



## **METODIKA pro nastavení, chod a aktualizaci specifické webové aplikace BOZP Kraje postavené na otevřených datech**

---

Konečný uživatel výsledku: **Ministerstvo práce a sociálních věcí Na Poříčním právu 1/376**  
**128 01 Praha 2**

**Název projektu/výzkumné potřeby:** Vývoj webové aplikace postavené na platformě otevřených dat k problematice bezpečnosti a ochrany zdraví při práci (BOZP)

**Číslo projektu/výzkumné potřeby:** V03-S4

**Řešitel výzkumné potřeby:** Výzkumný ústav bezpečnosti práce, v. v. i., Jeruzalémská 1283/9, 110 00 Praha 1 – Nové Město

**Doba řešení:** 1. 1. 2019 – 31. 12. 2020

**Důvěrnost a dostupnost:** veřejně přístupný



**Informace o autorském týmu:**

Mgr. Jiřina Ulmanová, DiS.

Mgr. Veronika Měrková, DiS.

Ing. Michal Hejdánek

**Další informace o projektu:**

Cílem výzkumného úkolu je vytvoření veřejně dostupné webové aplikace postavené na otevřených datech, které lze využít pro analýzy v oblasti bezpečnosti a ochrany zdraví při práci a prevence rizik.



Metodika je výsledkem výzkumné potřeby V03-S4 „Vývoj webové aplikace postavené na platformě otevřených dat k problematice bezpečnosti a ochrany zdraví při práci (BOZP)“, která je řešena Výzkumným ústavem bezpečnosti práce, v. v. i., v období 1/2019 až 12/2020 v rámci institucionální podpory MPSV ČR

### Členové projektového týmu

Mgr. Jiřina Ulmanová, DiS. (hlavní řešitelka), Kateřina Hrubá, PhDr. Irena Kuhnová, Mgr. Veronika Měrková, DiS., Kamila Myšková, Zdeňka Opletalová, Petra Růžičková, Marie Svobodová, Tereza Štanglerová

### Odborný garant MPSV

Ing. Zdeněk Cais

### Oponenti

**Ing. Martin Potančok, Ph.D.**, business a datový analytik, projektový manažer, Katedra informačních technologií VŠE

**Ing. Tomáš Zeman**, specialista na vývoj SW, spoluautor řady informačních systémů a aplikací zabývajících se oblastí bezpečnosti a ochrany zdraví při práci, nezávislý konzultant bez zaměstnaneckého poměru

Schválená „*METODIKA pro nastavení, chod a aktualizaci specifické webové aplikace BOZP Kraje postavené na otevřených datech*“

Praha 2020



## OBSAH

<b>SEZNAM ZKRATEK .....</b>	<b>6</b>
<b>ÚVOD.....</b>	<b>7</b>
<b>1. Cíl metodiky .....</b>	<b>8</b>
1.1 Novost postupů .....	8
1.2 Uplatnění metodiky .....	8
1.3 Ekonomické aspekty .....	8
<b>2. Metodika pro tvorbu katalogu otevřených datových sad (CKAN).....</b>	<b>9</b>
2.1 Získání a zpracování datových sad v otevřeném formátu .....	9
2.2 Popis otevřených datových sad – metadata.....	10
2.3 Software pro tvorbu katalogu - CKAN .....	11
2.4 Pracovní postup tvorby a údržby datového katalogu.....	14
2.4.1 Datové sady (DATASETS) .....	14
2.4.2 Organizace (ORGANIZATIONS) .....	21
2.4.3 Skupiny (GROUPS) .....	22
2.4.4 SHOWCASES.....	23
<b>3. Metodika pro tvorbu datového modelu a vizualizací (Tableau) .....</b>	<b>24</b>
3.1 Popis datových souborů .....	24
3.2 Datový model .....	26
3.3 Úprava zdrojových tabulek.....	27
3.3.1 Pivot.....	29
3.3.2 Čistící kroky.....	30
3.4 Spojení tabulek .....	32
3.4.1 Join.....	32
3.4.2 Union .....	34
3.5 Output .....	35
3.6 Dílčí datové modely .....	35
3.6.1 Ukazatele BOZP .....	36
3.6.2 Informace o krajích.....	37
3.6.3 Kontrolní činnost .....	37
3.6.4 Oceněné podniky.....	38
3.6.5 Instituce BOZP a zdravotnická zařízení.....	39



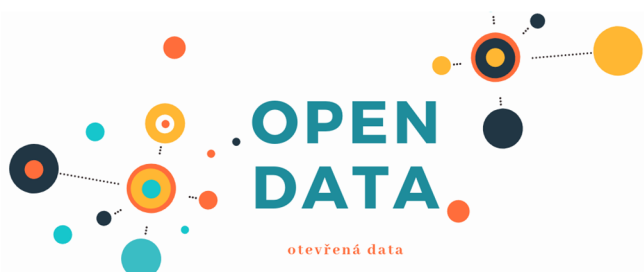
3.6.6	Nebezpečné události .....	39
3.7	Globální datový model .....	40
3.8	Tvorba vizualizací v prostředí nástroje Tableau .....	41
3.8.1	Tabulky a celá čísla .....	46
3.8.2	Mapy .....	49
3.8.3	Liniové grafy .....	52
3.8.4	Scatter plot .....	55
3.8.5	Dashboards .....	56
3.8.6	Publikace do webového prostředí .....	58
	<b>Použitá literatura .....</b>	<b>60</b>
	<b>Seznam obrázků .....</b>	<b>61</b>



## SEZNAM ZKRATEK

BOZP	Bezpečnost a ochrana zdraví při práci
CKAN	Comprehensive Knowledge Archive Network
ČSSZ	Česká správa sociálního zabezpečení
ČSÚ	Český statistický úřad
MAPIS	Major Accident Prevention Information System
MPSV	Ministerstvo práce a sociálních věcí
MZd	Ministerstvo zdravotnictví
NKOD	Národní katalog otevřených dat
SÚIP	Státní úřad inspekce práce
SW	Software
SZÚ	Státní zdravotní ústav
ÚZIS	Ústav zdravotnických informací a statistiky ČR
VÚBP, v. v. i.	Výzkumný ústav bezpečnosti práce, veřejná výzkumná instituce (v textu dále VÚBP)

## ÚVOD



Otevřená data jsou podle OECD předpokladem zvýšení konkurenceschopnosti České republiky a jsou obrovskou příležitostí současnosti. Veřejné i soukromé instituce disponují celou řadou datových a informačních zdrojů, kterých lze využít a efektivně je analyzovat a sdílet. Využívání a sdílení otevřených dat je zakotveno v činnosti celé řady mezinárodních i národních iniciativ a organizací. Na národní úrovni

jde např. o Strategický rámec rozvoje eGovernmentu 2014+, Digitální Česko v. 2.0 - Cesta k digitální ekonomice, Implementace PSI směrnice, Strategie vlády v boji s korupcí na období let 2013 a 2014, Akční plán Partnerství Otevřeného vládnutí. Na Portálu veřejné správy je dostupný Národní katalog otevřených dat, který umožňuje vyhledávat otevřené datové sady zveřejněné jednotlivými orgány veřejné správy v ČR. Otevřeným datům se věnuje projekt Ministerstva vnitra České republiky, který zprovoznilo portál pro ostatní orgány veřejné správy, které plánují otevřít svá data. Popisuje doporučený postup otevírání dat a nabízí vzorové publikační plány, které obsahují (pro jednotlivé typy orgánů veřejné správy) konkrétní datové sady k publikaci ve formátu otevřených dat. Pozadu nejsou ani ostatní ministerstva, která se také rozhodla publikovat svá data v otevřeném formátu. Je tedy zjevné, že se tomuto trendu Česká republika otevřela a že se snaží na něj reagovat. Totéž se děje i na mezinárodní úrovni. Na úrovni Evropské unie byl vytvořen Evropský portál veřejně přístupných dat, který je centralizovaným úložištěm dat, jež produkují evropské instituce a další subjekty EU a jejichž objem neustále narůstá.

V oblasti bezpečnosti a ochrany zdraví při práci (BOZP), která je multidisciplinárním oborem, existuje pro jejich využití obrovský potenciál. Analýzy dat o pracovní úrazovosti, nemocech z povolání, inspekčních činnostech, kvalitě pracovního života a dalších aspektech spojených s naším pracovním životem mohou pomoci směřovat národní politiku v oblasti BOZP do exponovaných a nejvíce rizikových průmyslových oblastí. Díky nim bude možné zaměřit prevenci, osvětu a propagaci BOZP na ohrožené skupiny pracovníků a zaměstnanců. S tím, jak se mění charakter práce, s nástupem digitalizace, automatizace a robotizace, nových forem práce či vznikem nových rizik je třeba pohlížet na oblast BOZP v širokém kontextu a analýzy dat a informací mohou sloužit jako objektivní a relevantní zdroj pro podporu preventivních činností k ochraně fyzického a duševního zdraví člověka.

Tato metodika je jedním z výstupů výzkumné potřeby institucionální podpory MPSV č. V03-S4 „Vývoj webové aplikace postavené na platformě otevřených dat k problematice bezpečnosti a ochrany zdraví při práci“, který řešil Výzkumný ústav bezpečnosti práce, v. v. i. (dále VÚBP), od ledna 2019 do prosince 2020. Primárním cílem bylo vytvoření veřejně dostupné webové aplikace postavené na otevřených datech, které lze využít pro analýzy v oblasti bezpečnosti a ochrany zdraví při práci a prevence rizik. Tento cíl byl definován v návaznosti na platnou koncepci VaVal rezortu MPSV a priority v oblasti bezpečnosti a ochrany zdraví při práci. Přispěje k vytvoření a doplnění potřebné odborné základny pro rozšíření a zefektivnění komunikačních a osvětových prostředků v oblasti BOZP.



## 1. CÍL METODIKY

### 1.1 Novost postupů

Cílem metodiky je poskytnout pracovní postup pro nastavení, chod a aktualizaci specifické webové aplikace BOZP Kraje postavené na otevřených datech, která je dostupná na adrese <http://opendata.vubp.cz>.

Webová aplikace BOZP Kraje využívá 2 softwarové platformy, jednak je to CKAN pro tvorbu katalogu otevřených dat, a dále Tableau, které je využíváno pro datové analýzy a tvorbu vizualizačních dashboardů. CKAN tvoří datovou základnu pro vizualizace v podobě tematických dashboardů.

Metodika je tedy rozdělena do dvou částí. První se věnuje sběru, tvorbě a popisu datových sad a následně údržbě datového katalogu. Druhá část je zaměřena na obecné principy tvorby datového modelu spolu s vizualizacemi a na praktický postup provádění jednotlivých kroků v rámci platformy Tableau.

### 1.2 Uplatnění metodiky

Předkládaná metodika je primárně určena pro správce a provozovatele webové aplikace BOZP Kraje a dále pro úzký okruh autorů, kteří pracují s platformou Tableau, konkrétně s programem Tableau Prep a Tableau Desktop. Tvorba datového modelu je vždy specifická pro konkrétní případ, avšak zná-li autor základní principy a práci s platformou, může datový model rozšířit o další zdroje, případně část datového modelu využít pro jiný případ. Co se tvůrců katalogu otevřených datových sad týče, v první řadě se bude jednat o pracovníky VÚBP, a pokud by došlo k masivnějšímu otevírání dat v rámci rezortu práce a sociálních věcí, budou se moci do tvorby katalogu zapojit i další poskytovatelé těchto otevřených datových sad.

### 1.3 Ekonomické aspekty

Díky metodice bude zaučení nových autorů rychlejší a zároveň požadavky na zdrojová data budou přesnější. Nekonzistence případných dalších datových zdrojů s datovým modelem půjde jednoduše odstranit tak, aby bylo možné připojit nové zdroje do funkčního prostředí. Nebude potřeba zaměstnávat datové analytiku, kteří by data zpracovávali, a sami poskytovatelé dat budou schopni v jasně definované struktuře data k analýzám připravit.





## 2. METODIKA PRO TVORBU KATALOGU OTEVŘENÝCH DATOVÝCH SAD (CKAN)

### 2.1 Získání a zpracování datových sad v otevřeném formátu

Úvodním krokem tohoto procesu je vytipování institucí, jejichž data chceme využívat a zjistit, zda data v otevřeném formátu publikují. Pokud je nezveřejňují a nezpřístupňují v otevřené podobě, je dalším krokem projít volně dostupné dokumenty, v nichž se námi potřebná a zajímavá data objevují. Ve většině případů lze data a informace nalézt ve výročních zprávách, zprávách o činnosti či v jiných materiálech, např. pravidelně publikovaných reportech o určitém ukazateli.

Stěžejním zdrojem o tom, jaké datové sady jsou v současné době k dispozici od státních a veřejných institucí, je Národní katalog otevřených dat veřejné správy ČR (dále NKOD). NKOD obsahuje katalogizační záznamy o datových sadách zpřístupněných veřejnými institucemi ČR, včetně odkazů ke stažení dat, která jsou uložena na webu příslušného poskytovatele dat.

Z pohledu BOZP jsou v NKOD zajímavé datové sady od těchto institucí:

- České správy sociálního zabezpečení (pracovní činnosti, invalidita a invalidní důchody, diagnózy, pracovní neschopnost),
- Českého statistického řadu (demografická data, vzdělávání, zaměstnání, nezaměstnanost, pohlaví, průměrné mzdy, pracovní neschopnost),
- Ministerstva vnitra (číselníky),
- Ministerstva zdravotnictví ČR (zdravotní služby).

Vhodné datové sady od výše uvedených institucí je třeba stáhnout do lokálního katalogu CKAN a opatřit sadou metadat. Pro zajištění aktuálnosti a kontinuity časových řad je třeba sledovat nově zveřejňovaná data v NKOD a postupně stahovat ta, která se do profilu aplikace hodí.

Pro obsáhnutí celého tematického spektra BOZP však existují další veřejné instituce, jejichž data jsou pro zpracování tematických analýz stěžejní. Jedná se o:

- Státní zdravotní ústav (SZÚ), která zpracovává data o nemocech z povolání,
- Státní úřad inspekce práce (SÚIP), který zpracovává data o pracovních úrazech a disponuje informacemi o realizovaných kontrolách.

S oběma institucemi VÚBP, v. v. i., úzce spolupracuje, ale ani jedna z uvedených institucí svá data v otevřeném formátu nezveřejňuje. Často jsou svázány legislativními a dalšími překážkami. Jelikož jsou ale jimi poskytované informace pro oblast BOZP jedny z nejdůležitějších, lze vytvořit vlastní datové sady. Obě instituce zveřejňují na svých webových stránkách výroční zprávy, zprávy o činnosti, zprávy z realizovaných kontrol apod. (agregovaná data), přičemž je možné potřebná data získat právě z nich. Proto byly vytvořeny vlastní otevřené datové sady z uvedených dokumentů, kde jsou data ve formátu tabulek zpracována a připravena pro využití v aplikaci.



Jedná se o:

- nemoci z povolání,
- pracovní úrazy,
- smrtelné pracovní úrazy,
- kontroly realizované orgány inspekce práce.

I v případě těchto dat je nezbytné sledovat, kdy je dané instituce zveřejňují, a následně je z dokumentů zpracovat a vytvořit nové otevřené datové sady.

Další data, která jsou do aplikace zahrnuta, jsou:

- závažné havárie a mimořádné události a objekty spadající pod zákon o prevenci závažných havárií,
- podniky oceněné z hlediska BOZP,
- adresáře institucí zabývajících se činnostmi souvisejícími s problematikou BOZP atd., např. poskytovatelé zdravotních služeb.

## 2.2 Popis otevřených datových sad – metadata

Každá datová sada je popsána souborem metadat, která obsahují jak informace o vlastních datech, tak informace o kontextu. Jedná se např. o zdrojovou instituci (původce dat), datum pořízení, licence atd. Metadata datové sady ji popisují bez ohledu na jejich konkrétní reprezentaci. Při publikování otevřených dat mají metadata klíčovou roli:

- představují dokumentaci dat, která je nutná pro jejich správnou interpretaci,
- pomocí metadat je možné v katalogu data podle určitých parametru vyhledávat.

Definované položky metadatového záznamu datové sady (datasetu) jsou:

- jméno/název,
- popis – o jaká data se jedná a jakou přináší informaci,
- štítky/tagy – věcný popis pomocí klíčových slov,
- formát (xls, csv, pdf),
- autor/zdroj (myšleno instituce) – jedná se o katalog vytvořený VÚBP,
- správce (vždy bude tato položka obsahovat VÚBP),
- naposledy aktualizováno,
- vytvořeno,
- periodičita aktualizace – jak často se budou data aktualizovat (v tomto případě půjde nejčastěji o roční periodicitu),
- časové pokrytí – zajištění kontinuity časových řad v rámci analýz.



Každý nově vložený dataset musí být vždy opatřen výše uvedenými metadaty. Datové sady budou díky tomu jednotně popisovány a bude snadné je vyhledávat. Se zvětšujícím se počtem datasetů bude možné se v katalogu pomocí metadat lépe orientovat.

**Nemoci z povolání**

Následovnici: 0

**Organizace**

**Státní zdravotní ústav**  
Státní zdravotní ústav je příspěvkovou organizací ministerstva zdravotnictví. Jeho postavení a úkoly jsou stanoveny § 86 zákona č. 258/2000 Sb., o ochraně veřejného zdraví a o... načíst další

**Data a zdroje**

- Nemoci z povolání v ČR v roce 2015
- Nemoci z povolání v ČR v roce 2016
- Nemoci z povolání v ČR v roce 2017
- Nemoci z povolání v ČR v roce 2018
- Nemoci z povolání v ČR v roce 2019

**Doplňující informace**

Pole	Hodnota
Autor	SZÚ
Správce	VÚBP
Naposledy aktualizováno	2. září 2020, 12:06 (UTC+02:00)
Vytvořeno	2. září 2020, 11:33 (UTC+02:00)
Periodicita aktualizace	roční
Časové pokrytí	2015-2019

Obrázek 1: Ukázka záznamu datové sady/datasetu

### 2.3 Software pro tvorbu katalogu - CKAN

Pro tvorbu lokálního datového katalogu je zvolen SW Comprehensive Knowledge Archive Network (CKAN). Jedná se o open-sourcový projekt, napsaný v programovacím jazyku Python na serverové straně a v programovacím jazyku JavaScript na frontendu. CKAN je po stažení okamžitě připravený k nasazení.

Tento software je užíván víceméně jako **standard pro katalogizaci dat** vládami a skupinami uživatelů po celém světě. Je to platforma mnoha oficiálních a komunitních datových portálů.



CKAN je přední světovou open-source platformou pro datové portály. Software CKAN je navržen tak, aby s ním mohli pracovat jak lidé, tak i další programy a aplikace. Za tímto účelem CKAN disponuje webovým rozhraním pro interakci s člověkem a aplikačním rozhraním (API<sup>1</sup>).

Záznamy o datech (datových množinách) lze do CKAN vkládat, upravovat a lze v nich také vyhledávat. Datové množiny v datovém katalogu lze **kategorizovat** pomocí skupin a štítků (tagů). Uživatelé mohou k záznamům o datových množinách (datasetech) přidávat svoje komentáře a mohou také datasety hodnotit.

Základní jednotkou pro popis dat v software CKAN je tzv. **dataset**, který obsahuje odkaz na datové zdroje (jejich URI<sup>2</sup>), prostřednictvím kterých lze získat datové množiny, a dále je v datasetu obsažen také popis těchto datových množin.

Každý dataset má několik **atributů/metadat**, jejichž hodnoty je možné při jeho zakládání nastavit či později editovat (titulek, název, URL adresa, domovská stránka, popis, licence, tagy, zdroje, skupiny, autor, verze, stav, e-mail autora).

**CKAN API** je rozhraní nejrozšířenější implementace datového katalogu CKAN. Je založené na formátu JSON<sup>3</sup>, a obsahuje dvě hlavní funkce, `package_list` a `package_show`. Toto rozhraní je implementováno samotným katalogem CKAN a také katalogem DKAN. Pro instituce, které nemohou nebo nechtějí provozovat celé katalogizační řešení, postačí, když budou CKAN API simulovat pomocí statických JSON souborů umístěných na odpovídající URL.

CKAN API umožňuje nastavit vazby mezi jednotlivými CKAN datasety. Všechny vazby, které má CKAN dataset k jiným CKAN datasetům jsou ukládány do atributu Vazby (Relationships).

Rozhraní CKAN API je pro NKOD rozšířeno o některé metadatové položky inspirované standardem DCAT-AP<sup>4</sup>. Bez nich bude metadatový záznam z hlediska NKOD neúplný. Jsou k dispozici vzorová rozšíření pro CKAN a rozšíření pro DKAN, která lze využít pro předávání úplných metadatových záznamů tímto rozhraním.

Uživatelská podpora CKAN týmu je dostupná prostřednictvím FAQ, blogu, wiki stránek. Podpora je zpoplatněna.

---

<sup>1</sup> API (Application Programming Interface) označuje v informatice rozhraní pro programování aplikací.

<sup>2</sup> URI (Uniform Resource Identifier) je textový řetězec s definovanou strukturou, který slouží k přesné specifikaci zdroje informací (ve smyslu dokument nebo služba), hlavně za účelem jejich použití pomocí počítačové sítě, zejména Internetu.

<sup>3</sup> JSON (JavaScript Object Notation) je odlehčený formát pro výměnu dat. Je jednoduše čitelný i zapisovatelný člověkem a snadno analyzovatelný i generovatelný strojem.

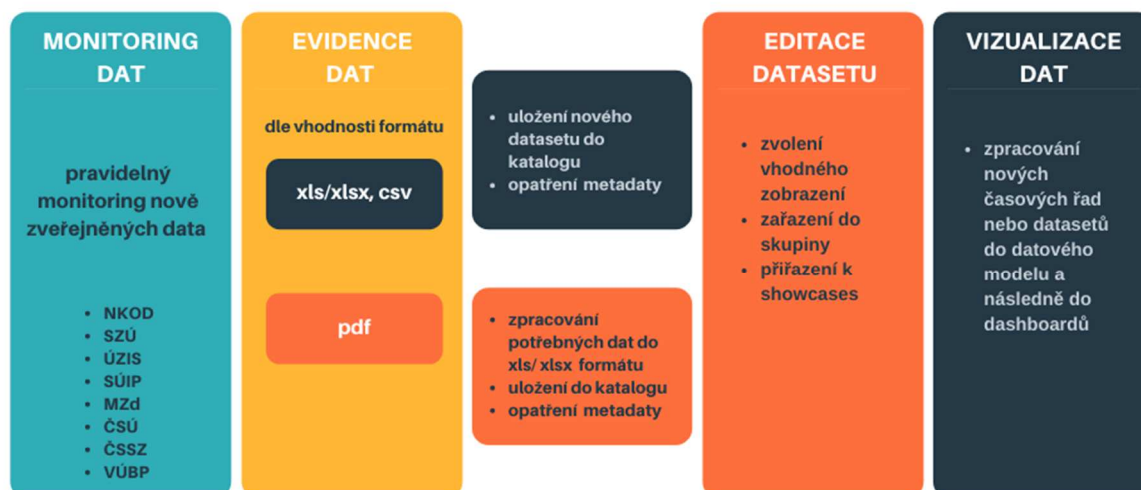
<sup>4</sup> DCAT-AP (DCAT Application Profile for Data Portals in Europe) je standard založený na propojených datech, který rozlišuje dva základní typy entit, datovou sadu, a její distribuci. Metadata datové sady ji popisují bez ohledu na konkrétní reprezentaci jejích dat.



Obrázek 2: Úvodní stránka CKANu

## 2.4 Pracovní postup tvorby a údržby datového katalogu

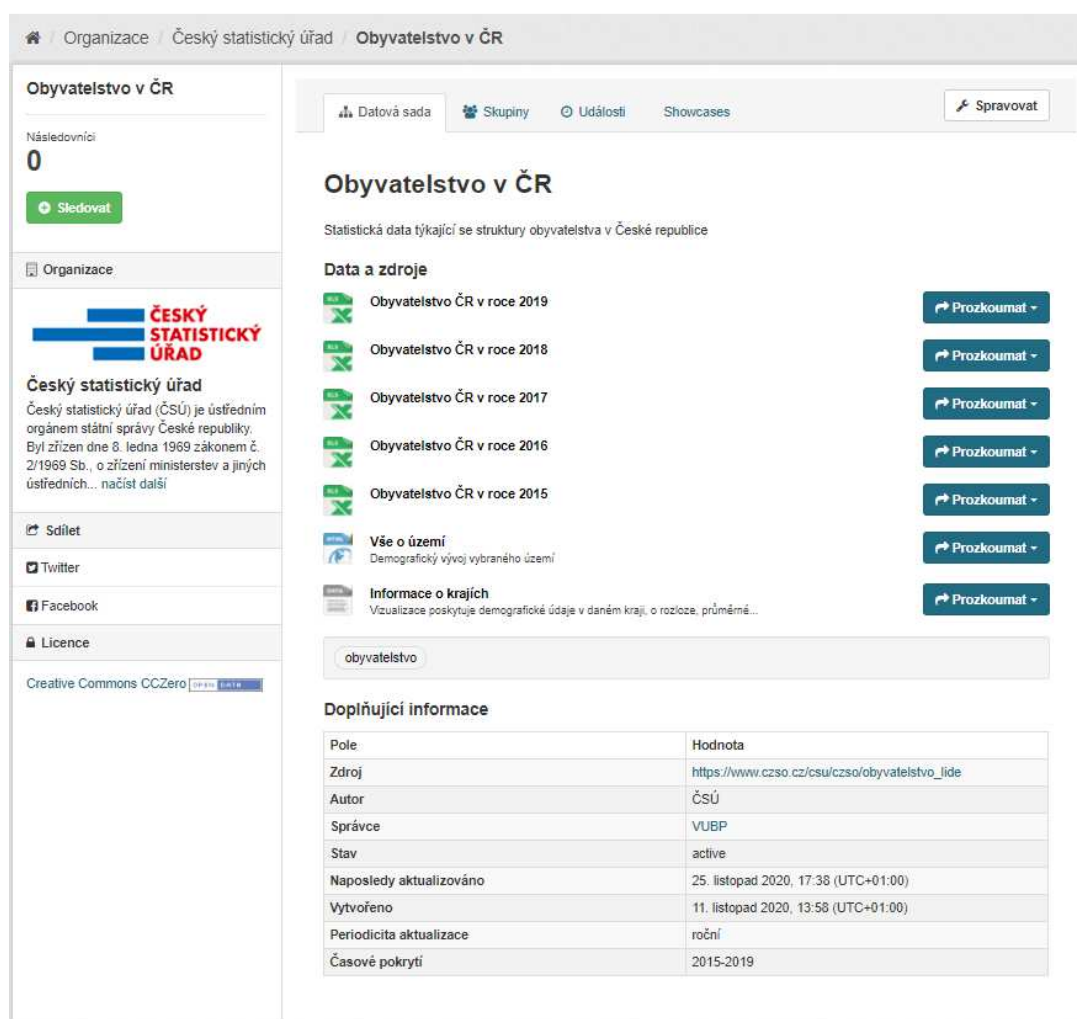
Základní jednotkou pro popis dat v software CKAN je tzv. dataset, který obsahuje odkaz na datové zdroje, prostřednictvím kterých lze získat datové množiny, a dále je v datasetu obsažen také popis těchto datových množin.



### 2.4.1 Datové sady (DATASETS)

Každý dataset má několik atributů (metadat), jejichž hodnoty je možné při jeho zakládání nastavit či později editovat (titulek, název, URL adresa, domovská stránka, popis, licence, tagy, zdroje, skupiny, autor, verze, stav, e-mail autora).

V případě této aplikace jde o vlastní datové soubory týkající se závažných havárií, pracovních úrazů, smrtelných pracovních úrazů, přehledu ocenění v BOZP a datové sady cizích institucí – SZÚ, ÚZIS, SÚIP, MZd, ČSÚ a ČSSZ.



Organizace / Český statistický úřad Obyvatelstvo v ČR

Obyvatelstvo v ČR

Následovníci: 0

Stledovat

Organizace

**ČESKÝ STATISTICKÝ ÚŘAD**

Český statistický úřad  
Český statistický úřad (ČSÚ) je ústředním orgánem státní správy České republiky. Byl zřízen dne 8. ledna 1969 zákonem č. 2/1969 Sb., o zřízení ministerstev a jiných ústředních... načíst další

Sdílet

Twitter

Facebook

Licence

Creative Commons CCZero [Open Data](#)

Datová sada Skupiny Události Showcases Spravovat

### Obyvatelstvo v ČR

Statistická data týkající se struktury obyvatelstva v České republice

#### Data a zdroje

- Obyvatelstvo ČR v roce 2019 [Prozkoumat](#)
- Obyvatelstvo ČR v roce 2018 [Prozkoumat](#)
- Obyvatelstvo ČR v roce 2017 [Prozkoumat](#)
- Obyvatelstvo ČR v roce 2016 [Prozkoumat](#)
- Obyvatelstvo ČR v roce 2015 [Prozkoumat](#)
- Vše o území [Prozkoumat](#)  
Demografický vývoj vybraného území
- Informace o krajích [Prozkoumat](#)  
Vizualizace poskytuje demografické údaje v daném kraji, o rozloze, průměrně...

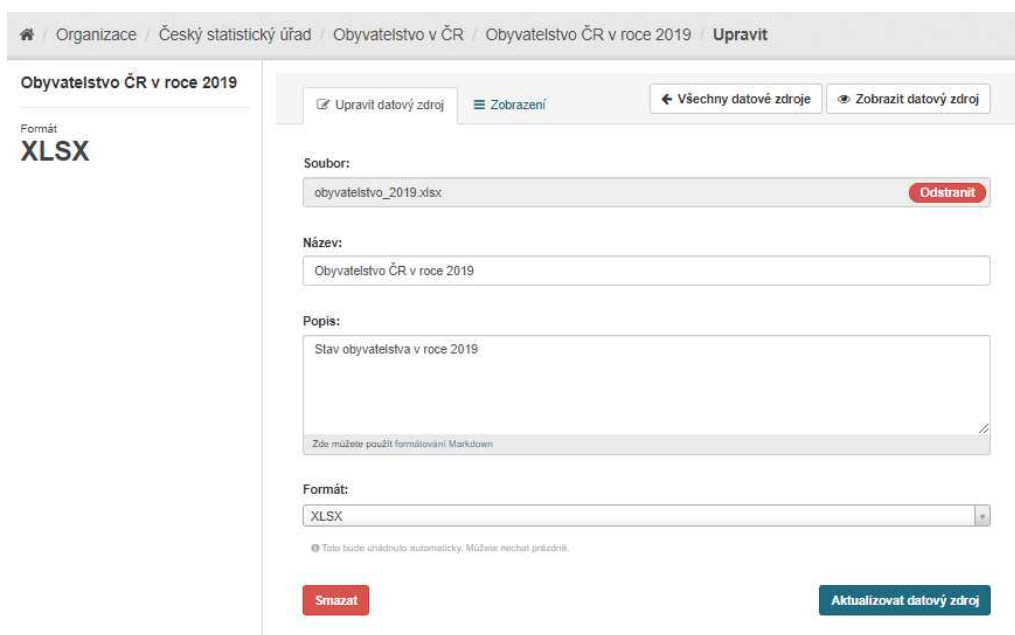
obyvatelstvo

#### Doplňující informace

Pole	Hodnota
Zdroj	<a href="https://www.czso.cz/csu/czso/obyvatelstvo_lide">https://www.czso.cz/csu/czso/obyvatelstvo_lide</a>
Autor	ČSÚ
Správce	VUBP
Stav	active
Naposledy aktualizováno	25. listopad 2020, 17:38 (UTC+01:00)
Vytvořeno	11. listopad 2020, 13:58 (UTC+01:00)
Periodicita aktualizace	roční
Časové pokrytí	2015-2019

Obrázek 3: Ukázka datasetu Obyvatelstvo v ČR

Datová sada je nejprve popsána výše uvedenými **metadaty** a následně je přidán **datový zdroj**. Každý datový set může obsahovat více datových zdrojů. Datový zdroj lze vložit **nahráním souboru** nebo uvedením **URL odkazu na zdroj**. Datový zdroj vlastních datasetů či upravované datasety jiných organizací jsou přímo uloženy v katalogu. Datasety ostatních institucí jsou v CKANu evidovány a popsány, avšak na datový zdroj je odkazováno. Tímto způsobem je v některých případech zajištěna automatická aktualizace.

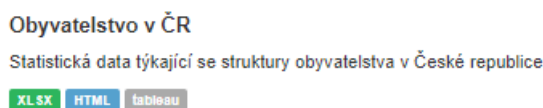


Obrázek 4: Editace datového zdroje

U datového zdroje se uvede jeho název, stručný popis a formát. Dále je nezbytné zvolit vhodné **zobrazení datového zdroje**, a to dle příslušného formátu. Základní zobrazení nabízí tyto možnosti:

- prohlížeč dat (data explorer) – mřížka, graf, mapa,
- obrázek – formáty jpeg, png,
- Tableau – zobrazení tematických dashboardů,
- webová stránka – webové stránky,
- PDF – embedované pdf soubory.

Jednotlivá zobrazení lze v rámci jednoho datového zdroje řadit dle zvoleného pořadí.



Obrázek 5: Ukázka vyhledaného datasetu s uloženými formáty

Zpracovaný dataset je dále přiřazen do předdefinovaných **skupin** (viz 2.4.3) a **showcases** (viz 2.4.4). Každý zobrazený dataset si může uživatel stáhnout či přejít na původní datový zdroj. Jednotlivé datasety lze také **sdílet prostřednictvím sociálních sítí** Facebook a Twitter.



## Ukázky typů zobrazení

[Organizace](#) / [Český statistický úřad](#) / [Obyvatelstvo v ČR](#) / [Obyvatelstvo ČR v roce 2019](#)

[Spravovat](#) [Stáhnout](#)

URL: [http://opendata.vubp.cz/cs\\_CZ/dataset/d068fee-8596-456e-9875-41a914a7b82d/resource/770ce08a-a2d6-4557-96a9-9cc69ef006af/download/obyvatelstvo\\_2019.xlsx](http://opendata.vubp.cz/cs_CZ/dataset/d068fee-8596-456e-9875-41a914a7b82d/resource/770ce08a-a2d6-4557-96a9-9cc69ef006af/download/obyvatelstvo_2019.xlsx)

Stav obyvatelstva v roce 2019

Stav obyvatelstva v roce 2019 Zakomponovat

15 records

KRAJ	OBYVATELSTVO CELKEM	muži	ženy	VĚK CELKEM	věk muži	věk ženy
Hlavní město Praha	1308632	638009	670623	41.92148976946918	40.4605...	43.3113...
Středočeský kraj	1369332	676696	692636	41.188851206281605	39.9696...	42.3800...
Jihočeský kraj	642133	317268	324865	42.65711542624347	41.3528...	43.9308...
Plzeňský kraj	584672	290226	294446	42.725066361994415	41.5198...	43.9129...
Karlovarský kraj	294896	145674	149222	42.89413895068092	41.5491...	44.2071...
Ústecký kraj	820789	407395	413394	42.029741504820365	40.6705...	43.3692...
Liberecký kraj	442356	217791	224565	42.11849505827885	40.6788...	43.5147...
Královéhradecký kraj	551021	271591	279430	43.08185985652089	41.5870...	44.5347...
Pardubický kraj	520316	257948	262368	42.447132127399506	41.0209...	43.8493...
Kraj Vysočina	509274	253061	256213	42.752105939042636	41.3925...	44.0948...
Jihomoravský kraj	1187667	582516	605151	42.375194814708166	40.8534...	43.8399...
Olomoucký kraj	632492	309620	322872	42.785424953991516	41.2351...	44.2720...
Zlínský kraj	582921	285883	297038	43.0835559192412	41.3936...	44.7099...
Moravskoslezský kraj	1203299	590516	612783	42.669036955902065	41.0559...	44.2234...
Česká republika	10649800	5244194	5405606	42.33463360814287	40.9036...	43.7229...

Obrázek 6: Ukázka zobrazení mřížka

U zobrazení **prohlížeč dat** lze z tabulkových dat vytvářet jednoduché grafy či na základě mapových podkladů generovat mapy.

Aby se data správně zobrazovala v náhledu mřížka, je třeba vytvořit jednoduché tabulky bez záhlaví a zbytečného formátování ve formátu xls/xlsx či použít dokumenty ve formátu CSV. V náhledu mřížka lze mezi jednotlivými daty vyhledávat a filtrovat.

## Vývoj počtu nelegálně pracujících osob v ...

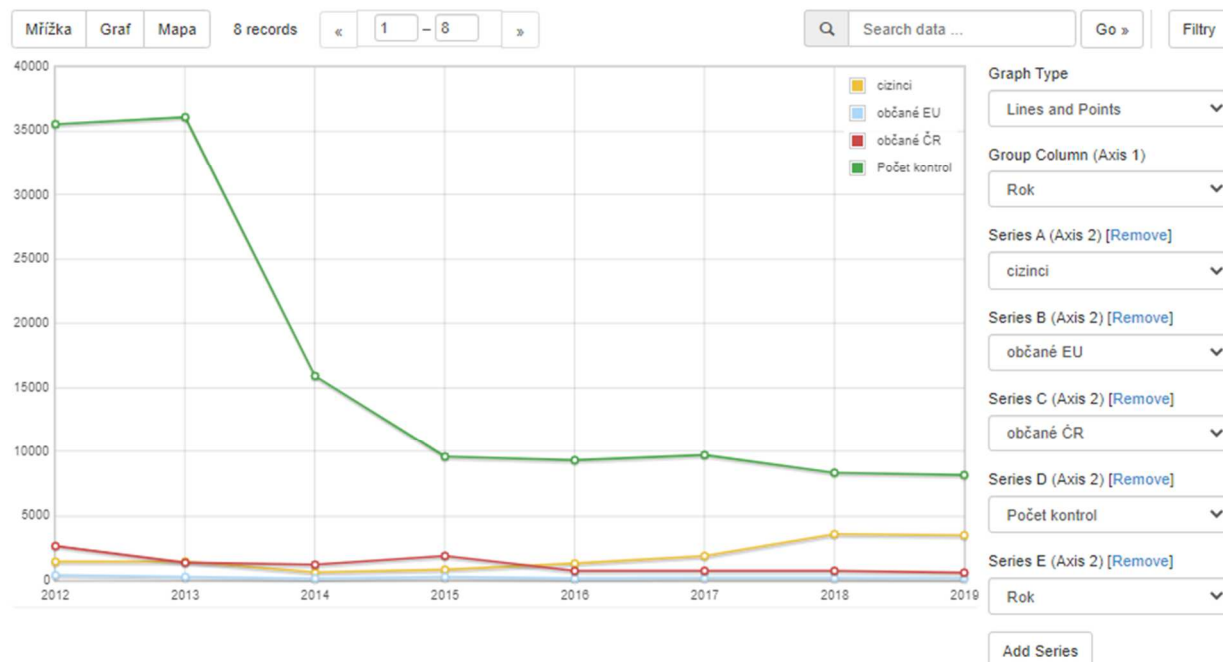
[Spravovat](#) [Stáhnout](#)

URL: [http://opendata.vubp.cz/cs\\_CZ/dataset/f08aaa7b-b1f7-4164-96f2-553401b5e07b/resource/cb9ea601-3329-46d9-964a-dc55788f3c94/download/vyvoj-poctu-zjisteny-ch-oso-...](http://opendata.vubp.cz/cs_CZ/dataset/f08aaa7b-b1f7-4164-96f2-553401b5e07b/resource/cb9ea601-3329-46d9-964a-dc55788f3c94/download/vyvoj-poctu-zjisteny-ch-oso-...)

Zatímco v roce 2018 byl oproti roku 2017 zaznamenán výrazný nárůst právě v počtu nelegálně zaměstnaných cizinců, v roce 2019 byl tento počet prakticky shodný s rokem předchozím. Počet zjištěných nelegálně zaměstnaných cizinců tak i nadále vysoce převyšuje počet nelegálně pracujících občanů ČR a EU. Pokračování tohoto trendu odráží přetrvávající situaci na trhu práce a nedostatek pracovních sil.

[Prohlížeč dat](#) [+ Tableau viewer](#)

Zakomponovat



Obrázek 7: Ukázka vygenerovaného grafu

Organizace / Český statistický úřad / Obyvatelstvo v ČR / Vše o území

**Vše o území** Spravovat Přejít na datový zdroj

URL: [https://vdb.czso.cz/vdbvo2/faces/cs/index.jspx?\\_af=profil-uzemi#w=](https://vdb.czso.cz/vdbvo2/faces/cs/index.jspx?_af=profil-uzemi#w=)

Demografický vývoj vybraného území

Demografický vývoj ve vybraném území Zakomponovat

**ČESKÝ STATISTICKÝ ÚŘAD** **Veřejná databáze** Časté dotazy Kontakty Nápověda

Statistiky **Vše o území** Vlastní výběr Ukazatele Registrace | Přihlášení

[Úvod](#) > [Vše o území](#) > [Abertamy \(okres Karlovy Vary\)](#)

**Abertamy (okres Karlovy Vary)** Výběr profilu a území Calý profil vybraného území

Demografický vývoj

**Stav obyvatel k 31.12.**

	2015	2016	2017	2018	2019
Počet obyvatel celkem	1 094	1 050	1 016	975	
v tom muži	551	533	520	507	
podle pohlaví ženy	543	517	496	468	
v tom 0-14	178	161	154	145	
ve věku (let) 15-64	770	737	709	662	
65 a více	146	152	153	168	
Průměrný věk	40,0	40,8	41,2	42,1	

Kód: PU-DEM-OB1/2

**Sňatky, rozvody, potraty**

	2015	2016	2017	2018	2019
Sňatky	1	6	4	5	
Rozvody	2	2	4	3	
Potraty	17	30	24	30	

Kód: PU-DEM-OB3/3

**Pohyb obyvatel**

	2015	2016	2017	2018	2019
Zvě narození	16	8	9	11	
Zemřelí	11	12	12	13	
Přistěhovalí	30	18	14	20	
Vystěhovalí	59	58	45	59	
Přírůstek (úbytek) přirozený	5	-4	-3	-2	
stěhováním	-29	-40	-31	-39	
celkový	-24	-44	-34	-41	

Kód: PU-DEM-OB2/3

**Zdroje**

- Obyvatelstvo ČR v ...
- Obyvatelstvo ČR v ...
- Obyvatelstvo ČR v ...
- Obyvatelstvo ČR v ...
- Obyvatelstvo ČR v ...
- Obyvatelstvo ČR v ...
- Vše o území**
- Informace o krajích

**Další informace**

Pole	Hodnota
Datum aktualizace dat	Listopad 23, 2020
Datum aktualizace metadat	Listopad 23, 2020
Vytvořeno	Listopad 23, 2020
Formát	HTML
Licence	Creative Commons CCZero

[Ukázat více](#)

[Sdílet](#) [Twitter](#)

Obrázek 8: Ukázka zobrazení webové stránky

Organizace / Český statistický úřad / Obyvatelstvo v ČR / Informace o krajích

## Informace o krajích

URI: [https://public.isbioau.com/profile/vubp.priha#Vizorna/Demografie\\_16063200857440/Krajinfo](https://public.isbioau.com/profile/vubp.priha#Vizorna/Demografie_16063200857440/Krajinfo)

Vizualizace poskytují demografické údaje v daném kraji, o názvu, průměrné mzdě a nezaměstnanosti

+ Vizualizace - info o krajích

\* demografické údaje \* nezaměstnanost \* průměrná mzda \* rozloha

Kraj info

Zakomponovat


Výběr kraje ČR

False

True

Kraj

Česká republika



Tableau

Zdroje

Obyvatelstvo ČR v ...

Obyvatelstvo ČR v ...

Obyvatelstvo ČR v ...

Obyvatelstvo ČR v ...

Obyvatelstvo ČR v ...

Vše o území

Informace o krajích

sdílet

Twitter

Facebook

### Další informace

Pole	Hodnota
Datum aktualizace dat	Listopad 25, 2020
Datum aktualizace metadat	Listopad 25, 2020
Vytvořeno	Listopad 25, 2020
Formát	tableau
Licence	Creativa Commons CCZero

Ukážet více

Obrázek 9: Ukázka zobrazení Tableau



## 2.4.2 Organizace (ORGANIZATIONS)

V tomto bloku jsou evidovány instituce, jejichž datové sady jsou v aplikaci využívány, viz výše.

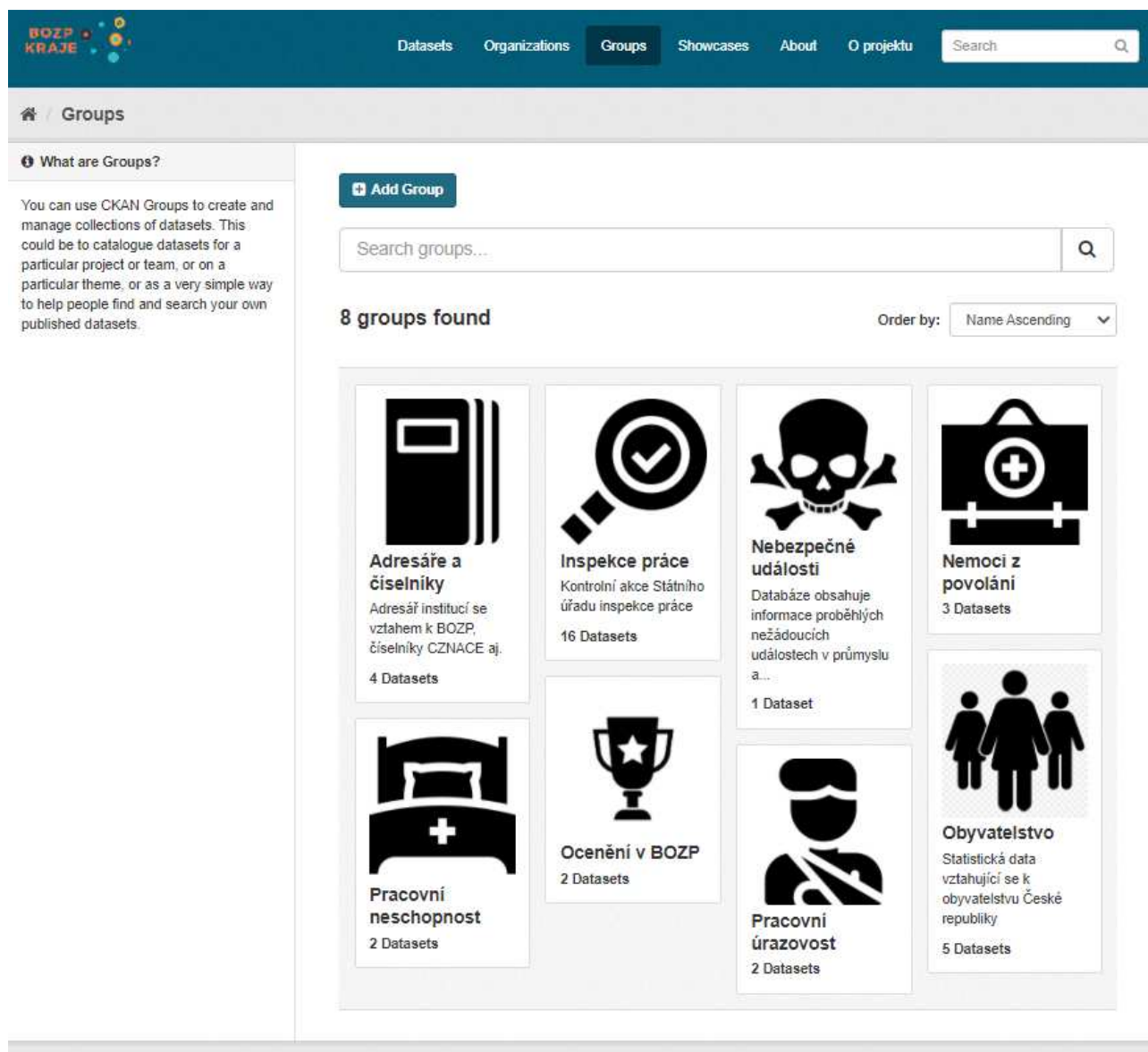
The screenshot shows the CKAN 'Organizace' page. The header includes navigation tabs: 'Datové sady', 'Organizace', 'Skupiny', 'Showcases', 'O nás', and 'O projektu'. A search bar is present with the text 'Vyhledat'. The main content area is titled 'Organizace' and contains a sidebar with the heading 'Co jsou organizace?' and a main section with a '+ Přidat organizaci' button and a search input 'Vyhledat organizace...'. Below the search bar, it states '8 organizací nalezeno' and 'Seřadit dle: Jména vzestupně'. The main content displays a grid of organization cards, each with a logo, name, brief description, and dataset count:

- Česká správa sociálního zařízení**: ČSSZ je největší a v rámci státní správy ČR zcela výjimečnou finančně správní... 3 datasety
- Český statistický úřad**: Český statistický úřad (ČSÚ) je ústředním orgánem státní správy České... 2 datasety
- Ministerstvo práce a sociálních věcí**: Do kompetence ministerstva patří především sociální politika (problematika... 0 datových sad
- Ministerstvo zdravotnictví**: Ministerstvo zdravotnictví 1 dataset
- Státní úřad inspekce práce**: Státní úřad inspekce práce a osm oblastních inspektorátů práce (OIP Hlavní... 7 datových sad
- Ústav zdravotnických informací a statistiky ČR**: Ústav zdravotnických informací a statistiky ČR („ÚZIS ČR“ nebo „Ústav“) je... 1 dataset
- Státní zdravotní ústav**: Státní zdravotní ústav je příspěvkovou organizací ministerstva zdravotnictví... 4 datasety
- Výzkumný ústav bezpečnosti práce**: Výzkumný ústav bezpečnosti práce, v. v. i., je stěžejní veřejnou výzkumnou... 4 datasety

Obrázek 10: Organizace v CKANu

### 2.4.3 Skupiny (GROUPS)

Tento blok obsahuje tematické skupiny pro snazší orientaci, např. pracovní úrazovost, pracovní neschopnost, adresáře, ocenění v BOZP atd.



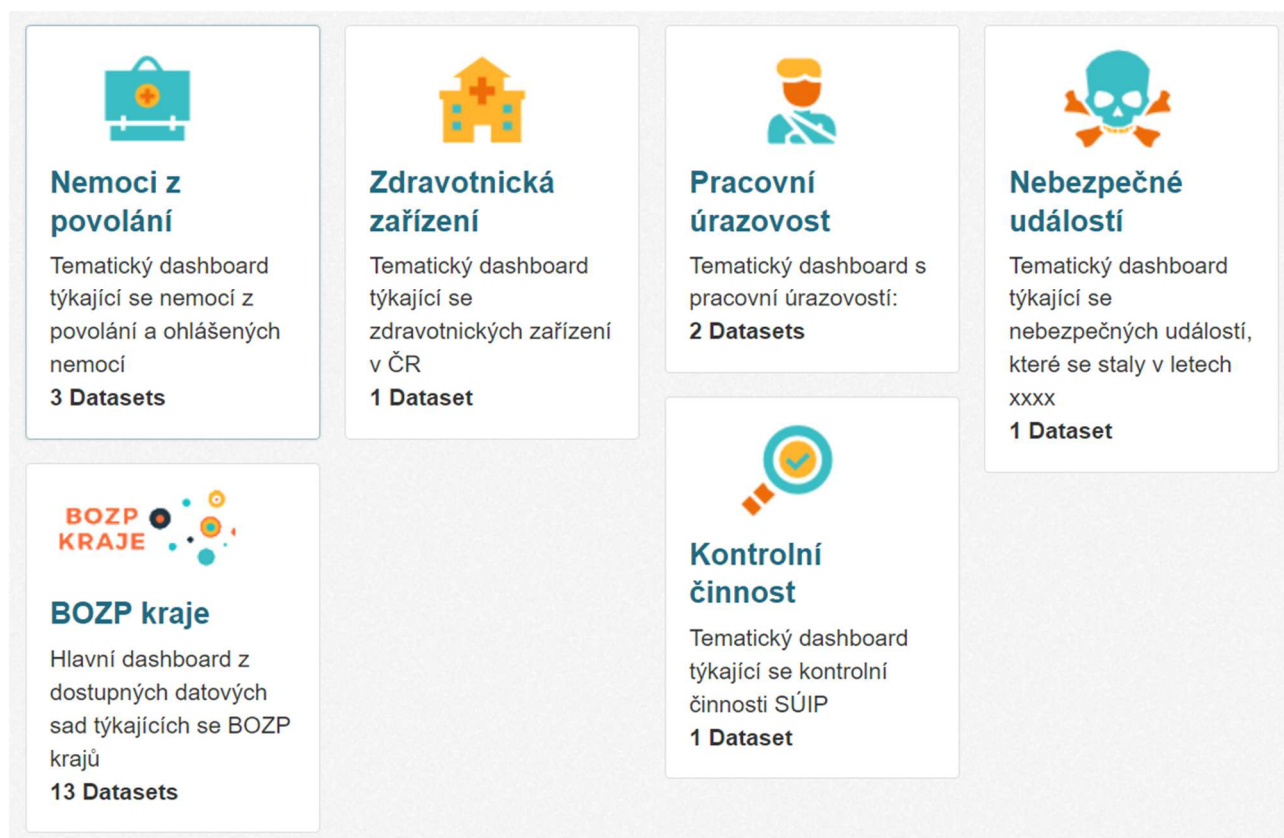
The screenshot shows the CKAN Groups interface. At the top, there is a navigation bar with 'Groups' selected. Below the navigation bar, there is a search bar and a list of 8 groups. Each group card includes an icon, a title, a brief description, and the number of datasets it contains.

Group Name	Description	Number of Datasets
Adresáře a číselníky	Adresář institucí se vztahem k BOZP, číselníky CZNACE aj.	4 Datasets
Inspekce práce	Kontrolní akce Státního úřadu inspekce práce	16 Datasets
Nebezpečné události	Databáze obsahuje informace proběhlých nežádoucích událostech v průmyslu a...	1 Dataset
Nemoci z povolání		3 Datasets
Pracovní neschopnost		2 Datasets
Ocenění v BOZP		2 Datasets
Pracovní úrazovost		2 Datasets
Obyvatelstvo	Statistická data vztahující se k obyvatelstvu České republiky	5 Datasets

Obrázek 11: Skupiny v CKANu

## 2.4.4 SHOWCASES

Tento blok je místem pro konkrétní tematické vizualizace. Obsahuje hlavní dashboard a dále dílčí vizualizace. Tyto vizualizace jsou zpřístupněny externím odkazem.



Obrázek 12: Showcases v CKANu

Všechny výše uvedené části CKANu budou průběžně rozšiřovány a aktualizovány podle toho, jak budou přibývat další publikovaná data zainteresovaných institucí.

Aktuálně je v katalogu uloženo:

- 35 datových sad (datasetů),
- 8 organizací,
- 8 tematických skupin,
- 6 showcases (tematických dashboardů) prezentujících data uložená v katalogu.



### 3. METODIKA PRO TVORBU DATOVÉHO MODELU A VIZUALIZACÍ (TABLEAU)

#### 3.1 Popis datových souborů

Zdrojovými daty pro webovou aplikaci - BOZP Kraje jsou otevřená data (open data). Nejčastěji ve formě xls, csv a případně i pdf souborů. Uložení v této formě je pro uživatele jednoduché a zároveň si každý uživatel může vytvořit v souborech vlastní strukturu, které nejlépe rozumí. Druhým důvodem, proč je tato forma oblíbená, je možnost rozdělit si data do více souborů (v případě xls do více sešitů), které lze poté odděleně analyzovat, a reportovat dále.

Na druhou stranu, chceme-li taková data analyzovat dohromady, vzniká několik problémů.

- Prvním je struktura dat, v csv nejsou data uspořádána do sloupců (vizuálně) ale jsou rozdělena oddělovačem, středníkem ( ; ), pomlčkou ( - ), čárkou ( , ) a dalšími typy. Nebo v případě údajů o datech se lze setkat s různými typy formátu data (DD-MM-RRRR, MM-RRRR, 1234-QQ-RRRR)<sup>5</sup>. V jedné z tabulek jsou hodnoty ukazatele za jednotlivé kraje rozděleny do sloupců, v druhé naopak do řádků.
- Druhým problémem může být množství dat. Mají-li být data v jedné tabulce xls, pak je třeba vzít v úvahu omezení Excelu na 16 384 sloupců a 1 048 576 řádků.
- Exportovat data z pdf souborů je často problematické, ne vždy stačí data zkopírovat a vložit (v mnoha případech se data zkopírují jako jeden řádek textu, který se musí náročně upravovat) a někdy se to neobejde bez ručního přepisování dat (v případě skenovaných pdf).

Největší výhodou otevřených dat je dostupnost, avšak jejich struktura je soubor od souboru jiná. Některé soubory vznikají vygenerováním dat v rámci určitého programu, jiné jsou ručně sestavenými sadami informací z různých zdrojů. Záleží také na tom, zda jsou otevřená data dostupná v tzv. surové formě nebo se jedná o určitou interpretaci dat.

Surová data jsou vhodná pro podrobnější analýzu, interpretovaná data jsou vhodná pro zjištění konkrétní informace či převzetí těchto dat do kontextu dalších dat (ad ruční sestavení o několik řádků výše).

Například data z Českého statistického úřadu (ČSÚ) lze získat na portálu této instituce jako předem definované agregace v určitém srovnání (např. tab. 9-105: Zaměstnanci a jejich průměrné hrubé měsíční mzdy podle krajů v roce 2018). Struktura této tabulky už je interpretací dat pro koncového uživatele, není však vhodná pro analytické nástroje typu Tableau.

---

<sup>5</sup> DD značí den, MM značí měsíc, QQ značí kvartál, RRRR značí rok



	A	B	C	D	E	F
1	<b>TRH PRÁCE</b>					<b>LABOUR MARKET</b>
2						
3	<b>9-105. Zaměstnanci a jejich průměrné hrubé měsíční mzdy podle krajů v roce 2018<sup>1)</sup></b>					
4	<i>Employees and their average gross monthly wages by region in 2018<sup>2)</sup></i>					
5						
6	ČR, kraje CR, Region	Průměrný evidenční počet zaměstnanců (tis. osob) <i>Average registered number of employees (thousand persons)</i>		Průměrná hrubá měsíční mzda (Kč) <i>Average gross monthly wage (CZK)</i>		
7		přepočtené počty <i>FTE persons</i>	fyzické osoby <i>Headcount</i>	přepočtené počty <i>FTE persons</i>	fyzické osoby <i>Headcount</i>	
8	Česká republika <sup>1)</sup> <i>Czech Republic<sup>1)</sup></i>	4 071,0	4 211,9	31 868	30 802	
9	Hl. m. Praha	823,7	861,2	40 059	38 316	
10	Středočeský	412,1	425,0	32 393	31 407	
11	Jihočeský	223,9	230,9	28 870	27 996	
12	Plzeňský	215,2	222,1	31 032	30 069	
13	Karlovarský	90,0	93,1	27 974	27 043	
14	Ústecký	251,0	257,9	29 290	28 507	
15	Liberecký	148,1	152,9	29 682	28 755	
16	Královéhradecký	202,2	208,9	29 622	28 668	
17	Pardubický	185,1	190,8	28 685	27 828	
18	Vysočina	175,7	180,7	29 256	28 447	
19	Jihomoravský	467,4	484,5	30 736	29 650	
20	Olomoucký	225,4	233,4	28 646	27 660	
21	Zlínský	210,1	216,1	28 427	27 633	
22	Moravskoslezský	438,3	451,5	28 796	27 954	
23						
24	<sup>1)</sup> pracovníštní metoda, předběžné údaje			<sup>2)</sup> The workplace method is applied here. Preliminary data.		
25	<sup>1)</sup> zahrnutí jsou i zaměstnanci sledovaných subjektů pracující mimo území ČR			<sup>2)</sup> Includes employees of of reporting units active outside the CR territory.		

Obrázek 13: Ukázka interpretovaných dat z ČSÚ

Analytické nástroje si s takovou tabulkou neporadí, proto je nutné taková data upravit již ve zdroji (nejčastěji s využitím programu Excel). Nejprve je třeba odstranit přebytečné informace - v tomto konkrétním případě prvních 5 řádků a řádky 24 a 25, dále zrušit sloučení buněk a předělat text v názvu sloupců (např. průměrný počet zaměstnanců přepočet na FTE), případně smazat anglické výrazy, pokud nejsou nutné pro koncového uživatele.

	A	B	C	D	E
1	Kraj	FTE	fyzické osoby	Průměrná mzda (FTE)	Průměrná mzda (FO)
2	Česká republika	4 071,0	4 211,9	31 868	30 802
3	Hl. m. Praha	823,7	861,2	40 059	38 316
4	Středočeský	412,1	425,0	32 393	31 407
5	Jihočeský	223,9	230,9	28 870	27 996
6	Plzeňský	215,2	222,1	31 032	30 069
7	Karlovarský	90,0	93,1	27 974	27 043
8	Ústecký	251,0	257,9	29 290	28 507
9	Liberecký	148,1	152,9	29 682	28 755
10	Královéhradecký	202,2	208,9	29 622	28 668
11	Pardubický	185,1	190,8	28 685	27 828
12	Vysočina	175,7	180,7	29 256	28 447
13	Jihomoravský	467,4	484,5	30 736	29 650
14	Olomoucký	225,4	233,4	28 646	27 660
15	Zlínský	210,1	216,1	28 427	27 633
16	Moravskoslezský	438,3	451,5	28 796	27 954

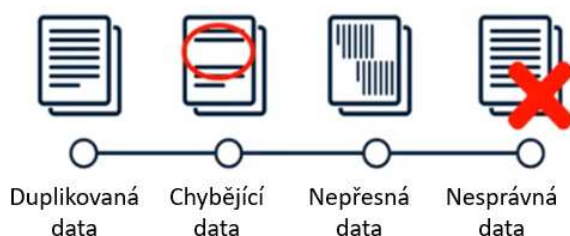
Obrázek 14: Upravená zdrojová tabulka

S takto upravenou tabulkou už si analytické nástroje poradí a lze s ní pracovat dále.

### 3.2 Datový model

Datový model je zjednodušeně řečeno schéma či popis, jaké datové zdroje jsou použity a jakým způsobem jsou propojeny, případně upraveny. U většiny nástrojů je nutné vždy vytvořit datový model a s ním následně pracovat jako s celkem. Platforma Tableau od verze 2020.2 dává uživateli možnost pracovat i s dílčími datovými modely (případně jen tabulkami samostatně) či datovým modelem jako celkem.

Zvláštní kapitolou při tvorbě datového modelu je čištění dat. Častokrát lze narazit na to, že tabulky mezi sebou nelze spojit, protože záznamy, které k sobě mají pasovat, nejsou stejné. Za nečisté datové zdroje lze považovat takové, které obsahují:



Obrázek 15: Nečistoty v datech

V řadě případů se na nekonzistenci v datech přichází právě ve fázi tvorby datového modelu, avšak nejsou výjimky, kdy se na to přijde až v průběhu analýzy. V takovém případě se musí celý proces opakovat, a ne vždy je to otázka několika kliknutí, či minut. Pro tvorbu datového modelu budou používány dva nástroje či produkty z platformy Tableau.

Prvním je nástroj Tableau Prep, který slouží k tvorbě datových modelů a zároveň obsahuje čisticí procesy. Velkou výhodou je zachování informací o provedených změnách (odstranění záznamů, přejmenování, změna typu atd.) a také vizuální forma datového modelu (tzv. flow).

Druhým je nástroj Tableau Desktop, který je primárně určen pro analýzu dat a tvorbu dashboardů. Umožňuje nahrát několik datových zdrojů současně, a ty mezi sebou propojit v případě, kdy je to potřeba. Je tedy možné analyzovat či tvořit nad jedním datovým zdrojem a případně tuto analýzu rozšířit o informace z jiného datového zdroje. Uživatel se tak nemusí vracet o několik kroků a přidávat zdroj do širokého datového modelu.

V rámci této metodiky jsou používány pojmy dílčí datový model a globální datový model.

**Dílčí datový model** bude vytvořen pro určitou tematickou část se společným druhem informací (např. pro pracovní neschopnost, pro informace o krajích, pro kontrolní činnosti apod.). Tyto dílčí datové modely budou tvořeny v Tableau Prep.

**Globální datový model** bude vytvořen v Tableau Desktop a bude obsahovat pouze informaci o spojení mezi dílčími datovými modely či tabulkami. Jedná se o určení jednoho či více polí (sloupců), které jsou oběma tabulkám/modelům společné, tímto krokem se taktéž určí úroveň agregace pro analýzu. Např. obsahuje-li jeden datový zdroj pole *Rok*, *Měsíc*, *Den* a druhý obsahuje pouze pole s názvem *Měsíc*, lze tyto dva zdroje spojit přes pole *Měsíc* a data, která budou analyzována, budou agregována na úroveň měsíce a na detail dnu se již nelze dostat.



Datový model je možno využít i pro jiné projekty či analýzy. Kromě této metodiky budou autorům dostupná i jednotlivá flow z Tableau Prep, která může později kdokoli, kdo se s tímto nástrojem chce naučit, prostudovat a zopakovat jednotlivé kroky.

### 3.3 Úprava zdrojových tabulek

Kromě již zmíněných úprav tabulek z ČSÚ pro tvorbu dílčího datového modelu Informace o krajích, bylo nutné upravit také tabulky s ukazateli BOZP, např. tabulka 4.3 PÚ odvětví 2015-2019.

	A	B	C	D	E	F	G	H	I	J	K	L					M								
												průměrné procento pracovní neschopnosti graf č. 4.3.3					průměrná doba trvání případu graf č. 4.3.4								
1	Časové řady ukazatelů pracovní úrazovosti s pracovní neschopnosti ve vybraných odvětvích v ČR v letech 2015-2019																								
2	odvětví (CZ-NACE) kromě J, K, L, M, N, O, R, S, U					počet pracovních úrazů s pracovní neschopností graf č. 4.3.1					četnost případů na 100 pojištěnců graf č. 4.3.2					průměrné procento pracovní neschopnosti graf č. 4.3.3					průměrná doba trvání případu graf č. 4.3.4				
3		2015	2016	2017	2018	2019	2015	2016	2017	2018	2019	2015	2016	2017	2018	2019	2015	2016	2017	2018	2019				
4	A. Zemědělství, lesnictví, rybnářství	2 626	2 592	2 613	2 608	2 473	2,36	2,36	2,36	2,35	2,25	0,397	0,390	0,371	0,384	0,383	61,28	60,54	57,46	59,48	61,93				
5	B. Těžba a dobývání	399	445	396	409	376	1,31	1,59	1,55	1,67	1,59	0,455	0,507	0,511	0,491	0,519	126,91	116,51	120,18	107,15	118,84				
6	C. Zpracovatelský průmysl	20 045	20 770	20 286	19 757	18 299	1,69	1,72	1,65	1,59	1,48	0,227	0,232	0,224	0,223	0,214	49,02	49,37	49,66	51,40	52,61				
7	D. Vým. a rozvod el. a j.e.	119	136	177	151	163	0,37	0,42	0,51	0,40	0,43	0,082	0,079	0,093	0,075	0,074	80,77	69,32	66,50	67,90	62,09				
8	E. Zás. vodou; čn. s odp.	1 007	968	1 085	1 038	1 073	1,76	1,69	1,89	1,79	1,84	0,269	0,254	0,283	0,289	0,282	55,88	55,04	54,71	58,92	56,16				
9	F. Stavebnictví	2 984	2 965	2 940	2 817	2 809	1,27	1,27	1,26	1,20	1,19	0,248	0,236	0,231	0,214	0,211	71,26	67,81	66,83	65,00	64,69				
10	G. Velkoobchod, maloobchod; opravy motor. voz.	4 876	5 277	5 392	5 326	5 347	0,80	0,85	0,86	0,84	0,86	0,116	0,120	0,122	0,120	0,123	52,79	51,67	51,73	51,70	52,38				
11	H. Doprava a skladování	4 142	4 332	4 604	4 574	4 477	1,48	1,50	1,55	1,53	1,49	0,272	0,256	0,274	0,277	0,279	67,29	62,60	64,73	65,99	68,45				
12	I. Ubytování, stravování a pohostinství	868	956	1 020	988	1 039	0,65	0,70	0,71	0,67	0,68	0,085	0,086	0,084	0,089	0,085	47,80	44,96	42,96	48,55	45,61				
13	P. Vzdělávání	1 303	1 334	1 404	1 353	1 454	0,38	0,39	0,40	0,37	0,39	0,054	0,057	0,053	0,053	0,058	51,57	53,61	48,22	51,99	54,69				
14	Q. Zdravotní a sociální péče	2 152	2 174	2 195	2 193	2 230	0,63	0,62	0,60	0,60	0,60	0,104	0,100	0,099	0,096	0,102	60,30	58,98	58,42	58,50	62,58				
15	ČR*	46 331	47 379	47 491	46 223	44 552	1,03	1,04	1,02	0,98	0,94	0,156	0,154	0,151	0,149	0,147	55,44	54,35	54,39	55,62	56,85				
16	*) Údaje se vztahují ke všem odvětvím.																								

Obrázek 16: Neupravená tabulka

Upraveno do následující struktury:

	A	B	C	D	E	F
1	odvětví (CZ-NACE) kromě J, K, L, M, N, O, R, S, U	Rok	počet pracovních úrazů s pracovní neschopností graf č. 4.3.1	četnost případů na 100 pojištěnců graf č. 4.3.2	průměrné procento pracovní neschopnosti graf č. 4.3.3	průměrná doba trvání případu graf č. 4.3.4
2	A. Zemědělství, lesnictví, rybnářství	2015	2 626	2,36	0,397	61,28
3	B. Těžba a dobývání	2015	399	1,31	0,455	126,91
4	C. Zpracovatelský průmysl	2015	20 045	1,69	0,227	49,02
5	D. Vým. a rozvod el. a j.e.	2015	119	0,37	0,082	80,77
6	E. Zás. vodou; čn. s odp.	2015	1 007	1,76	0,269	55,88
7	F. Stavebnictví	2015	2 984	1,27	0,248	71,26
8	G. Velkoobchod, maloobchod; opravy motor. voz.	2015	4 876	0,80	0,116	52,79
9	H. Doprava a skladování	2015	4 142	1,48	0,272	67,29
10	I. Ubytování, stravování a pohostinství	2015	868	0,65	0,085	47,80
11	P. Vzdělávání	2015	1 303	0,38	0,054	51,57
12	Q. Zdravotní a sociální péče	2015	2 152	0,63	0,104	60,30
13	ČR*	2015	46 331	1,03	0,156	55,44
14	A. Zemědělství, lesnictví, rybnářství	2016	2 592	2,36	0,390	60,54
15	B. Těžba a dobývání	2016	445	1,59	0,507	116,51
16	C. Zpracovatelský průmysl	2016	20 770	1,72	0,232	49,37
17	D. Vým. a rozvod el. a j.e.	2016	136	0,42	0,079	69,32
18	E. Zás. vodou; čn. s odp.	2016	968	1,69	0,254	55,04
19	F. Stavebnictví	2016	2 965	1,27	0,236	67,81
20	G. Velkoobchod, maloobchod; opravy motor. voz.	2016	5 277	0,85	0,120	51,67
21	H. Doprava a skladování	2016	4 332	1,50	0,256	62,60
22	I. Ubytování, stravování a pohostinství	2016	956	0,70	0,086	44,96
23	P. Vzdělávání	2016	1 334	0,38	0,057	53,61
24	Q. Zdravotní a sociální péče	2016	2 174	0,62	0,100	58,98
25	ČR*	2016	47 379	1,04	0,154	54,35
26	A. Zemědělství, lesnictví, rybnářství	2017	2 613	2,36	0,371	57,46
27	B. Těžba a dobývání	2017	396	1,55	0,511	120,18
28	C. Zpracovatelský průmysl	2017	20 286	1,65	0,224	49,66
29	D. Vým. a rozvod el. a j.e.	2017	177	0,51	0,093	66,50
30	E. Zás. vodou; čn. s odp.	2017	1 085	1,89	0,283	54,71
31	F. Stavebnictví	2017	2 940	1,26	0,231	66,83

Obrázek 17: Upravená tabulka

Úpravy:

- odstranění prvního a posledního řádku (název tabulky, poznámky),
- zrušení sloučených buněk, a zbarvení textu,
- vytvoření nového sloupce B s názvem Rok,
- zkopírování názvů odvětví (A3-A14), vložení 4krát (pod sebe),



- do sloupce *Rok* vložit rok 2015, zkopírovat pod sebe 11krát, pak pro další rok a opakovat stejný postup,
- vyjmutí hodnot ze sloupce D a vložení do sloupce C, stejně tak ze sloupce E, F a G,
- odstranění sloupce D-G a opakovat předchozí krok, akorát s tím, že se posuneme o 1 sloupec vpravo,

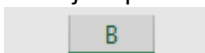
Výsledkem bude tabulka s 6 sloupci (název odvětví, rok, 4 sloupce s ukazateli).

TIP

Pro kopírování stejného názvu lze v programu Excel zatáhnout za spodní roh (čtvereček) a táhnout myší dolů



Místo sloučení buněk je lepší buňky roztáhnout, dvojitým kliknutím se sloupec rozšíří dle nejdelšího textu v daném sloupci



V případě že šlo o tabulku s jedním ukazatelem (např. 6.3 četnost PÚ ženy 2019), byl vytvořen nový sloupec.

	A	B	C	D	E	F
1	odvětví (CZ-NACE)	ženy/všichni	Četnost PÚ u žen			
2	A. Zemědělství, lesnictví, zemědělné služby	ženy	2,00			
3	B. Těžba a dobývání	ženy	0,61			
4	C. Zpracovatelský průmysl	ženy	0,98			
5	D. Výroba a rozvod elektřiny a tepla	ženy	0,23			
6	E. Zásobování vodou; činnosti v oblasti nemovitosti	ženy	0,62			
7	F. Stavebnictví	ženy	0,22			
8	G. Velkoobchod, maloobchod; opravy a výměna motorových vozidel	ženy	0,69			
9	H. Doprava a skladování	ženy	1,22			
10	I. Ubytování, stravování a pohostinství	ženy	0,67			
11	J. Informační a komunikační	ženy	0,06			
12	K. Peněžnictví a pojišťovnictví	ženy	0,12			
13	L. Činnosti v oblasti nemovitosti	ženy	0,18			
14	M. Profesní, vědecké a administrativní	ženy	0,13			
15	N. Administrativní a podpůrné	ženy	0,63			
16	O. Veřejná správa a obrana; mezinárodní organizace	ženy	0,26			
17	P. Vzdělávání	ženy	0,43			
18	Q. Zdravotní a sociální péče	ženy	0,57			
19	R. Kulturní, zábavní a rekreační	ženy	0,40			
20	S. Ostatní činnosti	ženy	0,30			
21	nezjištěno	ženy	0			
22	ČR	ženy	0,62			
23	A. Zemědělství, lesnictví, zemědělné služby	všichni	2,25			
24	B. Těžba a dobývání	všichni	1,59			
25	C. Zpracovatelský průmysl	všichni	1,48			
26	D. Výroba a rozvod elektřiny a tepla	všichni	0,43			
27	E. Zásobování vodou; činnosti v oblasti nemovitosti	všichni	1,84			
28	F. Stavebnictví	všichni	1,19			
29	G. Velkoobchod, maloobchod; opravy a výměna motorových vozidel	všichni	0,86			
30	H. Doprava a skladování	všichni	1,49			
31	I. Ubytování, stravování a pohostinství	všichni	0,68			
32	J. Informační a komunikační	všichni	0,09			
33	K. Peněžnictví a pojišťovnictví	všichni	0,10			
34	L. Činnosti v oblasti nemovitosti	všichni	0,30			
35	M. Profesní, vědecké a administrativní	všichni	0,21			

	A	B	C	D	E	F
1	Četnost pracovních úrazů s pracovní neschopností žen (na 100 pracovních žen)	ženy	všichni			
2	odvětví (CZ-NACE)	ženy	všichni			
3	A. Zemědělství, lesnictví, zemědělné služby	2,00	2,25			
4	B. Těžba a dobývání	0,61	1,59			
5	C. Zpracovatelský průmysl	0,98	1,48			
6	D. Výroba a rozvod elektřiny a tepla	0,23	0,43			
7	E. Zásobování vodou; činnosti v oblasti nemovitosti	0,62	1,84			
8	F. Stavebnictví	0,22	1,19			
9	G. Velkoobchod, maloobchod; opravy a výměna motorových vozidel	0,69	0,86			
10	H. Doprava a skladování	1,22	1,49			
11	I. Ubytování, stravování a pohostinství	0,67	0,68			
12	J. Informační a komunikační	0,06	0,09			
13	K. Peněžnictví a pojišťovnictví	0,12	0,10			
14	L. Činnosti v oblasti nemovitosti	0,18	0,30			
15	M. Profesní, vědecké a administrativní	0,13	0,21			
16	N. Administrativní a podpůrné	0,63	0,84			
17	O. Veřejná správa a obrana; mezinárodní organizace	0,26	0,41			
18	P. Vzdělávání	0,43	0,39			
19	Q. Zdravotní a sociální péče	0,57	0,60			
20	R. Kulturní, zábavní a rekreační	0,40	0,48			
21	S. Ostatní činnosti	0,30	0,34			
22	nezjištěno	-	-			
23	ČR	0,62	0,94			

Obrázek 18: Úprava před a po

U dalších tabulek ve většině případů stačilo zrušit sloučené buňky a odstranit poté prázdné sloupce.

Dále je možné si zdrojové soubory dávat do podsložek, aby autor při jejich nahrávání do Tableau Prep nehledal, kam byly uloženy nebo aby se na některou tabulku nezapomnělo. Např. pro kontroly byla vytvořena složka *Kontroly\_průmysl* a *Kontroly\_zařízení*, kde jsou uloženy xls soubory k těmto vybraným kontrolám.

### 3.3.1 Pivot

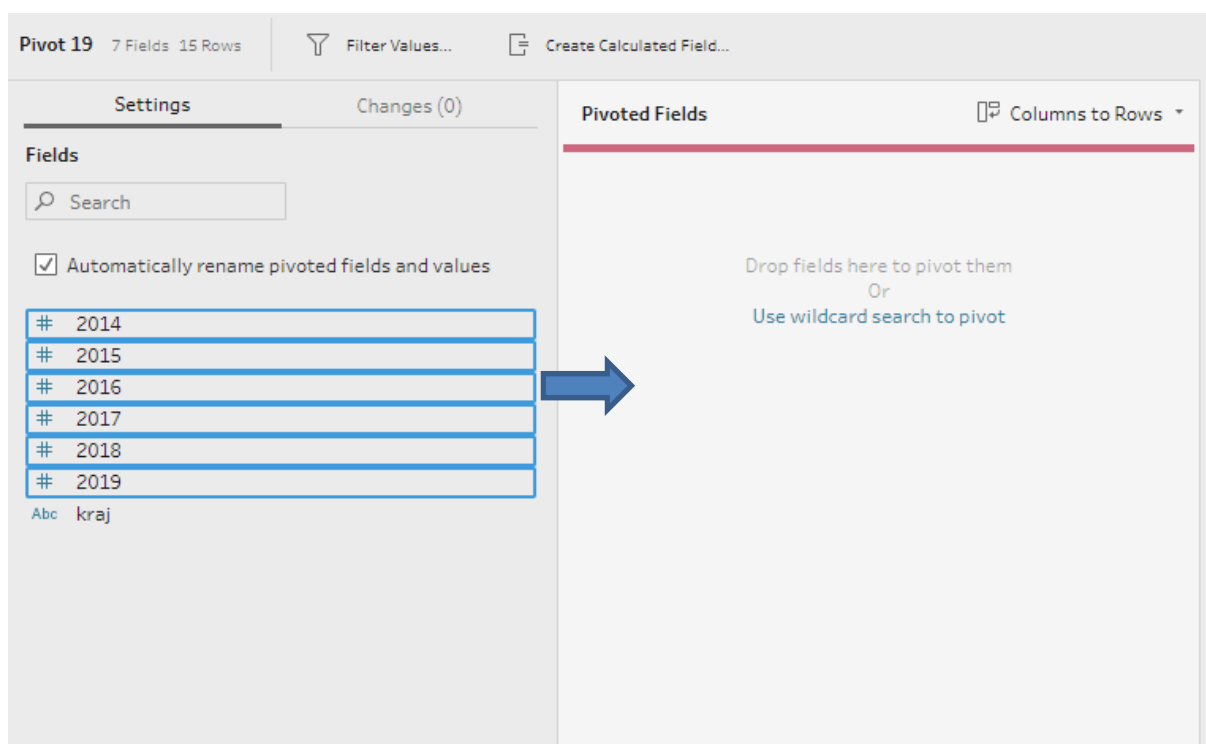
Častou úpravou zdrojových tabulek po nahrání do Tableau Prep byla funkce *Pivot*. Ta slouží k tomu, aby přenesla vícero sloupců do čitelnější struktury dvou sloupců – jeden s názvy a druhý s hodnotami.



Obrázek 19: Označení Pivot v Tableau Prep

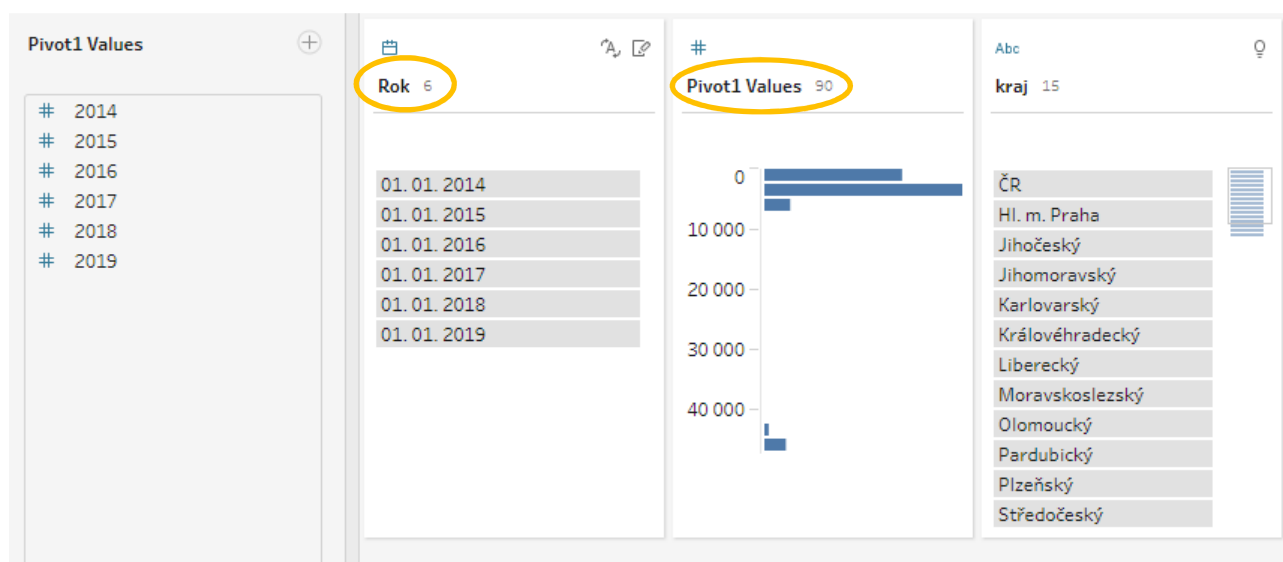
Například tabulka Počet PU s PN obsahuje 7 sloupců - Kraj, 2014, 2015, 2016, 2017, 2018, 2019.

Žádoucí stav pro analýzu je mít sloupce tři. V prostředí Tableau Prep lze označit sloupce 2014-2019 a přetáhnout je do prostoru *Pivot Fields*.



Obrázek 20: Pivot, krok 1

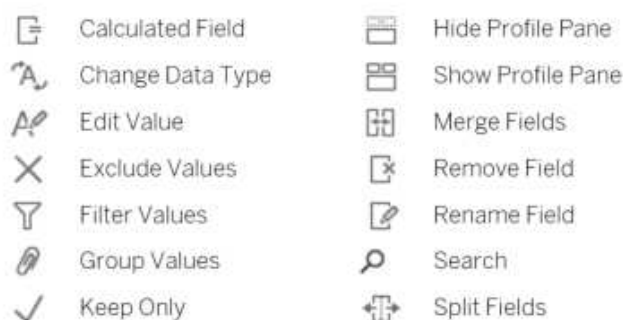
Tím se vytvoří dva sloupce – *Pivot Names* a *Pivot Values*. Sloupec *Pivot Names* bude obsahovat rok a sloupec *Pivot Values* danou hodnotu pro daný *Rok* a *Kraj*. Sloupce se dále přejmenují na *Pivot Names na Rok* a *Pivot Values na PU s PN* (tuto změnu lze provést také v čistícím kroku – následující kapitola).



Obrázek 21: Pivot, krok 2

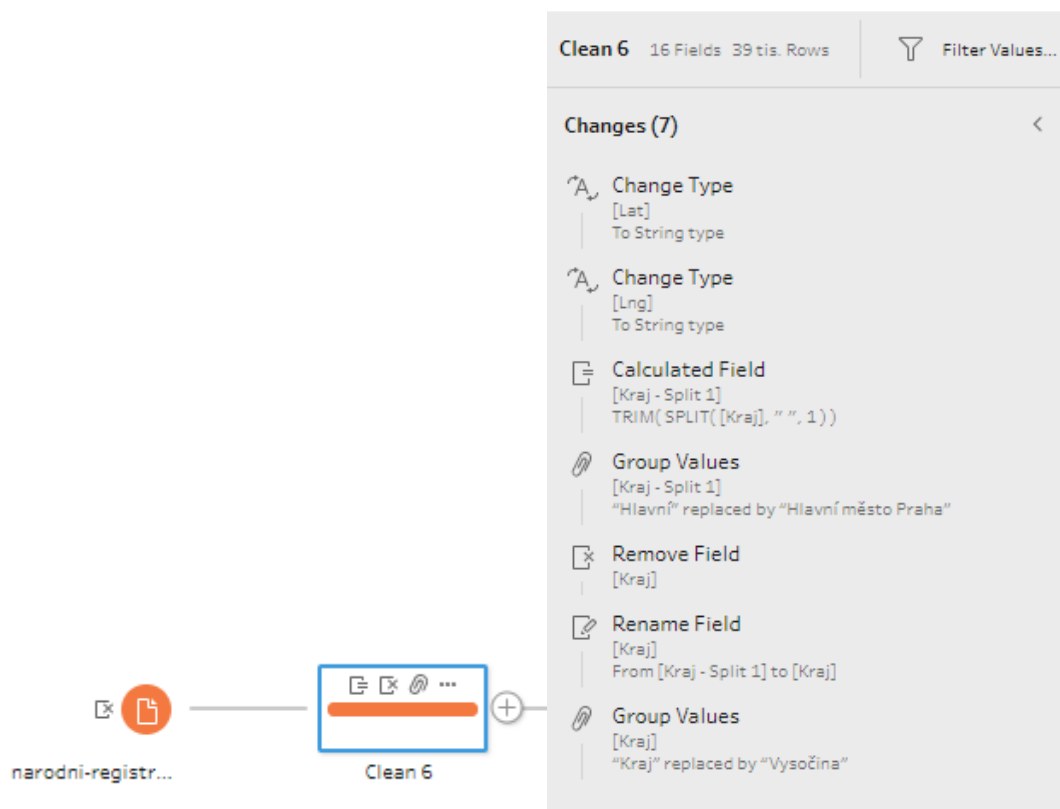
### 3.3.2 Čistící kroky

U žádného datového zdroje nelze mít 100 % čistotou či nekonzistencí. Ta může vzniknout na straně zadavatele hodnot/tvůrcem zdroje, nebo také importem dat do používaného nástroje. Např. u Tableau se lze často setkat s tím, že se přidělí datový typ sloupci s čísly jako tzv. *Number* (značeno #). Kontext dat ale může být jiný a číslo je v některých případech vhodnější jako tzv. *String* (značeno Abc). Typicky to takto bývá u sloupce PSČ, hodnota 28401 neznamená *dvacet osm tisíc čtyři sta jedna*, ale jde o kódové označení oblasti (v tomto případě Kutné Hory).



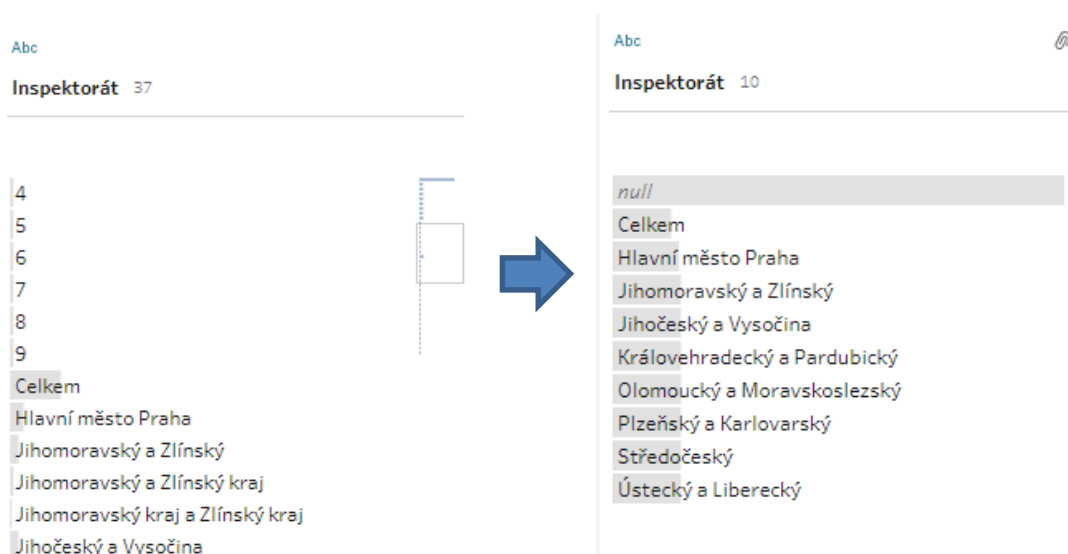
Obrázek 22: Možnosti čistících procesů v Tableau Prep

V rámci tvorby datových modelů bylo provedeno mnoho čistících kroků, ne však všechny výše uvedené na obrázku. Zároveň v každém kroku je vizuálně znázorněno, jaké úpravy byly třeba u daného zdroje - ikonami u daného čistícího kroku a taktéž v poli *Changes*.



Obrázek 23: provedené čistící procesy u vybraného datového zdroje

Po spojení dvou nebo více datových zdrojů (*Union*, nebo *Join* – bude vysvětleno v následující kapitole) občas dojde k tomu, že sice názvem stejný sloupec obsahuje trochu rozdílné hodnoty. Např. při spojení datových zdrojů o kontrolní činnosti získáme sloupec *Inspektorát*, ten však po spojení v Tableau Prep obsahuje 37 hodnot. Jednou se jedná o číslo inspektorátu, poté je o to jaké kraje má v gesci (ovšem někdy název obsahuje slovo „kraj“ jindy nikoliv). Proto zvolíme čistící krok „Group values“.



Obrázek 24: Group values

Posledním používaným čistícím procesem je rozdělení pole do více sloupců – *Split values*. Tato funkce byla použita v případě rozdělení názvu krajů, např. Středočeský kraj. Jednodušším pro tvorbu datového modelu je pracovat pouze s pojmem *Středočeský*, čili rozdělit sloupce tak, aby zůstal pouze první záznam (rozdělení se provede zvolením oddělovače ‚mezerník‘).

### 3.4 Spojení tabulek

Zdrojové tabulky, které jsou základem datového modelu, je třeba mezi sebou propojovat a to pomocí příkazů *Join* a *Union*.

#### 3.4.1 Join

- 1) Prvním typem propojení je spojení dvou tabulek vedle sebe, přičemž musí být jeden sloupec společný oběma tabulkám.



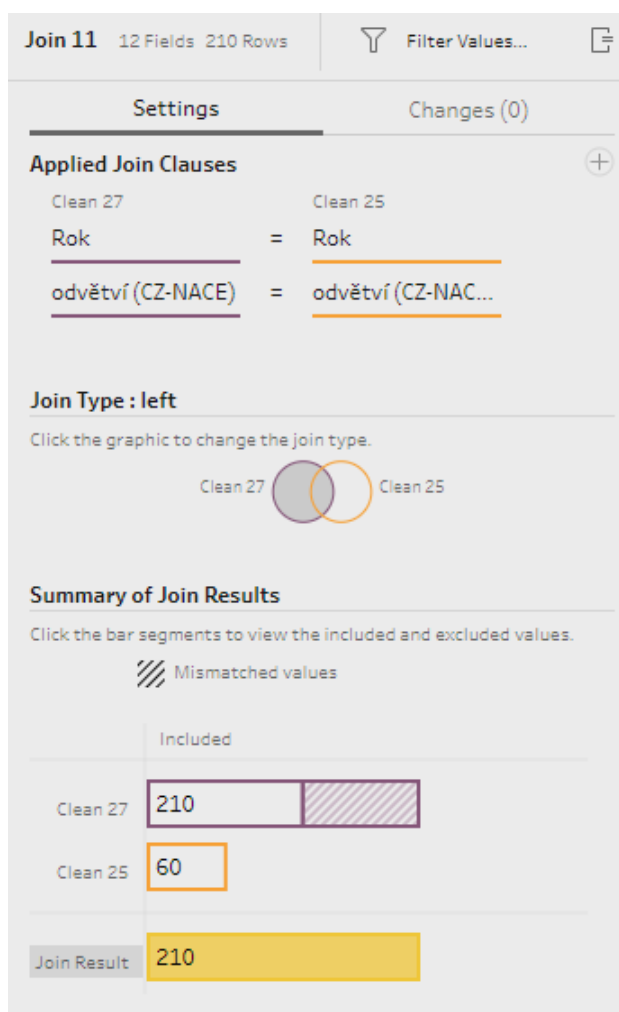
Obrázek 25: Typy spojení – join

Tableau Prep umožňuje všechny typy spojení. Uživatel volí, zda má být ve výsledné tabulce celá levá tabulka a doplňující informace z pravé, pouze pro záznamy, u kterých se shoduje – *Left join*, či opačná situace – *Right join*. Nebo pouze záznamy společné oběma tabulkám – *Inner join*. Další možností je i výběr záznamů, které se



neshodují – takové spojení však nedává moc smysl, a v případě datového modelu pro webovou aplikaci použít není.

Na tomto místě je důležité upozornit na situaci, kdy špatně zvoleným spojovacím sloupcem či sloupci, může dojít k násobení počtu řádků tzv., ke kartézskému součinu. Tableau Prep tomuto předchází tím, že zobrazí uživateli výsledný počet řádků po spojení.



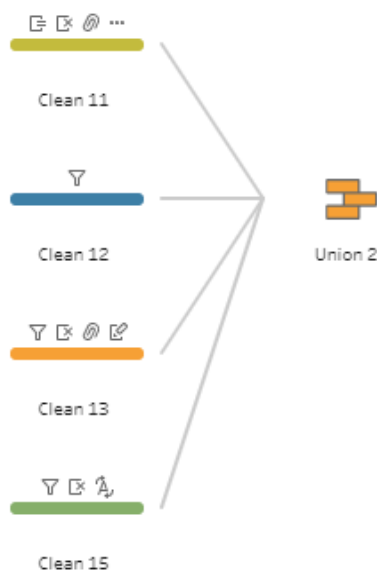
Obrázek 26: Join – definice spojení

*Join Clauses* (viz obr. 26) je volba spojovacího pole, v tomto konkrétním případě je nutné zvolit dvě spojovací pole (sloupce), při zvolení pouze jednoho by došlo k namnožení dat. Dále se volí typ spojení – v tomto případě *Left join*. Poslední část informuje o počtu řádků z jednotlivých tabulek a výsledný počet řádků.

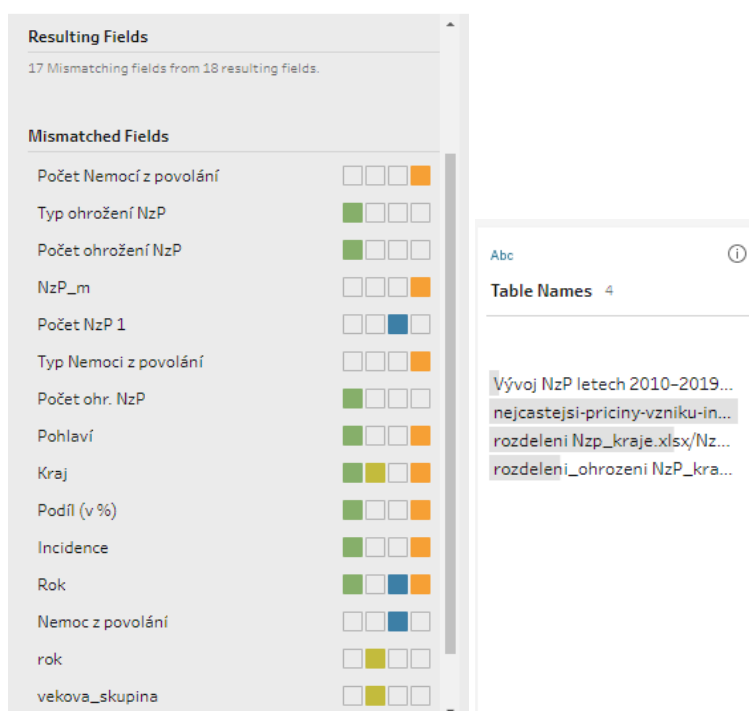
### 3.4.2 Union

Druhým typem propojení tabulek je spojení Union. Toto spojení uspořádává tabulky pod sebe, teoreticky nemusí mít ani jeden sloupec stejný, a přesto vytvoří výslednou tabulku (s tím, že do chybějících, resp. nespojených sloupců doplní *null* hodnoty). Prakticky by uživatel takové tabulky nespojoval, jelikož nemají spolu nic společného. Tableau Prep na rozdíl od Desktop (který tuto funkcionalitu má také) umí propojit tabulky funkcí *Union* z různých datových zdrojů (databáze, xls, cloud), zatímco Desktop umí propojit funkcí *Union* pouze tabulky stejného typu.

Zároveň je možné spojit více tabulek v jenom kroku. Názorným příkladem je spojení dat za roky, tabulku 2018, 2019 a 2020, kdy byl použit příkaz *Union* pod sebe. Předpokladem je, že tabulky mají identickou strukturu sloupců. Vznikne stejně široká, avšak dlouhá tabulka. V případě, že v roce 2020 byl do tabulky přidán nový sloupec, zobrazí se po spojení v tomto sloupci u roků 2019 a 2018 *null* hodnoty.



Obrázek 27: Union – spojení více zdrojů



Obrázek 28: Union – spojené tabulky

Obrázky znázorňují spojení 4 tabulek v Tableau Prep, přičemž v tomto případě se spojují tabulky, které mají různé sloupce, a jen u dvou sloupců se shodují. Těmi sloupci jsou *Kraj* a *Rok*, přičemž v jednom datovém zdroji je informace pouze o *Kraji*, ale nemá sloupec *Rok* (žlutá), a u druhého datového zdroje je to naopak (modrá). Zároveň v Tableau Prep vznikne nový sloupec *Table Names*, obsahující informaci, z kterého datového zdroje záznamy pocházejí. Díky tomu lze s daty dále pracovat a analyzovat je i samostatně (filtrováním tohoto sloupce).

### 3.5 Output

Výsledkem datového modelu je v určitém smyslu velká tabulka, tzv. *output*. V Tableau Prep jsou pouze dvě možnosti první je *Tableau data extract (.hyper, .tde)*, druhou je *Comma separated values (.csv)*.

Výhoda Tableau data extraktů je především v tom, že zlepšují výkon při načítání dat během jejich analýzy v Tableau Desktop. Omezením je naopak to, že ne každý analytický nástroj dokáže s takovým souborem pracovat, zatímco s *.csv* lze víceméně pracovat v jakémkoliv programu.

### 3.6 Dílčí datové modely

Zdrojová data jsou rozdělena do 6 skupin a z nich jsou poté vytvořeny dílčí datové modely. 4krát byl zvolen output ve formě *.hyper* a u 3 *.csv*. S každým datovým modelem bude v Tableau Desktop možné pracovat



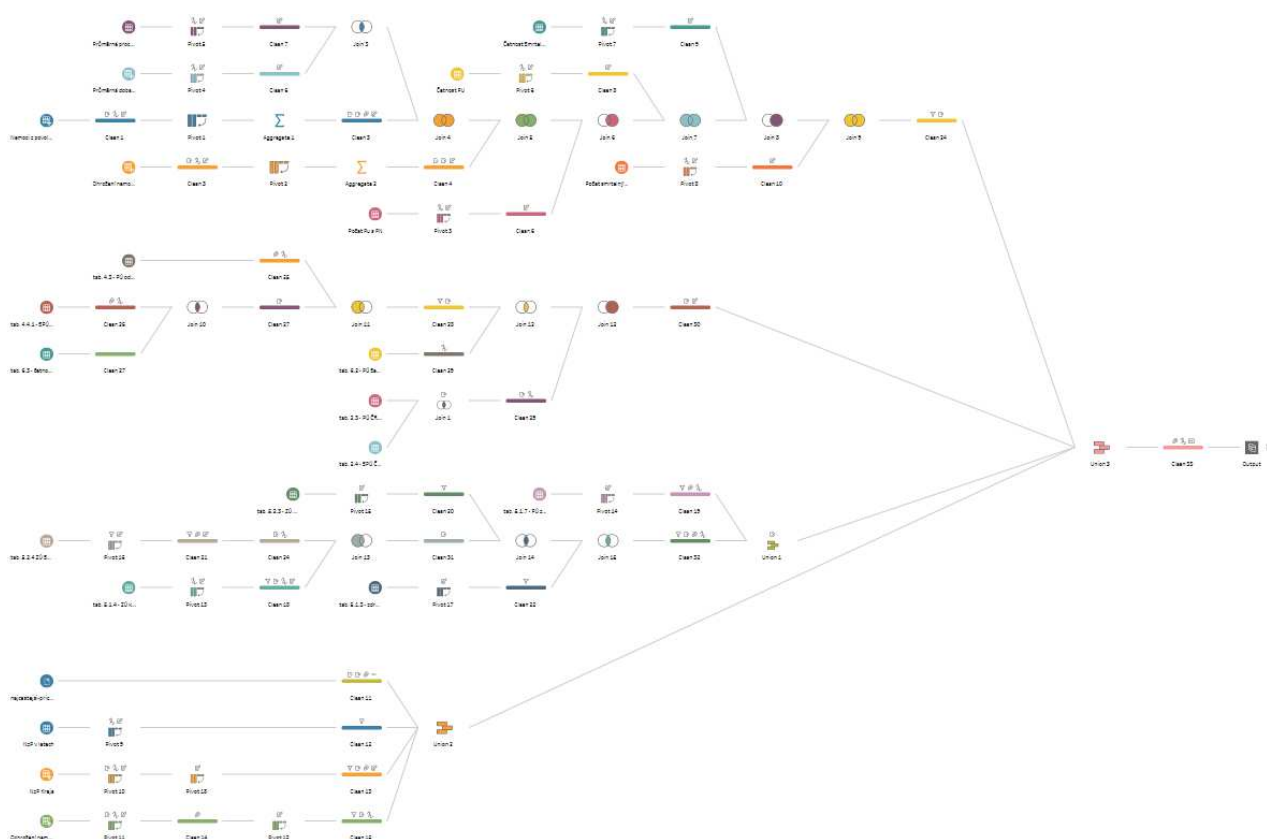
samostatně. Postačí spojit informaci do přehledového dashboardu, které podniky byly oceněné v jednotlivých krajích a zároveň ke kolika případům pracovních úrazů v daném kraji došlo.

Ve webové aplikaci jsou tyto dílčí datové modely:

- ukazatele BOZP,
- informace o krajích,
- kontrolní činnost,
- oceněné podniky,
- instituce BOZP a zdravotnické instituce,
- nebezpečné události.

### 3.6.1 Ukazatele BOZP

Počtem zdrojových tabulek, a také hlavním tématem webové aplikace, jsou ukazatele BOZP. Tento datový model čítá 23 zdrojů dat (xls a csv soubory), u nichž byly provedeny funkce *Pivot*, *Čistící kroky*, *Join* a *Union*. Výsledný output je Tableau data extrakt .hyper.

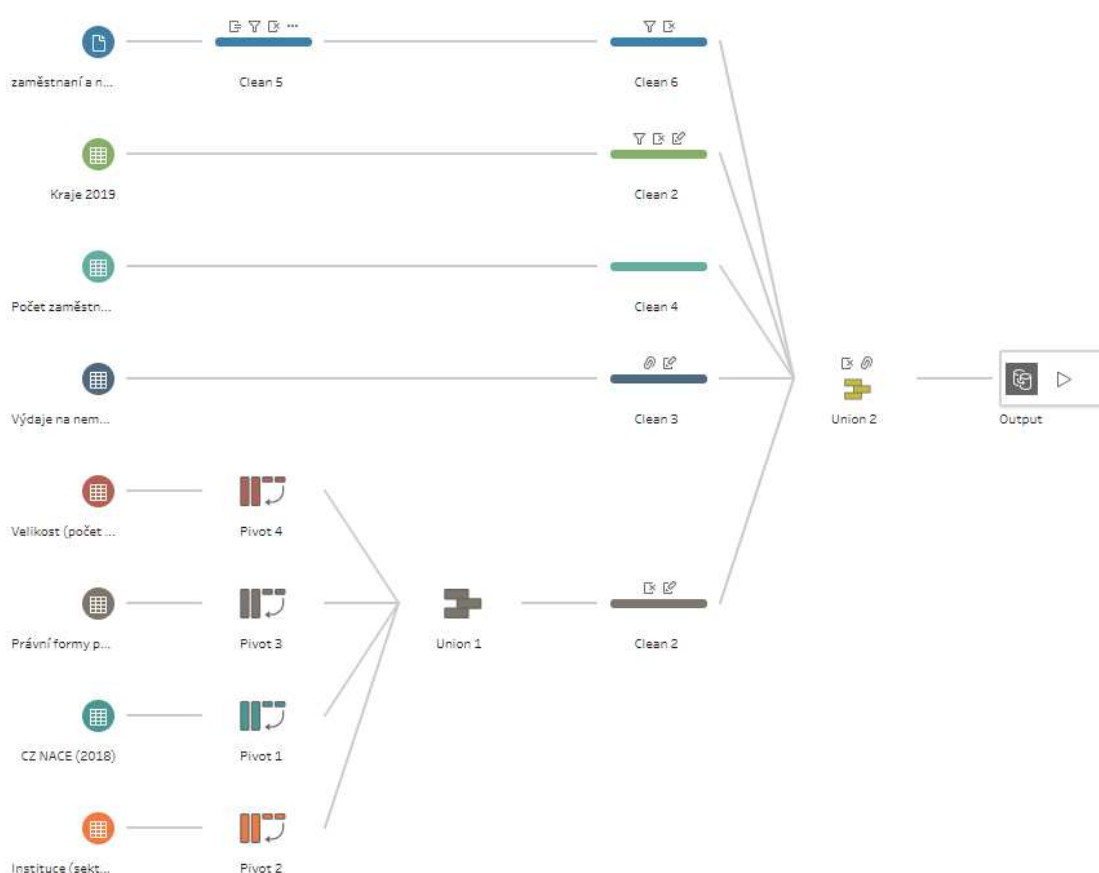


Obrázek 29: Dílčí datový model – ukazatele BOZP

### 3.6.2 Informace o krajích

Informace o BOZP ukazatelích v krajích je vhodné doplnit o údaje obecnějšího charakteru jako počet obyvatel daného kraje, rozloha, průměrný věk apod.

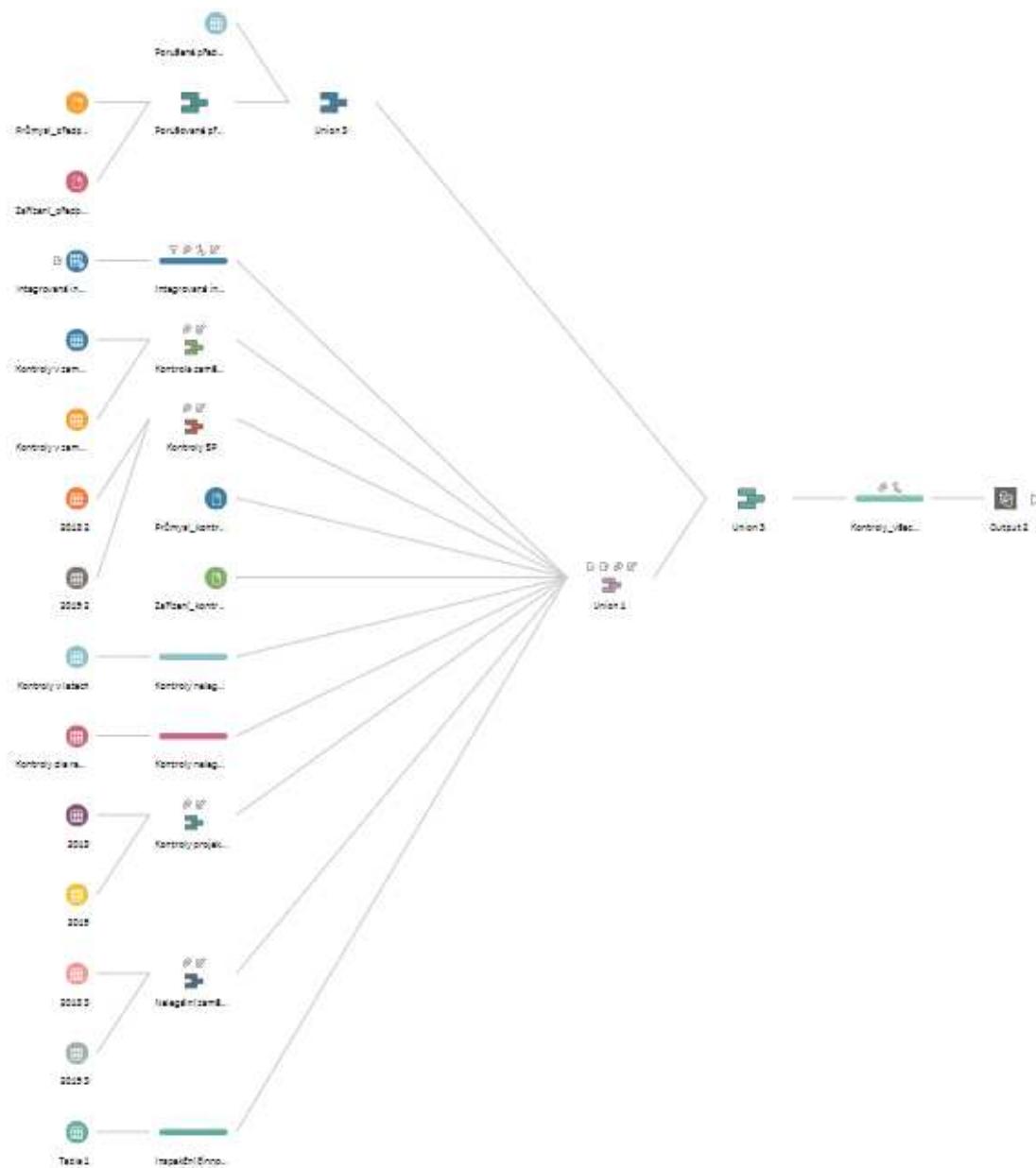
U tohoto datového modelu bylo více práce u zdrojových tabulek, především těch z ČSÚ. Samotné spojení dat bylo provedeno funkcí *Union*. Čistící kroky byly jen na přejmenování polí či sloučení některých hodnot.



Obrázek 30: Dílčí datový model – informace o krajích

### 3.6.3 Kontrolní činnost

Ke kontrolní činnosti je k dispozici celkem 17 zdrojových tabulek. U těchto zdrojových dat byla pro přehlednost provedena úprava v Tableau Prep a spojena příkazem *Union*. Toto řešení je nakonec zvoleno pro celkové spojení.



Obrázek 31: Dílčí datový model – kontrolní činnost

### 3.6.4 Oceněné podniky

Třetí kategorií byla data o podnicích oceněných v programu *Bezpečný podnik* a programu *Podnik podporující zdraví (PPZ)*. Spojení je provedeno pomocí funkce *Union*, aby bylo odlišeno, o jaké ocenění se jedná (při spojení *Join* by tato informace nebyla zachycena). Čistící procesy byly složitější u dat z programu PPZ, jelikož



sloupec *Rok* neobsahoval pouze jeden rok, ale více let, musel být rozdělen a následně funkcí *Pivot* znova spojen do lepší struktury. Stejný proces je třeba provést u sloupce *Kategorie*. Výsledným outputem je *hyper*.

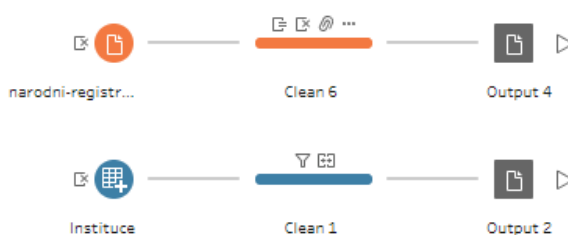


Obrázek 32: Dílčí datový model – oceněné podniky

### 3.6.5 Instituce BOZP a zdravotnická zařízení

Kontaktní informace ve webové aplikaci nemohou chybět, přičemž jde v první řadě o kontaktní údaje na instituce, které mají co dočinění s BOZP. V druhé řadě se vzhledem k datům o pracovních úrazech mohou uživatelům aplikace hodit informace o zdravotnických zařízeních v jejich okolí.

Čistící kroky, provedené ve zdroji o institucích BOZP, se týkají pouze filtrování záznamů ve sloupci *Kraj*, které tam nepatří. Dva sloupce obsahují stejnou informaci – datová schránka, ty jsou sjednoceny do jednoho. Čistící kroky u zdravotnických institucí se jako v předchozích případech věnovaly sloupci *Kraj*. Výslednými outputy jsou .csv soubory.

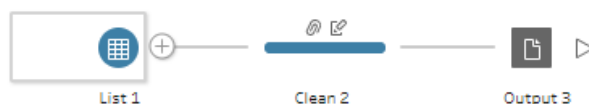


Obrázek 33: Dílčí datové modely – instituce BOZP a zdravotnická zařízení

### 3.6.6 Nebezpečné události

V průběhu řešení došlo na rozšíření datového modelu o data z portálu Major Accident Prevention Information System (MAPIS; tematicky zaměřený na prevenci závažných havárií). Po kliknutí otevře aplikace uživateli okno s webovou stránkou MAPIS a detailem události.

Z pohledu datového modelu se jedná o samostatný datový zdroj, který nedávalo smysl spojovat do některého z předchozích. Potřebné však bylo data vyčistit, především názvy krajů sjednotit do stejné struktury. Výsledným outputem je csv soubor.



Obrázek 34: Dílčí datový model – Nebezpečné události

### 3.7 Globální datový model

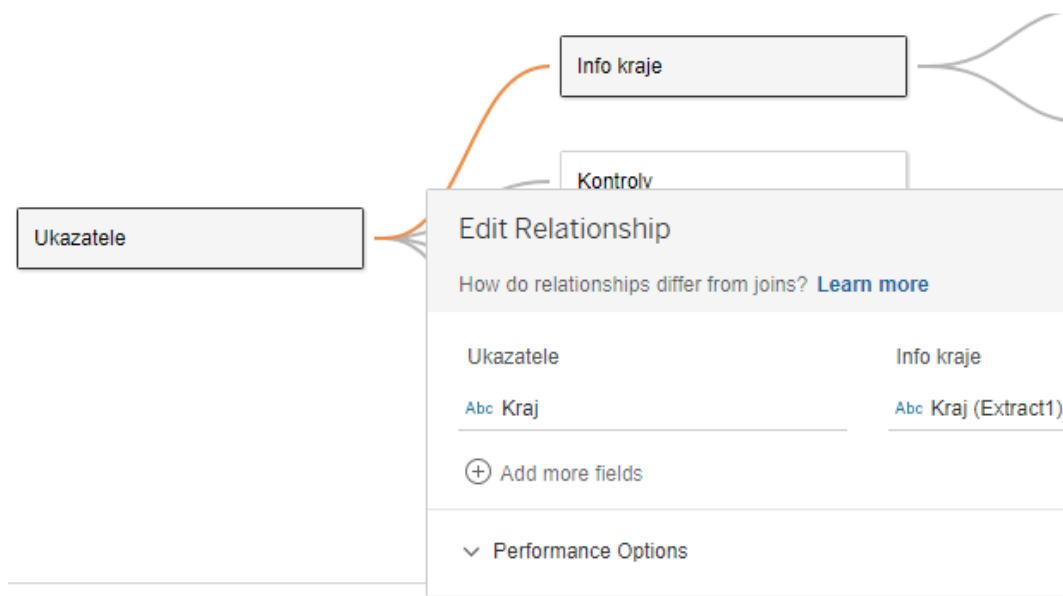
K propojení dílčích datových modelů, vytvořených v Tableau Prep, dochází až v nástroji Tableau Desktop. Přičemž toto propojení není fixní a nevytváří ve výsledku jednu velkou tabulku. Na rozdíl od spojení tabulek formou *Join* nebo *Union* je v Desktop využito jiné technologie a propojení tabulek se zde nazývá *Relationship*. Tato technologie umožňuje s jednotlivými tabulkami pracovat nezávisle a pouze v případě společné analýzy tyto tabulky k sobě spojí na úrovni zadaného *Relationship* klíče.



Obrázek 35: Globální datový model

Ve všech případech - až na jediný - je zvoleným klíčem pole *Kraj*. Onou jedinou výjimkou je propojení na tabulku *Kontroly*, zde je zvoleno pole *Rok*. Tabulka *Kontroly* nemá pole *Kraj*, ale *Inspektorát*, který má odpovědnost za kraj. Toto určení nelze rozdělit, aby bylo možno získat počet kontrol za Jihomoravský kraj a za Zlínský kraj zvlášť.





Obrázek 36: Relationship – propojení tabulek

### 3.8 Tvorba vizualizací v prostředí nástroje Tableau

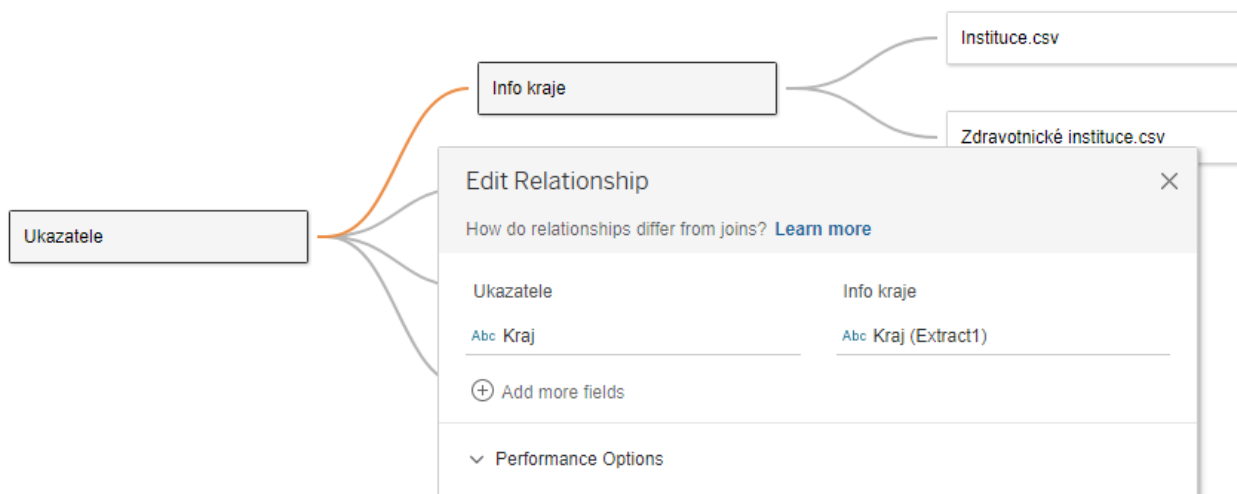
Při otevření nástroje Tableau Desktop se zobrazí pracovní plocha, kde se nachází po levé straně připojení na datové zdroje, uprostřed naposledy otevřená Tableau soubory a vpravo jsou odkazy na výuková videa a další zdroje.

#	Číslo	Abc	#	#	#	#	#	#	#	#
Číslo	NzP_m	Pohlaví	Number of Rows (...)	Počet PU s FN	Průměrná doba P...	Počet shr. NzP	Počet smrtelných ...	Počet SPÚ	Počet ohrožení NzP	Počet
null	null	null	1	null	null	null	null	null	null	null
null	null	null	1	null	null	null	null	null	null	null
null	null	null	1	null	null	null	null	null	null	null
null	null	null	1	null	null	null	null	null	null	null
null	null	Ženy	1	null	null	2	null	null	null	null

Obrázek 37: Pracovní plocha – Data Source

**DATASOURCE** je lokálně uložený soubor či databáze, případně již zhotovený datový model:

- způsob připojení – *Live* nebo *Extract*,
- práce s metadaty – úprava názvů sloupců, výběr datového typu apod.,
- tvorba globálního datového modelu – *Connections* – *Add*,
- spojení tabulek pomocí *Relationship* (nefixní spojení).



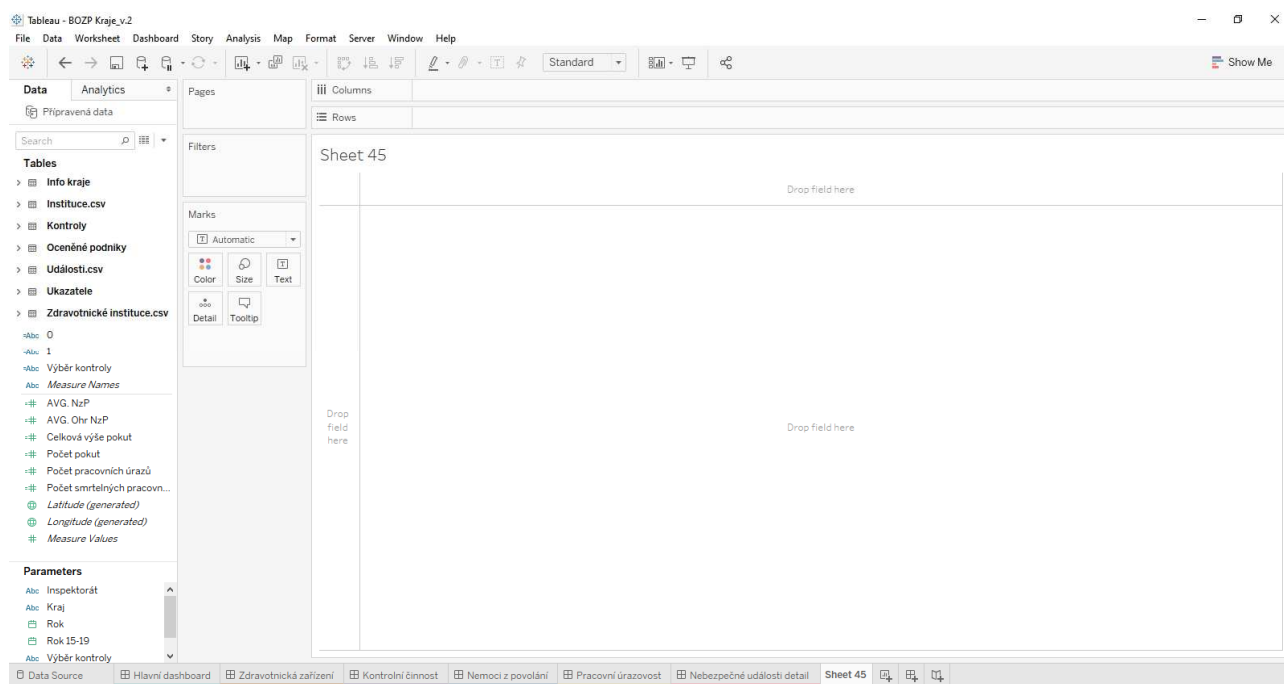
Obrázek 38: Nefixní spojení tabulek a editační okno

## WORKSHEET

Tableau automaticky rozdělí data na dimenze (v horní části označené modře) a ukazatele (ve spodní části označení zeleně) v každé tabulce zvlášť. V případě, že uživatel tvoří další sloupce pomocí *Calculated field*, zobrazí se buďto v dané tabulce (pokud je složen z dat této tabulky), nebo pod všemi tabulkami jako nezávislá data (vytvořené bez dat z tabulek nebo složené z dat z vícero tabulek).

S každým datovým polem/sloupcem lze provádět určité akce, nejčastěji:

- *Duplicate*,
- *Default Properties*,
- *Aliases*,
- *Hierarchy*.



Obrázek 39: Pracovní plocha – worksheet

Tabelau zachovává logiku pracovních ploch, a proto se v levé části nachází data, se kterými uživatel pracuje – metadata (názvy sloupců a jejich datový typ).

Další okna pracovní plochy jsou:

- *Pages* - slouží ke stránkování pohledu (známý je například pohled na měnící se graf v čase),
- *Filters* - slouží k filtrování dat v analýze, nemusí vždy nutně jít o data použitá ve vizualizaci,
- *Marks* - slouží k vizualizaci dat, obsahuje volbu typu grafu, barvu, velikost, text, detail, tooltip a v určitých případech i další možnosti (např. shapes),
- *Rows* a *Columns* (zde se odehrává nejvíce pracovních kroků),
- *Show me* - 24 vizualizačních vzorů s nápovědou.

**Pro manipulaci s daty slouží:**

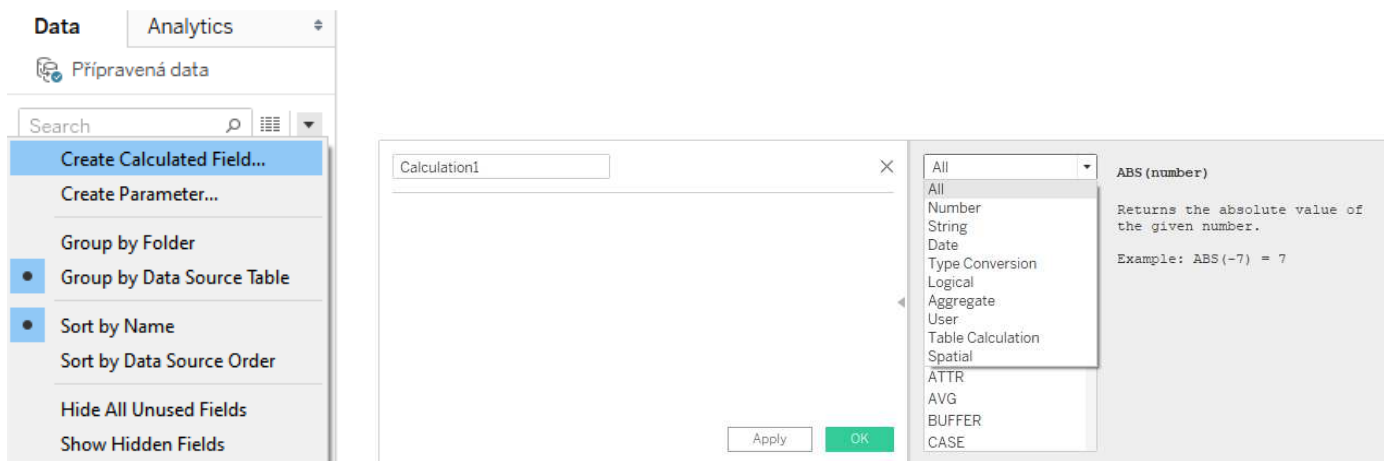
- *Add to Sheet*,
- *Rows*, *Columns* nebo *Marks* - přetáhnutí datového pole do okna,
- přetažení pole přímo na pracovní plochu s vizualizací.

**Analytické funkce** - *Data* na *Analytics*.

## POJMY FUNKCIONALIT

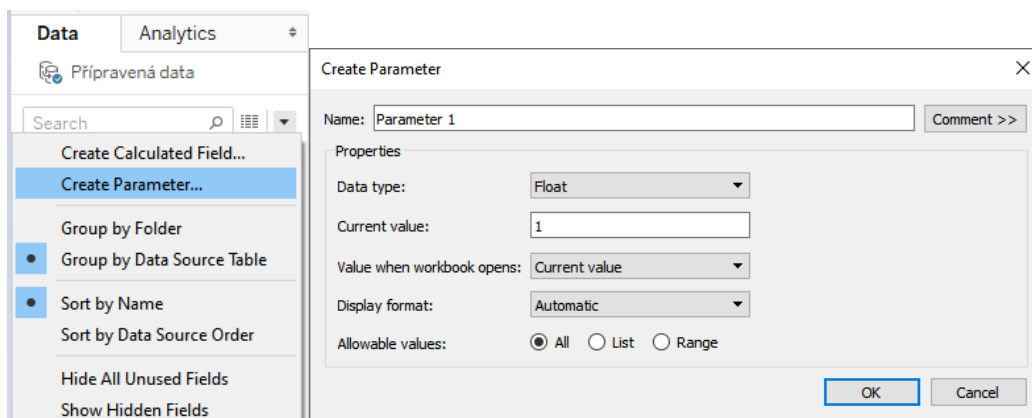
Každý softwarový nástroj obsahuje některé funkcionality a pojmy, se kterými se musí uživatel seznámit, pochopit k čemu slouží a jak s nimi pracovat.

- **Calculated field (CF)** – určen k tvorbě nových datových polí.



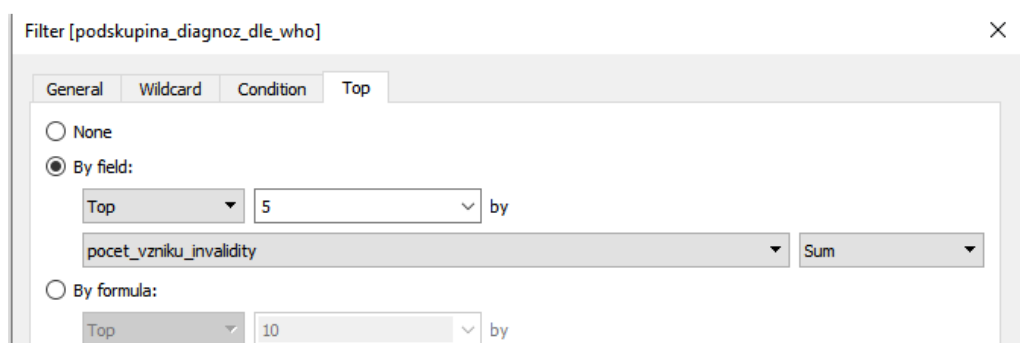
Obrázek 40: Calculated field

- **Parameter** - určen pro zjednodušení interaktivity koncovému uživateli; autor naopak může nabídnout uživateli možnost volby zobrazení, na rozdíl od filtru je ale možná vždy pouze jedna volba; pro správnou funkci je nezbytné vytvořit CF.



Obrázek 41: Parametr

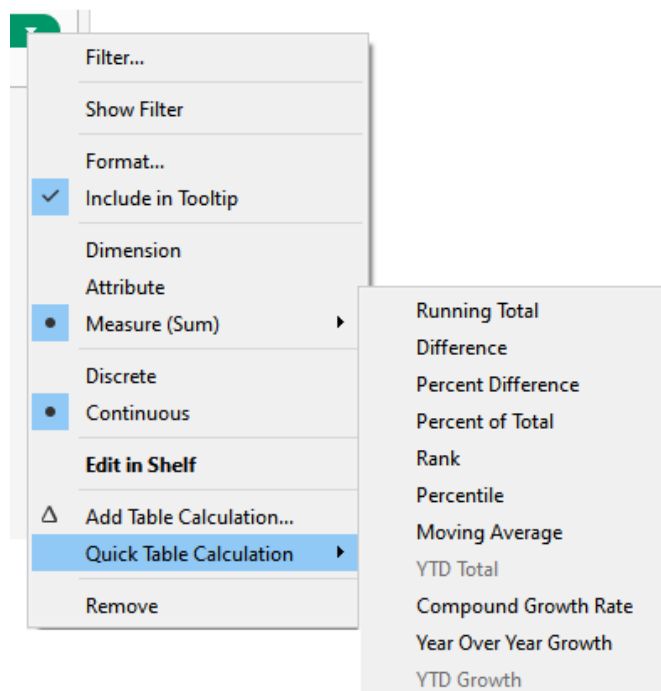
- **Filters** – určen pro filtrování obsahu v Tableau; může mít dvě roviny - první omezuje data, která ve vizualizaci budou zobrazena, druhá rovina je filtrem pro koncového uživatele, který takto interaguje s daty a sám si vybírá, jak mají být data filtrována; filtr na dimenzi může být klasicky výběr (*General*) pro jeden či několik záznamů; případně může být s určitou podmínkou, např. obsahuje-li nějaký záznam x, pak ať je vyloučen, nebo naopak zachován (*Wildcard*); jinou podmínkou (často využívanou) je Top x záznamů.



Obrázek 42: Filtr Top 5

Ve webové aplikaci BOZP Kraje jsou nad opendaty nejvíce použity *General filtry*, jednou *Top*. Je-li použito v jednom worksheetu více filtrů, filtry na sebe reagují a ovlivňují se.

- *Default properties* - zjednoduší a zrychlí práci autorům, především při formátovacích úkonech. Tato funkcionality zachovává formátování a další akce, pokud je daná dimenze nebo ukazatel použit.
- *Group* - slučování záznamů (*Group*)
- *Table Calculation* - speciální předdefinované *Calculated fields* (nejčastěji používané – záložka *Quick Table Calculation*); při kliknutí na *Add Table Calculation* se otevře okno, kde si uživatel může nadefinovat, jakým způsobem bude proveden výpočet napříč celou tabulkou.



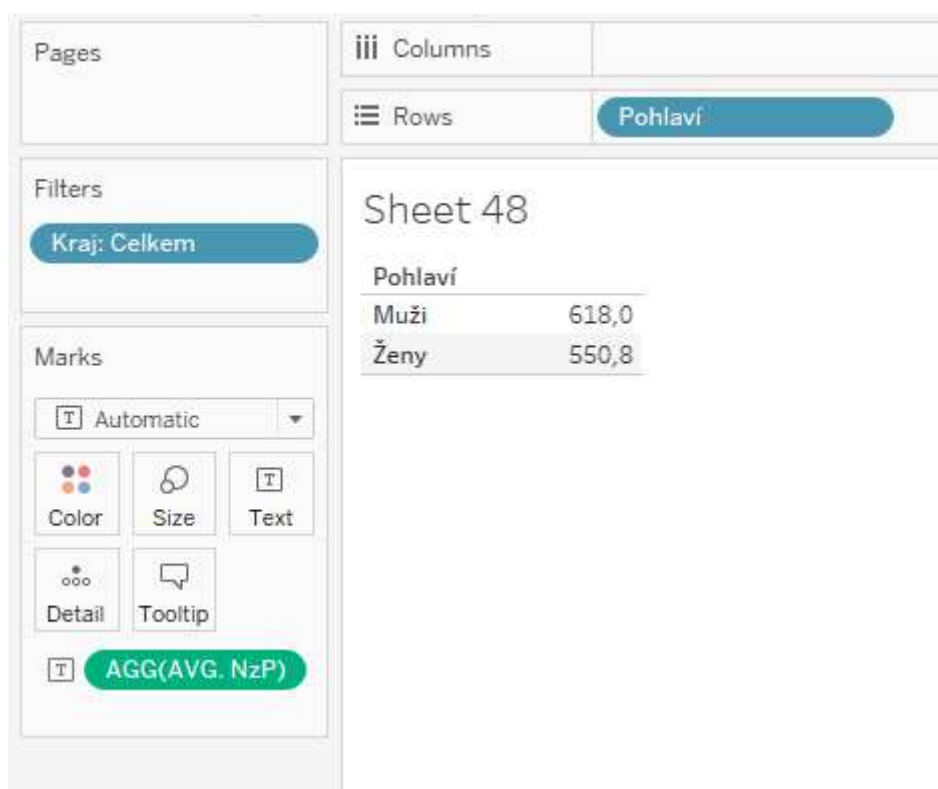
Obrázek 43: Table Calculation

### 3.8.1 Tabulky a celá čísla

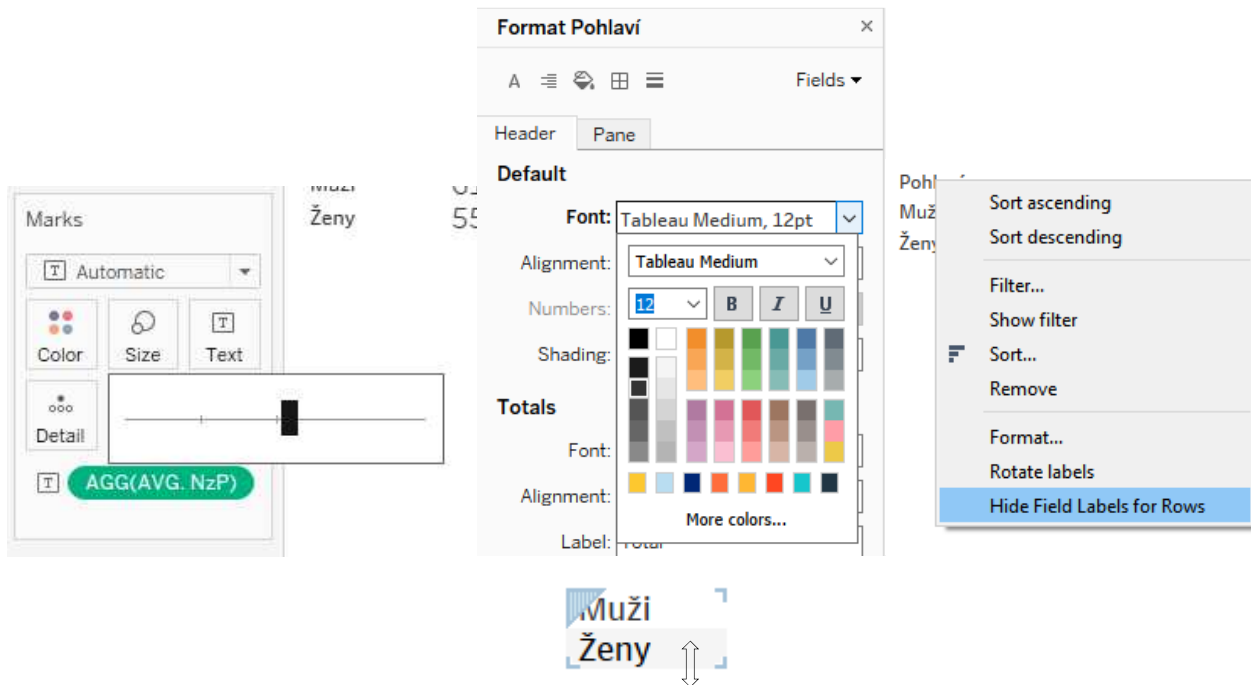
Jedním z nejsrozumitelnějších typů vizualizací dat jsou tabulky a celá čísla, kterými lze zdůraznit informaci o celkovém součtu či průměru skupiny dat.

**Zde je uveden příklad – průměrný počet nemocí z povolání:**

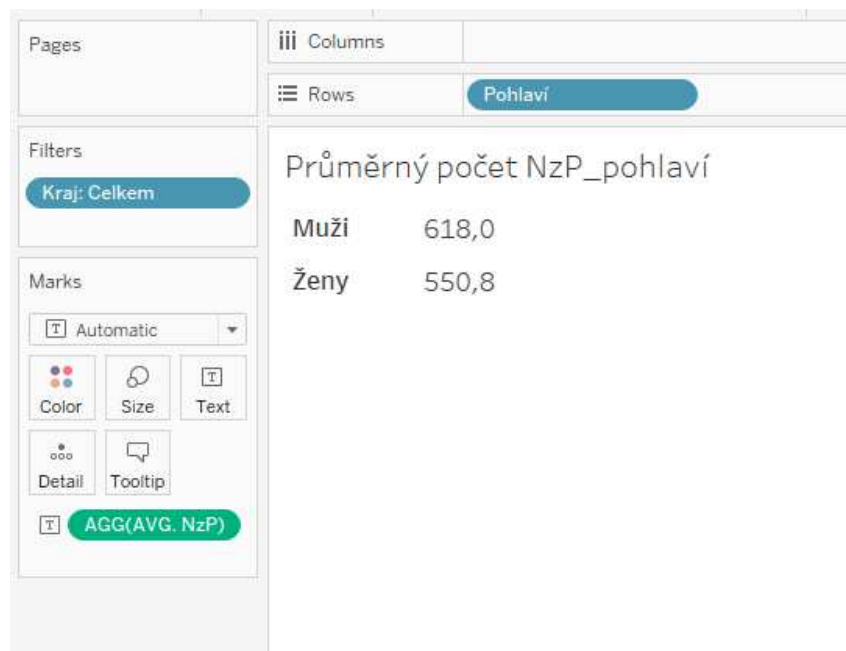
- vypočtení průměru pomocí CF (ukazatel NzP vydělen 5-ti),
- přetažení CF v poli *Marks* na *Text*,
- přetažení dimenze *Pohlaví* do *Rows*,
- přetažení dimenze *Kraj* do *Filters*,
- kliknutí na jedno z čísel a zvolení akce *Format* (*Format Shading*, *Row Banding*, *Format Borders*, *Row Divider*),
- provedení dalších formátovacích úpravy (*Size*, *Hide Fields Labels for Rows* atp.),
- uložení vytvořeného worksheetu.



Obrázek 44: Tvorba vizualizace – samostatné číslo



Obrázek 45: Formátování



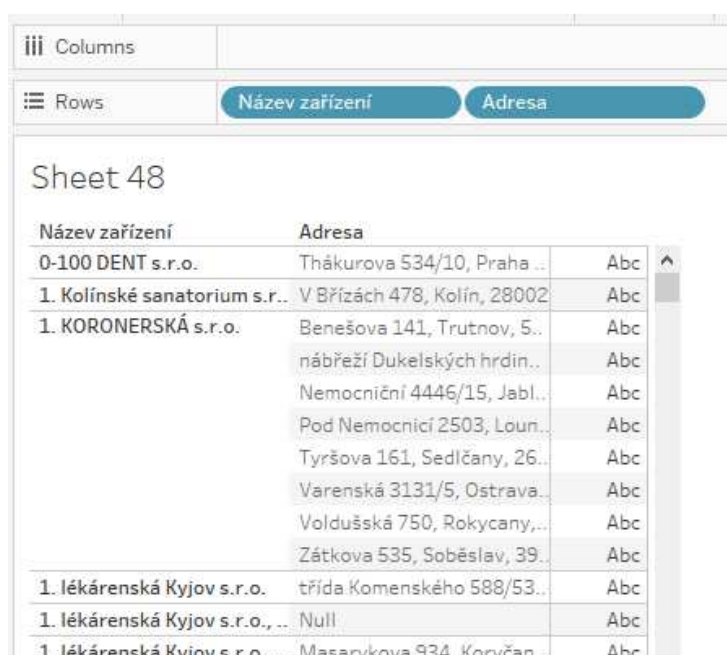
Obrázek 46: Výsledná vizualizace – samostatné číslo

### Dalším příkladem je tabulka se zdravotnickými zařízeními:

- vytvoření tabulky bez ukazatelů, obsahuje pouze dvě dimenze *Název* a *Adresu zařízení*,
- vytvoření pole *Adresa* (sloupce obec, ulice atd.),
- přetažení pole *Název zařízení* a *Adresa* do *Rows*,
- uzpůsobení velikosti textu (*Marks – Polygon*),
- vytvoření *Group* na datovém poli *Druh zařízení* a seskupení zařízení, která jsou si svojí činností podobná,
- aktivace filtru na vytvořené skupině – *Show Filter*.



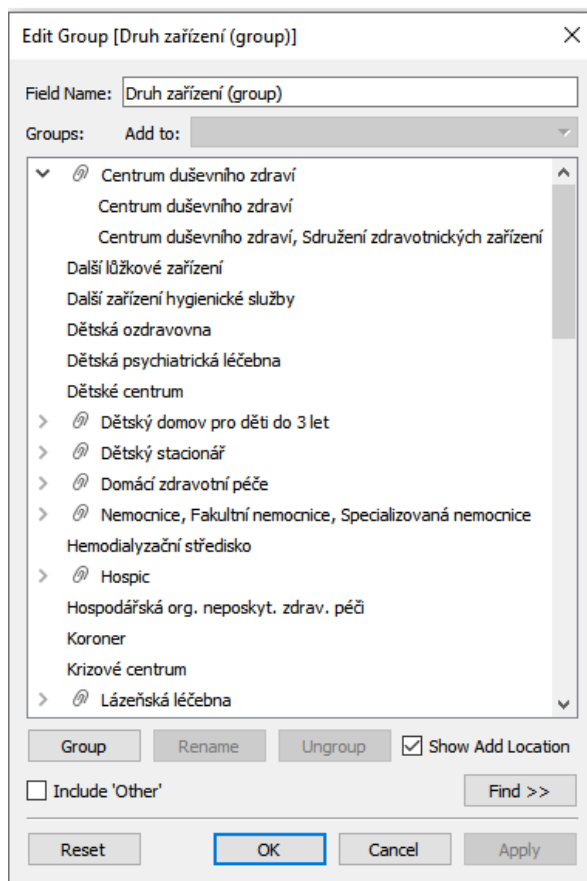
Obrázek 47: CF – Adresa



Název zařízení	Adresa	
0-100 DENT s.r.o.	Thákurova 534/10, Praha ..	Abc
1. Kolínské sanatorium s.r.o.	V Břizách 478, Kolín, 28002	Abc
1. KORONERSKÁ s.r.o.	Benešova 141, Trutnov, 5..	Abc
	nábřeží Dukelských hrdin..	Abc
	Nemocniční 4446/15, Jabl..	Abc
	Pod Nemocnicí 2503, Loun..	Abc
	Tyršova 161, Sedlčany, 26..	Abc
	Varenská 3131/5, Ostrava..	Abc
	Voldušská 750, Rokycany,...	Abc
	Zátkova 535, Soběslav, 39..	Abc
1. lékárenská Kyjov s.r.o.	třída Komenského 588/53..	Abc
1. lékárenská Kyjov s.r.o., ...	Null	Abc
1. lékárenská Kyjov s.r.o., ...	Masarykova 934, Koryčan..	Abc

Obrázek 48: Tabulka – krok 2





Obrázek 49: Group – druh zařízení

### 3.8.2 Mapy

Zobrazení dat na mapových podkladech je velmi oblíbeným a efektivním vizualizačním vzorem. Pro vytvoření takové vizualizace v Tableau postačí přiřadit některému datovému poli geografickou roli. Ve zdrojových datech jsou těmito poli *Kraj* a *PSC*.

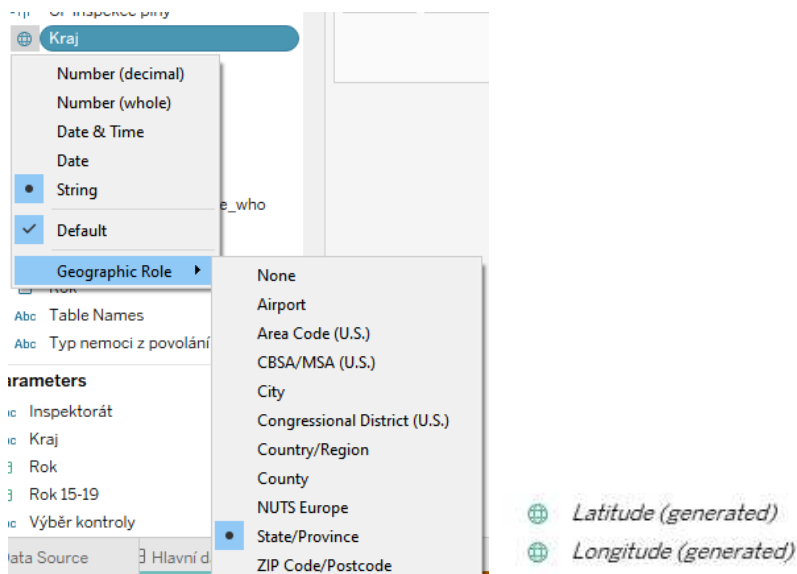
**Zde je uveden příklad vytvoření mapového podkladu:**

- zvolení *ZIP Code/Postcode* u datového pole *PSC* (automatické přiřazení údajů zeměpisné šířky a délky každému řádku v dané tabulce),
- kliknutí na datové pole *Kraj* a z nabídky *Show Me* a výběr požadované mapy,
- upravení pozadí mapy - *Map Layers*.

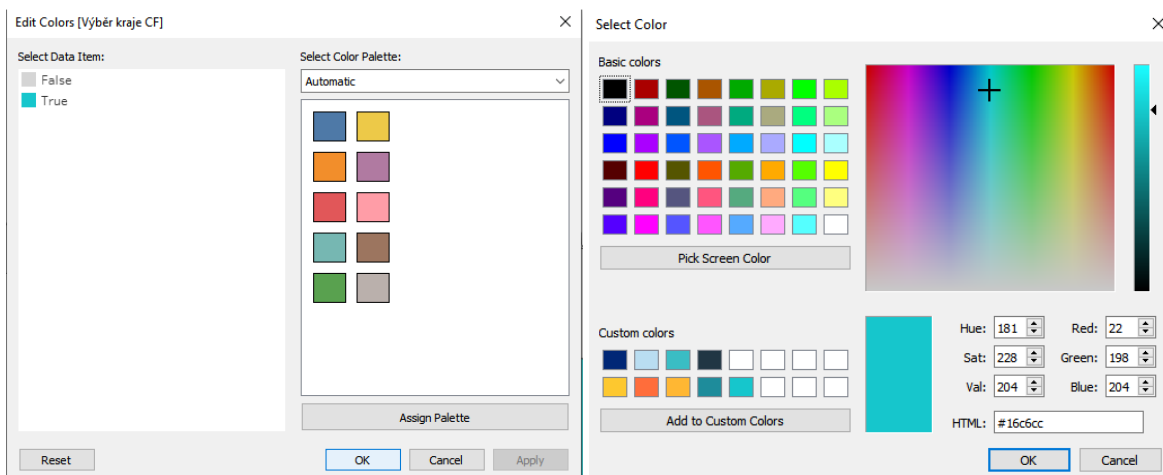
Dále je třeba uzpůsobit mapu tak, aby koncový uživatel mohl vybírat dle zvoleného kraje a filtrovat tak data:

- vytvoření *Parameter* a použití v CF pro výběr (využití *Add values from pole Kraj*),
- vytvoření dalšího CF,
- přetažení prvního CF v poli *Marks* na *Color*,
- uzpůsobení barvy mapy při výběru - *Worksheet – Action – Add Action – Change Parameter*,

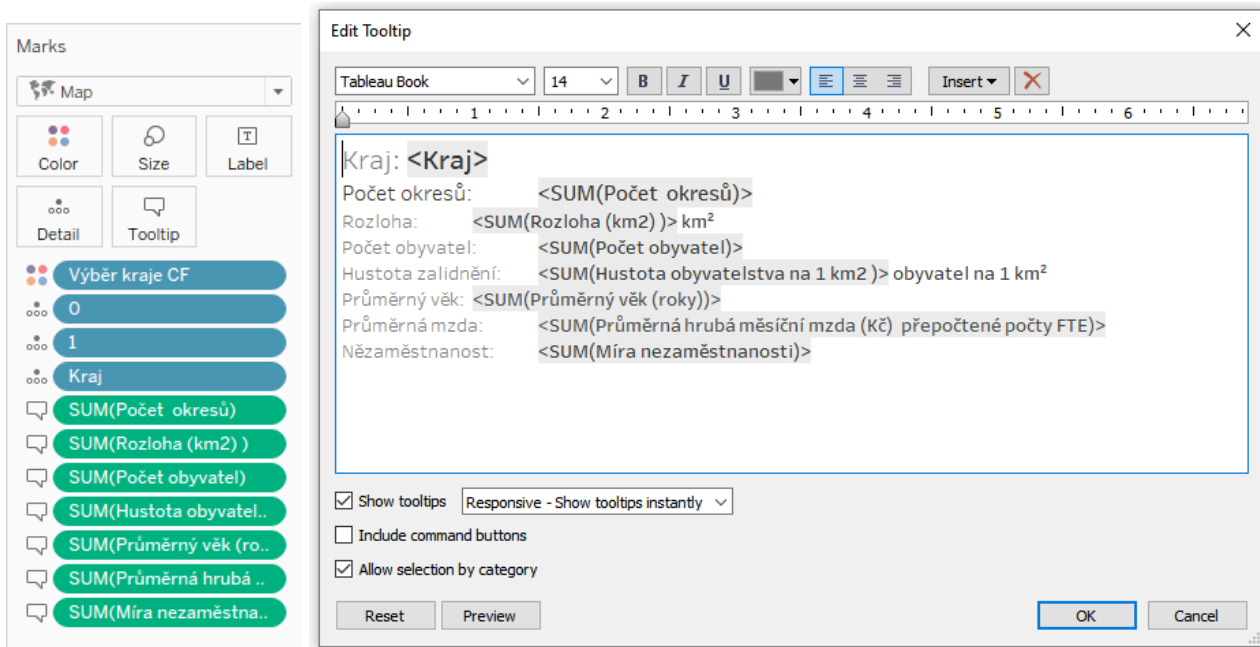
- nastavení polí - *Target parameter* – Kraj, *Field* – Kraj,
- přetažení obou CF v poli *Marks* do *Detail* - upravení *Tooltipu* (zobrazení základních informací o daném kraji).



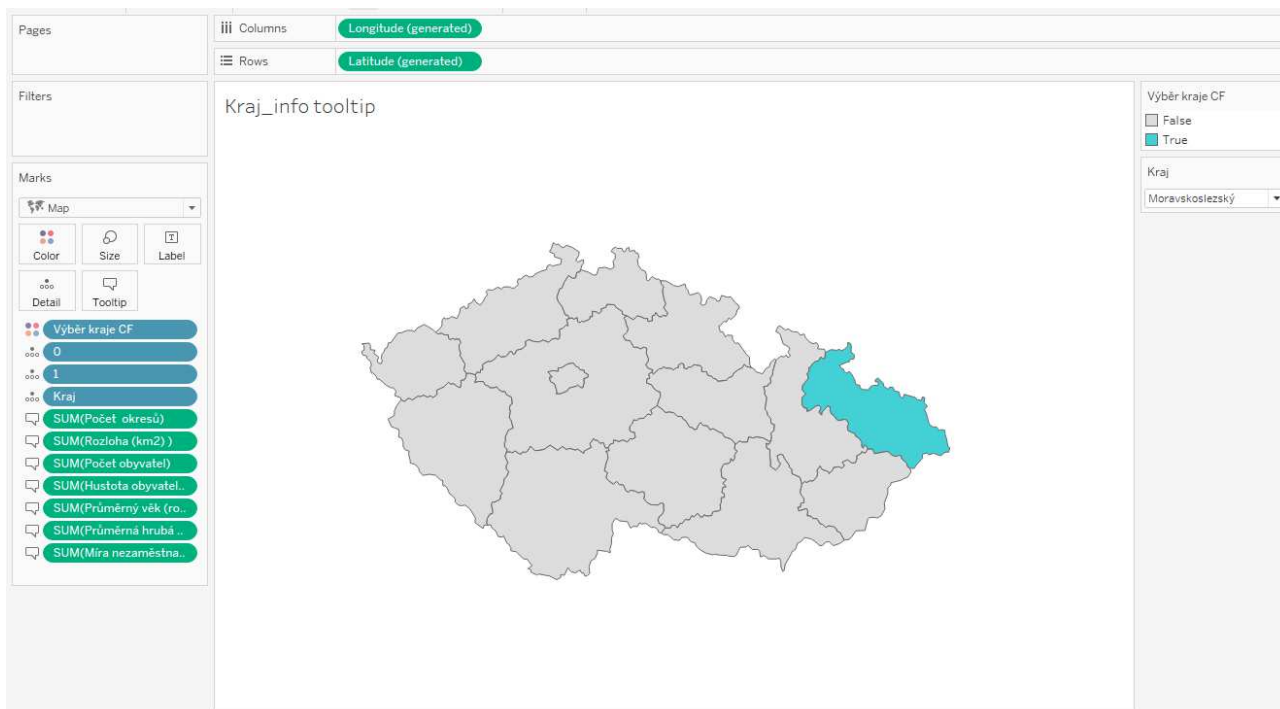
Obrázek 50: Nastavení geografické role poli Kraj, vygenerovaná datová pole



Obrázek 51: Výběr barev



Obrázek 52: Tooltip



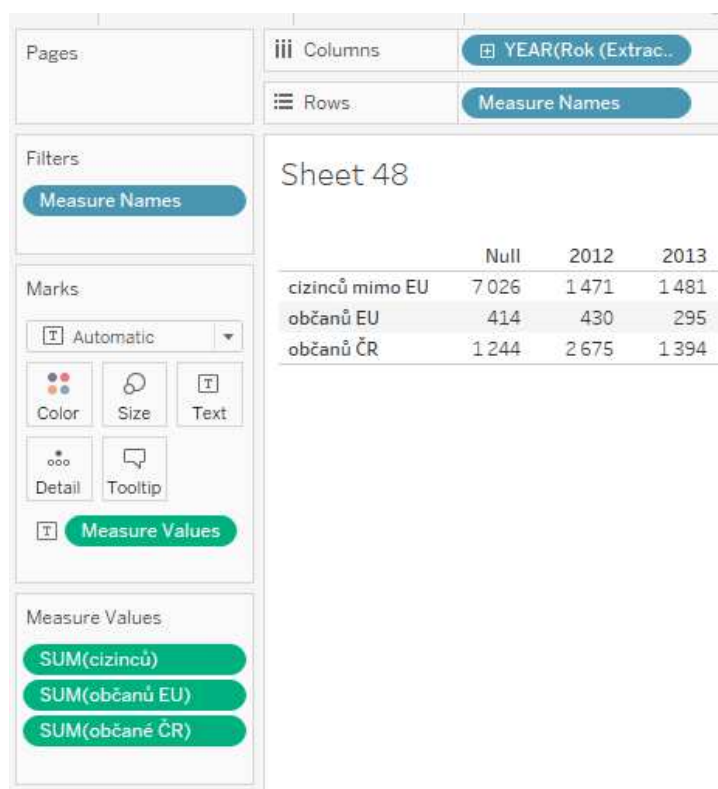
Obrázek 53: Vytvořená mapa

### 3.8.3 Liniové grafy

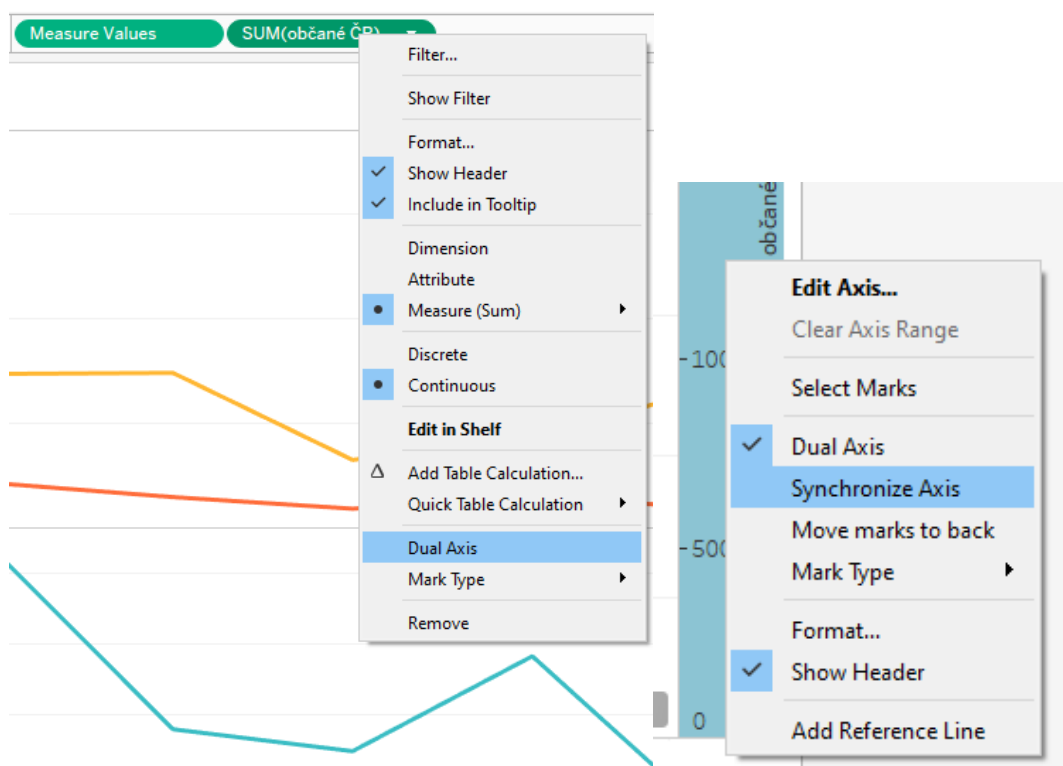
K tvorbě liniového grafu je nutné mít mezi datovými poli alespoň jednu dimenzi ve formátu data. Tableau si automaticky vytvoří hierarchii *Rok-Kvartál-Měsíc-Den* (případně i *Hodiny-Minuty-Vteřiny*).

Zde je uveden příklad liniového grafu – výsledky kontroly nelegálního zaměstnávání s přidáním tooltipu s informací o původu země cizinců:

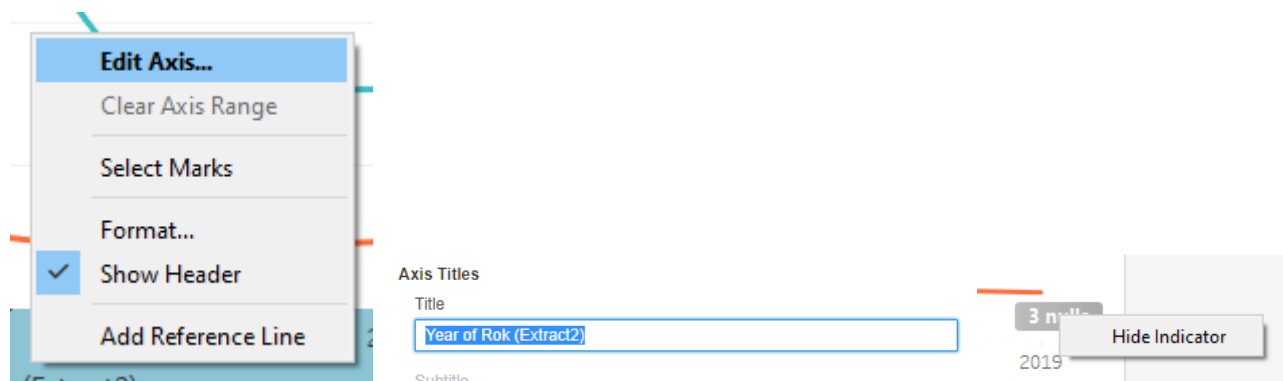
- výběr datového pole *Rok* a jeho přetažení do *Columns*, dvojklikem výběr ukazatele *Cizinci, Občané EU, Občané ČR*,
- přesunutí datového pole *Measure Values* z pole *Marks* do *Rows* a pole *Measure Names* do pole *Marks* do *Color*,
- přesunutí pole *Občané ČR* z pole *Marks* do *Rows* vedle pole *Measure Values* – zvolení akce *Dual axis* – *Synchronize Axis* a *Show Header*,
- nastavení názvu os (*Edit Axis*),
- schování ukazatele (*Hide Indicator*).



Obrázek 54: Liniový graf krok 1



Obrázek 55: Dual axis



Obrázek 56: Editace osy, vymazání názvu, Hide null

Dále je třeba vytvořit nový *worksheet*, který bude použit pro detailnější informace v *Tooltip*:

- vytvoření *Group* z datového pole *Státní příslušnost*, označení států, které patří do EU a kliknutí na tlačítko *Group*,
- přetažení datového pole *Rok*, *Státní příslušnost*, *Státní příslušnost (group)*, ukazatel *Počet osob*,
- přetažení s podržením CTRL pole *YEAR (Rok ..)*, výběr roků 2018 a 2019 - název „Null“ a *Exclude*,
- seřazení tabulky (Mimo EU, EU), skrytí názvu *Státní příslušnost* a *Rok (Hide Field Label for Rows)*,
- upravení formátu a uložení *worksheetu*,

- v předchozím worksheetu je třeba kliknout v okně *Marks* na *Tooltip* (*Insert – Sheet – tooltip: státní příslušnost*).

Státní přísl..	Státní příslušn..	2018	2019
Null	Null		
EU	✓ Keep only	47	21
	✗ Exclude	3	
	Hide	40	34
		1	2
	Format...	49	36

#### Edit Group [Státní příslušnost (

Field Name: Státní příslušnost (

Groups: Add to:

- Null
- > Mimo EU
- ▼ EU
  - Bulharsko
  - Lotyšsko
  - Maďarsko
  - Polsko
  - Rumunsko
  - Řecko
  - Slovensko

Obrázek 57: Group – státní příslušnost

Columns: YEAR(Rok (Extrac...))

Rows: Státní příslušnost (gr...), Státní příslušnost

Sheet 49

Státní přísl..	Státní příslušn..	Null	2012	2013	2014	2015	2016	2017	2018	2019
Null	Null									
EU	Bulharsko								47	21
	Lotyšsko								3	
	Maďarsko								40	34
	Polsko								1	2
	Rumunsko								49	36
	Řecko								1	
	Slovensko								84	97
Mimo EU	Bělorusko								23	13
	Filipíny									13
	Gruzie								16	16
	Indie								10	
	Moldavsko								197	246
	Mongolsko									14
	Nepál								8	
	Ostatní země								69	127
	Rusko								21	12
	Srbsko								59	12
	Ukrajina								2873	3015
	Uzbekistán								34	
	Vietnam								285	62

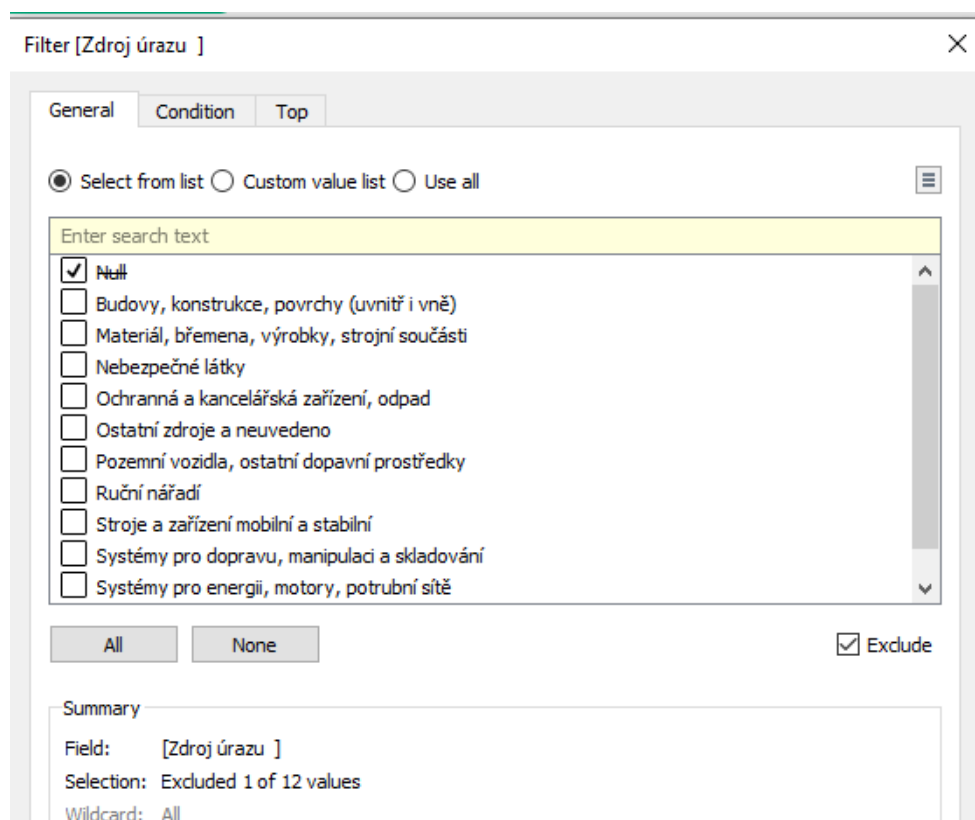
Obrázek 58: Tabulka do tooltipu

### 3.8.4 Scatter plot

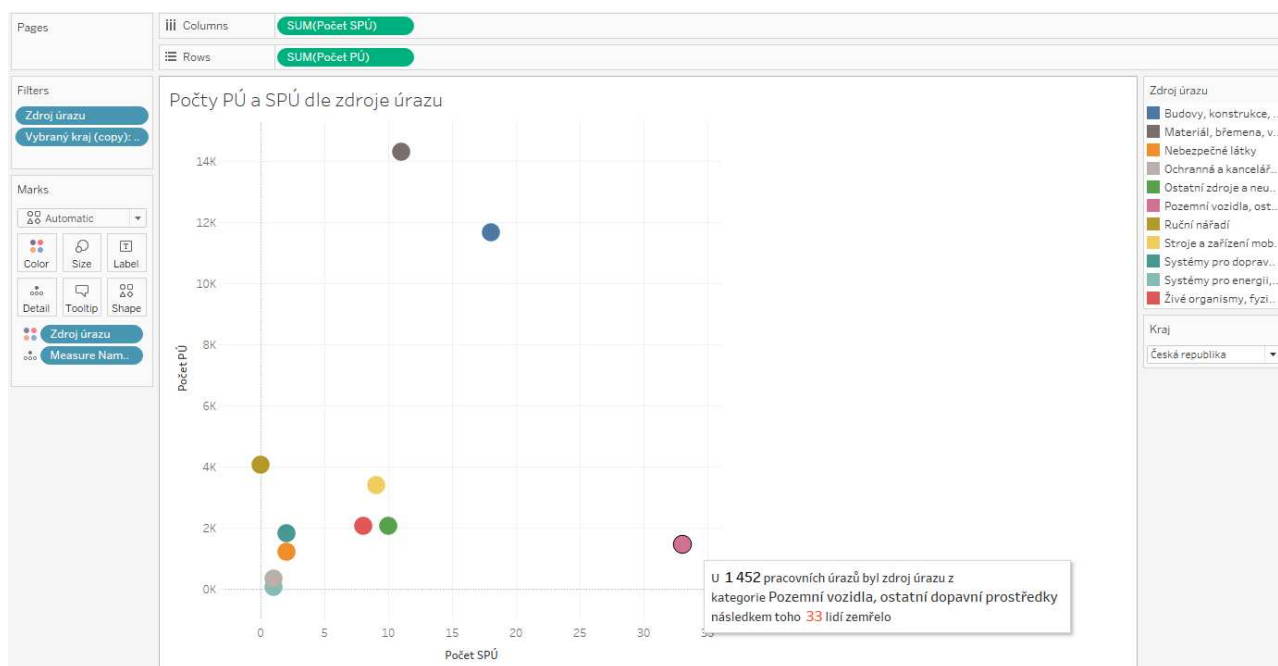
Zajímavým typem vizualizace je tzv. scatter plot, který je využíván pro porovnání dvou ukazatelů mezi sebou.

**Zde je uveden příklad – porovnání zdrojů úrazů v ukazatelích počtu pracovních úrazů a počtu smrtelných pracovních úrazů:**

- přetažení ukazatele *Počet PÚ* do *Rows*, ukazatele *Počet SPÚ* do *Columns*,
- přetažení dimenze *Zdroj úrazu* v poli *Marks* do *Color* (výběr *Circle* místo *Automatic*),
- přetažení s CTRL z pole *Marks* pole *Zdroj úrazu* do *Filters*,
- kliknutí na tlačítko *None* a označit *Null*, zakliknutí *Exclude*,
- přetažení vytvořeného CF *Vybraný kraj* do pole *Filters* a vybrat *True*,
- upravení Tooltipu.



Obrázek 59: Exclude Null pomocí filtrace dimenze



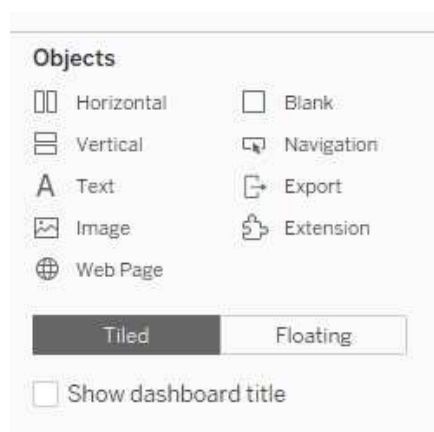
Obrázek 60: Scatter plot

### 3.8.5 Dashboardy

Dashboard je jedna plocha složená z vizualizací, doplňkových textů a dalších prvků, které pomáhají uživateli interagovat s daty či se přepínat mezi jednotlivými dashboardy. Díky těmto prvkům a provázání jednotlivých vizualizací mezi sebou (je-li to vhodné a datově podpořené) lze vytvořit webovou aplikaci.

Tableau Desktop má předpřipravené velikosti zobrazení pro různá zařízení. Po kliknutí na *Device Preview* se zobrazí lišta, kde si může autor vybrat a přizpůsobit velikost pracovní plochy (sekce *Size*).

V levé části se nachází výčet vytvořených worksheetů (vizualizací) a pod ním sekce *Objects*.



Obrázek 61: Objects



Objekty slouží ke vkládání obsahu do dashboardu (*Horizontal* a *Vertical*):

- *Navigation* - objekt sloužící k přesunu na jiný dashboard nebo worksheet,
- *Export* - může mít také podobu textu či obrázku a jeho cílem je uživateli dát možnost exportovat dashboard jako obrázek do PDF nebo PowerPointu,
- *Image* objekt - pro vkládání obrázků do dashboardu,
- *Web Page* - pro vložení okna s webovou stránkou, zobrazenou přímo v dashboardu,
- *Extension* – integrace dalších nástrojů do dashboardu.

Při vkládání objektu do dashboardu má autor možnost zvolit mezi *Tiled* a *Floating*. *Floating* objekt je nezávislý a lze ho umístit kamkoliv na pracovní plochu, případně překrývat *Tiled* objekt. Hranice tohoto objektu mohou být i mimo pracovní plochu. *Tiled* objekt je fixně daný do pracovní plochy a nemůže vyčnívat mimo ní, zároveň nemůže překrývat jiné objekty.

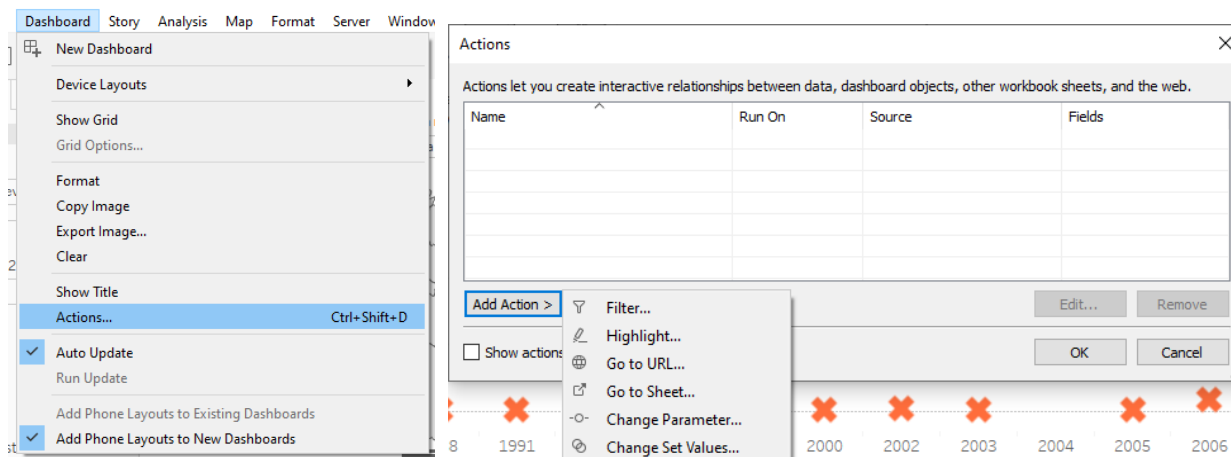
**Postup pro sestavení dashboardu** je následující:

- náčrt rozložení worksheetů,
- zvolení velikosti,
- přetažení horizontálního a vertikálního objektu (nastavení ohraničení v sekci *Layout*),
- formátování objektů,
- upravení záložek (*Hide All Sheets*).



Obrázek 62: Vkládání objektů do dashboardu

Funkce **Actions** je hlavní funkcionalitou s přidanou hodnotou pro koncového uživatele, která provádí interakce mezi worksheety či dalšími objekty.

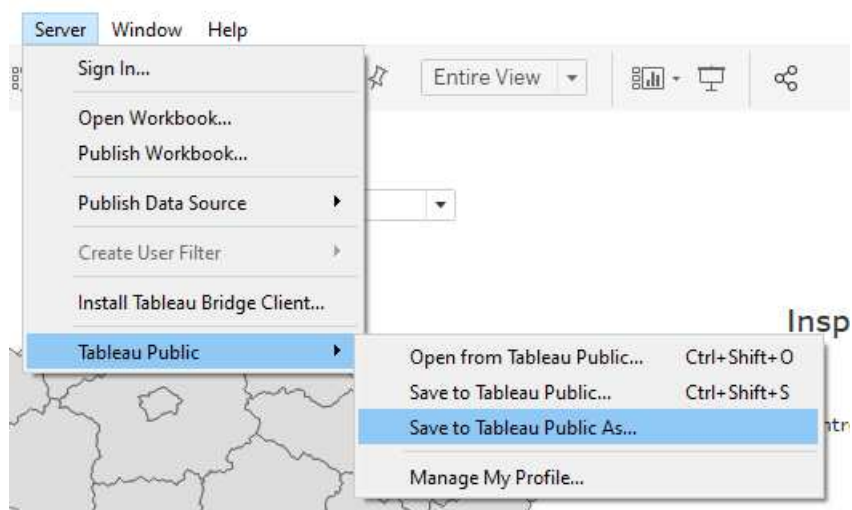


Obrázek 63: Actions

### 3.8.6 Publikace do webového prostředí

Pro publikování do online prostředí je nutné mít založen účet na stránkách <https://public.tableau.com/en-us/s/>.

Před umístěním dashboardů do online prostředí je třeba nastavit všechny dashboardy do defaultního zobrazení, které uživatel uvidí jako první (*Server – Tableau Public – Save to Tableau Public As*).



Obrázek 64: Publikace do online prostředí



Save Workbook to Tableau Public

Workbook Title

BOZP Kraje\_v.1

Cancel Save

Obrázek 65: Save to Tableau Public

Dále je vhodné provést nastavení uživatelských oprávnění (stažení souborů, zobrazení jednotlivých záložek, sdílení, embedování atp.) a zkontrolovat, zda se vše správně zobrazuje.

Toolbar Settings

- Show view controls *Undo, Redo, Revert*
- Show author profile link
- Allow others to download or explore and copy this workbook and its data

Other Settings

- Show workbook sheets as tabs

Obrázek 66: Nastavení „Edit details“



## POUŽITÁ LITERATURA

- *Akční plán České republiky Partnerství pro otevřené vládnutí 2018-2020* [online]. Praha: Úřad vlády České republiky, 2016 [cit. 2020-11-30]. Dostupný z: <https://www.databaze-strategie.cz/cz/urad-vlady/strategie/akcni-plan-cr-partnerstvi-pro-otevrene-vladnuti-na-obdobi-let-2016-az-2018?typ=o>.
- *Digitální Česko v. 2.0: cesta k digitální ekonomice* [online]. Vláda ČR, 2011 [cit. 2020-11-30]. Dostupný z: [https://www.vlada.cz/assets/media-centrum/aktualne/Digitalni-Cesk, June 2013 \[cit. 2020-02-14\]. Dostupný z: o-v--2-0\\_120320.pdf](https://www.vlada.cz/assets/media-centrum/aktualne/Digitalni-Cesk, June 2013 [cit. 2020-02-14]. Dostupný z: o-v--2-0_120320.pdf).
- *Národní katalog otevřených dat* [online]. Ministerstvo vnitra [cit. 2020-02-13]. Dostupný z: <https://data.gov.cz/datov%C3%A9-sady>.
- NOVOTNÝ, Ota; POUR, Jan; SLÁNSKÝ, David. 2005. *Business Intelligence*. Praha: Grada Publishing, 2005. ISBN 80-247-1094-3.
- *Open Government Data* [online]. OECD [cit. 2020-11-30]. Dostupný z: <https://www.oecd.org/digital/digital-government/open-government-data.htm>.
- POTANČOK, Martin; POUR, Jan; CHRAMOSTOVÁ, Veronika. 2020. *Business analytika v praxi*. Praha: Oeconomica, 2020. ISBN 978-80-245-2382-8.
- SLÁNSKÝ, David, 2018. *Data and Analytics for the 21st Century: Architecture and Governance*. Praha: Professional Publishing. ISBN 978-80-88260-16-5.
- *Strategická rámeček rozvoje veřejné správy České republiky pro období 2014-2020* [online]. Ministerstvo vnitra ČR, listopad 2016 [cit. 2020-11-30]. Dostupný z: <http://www.mvcr.cz/soubor/strategicky-ramecek-rozvoje-verejne-spravy-v-cr-pro-obdobi-2014-2020.aspx>.
- *Strategie vlády v boji s korupcí na období let 2013 a 2014* [online]. Úřad vlády České republiky, 2013 [cit. 2020-11-30]. Dostupný z: <https://www.vlada.cz/assets/protikorupcni-strategie-vlady/na-leta-2013-2014/Strategie-2013-a-2014---aktualni-verze.pdf>.
- ULMANOVÁ, Jiřina; MĚRKOVÁ, Veronika. *Využití otevřených dat pro prezentaci a vizualizaci informací z oblasti BOZP. Časopis výzkumu a aplikací v profesionální bezpečnosti* [online]. 2019, roč. 12, č. 3-4. Dostupný z: <https://www.bozpinfo.cz/josra/vyuziti-otevrenych-dat-pro-prezentaci-vizualizaci-informaci-z-oblasti-bozp>. ISSN 1803-3687.



## SEZNAM OBRÁZKŮ

Obrázek 1: Ukázka záznamu datové sady/datasetu .....	11
Obrázek 2: Úvodní stránka CKANu .....	13
Obrázek 3: Ukázka datasetu Obyvatelstvo v ČR .....	15
Obrázek 4: Editace datového zdroje .....	16
Obrázek 5: Ukázka vyhledaného datasetu s uloženými formáty .....	16
Obrázek 6: Ukázka zobrazení mřížka .....	17
Obrázek 7: Ukázka vygenerovaného grafu .....	18
Obrázek 8: Ukázka zobrazení webové stránky .....	19
Obrázek 9: Ukázka zobrazení Tableau .....	20
Obrázek 10: Organizace v CKANu .....	21
Obrázek 11: Skupiny v CKANu .....	22
Obrázek 12: Showcases v CKANu .....	23
Obrázek 13: Ukázka interpretovaných dat z ČSÚ .....	25
Obrázek 14: Upravená zdrojová tabulka .....	25
Obrázek 15: Nečistoty v datech .....	26
Obrázek 16: Neupravená tabulka .....	27
Obrázek 17: Upravená tabulka .....	27
Obrázek 18: Úprava před a po .....	28
Obrázek 19: Označení Pivot v Tableau Prep .....	29
Obrázek 20: Pivot, krok 1 .....	29
Obrázek 21: Pivot, krok 2 .....	30
Obrázek 22: Možnosti čisticích procesů v Tableau Prep .....	30
Obrázek 23: provedené čisticí procesy u vybraného datového zdroje .....	31
Obrázek 24: Group values .....	32
Obrázek 25: Typy spojení – join .....	32
Obrázek 26: Join – definice spojení .....	33
Obrázek 27: Union – spojení více zdrojů .....	34
Obrázek 28: Union – spojené tabulky .....	35
Obrázek 29: Dílčí datový model – ukazatele BOZP .....	36



Obrázek 30: Dílčí datový model – informace o krajích.....	37
Obrázek 31: Dílčí datový model – kontrolní činnost .....	38
Obrázek 32: Dílčí datový model – oceněné podniky .....	39
Obrázek 33: Dílčí datové modely – instituce BOZP a zdravotnická zařízení.....	39
Obrázek 34: Dílčí datový model – Nebezpečné události .....	40
Obrázek 35: Globální datový model .....	40
Obrázek 36: Relationship – propojení tabulek .....	41
Obrázek 37: Pracovní plocha – Data Source.....	42
Obrázek 38: Nefixní spojení tabulek a editační okno .....	42
Obrázek 39: Pracovní plocha – worksheet .....	43
Obrázek 40: Calculated field.....	44
Obrázek 41: Parametr.....	44
Obrázek 41: Filtr Top 5 .....	45
Obrázek 43: Table Calculation .....	45
Obrázek 44: Tvorba vizualizace – samostatné číslo .....	46
Obrázek 46: Formátování .....	47
Obrázek 46: Výsledná vizualizace – samostatné číslo .....	47
Obrázek 47: CF – Adresa.....	48
Obrázek 48: Tabulka – krok 2 .....	48
Obrázek 49: Group – druh zařízení.....	49
Obrázek 50: Nastavení geografické role poli Kraj, vygenerovaná datová pole.....	50
Obrázek 51: Výběr barev .....	50
Obrázek 52: Tooltip .....	51
Obrázek 53: Vytvořená mapa .....	51
Obrázek 54: Liniový graf krok 1 .....	52
Obrázek 55: Dual axis .....	53
Obrázek 56: Editace osy, vymazání názvu, Hide null.....	53
Obrázek 57: Group – státní příslušnost .....	54
Obrázek 58: Tabulka do tooltipu .....	54
Obrázek 59: Exclude Null pomocí filtrace dimenze .....	55
Obrázek 60: Scatter plot.....	56
Obrázek 61: Objects .....	56



Obrázek 62: Vkládání objektů do dashboardu .....	57
Obrázek 63: Actions.....	58
Obrázek 64: Publikace do online prostředí .....	58
Obrázek 65: Save to Tableau Public .....	59
Obrázek 66: Nastavení „Edit details“ .....	59

