

Вовед во

ЧИСТЕЊЕ ПОДАТОЦИ



D.A.T.A.

Data Analytics for Transparency and Accountability

МЕТАМОРФОЗИС

Фондација за интернет и општество



This project is funded by the European Union.



ZaštoNe?



The "Accountability, Technology and Institutional Openness Network in South East Europe - ACTION SEE" project is implemented by Metamorphosis Foundation, Westminster Foundation for Democracy, CRJA – Center for Research, Transparency and Accountability, Citizens Association Why not?, Center for Democratic Transition, Open Data Kosovo (ODK) and Levizja Mjaft!.

What is a
TABLE?

HINT: IT'S NOT WHAT YOUR LAPTOP IS ON RIGHT NOW!

Неформална дефиниција

Судир на зборови

Извештаи VS. Табели

Product Sales

Sales Date	Subcategory	Product	Quantity	Sales
January 05, 2009	Accessories	Carrying Case	68	\$9,224.60
		Mini Battery Charger	44	\$1,056.00
			112	\$10,980.60
	Digital	Slim Digital	44	\$8,357.80
			44	\$8,357.80
			156	\$19,338.40
January 11, 2009	Accessories	Lens Adapter	17	\$1,147.50
			17	\$1,147.50
	Digital	Advanced Digital	39	\$7,234.50
			39	\$7,234.50
	Total		56	\$8,382.00
Total			576	\$106,920.40

Ова е извештај

C2		= Blue Sports				
A	B	C	D	E	F	G
1						
2	Company	Blue Sports				
3						
4					Data	
5	Payment method	OrderNo	Ship date	Tax rate	Items total	Amount paid
6	Cash	1083	05.10.89	0,00	11 164,80	11 164,80
7		1283	12.30.94	0,00	7 134,00	7 134,00
8	<i>Cash Sum</i>				18 298,80	18 298,80
9	Credit	1012	05.20.88	0,00	5 201,00	5 201,00
10		1057	02.19.89	0,00	1 975,00	1 975,00
11		1061	03.04.89	0,00	24 277,30	24 277,30
12		1091	06.01.89	0,00	1 950,00	1 950,00
13		1161	06.04.94	0,00	102 453,60	102 453,60
14		1212	11.14.94	0,00	3 975,75	3 975,75
15		1261	12.11.94	0,00	1 999,00	1 999,00
16	<i>Credit Sum</i>				141 831,65	141 831,65
17	MC	1168	07.04.94	0,00	104,00	104,00
18	<i>MC Sum</i>				104,00	104,00
19	Visa	1148	03.03.94	0,00	5 011,00	5 011,00
20	<i>Visa Sum</i>				5 011,00	5 011,00
21	Grand Total				165 245,45	165 245,45
22						
23						

Ова исто така е извештај

Извештаи!

Луѓето ги сакаат

- Тие се наменети да ги читаат **луѓе**
- Презентираат **консолидирани** резултати
- Вообичаено **компајлираат** податоци кои дошле од
[се надеваме] **чисти табели**
- Тие се **лоши** за **компјутерска анализа**

	A	B	C	D	E	F
1	Датум	ЕУРОДИЗЕЛ БС (Д-Е V)	ЕУРОСУПЕР БС - 95	ЕУРОСУПЕР БС - 98	Барел нафта	
2	2007	52.50	67.00	68.00	70.158	
3	2008	76.00	77.50	78.50	133.179	
4	2009	49.50	63.00	64.00	69.661	
5	2010	58.50	69.00	70.50	74.406	
6	2011	66.50	73.50	75.00	117.738	
7	2012	67.50	78.00	80.00	98.063	
8	2013	66.50	79.00	80.50	103.173	
9	2014	57.00	81	82.50	109.831	
10	2015	56.50	71.50	73.50	62.438	
11	2016	50.00	63.50	65.00	49.268	
12	2017	48.00	62.50	64.50	48.706	
13	2018	65.00	73.50	75.50	78.687	
14						
15						

Погледнете! Табела!

fx Italy						
	A	B	C	D	E	F
1	Ano	Cidade	País	Ouro	Prata	Bronze
2	1896	Atenas	Áustria	2	1	2
3	1896	Atenas	França	5	4	2
4	1896	Atenas	Hungria	2	1	3
5	1896	Atenas	Grã Bretanha	2	3	2
6	1896	Atenas	Suíça	1	2	0
7	1896	Atenas	Estados Unidos	11	7	2
8	1896	Atenas	Equipe mista	1	0	1
9	1896	Atenas	Dinamarca	1	2	3
10	1896	Atenas	Austrália	2	0	0
11	1896	Atenas	Alemanha	6	5	2
12	1896	Atenas	Grécia	10	18	19

УШТЕ ЕДНА ТАБЕЛА!

Табели

Компјутерите ги обожаваат

- Машински читливи
- Тие се наменети да „ПОМИНАТ“ низ апликациите кои анализираат податоци
- Тие овозможуваат создавање ...**ИЗВЕШТАИ!**
- Тие се **добри** за **компјутерска анализа**

Структура на табела

The image shows a screenshot of an Excel spreadsheet with a table structure. The table has 6 columns (A-F) and 15 rows (1-15). The data is as follows:

	A	B	C	D	E	F
1	Датум	ЕУРОДИЗЕЛ БС (Д-Е V)	ЕУРОСУПЕР БС - 95	ЕУРОСУПЕР БС - 98	Барел нафта	
2	2007	52.50	67.00	68.00	70.158	
3	2008	76.00	77.50	78.50	133.179	
4	2009	49.50	63.00	64.00	69.661	
5	2010	58.50	69.00	70.50	74.406	
6	2011	66.50	73.50	75.00	117.738	
7	2012	67.50	78.00	80.00	98.063	
8	2013	66.50	79.00	80.50	103.173	
9	2014	57.00	81	82.50	109.831	
10	2015	56.50	71.50	73.50	62.438	
11	2016	50.00	63.50	65.00	49.268	
12	2017	48.00	62.50	64.50	48.706	
13	2018	65.00	73.50	75.50	78.687	
14						
15						

Red arrows point to columns B, C, D, and E. Blue arrows point to rows 2 through 13.

Структура на табела

	A	B	C	D	E	F
1	Ano	Cidade	País	Ouro	Prata	Bronze
2	1896	Atenas	Áustria	2	1	2
3	1896	Atenas	França	5	4	2
4	1896	Atenas	Hungria	2	1	3
5	1896	Atenas	Grã Bretanha	2	3	2
6	1896	Atenas	Suíça	1	2	0
7	1896	Atenas	Estados Unidos	11	7	2
8	1896	Atenas	Equipe mista	1	0	1
9	1896	Atenas	Dinamarca	1	2	3
10	1896	Atenas	Austrália	2	0	0
11	1896	Atenas	Alemanha	6	5	2
12	1896	Atenas	Grécia	10	18	19

Примарен клуч!

Колони

Структура на табела

	A	B	C	D	E	F
1	Ano	Cidade	País	Ouro	Prata	Bronze
2	1896	Atenas	Áustria	2	1	2
3	1896	Atenas	França	5	4	2
4	1896	Atenas	Hungria	2	1	3
5	1896	Atenas	Grã Bretanha	2	3	2
6	1896	Atenas	Suíça	1	2	0
7	1896	Atenas	Estados Unidos	11	7	2
8	1896	Atenas	Equipe mista	1	0	1
9	1896	Atenas	Dinamarca	1	2	3
10	1896	Atenas	Austrália	2	0	0
11	1896	Atenas	Alemanha	6	5	2
12	1896	Atenas	Grécia	10	18	19

Колона: Година
Ред: 4
Вид: Датум (Година)

Колона: Држава
Ред: 7
Вид: Текст

Интегритет и предвидливост

	A	B	C	D	E
1	PROGRAMADO Nº VOOS		ATRASADOS (% ATRASADOS Nº		CANCELADOS (%)
2	TOTALIS	1534	44 (2.9 %)	3 (0.2 %)	85 (5.5 %)
3	Altamira-PA	7	0 (0 %)	0 (0 %)	0 (0 %)
4	Aracaju-SE	11	2 (18.2 %)	0 (0 %)	1 (9.1 %)
5	Belém-PA	39	0 (0 %)	0 (0 %)	0 (0 %)
6	Belo Horizonte -	8	0 (0 %)	0 (0 %)	0 (0 %)
7	Belo Horizonte -	104	5 (4.8 %)	0 (0 %)	10 (9.6 %)
8	Boa Vista-RR	4	0 (0 %)	0 (0 %)	0 (0 %)
9	Brasília-DF	129	5 (3.9 %)	0 (0 %)	10 (7.8 %)
10	Campina Grande	1	0 (0 %)	0 (0 %)	0 (0 %)
11	Campinas-SP	117	0 (0 %)	0 (0 %)	20 (17.1 %)
12	Campo Grande-MS	19	0 (0 %)	0 (0 %)	0 (0 %)
13	Campos dos Go	1	0 (0 %)	0 (0 %)	0 (0 %)
14	Carajás-PA	2	0 (0 %)	0 (0 %)	0 (0 %)
15	Corumbá-MS	1	0 (0 %)	0 (0 %)	0 (0 %)
16	Criciúma-SC	2	0 (0 %)	0 (0 %)	0 (0 %)
17	Cruzeiro Do Sul-	1	0 (0 %)	0 (0 %)	0 (0 %)
18	Cuiabá-MT	38	1 (2.6 %)	0 (0 %)	0 (0 %)
19	Curitiba-PR	67	1 (1.5 %)	0 (0 %)	1 (1.5 %)
20	Florianópolis-SC	31	1 (3.2 %)	0 (0 %)	3 (9.7 %)
21	Fortaleza-CE	50	2 (4 %)	0 (0 %)	6 (12 %)
22	Foz Do Iguaçu-P	16	0 (0 %)	0 (0 %)	0 (0 %)
23	Goiânia-GO	31	1 (3.2 %)	0 (0 %)	0 (0 %)

Вредност: Campinas-SP
Колона: Град или Држава?

Апсолутна вредност или проценти?

А што со заградите?

И празните места?

Интегритет и предвидливост

Овие две работи секогаш треба да доаѓаат во пакет

- Секоја колона треба да има **само еден** вид на **податоци**
 - Текст, бројки, дати, нули, координати...
- Идентични линии треба да се консолидираат во една
- Истиот ентитет треба да биде внесен прецизно: “Врата” не е “ Врата”
- Податоците во истиот ред припаѓаат на истиот регистар

Чести формати

Всушност, се е варијација на тема

- CSV
 - Comma (или "Character") Separated Values

Comma, separated, values...or any other delimiter (TSV, TXT etc)!

Чести формати

Краток опис

- Колони се разделени од раздвојувач
- Линиите се разделени од линиски раздвојувач (кога притискате на “enter”)
- Податоците во првиот ред вообичаено ги содржат имињата на колоните
- Нема ограничување во бројот на регистри!

	A	B	C	D	E	F
1	Датум	ЕУРОДИЗЕЛ БС (Д-Е V)	ЕУРОСУПЕР БС - 95	ЕУРОСУПЕР БС - 98	Барел нафта	
2	2007	52.50	67.00	68.00	70.158	
3	2008	76.00	77.50	78.50	133.179	
4	2009	49.50	63.00	64.00	69.661	
5	2010	58.50	69.00	70.50	74.406	
6	2011	66.50	73.50	75.00	117.738	
7	2012	67.50	78.00	80.00	98.063	
8	2013	66.50	79.00	80.50	103.173	
9	2014	57.00	81	82.50	109.831	
10	2015	56.50	71.50	73.50	62.438	
11	2016	50.00	63.50	65.00	49.268	
12	2017	48.00	62.50	64.50	48.706	
13	2018	65.00	73.50	75.50	78.687	
14						
15						

Создавање на CSV

Многу лесно

Датум,ЕУРОДИЗЕЛ БС (Д-Е V),ЕУРОСУПЕР БС - 95,ЕУРОСУПЕР БС - 98,Барел нафта
2007,52.50,67.00,68.00,70.158

Датум;ЕУРОДИЗЕЛ БС (Д-Е V);ЕУРОСУПЕР БС – 95;ЕУРОСУПЕР БС – 98;Барел нафта
2007;52.50;67.00;68.00;70.158

Создавање на CSV

Многу лесно

```
Датум|ЕУРОДИЗЕЛ БС (Д-Е V)|ЕУРОСУПЕР БС – 95|ЕУРОСУПЕР БС – 98|Барел нафта  
2007|52.50|67.00|68.00|70.158
```

Датум	ЕУРОДИЗЕЛ БС (Д-Е V)	ЕУРОСУПЕР БС – 95	ЕУРОСУПЕР БС – 98	Барел нафта
2007	52.50	67.00	68.00	70.158

Најчестите разделувачи

Можете да користите и други ако
НАВИСТИНА сакате

,

;

|

“tab”

...но има и други!

За крај

Работите не се секогаш црно-бели

- Нема точен и погрешен начин, само најсоодветен
 - **Извештаите** се за луѓе
 - На **машините** им требаат **табели!**
- Постојат многу недоразбирања и несогласувања!
- Останете позитивни!
- Бидете ригорозни кога чистите!

Неколку алатки за чистење податоци

Бесплатни се!

- [Google Sheets](#)
- [Data Wrangler](#)
- [Open Refine](#)

СО СРЕЌА ЧИСТЕЊЕ!



D.A.T.A.

Data Analytics for Transparency and Accountability

МЕТАМОРФОЗИС 

Фондација за интернет и општество



This project is funded by the European Union.



ZaštoNe?



The "Accountability, Technology and Institutional Openness Network in South East Europe - ACTION SEE" project is implemented by Metamorphosis Foundation, Westminster Foundation for Democracy, CRJA – Center for Research, Transparency and Accountability, Citizens Association Why not?, Center for Democratic Transition, Open Data Kosovo (ODK) and Levizija Mjajt!.