA IELA	6		0604342	0	1
Agļona	1-78067ATVK:78...		KALNA IELA	6		0604342	0	1
Jaunagļona	1-78067ATVK:78...		PARKA IELA	6		0604342	0	1
Agļona	1-78067ATVK:78...		ALEJAS IELA	6		0604342	0	2
Agļona	1-78067ATVK:78...		ALOŽA BROKA ...	6		0604342	1	2
Agļona	1-78067ATVK:78...		CIRIŠA IELA	6		0604342	0	2
Agļona	1-78067ATVK:78...		CIRIŠA IELA	6	7	0604342	0	2
Agļona	1-78067ATVK:78...		JAUDZEMU IELA	6		0604342	1	2

Adrešu saraksts (VARIS)

ATVK	Ciems	Majas_nosaukums	iele	Majas_Nr	Korpuss	Adreses_kods
0604342	Agļona	JAUDZEMU IELA	6			102143174
0604342	Agļona	JAUDZEMU IELA	6A			106192674

Manuāli sameklē pareizo adresi VAR un to piesaista



Centrālā statistikas
pārvalde

	Personas	% no kopējā iedzīvotāju skaita
Automātiski savienotas	2 080 803	88
Manuāli savienotas (t.sk. automātiski savienotas bez koordinātām, vairāk nekā viena automātiski pievienota adrese un izlabotas kļūdas automātiskajā savienošanā)	296 070	12
Adreses nebija iespējams ģeokodēt vai eksistence TS kritiskajā brīdī ir apšaubāma	510	0



Centrālā statistikas
pārvalde

