

UNIVERZITA PARDUBICE

FAKULTA EKONOMICKO-SPRÁVNÍ

DIPLOMOVÁ PRÁCE

2025

Bc. Petr Kábele

Univerzita Pardubice
Fakulta Ekonomicko-správní

Analýza a vizualizace IoT dat pro podporu rozhodování

Diplomová práce

2025

Bc. Petr Kábele

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2024/2025

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Bc. Petr Kábele**
Osobní číslo: **E23044**
Studijní program: **N0613A140041 Aplikovaná informatika – Data Science pro business**
Téma práce: **Analýza a vizualizace IoT dat pro podporu rozhodování**
Zadávající katedra: **Ústav systémového inženýrství a informatiky**

Zásady pro vypracování

Cílem práce je analýza a vizualizace vybraných IoT dat pro podporu rozhodování. Práce bude obsahovat přehled současných přístupů k řešení této problematiky, volbu ukázkových IoT dat, jejich analýzu a vizualizaci výsledků.

Osnova:

- Internet věcí (IoT) jako zdroj dat včetně prostorových.
- Metody sběru, ukládání a analýzy dat v prostředí IoT – základní charakteristika.
- Výběr vhodných ukázkových IoT datových sad a formulace příkladů problémů, které s jejich využitím mohou být řešeny.
- Analýza ukázkových dat a vyhodnocení a vizualizace výsledků.

Rozsah pracovní zprávy: **cca 50 stran**
Rozsah grafických prací:
Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam doporučené literatury:

Demers, Michael N. *Fundamentals of Geographical Information Systems*. Hoboken, New Jersey, USA : John Wiley & Sons Inc, 2008. 978-0470129067.
Goodchild, Michael F., Rhind, David W. , Maguire, David J., Longley, Paul A. *Geographic Information Science and Systems, 4th Edition*. Hoboken, New Jersey, USA : John Wiley & Sons Inc, 2015. 978-1118676950.
Tomlinson, Roger. *Thinking About GIS: Geographic Information System Planning for Managers*. místo neznámé : Esri Press, 2011. 978-1589482739.
Wang, Fahui. *Quantitative Methods and Socio-Economic Applications in GIS*. Boca Raton, Florida, USA : CRC Press, 2014. 978-1466584723.

Vedoucí diplomové práce: **prof. Ing. Jitka Komárková, Ph.D.**
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: **1. září 2024**
Termín odevzdání diplomové práce: **30. dubna 2025**

prof. Ing. Jan Stejskal, Ph.D. v.r.
děkan

L.S.

prof. Ing. Petr Hájek, Ph.D. v.r.
garant studijního programu

V Pardubicích dne 1. září 2024

Prohlašuji:

Práci s názvem *Analýza a vizualizace IoT dat pro podporu rozhodování* jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnicí Univerzity Pardubice č. 7/2019 Pravidla pro odevzdávání, zveřejňování a formální úpravu závěrečných prací, ve znění pozdějších dodatků, bude práce zveřejněna prostřednictvím Digitální knihovny Univerzity Pardubice.

V Pardubicích dne 30.4. 2025

Bc. Petr Kábele v. r.

PODĚKOVÁNÍ

Na tomto místě bych chtěl v první řadě poděkovat paní prof. Ing. Jitce Komárkové, Ph.D. za odborné vedení, ochotu, trpělivost a vstřícnost, kterou mi po celou dobu vypracování mé diplomové práce poskytovala. Zároveň bych chtěl také poděkovat celé své rodině a partnerce Barboře Matykové, kteří mi po celou dobu byli velkou oporou.

ANOTACE

Cílem této diplomové práce je *Analyzovat a vizualizovat IoT data pro podporu rozhodování* v konkrétně vymezených oblastech. První část práce se věnuje teoretickému vymezení základních pojmů, jako jsou otevřená data, IoT zařízení a analytické přístupy k jejich zpracování. Následuje praktická část, která zahrnuje analýzu prostorové a časové distribuce uživatelů Twitteru na území Spojených států amerických a zpracování simulovaných IoT dat zaměřených na oblast Greater Melbourne v Austrálii. Dále je vytvořena syntetická datová sada pro modelové scénáře v logistice. Součástí práce je předzpracování dat, aplikace vybraných analytických metod a vizualizace výsledků pomocí dostupných nástrojů. Diplomová práce je zakončena shrnutím dosažených výsledků a diskusí o omezeních použitých dat a metod.

KLÍČOVÁ SLOVA

Internet věcí, otevřená data, GIS, Twitter, IoT zařízení, logistika, vizualizace dat, syntetická data, rozhodování, optimalizace tras, Austrálie, USA

TITLE

Analysis and Visualization of IoT Data for Decision Support

ANNOTATION

The objective of this master thesis is to *Analyze and visualize IoT data to support decision-making* in specifically defined areas. The first part of the thesis focuses on the theoretical background, including concepts such as open data, IoT devices, and analytical approaches to data processing. The practical part includes the analysis of digital activity based on Twitter data from the United States, the processing of simulated IoT data related to the Greater Melbourne area in Australia, and the creation of a synthetic dataset for modeling logistics scenarios. The thesis involves data preprocessing, the application of selected analytical methods, and the visualization of results using available tools. The thesis concludes with a summary of the achieved results and a discussion of the limitations of the used data and methods.

KEYWORDS

Internet of Things, open data, GIS, Twitter, IoT devices, logistics, data visualization, synthetic data, decision-making, route optimization, Australia, USA

OBSAH

Úvod	9
1 Internet věcí jako zdroj dat včetně prostorových	10
1.1 Typy dat v oblasti IoT	10
1.2 Využití IoT v dopravě.....	11
2 Metody sběru, ukládání a analýzy dat v prostředí IoT – základní charakteristika	13
2.1 Metody sběru IoT dat	13
2.1.1 Přenos dat pomocí komunikačních protokolů	13
2.1.2 Kvalita přenosu a integrace IoT dat.....	13
2.2 Metody a technologie pro ukládání dat v prostředí IoT	15
2.2.1 Centralizované a distribuované úložiště dat	15
2.2.2 Technologie ukládání dat v IoT prostředí	15
2.3 Metody analýzy v prostředí IoT dat	16
2.3.1 Předzpracování dat pro analýzu a vizualizaci	16
2.3.2 Přístupy k analýze v prostředí internetu věcí	18
2.3.3 Strojové učení a umělá inteligence v IoT	19
2.4 Výzvy v oblasti sběru, ukládání a analýzy IoT dat	20
2.4.1 Zabezpečení a ochrana soukromí dat	20
2.4.2 Škálovatelnost systémů	21
2.4.3 Standardizace a kompatibilita zařízení	22
2.4.4 Energetická efektivita a napájení IoT zařízení	22
2.4.5 Energetická účinnost komunikačních protokolů	23
2.5 Otevřená data – jejich zdroje dat a práce s nimi	24
3 Zvolený přístup k analýze IoT dat a vizualizaci výsledků	27
3.1 Obecný postup.....	27
3.2 Použité nástroje a prostředí pro analýzu dat	28
3.2.1 Role GIS při analýze prostorových dat	28
3.2.2 Prostředí Python	28
3.2.3 Prostředí ArcMap.....	29
4 Ukázkové příklady analýzy dostupných dat a vyhodnocení a vizualizace výsledků	30
4.1 Příklad 1 – Časoprostorová analýza uživatelů Twitteru v USA.....	30
4.1.1 Popis použité datové sady	31
4.1.2 Předzpracování dat	31
4.1.3 Časová analýza v Pythonu.....	32
4.1.4 Analýza prostorové distribuce uživatelů Twitteru	35

4.1.5 Závěr a interpretace prostorové distribuce uživatelů Twitteru	36
4.2 Příklad 2 – Prostorová analýza vztahu mezi uživateli a servery v oblasti Melbourne....	38
4.2.1 Datová sada uživatelů – users-aus.csv.....	39
4.2.2 Datová sada serverů – site.csv	39
4.2.3 Výběr vhodné oblasti pro prostorovou analýzu.....	40
4.2.4 Předzpracování dat pro analýzu oblasti Melbourne	43
4.2.5 Analýza prostorového pokrytí v oblasti Melbourne	44
4.2.6 Závěr a interpretace prostorové analýzy oblasti Melbourne	47
4.3 Příklad 3 – Modelový příklad využití syntetických IoT dat a prostorové analýzy efektivitu trasy.....	56
4.3.1 Vytvoření a charakteristika syntetické datové sady	57
4.3.2 Základní analýza a vizualizace syntetických dat.....	59
4.3.3 Grafická analýza hlavních proměnných	60
4.3.4 Prostorová analýza tras a výpočet parametrů	63
4.3.5 Závěr a interpretace výsledků.....	67
Závěr	70
Seznam použité literatury	72
Seznam příloh.....	78

SEZNAM OBRÁZKŮ

Obrázek 1: Základní architektura pro zpracování dat ze senzorů IoT, fúzi dat a analýzu dat ..	14
Obrázek 2: Grafické zobrazení tzv. "Man-in-the-middle" útoku.....	21
Obrázek 3: Denní aktivita uživatelů Twitteru podle hodin v USA	32
Obrázek 4: Počet tweetů podle časové zóny.....	33
Obrázek 5: Denní rytmus tweetování podle časové zóny v USA.....	34
Obrázek 6: Počet tweetů podle dnů v USA	34
Obrázek 7: Prostorová distribuce uživatelů Twitter v USA.....	37
Obrázek 8: Prostorová distribuce uživatelů Twitter – východ.....	38
Obrázek 9: Rozložení uživatelů v prostoru.....	41
Obrázek 10: Rozložení serverů v prostoru	41
Obrázek 11: Vymezení metropolitní oblasti Greater Melbourne	48
Obrázek 12: Hustota uživatelů v oblasti Melbourne (2018).....	49
Obrázek 13: Hustota serverů v oblasti Melbourne (2018).....	50
Obrázek 14: Rozložení uživatelů v oblasti Melbourne (2018).....	51
Obrázek 15: Rozložení serverů v oblasti Melbourne (2018).....	52
Obrázek 16: Počet uživatelů na server v oblasti Melbourne (2018).....	54
Obrázek 17: Analýza dostupnosti – vzdálenost k nejbližšímu serveru.....	55
Obrázek 18: Prostorová dostupnost serverů z pohledu uživatelů v oblast Melbourne (2018) ..	56
Obrázek 19: Hmotnost nákladu vs Spotřeba paliva.....	61
Obrázek 20: Délka trasy vs Emise CO ₂	62
Obrázek 21: Hmotnost nákladu vs Emise CO ₂	62
Obrázek 22: Vizualizace dopravních tras Hradec Králové – Trutnov	68

SEZNAM TABULEK

Tabulka 1: Datový slovník uživatelů Twitter	31
Tabulka 2: Datová sada uživatelů	39
Tabulka 3: Datová sada serverů	40
Tabulka 4: Kvantitativní porovnání podle států.....	42
Tabulka 5: Přehled sloupců syntetické datové sady.....	58
Tabulka 6: Základní statistiky syntetické datové sady.....	60
Tabulka 7: Výsledky délky a sklonu tras	64
Tabulka 8: Přehled výsledků prostorové analýzy variant tras	66
Tabulka 9: Výsledky parametrů pro zatížení 1000 kg	67

ÚVOD

Rozvoj technologií internetu věcí (IoT) a rostoucí dostupnost prostorových dat otevírají nové možnosti pro analýzu a podporu rozhodování. Data generovaná senzory, mobilními zařízeními a dalšími zdroji obsahují nejen časové, ale i geografické informace, které mohou být využity pro lepší pochopení různých jevů a optimalizaci provozních činností. Pro efektivní využití těchto dat je však nutné zvládnout celý proces od jejich sběru a předzpracování až po analýzu a vhodnou formu vizualizace.

Cílem této práce je analyzovat a vizualizovat IoT data s důrazem na jejich využití v rozhodovací podpoře. Práce kombinuje využití otevřených datových zdrojů a modelově vytvořených syntetických dat, přičemž k analýze a prostorovému zpracování dat jsou využívány běžně dostupné nástroje. Důraz je kladen na praktickou aplikaci základních metod v dostupných softwarových prostředích a realistické zhodnocení možností i omezení takto prováděných analýz.

V rámci práce byly realizovány tři samostatné příklady. První část se zaměřila na časoprostorovou analýzu aktivity uživatelů Twitteru na území Spojených států amerických, a to na základě rozsáhlého otevřeného datasetu. Druhá část se věnovala analýze dostupnosti serverové infrastruktury v oblasti Greater Melbourne, kde byla kombinována simulace rozmístění uživatelů s reálnými daty o lokalizaci serverů. Třetí část práce se zaměřila na modelový příklad logistické analýzy v Královéhradeckém kraji, kdy byl vytvořen syntetický dataset jízd a následně analyzovány variantní trasy v prostředí ArcMapu.

Prostorové analýzy a kartografické vizualizace byly realizovány převážně v prostředí ArcMap s využitím základních analytických funkcí, přičemž pro práci s daty byl použit i Python. Práce kladla důraz na procesní stránku analýzy – od přípravy dat, jejich zpracování až po interpretaci výsledků – a reflektovala reálné limity související s kvalitou, úplností a strukturou dostupných dat. Výstupy byly vizualizovány formou map a grafů, které měly sloužit jako názorný podklad pro rozhodování v jednotlivých scénářích.

Hlavní ambicí práce bylo ukázat, že i při využití otevřených či syntetických dat a základních analytických nástrojů je možné získat užitečné poznatky využitelné pro rozhodovací procesy. Zároveň práce upozornila na nezbytnost pečlivého přístupu k datům a na vědomí omezení, která s jejich charakterem souvisejí.

1 INTERNET VĚCÍ JAKO ZDROJ DAT VČETNĚ PROSTOROVÝCH

Internet věcí (IoT) představuje propojení fyzických zařízení a senzorů prostřednictvím internetu, což umožňuje shromažďování, přenos a analýzu dat v reálném čase. Tento koncept zahrnuje různá zařízení, která bez nutnosti lidského zásahu interagují a předávají informace, což dnes proniká do oblastí jako průmysl, zdravotnictví a stále více i do dopravy (IoTPORT, 2020).

1.1 Typy dat v oblasti IoT

S rostoucím počtem IoT zařízení roste i množství dat, která tato zařízení neustále generují. Aby bylo možné tato data efektivně využít, je důležité je správně klasifikovat a analyzovat. Existuje obrovské množství různých typů sensorových dat, která se liší svou strukturou, vlastnostmi, nároky na zpracování i účelem, pro který jsou sbírána. Neexistuje však žádné jednotné členění, které by platilo univerzálně pro všechny systémy. Správně třídění a porozumění jednotlivým kategoriím dat je proto klíčové pro jejich efektivní využití v různých IoT oblastech (Abba Ari a kol., 2024).

Nejběžnějším typem IoT dat jsou **senzorová data**, která tvoří základ celého ekosystému chytrých zařízení. Sensory neustále sledují okolní prostředí, měří různé hodnoty a umožňují automatizaci procesů bez zásahu člověka. Díky nim mohou chytré systémy rychle reagovat na změny, ať už jde o úpravu teploty v místnosti, optimalizaci průmyslových strojů nebo predikci údržby zařízení. Dle Abba Ari a kol. (2024) existuje několik klíčových kategorií sensorových dat, které se liší podle účelu, způsobu sběru a metod zpracování.

Enviromentální sensorová data zahrnují měření fyzikálních veličin v prostředí, jako jsou teplota, tlak, kvalita vzduchu, vlhkost apod. Používají se například v chytrých domácnostech, meteorologii a zemědělství (Abba Ari a kol., 2024).

Průmyslová a automatizační sensorová data sledují vibrace strojů, teplotu motorů nebo spotřebu energie. Pomáhají odhalovat závady dříve, než se stanou problémem, což šetří náklady na opravy (Abba Ari a kol., 2024).

Pohybová a polohová sensorová data využívají akcelerometry, gyroskopy a GPS senzory pro sledování pohybu a polohy zařízení. Tyto senzory se běžně využívají v chytrých telefonech, autonomních vozidlech a nositelných zařízeních (Abba Ari a kol., 2024).

Energetická a síťová sensorová data měří spotřebu elektřiny, vody, plynu nebo stav baterií v IoT zařízeních. V chytrých městech pomáhají zlepšovat distribuci energie a snižovat zbytečné plýtvání. (Abba Ari a kol., 2024).

Bezpečnostní a identifikační sensorová data využívají RFID čipy, NFC a biometrické senzory pro kontrolu přístupu nebo platby. Využívají se například ve firemních budovách pro vstup pomocí čipové karty nebo v logistice pro sledování pohybu zásilky (Abba Ari a kol., 2024).

1.2 Využití IoT v dopravě

V dopravě IoT využívá senzory ke sledování parametrů, jako jsou rychlost, poloha či stav vozidla. Díky kombinaci těchto dat a IoT algoritmů lze efektivně řídit dopravní tok, optimalizovat energetickou spotřebu a zajišťovat bezpečnost na silnicích, čímž se zvyšuje plynulost dopravy a snižuje riziko nehod. Jednou z hlavních výzev, na kterou se IoT v dopravě zaměřuje, je prevence nehod a efektivní řízení provozu. Díky propojeným vozidlům (tzv. „smart cars“), která komunikují mezi sebou i s okolní infrastrukturou, mohou systémy poskytovat v reálném čase data o aktuální dopravní situaci. Tato data zahrnují informace o překážkách, stavu vozovky či hustotě provozu, což přispívá k bezpečnějšímu a plynulejšímu dopravnímu provozu (Nawara a Kashef, 2021; Shah a kol., 2022).

IoT v dopravě využívá různé úrovně automatizace, od asistovaných řídicích systémů (ADAS), které upozorňují na potenciální rizika, až po zcela autonomní řízení, kde vozidla zvládají cestu bez zásahu člověka. Zároveň IoT přispívá k lepšímu porozumění dynamiky okolního prostředí a stavů vozidel pomocí senzorů sledujících například tlak v pneumatikách nebo spotřebu paliva, což přispívá k prevenci nečekaných poruch. Tímto způsobem IoT nejen zvyšuje bezpečnost, ale také podporuje rozhodování řidičů a umožňuje efektivnější využití dopravní infrastruktury (Shah a kol., 2022).

Dalším přínosem IoT v dopravě je podpora prediktivní údržby, která umožňuje sledování stavu vozidel a infrastruktury pomocí různých senzorů. Tyto senzory dokážou zaznamenávat informace o opotřebením součástí, vibracích či teplotních změnách, což pomáhá předcházet poruchám a snižuje náklady na údržbu. Kombinací těchto dat s algoritmy strojového učení mohou systémy předpovědět, kdy je potřeba údržba nebo výměna komponent, čímž se zvyšuje bezpečnost a efektivita dopravy (Safari Bazargani a kol., 2021).

V článku Nawara a Kashef (2021) popsali přínos IoT pro systémy doporučení (IoT-CARS). Tyto systémy shromažďují data o preferencích uživatelů, jejich poloze a okolních podmínkách,

aby poskytovaly personalizovaná doporučení, například ohledně ideální trasy nebo vhodného dopravního prostředku v závislosti na aktuální situaci. IoT-CARS tak zlepšují uživatelskou zkušenost a pomáhají optimalizovat plynulost dopravy.

IoT přináší do dopravy různé inovativní aplikace, které zlepšují bezpečnost a efektivitu provozu. Chytré parkovací systémy využívají senzory ke sledování obsazenosti parkovacích míst, což řidičům pomáhá rychleji najít volná místa a snižuje tak dopravní zácpy. Ovládání dopravních světel a křižovatek v reálném čase pomocí IoT umožňuje přizpůsobit načasování světel aktuálnímu provozu, což zlepšuje plynulost a zkracuje čekací doby na semaforech. Díky všem těmto aplikacím IoT se doprava stává plynulejší, bezpečnější a efektivnější (Oladimeji a kol., 2023).

2 METODY SBĚRU, UKLÁDÁNÍ A ANALÝZY DAT V PROSTŘEDÍ IOT – ZÁKLADNÍ CHARAKTERISTIKA

Internet věcí přináší revoluci do způsobu sběru a využívání dat. Zařízení IoT vybavená senzory mohou automaticky sledovat a shromažďovat obrovské množství informací z různých prostředí, což přináší přesné a aktuální poznatky potřebné pro zlepšení rozhodovacích procesů a provozní efektivity v oblastech jako průmysl, doprava, zdravotnictví nebo chytrá města. Aby tato data měla skutečnou hodnotu, nestačí je jen sbírat – důležité je také zajistit jejich spolehlivý přenos, bezpečné ukládání a efektivní analýzu. Klíčovou roli zde hrají moderní metody sběru dat, které umožňují pracovat s daty rychle, přesně a bezpečně (Dusun, 2024).

2.1 Metody sběru IoT dat

Základem sběru IoT dat jsou senzory, které monitorují různé fyzikální jevy a převádějí je na digitální data. Tyto senzory jsou zásadní součástí IoT ekosystému, protože umožňují sběr informací o zařízeních, prostředí nebo uživateli. Mezi příklady běžně používaných IoT senzorů patří například: (Dusun, 2024)

- **Teplotní senzory** – sledují teplotu okolí, využívají se v meteorologii, průmyslu nebo chytrých domácnostech.
- **Senzory spotřeby energie** – měří spotřebu energie, vody a plynu, pomáhají optimalizovat provoz průmyslových systémů nebo chytrých domácností.
- **Pohybové senzory** – detekují přítomnost a pohyb objektů, uplatňují se v bezpečnostních systémech, chytrém osvětlení nebo autonomních vozidlech.

2.1.1 Přenos dat pomocí komunikačních protokolů

Po sesbírání je nutné data efektivně přenést do centralizovaných systémů. IoT využívá různé komunikační protokoly, jako jsou Wi-Fi, Bluetooth, Zigbee nebo specializované sítě typu LoRa a NB-IoT. Výběr vhodného protokolu závisí na nárocích aplikace – přenosové rychlosti, dosahu nebo energetické efektivitě (WebbyLab, 2025). Jedním z příkladů zařízení, které usnadňuje efektivní přenos dat v prostředí IoT, je DSGW-510 KNX Zigbee Gateway. Díky podpoře více protokolů může tato jednotka efektivně řídit komunikaci mezi zařízeními a tím zvyšovat flexibilitu a efektivitu sítě (Dusun, 2025).

2.1.2 Kvalita přenosu a integrace IoT dat

Po sběru sensorových dat je klíčové zajistit jejich spolehlivý přenos a následnou integraci do IoT infrastruktury. Zatímco výběr vhodného komunikačního protokolu (např. MQTT, Zigbee)

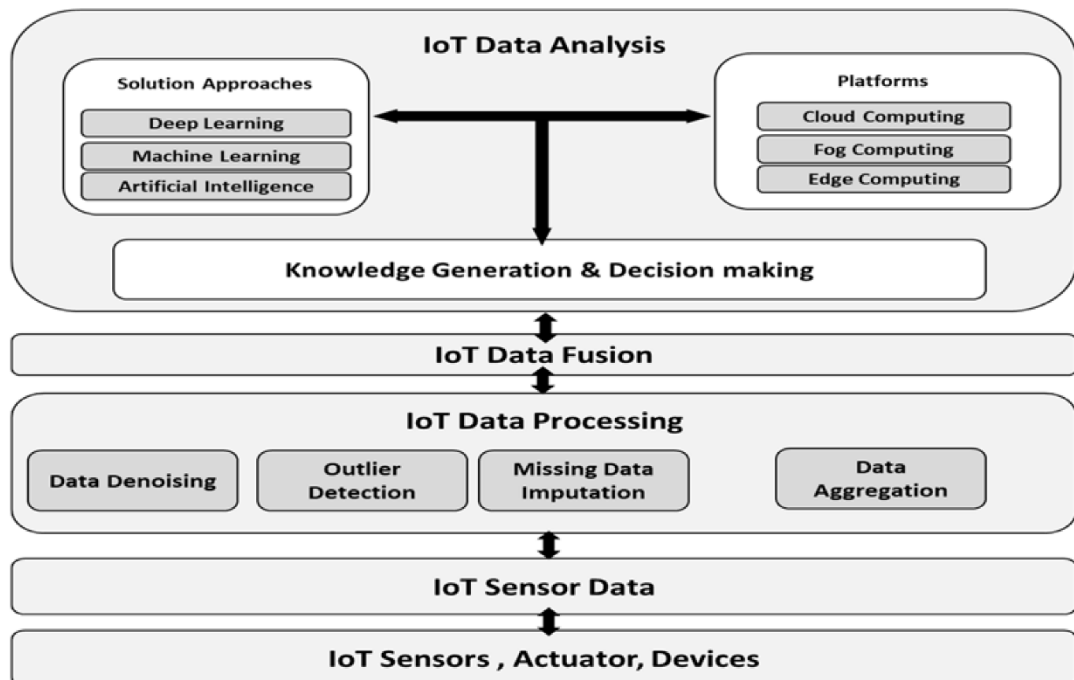
ovlivňuje efektivitu přenosu, neméně důležitá je i kvalita a přesnost přenášených dat, které hrají klíčovou roli při jejich následném zpracování a analýze (WebbyLab, 2025).

Mezi klíčové faktory ovlivňující kvalitu přenosu IoT dat patří: (Abba Ari a kol., 2024)

- **latence a rychlost přenosu** – zpoždění může být problémem pro aplikace vyžadující okamžitou odezvu (např. chytré dopravy, průmyslová robotika),
- **šum a nepřesnosti** – senzory mohou produkovat zkreslená nebo chybná data, která je nutné filtrovat či doplňovat,
- **ztrátovost a bezpečnost dat** – přenášená data mohou být ovlivněna výpadky, rušením signálu nebo kybernetickými útoky.

Pro zvýšení přesnosti a spolehlivosti se často využívají techniky fúze dat, které kombinují informace z více senzorů. Např. Kalmanovy filtry nebo Bayesovské metody pomáhají snižovat šum a ověřit kvalitu vstupních dat (Krishnamurthi a kol., 2020).

Následující schéma (viz Obrázek 1) znázorňuje jednotlivé fáze zpracování IoT dat – od odstranění šumu (Data Denoising) a detekce odlehlých hodnot (Outlier Detection) po doplnění chybějících dat (Missing Data Handling) (Krishnamurthi a kol., 2020).



Obrázek 1: Základní architektura pro zpracování dat ze senzorů IoT, fúzi dat a analýzu dat

Zdroj: (Krishnamurthi a kol., 2020)

2.2 Metody a technologie pro ukládání dat v prostředí IoT

Internet věcí vytváří rozsáhlé objemy dat, které vyžadují zdatné a spolehlivé strategie pro ukládání a správu. Vzhledem k tomu, že množství zařízení a aplikací roste, je nezbytné navrhnout systémy, které zaručují nejen bezpečné uchovávání dat, ale také jejich rychlé zpracování a okamžitou dostupnost. Převládající metodiky pro ukládání dat v IoT zahrnují centralizované a distribuované techniky, z nichž každá má odlišné výhody a specializované aplikace (Cruz Huacarpuma a kol., 2017; Liu a kol., 2019).

2.2.1 Centralizované a distribuované úložiště dat

Centralizované ukládání dat, často realizované prostřednictvím cloudových platform jako Microsoft Azure, umožňuje agregaci dat v jediném úložišti, které následně podporuje jejich analýzu a správu. Tato metodika poskytuje škálovatelnost, integraci různorodých aplikací IoT a spolehlivé zabezpečení. Díky konsolidovanému rámci je vhodná zejména pro aplikace vyžadující robustní zpracování dat. Naopak, tento model může být omezen vyšší latencí a závislostí na připojení k internetu (Liu a kol., 2019).

Distribuované úložiště dat představuje flexibilnější alternativu, kdy jsou data uložena v několika decentralizovaných uzlech. Tato strategie zvyšuje dostupnost dat, snižuje zátěž centrálních serverů a odolává výpadkům. Je obzvláště vhodná pro aplikace IoT s požadavkem na nízkou latenci a vysokou dostupnost v reálném čase, například u průmyslových monitorovacích systémů nebo v městské infrastruktuře. Přístup je však složitější na implementaci a řízení (Cruz Huacarpuma a kol., 2017).

Kombinace centralizovaných a distribuovaných metod prostřednictvím hybridních modelů umožňuje využívat výhod obou přístupů. Centralizované modely nabízejí analytickou podporu, bezpečnost a dlouhodobé ukládání, zatímco distribuované systémy zajišťují nízkou latenci a dostupnost v reálném čase. Hybridní přístupy, například kombinace NVM (Non-Volatile Memory) a SSD, umožňují efektivní ukládání časových dat a optimalizaci výkonu i redundance (Cai a kol., 2021).

2.2.2 Technologie ukládání dat v IoT prostředí

Mezi nejvýznamnější technologie patří cloudové platformy, jako jsou Microsoft Azure, AWS a Google Cloud, které nabízejí centralizované ukládání dat s podporou pokročilé analýzy. Distribuované databáze, například Apache Cassandra, umožňují nízkou latenci a vysokou dostupnost dat v reálném čase. Edge computing zařízení, jako je Raspberry Pi nebo NVIDIA

Jetson, umožňují lokální zpracování dat a minimalizují závislost na centrálních systémech (Jyothis, 2022; Borra, 2024).

Microsoft Azure IoT nabízí výkonné nástroje pro správu, analýzu a bezpečné ukládání dat. Technologie jako Azure IoT Hub podporují spolehlivou komunikaci mezi zařízeními a cloudem, což je zásadní pro aplikace vyžadující rychlé a přesné zpracování dat. Platforma Azure IoT Central umožňuje rychlé nasazení projektů díky šablonám a integrovaným analytickým nástrojům. Pro oblasti s omezeným připojením je vhodným řešením Azure IoT Edge, který umožňuje analýzu dat přímo na zařízení (Borra, 2024).

Apache Cassandra je open-source NoSQL databáze optimalizovaná pro rozsáhlé datové objemy s důrazem na dostupnost a škálovatelnost. Díky své peer-to-peer architektuře bez jednoho bodu selhání je vhodná pro IoT aplikace s požadavkem na nízkou latenci a spolehlivost. Cassandra využívá replikaci mezi uzly a mechanismy pro udržení konzistence. Flexibilní ukládání dat do sloupcových struktur je ideální pro záznamy časových řad. Díky podpoře jazyků jako Python nebo Java se snadno integruje do aplikací. V praxi se uplatňuje hlavně v prediktivní údržbě a zpracování dat generovaných chytrými městy (Jyothis, 2022).

2.3 Metody analýzy v prostředí IoT dat

Internet věcí produkuje obrovské množství dat z různých senzorů, chytrých zařízení a průmyslových systémů. Aby tato data měla skutečnou hodnotu, je potřeba je zanalyzovat – to umožňuje zlepšení výkonu zařízení, automatizaci procesů a rychlejší, efektivnější rozhodování. Bez kvalitní analýzy by šlo jen o nestructurovaný tok informací, který by se dal jen těžko využít. Velkou výzvou je proto najít způsoby, jak data ukládat, zpracovávat a vyhodnocovat tak, aby to bylo co nejrychlejší, efektivní a dobře přizpůsobitelné různým oblastem (Mijuskovic a kol., 2021; Andriulo a kol., 2024).

2.3.1 Předzpracování dat pro analýzu a vizualizaci

Předzpracování dat je důležitým krokem při práci s daty z prostředí internetu věcí. Data z IoT zařízení vznikají často v reálném čase, a proto mohou být neúplná, nepřesná nebo obsahovat chyby. Aby je bylo možné dále analyzovat a vizualizovat, je potřeba je nejprve upravit do vhodné podoby. Jak upozorňuje článek (Sirisha a kol., 2023), kvalita těchto dat má zásadní vliv na celý analytický proces. Je zásadní, aby byla data přesná, úplná, konzistentní a relevantní.

V situacích, kdy data pocházejí z různých senzorů a zařízení s odlišnou kvalitou výstupů, je práce s nimi ještě náročnější. Jak popisuje (Das a kol., 2024), v takových případech se vyplatí použít metody, které pomáhají vybírat spolehlivé zdroje dat, spojovat datové toky a zaměřit se

jen na ty informace, které jsou skutečně důležité. Ti navrhují řešení UniPreCIS, které umožňuje zpracovávat data přímo na zařízeních na okraji sítě (edge), díky čemuž se zrychluje celý proces. Předzpracování dat je také důležité při využívání metod strojového učení. Podle Krzysztońa a kol. (2024) může mít například odstranění duplicit, doplnění chybějících hodnot, úprava měřítek apod. zásadní vliv na to, jak dobře bude následná analýza fungovat. Tyto kroky jsou běžnou součástí přípravy dat a jejich správné provedení je nezbytné pro dosažení spolehlivých výsledků.

Kvalita dat je základním předpokladem pro smysluplné využití datové analýzy. Mezi nejdůležitější vlastnosti patří: (Sirisha a kol., 2023)

- **přesnost** – jak dobře data odpovídají reálným hodnotám,
- **úplnost** – zda jsou všechny potřebné informace zaznamenány,
- **konzistence** – jestli spolu data vzájemně souhlasí a neobsahují rozpory,
- **včasnost** – jak aktuální jsou data ve vztahu ke konkrétním potřebám,
- **relevance** – zda jsou dat opravdu užitečná pro zamýšlený účel.

Předzpracování dat v IoT slouží k tomu, aby byla surová data upravena do takové podoby, která je vhodná pro další analýzu nebo vizualizaci. Jde o soubor kroků, jejichž cílem je zvýšit kvalitu a srozumitelnost dat. Mezi nejčastější metody patří následující postupy: (Sirisha a kol., 2023; Krzysztoń a kol., 2024)

a) Čistění dat

Odstraňují se chybné, neúplné či nadbytečné záznamy – např. duplicitní řádky, chybějící hodnoty nebo nesmyslné GPS souřadnice.

b) Úprava formátu dat (transformace)

Data se převádí do jednotné a zpracovatelné formy. Jde například o převod textových kategorií na čísla nebo sjednocení časových údajů.

c) Zjednodušení a výběr dat

Redukce objemu dat bez ztráty podstatných informací – např. odstranění nepotřebných sloupců, výběr záznamů dle časové nebo místní relevance.

d) Sjednocení dat z různých zdrojů (integrace)

Pokud data pocházejí z více zařízení, je nutné je sjednotit. Tento krok zahrnuje sladění formátu, struktury i časových údajů, aby bylo možné s daty pracovat jako s jedním celkem.

e) Diskretizace a vzorkování

Hodnoty se převádějí do kategorií (např. teplota na nízká/střední/vysoká). Vzorkování slouží k výběru menšího, ale reprezentativního souboru dat.

Zavedení těchto kroků má zásadní význam pro zvýšení spolehlivosti a hodnoty analytických a vizualizačních procesů, které jsou na kvalitě vstupních dat silně závislé (Sirisha a kol., 2023; Krzysztoń a kol., 2024).

2.3.2 Přístupy k analýze v prostředí internetu věcí

V prostředí IoT existují dva přístupy k analýze dat: cloud computing a edge computing. **Cloud computing** využívá vzdálené servery, jako jsou Google Cloud, Microsoft Azure nebo AWS, kde dochází k centralizovanému zpracování dat. Tento přístup je ideální pro náročné operace, jako je využití strojového učení nebo práce s velkými objemy dat. Nevýhodou ale může být vyšší latence a nutnost stabilního internetového připojení (Andriulo a kol., 2024).

Naopak **Edge computing** přesouvá analýzu dat co nejbližší k místu jejich vzniku, tedy přímo na IoT zařízení nebo na blízké servery. Díky tomuto přístupu se snižuje latence a data se mohou zpracovávat v reálném čase. Omezení přináší v oblasti výpočetního výkonu zařízení, a proto se často kombinuje s cloud computingem (Mijuskovic a kol., 2021).

Kombinace Cloud a Edge computingu spojuje výhody obou přístupů – edge computing zajišťuje rychlé reakce, zatímco cloud poskytuje výkonné analytické nástroje a dlouhodobé ukládání dat (Mijuskovic a kol., 2021; Andriulo a kol., 2024).

Real-time vs. dávkové zpracování dat

Real-time analýza dat v prostředí internetu věcí hraje klíčovou roli ve všech aplikacích, které vyžadují okamžitou reakci na přijatá data. Patří sem například autonomní vozidla, zdravotnické monitorovací systémy nebo průmyslová automatizace, kde i minimální zpoždění může mít zásadní dopad. Tento přístup umožňuje okamžité zpracování informací, aniž by bylo nutné je nejdříve ukládat a později analyzovat. Díky tomu lze rychle detekovat změny, předcházet poruchám a efektivně reagovat na aktuální situace (Abolhassani Khajeh a kol., 2023).

Real-time analýzu lze rozdělit do dvou hlavních typů, které se liší podle toho, jak přísné jsou požadavky na rychlost zpracování dat. **Hard real-time** – u těchto systémů je okamžité

zpracování dat klíčové. Příkladem mohou být bezpečnostní systémy v autonomních vozidlech nebo důležité zdravotnické přístroje (Malik a kol., 2019).

Soft real-time – tyto systémy sice také vyžadují rychlé zpracování, ale menší zpoždění neohrozí jejich funkcionalitu. Typickým příkladem může být chytrý termostat, které reguluje teplotu v domě. Pokud by reagoval o pár sekund později, nebude to mít kritický dopad (Malik a kol., 2019).

Díky rostoucímu využití edge computingu se real-time analýza stává ještě efektivnější. Data se totiž zpracovávají přímo tam, kde vznikají, což znamená rychlejší reakce a menší zpoždění. Zároveň není nutné posílat obrovské množství dat na cloud, což nejen šetří síťové zdroje, ale také zvyšuje bezpečnost, protože důležité informace zůstávají v místě vzniku a nejsou tolik vystaveny nežádoucím útokům (Abolhassani Khajeh a kol., 2023).

Dávkové zpracování dat je osvědčený způsob práce s velkými objemy dat. Funguje tak, že data se nejdříve nahromadí a pak se zpracují najednou ve větších částech. Ideálním příkladem je Apache Hadoop, populární open-source technologie, která se využívá v cloudu i na vlastních serverech. Výhodou dávkového zpracování je vysoká propustnost a efektivita při analýze velkého množství dat. Na druhou stranu má i svou nevýhodu – protože se data zpracovávají ve větších intervalech, výsledky nejsou dostupné hned (Kolajo a kol., 2019).

2.3.3 Strojové učení a umělá inteligence v IoT

Využití umělé inteligence (AI) a strojového učení v oblasti internetu věcí přináší revoluci v tom, jak se zpracovávají data. Díky těmto technologiím dokážeme automaticky rozpoznávat vzorce, předvídat budoucí chování a efektivně řídit systémy v reálném čase. S rostoucím množstvím dat generovaných IoT zařízeními přestávají tradiční analytické nástroje a metody stačit, z tohoto důvodu se stále více využívají pokročilé AI algoritmy, které umožňují efektivnější zpracování a interpretaci dat (Al-Ansi a kol., 2024).

Umělá inteligence v IoT nachází uplatnění v širokém spektru oblastí, včetně chytrých měst, kde pomáhá optimalizovat dopravu, řídit svoz odpadu a zlepšovat energetickou účinnost (Alahi a kol., 2023). Ve zdravotnictví umožňuje přesnější diagnostiku a personalizovanou péči (Shumba a kol., 2023). V průmyslu pak AI přispívá k prediktivní údržbě strojů a efektivnější výrobě (Ahmmed a kol., 2024).

Například při správě odpadu mohou AI modely předpovědět množství produkovaného odpadu a optimalizovat svozové trasy, čímž se snižují náklady i uhlíková stopa (Lakhout, 2025).

Navzdory výhodám přináší implementace umělé inteligence v IoT i výzvy, zejména v oblasti zabezpečení dat, výpočetních nároků a odezvy zpracování. Proto se stále častěji využívají hybridní přístupy, které kombinují edge computing a cloud AI, což umožňuje rychlé lokální zpracování dat v kombinaci s robustními modely trénovanými v cloudu (Shumba a kol., 2023).

Geoprostorová umělá inteligence (GeoAI), je obor, který propojuje strojové učení a hluboké učení s prostorovými daty, aby bylo možné automatizovaně analyzovat jevy v geografickém prostoru. Jak uvádí Mai a kol. (2025), rozvoj GeoAI je úzce spojen s tím, že dnes máme k dispozici velké množství dat o poloze. Důležitou roli zde hrají data, která kombinují informace o tom, kde a kdy se něco děje. Díky tomu je možné sledovat, jak se určité jevy mění v čase a prostoru.

GeoAI se využívá především tam, kde je potřeba najít opakující se vzory, porozumět vztahům mezi místem a časem nebo předpovídat další vývoj. Praktickým příkladem může být sledování pohybu lidí, analýza dopravních zácp nebo předpovídání rozšíření znečištění v konkrétní oblasti (Marasinghe a kol., 2024).

2.4 Výzvy v oblasti sběru, ukládání a analýzy IoT dat

S rostoucím rozvojem IoT se významně zvyšuje množství dat generovaných a přenášených mezi propojenými zařízeními. Na druhou stranu ale sběr IoT dat přináší také několik zásadních výzev, které je třeba řešit, aby byla zajištěna spolehlivost, bezpečnost a udržitelnost těchto systémů. Mezi nejvýznamnější výzvy patří: (Ahmad a kol., 2021)

- zabezpečení a ochrana soukromí dat,
- škálovatelnost systémů pro zvládnutí obrovských objemů dat (viz kap. 0),
- otázky kompatibility a standardizace mezi různými zařízeními (viz kap. 2.4.3),
- energetická efektivita IoT zařízení (viz kap. 2.4.4).

2.4.1 Zabezpečení a ochrana soukromí dat

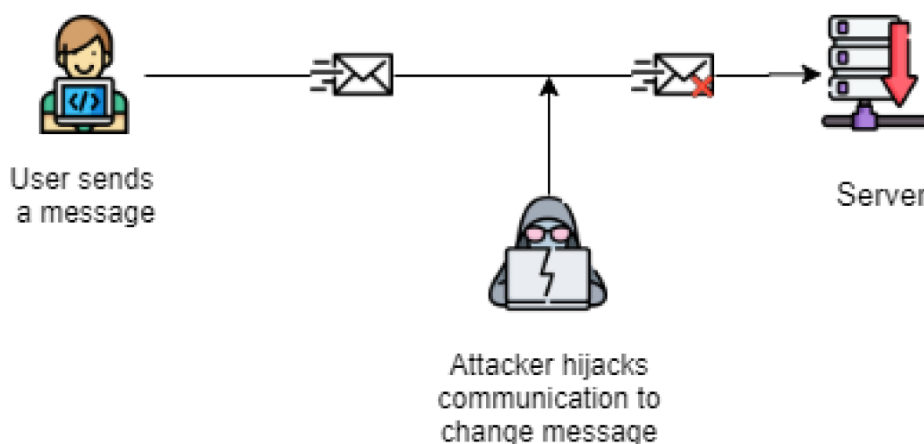
V prostředí IoT založeném na cloud computingu je zajištění bezpečnosti a ochrany soukromí dat klíčovou výzvou. Data jsou často přenášena mezi různými zařízeními a cloudovými službami, což zvyšuje riziko jejich zneužití. Mezi největší hrozby patří neoprávněný přístup, útoky typu Denial of Service (DoS) a narušení integrity dat. DoS útoky představují vážné nebezpečí, protože útočník může zahltnout systém obrovským množstvím požadavků, což může vést k vyřazení zařízení z provozu a narušení služeb (Ahmad a kol., 2021).

Bezpečnostní opatření

Abychom těmto útokům předešli, je zásadní implementovat kvalitní bezpečnostní opatření, jako je **monitorování síťového provozu**, které pomáhá detekovat neobvyklé vzorce a potencionální hrozby v reálném čase. Doporučuje se také **použití bezpečnostních filtrů**, které dokážou blokovat podezřelý provoz, a technologií, jako jsou **firewally**, jež mohou rozkládat zátěž a bránit přetížení systému (Ahmad a kol., 2021).

Specifické hrozby a prevence

Jednou ze specifických hrozeb je tzv. „Man-in-the-Middle“ (MitM) útok (viz Obrázek 2), kdy útočník získává přístup k datům tím, že se vkládá mezi komunikující zařízení a narušuje přenos informací. K prevenci takových útoků je nutné implementovat silné šifrování přenosů dat a zabezpečit síť pomocí bezpečnostních protokolů, jako je TLS (Transport Layer Security). Nezbytné je také vzdělávání uživatelů o bezpečnostních postupech a zavádění bezpečnostních politik, které podporují odpovědné používání IoT zařízení (Ahmad a kol., 2021).



Obrázek 2: Grafické zobrazení tzv. "Man-in-the-middle" útoku

Zdroj: (Ahmad a kol., 2021)

Dalším důležitým aspektem, který souvisí se zabezpečením IoT dat, je jejich publikace prostřednictvím otevřených datových platform. Někteří poskytovatelé se obávají možného zneužití nebo nesprávné interpretace zveřejněných dat, což může vést k omezené ochotě data sdílet nebo k jejich poskytování v nižší kvalitě. Tyto obavy je proto nutné zohlednit při navrhování systémů pro správu a sdílení otevřených IoT dat (Pavlík a kol., 2020).

2.4.2 Škálovatelnost systémů

Škálovatelnost je zásadní při sběru dat v systémech internetu věcí, zejména protože počet připojených zařízení a objem dat neustále roste. Jak uvádí studie Saidu a kol. (2025), tradiční

centralizované systémy často narážejí na limity, což může způsobit zpomalení sítě a snížení jejího výkonu.

Možná řešení pro škálovatelnost

Řešením může být integrace blockchainu, který poskytuje decentralizovanou strukturu pro bezpečné a efektivní ukládání i zpracování dat. Díky této decentralizaci se zátěž rovnoměrně rozprostře mezi více uzlů v síti, což zlepšuje škálovatelnost celého systému. Navíc, pokud se blockchain zkombinuje s umělou inteligencí, je možné ještě lépe optimalizovat sběr i analýzu dat, což dále posiluje efektivitu a škálovatelnost IoT systémů (Saidu a kol., 2025).

2.4.3 Standardizace a kompatibilita zařízení

S rostoucím počtem zařízení a různých platform se objevuje **výzva v podobě standardizace**. IoT ekosystém zahrnuje obrovské množství zařízení od různých výrobců, které často používají odlišné komunikační protokoly a datové formáty. To může ztížit jejich vzájemnou kompatibilitu a efektivní spolupráci. Pokud chybí jednotné standardy, může to vést k problémům s propojením, bezpečností a celkovou správou dat. Proto je důležité zavést sjednocené rámce a protokoly, které zajistí hladký a spolehlivý tok informací napříč celým IoT systémem (Saidu a kol., 2025).

2.4.4 Energetická efektivita a napájení IoT zařízení

Zařízení IoT se často používají v odlehlých oblastech s omezeným přístupem k energii, a proto je klíčová efektivita jejího využití. Většina zařízení spoléhá na baterie nebo na obnovitelné zdroje, přičemž prodloužení jejich životnosti je zásadní pro provozní i ekonomickou udržitelnost (Pandey a kol., 2024).

Moderní zařízení proto využívají dostupné zdroje co nejúspornějším způsobem – minimalizují spotřebu a prodlužují provozní dobu. Jednou z hlavních strategií je nasazení pokročilých **bateriových systémů**, které zajistí stabilní napájení i v náročných podmínkách. Hlavní omezení baterií spočívá v omezené kapacitě a nutnosti častého dobíjení, což zvyšuje náklady na údržbu. Stále větší pozornost se proto věnuje inovativním technologiím, které tyto limity překonávají. (Pandey a kol., 2024).

Solární panely jsou efektivní formou napájení zejména v oblastech s dostatečným slunečním zářením. Umožňují konverzi solární energie na elektřinu a nacházejí uplatnění například v zemědělství, průmyslu či při monitoringu odlehlých lokalit. Kombinace solárních panelů a inteligentních systémů řízení energie zvyšuje efektivitu a umožňuje optimalizaci (Phung a kol., 2017).

Piezelektrické materiály představují základní technologii pro napájení zařízení IoT v prostředí charakterizovaném přítomností mechanických vibrací, tlaku nebo různých forem mechanické energie. Tyto materiály mají schopnost přeměnit mechanické namáhání na elektrickou energii, což je činí výjimečně vhodnými pro autonomní systémy v průmyslových kontextech. (Shirvanimoghaddam a kol., 2019).

Další možností je **využití rádiové frekvence (RF)**, která přeměňuje rádiové vlny (např. Wi-Fi nebo mobilní signály) na elektrickou energii. Tato technologie je vhodná zejména pro městské oblasti s hustým pokrytím signálem a umožňuje napájet IoT zařízení i tam, kde nejsou dostupné mechanické zdroje energie (Shirvanimoghaddam a kol., 2019).

2.4.5 Energetická účinnost komunikačních protokolů

Energetická účinnost komunikačních protokolů má zásadní význam pro zajištění trvalé funkčnosti zařízení IoT, zejména těch, které jsou závislé na baterii nebo alternativních zdrojích energie. Uvážlivý výběr protokolu může hluboce ovlivnit energetický výdej spojený s přenosem dat, čímž se prodlouží provozní životnost zařízení. Protokoly, které jsou nejčastěji využívány na aplikační vrstvě sítí IoT, zahrnují MQTT, MQTT-SN, CoAP a HTTP. Tyto běžně používané protokoly byly podrobeny analýze z hlediska spotřeby energie v prostředí simulovaném pomocí nástrojů COOJA a Powertrace na operačním systému Contiki: (Shahrokhi a Ahmadi, 2023)

- **MQTT (Message Queuing Telemetry Transport):** Protokol založený na TCP/IP, usnadňující spolehlivou komunikaci mezi zařízeními, avšak s vyšší energetickou náročností kvůli trvalému připojení.
- **MQTT-SN (Message Queuing Telemetry Transport For Sensor Networks):** Odlehčená verze MQTT pro senzorové sítě, eliminuje závislost na TCP/IP a je vhodná pro prostředí s nízkou spotřebou energie.
- **CoAP (Constrained Application Protocol):** Lehký protokol využívající UDP, optimalizovaný pro zařízení s omezenými zdroji a vhodný pro přenos malých datových paketů.
- **HTTP (HyperText Transfer Protocol):** Tradiční, ale energeticky náročný a pro IoT málo vhodný kvůli své složitosti a vysoké spotřebě.

Simulace ukázaly, že **MQTT-SN** nabízí nejlepší výsledky z hlediska úspory energie, což z něj činí vhodnou volbu pro senzory. **CoAP** dosahuje rovněž dobrých výsledků díky využití UDP.

Naopak **MQTT** a zejména **HTTP** vykazují vyšší spotřebu, což je činí méně vhodnými pro energeticky úsporné aplikace IoT. (Shahrokhi a Ahmadi, 2023).

Jedním ze způsobů, jak se vyrovnat s uvedenými výzvami, je širší využívání **otevřených datových zdrojů**. Ty mohou zjednodušit přístup k datům, snížit náklady a usnadnit jejich integraci i následnou analýzu. Jak však upozorňuje (Pavlík a kol., 2020), i otevřená data v České republice narážejí na určité překážky – patří mezi ně například roztržitost zdrojů, nejednotné formáty nebo kolísající kvalita a úplnost.

2.5 Otevřená data – jejich zdroje dat a práce s nimi

Otevřená data představují takový typ informací, které jsou veřejně dostupné a každý je může dále používat, upravovat a sdílet bez výrazných právních nebo technických omezení. Obvykle jsou poskytována ve formátech, které umožňují jejich snadné zpracování počítačem, což usnadňuje jejich opakované využití napříč různými oblastmi – ať už jde o státní správu, soukromý sektor nebo jednotlivé občany (Boček a kol., 2012).

Hlavním cílem otevřených dat není jen zpřístupnit informace samotné, ale podpořit jejich širší využití nad rámec původního účelu. Díky tomu mohou otevřená data přispět k vyšší transparentnosti veřejných institucí, podpoře inovací a k lepší informovanosti při rozhodování. Data tak mohou sloužit jako základ pro tvorbu nových služeb, aplikací či produktů, které staví na veřejně sdílených údajích (Boček a kol., 2012).

Formáty a dostupnost dat

V rámci IoT systémů existuje široká škála používaných datových formátů. Nejpoužívanějšími jsou zejména JSON (JavaScript Object Notation), XML (eXtensible Markup Language) a CSV (Comma-Separated Values), které jsou známé svou jednoduchostí a širokou podporou v různých programovacích jazycích. Pro potřeby IoT aplikací, kde je nutné minimalizovat objem přenášených dat, jsou rovněž využívány kompaktnější formáty, jako například CBOR (Concise Binary Object Representation) nebo Protocol Buffers. V oblasti IoT není k dispozici jednotný standard datového formátu napříč různými zařízeními a platformami. To může vést k problémům při integraci systémů, kdy je nutné zajistit převody mezi různými formáty a zajistit kompatibilitu. Různorodost formátů zároveň reflektuje potřeby jednotlivých aplikací – například potřebu lidské čitelnosti dat (JSON, XML) versus požadavek na nízkou přenosovou režii v omezených sítích (Domínguez-Bolaño a kol., 2022).

Příklady open data portálů a datových zdrojů pro IoT

V rámci České republiky patří mezi hlavní poskytovatele otevřených dat zejména města Praha, Plzeň a Brno. Nejvíce datových souborů je dostupných v Praze (237 souborů), následovanou Plzní (163) a Brnem (149). Ačkoliv Praha nabízí nejširší spektrum otevřených dat, z pohledu počtu aplikací, které na těchto datech staví, je úspěšnější Brno – zde bylo evidováno 78 aplikací, zatímco v Praze pouze 10 (Česko v datech, 2019).

Mezi typické příklady aplikací založených na otevřených datech patří například přehledy kulturních a sportovních akcí, mapy parkovacích míst, monitoring kvality ovzduší nebo aplikace zaměřené na hospodaření městských částí. Tyto datové sady a aplikace tvoří důležitý základ pro rozvoj konceptu chytrých měst a mohou být dále využity také v projektech zaměřených na internet věcí, například v oblasti dopravy, životního prostředí nebo veřejných služeb (Česko v datech, 2019).

Integrace dat z různých open sources

Integrace dat v prostředí internetu věcí představuje jednu z klíčových výzev, a to především kvůli různorodosti použitých zařízení, datových formátů a komunikačních protokolů. Text upozorňuje, že v praxi se běžně kombinují zařízení a aplikace využívající různé protokoly, jako jsou MQTT (Message Queuing Telemetry Transport), CoAP (Constrained Application Protocol) nebo HTTP. Pro vyřešení tohoto problému je nezbytné využívat takzvané IoT gateway zařízení, která zajišťují převod dat z různých zdrojů do sjednoceného formátu vhodného pro další zpracování. Kromě samotné transformace dat mohou gateway zařízení zároveň provádět i další funkce, například filtrování, agregaci přenášených dat nebo zajištění základní úrovně bezpečnostních kontrol. (Domínguez-Bolaño a kol., 2022).

Výhody a limity využití open dat v IoT

Otevřená data představují významnou příležitost jak pro veřejný sektor, tak pro podnikatelské prostředí. Jejich klíčovou výhodou je volná dostupnost pro všechny uživatele a skutečnost, že jejich hodnota roste s každým dalším využitím. Otevřená data podporují vznik nových služeb a aplikací, stimulují inovace a přispívají ke zvýšení transparentnosti činnosti veřejné správy. Zároveň mohou vést ke snížení provozních nákladů státních institucí a otevírají nové možnosti v různých odvětvích, například v oblasti dopravy, zdravotnictví nebo finančních služeb (Boček a kol., 2012).

V oblasti internetu věcí nabízejí síťově propojené senzory účinnou alternativu ke klasickým zdrojům dat, protože umožňují sběr lokalizovaných a aktuálních informací přímo z terénu.

Mezi hlavní výhody patří vysoká přesnost měření a možnost flexibilního nastavení frekvence sběru dat podle konkrétních potřeb projektu. Nasazení vlastních IoT sítí však současně přináší určité nároky – zejména vyšší počáteční investice a nutnost odborného nastavení a správy zařízení. Výzvou je také dostupnost pokrytí v odlehlých oblastech, kde je často zapotřebí individuálních technických řešení (Pavlík a kol., 2020).

Jak upozorňuje (Hladík, 2024) ve své disertační práci, otevřená data sice poskytují užitečný základ pro vizualizace a prostorové analýzy, jejich praktické využití ale často naráží na určitá omezení – například co se týče úrovně detailu, aktuálnosti nebo úplnosti. Na příkladu analýzy dopravních přestupků v Brně autor ukazuje, že výsledky založené pouze na otevřených datech mohou nabídnout spíše zjednodušený pohled na realitu, protože často chybějí doplňující informace o okolnostech konkrétních událostí. Přesto ale dochází k závěru, že i data s omezenou vypovídací hodnotou mohou být cenným zdrojem – pokud si je analytik vědom jejich limitací a přistupuje k nim kriticky.

Možnosti využití syntetických dat

Syntetická data nabízejí alternativní zdroj informací v případech, kdy je získání reálných dat obtížné, rizikové nebo finančně náročné. Umožňují simulovat široké spektrum situací, které by bylo v reálném prostředí těžké nebo nemožné zachytit. Typickým příkladem využití je oblast autonomního řízení, kde syntetické datasey slouží jako doplněk nebo náhrada reálných dat při trénování a testování algoritmů. Výhodou syntetických dat je možnost generovat velké množství přesně anotovaných záznamů, nastavovat podmínky sběru (například různé počasí, denní doby nebo hustotu provozu) a zajišťovat zároveň konzistenci i potřebnou variabilitu scénářů. Tím syntetická data podporují robustnost vyvíjených systémů, urychlují jejich vývoj a umožňují testování v extrémních či málo častých podmínkách, které by jinak bylo obtížné reálně zaznamenat (Song a kol., 2024).

3 ZVOLENÝ PŘÍSTUP K ANALÝZE IOT DAT A VIZUALIZACI VÝSLEDKŮ

3.1 Obecný postup

Internet věcí výrazně proměnil práci s daty – od jejich sběru až po využití ve specifických odvětvích, jako je doprava či energetika. IoT generuje rozsáhlé množství dat, které mohou sloužit jako cenný zdroj pro analýzu a rozhodování, ale zároveň přinášejí nové výzvy, zejména v oblasti jejich zpracování a interpretace. (IoTPORT, 2020).

Aby tato data měla skutečný přínos pro rozhodování, je potřeba je prezentovat v jasné a srozumitelné podobě. Jak uvádí (Al-Ali a kol., 2024), vizualizace IoT dat hraje klíčovou roli při odhalování trendů a problémů, které by ve své surové podobě zůstaly skryté – například v chytrých městech nebo energetických sítích. Vhodná vizualizace umožňuje sledovat aktuální situaci i predikovat vývoj na základě historických dat.

Pro efektivní analýzu je třeba data upravit a zorganizovat. Zpracování zahrnuje několik kroků – od sběru po interpretaci. Vizualizace pak převádí výsledky do přehledné podoby, která usnadňuje jejich pochopení a využití při rozhodování (Al-Ali a kol., 2024).

Dostupnost dat sama o sobě nezaručuje jejich využitelnost. Klíčové je nejen sběr, ale i jejich smysluplné zpracování a interpretace. V práci se postupně vytvářel praktický rámec pro práci s daty, který se osvědčil při různých typech úloh – ať už šlo o modelová nebo reálná data. Nejde o striktně definovanou metodiku, ale spíše o obecný přístup. Základem tohoto přístupu je vždy jasné vymezení problému, který má být pomocí dat řešen, a výběr vhodných datových zdrojů – ať už reálných nebo syntetických. Důležité je i prostorové ukotvení dat, jejich propojení a obohacení o další atributy, které podporují interpretaci. V praxi se osvědčila kombinace kvantitativních metod (např. výpočty, četnosti) s vizualizacemi, které znázorňují nejen hodnoty, ale zejména vztahy, prostorové rozložení nebo časový vývoj. Výsledkem jsou přehledné a analyticky přesné výstupy využitelné při rozhodování – např. při hledání optimálního umístění infrastruktury nebo plánování dalších kroků (Al-Ali a kol., 2024).

Práce s otevřenými daty ukázala, že mohou být velmi užitečné pro různé analýzy. Zároveň se ale ukázalo, že s nimi mohou být i určité problémy – například chybějící nebo neúplné údaje, různý formát dat nebo nedostatek doprovodných informací. Tyto věci mohou ovlivnit kvalitu výsledků a často je potřeba data upravit nebo je spojit s vlastními vstupy. V této práci se jako nejvhodnější řešení ukázala kombinace veřejně dostupných dat a vlastního syntetického modelu.

3.2 Použité nástroje a prostředí pro analýzu dat

Pro zpracování, analýzu a vizualizaci dat byla v rámci této práce využita kombinace několika nástrojů a softwarových prostředí, která společně umožňovala efektivní práci s tabulkovými i prostorovými daty. Volba konkrétní nástrojů byla určena charakterem dostupných dat, typem prováděných analýz a požadavky na výslednou vizualizaci.

Díky kombinaci analytických funkcí v prostředí **Python** (Google Colab) a prostorových nástrojů v **ArcMap** bylo možné propojit zpracování dat, provedení výpočtů i tvorbu kartografických výstupů v jednotném navazujícím pracovním postupu.

3.2.1 Role GIS při analýze prostorových dat

Geografické informační systémy (GIS) jsou komplexní nástroje určené pro sběr, správu, analýzu a vizualizaci dat s geografickou vazbou. Nejedná se pouze o software, ale o propojení technologií, dat, lidí a procesů, které společně umožňují efektivně pracovat s informacemi v prostorovém kontextu. Správné využití možností GIS vyžaduje znalost základních geografických principů, schopnost prostorového uvažování a práci s digitálními daty. Moderní GIS nástroje dnes umožňují nejen analýzu geografických jevů, ale také tvorbu prostorových modelů, predikcí a podporu rozhodovacích procesů v celé řadě oblastí – od životního prostředí a územního plánování až po dopravu a logistiku (DeMers, 2009).

3.2.2 Prostředí Python

Pro analýzu IoT dat bylo využito cloudové prostředí **Google Colab**, které umožňuje spouštět kód v jazyce Python (verze 3.11) přímo v internetovém prohlížeči. Výhodou tohoto řešení je, že není potřeba instalovat žádný software na vlastní zařízení – vše běží online, což usnadňuje práci a umožňuje přístup k projektu odkudkoliv.

Google Colab podporuje řadu knihoven určených pro práci s daty, jejich vizualizaci i prostorovou analýzu. V rámci této práce byly využity například tyto knihovny:

- **pandas** – pro práci s tabulkovými daty,
- **matplotlib a seaborn** – pro tvorbu přehledných grafů,
- **geopandas** – pro zpracování prostorových dat a shapefile souborů,
- **openpyxl** – pro načítání a ukládání dat ve formátu Excel.

Knihovny, které nejsou k dispozici automaticky, lze jednoduše doinstalovat pomocí příkazu *!pip install*, který umožňuje rychlou instalaci přímo v prostředí.

V některých specifických případech, zejména při práci s rozsáhlejšími datovými soubory, jako byl například dataset tweetů, bylo kromě cloudového prostředí využito také **lokální prostředí ve Windows**. Tento přístup umožnil efektivnější manipulaci s velkým objemem dat a snadnější přístup k lokálním adresářovým strukturám. Python byl kromě analýzy dat využit také při vytváření syntetické datové sady pro modelový scénář v logistice.

3.2.3 Prostředí ArcMap

Pro vizualizaci výsledků a práci s prostorovými daty bylo využito desktopové prostředí ArcMap (verze 10.8.2), které je součástí softwarového balíku ArcGIS Desktop. Tento nástroj nabízí široké možnosti pro zpracování vektorových i rastrových dat, včetně pokročilých funkcí pro prostorovou analýzu, vizualizaci a kartografickou úpravu mapových výstupů.

V rámci této práce byly v prostředí ArcMap realizovány například tyto úkony:

- tvorba a editace prostorových vrstev, například bodových dat, liniových tras nebo polygonálních oblastí;
- import a zpracování otevřených geografických dat, jako jsou komunikace, vrstevnice nebo digitální modely reliéfu;
- aplikace prostorových analytických nástrojů;
- finální úprava výstupu do podoby kartograficky čistých map s přehlednou legendou, měřítkem a popisky.

4 UKÁZKOVÉ PŘÍKLADY ANALÝZY DOSTUPNÝCH DAT A VYHODNOCENÍ A VIZUALIZACE VÝSLEDKŮ

Tato kapitola představuje vybrané příklady zpracování a analýzy dat, které čerpají z různých typů zdrojů – od reálných dat ze sociálních sítí až po synteticky vygenerovaný dataset simulující provoz logistické firmy. Cílem těchto případových studií bylo nejen demonstrovat technické možnosti prostorové a časové analýzy, ale zároveň ukázat jejich praktický potenciál při podpoře rozhodování v různých oblastech.

Jednotlivé ukázky zastupují rozdílné přístupy a kontexty – od analýzy chování uživatelů Twitteru na území USA, přes zpracování prostorových dat simulujících digitální infrastrukturu v oblasti Greater Melbourne, až po prostorovou a provozní analýzu tras mezi Hradcem Králové a Trutnovem na základě synteticky vytvořeného datasetu. Každý z příkladů slouží jako ilustrace toho, jak lze konkrétní metody aplikovat v praxi i při práci s daty různé kvality a původu.

Je důležité zdůraznit, že se nejedná o komplexní případové studie, ale o demonstrační přístupy, které lze dále rozvíjet a přizpůsobovat konkrétním potřebám. Ukázky zároveň poukazují na některé limity práce s otevřenými či simulovanými daty – například na rozdílnou kvalitu dat, jejich úplnost, nebo omezenou přesnost některých syntetických datových modelů. Tyto okolnosti jsou běžnou součástí analytického procesu a podtrhují důležitost schopnosti přizpůsobit použitou metodu dostupným datovým možnostem a vyvodit z nich smysluplné závěry.

4.1 Příklad 1 – Časoprostorová analýza uživatelů Twitteru v USA

Cílem této části práce je analyzovat a vizualizovat časoprostorová data pocházející z platformy Twitter (dnes již platforma X), která byla shromážděna během jednoho týdne v lednu 2013 na území Spojených států amerických. Dataset zahrnuje více než 14 milionů tweetů, přičemž každý záznam obsahuje přesnou GPS polohu, časovou značku a informaci o časovém pásmu.

Analýza se zaměřuje na identifikaci denních vzorců aktivity uživatelů, porovnání časových rozdílů mezi regiony a zhodnocení prostorového rozložení tweetů. Výstupy mohou posloužit jako užitečný podklad pro rozhodování v různých oblastech – například při krizovém řízení, plánování dopravní a komunikační infrastruktury nebo modelování chování obyvatelstva.

Zpracování dat probíhá v prostředí Pythonu s využitím specializovaných knihoven, zatímco výsledná vizualizace je realizována v nástroji ArcMap.

Tato část práce slouží jako praktická ukázka toho, jak lze pomocí otevřených velkoobjemových dat a prostorové analýzy vytvářet informačně hodnotné výstupy. Kombinace datové analytiky a geografických informačních systémů zde umožňuje přetvořit surová data v přehledné vizualizace, které mohou najít uplatnění jak ve veřejném sektoru, tak v komerční sféře.

4.1.1 Popis použité datové sady

Dataset je publikovaný na platformě UC Irvine Machine Learning Repository (UC Irvine Machine Learning Repository, 2013) a obsahuje následující sloupce (viz Tabulka 1):

Tabulka 1: Datový slovník uživatelů Twitter

Zdroj: Vlastní zpracování

Sloupec	Popis
Longitude	Zeměpisná délka (GPS souřadnice)
Latitude	Zeměpisná šířka (GPS souřadnice)
Timestamp	Čas odeslání tweetu ve formátu (vždy v časové zóně CST)
Timezone	Kód časové zóny uživatele (1 = Eastern, 2 = Central, 3 = Mountain, 4 = Pacific)

4.1.2 Předzpracování dat

Načtení a kontrola dat

Nejprve byl dataset načten pomocí knihovny pandas, což umožnilo rychlý přehled o jeho struktuře a typech obsažených dat. Klíčovým prvkem byl sloupec *timestamp*, který obsahoval čas odeslání tweetu. Tento čas byl uložen v časové zóně CST (UTC−6), a to bez ohledu na skutečnou polohu uživatele. Pro další časovou analýzu bylo proto nezbytné tuto hodnotu převést na lokální čas odpovídající dané geografické poloze.

Převod časové značky a lokalizace

Nejdříve byly hodnoty ve sloupci *timestamp* převedeny do formátu *datetime* a uloženy do nového sloupce *datetime_cst*. Jelikož původní čas byl v čase CST (UTC−6), byl následně přepočítán na univerzální koordinovaný čas (UTC). Poté byl na základě sloupce *timezone* každému záznamu přiřazen odpovídající časový posun (například +1 hodina pro Eastern Time nebo −1 hodina pro Mountain Time). Tento posun byl využit k výpočtu lokálního času (*local_datetime*), který byl klíčový pro správnou časovou analýzu napříč různými oblastmi USA.

Z přepočítaného lokálního času byly dále odvozeny dvě nové proměnné:

- *local_hour* – hodina odeslání tweetu (0–23), použitá při analýze denní aktivity,
- *local_date* – datum ve formátu YYYY-MM-DD, sloužící k agregaci podle jednotlivých dní.

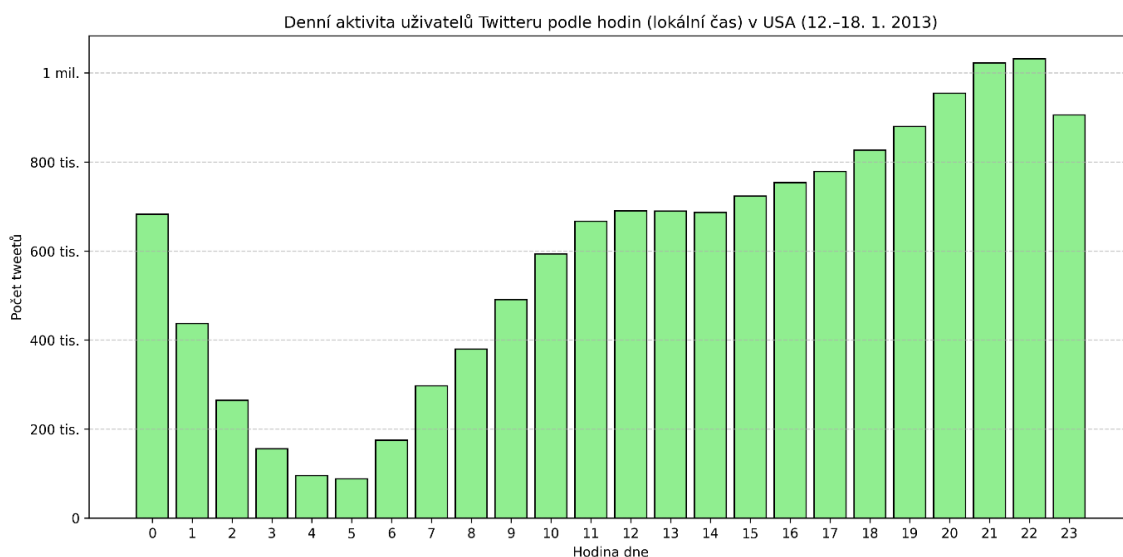
Tyto proměnné tvořily základ pro následné analýzy i vizualizace časových vzorců chování uživatelů. Kompletní Python kód použitý pro převod časové značky a odvození proměnných je k dispozici v příloze (viz Příloha A).

Filtrování okrajových dnů

V datasetu se nacházely také záznamy z 11. a 19. ledna 2013, které ale nepředstavovaly kompletní dny – obsahovaly jen malý zlomek tweetů oproti ostatním dnům. I když tyto záznamy zůstaly zachovány v datech, nebyly zahrnuty do časových vizualizací. Analýza se tedy zaměřila pouze na období od 12. do 18. ledna 2013, které poskytovalo úplný a reprezentativní přehled.

4.1.3 Časová analýza v Pythonu

V úvodní fázi časové analýzy byla sledována aktivita uživatelů Twitteru během jednotlivých hodin dne. Z dostupných dat byl vytvořen souhrn četností tweetů podle proměnné *local_hour*, která udává hodinu odeslání tweetu v lokálním čase uživatele. Výsledkem je sloupcový graf zobrazující aktivitu v časovém rozmezí od 0 do 23 hodin, díky kterému je možné identifikovat neaktivnější části dne (viz Obrázek 3).

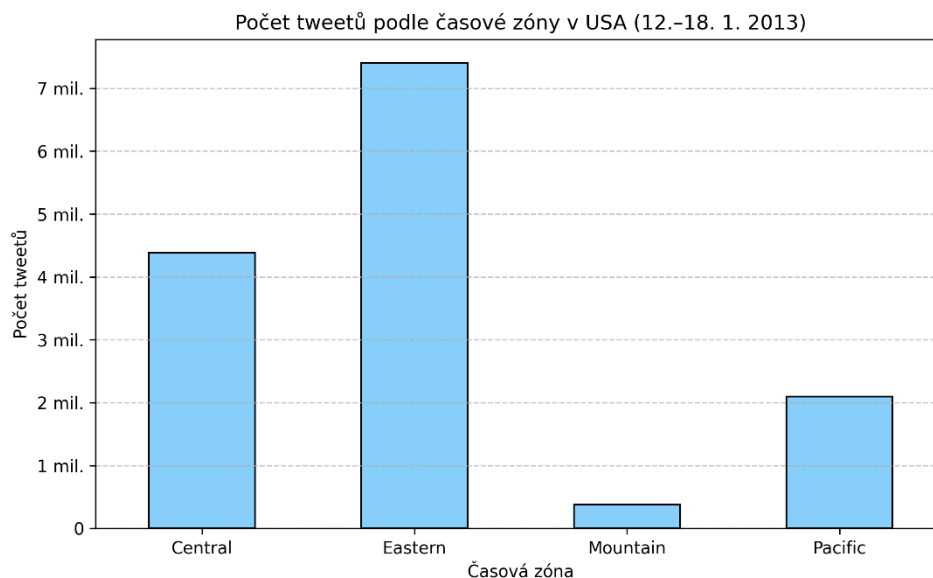


Obrázek 3: Denní aktivita uživatelů Twitteru podle hodin v USA

Zdroj: Vlastní zpracování

Kompletní Python kód použitý pro zpracování grafu denní aktivity je k dispozici v příloze (viz Příloha B).

Dále byly tweetové záznamy rozděleny podle časových zón uživatelů. Agregací celkového počtu tweetů v každé zóně vznikl přehledný srovnávací graf objemu aktivity napříč regiony USA (viz Obrázek 4).



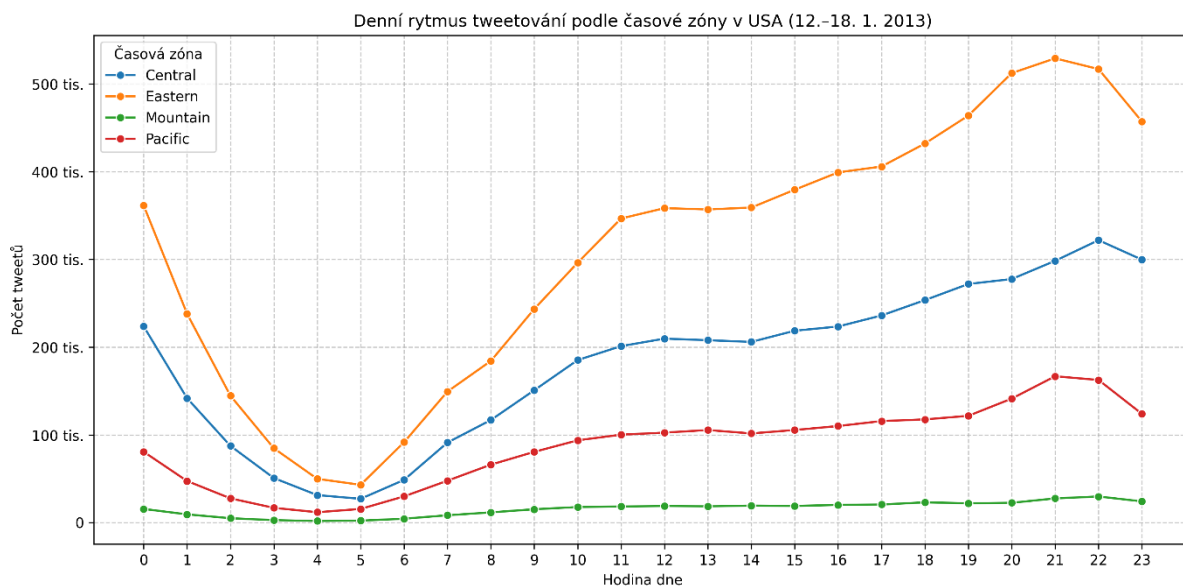
Obrázek 4: Počet tweetů podle časové zóny

Zdroj: Vlastní zpracování

Kompletní Python kód použitý pro zpracování grafu počtu tweetů je k dispozici v příloze (viz Příloha D).

Denní aktivity podle časových pásem

Další krok se zaměřil na porovnání denního rytmu napříč čtyřmi hlavními časovými zónami USA – Eastern, Central, Mountain a Pacific. Pro každé časové pásmo byla spočtena průměrná denní aktivity, tedy průměrný počet tweetů v jednotlivých hodinách dne. Výsledný spojnicový graf ukazuje, zda mají obyvatelé jednotlivých zón podobné návyky v časech, kdy na Twitter nejčastěji přispívají (viz Obrázek 5).



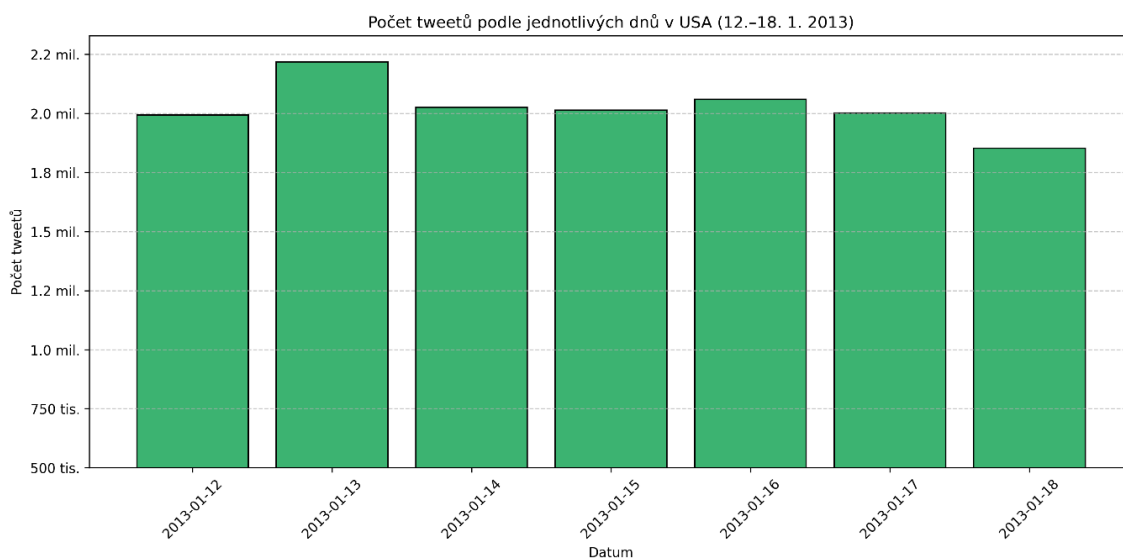
Obrázek 5: Denní rytmus tweetování podle časové zóny v USA

Zdroj: Vlastní zpracování

Kompletní Python kód použitý pro zpracování grafu denní aktivity podle časových pásem je k dispozici v příloze (viz Příloha C).

Aktivita v průběhu dní

Kromě hodinové aktivity byla sledována také denní distribuce tweetů během analyzovaného týdne. Tweetovací aktivita byla seskupena podle proměnné `local_date`, která udává kalendářní den. Tím vznikl přehled o tom, jak se objem sdíleného obsahu měnil v průběhu jednotlivých dnů od 12. do 18. ledna 2013 (viz Obrázek 6).



Obrázek 6: Počet tweetů podle dnů v USA

Zdroj: Vlastní zpracování

Kompletní Python kód použitý pro zpracování grafu počtu tweetů podle jednotlivých dnů je k dispozici v příloze (viz Příloha E).

Shrnutí výsledků časové analýzy

Časová analýza přinesla první důležité poznatky o chování uživatelů Twitteru v průběhu dne i týdne. Ukázalo se, kdy jsou uživatelé nejaktivnější a zda se jejich chování liší v závislosti na časové zóně. Tyto výstupy tvoří základní rámec pro následující prostorovou analýzu, která se bude zabývat geografickým rozmístěním tweetů.

4.1.4 Analýza prostorové distribuce uživatelů Twitteru

Tato kapitola shrnuje postup a nástroje použité pro prostorovou analýzu aktivity uživatelů Twitteru na území USA. Vychází z datasetu obsahujícího přes 14 milionů tweetů publikovaných mezi 12.–18. lednem 2013. Každý záznam obsahuje polohu, čas odeslání a časovou zónu, což umožňuje prostorovou analýzu uživatelského chování. Tyto informace mohou být prakticky využity např. při plánování krizových opatření nebo infrastruktury.

Před zahájením analýzy byl připraven mapový podklad – polygon území USA stažený z ArcGIS Online, který sloužil k vymezení oblasti a prostorovým operacím. Analýza byla rozdělena do dvou částí: nejprve byla vytvořena mapa znázorňující hustotu tweetů za celé sledované období. Následně byla provedena analýza konkrétního časového okamžiku (22:00 východního času), jejímž cílem bylo porovnat celkovou aktivitu s konkrétním časovým stavem.

Celková hustota aktivity během týdne

Prvním krokem prostorové analýzy bylo zobrazení celkové intenzity aktivity uživatelů Twitteru za sledované období. Do prostředí ArcMap byla importována bodová vrstva obsahující všechny tweetové záznamy za období 12.–18. ledna 2013. Pro vizualizaci prostorové koncentrace tweetů byl využit nástroj Kernel Density, který umožňuje přehledně zobrazit místa s vyšší aktivitou v podobě rastrové hustotní mapy.

Parametry analýzy byly nastaveny takto:

- Output cell size: 20 000 metrů
- Search radius: 50 000 metrů

Výsledný rastr byl následně rozdělen do tří kategorií (nízká, střední, vysoká hustota) pomocí klasifikační metody Natural Breaks (Jenks), která zohledňuje přirozené seskupení hodnot v datech. Do mapy byly doplněny administrativní hranice států a názvy vybraných měst, což

napomáhá lepší orientaci ve výsledcích. Tato vizualizace poskytuje celkový přehled o prostorovém rozložení aktivity na Twitteru během jednoho týdne.

Noční aktivita ve východní časové zóně

Výběr časového pásma Eastern Time a konkrétního času 22:00 nebyl náhodný. Vyšel z předchozí časové analýzy, která ukázala, že právě v tomto časovém okně docházelo k nejvyšší intenzitě tweetování. Zaměřením se na období s maximální aktivitou se podařilo zvýšit relevanci výsledné prostorové vizualizace a lépe zachytit chování uživatelů v čase.

V této fázi byla provedena detailnější časoprostorová analýza zaměřená pouze na tuto specifickou kombinaci časové zóny a hodiny. Nejprve byly z celé datové sady vybrány pouze záznamy odpovídající časové zóně Eastern a času 22:00, což představovalo přibližně 500 tisíc tweetů.

Pro tuto vybranou podmnožinu byla opět aplikována metoda Kernel Density, přičemž parametry analýzy byly nastaveny takto:

- Output cell size: 5 000 metrů
- Search radius: 20 000 metrů

Výsledný rastrový výstup byl následně oříznut polygonem, který reprezentuje hranice časové zóny Eastern, a to pomocí nástroje Clip. Výsledná mapa byla doplněna o hranice států spadajících do této oblasti a o malou přehledovou mapku, která znázorňuje polohu v rámci celých Spojených států.

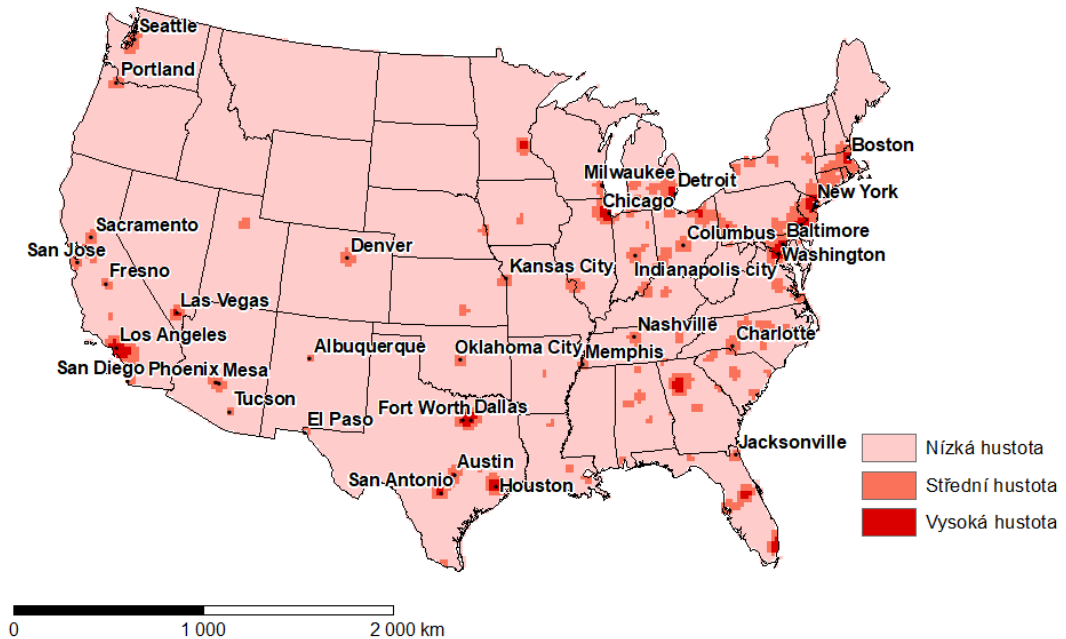
4.1.5 Závěr a interpretace prostorové distribuce uživatelů Twitteru

Tato kapitola shrnuje hlavní poznatky z prostorové analýzy distribuce uživatelů Twitteru na území Spojených států amerických. Analýza byla provedena na základě rozsáhlého datasetu obsahujícího více než 14 milionů tweetů publikovaných v období od 12. do 18. ledna 2013. Hlavním cílem bylo identifikovat oblasti s nejvyšší digitální aktivitou a naznačit prostorové vzorce komunikace.

Výsledky ukázaly, že nejvyšší koncentrace aktivity se soustředila do hustě zalidněných městských oblastí, zejména na severovýchodním pobřeží (New York, Washington, Boston) a v dalších velkých metropolitních centrech jako Chicago, Los Angeles nebo Houston.

Naopak v řídce osídlených oblastech středozápadních a horských států byla zaznamenána nižší intenzita aktivity. Tento jev lze vysvětlit nejen nižší hustotou obyvatelstva, ale také rozdíly v dostupnosti digitální infrastruktury a socioekonomických podmínkách (viz Obrázek 7).

PROSTOROVÁ DISTRIBUCE UŽIVATELŮ TWITTERU V USA (12.- 18. 1. 2013)



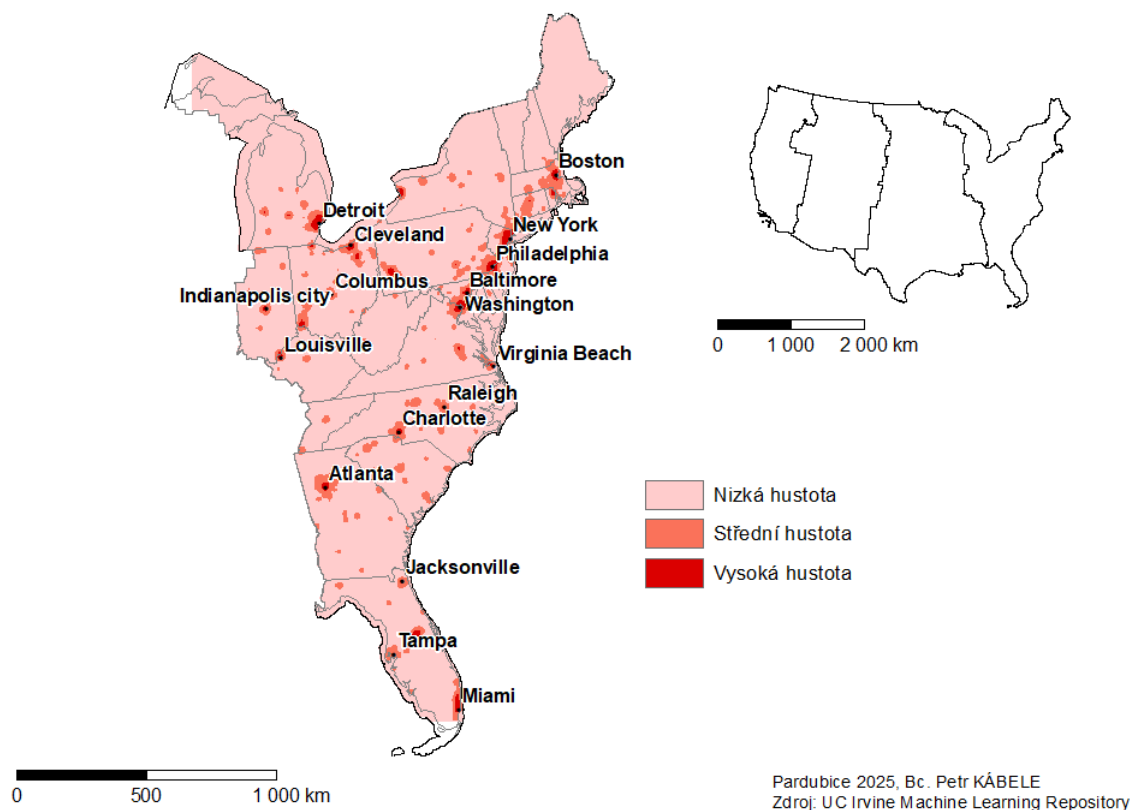
Pardubice 2025, Bc. Petr KÁBELE
Zdroj: UC Irvine Machine Learning Repository

Obrázek 7: Prostorová distribuce uživatelů Twitter v USA

Zdroj: Vlastní zpracování

Následně byla provedena detailnější analýza zaměřená na aktivitu v čase 22:00 východního času (Eastern Time Zone). Tento čas byl vybrán na základě předchozí časové analýzy jako období s nejvyšší aktivitou. Výsledky ukázaly, že v tomto čase se aktivita výrazně soustřeďovala do východní a jihovýchodní části USA. Mezi hlavní centra patřila města jako New York, Washington, Atlanta, Miami a Philadelphia (viz Obrázek 8).

PROSTOROVÁ DISTRIBUCE UŽIVATELŮ TWITTERU V ROCE 2013 (22:00, EASTERN TIME ZONE - VÝCHODNÍ A JIHOVÝCHODNÍ USA)



Obrázek 8: Prostorová distribuce uživatelů Twitter – východ

Zdroj: Vlastní zpracování

Tato analýza ukazuje, jak mohou prostorová data podpořit porozumění digitální komunikaci. Vizualizace prostorových vzorců se uplatní například při plánování rozvoje digitálních služeb, cílení marketingových kampaní nebo při návrhu optimalizace síťové infrastruktury.

Pro hlubší a přesnější analýzu by však bylo vhodné pracovat s detailnějšími daty, například o kvalitě připojení, použitých zařízeních, lokalizační přesnosti nebo demografickém profilu uživatelů. Takové údaje by umožnily nejen přesnější prostorové modelování, ale i lepší porozumění faktorům, které digitální aktivitu ovlivňují. Navíc by bylo možné simulovat zatížení sítí v čase a efektivněji plánovat jejich rozvoj podle reálné poptávky i očekávaných změn v chování uživatelů.

4.2 Příklad 2 – Prostorová analýza vztahu mezi uživateli a servery v oblasti Melbourne

Pro účely této ukázkové analýzy byly zvažovány dva datové soubory popisující rozmístění uživatelů internetu na území Austrálie (GitHub, 2018). V úvodní fázi byl analyzován dataset

users_au.csv, který obsahoval unikátní lokalizace uživatelů napříč celou zemí. Pro detailnější zhodnocení prostorových vztahů v oblasti Greater Melbourne byl však následně zvolen datový soubor *users-melbmetro-generated.csv* s rozmístěnými body uvnitř polygonu metropolitní oblasti.

Tento přístup umožnil přesnější analýzu prostorové dostupnosti infrastruktury a jejího vztahu k přítomnosti uživatelů. Hlavním cílem bylo identifikovat oblasti s vyšší koncentrací uživatelů a posoudit, zda je v těchto lokalitách zajištěna odpovídající síťová podpora prostřednictvím edge serverů.

Je však důležité zdůraznit, že datová sada *users-melbmetro-generated.csv* nepředstavuje reálná data, ale byla vytvořena rovnoměrným generováním bodů na základě polygonu oblasti Melbourne. Tento typ simulace je běžně využíván v případech, kdy nejsou dostupné autentické datové zdroje, a slouží k objektivnímu testování přístupů hodnotících přetížení, pokrytí nebo efektivitu rozmístění digitálních technologií v konkrétním území.

4.2.1 Datová sada uživatelů – users-aus.csv

Tento dataset obsahuje anonymizované informace o poloze uživatelů internetu na základě jejich IP adres. Každý záznam odpovídá jedné IP adrese a jejímu přibližnému umístění. Mezi obsažené údaje patří (viz Tabulka 2):

Tabulka 2: Datová sada uživatelů

Zdroj: Vlastní zpracování

Sloupec	Popis
IP	Anonymizovaná IP adresa uživatele
Latitude, Longitude	Zeměpisná šířka a délka
PostCode	Poštovní směrovací číslo
City	Název města
State	Australský spolkový stát (např. NSW, VIC, WA)
Country	Kód země (ve všech případech „AU“)

I přesto, že dataset neobsahuje časové ani demografické informace, umožňuje vytvořit základní přehled o prostorovém rozmístění uživatelů a identifikovat oblasti s vyšší koncentrací.

4.2.2 Datová sada serverů – site.csv

Druhá využitá datová sada obsahuje údaje o poloze tzv. edge serverů, které představují fyzickou infrastrukturu zajišťující dostupnost digitálních služeb. Každý záznam odpovídá jednomu síťovému zařízení a obsahuje tyto informace (viz Tabulka 3):

Sloupec	Popis
Side_ID	Jedinečný identifikátor serveru
Latitude, Longitude	Souřadnice umístění zařízení
Name	Název nebo adresa dané lokality
State	Australský spolkový stát
Postcode	Poštovní směrovací číslo
Site_precision	Informace o přesnosti určení polohy
Elevation, HCIS_L2	Doplňkové technické údaje (např. nadmořská výška)

Spojením této datové sady s informacemi o uživateli lze analyzovat, do jaké míry odpovídá rozmístění infrastruktury potřebám uživatelů. Výsledky mohou napovědět, zda některé oblasti trpí nedostatečným pokrytím nebo jsou naopak efektivně obslouženy.

Oba datasety byly následně analyzovány a vizualizovány za účelem určení oblastí s vysokou hustotou uživatelů a odpovídajícím rozmístěním edge serverů. Tento krok tvořil podklad pro výběr nejvhodnější oblasti k podrobnější prostorové analýze, která je dále rozvedena v následující kapitole.

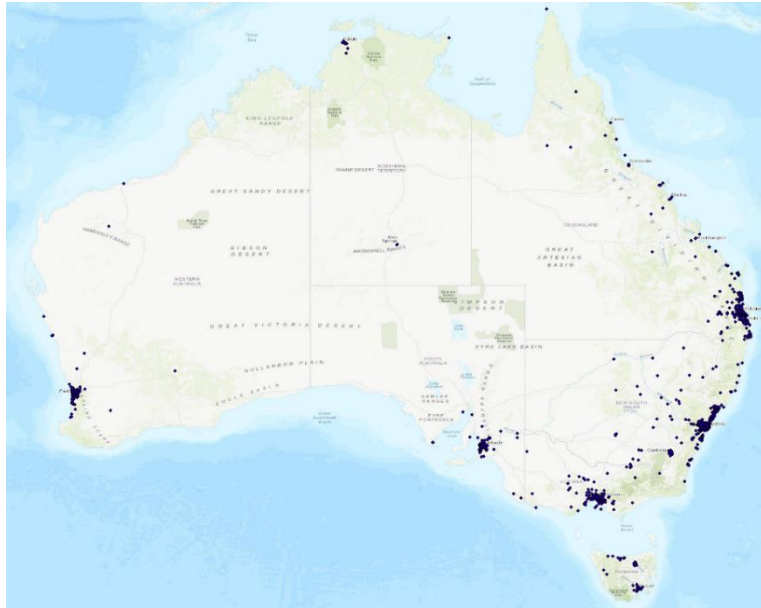
4.2.3 Výběr vhodné oblasti pro prostorovou analýzu

Před samotnou prostorovou analýzou bylo nutné stanovit, na jakou oblast Austrálie se analýza zaměří. Jelikož datová sada obsahuje více než 95 tisíc záznamů o umístění edge serverů a přibližně 4 tisíce záznamů o uživatelských IP adresách, je zřejmé, že datové vzorky jsou velmi nesouměrné. Z tohoto důvodu bylo nezbytné zvolit reprezentativní oblast, která bude zároveň datově dostatečně bohatá a analyticky relevantní.

Prostorová distribuce dat

V rámci předběžné vizualizace byly do prostředí ArcMap načteny dvě samostatné vrstvy – první obsahující informace o uživatelských IP adresách (modře), druhá zobrazující umístění edge serverů (zeleně). Tyto vrstvy byly následně zobrazeny nad topografickou mapou Austrálie, přičemž bylo možné pozorovat následující jevy.

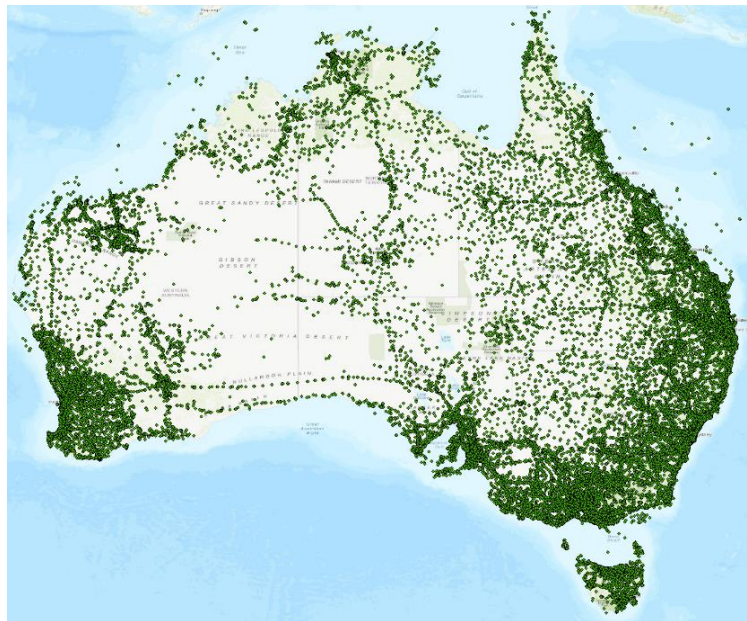
Uživatelé (modré body) jsou koncentrováni především v oblastech jihovýchodního pobřeží, zejména v okolí měst Sydney, Melbourne a Brisbane (viz Obrázek 9).



Obrázek 9: Rozložení uživatelů v prostoru

Zdroj: Vlastní zpracování

Servery (zelené body) pokrývají v podstatě celé území Austrálie, přičemž se rovněž koncentrují v oblastech s vyšší hustotou obyvatelstva (viz Obrázek 10).



Obrázek 10: Rozložení serverů v prostoru

Zdroj: Vlastní zpracování

Kvantitativní porovnání podle států

Pro objektivní výběr oblasti vhodné pro detailní analýzu byla provedena agregace dat podle australských spolkových států. Pro každý stát byly spočítány následující hodnoty:

- Počet edge serverů
- Počet zaznamenaných uživatelů
- Rozloha státu (v km²)
- Počet uživatelů na 1 km²
- Počet serverů na 1 km²
- Poměr serverů vůči uživatelům (čím vyšší, tím vyšší hustota infrastruktury)

Rozlohy jednotlivých států byly převzaty z oficiálních statistik Australian Bureau of Statistics. Výsledky jsou uvedeny v následující tabulce (viz Tabulka 4).

Tabulka 4: Kvantitativní porovnání podle států

Zdroj: Vlastní zpracování

Stát	Servery	Uživatelé	Rozloha (km ²)	Uživatelé/rozloha	Servery/rozloha	Poměr serv./uživ.
NSW	24 571	1 420	800 642	0,00177	0,03068	17,30
VIC	18 163	865	227 444	0,00380	0,07986	20,99
QLD	21 575	816	1 730 648	0,00047	0,01246	26,44
WA	15 931	397	2 529 875	0,00016	0,00630	40,13
SA	7 682	374	983 482	0,00038	0,00788	20,54
TAS	3 213	131	68 401	0,00192	0,04698	24,53
ACT	931	117	2 358	0,04962	0,39483	7,96
NT	3 137	57	1 349 129	0,00004	0,00233	55,05

Výběr vhodné oblasti

Z analýzy hodnot uvedených v tabulce vyplynulo několik důležitých poznatků, které pomohly určit nejvhodnější lokalitu pro podrobnější prostorovou analýzu.

Ačkoliv **Australian Capital Territory (ACT)** vykazuje nejvyšší hustotu serverů i uživatelů na jednotku plochy, ale kvůli malé rozloze a specifickému charakteru (hlavní město Canberra) nebyl pro další analýzu vhodný.

Naopak Northern Territory (NT) byl naopak vyhodnocen jako nejméně vhodný – kvůli nízké hustotě uživatelů i serverů a zároveň nejvyššímu poměru serverů na jednoho uživatele, což svědčí o nerovnoměrné infrastruktuře.

Nejvhodnějším kandidátem pro detailnější zkoumání byla vybrána **Victoria (VIC)**. Má vysokou hustotu uživatelů, rozsáhlou infrastrukturu s hustou sítí serverů a čtvrtý nejnižší poměr uživatel/server ze všech států. Díky přítomnosti velkého města Melbourne je oblast zároveň prakticky i analyticky relevantní. Z těchto důvodů byla pro navazující prostorovou analýzu zvolena právě oblast státu Victoria, konkrétně město Melbourne.

4.2.4 Předzpracování dat pro analýzu oblasti Melbourne

Na základě výsledků kvantitativní analýzy australských států (viz předešlá kapitola) byla pro detailní prostorovou analýzu zvolena oblast státu Victoria (VIC), konkrétně metropolitní oblast města **Melbourne**. Cílem této fáze je připravit obě datové sady – uživatele a servery – tak, aby bylo možné analyzovat jejich prostorovou distribuci a vzájemný vztah.

Výběr datových sad

Pro tuto analýzu byly použity následující datové sady:

- **Uživatelé:** *users-melbmetro-generated.csv* – dataset obsahuje přibližně 130 000 záznamů o rovnoměrně rozložených uživateli v oblasti Melbourne. Je založena na simulaci IP adres a prostorové distribuce, nikoli na reálném sběru. Obsahuje pouze zeměpisné souřadnice (Latitude, Longitude), které jsou však pro účely prostorové analýzy a vizualizace plně dostačující. Takový dataset slouží jako modelový podklad pro testování metod hodnocení prostorového pokrytí digitální infrastruktury, zejména v případech, kdy nejsou dostupná autentická data.
- **Servery:** *site.csv* – kompletní australská datová sada edge serverů (přes 95 000 záznamů), ze které budou vybrány pouze ty záznamy, které geograficky spadají do oblasti Melbourne.

Postup předzpracování dat pro následnou analýzu zahrnoval:

- načtení a zobrazení dat,
- základní kontrolu kvality dat (identifikace duplicitních záznamů, přehled chybějících hodnot),
- filtraci datové sady serverů na oblast Melbourne

Načtení datového souboru a jeho zobrazení

Oba soubory byly do prostředí Google Colab nahrány pomocí modulu *google.colab.files*, který umožňuje snadný import dat přímo ze souborového systému. Po úspěšném nahrání byl dataset

načten jako *DataFrame* s využitím funkce *read_csv()* z knihovny *pandas*, což umožnilo další zpracování a analýzu dat v prostředí Pythonu).

Struktura datové sady uživatelů odpovídá očekávání. Celkem obsahuje 131 312 záznamů, přičemž každý záznam je určen dvojicí hodnot – zeměpisné šířky (*Latitude*) a délky (*Longitude*). Oba sloupce jsou typu *float64*. Výstup k dispozici v příloze (viz Příloha I).

Struktura datové sady *site.csv* obsahuje 95 562 záznamů o umístění edge serverů v Austrálii. Mezi klíčové sloupce patří *LATITUDE* a *LONGITUDE*, *STATE* (stát), *POSTCODE* (PSČ) a další doplňkové technické informace. Výstup k dispozici v příloze (viz Příloha J).

Základní kontrola kvality dat

Následně byla u obou datových sad (*users-melbmetro-generated.csv* a *site.csv*) provedena kontrola kvality dat. Ta zahrnovala přehled chybějících hodnot, identifikaci duplicitních záznamů a ověření, zda se zeměpisné souřadnice nacházejí v očekávaném rozsahu. Pro oba datasey byl použit totožný kód, který umožnil jednotné zpracování.

Při základní kontrole datové sady uživatelů bylo ověřeno, že dataset neobsahuje žádné chybějící hodnoty ani duplicitní záznamy. Sloupce *Latitude* a *Longitude* jsou kompletní a typu *float64*. Hodnoty souřadnic se nachází v rámci reálného území Austrálie – konkrétně zeměpisná šířka se pohybuje v rozmezí od $-38,48^\circ$ do $-37,41^\circ$, zatímco zeměpisná délka od $144,46^\circ$ do $145,74^\circ$. Tato data tak plně odpovídají metropolitní oblasti Melbourne. Výstup k dispozici v příloze (viz Příloha K). U datové sady serverů (*site.csv*) kontrola ukázala, že 291 záznamů postrádá údaj o státě (*STATE*) a 9648 záznamů nemá vyplněno poštovní směrovací číslo (*POSTCODE*). Dále chybí hodnoty také ve sloupcích *SITE_PRECISION* (2 záznamy), *HCIS_L2* (355 záznamů) a výrazné množství chybějících údajů bylo identifikováno ve sloupci *ELEVATION* (71 100 záznamů). Výstup k dispozici v příloze (viz Příloha L).

4.2.5 Analýza prostorového pokrytí v oblasti Melbourne

Datové soubory o uživatelích a serverech (ve formátu CSV) byly importovány do ArcMap. Pomocí nástroje *Display XY Data* byly převedeny na bodové vrstvy podle zadaných souřadnic. Výsledné vrstvy zobrazily prostorové rozložení dat a potvrdily geografické vymezení oblasti Melbourne.

Vymezení zájmové oblasti a ořez dat

Prostorová analýza byla zahájena vymezením cílové oblasti, konkrétně metropolitní zóny Melbourne. Tato lokalita byla zvolena vzhledem k vysoké koncentraci jak uživatelů, tak

serverové infrastruktury, a zároveň pro svou vhodnost k posouzení dostupnosti a rozmístění digitální sítě. Polygon znázorňující oblast Greater Melbourne byl vyhledán prostřednictvím nástroje Search v prostředí ArcGIS Online, následně stažen a importován do ArcMap, kde sloužil jako klíčový prvek pro prostorový výběr.

Pomocí nástroje Clip byly následně upraveny obě hlavní datové vrstvy – tedy vrstva serverů i vrstva uživatelů – tak, aby zahrnovaly pouze záznamy nacházející se uvnitř hranic polygonu Greater Melbourne. Tímto způsobem bylo zajištěno, že veškeré navazující analýzy budou pracovat výhradně s daty odpovídajícími zájmové oblasti.

Analýza hustoty výskytu

Pro detailnější prostorovou analýzu byly na obě datové vrstvy – uživatelů a serverů – aplikovány nástroje hustotní analýzy využívající metodu Kernel Density. Tento postup generuje rastrovou vrstvu, která vizuálně zachycuje koncentraci bodových prvků v prostoru a umožňuje identifikaci oblastí s vyšší či nižší poptávkou po síťové infrastruktuře.

V prostředí ArcMap byl nástroj Kernel Density nejprve použit na vrstvu uživatelů, následně pak na vrstvu serverů. V obou případech byly nastaveny totožné parametry: velikost výstupní buňky 100 metrů a vyhlazovací poloměr 1000 metrů. Tyto hodnoty byly zvoleny tak, aby zajistily vyvážený poměr mezi mírou generalizace a zachycením jemnějších prostorových rozdílů.

Výstupem byly dvě samostatné rastrové vrstvy – jedna zobrazující hustotu výskytu uživatelů, druhá znázorňující prostorovou koncentraci serverové infrastruktury. Pro klasifikaci hodnot byla zvolena metoda Natural Breaks (Jenks). Tyto mapy poskytly důležitý přehled o prostorovém rozložení obou typů prvků a umožnily identifikovat území, kde dochází k disproporcí mezi poptávkou a dostupnými kapacitami.

Analýza vzdálenosti k nejbližšímu serveru

Dalším použitým analytickým nástrojem byla funkce „Near“, která umožňuje vypočítat vzdálenost mezi prvky dvou různých vrstev – v tomto případě mezi každým bodem uživatele a nejbližším bodem reprezentujícím edge server. V prostředí ArcMap byl tento nástroj dostupný ve složce Analysis Tools > Proximity > Near.

Jako vstupní vrstva byla specifikována vrstva uživatelů, zatímco vrstva edge serverů byla zadána jako Near Features. Výpočet byl proveden s výchozím nastavením, tedy bez určení

maximální vyhledávací vzdálenosti. Výstupem analýzy byl nový atribut ve vrstvě uživatelů s názvem NEAR_DIST, který udával vzdálenost k nejbližšímu serveru v mapových jednotkách. Tato metoda umožnila statisticky vyhodnotit, jak daleko se jednotliví uživatelé nacházeli od nejbližšího dostupného serveru. Byly vypočteny základní statistické charakteristiky, jako je průměrná, minimální a maximální vzdálenost.

Prostorová agregace dat

Za účelem kvantitativního vyhodnocení vztahu mezi rozmístěním uživatelů a serverů byla provedena analýza využívající pravidelnou polygonovou mřížku (fishnet). Nejprve byla vytvořena síť složená z polygonů pokrývajících celé území Greater Melbourne. Mřížka byla navržena ve velikosti 10×15 buněk, čímž vznikly prostorové jednotky vhodné pro agregaci bodových prvků.

Na vytvořenou síť byly následně aplikovány obě datové vrstvy – vrstva uživatelů a vrstva serverů. Pomocí nástroje Spatial Join bylo pro každou buňku mřížky vypočteno množství bodů, které do ní prostorově spadají. Získané hodnoty byly zapsány do atributové tabulky jednotlivých polygonů. Pro obě vrstvy byl výsledek zaznamenán ve sloupci Join_Count.

Výsledné vrstvy byly následně tematicky vizualizovány. Pro klasifikaci hodnot byla zvolena metoda Natural Breaks (Jenks), která rozdělila data do přirozených intervalů a umožnila přehledné zobrazení prostorového rozložení bodových prvků v rámci jednotlivých buněk mřížky.

Výpočet poměru uživatelů na server

Na základě prostorové agregace dat bylo možné přistoupit k výpočtu indikátoru vyjadřujícího míru zatížení serverové infrastruktury, konkrétně poměru počtu uživatelů na jeden server. Tento výpočet byl proveden na úrovni jednotlivých buněk fishnet mřížky, v nichž byly předem agregovány hodnoty z obou bodových vrstev.

V rámci atributové tabulky vznikl nový sloupec uzivatele_na_server, který udával počet uživatelů připadajících na jeden server v dané prostorové jednotce. V buňkách, kde nebyl zaznamenán žádný server, byla výsledná hodnota ponechána jako nulová nebo nevyplněná (null), aby nedošlo ke zkreslení celkových výsledků.

Vypočtené hodnoty byly následně klasifikovány pomocí metody Natural Breaks (Jenks) do pěti kategorií a vizualizovány formou tematické mapy. Výsledná vizualizace poskytla detailní

pohled na prostorové rozložení zatížení infrastruktury, přičemž nejvíce přetížené oblasti byly zřetelně zvýrazněny červenou barvou.

Analýza prostorové dostupnosti serverů

Jako doplňující prostorová analýza byla provedena metoda Buffer, jejímž cílem bylo vizuálně i analyticky znázornit dostupnost serverové infrastruktury v rámci definovaného dosahu. Pro každý bod reprezentující edge server byla vytvořena kruhová zóna o poloměru 1000 metrů, čímž vznikla vrstva vyjadřující teoretické pokrytí.

Pomocí nástroje Dissolve byly jednotlivé zóny sloučeny do jednoho polygonu, který znázorňoval souhrnnou oblast pokrytí dané infrastruktury.

Následně byla aplikována funkce Select Layer by Location, jejímž účelem bylo identifikovat uživatele, kteří se nacházejí mimo uvedenou oblast pokrytí. Výsledkem byla samostatná vrstva obsahující pouze ty body, jež neležely v žádném vytvořeném bufferu, a tedy nebyly v blízkosti žádného serveru.

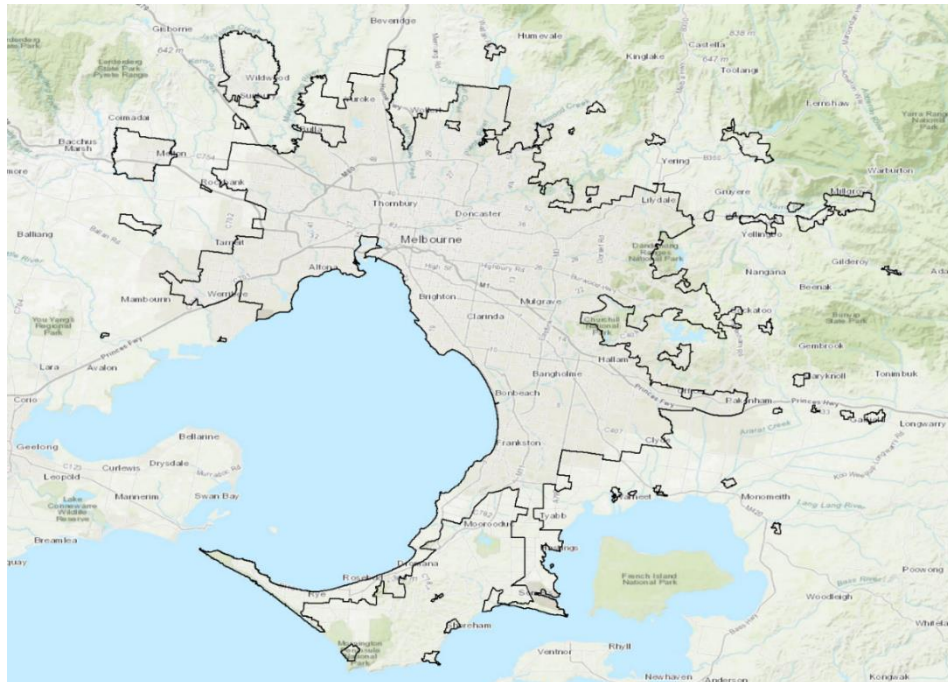
Tato analýza poskytla důležitý podklad pro vyhodnocení lokalit s nedostatečným pokrytím z pohledu koncových uživatelů.

4.2.6 Závěr a interpretace prostorové analýzy oblasti Melbourne

Tato kapitola shrnuje prostorové analýzy provedené nad datovými vrstvami reprezentujícími uživatele a servery v oblasti Melbourne. Po vymezení cílového území prostřednictvím polygonu získaného z ArcGIS Online (viz Obrázek 11) a následném ořezu vstupních dat pomocí nástroje Clip byly aplikovány různé nástroje prostorové analýzy. Tyto postupy umožnily podrobněji porozumět prostorovému rozložení digitální infrastruktury a uživatelského zatížení. Je však třeba zdůraznit, že datová vrstva uživatelů byla vytvořena simulací prostorového rozložení na základě IP rozsahů a modelového přístupu. Přesto lze výstupy považovat za relevantní ilustraci potenciální poptávky v prostředí, kde nejsou k dispozici podrobná reálná data o uživatelích.

Cílem jednotlivých kroků bylo nejen vizuálně a kvantitativně zachytit prostorové vzorce, ale především vytvořit podklad pro rozhodování – například v souvislosti s plánováním rozšíření infrastruktury, optimalizací pokrytí nebo identifikací lokalit s omezeným přístupem ke službám. Výsledky analýzy mají potenciál podpořit optimalizační rozhodnutí z různých perspektiv – z hlediska kvality poskytovaných služeb (zajištění dostatečného pokrytí a snížení přetížení), ekonomické efektivity (cílené investice do oblastí s největším dopadem) i environmentálních aspektů (minimalizace potřeby výstavby nové infrastruktury v oblastech

s dostatečným pokrytím). Tyto poznatky tak mohou výrazně přispět k efektivnějšímu a udržitelnějšímu rozvoji digitální infrastruktury.



Obrázek 11: Vymezení metropolitní oblasti Greater Melbourne

Zdroj: Vlastní zpracování

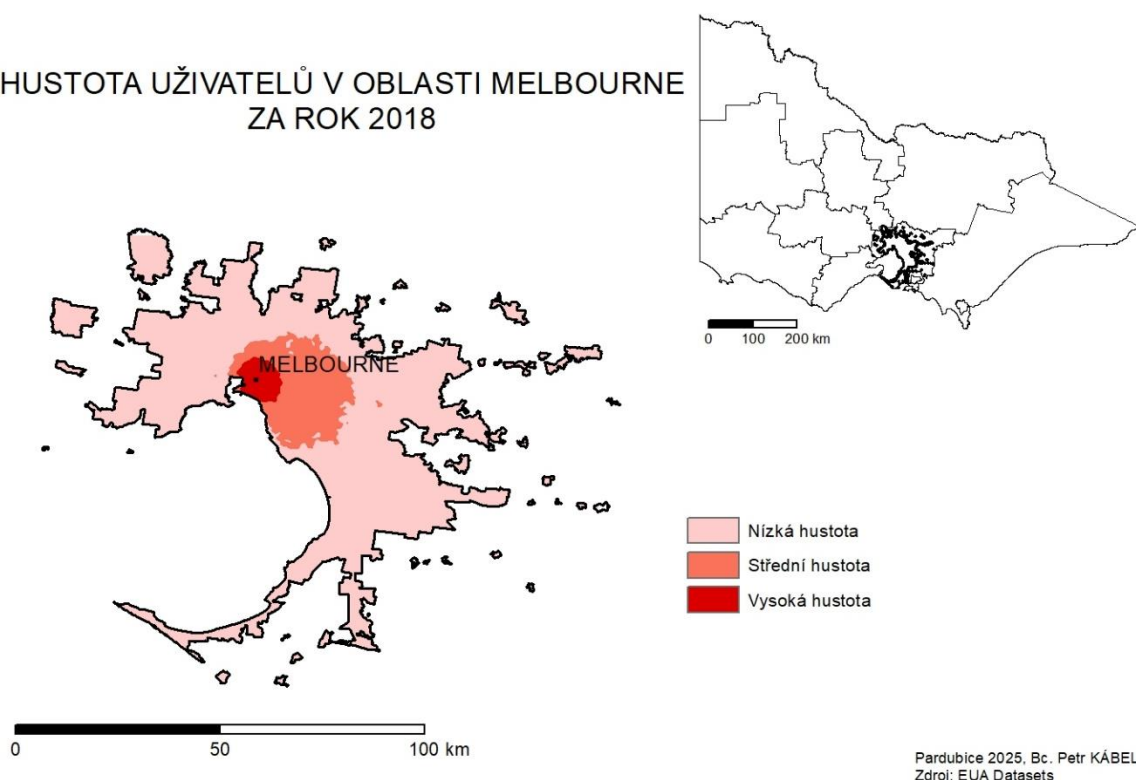
Hustota uživatelů – nástroj Kernel Density

Metoda Kernel Density byla využita k posouzení prostorové koncentrace **uživatelů** v oblasti Greater Melbourne. Tento analytický přístup umožnil identifikovat zóny s nejvyšším výskytem uživatelských bodů a tím lépe porozumět prostorovému rozložení poptávky po digitálních službách. V rámci nastavení analýzy byla zvolena velikost buňky 100 metrů a vyhlazovací poloměr 1000 metrů, což umožnilo zachytit jemné prostorové rozdíly bez ztráty čitelnosti.

Výsledky ukázaly, že nejvyšší koncentrace uživatelů se nachází v centru města Melbourne, což odpovídá vyšší hustotě osídlení i nárokům na dostupnost služeb. Tyto poznatky mohou sloužit jako relevantní podklad pro plánování rozšíření digitální infrastruktury v souladu s lokalizovanou poptávkou.

Naopak v periferních částech metropolitní oblasti byla zaznamenána nižší hustota výskytu uživatelů. Tento jev může souviset s menší hustotou osídlení, ale také může signalizovat nedostatečné pokrytí digitálními službami – například omezenou dostupností vysokorychlostního připojení nebo absencí blízkých serverových uzlů (viz Obrázek 12).

HUSTOTA UŽIVATELŮ V OBLASTI MELBOURNE ZA ROK 2018



Obrázek 12: Hustota uživatelů v oblasti Melbourne (2018)

Zdroj: Vlastní zpracování

Hustota serverů – nástroj Kernel Density

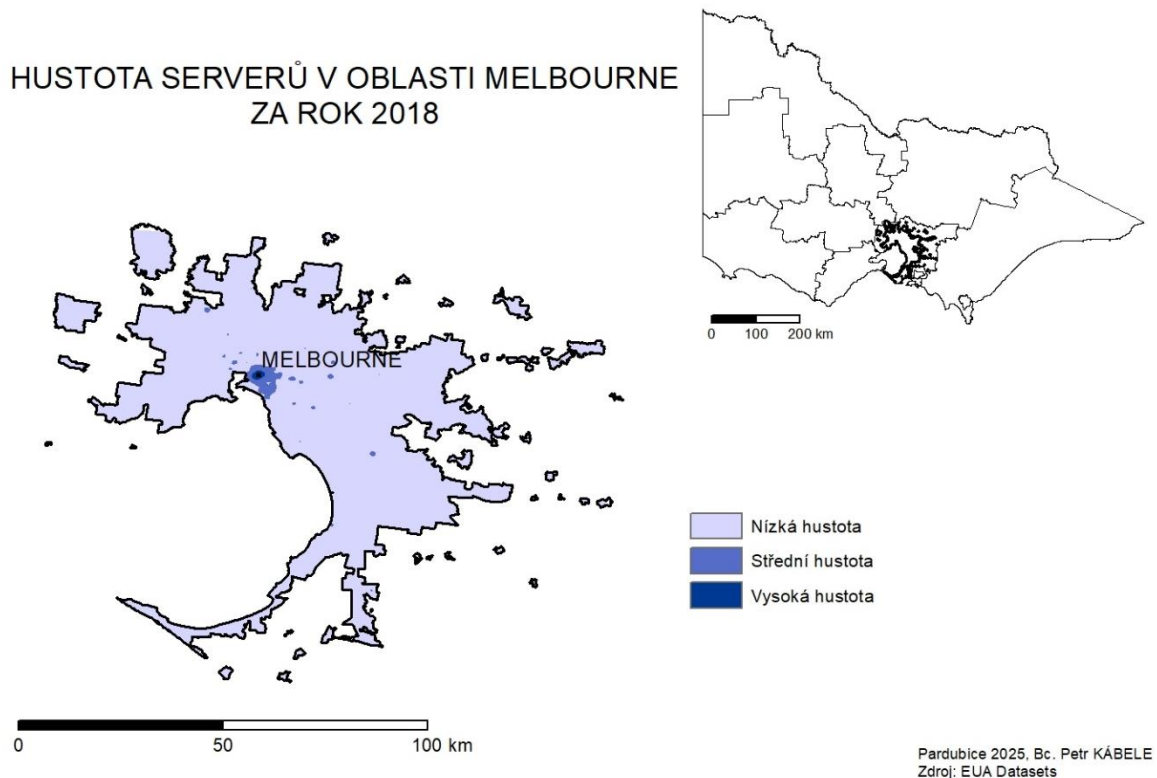
Metoda Kernel Density byla následně aplikována také na vrstvu **edge serverů**, a to za účelem zjištění prostorové koncentrace síťové infrastruktury v oblasti Greater Melbourne. Cílem této analýzy bylo identifikovat oblasti s vyšší mírou infrastruktury a porovnat jejich rozmístění s koncentrací uživatelů. Použitím stejného analytického postupu, jako u dat o uživateli, bylo možné vytvořit srovnatelný výstup, který usnadnil identifikaci případných disproporcí mezi nabídkou a poptávkou po digitálních službách. Parametry analýzy byly opět nastaveny na velikost buňky 100 metrů a vyhlazovací poloměr 1000 metrů.

Výsledky ukázaly, že nejvyšší koncentrace serverů se nachází v centru Melbourne, kde dochází k překryvu s oblastmi s vysokou hustotou uživatelů. Tento stav naznačuje, že infrastruktura v těchto místech do jisté míry odpovídá lokalizované poptávce. Zároveň bylo zjištěno, že serverové uzly jsou rovnoměrněji rozmístěny i do dalších částí metropolitní oblasti, na rozdíl od uživatelů, jejichž koncentrace je spíše lokalizovaná.

V některých územích s výraznou hustotou uživatelů se však ukázala nízká koncentrace serverů, což může naznačovat potenciální problémy s nedostatečným pokrytím nebo přetížením infrastruktury. Naopak v některých periferních částech Melbourne byla zaznamenána

přítomnost serverových uzlů i přesto, že se v těchto oblastech nachází relativně málo uživatelů. Tyto servery mohou fungovat jako záložní body nebo být součástí strategického rozmístění infrastruktury za účelem rovnoměrného rozložení zátěže.

Z hlediska plánování je vhodné dále analyzovat efektivitu takto rozmístěné infrastruktury, a v případě potřeby zvážit její přeskupení nebo posílení v lokalitách s vyšší poptávkou po digitálních službách (viz Obrázek 13).



Obrázek 13: Hustota serverů v oblasti Melbourne (2018)

Zdroj: Vlastní zpracování

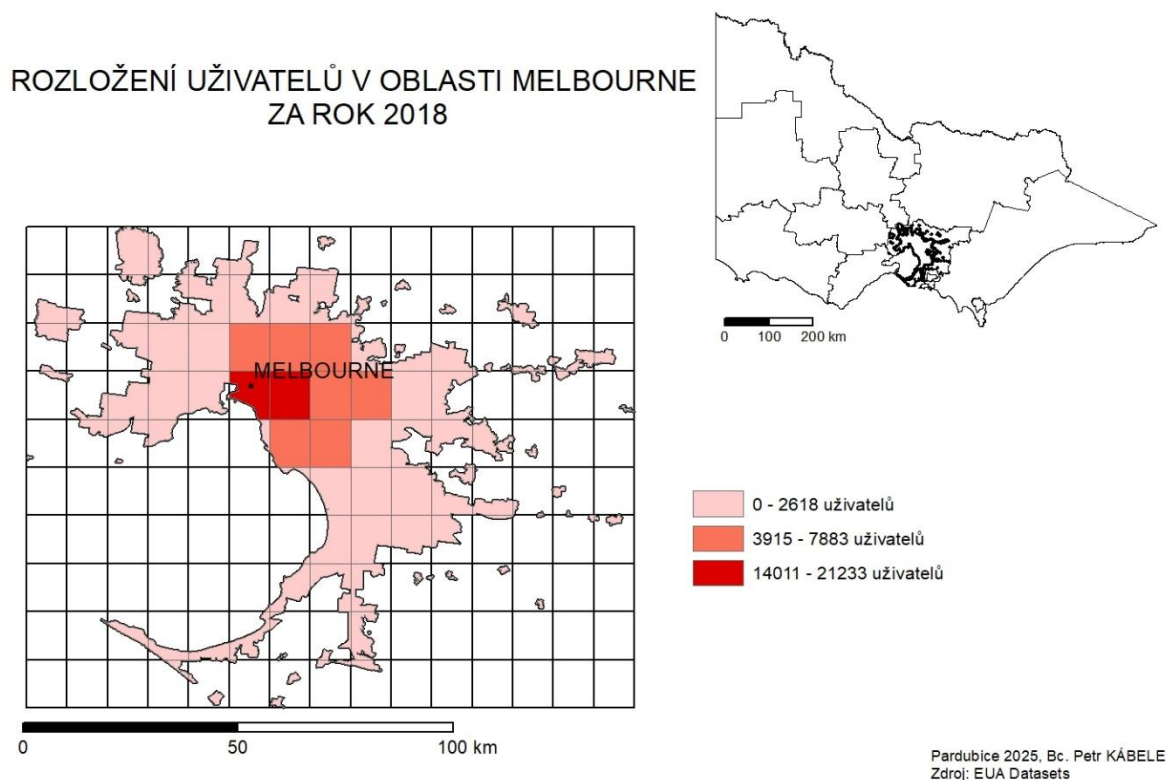
Rozložení uživatelů v jednotlivých buňkách sítě (Fishnet)

Pro detailní zhodnocení prostorového rozložení uživatelů byla oblast Greater Melbourne rozdělena do mřížky o rozměrech 10×15 buněk, a to pomocí nástroje *Fishnet*. Do jednotlivých buněk byly následně pomocí nástroje *Spatial Join* agregovány počty uživatelských bodů podle jejich geografické polohy. Výsledkem této operace byla mapa s kvantitativním přehledem, která ukazuje počet uživatelů v každé prostorové jednotce.

Výstupní vizualizace umožnila přesné porovnání hustoty výskytu uživatelů v rámci celého území. Nejvyšší koncentrace byla zaznamenána ve střední části Melbourne, kde hodnoty přesahovaly 20 000 uživatelů na jednu buňku. Oproti tomu v okrajových částech metropolitní oblasti se počty pohybovaly pod hranicí 500 uživatelů na buňku. Tento kontrast poukázal na

výraznou nerovnoměrnost v rozložení poptávky po digitálních službách – s největším zatížením v centrálních oblastech a výrazně nižší zátěží v periferních částech.

Tato analýza poskytla základ pro pochopení prostorového rozmístění uživatelů. Zjištěné rozdíly v hustotě představují důležitý podklad pro plánování digitální infrastruktury, neboť pomáhají přesněji identifikovat oblasti s nejvyšším potenciálním zatížením serverové sítě. Výsledná vrstva zároveň slouží jako referenční vstup pro navazující analýzy, například při výpočtu poměru uživatelů na jeden server (viz Obrázek 14).



Obrázek 14: Rozložení uživatelů v oblasti Melbourne (2018)

Zdroj: Vlastní zpracování

Rozložení serverů v jednotlivých buňkách sítě (Fishnet)

Prostorová analýza byla rozšířena o kvantifikaci rozložení edge serverů pomocí čtvercové mřížky Fishnet (viz Obrázek 15), která pokrývala celé území Greater Melbourne. Stejně jako v případě uživatelů byla síť tvořena 10×15 buňkami, přičemž do každé z nich byly pomocí nástroje *Spatial Join* agregovány záznamy serverů podle jejich geografické polohy.

Výsledná vizualizace odhalila výrazné rozdíly v rozmístění infrastruktury. Největší koncentrace serverů byla zaznamenána v centrální části Melbourne, kde jedna buňka obsahovala až 2 414 serverů. Další oblasti s vysokou hustotou infrastruktury se pohybovaly

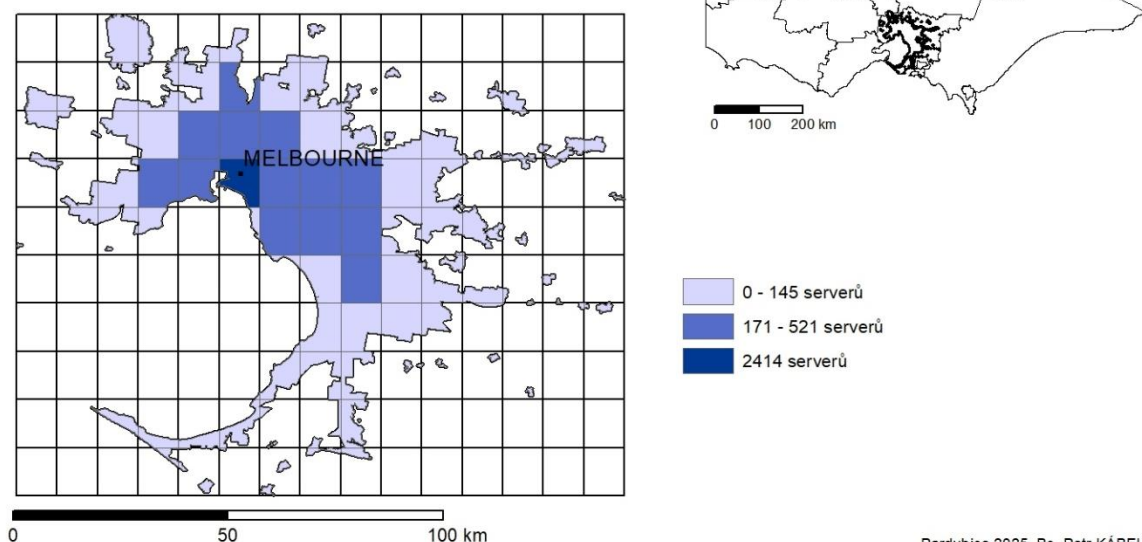
v rozmezí 283 až 521 serverů, zatímco v okrajových částech aglomerace se některé buňky vyznačovaly minimální infrastrukturou – pouze 0 až 18 serverů.

Ve srovnání s výsledky předchozí analýzy uživatelů, kde centrální buňky dosahovaly až přes 20 000 uživatelů, se ukazuje, že v některých oblastech existuje poměrně dobrá shoda mezi poptávkou a nabídkou – tedy vysoká koncentrace uživatelů je vyvážena odpovídající serverovou kapacitou. Přesto však existují místa, kde je tato rovnováha narušena.

Například v několika buňkách východního a jihovýchodního předměstí Melbourne byla zaznamenána vyšší koncentrace uživatelů, ale bez odpovídající infrastruktury. Tyto oblasti mohou být v reálném prostředí vystaveny vyššímu zatížení nebo horší dostupnosti služeb. Naopak v několika buňkách s velmi nízkým počtem uživatelů byla přítomnost serverů relativně vysoká, což může naznačovat různá strategická rozhodnutí nebo rezervu pro budoucí rozvoj.

Tato analýza poukázala na fakt, že rozmístění digitální infrastruktury není vždy přesně sladeno s prostorovým rozložením uživatelů. Zjištěné disproporce mezi poptávkou a nabídkou mohou sloužit jako podklad pro optimalizaci rozmístění edge serverů – ať už formou posílení infrastruktury v přetížených oblastech, nebo přesměrováním kapacit z méně zatížených lokalit.

ROZLOŽENÍ SERVERŮ V OBLASTI MELBOURNE ZA ROK 2018



Obrázek 15: Rozložení serverů v oblasti Melbourne (2018)

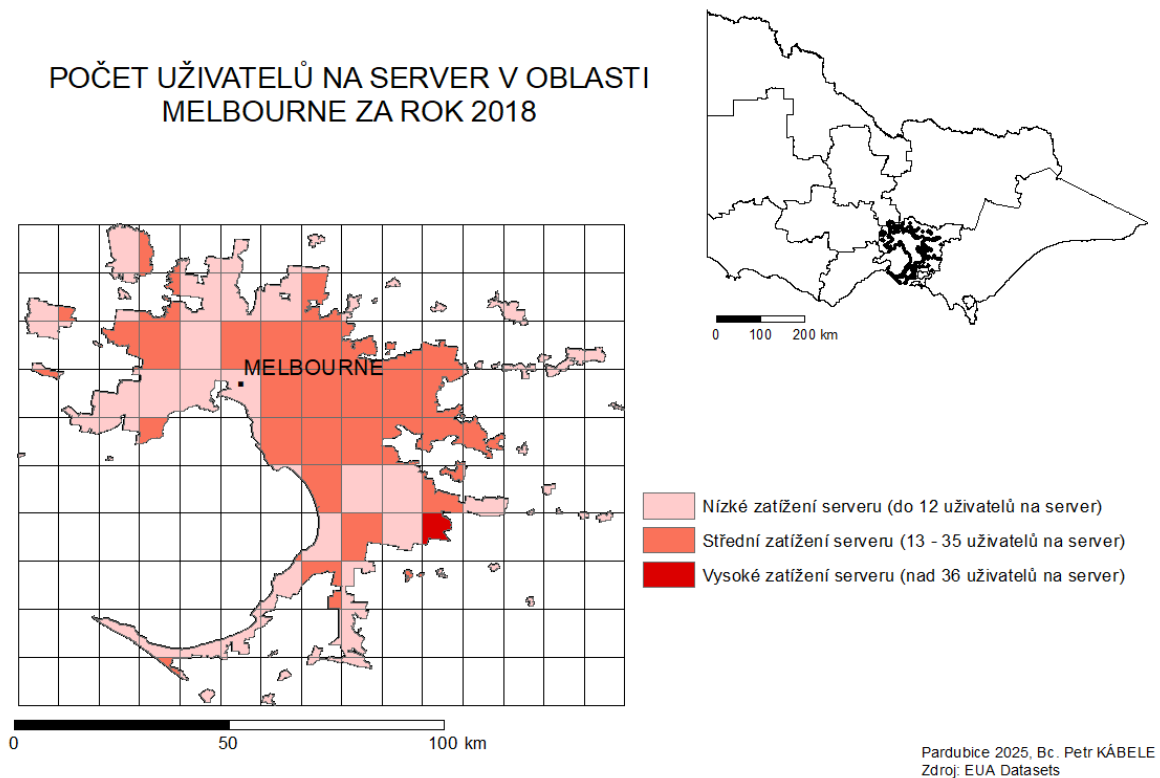
Zdroj: Vlastní zpracování

Poměr uživatelů na server

V další fázi analýzy byl proveden výpočet poměru mezi počtem uživatelů a počtem edge serverů v jednotlivých buňkách fishnet mřížky. Tento ukazatel slouží jako nástroj pro zhodnocení rovnováhy mezi dostupnou infrastrukturou a její skutečnou zátěží. Jinými slovy, poměr uživatelů na jeden server poskytuje přehled o tom, kde může docházet k přetížení sítě a kde je naopak kapacita využívána efektivně.

Výsledky ukázaly, že zatímco některé části oblasti Greater Melbourne vykazují přiměřený poměr – například centrální části města, kde je vysoká hustota infrastruktury – jiné regiony, zejména na jihovýchodním a východním okraji, vykazují výrazné zatížení. V těchto buňkách byl zaznamenán nízký počet serverů, které musejí pokrýt relativně vysoký počet uživatelů. Taková disproporce může v reálném provozu znamenat nižší kvalitu služeb nebo vyšší latenci.

Výsledná tematická mapa ukazující rozložení zatížení poskytla důležitý podklad pro lokalizaci problémových oblastí. Tyto informace mohou být dále využity při plánování rozšiřování infrastruktury – například identifikací konkrétních míst, kde by nasazení nového edge serveru přispělo ke zlepšení dostupnosti digitálních služeb. Zvláštní pozornost by měla být věnována buňkám s nejvyšším poměrem, tedy těm, kde je více než 36 uživatelů na jeden server, neboť právě zde je riziko přetížení nejvyšší. Naopak oblasti se středním zatížením (13–35 uživatelů/server) a zejména ty s nízkým zatížením (do 12 uživatelů/server) vykazují lepší rovnováhu mezi poptávkou a kapacitou infrastruktury (viz Obrázek 16).



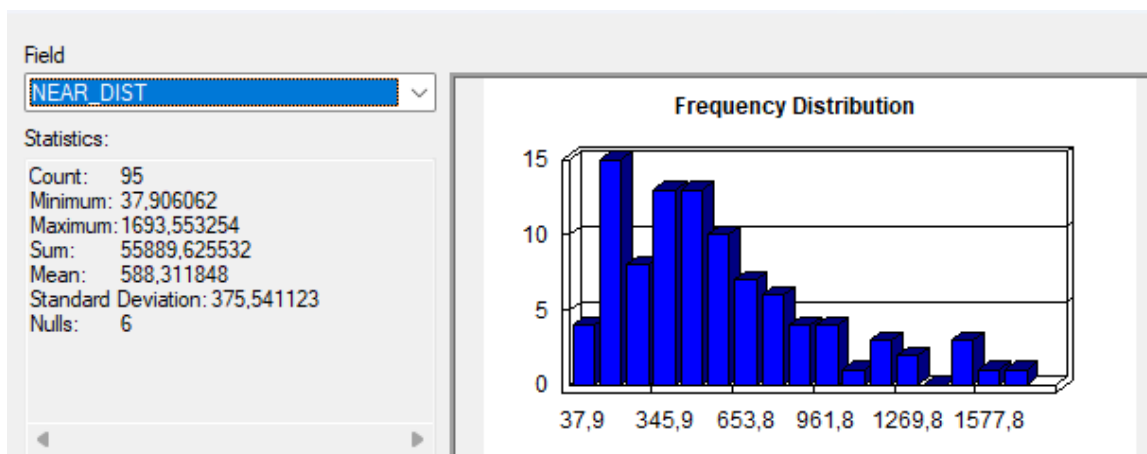
Obrázek 16: Počet uživatelů na server v oblasti Melbourne (2018)

Zdroj: Vlastní zpracování

Analýza dostupnosti – vzdálenost k nejbližšímu serveru

Pro kvantifikaci dostupnosti digitální infrastruktury byla použita funkce *Near*, která pro každý bod uživatele spočítala vzdálenost k nejbližšímu edge serveru. Výsledkem je přehled prostorových rozdílů v rámci oblasti Greater Melbourne.

Na základě histogramu vytvořeného v ArcGIS lze konstatovat, že většina uživatelů má server ve vzdálenosti 500–600 metrů. Průměrná vzdálenost činila 588,3 m, nejkratší 37,9 m a nejdelší přesahovala 1 693 m. Centrum Melbourne vykazuje velmi dobrou dostupnost, což souvisí s vyšší koncentrací infrastruktury. Naopak v okrajových oblastech byly identifikovány delší vzdálenosti, které mohou ovlivnit kvalitu služeb (např. vyšší latence). Tato zjištění je vhodné zohlednit při plánování rozvoje infrastruktury – zejména v oblastech s dlouhým průměrným dosahem ke stávajícím serverům (viz Obrázek 17).



Obrázek 17: Analýza dostupnosti – vzdálenost k nejbližšímu serveru

Zdroj: Vlastní zpracování

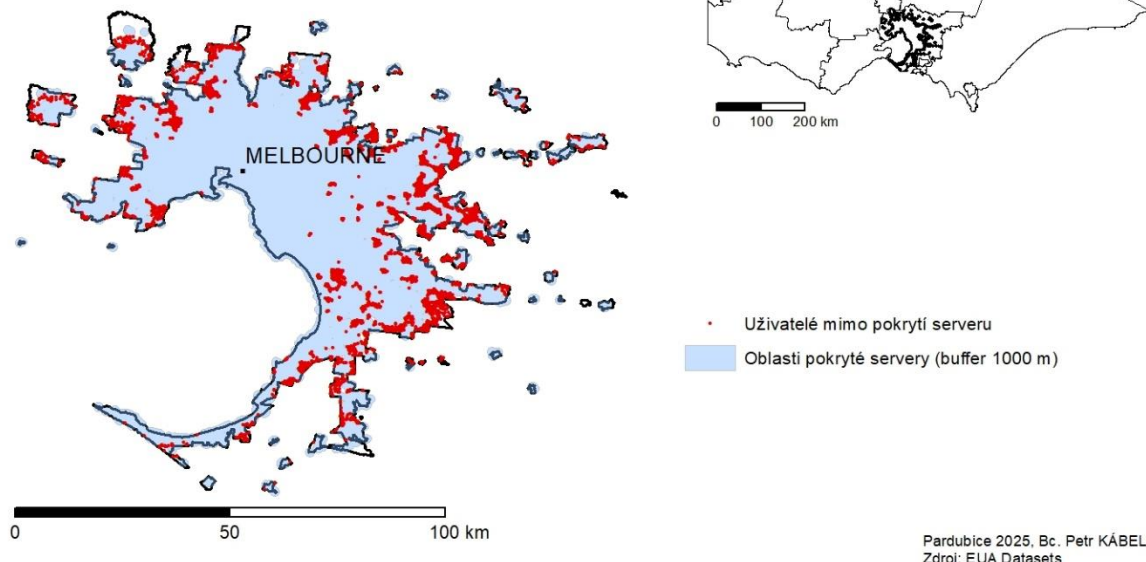
Pokrytí uživatelů ve vzdálenosti do 1000 m od serverů

Tato část analýzy bezprostředně navazuje na předchozí hodnocení dostupnosti serverové infrastruktury pomocí nástroje *Near*, který pro každý bod reprezentující uživatele spočítal vzdálenost k nejbližšímu edge serveru. Zjištěná průměrná vzdálenost činila přibližně 588 metrů, přičemž v některých okrajových částech oblasti Melbourne přesahovala 1 500 metrů.

Pro vizuální a prostorové vyhodnocení těchto výsledků byl dále využit nástroj *Buffer*, který umožnil vytvořit zóny o poloměru 1 000 metrů kolem každého serverového bodu. Do těchto zón byly následně prostorově přiřazeny body uživatelů, čímž bylo možné vyhodnotit, jaká část uživatelské populace se nachází v oblasti s dostatečným pokrytím.

Výsledky ukazují, že většina uživatelů v oblasti Greater Melbourne spadá do pokrytí existující infrastrukturou, zejména v centrálních a hustě osídlených částech města. V některých oblastech, zejména na jihu a jihovýchodě, však analýza odhalila body mimo dosah 1 000 metrů od nejbližšího serveru. Tyto zóny představují potenciálně nedostatečně pokrytá území, která mohou být ohrožena zhoršenou kvalitou připojení. Identifikace těchto míst přináší konkrétní podklad pro strategické rozhodování o rozšíření edge infrastruktury a zlepšení síťové dostupnosti v méně obslužených lokalitách (viz Obrázek 18).

PROSTOROVÁ DOSTUPNOST SERVERŮ Z POHLEDU UŽIVATELŮ V OBLASTI MELBOURNE ZA ROK 2018



Obrázek 18: Prostorová dostupnost serverů z pohledu uživatelů v oblasti Melbourne (2018)

Zdroj: Vlastní zpracování

Tato analýza slouží jako názorná ukázka toho, jak lze využít dostupná prostorová data ke zkoumání rozložení uživatelů a síťové infrastruktury. Použité parametry a analytické metody byly zvoleny tak, aby ilustrovaly možnosti prostorové analýzy, aniž by bylo nutné pracovat s detailními technickými specifikacemi. Výsledky poskytují základní orientaci v prostorových vzorcích rozložení a mohou sloužit jako odrazový můstek pro hlubší analýzy nebo další výzkum.

Pro přesnější hodnocení síťové infrastruktury by však bylo přínosné pracovat s podrobnějšími daty – například o typu připojení, jeho kvalitě, přenosové rychlosti, objemu dat nebo odezvě sítě. Tyto doplňující informace by umožnily posoudit nejen to, kde se infrastruktura nachází, ale také jak dobře funguje v praxi. Díky tomu by bylo možné zaměřit se na konkrétní problémy, jako je přetížení sítě v určitých lokalitách, plánovat její rozvoj efektivněji nebo modelovat chování sítě při výpadech či nárazovém nárůstu poptávky.

4.3 Příklad 3 – Modelový příklad využití syntetických IoT dat a prostorové analýzy efektivity trasy

Tato část práce se věnuje modelovému příkladu z oblasti chytré logistiky a ukazuje, jak lze i bez přístupu k reálným historickým datům navrhnout a provést smysluplnou analýzu. Hlavní

myšlenkou je propojit synteticky vytvořená data s prostorovou analýzou a vytvořit tak prakticky využitelný rámec pro podporu rozhodování, například v oblasti dopravního plánování nebo správy vozového parku.

Představovaný scénář vychází ze situace dopravní společnosti, která provozuje flotilu vozidel značky Ford Transit a usiluje o nalezení nejefektivnější trasy mezi dvěma lokalitami – konkrétně mezi Hradcem Králové a Trutnovem. Kromě délky trasy jsou při hodnocení zohledněny také další faktory, jako je spotřeba paliva, výše provozních nákladů nebo objem emisí CO₂.

Vzhledem k tomu, že reálná provozní data nebyla k dispozici, byl sestaven syntetický dataset obsahující 200 záznamů, které reprezentují různé jízdní situace. Jednotlivé jízdy se liší například hmotností nákladu, délkou trasy nebo sklonem terénu. Přestože data byla generována uměle, respektují reálné vztahy mezi sledovanými proměnnými – například vyšší hmotnost nebo větší sklon zvyšují spotřebu paliva i emise. Díky tomu je možné s tímto modelem reálně pracovat, srovnávat různé varianty tras a formulovat doporučení pro rozhodování.

Prostorová část analýzy byla realizována v prostředí ArcMap za využití nástroje Network Analyst a geografických dat z databáze ArcČR 500. Na základě těchto podkladů byly navrženy tři varianty tras mezi Hradcem Králové a Trutnovem a pro každou byla vypočtena délka i průměrný sklon. Tyto informace byly následně propojeny se syntetickými daty o jízdách a sloužily jako podklad pro výpočet spotřeby, emisí i celkových nákladů spojených s danou trasou.

4.3.1 Vytvoření a charakteristika syntetické datové sady

Z důvodu absence reálných provozních dat byla v prostředí Pythonu vytvořena syntetická datová sada obsahující celkem 200 záznamů simulovaných jízd vozidla Ford Transit. Každý záznam zahrnuje kombinaci technických, provozních, ekonomických a environmentálních proměnných, přičemž hodnoty byly navrženy tak, aby odrážely běžné provozní podmínky a vzájemné vztahy mezi proměnnými.

I když se nejedná o data získaná z reálného provozu (např. pomocí GPS nebo palubních jednotek), generované hodnoty respektují logiku fungování dopravních systémů:

- Spotřeba paliva se zvyšuje s rostoucím sklonem trasy a hmotností nákladu,
- Emise CO₂ jsou vypočteny na základě emisního faktoru vztaženého ke spotřebě,
- Náklady na jízdu vycházejí z celkové spotřeby a modelové ceny paliva (36 Kč za litr),

Doplňkové proměnné jako teplota vzduchu nebo tlak v pneumatikách odpovídají typickým hodnotám z běžných IoT senzorů, i když nejsou přímo využívány ve výpočtech – zvyšují však věrohodnost a komplexnost datové sady.

Tento syntetický dataset poskytuje základ pro testování analytických postupů a umožňuje simulovat různé scénáře v prostředí plánování a chytré logistiky (viz Tabulka 5).

Tabulka 5: Přehled sloupců syntetické datové sady

Zdroj: Vlastní zpracování

Sloupec	Popis
ID jízdy	Jedinečný identifikátor jízdy
Typ dodávky	Ford Transit
Pohon	Nafta
Trasa	Označení varianty trasy (např. Trasa 1–3)
Délka trasy (km)	Dle variant vytvořených v ArcMap
Průměrný sklon (%)	Spočten z vrstevnice DMR
Hmotnost nákladu (kg)	Náhodně přiřazeno (500–1400 kg)
Spotřeba (l/100 km)	Vypočtena na základě sklonu a hmotnosti
Spotřeba celkem (l)	$(\text{Spotřeba na 100 km} / 100) \times \text{délka trasy}$
Náklady (Kč)	$\text{Spotřeba celkem} \times 36 \text{ Kč/l}$
Emise (g CO ₂ /km)	Odvozeny ze spotřeby
Emise celkem (kg CO ₂)	$\text{Emise/km} \times \text{délka trasy} / 1000$
Teplota vzduchu	Simulovaná sensorová hodnota
Tlak v pneu	Simulovaná sensorová hodnota

Pro lepší přehlednost a reprodukovatelnost celého postupu je kompletní ukázka použitého kódu přiložena v příloze této práce. Kód zahrnuje všechny klíčové kroky – od načítání a předzpracování dat, přes časovou a prostorovou analýzu, až po tvorbu vizualizací.

Tvorba syntetického datasetu a vztahy mezi proměnnými

Pro potřeby této práce byl připraven syntetický dataset, který simuluje sběr provozních dat z logistické firmy využívající IoT zařízení instalovaná ve vozidlech. Cílem bylo vytvořit realistický model jízd, jejichž parametry odpovídají typickému provozu lehkých užitkových vozidel v reálných podmínkách.

Při generování dat byly uplatněny následující principy a logické vztahy mezi proměnnými:

- Spotřeba paliva byla ovlivněna zejména dvěma faktory – průměrným sklonem trasy a hmotností nákladu.
 - Se stoupajícím sklonem trasy se spotřeba zvyšovala o 0,02 l/100 km na každý 1 % sklonu.
 - Na každých 100 kg nákladu připadalo zvýšení spotřeby o 0,05 l/100 km.
- Emise CO₂ byly odvozeny přímo ze spotřeby paliva pomocí přibližného přepočtu – na základě hodnoty 26,5 gramů CO₂/km při spotřebě 1 litru nafty na 100 km.
- Náklady na jízdu byly vypočteny z aktuální spotřeby při předpokládané ceně paliva 36 Kč za litr.

Pro zvýšení realismu byla do datového souboru zahrnuta také simulovaná sensorová data, například:

- teplota motoru,
- tlak v pneumatikách,
- míra vibrací zaznamenaná při pohybu.

Výsledný dataset tak umožňuje nejen provádět základní analýzy spotřeby, emisí a nákladů, ale také ilustruje, jakým způsobem mohou data z IoT senzorů přispět k hlubšímu porozumění provoznímu chování vozidel. Je třeba zdůraznit, že dataset byl vytvořen synteticky a slouží výhradně jako modelový příklad. Přesto však umožňuje názorně demonstrovat různé analytické přístupy a propojení prostorových a provozních dat při podpoře rozhodování v oblasti chytré logistiky. Kompletní Python kód k dispozici v příloze (viz Příloha F).

4.3.2 Základní analýza a vizualizace syntetických dat

Po vygenerování syntetického datasetu byla provedena jeho základní analýza. Cílem této části bylo získat první přehled o struktuře dat, jejich rozložení a klíčových charakteristikách, které budou následně využity při detailnější analýze a vizualizaci.

Analýza se zaměřila především na hlavní proměnné související s přepravou a provozem vozidla, tedy na délku trasy, průměrný sklon, spotřebu paliva a emise CO₂. Tyto veličiny mají přímou vazbu na cílový úkol analýzy – posouzení efektivity jednotlivých tras a odhad provozních nákladů.

Základní statistické ukazatele

Pro vybrané proměnné byly vypočítány základní statistiky: průměr, medián, směrodatná odchylka, minimální a maximální hodnota. Shrnutí těchto údajů je uvedeno v následující tabulce (viz Tabulka 6):

Tabulka 6: Základní statistiky syntetické datové sady

Zdroj: Vlastní zpracování

Proměnná	Průměr	Medián	Směr. odchylka	Minimum	Maximum
Vzdálenost (km)	46,24	43,89	21,83	9,00	110,15
Průměrný sklon (%)	4,89	4,81	2,05	1,00	10,09
Spotřeba na 100 km (l)	8,21	8,20	0,49	7,31	9,06
Emise CO ₂ (kg/km)	0,22	0,22	0,013	0,20	0,24

Výsledky ukazují, že hodnoty jednotlivých proměnných odpovídají realistickému provoznímu prostředí užitkových vozidel. Průměrné hodnoty spotřeby a emisí vykazují relativně konzistentní rozložení, což potvrzuje vhodnost datové sady pro další modelování a srovnávací analýzy. Rozptyl hodnot délky trasy i průměrného sklonu zároveň naznačuje dostatečnou variabilitu jízd, což umožňuje analyzovat různé provozní podmínky a jejich dopady na spotřebu a emise. Kompletní Python kód k dispozici v příloze (viz Příloha G).

4.3.3 Grafická analýza hlavních proměnných

Vizualizace vztahů mezi hlavními parametry

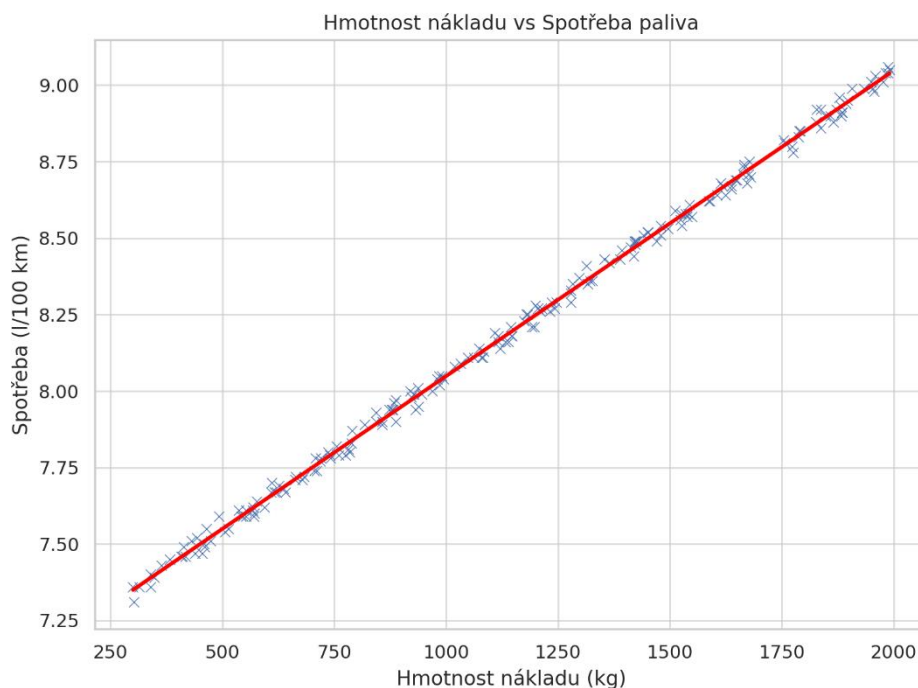
V rámci základní analýzy syntetického datasetu byla provedena také vizualizace vztahů mezi vybranými provozními a environmentálními parametry. Konkrétně byly zkoumány následující dvojice:

- hmotnost nákladu vs. spotřeba paliva,
- délka trasy vs. emise CO₂,
- hmotnost nákladu vs. emise CO₂.

Cílem této části bylo ověřit, zda se v generovaných datech projevují očekávané souvislosti mezi provozními vlastnostmi vozidla a jeho ekologickou stopou.

Hmotnost nákladu vs. spotřeba paliva.

První graf znázorňuje vztah mezi hmotností přepravovaného nákladu a průměrnou spotřebou paliva na 100 km. Je patrné, že spotřeba paliva s rostoucí hmotností nákladu soustavně stoupá, a to přibližně lineárním způsobem. Tento výsledek odpovídá běžné provozní logice – těžší vozidlo vyžaduje více energie k pohybu, což se promítá do vyšší spotřeby. Přidaná lineární trendová čára tento vztah vizuálně potvrzuje (viz Obrázek 19).

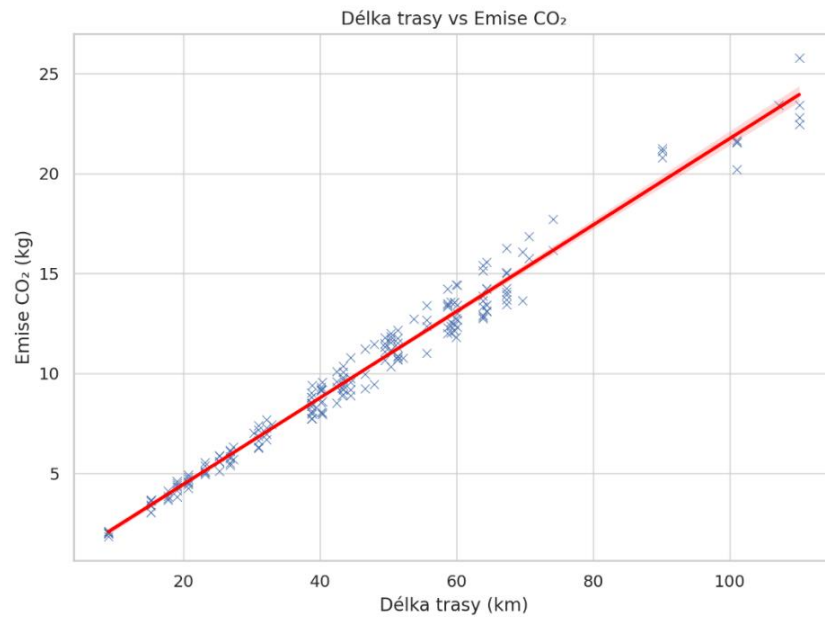


Obrázek 19: Hmotnost nákladu vs Spotřeba paliva

Zdroj: Vlastní zpracování

Délka trasy vs. emise CO₂

Druhý graf zachycuje vztah mezi celkovou délkou trasy a produkcí emisí CO₂. Také zde je patrná silná lineární závislost – čím delší je absolvovaná trasa, tím vyšší je množství vyprodukovaných emisí. Tento trend vychází z přímé vazby mezi ujetou vzdáleností, spotřebou paliva a emisemi, a tedy odpovídá očekávanému provoznímu chování (viz Obrázek 20).

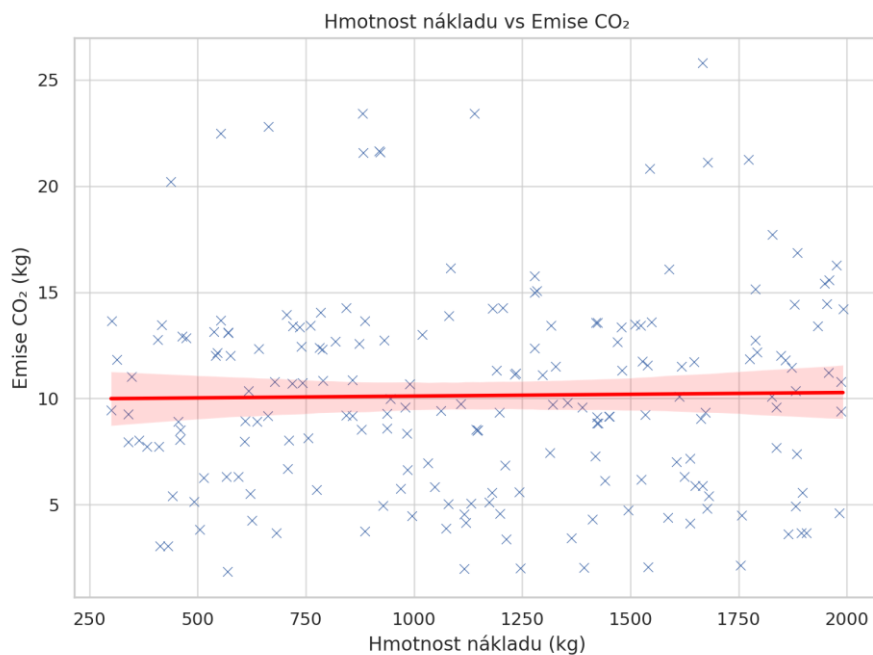


Obrázek 20: Délka trasy vs Emise CO₂

Zdroj: Vlastní zpracování

Hmotnost nákladu vs. emise CO₂

Třetí graf ukazuje vztah mezi hmotností nákladu a množstvím emisí CO₂. Ačkoliv zde není závislost tak jednoznačná jako v předchozích případech, lze pozorovat mírně vzrůstající trend – vyšší náklad bývá spojen s vyšší produkcí emisí. Rozptyl hodnot je poměrně velký, což naznačuje, že na celkovou produkci emisí má vliv více faktorů, například také délka trasy nebo průměrný sklon terénu (viz Obrázek 21).



Obrázek 21: Hmotnost nákladu vs Emise CO₂

Zdroj: Vlastní zpracování

Výsledky základní grafické analýzy ukazují, že i synteticky generovaná data mohou věrně odrážet očekávané provozní souvislosti. U klíčových proměnných, jako je hmotnost nákladu, délka trasy, spotřeba paliva a produkce emisí CO₂, byly potvrzeny logické trendy a vztahy.

Zatímco některé vztahy (např. mezi délkou trasy a emisemi CO₂) jsou vysoce předvídatelné, jiné (např. vliv hmotnosti nákladu na emise) vykazují větší rozptyl hodnot, což odpovídá reálnému provozu, kde na výsledné parametry působí více faktorů současně. Kompletní Python kód k dispozici v příloze (viz Příloha M).

4.3.4 Prostorová analýza tras a výpočet parametrů

Po provedení základní statistické analýzy a vyhodnocení vztahů mezi klíčovými proměnnými v rámci syntetického datasetu následuje praktická aplikace v podobě prostorové analýzy reálných tras. Cílem této části je ukázat, jak se geografické faktory – zejména délka trasy a její sklon – promítají do provozních parametrů, jako jsou spotřeba paliva, produkce emisí CO₂ a celkové přepravní náklady.

Analýza vychází ze simulace variant tras v oblasti východních Čech, konkrétně mezi městy Hradec Králové a Trutnov. Každá z tras byla navržena tak, aby se odlišovala délkou i výškovým profilem a bylo možné demonstrovat, jak různé geografické podmínky ovlivňují efektivitu dopravy.

Prostorová část analýzy byla realizována v prostředí geografického informačního systému ArcMap. Pomocí dostupných vrstev silniční sítě a vrstevnic byl pro každou variantu trasy vypočítán nejen její průběh, ale také průměrný sklon terénu.

Na základě těchto prostorových dat pak byly pro každou trasu odhadnuty provozní ukazatele – spotřeba paliva, emise CO₂ a náklady na přepravu. Výpočty navazují na vztahy a parametry dříve nastavené v syntetickém datasetu a umožňují tak zhodnocení výhodnosti jednotlivých tras z hlediska nákladů i environmentálního dopadu.

Tvorba tras a získání prostorových dat v ArcMap

Prostorová analýza variant tras mezi Hradcem Králové a Trutnovem byla realizována v prostředí ArcMap, které umožňuje efektivní kombinaci síťové a prostorové analýzy. Podkladem pro tvorbu tras byly vektorové vrstvy ArcČR 500 (silniční síť, vrstevnice) a rastrová vrstva Slope, která reprezentuje sklon terénu v procentech.

V prvním kroku byla pomocí nástroje Network Analyst vygenerována nejkratší trasa mezi městy – takzvaná prostřední varianta. Následně byly vytvořeny dvě alternativní varianty (levá

a pravá) přidáním průjezdních bodů, které trasy odklonily jinými směry. Tím vznikly tři odlišné možnosti přepravy, lišící se délkou i výškovým profilem.

Aby bylo možné získat detailní informace o sklonu podél tras, byla využita kombinace nástrojů:

- Feature Vertices to Points – tento nástroj vytvořil body podél linií tras v pravidelných vzdálenostech,
- Extract Values to Points – pro každý bod byla následně přiřazena hodnota sklonu terénu z rastrové vrstvy Slope.

Výsledkem byl soubor bodů s přiřazenou hodnotou sklonu, souřadnicemi a délkou příslušného úseku. Průměrný sklon trasy byl následně vypočítán jako průměr hodnot sloupce *Rastervalu*.

Tento postup umožnil nejen získat celkové délky tras, ale především kvantifikovat jejich náročnost z pohledu sklonu terénu.

Výsledky délky a sklonu tras

Pro každou variantu trasy byly určeny následující základní parametry (viz Tabulka 7):

Tabulka 7: Výsledky délky a sklonu tras

Zdroj: Vlastní zpracování

Varianta trasy	Délka trasy (km)	Průměrný sklon (%)	Minimální sklon (%)	Maximální sklon (%)
Prostřední	48,2	4,29	0,35	17
Pravá	51,45	4,18	0,35	17
Levá	51,95	3,96	0,35	17

Výsledky ukazují, že prostřední trasa je nejkratší, ale má mírně vyšší průměrný sklon než ostatní varianty.

Výpočet provozních parametrů (spotřeba paliva, náklady, emise)

Na základě prostorových parametrů jednotlivých variant tras – konkrétně délky a průměrného sklonu – byly následně spočítány klíčové provozní ukazatele: spotřeba paliva, náklady na přepravu a produkce emisí CO₂.

Výpočty navazují na metodiku popsanou při tvorbě syntetického datasetu (viz kapitola XY) a respektují vztahy mezi fyzikálními a provozními charakteristikami vozidla. Postup byl založen na následujících principech:

Výpočet spotřeby paliva

- Základní spotřeba vozidla na rovině byla stanovena na 7,5 l/100 km, což odpovídá běžné spotřebě lehkých užitkových vozidel.
- Na tuto základní hodnotu byly aplikovány dva korekční faktory:
 - Korekce za sklon terénu: spotřeba byla navýšena o 0,02 l/100 km za každý 1 % průměrného sklonu trasy.

$$\mathbf{Spotreba_sklon} = 0,02 \times Prumerny_sklon (\%)$$

- Korekce za hmotnost nákladu: spotřeba byla dále zvýšena o 0,05 l/100 km za každých 100 kg nad základní hmotnost vozidla.

$$\mathbf{Spotreba_hmotnost} = 0,05 \times \left(\frac{Hmotnost_nakladu (kg)}{100} \right)$$

Celková spotřeba paliva v l/100 km pak byla vypočtena následovně:

$$\mathbf{Spotreba_celkem (l/100 km)} = 7,5 + Spotreba_sklon + Spotreba_hmotnost$$

Pro každou trasu a hmotnost byla následně dopočítána celková spotřeba paliva (v litrech) pro danou délku trasy:

$$\mathbf{Spotreba_celkem (l)} = \frac{Spotreba_celkem (l/100 km) \times Delka_trasy (km)}{100}$$

Výpočet nákladů na palivo

Cena paliva byla stanovena na **36 Kč/litr**.

Celkové náklady na přepravu byly spočítány jako:

$$\mathbf{Naklady (Kc)} = Spotreba_celkem (l) \times 36$$

Výpočet emisí CO₂

Produkce emisí byla počítána přímo ze spotřeby paliva, s využitím emisního faktoru:

- 26,5 g CO₂/km při spotřebě 1 l/100 km.

Výpočet produkce emisí na 1 km:

$$\mathbf{Emise (g CO_2/km)} = 26,5 \times \frac{(Spotreba_celkem (l/100 km))}{1}$$

Celková produkce emisí CO₂ pro celou trasu:

$$\mathbf{Emise_celkem (kg CO_2)} = \left(\frac{Emise (g CO_2/km) \times Delka_trasy (km)}{1000} \right)$$

Přehled výsledků prostorové analýzy variant tras

Výsledky výpočtů provozních parametrů – tedy spotřeby paliva, nákladů na přepravu a produkce emisí CO₂ – pro jednotlivé varianty tras a různé hmotnosti nákladu jsou shrnuty v následující tabulce (viz Tabulka 8):

Tabulka 8: Přehled výsledků prostorové analýzy variant tras

Zdroj: Vlastní zpracování

Trasa	Hmotnost nákladu (kg)	Délka trasy (km)	Průměrný sklon (%)	Spotřeba (l/100 km)	Spotřeba celkem (l)	Náklady (Kč)	Emise (g CO ₂ /km)	Emise celkem (kg CO ₂)
Trasa 1	600	48,2	4,29	8,38	4,04	145,44	225	10,85
Trasa 1	1000	48,2	4,29	8,48	4,09	147,24	228	10,99
Trasa 1	1400	48,2	4,29	8,58	4,13	148,92	231	11,13
Trasa 2	600	51,45	4,18	8,28	4,26	153,36	223	11,48
Trasa 2	1000	51,45	4,18	8,38	4,31	155,16	226	11,63
Trasa 2	1400	51,45	4,18	8,48	4,36	156,96	229	11,78
Trasa 3	600	51,95	3,96	8,18	4,25	152,76	221	11,48
Trasa 3	1000	51,95	3,96	8,28	4,30	154,56	224	11,63
Trasa 3	1400	51,95	3,96	8,38	4,35	156,36	227	11,78

Vizualizace tras a příprava mapového výstupu

Pro lepší znázornění analyzovaných variant tras byla připravena prostorová vizualizace v prostředí ArcMap. Cílem bylo graficky zobrazit tři simulované trasy mezi městy Hradec Králové a Trutnov, přičemž jednotlivé varianty byly odlišeny podle jejich průběhu a délky.

Každá z tras byla reprezentována samostatnou linií vrstvou a barevně odlišena následovně:

- Modrá barva: Trasa 3 (levá varianta),
- Zelená barva: Trasa 2 (pravá varianta),
- Červená barva: Trasa 1 (prostřední varianta).

Součástí mapového výstupu byla také základní silniční síť a označení klíčových bodů – počátečního bodu („START“) a cílového bodu („CÍL“). Pro vymezení zájmového území byla využita hranice Královéhradeckého kraje, což umožnilo soustředit vizualizaci na relevantní oblast.

Výsledná mapa slouží jako přehledný podklad pro následné porovnání jednotlivých tras z hlediska provozních parametrů, jako je spotřeba paliva, produkce emisí CO₂ a přepravní náklady.

4.3.5 Závěr a interpretace výsledků

Tato kapitola shrnuje hlavní poznatky z prostorové analýzy tří variant tras mezi městy Hradec Králové a Trutnov. Analýza byla provedena na základě syntetického modelového datasetu, který simuloval reálné provozní podmínky lehkých užitkových vozidel vybavených IoT zařízeními. Hlavním cílem bylo porovnat varianty tras z hlediska jejich délky, průměrného sklonu a odvozených provozních parametrů – spotřeby paliva, emisí CO₂ a celkových nákladů.

Výpočet základních parametrů tras

V prostředí ArcMap byla pomocí nástroje Network Analyst vytvořena tři trasy, jejichž délky a průměrné sklony byly vypočteny následovně:

- Trasa 1 (prostřední, červená): délka 48,2 km, průměrný sklon 4,29 %.
- Trasa 2 (pravá, zelená): délka 51,45 km, průměrný sklon 4,18 %.
- Trasa 3 (levá, modrá): délka 51,95 km, průměrný sklon 3,96 %.

Odhad provozních parametrů

Na základě vypočtených parametrů byla pro každou trasu odhadnuta spotřeba paliva a celkové náklady. Výpočty byly provedeny pro tři různé hmotnosti nákladu: 600 kg, 1000 kg a 1400 kg. Hodnoty spotřeby paliva byly upravovány s ohledem na délku trasy, průměrný sklon a zatížení vozidla podle definovaných vztahů v syntetickém datasetu.

Přehled vybraných výsledků ukazuje, že při všech hmotnostech nákladu dosahuje nejlepší provozní parametry prostřední trasa (modrá). Výsledky pro zatížení 1000 kg jsou uvedeny v následující tabulce (viz Tabulka 9):

Tabulka 9: Výsledky parametrů pro zatížení 1000 kg

Zdroj: Vlastní zpracování

Trasa	Délka (km)	Průměrný sklon (%)	Spotřeba celkem (l)	Náklady (Kč)
Trasa 1	48,2	4,29	4,09	147,24
Trasa 2	51,45	4,18	4,31	155,16
Trasa 3	51,95	3,96	4,30	154,56

Podobný trend byl pozorován i při zatížení 600 kg a 1400 kg. Při vyšším zatížení (1400 kg) došlo k mírnému zvýšení spotřeby paliva o přibližně 0,1 l/100 km u všech tras, přičemž relativní rozdíly mezi variantami zůstaly zachovány.

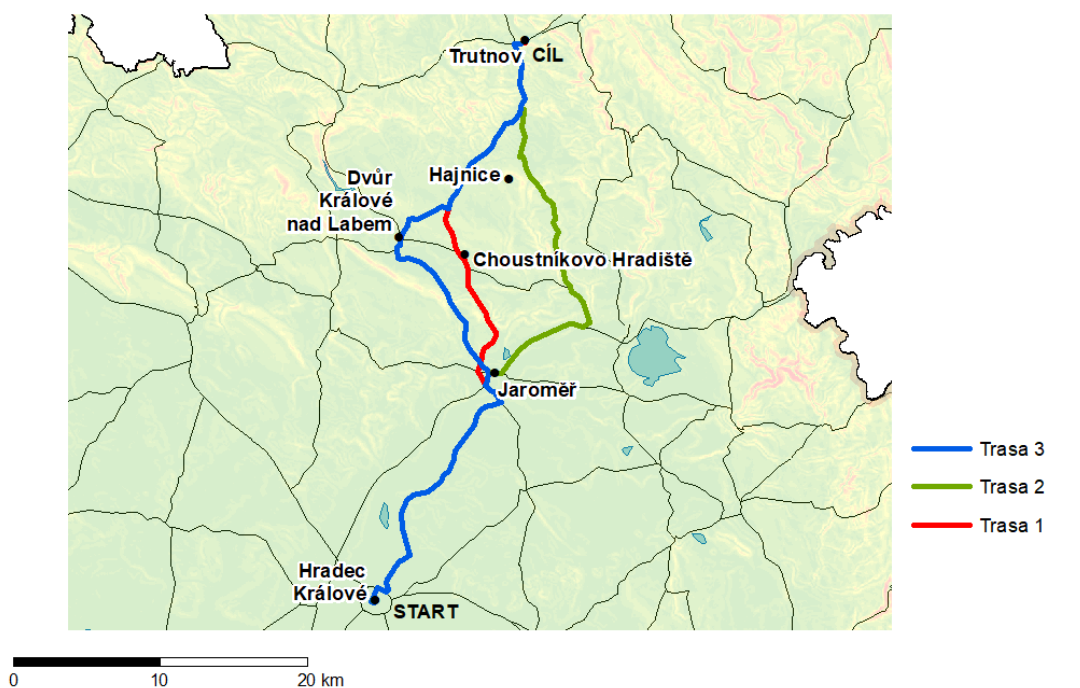
Z výsledků je patrné, že prostřední trasa (trasa 1) dosahuje nejlepších hodnot z hlediska délky, spotřeby paliva i nákladů při všech analyzovaných hmotnostech nákladu. Rozdíly v průměrném sklonu mezi trasami jsou relativně malé (v rozmezí přibližně 0,3 %), což naznačuje, že v tomto případě měla na provozní parametry větší vliv délka trasy než samotný sklon.

Emise CO₂ se u jednotlivých tras pohybovaly v rozmezí 224–229 g/km, což odpovídá očekávaným hodnotám pro analyzovaný typ vozidla a přepravní zátěž.

Vizualizace tras

Výsledné trasy byly vizualizovány v prostředí ArcMap. Pro vizualizaci byly jednotlivé varianty tras barevně odlišeny: prostřední trasa byla znázorněna červeně, pravá trasa zeleně a levá trasa modře. Trasy byly zakresleny přes podkladovou mapu s vrstevnicemi a hlavní silniční sítí, aby bylo možné lépe vnímat nejen samotný směr tras, ale také jejich vztah ke geografickému prostředí oblasti. Cílem vizualizace bylo umožnit rychlé a přehledné porovnání navržených variant a podpořit interpretaci výsledků prostorové analýzy (viz Obrázek 22).

SIMULACE DOPRAVNÍCH TRAS HRADEC KRÁLOVÉ – TRUTNOV



Pardubice 2025, Bc. Petr KÁBELE
Zdroj: ArcČR 500

Obrázek 22: Vizualizace dopravních tras Hradec Králové – Trutnov

Zdroj: Vlastní zpracování

Závěrem lze konstatovat, že syntetický dataset umožnil ilustrovat základní vztahy mezi geografickými charakteristikami tras a provozními parametry, jako jsou spotřeba paliva, emise CO₂ a přepravní náklady. Výsledky ukazují, že v daném modelovém případě byla nejefektivnější prostřední trasa.

Je však třeba dodat, že analýza byla založena na řadě zjednodušujících předpokladů. Nebyly zohledněny dynamické faktory, jako je aktuální dopravní situace, sezónní vlivy, kvalita povrchu nebo detailní výškový profil jednotlivých úseků. Výpočty vycházely ze stanovených konstantních parametrů, například fixní ceny paliva. Dále byla mezi vybranými proměnnými vypočtena korelace, nicméně vzhledem k syntetickému charakteru dat nebyly získané hodnoty zcela ideální. Kompletní přehled korelačních hodnot je uveden v příloze (viz Příloha H).

Přes tato omezení analýza názorně ukázala, jak lze propojit prostorová data a provozní charakteristiky vozidel pro účely rozhodování. Pokud by bylo k dispozici více dat o skutečném provozu, bylo by možné model zpřesnit například o rychlostní profily, detailní analýzu spotřeby na různých typech terénu nebo predikce vlivu dopravních omezení na provozní náklady.

ZÁVĚR

Diplomová práce byla zaměřena na analýzu a vizualizaci dat v oblasti internetu věcí (IoT), včetně prostorových dat, s cílem ukázat jejich využitelnost pro podporu rozhodovacích procesů. Vzhledem k dostupnosti datových zdrojů byla analýza realizována kombinací otevřených dat a synteticky vytvořené datové sady.

Práce pokrývá všechny hlavní kroky práce s daty – od sběru a předzpracování přes vlastní analýzu až po prostorovou vizualizaci. Použity byly běžně dostupné nástroje, zejména Python a ArcMap, a zvolený přístup odpovídal snaze efektivně využít dostupné možnosti bez nasazení pokročilých analytických metod. Použité metody a nástroje byly zvoleny s ohledem na dostupné možnosti a praktickou aplikovatelnost v běžném analytickém prostředí.

V praktické části byly použity tři ukázkové datové sady, na kterých byly předvedeny různé možnosti analýzy s využitím nástrojů ArcMap. Každá sada se zaměřuje na jiný typ problému a ukazuje, jak lze dostupná data využít v různých analytických scénářích.

První příklad se věnuje časoprostorové analýze aktivity uživatelů Twitteru na území Spojených států amerických. Byl využit otevřený dataset obsahující více než 14 milionů tweetů z ledna 2013. Data byla nejprve předzpracována převodem časových značek do místního času jednotlivých časových pásem USA. Časová analýza potvrdila nejvyšší aktivitu uživatelů ve večerních hodinách, zejména kolem 22:00 místního času, přičemž největší koncentrace aktivity byla v časovém pásmu Eastern. Prostorová analýza provedená nástrojem Kernel Density v prostředí ArcMap znázornila rozložení aktivity napříč Spojenými státy, s nejvyššími hustotami v hustě osídlených oblastech. Tato část ukázala, že i otevřená data lze při správném předzpracování využít k základní prostorově-časové analýze digitální aktivity.

Ve druhé části byla analyzována prostorová dostupnost serverové infrastruktury v oblasti Greater Melbourne v Austrálii. Byl využit simulovaný dataset uživatelů a reálná data o poloze serverů. Pomocí nástroje fishnet byla agregována data a vypočítán poměr uživatelů na server v jednotlivých prostorových buňkách. Analýza ukázala, že i při relativně rovnoměrném rozmístění uživatelů mohou vznikat oblasti s vyšší pravděpodobností přetížení infrastruktury, zejména na okrajích městské oblasti. Výsledky této části ilustrovaly možnosti využití základních prostorových operací při plánování kapacit digitální infrastruktury.

Třetí část práce byla věnována modelovému příkladu prostorové analýzy variant tras v logistice. Vytvořený syntetický dataset obsahoval 200 jízd mezi městy v Královéhradeckém kraji a zahrnoval parametry jako délka trasy, průměrný sklon, hmotnost nákladu, spotřeba

paliva, náklady a emise CO₂. Nejprve byla provedena základní analýza vztahů mezi těmito parametry. Následně byly v prostředí ArcMap navrženy tři varianty tras mezi Hradcem Králové a Trutnovem, pro které byla spočítána délka, průměrný sklon a provozní ukazatele při různých hmotnostech nákladu. Výsledky ukázaly, že prostřední trasa (směr Jaroměř – Choustníkovo Hradiště) byla ve všech případech provozně nejvýhodnější, přičemž rozdíly mezi variantami byly vzhledem k podobnému geografickému prostředí relativně malé. Tento příklad ukázal, jak lze i se syntetickými daty simulovat rozhodovací scénáře v oblasti logistiky.

Práce zároveň upozorňuje na limity spojené s využíváním otevřených a syntetických dat. Klíčovou roli hrála kvalita vstupních dat a pečlivé předzpracování. Použitá otevřená data vyžadovala úpravy a kontrolu konzistence, zatímco syntetická data musela být nastavena tak, aby realisticky zachycovala základní provozní vztahy.

Hlavní omezení práce spočívá v charakteru použitých dat – syntetická data nemohou plně vystihnout složitost reálných provozních podmínek, a absence pokročilejších metod omezila možnosti hlubší predikce nebo optimalizace. Přesto se podařilo ukázat základní vztahy mezi prostorovými charakteristikami, provozními parametry a jejich možnou vizualizací.

Do budoucna by bylo možné rozšířit analýzu například o využití reálných dat ze senzorových sítí v reálném čase, aplikaci metod strojového učení, zapojení environmentálních nebo ekonomických faktorů, nebo realizaci podobných analýz v open-source GIS prostředích.

Závěrem lze shrnout, že práce v rámci možností dostupných dat splnila stanovené cíle, ukázala postupy práce s IoT daty a nabídla realistický pohled na možnosti i omezení těchto přístupů v podpoře rozhodování.

SEZNAM POUŽITÉ LITERATURY

- [1] ABBA ARI, Ado Adamou, Hamayadji Abdoul AZIZ, Arouna Ndam NJOYA, Moussa ABOUBAKAR, Assidé Christian DJEDOUBOUM, Ousmane THIARE a Alidou MOHAMADOU, 2024. Data collection in IoT networks: Architecture, solutions, protocols and challenges. *IET Wireless Sensor Systems* [online]. 14(4), 85-110 [cit. 2025-02-18]. ISSN 2043-6386. Dostupné z: doi:10.1049/wss2.12080
- [2] ABOLHASSANI KHAJEH, Sima, Morteza SABERIKAMARPOSHTI a Amir Masoud RAHMANI, 2023. Real-Time Scheduling in IoT Applications: A Systematic Review. *Sensors* [online]. 23(1), 232 [cit. 2025-02-24]. ISSN 1424-8220. Dostupné z: doi:10.3390/s23010232
- [3] AHMAD, Waqas, Aamir RASOOL, Abdul Rehman JAVED, Thar BAKER a Zunera JALIL, 2021. Cyber Security in IoT-Based Cloud Computing: A Comprehensive Survey. *Electronics* [online]. 11(1) [cit. 2025-03-26]. ISSN 2079-9292. Dostupné z: doi:10.3390/electronics11010016
- [4] AHMMED, Md Sazol, Sriram Praneeth ISANAKA a Frank LIOU, 2024. Promoting Synergies to Improve Manufacturing Efficiency in Industrial Material Processing: A Systematic Review of Industry 4.0 and AI. *Machines* [online]. 12(10) [cit. 2025-03-12]. ISSN 2075-1702. Dostupné z: doi:10.3390/machines12100681
- [5] AL-ALI, A.R., Ragini GUPTA, Imran ZUALKERNAN a Sajal K. DAS, 2024. Role of IoT technologies in big data management systems: A review and Smart Grid case study. *Pervasive and Mobile Computing* [online]. 100 [cit. 2025-04-15]. ISSN 15741192. Dostupné z: doi:10.1016/j.pmcj.2024.101905
- [6] AL-ANSI, Abdullah M., Askar GARAD, Mohammed JABOUB a Ahmed AL-ANSI, 2024. Elevating e-government: Unleashing the power of AI and IoT for enhanced public services. *Heliyon* [online]. 10(23) [cit. 2025-03-10]. ISSN 24058440. Dostupné z: doi:10.1016/j.heliyon.2024.e40591
- [7] ALAHI, Md Eshrat E., Arsanchai SUKKUEA, Fahmida Wazed TINA, Anindya NAG, Wattanapong KURDTHONGMEE, Korakot SUWANNARAT a Subhas Chandra MUKHOPADHYAY, 2023. Integration of IoT-Enabled Technologies and Artificial Intelligence (AI) for Smart City Scenario: Recent Advancements and Future Trends. *Sensors* [online]. 23(11), 5206 [cit. 2025-03-12]. ISSN 1424-8220. Dostupné z: doi:10.3390/s23115206

- [8] ANDRIULO, Francesco Cosimo, Marco FIORE, Marina MONGIELLO, Emanuele TRAVERSA a Vera ZIZZO, 2024. Edge Computing and Cloud Computing for Internet of Things: A Review. *Informatics* [online]. 11(4), 71 [cit. 2025-02-11]. ISSN 2227-9709. Dostupné z: doi:10.3390/informatics11040071
- [9] BOČEK, Jan, Jakub MRÁČEK a Jindřich MYNARZ, 2012. *Otevřená data: příležitost pro Českou republiku* [online]. Praha: Fond Otakara Motejla [cit. 2025-04-15]. ISBN 978-80-87725-03-0.
- [10] BORRA, Praveen, 2024. Impact and Innovations of Azure IoT: Current Applications, Services, and Future Directions. *International Journal of Recent Technology and Engineering (IJRTE)* [online]. 13(2), 21-26 [cit. 2025-01-30]. ISSN 22773878. Dostupné z: doi:10.35940/ijrte.B8111.13020724
- [11] CAI, Tao, Peiyao LIU, Dejiao NIU, Jiancong SHI, Lei LI a Shah NAZIR, 2021. The Embedded IoT Time Series Database for Hybrid Solid-State Storage System. *Scientific Programming* [online]. 2021, 1-13 [cit. 2025-01-30]. ISSN 1875-919X. Dostupné z: doi:10.1155/2021/9948533
- [12] CRUZ HUACARPUMA, Ruben, Rafael DE SOUSA JUNIOR, Maristela DE HOLANDA, Robson DE OLIVEIRA ALBUQUERQUE, Luis GARCÍA VILLALBA a Tai-Hoon KIM, 2017. Distributed Data Service for Data Management in Internet of Things Middleware. *Sensors* [online]. 17(5), 977 [cit. 2025-01-30]. ISSN 1424-8220. Dostupné z: doi:10.3390/s17050977
- [13] ČESKO V DATECH, 2019. Chytrá města a jejich data. *Česko v datech* [online]. [cit. 2025-04-28]. Dostupné z: <https://www.ceskovdatech.cz/clanek/131-chytra-mesta-a-jejich-data/>
- [14] DAS, Anirban, Navlika SINGH a Suchetana CHAKRABORTY, 2024. UniPreCIS: A data preprocessing solution for collocated services on shared IoT. *Future Generation Computer Systems* [online]. 153, 543-557 [cit. 2025-03-21]. ISSN 0167739X. Dostupné z: doi:10.1016/j.future.2023.11.029
- [15] DEMERS, Michael N., 2009. *Fundamentals of geographic information systems* [online]. 4th ed. Hoboken: John Wiley. ISBN 978-0-470-12906-7.
- [16] DOMÍNGUEZ-BOLAÑO, Tomás, Omar CAMPOS, Valentín BARRAL, Carlos J. ESCUDERO a José A. GARCÍA-NAYA, 2022. An overview of IoT architectures, technologies, and existing open-source projects. *Internet of Things* [online]. 20 [cit. 2025-04-20]. ISSN 25426605. Dostupné z: doi:10.1016/j.iot.2022.100626

- [17] DUSUN, 2024. How IoT Data Collection Works. *Dusun* [online]. [cit. 2024-12-21]. Dostupné z: <https://www.dusuniot.com/blog/iot-data-collection/>
- [18] DUSUN, 2025. DSGW-510 KNX Zigbee Gateway (Wi-Fi/ Ethernet/RS485/RS232/M-Bus/CAN). *Dusun* [online]. [cit. 2025-01-25]. Dostupné z: <https://www.dusuniot.com/product/dsgw-510-knx-zigbee-wifi-ethernet-gateway/>
- [19] GITHUB, 2018. EUA Datasets. *GitHub* [online]. [cit. 2025-04-29]. Dostupné z: <https://github.com/PhuLai/eua-dataset>
- [20] HLADÍK, Jiří, 2024. *Prostorové analýzy a vizualizace dat ze sociálních sítí* [online]. Brno [cit. 2025-04-22]. Dostupné z: <https://is.muni.cz/th/gvoyr/?studium=246504;lang=cs>. Disertace. Masarykova univerzita.
- [21] IOTPORT, 2020. Co je to IoT? *IoTPORT* [online]. [cit. 2024-09-11]. Dostupné z: <https://www.iotport.cz/iot-novinky/ostatni-clanky-o-iot/co-to-je-iot>
- [22] JYOTHI, J, 2022. Cassandra is a Better Option for Handling Big Data in a No-SQL Database. *International Journal of Research Publication and Reviews* [online]. 3(9), 880-883 [cit. 2025-01-31]. ISSN 25827421. Dostupné z: doi:10.55248/gengpi.2022.3.9.27
- [23] KOLAJO, Taiwo, Olawande DARAMOLA a Ayodele ADEBIYI, 2019. Big data stream analysis: a systematic literature review. *Journal of Big Data* [online]. 6(47) [cit. 2025-03-10]. ISSN 2196-1115. Dostupné z: doi:10.1186/s40537-019-0210-7
- [24] KRISHNAMURTHI, Rajalakshmi, Adarsh KUMAR, Dhanalekshmi GOPINATHAN, Anand NAYYAR a Basit QURESHI, 2020. An Overview of IoT Sensor Data Processing, Fusion, and Analysis Techniques. *Sensors* [online]. 20(21), 6076 [cit. 2025-01-25]. ISSN 1424-8220. Dostupné z: doi:10.3390/s20216076
- [25] KRZYSZTOŃ, Emanuel, Izabela ROJEK a Dariusz MIKOŁAJEWSKI, 2024. A Comparative Analysis of Anomaly Detection Methods in IoT Networks: An Experimental Study. *Applied Sciences* [online]. 14(24), 11545 [cit. 2025-03-21]. ISSN 2076-3417. Dostupné z: doi:10.3390/app142411545
- [26] LAKHOUIT, Abderrahim, 2025. Revolutionizing urban solid waste management with AI and IoT: A review of smart solutions for waste collection, sorting, and recycling. *Results in Engineering* [online]. 25 [cit. 2025-03-12]. ISSN 25901230. Dostupné z: doi:10.1016/j.rineng.2025.104018
- [27] LIU, Yu, Kahin AKRAM HASSAN, Magnus KARLSSON, Zhibo PANG a Shaofang GONG, 2019. A Data-Centric Internet of Things Framework Based on Azure Cloud. *IEEE Access* [online]. 7, 53839-53858 [cit. 2025-01-29]. ISSN 2169-3536. Dostupné z: doi:10.1109/ACCESS.2019.2913224

- [28] MAI, Gengchen, Yiqun XIE, Xiaowei JIA, a kol., 2025. Towards the next generation of Geospatial Artificial Intelligence. *International Journal of Applied Earth Observation and Geoinformation* [online]. 136 [cit. 2025-03-21]. ISSN 15698432. Dostupné z: doi:10.1016/j.jag.2025.104368
- [29] MALIK, Sehrish, Shabir AHMAD, Israr ULLAH, Dong Hwan PARK a DoHyeun KIM, 2019. An Adaptive Emergency First Intelligent Scheduling Algorithm for Efficient Task Management and Scheduling in Hybrid of Hard Real-Time and Soft Real-Time Embedded IoT Systems. *Sustainability* [online]. 11(8), 2192 [cit. 2025-02-25]. ISSN 2071-1050. Dostupné z: doi:10.3390/su11082192
- [30] MARASINGHE, Raveena, Tan YIGITCANLAR, Severine MAYERE, Tracy WASHINGTON a Mark LIMB, 2024. Towards Responsible Urban Geospatial AI: Insights From the White and Grey Literatures. *Journal of Geovisualization and Spatial Analysis* [online]. 8(24) [cit. 2025-03-21]. ISSN 2509-8810. Dostupné z: doi:10.1007/s41651-024-00184-2
- [31] MIJUSKOVIC, Adriana, Alessandro CHIUMENTO, Rob BEMTHUIS, Adina ALDEA a Paul HAVINGA, 2021. Resource Management Techniques for Cloud/Fog and Edge Computing: An Evaluation Framework and Classification. *Sensors* [online]. 21(5), 1832 [cit. 2025-02-11]. ISSN 1424-8220. Dostupné z: doi:10.3390/s21051832
- [32] NAWARA, Dina a Rasha KASHEF, 2021. Context-Aware Recommendation Systems in the IoT Environment (IoT-CARS)—A Comprehensive Overview. *IEEE Access* [online]. 9, 144270-144284 [cit. 2025-02-18]. ISSN 2169-3536. Dostupné z: doi:10.1109/ACCESS.2021.3122098
- [33] OLADIMEJI, Damilola, Khushi GUPTA, Nuri Alperen KOSE, Kubra GUNDOGAN, Linqiang GE a Fan LIANG, 2023. Smart Transportation: An Overview of Technologies and Applications. *Sensors* [online]. 23(8), 3880 [cit. 2024-12-05]. ISSN 1424-8220. Dostupné z: doi:10.3390/s23083880
- [34] PANDEY, Arvind Kumar, V. SELVAKUMAR, P. LAVANYA, S. LAKSHMI PRABHA, S. UMA MAGESHWARI, K. BAPAYYA NAIDU a Rachna SRIVASTAVA, 2024. Optimizing Power Management in IoT Devices Using Machine Learning Techniques. *Journal of Electrical Systems* [online]. 20(5s), 2929-2940 [cit. 2025-03-30]. ISSN 1112-5209. Dostupné z: doi:10.52783/jes.3331
- [35] PAVLÍK, Jan, Markéta HRNČÍROVÁ, Michal STOČES, Jan MASNER a Jiří VANĚK, 2020. Usability of IoT and Open Data Repositories for Analyzing Water Pollution. A Case

- Study in the Czech Republic. *ISPRS International Journal of Geo-Information* [online]. 9(10) [cit. 2025-04-07]. ISSN 2220-9964. Dostupné z: doi:10.3390/ijgi9100591
- [36] PHUNG, Manh Duong, Michel DE LA VILLEFROMOY a Quang HA, 2017. Management of solar energy in microgrids using IoT-based dependable control. 2017 20th *International Conference on Electrical Machines and Systems (ICEMS)* [online]. IEEE, 1-6 [cit. 2025-04-07]. ISBN 978-1-5386-3246-8. Dostupné z: doi:10.1109/ICEMS.2017.8056441
- [37] SAFARI BAZARGANI, Jalal, Abolghasem SADEGHI-NIARAKI a Soo-Mi CHOI, 2021. A Survey of GIS and IoT Integration: Applications and Architecture. *Applied Sciences* [online]. 11(21), 10365 [cit. 2024-11-23]. ISSN 2076-3417. Dostupné z: doi:10.3390/app112110365
- [38] SAIDU, Yahaya, Shuhaida Mohamed SHUHIDAN, Dahiru Adamu ALIYU, Izzatdin ABDUL AZIZ a Shamsuddeen ADAMU, 2025. Convergence of Blockchain, IoT, and AI for Enhanced Traceability Systems: A Comprehensive Review. *IEEE Access* [online]. 13, 16838-16865 [cit. 2025-02-10]. ISSN 2169-3536. Dostupné z: doi:10.1109/ACCESS.2025.3528035
- [39] SHAH, Kunj, Chaitanya SHETH a Nishant DOSHI, 2022. A Survey on IoT-Based Smart Cars, their Functionalities and Challenges. *Procedia Computer Science* [online]. 210, 295-300 [cit. 2024-10-13]. ISSN 18770509. Dostupné z: doi:10.1016/j.procs.2022.10.153
- [40] SHAHROKHI, Amirhossein a Mahmood AHMADI, 2023. Power Evaluation of IOT Application Layer Protocols. 2023 *7th International Conference on Internet of Things and Applications (IoT)* [online]. IEEE, 2023-10-25, 1-7 [cit. 2025-04-12]. ISBN 979-8-3503-6941-0. Dostupné z: doi:10.1109/IoT60973.2023.10365351
- [41] SHIRVANIMOGHADDAM, Mahyar, Kamyar SHIRVANIMOGHADDAM, Mohammad Mahdi ABOLHASANI, Majid FARHANGI, Vahid ZAHIRI BARSARI, Hangyue LIU, Mischa DOHLER a Mino NAEBE, 2019. Towards a Green and Self-Powered Internet of Things Using Piezoelectric Energy Harvesting. *IEEE Access* [online]. 7, 94533-94556 [cit. 2025-04-12]. ISSN 2169-3536. Dostupné z: doi:10.1109/ACCESS.2019.2928523
- [42] SHUMBA, Angela-Tafadzwa, Teodoro MONTANARO, Ilaria SERGI, a kol., 2023. Wearable Technologies and AI at the Far Edge for Chronic Heart Failure Prevention and Management: A Systematic Review and Prospects. *Sensors* [online]. 23(15) [cit. 2025-03-12]. ISSN 1424-8220. Dostupné z: doi:10.3390/s23156896

- [43] SIRISHA, N., M. GOPIKRISHNA, P. RAMADEVI, Raveendranadh BOKKA, K.V.B. GANESH a M. Kalyan CHAKRAVARTHI, 2023. IoT-based data quality and data preprocessing of multinational corporations. *The Journal of High Technology Management Research* [online]. 34(2), 100477 [cit. 2025-03-21]. ISSN 10478310. Dostupné z: doi:10.1016/j.hitech.2023.100477
- [44] SONG, Zhihang, Zimin HE, Xingyu LI, a kol., 2024. Synthetic Datasets for Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Vehicles* [online]. 9(1), 1847-1864 [cit. 2025-04-22]. ISSN 2379-8904. Dostupné z: doi:10.1109/TIV.2023.3331024
- [45] UC IRVINE MACHINE LEARNING REPOSITORY, 2013. Twitter Geospatial Data. *UC Irvine Machine Learning Repository* [online]. [cit. 2025-04-29]. Dostupné z: <https://archive.ics.uci.edu/dataset/1050/twitter+geospatial+data>
- [46] WEBBYLAB, 2025. How Internet of Things Data Collection Works: A Full Guide. *WebbyLab* [online]. [cit. 2025-01-20]. Dostupné z: <https://webbylab.com/blog/iot-data-collection/>

SEZNAM PŘÍLOH

Příloha A Převod časového pásma a odvození proměnných.....	I
Příloha B Graf denní aktivity uživatelů Twitteru podle hodin.....	II
Příloha C Modelování v Pythonu.....	III
Příloha D Modelování v Pythonu	IV
Příloha E Modelování v Pythonu.....	V
Příloha F Tvorba syntetického datasetu	VI
Příloha G Modelování v Pythonu	VIII
Příloha H Korelační matice.....	IX
Příloha I Přehled sloupců a datových typů uživatelů.....	XI
Příloha J Přehled sloupců a datových typů serverů	XI
Příloha K Výstup základní kontroly dat datové sady uživatelů.....	XI
Příloha L Výstup základní kontroly dat datové sady serverů	XII
Příloha M Grafická analýza hlavních proměnných syntetického datasetu	XIII

Příklad 1 – data Twitter

Python – převod časového pásma a odvození proměnných

```
1 import pandas as pd
2
3 # Načtení původního datasetu (např. po časové analýze)
4 df = pd.read_csv("twitter.csv")
5
6 # Převod timestamp na datetime (CST)
7 df["datetime_cst"] = pd.to_datetime(df["timestamp"], format="%Y%m%d%H%M%S")
8
9 # Převod timezone na posun vůči CST
10 tz_shift = {1: +1, 2: 0, 3: -1, 4: -2}
11 df["tz_offset"] = df["timezone"].map(tz_shift)
12
13 # Výpočet lokálního času a hodiny
14 df["local_datetime"] = df["datetime_cst"] + pd.to_timedelta(df["tz_offset"], unit="h")
15 df["local_hour"] = df["local_datetime"].dt.hour
16
17 # Přesný přetyp souřadnic na čísla (float), odstranění chyb
18 df["longitude"] = pd.to_numeric(df["longitude"], errors="coerce")
19 df["latitude"] = pd.to_numeric(df["latitude"], errors="coerce")
20 df = df.dropna(subset=["longitude", "latitude"]) # odstranění NaN hodnot
21
22 # Výběr sloupců pro výstup (pro ArcMap)
23 columns_for_export = [
24     "longitude", "latitude", "timestamp", "timezone",
25     "datetime_cst", "tz_offset", "local_datetime", "local_hour"
26 ]
27 df = df[columns_for_export]
28
29 # Uložení CSV
30 df.to_csv("tweets_clean.csv", index=False, encoding="utf-8")
```

Příloha A Převod časového pásma a odvození proměnných

Příklad 1 – data Twitter

Python – Graf denní aktivity uživatelů Twitteru podle hodin

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 from matplotlib import ticker
4
5 # Načtení dat
6 df = pd.read_csv("tweets_clean.csv")
7
8 # Převod na datetime objekt
9 df["local_datetime"] = pd.to_datetime(df["local_datetime"])
10
11 # Extrakce hodiny z lokálního času
12 df["local_hour"] = df["local_datetime"].dt.hour
13
14 # Spočítání výskytů podle hodiny
15 hour_counts = df["local_hour"].value_counts().sort_index()
16
17 # Vykreslení grafu
18 plt.figure(figsize=(12, 6))
19 plt.bar(hour_counts.index, hour_counts.values, color="lightgreen", edgecolor="black")
20
21 plt.title("Denní aktivita uživatelů Twitteru podle hodin (lokální čas) v USA (12.-18. 1. 2013)")
22 plt.xlabel("Hodina dne")
23 plt.ylabel("Počet tweetů")
24
25 # Vlastní formátování osy Y: „tis.“ a „mil.“
26 def custom_formatter(x, _):
27     if x >= 1_000_000:
28         return f"{int(x / 1_000_000)} mil."
29     elif x >= 1_000:
30         return f"{int(x / 1_000)} tis."
31     else:
32         return str(int(x))
33
34 plt.gca().yaxis.set_major_formatter(ticker.FuncFormatter(custom_formatter))
35
36 plt.xticks(range(24))
37 plt.grid(axis="y", linestyle="--", alpha=0.7)
38
39 # Uložení grafu
40 plt.tight_layout()
41 plt.savefig("tweets_by_hour_local.png", dpi=300)
42 plt.show()
```

Příloha B Graf denní aktivity uživatelů Twitteru podle hodin

Modelování v Pythonu

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import matplotlib.ticker as ticker
4 import seaborn as sns
5
6 # Načtení dat
7 df = pd.read_csv("tweets_clean.csv")
8
9 # Převod local_datetime zpět na datetime
10 df["local_datetime"] = pd.to_datetime(df["local_datetime"])
11
12 # Přiřazení názvů časových zón
13 zone_labels = {1: "Eastern", 2: "Central", 3: "Mountain", 4: "Pacific"}
14 df["timezone_label"] = df["timezone"].map(zone_labels)
15
16 # Vytvoření kontingenční tabulky s počty tweetů
17 pivot = df.groupby(["timezone_label", "local_hour"]).size().reset_index(name="count")
18
19 # Vykreslení čárového grafu
20 plt.figure(figsize=(12, 6))
21 sns.lineplot(data=pivot, x="local_hour", y="count", hue="timezone_label", marker="o")
22
23 plt.title("Denní rytmus tweetování podle časové zóny v USA (12.-18. 1. 2013)")
24 plt.xlabel("Hodina dne")
25 plt.ylabel("Počet tweetů")
26
27 plt.xticks(range(0, 24))
28 plt.grid(True, linestyle="--", alpha=0.7)
29
30 # Vlastní formátování osy Y
31 def custom_formatter(x, _):
32     if x >= 1_000_000:
33         return f"{int(x / 1_000_000)} mil."
34     elif x >= 1_000:
35         return f"{int(x / 1_000)} tis."
36     else:
37         return str(int(x))
38
39 plt.gca().yaxis.set_major_formatter(ticker.FuncFormatter(custom_formatter))
40
41 plt.legend(title="Časová zóna")
42
43 # Uložení grafu
44 plt.tight_layout()
45 plt.savefig("tweets_by_hour_and_timezone.png", dpi=300)
46 plt.show()
```

Příloha C Modelování v Pythonu

Modelování v Pythonu

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import matplotlib.ticker as ticker
4
5 # Načtení dat
6 df = pd.read_csv("tweets_clean.csv")
7
8 # Mapování kódů časových zón na názvy
9 zone_labels = {1: "Eastern", 2: "Central", 3: "Mountain", 4: "Pacific"}
10 df["timezone_label"] = df["timezone"].map(zone_labels)
11
12 # Sečtení tweetů podle časové zóny
13 tweets_by_zone = df["timezone_label"].value_counts().sort_index()
14
15 # Vykreslení grafu
16 plt.figure(figsize=(8, 5))
17 tweets_by_zone.plot(kind="bar", color="lightskyblue", edgecolor="black")
18 plt.title("Počet tweetů podle časové zóny v USA (12.-18. 1. 2013)")
19 plt.xlabel("Časová zóna")
20 plt.ylabel("Počet tweetů")
21 plt.xticks(rotation=0)
22 plt.grid(axis="y", linestyle="--", alpha=0.7)
23
24 # Formátování osy Y na „x tis.“ nebo „x mil.“ bez desetinných míst
25 def custom_formatter(x, _):
26     if x >= 1_000_000:
27         return f"{int(x / 1_000_000)} mil."
28     elif x >= 1_000:
29         return f"{int(x / 1_000)} tis."
30     else:
31         return str(int(x))
32
33 plt.gca().yaxis.set_major_formatter(ticker.FuncFormatter(custom_formatter))
34
35 # Uložení grafu
36 plt.tight_layout()
37 plt.savefig("tweets_by_timezone.png", dpi=300)
38 plt.show()
```

Příloha D Modelování v Pythonu

Modelování v Pythonu

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 from matplotlib import ticker
4
5 # Načtení dat
6 df = pd.read_csv("tweets_clean.csv")
7 df["local_datetime"] = pd.to_datetime(df["local_datetime"])
8
9 # Filtrovat jen požadované období
10 df = df[(df["local_datetime"] >= "2013-01-12") & (df["local_datetime"] < "2013-01-19")]
11
12 # Převést na datum bez času
13 df["local_date"] = df["local_datetime"].dt.date
14
15 # Spočítat tweety podle data
16 date_counts = df["local_date"].value_counts().sort_index()
17
18 # Vykreslení grafu
19 plt.figure(figsize=(12, 6))
20 plt.bar(date_counts.index.astype(str), date_counts.values, color="mediumseagreen", edgecolor="black")
21
22 plt.title("Počet tweetů podle jednotlivých dnů v USA (12.-18. 1. 2013)")
23 plt.xlabel("Datum")
24 plt.ylabel("Počet tweetů")
25
26 # Nastavení minimální hodnoty osy Y na 500 000
27 plt.ylim(bottom=500_000)
28
29 # Vlastní formátování osy Y
30 def custom_formatter(x, _):
31     if x >= 1_000_000:
32         return f"{x / 1_000_000:.1f} mil."
33     elif x >= 1_000:
34         return f"{int(x / 1_000)} tis."
35     else:
36         return f"{int(x)}"
37
38 plt.gca().yaxis.set_major_formatter(ticker.FuncFormatter(custom_formatter))
39
40 plt.xticks(rotation=45)
41 plt.grid(axis="y", linestyle="--", alpha=0.7)
42 plt.tight_layout()
43 plt.savefig("tweets_by_day_filtered.png", dpi=300)
44 plt.show()
```

Příloha E Modelování v Pythonu

Python – tvorba syntetického datasetu

```
1 import pandas as pd
2 import numpy as np
3 import random
4
5 # Nastaveni nahodneho seedu pro opakovatelnost
6 np.random.seed(42)
7 random.seed(42)
8
9 # Definice mest v Kralovehradeckem kraji
10 mesta = ['Hradec Kralove', 'Trutnov', 'Nachod', 'Jicin', 'Rychnov nad Kneznou', 'Dvur Kralove nad Labem',
11          'Jaromer', 'Nova Paka', 'Broumov', 'Kostelec nad Orlici']
12
13 # Funkce na generovani nahodnych souradnic kolem daneho mesta
14 def generuj_souradnice(mesto):
15     lat = {
16         'Hradec Kralove': 50.209,
17         'Trutnov': 50.561,
18         'Nachod': 50.416,
19         'Jicin': 50.437,
20         'Rychnov nad Kneznou': 50.168,
21         'Dvur Kralove nad Labem': 50.435,
22         'Jaromer': 50.363,
23         'Nova Paka': 50.490,
24         'Broumov': 50.586,
25         'Kostelec nad Orlici': 50.123
26     }
27     lon = {
28         'Hradec Kralove': 15.832,
29         'Trutnov': 15.912,
30         'Nachod': 16.161,
31         'Jicin': 15.351,
32         'Rychnov nad Kneznou': 16.278,
33         'Dvur Kralove nad Labem': 15.813,
34         'Jaromer': 15.927,
35         'Nova Paka': 15.514,
36         'Broumov': 16.331,
37         'Kostelec nad Orlici': 16.210
38     }
39     return lat[mesto] + np.random.uniform(-0.01, 0.01), lon[mesto] + np.random.uniform(-0.01, 0.01)
40
41 # Pocet zaznamu
42 n = 200
43
44 # Generovani dat
45 data = []
```

Příloha F Tvorba syntetického datasetu

Python – tvorba syntetického datasetu – část 2.

```
46
47 for i in range(1, n+1):
48     start_mesto = random.choice(mesta)
49     cil_mesto = random.choice([m for m in mesta if m != start_mesto])
50     start_lat, start_long = generuj_souradnice(start_mesto)
51     cil_lat, cil_long = generuj_souradnice(cil_mesto)
52     delka_trasy = np.random.uniform(10, 100)
53     sklon_prumer = np.random.uniform(0, 7)
54     hmotnost_nakladu = np.random.uniform(300, 1200)
55     spotreba_na_100km = 7.5 + 0.03 * sklon_prumer + 0.001 * hmotnost_nakladu
56     spotreba_l = (spotreba_na_100km / 100) * delka_trasy
57     cena_nafta_kc_l = 36
58     cena_jizdy = spotreba_l * cena_nafta_kc_l
59     emise_co2_g_km = 2.65 * spotreba_na_100km * 10
60     emise_co2_kg = (emise_co2_g_km / 1000) * delka_trasy
61     vibrace = np.random.uniform(0.1, 2.5)
62     tlak_v_pneumatikach = np.random.uniform(2.2, 2.8)
63     teplota_motoru = np.random.uniform(80, 105)
64
65     data.append({
66         'ID_jizdy': i,
67         'Start_mesto': start_mesto,
68         'Cil_mesto': cil_mesto,
69         'Start_lat': round(start_lat, 6),
70         'Start_long': round(start_long, 6),
71         'Cil_lat': round(cil_lat, 6),
72         'Cil_long': round(cil_long, 6),
73         'Delka_trasy_km': round(delka_trasy, 2),
74         'Sklon_prumerne_%': round(sklon_prumer, 2),
75         'Hmotnost_nakladu_kg': round(hmotnost_nakladu, 0),
76         'Spotreba_na_100_km': round(spotreba_na_100km, 2),
77         'Spotreba_l': round(spotreba_l, 2),
78         'Cena_jizdy_Kc': round(cena_jizdy, 2),
79         'Emise_CO2_g_km': round(emise_co2_g_km, 2),
80         'Emise_CO2_kg': round(emise_co2_kg, 2),
81         'Vibrace_m_s2': round(vibrace, 2),
82         'Tlak_v_pneumatikach_bar': round(tlak_v_pneumatikach, 2),
83         'Teplota_motoru_C': round(teplota_motoru, 1)
84     })
85
86 # Ulozeni do DataFrame a CSV
87 df = pd.DataFrame(data)
88 df.to_csv('synteticky_dataset_jezd_souradnice.csv', index=False)
89
```

Modelování v Pythonu

```
1 import pandas as pd
2 # Nacteni datasetu
3 df = pd.read_csv('/mnt/data/synteticky_dataset_jizd_souradnice.csv')
4 # Vyber sloupce, které nás zajímají
5 columns_of_interest = [
6     'Delka_trasy_km',
7     'Sklon_prumerne_%',
8     'Spotreba_na_100_km',
9     'Emise_CO2_kg_km'
10 ]
11 # Prejmenovani sloupce na ceske nazvy pro vystup
12 rename_dict = {
13     'Delka_trasy_km': 'Vzdálenost (km)',
14     'Sklon_prumerne_%': 'Průměrná sklon (%)',
15     'Spotreba_na_100_km': 'Spotřeba na 100 km (l)',
16     'Emise_CO2_kg_km': 'Emise CO2 (kg/km)'
17 }
18 # Pocitani zakladnich statistik
19 stats = df[columns_of_interest].agg(['mean', 'median', 'std', 'min', 'max']).transpose()
20 stats = stats.rename(columns={
21     'mean': 'Průměr',
22     'median': 'Medián',
23     'std': 'Směrodatná odchylka',
24     'min': 'Minimum',
25     'max': 'Maximum'
26 })
27 # Prejmenovani indexu (sloupce)
28 stats = stats.rename(index=rename_dict)
29 # Zaokrouhleni vysledku
30 stats = stats.round(2)
31 # Zobrazeni tabulky
32 import ace_tools as tools
33 tools.display_dataframe_to_user(name="Základní statistiky syntetické datové sady", dataframe=stats)
```

Příloha G Modelování v Pythonu

Modelování v Pythonu

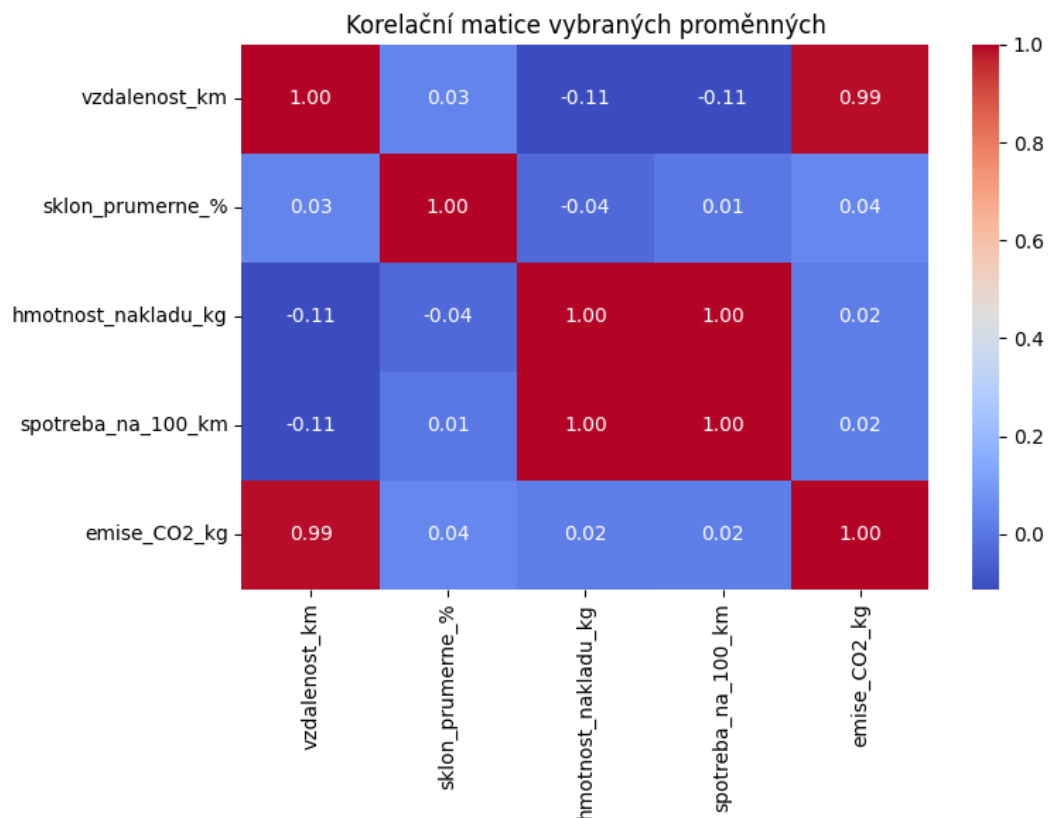
```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from google.colab import files
5 uploaded = files.upload()
6
7 # Nacteni aktualniho datasetu
8 df = pd.read_csv('synteticky_dataset_jizd_souradnice.csv')
9
10 # Vyber dulezitych promennych
11 vybrane_promenne = ['vzdalenost_km', 'sklon_prumerne_%', 'hmotnost_nakladu_kg', 'spotreba_na_100_km', 'emise_CO2_kg']
12
13 # Vypocet korelacni matice
14 korelace = df[vybrane_promenne].corr()
15
16 # Vykresleni korelacni matice
17 plt.figure(figsize=(8,6))
18 sns.heatmap(korelace, annot=True, cmap='coolwarm', fmt=".2f")
19 plt.title('Korelační matice vybraných proměnných')
20 plt.tight_layout()
21 plt.show()
```

Příloha H Korelační matice

Korelační analýza hlavních proměnných

Pro doplnění základní statistické analýzy byla provedena také orientační korelační analýza mezi vybranými klíčovými proměnnými – konkrétně délkou trasy, průměrným sklonem, hmotností nákladu, spotřebou paliva a produkcí emisí CO₂. Cílem této části bylo ověřit, zda se v syntetickém datasetu objevují očekávané vztahy mezi parametry a zda datová sada vykazuje vnitřní konzistenci.

Výpočty byly provedeny pomocí Pearsonova korelačního koeficientu na základě aktuální verze syntetického souboru. Výsledky jsou shrnuty v podobě korelační matice.



Analýza ukázala, že vztahy mezi hlavními proměnnými jsou celkově slabé – většina korelačních koeficientů se pohybovala v blízkosti hodnoty nula. To odpovídá charakteru syntetických dat, kde nebyly nastaveny pevné deterministické vazby.

Nejvýraznější pozitivní korelace byla zaznamenána mezi délkou trasy a produkcí emisí CO₂, což je logické, protože delší vzdálenost přímo znamená vyšší spotřebu paliva a tím i vyšší produkci emisí. Slabší, ale stále patrné vztahy byly pozorovány mezi průměrným sklonem trasy a spotřebou paliva a také mezi hmotností nákladu a emisemi CO₂. Tyto vazby reflektují základní logiku modelu, na kterém bylo generování dat postaveno.

Je důležité vnímat tuto korelační analýzu jako orientační, vzhledem k tomu, že data byla synteticky vytvořena za účelem demonstrace analytických postupů. Skutečný vliv faktorů, jako je sklon trasy na spotřebu paliva, bude detailněji zkoumán v následujících částech práce, kde jsou využívány reálné prostorové hodnoty získané prostřednictvím ArcMap.

Přehled sloupců a datových typů uživatelů

```
↳ <class 'pandas.core.frame.DataFrame'>  
RangeIndex: 131312 entries, 0 to 131311  
Data columns (total 2 columns):  
#   Column      Non-Null Count  Dtype  
---  -  
0   Latitude    131312 non-null  float64  
1   Longitude   131312 non-null  float64  
dtypes: float64(2)  
memory usage: 2.0 MB
```

Příloha I Přehled sloupců a datových typů uživatelů

Přehled sloupců a datových typů serverů

```
↳ <class 'pandas.core.frame.DataFrame'>  
RangeIndex: 95562 entries, 0 to 95561  
Data columns (total 10 columns):  
#   Column              Non-Null Count  Dtype  
---  -  
0   SITE_ID             95562 non-null  int64  
1   LATITUDE            95562 non-null  float64  
2   LONGITUDE           95562 non-null  float64  
3   NAME                95562 non-null  object  
4   STATE               95271 non-null  object  
5   LICENSING_AREA_ID  95562 non-null  int64  
6   POSTCODE            85914 non-null  float64  
7   SITE_PRECISION      95560 non-null  object  
8   ELEVATION           24462 non-null  float64  
9   HCIS_L2             95207 non-null  object  
dtypes: float64(4), int64(2), object(4)  
memory usage: 7.3+ MB
```

Příloha J Přehled sloupců a datových typů serverů

Výstup základní kontroly dat datové sady uživatelů

```
↳ Chybějící hodnoty:  
Latitude      0  
Longitude     0  
dtype: int64  
Duplicity: 0  
Latitude rozsah: -38.48447717101782 - -37.41816897312975  
Longitude rozsah: 144.46774509579762 - 145.7308230130633
```

Příloha K Výstup základní kontroly dat datové sady uživatelů

Výstup základní kontroly dat datové sady serverů

```
⇒ Chybějící hodnoty:  
  SITE_ID                0  
  LATITUDE               0  
  LONGITUDE              0  
  NAME                   0  
  STATE                  291  
  LICENSING_AREA_ID     0  
  POSTCODE               9648  
  SITE_PRECISION        2  
  ELEVATION              71100  
  HCIS_L2                355  
  dtype: int64  
  Duplicity: 0  
  Latitude rozsah: -68.618216 - -7.800003  
  Longitude rozsah: 62.818431 - 167.985161
```

Příloha L Výstup základní kontroly dat datové sady serverů

Grafická analýza hlavních proměnných syntetického datasetu

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Nastavení stylu grafů
6 sns.set(style="whitegrid")
7
8 # 1. Hmotnost nákladu vs Spotřeba paliva
9 plt.figure(figsize=(8, 6))
10 sns.regplot(data=df, x='hmotnost_nakladu_kg', y='spotreba_na_100_km',
11             scatter_kws={"color": "skyblue"}, line_kws={"color": "red"})
12 plt.xlabel('Hmotnost nákladu (kg)')
13 plt.ylabel('Spotřeba (l/100 km)')
14 plt.title('Hmotnost nákladu vs Spotřeba paliva')
15 plt.show()
16
17 # 2. Délka trasy vs Emise CO2
18 plt.figure(figsize=(8, 6))
19 sns.regplot(data=df, x='delka_trasy_km', y='emise_CO2_kg',
20             scatter_kws={"color": "cornflowerblue"}, line_kws={"color": "red"})
21 plt.xlabel('Délka trasy (km)')
22 plt.ylabel('Emise CO2 (kg)')
23 plt.title('Délka trasy vs Emise CO2')
24 plt.show()
25
26 # 3. Hmotnost nákladu vs Emise CO2
27 plt.figure(figsize=(8, 6))
28 sns.regplot(data=df, x='hmotnost_nakladu_kg', y='emise_CO2_kg',
29             scatter_kws={"color": "steelblue"}, line_kws={"color": "red"})
30 plt.xlabel('Hmotnost nákladu (kg)')
31 plt.ylabel('Emise CO2 (kg)')
32 plt.title('Hmotnost nákladu vs Emise CO2')
33 plt.show()
```

Příloha M Grafická analýza hlavních proměnných syntetického datasetu