

# Štatistiky dátovej kvality vo verejnej správe

*Projekt*

**Zlepšenie využívania údajov vo verejnej správe**

Zmluva o dielo č. 321/2018

**Výstup č.2**

## Zoznam skratiek

RU	Register úpadcov
RA	Register adries
RPO	Register právnických osôb
N/A	Not Applicable - Neaplikovateľné
KPI	Key Performance Indicator – kľúčový ukazovateľ výkonnosti
ID	Identifikačné číslo
UPVII	Úrad podpredsedu vlády pre investície a informatizáciu Slovenskej republiky
RFO	Register fyzických osôb

## Obsah

1	Manažérske zhrnutie pre štatistiky dátovej kvality	3
2	Vypočítané hodnoty základnej štatistiky	4
3	Grafické zobrazenie vypočítaných hodnôt základnej štatistiky	6
4	Záverečné zhodnotenie základnej štatistiky	9
5	Zoznamy	11
5.1	Zoznam tabuliek	11
5.2	Zoznam obrázkov a grafov	11

## 1 Manažérske zhrnutie pre štatistiky dátovej kvality

Je potrebné rozlišovať medzi dátovými štatistikami a štatistikami dátovej kvality.

Pri štatistikách dátovej kvality nám ide primárne o štatistiku a štatistické rozdelenie ukazovateľov dátovej kvality.

Pre lepšiu zrozumiteľnosť a ľahšiu ilustráciu sme pre jednotlivé merané registre a k nim vybrané ukazovatele dátovej kvality uviedli priemernú hodnotu, strednú hodnotu, najčastejšiu hodnotu, minimum a maximum.

Z takto vypočítaných hodnôt základnej štatistiky sa dajú identifikovať nové súvislosti, vzťahy a rozdelenia, ktoré nie sú rozpoznateľné v jednočíselnom sumárnom KPI pre daný konkrétny ukazovateľ dátovej kvality pre konkrétny dátový register.

Tieto štatistické údaje pre ukazovatele dátovej kvality sú prezentované v číselnej aj v grafickej forme – pre ich lepšiu zrozumiteľnosť a názornosť.

Štatistiky dátovej kvality môžu byť aj prvotným indikátorom na prioritizáciu kapacít a zdrojov alokovaných na zlepšenie dátovej kvality aj na detekciu najproblematickejších segmentov v danom datasete.

Vyššie formy štatistických ukazovateľov ako sú typ rozdelenia dát, lineárne a nelineárne korelácie a kvantitatívne väzby je odporúčané postupne aplikovať až po systematickom zavedení merania dátovej kvality na základnej štandardnej úrovni vo všetkých rezortoch.

## 2 Vypočítané hodnoty základnej štatistiky

Výpočet základnej štatistiky vychádza z nameraných KPI hodnôt v rámci profilácie pre vybrané registre. Tieto hodnoty sú uvedené v troch tabuľkách nižšie.

### Za Register adries – RA (vzorka 8 tabuliek)

Ukazovateľ	ID KPI	Minimum [%]	Maximum [%]	Priemerná hodnota [%]	Stredná hodnota [%]	Najčastejšia hodnota [%]
3.1 Dodržiavanie formátu atribútu	3.1a	88,89	100,00	99,16	100,00	100,00
4.2 Vyplnenosť povinného údajja	4.2a	73,48	100,00	99,24	100,00	100,00
4.3 Vyplnenosť nepovinného údajja	4.3a	00,01	100,00	73,98	99,17	100,00
5.2 Unikátnosť hodnoty	5.2a	99,38	100,00	99,96	100,00	100,00
	5.2b	00,00	33,69	2,26	0,12	1,25
8.1 Kompletnosť referenčného identifikátora	8.1a	100,00	100,00	100,00	100,00	100,00

**Tabuľka 1: Vypočítané hodnoty základnej štatistiky za Register adries**

### Za Register úpadcov – RU (vzorka 7 tabuliek)

Ukazovateľ	ID KPI	Minimum [%]	Maximum [%]	Priemerná hodnota [%]	Stredná hodnota [%]	Najčastejšia hodnota [%]
3.1 Dodržiavanie formátu atribútu	3.1a	0,00	100,00	95,61	100,00	100,00
4.2 Vyplnenosť povinného údajja	4.2a	0,00	100,00	48,73	36,95	100,00
4.3 Vyplnenosť nepovinného údajja	4.3a	0,00	100,00	49,87	37,15	100,00
5.2 Unikátnosť hodnoty	5.2a	0,00	100,00	17,58	1,72	0,00
	5.2b	0,00	28,57	4,35	1,41	5,71
8.1 Kompletnosť referenčného identifikátora	8.1a	100,00	100,00	100,00	100,00	100,00

**Tabuľka 2: Vypočítané hodnoty základnej štatistiky za Register úpadcov**

### Za Register právnických osôb – RPO (vzorka 7 tabuliek)

Ukazovateľ	ID KPI	Minimum [%]	Maximum [%]	Priemerná hodnota [%]	Stredná hodnota [%]	Najčastejšia hodnota [%]
3.1 Dodržiavanie formátu atribútu	3.1a	85,32	100,00	99,56	100,00	100,00
4.2 Vyplnenosť povinného údajja	4.2a	0,02	100,00	85,98	100,00	100,00

Tento projekt je podporený z Európskeho sociálneho fondu.

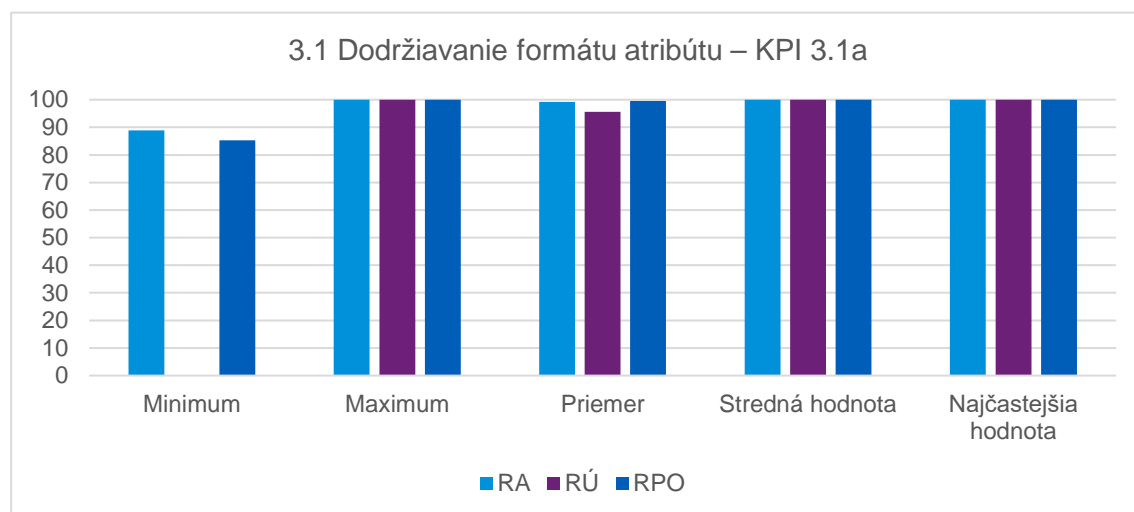
Ukazovateľ	ID KPI	Minimum [%]	Maximum [%]	Priemerná hodnota [%]	Stredná hodnota [%]	Najčastejšia hodnota [%]
4.3 Vyplnenosť nepovinného údaja	4.3a	0,02	100,00	85,98	100,00	100,00
5.2 Unikátnosť hodnoty	5.2a	0,00	100,00	25,27	0,37	0,00
	5.2b	0,00	8,33	1,57	0,03	0,00
8.1 Kompletnosť referenčného identifikátora	8.1a	100,00	100,00	100,00	100,00	100,00

**Tabuľka 3: Vypočítané hodnoty základnej štatistiky za Register právnických osôb**

### 3 Grafické zobrazenie vypočítaných hodnôt základnej štatistiky

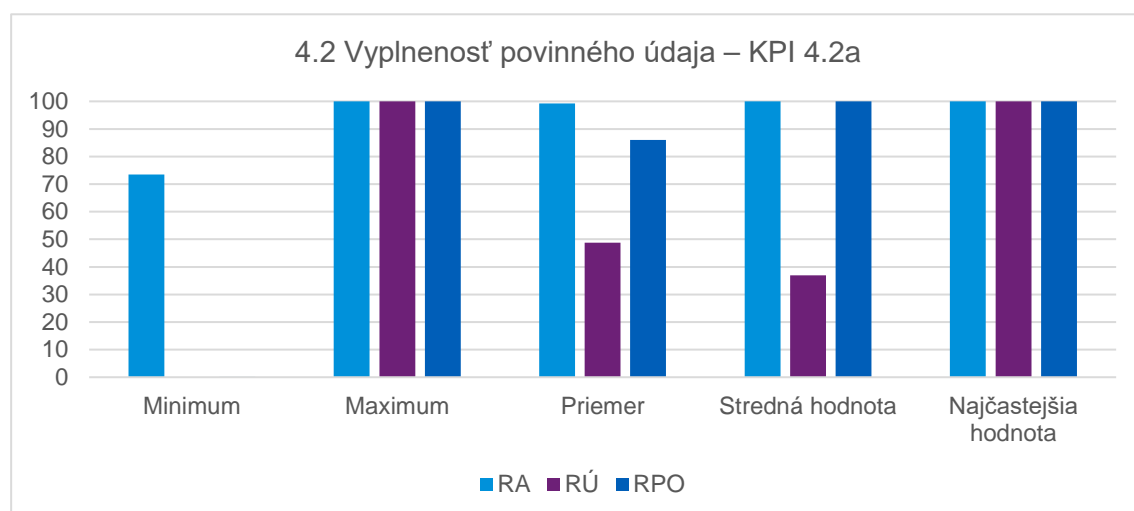
Vypočítané hodnoty základnej štatistiky sú zobrazené nižšie. Každý graf sa zameriava na jedno KPI a vizualizácia poskytuje pohľad na porovnanie medzi vybranými registrami.

#### *Ukazovateľ 3.1 Dodržiavanie formátu atribútu: KPI 3.1a – základná štatistika*



**Obrázok 1: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 3.1 Dodržiavanie formátu atribútu – KPI 3.1a**

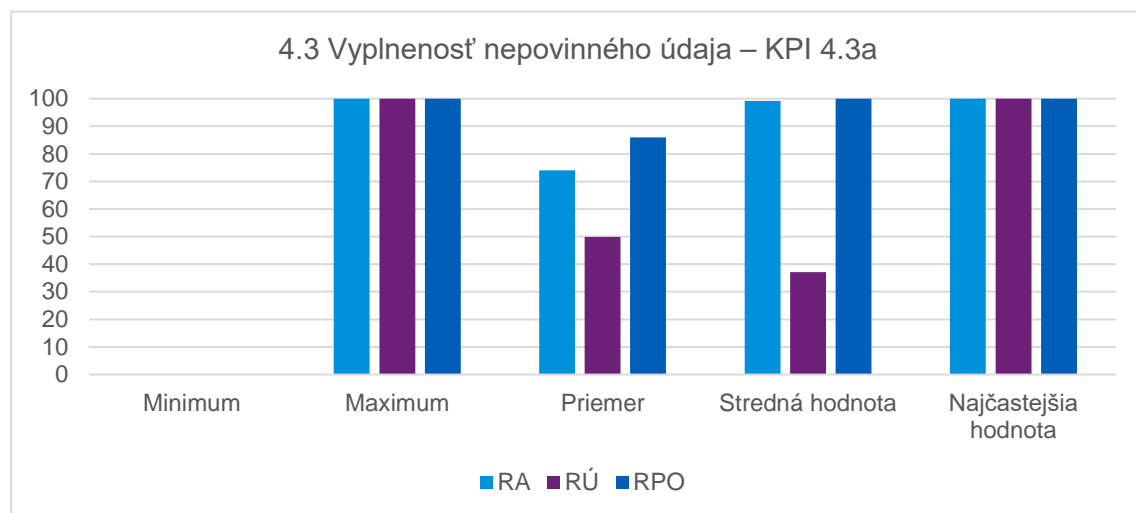
#### *Ukazovateľ 4.2 Vyplnenosť povinného údajá: KPI 4.2a - základná štatistika*



Tento projekt je podporený z Európskeho sociálneho fondu.

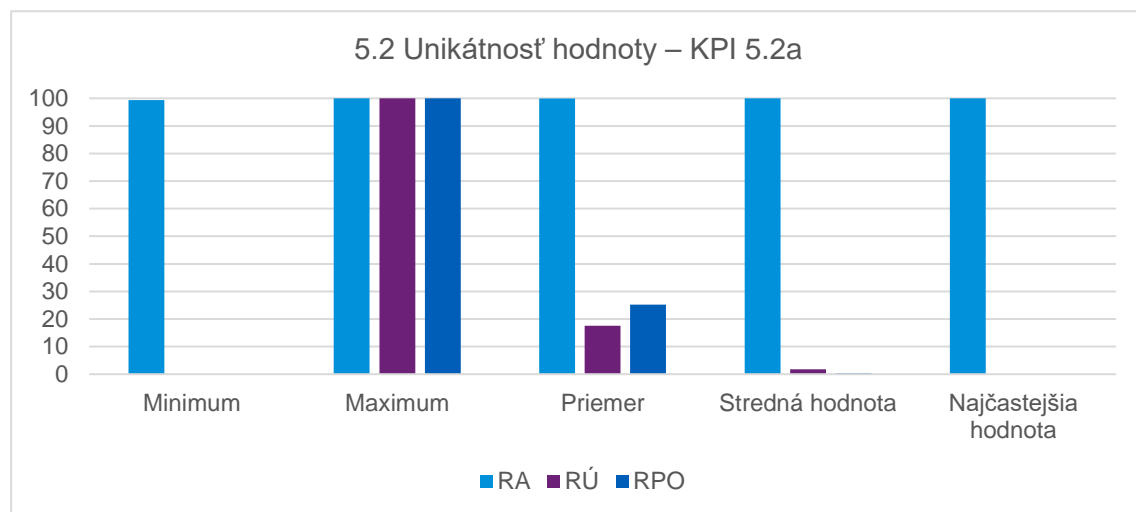
**Obrázok 2: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 4.2  
Vyplnenosť povinného údajá – KPI 4.2a**

**Ukazovateľ 4.3 Vyplnenosť nepovinného údajá: KPI 4.3a - základná štatistika**



**Obrázok 3: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 4.3  
Vyplnenosť nepovinného údajá – KPI 4.3a**

**Ukazovateľ 5.2 Unikátnosť hodnoty: KPI 5.2a - základná štatistika**

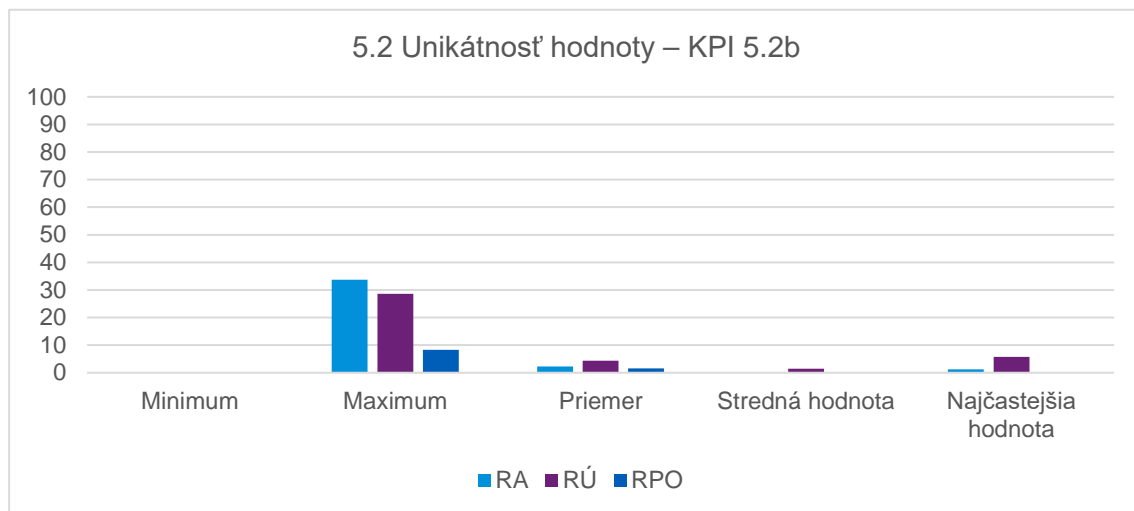


**Obrázok 4: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 5.2  
Unikátnosť hodnoty – KPI 5.2a**

Tento projekt je podporený z Európskeho sociálneho fondu.

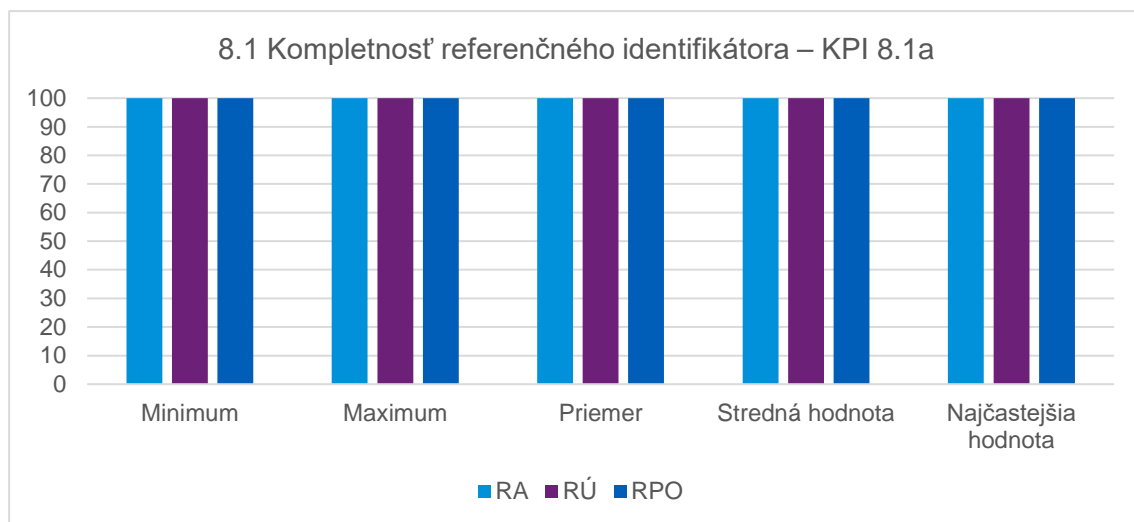


**Ukazovateľ 5.2 Unikátnosť hodnoty: KPI 5.2b (multiplicita) - základná štatistika**



**Obrázok 5: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 5.2 Unikátnosť hodnoty – KPI 5.2b (multiplicita)**

**Ukazovateľ 8.1 Kompletnosť referenčného identifikátora: KPI 8.1a - základná štatistika**



**Obrázok 6: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 8.1 Kompletnosť referenčného identifikátora – KPI 8.1a**

## 4 Záverečné zhodnotenie základnej štatistiky

Základná štatistika dátovej kvality vo verejnej správe vychádza z výsledkov merania dátovej kvality pre KPI v rámci profilácie. Zvolený postup merania bol založený na odhadnutých biznis pravidlách.

Tak ako pri meraní v rámci profilácie, aj pri výsledkoch základnej štatistiky vyplýva, že dôsledkom určeného postupu merania je vznik značnej rezervy v presnosti merania dátovej kvality. Nie každý atribút potrebuje mať definovaný formát, povinnosť vyplňania, unikátnosť hodnoty, multiplicitu hodnoty alebo určenie referenčného identifikátora.

Biznis pravidlá pomáhajú určiť kritéria dátovej kvality pre jednotlivé atribúty a mali by byť formálne definované, dokumentované, zdieľané a schválené. Podľa biznis pravidiel je potom možné určiť na ktorých atribútoch je potrebné merať konkrétne KPI. Tým sa viditeľne zvyšuje miera kvality výsledkov štatistiky dátovej kvality vo verejnej správe.

V tabuľke nižšie je **zobrazené porovnanie priemerných hodnôt** v základnej štatistike pre všetky vypočítané KPI v rámci profilácie vo vybraných registroch. Odhadnuté biznis pravidlá v rámci profilácie sledovali rovnaký postup odhadu pre všetky vybrané registre. Z výsledkov má najväčšie rezervy na zlepšovanie RU – Register úpadcov.

Ukazovateľ	ID KPI	RU - Priemerná hodnota [%]	RA - Priemerná hodnota [%]	RPO - Priemerná hodnota [%]
3.1 Dodržiavanie formátu atribútu	3.1a	95,61	99,16	99,56
4.2 Vyplnenosť povinného údajja	4.2a	48,73	99,24	85,98
4.3 Vyplnenosť nepovinného údajja	4.3a	49,87	73,98	85,98
5.2 Unikátnosť hodnoty	5.2a	17,58	99,96	25,27
	5.2b	4,35	2,26	1,57
8.1 Kompletnosť referenčného identifikátora	8.1a	100,00	100,00	100,00

**Tabuľka 4: Porovnanie priemerných hodnôt v základnej štatistike pre všetky vypočítané KPI v rámci profilácie vo vybraných registroch**

**Poznámka:** Farebné označenie zmeranej hodnoty KPI 2.1a je určené prahovými hodnotami KPI. Odporúčané prahové hodnoty sú nasledovné:

- Červená: pod 50 %
- Oranžová: 50-85 %
- Zelená: nad 85 %

V tabuľke nižšie je **zobrazené porovnanie minimálnych hodnôt** v základnej štatistike pre všetky vypočítané KPI v rámci profilácie vo vybraných registroch. Odhadnuté biznis pravidlá v rámci profilácie sledovali rovnaký postup odhadu pre všetky vybrané registre.

Tento projekt je podporený z Európskeho sociálneho fondu.

Keďže vyplnenosť nepovinného údaja je označovaný ako doplnkový ukazovateľ v rámci určenia adeptov na budúce analytické spracovanie, z výsledkov majú všetky vybrané registre značné rezervy na zlepšovanie.

Ukazovateľ	ID KPI	RU - Minimum [%]	RA - Minimum [%]	RPO - Minimum [%]
3.1 Dodržiavanie formátu atribútu	3.1a	0,00	88,89	85,32
4.2 Vyplnenosť povinného údaja	4.2a	0,00	73,48	0,02
4.3 Vyplnenosť nepovinného údaja	4.3a	0,00	00,01	0,02
5.2 Unikátnosť hodnoty	5.2a	0,00	99,38	0,00
	5.2b	0,00	00,00	0,00
8.1 Kompletnosť referenčného identifikátora	8.1a	100,00	100,00	100,00

**Tabuľka 5: Porovnanie minimálnych hodnôt v základnej štatistike pre všetky vypočítané KPI v rámci profilácie vo vybraných registroch**

**Poznámka:** Farebné označenie zmeranej hodnoty KPI 2.1a je určené prahovými hodnotami KPI. Odporúčané prahové hodnoty sú nasledovné:

- Červená: pod 50 %
- Oranžová: 50-85 %
- Zelená: nad 85 %

## **5 Zoznamy**

### **5.1 Zoznam tabuliek**

Tabuľka 1: Vypočítané hodnoty základnej štatistiky za Register adries .....	4
Tabuľka 2: Vypočítané hodnoty základnej štatistiky za Register úpadcov .....	4
Tabuľka 3: Vypočítané hodnoty základnej štatistiky za Register právnických osôb .....	5
Tabuľka 4: Porovnanie priemerných hodnôt v základnej štatistike pre všetky vypočítané KPI v rámci profilácie vo vybraných registroch .....	9
Tabuľka 5: Porovnanie minimálnych hodnôt v základnej štatistike pre všetky vypočítané KPI v rámci profilácie vo vybraných registroch .....	10

### **5.2 Zoznam obrázkov a grafov**

Obrázok 1: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 3.1 Dodržiavanie formátu atribútu – KPI 3.1a .....	6
Obrázok 2: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 4.2 Vyplnenosť povinného údaja – KPI 4.2a.....	7
Obrázok 3: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 4.3 Vyplnenosť nepovinného údaja – KPI 4.3a.....	7
Obrázok 4: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 5.2 Unikátnosť hodnoty – KPI 5.2a.....	7
Obrázok 5: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 5.2 Unikátnosť hodnoty – KPI 5.2b (multiplicita).....	8
Obrázok 6: Grafické zobrazenie vypočítaných hodnôt základnej štatistiky - 8.1 Kompletnosť referenčného identifikátora – KPI 8.1a.....	8