## 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.



**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.

****

**Otázka 9: ženská nebo mužská profese**

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

****

**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.



**Otázka 13: Pohlaví**

U otázky 13 byly odstraněny tyto chybné zápisy:



**Otázka 15: Ekon. činnost**



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 zahrnuji 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.



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**.