

## Čištění dat

Pracovat budeme s datasetem v2 (aktuální k datu 16. března). Jak probíhalo čištění dat?

### Otázka č. 3.

U otázky č. 3 došlo ke změně – tabulka byla nachystána ve špatném formátu a měli jste pouze jeden sloupeček na odpověď. Upravila jsem tedy otázku 3 tak, aby měla každá možnost svůj sloupec. Odpovědi jsou kódovány hodnotou „1“, pokud respondent danou odpověď zvolil (respondenti mohli zvolit více odpovědí).

U respondentů č. 81 a 82 byly uvedeny kódy „4“, které ale podle tabulky pro danou proměnnou neexistují. Tyto odpovědi byly tedy smazány a nahrazeny variantou „nevím“ jako nejpravděpodobnější možnou variantou (varianta „nevím“ byla zároveň čtvrtou variantou, což mohlo způsobit omyl při zápisu dat.

	1
MZK	1 a 3 a MZK
.	1
1	-1
.	3
1	-1
.	1 a 3
1	4
1	4
	3

### Otázka 4: znalost knihovníků

U otázky č. 4 došlo k chybnému vyplnění ve třech případech. Ačkoliv u otázky byly vyplněny instrukce k vyplnění s výběrem jen jedné možné odpovědi („Pokud pro vás platí více možností, vyberte vždy první z nich“), někteří tazatelé zapsali více variant.

- Varianta „1,3“ byla nahrazena kódem „1“.
- Varianta „2 a 3“ byla nahrazena kódem „2“.
- Kód „-1“ nebyl definován, odpověď byla smazána.

4_znalostknhovniku				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	,3	,3	,3
-1	1	,2	,2	,5
1	8	1,4	1,4	1,9
1,3	1	,2	,2	2,1
2	57	9,8	9,8	11,9
2 a 3	2	,3	,3	12,2
3	188	32,4	32,4	44,6
4	319	54,9	54,9	99,5
6	2	,3	,3	99,8
8	1	,2	,2	100,0
Total	581	100,0	100,0	

### Otázka 9: ženská nebo mužská profese

U otázky 9 je jeden chybný zápis – objevila se hodnota „4“. Hodnota byla smazána jako chyba.

9_genderprofese					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	353	60,8	61,0	61,0
	2	16	2,8	2,8	63,7
	3	209	36,0	36,1	99,8
	4	1	,2	,2	100,0
	Total	579	99,7	100,0	
Missing	System	2	,3		
Total		581	100,0		

### Otázka 10: Potřebné vzdělání:

U otázky 10 se objevil chybný zápis, kdy bylo umožněno respondentovi zvolit dvě možnosti, ačkoliv si měl vybrat jen jednu. Protože nelze určit, jakou možnost by respondent zvolil, pokud by měl jednu jednu možnou odpověď, kód byl smazán.

10_potrebnevzdelani					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		2	,3	,3	,3
	1	11	1,9	1,9	2,2
	2	301	51,8	51,8	54,0
	2,3	1	,2	,2	54,2
	3	198	34,1	34,1	88,3
	4	35	6,0	6,0	94,3
	5	19	3,3	3,3	97,6
	jak ve které knihovně	1	,2	,2	97,8
	je to jedno, jde o to, zda má rád knihy	1	,2	,2	97,9
	Je to jedno, pokud má k	1	,2	,2	98,1

### Otázka 13: Pohlaví

U otázky 13 byly odstraněny tyto chybné zápisy:

13_pohlavi					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	236	40,6	40,7	40,7
	2	341	58,7	58,8	99,5
	6	1	,2	,2	99,7
	7	1	,2	,2	99,8
	10	1	,2	,2	100,0
	Total	580	99,8	100,0	
Missing	System	1	,2		
Total		581	100,0		

### Otázka 15: Ekon. činnost

15_ekoncinnost					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		1	,2	,2	,2
	1	220	37,9	37,9	38,0
	1 a 3	1	,2	,2	38,2
	2	61	10,5	10,5	48,7
	2,8	1	,2	,2	48,9
	3	35	6,0	6,0	54,9
	4	25	4,3	4,3	59,2
	5	13	2,2	2,2	61,4
	6	46	7,9	7,9	69,4
	7	10	1,7	1,7	71,1
	8	164	28,2	28,2	99,3
	Mateřská dovolená	1	,2	,2	99,5
	podnikatel'	1	,2	,2	99,7
	V léčení	1	,2	,2	99,8
	Zaměstnanec na HPP a OSVČ na VPP	1	,2	,2	100,0
	Total	581	100,0	100,0	

Otázka na ekonomickou činnost byla upravena následovně:

- Varianta „1 a 3“ – překódováno na 1
- Varianta „2,8“ – překódováno na 2 (v závorce bylo upřesněno, že 2 zahrnují i pracující studenty)
- Hodnota „Mateřská dovolená“ – překódováno na „5“
- Hodnota „Zaměstnanec na HPP a OSVČ na VPP“ – překódováno na „1“ (hlavní činnost)
- Hodnota „Podnikatel“ – překódováno na „3“ (OSVČ = podnikatel)
- Hodnota „V léčení“ – překódováno na „9“ – jiné

## Respondenti č. 390 - 399

Prostým pohledem na data lze odhalit ještě anomálii u respondentů č. 390 až 399.

9	-	16500
-	-	15000
6	Fanoušek Black Books	17 000
3		14 000
6		15 500
2		18 000
3	Fanoušek Black Books	16000
5	Fanoušek Black Books	13000
2		15000
5	Fanoušek Black Books	15000
5	Fanoušek Black Books	15000
2	Fanoušek Black Books	16000
6		12500

Tito respondenti shodně uváděli stejnou odpověď u otázky č. 5. Pravděpodobnost, že by stejnou odpověď zvolilo nezávisle na sobě více lidí, je minimální, proto můžeme pojmout podezření. Další kontrolou zjistíme chybně vyplněnou otázku č. 5 – respondenti neseřadili důvody podle důležitosti, jak bylo uvedeno v zadání (na jedné pozici je více důvodů atd.). Respondenti mají navíc velmi shodné charakteristiky. Z těchto důvodů **všechny respondenty daného tazatele vyřazujeme z datasetu.**