



Centrālā statistikas
pārvalde

Alternatīvo datu vākšanas rīku izmantošana Patēriņa cenu indeksu vajadzībām

Patēriņa cenu indeksu daļas vadītāja vietnieks
Dāvis Liberts

28. Septembris 2018



Centrālā statistikas
pārvalde

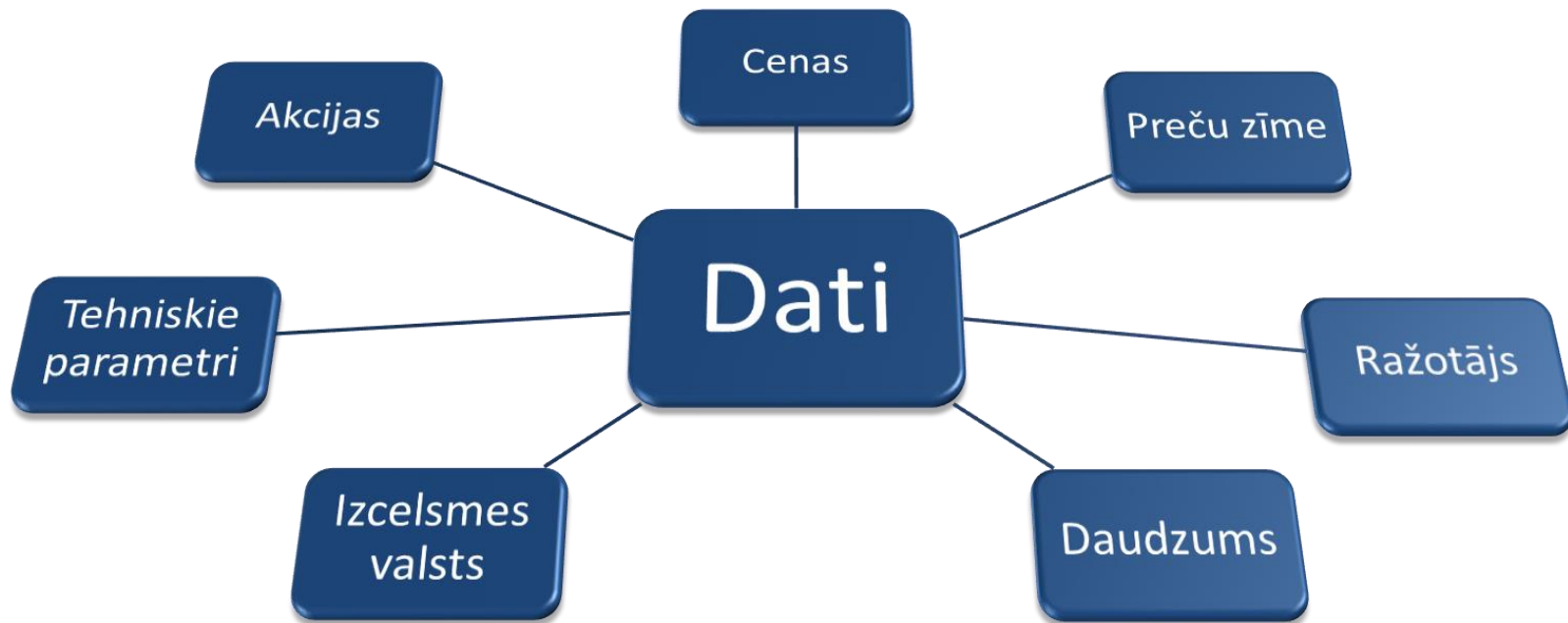
Šajā prezentācijā

- Tendences cenu vākšanā
- Kas ir automātiskā datu vākšana (*web-scraping*)
- Ko esam paveikuši līdz šim
- Turpmākie soļi un izaicinājumi



Centrālā statistikas
pārvalde

Informācija, kas nepieciešama cenu pārmaiņu atspoguļošanai





Cenu reģistrēšana ikdienā

Centrālā statistikas
pārvalde



Procesora tips	€ 832,00	€ 891,00	€ 873,00	€ 817,00	€ 832,00	
Intel Core™ i3 Intel Core™ i5 Intel Core™ i7 AMD Intel Pentium Intel Atom Intel Core™ i7 AMD A series Pārskats un salīdz.	HP OVEN 18-04013 (22M93EA) WH13P Procesors: Intel Core i3 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 832,00 €	LENOVO Y330 (81P400CFL1) Procesors: Intel Core Operatīvā atmiņa: 8 GB Cietās diska: 500 GB Cena: 891,00 €	Apple MacBook Air 13 (M0D222AL) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 128 GB SSD Cena: 873,00 €	Procesors: Intel Core i7 Operatīvā atmiņa: 8 GB Cietās diska: 256 GB SSD Cena: 817,00 €	Acer Aspire 5 Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 256 GB SSD Cena: 832,00 €	
Operatīvā atmiņa (RAM)	4 GB 8 GB 12 GB 16 GB 32 GB	ASUS V870UD-DH570T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 770,00 €	Lenovo IdeaPad Y330 (81WH0D1M4G) Procesors: Intel Core i3 Operatīvā atmiņa: 8 GB Cietās diska: 500 GB Cena: 830,00 €	HP 15-SD248EA (22C0EAAB/L) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 819,00 €	HP 320G2ES Operatīvā atmiņa: 8 GB Cietās diska: New Information Cena: 816,00 €	Lenovo S31-90 (82M01SC40) Win10 Home Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 830,99 €
Cietās diska	1 TB 800 GB 128 GB (SSD) 256 GB (SSD) 32 GB (SSD) Tas informācijai 64 GB (M.2) 2 TB (128 GB) Pārskats un salīdz.	ASUS V870UD-E083T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 829,99 €	Lenovo Yoga 830-14ICE (82B00381WH13) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 829,00 €	ASUS V870UD-E083T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 827,00 €	ASUS V870UD-E083T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 826,27 €	ASUS A516-B1G (82C0L41313) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 256 GB Cena: 806,27 €
Videokarte (tips)	Integrēta Nvidia GeForce AMD Radeon AMD Integrēta Nvidia Quadro	HP 258 G4 (F7V13) (EAKABE) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 500 GB Cena: 829,99 €	ASUS V870UD-E083T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 829,00 €	Lenovo Yoga 830-14ICE (82B00381WH13) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 827,00 €	ASUS V870UD-E083T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 826,27 €	ASUS A516-B1G (82C0L41313) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 256 GB Cena: 806,27 €
Operētājsistēma	Windows 10 Windows 10 Home DOS Windows Enclous OS Mac OS Windows 10 Pro Windows 10 Home Pārskats un salīdz.	ASUS V870UD-E083T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 829,99 €	Lenovo Yoga 830-14ICE (82B00381WH13) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 829,00 €	ASUS V870UD-E083T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 827,00 €	ASUS V870UD-E083T Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 1 TB Cena: 826,27 €	ASUS A516-B1G (82C0L41313) Procesors: Intel Core i5 Operatīvā atmiņa: 8 GB Cietās diska: 256 GB Cena: 806,27 €





Centrālā statistikas
pārvalde

Izaicinājumi cenu vākšanā

- Arvien plašāks preču klāsts veikalos
- Strauji augoša tehnoloģiju attīstība
- Nepieciešamas specifiskas tehniskās zināšanas
- Preču kvalitātes izvērtēšana un salīdzināšana

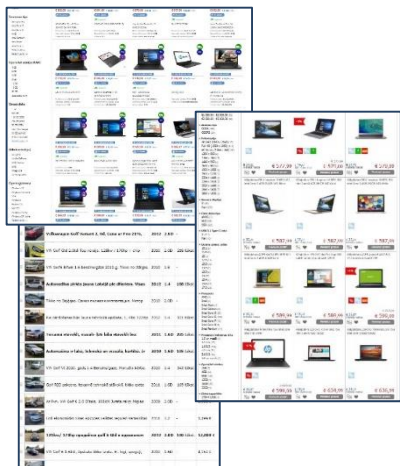


Centrālā statistikas
pārvalde

Automātiskā cenu vākšana

WEB-SCRAPING

DATU GLABĀŠANA



DATU VALIDĀCIJA

DATU ANALĪZE



Centrālā statistikas
pārvalde

Web-scraping darbības princips

Automātiski
tiek ievadīti
nepieciešam
parametri

The screenshot shows a sidebar on the left with filter categories: **Procesora tips** (Intel Celeron, Intel Core i3, Intel Core i5, AMD, Intel Pentium, Intel Atom, Intel Core i7, AMD A series, Pārēdi: vairāk...), **Operatīvā atmiņa (RAM)** (4 GB, 6 GB, 8 GB, 12 GB, 16 GB, 32 GB, 40GB DDR3 RAM), **Cietais disks** (1 TB, 500 GB, 128 GB (SSD), 256 GB (SSD), 32 GB (SSD), Nav informācijas, 64 GB (M.2), 1 TB + (128 SSD), Pārēdi: vairāk...), **Videokarte (tips)** (Integrēta, Nvidia GeForce, AMD Radeon, AMD, Integrēta, Nvidia Quadro), and **Operētājsistēma** (Windows 10, Windows 10 Home, DOS, Windows, Endless OS, Mac OS, Windows 10 Pro, Windows XP, Pārēdi: vairāk...). The main area displays a grid of laptop products with their prices and specifications. One product, an HP 3Q4EE5, is highlighted with a red box.

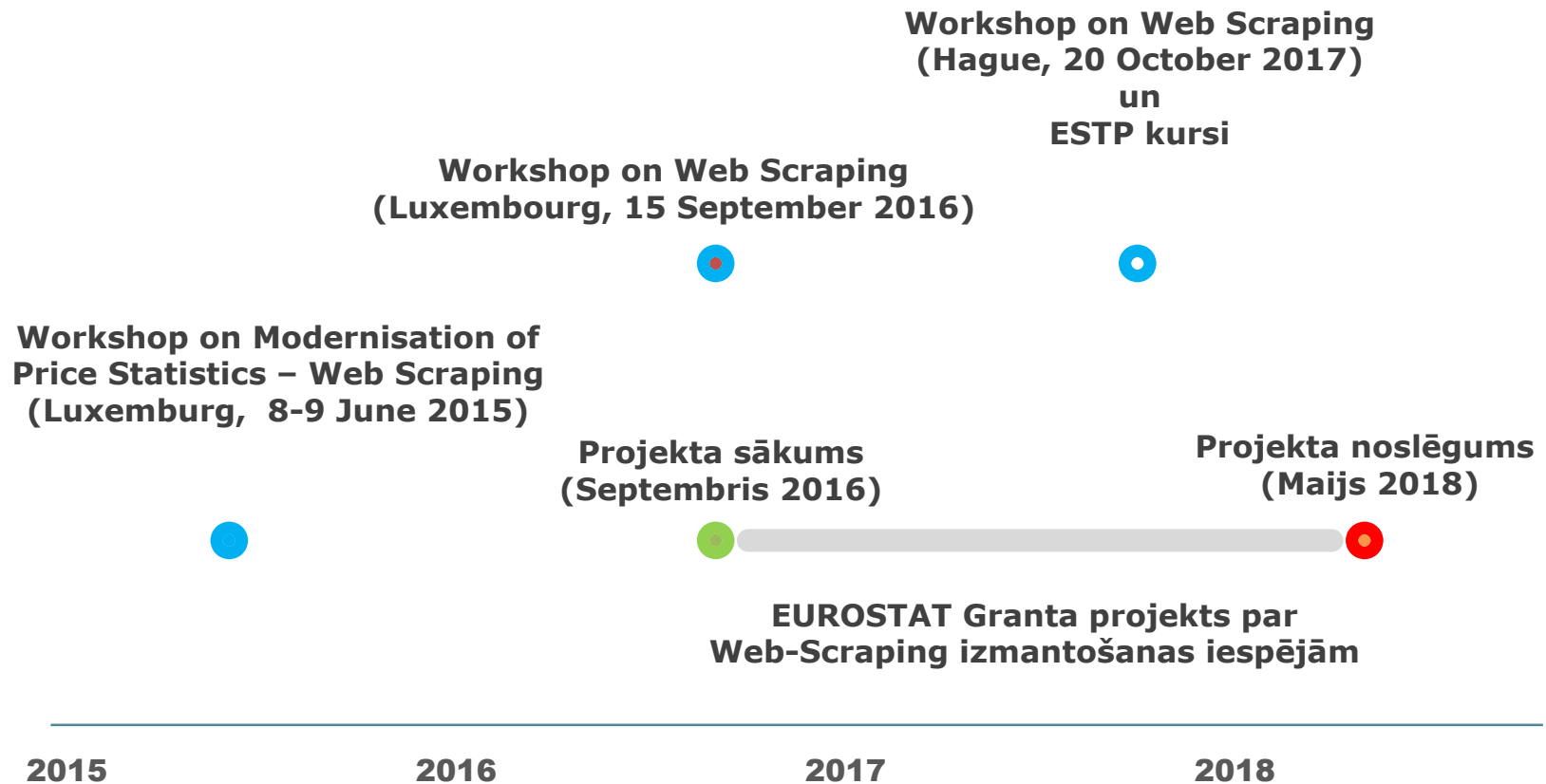
Tiek atlasīta
kāda konkrēta
prece vai preču
grupa

Atlasītie dati tiek strukturizēti un saglabāti tālākai datu analīzei



Centrālā statistikas
pārvalde

Web-scraping izpētes gaita





Centrālā statistikas
pārvalde

Līdz šim paveikts

- Dalība Eurostat Granta projektā
- Izvērtēti pieejamie *web-scraping* rīki un citu valstu pieredze
- Ar CSP IT nodaļas atbalstu izstrādāta *web-scraping* programma
(datu vākšanas un pamata analīzes funkcijas)
- Automatizēti vācam testa datus tālākai izpētei
(dzīvokļu īre, lietotas automašīnas, planšetdatori)



Centrālā statistikas pārvalde

WebScraping - 1.0.0.45

UNIT / REFRESH

CARS 3-5

Scraping

FLATS

DELETED READ VIEW WEB PAGES

Active	Description	Web_Page	Year_From	Year_To
<input checked="" type="checkbox"/>	Audi A4	/transport/cars/audi/a4/sell/	2013	2015
<input checked="" type="checkbox"/>	Audi A6	/transport/cars/audi/a6/sell/	2013	2015
<input checked="" type="checkbox"/>	BMW 320	/transport/cars/bmw/320/sell/	2013	2015
<input checked="" type="checkbox"/>	BMW X5	/transport/cars/bmw/x5/sell/	2013	2015
<input checked="" type="checkbox"/>	FIAT 500	/transport/cars/flat/500/sell/	2013	2015
<input type="checkbox"/>	Ford Fiesta	/transport/cars/ford/fiesta/sell/	2013	2015
<input type="checkbox"/>	Ford Focus	/transport/cars/ford/focus/sell/	2013	2015
<input type="checkbox"/>	Go!6	/transport/cars/volkswagen/go!6/sell/	2013	2015
<input checked="" type="checkbox"/>	Go!7	/transport/cars/volkswagen/go!7/sell/	2013	2015
<input checked="" type="checkbox"/>	Nissan Qashqai	/transport/cars/nissan/qashqai/sell/	2013	2015
<input type="checkbox"/>	Opel Insignia	/transport/cars/opel/insignia/sell/	2013	2015
<input checked="" type="checkbox"/>	Passat B7	/transport/cars/volkswagen/passat-b7/sell/	2013	2015
<input type="checkbox"/>	Polo	/transport/cars/volkswagen/polo/sell/	2013	2015
<input checked="" type="checkbox"/>	Skoda Octavia	/transport/cars/skoda/octavia/sell/	2013	2015
<input checked="" type="checkbox"/>	Toyota Auris	/transport/cars/toyota/auris/sell/	2013	2015
<input checked="" type="checkbox"/>	Toyota Rav-4	/transport/cars/toyota/rav-4/sell/	2013	2015
<input checked="" type="checkbox"/>	Toyota Yaris	/transport/cars/toyota/yaris/sell/	2013	2015
<input checked="" type="checkbox"/>	Volkswagen Passat	/transport/cars/volkswagen/passat-b5/sell/	2015	2015
<input checked="" type="checkbox"/>	Volkswagen Passat B7	/transport/cars/volkswagen/passat-b7/sell/	2013	2014
<input checked="" type="checkbox"/>	Volvo V60	/transport/cars/volvo/v60/sell/	2013	2015
<input checked="" type="checkbox"/>	Volvo XC60	/transport/cars/volvo/xc60/sell/	2013	2015

Current user: CSPDOMDLiberts | Page count: 0 | Row count: 0 | Deleted rows: 0 | Deleted duplicates: 0 | Reading time:

WebScraping

UNIT / REFRESH

CARS 9-11

Scraping

WEB DATA DELETED READ VIEW WEB PAGES

Date	Selected Count	Segments	Year
03.04.2018 12:14:11	814	Audi A4	2007
09.05.2018 9:36:23	815		
02.05.2018 13:00:23	824		
25.04.2018 10:13:28	838		
18.04.2018 9:50:03	806		
10.04.2018 13:02:42	789		
03.04.2018 12:14:11	814		
20.03.2018 15:52:42	836		
14.03.2018 9:33:45	802		
07.03.2018 11:11:42	782		
28.02.2018 10:00:27	757		
21.02.2018 9:48:24	792		
14.02.2018 10:06:15	784		
08.02.2018 11:21:05	764		
07.02.2018 11:38:36	771		
31.01.2018 9:40:13	1030		
17.01.2018 13:28:06	1073		
10.01.2018 15:29:12	930		
03.01.2018 9:31:44	1009		
20.12.2017 9:28:23	1038		
13.12.2017 11:16:02	1060		
22.11.2017 10:12:39	1051		
15.11.2017 12:34:51	1068		
07.11.2017 10:06:04	1318		
01.11.2017 9:58:34	1047		
25.10.2017 17:26:04	1060		
18.10.2017 10:27:29	1002		
04.10.2017 9:09:20	993		
27.09.2017 15:48:24	1001		
21.09.2017 9:24:28	993		

WebScraping - 1.0.0.45

UNIT / REFRESH

CARS 3-5

Scraping

WEB DATA DELETED READ VIEW WEB PAGES

Export to CSV Import DB

Marka	Modelis	Mot_tips	Nobraukums	Cena	Virs_tips	Ist_gads	Azumkriba	Tehn_apgk	Skaidrums	State	
18	Volkswagen	Passat (B7)	1.6 dīzeļis	246 000	7 290 €	Universāts	2012 februāris	Manuāla 6 ātr.	04.2019	Pārdošu vai M.	https://www.s...
19	Volkswagen	Passat (B7)	2.0 dīzeļis	264 040	8 800 €	Universāts	2012 majs	Manuāla 6 ātr.	12.2015	Volkswagen	https://www.s...
20	Volkswagen	Passat (B7)	2.0 dīzeļis	11 290 €	Universāts	2012 decembris	Automāts	09.2015	Highline 2.0L	https://www.s...	
21	Volkswagen	Passat (B7)	2.0 dīzeļis	192 000	9 595 €	Universāts	2012 aprīlis	Manuāla 6 ātr.	05.2019	Propana vārs	https://www.s...
22	Volkswagen	Passat (B7)	2.0 dīzeļis	174 000	16 890 €	Universāts	2014 marts	Automāts 6 ātr.	05.2019	Karlo Motors	https://www.s...
23	Volkswagen	Passat (B7)	2.0 dīzeļis	184 000	16 950 €	Universāts	2013	Automāts	10.2015	Karlo Motors	https://www.s...
24	Volkswagen	Passat (B7)	2.0 dīzeļis	149 000	13 450 €	Universāts	2014	Automāts	06.2015	Moller Auto p.	https://www.s...
25	Volkswagen	Passat (B7)	2.0 dīzeļis	70 800	11 300 €	Sedāns	2013 augusts	Automāts 6 ātr.	01.2019	Pārdošu VW	https://www.s...
26	Volkswagen	Passat (B7)	2.0 dīzeļis	300 600	7 800 €	Universāts	2012 jūnijs	Automāts	02.2019	Skaidrums N.	https://www.s...
27	Volkswagen	Passat (B7)	2.0 dīzeļis	219 000	9 790 €	Universāts	2011 majs	Automāts	04.2019	Highline mod.	https://www.s...
28	Volkswagen	Passat (B7)	2.0 dīzeļis	160 475	13 900 €	Universāts	2013 novembris	Automāts	05.2019	Pārdošu Motis	https://www.s...
29	Volkswagen	Passat (B7)	2.0 dīzeļis	150 000	13 900 €	Universāts	2013 septem.	Automāts	03.2019	Pārdošu Motis	https://www.s...
30	Volkswagen	Passat (B7)	1.4 benzīns	112 500	8 400 €	Sedāns	2012 februāris	Automāts 7 ātr.	03.2019	Masuma s.y.m.	https://www.s...
31	Volkswagen	Passat (B7)	1.4 benzīns	186 351	16 500 €	Sedāns	2014 aprīlis	Automāts	04.2019	Uzdevums pār.	https://www.s...
32	Volkswagen	Passat (B7)	2.0 dīzeļis	130 850	17 500 €	Universāts	2014 aprīlis	Automāts	08.2018	Volkswagen	https://www.s...
33	Volkswagen	Passat (B7)	1.4 benzīns	140 200	9 800 €	Sedāns	2013 februāris	Automāts 7 ātr.	03.2019	Tiek pārdošs	https://www.s...
34	Volkswagen	Passat (B7)	2.0 dīzeļis	260 331	8 850 €	Sedāns	2011	Automāts 6 ātr.	03.2019	Highline-Limo.	https://www.s...
35	Volkswagen	Passat (B7)	1.8 dīzeļis	189 750	8 850 €	Universāts	2013 septem.	Manuāla 6 ātr.	04.2019	VW Passat Bl.	https://www.s...
36	Volkswagen	Passat (B7)	2.0 dīzeļis	221 124	9 300 €	Universāts	2012	Manuāla 6 ātr.	04.2019	Passat 125K.	https://www.s...
37										Auto lab.taps	https://www.s...
38										Pārdošu perf.	https://www.s...
39										Pārdošu VW P.	https://www.s...
40										Pārdošu Pass.	https://www.s...
41										Automāts, Ce.	https://www.s...
42										Autorepublik.	https://www.s...
43										Six Automobile.	https://www.s...
44										Highline 2.0T.	https://www.s...
45										Volkswagen	https://www.s...
46										Takro no VWL.	https://www.s...
47										Pārdošu Pass.	https://www.s...
48										VW Passat H.	https://www.s...
49										Pārdošu Pass.	https://www.s...
50										VW Passat 2.	https://www.s...
51										VW Passat 2.	https://www.s...
52										Pārdošu vai M.	https://www.s...

Current user: CSPDOMDLiberts | Page count: 4 | Row count: 109 | Deleted rows: 0 | Deleted duplicates: 0 | Reading time: 00:00:49:976

WebScraping - 1.0.0.45

UNIT / REFRESH

CARS 9-11

Scraping

WEB DATA DELETED READ VIEW WEB PAGES

Date	Selected Count	Segments	Year
03.04.2018 12:14:11	814	Audi A6	2007

SEGMENTS

SEGMENTS	Datums	AVG_Cena	Skaitis	MIN_Cena	MAX_Cena
Audi A6	09.05.2018 09:36	6461.37	38	1000	8900
Audi A6	02.05.2018 13:00	6706.91	32	1000	8900
Audi A6	25.04.2018 10:13	6681.57	35	4990	8150
Audi A6	18.04.2018 09:50	6657.97	37	4990	8150
Audi A6	10.04.2018 13:02	6657.28	46	4990	7950
Audi A6	03.04.2018 12:14	6699.22	46	4700	7995
Audi A6	20.03.2018 15:52	6827.19	43	5350	8500
Audi A6	14.03.2018 09:33	6894.11	46	4400	8500
Audi A6	07.03.2018 11:11	6819.42	36	4990	8500
Audi A6	28.02.2018 10:00	6690.90	39	4990	8300
Audi A6	21.02.2018 09:48	6850.38	37	4995	7995
Audi A6	14.02.2018 10:06	6869.44	41	5350	7995
Audi A6	08.02.2018 11:21	6918.70	43	5550	8000
Audi A6	07.02.2018 11:38	7069.15	48	5550	9500



Centrālā statistikas
pārvalde

Priekšrocības

- *Web-scraping* ļauj iegūt vairāk datu nekā tas iespējams manuāli
- Iespēja iegūt datus pat pa stundām, dienām vai nedēļām
- Cenu vākšanas laiks:

Preču grupa	Manuāli vācot cenas	Web-scraping
3-5 g lietotas automašīnas	~ 3 stundas	~ 10 minūtes
9-11 g lietotas automašīnas	~ 2 stundas	~ 21 minūte
Dzīvokļu īres cenas	~ 5 stundas	~ 30 sekundes

- *Web-scraping* viennozīmīgi negarantē labāku datu kvalitāti, bet noteikti daudz lielāku datu apjomu
- Papildus informācija par tirgus tendencēm



Centrālā statistikas
pārvalde

Datu analīzes iespēju piemēri

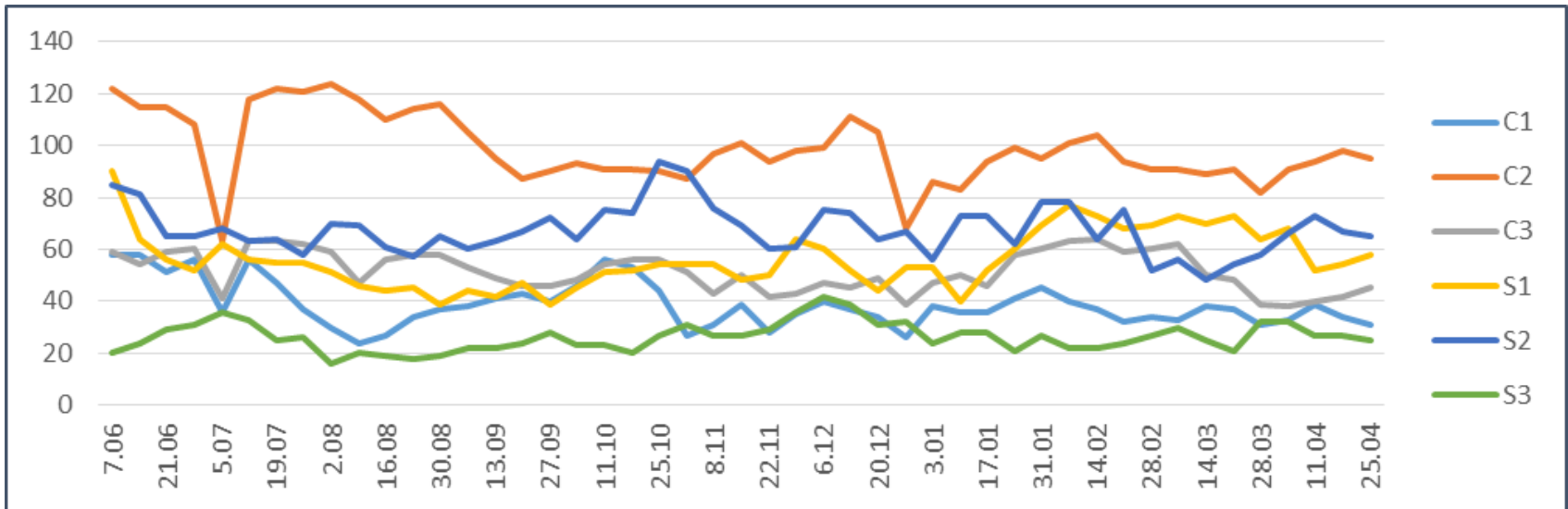


Planšetdatoru pieejamība tirdzniecības vietās



Centrālā statistikas
pārvalde

Datu analīzes iespēju piemēri



Dzīvokļu ģres sludinājumu skaits sadalījumā pa segmentiem



Centrālā statistikas
pārvalde

Izaicinājumi

- Tehnoloģiski sarežģīts datu vākšanas process
- *Web-scraping* rīks jāpielāgo katram veikalam
- Nepieciešama pastāvīga procesa uzraudzība un IT atbalsts
(tehniskas problēmas, izmaiņas mājas lapās)
- Metodoloģiski izaicinājumi
(Izlases apjoms, cenu vākšanas biežums, datu segmentēšana, datu integrēšana patēriņa cenu indeksu aprēķinā)
- Lielāks uzsvars no cenu vākšanas pāriet uz datu analīzi



Centrālā statistikas
pārvalde

Cenu reģistrēšanas process pārmaiņas

CENU VĀKŠANA MANUĀLI



Cenu reģistratori



Datu analītiķi

CENU VĀKŠANA AR WEB-SCRAPING



Pastāvīgs IT
atbalsts



Centrālā statistikas
pārvalde

Turpmākie soļi

- Turpināt datu vākšanu testa režīmā padziļinātai analīzei
- Maksimāli automatizēt *web-scraping* procesu
- Metodoloģiskie izaicinājumi
(Izlases apjoms, cenu vākšanas biežums, datu segmentēšana, datu integrēšana patēriņa cenu indeksu aprēķinā)
- Paplašināt ar *web-scraping* vākto preču un pakalpojumu cenu klāstu



Centrālā statistikas
pārvalde

Paldies par Jūsu uzmanību!

Davis.Liberts@csb.gov.lv