

# Metodický postup řešení Finální databáze 4

**Objednatel:** Česká republika – Ministerstvo vnitra  
**Sídlo:** Nad Štolou 936/3, 170 34 Praha 7  
**IČ:** 00007064

**Zpracovatel:** INTENS Corporation s.r.o.  
**Sídlo:** Za Brumlovkou 266/2, Michle, 140 00 Praha 4  
**IČ:** 28435575

Zpracováno v rámci projektu „Zlepšení podmínek pro decentralizaci a dostupnost veřejné správy v území“, reg. č. GG-PDP1-001, financovaného z fondů EHP a Norska 2014–2021.

## **Obsah**

Obsah.....	2
1 Úvod.....	3
2 Manažerské shrnutí.....	4
3 Vstupní data.....	10
3.1 Signalizační data.....	10
3.2 Základní soubor SIM.....	10
3.3 Územní dimenze.....	12
3.4 Časová dimenze.....	12
3.5 Anonymizace dat.....	14
4 Proces získávání dat a způsob měření.....	16
4.1 Georeferencování signalizačních dat.....	16
4.2 Cellmapping – přemapování.....	17
4.3 Čištění dat.....	18
4.4 Detekce míst pobytu.....	18
4.5 Metoda přiřazování příznaků místům uživatele.....	21
4.5.1 Parametry pro přidělování příznaků.....	21
4.5.2 Příznaky.....	21
4.5.3 Hierarchie příznaků.....	25
5 Korekce reprezentativity redukované báze SIM karet.....	27
5.1 Výpočet koeficientů korekce reprezentativity na úrovni obcí.....	29
5.2 Způsob aplikace koeficientů korekce reprezentativity.....	33
6 Projekce na populaci.....	35
7 Sledované jevy.....	36
7.1 Tabulka č. 1 - Data o počtu obvykle bydlícího obyvatelstva.....	36
7.1.1 Porovnání výstupů s konvenčními zdroji.....	37
7.2 Tabulka č. 2 - Data o meziobecní dojížděce.....	46
7.3 Tabulka č. 3 - Denní chod počtu přítomného obyvatelstva.....	47

## 1 Úvod

Dokument „Metodický postup řešení“ je jedním z výstupů veřejné zakázky „Nákup geolokačních dat mobilních operátorů k realizaci projektu Zlepšení podmínek pro decentralizaci a dostupnost veřejné správy v území“, která byla realizována v rámci projektu „Zlepšení podmínek pro decentralizaci a dostupnost veřejné správy v území“, reg. č. GG-PDP1-001, který je finančně podpořen z fondů EHP a Norska 2014–2021.

Finální databáze byly vytvořeny v souladu s parametry a postupem definovaným v dokumentu Finální návrh řešení, který je závazný pro dodávky Finálních databází 1, 2, 3 a 4, definovaných ve Smlouvě o poskytování a zpracování geolokačních dat mobilních operátorů, uzavřené 1. 12. 2021.

Výstupní databáze, jsou zpracovány tak, aby bylo možné je otevřít a editovat v MS Excel (ve formátu .xlsx). Z důvodu systémového omezení počtu řádků v této aplikaci jsou soubory v případě Tabulky č. 2 a Tabulky č. 3 rozděleny do více souborů. V případě Tabulky č. 2 bylo toto dělení provedeno také dle typů cest. Datový fond Finálních databází je tvořen csv soubory, které nemusí být, a tedy nejsou tímto způsobem děleny.

Ve všech zpracovaných tabulkách jsou obce identifikovány pomocí totožného ukazatele, kterým je kód obce, dle aktuálního číselníku obcí (CISOB) vedeného ČSÚ.

Dodávka výsledných dat (Finální databáze za celé území Česka) dále obsahuje:

- metodický postup řešení podrobně popisující analytický postup získání dat (včetně všech operací s daty), popisující výsledná data, jejich validitu, srovnání s konvenčními zdroji dat a interpretaci výsledných dat. – tento dokument
- mapovou přílohu ke každé tabulce, která obsahuje základní zobrazení sledovaných jevů prostřednictvím tematických map: - příloha tohoto dokumentu
  - o Základní kartogram počtu rezidentů – z Tabulky č. 1
  - o Porovnání vůči ČSÚ dle % - z Tabulky č. 1
  - o Nejčastější směr vyjížďky T1 – z Tabulky č. 2
  - o Rozdíl počtu přítomných ve 12:00 a 00:00 ve vybraném dni v týdnu (středa) – z Tabulky č. 3

## 2 Manažerské shrnutí

Hlavním cílem projektu realizace veřejné zakázky „Nákup geolokačních dat mobilních operátorů k realizaci projektu Zlepšení podmínek pro decentralizaci a dostupnost veřejné správy v území“ je vytvořit datově-analytické podklady v podobě datových sad popisujících mobilitu obyvatel v takové podobě, aby z nich bylo možné odvozovat návrhy na optimální rozmístění služeb veřejné správy v území, které v sobě bude odrážet přirozené koncentrační procesy každodenní mobility občanů.

Hlavním výstupem projektu jsou tři druhy datových sad označovaných jako tabulky. Obsah a struktura těchto tří tabulek byly stanoveny zadavatelem projektu v rámci zadání výběrového řízení. V rámci fáze PoC byla struktura těchto tabulek optimalizována tak, aby vhodným způsobem reflektovala metodické postupy zpracovatele včetně případných omezení vyplývajících z požadavků na anonymizaci dat.

Jedná se o tyto tři druhy tabulek:

- **Tabulka č. 1 - Data o počtu obvykle bydlicího obyvatelstva**
- **Tabulka č. 2 - Data o meziobecní dojížděcí**
- **Tabulka č. 3 - Denní chod počtu přítomného obyvatelstva**

Výstupy projektu jsou počítány pro 4 vybraná období v celkové délce 4 týdnů za každé z těchto vybraných období. Jedná se o tato období:

- **Podzim 2021** - celkový rozsah 28 dní vybraných během září a října 2021
- **Jaro 2022** - celkový rozsah 28 dní vybraných od března do června 2022
- **Letní prázdniny 2022** - celkový rozsah 28 dní vybraných během července a srpna 2022
- **Zima 2022/2023** - celkový rozsah 28 dní vybraných během prosince 2022 a ledna 2023

Období Podzim 2021 a Jaro 2022 byla vybrána jako typická období a ve výběru dní byly vyloučeny z pohledu mobility netypické dny - státní svátky připadající na pracovní den, dny prázdnin a pracovní dny, které mohly být ovlivněny blízkým sousedstvím s těmito netypickými dny, např. v důsledku efektu prodloužených víkendů na chování jednotlivců a případně celých rodin akcentovaného případnými dny ředitelského volna udělovaného v rámci kompetence ředitelů středních a základních škol.

Období Letních prázdnin 2022 a zimní období 2022/2023 byla vybrána jako typická období reprezentující mobilitní chování české populace během prázdnin a svátků.

Specifika mobility obyvatel pohybujících se na území ČR během zimního měřeného období jsou uvedena v kapitole 3.4.

V rámci realizace projektu byly použity analytické kategorie (příznaky) druhů prostorových vztahů uživatelů mobilních telefonů charakterizující dojíždětko-vyjíždětkové vztahy mezi obcemi vytvořené dle požadavků zadání výběrového řízení a metodicky upřesněné v rámci

fáze PoC. Tyto příznaky, tak jak jsou stručně popsány níže, představují základní analytické kategorie vyhodnocující opakované či neopakované prostorové vzorce mobility uživatelů mobilních telefonů identifikované pomocí tzv. příznaků v rámci 28 dní každého sledovaného období v prostorové agregaci na úrovni obcí.

Byly použity následující základní analytické kategorie pro analýzy mobility:

- Kategorie bydlících v obci trvale nebo opakovaně přenocujících
  - Rezidenti (**D – domov**)
  - Druhé bydlení (**D2 – druhý domov**) – specifický typ dojížděky, kdy rezident jedné obce pravidelně nocuje v jiné obci. Typově se jedná zejména o víkendové pobyty na chalupách či chatách, dojížděka studentů či pracovníků vícedenního charakteru apod.
- Dojíždějící opakovaně
  - **Typ 1 (T1) = dojížděka za prací a do škol** – nejintenzivnější typ dojížděky.
  - **Typ 2 (T2) = intenzivní dojížděka za službami** – typ dojížděky, který lze připodobnit k pravidelné dojížděce za službami, a to zejména nákupy, dále pak kulturní, sociální a sportovní služby, popřípadě návštěvy příbuzných a přátel aj.
  - **Typ 3 (T3) = občasná dojížděka za službami** – nejslabší dojížděková vazba, u které lze identifikovat pravidelnost v chování a lze ji proto považovat za cílenou a periodickou dojížděku. Lze ji připodobnit k nedenní dojížděce za specifickými službami, jako jsou kulturní, sociální, sportovní a jiné služby, popřípadě lékařská péče či návštěva úřadů či návštěvy příbuzných a přátel.
- Dojíždějící neopakovaně
  - **Nocující návštěvník (PN)** – osoba, která v dané obci není rezidentem ani pro ni není tato obec místem pravidelné dojížděky či druhým bydlením T1, T2, T3, ani D2, avšak přesto zde přenocuje.
  - **Návštěvník (N)** - osoba jednorázově navštěvující danou obec a setrvávající zde při jedné návštěvě alespoň 3 hodiny. Zároveň se však nesmí jednat o žádnou z předešlých tří kategorií opakovaně dojíždějících (T1, T2, T3) ani o PN.
- Pomocné analytické kategorie
  - **Na cestě** – specifická kategorie pro osoby, které nepobývají v dané hodině v žádné obci v rámci výše uvedených režimů
  - **Rezidenti bez SIM karty** – specifická dopočtová kategorie vytvořená pro potřeby interpretační korekce agregovaných výroků o mobilitě pro osoby, které nemají SIM kartu, tudíž jejich chování nemůže být zastoupeno ve výsledcích na základě signalizačních dat

Tyto výše uvedené základní analytické kategorie druhů prostorových vztahů byly identifikovány zadavatelem a v rámci PoC byly upřesněny v rámci dialogu zadavatele a zhotovitele volbou vhodné parametrizace nastavené tak, aby umožnily srozumitelným a

metodicky podloženým způsobem analyzovat mobilitu obyvatel ČR. Výše uvedené analytické kategorie jsou vytvářeny agregací tzv. příznaků přidělených jednotlivým uživatelům mobilních telefonů v procesu zpracování dávek dat za 28 dní příslušných jednotlivým obdobím sběru dat. Přesné nastavení podmínek přidělení příznaku uživateli z hlediska stráveného času, počtu návštěv, denní doby přítomnosti či možnosti přidělení příznaku více obcím je uvedeno v samostatné kapitole. Přidělování příznaků je hierarchické. Nejprve je přidělen domov, který musí být právě jeden. Po nalezení příznaku domova je v jedné hierarchické linii přidělován nejprve nejvýše jeden příznak T1 a následně postupně příznaky s možným vícenásobným výskytem T2, T3, PN, N. V druhé hierarchické linii je k příznaku domova hledán nejvýše jeden příznak D2. Vzhledem ke dvou hierarchickým liniím přiřazování příznaků, může docházet k překryvu příznaku D2 s dalšími příznaky z první (základní) hierarchické řady. Ve třetí hierarchické linii jsou také přiřazovány příznaky spojené s pracovní dobou – T1vPD, T2vPD, T3vPD a NvPD. Pomocné analytické kategorie – Na cestě a Rezidenti bez SIM – jsou dopočítávány specifickým způsobem dle metodiky pro tabulky, kde je to požadováno.

Hlavním datovým vstupem zpracovatele při realizaci projektu jsou signalizační data mobilních operátorů. Zpracovatel projektu v rámci projektu využil data všech tří dominantních mobilních operátorů – O2, T-Mobile a Vodafone, což maximálně možným způsobem eliminuje riziko vychýlení obrazu mobility občanů v důsledku možných specifických odchylek v časoprostorovém rozmístění uživatelské báze jednotlivých mobilních operátorů.

Datovými zdroji zpracovatele pro geografické zpracování a prostorovou interpretaci signalizačních dat jsou:

- mapy pokrytí mobilních sítí všech druhů mobilních sítí všech tří operátorů (tzv. cell mapy)
- vektorové administrativní mapy České republiky v rozlišení do úrovně obcí
- mapa zastavěných území jednotlivých obcí umožňující zpřesněný rozklad pohybových vzorců uživatelů mobilních telefonů z prostorových jednotek cell map mobilních sítí do prostorových jednotek administrativních území obcí

Datovými zdroji pro celkovou kalibraci výsledků, jejich strukturální korekturu a upřesnění prostorového rozkladu z buněk sítě do obcí jsou údaje o počtu trvale bydlících obyvatel v jednotlivých obcích ČR dle ČSÚ ze sčítání v roce 2021 dopočtené dle standardní metodiky ČSÚ k datu 1.1. 2022.

Dalším pomocným zdrojem pro strukturální korekturu výsledků jsou data ČSÚ o věkové struktuře trvale bydlících v jednotlivých obcích dle dat ČSÚ ze sčítání v roce 2021 a údaje o míře využívání mobilních telefonů v celé populaci ČR stanovené pro jednotlivé věkové skupiny dle reprezentativního rozsáhlého dotazování českých domácností v roce 2021. Strukturální korekturou se rozumí eliminace hlavních očekávaných zkreslení chování báze uživatelů SIM karet od chování populace jako celku bez ohledu na vlastnictví telefonu ve výsledcích.

Proces zpracování ze signalizačních dat na straně mobilních operátorů jakožto vstupních dat na výstupní data v podobě 3 tabulek (sestavených pomocí základních analytických kategorií vzniklých agregací příznaků identifikovaných pro jednotlivé uživatele mobilních telefonů přes územní jednotky obcí) zahrnuje tyto kroky v posloupnosti dle uvedeného pořadí:

1. Očištění dat od technických SIM karet a ponechání pouze dat, které byly generovány relevantními SIM kartami umožňující hlasové služby (odfiltrování čistě datových SIM a SIM v zařízeních, které nejsou mobilními telefony, neaktivních SIM, SIM, u nichž nebylo možné přiřadit domov). (Poznámka: Do procesu zpracování nevstupují rovněž signalizační data vztahující se k SIM kartám, k jejichž užití nebyl udělen souhlas. Jedná se o dva druhy SIM karet – jednak SIM karty skupiny uživatelů s individuálními či hromadnými (rámcovými) smlouvami s mobilními operátory, kteří neudělili souhlas se zpracováním dat, a dále pak skupinu uživatelů, kteří mají smlouvy s virtuálními operátory. Zákazníci virtuálních operátorů nemají přímý smluvní vztah s mobilními operátory účastnících se tohoto projektu a nejsou tedy součástí jejich datové báze. Případné odchylky výsledků v důsledku nerovnoměrného územního rozdělení SIM karet virtuálních operátorů ve vztahu k celkové populaci SIM karet jsou ve výsledcích kompenzovány pomocí vhodných korekčních mechanismů.)
2. Shlukování sekvencí signalizačních transakcí jednotlivých uživatelů klastrů. V rámci tohoto kroku je řešena problematika přeskokování signálu mobilního telefonu mezi blízkými buňkami. Časově blízké transakce v různých blízkých buňkách jsou sdružovány do jednoho klastru skládajícího se z více buněk sítě.
3. Přemapování klastrů z nativního georeferencování pomocí Cell-ID na prostorové referencování příslušnosti k administrativním územím obcí. V rámci tohoto kroku je používán dvoustupňový systém převodu, kdy v prvním kroku jsou klastry rozloženy kvótně dle podílu obyvatel bydlících v obcích spadajících pod buňky mobilní sítě. V druhém kroku jsou následně realokována přiřazená bydliště v rámci blízkých buněk v případech, kdy je možné zlepšit lokální rozložení bydlících v rámci sousedních obcí tak, aby se po realokaci snížila odchylka počtu bydlišť přiřazených obcím v rámci prvního kroku alokace ve srovnání s referenčními daty ČSÚ o počtu trvale bydlících v obcích ČR.
4. Identifikace mobilních vzorců uživatelů mobilních telefonů pomocí přidělení příznaků vztahujících se k základním analytickým kategoriím mobility. Přidělení příznaků D, D2, T1, T2, T3, PN, N každému uživateli a druhé doplňkové sady příznaků v pracovní době (T1vPD, T2vPD, T3vPD, NvPD).
5. Strukturální korektura pomocí korekce reprezentativity jednotlivých uživatelů dle obce bydliště v databázi mobilních vzorců. V závislosti na předpokládaném počtu osob držících více než jeden mobilní telefon v dané obci je síla hlasu mobilního vzorce každého jednoho obyvatele dané obce korigována tak, aby každý mobilní vzorec

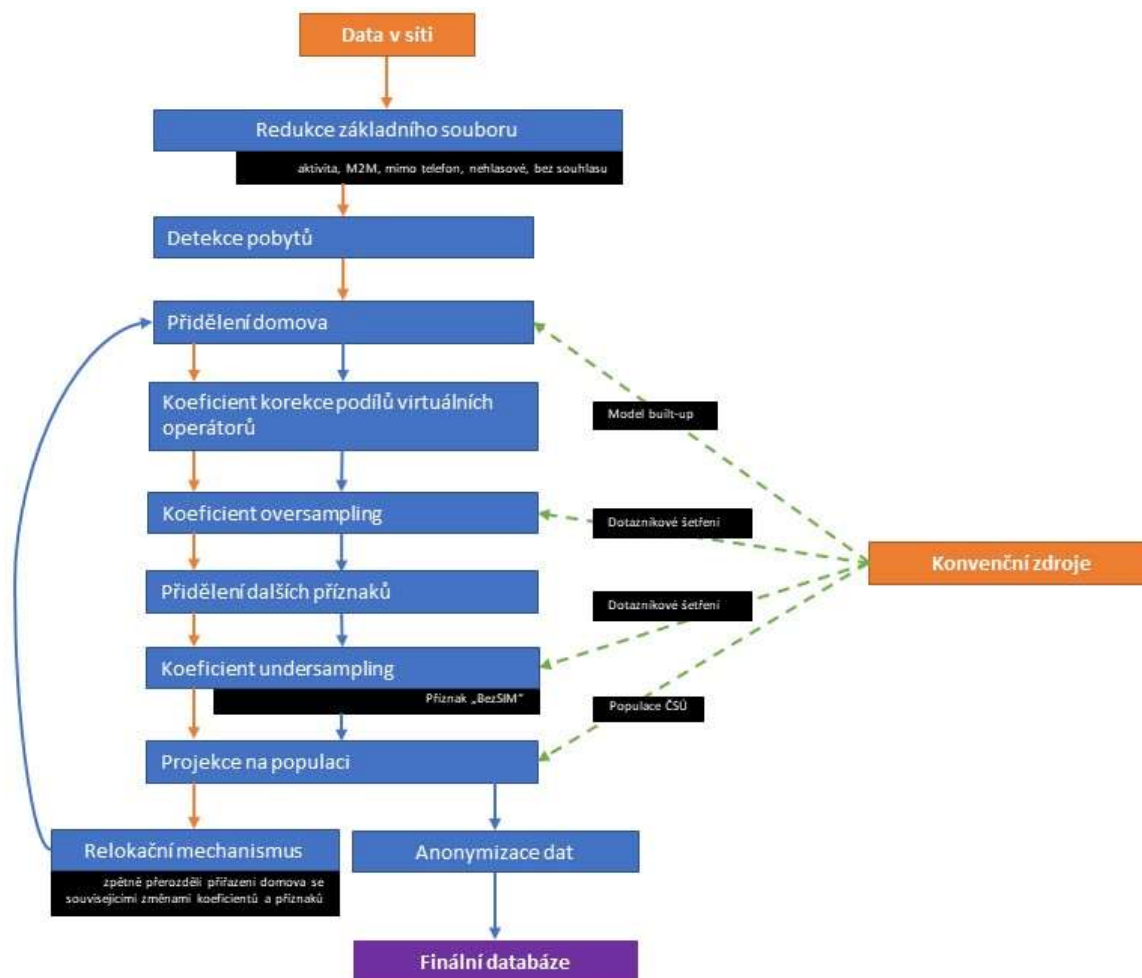
jednoho mobilního telefonu zastupoval v průměru pro danou obec právě jednoho obyvatele obce držícího mobilní telefon.

6. Výpočet exportních tabulek č. 1, č. 2 a č. 3 z databází mobilních vzorců vztažených k jednotlivým SIM kartám na straně mobilních operátorů do podoby agregátů analytických kategorií na bázi klasifikovaných příznaků vztažených k jednotlivým obcím.
7. Sloučení exportních tabulek č. 1, č. 2 a č. 3 od všech tří mobilních operátorů v prostředí integrátora dat.
8. Dopočet kategorie předpokládaných počtů osob v obcích bez mobilního telefonu dle příslušných koeficientů pro jednotlivé obce.
9. Anonymizace dat
10. Kalibrace dat strukturálně korigovaných dat na předpokládanou velikost celé české populace.

Celý proces na high-level úrovni je znázorněn na následujícím schématu. Oranžové šipky znázorňují proces zpracování až do dosažení fáze aplikace relokačního mechanismu, kdy dochází k lokálním korekcím přidělení domova. Následně proces pokračuje ve směru modrých šipek. Zelené šipky znázorňují vstupy konvenčních zdrojů.



Obrázek 1: High-level schéma zpracování dat



### **3 Vstupní data**

#### **3.1 Signalizační data**

Základním zdrojem pro mapování mobility a počtu obyvatel na území ČR dle požadavků projektu jsou signalizační data mobilních operátorů O2, Vodafone a T-Mobile. Jedná se o data generovaná technologickým vybavením sítě, která jsou využívána pro správnou funkci sítě (management), dosažitelnost mobilních zařízení (mobility management) a účtování (billing). Informace o uživateli mobilní sítě je přítomna zejména v případě, že je mobilní telefon aktivní, tj. uživatel uskuteční odchozí hovor, uživatel přijme příchozí hovor, probíhá komunikace pomocí SMS, MMS nebo je využíváno datové spojení. I v případě, že mobilní telefon uživatele není aktivní, jsou v signalizačních datech obsaženy informace o přítomnosti mobilního telefonu v konkrétních částech sítě a přepínání telefonu (bez účasti uživatele) do jiného technického prvku sítě.

Mobilní telefon nehlásí svou polohu při každé změně polohy nebo kdykoli je v dosahu signálu nového vysílače, nýbrž pouze ve specifických případech. V rámci jedné územní oblasti seskupujících více buněk tedy není přesná poloha neaktivního mobilního telefonu známa.

Jednou za stanovený interval mobilní telefon oznámí svou aktivní buňku mobilní sítě a tato událost se propíše do signalizace. Uvedený interval je v sítích mobilních operátorů nastavený typicky okolo 30 minut v závislosti na generaci technologie. Telefon, který svou polohu po tomto intervalu neohlásí, je z pohledu mobilní sítě „vypnutý“ a po technologických intervalech odstraněn z provozních registrů sítě až do jeho opětovného „zapnutí“.

Signalizační data ze své podstaty neobsahují informace o mobilním telefonu (přesněji SIM kartě), pokud je přihlášena do sítě jiného mobilního operátora, například v zahraničí během roamingu.

#### **3.2 Základní soubor SIM**

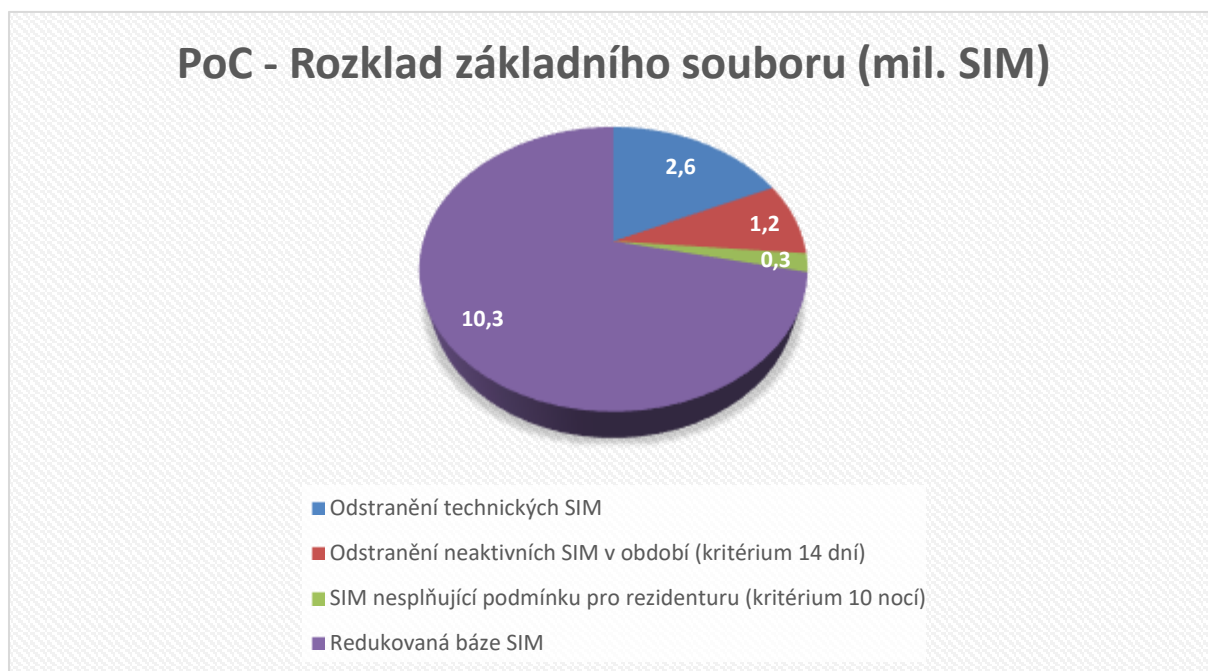
Základní soubor = souhrn SIM karet, o kterých má dodavatel k dispozici informace ze sítě a využívají hlasové služby (tzn., že základní soubor již neobsahuje SIM karty typu M2M či SIM karty užívající výhradně datové služby) a k jejichž zpracování byl uživateli udělen souhlas se zpracováním dat. Data o uživateli, kteří neudělili souhlas se zpracováním dat (týká se i zahraničních SIM), nejsou součástí základního souboru. Případné odchylky výsledků v důsledku nerovnoměrného územního rozdělení SIM karet, k jejichž zpracování nebyl udělen souhlas ve vztahu k celkové populaci SIM karet jsou ve výsledcích kompenzovány pomocí vhodných korekčních mechanismů.)

Redukce základního souboru = nutné odstranění SIM karet, které jsou vzhledem k účelu projektu přebytečné.

- SIM karty v jiných zařízeních než mobilní telefony – odstranění tzv. technických SIM karet v jiných zařízeních (tablety, hodinky aj.) je provedeno s využitím informací v signalizačních datech o typu zařízení. Spadají sem i vydané, ale v měřeném období nepřihlášené SIM, u kterých není možné specifikovat jejich typ.
- Neaktivní SIM karty – SIM karta musí vykazovat aktivitu (záznamy v síti) alespoň v polovině sledovaných dnů (tj. aktivita alespoň ve 14 dnech)
- SIM s nepřirazenou rezidenční obcí – Nelze-li SIM kartě přiřadit rezidenční obec (místo domova) neexistuje pro ni základní referenční údaj a není s ní možné dále pracovat. SIM bez přiřazené rezidenční obce nebudou do analýzy využity. Tento bod přímo souvisí s podmínkami pro určení Domova (rezidenční obce) viz níže.

Redukce báze SIM pro zvolené období je patrná na následujícím diagramu, přičemž signalizační data ze SIM karet zahraničních operátorů do zpracování nevstupují. Způsob redukce vícenásobných SIM karet uživatelů je uveden v kapitole „Korekce reprezentativity redukované báze SIM karet“.

Obrázek 2: Rozklad základního souboru SIM



Výsledná redukováná báze SIM v jednotlivých sběrech (FD1 – FD4) mírně fluktuuje (10,276 mil; 10,300 mil; 10,287 mil; 10,316 mil). Ke stálosti tohoto ukazatele v čase přispívají následující faktory:

- Parametrické nastavení redukce báze
  - o proces redukce báze a odstranění SIM včetně stanovení kritérií neaktivity byl nastaven v rámci fáze Proof of concept s využitím expertních znalostí řešitelského týmu a ve spolupráci s Objednatelém tak, aby vytvářel co nejlepší

základ pro sledování požadovaných příznaků. Následná realizace 4 sběrů ukázala, že toto parametrické nastavení současně vede na dlouhodobě stabilní redukovanou bázi SIM.

- Ze struktury rozkladu základního souboru vyplývá postupné konvergování k výsledné hodnotě. Jinými slovy v každém redukčním kroku je odstraněno méně SIM karet než v předchozím. Porovnání sběrů ukazuje, že zejména poslední krok odstranění nepřířaditelných SIM k rezidentuře vede na stabilní bázi.
- Do jednotlivých redukčních kroků posuzování aktivity/neaktivity dále vstupuje mj. strategie výběru měřeného období (souvislé, nesouvislé). To může mít vliv na odlišnou strukturu redukce v měřených obdobích.
- Projektu se účastní všichni dominantní operátoři (O2, T-mobile, Vodafone). Výhodou je, že jsou v datech fakticky eliminovány přesuny zákazníků mezi operátory.
- Trend růstu velikosti výchozí báze může být způsoben marketingovým tlakem operátorů na vydávání větších a větších počtů SIM karet, přestože strana poptávky (zejména reprezentovaná aktivně používanými hlasovými SIM kartami) je nasycena.
- Báze nezahrnuje zahraniční SIM karty. Problematika migrace a její možné vlivy na datový soubor je blíže popsána v kap. 3.4.

Celková velikost základní báze i velikost redukované báze jsou pouze dvěma z mnoha charakteristik datové základny v komplexním procesu zpracování signalizačních dat. Tyto ukazatele přinášejí představu o velikosti vstupní datové báze, není je však vhodné používat pro účely interpretace počtu přítomných, bydlících obyvatel apod. Snahou projektu bylo relevantní vlivy promítnout do navazujícího procesu zpracování dat, což dokumentuje metodika projektu.

### 3.3 Územní dimenze

Zájmovým územím je celá Česká republika v podrobnosti do úrovně obcí. Praha a další územně členěná města jsou považována vždy za 1 obec. Česká republika je rozdělena do 6254 obcí, za které budou data poskytnuta. Údaje připadající na 4 vojenské újezdy jsou individuálně řešeny tak, aby neměly žádné rezidenty a aby údaje za okolní obce nebyly zkresleny. Údaje za tyto 4 vojenské újezdy nejsou součástí datasetu.

### 3.4 Časová dimenze

Měření sleduje úsek 28 dnů. Období stanovil Objednatel po konzultacích s Dodavatelem s cílem vyhnout se termínům ovlivněných technickými problémy v síti. Tímto postupem je možné eliminovat zásadní výpadky v sítích operátorů poskytujících data.<sup>1</sup> Dalším kritériem,

<sup>1</sup> Nicméně i mimo tyto významné výpadky na úrovni sítí jednotlivých operátorů může docházet k lokálnímu a časově omezenému nestandardnímu chování sítě vyplývajícího z provozně-technologických aspektů konkrétních vysílačů. V důsledku se tak např. v hodinových řezech (tj. Tabulka č. 3) mohou u některých obcí vyskytovat lokální odchylky od očekávaného průběhu sledovaných kategorií.

které vstupuje do volby období, byla specifická konkrétních dnů mimo běžný týdenní cyklus, jako jsou dny státních svátků, prázdnin apod. a specifika daného časového období.

Pro Finální databázi 4 bylo stanoveno období 28 dnů, a to 7.12.2022 – 20.12.2022, 29.12.2022 – 6.1.2023 a 14.1.2023 – 18.1.2023.

Měřené období bylo stanoveno po vzájemné konzultaci Dodavatele a Objednatele a reflektuje výskyt technických překážek u provozovatelů sítě, v jejichž důsledku jsou údaje v některých dnech pro potřeby měření nevyužitelné. Zároveň bylo zapotřebí vyhnout se dnům vánočních svátků, ve kterých lze očekávat jiný, výrazně acyklický charakter dojížděkového chování mimo týdenní a delší cykly, než na který je tento sběr dat s frekvenčně zaměřenou metodikou zacílen. Z toho důvodu nebylo možné koncipovat měřené období jako souvislé čtyřtýdenní období, jak bylo původně zamýšleno dle Přílohy č. 2 Smlouvy.

Mobilita obyvatel pohybujících se na území ČR v zimě roku 2022/2023 byla, stejně jako v předešlých dvou sběrech poznamenána eskalací válečného konfliktu na Ukrajině v důsledku invaze na další území Ukrajiny ze strany Ruska dne 24. 2. 2022. Tato událost ovlivnila mobilitu obyvatel pohybujících se na území ČR v několika rovinách:

- Ve dnech následujících po únorové eskalaci válečného konfliktu přišla z Ukrajiny výrazná vlna uprchlíků, kteří se z pohledku jejich chování v mobilních sítích na území ČR mohli „domestikovat“ - získat a začít užívat české SIM karty ať již se jednalo o SIM karty získané darem nebo samostatným pořízením SIM karty
- Příliv uprchlíků z Ukrajiny pokračoval i v dalších jarních a letních měsících, které spadaly do měřené období
- V opačném směru proti nově příchozím z Ukrajiny naopak došlo k odchodu ukrajinských mužů či případně žen, jejichž status se z pohledku vlastnictví SIM karty mohl jevit v rámci ČR jako domácí - odešli Ukrajinci držící české SIM karty zpět na Ukrajinu
- V průběhu jarního a letního období spadajícího do analyzovaného období rovněž docházelo k odchodům nově příchozích Ukrajinců a naopak k návratu v ČR dlouhodobě pobývajících Ukrajinců zpět do ČR. Tyto příchody a odchody mohly být i opakované.
- V průběhu zimního období pravděpodobně pokračovaly obousměrné pohyby ukrajinských občanů mezi ČR a Ukrajinou. Jako návratový faktor mohl začít působit relativní stabilizace bezpečnostní situace, na druhou stranu ovšem mohly působit mimořádně obtížné životní podmínky na Ukrajině a riziko další eskalace konfliktu.

Přesná kvantifikace těchto jevů není známá, existují pouze odhady různé kvality dílčích komponent těchto jevů a jejich časového průběhu, stejně jako nejsou v dostatečné míře známy „průměty“ těchto jevů do signalizačních dat v důsledku „analytické domestikace“ mobility ukrajinských občanů držení české SIM karty. V důsledku toho nelze identifikovat žádný dostatečně odůvodněný návrh korekce těchto jevů spojených s válkou na Ukrajině, lze pouze

konstatovat, že tento fenomén je v datech přítomen, ale jeho agregovaný význam a struktura jeho složek je neznámá.

Není známa ani odlišnost souhrnného výsledku těchto jevů ve srovnání s jarním a letním obdobím, kdy různé faktory této mobility mohly mít jinou váhu.

Rovněž je nutné poznamenat, že celkový efekt mezinárodních dovolených českých občanů a pohybů občanů Ukrajiny držících českou SIM kartu působí na celkovou velikost redukované báze SIM karet v souhrnu a tyto složky nelze od sebe navzájem oddělit.

Měřené období je specifické z pohledu zařazení dnů kolem konce roku. To se promítlo mj. do denního chodu přítomného obyvatelstva v některých rekreačních oblastech. Vysoká návštěvnost na konci roku se projevuje ve skocích počtu přítomných, které jsou zřejmé zejména na přelomu středy/čtvrťky v průměrovaném týdnu. To je způsobeno kontrastem „běžných“ pracovních střed ve výběru a specifickým čtvrtkem 29.12., v kterém se již významně promítá přítomnost návštěvníků/rekreatů před Silvestrem v rekreačních destinacích.

Stanovená kritéria aktivity v konečném důsledku po odstranění 4,1 mil. SIM splnilo 10,3 mil SIM karet. Redukovaná báze je tak na stejné úrovni, jako u předcházejících Finálních databází. Rozdíly lze však spatřit ve skladbě SIM karet, které byly redukovány z důvodu nenaplnění kritérií. Možnou příčinou může být nejen podíl vydaných (ale v měřeném období neaktivních) SIM, nebo zaznamenáním aktivity SIM ve smyslu jejich (ne)přítomnosti v ČR.

### 3.5 Anonymizace dat

Data byla anonymizována do té míry, aby nebylo možné ztotožnit konkrétního jedince s konkrétním údajem.

Z důvodu anonymizace je stanovena kritická hodnota pro zveřejnění údaje. Kritickou hodnotou pro zveřejnění údaje je 7, hodnota 0 značí neexistenci vztahu je vymezena zvlášť. Tzn., že jsou rozlišeny 3 případy:

- 7 a více = zveřejněná hodnota
- 1 – 6,9 = označeno symbolem „xx“
- 0 = vztah/jev neexistuje („nikdo nedojíždí“)

Číselný údaj v Tabulce č. 1 s názvem „Vyjíždějící za prací/školou“ představuje pro danou obec součet všech neanonymizovaných vztahů s typem cesty 1 uvedených v Tabulce č. 2. Jinak řečeno, nejsou v něm započítány hodnoty označené „xx“.

Sloupec s názvem „T1 anon.“ V Tabulce č. 1 u každé obce představuje počet relací cesty typu 1 z Tabulky č. 2, které byly anonymizovány. Jako důsledek je tedy tento ukazatel v Tabulce č. 1 ponížěn o hodnoty nevyjádřených (anonymizovaných) směrů.

Na ukazatel celkového počtu rezidentů (údaj v Tabulce č. 1) a „Celkový počet přítomných osob“ v Tabulce č. 3 bylo aplikováno zarušení, které je vzhledem k hodnotě v % (v jednotkách)

---

i absolutním vyjádření ve vztahu k celkové hodnotě nevýznamné.<sup>2</sup> Účelem je zajistit nemožnost dopočtu konkrétních malých hodnot v nevyjádřených kategoriích v datasetu.

---

<sup>2</sup> Konkrétní metoda anonymizace není uvedena z důvodu možného využití uvedených informací pro pokusy o deanonymizaci datasetu.

## **4 Proces získávání dat a způsob měření**

Za pobyt v jedné buňce jsou považovány alespoň 2 po sobě následující přihlášení ke stejné buňce BTS. Buňky BTS jsou následně spojovány do klastrů, přičemž pohyb mezi buňkami v rámci jednoho klastru je považován za pokračování pobytu. Tím jsou odstraněny nežádoucí vlivy překmitávání. Klastry včetně elementárních klastrů v podobě jedné transakce v jedné BTS jsou po převedení na územní příslušnost k jednotlivým obcím následně shlukovány dle území obcí do podoby pobytů v obci.

Převod klastrů na územní příslušnost k jednotlivým obcím je prováděno na základě geostatistického modelu přepočtu z polygonů území pokrývaných jednotlivými BTS na polygony území jednotlivých obcí.

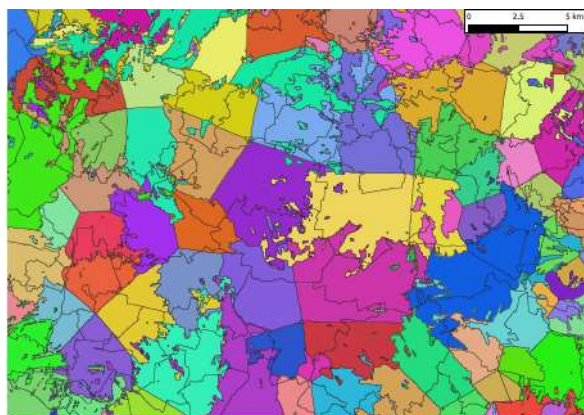
### **4.1 Georeferencování signalizačních dat**

Nedílnou a stěžejní součástí zpracování telekomunikačních signalizačních dat je přiřazení polohové informace signalizačním datům. Pro uvedený proces jsou k dispozici pouze číslo buňky CGI a mapy pokrytí území ČR jednotlivými buňkami. Na základě přijímaných dat neexistuje žádné jiné zpřesnění polohy v buňce. Jelikož data obsahují pouze číslo buňky, jiné existující technologie poskytující přesnější lokalizaci mobilního telefonu, jako například Cell-ID+TA/RTT, AECID, U-TDoA, OTDoA, E-OTD, nejsou použity a nelze je využít ani dodatečně.

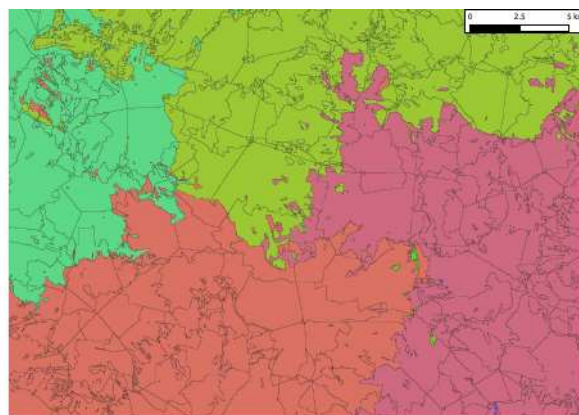
Nejbližší aproximací tvaru buňky, kdy buňkou se rozumí území obsluhované jednou unikátně identifikovatelnou směrovou anténou umístěnou na jedné vysílací věži, poskytuje mobilní operátor ve formě map pokrytí (cellmap). Jedná se o schematicky vyznačené geografické území na mapě ČR, které je v ideálním případě dle modelů a měření mobilního operátora pokryto konkrétní směrovou anténou umístěnou na (fyzickém) vysílači. Ukázka výřezu mapy pokrytí je k dispozici na obrázku níže.



Obrázek 3: Ukázka buněk mobilní sítě. Vlevo jsou zobrazeny buňky mobilní sítě s vybarvením podle čísla buněk cell global identity. Vpravo jsou vyobrazeny ty samé buňky, ale s vybarvením podle tracking area code



(a) Cell Global Identity



(b) Tracking Area Code

Mapy pokrytí jsou poskytovány ke zpracování signalizačních dat obvykle jednou měsíčně pro každou síťovou technologii. Každá mapa reflektuje změny za uplynulý měsíc: obsahuje nové mobilní buňky a jsou z ní odstraněny buňky staré. Síťové technologie rozlišujeme v síti mobilních operátorů v současné době následovně: GSM, FDD, LTE800, LTE1800, LTE2100, LTE2600. I v rámci sítě jednoho operátora se síťové technologie vzájemně na území ČR překrývají. Některé síťové technologie jsou „plošné“, tedy pokrývají celé území ČR, zatímco jiné jsou lokálního charakteru, například LTE sítě pouze ve velkých městech a jejich okolí. Existují tedy místa, která jsou pokryta více typy síťové technologie a místa, kde stále ještě dominuje GSM a FDD. Každé buňce mobilní sítě je přiřazeného operátorem unikátní číslo CGI, cell global identity, které sjednocuje různé druhy identifikátorů buněk napříč technologiemi.

## 4.2 Cellmapping – přemapování

Poloha mobilního telefonu je z výše uvedených důvodů aproximována na obec v procesu nazvaném cellmapping. Pro účely mapování buněk na administrativní jednotky je vytvořena podpůrná vrstva obcí s rozdělením obcí na zastavěnou část obcí (muni build-up areas) a zbylou nezastavěnou část. V rámci rozpočítávání prostorového rozložení obyvatelstva jsou při mapování buněk obce počítány průniky obcí v rozsahu jejich zastavěných částí ČR (nezastavěné části jsou ignorovány) s každou jednotlivou mapou pokrytí pro každou technologii. Na základě velikosti plochy překryvu celé nebo části buňky s celou nebo částí plochy build-up area polygonu obce, a v kombinaci s očekávaným počtem obyvatel v obcích dle ČSÚ, jsou vypočteny distribuce pravděpodobnosti udávající přítomnost SIM karty v libovolné obci pro každé jednotlivé číslo buňky.

Takto vytvořené distribuce vstupují do všech následných výpočtů a jsou kombinovány s pokročilými technikami shlukování míst pro uživatele mobilní sítě za účelem zpřesnění lokalizace.

### 4.3 Čištění dat

Protože se buňky v mobilní síti mění, nové se přidávají a staré ubírají, a navíc dochází ke změnám v technologii sítě a vybavení i nastavení core network, je třeba před dalším použitím signalizačních dat tato data očistit od nepravděpodobných přesunů. Analýzou vstupních dat jsou zjišťovány opakované i chybové záznamy, které jsou důsledkem hardwarové chyby samostatného dodavatele síťového řešení v síti mobilního operátora. Na data je aplikován čistící algoritmus, který odstraňuje záznamy se shodným časovým razítkem, protože nelze určit, která buňka je správná. Algoritmus dále odstraňuje události, které vznikly příliš brzy po sobě a týkají se nepřírodně vzdálených buněk. Odstraňovány jsou také anomálie v datech identifikovatelné např. skokovými nárůsty/poklesy při detailní analýze denních průběhů. Tímto procesem jsou data očištěna od hrubých strukturálních vad a vstupují do části shlukování buněk pro jednotlivé uživatele.

### 4.4 Detekce míst pobytu

Je dobře známým faktem, že mobilní telefon se může přepínat (přeladovat) na okolní buňky bez ohledu na to, zda je v pohybu nebo ne. Tento efekt je známý jako ping-pong nebo cell jitter. Spolehne-li se pouze na aktuální číslo buňky a k ní příslušnou pozici na mapě, mění v tomto pohledu uživatel svou polohu v řádech sekund na vzdálenosti několika set metrů po kilometry, což zcela jistě neodpovídá realitě. Z tohoto důvodu je zapotřebí cell jitter odstranit. Protože jsou statistiky pohybu obyvatelstva požadovány vzhledem k administrativnímu členění České republiky, je dalším krokem ve zpracování mobilních dat přiřazení míst uživateli mobilního telefonu. Místo pro uživatele definujeme jako obec, jíž přísluší konkrétní množina buněk mobilní sítě a na níž má uživatel konzistentní chování v rámci zkoumaného období.

Vlastní detekce míst pobytu je prováděna na dvou úrovních:

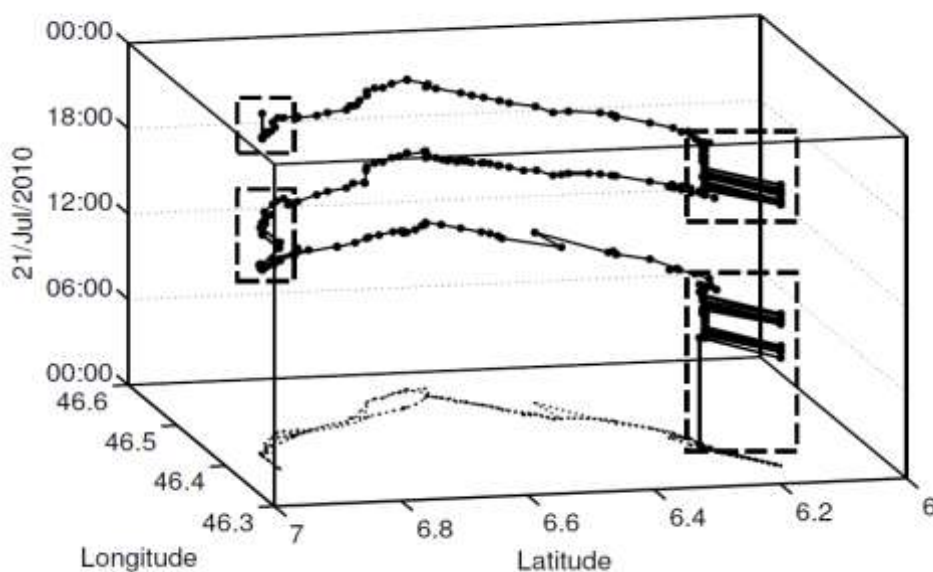
#### Ad1) Shlukovací algoritmus

Pro určení míst elementární přítomnosti (pobytu) v rámci buněk je aplikován mj. shlukovací algoritmus. Ten shlukuje buňky prioritně podle času na nich stráveném v případě, že jsou dle definované metriky dostatečně blízko. Oproti běžně používaným řešením není použita v tomto kroku metrika eukleidovská, tedy vzdálenost mezi buňkami, nýbrž časová, kdy je podobnost buněk definována rychlým přepnutím mezi buňkami.

Signalizační data jsou poté prolnta s mapovým pokrytím s přihlédnutím k časové známce záznamu a časové známce mapy pokrytí. Platí, že informace o pokrytí jsou vztahené k nejbližší časové známce dat. Výsledným shlukům buněk jsou přiřazeny dle předpočítaného cellmappingu kódy obcí na základě maximality sečtené distribuce příslušnosti všech CGI v místě k obci, převážným časem stráveným v každé buňce. Tím je dosaženo vyšší robustnosti přiřazení obcí, které zároveň sleduje časové charakteristiky výskytu uživatele v buňce. Výsledná místa jsou dále shlukována podle přiřazeného čísla obce a časové charakteristiky výskytu na místě. Motivací je častá změna buněk v síti a instalace nových technologií, které se promítají do map pokrytí. Je-li na určitém místě změněna topologie sítě a jsou-li přidány nebo odebrány buňky, získává shluk buněk sice stejné umístění, ale jiné pořadové číslo a je tak v dosavadních krocích jiným místem. Shlukování míst podle obce a časové charakteristiky výskytu částečně kompenzuje dopad změn v topologii sítě. Trajektorie uživatele v mobilní síti je po detekci míst nadále popisována nikoli přeskoky mezi buňkami, ale změnami míst v čase. Tím dochází k efektivní komprimaci trajektorií a zároveň obohacení o informaci o geografické poloze.

*Obrázek 4: Příklad detekce míst. Pohyb uživatele mobilní sítě je zanesen do 3D space-time cube.*

*Algoritmus detekuje místa, která ač mají rozdílnou polohu na mapě, dochází mezi nimi k tak rychlému přepnutí z místa na místo, že musí tvořit jedno místo. Na obrázku jsou vidět dvě místa (na pravé a na levé straně) a tři přesuny mezi nimi během jednoho dne. Tečkovaná čára je ortogonální projekcí trajektorie v čase na mapu.*



## Ad 2) Aplikace úprav výpočtu

V návaznosti na přidělení příznaků, viz kapitoly uvedené níže, je v druhém stupni aplikována úprava výpočtu vycházející z příslušnosti rezidentů k obcím. Vzhledem k technologickým limitům signalizačních dat v kontextu požadované územní a časové granularity dochází na úrovni přiřazování k území jednotlivých obcí k nejednoznačným situacím. Při porovnání počtů rezidentů přiřazených k obcím na základě zpracování signalizačních dat s konvenčními datovými zdroji popisujícími obdobné jevy (počet obyvatel dle ČSÚ) dochází na úrovni obcí k rozdílům. Přestože jsou vzniky těchto rozdílů očekávatelné z povahy dat, bylo po dohodě s Objednatelem přistoupeno k úpravě výpočtu, a to ve smyslu aplikování relokačního mechanismu. V tomto kontextu bylo v průběhu projektu Objednatelem sděleno kritérium úměrnosti rozdílů, a to v metrice procentuální odchylky počtu rezidentů (stanovených pro danou obec dle *maxima stráveného času za celé vyhodnocovací období v Noci s minimem 10 nocí*) oproti údajům o bydlících z ČSÚ.

Cílem relokačního mechanismu je hledat obce, které mají v uvedené metrice největší odchylku a následně tuto odchylku korigovat v kontextu jejich nejbližšího okolí. Plánovány jsou přesuny primárně mezi obcemi ležícími v nejbližším okolí s odchylkami na opačnou stranu od průměrné hodnoty v rámci obcí identifikovaných jako nejbližší okolí. Přesuny jsou na nejnižší úrovni plánovány v granularitě jednotlivých built-up areas (příslušících konkrétním obcím) s rezidenty, přičemž se postupuje dle kritéria vzdálenosti (preferovány jsou built-up areas, která jsou nejbližší cílové obci). Z hlediska vzdálenosti je horním limitem přesunu obec sousedící se sousední obcí (tento limit byl stanoven s ohledem na podobu územních hranic obcí, kdy i soused souseda může být velice blízko řešené obce) a vzdálenostní podmínka, která zamezuje vzdáleným přesunům u rozlehlých obcí v sousedství Prahy, Brna apod.

Po automatizovaném naplánování přesunů dochází v dalším kroku k tzv. manuální korekci, kdy dochází ke korekci obcí/oblastí, kde automatický mechanismus nenaplánoval přesuny dle expertního očekávání. Tzn. na jeho výstupu se vyskytovaly oblasti např. kde sousedí významně nadhodnocené a podhodnocené obce, kde se relokační mechanismus aplikoval vyšší míru agresivity přesunu apod. Tento krok byl řešen v úzké spolupráci Objednatele s Dodavatelem.

V rámci Finální databáze 4 se relokace týkala 1,3 % populace.

Pobyty (v pojetí této zakázky) = v rámci zpracování zakázky jsou za pobyt považovány jen ty pobyty, kdy mezi prvním a posledním přihlášením ve stejné obci uplynulo alespoň 30 minut. Pobyt počíná prvním přihlášením k BTS a končí posledním přihlášením k BTS.

Přechodový interval = interval mezi dvěma pobyty, tj. období mezi posledním přihlášením k jedné buňce BTS (kde proběhl pobyt) a prvním přihlášením k následující buňce BTS (kde

proběhl pobyt). Přechodový interval není započítáván k žádnému z pobytů (předcházejícímu ani následujícímu).

V případě, že mezi dvěma po sobě následujícími přihlášeními ke stejné buňce BTS dojde k prodlevě delší, než je 1 hodina, je na tento interval pohlíženo jako na přechodový. Dojde tedy k přerušení dosavadního pobytu posledním přihlášením a případnému započítání nového pobytu následujícímu po přihlášení.

#### 4.5 Metoda přiřazování příznaků místům uživatele

Na základě znalosti pobytu každé SIM karty v jednotlivých obcích během sledovaného období jsou každé SIM kartě přiřazovány tzv. příznaky (labely).

Příznak = určitý vztah mezi konkrétní SIM kartou a konkrétní obcí. SIM karta dle svého výskytu má různé vztahy k různým obcím. Ke každé z obcí má však SIM karta zpravidla jen jeden příznak (existuje jedna výjimka z tohoto pravidla, viz dále).

##### 4.5.1 Parametry pro přidělování příznaků

Časové kategorie:

- Noc: 22:00 – 6:00
- Pracovní doba: 6:00 – 18:00
- Den týdne:
  - Pracovní den: Po, Út, St, Čt, Pá
  - Víkendový den: So, Ne
- Den týdne přenocování: (pozn.: „Út“ znamená noc z pondělí na úterý, rozšířením okna přes půlnoc je nutno interpretovat přes hranici kalendářního dne)
  - Noc všedního dne: Út, St, Čt, Pá
  - Noc víkendového dne: So, Ne, Po
- Časové kategorie exportních časových řezů (pro potřeby Tabulky č. 3)
  - hodinové řezy

Pomocné časové kategorie („plovoucí časová okna“):

- Přenocování: – podmínka Přenocování – minimálně 3 hodiny souvislého pobytu v rámci jedné noci (pro atributy PN a D2); během jedné noci je možné přenocovat jen v 1 obci, v případě že je splněno kritérium vůči dvěma obcím, bude za obec přenocování považována ta, kde byl pobyt delší.
- Návštěva: podmínka minimální souvislé délky pobytu Návštěvy v rámci kalendářního dne – 3 hodiny

##### 4.5.2 Příznaky

Rezident (D – domov) – jedná se o základní příznak určující obec, kterou lze považovat za místo bydliště. **Každá SIM karta ve sledovaném souboru musí mít přiřazeno právě jedno místo, kde je rezidentem.** Není-li možné přiřadit rezidenční obec, SIM karta je z analyzovaného souboru vyřazena. Je-li obec pro jedince rezidenční, nemůže již vůči této obci mít přidělen žádný jiný atribut.

Rezidenční obec (domov) je ta obec, kde jedinec **tráví ve sledovaném měření nejvíce času v Noci.**

*D – domov*

- *Podmínka: obec s maximem stráveného času za celé období v Noci. Zároveň musí být dostupné záznamy alespoň z 10 Nocí.*

Dojíždějí – jsou definovány **3 typy dojížděky** rozlišené dle jejich intenzity (počet návštěv a celkový strávený čas).

Typ 1 (T1) = dojížděka za prací a do škol – nejintenzivnější typ dojížděky. **Každá SIM karta může mít přiřazenu pouze jednu či žádnou obec dojížděky T1.**

Příznak T1 je přidělen těm SIM kartám, které ve sledovaném období **danou obec navštíví alespoň ve 13 dnech a celkem zde stráví alespoň 50 hodin.** V případě, že pro SIM kartu neexistuje žádná obec (vyjma rezidenční obce), ve které by se vyskytovala s takto stanovenou intenzitou, není tento příznak udělen vůči žádné obci. V případě, že stanovené podmínky splňuje SIM karta vůči většímu množství obcí, příznak T1 je přiřazen k té obci, kde má uživatel T1vPD. Ostatní takové obce jsou pak zahrnuty do dojížděky typu 2 (T2).

**T1 – Dojížděka 1. typu - (typ cesty = 1)**

- *Identifikace kandidátů dle třech kritérií, které musí platit současně:*
  - *Podmínka 1: obec s minimálně 13 pobyty v obci v různých dnech z celého období*
  - *Podmínka 2: celková doba pobytu v obci minimálně 50 h z rámci celého období (tj. ve dne i v noci bez ohledu na pracovní dobu a období pro možné načtení přenocování)*
- *Výběr vhodnějšího kandidáta:*
  - *v případě více kandidátů se vybere kandidát s přiřazeným příznakem T1vPD*
  - *pro T1vPD se v případě více kandidátů vybere kandidát s více pobyty v obci v různých dnech z celého období, při stejném počtu je dalším kritériem větší hodnota součtu celkové délky všech pobytů v obci v celém období.*

Typ 2 (T2) = intenzivní dojížděka za službami – typ dojížděky, který lze připodobnit k pravidelné dojížděce za službami, a to zejména nákupy, dále pak kulturní, sociální a sportovní služby, popřípadě návštěvy příbuzných a přátel aj. Každá SIM karta může mít přiřazen tento příznak vůči několika obcím.

Příznak T2 je přiřazen jen těm SIM kartám, které nejsou již zahrnuty v kategorii T1 a zároveň ve sledovaném období **danou obec navštívily alespoň 4krát a celkem zde strávily alespoň 8 hodin.**

**T2 – Dojíždka 2. typu - (typ cesty = 2)**

- *Všechny kandidátské obce splňují obě dvě kritéria:*
  - *Podmínka 1: obec s minimálně 4 pobyty v různých dnech z celého období*
  - *Podmínka 2: celková doba pobytu v obci minimálně 8 h z celého období*

Typ 3 (T3) = občasná dojíždka za službami – nejslabší dojíždková vazba, u které lze identifikovat pravidelnost v chování a lze ji proto považovat za cílenou a periodickou dojíždku. Lze ji připodobnit k nedenní dojíždce za specifickými službami, jako jsou kulturní, sociální, sportovní a jiné služby, popřípadě lékařská péče či návštěva úřadů či návštěvy příbuzných a přátel. Jedna SIM karta může mít přidělen tento příznak vůči několika obcím.

Příznak občasné dojížděcího za službami je přiřazen pouze tehdy, pokud nebyl již přidělen příznak T1 či T2 vůči dané obci a zároveň ve sledovaném období **byla obec navštívena alespoň 2krát a celkem zde byly stráveny alespoň 4 hodiny.**

**T3 – Dojíždka 3. typu - (typ cesty = 3)**

- *Všechny kandidátské obce splňující obě podmínky:*
  - *Podmínka 1: obec s minimálně 2 pobyty v různých dnech z celého období*
  - *Podmínka 2: celková doba pobytu v obci minimálně 4 h (součet za období)*

Druhé bydlení (D2) – specifický typ dojíždky, kdy rezident jedné obce pravidelně nocuje v jiné obci. Typově se jedná zejména o víkendové pobyty na chalupách či chatách, dojíždka studentů či pracovníků vícedenního charakteru apod. Každá **SIM může mít stanovenou jen jednu či žádnou obec s příznakem D2.** V případě, že níže stanoveným parametrům vyhovuje více obcí, příznak D2 je přiřazen vůči té obci, kde přenocoval nejčastěji, v případě shody, kde strávil celkově více času v noci.

Příznak D2 je přiřazen, pokud proběhne alespoň 6 přenocování během sledovaného období. Přičemž tato přenocování musí být uskutečněna v pravidelné periodicitě, tj. všechna přenocování musí být uskutečněna ve víkendových nocích nebo v nocích všedních dnů.

Z výše uvedených kritérií a zvolených parametrů nastavení je zřejmé, že pokud bude SIM kartě přidělen příznak D2, může být stejnému místu uživatele přidělen příznak v jiné hierarchické

linii příznaků. Při zvolené parametrizaci příznaků může docházet k překryvu mezi D2 a T2 či T3. Příznak D2 nebude figurovat v součtech (zejména Tabulka č. 3) jednotlivých příznaků, jelikož SIM karty již budou započítány v rámci příznaku T2 nebo T3.

**D2 – Druhé bydlení - (typ cesty = 4)**

- *Identifikace kandidátů dle dvou kritérií, musí platit jedno z nich:*
  - *minimálně 6x za celé období Přenocování v Noc Všedního dne v jedné obci*
- *NEBO*
  - *minimálně 6x za celé období Přenocování v Noc Víkendového dne v jedné obci*
- *Výběr vhodnějšího kandidáta:*
  - *v případě více kandidátů se vybere kandidát s více přenocováními (součet přenocování v Noci Všedního dne a v Noci Víkendového dne), při stejném počtu je dalším kritériem větší hodnota součtu celkové délky všech Přenocování v místě (délkou přenocování je maximální souvislá délka přenocování v rámci období v rámci něhož může být plněna podmínka na přenocování v dané noci)*

Nocující návštěvník – SIM, která v dané obci není rezidentem ani pro ni není místem pravidelné dojíždky či druhým bydlením T1, T2, T3, ani D2, avšak přesto zde přenocuje. Jedné osobě/SIM kartě může být přiřazeno více míst, kde je nocujícím návštěvníkem.

Nocujícím návštěvníkem je ta osoba, která v obci ve sledovaném čtyřtýdenním období alespoň 1x přenocovala a zároveň nemá přidělený vůči obci žádný z předešlých příznaků.

**PN – Přenocující návštěvník - (typ cesty = 5)**

- *Všechny přípustné kandidátské obce splňující podmínku minimálně jednoho přenocování*

Návštěvník – osoba jednorázově navštěvující danou lokalitu a setrvávající zde při jedné návštěvě alespoň 3 hodiny. Zároveň se však nesmí jednat o žádnou z předešlých tří kategorií příznaků – T1, T2, T3 ani o PN. Jedné SIM kartě může být přiřazeno více obcí.

**N – Návštěvník - (typ cesty = 6)**

- *Všechny přípustné kandidátské obce splňující podmínku s délkou jednoho pobytu minimálně 3 hodiny (pobyt souvisle min 3 h)*



Pracovní doba – v tabulce meziobecních dojížděkových vztahů – Tabulka č. 2 (dále podrobněji vysvětleno) jsou definované příznaky sledovány rovněž jen pro časový úsek reprezentující pracovní dobu. Jsou tedy vymezeny další příznaky T1vPD, T2vPD, T3vPD a NvPD.

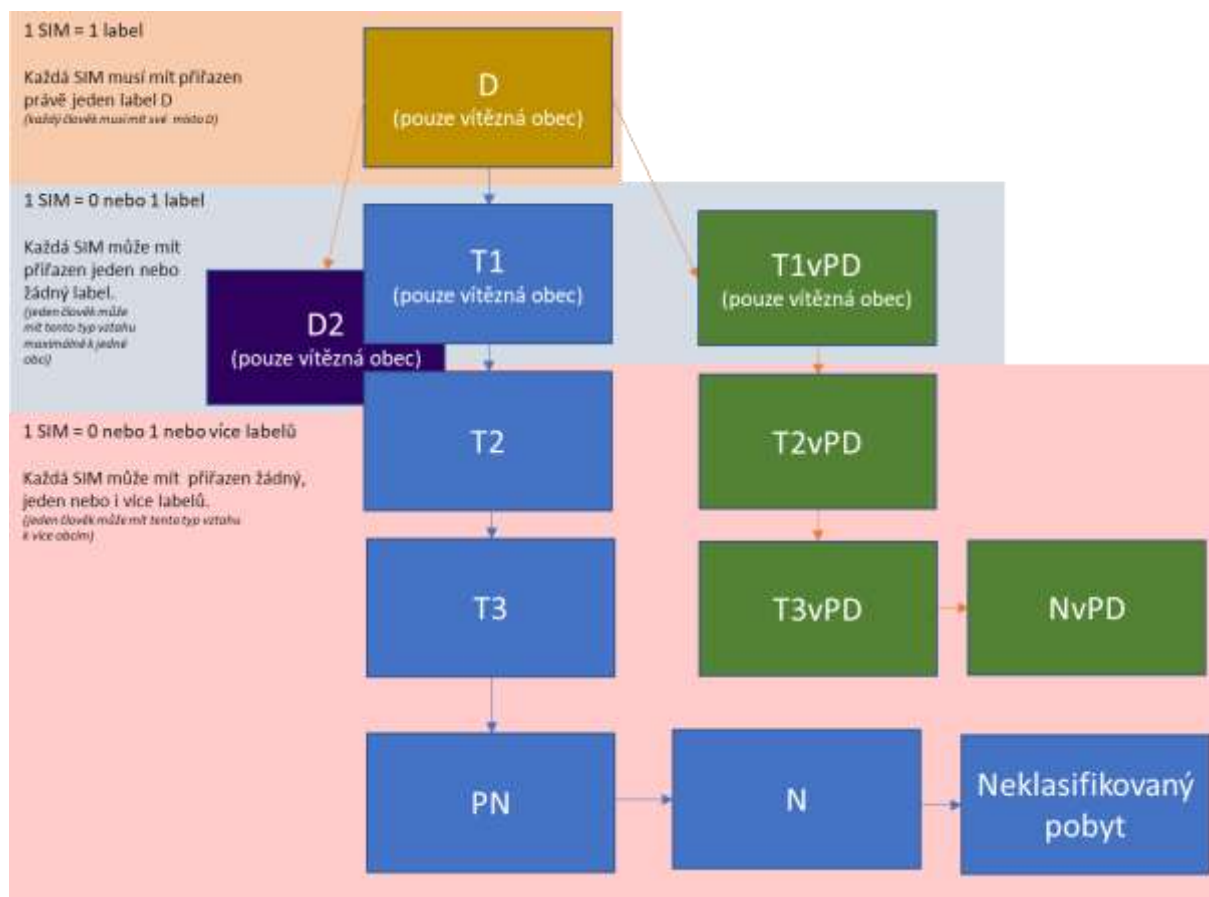
Při přidělování příznaků T1vPD, T2vPD, T3vPD, NvPD je postupováno analogicky dle stejných parametrů jako v případě T1, T2, T3, N s tím, že klasifikační časové okno je zúženo z celého dne na pracovní dobu (6:00 – 18:00)

Pro příznaky PN a D2 nelze z logiky věci přidělovat pro časové okno pracovní doby. V Tabulce č. 2 tak nebude příslušná kolonka vyplněna.

### 4.5.3 Hierarchie příznaků

Z výše uvedeného popisu jednotlivých příznaků je zřejmá neslučitelnost některých příznaků a způsob řešení, pokud SIM karta splňuje podmínky pro přidělení více příznaků. Jednotlivé příznaky mají mezi sebou hierarchické vztahy. Pokud tedy splňují podmínky pro udělení více příznaků, vždy je udělen ten hierarchicky vyšší. Schéma přidělování příznaků je uvedeno na obrázku níže.

Obrázek 5: rozhodovací schéma pro přidělování příznaků (labelů) a popis postupu.



Pro každého uživatele a každou obec, v níž má pobyt, zkontroluj, jestli platí podmínka. Když platí, přiřaď label, jinak pokračuj ve směru šipek.

V případě že uživatel má vůči obci „zlatý“ label D nemá již vůči stejné obci žádné jiné labely modré, zelené ani fialové barvy.

Pokud není label D přiřazen, bude přiřazen:

1 label modré barvy

1 nebo 0 labelů fialové barvy

1 nebo 0 labelů zelené barvy

Fialový label D2 (pokud bude přidělen) bude vždy současně s labelem T1 nebo T2.

V případě přiřazení labelu D, T1, T1vPD a D2 vyber po nalezení možných kandidátů pro každého uživatele tu obec, která kritérium labelu plní nejlépe (vyber vítěznou obec). Vyřazené kandidátské obce zařaď do dalšího přiřazování labelů (k těmto obcím může SIM karta mít jiné řádově nižší label).

Label D bude mít přiřazena každá SIM. Labely T1, T1 v PD a D2 mohou být přiřazeny maximálně vůči 1 obci. V případě ostatních labelů může mít 1 SIM karta label přiřazen vůči několika obcím.

## 5 Korekce reprezentativity redukované báze SIM karet

V rámci redukované báze SIM karet je nutno ošetřit tři fenomény, které by mohly vést ke strukturálním odlišnostem mobilityního chování redukované báze SIM karet od mobilityního chování celé populace. Jedná se o tyto tři jevy:

- V redukované bázi nejsou přítomny SIM karty virtuálních operátorů, což může vést k oblastním disparitám v důsledku výrazně odlišného podílu virtuálních operátorů v rámci oblastí.
- V redukované bázi jsou rozdílným způsobem zastoupeni uživatelů různých věkových skupin v důsledku rozdílného sklonu k užívání více než jedné SIM karty
- V redukované bázi nejsou zastoupeny osoby bez mobilního telefonu.

Zatímco o mobilitě osob, které nemají mobilní telefon, nelze na základě signalizačních dat říci nic („hlas“ jejich prostorového chování není v signalizačních datech „slyšet“), naopak „hlas“ prostorového chování osob, které mají dva a více telefonů, je slyšet více než by měl být. Analogicky, hlas uživatelů virtuálních operátorů není slyšet vůbec, protože jejich data nejsou přítomna v redukované bázi dat. Všechny tři výše uvedené jevy ovlivňující strukturální vlastnosti redukovaného souboru je třeba vhodným způsobem korigovat tak, aby byly případné deficity reprezentativity vzorců chování přítomných v signalizačních datech korigovány v rámci převodu zpracovatelných signalizačních dat přítomných v rámci redukované báze dat do výsledných datových sad o mobilitě osob v požadovaných výstupních formátech.

Vzhledem k tomu, že z principu věci nelze doplnit chybějící data o mobilitě pro osoby bez SIM karty ze samotných signalizačních dat, je tento strukturální deficit signalizačních dat řešen dopočtem počtu osob bez SIM karet v jednotlivých obcích pomocí koeficientu podílu obyvatel obce bez SIM karty stanoveného pro každou jednotlivou obec. Naopak síla hlasu prostorového chování osob, které mají dva a více telefonů, je aplikací koeficientu nadhodnocení síly hlasu potlačena. Korekce oblastních disparit v důsledku různého podílu virtuálních operátorů na celkové bázi uživatelů mobilních telefonů je realizována pomocí koeficientu korekce podílů mobilních operátorů.

### **Korekce počtu osob v obci pomocí koeficientu korekce podílů mobilních operátorů**

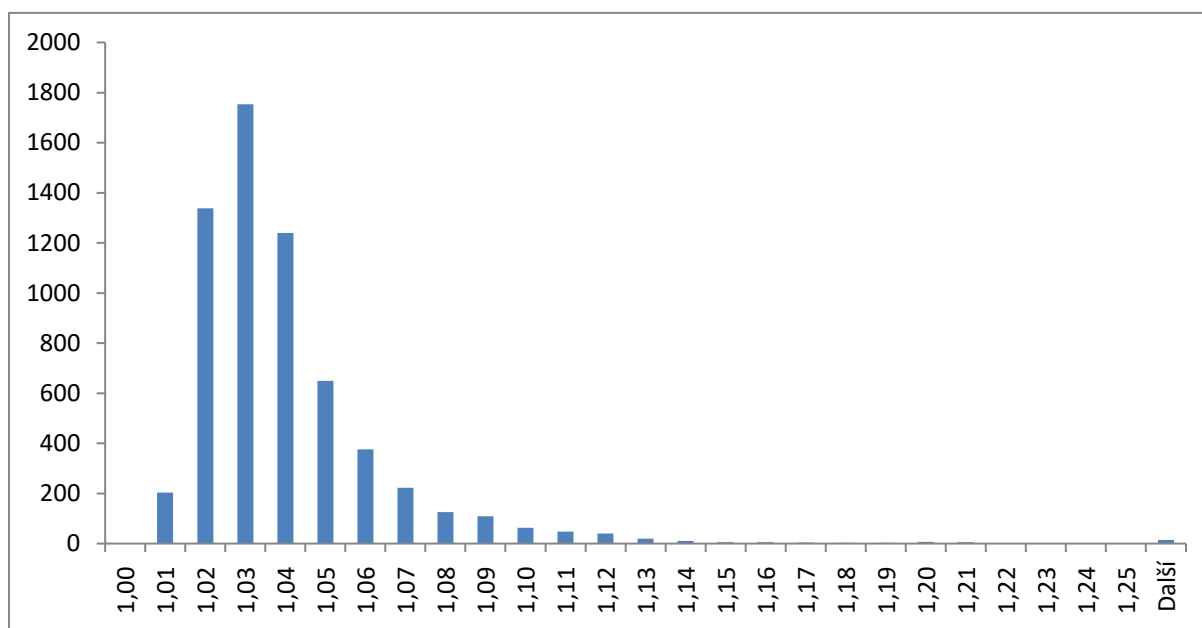
Do zpracování vstupují SIM karty tří dominantních mobilních operátorů, kteří mají tržní zastoupení 92,4 %.<sup>3</sup> Provedená analýza dat prokázala, že podíl virtuálních operátorů může lokálně výrazněji kolísat, a proto bylo přistoupeno k ošetření tohoto jevu korekčním koeficientem (z důvodu nemožnosti přímého využití dat virtuálních operátorů pro tento projekt).

---

<sup>3</sup> Český telekomunikační úřad [online]. Copyright © [cit. 07.12.2022]. Dostupné z: <https://www.ctu.cz/sites/default/files/obsah/stranky/424163/soubory/vyrocnizpravaceskehotelekomunikacnihouraduzarok2021.pdf>

Koeficient korekce podílu virtuálních operátorů je počítán na základě vyhodnocení podílu objemu dat generovaných virtuálními operátory na celkovém objemu dat v signalizaci. Tento podíl je stanoven pouze dle podílu celkového objemu základního souboru dat bez jakýchkoliv filtrací na všech datech v signalizaci bez zohlednění typu signalizačních transakcí. Koeficient korekce podílu virtuálních operátorů je počítán na základě analýzy vzorku signalizačních dat a vyjadřuje celkový podíl virtuálních operátorů v dané obci. Následně je aplikován na počty SIM karet redukované báze přiřazené jednotlivým obcím dle přidělených příznaků domova.

Obrázek 6: Distribuce koeficientu korekce podílu virtuálních operátorů



Tímto způsobem se podařilo do dat promítnout virtuální operátory, kteří společně s operátory vstupujícími do projektu tvoří 98 % celkového tržního podílu, jak je udáván ČTÚ.

### Podíl osob bez SIM karty

Některé skupiny obyvatel nemají žádnou SIM kartu – problém řešen aplikací koeficientu podílu osob bez SIM karty. V rámci populace ČR se dle průzkumu jedná o 12 % obyvatel bez SIM karty (koeficient podílu osob bez SIM karty za obyvatelstvo ČR je 12 %), tudíž 88 % obyvatel ČR má alespoň jednu SIM kartu. Na úrovni věkových skupin je koeficient podílu osob bez SIM karty relativně konstantní v širokém věkovém pásmu od 15 do 84 let – pohybuje se mezi 4 až 7 %. Pouze s velmi vysokým věkem nad 85 let se podíl osob bez SIM karty prudce zvyšuje až na 24 %. Velmi malé děti do 4 let nemají téměř žádné SIM karty, v kategorii do 5-9 je bez SIM karty 58 % osob, v kategorii 10-14 je bez SIM karty již jen 10 % osob.

### Koeficient nadhodnocení síly hlasu uživatelů mobilních telefonů pro vícenásobné SIM karty jimi držené

Některé skupiny obyvatel mají naopak dvě nebo i více hlasových SIM karet – problém řešen aplikací koeficientu nadhodnocení síly hlasu. Tento koeficient je dán podílem SIM karet na

počet osob, které tyto SIM karty mají v držení (počet osob disponujících alespoň jednou SIM kartou). V rámci populace ČR připadá na jednoho uživatele držícího alespoň jednu SIM kartu v průměru 1,18 SIM karty. Na úrovni celé populace ČR je tedy koeficient nadhodnocení síly hlasu 1,18. Nejvíce nadprůměrný počet SIM karet na osobu připadá osobám ve věku 35 až 59 let – minimálně 1,26 SIM karty na osobu. Vícenásobný výskyt SIM karet je však stále záležitost mnoha generací – lidí ve věku od 25 do 79 let – v těchto skupinách připadá minimálně 1,17 SIM karet na osobu. Užívání více SIM karet jednou osobou je velmi vzácné ve věkové skupině do 9 let - koeficient je blízký číslu 1. Rovněž na opačném konci věkového spektra ve věkové skupině od 80 let je výskyt více SIM karet relativně vzácný.

Vzhledem k tomu, že není možné přímo identifikovat duplicity mobilitního chování v rámci signalizačních dat spojené s existencí „zdvojeného“ případně i vícenásobného vlastnictví hlasových SIM karet jednou osobou a následně je odfiltrovat, je třeba tento fenomén vychylující reprezentativitu dat řešit strukturální korekturou mobilitních řetězců obsažených v balíku signalizačních dat redukované báze SIM. Tato strukturální korektura se provádí zeslabením váhy hlasu každé osoby přiřazené dle obce bydliště pomocí aplikace koeficientu nadhodnocení síly hlasu obyvatel obce držících alespoň jednu SIM kartu stanoveného pro každou jednotlivou obec.

### **Datové zdroje pro výpočty koeficientů korekce reprezentativity**

Pomocná datová sada koeficientů korekce reprezentativity je tvořena tabulkou dvou koeficientů – koeficientu podílu obyvatel obce bez SIM karty a koeficientu nadhodnocení síly hlasu obyvatel obce držících alespoň jednu SIM kartu.

Informace o specifickém chování věkových skupin české populace vychází z dat získaných v rámci robustního průzkumu trhu telekomunikačních služeb pro spotřebitelský trh na přelomu roku 2021 a 2022 metodou dotazníkového šetření provedeného v českých domácnostech. V rámci tohoto výzkumu byla řešena také problematika využití SIM karet – vlastnictví jedné nebo více hlasových SIM karet.

Dalším zdrojem dat pro stanovení obou koeficientů korekce reprezentativity na úrovni obcí je počet trvale bydlících obyvatel v obcích rozdělen do věkových skupin v kroku 5 let získaný od ČSÚ. Pro finální databáze jsou použita data popisující stav k 31.12.2021.

Zdrojem dat pro výpočet koeficientů korekce podílu virtuálních operátorů je časově omezený vzorek velikosti signalizačních dat v územním rozlišení datových sběrnic v mobilních sítích, na nichž je možné realizovat měření celkového objemu procházejících signalizačních dat.

### **5.1 Výpočet koeficientů korekce reprezentativity na úrovni obcí**

Koeficient korekce podílu obyvatel obce bez SIM karty a koeficient nadhodnocení síly hlasu obyvatel obce držících alespoň jednu SIM kartu jsou vypočteny pro každou obec následujícím způsobem:

Od ČSÚ byly převzaty údaje o počtu trvale bydlících v obcích v kroku 5 let. Jedná se o celkem 18 věkových kategorií: 0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, 85+.

Na základě průzkumu trhu bylo spočítáno 18 dílčích koeficientů podílů osob bez SIM karet a 18 koeficientů nadhodnocení síly hlasu v 18 věkových kohortách v rámci celé české populace aproximované reprezentativním vzorkem.

Pro každou obec byly koeficienty podílu osob bez SIM karet a koeficientů nadhodnocení síly hlasu obyvatel obce spočteny tímto způsobem:

Pro **každou věkovou skupinu** v obci byl dle počtu obyvatel obce ve věkové skupině vypočten pravděpodobný počet osob se SIM kartou v dané věkové skupině, tj. např. obec X měla 100 osob ve věkové skupině 70-74 let, při podílu osob bez SIM karty ve věkové kohortě - tj. koeficientu podílu osob bez SIM karty v této věkové skupině ve výši 6,92 % bylo vypočteno, že pravděpodobný počet osob s alespoň jednou SIM kartou v této věkové skupině je 93,08 osob a 6,92 je pravděpodobný počet osob v této věkové skupině bez SIM karty.

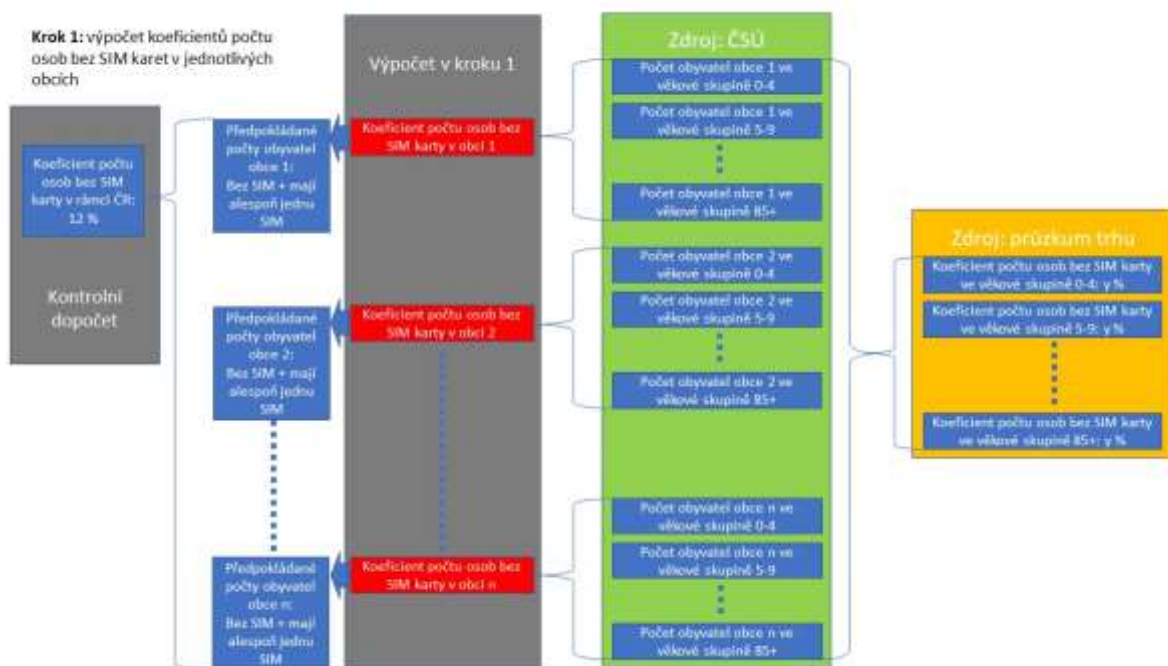
Následně byl pro **každou věkovou skupinu** v obci vypočten předpokládaný počet SIM karet v dané věkové skupině, tj. např. pro obec X čítající 100 osob, z toho 93,08 s minimálně jednou SIM kartou, byl dle koeficientu nadhodnocení síly hlasu ve věkové kohortě ve výši 1,201 pro tuto věkovou kategorii (tj. průměrného počtu SIM karet v dané věkové skupině připadajících na jednu osobu mající alespoň jednu SIM kartu) vypočten předpokládaný počet SIM karet v držení osob této věkové skupiny v této obci, tedy v uvedeném příkladu se jedná o  $93,08 * 1,201 = 111,80$  SIM karet.

Koeficient podílu obyvatel obce bez SIM karty pro obec X byl následně spočítán jako podíl předpokládaného počtu obyvatel obce bez SIM karet napříč všemi věkovými skupinami v obci X na celkové populaci obce X.

Koeficient nadhodnocení síly hlasu obyvatel obce držících alespoň jednu SIM kartu byl pro obec X následně spočítán jako podíl předpokládaného počtu SIM karet v držení všech obyvatel obce napříč všemi věkovými skupinami v obci X a předpokládaného počtu obyvatel obce užívajících minimálně jednu SIM kartu.

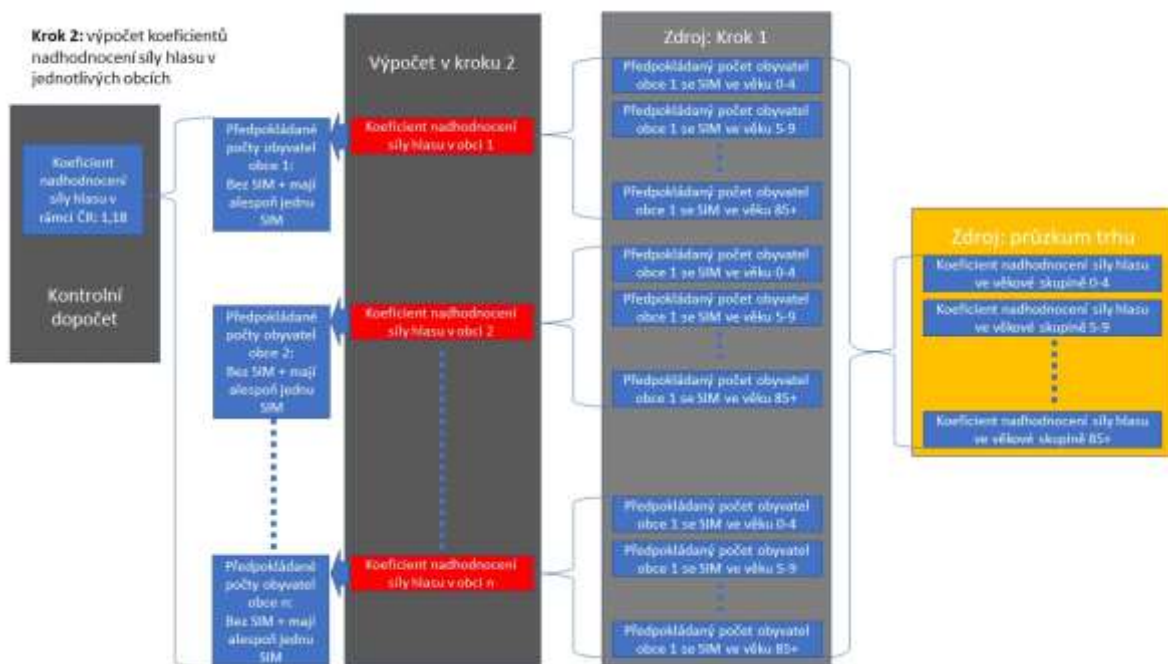
Proces výpočtu koeficientu podílu obyvatel bez SIM karty v obci a koeficientu nadhodnocení síly hlasu v obci je znázorněn na následující infografice v podobě dvou kroků. Prvním krokem je výpočet koeficientu podílu obyvatel bez SIM karty v obci.

Obrázek 7: Výpočet koeficientu osob bez SIM v obci



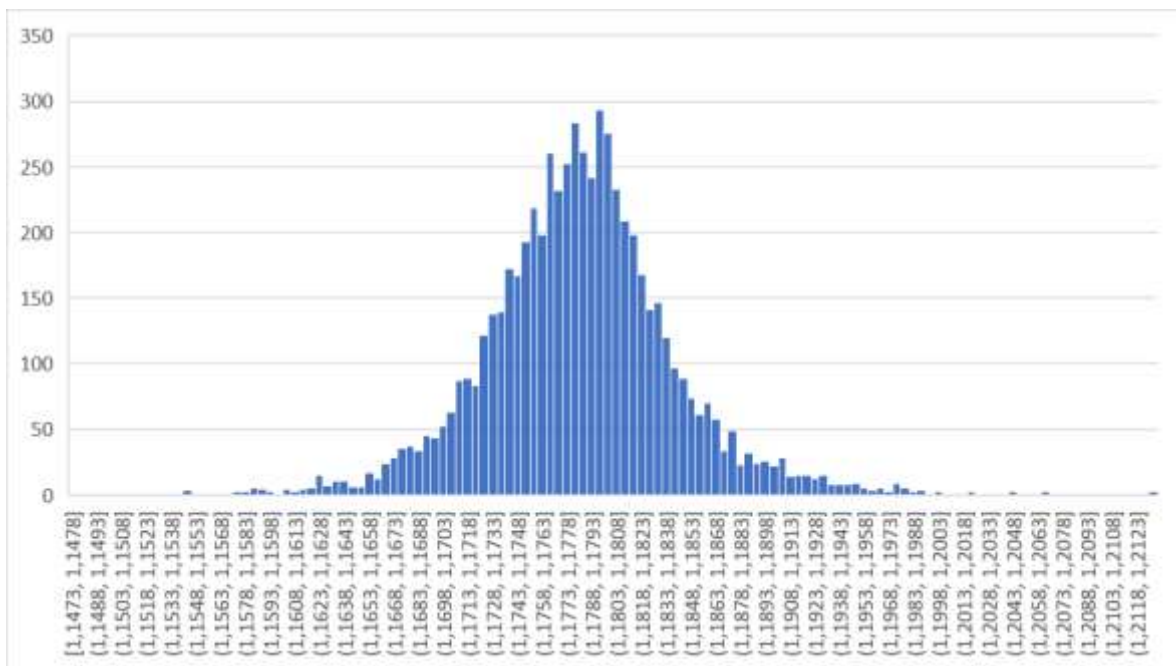
Druhým krokem je výpočet koeficientu nadhodnocení síly hlasu v obci.

Obrázek 8: Výpočet koeficientu nadhodnocení síly hlasu v obci



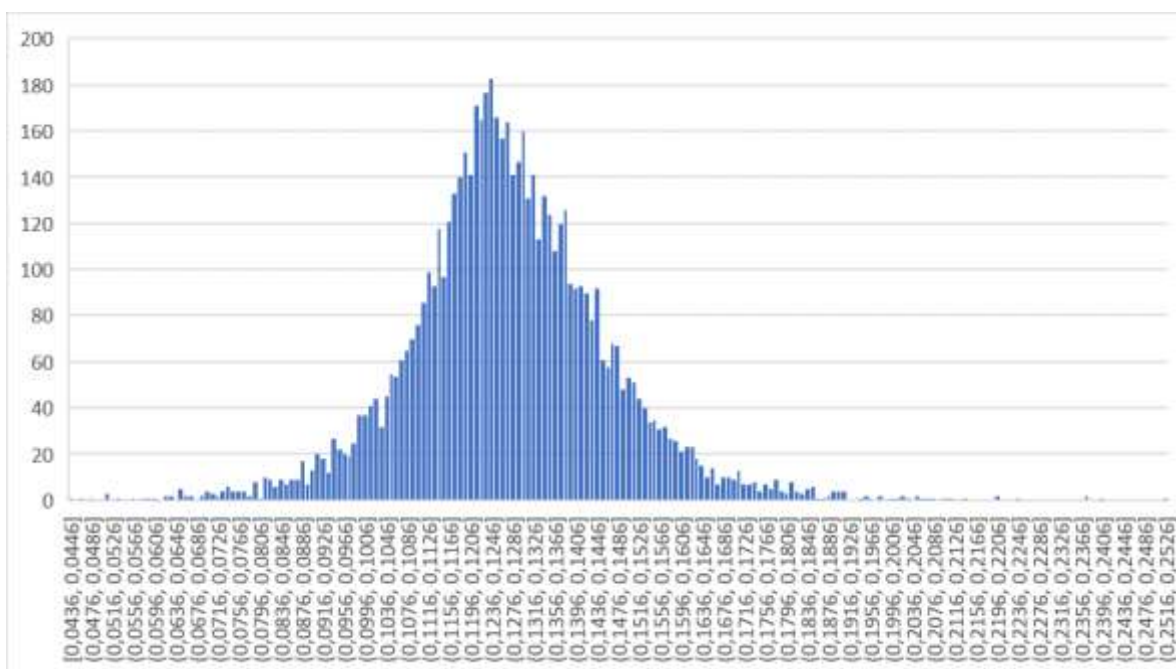
Koeficient nadhodnocení síly hlasu obyvatel obce držících alespoň jednu SIM kartu nabývá hodnot od 1,14 do 1,21 viz následující graf, z kterého je patrná jeho distribuce hodnot u obcí v rámci ČR.

Obrázek 9: Distribuce koeficientu nadhodnocení síly hlasu



Koeficient podílu obyvatel obce bez SIM karty nabývá hodnot od 4,3 % do 25,2 %. Distribuce je patrná z grafu níže.

Obrázek 10: Distribuce koeficientu podílu obyvatel obce bez SIM





## 5.2 Způsob aplikace koeficientů korekce reprezentativity

Jako první je do procesu strukturální korekce reprezentativity aplikován koeficient nadhodnocení síly hlasu obyvatel obce držící SIM karty, tedy koeficient, který je v rámci pomocné datové sady koeficientů vypočítán jako druhý v pořadí. Tento koeficient nadhodnocení síly hlasu je použit ke korekci (zeslabení) velikosti příspěvku („síly hlasu“) veškerých nalezených matic prostorových vazeb všech SIM karet přítomných v redukované bázi pro každou SIM kartu přítomnou v této bázi.

Koeficient nadhodnocení síly hlasu je aplikován následujícím způsobem pro každého uživatele zahrnutého v redukované bázi v podobě multidimenzionálních dat o mobilitě uživatelů mobilních telefonů:

U každého uživatele je prvnímu místu s příznakem „D“ označující místo Domova přiřazen koeficient nadhodnocení síly hlasu odpovídající obci, do níž byl uživatel přiřazen jako rezident dle příslušných kritérií na stanovení příznaku Domov a „základní (vstupní) hlas“ příznaku v podobě čísla 1 bude vydělen koeficientem nadhodnocení síly hlasu příslušnému této obci. Koeficient nadhodnocení síly hlasu je vždy z principu větší nebo roven 1. Čím je míra nadhodnocení síly hlasu v obci vyšší, tím slabší síla korigovaného hlasu je uživateli přidělena, tj. je-li pro přiřazenou obec rezidence spočítán koeficient nadhodnocení síly hlasu ve výši 1,2, má výsledná síla korigovaného hlasu tohoto uživatele přiřazenu hodnotu  $1/1,2 = 0,833$ .

U každého dalšího místa s příznakem přiděleného uživateli bude hodnota síly hlasu přidělená uživateli děděna z hodnoty přidělené dle místa s příznakem „D“ podle hodnoty korigovaného hlasu vypočteného pro obec rezidentství. Díky tomuto dědění hodnoty korigované síly hlasu místa s příznakem bude zajištěno propsání korekce síly hlasu do všech míst s příznakem přidělených každému uživateli (korekce síly hlasu dynamického chování uživatele).

Tímto způsobem bude korigováno celkové mobilní chování každého uživatele, a to jak ve vztahu k obci bydliště, tak ve vztahu ke všem ostatním jemu přiřazeným místům s příznakem. Tímto způsobem budou tedy korigovány veškeré vztahy mezi místy s příznakem každé jednotlivé SIM karty v redukované bázi. Díky této korekci bude redukováná báze SIM karet převedena na redukovanou bázi uživatelů SIM karet (uživatelů držících alespoň jednu SIM kartu).

Korekce podílu virtuálních operátorů pomocí koeficientů korekce je provedena obdobně jako u koeficientu nadhodnocení síly hlasu – zesílením hlasu každého uživatele dle koeficientu obce s příznakem D.

Druhý z koeficientů korekce reprezentativity – koeficient podílu obyvatel obce bez SIM karet – je aplikován pro každou obec, a to v podobě dopočtu hodnot nazvaných „počet obyvatel bez SIM“ ve sloupci H Tabulky č. 1. Dopčet hodnot ukazatele „počet obyvatel bez SIM“ je proveden následujícím způsobem:

- Na základě zjištění skutečného celkového počtu SIM karet v redukované bázi přiřazených k obci dle kritérií rezidence (opatřeny příznakem „D“) je pomocí koeficientu nadhodnocení síly hlasu pro danou obec zjištěn přepočtený počet osob přiřazených obci jako rezidenti mající alespoň jednu SIM kartu.
- Pomocí koeficientu podílu obyvatel bez SIM karty v obci je stanoven dopočet ukazatele „počet obyvatel v obci“ jakožto komplementu počtu rezidentů majících alespoň jednu SIM kartu zjištěných z redukované báze v rámci předchozího kroku.

V rámci tohoto postupu je počet obyvatel bez SIM vázán přes zafixované procento předpokládaného počtu uživatelů bez SIM na skutečný počet uživatelů majících alespoň jednu SIM kartu odvozený z redukované báze signalizačních dat. Smyslem tohoto dynamického fixování počtu osob bez SIM na počet osob s alespoň jednou SIM kartou, je udržení proporce mezi kumulovanou silou hlasu osob držících SIM kartu a silou hlasu (respektive „silou mlčení“) dopočtených obyvatel obce bez SIM karty.

V rámci PoC v důsledku použití koeficientů došlo k přepočtu redukované báze čítající 10,3 mil. SIM karet na soubor o velikosti 8,7 mil. jedinců (ošetření nadhodnocení síly hlasu uživatelů mobilních telefonů v důsledku pro vícenásobné SIM karty jimi držené) a na druhé straně ke stanovení souboru kategorie „bez SIM“ o velikosti 1,4 mil. Ve výsledku tedy do dalšího kroku projekce na populaci vstupoval soubor o velikosti 10,1 mil předpokládaných osob odvozených ze SIM karet po zohlednění koeficientů korekce reprezentativity.

## 6 Projekce na populaci

Výsledné databáze musí vypovídat o prostorových vzorcích chování obyvatel České republiky. Výsledky analyzovaného vzorku (vypovídající o SIM kartách) jsou po výše popsáných korekcích koeficienty nadhodnocení síly hlasu a podílu obyvatel bez SIM proto přepočteny na celkovou populaci.

Finální přepočet na populaci je realizován následujícím způsobem. V případě Tabulky č. 1 je z počtu rezidentů za všechny obce (před provedením anonymizace a po korekci síly hlasu aplikací koeficientů nadhodnocení síly hlasu a dopočtu počtu obyvatel bez SIM karty) a cílové hodnoty populace v ČR stanoven koeficient přepočtu na populaci ( $P_k$ ) dle následujícího vzorce:

$$P_k = \text{pop\_ČSÚ} / \sum \text{počtu rezidentů v obcích}$$

Do sumy počtu rezidentů v obcích jsou započítáni i rezidenti bez SIM. Tento koeficient přepočtu na populaci je pak aplikován na všechny ukazatele Tabulek č. 1 a č. 2. V případě finální databáze tento koeficient dosahuje hodnoty 1,02 (po zaokrouhlení; *jako podíl 10,5 mil obyvatel dle ČSÚ a 10,3 mil rezidentů*).

V případě Tabulky č. 3 je přepočet na populaci počítán pro každou hodinu zvlášť, dle následujícího vzorce:

$$P_{k_h} = (\text{pop\_ČSÚ} - \sum \text{rezidenti bez SIM}) / \sum \text{počet přítomných}_h$$

*Pozn. pro účely stanovení koeficientu  $P_{k_h}$  do „počet\_přítomných<sub>h</sub>“ vstupuje i hodnota v kategorii „na cestě“ v dané hodině, zatímco SIM v neurčeném stavu nikoliv (neurčený stav reprezentuje SIM, o kterých z pohledu signalizačních dat nevíme v daném časovém úseku nic). V důsledku je tak množina „neurčeno“ uniformě „rozpuštěna“ do všech ukazatelů v dané hodině.*

Pro finální databáze byla použita dostupná data ČSÚ o „Počtu obyvatel v obcích České republiky, tzn. k 1. 1. 2022“<sup>4</sup>, a o počtu trvale bydlících dle věkových skupin k 31.12.2021. K těmto datům se tak vztahují všechny údaje a grafy uvedené v tomto dokumentu.

V důsledku aplikace koeficientů korekce reprezentativity, projekce na populaci a anonymizace dochází k tomu, že ve finálních databázích se vyskytují hodnoty od 7,8 výše. Je to způsobeno tím, že při zpracování signalizačních dat se stále počítá s „celými“ SIM kartami. V datech před anonymizací (tzn. po aplikaci koeficientů korekce reprezentativity) se tak ve spojených datech po zaokrouhlení vyskytuje nejnižší hodnota 0,9 a její násobky. U násobků postupně roste jejich rozptyl, což je dáno různými přepočtovými koeficienty pro konkrétní obce. Přesto jsou i ve finálních datech patrné skoky v hodnotách (tzn. některá desetinná čísla se v souboru nevyskytují). Nejedná se o chybu, ale o vlastnost dat vyplývající ze skutečnosti, že při zpracování dat nedochází k dělení 1 SIM karty do více obcí.

<sup>4</sup> Dostupné z <https://www.czso.cz/csu/czso/pocet-obyvatel-v-obcich-k-112021>

## 7 Sledované jevy

### 7.1 Tabulka č. 1 - Data o počtu obvykle bydlícího obyvatelstva

Jedná se o tabulku obsahující statistické údaje o předpokládaném počtu rezidentů každé obce Česka. Obsah databáze je prezentován v Tabulce č. 1.

Sloupec C – „Počet rezidentů (počet obyvatel)“ - počet osob s příznakem rezident (D) v této obci. Tj. počet osob, pro které je obec definovaná ve sloupci A rezidenční obcí. Na hodnoty bylo z důvodu anonymizace aplikováno zarušení, které je však vzhledem k hodnotě v % (v jednotkách) i absolutním vyjádření ve vztahu k celkové hodnotě nevýznamné.

Sloupec D – „Nevyjíždějící za službami ani za prací/školou“ - nevyjíždějící rezidenti. Jedná se o počet rezidentů obce, kteří nemají vůči jiné obci přidělen žádný příznak dojížděky (nemá T1, T2, ani T3).

Sloupec E – „Občasně vyjíždějící za službami“ - občasně vyjíždějící za službami. Jedná se o počet rezidentů obce, kteří mají vůči jiné obci příznak dojížděky 3. typu (T3), ale nemají vůči žádné obci příznaky 2. či 1. typu (T1 ani T2). Jinými slovy je pro tyto SIM karty, jejichž nejvyšší přidělený příznak je T3.

Sloupec F – „Intenzivně vyjíždějící za službami“ - jedná se o počet rezidentů obce, kteří mají vůči jiné obci příznak dojížděky 2. typu T2, ale nemají vůči žádné obci příznak 1. typu (T1). Jinými slovy je pro tyto SIM karty, jejichž nejvyšší přidělený příznak je T2.

Sloupec G – „Vyjíždějící za prací/školou“ - intenzivně vyjíždějící za službami. Jedná se o počet rezidentů obce, kteří mají vůči jiné obci příznak dojížděky 1. typu (T1). Číselný údaj představuje pro danou obec součet všech neanonymizovaných vztahů s typem cesty 1 uvedených v Tabulce č. 2. Jinak řečeno, nejsou v něm započítány relace s hodnotou „xx“ v Tabulce č. 2.

Sloupec H – „Bez SIM“ - počet obyvatel bez SIM. Jedná se o počet lidí bez mobilních telefonů stanovený na základě demografických ukazatelů o počtu rezidentů obce dle ČSÚ, o míře využívání mobilních telefonů ve věkových skupinách dle průzkumu a nalezeného počtu rezidentů obce dle báze signalizačních dat (tj. samotná signalizační data jsou jedním ze zdrojů dat pro výpočet tohoto čísla, ale nikoliv zdrojem jediným a přímo aplikovaným). Přesná konstrukce tohoto parametru je popsána v části tohoto dokumentu nazvané „Korekce reprezentativity redukované báze SIM karet.“

Sloupec I – „Druhé bydlení“ (D2). Jedná se o počet rezidentů jiných obcí, pro které je obec definovaná sloupcem A obcí druhého bydlení.

Sloupec J – „Nocující návštěvníci“ - jedná se o počet rezidentů jiných obcí, kteří vůči obci definované sloupcem A naplňují parametry pro označení za přenocujícího návštěvníka (PN).

Sloupec K – „Návštěvníci“ - jedná se o počet rezidentů jiných obcí, kteří vůči obci definované sloupcem A naplňují parametry pro označení návštěvníka (N).

Sloupec L – „Počet T1 anon.“ - počet anonymizovaných relací T1 (tzn. párů zdrojová obec – cílová obec) z Tabulky č. 2, které byly anonymizovány.

Sloupec C v datech před krokem anonymizace odpovídá součtu hodnot ve sloupcích D-H. Ve výsledných tabulkách se může mírně lišit v důsledku zaokrouhlení dílčích hodnot či v důsledku anonymizace (včetně aplikace zarušení hodnot, viz kapitola „Anonymizace dat“).

Tabulka č. 1: Podoba výsledné databáze údajů o počtu rezidentů

Data o obvyklé bydlicím obyvatelstvu - počty rezidentů											
Obec	Kód obce	Počet rezidentů (počet obyvatel)	Z toho					Druhé bydlení	Nocující návštěvníci	Návštěvníci	Počet T1 anon.
			Nevyjíždějící za službami a za prací/školou	Občasně vyjíždějící za službami	Intenzivně vyjíždějící za službami	Vyjíždějící za prací/školou	Bez SIM				
Adamov (Blansko)	581291	4180,0	710,0	460,0	980,0	1450,0	580,0	120,0	430,0	1670,0	90,0
Archlebov (Hodonín)	586030	750,0	110,0	100,0	190,0	240,0	110,0	70,0	320,0	480,0	40,0
Arnešovice (Pelhřimov)	509388	60,0	10,0	10,0	10,0	20,0	10,0	xx	40,0	40,0	20,0
Arnolec (Jihlava)	586854	160,0	10,0	30,0	50,0	30,0	40,0	10,0	50,0	100,0	40,0
Babice (Třebíč)	590274	130,0	20,0	20,0	40,0	30,0	20,0	20,0	80,0	110,0	20,0

Ve sloupcích C, D, E, F, G, I, J a K, ve kterých se vyskytují symboly značící nezveřejnitelné hodnoty, je na konci tabulky uveden součet osob, které se pod těmito hodnotami skrývají. Součet bude zobrazen u všech sloupců, kde je 7 a více anonymizovaných hodnot ve sloupci.

### 7.1.1 Porovnání výstupů s konvenčními zdroji

V rámci projektu byl ukazatel počtu rezidentů/obyvatel v obcích stanovován z přítomnosti SIM karet v nočním období po dobu 4 týdnů zvoleného měřeného období.<sup>5</sup> Při interpretaci dat je nutné respektovat odlišnost metody zpracování signalizačních dat a metody, kterou používá ČSÚ. I když je označení výstupních ukazatelů (tzn. počet obyvatel/rezidentů) v obou případech podobné (popisují stejný fenomén), jejich hodnoty se mohou významně lišit. Níže je pro přehlednost v tabulce shrnuta základní odlišnost vstupů pro stanovení počtu obyvatel dle projektu a dle ČSÚ.

Rezidenti/obyvatelé dle definice projektu	Obyvatelé dle ČSÚ
<ul style="list-style-type: none"> <li>- Vstupní báze jsou data z aktivních/zapnutých mobilních telefonů</li> <li>- Počet obyvatel zde vychází z předpokladu, že člověk obvykle tráví noc doma</li> </ul>	<ul style="list-style-type: none"> <li>- Primární vstupní báze jsou obdržené odpovědi, které lidé uvedli při celostátním sčítání lidí, domů a bytů.</li> <li>- výsledný počet obyvatel je výstup zpracování dat ze sčítání</li> </ul>

<sup>5</sup> Metoda a proces zpracování signalizačních dat jsou blíže popsány v předcházejících kapitolách

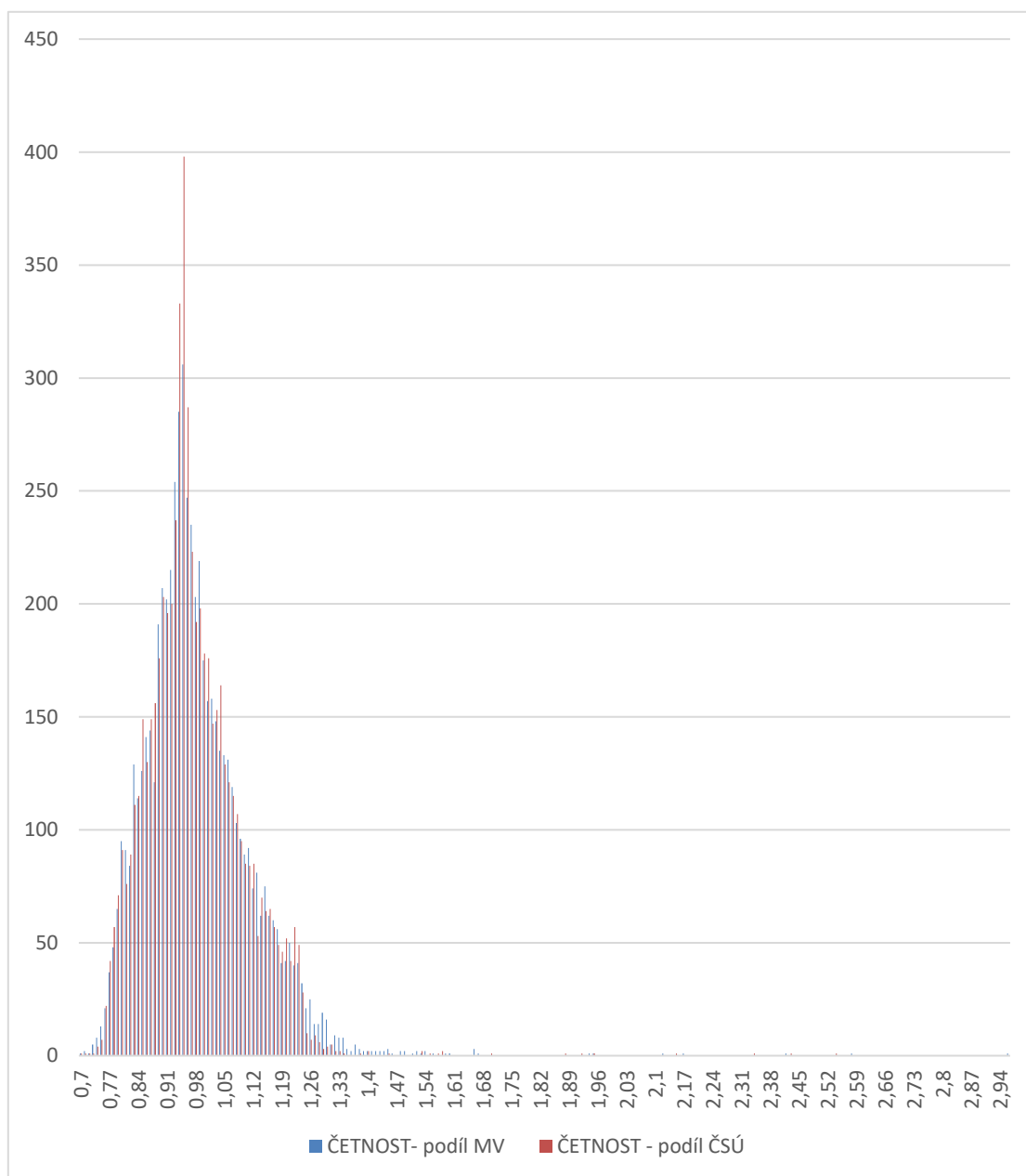
<ul style="list-style-type: none"> <li>- Příznak rezidentství je přiřazen SIM kartám, které splnily kritéria aktivity a ve zvoleném měřeném období (28 dní) byly v noci detekovány majoritně na území dané obce</li> <li>- Na počet SIM karet jsou následně aplikovány přepočtové koeficienty na populaci</li> </ul>	<p>obyvatelstva, kde dotazování odpovídají na otázku: „Kde bydlíte? Uveďte místo svého obvyklého bydliště, tj. kde skutečně bydlíte nebo dlouhodobě žijete“.</p> <ul style="list-style-type: none"> <li>- Zjišťován je stav ke konkrétnímu datu (v SLDB 2021 to byl 26.3.2021)</li> </ul>
--	---

Snahou projektu bylo nastavit metodiku a parametry ukazatelů projektu tak, aby ukazatel počtu residentů reprezentoval stejný jev jako počet obyvatel dle ČSÚ. Nominálně by se tak hodnoty obou ukazatelů měly blížit. Vycházíme tedy z předpokladu, že lidé obvykle bydlí tam, kde tráví čas v noci, tzn. kde přespávají. Přesto je nutné při porovnávání respektovat diametrálně odlišné způsoby, kterými byl počet obyvatel v obcích zjišťován (přítomnost SIM vs. dotazníkové šetření).

V souladu s požadavky na projekt byly výstupy počtů obyvatel porovnávány s konvenčními zdroji dat o obyvatelstvu, primárně se jedná o Počet obyvatel v obcích České republiky dle ČSÚ, sekundárně pak o evidenci obyvatel dle Ministerstva vnitra. Odlišnost metodik sběru dat se odráží ve vzájemném porovnávání databází. Různé metodiky se projevují mj. i při vzájemném porovnání databází ČSÚ a MV (mezi sebou), kdy se ke stejnému datu počty residentů v obci liší i o 35 %.

Rozdělení obcí dle jejich odchylky od ČSÚ, respektive evidence MV dle ukazatele počtu obyvatel jsou znázorněny na následujícím grafu. Na ose y je vyneseno počet obcí, osa x pak reprezentuje podíl hodnoty počtu residentů ze signalizačních dat oproti hodnotám ČSÚ, respektive hodnotám evidence MV. Z grafu je patrná podobná distribuce odchylek vůči oběma konvenčním datasetům.

Obrázek 11: Míra souladu s daty ČSÚ, MV

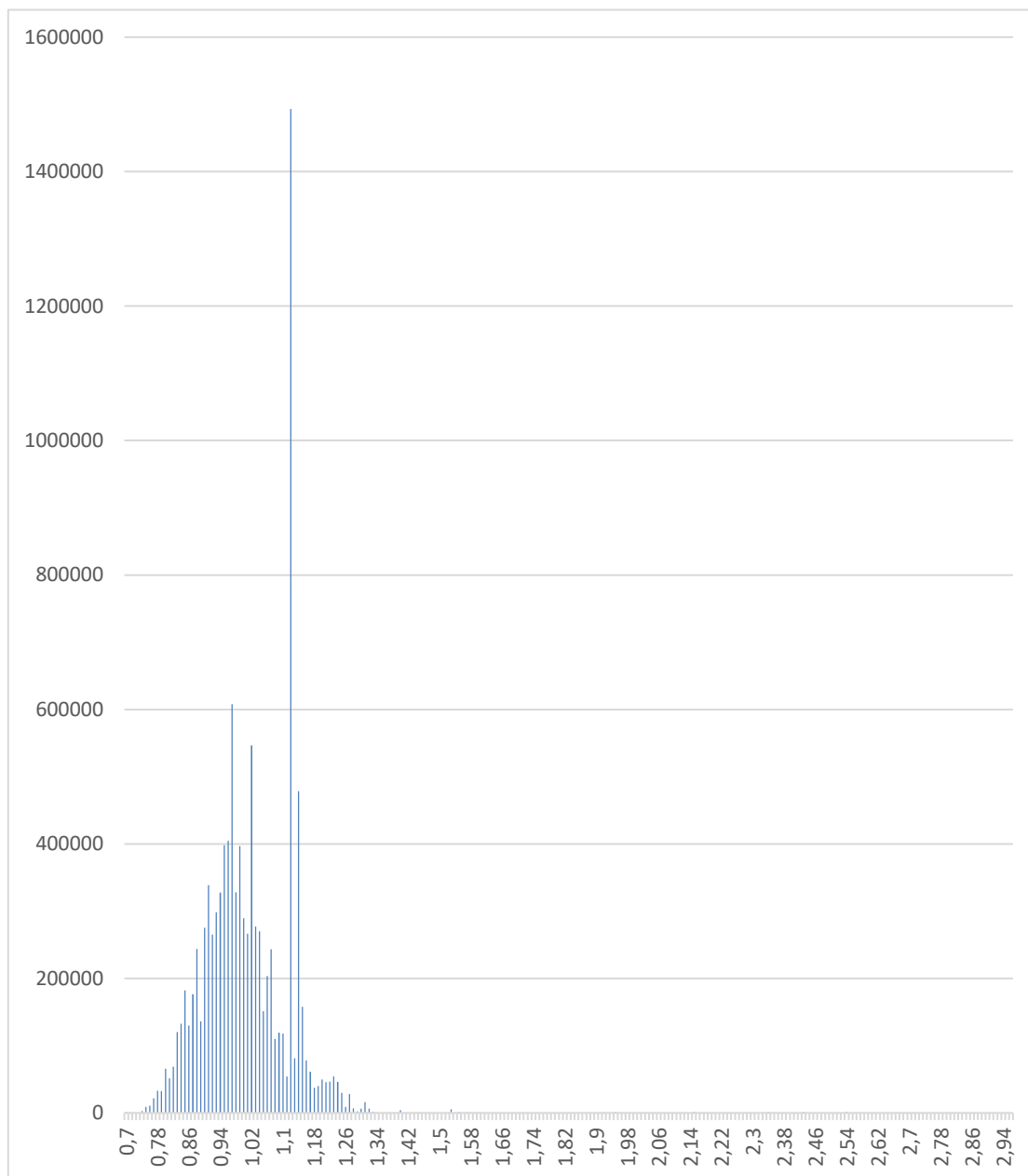


V pásmu odchylky větší než 25 % od ČSÚ se nachází 73 obcí, což představuje 1,2 % z celkového počtu 6254 obcí.

Následující graf navazuje na graf míry souladu s daty ČSÚ a evidence ministerstva vnitra. Osa y reprezentuje počet obyvatel. Na ose x jsou vyneseny podíly počtu rezidentů (stanovených ze signalizačních dat) oproti počtu obyvatel dle ČSÚ. Graf tak reprezentuje rozložení obyvatelstva dle jeho rezidentury v obcích, které jsou řazeny dle jejich odchylky od ČSÚ. V intervalu odchylky  $\pm 25\%$  (na ose x 0,75 – 1,25) se nachází

98,9 % populace ČR. Tzn. že pouze 1,1 % obyvatel ČR je alokováno v obcích, jejichž počet rezidentů se oproti hodnotě počtu obyvatel dle ČSÚ liší o více, než 25 %.

Obrázek 12: Obyvatelstvo v obcích dle odchylek od ČSÚ



Při porovnání oproti ČSÚ je nutné brát v potaz jak procentuální odchylku obou hodnot, tak absolutní hodnoty, ke kterým je z povahy projektu ve smyslu zdrojových dat a požadované granularity nutné přihlížet. V menších obcích, může i nominálně malá odchylka počtu osob představovat vyšší relativní odchylku (jelikož je vztažena k celkovému počtu obyvatel obce). Odchylky mohou být vedle odlišné metodiky sčítání



(signalizačních) dat oproti konvenčním zdrojům způsobeny jevy, které vyplývají z možností zpracování signalizačních dat, technicko-technologických omezení a zvolené metodě výpočtu při požadované časové a územní granularitě.

V tabulce níže je uveden přehled porovnání pro krajská města. Při porovnávání vůči ČSÚ datové výstupy projektu odpovídají odbornému očekávání, které je možno si vytvořit na základě znalosti trendů regionálních ekonomik, trendů trhu nemovitostí, trhů práce, meziregionální migrace a dalších faktorů ovlivňujícího rozložení obyvatelstva v krajských městech mezi sebou a ve vztahu k jejich spádovým oblastem. Z tabulky je patrné několik fenoménů. Předně města Praha a Brno jsou před ostatními krajskými městy v naskoku "podregistrace", tedy že v těchto městech ve skutečnosti "bydlí" (ve smyslu definice *Domova* v rámci projektu) více obyvatel, než uvádí ČSÚ. Je nutno mít na paměti, že kritérium definice *Domova* připouští jako "žijící ve městě" i studenty vysokých škol a studenty středních škol na internátech. Např. k výrazně vyššímu podílu ve městě Olomouc přispívá kombinace použité metodiky a zvoleného časového období – do bydlících jsou započítáni i studenti VŠ. Naproti tomu např. situace v Ostravě odpovídá neutrálnímu stavu (podíl ve Finální databázi 4 je 0,97), neboť město čelí desetiletí trvající restrukturalizaci ekonomiky v kombinaci s refokusem ekonomiky města díky jeho velikosti na významné středisko služeb nadregionálního významu. Další krajská města se pak dle podílu nacházejí na obdobné úrovni mezi neutrálním stavem a hodnotami center Prahy a Brna (výjimku tvoří specifické Ústí nad Labem).

Tab – Porovnání s ČSÚ – krajská města

Obec	ID	Obyvatelstvo ČSÚ	Obyvatelstvo data FD4	Podíl FD1	Podíl FD2	Podíl FD3	Podíl FD4
<b>Praha</b>	554782	1 275 406	1 422 724	1,14	1,15	1,03	<b>1,12</b>
<b>Brno</b>	582786	379 466	431 043	1,16	1,14	1,06	<b>1,14</b>
<b>Ostrava</b>	554821	279 791	271 217	1,00	0,98	0,95	<b>0,97</b>
<b>Plzeň</b>	554791	168 733	172 643	1,08	1,07	0,99	<b>1,02</b>
<b>Liberec</b>	563889	102 951	105 100	1,02	1,04	0,98	<b>1,02</b>
<b>Olomouc</b>	500496	99 496	114 892	1,20	1,17	1,11	<b>1,15</b>
<b>České Budějovice</b>	544256	93 426	95 677	1,06	1,05	0,96	<b>1,02</b>
<b>Hradec Králové</b>	569810	90 596	93 636	1,04	1,04	0,97	<b>1,03</b>
<b>Ústí nad Labem</b>	554804	90 378	82 086	0,91	0,91	0,85	<b>0,91</b>
<b>Pardubice</b>	555134	88 520	91 914	1,05	1,06	0,98	<b>1,04</b>
<b>Zlín</b>	585068	72 973	77 717	1,07	1,06	1,03	<b>1,07</b>
<b>Jihlava</b>	586846	50 108	52 880	1,07	1,07	1,01	<b>1,06</b>
<b>Karlovy vary</b>	554961	45 500	45 159	0,99	1,01	0,99	<b>0,99</b>

Naproti tomu největší extrém v souboru dle % odchylky od ČSÚ mají následující obce, u kterých byly identifikovány pravděpodobné příčiny a faktory přispívající k vysoké míře odchylky od metriky ČSÚ.

Tab – Porovnání s ČSÚ – odchýlené obce

Kód	Obec	Počet obyvatel dle ČSÚ	Stanovený počet obyvatel	Podíl	Významný podíl areálů s nepřetržitým provozem	Rekreační oblast, lázně	Širší nadhodnocená oblast	Širší podhodnocená oblast	Příhraničí	Nízká absolutní odchylka
538141	Čestlice	684	1736	2,54	x		x			
579581	Pec pod Sněžkou	599	1456	2,43		x				
598267	Modletice	528	1234	2,34	x		x			
579742	Špindlerův Mlýn	979	2103	2,15		x				
533572	Ovčáry	849	1655	1,95	x					
539171	Dobrovíz	572	1100	1,92	x		x			
579505	Malá Úpa	144	271	1,88		x				
597473	Karlova Studánka	175	297	1,70		x				
545597	Lipno nad Vltavou	639	1013	1,58		x				
550337	Kvilda	125	197	1,58		x				
579351	Janské Lázně	655	1026	1,57		x				
539660	Slapy	854	1323	1,55		x				
542148	Modrava	99	152	1,53		x				
538311	Klečany	3601	5501	1,53	x		x			
563536	Bedřichov	345	500	1,45		x				
538264	Jenštejn	1360	1904	1,40			x			
576425	Kvasiny	1534	2141	1,40	x					
506486	Boží Dar	227	314	1,38		x				
530964	Bludov	24	32	1,34						x
571598	Zájezd	124	166	1,33			x			
539317	Jeneč	1299	1722	1,33	x		x			
539236	Horoměřice	4511	5942	1,32			x			
530522	Rabyně	268	353	1,32		x	x			
531723	Rudná	5090	6686	1,31			x			
529974	Krňany	480	630	1,31		x	x			
540048	Buš	349	457	1,31		x	x			
539309	Chýně	4309	5637	1,31			x			
539732	Štěchovice	2116	2766	1,31		x	x			
538981	Velké Popovice	3110	4056	1,30			x			
539562	Pohoří	382	498	1,30			x			
532711	Pavlov	191	249	1,30			x			

Kód	Obec	Počet obyvatel dle ČSÚ	Stanovený počet obyvatel	Podíl	Významný podíl areálů s nepřetržitým provozem	Rekreační oblast, lázně	Širší nadhodnocená oblast	Širší podhodnocená oblast	Příhraničí	Nízká absolutní odchylka
531146	Drahelčice	1270	1651	1,30			x			
539511	Okrouhlo	746	964	1,29			x			
539546	Petrov	760	981	1,29		x	x			
598283	Sedlec	429	553	1,29			x			
538850	Svojetice	1219	1566	1,28			x			
539457	Líšnice	787	1009	1,28			x			
539686	Statenice	1547	1981	1,28			x			
539015	Vodochody	724	927	1,28			x			
539261	Hvozdnice	557	713	1,28			x			
533254	Černé Voděrady	340	435	1,28			x			
538329	Klíčany	558	711	1,27			x			
530263	Nespeky	765	972	1,27			x			
513288	Mrzky	170	216	1,27			x			
539252	Hradištko	2306	2927	1,27		x	x			
571326	Lichoceves	396	503	1,27			x			
530921	Vranov	436	553	1,27			x			
539244	Hostivice	8777	11131	1,27			x			
540897	Nové Dvory	269	341	1,27		x	x			
539627	Roztoky	8611	10905	1,27			x			
538612	Petřikov	604	764	1,26			x			
529745	Hvězdonice	330	417	1,26			x			
513539	Velká Lečice	185	234	1,26			x			
538604	Panenské Břežany	621	783	1,26			x			
539333	Jílové u Prahy	4919	6193	1,26			x			
533548	Nučice	366	460	1,26			x			
533874	Vlkančice	210	264	1,26			x			
532991	Úhonice	1124	1411	1,25			x			
539163	Davle	1765	2213	1,25		x	x			
599719	Tehovec	640	802	1,25			x			
533718	Stříbrná Skalice	1494	1870	1,25			x			
532614	Vodslivý	100	125	1,25			x			
536148	Javorník	2607	1955	0,75				x	x	
544370	Lidečko	1810	1356	0,75				x	x	
551864	Dívčí Hrad	298	222	0,75				x	x	
564036	Habartice	473	352	0,74				x	x	
597201	Bohušov	379	281	0,74				x	x	

Kód	Obec	Počet obyvatel dle ČSÚ	Stanovený počet obyvatel	Podíl	Významný podíl areálů s nepřetržitým provozem	Rekreační oblast, lázně	Širší nadhodnocená oblast	Širší podhodnocená oblast	Příhraničí	Nízká absolutní odchylka
512176	Hrádek	1912	1409	0,74				x	x	
597716	Osoblaha	1064	782	0,74				x	x	
551872	Hlinka	203	148	0,73				x	x	
598232	Hrčava	243	175	0,72				x	x	
511935	Bukovec	1367	971	0,71				x	x	
512028	Písek	1877	1323	0,70				x	x	

Z povahy signalizačních dat a metody výpočtu vyplývají následující omezení, respektive situace, ke kterým může lokálně u obcí docházet buď jednotlivě, nebo v kombinacích.

- *Selhání klasifikace příznaků – přidělení příznaků je definováno časoprostorovými podmínkami, které mohou splňovat i osoby, pro které nebyl příznak (nejen) rezident primárně navržen*
  - **Významný podíl areálů s nepřetržitým provozem** – obecně se jedná o obce s výrazným podílem nerezidenční funkce území v obci. Jedná se o situace vymykající se typickému režimu, kdy lidé tráví noc v místě bydliště. Jedná se např. směnný provoz apod.
  - **Rekreační oblast, lázně, tábory** (vliv cestovního ruchu) – s odkazem na definici domova/rezidentů mohou být tímto příznakem označeny také osoby, které v obci tráví např. dlouhodobější dovolenou. Tento jev může nabývat lokálně na významu v rekreačních lokalitách.
  - Změna povahy časoprostorového rytmu práce v důsledku COVIDu – vzhledem k nestandardní situaci, kterou vyvolala epidemie COVIDu v populaci během měřených období, může docházet ke změnám oproti očekávanému chování. Typickým příkladem může být „remote“ práce z rekreačních objektů apod. Tyto jevy se pak odlišně projevují ve statistikách ČSÚ a měřených obdobích.

- *Odchylky vyplývající z modelu územního rozpočtu*
  - o **Širší nadhodnocená/podhodnocená oblast** – při interpretaci dat je vzhledem k detailu územní granularity třeba přihlídnout k situaci v okolních obcích, tedy neposuzovat obec izolovaně. Může docházet k případům (např. v rekreačních oblastech), kde se širší oblasti chovají při porovnání oproti ČSÚ homogenně. V ojedinělých případech se pak může objevovat sousedství významně nadhodnocené a podhodnocené obce. Tyto jevy mohou vznikat u obcí, kde přiřazování rezidentů do obcí včetně jejich relokace dosáhne definovaných omezení během procesu zpracování dat. Tato omezení (podmínky) byla nastavena systémově, homogenně pro celé území a vyplývají zejména z map pokrytí, modelu přepočtu a konzervativně nastaveného tzv. relokačního mechanismu.
  - o **Příhraničí** (vliv geografická lokalizace obce – vnitrozemí/příhraničí) – Údaje o příhraničních oblastech mohou být lokálně poznamenány jevem, kdy se mobilní telefony uživatelů připojují k zahraničním sítím. Tím je narušeno vzorkování dat. Z pohledu metodiky tak uživatelé upadají do neurčeného stavu a v důsledku nemusí splnit podmínky pro přiřazení příznaků. Nicméně analýza SIM, jimž nebyl přiřazen domov, nepotvrdila, že by se jednalo o systémový problém.
  - o Vliv územní velikosti obce oproti velikosti buněk – Pokud jedna buňka základnové stanice pokrývá zcela nebo částečně více obcí, dochází k rozpočítávání dle zastavěných ploch v souladu s popsanou metodikou. V těchto případech neexistuje jednoznačný způsob, jak přesněji určit, v které obci se mobilní telefony uživatelů nacházejí. Byly nasazeny modely, jak se s těmito situacemi vypořádat, například aplikace modelu rozpočtu do území podle zastavěných ploch nebo relokační mechanismus, ale i tyto modely mohou lokálně selhávat (např. vychýlení počtu rezidentů vlivem lokální nehomogenity hustoty obyvatel v zastavěné ploše pod jednou buňkou).
- *Vliv zvolené metriky sledování odchylek vůči ČSÚ*
  - o **Nízká absolutní odchylka** – i malý nominální rozdíl počtu obyvatel v porovnání s ČSÚ představuje vysokou % odchylku v případech obcí s nízkým počtem obyvatel.

Nahodnocené obce se ve velké míře nacházejí ve Středočeském kraji s koncentrací v okolí hl. m. Prahy, a to zejména v jižním a jihovýchodním směru. Vedle toho jsou patrné významně vyšší podíly v rekreačních oblastech a v obcích s významným podílem průmyslové nebo skladové zástavby.

Naproti tomu podhodnocené obce se nacházejí zejména v příhraničních oblastech Třinecka a Jesenicka. V periferních oblastech, typicky v kopcovitých/hornatých

výběžcích hranic může být narušeno vzorkování dat, a tedy zde může být menší šance na přiřazení pobytů a v důsledku i příznaků „rezident“.

Základní kartogram počtu rezidentů je uveden v mapové příloze č. 1, kde jsou tmavší barvou znázorněny obce s vyšším počtem rezidentů v absolutní hodnotě. Na toto zobrazení potom navazuje mapová příloha č. 2, která představuje grafické znázornění porovnání těchto hodnot vůči údajům ČSÚ vycházející z podílu *počet rezidentů/počet obyvatel dle ČSÚ*. Modře jsou znázorněny obce, u kterých byl naměřen nižší počet rezidentů, než uvádí ČSÚ, v červené škále potom naopak obce s vyšším počtem rezidentů oproti údajům ČSÚ. Bíle potom obce s odchylkou do 15 %, (tzn. podíl 0,85 - 1,15).

## 7.2 Tabulka č. 2 - Data o meziobecní dojíždě

Jedná se o **upravenou formu OD (Origin-Destination) matice popisující směrově určenou pravidelnou meziobecní vyjížděku/dojížděku**. V databázi je identifikována **zdrojová obec, cílová obec a typ uskutečňované cesty**. Databáze zahrnuje všechny meziobecní relace (hlavní i vedlejší dojížděkové/vyjížděkové směry), které nemají nulovou hodnotu. Obsah databáze je prezentován v Tabulce č. 2.

Sloupec E – „*Typ cesty*“ - stanovuje typ cesty dle definovaných 3 typů dojížděky a dalších příznaků (druhé bydlení, nocující návštěvník, návštěvník). Typ cesty 1 = T1; 2 = T2; 3 = T3; 4 = D2; 5 = PN; 6 = N.

Sloupec F – „*Počet osob*“ – počet rezidentů zdrojové obce (sloupec A), kteří mají přiřazen příznak vůči cílové obci (sloupec C) odpovídající danému typu cesty (1-6)

Sloupec G – „*Počet osob (PO-PÁ, 6-18)*“ - počty rezidentů zdrojové obce (sloupec A) uskutečňujících cesty během obvyklé pracovní doby. Zobrazuje stejný ukazatel jako sloupec F, s tím rozdílem, že v tomto případě je časové okno zúženo jen na pracovní dobu, tj. ve **všední dny v době od 6 do 18 hodin**, přičemž pobyt v cílové obci může být zahájen před tímto intervalem a ukončen po intervalu. Jinými slovy do této statistiky jsou počítány jen ti rezidenti, kteří daná kritéria pro přidělení atributů splňují, přesto že jsou započítávány pouze ty cesty/pobyty, které byly uskutečňovány ve stanoveném intervalu. Přidělované příznaky T1vPD, T2vPD, T3vPD, NvPD.

Tabulka č. 2: Podoba výsledné databáze údajů o meziobecní dojížděcí

A	B	C	D	E	F	G
Data o meziobecní dojížděcí						
Zdrojová obec	Kód ZO	Cílová obec	Kód CO	Typ cesty	Počet osob	Počet osob (PO-PÁ, 6-18)
Adamov (Blansko)	581291	Babice nad Svitavou (Brno-venkov)	582794	1	10,0	xx
Adamov (Blansko)	581291	Babice nad Svitavou (Brno-venkov)	582794	2	30	20
Adamov (Blansko)	581291	Babice nad Svitavou (Brno-venkov)	582794	3	30,0	20,0
Adamov (Blansko)	581291	Babice nad Svitavou (Brno-venkov)	582794	4	xx	0
Adamov (Blansko)	581291	Babice nad Svitavou (Brno-venkov)	582794	5	10,0	0,0
Adamov (Blansko)	581291	Babice nad Svitavou (Brno-venkov)	582794	6	20	10
Adamov (Blansko)	581291	Babice u Rosic (Brno-venkov)	582808	2	xx	0,0
Adamov (Blansko)	581291	Babice u Rosic (Brno-venkov)	582808	3	0	0
Adamov (Blansko)	581291	Bačice (Třebíč)	590282	2	0,0	0,0
Adamov (Blansko)	581291	Batelov (Jihlava)	586862	2	0	0
Adamov (Blansko)	581291	Batelov (Jihlava)	586862	3	0,0	0,0
Adamov (Blansko)	581291	Běhařovice (Znojmo)	593737	2	xx	xx
Adamov (Blansko)	581291	Běhařovice (Znojmo)	593737	3	10,0	xx
Adamov (Blansko)	581291	Běhařovice (Znojmo)	593737	5	30	20

Součet počtu rezidentů vyjíždějících z dané obce za práci/školou (typ 1) do všech směrů, je roven počtu vyjíždějících za práci/školou definovaných v Tabulce č. 1, sloupec G. Rozdíly jsou způsobené provedenou anonymizací a zaokrouhlováním hodnot.

Ukázka grafické reprezentace Tabulky č. 2 je uvedena v mapové příloze č. 3, kde jsou znázorněny dominantní vazby T1 všech obcí. Z mapy jsou patrná základní spádová území dle nejčastějších dojížděk T1 a centra těchto oblastí.

### 7.3 Tabulka č. 3 - Denní chod počtu přítomného obyvatelstva

Jedná se o rozsáhlou databázi, která **v každé obci Česka sleduje počet obvykle přítomného obyvatelstva během jednotlivých dnů v týdnu**. Za každou z 6 254 obcí je v databázi k dispozici denně 24hodinových intervalů, za 7 dní v týdnu, tj. celkem 168 záznamů ke každé obci a každému sledovanému údaji. Databáze obsahuje statistické údaje na základě sledování během celého měření (28 dní) následně přepočítané na průměrné hodnoty výskytu v daný den a hodinu).

Na základě klasifikace pobytů je každé analyzované SIM kartě přiřazeno pro každý měřený hodinový interval v 28 dnech jedno místo pobytu dle maximálního stráveného času. Za každý den v týdnu a každou hodinu budou tedy k dispozici 4 údaje o obci výskytu (jelikož jsou 4 sledované týdny). Při vytváření průměrného týdne výskytu SIM karty má jeden výskyt v obci hodnotu 0,25. Každá SIM karta tak ovlivní průměrný interval poměrnou částí svých 4 naměřených obcí dominantního výskytu. Celkový počet přítomných osob v obci v daný časový interval je tak utvářen metodou bottom-up.

V případě, že v konkrétní hodinový úsek nebyl u SIM karty zaznamenán žádný alespoň 30minutový pobyt (ani část takového pobytu přesahující z jiných intervalů), je SIM karta

přičtena k souhrnné kategorii „na cestě“. V případě, že za daný hodinový úsek nejen nebyl zaznamenán pobyt, ale nebyly v geolokační síti detekovány žádné záznamy, je na SIM kartu pohlíženo jako na neaktivní/vypnutou a není v tomto intervalu započítána do žádné kategorie.

Stejně jako u předchozích tabulek, uvádí Tabulka č. 3 hodnoty po projekci počtu SIM karet na populaci.

Sloupec E – „*Celkový počet přítomných osob*“ - jedná se o **počet osob obvykle přítomných v obci v daném intervalu**. Sloupec v datech před krokem anonymizace odpovídá součtu hodnot ve sloupcích F-L. Ve výsledných tabulkách se může mírně lišit v důsledku zaokrouhlení dílčích hodnot či v důsledku anonymizace (včetně aplikace zarušení hodnot, viz kapitola „Anonymizace dat“).

Sloupec F – „*Rezidenti*“ - počet osob, které měly v daném intervalu dominantní pobyt v dané obci a které mají **vůči dané obci příznak rezidenta (D)**.

Sloupce G – „*Dojíždějící za prací/školou*“ - počet osob, které měly v daném intervalu dominantní pobyt v dané obci a které mají **vůči dané obci příznak dojíždějícího typ 1 (T1)**.

Sloupce H – „*Intenzivní dojíždka za službami*“ - počet osob, které měly v daném intervalu dominantní pobyt v dané obci a které mají **vůči dané obci příznak dojíždějícího typ 2 (T2)**.

Sloupce I – „*Občasná dojíždka za službami*“ - počet osob, které měly v daném intervalu dominantní pobyt v dané obci a které mají **vůči dané obci příznak dojíždějícího typ 3 (T3)**.

Sloupec J – „*Nocující návštěvníci*“ - počet osob, které měly v daném intervalu dominantní pobyt v dané obci a které mají **vůči dané obci příznak přenocujícího návštěvníka (PN)**.

Sloupec K – „*Návštěvníci*“ - počet osob, které měly v daném intervalu dominantní pobyt v dané obci a které mají **vůči dané obci příznak návštěvníka (N)**.

Sloupec L – „*Tranzitující a jiné*“ - počet osob, které měly v daném intervalu dominantní pobyt v dané obci a které nemají **vůči dané obci žádný příznak**. Jedná se o takzvaný neklasifikovaný pobyt, kdy SIM karta má v obci pobyt alespoň 30 minut, avšak nemá vůči obci žádný příznak přidělen.

Sloupec M – „*Druhé bydlení*“ - počet osob, které měly v daném intervalu dominantní pobyt v dané obci a které mají **vůči dané obci příznak druhé bydlení (D2)**. Tento příznak je doplňkový a nefiguruje v součtech, jelikož již jsou SIM karty započítány v rámci příznaků T2 a T3. Sloupec M tedy není využit pro výpočet průměrného počtu přítomných osob (sloupec E)

K přepočtu na populaci dochází za jednotlivé časové úseky, viz kap. Projekce na populaci.



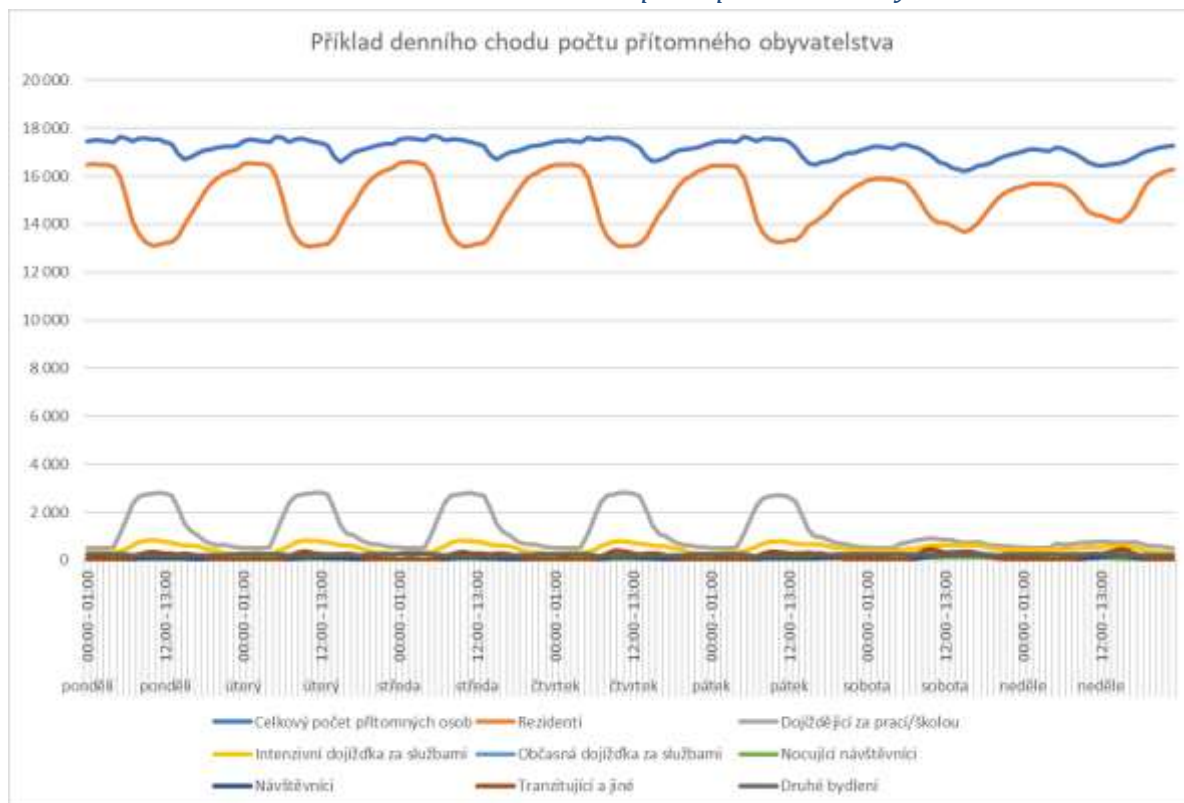
Tabulka č. 3: Podoba výsledné databáze údajů o denním chodu počtu přítomného obyvatelstva

Obec	Kód Obce	Den v týdnu	Časový interval	Celkový počet přítomných osob	z toho							Druhé bydlení
					Rezidenti	Dojíždějící za prací/školou	Intenzivní dojízdka za službami	Občasná dojízdka za službami	Nocující návštěvníci	Návštěvníci	Tranzitující a jiné	
Adamov (Blansko)	581291	pondělí	00:00 - 01:	3672,8	3536,3	52,5	59,0	16,0	9,0	0,0	XX	50,5
Adamov (Blansko)	581291	pondělí	01:00 - 02:	3679,6	3544,2	52,7	58,1	15,5	9,0	0,0	XX	50,9
Adamov (Blansko)	581291	pondělí	02:00 - 03:	3682,3	3546,0	53,6	58,5	15,5	8,8	0,0	XX	50,5
Adamov (Blansko)	581291	pondělí	03:00 - 04:	3670,6	3541,6	55,2	58,3	15,5	XX	0,0	XX	51,3
Adamov (Blansko)	581291	pondělí	04:00 - 05:	3636,2	3503,7	60,2	57,1	15,2	XX	0,0	XX	51,7
Adamov (Blansko)	581291	pondělí	05:00 - 06:	3410,6	3231,5	107,6	55,8	15,0	XX	XX	XX	48,4
Adamov (Blansko)	581291	pondělí	06:00 - 07:	3174,4	2887,9	188,8	67,3	20,3	XX	XX	10,1	44,5
---												
Přeskače (Znojmo)	550078	pátek	15:00 - 16:	82,4	73,5	XX	8,9	XX	XX	XX	XX	XX
Přeskače (Znojmo)	550078	pátek	16:00 - 17:	89,8	78,7	XX	11,1	XX	XX	XX	XX	XX
Přeskače (Znojmo)	550078	pátek	17:00 - 18:	92,7	82,8	XX	9,9	XX	XX	XX	XX	XX
Přeskače (Znojmo)	550078	pátek	18:00 - 19:	104,7	83,9	9,7	11,1	XX	XX	XX	XX	XX
Přeskače (Znojmo)	550078	pátek	19:00 - 20:	109,9	88,4	10,0	11,4	XX	XX	XX	XX	XX
---												
NA CESTĚ	1	pondělí	00:00 - 01:	35336,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	35336,0
NA CESTĚ	1	pondělí	01:00 - 02:	19577,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	19577,0
NA CESTĚ	1	pondělí	02:00 - 03:	19805,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	19805,5
NA CESTĚ	1	pondělí	03:00 - 04:	26565,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	26565,0
NA CESTĚ	1	pondělí	04:00 - 05:	61679,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	61679,4
NA CESTĚ	1	pondělí	05:00 - 06:	166304,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	166304,0
NA CESTĚ	1	pondělí	06:00 - 07:	261724,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	261724,3

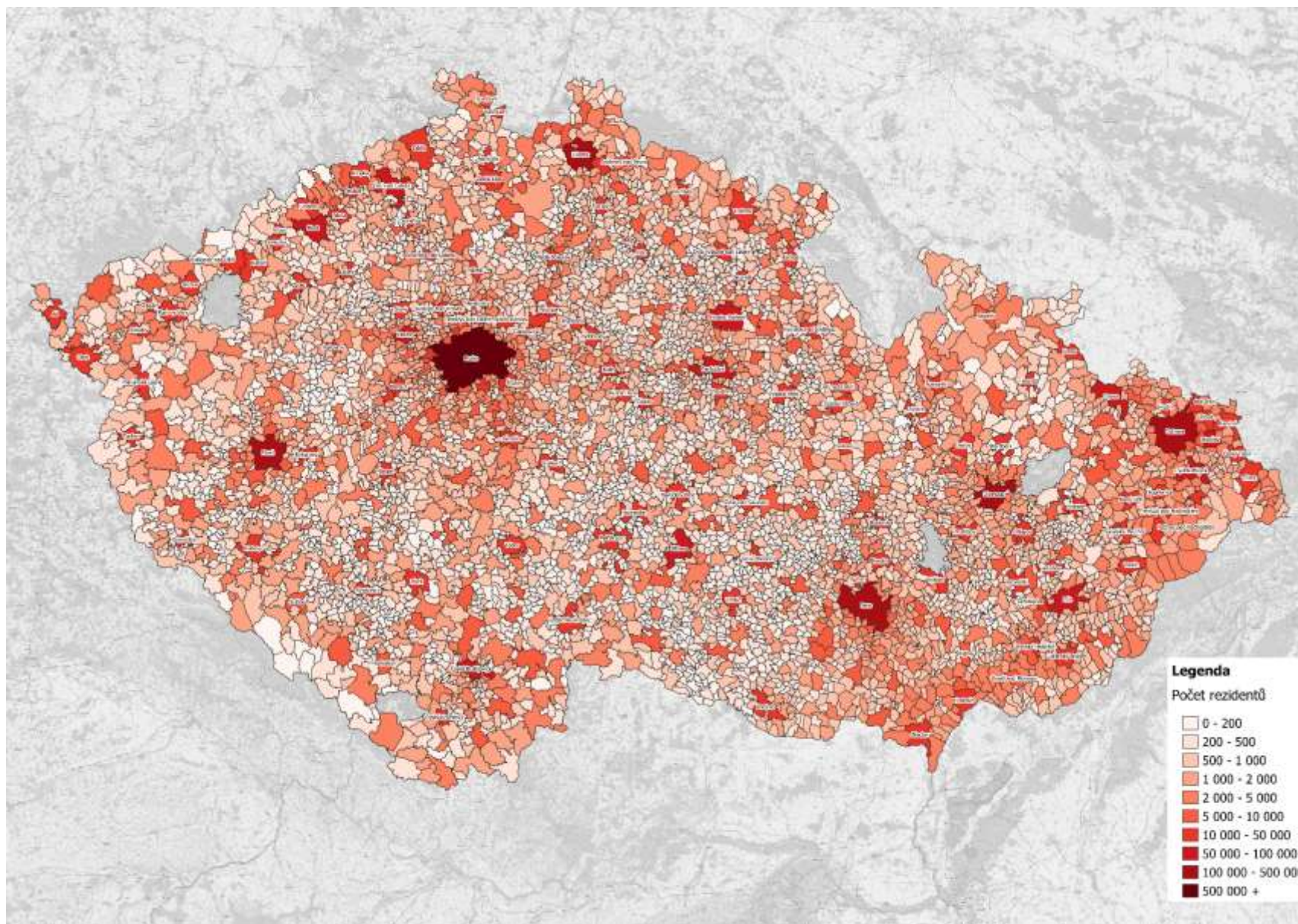
Grafická reprezentace Tabulky č. 3, konkrétně rozdílu počtu přítomných v noci a ve dne je uvedena v Mapové příloze č. 4. Konkrétně se jedná o procentuální odchylku počtu přítomných ve středu 12:00 – 13:00 oproti časovému intervalu 00:00 – 01:00. Záporné hodnoty představují pokles v polední hodinu oproti půlnoci, kladné hodnoty pak vyšší počet přítomných v poledne oproti půlnoci. V mapě jsou tak patrné oblasti v okolí velkých měst, kde dochází k odlivu počtu přítomných ve prospěch těchto center.

Na obrázku níže je uveden graficky znázorněný průběh počtu přítomných u vybrané obce v průběhu týdne. Modře je znázorněn celkový počet přítomných osob, který se skládá ze všech ostatních příznaků (vyjma Druhého bydlení). Z grafu je patrná denní fluktuace počtu rezidentů, kdy během dne jejich počet klesá vlivem vyjížděk mimo obec během dne. V nočních hodinách pak jejich počet roste, protože se opět vrací do obce na přespaní. Opačně se chová křivka počtu dojíždějících (v absolutním počtu patrné zejména dojízdky za prací a školou a za službami). Osoby s těmito příznaky jsou rezidenti v jiných obcích a ve sledované obci jsou přítomni během dne a na noc se obvykle vrací zpět do obce své rezidentury. V grafu je navíc patrný odlišný rytmus o víkendu, kdy je pokles rezidentů během dne na jedné straně méně výrazný, na druhé straně v noci jejich počet nedosahuje takových hodnot jako během týdne, což může být způsobeno např. jejich odjezdem na celý víkend mimo obec. Počty dojíždějících jsou pak ve zvolené obci o víkendu výrazně nižší.

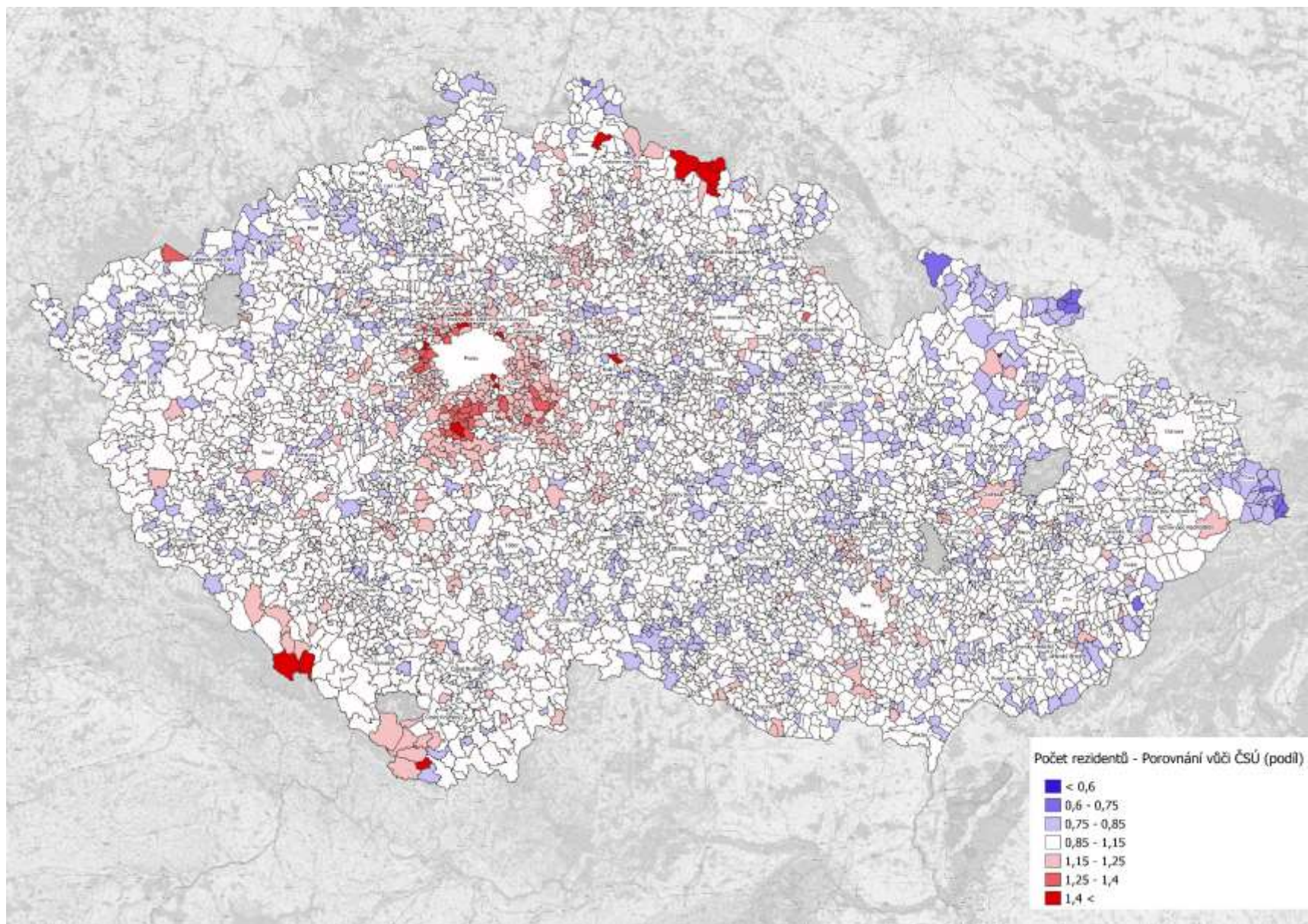
Obrázek 14: Příklad denního chodu počtu přítomného obyvatelstva



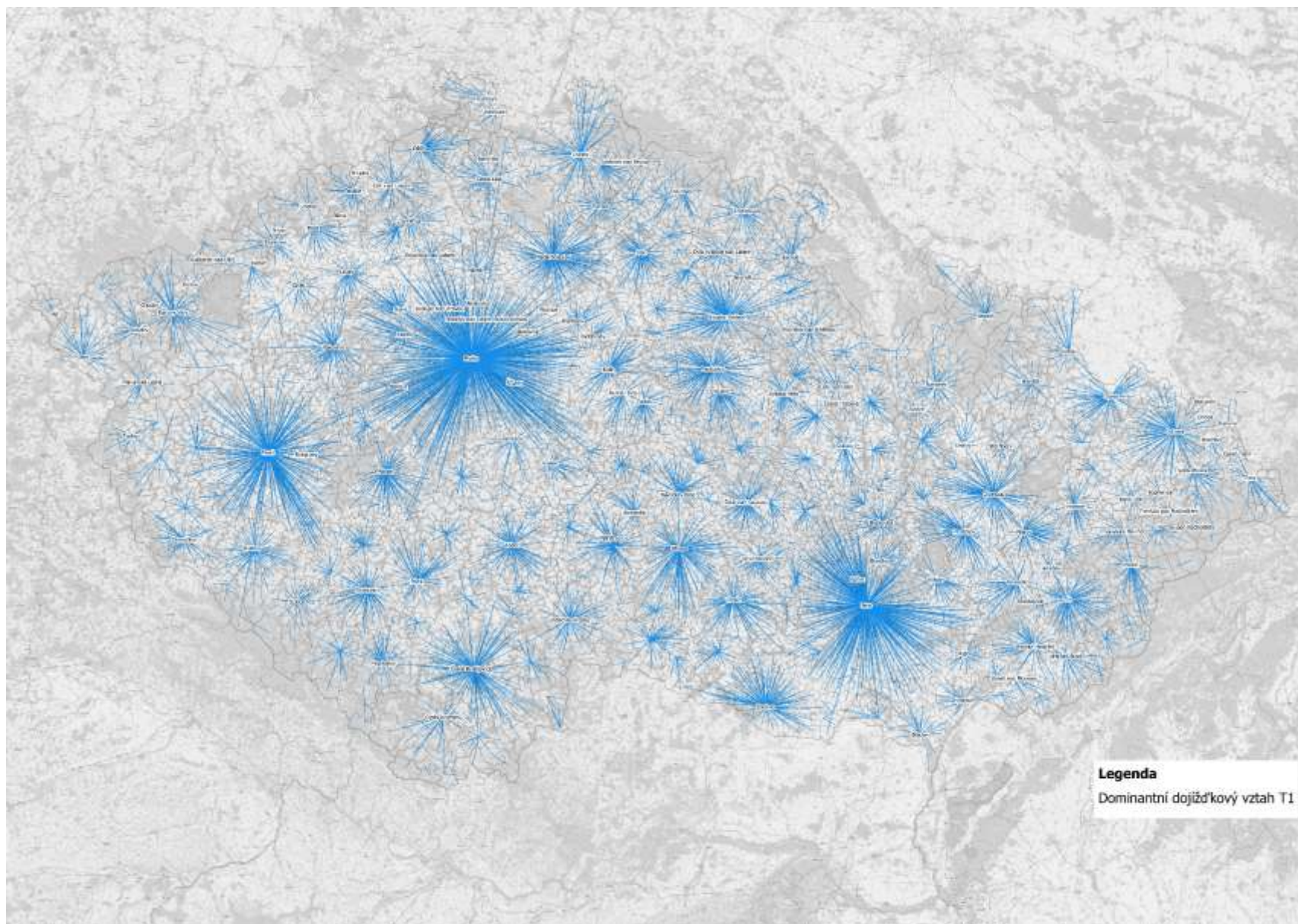
Mapová příloha č. 1 – Základní kartogram počtu rezidentů



Mapová příloha č. 2 – Porovnání vůči ČSÚ dle %



Mapová příloha č. 3 – Nejčastější směr vyjíždky T1



Mapová příloha č. 4 – Rozdíl počtu přítomných ve 12:00 a 00:00 ve vybraném dni v týdnu (středa)

