



StatSoft

# Poznejte nástroj Reportovací tabulky

Reportovací tabulky jsou zbrusu novým nástrojem verze 12 a prodávají se jako samostatný modul. Cílem tohoto článku je seznámit Vás s tímto úžasným nástrojem a pomoci Vám zvládnout jeho ovládání v takovém rozsahu, abyste byli schopni ho ihned začít používat.

Reportovací tabulky jsou komplexním interaktivním nástrojem pro tvorbu pokročilých tabulkových výstupů. Jde o výpočet popisných charakteristik pro kategorie a kombinace kategorií. Tyto výsledky jsou výraznou nástavbou a rozšířením klasických kontingenčních tabulek. To, co jste v minulosti museli počítat přes mezivýsledky a nejrůznější operace s daty můžete nyní udělat přímo. Nejlépe si vše představit na následující motivaci.

		Credit Rating	
		bad	good
Balance of Cu	<= \$300	22%	78%
	>\$300	12%	88%
	no balance	39%	61%
	no running acc	49%	51%
Total w/o Missing		30%	70%
Payment of Pr	hesitant	63%	38%
	no previous cri	32%	68%
	no problems w paid back	32%	68%
	problematic ru	17%	83%
Total w/ Missing		57%	43%
Total w/ Missing		30%	70%
Purpose of Cre	business	42%	58%
	furniture	22%	78%
	household app	36%	64%
	new car	17%	83%
	other	38%	62%
	repair	44%	56%
	retraining	35%	65%
television	33%	67%	
used car	32%	68%	
vacation	11%	89%	
Total w/ Missing		30%	70%

## Motivace

Abyste si hned na začátek udělali představu, co si máte pod pojmem reportovací tabulky představit, ukážeme naprosto triviální příklad. Máme data, kde máme 4 proměnné (rok, měsíc, den a výdělek), může tedy jít o data o prodejkách či obzatech firmy, přeneseně také třeba o počty návštěv na Vašem webu, počty reklamací, počty návštěvníků obchodu, atd.

	1	2	3	4
	rok	měsíc	den	Výdělek
1	2011	1	14	649
2	2011	1	15	930
3	2011	1	17	1091
4	2011	1	19	485
5	2011	1	22	1490
6	2011	1	28	733
7	2011	2	2	87
8	2011	2	2	1470
9	2011	2	2	1377
10	2011	2	3	390
11	2011	2	11	482
12	2011	2	12	129
13	2011	2	13	1164

Máme tedy takováto data a chtěli bychom vytvořit jednoduchou tabulku, kde bude souhrn výdělků za jednotlivé roky a měsíce a to jak průměrný výdělek, tak i celkový, jakož i počet položek v daném měsíci. Výsledkem je tedy například tabulka na další straně (vlevo nahoře) vzniklá právě použitím modulu Reportovací tabulky. Nutno poznamenat, že toto je pouze rychlý vlastně defaultně nastavený výstup. Pokud si člověk trochu pohraje s formátováním a nastavením, může být výsledek podobný například tabulce vpravo nahoře na další straně.

				Průměrný výdělek	Celkový výdělek	Počet položek	
rok	2011	měsíc	1	896,33	5378,00	6	
			2	855,58	10267,00	12	
			3	854,83	5129,00	6	
			4	1213,80	6069,00	5	
			5	968,80	4844,00	5	
			6	751,20	7512,00	10	
			7	769,40	3847,00	5	
			8	1227,20	6136,00	5	
			9	742,00	3710,00	5	
			10	841,10	8411,00	10	
			11	1311,60	19674,00	15	
			12	870,30	17406,00	20	
	<b>Total w/ Missing</b>				<b>945,99</b>	<b>98383,00</b>	<b>104</b>
	2012	měsíc	1	1046,33	9417,00	9	
			2	1033,12	17563,00	17	
			3	635,63	5085,00	8	
			4	979,10	9791,00	10	
			5	1010,91	11120,00	11	
			6	1119,45	22389,00	20	
			7	1033,70	10337,00	10	
			8	1111,10	11111,00	10	
			9	877,50	8775,00	10	
			10	934,88	15893,00	17	
			11	852,92	21323,00	25	
			12	805,71	22560,00	28	
	<b>Total w/ Missing</b>				<b>944,94</b>	<b>165364,00</b>	<b>175</b>
	2013	měsíc	1	1325,50	2651,00	2	
			2	361,25	1445,00	4	
			3	812,00	1624,00	2	
			4	687,00	1374,00	2	
			5	768,50	1537,00	2	
			6	694,50	2778,00	4	
			7	847,50	1695,00	2	
			8	882,00	1764,00	2	
			9	1007,50	2015,00	2	
	<b>Total w/ Missing</b>				<b>767,41</b>	<b>16883,00</b>	<b>22</b>
<b>Total w/ Missing</b>				<b>932,33</b>	<b>280630,00</b>	<b>301</b>	

				Průměrný výdělek	Celkový výdělek	Počet položek
2011	měsíc	1	896,33	5 378	6	
		2	855,58	10 267	12	
		3	854,83	5 129	6	
		4	1213,80	6 069	5	
		5	968,80	4 844	5	
		6	751,20	7 512	10	
		7	769,40	3 847	5	
		8	1227,20	6 136	5	
		9	742,00	3 710	5	
		10	841,10	8 411	10	
		11	1311,60	19 674	15	
		12	870,30	17 406	20	
<b>Celkově</b>				<b>945,99</b>	<b>98 383</b>	<b>104</b>
2012	měsíc	1	1046,33	9 417	9	
		2	1033,12	17 563	17	
		3	635,63	5 085	8	
		4	979,10	9 791	10	
		5	1010,91	11 120	11	
		6	1119,45	22 389	20	
		7	1033,70	10 337	10	
		8	1111,10	11 111	10	
		9	877,50	8 775	10	
		10	934,88	15 893	17	
		11	852,92	21 323	25	
		12	805,71	22 560	28	
<b>Celkově</b>				<b>944,94</b>	<b>165 364</b>	<b>175</b>
2013	měsíc	1	1325,50	2 651	2	
		2	361,25	1 445	4	
		3	812,00	1 624	2	
		4	687,00	1 374	2	
		5	768,50	1 537	2	
		6	694,50	2 778	4	
		7	847,50	1 695	2	
		8	882,00	1 764	2	
		9	1007,50	2 015	2	
<b>Celkově</b>				<b>767,41</b>	<b>16 883</b>	<b>22</b>
<b>Celkově</b>				<b>932,33</b>	<b>280 630</b>	<b>301</b>

## Kde reportovací tabulky najdete?

Funkcionalitu reportovacích tabulek najdete v zobrazení typu *Pás karet*, v záložce *Data*. V klasickém rozhraní tlačítko pro spuštění reportovacích tabulek zatím nenajdete.

Je potřeba poznamenat, že modul Reportovací tabulky je k dispozici i v [trial verzi programu STATISTICA 12](#), nic Vám tedy nebrání si nové funkcionality vyzkoušet na vlastní kůži.

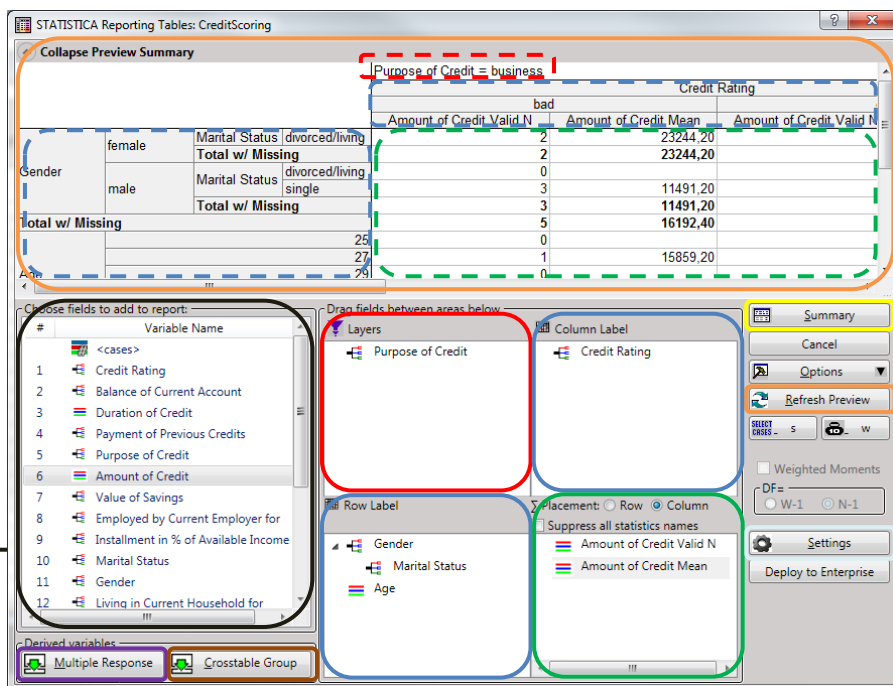
## Popis dialogového okna

Dialogové okno reportovacích tabulek je mírně odlišné od ostatních, na které jste zatím zvyklí. Rozdílnost je v jeho interaktivitě a také v tom, že je vše vidět vlastně najednou v jednom dialogu. Pojdme si vysvětlit, co jednotlivé části dialogového okna znamenají. Pro tyto účely si otevřeme bohatší a zajímavější datový soubor *CreditScoring.sta* (soubor v příkladech programu *STATISTICA*, *Soubor->Otevřít příklady->Datasets*).

**Náhled** – po zmáčknutí tlačítka *Refresh Preview* se objeví aktuální náhled výsledné tabulky, aniž by tabulka byla opravdu vyexportována jako nový soubor. Náhled je zde hlavně pro vizuální kontrolu, jestli je vše nastaveno tak, jak má.

**Souhrn** – vytvoří tabulku či tabulky (pokud je zvoleno nějaké dělení podle vrstev) ve formátu tabulky *STATISTICA*. Dialogové okno zůstane minimalizováno dole na liště

Jsou zde vyobrazeny proměnné, které máme k dispozici v aktuální datové tabulce, s kterou modul počítá. Je možné je přetáhnout do oken vpravo přetažením (metodou drž a puš tam, kde potřebuješ), přetažené proměnné budou použity v analýze. Po dvojkliku na proměnnou (v této oblasti) je pro možnosti výstupu možné ubírat a přidávat kategorie a měnit jejich pořadí (toto pořadí je zohledněno, pokud se má později proměnná řadit podle uživatelem definovaného pořadí).



**Dělení v řádcích/sloupcích** – každý řádek/sloupec výstupu bude obsahovat jednu kategorii podle zvolených grupovacích proměnných. Je možné řazení pod sebe i hierarchicky – obojí lze tvořit přetahováním proměnných. Rozdíl mezi hierarchickým řazením a řazením pod sebe je zřejmý při pohledu na nastavení příkladu a náhled tabulky. Dvojklik na proměnnou otevírá možnosti formátování a vůbec nastavení vzhledu pro oblasti, které tato proměnná ovlivňuje.

**Volba vrstev** – podle proměnných (a jejich kategorií) v oblasti pro vrstvy se soubor rozdělí a na každém rozdeleném kuse se spočte tabulka na základě ostatních nastavených voleb. Výsledkem je tedy série tabulek a ne jen jedna tabulka – ta by vznikla, kdyby do okna vrstev nebyla přetažena žádná proměnná.

**Okno statistik** – zde si volíme, co chceme v tabulce za výsledné statistiky (dvojklikem vyvoláme dialog, ve kterém je možné volit charakteristiku, stejně jako formátování výsledných hodnot ve výstupu). Statistika se spočítá pro každou kategorii určenou řádky, sloupci i vrstvami zvlášť.

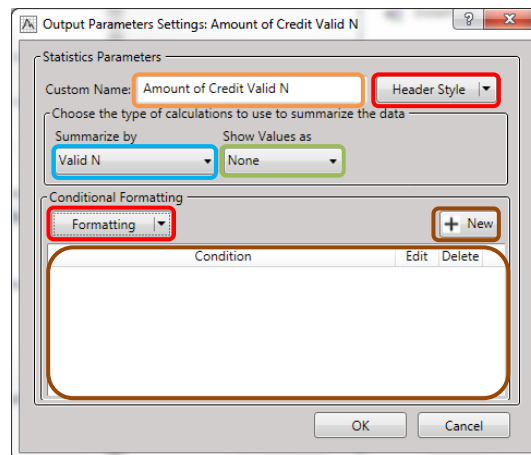
Tlačítko **Mnohonásobné odpovědi** – je jím možné vytvořit proměnnou s mnohonásobnými odpověďmi ze stávajících proměnných. Tuto proměnnou lze poté použít dále. Soubor na vyzkoušení může být např. soubor z našeho [článku o Sportce](#) nebo [Fastfood.sta](#) v příkladech programu, zde si můžete vytvořit proměnnou pro jídla, auta i burgery.

**Crosstable Group** – po výběru proměnných z nich vytvoří dvě nové proměnné: jedna představuje proměnné, které byly vybrány a druhá sjednocené skupiny vyskytující se ve všech vybraných proměnných dohromady, pokud chceme tuto druhou proměnnou využívat, je potřeba vybrat proměnné, u kterých dává sjednocení smysl (nejlépe tedy proměnné se stejnými kategoriemi). Sjednocení probíhá na základě numerických hodnot, ne textových popisků. Nejčastější využití obou proměnných je vložení proměnné s proměnnými do řádků a proměnné s úrovněmi do sloupců. Vhodným souborem na vyzkoušení je [Adstudy.sta](#) v příkladech, kde si sloučíte jednotlivá měření.

**Nastavení** – Jedná se o nastavení parametrů charakteristik, které se počítají, například jde o hladinu spolehlivosti pro interval spolehlivosti, procento useknutí pro useknutý průměr,... Také je možné zde napsat znění infoboxu v pravém horním rohu výsledné tabulky.

Pro volbu požadovaného výpočtu (použité statistiky) je potřeba pokliktat na veličinu v okně statistik, objeví se dialogové okno vpravo. Zde si lze přepsat název proměnné, formátování buněk a příslušné hlavičky souvisejících s touto charakteristikou, stejně tak podmíněné formátování (o něm bude řeč později). Nejdůležitější je ale výběr počítané charakteristiky a unikátní možnost zobrazovat výsledky jako procenta podle vybrané volby. Následuje široký přehled statistik a možnosti z čeho se dají počítat procenta.

Valid N	Low CI Limit for SD	% Row Total
Sum of Weights	Median	% Column Total
Sum	Mode	% Page Total
Mean	Percentile 1	% Grand Total
Std. Deviation	Percentile 2	
Min	Quartile 25	
Max	Quartile 75	
Coefficient of Variation	Range	
Geometric Mean	Skewness	
Grubbs P Value	Std. Error of Kurtosis	
Grubbs Test	Std. Error of Mean	
Harmonic Mean	Std. Error of Skewness	
High CI Limit for Mean	Trimmed Mean	
High CI Limit for SD	Valid Observations	
Kurtosis	Variance	
Low CI Limit for Mean	Winsorized Mean	



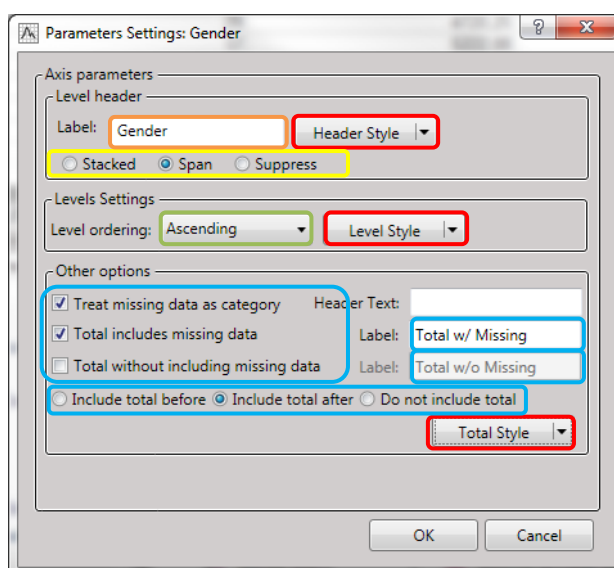
## Využití

- Reportovací tabulky mají všechny potřebné předpoklady, aby splnily funkci základního nástroje pro prvotní pohled na data.
- Zároveň si jimi můžete výrazně zjednodušit práci při vytváření klidně i jednoduchých „kontingenčních“ tabulek obsahujících průměry či jakékoli jiné charakteristiky.
- Jsou významným pomocníkem, pokud potřebujeme sumarizovat velké množství dat v jednoduše srozumitelném formátu.
- Mohou výhodně posloužit i jako mezivýsledek (jako nástroj pro úpravu a agregaci dat), jejichž použití je možné v následných analýzách. Výhodně se dá také využít výsledek ve formě procentuálních charakteristik.
- S použitím maker a nastavením formátování lze vytvářet automatické reporty s minimálním úsilím, především pokud mají být reporty pravidelné a vždy se stejnou strukturou.

## Vzhled

Reportovací tabulky jsou nástrojem, jehož možnosti zdaleka nemusí končit u vytvoření tabulky výsledků, v modulu je možné přímo nastavit různé formátování a spolu s ním i co chceme zobrazit a jak. Dále lze například nastavit, jak se mají jmenovat sloupce či celkové souhrny. Na všechny proměnné v řádcích i sloupcích dialogu reportovací tabulky lze pokliktat a zobrazit okno možností vpravo:

Zde lze nastavit název a formátování hlavičky, úroveň i souhrnu, zvolit pořadí faktorů, skrýt či zobrazit hlavičku, určit vlastnosti souhrnu, určit mu popis, či jej vůbec nezobrazovat.



Výsledkem může být tedy například barevná tabulka v úvodním příkladu.

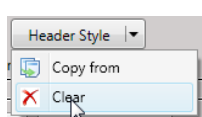
Nebudeme zde již zabíhat do dalších detailů, voleb a kombinací nastavení pro formátování výstupu je nepřehledné množství a dovolí Vám vytvořit tabulky přesně podle Vašich představ. Jediné, co bychom Vám ještě rádi ukázali je několik rad a triků.

### Doporučení a triky:

- Smísit** Při každém výběru podkladové barvy je možné vybrat možnost pomocí zaškrtnutí **Smísit (Blend)**, která dovolí prolínání barev, viz obrázek:

8	14	22
15	13	28
23	27	50

- Pozor!** V dialogích Reportovacích tabulek jsou vždy 3 polohy zaškrtnutí: **zapnuto, default, vypnuto**. Někdy tedy pro vypnutí nestačí kliknout jen jednou!



Některá nastavení formátu výstupu (například ohraničení) jsou k dispozici v mnoha dialogích. Některé se vzájemně vylučují a nemohou mít efekt zároveň (například ohraničení v hlavičce pro sloupec a zároveň ohraničení v hlavičce pro řádky). V tomto případě se hodí vědět, že celé nastavení tlačítka se spolehlivě vymaže klepnutím na **Clear** u daného tlačítka.

Obecně doporučujeme volit nastavení ohraničení (hraničních čar) jen v jednom nastavení. Pokud je to možné vyvarujte se používání ohraničení ve více místech zároveň.

Použití ohraničení i jen částečného u hlaviček Vám výrazně sníží další možnosti, některá další ohraničení nemají totiž poté efekt.

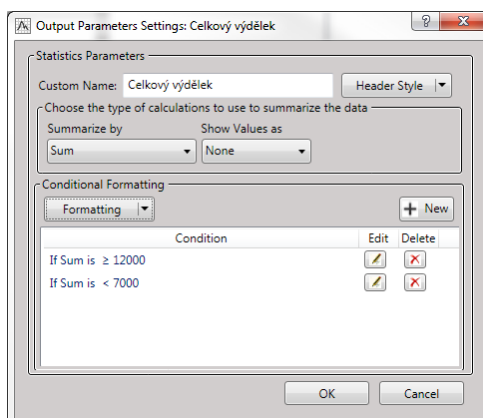
Ohraničení pro úrovně a souhrny je možné použít dohromady. Pokud použijeme ohraničení u hlavičky, souhrn i úrovně již nebudou mít efekt.

Pokud je aktivních více nastavení, je zde hierarchie, kde hlavní jsou nastavení sloupce, pak nastavení statistik a nakonec řádků.

Doporučujeme si vždy při vytvoření složitější tabulky uložit také makro (o tom, jak to udělat, bude řeč níže), umožní Vám to totiž později vyvolat dialog Reportovacích tabulek se všemi nastaveními, poté můžete pokračovat dále v práci.

## Podmíněné formátování

Modul nabízí i jedinečnou možnost podmíněného formátování. To znamená, že daná hodnota změní svůj vzhled podle nastaveného formátování a podmínek. Například pro příklad z úvodu vidíme, že pro sloupec Celkový výdělek jsou nastaveny dvě podmínky, které obarvují výstup. Dialog jsme vyvolali dvojklikem na proměnnou **Celkový výdělek**. Položky jsme přidali tlačítkem **New** a formátování jsme nastavili v dialogu vyvolaném tlačítkem **Edit** u příslušné podmínky.



Celkový výdělek
5 378
10 267
5 129
6 069
4 844
7 512
3 847
6 136
3 710
8 411
19 674

## Nahrávání maker

Stejně jako i u ostatních analýz programu *STATISTICA*, i zde je možné jednoduše získat makro, které bude vytvářet Vámi nadefinovanou tabulku jedním stisknutím tlačítka bez jakéhokoli nového nastavování. Pokud potřebujete dělat reporty a tabulky pravidelně, pak je toto volba pro Vás. Makro právě nastavené analýzy reportovací tabulky najdete zde:

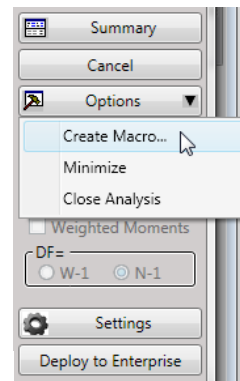
Pokud již máte otevřené makro a potřebovali byste vyvolat dialog s aktuálním nahraným nastavením, je to možné, je ale potřeba do zdrojového kódu makra přidat příkaz:

```
newanalysis.Visible = True
```

A to nejlépe před řádek `End Sub`:

```
oReportingTable.CaseWeight.Variable = 0
oReportingTable.WeightedMoments = False
oReportingTable.DfEqualsWMinus1 = False

newanalysis.RouteOutput(oReportingTable.Summary).Visible = True
newanalysis.Visible = True
End Sub
```



## Shrnutí

Na závěr si shrňme, co všechno nový modul dokáže. *STATISTICA Reportovací tabulky* přináší tyto funkcionality:

- vypočítat výsledky pro skupiny kategorizované podle různých proměnných sloupcově, řádkově i ve vrstvách
- možnost kategorizování podle mnoha proměnných najednou, tyto proměnné mohou být navíc do sebe hierarchicky vnořeny
- velké množství dostupných statistik pro reportování (jde o všechny charakteristiky přístupné v popisných statistikách softwaru *STATISTICA* a některé další), je možné také reportovat v jedné tabulce více charakteristik najednou (například průměr spolu s useknutým průměrem a mediánem)
- možnost plné kontroly nad formátováním výstupu (ohraničení, barvy a velikosti písma, pozadí, prolínání barev)
- výstupy jsou ve formě tabulky formátu .sta, jsou tedy použitelné jako vstup pro následné analýzy
- možnost podmíněného formátování - je možné nadefinovat pravidla pro formátování buňky na základě výsledné hodnoty v dané buňce
- během tvoření tabulky je možné generovat si náhled, finální výsledky ve formě tabulky lze vygenerovat stiskem tlačítka
- podpora dat typu mnohonásobných odpovědí
- možnost nahrávání makra pro pravidelné vytváření analogických reportů
- je k vyzkoušení v [trial verzi](#)

		Duration of Credit Mean	Duration of Credit Median
Gender	female	12,00	12,00
	male	24,00	24,00
Marital Status	divorced/living apart	24,00	24,00
	divorced/living apart/married	12,00	12,00
	married/widowed	12,00	12,00
	single	24,00	24,00
Balance of Current Account	<= \$300	12,00	12,00
	>\$300	19,25	11,00
	no balance	20,25	21,00
	no running account	21,70	21,00
	2-4	12,00	12,00

		Dolní kvartil	Medián	Horní kvartil
Přisun tabáku	malý	nizka	120	128
		středni	124	130
	středni	nizka	126	134
		středni	122	130
Souhrn	nizka	122	128	
	středni	124	134	
	středni	127	138	
	středni	124	134	
Souhrn	nizka	126	134	
	středni	122	140	
	středni	122	140	
	středni	122	140	

Levels of MEASURES										
	0	1	2	3	4	5	6	7	8	9
MEASURE01	1	2	3	3	2	7	8	14	1	9
MEASURE02	4	5	8	3	4	5	6	5	5	5
MEASURE03	4	4	11	6	3	3	7	5	4	3
MEASURE04	1	3	5	4	6	5	2	8	10	6
MEASURE05	5	8	5	4	3	10	5	5	4	1

Věříme, že budete z Reportovacích tabulek nadšení stejně, jako jsme my. V případě jakýchkoli dotazů nás neváhejte kontaktovat emailem na adresu [support@statsoft.cz](mailto:support@statsoft.cz).