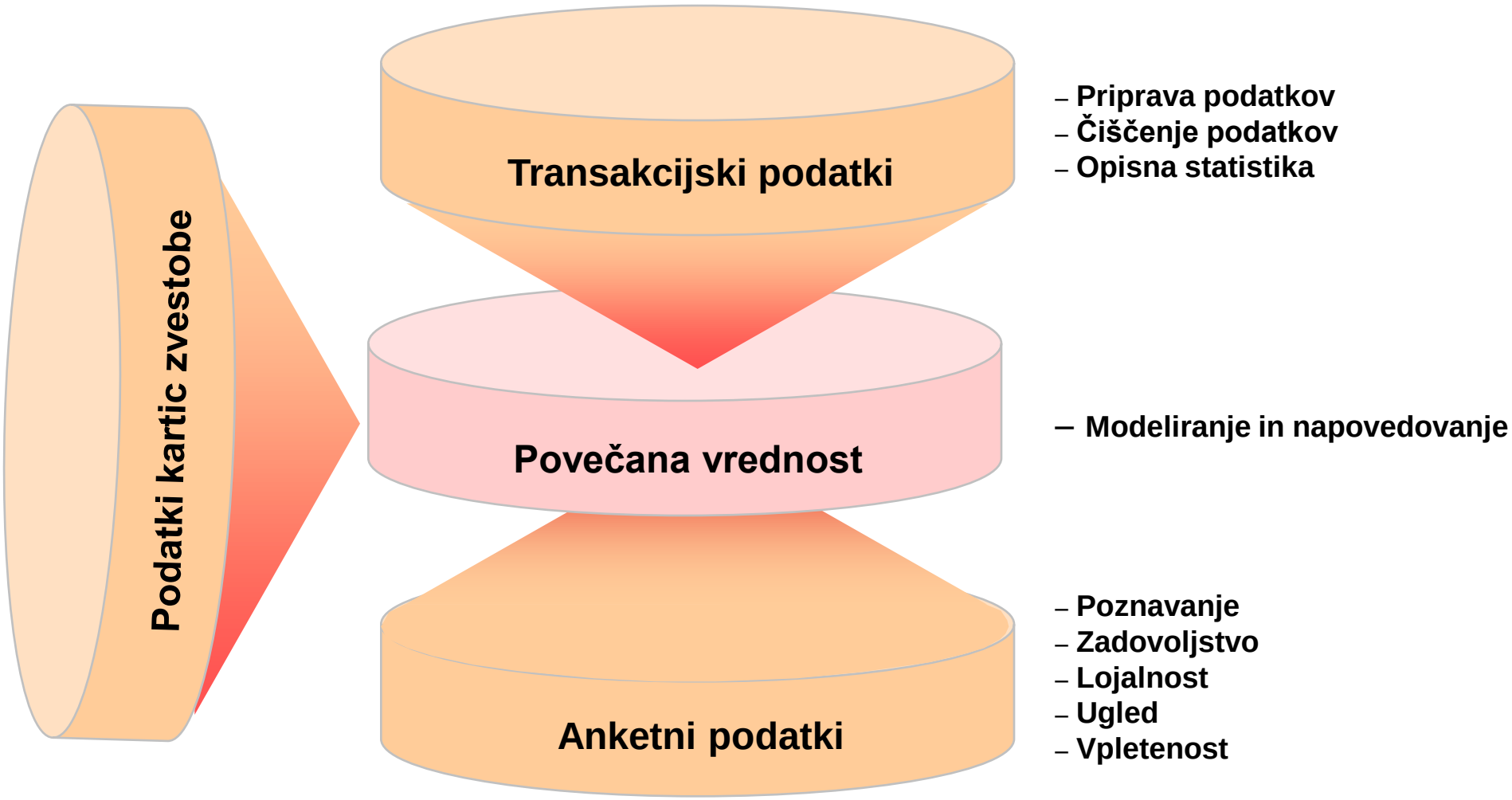
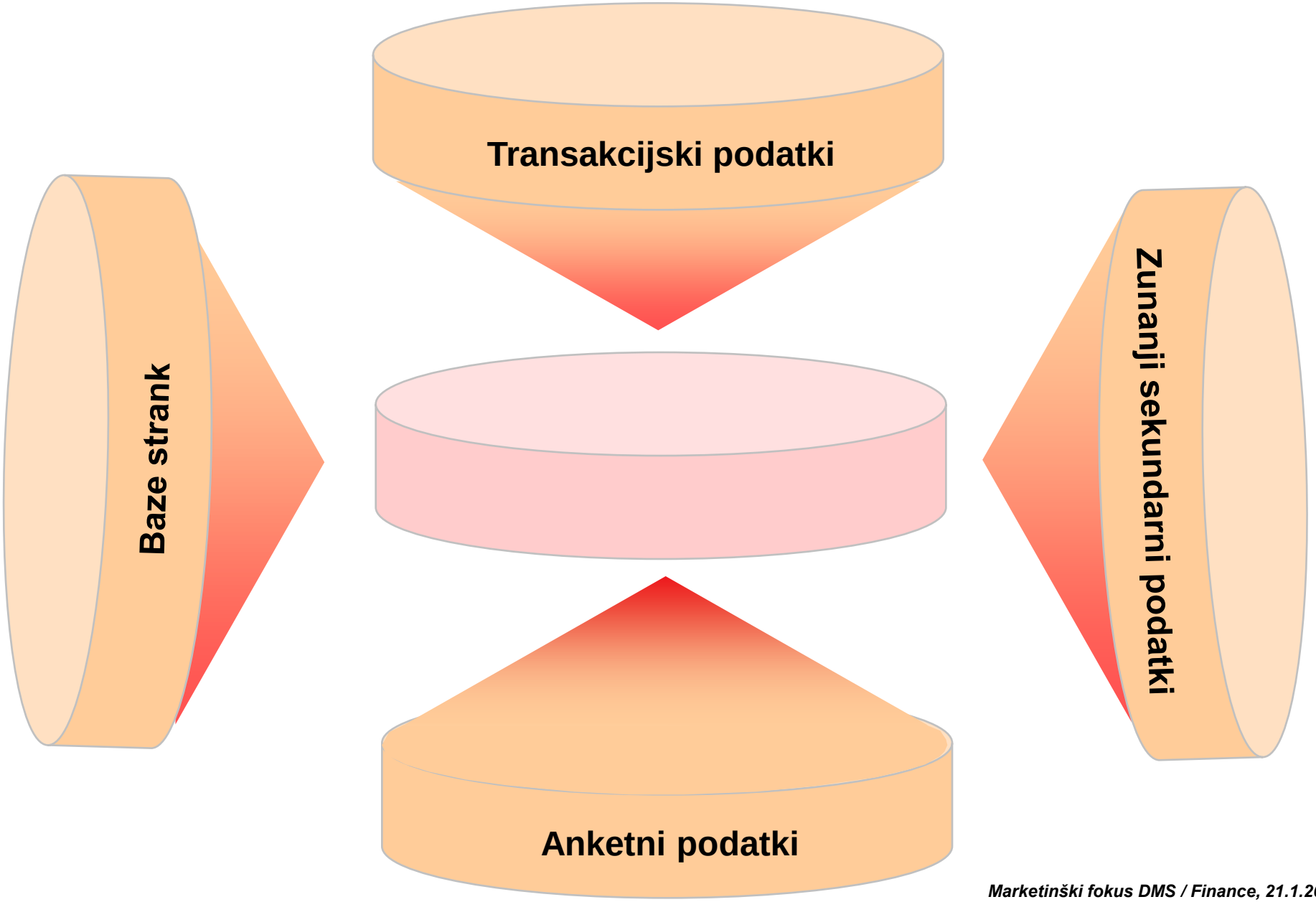
Povezovanje baz podatkov na primeru Monitorja in Kaj je novega na slovenskem internetu?

Zenel Batagelj, CATI

Korak naprej – Data fussion



Splošni model z vidika raziskovalca



Cilji povezovanj baz

- **1. Skupina: Napovedovanje**
  - Kateri kupci so najmanj zvesti?
  - Kateri kupci so najmanj zadovoljni?

- **2. Skupina: Opisovanje**
  - Kdo so kupci določenih produktov?
  - Koliko piva kupijo bralci One?

- **3. Skupina: Posploševanje**
  - Življenjski stil

Povezovanje podatkov

- Problem:
  - Imamo dve bazi podatkov, ki bi ju radi združili v skupno bazo, tj. eno “nalepili” na drugo.

- Zakaj je to dobro?
  - V vseh bazah nimamo vseh informacij in bi jih radi povezali
  - Nekaterih stvari sploh ni treba spraševati, ker jih že imamo v bazah (npr. natančna potrošnja po proizvodih)

Torej: obstajajo različne stopnje povezljivosti podatkov

- ID – enolična identifikacija
  - tu praktično ni težav
- Alternativni identifikacijski podatki (ime, priimek, naslov)
  - problem enoličnosti identifikacije
- Katerikoli skupni podatki
  - še večji problem z enoličnostjo ident.
- Skupnih točk ni
  - povezovanje ni možno

Primer 1: Napovedovanje

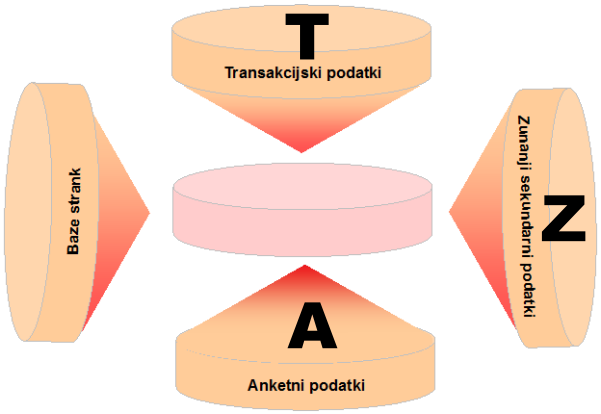
- Problem:
  - Imamo raziskavo zadovoljstva, ki jo izvajamo med našimi potrošniki, rezultate (npr. indeks zadovoljstva) bi radi posplošili na celotno bazo potrošnikov
- Rešitev:
  - Pri vsaki izvedeni anketi potrebujemo tudi ID potrošnika v bazi (če jo ima)
  - Bazi tako povežemo
  - Zadovoljstvo “modeliramo” na podlagi vseh podatkov v transakcijski bazi

Primer 2: Opisovanje

- Problem:
  - Kaj berejo kupci moje trgovske BZ?
- Rešitev:
  - V bazi potrošnikov imamo podatke o njihovi demografiji, kje živijo in še nekaj dodatnih skupnih podatkov
  - Te (skupne) podatke imamo tudi v medijski raziskavi (npr. NRB)
  - Bazi preko skupnih podatkov povežemo
  - Za vsakega kupca imamo tudi ocenjeno medijsko potrošnjo

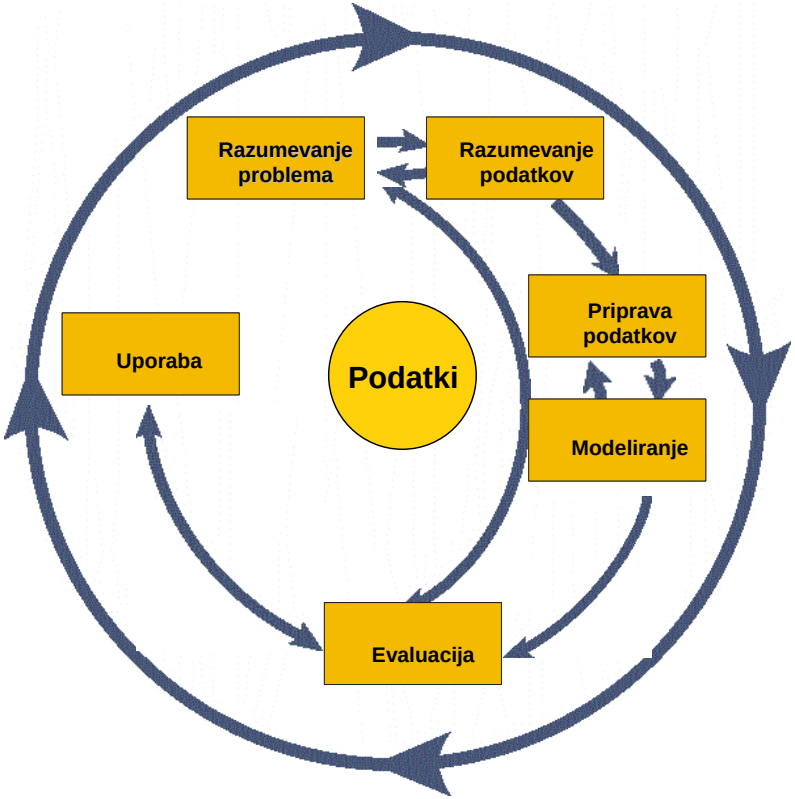
Primer 3: Posploševanje

- Problem:
  - Ugotoviti doseg in strukturo uporabnikov slovenskih spletnih strani
- Rešitev:
  - **Transakcije**: enomesečno beleženje vseh obiskov slovenskih strani, unikatna identifikacija uporabnikov prek “cookija”
  - **Anketa**: spletna anketa za socio-psiho-demo-lajf-stilijo
  - **Zunanje**: telefonska anketa za eksterno evaluacijo vzorcev



Proces...

1. Priprava podatkov
   - Zajem
   - Razumevanje
   - Čiščenje
   - Predelava
2. Povezovanje baz
3. Modeliranje / doseganje reprezentativnosti - uteževanje
4. Evaluacija
5. Končni rezultat



Primer transakcijske baze - splet

MARKETINŠKI FOKUS

CATIWWW.SAV - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

2093 :

	sysmid	i_date	i_time	i_act	i_site	i_ad	var	var	var	var	var	var	var	var	var
2979	3000000038181455	02-NOV-02	1:43:08	I	7ft79n17	meyn29									
2980	3000000038181455	02-NOV-02	1:43:08	I	7ft79n17	o4q71fu9									
2981	3000006204235151	02-NOV-02	1:43:21	I	qdyconw	o4q71fu9									
2982	3000006204235151	02-NOV-02	1:43:22	I	qdyconw	3ddh1raf									
2983	3000000048013137	02-NOV-02	1:43:23	I	r1gmuyx	3ddh1raf									
2984	3000000048013137	02-NOV-02	1:43:33	I	r1gmuyx	3ddh1raf									
2985	3000000997113931	02-NOV-02	1:43:34	I	0aol2shl	yr8ak4m									
2986	3000000495181956	02-NOV-02	1:43:37	I	0nu5e1i1	3ddh1raf									
2987	3000000495181956	02-NOV-02	1:43:37	I	0nu5e1i1	yr8ak4m									
2988	3000000997113931	02-NOV-02	1:43:39	C	0aol2shl	yr8ak4m									
2989	3000000545195233	02-NOV-02	1:43:39	I	tr6c95m	uc2mtan									
2990	3000006204235151	02-NOV-02	1:43:47	I	qdyconw	yr8ak4m									
2991	3000006204235151	02-NOV-02	1:43:48	I	qdyconw	o4q71fu9									
2992	3000006204235151	02-NOV-02	1:43:49	I	qdyconw	3ddh1raf									
2993	3000006204235151	02-NOV-02	1:43:50	I	qdyconw	yr8ak4m									
2994	3000001994194705	02-NOV-02	1:43:51	I	7sgghz7	meyn29									
2995	3000000048013137	02-NOV-02	1:43:51	I	r1gmuyx	qvdsb9d									
2996	3000000387235743	02-NOV-02	1:44:00	I	jy23p2ae	qvdsb9d									
2997	3000000048013137	02-NOV-02	1:44:00	I	r1gmuyx	3ddh1raf									
2998	3000000048013137	02-NOV-02	1:44:14	I	r1gmuyx	o4q71fu9									
2999	3000001994194705	02-NOV-02	1:44:15	I	7sgghz7	meyn29									
3000	3000000261153418	02-NOV-02	1:44:24	I	i7rcbsc2	meyn29									
3001	3000000261153418	02-NOV-02	1:44:24	I	i7rcbsc2	uc2mtan									
3002	3000006204235151	02-NOV-02	1:44:34	I	qdyconw	o4q71fu9									
3003	3000006204235151	02-NOV-02	1:44:35	I	qdyconw	yr8ak4m									
3004	3000000154192718	02-NOV-02	1:44:42	I	kwq30av	3ddh1raf									
3005	3000000154192718	02-NOV-02	1:44:42	I	kwq30av	o4q71fu9									
3006	3000006204235151	02-NOV-02	1:44:43	I	qdyconw	i2w65ze									
3007	3000000048013137	02-NOV-02	1:44:47	I	r1gmuyx	yr8ak4m									
3008	3000001865103832	02-NOV-02	1:44:54	I	kwq30av	o4q71fu9									
3009	3000006847160219	02-NOV-02	1:44:55	I	kwq30av	o4q71fu9									
3010	3000000001025616	02-NOV-02	1:44:56	I	i7rcbsc2	3ddh1raf									
3011	3000006204235151	02-NOV-02	1:45:04	I	qdyconw	o4q71fu9									
3012	3000000545195233	02-NOV-02	1:45:09	I	tr6c95m	uc2mtan									
3013	3000006204235151	02-NOV-02	1:45:11	I	qdyconw	qvdsb9d									
3014	3000000545195233	02-NOV-02	1:45:11	I	tr6c95m	meyn29									
3015	3000001865103832	02-NOV-02	1:45:12	I	kwq30av	o4q71fu9									
3016	3000006847160219	02-NOV-02	1:45:12	I	kwq30av	o4q71fu9									
3017	3000006204235151	02-NOV-02	1:45:13	I	qdyconw	o4q71fu9									

cca. 50 mio takih zapisov

Primer obdelane transakcije baze

MARKETINŠKI FOKUS

i_adic.sav - SPSS Data Editor

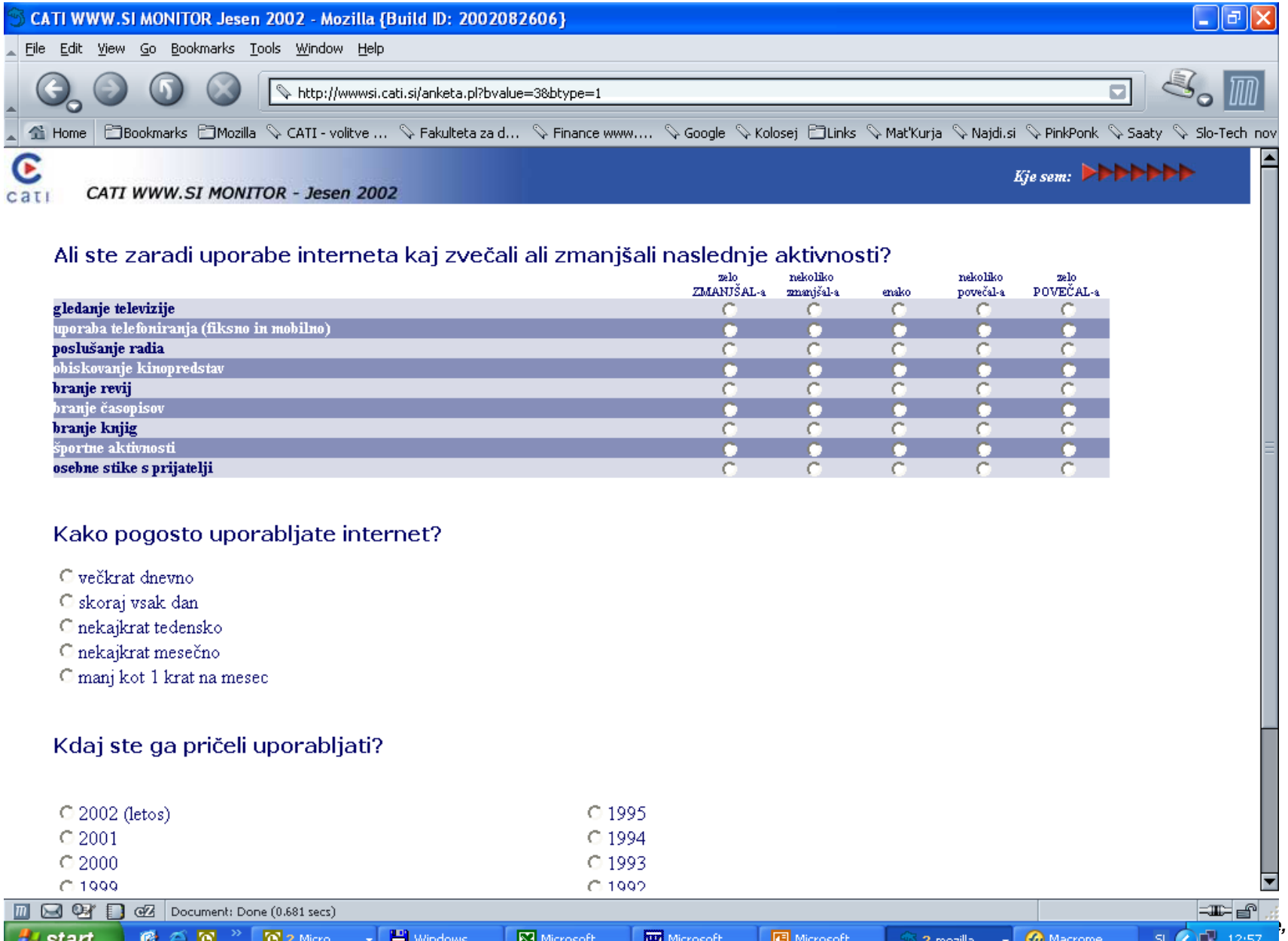
File Edit View Data Transform Analyze Graphs Utilities Window Help

159 : i.114 | 16

	sysmid	i.101	i.102	i.103	i.104	i.105	i.106	i.107	i.108	i.109	i.110	i.111	i.112	i.113	i.114	i.115
120	3000000017200405	,	,	,	1	,	,	9	9	6	10	,	,	,	,	
121	3000000018080719	,	,	,	,	,	,	,	,	,	,	,	,	1	,	
122	3000000018105502	,	,	9	43	,	28	3	6	3	4	,	,	,	,	
123	3000000018173931	,	,	,	4	,	3	2	22	3	30	,	,	,	,	
124	3000000018201511	,	,	,	,	,	,	,	6	,	,	,	,	,	,	
125	3000000018204108	,	,	,	,	,	,	,	,	,	,	,	,	,	,	
126	3000000019102257	,	,	,	1	,	,	,	,	25	,	,	,	3	,	
127	3000000019125257	,	,	,	,	,	,	,	1	,	1	,	,	,	1	
128	3000000019134327	,	,	,	,	,	,	1	3	2	1	,	,	,	,	
129	3000000019144412	,	,	,	,	,	,	,	,	,	,	,	,	,	,	
130	3000000019184505	,	,	,	,	,	,	9	1	,	2	,	,	,	,	
131	3000000020130247	,	,	,	13	,	,	,	3	4	5	,	,	,	,	
132	3000000020204902	,	,	,	,	,	,	,	,	,	,	,	,	,	,	
133	3000000021080823	,	,	,	,	,	,	,	,	,	3	,	,	,	,	
134	3000000021170424	,	,	,	5	,	,	1	1	,	3	1	,	,	,	
135	3000000021181153	,	,	,	2	,	,	,	,	,	,	,	,	,	,	
136	3000000021194026	,	,	,	,	,	,	4	3	1	1	,	,	,	,	
137	3000000022100256	,	,	2	98	,	,	26	25	17	36	,	,	,	,	
138	3000000022105808	,	,	,	,	,	,	,	,	5	,	,	,	,	,	
139	3000000022140532	,	,	,	,	,	,	1	1	,	3	,	,	1	,	
140	3000000022161605	,	,	3	20	,	,	5	8	3	4	,	,	,	,	
141	3000000022161616	,	,	5	34	,	1	1	4	10	,	,	,	6	29	
142	3000000022174112	,	,	,	2	2	,	15	20	3	11	,	1	1	,	
143	3000000022183231	,	,	,	2	,	,	,	16	,	16	,	,	,	,	
144	3000000022195421	,	,	,	,	,	,	3	,	,	,	,	,	,	,	
145	3000000022230250	,	,	3	4	,	,	,	2	2	2	,	,	,	1	
146	3000000023221137	,	,	2	3	,	,	,	,	,	2	,	,	,	,	
147	3000000024080907	,	,	20	6	,	,	15	12	,	7	,	,	,	11	
148	3000000024110758	,	,	,	11	,	,	2	4	2	5	,	,	,	,	
149	3000000024130318	,	,	1	1	,	,	4	4	,	4	,	,	,	,	
150	3000000024162022	,	,	1	,	,	,	,	,	,	,	,	,	,	,	
151	3000000024175927	,	,	,	30	2	,	1	6	1	1	1	3	,	5	
152	3000000024181224	,	,	,	2	,	,	,	,	1	,	,	,	,	,	
153	3000000025105456	,	,	,	,	,	,	,	2	,	,	,	,	,	,	
154	3000000025154000	,	,	,	4	,	,	,	3	1	3	,	,	,	,	
155	3000000026114124	,	,	,	,	,	,	2	4	1	5	,	,	,	,	
156	3000000026125540	,	,	1	,	,	,	,	2	1	3	,	,	,	,	
157	3000000026140023	,	,	,	,	,	,	,	,	,	,	,	,	,	,	

...spletna anketa (vprašalnik)

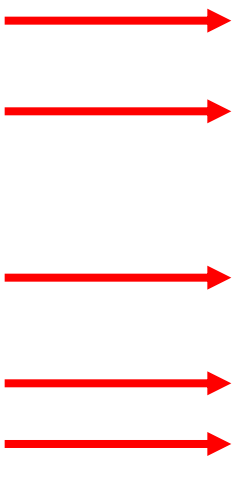
MARKETINŠKI FOKUS



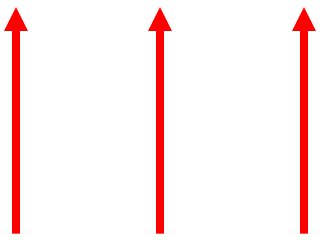
WWW.SI Monitor baze

MARKETINŠKI FOKUS

ID_cookie	P1	P2	...
253569831289	1	0	
253569831290	0	0	
253569831291	1	0	
253569831292	0	1	
253569831293	0	0	
253569831294	1	0	
253569831295	0	1	
253569831296	1	0	
253569831297	0	0	
253569831298	1	1	
253569831299	0	0	
253569831300	0	0	
253569831301	1	1	
253569831302	0	0	
253569831303	1	0	
253569831304	0	0	



ID_survey	SD1	SD2	SD3	...
253569831290	5	3	3	
253569831293	3	2	2	
253569831298	5	4	3	
253569831301	3	4	5	
253569831303	4	4	1	



SD1	SD2	SD3	...
3	2	2	
4	4	1	
5	4	3	
5	3	3	

Različni načini povezovanj baz

- “Prazne” enote izpustimo
- *Data fussion* - običajni:
  - Na podlagi podobnosti transakcijskih podatkov iščemo tako enote z demografskimi podatki, ki je najbolj podobna tistim z manjkajočimi
- *Data fussion* - napredni:
  - Na podlagi podobnosti transakcijskih podatkov z naprednimi modeli napovedujem vse prazne vrednosti spremenljivk

Reprezentativnost

- = povezana baza mora imeti enako strukturo kot:
  - rezultati transakcijske baze (npr. obiskanost strani, nakupne navade)
  - eksterni rezultati (npr. demografija)
- Rešitev:
  - Poglobljena analiza razlik z multivariatnimi metodami
  - Kompleksno uteževanje baze glede na vse vire podatkov, tj. transakcijske in eksterne
  - Metoda: *izpopolnjena metoda raking*

Kaj torej imamo?

- Za vsako opazovano spletno stran njen doseg
- Profil uporabnikov vseh spletnih strani
- Pri čemer je vzorec reprezentativen glede na obiskovanje in demografske lastnosti uporabnikov

...in BTW ...kaj je novega na Internetu?

Politično gledano...

- Investicije v spletno oglaševanje so majhne
- Morda je razlog tudi pomanjkanje podatkov?
- Raziskava uporablja edinstveno in zelo napredno metodologijo prilagojeno za male trge, ki je relativno poceni
- Spletne strani so se povezale in aktivno sodelovale pri izvedbi raziskave obiskanosti spletnih strani
- Spletne strani so primerjale rezultate s svojimi, potrdile tako metodologijo kot rezultate
- Ključno: vse strani so bile merjene enako.

Statistično gledano...

- Uporablja internet mesečno 580.000 ljudi
- Dosežejo slovenske spletne strani 450.000 slovenskih uporabnikov mesečno
- Novembra so strani zabeležile 53 mio prikazov
- Uporabniki so: mlajši aktivni ljudje z višjimi dohodki, višjo izobrazbo...
- Tri spletne strani (Najdi.si, Matkurja, SIOL) dosežejo že več kot 100.000 uporabnikov mesečno, prva dva tudi tedensko prek 100.000
- Največji spletni mediji dosežejo cca. 50.000 oseb mesečno
- Tudi najmanjše strani, ki jih večinoma niti ne poznamo dosežejo preko 2000 oseb

mesečni (28-dnevni dosegi)

ime strani	število ljudi	doseg med uporabniki interneta	doseg v populaciji	prikazov	14-dnevni doseg štev. ljudi	tedenski doseg štev.ljudi
Najdi.si	282.563	48,8%	16,7%	16.174.945	187.945	135.578
Matkurja.com	250.347	43,3%	14,8%	7.497.057	164.368	111.812
SiOL	128.439	22,2%	7,6%	3.780.394	82.654	57.210
Email.si	86.769	15,0%	5,1%	2.250.251	52.781	35.147
TIS telekom	72.902	12,6%	4,3%	1.454.526	45.029	28.097
24ur.com	70.670	12,2%	4,2%	2.320.636	43.906	28.795
Mobitel	67.948	11,7%	4,0%	1.254.471	36.261	21.735
Slowwwenia.com	58.024	10,0%	3,4%	733.501	34.290	20.820
Večer	46.955	8,1%	2,8%	1.036.801	27.781	16.667
Mobisux	45.376	7,8%	2,7%	3.038.928	26.268	16.829
Pinkponk	44.715	7,7%	2,6%	222.144	25.313	15.371
Mladina	42.608	7,4%	2,5%	631.529	23.623	13.634
Bolha.com	40.253	7,0%	2,4%	2.123.424	22.128	13.015
Buy.si	36.079	6,2%	2,1%	327.303	21.087	12.543
Strani RS	33.710	5,8%	2,0%	237.267	19.279	11.180
RTV Slovenija	32.824	5,7%	1,9%	963.350	18.904	11.390
Finance	31.278	5,4%	1,8%	1.963.030	20.414	13.549
Kolosej	30.840	5,3%	1,8%	308.398	16.805	9.854
Dnevnik.si	27.857	4,8%	1,6%	481.029	14.968	10.907
Big Bang	25.014	4,3%	1,5%	567.300	12.496	6.834
Shopolina	23.289	4,0%	1,4%	195.199	12.816	6.428
Slo-tech	23.254	4,0%	1,4%	1.053.748	13.266	8.174
Megaklik	22.750	3,9%	1,3%	389.727	12.000	6.881
Delo	22.341	3,9%	1,3%	323.275	13.137	8.358
Merkur	18.407	3,2%	1,1%	319.980	8.782	4.306
Si21.com	17.279	3,0%	1,0%	122.764	9.633	5.508
Eon.si	16.716	2,9%	1,0%	187.108	9.848	5.837
Kulinarika.net	12.687	2,2%	0,7%	324.223	7.105	4.175
Poljub	11.943	2,1%	0,7%	625.237	7.590	4.348

Splošno gledano...

- Je ta raziskava prvi in verjetno eden redkih javnih primerov aplikacije analiz velikih transakcijskih baz in povezovanja z anketnimi bazami
- Raziskava se bo še nadgradila s podatki drugih kompatibilnih raziskav (*data fussion*)

In definitivno...

... je te prezentacije konec ;)

Zenel Batagelj

Zb@cati.si

T: 2410072

http://wwwsi.cati.si