

**ZÁPADOČESKÁ UNIVERZITA V PLZNI**

**FAKULTA EKONOMICKÁ**

**Bakalářská práce**

**Využití dat pro řízení městské dopravy**

**Data usage for urban traffic management**

**Ondřej Nedvěd**

**Plzeň 2023**

## **Čestné prohlášení**

Prohlašuji, že jsem bakalářskou práci na téma

*„Využití dat pro řízení městské dopravy“*

vypracoval/a samostatně pod odborným dohledem vedoucí/vedoucího bakalářské práce za použití pramenů uvedených v příložené bibliografii.

Plzeň dne 24.04.2023

v. r. *Ondřej Nedvěd*

## **Zásady pro vypracování práce**

1. Zpracujte teoretický úvod do problematiky využití dat v hromadné dopravě.
2. Charakterizujte možnosti sběru a současné využití dat při řízení MHD.
3. Ve spolupráci s konkrétním dopravním podnikem a s využitím příslušných dat (pro vybranou část sítě provozu MHD) analyzujte stávající stav.
4. Zhodnoťte zjištění poznatky a navrhněte doporučení pro další praxi.

## **Poděkování**

Chtěl bych tímto poděkovat vedoucímu bakalářské práce, Ing. Zdeňku Kresovi, za odborné vedení, ochotu a osobní přístup, který mi při zpracování této práce věnoval. Dále bych rád poděkoval Ing. Adamu Šťastnému z PMDP za ochotu při poskytování informací.

# Obsah

<b>Úvod</b> .....	<b>6</b>
<b>1 Data</b> .....	<b>7</b>
1.1 Klasifikace dat .....	7
1.2 Sběr dat a problémy v datovém sběrníku .....	8
1.3 Big data.....	9
1.4 Využití dat .....	11
1.5 Ochrana soukromí .....	13
<b>2 Městská hromadná doprava</b> .....	<b>16</b>
2.1 Klasifikace .....	16
2.2 Měření kvality dopravy .....	18
<b>3 Využití dat v městské hromadné dopravě</b> .....	<b>22</b>
3.1 Nástroje sběru .....	22
3.2 Současné využívání .....	27
<b>4 Využívání dat v PMDP</b> .....	<b>29</b>
4.1 Plzeňské městské dopravní podniky .....	29
4.2 Nástroje sběru a využívání dat.....	30
4.2.1 Systém automatického počítání cestujících (APC).....	30
4.2.2 Systém Sprinter.....	31
4.3 Popis problému .....	31
4.4 Popis vytvořeného programu .....	35
4.5 Analýza dat .....	49
4.5.1 Tramvaje.....	50
4.5.2 Trolejbusy.....	59
<b>Závěr</b> .....	<b>68</b>

<b>Seznam použitých zdrojů .....</b>	<b>70</b>
<b>Seznam tabulek.....</b>	<b>74</b>
<b>Seznam obrázků .....</b>	<b>75</b>
<b>Seznam příloh.....</b>	<b>77</b>
<b>Přílohy</b>	
<b>Abstrakt</b>	
<b>Abstract</b>	

# Úvod

Dnešní doba je charakterizována obrovským množstvím dat, která i díky stále dostupnějším možnostem sběru stále více přibývají. O to více je však důležité se zaměřit na zpracování a využití dat, což platí i pro oblast městské hromadné dopravy. Městská hromadná doprava představuje základní službu, která umožňuje efektivní přepravu osob v rámci měst a aglomerací. V posledních desetiletích dochází k růstu urbanizace a zvýšení dopravní zátěže ve městech, což klade na veřejnou hromadnou dopravu stále větší nároky. V tomto kontextu je důležité, aby dopravní podniky a městské správy efektivně využily dostupná data k optimalizaci a zlepšování služeb, které nabízejí cestujícím.

Tato práce se zabývá využitím dat v městské hromadné dopravě, s cílem poskytnout přehled o sběru a využití dat v MHD, analyzovat provozní data a identifikovat možnosti zlepšení dopravního systému. Práce tak bude demonstrovat, jak efektivní zpracování a využití dat může přispět k lepšímu pochopení provozu hromadné dopravy, identifikaci problémových oblastí a nalezení efektivních řešení. Na základě zjištěných výsledků budou navrženy doporučení pro zlepšení veřejné hromadné dopravy.

V teoretické části se práce zaměřuje na základní pojmy týkající se dat, sběru dat, ochrany soukromí a využití dat v kontextu veřejné hromadné dopravy. V praktické části je popsáno vytvoření vlastního nástroje pro analýzu a vizualizaci dat pro městské dopravní podniky, který je následně aplikován na případovou studii tramvajové a trolejbusové dopravy v Plzni za vybrané období.

# 1 Data

V této kapitole bude čtenář obeznámen s tím, jak jsou data definována a s jejich klasifikací, možnostmi jejich sběru a s případnými problémy, které mohou při práci s daty nastat. Dále pak bude vysvětlen pojem „Big Data“ a základní pilíře, které „Big Data“ tvoří. V části využití dat bude vysvětleno, jak samotná data připravit a následně využít k získání informací a přidané hodnoty. V závěru kapitoly pak bude představena problematika ochrany soukromí při práci s daty, které se v posledních letech věnuje stále více pozornosti.

Pojem data označuje soubor faktů, které jsou většinou získány skrze experimenty, pozorování, transakce nebo zkušenosti. Lze na ně pohlížet jako nejnižší úroveň abstrakce, ze které informace a následně znalosti získávají. (Sharda et al., 2021)

Rowley (2007) zkoumá různé definice informací a jejich vztah k datům. Obecně lze říci, že informace jsou považovány za data, která prošla procesem přeměny, zahrnující klasifikaci, změnu uspořádání, provádění výpočtů, výběr a agregaci a byla uspořádána nebo strukturována, takže jsou smysluplná, hodnotná, užitečná a relevantní pro určitý účel nebo kontext. Tento proces přeměny může být řízen jak lidmi, tak informačními systémy, nebo obojím. Význam informací je navíc subjektivní a určuje jej příjemce, který jim přiřazuje význam na základě svých zkušeností nebo vztahů s jinými údaji.

Zatímco informace jsou často definovány vzhledem k datům, definování znalostí je často složitější. Obecně je lze chápat jako kombinace informací, porozumění, schopností, zkušeností, dovedností a hodnot, které jsou zpracovány a organizovány tak, aby byly užitečné a relevantní pro určitý účel nebo kontext. (Rowley, 2007)

## 1.1 Klasifikace dat

Data se prvotně dělí na strukturovaná a nestrukturovaná, případně semistrukturovaná. Nestrukturovaná a semistrukturovaná data jsou taková, která svým formátem nejsou vhodná pro datovou analýzu počítačem (může se jednat například o obrázky, webový kontent nebo zvuky). Strukturovaná data jsou pak uspořádaná a vhodná k další analýze počítačovými algoritmy. Data dále můžeme dělit na nominální, ordinální, intervalová a poměrová. (Sharda et al., 2021)

- Nominální data se používají pro kategorizaci nebo řazení jednotek do skupin podle určitého kritéria. Mohou být reprezentována pomocí binomiálních hodnot, například pro hodnoty „ano / ne“, „dobrý / špatný“, ... nebo vícenomiální hodnoty pro vyjádření více možných hodnot, například rodinný status (ženatý, rozvedený, svobodný, ...). (Shaughnessy et al., 2000; Turban et al., 2010)
- Ordinální data se stejně jako data nominální využívají k řazení jednotek podle určitého kritéria. Jednotlivým měřením se přiřadí skóre, které reprezentuje pořadí sledovaných prvků, oproti nominálním datum tak poskytují možnost porovnávání mezi sledovanými prvky podle jejich pořadí. Může se jednat například o dosažené vzdělání (základní / střední / vyšší odborné / vysoké). (Shaughnessy et al., 2000; Turban et al., 2010)
- Intervalová data se používají pro změření velikosti nebo intenzity určitého jevu nebo charakteristiky. Dále také umožňují výpočet průměrů a odchylek. Je důležité zmínit, že hodnota 0 neznamena absolutní absenci charakteristiky. Například při měření teploty ve stupnicích Celsia nebo Fahrenheita nelze říct, že 5° je polovina teploty 10°, protože 0 na těchto stupnicích teploty nemá absolutní význam. (Shaughnessy et al., 2000)
- Poměrová data se stejně jako data intervalová využívají pro změření velikosti nebo intenzity určitého jevu nebo charakteristiky. Na rozdíl od dat intervalových jsou však měřeny takovým způsobem, že 0 reprezentuje absolutní absenci dané vlastnosti. Příkladem by bylo při měření teploty využití Kelvinovi stupnice, kde 0 znamená absenci jakékoliv kinetické energie jednotlivých částic. (Shaughnessy et al., 2000; Turban et al., 2010)

## 1.2 Sběr dat a problémy v datovém sběrníku

Dříve se data sbírala hlavně přes dotazníky, či experimenty a pozorování. Dnes se sbírají hlavně pomocí technologií – internet, senzory, ... Stále více se však využívá oportunistické snímání dat, které využívá data na zaběhlých systémech, ale využívá je pro jiné účely. Hlavním takovým systémem se staly mobilní telefony, které díky stále přítomnému používání nahradily potřebu senzorů vybudovaných za specifickým účelem. Stále častěji zanechávají digitální stopy kamkoliv jdou, jak dobrovolně, tak nedobrovolně, skrze jejich přenosná zařízení. Každá aktivita, jako je posílání zpráv



a emailů, telefonní hovory, vyhledávání na internetu, platby kartou, ... poskytují informace o poloze v daném čase. Díky těmto faktorům je množství dostupných dat mnohem větší, ale také se tím zlepšuje jejich kvalita a integrita. (Crist et al., 2015; Sharda et al., 2021)

Při práci se sesbíranými daty mohou nastat některé problémy, na které je třeba brát ohledy, aby negativně neovlivnily následné odhady a rozhodování na nich založené.

Může se jednat například o nedostatek dostupných dat, nebo vysoké náklady na jejich získání. Při použití subjektivních metod jako jsou dotazníky může dojít k poskytnutí nepřesných či dokonce nesprávných údajů. Dalšími možnými problémy jsou pak důvěryhodnost dat, měkkost údajů (soft data), či moc velké množství dat, které může vést k přehlcení informacemi. (Sharda et al., 2021)

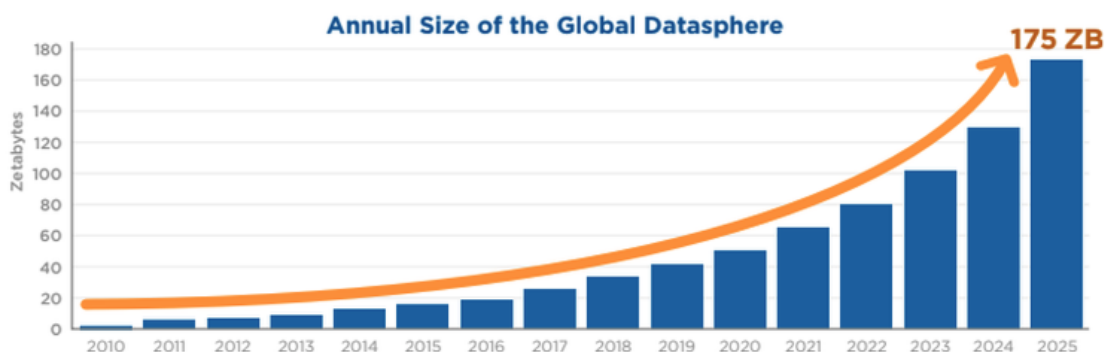
### 1.3 Big data

Díky velkému množství dostupných dat se stále častěji můžeme setkat s pojmem Big data. Big data nejsou určena specifickou velikostí a postrádají jednotnou definici, De Mauro et al. (2016, s.7) je navrhuje definovat na základě nejčastěji se souhlasně vyskytujících prvků napříč různými definicemi jako "informační aktivum charakterizované tak velkým objemem, různorodostí a rychlostí, že je potřeba speciální technologie a analytických metod pro jejich transformaci k užítku." Objem, různorodost a rychlost jsou často označovány jako tři základní pilíře big dat neboli "tři V" z anglického *volume*, *velocity* a *variety*:

- Objem: je s pojmem Big Data spojován nejčastěji. Podle Turner et al. (2014) byl objem dat v digitálním vesmíru v roce 2013 4,4 Zettabajtů ( $4,4 \times 10^{21}$  Bajtů). Rydning et al. (2018) uvádí, že toto číslo vzrostlo na 33 Zettabajtů ( $33 \times 10^{22}$  Bajtů) v roce 2018 a odhaduje růst na 175 Zettabajtů ( $1,75 \times 10^{23}$  Bajtů) do roku 2025 (podrobněji přibližuje Obrázek č. 1). K exponenciálnímu růstu množství dat přispělo několik faktorů. Hlavními jsou zejména zvýšená výpočetní síla a navýšení množství senzorů, které data sbírají. Většina sbíraných dat se týká lidí a událostí (senzory automobilů, platební transakce, text a obrázky ze sociálních sítí, informace o poloze, ...), tyto data pak tedy lze využít k lepšímu porozumění a odhadu lidského chování. Dříve nadměrné množství dat způsobovalo problémy s úložištěm, jak po technické, tak finanční

stránce. Ale se zlepšující se technologií a snižováním cen datových úložišť, se tyto problémy dostaly do pozadí. Hlavním problémem se stalo určení „užitečných“ dat, skrývajících se ve velkém množství nerelevantních, neužitečných dat. Dalším problémem je pak to, jak z těchto dat získat nějakou přidanou hodnotu (Crist et al., 2015; Sharda et al., 2021)

Obrázek č. 1: Růst velikosti digitálního vesmíru



Zdroj: Rydning et al. (2018)

- **Různorodost:** velký objem dat a kombinace různých zdrojů, které se k získávání těchto dat používají, způsobuje další výzvu v souvislosti s jejich zpracováním. Různorodost formátu, typu a struktury dat, se s rozšiřováním nových technologií a nárůstem množství senzorů pouze zvyšuje. Zikopoulos a Eaton (2011) zmiňují, že tato různorodost je výzvou, protože tradiční systémy mají problémy analyzovat a pochopit jiná než strukturovaná data. Historicky se v analýze dat zaměřovalo pouze na strukturovaná data, která dnes tvoří okolo 20 procent všech dat, ale faktem je, že 80 procent dat je nestrukturovaných, případně semistrukturovaných a tento poměr se stále zvyšuje. Různorodost je fundamentální změnou oproti tradiční analýze dat a k využití maximálního potenciálu dat je nutné mít možnost skladovat i analyzovat veškeré typy dat.
- **Rychlost:** nejedná se pouze o rychlost, s jakou jsou nová data generována, ale také jak rychle musí být zpracována. Je to dáno tím, že data postupem času ztrácejí na své hodnotě. Jak rychle hodnotu ztrácí závisí na specifikách jednotlivých dat a podle toho je důležité přizpůsobit rychlost jejich ukládání a analýzy a – často se jedná o nejnáročnější část práce s Big Daty. (Sharda et al., 2021).

- Pravdivost: pojednává o záruce kvality nebo kredibility sesbíraných dat. Je závislá na důvěryhodnosti zdroje, kredibilitě informace a pravdivosti obsahu. (Moutassem et al., 2022).

Na rozdíl od "třech V", pravdivost není vždy uváděna jako jeden z pilířů Big dat, je nutné ji však při analýze dat brát v potaz. Jak tvrdí Schroeck et al. (2012): „Zahrnutí pravdivosti jako čtvrtého základního atributu Big dat klade důraz na důležitost práce s neurčitostí, která je pro některé typy dat inherentní. Snažit se o získání dat s vysokou kvalitou je důležitou součástí big dat a také velikou výzvou, avšak i nejlepší metody na očišťování dat nemohou odstranit inherentní nepředvídatelnost některých dat a je tedy nutné brát pravdivost jako jejich další dimenzi.“

## 1.4 Využití dat

Při práci se samotnými daty je nutné je nejprve očistit od špatně zadaných hodnot, či extrémních údajů. Ty mohou vzniknout v důsledku ztráty informace, nebo v důsledku vynechání odpovědi účastníků studie. Chybějící data vedou ke zmenšení vzorku, případně mohou vést ke zkreslení dat. (Kwak & Kim, 2017)

Chybějící data lze rozdělit podle způsobu výskytu, Little a Rubin (1987) je rozdělili do těchto kategorií:

- MCAR z anglického „missing completely at random“, jedná se o chybějící data, kde důvod ztráty není nijak spojen s žádnou pozorovanou, či jakoukoli další proměnnou. Chybějící data jsou tak zcela náhodná a nezpůsobují zkreslení statistické analýzy.
- MAR z anglického „missing at random“ jsou chybějící data, která jsou závislá na pozorovaných proměnných, ale jsou nezávislá na chybějící proměnné samotné.
- NMAR z anglického „not missing at random“ jsou chybějící data, kde důvod nepřítomnosti je závislý na samotné hodnotě chybějících dat. Tato chybějící data je nejobtížnější nahradit, ale zároveň způsobují největší zkreslení, takže je nelze ignorovat.

Scheffer (2002) diskutuje výhody a nevýhody různých přístupů k odstranění chybějících dat:

- Odstranění: jedná se o kompletní vymazání záznamu s chybějícími daty relevantními pro danou analýzu. Tento postup by měl být zvolen pouze pokud se jedná o MCAR. U MAR a NMAR by pouhé odstranění hodnoty vedlo ke zkreslení dat.
- Doplnění hodnoty: výhoda doplnění je, že minimalizuje zkreslení a využívá data na jejichž sesbírání byly vynaloženy náklady. Doplněním hodnot se také dosáhne čtvercového data setu, se kterým pak lze pracovat pomocí běžného softwaru a metod. Doplněná data však musí odrážet, že se nejedná o opravdová data. To musí být zohledněno při počítání odhadu rozptylu. Doplnění lze rozdělit na Single Imputation(SI) a Multi Imputation(MI), kde SI je méně náročná, ale také méně přesná a měla by se používat maximálně pro MAR, kde maximálně 10 % hodnot chybí.

Druhý problém, který je třeba před prací s daty vyřešit, jsou extrémní údaje. Jedná se o data s hodnotami, které se abnormálně vymykají celkovému vzorci rozdělení proměnných. Tyto údaje mohou vzniknout špatně zadanou odpovědí, nebo chybou zařízení. Bez řádného ošetření extrémních hodnot pak dochází ke zkreslení statistických dat. (Kwak & Kim, 2017)

Přístupů k problematice odstranění extrémních hodnot je mnoho. Při výběru správného postupu a vhodného algoritmu na jejich odhalení závisí na konkrétním případě a na typu dat (podrobněji se problematice věnují Dixon 1953; Ilango et al. 2012; Gupta et al. 2014; Zhang et al. 1996). Obecně se však jedná o relativně přímý proces:

- Určení hladiny významnosti,
- vybrání vhodného postupu pro výpočet odchylky dat,
- v závislosti na situaci označit data s extrémními hodnoty za chybná měření, nebo za význačné jedince

Po ošetření a analýze dat, je často nutné komunikovat výsledky šetření se stakeholdery nebo širokou veřejností. Ke komunikaci těchto dat se používá datová vizualizace.

Zvolení vhodné vizualizace je pro reprezentaci dat klíčové. Každý typ dat vyžaduje jedinečný přístup a metodu vizualizace, aby bylo zajištěno, že budou předány znalosti, které obsahují. Dalším důležitým aspektem je uvědomit si, pro koho je daná vizualizace tvořena (tj. tvořit ji tak, aby se v ní odrážely informace, které jsou relevantní). V současné době musí velké množství manažerů a lidí ve vedoucích pozicích činit několik rozhodnutí

denně, často jen na základě poskytnutých vizualizací. Je tedy důležité zajistit, aby šíře a rozsah vizualizace nevynechávala důležité informace, nepřeháněla některé údaje a aby vizualizace přesně odrážela informace obsažené v datech. (Azzam et al., 2013; Burnay et al., 2019)

Vizualizace dat není nástroj pouze pro komunikaci dat s ostatními. Právě sama vizualizace je často používána například pro identifikaci extrémních hodnot, ale také pro nalezení vzorců chování, či odhalení neobvyklých distribucí v datech (které by běžné statistiky nebo modely nemusely odhalit). Správně vytvořená vizualizace pak může hrát zásadní roli při analýze komplexních modelů, ve kterých může pomoci diagnostikovat silné a slabé stránky. (Unwin, 2020)

## 1.5 Ochrana soukromí

Soukromí můžeme definovat jako privilegium kontroly nad tím, jaké osobní informace jsou sbírány a používány. Soukromí informace je pak schopnost skrýt šíření informace před jedinci, se kterými danou informaci nechceme sdílet. (Porambage et al., 2016)

Soukromí je základní lidské právo, které je v mnoha zemích chráněno zákonem. S rostoucí závislostí na technologiích a se šířením dat je důležité zajistit, aby bylo osobní soukromí respektováno a chráněno. To je zejména relevantní v kontextu veřejné dopravy, jejíž provozovatelé mohou shromažďovat velké množství dat o osobách, které tyto služby využívají.

Tyto obavy vychází zejména z množství dat, které je sbíráno bez vědomí a souhlasu dané osoby. I z dat, které se zdají být anonymní, se pomocí křížových odkazů na další zdroje dat v okolí dá dohledat specifická osoba a její lokace, na což může být pohlíženo jako na narušení osobního soukromí. (Crist et al., 2015)

V roce 1980 OECD (2022) zavedla pokyny, které by se měly dodržovat při zacházení s daty. Tyto pokyny udávají, že by se měly dodržovat následující principy:

- Omezení sběru: shromažďování osobních údajů by mělo být omezeno a všechny tyto údaje by měly být získávány zákonnými a spravedlivými prostředky a případně s vědomím nebo souhlasem subjektu údajů.
- Kvalita dat: Osobní údaje by měly odpovídat účelům, pro které mají být použity, a v rozsahu nezbytném pro tyto účely by měly být přesné, úplné a aktualizované.

- Specifikace účelu: Účely, pro které jsou osobní údaje shromažďovány, by měly být specifikovány nejpozději v okamžiku shromažďování údajů a následné použití by mělo být omezeno na splnění těchto účelů nebo jiných účelů, které nejsou s těmito účely neslučitelné a které jsou specifikovány při každé změně účelu.
- Omezení použití: Osobní údaje by neměly být zveřejňovány, ani jinak používány k jiným účelům než pro účel, pro který byly specifikovány, s výjimkou udělení souhlasu a moci zákona.
- Záruka bezpečnosti: Osobní údaje by měly být chráněny přiměřenými bezpečnostními opatřeními proti rizikům, jako je ztráta nebo neoprávněný přístup k údajům, jejich zničení, použití, změna nebo zveřejnění.
- Otevřenost: Existence obecné politiky otevřenosti o vývoji a postupech v oblasti osobních údajů. Jednotlivec by měl mít přístup k informacím o jeho osobních údajích a o tom, kdo je má k dispozici a kdo je používá.
- Individuální účast: Jedinec by měl mít právo vědět, zda správce údajů má údaje o něm a v případě žádosti je správce nucen ve srozumitelné formě k těmto datům poskytnut přístup. Při neposkytnutí má jedinec právo napadnout neochotu poskytnutí dat, nebo jejich nepřesnost a požadovat, aby byla data napravena nebo odstraněna.
- Odpovědnost: Správce údajů by měl být odpovědný za dodržování výše zmíněných opatření.

Dodržování těchto zásad však neznamená, že přístup k ochraně dat je univerzální. Právní rámce států přistupují k ochraně dat nejednotně, a tak nelze uplatnit jednotný přístup a ochrana soukromí musí být vždy řešena v kontextu daného právního rámce, což může představovat výzvu pro organizace, která se snaží poskytovat služby a shromažďovat data na mezinárodní úrovni. Miranda-Segundo et al. (2020) také upozorňuje na možné rozdíly v legislativním přístupu k ochraně napříč odvětvími v rámci jednoho státu.

Evropská unie je často považována za vůdce v oblasti ochrany dat. Dopady změn v legislativě Evropské unie o ochraně dat jsou nesporné. Několik států bylo nutno v reakci měnit a modernizovat stávající legislativu a GDPR, zavedený v roce 2018, je obecně považován za „zlatý standard“. (Mantelero, 2021)

Hlavním cílem GDPR a dříve zavedené směrnice 95/46/EC je harmonizace právních předpisů napříč členskými zeměmi na ochranu údajů, předávání osobních údajů zemím mimo Evropskou unii a zlepšení práv jedinců, (jako možnost vymazání a přístup k osobním datům). Společnosti tak musí mít souhlas k pořizování a používání osobních dat a jsou nuceny implementovat dostatečná opatření na zabezpečení sesbíraných dat. (Miranda-Segundo et al., 2020; Regulation, 2016)

Navzdory překážkám a znepokojením nad narušováním soukromí, sbírání dat má nepopiratelný přínos pro zlepšení bezpečnosti, provozu a plánování dopravy. (Crist et al., 2015)

Sbírání dat je tedy klíčové pro zlepšování dopravy, avšak za rizika narušení práv, Crist et al. (2015) doporučuje v rámci prevence implementaci řady strategií na anonymizaci a de-identifikaci s využitím co nejvíce možnými robustními techniky, avšak upozorňuje, že ani jejich využití nemůže kompletně zaručit anonymitu a vyloučit zneužití získaných dat.

## 2 Městská hromadná doprava

Obecně můžeme městskou hromadnou dopravu definovat jako typ transportu ve městě, který se většinou pohybuje po stanovených trasách, na kterých zastavuje v určeném místě a čase a je dostupný pro všechny, kdo jsou ochotni zaplatit stanovené jízdné. Počátky organizované městské hromadné dopravy můžeme najít už v 16. století, kdy se prvky využívané jen pro přepravu mezi městy, začínají používat i ve městech samotných. Mezi první rozšířené městské dopravní prostředky patřily trajekty, zejména v Londýně a Benátkách. Dalším prostředkem byla jízda na koni, s postupem času se však začala přidávat různá tažná zařízení, aby mohlo být přepraveno více zákazníků. (Vuchic, 2007) V roce 1881 byla v Německu uvedena do provozu první tramvaj tažená koňmi, která se během počátku 19. století stala běžnou součástí měst. Tramvaje se postupem času elektrifikovaly a daly tak možnost městům se rozšiřovat. Místa umístěná poblíž zastávek se začala zhodnocovat a zlepšila se dostupnost pracovišť, která už nemusela být v docházkové vzdálenosti. (Knowles, 2020; Kupaysinov, 2022)

Městská hromadná doprava tak nepředstavuje pouze způsob přepravy z bodu A do bodu B, ale mnohem více. Jedná se o vlivnou sílu, která pomáhá tvarovat města a zároveň vytváří sociálně, environmentálně a ekonomicky udržitelné komunity. (Van Lierop et al., 2018)

Ashmore et al. (2019) argumentuje i impakt městské hromadné dopravy na status lidí napříč různými kulturami, kde je sice jízda metrem v Číně brána jako pozitivní, avšak pohled na cestující v autobusu, který je spojován s chudobou, je často negativní. Toto může mít daleké implikace a je proto důležité zajistit takové prostředí, kde všechny skupiny budou motivovány využívat MHD. (Daniels & Mulley, 2013)

### 2.1 Klasifikace

Městská hromadná doprava může být klasifikována mnoha způsoby. Výběr vhodného typu závisí na konkrétních potřebách systému. Je důležité si stanovit cíl, kterého se musí dosáhnout a následně se rozhodnout, který typ nejlépe vyhovuje danému problému. V případných analýzách s už vytvořenou infrastrukturou se musí brát ohledy na různá omezení jednotlivých kategorií.



Jednou z těchto klasifikací je rozdělení městské hromadné dopravy podle přednosti v jízdě, kterou Vuchic (2007) dělí do 3 kategorií:

- Kategorie A: Plná přednost v jízdě a bez žádných dalších omezení v podobě různých přechodů s dalšími účastníky provozu. Může se jednat například o metro či nadzemní dráhu.
- Kategorie B: Plná přednost v jízdě, fyzické oddělení od dalších účastníků provozu, avšak jsou zde omezení v podobě různých přejezdů, ve kterých se cesty protínají s dalšími účastníky provozu. Jedná se například o spoje vlaků v rámci města. Dále se může také jednat o tramvaje, které ale mohou částečně spadat i do kategorie C, v závislosti na implementaci.
- Kategorie C: Zde je hromadná doprava plně integrovaná s ostatní dopravou a využívá stejných komunikací a pravidel. MHD může být například zvýhodněna preferenčním pruhem, který však není oddělen žádnou fyzickou překážkou, ale pouze dělicí čarou. Jedná se například o autobusy.

Dalším možným rozdělením je členění vozidel podle typu pohonu, a to na motorová, elektrická a kombinovaná. Elektrická vozidla lze dále rozlišovat na vozidla závislá na stálém přívodu proudu, bateriová a vozidla s pohony založenými na principu palivového článku. Každý typ pohonu má specifika, která je nutné zvážit při rozhodování o vhodnosti pro konkrétní situaci.

Výhodou elektrického pohonu je nulová produkce škodlivých látek, čímž vozidla přispívají k lepšímu ovzduší. Je nutné však podotknout, že to nutně neznamená, že jsou více ekologická. Produkce skleníkových plynů bude záležet na způsobu produkce elektřiny, kterou budou vozidla využívat. Další výhodou elektrického pohonu je tichost provozu, čímž se snižuje hlukové znečištění. (Kalghatgi, 2018; Tzeng et al., 2005)

I když je tichost elektrického motoru uvedena jako jedna z jeho výhod, je důležité poznamenat, že tišší typ pohonu nezajišťuje tišší běh vozidla. Například Michta a Haniszewski (2018) ve své studii uvádí průměrnou hlasitost na místě pasažéra tramvaje během provozu na ~75 dB, zatímco průměrnou hlasitost autobusu na místě pasažéra uvádí na ~60 dB.

Většina benefitů z elektrických pohonů platí pro všechny typy, nevýhody jsou však závislé na konkrétním typu pohonu. U elektrických baterií je nevýhoda omezený dojezd a dlouhé dobíjení baterií, které při špičce mohou přispívat k zatížení rozvodných sítí.

Produkce současných baterií je také závislá na těžbě lithia, která má negativní vliv na životní prostředí. (Clairand et al., 2019) Nabíjecí cyklus také poškozují kapacitu baterie, která postupem času tedy degraduje a má jen omezenou životnost. (Wang et al., 2016)

Dlouhý nabíjecí cyklus se dá řešit pomocí výměnných stanic baterií, kde namísto dobíjení je použitá baterie vyjmuta a nahrazena baterií nabitou. Celý tento proces se odehrává v řádu minut. Další možnou výhodou tohoto systému by byla možnost nabíjet baterie přes noc, při nižších cenách energií a nenabíjet tak v dobách vysokého zatížení. Výměna baterií má však také nedostatky. Baterie je nutné upravit, aby je bylo možné nahrazovat a musí být dobře přístupné, což omezuje možnost umístění baterie ve vozidle. Další nevýhodou jsou samotné náklady spojené s výstavbou výměnných stanic a množstvím baterií navíc, které je nutno mít. (Ahmad et al., 2020)

Pohonů na principu palivových článků existuje více druhů a většina z nich využívá vodík jako hlavní palivo. Nevýhodou palivových článků je extrakce samotného vodíku, nákladná infrastruktura potřebná ke zkapalnění, transportu a následnému uložení. Tento proces je také velice neefektivní, energie využitá pro pohon vozidla tvoří přibližně 30–40 % původní energie vynaložené k získání vodíku. (Kalghatgi, 2018; Kirubakaran et al., 2009; Manoharan et al., 2019)

Hlavní nevýhodou vozidel závislých na přímém proudu pak jsou vysoké náklady na výstavbu potřebné infrastruktury a nízká flexibilita vozidel. (Vuchic, 2007)

Co se nákladů na běžný provoz týče, Vuchic (2007) udává, že nelze určit provozně výhodnější pohon, vzhledem ke stálým změnám cen potřebných komodit a rozdílným cenám v závislosti na geografii.

Žádný pohon tak není absolutně lepší než ostatní a vhodný výběr bude záviset na infrastruktuře a specifických potřebách. Výběr správného pohonu je důležitý, aby bylo dosaženo optimálního výkonu a minimalizovaly se dané nevýhody.

## **2.2 Měření kvality dopravy**

Měření kvality městské hromadné dopravy je důležité pro určení, jak dobře služby hromadné dopravy splňují potřeby a preference cestujících a jak efektivně jsou provozovány. Existuje mnoho různých metod hodnocení kvality veřejné hromadné dopravy, které se dají rozdělit do několika hlavních oblastí: kvalitu služeb, dostupnost a přístupnost, bezpečnost a pohodlí, efektivnost a účinnost, environmentální faktory

a externality. Každá z těchto oblastí přináší specifické výzvy a cíle, které jsou nutné pro dosažení celkového zlepšení kvality městské hromadné dopravy.

Vuchic (2007) vymezuje několik elementárních atributů, na základě kterých lze pak provádět další výpočty a analýzy kvality dopravy, které se zaměřují hlavně na efektivnost a účinnost, lze však tyto atributy využít i pro analýzy bezpečnosti, pohodlí a dostupnosti. Těmito elementárními atributy jsou:

Rychlost:

$$v = \frac{s}{t} \quad (1)$$

Pomalost:

$$v^{-1} = \frac{t}{s} \quad (2)$$

Hustota:

$$k = \frac{u}{s} \quad (3)$$

Rozestupy:

$$s_s = \frac{s}{u} \quad (4)$$

Frekvence:

$$f = \frac{u}{t} \quad (5)$$

kde:  $u$  ... počet objektů

$s$  ... vzdálenost

$t$  ... čas

Dále pak definuje několik z nich plynoucích ukazatelů kvality:

$$w = u * s \quad (7)$$

Kde:  $w$  ... *dopravní výkon*

Dopravní výkon se měří za časové období. Většinou rok, ale může se měřit i na měsíční, denní nebo hodinové bázi.

$$P_v = C_v * V_c \quad (8)$$

Kde:  $P_v$  ... produktivita vozidla

$C_v$ ... kapacita vozidla

$V_c$  ... průměrná rychlost vozidla včetně stání

Produktivita vozidla udává práci vykonanou vozidlem za jednotku času při běžném provozu.

$$P_L = N * P_v = N * C_v * V_c \quad (9)$$

Kde:  $P_L$ ... produktivita linky

$N$  ... počet vozidel na dané lince

Produktivita linky udává práci vykonanou všemi vozidly na dané lince za jednotku času při běžném provozu.

Vzhledem k možnému využití různých jednotek jak na vstupu (peníze, energie, práce, ...), tak na výstupu (osobo-kilometry, vozidlo-kilometry, ...), definuje efektivitu obecně jako:

$$\text{Poměr efektivnosti} = \frac{\text{množství vyprodukovaného výstupu}}{\text{množství spotřebovaných zdrojů}} \quad (10)$$

Stejně obecně je pak nadefinována míra spotřeby, která udává množství využitého zdroje v procesu vůči specifickému výstupu daného procesu. Tato funkce je inverzní k funkci poměru efektivnosti a je definována takto:

$$\text{Míra spotřeby} = \frac{\text{množství využitého zdroje}}{\text{množství vyprodukovaného výstupu}} = \frac{1}{\text{poměr efektivnosti}} \quad (11)$$

Kvalitu městské hromadné dopravy nelze však měřit pouze „tvrdými“ údaji přímo vyplývajícími z měření jejího provozu. Při rozhodování o změnách za účelem nějaké optimalizace je důležité vzít v potaz těžko měřitelné údaje, většinou v podobě „měkkých“ údajů, jako je spokojenost zákazníků a externality, který provoz městské hromadné dopravy přináší.

O kvantifikování spokojenosti usiluje například NPS (Net Promoter Score), který je využíván velkým množstvím společností jako hlavní indikátor spokojenosti zákazníka. (Rockes, 2016) a je založen na principu formuláře, dotazujícího se vzorku zákazníků na to, s jakou pravděpodobností by doporučili společnost někomu jinému. NPS skóre pak udává rozdíl mezi zákazníky, kteří by společnost doporučili a kteří ne. (Ritson, 2006) Fisher a Kordupelski (2019) upozorňují, že ač je NPS dobrý na měření spokojenosti

stávajících zákazníků, tak neposkytuje vhodné informace pro určení směru dalšího zlepšení a neposkytuje informace potřebné k získání nových zákazníků. NPS má také problémy se zachycením nespokojených zákazníků (Bitencourt et al., 2023). Esklidsen a Kristensen (2011) poukazují na obecnou nespolehlivost tohoto systému a doporučují společnostem přehodnocení využívání této metriky.

Rozvoj a motivace veřejné dopravy přináší značné výhody, protože cestování automobilem je spojeno s vysokými náklady. Zlepšování kapacity veřejné hromadné dopravy a propagace alternativních způsobů cestování přispívají k přilákání cestujících, kteří by jinak volili osobní automobil, a tím pomáhají snižovat dopravní zácpy, zlepšovat životní prostředí a ušetřit náklady. Jinak řečeno, samotný provoz veřejné hromadné dopravy nepřináší velké výhody sám o sobě. Většina výhod závisí na tom, jak se veřejná doprava využívá, jak dobře se přizpůsobuje potřebám a preferencím zákazníků, na množství nahrazených aut a na dalších přínosech a úsporách, které z tohoto vyplývají. (Litman, 2015)

Litman (2015) dále udává tabulku dopadů, které by se měly zvážit při evaluaci konkrétních projektů. Vizte příloha A.

Měření těchto kvalit nemá univerzální přístup, je důležité se rozhodnout, které z těchto dopadů jsou důležité pro řešený projekt a následně zvolit vhodné metriky a nástroje pro každou z těchto oblastí a získat data, která umožní jejich vyhodnocení. Výsledky měření kvality pak mohou zajistit, že služby veřejné dopravy jsou zaměřeny na splnění daných cílů a poskytují co nejlepší hodnotu pro cestující a společnost.

### 3 Využití dat v městské hromadné dopravě

Tradičně byla veřejná doprava závislá na manuálních dotaznících, které sbírala data od cestujících (tato data jsou vhodná pro získání údajů o socio-demografických charakteristik respondentů jako je věk, příjem a zaměstnání spolu s účelem cesty). Nicméně tyto dotazníky jsou finančně náročné, což vede k malým vzorkům a málo časté obnově dat. Dále je zde riziko chyby v odpovědích a ve statistické zaujatosti. Hlavním problémem těchto dat je však zachycení pouze jednoho okamžiku v čase a z toho vycházející nemožnost zachytit krátkodobé či dlouhodobé variace v používání veřejné dopravy. (Zannat & Choudhury, 2019)

Některá data mohou být využívána za specifickým účelem – například výstavba senzorů na měření rychlosti. Nicméně, většina hodnoty těchto dat se skrývá ve využití v kombinaci s daty z dalších zdrojů. Tato data, ani jejich specifické využití nemusí být konkrétně definováno. Jedná se spíše o digitální fragmenty informací, které jsou zanechány z interakcí s informačními systémy a infrastrukturou. Kombinace těchto dat ukazuje nečekané a neobjevené vzorce v běžných životech jedinců a společnosti. Z této skutečnosti plynou také nová rizika, kde využívání těchto vzorců by mohlo vést k manipulaci a zneužití dat jedinců. Stále narůstající množství a rychlost dat poskytuje jak nové příležitosti, tak také nové hrozby a výzvy. (Crist et al., 2015)

#### 3.1 Nástroje sběru

Výběr vhodného nástroje pro sběr dat závisí na cíli a potřebách plánování a řízení veřejné dopravy. Je důležité zvážit, jaké informace jsou potřeba získat, a vybrat nástroj, který je nejlépe schopen tyto informace poskytnout.

Jak bylo řečeno dříve, dotazníky jsou původním způsobem sběru dat o veřejné dopravě. Jedná se o získaný vzorek z cílené skupiny lidí, kteří byli ochotni podstoupit daný dotazník. Data z dotazníků mají své nevýhody. Jsou většinou kapitálově a pracovně náročné. I přes velké investice a velké množství práce však nedosahují velikosti vzorku, který poskytují automatické systémy na sbírání dat.

I dnes mají však stále využití, například kvůli množství faktorů, které ovlivňují lidské chování, které nejsme schopni zachytit pomocí matematických a automatických systémů. Nebo pro kombinaci s automaticky sesbíranými daty, které se tak rozšíří o části, které

nedokážou sami zachytit. Dále mohou taky být využívány k učení různých algoritmů, díky velkému množství dotazníků, které bylo dříve provedeno a zkoumáno. (Ge et al., 2021)

Dalším klasickým způsobem počítání pasažérů je manuální počítání cestujících zaměstnanci. Tento způsob, stejně jako dotazníky, je velice nákladný, poskytuje malý vzorek a je potenciálně nepřesný, vzhledem k častému využití méně kvalifikovaných pracovníků i rychlému střídání velkého počtu cestujících ve vozidle. Již první generace systémů automatického počítání se ukázala jako přesnější varianta. Systém manuálního počítání je stále občas využíván, hlavně při implementaci nových automatických systémů pro ověření jejich funkčnosti. (Siebert & Ellenberger, 2020)

Jedním z automatických systémů na počítání pasažérů je využití cestovních karet, jejichž primárním cílem je automatické vybírání poplatků za jízdu. Tyto karty fungují na radiofrekvenční identifikaci, při které cestující musí kartou pípnout při vstupu a i výstupu. Existují různé druhy karet, od základních, které snímají pouze informace při vstupu, jako je čas nástupu, zastávka a způsob transportu až po mnohem sofistikovanější, jako například London Oyster card, která sbírá data jak nástupů, tak výstupů, což otevírá možnosti pro další využití dat, které se nelimituje pouze na sbírání cestovného. (Gordon et al., 2013; Zannat & Choudhury, 2019). Je nutné poznamenat, že mnoho systémů fungujících na systému cestovních karet povoluje placení hotovostí nebo papírovým jízdním dokladem, a cestující využívající tyto způsoby úhrady tak nejsou evidováni, což vede k nepřesnostem v sesbíraných datech. Evidenci tak mohou unikat například určité skupiny cestujících, které se neadaptovaly na moderní systém. V České republice není uplatňován princip radiofrekvenční identifikace a v případě zájmu o data z cestovních karet tak dopravce musí cestujícím uložit povinnost označovat nástup a výstup přiložením karty k terminálu (tento princip uplatňuje dopravní podnik v Pardubicích).

Automatická počítadla pasažérů (APC z anglického automatic passenger counters) počítají pasažéry automaticky při nástupu a výstupu. APC zajišťuje podobná data jako data získaná z cestovních karet. Výhodou je však absence nutnosti používání karty. Cestovní karty se nepoužívají všude, takže se jedná o technologii více všestrannou, pro kterou není potřeba další zaběhlé infrastruktury. Dále klade menší nároky na samotné cestující, kteří nemusí používat kartu a jsou počítáni pasivně. (Ge et al., 2021)

APC systémy lze rozdělit na přímé a nepřímé. Nepřímé systémy lze dále rozdělit na technologie založené na hmotnosti a technologie založené na mobilních zařízeních. Technologie založené na hmotnosti fungují na principu změny celkové hmotnosti vozidla a odhadu počtu pasažérů pomocí výpočtu. Evidence pomocí mobilního zařízení je založeno na počítání množství Wi-Fi, Bluetooth, nebo jiných signálů zachycených ve vozidle. Hlavní výhodou těchto systémů jsou nízké náklady. Hlavní nevýhodou je nepřesnost, kde výpočet je pouhý odhad. Systém na bázi hmotnosti navíc zachycuje pouze stav, nepodává tak informace o toku pasažérů. (Myrvol et al., 2017; Olivo et al., 2019)

Přímé systémy používají primárně tři technologie. Může se jednat o APC senzory založené na infračervené technologii, které buď snímají odrazy objektů (mají však problémy s černou barvou) nebo snímají změny tepla (tyto senzory mají nevýhodu v tom, že bez pohybu nemůže být žádná detekce). Pro nejlepší přesnost je vhodné využít kombinaci obou senzorů. Druhou technologií, která se využívá pro APC, jsou šlapací podložky, které jsou umístěny na vchodech a zachycují šlápnutí pasažérů při výstupu a nástupu. (Perk et al., 2003). Nejnovější technologie, která se v souvislosti s APC používá, je počítání na základě videozáznamu. Jednou z výhod této technologie je, že poskytuje nejen informace o toku pasažérů na dané zastávce, ale také umožňuje sledování výstupu a nástupu konkrétních pasažérů. Tyto dodatečné informace jsou poté využitelné pro podrobnější sledování toku pasažérů a analýzu cestovního chování. (Hsu et al., 2020) Je důležité zdůraznit, že při využívání této technologie je nutné brát v úvahu ochranu soukromí a dat. Sběr a zpracování dat by měly být v souladu s platnými právními předpisy a měly by respektovat právo jednotlivců na ochranu soukromí.

Další důležitá data pro analýzy provozu MHD představuje poloha vozidla. Automatická poloha vozidla (AVL z anglického automatic vehicle location, také AVM z anglického automatic vehicle monitoring) označuje systémy na sledování pozice každého vozidla v reálném čase a odesíláním informací o daném vozidle do centrály. Nejvíce využívanou technologií pro sledování je GPS, která funguje na základě sítě satelitů obíhajících Zemi, které posílají signál na Zem a speciální přijímač umístěný na vozidle přijme tento signál a vypočítá pomocí něj zeměpisné souřadnice vozidla. AVL systémy jsou velice rozšířené ve většině světa pro sledování hromadné dopravy, protože poskytují data v reálném čase, většina technologií využívá obnovu polohy každých 10 nebo každých 30 sekund. (Ge et al., 2021)



Relativně nově využívanou technologií pro sběr informací v dopravě se stal také LiDAR z anglického light detection and ranging (světelná detekce a měření rozsahu), kdy vozidlo měří vzdálenosti okolních objektů pomocí doby šíření laserového paprsku. Tyto informace, v kombinaci s informacemi z GPS, pak poskytují cenné informace o stavu dopravy v reálném čase. (Cui et al., 2019; Zheng et al., 2019)

Další možností získávání dat jsou mobilní telefony, které jsou velice rozšířené a jejich používání se pouze zvyšuje, podle Bachir et al. (2019) se jedná o největší zdroj dopravních dat, který však zatím nebyl plně integrován do plánovacích modelů. Jednou z překážek je náročnější očištění dat a následné rozlišení cestujících využívající MHD a řidičů aut.

Data z chytrých telefonů můžeme rozdělit na 2 zdroje – telekomunikační a na data z aplikací.

Telekomunikační data jsou dostupná tím, že telefony musí navazovat spojení, aby mohly přijímat a odesílat zprávy a hovory, tato spojení jsou zaznamenávána mobilními operátory. Díky velkému množství stále připojených lidí, tyto záznamy poskytují relativně dobrý odhad toho, kde se majitelé těchto zařízení nachází. (Morlion, 2021)

V současné době se v městské hromadné dopravě využívají dva typy telekomunikačních dat – Call Detail Record (CDR) a Global System for Mobile Communication (GSM) data. CDR data poskytují vykonávanou činnost (hovor, zpráva, přístup na internet) a k tomu poskytují odpovídají čas a lokaci mobilní věže, která žádost zprostředkovává. GSM data jsou generována kontinuálně, dokud je zařízení aktivní. GSM data tím pádem obsahují větší množství informací, ale jsou obtížněji dostupná. CDR data jsou snadno dostupná, vzhledem k tomu, že jsou ukládána telefonními společnostmi k vyúčtovacím účelům. (Zannat & Choudhury, 2019)

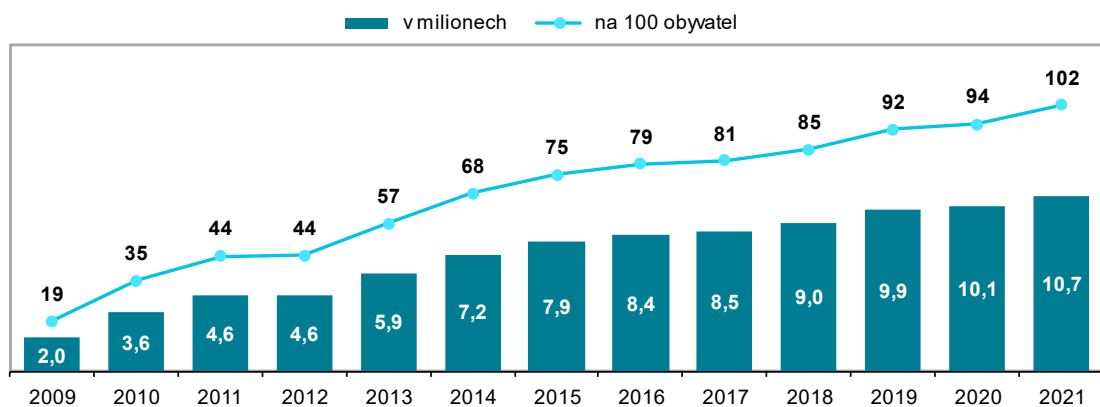
Původně byla data získávána z aplikací určených pro navigaci v dopravě jediným zdrojem pro účely hromadné dopravy. Nicméně, postupem času se začaly objevovat další typy aplikací, které prodávají data svých uživatelů. Mezi první takové aplikace patřily například aplikace zaměřené na měření aktivity, jako jsou Strava nebo Garmin, kde uživatelé úmyslně sdílejí svou polohu a další informace, které mohou být použity pro účely hromadné dopravy. Nyní se však stále více běžných aplikací snaží sbírat a ukládat co nejvíce údajů o uživateli, včetně jeho polohy. Tento sběr údajů však není často transparentní a mohou tak nastat problémy ohledně ochrany soukromí. (Morlion, 2021)

Velké množství mobilních zařízení dává přístup k velkému množství cenných dat, je však nutné v dalších analýzách a rozhodováních zahrnout, zda je zastoupení různých skupin rovnoměrné. Jak uvádí Český statistický úřad (2020), v roce 2020 používalo chytrý telefon 69,6 % lidí starších 16 let, což poskytuje dobrý vzorek populace, avšak zastoupení občanů starších 65 let bylo pouze 19,8 %.

Zannat a Choudhury (2019) poukazují na to, že aplikační data jsou obzvláště náchylná na zaujatost a je vhodné je spíše využívat na odhad fluktuací při speciálních událostech jako například olympiáda nebo jiné větší sportovní události.

Ze současného vývoje, který je možno sledovat na obrázku č. 3, lze odhadovat, že používání mobilních telefonů bude stále rozšířenější a s postupem času se reprezentace jednotlivých skupin bude vyrovnávat.

Obrázek č. 2: Účastníci s mobilním přístupem k internetu v Česku



Zdroj: Český statistický úřad (2022)

Ge et al. (2021) zmiňuje další možné zdroje, kterým se zatím nevěnuje moc pozornosti. Radí mezi ne například:

- Počasí – data o počasí mají vliv na dopravu, kvůli vlivu počasí na nabídku a poptávku po dopravních službách. To by v budoucnu mohlo přinést možnosti dynamickému řízení dopravy, kde by frekvence spojů byla v reálném čase měněna v závislosti na počasí.

Ostatní doprava – data získaná z osobních automobilů a dalších senzorů na silnicích by mohla být zakomponována do získaných dat z hromadné dopravy. V případě dynamicky řízených systémů by tato data opět mohla být využita k podávání informací v reálném čase. Toto by bylo velmi užitečné například u autobusů, kde okolní doprava je hlavní příčina jejich zpoždění. Dnes někteří dopravci ukazují momentální zpoždění daného

autobusu, ale do výpočtů už nezahrnují hustotu dopravy, kterou autobus bude muset projet, takže se informování omezuje na postupné navyšování velikosti zpoždění. Se zakomponovanými informacemi o celkové dopravní situaci by mohl být vypočítán relativně přesný odhad toho, jaký bude mít autobus zpoždění při příjezdu na danou stanici.

### 3.2 Současné využívání

Využití dat v dopravě není jen o jejich sběru. Hlavní přidaná hodnota pramení ze schopnosti získání nových poznatků z těchto dat, na základě kterých se pak mohou dělat informovaná rozhodnutí vedoucí k zlepšení kvality a efektivity veřejné dopravy. Samotné využití a způsob analýzy bude záviset na způsobu sběru a na struktuře dat. Následují některé příklady (zejména zahraniční) reprezentující, jak se data k těmto účelům v současnosti využívají:

- Zvyšování spokojenosti zákazníků: ze sesbíraných dat lze zjistit, které části provozu jsou zákazníky kritizovány a se kterými jsou nespokojeni (např. přetížené spoje, chybějící spojení v konkrétní čas apod.). Tato data lze získat buď marketingovými průzkumy nebo odvozením na základě chování zákazníků. Na základě toho pak lze identifikovat oblasti služby, které potřebují zlepšit anebo naopak služby, na které jsou vynakládány zbytečně vysoké prostředky. (Becker a Albers, 2016)
- Optimalizace dopravní sítě v Londýně: Analýzou informací získaných z londýnských cestovních karet, bylo možné získat kontextové znalosti o trasách a stanicích, odhad přestupujících osob mezi nimi a identifikovat neočekávané vzorce. Tyto informace lze využít k posouzení navrhovaných změn na trasách, jako zvýšení kapacity či frekvence na určitých segmentech, ale i umístění zastávek. Analýzu změn v chování zákazníků při cenách jízdného lze využít pro optimalizaci výnosů, a též kvality a spolehlivosti služeb. (Seaborn et al., 2009)
- Měření spokojenosti v čínské veřejné dopravě: Na základě upravené americké teorie o spokojenosti zákazníka (ACSI, 2023), byl vytvořen hodnotící index, který byl dále aplikován na shromážděná data. Z analýzy dat vyplývají služby, které mají přímý pozitivní vliv na spokojenost cestujících, jejich loajalitu a vnímanou hodnotu služby. Aspekty negativně ovlivňující spokojenost

zákazníka byly z dat také získány. Na základě zjištěných poznatků pak byla podána doporučení pro zlepšení služeb. (Zhang et al., 2019)

- Využití dat z cestovních karet ve městě Kóči: Z dat získaných v přestupních uzlech bylo možné zjistit a zhodnotit efektivitu jednotlivých přestupních uzlů. Jako tři hlavní parametry efektivnosti byly uvažovány efektivita přestupů, závislost přestupu a spolehlivost přestupu. Na základě kritérií jsou identifikovány přestupní uzly s nejvyšší a nejnižší efektivností. Dále z dat vyplývá, že vysoký počet cestujících v daném uzlu nezajišťuje vysokou efektivnost a je nutno tak brát v potaz i další faktory, jako je například četnost služeb. (Nishiuchi et al., 2015)
- Analýza dopravních dat v Buenos Aires: Analýzou veřejných dat o jízdách v kombinaci s daty o poloze v reálném čase byly získány cenné informace, které se dají využít na optimalizaci jízd, vytváření prediktivních modelů, které odhadují zpoždění, nebo pro navrhování nových tras. Modely lze také použít k identifikaci problémových úseků a na jejich případné úpravy. (Godfrid et al., 2022)

## 4 Využívání dat v PMDP

### 4.1 Plzeňské městské dopravní podniky

Plzeňské městské dopravní podniky, a.s. (PMDP) jsou provozovatelem městské veřejné dopravy na území města Plzně a v jeho blízkém okolí. Samotná tradice MHD v Plzni funguje od roku 1899 a ročně přepraví více než 110 miliónu cestujících. Kromě provozu MHD, PMDP zajišťuje provoz městského parkovacího systému, veřejného osvětlení a světelné signalizace, službu carsharing Karkulka a službu Senior Expres. (PMDP, 2023a)

PMDP provozují 3 základní druhy dopravních prostředků (PMDP, 2021; PMDP, 2023a):

- Tramvaje: mají 3 linky a jsou výhradně na elektrický pohon. Dohromady PMDP disponuje 108 vozidly, z nichž 81 % je nízkopodlažních a 15 % disponuje klimatizací salonu pro cestující.
- Trolejbusy: mají 10 různých linek, primárně fungují na elektřinu napojením na trakční vedení. Většina vozidel je opatřena alternativním pohonem (starší vozidla dieselaagregátem, novější pak trakčními bateriemi. Dohromady PMDP disponuje 102 trolejbusy, z nichž 100 % je nízkopodlažních a 22 % je klimatizováno.
- Autobusy: mají 33 linek. PMDP disponují 102 vozidly, z nichž 100 % je nízkopodlažních a 30 % je klimatizováno.

Díky velkému množství tramvají a trolejbusů jsou dvě třetiny dopravního výkonu zajištěny elektrickou dopravou, která je šetrná k životnímu prostředí. Do budoucna je plánováno tento podíl dále navyšovat.

Páteční linky obvykle začínají provoz po čtvrté hodině ranní a končí před půlnocí. Od půlnoci do čtyř pak jezdí noční autobusy, které zajišťují základní obslužnost města a jsou označeny NX, kde X je číslo, na rozdíl od běžných linek, které jsou označeny pouze číslem. Momentálně PMDP provozují devět nočních linek, z nichž v osmi z nich je zajišťována doprava autobusem a na jedné lince zajišťuje dopravu trolejbus. Podle čísla lze rozeznat i jednotlivé typy vozidla na běžných denních linkách, kde čísla tramvají nabývají hodnot  $X \leq 4$ , čísla trolejbusů nabývají hodnot  $X \geq 10 \wedge X \leq 19$  a čísla

autobusů nabývají hodnot  $X \geq 20$ . Momentálně PMDP provozují 46 různých denních linek. (PMDP, 2023a)

## **4.2 Nástroje sběru a využívání dat**

PMDP disponují několika nástroji na sběr dat. Dopravní oddělení, které připravuje rozsah provozu, ale i konkrétní jízdní řády, pak využívá dva základní nástroje. Prvním je systém automatického počítání cestujících od firmy Abirail, s. r. o., druhým pak dispečerský systém Sprinter vyvíjený firmou HERMAN SYSTEMS, s.r.o.

### **4.2.1 Systém automatického počítání cestujících (APC)**

APC je využíváno napříč jednotlivými trakcemi a momentálně je nainstalováno cca 40 % vozového parku. Nově dodávané vozy mají APC jako součást standardní výbavy.

PMDP využívají pro sčítání lidí senzory nad dveřmi a jejich vyžadovaná přesnost je 95 %. Momentálně se používají 3 různé technologie senzorů. Infrsenzory Dilax, které fungují na principu přerušování paprsku, dále infrakamery Dilax, nově také stereokamery VivoTek, fungující na principu zachycení hlavy a sledování pohybů těla vůči žluté kontrastní liště (nástupní hraně) dveří. Stereokamery lze také využít jako standardní dveřní kamery pro snímání cestujících v jejich prostoru (zejména pro řidiče). APC systém komunikuje v průběhu jízdy s palubním počítačem Herman EPIS. Po zapnutí vozidla ve vozovně se pak nasbíraná data odesílají přes router a Wi-Fi síť na server PMDP a jsou zobrazitelná přes aplikaci ABIRUN.

Kromě počítání cestujících za pomoci APC je také využíváno manuální sčítání brigádníky a praktikanty, za pomoci kterých se pak ověřuje dosažení vyžadované přesnosti automatických systémů.

Do aplikace ABIRUN se nahrávají základní číselníky (vozidla, zastávky, řidiči) a platný grafikon. Systém si páruje nasbíraná data z vozidla se spoji podle platného grafikonu, k ověření správnosti využívá také data z GPS.

O nasazování vozů s APC a zajišťování optimálního pokrytí na konkrétních linkách se stará dopravní oddělení. Získaná data už byla využita ke změnám – například k nové koncepci víkendového provozu, a to včetně dřívějšího nástupu zkráceného intervalu na linkách 11 a 13 vlivem návštěvnosti obchodních center ještě před polednem. Data

se také využívají k ověření stížností na přeplněnost a případná rozhodnutí o posílení spojů.

System také ukládá zpoždění vozidel, na základě kterého se pak upravují jízdní doby, případně ukazuje, jaký vliv má uzavření ulic osobní dopravě (aktuálně např. Americká třída).

#### **4.2.2 System Sprinter**

System Sprinter je používán pro dispečerský dohled. Sprinter využívá především dispečink PMDP pro sledování GPS polohy spojů, jejich zpoždění a při řešení mimořádných událostí. Tento system zároveň ukládá informace o vozidlech, jako jsou například časy příjezdů a odjezdů ze zastávek.

Na základě metodiky Ing. Vojtěcha Novotného, Ph.D. z Fakulty dopravní ČVUT (Novotný, 2017) jsou posuzována zdržení v každém hodnoceném úseku na základě tzv. ideální jízdní doby (průjezdy brzo ráno a pozdě večer). Pro každou hodinu nebo úseky provozu (celý den, ranní/odpolední špička, sedla) se vypočítá známka provozu, která udává rozdíl zdržení na průjezdu v příslušném intervalu oproti jízdní době ideální. Následně se vypočítá kombinovaná známka rychlosti a spolehlivosti (tj. zdržení na průjezdu a jak moc velký rozptyl v příslušném intervalu tyto hodnoty mají a určí se výsledná známka). Data se vyexportují ze systému Sprinter a následně se předvedou pomocí maker naprogramovaných v excelovském souboru a vytvoří zčásti interaktivní mapku s grafy pro jednotlivé úseky.

PMDP momentálně nedisponují schopností zachytit toky pasažérů. Nově instalované kamery fungující na principu sledování těla vůči liště dveří by mohly poskytovat možnost tohoto využití. Náročnost implementace a přínosy tohoto systému zatím nebyly zkoumány a představují tak jednu z možných oblastí pro další výzkum. (Ing. A. Šťastný, osobní komunikace, 5. 4. 2023)

### **4.3 Popis problému**

Současně se v PMDP využívá k reprezentaci kapacity jednotlivých segmentů prezentace v prostředí Power Point, kde je nahraná fotka mapy města Plzně, na které jsou vytvořené objekty reprezentující jednotlivé traťové segmenty a jejich kapacitu. Pro každou trakci

(autobusy, tramvaje, trolejbusy) je vytvořen vlastní snímek. Prezentace je zobrazena na obrázku č. 3.

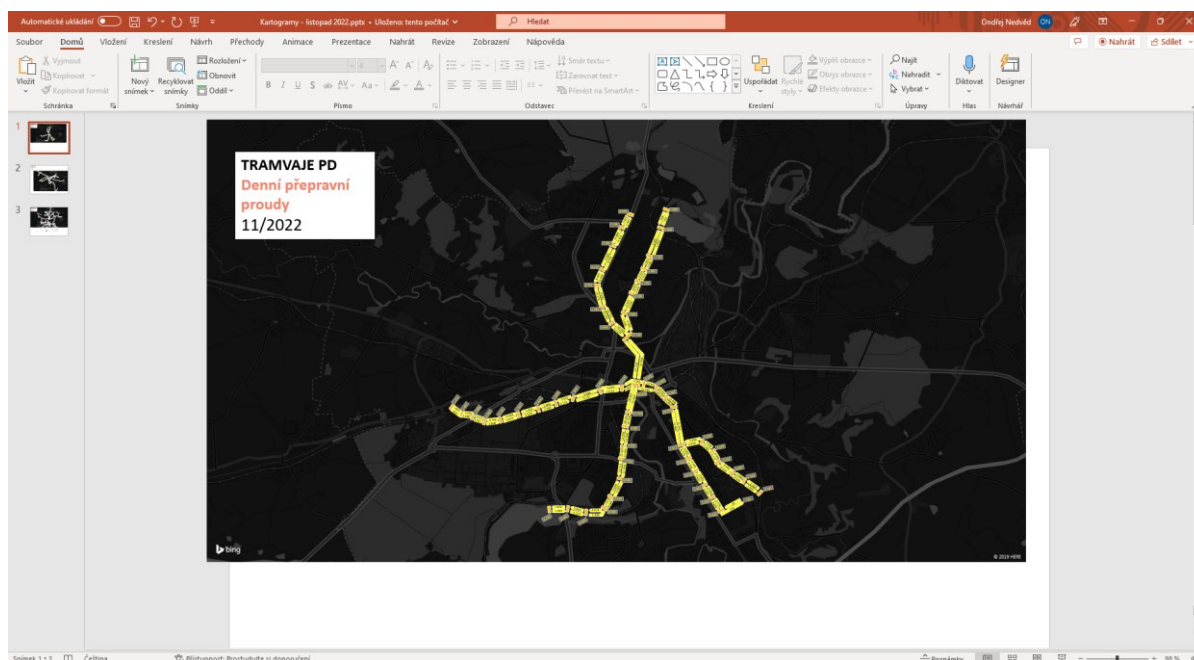
Data reprezentující kapacitu v segmentech jsou pak ručně přepisována ze zdrojového souboru do každého objektu zvlášť. Tento postup rozhodně nelze považovat za ideální, jelikož představuje náročnou a zdlouhavou práci, ve které je snadné se dopustit chyby.

Objekty také nejsou nijak uchyceny na mapový podklad, tedy v případě posunu mapy segmenty zůstávají na místě a tím pádem neodpovídají reálnému umístění tratí na mapě. To je problematické zejména při práci s větším přiblížením, kdy uživatel při intuitivním pokusu a posunutí mapy pohne pouze mapovým podkladem, ale objekty zůstanou na místě. Tento nedostatek je znát i při vkládání dat, kde je nutné relativně velkého přiblížení, aby bylo do objektu možné kliknout a data vepsat. V neposlední řadě pak také není možné obrázek klasickým způsobem kopírovat a poslat, jelikož při pokusu o zkopírování se zkopíruje pouze mapový podklad, případně jednotlivý segment, který byl zrovna označen. Mapový podklad je také starší a v případě autobusů a do menší míry trolejbusů vybíhají segmenty pryč z mapového podkladu. Tento problém je vidět na obrázcích číslo 4 a 5.

Části tratě, které jsou společné pro více linek, se zobrazují kumulativně. Tento přístup je pro PMDP preferovaný, protože kvalita celého segmentu je důležitější než kvalita jednotlivých linek. Tento způsob vizualizace tak nezachycuje granularitu dat, která jsou dostupná a tato chybějící informace by mohla mít za následek chybné určení vzorců chování nebo neidentifikaci problému.



Obrázek č. 3: Reprezentace kapacity segmentů v programu Power Point (tramvaje)



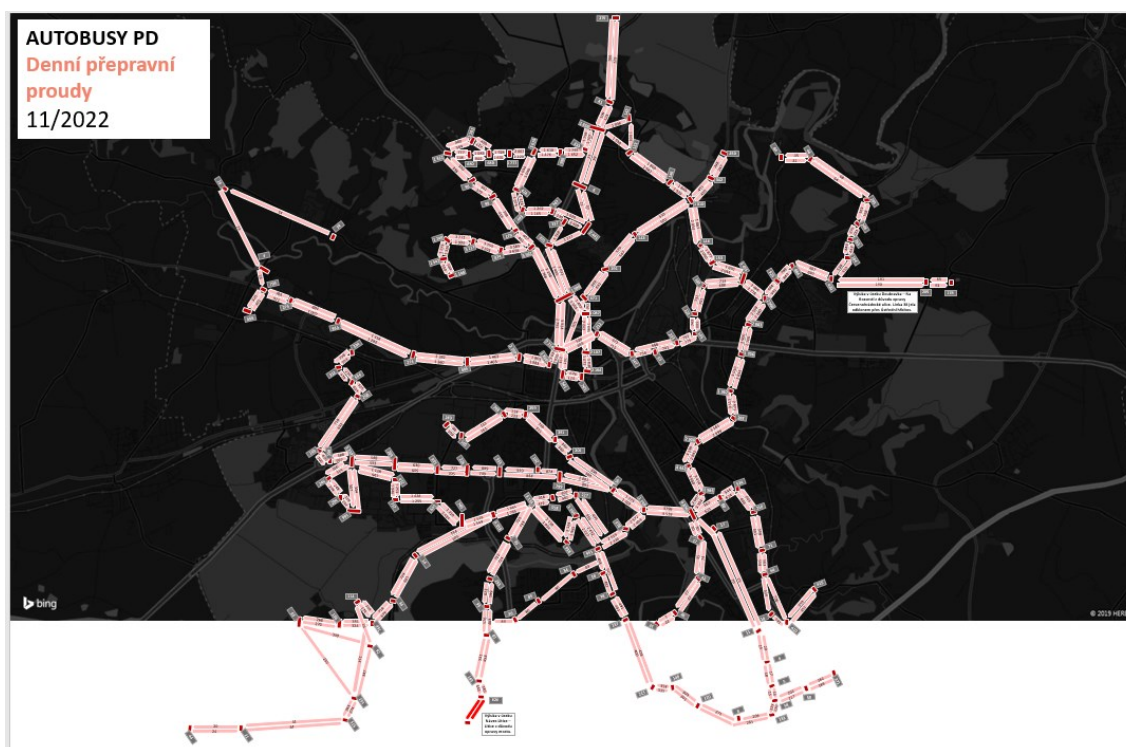
Zdroj: PMDP, 2023b

Obrázek č. 4: Vizuální reprezentace kapacity segmentů pro trolejbusy



Zdroj: PMDP, 2023b

Obrázek č. 5: Vizuální reprezentace kapacity segmentů pro autobusy



Zdroj: PMDP, 2023b

Dalším problémem je současný postup při identifikaci problematických segmentů na základě více ukazatelů kvality přes vyexportované tabulky dat z excelu. Pro Excel není vytvořené žádné makro či přímo daný postup. Identifikace těchto segmentů je tak dělána přes hledání a filtrování excelovských tabulek a pro každá nová data je postup nutné opakovat. Tento postup je časově náročný, dále je také nutné ručně ošetřit veškeré společné segmenty, aby v datech vystupovaly společně.

Posledním problémem je pak samotná prezentace těchto dat. Tabulka s hodnotami zpoždění a jiných metrik, je bez vizuální prezentace umístění a bez dalšího kontextu často obtížná na porozumění. Kvalita vizualizace je v tomto ohledu obzvláště důležitá, protože se velmi často využívá pro prezentaci problematických úseků veřejné dopravy při jednání orgánů Magistrátu města Plzně.

S ohledem na stávající ne zcela uspokojivou úroveň práce s daty v PMDP byl v rámci této práce vytvořen program, který dokáže řešit jak vizuální, tak analytické nedostatky stávajícího řešení.

## 4.4 Popis vytvořeného programu

Pro vytvoření tohoto programu bylo autorem této práce vybráno prostředí JupyterLab. Jedná se o interaktivní vývojové prostředí pro výpočty, analýzu a datovou vizualizaci. Je vyvinuto jako open-source projekt a poskytuje širokou škálu nástrojů. Umožňuje práci v několika programovacích jazycích, jako například Python, R, Julia a další. (Jupyter, 2023) Hlavními důvody pro jeho zvolení byla právě možnost interaktivní vizualizace a velké množství využitelných knihoven v open-source prostředí. Cílem je vytvoření programu, který vyřeší problém s ručním zadáváním dat, nevhodným mapovým podkladem, zachování detailu informace a zároveň vytvoří prostředí, ve kterém bude možné data vizuálně prezentovat.

Jako zdrojový soubor je využíván Microsoft Excel, který se pak pouze do JupyterLab nahraje a program z excelovského souboru přečte data. Každý subsystém dopravy má dva zdrojové Excely, první z nich obsahuje data pro jednotlivé mezistaniční úseky (segmenty). Jak lze vidět na obrázku č. 6, každý segment má ve svém zdrojovém souboru informace o lince, do které patří, souřadnice zastávek (výchozí a koncové pro každý segment), jméno zastávky na začátku segmentu, směr jízdy a pak datové ukazatele: průměrnou rychlost, kapacitu, známku (dle metodiky Ing. Novotného) a zdržení. Struktura exportovaných dat byla při vývoji programu brána v potaz a pro přidání nových dat tak stačí data ze zdrojových souborů zkopírovat pomocí „svyhledat“ nebo jiné podobné funkce. Druhý zdrojový Excel, jehož struktura je zobrazena na obrázku č. 7, pak obsahuje informace o linkách. Vkládání dat přes Excel bylo autorem práce zvoleno pro uživatelskou přívětivost, kdy celé vkládání dat je řešeno přes již zaběhlý způsob. Pokud by uživatel nechtěl jen upravovat data jednotlivých segmentů, ale chtěl by například přidat novou zastávku (segment), nebo celou novou linku, tak je stačí zapsat do Excelu ve správném formátu a po nahrání dat je program zobrazí. Vzhled prostředí se zvýrazněným tlačítkem pro nahrání souborů a spuštění programu je vidět na obrázku č. 8.

Obrázek č. 6: Datová struktura pro nahrání segmentů

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	line_num	line_dir	start_x	start_y	end_x	end_y	start_name	dir_name	avg_speed	capacity	grade	delay
2	1	1	A	49.726451	13.406965	49.726046	13.404585	Vozovna	Bolevec	20	500	1	50
3	2	1	A	49.726046	13.404585	49.724890	13.399463	Vozovna Slo	Bolevec	15	1000	1	1
4	3	1	A	49.724890	13.399463	49.727904	13.396727	Slovany	Bolevec	60	1500	1	2
5	4	1	A	49.727904	13.396727	49.729637	13.395193	Vresova	Bolevec	30	1700	2	5
6	5	1	A	49.729637	13.395193	49.733393	13.391771	Olsova	Bolevec	20	1800	3	48
7	6	1	A	49.733393	13.391771	49.735944	13.389501	Liliova	Bolevec	40	2290	4	16

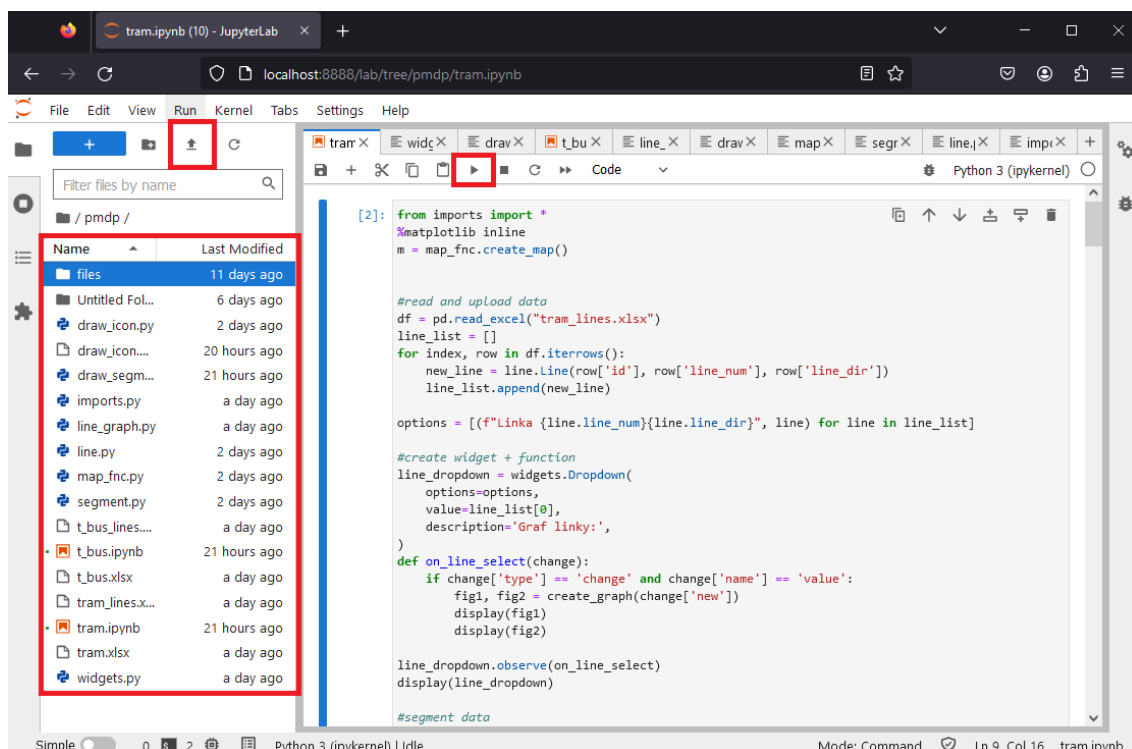
Zdroj: Vlastní zpracování, 2023

Obrázek č. 7: Datová struktura pro definici linek

A	B	C
id	line_num	line_dir
1	1	A
2	1	B
3	2	A
4	2	B
5	4	A
6	4	B

Zdroj: Vlastní zpracování, 2023

Obrázek č. 8: Vzhled Jupyter Lab, tlačítka na nahrání, spuštění a nahrané soubory



Zdroj: Vlastní zpracování, 2023

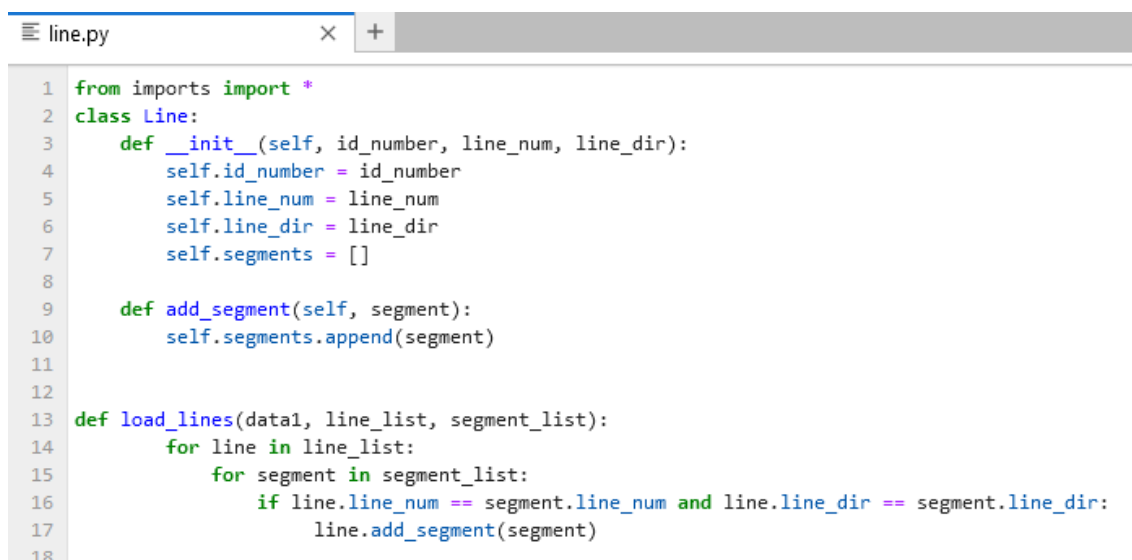
Celý projekt je složen ze 2 hlavních souborů „ipynb“, 4 zdrojových Excelů a ze 7 „py“ (python skript) souborů. Soubory s příponou ipynb (Interactive Python Notebook) jsou

specifické pro Jupyter prostředí a dovolují právě interaktivní vizualizace. Tyto dva soubory v projektu jsou pak „tram.ipynb“, který vytváří vizualizace pro tramvaje a „t\_bus.ipynb“, který se stará o vizualizace pro trolejbusy. Samotné soubory lze vidět vyznačené na obrázku č. 8. Dodatečné soubory, jako například obrázky (ikony), jsou umístěny ve složce „files“.

Po nahrání dat a spuštění, program data načte a přiřadí objektům vytvořeným v programu zavoláním funkcí „load\_lines“ a „load\_segments“.

Linky reprezentuje třída „Line“, v souboru „line.py“. „Line“ má pak data odpovídající datům linek v Excelu, ale navíc si „drží“ list segmentů. Soubor „line.py“ také obsahuje výše zmíněnou funkci „load\_lines“. Soubor „line.py“ je zobrazen na obrázku č. 9.

Obrázek č. 9: Soubor „line.py“ s třídou „Line“ a funkcí „load\_lines“



```
1 from imports import *
2 class Line:
3     def __init__(self, id_number, line_num, line_dir):
4         self.id_number = id_number
5         self.line_num = line_num
6         self.line_dir = line_dir
7         self.segments = []
8
9     def add_segment(self, segment):
10        self.segments.append(segment)
11
12
13 def load_lines(data1, line_list, segment_list):
14     for line in line_list:
15         for segment in segment_list:
16             if line.line_num == segment.line_num and line.line_dir == segment.line_dir:
17                 line.add_segment(segment)
18
```

Zdroj: Vlastní zpracování, 2023

Jednotlivé mezistaniční úseky jsou pak reprezentovány třídou „Segment“ v souboru „segment.py“. Každý segment má odpovídající hodnoty jako v Excelovské tabulce, k nahrání těchto dat do objektu slouží funkce „load\_segments“ vizte obrázek č. 10.

Obrázek č. 10: Soubor „segment.py“ s třídou „Segment“ a funkcí „load\_segments“

```
segment.py
1 from imports import *
2
3 class Segment:
4     def __init__(self, id_number, line_num, line_dir, start_xy, end_xy, start_name, end_name, avg_speed, capacity,
5     grade, delay):
6         self.id_number = id_number
7         self.line_num = line_num
8         self.line_dir = line_dir
9         self.start_xy = start_xy
10        self.end_xy = end_xy
11        self.start_name = start_name
12        self.end_name = end_name
13        self.avg_speed = avg_speed
14        self.capacity = capacity
15        self.grade = grade
16        self.delay = delay
17
18 def load_segments(data1, segment_list):
19     for datum in data1:
20         id_number, line_num, line_dir, start_x, start_y, end_x, end_y, start_name, end_name, avg_speed, capacity,
21         grade, delay = datum
22         start_xy = (start_x, start_y)
23         end_xy = (end_x, end_y)
24         segment1 = segment.Segment(id_number, line_num, line_dir, start_xy, end_xy, start_name, end_name, avg_speed,
25         capacity, int(grade), int(delay))
26         segment_list.append(segment1)
```

Zdroj: Vlastní zpracování, 2023

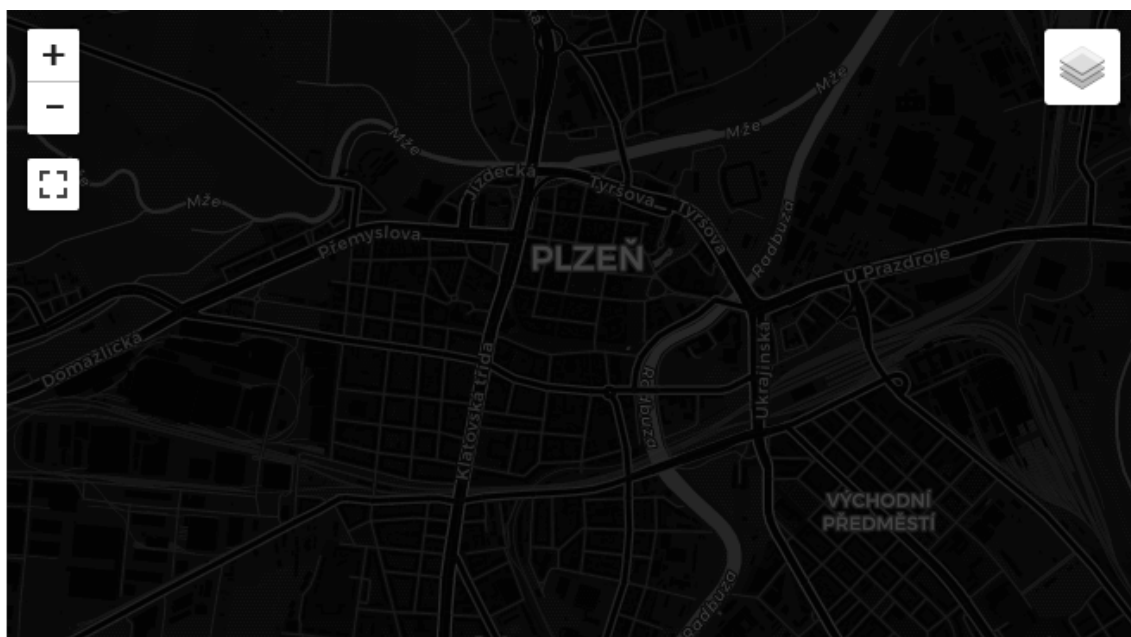
Následně probíhá vytvoření mapy, pomocí funkce „create\_map“ v souboru „map\_fnc“, která využívá knihovny „ipyleaflet“, která používá používá OpenStreetMap. Příslušný zdrojový soubor lze vidět na obrázku č. 11. Kromě vytvoření mapy, funkce také vycentruje mapu na dané souřadnice Plzně, přidá možnost přiblížení, maximalizace na celou obrazovku, satelitní a tmavou vrstvu a možnost jejich ovládání. Tmavá vrstva je pak dána jako primání, protože je nejpřehlednější pro datovou vizualizaci. Vytvořená mapa je zobrazena na obrázku č. 12.

Obrázek č. 11: Soubor „map\_fnc.py“ s funkcí „create\_map“

```
map_fnc.py x +
1 from ipyleaflet import Map, Marker, LayerGroup, basemaps, FullScreenControl, basemap_to_tiles, Icon,
  AntPath, Polyline, Popup, LayersControl
2 from imports import *
3
4 def create_map():
5     center = (49.7477, 13.3777)
6     m = Map(center=center, zoom=15)
7     m.scroll_wheel_zoom = True
8     dark = basemap_to_tiles(basemaps.CartoDB.DarkMatter, "Dark")
9     satelit = basemap_to_tiles(basemaps.Esri.WorldImagery)
10    m.add_layer(satelit)
11    m.add_layer(dark)
12    m.min_zoom = 11
13    m.add_control(FullScreenControl())
14
15    layer_control = LayersControl(
16        position='topright',
17        basemaps={
18            'Dark': dark,
19            'Satelit': satelit
20        }
21    )
22    m.add_control(layer_control)
23
24    return m
```

Zdroj: Vlastní zpracování, 2023

Obrázek č. 12: Mapa s tmavou vrstvou



Zdroj: Vlastní zpracování, 2023

Program dále vypočítá několik statistických ukazatelů, které budou dále potřebné pro další výpočty, jako jsou minimální a maximální hodnoty měřených ukazatelů a směrodatné odchyly (vizte obrázek č. 13).

Obrázek č. 13: Výpočet statistik

```
capacities = [seg.capacity for seg in segment_list if not np.isnan(seg.capacity)]
avg_speeds = [seg.avg_speed for seg in segment_list if not np.isnan(seg.avg_speed)]
#Průměr a směrodatná odchylka rychlosti a kapacity segmentů
capacity_mean = np.mean(capacities) if capacities else 0
capacity_std = np.std(capacities) if capacities else 0
avg_speed_mean = np.mean(avg_speeds) if avg_speeds else 0
avg_speed_std = np.std(avg_speeds) if avg_speeds else 0
min_capacity = min(segment.capacity for segment in segment_list)
max_capacity = max(segment.capacity for segment in segment_list)
min_avg_speed = min(segment.avg_speed for segment in segment_list)
max_avg_speed = max(segment.avg_speed for segment in segment_list)
min_delay = min(segment.delay for segment in segment_list)
max_delay = max(segment.delay for segment in segment_list)
```

Zdroj: Vlastní zpracování, 2023

Pro vizualizaci segmentů se v hlavní třídě zavolají funkce „ant“ a „poly“. Obě tyto funkce jsou v souboru „draw\_segment.py“. Obě funkce berou jako parametry data o segmentech, mapovou vrstvu do které bude vytvořená vizualizace uložena, průměrnou hodnotu a směrodatnou odchylku kapacity segmentů. Funkci „ant“ přibližuje obrázek č. 14.

Obrázek č. 14: Soubor „draw\_segment.py“ s funkcí „ant“

```
draw_segment.py x +
1 from ipyleaflet import Map, Marker, basemaps, basemap_to_tiles, Icon, AntPath, Polyline
2 from imports import *
3
4 def ant(segment_list, ant_group, cap_mean, cap_std):
5     for segment in segment_list:
6         layer = line_exists(segment.start_xy, segment.end_xy, ant_group)
7         if layer is not None:
8
9             layer.grade = round(((layer.grade*layer.counter) + segment.grade)/(layer.counter+1))
10            layer.color = assign_color(layer.grade)
11            layer.capacity = segment.capacity+layer.capacity
12            layer.weight = assign_weight(layer.capacity, cap_mean, cap_std)
13            layer.avg_speed = ((layer.avg_speed*layer.counter) + segment.avg_speed)/(layer.counter+1)
14            layer.delay = calc_delay(layer.avg_speed)
15            layer.counter += 1
16
17        else:
18
19            pulse_color='#ffffff'
20            color = assign_color(segment.grade)
21            weight = assign_weight(segment.capacity, cap_mean, cap_std)
22            delay = calc_delay(segment.avg_speed)
23            ant_path = draw_ant(segment, color, weight, pulse_color, delay)
24            assign_stats(ant_path, segment)
25            ant_group.add_layer(ant_path)
26
```

Zdroj: Vlastní zpracování, 2023

Funkce prochází všemi segmenty, nad kterými nejdříve volá funkci „line\_exists“ ve stejném souboru (ukázána na obrázku č. 15), která na základě souřadnic ověří, jestli už segment se stejnými souřadnicemi nebyl vytvořen.



Obrázek č. 15: Funkce „line\_exists“

```
135 def line_exists(start_xy, end_xy, group_layer):
136     for layer in group_layer.layers:
137         if (layer.locations == [start_xy, end_xy]) or (layer.locations == [end_xy, start_xy]):
138             return layer
139     return None
140
```

Zdroj: Vlastní zpracování, 2023

Pokud segment neexistuje, na základě známky segmentu a pomocí funkce „assign\_color“, se přiřadí barva. Postup přiřazení ve funkci lze vidět na obrázku č. 16.

Obrázek č. 16: Funkce „assign\_color“

```
85 def assign_color(grade):
86     if grade == 1:
87         color = '#009933' #tmavě zelená
88     elif grade == 2:
89         color = '#66ff66' #světle zelená
90     elif grade == 3:
91         color = '#b3b3b3' #šedá
92     elif grade == 4:
93         color = '#ff9900' #oranžová
94     elif grade == 5:
95         color = '#ff0000' #červená
96     else:
97         color = 'black'
98     return color
99
```

Zdroj: Vlastní zpracování, 2023

Na základě obsazenosti, odchylky od průměru a pomocí funkce „assign\_weight“ se přiřadí tloušťka (vyšší obsazenost bude vykreslena tširší čarou). Syntaxe funkce „assign\_weight“ je viditelná na obrázku č. 17.

Obrázek č. 17: Funkce „assign\_weight“

```
100 def assign_weight(capacity, cap_mean, cap_std):
101     if capacity < cap_mean - 2 * cap_std:
102         weight = 1
103     elif capacity < cap_mean - cap_std:
104         weight = 2
105     elif capacity < cap_mean - 0.5*cap_std:
106         weight = 3
107     elif cap_mean - 0.5* cap_std <= capacity <= cap_mean + 0.5*cap_std:
108         weight = 4
109     elif capacity <= cap_mean + 1 * cap_std:
110         weight = 5
111     elif capacity <= cap_mean + 2 * cap_std:
112         weight = 6
113     else:
114         weight = 7
115     return weight
```

Zdroj: Vlastní zpracování, 2023

Podle průměrné rychlosti a pomocí funkce „calc\_delay“ se přiřadí rychlost pohybu vytvořené čáry (jedná se o čáry, se světlými prvky znázorňující pohyb), výpočet lze vidět na obrázku č. 18. Delay v tomto případě neoznačuje zdržení, ale rychlost pohybu čáry.

Obrázek č. 18: Funkce „calc\_delay“

```
54 def calc_delay(avg_speed):
55     delay = int(((80 - avg_speed) / 80) * (6000 - 300) + 300)
56     return delay
```

Zdroj: Vlastní zpracování, 2023

Nakonec se zavolá funkce „draw\_ant“ viditelná na obrázku č. 19, která vezme předešlé vypočtené hodnoty a na jejich základě vytvoří pohyblivou čáru pomocí knihovny „ipyleaflet“ a přidá ji do mapové vrstvy „ant\_group“.

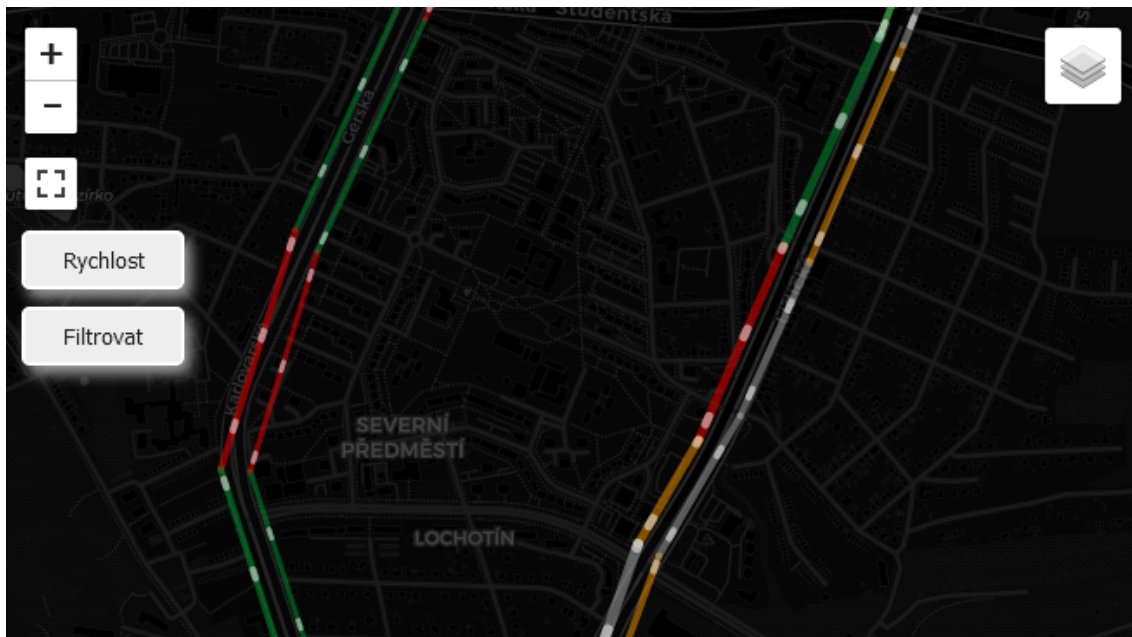
Obrázek č. 19: Funkce „draw\_ant“

```
29 def draw_ant(segment, color, weight, pulse_color, delay):
30     ant_path = AntPath(
31     locations = [segment.start_xy, segment.end_xy],
32     color=color,
33     weight=weight,
34     delay = delay, #Ant speed
35     dash_array=[5, 50],
36     )
37     return ant_path
```

Zdroj: Vlastní zpracování, 2023

V případě, že linka už existuje, vrátí funkce „line\_exists“ existující linku na daných souřadnicích a pouze se upraví hodnoty dané linky. Přepočítání hodnot je buď součtem, nebo aritmetickým průměrem v závislosti na konkrétní hodnotě, což lze vidět na obrázku č. 14. Vykreslené čáry na mapě jsou zobrazeny na obrázku č. 20.

Obrázek č. 20: Zobrazení mapy po proběhnutí funkce „ant“



Zdroj: Vlastní zpracování, 2023

Obrázek č. 21: Funkce „poly“

```

58 def poly(segment_list, poly_group, cap_mean, cap_std):
59
60     for segment in segment_list:
61         layer = line_exists(segment.start_xy, segment.end_xy, poly_group)
62         if layer is not None:
63             update_popup(segment, layer)
64             layer.grade = round(((layer.grade*layer.counter) + segment.grade)/(layer.counter+1))
65             layer.color = assign_color(layer.grade)
66             layer.capacity = segment.capacity+layer.capacity
67             layer.weight = assign_weight(layer.capacity, cap_mean, cap_std)
68             layer.avg_speed = ((layer.avg_speed*layer.counter) + segment.avg_speed)/(layer.counter+1)
69             layer.delay2 = segment.delay + layer.delay2 # pro počítání osobohod nás zajímá celkové
zpoždění na segmentu
70             layer.counter += 1
71
72         else:
73             color = assign_color(segment.grade)
74             weight = assign_weight(segment.capacity, cap_mean, cap_std)
75             polyLine = draw_poly(segment, color, weight)
76             assign_stats(polyLine, segment)
77             polyLine.popup = widgets.HTML(value=f"Linka č.{segment.line_num}:<br>Obsazenost:
{segment.capacity}<br>Průměrná rychlost: {segment.avg_speed} km/h<br>zpoždění: {segment.delay} s
<br>známka: {segment.grade}")
78
79             poly_group.add_layer(polyLine)

```

Zdroj: Vlastní zpracování, 2023

Funkce „poly“, viditelná na obrázku č. 21, funguje na stejném principu jako funkce „ant“, ale namísto pohybujících čar vykresluje čáry statické. Dalším rozdílem pak je přiřazení do jiné mapové vrstvy „poly\_group“ a přiřazení štítku s informacemi, který se při kliknutí na čáru zobrazí. V případě existence čáry se funkce opět chová podobně jako funkce „ant, s rozdílem zavolání funkce „update\_popup“, která přidá do informativního štítku informace s novým segmentem. Funkce je zobrazena na obrázku č. 22.

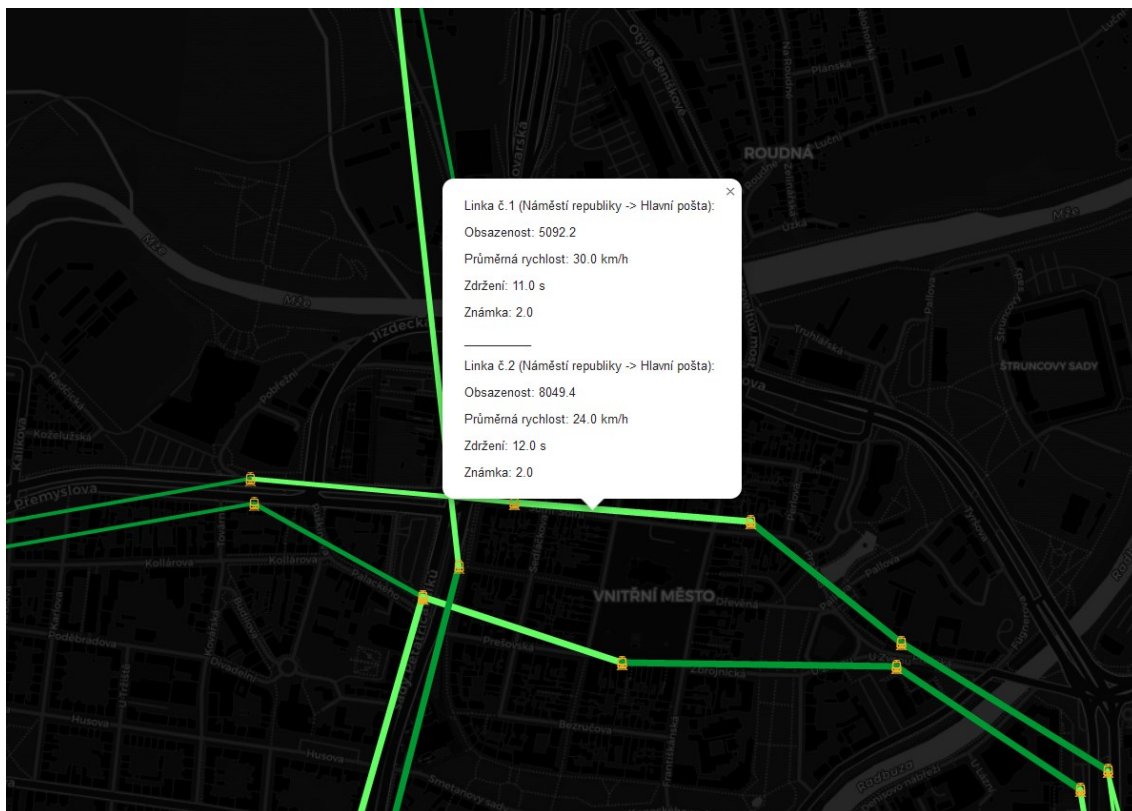
Obrázek č. 22: Funkce „update\_popup“

```
39 def update_popup(segment, layer):
40     old_content = layer.popup.value
41
42     new_content = (
43         f" _____ <br>"
44         f"Linka č.{segment.line_num} ({segment.start_name} -> {segment.dir_name}):<br> "
45         f"Obsazenost: {segment.capacity:.1f}<br> "
46         f"Průměrná rychlost: {segment.avg_speed:.1f} km/h<br> "
47         f"Zdržení: {segment.delay:.1f} s <br>"
48         f"Známka: {segment.grade:.1f} "
49     )
50
51     updated_content = old_content + "<br>" + new_content
52     layer.popup.value = updated_content
```

Zdroj: Vlastní zpracování, 2023

Na mapě se tak zobrazí kombinované hodnoty linek na daném segmentu, detail informací je pak však zanechán a je dostupný přes kliknutí na daný segment. Vykreslení statických čar a zachování informací na štítku je viditelné na obrázku č. 23.

Obrázek č. 23: Vykreslení čar pomocí funkce „poly“ a zobrazení štítku s informacemi



Zdroj: Vlastní zpracování, 2023

Pro možnost přepnutí mezi statickým a dynamickým/pohyblivým zobrazením existuje tlačítko „Rychlost“, které je viditelné na mapě na obrázku č. 20. Tlačítko je vytvořeno pomocí knihovny „ipywidgets“ a při kliknutí zavolá funkci „toggle\_poly\_group“, která nastaví statickým linkám maximální průhlednost. Statické linky jsou vykresleny přes linky pohyblivé, zvýšením jejich průhlednosti na maximum se tak zviditelní pohyblivé čáry, které jsou uloženy jako spodnější vrstva. Při opětovném kliknutí se statickým linkám nastaví průhlednost zpět na minimum. Kód funkce a vytvoření tlačítka jsou viditelné na obrázku č. 24.

Obrázek č. 24: Funkce „toggle\_poly\_group“ a vytvoření tlačítka „Rychlost“

```
def toggle_poly_group(button):  
    if ant_group not in m.layers:  
        m.add_layer(ant_group)  
    current_opacity = poly_group.layers[0].opacity  
    new_opacity = 0 if current_opacity == 1 else 1  
    for layer in poly_group.layers:  
        layer.opacity = new_opacity  
  
button = widgets.Button(description="Rychlost", layout=Layout(width='90px', height='30px'))  
button.on_click(toggle_poly_group)
```

Zdroj: Vlastní zpracování, 2023

Dalším krokem je vytvoření tlačítka „filtrvat“ na mapě. Při stisknutí tlačítka se objeví možnosti filtrování na mapě. Možnost filtrování je podle známky, kapacity, rychlosti, směru a zdržení. Vytvoření filtrů je zobrazeno na obrázcích č. 25 a 26.

Obrázek č. 25: Vytvoření filtru podle známky, směru a kapacity

```
#filtr pro známky a směr
checkbox1 = Checkbox(description='1', value=True)
checkbox2 = Checkbox(description='2', value=True)
checkbox3 = Checkbox(description='3', value=True)
checkbox4 = Checkbox(description='4', value=True)
checkbox5 = Checkbox(description='5', value=True)
checkbox6 = Checkbox(description='A', value=True)
checkbox7 = Checkbox(description='B', value=True)
#filtrování podle kapacity
max_capacity = max(layer.capacity for layer in poly_group.layers)
capacity_slider = widgets.FloatRangeSlider(
    value=[min_capacity, max_capacity],
    min=min_capacity,
    max=max_capacity,
    step=10,
    readout_format='.0f',
    style={'description_width': 'initial'}
)
```

Zdroj: Vlastní zpracování, 2023

Obrázek č. 26: Vytvoření filtru podle rychlosti a zdržení

```
#filtr pro rychlost
speed_slider = widgets.FloatRangeSlider(
    value=[min_avg_speed, max_avg_speed],
    min=min_avg_speed,
    max=max_avg_speed,
    step=0.1,
    readout_format='.1f',
    style={'description_width': 'initial'}
)
#filtr pro zpoždění
max_delay = max(layer.delay2 for layer in poly_group.layers)
delay_slider = widgets.FloatRangeSlider(
    value=[min_delay, max_delay],
    min=min_delay,
    max=max_delay,
    step=1,
    readout_format='.1f',
    style={'description_width': 'initial'}
)
```

Zdroj: Vlastní zpracování, 2023

Všechna tlačítka pak následně při uživatelem vyvolané změně volají funkci „update opacity“ zobrazenou na obrázku č. 27, která podle nastavených filtrovacích parametrů nastaví neodpovídajícím segmentům průhlednost na maximum. Zobrazení segmentů po využití filtru je zobrazeno na obrázku č. 28.

Obrázek č. 27: Funkce „update\_opacity“

```
def update_opacity(change):
    selected_grades = [int(checkbox.description) for
                      checkbox in [checkbox1, checkbox2, checkbox3, checkbox4, checkbox5] if checkbox.value]
    selected_line_dirs = [checkbox.description for
                        checkbox in [checkbox6, checkbox7] if checkbox.value]

    selected_min_capacity, selected_max_capacity = capacity_slider.value

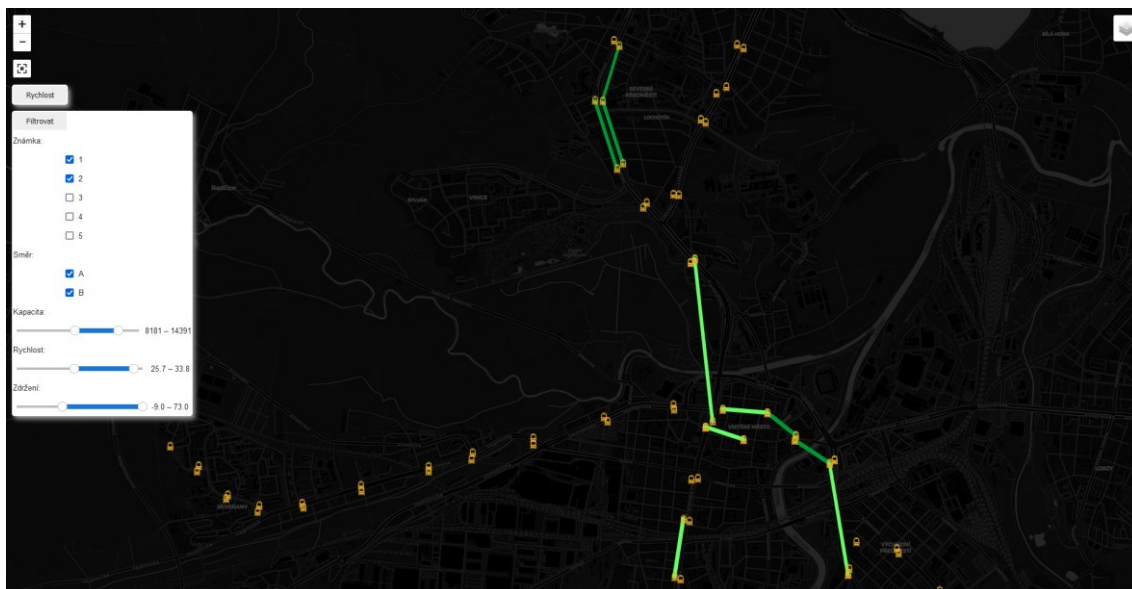
    selected_min_speed, selected_max_speed = speed_slider.value
    selected_min_delay, selected_max_delay = delay_slider.value

    if ant_group in m.layers:
        m.remove_layer(ant_group)

    for layer in poly_group.layers:
        layer_opacity = 0
        if (layer.grade in selected_grades) and
            (layer.line_dir in selected_line_dirs) and
            (selected_min_capacity <= layer.capacity <= selected_max_capacity) and
            (selected_min_speed <= layer.avg_speed <= selected_max_speed) and
            (selected_min_delay <= layer.delay2 <= selected_max_delay):
            layer_opacity = 1
        layer.opacity = layer_opacity
```

Zdroj: Vlastní zpracování, 2023

Obrázek č. 28: Mapa s využitím filtrovací funkce



Zdroj: Vlastní zpracování, 2023

Program také disponuje souborem „line\_graph“, který obsahuje několik funkcí na vykreslení grafů. Funkce „create\_graph“ vytváří dva grafy pro linku v jednom směru.

První graf dává údaje o zdržení a kapacitě na jednotlivých mezistaničních úsecích. Druhý graf pak informuje o zdržení v osobo-hodinách pro daný úsek a akumulované zdržení. Dále jsou v souboru funkce „line\_stats“ a „line\_stats\_bus“, jež obě vykreslují bodový graf, kde na ose x je zdržení a na ose y je kapacita. Každý bod je také barevně rozlišen podle známky. Rozdílem mezi těmito funkcemi je, že první funkce se využívá pro tramvaje, kde mají body pro každou tramvajovou linku jiné grafické znázornění. U trolejbusů jsou všechny body vykresleny stejným grafickým znakem, vzhledem k počtu linek by vykreslení několika různými znaky nebylo přehledné. Funkce „box\_and\_whiskers“ vytváří boxové grafy pro nalezení extrémních hodnot. Funkce „plot\_segments“ vytváří bodový graf pro oba směry a následně počítá jejich korelační koeficient. Poslední funkcí v souboru je „performance\_table“. Tato funkce ve výstupu vytvoří pomocí textových znaků souhrnnou tabulku s výkonnostními metrikami všech linek dané traktice. Funkce dále také uloží výstup do souboru „performance\_table.csv“, který se vytvoří v souborech v JupyterLab, kde je možné ho prohlížet, případně stáhnout a importovat do Excelu pro další analýzu. Vygenerovanou tabulku se souborem lze vidět na obrázku č. 29. Soubor „line\_graph“ spolu s kódem ke všem funkcím a textovou formou tabulky je dostupný v příloze B.

Obrázek č. 29: Vygenerovaný „performance\_table.csv“

	Linka č.	Směr	Průměrná rychlost(km/h)	Celkové zdržení(osobohod)	Připravených osob	Zdržení na osobu(sec)	Zdržení na zastávku(sec)
1	10	A	26.05	14.55	9355	5.60	0.77
2	10	B	19.74	9.33	7988	4.20	0.52
3	11	A	28.42	23.06	16863	4.92	1.10
4	11	B	26.75	26.50	17227	5.54	1.26
5	12	A	28.03	37.51	29946	4.51	1.50
6	12	B	27.17	60.29	32807	6.62	2.41
7	13	A	25.80	35.12	33301	3.80	1.25
8	13	B	22.10	41.43	28367	5.26	1.59
9	14	A	25.22	3.40	2506	4.88	0.26
10	14	B	22.26	4.04	2917	4.99	0.31
11	15	A	26.64	45.60	35369	4.64	2.17
12	15	B	25.77	60.90	30880	7.10	2.77
13	16	A	22.69	111.51	52273	7.68	6.20
14	16	B	21.79	150.57	57238	9.47	7.53
15	17	A	26.11	12.05	6163	7.04	0.57
16	17	B	25.36	15.02	8679	6.23	0.65
17	18	A	27.45	3.40	2213	5.53	0.31
18	18	B	27.61	4.83	2291	7.59	0.48
19	19	A	0.00	0.00	0	0.00	0.00
20	19	B	0.00	0.00	0	0.00	0.00

Zdroj: Vlastní zpracování, 2023

Vytvořený program tak řeší předem stanovené problémy, kde se data nemusí zadávat ručně a segmenty zůstávají na správné lokaci při pohybu mapy. Program také disponuje analytickými nástroji v podobě vykreslení grafů na zadaných datech a v grafické reprezentaci rychlosti, známky kvality dopravy a kapacity. V neposlední řadě je možné



filtr zabudovaný v mapě využít jak k analytickým potřebám, tedy k filtrování úseků podle zvolených údajů, tak zároveň jako nástroj komunikační, kde vyfiltrované výsledky analýzy jsou hned vizuálně připravené.

Program v současné formě disponuje několika nedostatky:

- Filtrování segmentů na mapě je možné jen se statickými objekty. Při filtrování je tak nutné pohyblivé čáry odstranit. Po odstranění je lze vrátit opětovným spuštěním programu.
- Nemožnost zobrazení všech informačních štítků (pro všechny mezistaniční úseky) najednou – s momentálním vzhledem štítků nepřehledné.
- Ikonky zastávek zabírají dost RAM paměti. Spolu s vykreslováním velkého množství linek to může způsobovat na starších, či méně výkonnějších strojích problémy. U trolejbusů, kde je hodně zastávek, se z tohoto důvodu momentálně nevykreslují. Toto omezení je způsobeno zejména během programu ve webovém prohlížeči, v případě budoucího vytvoření samostatné aplikace bude tento problém odstraněn.
- V prohlížečích jiných, než je Mozilla Firefox může docházet u pohyblivých čar k „problívání“. V prohlížeči Google Chrome by toto mělo být opraveno při příští aktualizaci prohlížeče.

## 4.5 Analýza dat

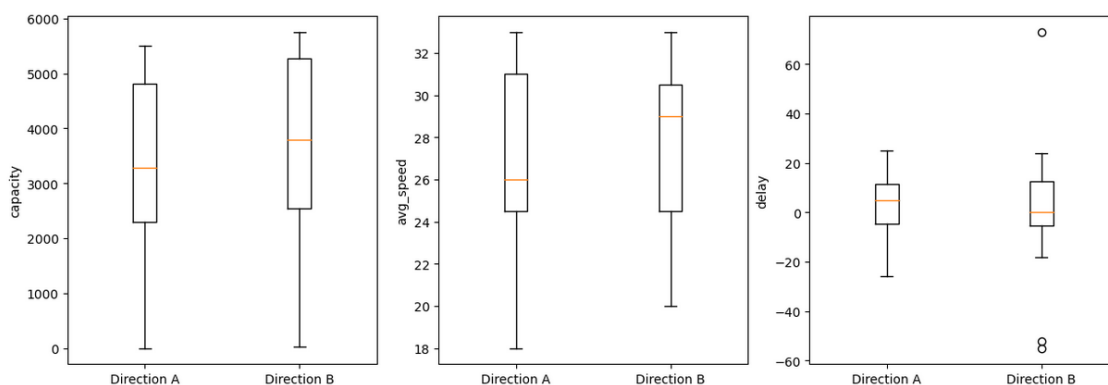
Vytvořený nástroj poslouží k analýze stávajícího stavu pro vybranou část sítě provozu MHD v Plzni. Pro tuto analýzu byly konkrétně vybrány plzeňské tramvajové a trolejbusové linky, jež celkově zajišťují 2/3 dopravního výkonu MHD v Plzni. Pro podrobnější analýzu části stávající sítě MHD byla do excelů nahrána průměrná denní data za pracovní dny v měsíci březnu 2023 (všechna níže prezentovaná data tedy odpovídají jednomu „průměrnému“ dni). Data pro nově zavedenou linku trolejbusu č. 19 nebyla v době zpracování analýzy dostupná, ale linka již je v Excelu nadefinována a bude možné v budoucnu sledovat její výkonnost s nově dostupnými daty. Linka 19 je nicméně provozována v rozsahu jednoho páru spojů v pracovní dny, a tak v současnosti zastává doplňkovou roli.

Po nahrání excelů s daty pak stačí spustit „tram.ipynb“ pro tramvaje a „t\_bus.ipynb“ pro trolejbusy a veškeré potřebné údaje program vygeneruje.

#### 4.5.1 Tramvaje

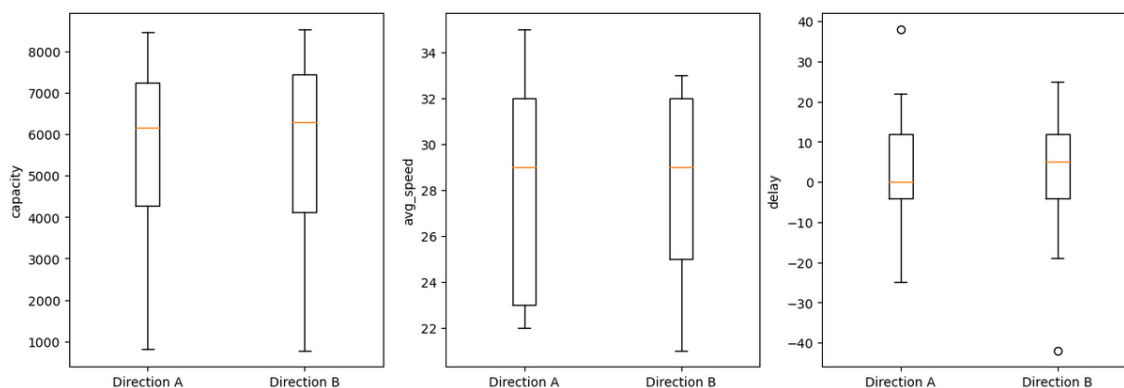
Analýza začne výstupem programu „tram.ipynb“. Prvním krokem analýzy je nalézt případné chybné hodnoty. Extrémní hodnoty lze hledat ve vygenerovaných „krabicových grafech“, které jsou zobrazeny na obrázcích č. 30, 31 a 32 (průměrná rychlost u tramvají není PMDP analyzována z důvodů oddělené jízdní dráhy a nebude tedy pro další analýzu brána v potaz).

Obrázek č. 30: Krabicové grafy linky č. 1



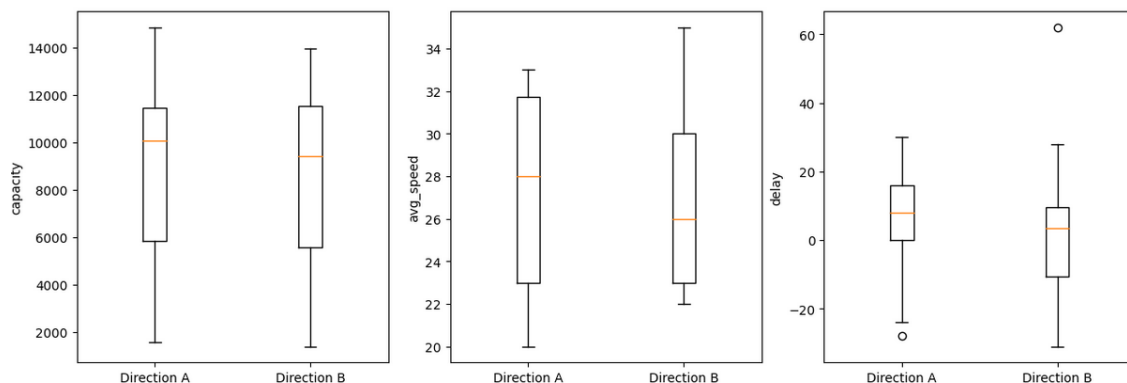
Zdroj: Vlastní zpracování, 2023

Obrázek č. 31: Krabicové grafy linky č. 2



Zdroj: Vlastní zpracování, 2023

Obrázek č. 32: Krabicové grafy linky č. 4

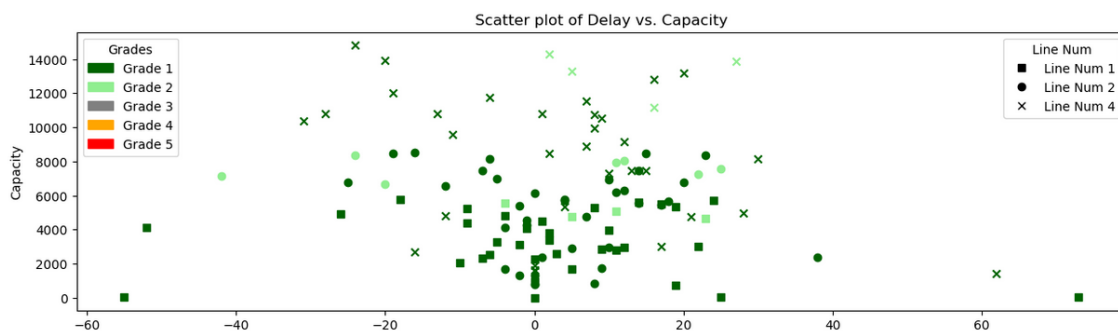


Zdroj: Vlastní zpracování, 2023

Z grafů lze vidět, že v obsazenosti (a rychlosti) se nenachází žádné extrémní hodnoty u žádné z linek. Každá linka má však extrémní hodnoty ve zdržení. V dopravě existuje velká variabilita a hodnoty získané v datech jsou průměrné hodnoty velkého množství měření. Na základě těchto faktů a konzultace s panem Ing. Šťastným z PMDP nelze považovat tyto hodnoty za chybná měření, ale spíše za extrémní hodnoty vycházející ze specifických podmínek daného úseku. V rámci další analýzy byly vyřazeny konečné zastávky, kde často dochází k těmto extrémním hodnotám z různých důvodů (procházení vozidla, vyčkávání na čas odjezdu, ...).

Z těchto grafů lze vyčíst, že linka č. 4 přepravuje nejvíce pasažérů, dále pak linka č. 2 a linka č. 1 nejméně. Zdržení se u linek pohybuje v relativně podobném rozpětí, tyto informace potvrzuje další z vygenerovaných grafů, viditelný na obrázku č. 33. Na grafu je patrné, že většina segmentů má známku kvality přepravy 1 a zbylé segmenty mají známku 2. Žádné segmenty tedy nemají za sledované období známku 3 nebo horší.

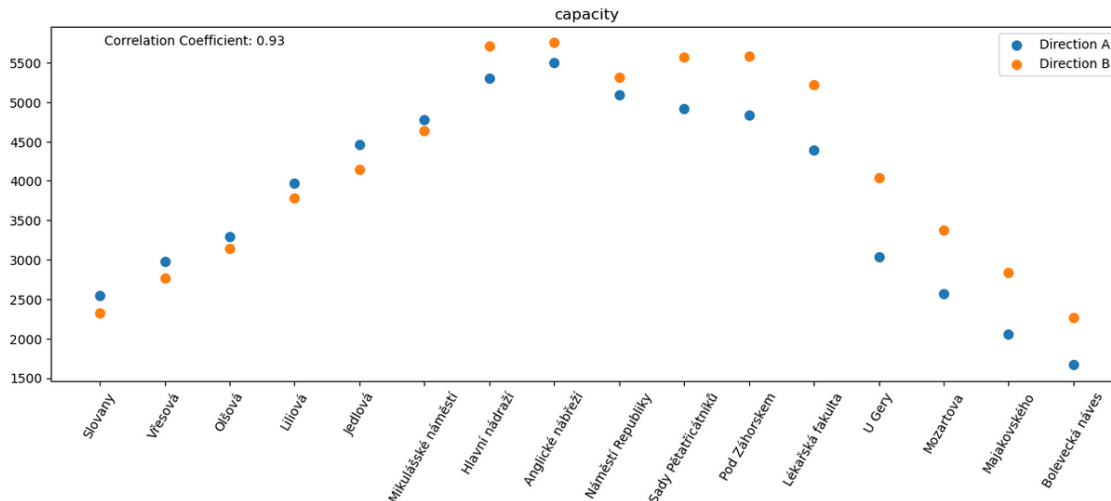
Obrázek č. 33: Bodový graf zobrazující kapacitu a zpoždění jednotlivých segmentů



Zdroj: Vlastní zpracování, 2023

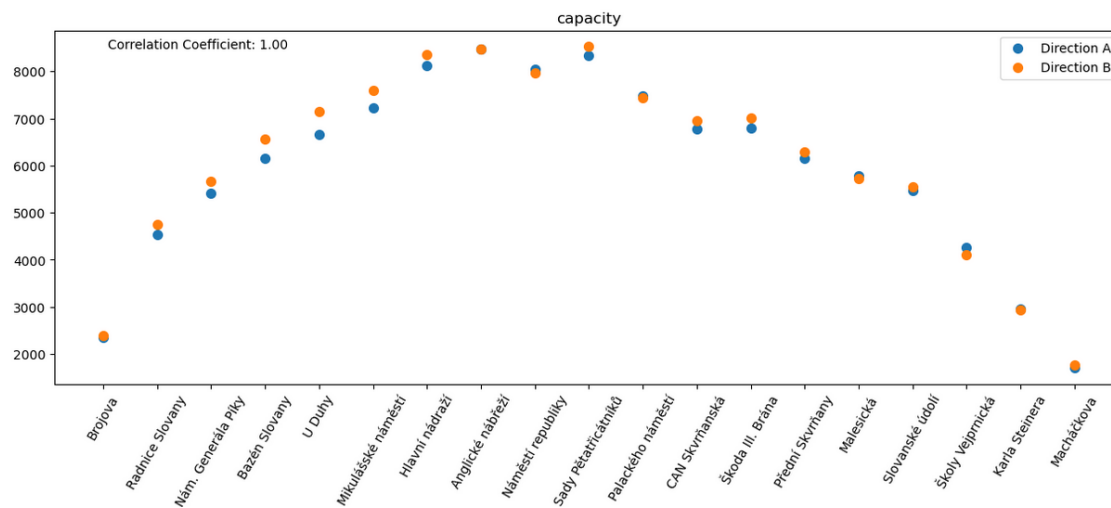
V grafech na obrázcích č. 34, 35 a 36 lze sledovat korelaci obsazenosti jednotlivých linek v závislosti na směru. Korelace je podle předpokládání u všech linek velice silná.

Obrázek č. 34: Závislost obsazenosti linky č. 1 na směru



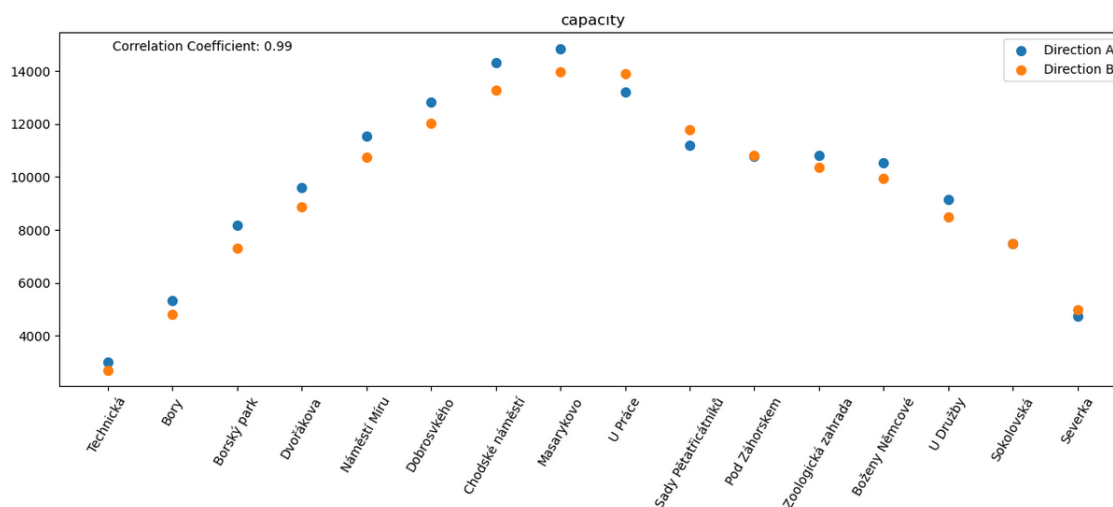
Zdroj: Vlastní zpracování, 2023

Obrázek č. 35: Závislost obsazenosti linky č. 2 na směru



Zdroj: Vlastní zpracování, 2023

Obrázek č. 36: Závislost obsazenosti linky č. 4 na směru

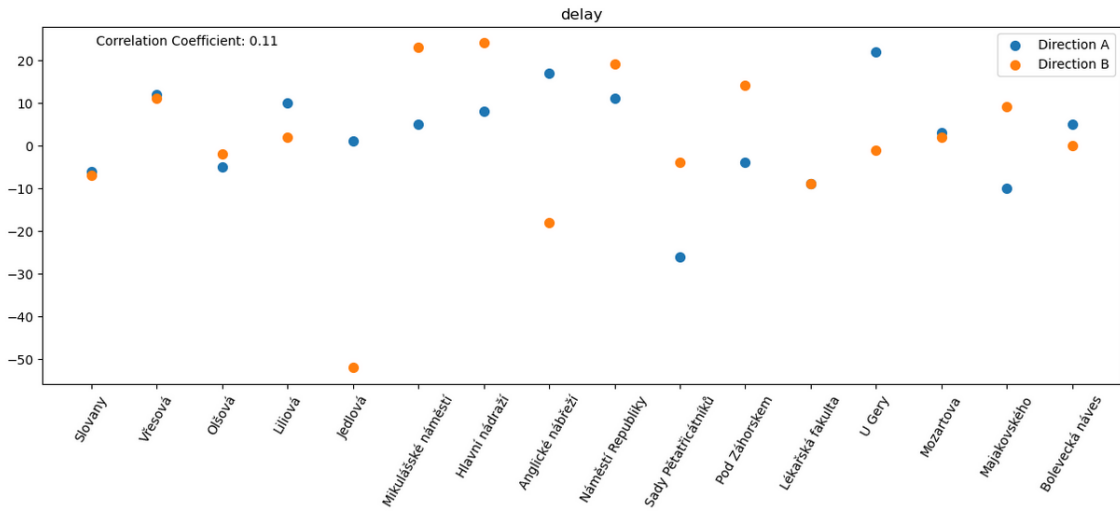


Zdroj: Vlastní zpracování, 2023

Další údaje pak poskytují korelace mezi směry pro zdržení, kde u každé linky lze pozorovat relativně malou korelaci.<sup>1</sup> Tato data lze považovat za překvapivá. Tramvaje jsou převážně vedeny po oddělené jízdni dráze a lze tak předpokládat menší vliv externích faktorů. Tato data však poukazují na to, že externí faktory zde hrají roli. Je zde také důležité zmínit, že tramvaje se v obou směrech nepohybují vždy stejnou trasou, jedná se například o segment Anglické nábřeží ↔ Náměstí Republiky u linek č. 1 a 2. Korelační koeficient je značně nižší u linek č. 1 (0,11) a 2 (0,21) než u linky č. 4 (0,43). Jedním z možných faktorů, který by toto mohl způsobovat je sdílení některých segmentů s běžnou dopravou a tím zvýšený vliv externích faktorů. Dalším možným faktorem může být také, že linka č. 4 se pohybuje v obou směrech po stejné trase. Bohužel úseky sdílené se zbytkem dopravy jsou v případě linek č. 1 a 2 zároveň také úseky s jinou trasou a nelze tak z těchto dat určit, který faktor hraje větší roli. Rozdíl mezi hodnotami lze pozorovat u linky č. 1 na obrázku č. 37 na úsecích Anglické nábřeží ↔ Náměstí Republiky a u linky č. 2 na obrázku č. 38 na úsecích U Duhy ↔ Mikulášské náměstí a Anglické nábřeží ↔ Náměstí Republiky. Rozdíly u linky č. 4 jsou zobrazeny na obrázku č. 39. Obecně lze také sledovat velké rozdíly u zastávek, které se nachází na místech se světelně řízenou křižovatkou, kde každý směr může mít jiné intervaly. Jedná se například o úseky Jedlová ↔ Mikulášské náměstí (37), Sady Pětatřicátníků ↔ Pod Záhorskem / Palackého náměstí (37, 38, 39), či Borský park ↔ Dvořákova (39).

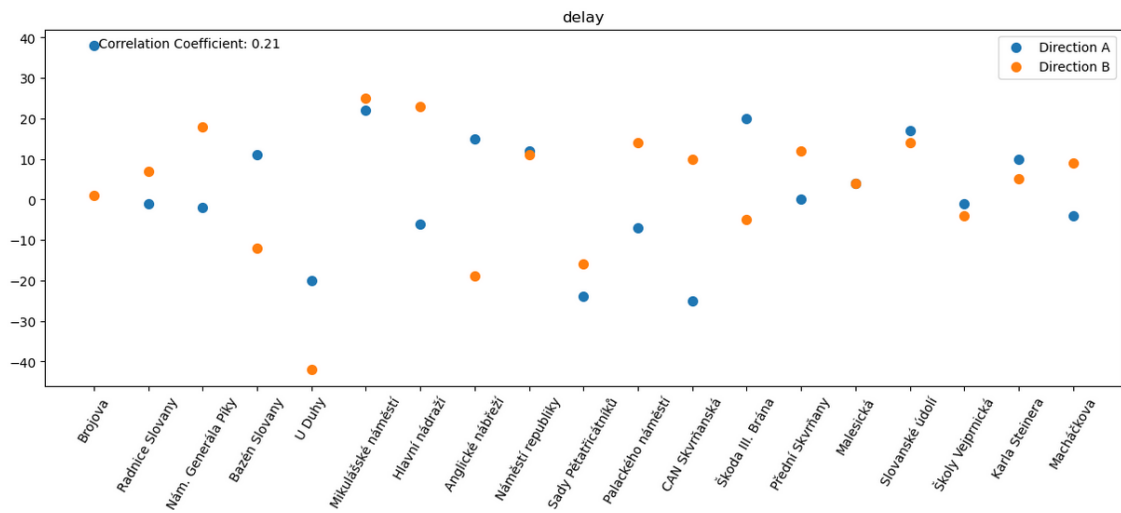
<sup>1</sup> Korelace zde vyjadřuje, jak směr jízdy ovlivňuje hodnoty na daných úsecích. Hodnota korelačního koeficientu rovna 1 by značila, že se jednotlivé úseky linky chovají z pohledu zdržení pro oba směry totožně, a tedy že na směru jízdy při plánování dopravy nezáleží. Hodnota rovna 0 pak vyjadřuje pravý opak.

Obrázek č. 37: Závislost zdržení linky č. 1 na směru



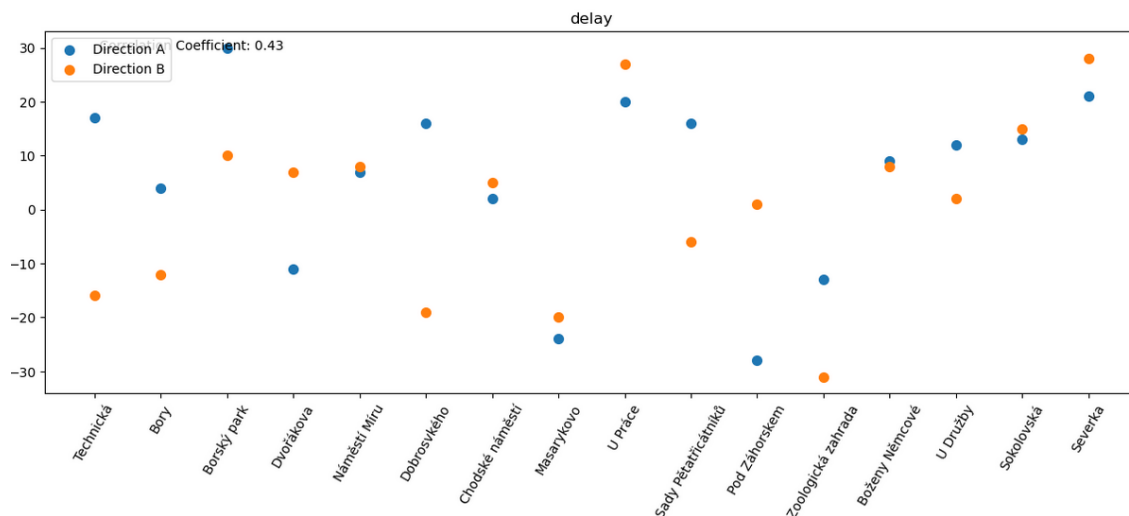
Zdroj: Vlastní zpracování, 2023

Obrázek č. 38: Závislost zdržení linky č. 2 na směru



Zdroj: Vlastní zpracování, 2023

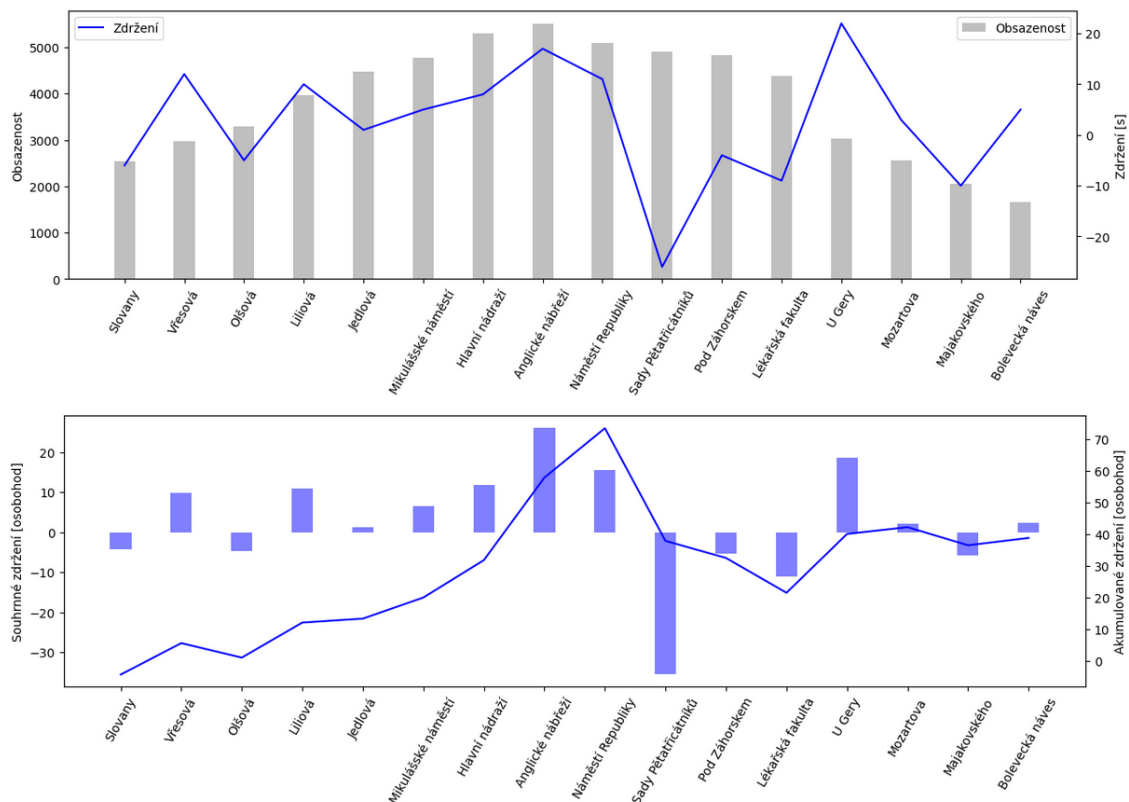
Obrázek č. 39: Závislost zdržení linky č. 4 na směru.



Zdroj: Vlastní zpracování, 2023

Dalším vygenerovaným statistickým ukazatelem pro oba směry každé linky jsou grafy zobrazující zdržení a obsazenost na daném úseku. Na obrázku č. 40 lze vidět grafy zobrazující data linky č. 1 ve směru Bolevec. První graf zobrazuje obsazenost pomocí sloupců (osa y na levé straně) a zdržení pomocí spojené čáry (osa y na pravé straně). Druhý graf pak zobrazuje souhrnné zdržení, tedy počet cestujících na daném úseku vynásobený zdržením daného úseku. Tuto hodnotu ukazují sloupce a lze ji vyčíst v ose y na levé straně (v osobohodinách). Měření v osobohodinách je pro plánování dopravy důležité, protože udává hromadné zdržení cestujících. Čára ve spodním grafu pak znázorňuje součet sloupců, tedy udává celkové zdržení od prvního úseku, také v osobohodinách (osa y na pravé straně). Z druhého grafu vyplývá, že nejvíce problematickými úseky linky č. 1 ve směru na Bolevec jsou Anglické nábřeží → Náměstí Republiky, U Gery → Mozartova a Náměstí Republiky → Sady Pětatřicátníků. Také je patrné, v zastávce Hlavní pošta (v grafech Sady Pětatřicátníků, směr A, u linek 1 a 2) dochází k velkému dorovnávání zpoždění. Celkově naakumulované zdržení je pak 38,8 osobohodin.

Obrázek č. 40: Kapacita a zdržení linky č. 1, směr A (Bolevec)



Zdroj: Vlastní zpracování, 2023

Ve směru Slovany je naakumulované zdržení značně menší a činí pouze 21,9 osobohodin a to i přes to, že se zde nachází nejhorší úsek na celé lince 1, úsek od Anglického nábřeží, kde dochází ke zdržení necelých 40 osobohodin. Další problematické úseky jsou pak od zastávky Hlavní nádraží, Sady Pětatřicátníků a Lékařská fakulta. Důvodem celkové nízkého akumulovaného zdržení jsou úseky od Náměstí Republiky, kde dochází ke zdržení - 25 osobohodin a na úseku od Mikulášského náměstí, kde tramvaj „dožene“ v průměru 55 vteřin a způsobuje tak pokles akumulovaného zdržení o 60 osobohodin.

Linka č. 2 ve směru na Světovar má celkově naakumulované zdržení 53,2 hodin. Nejvíce problematické segmenty jsou od zastávek Mikulášské náměstí, Anglické nábřeží a Škoda III. brána, kde dochází ke zdržení v rozmezí 30 - 40 osobohodin. Stejně jako u linky č. 1, na zastávce Hlavní pošta dochází k dorovnání zpoždění téměř 60 osobohodin. Další zastávky, kde dochází k umazání velké ztráty zpoždění jsou CAN Skvrňanská (cca 45 osobohodin) a U Duhy (téměř 40 osobohodin). V opačném směru má linka celkově naakumulované zdržení 72,1 hodin, kde stejně jako u linky č. 1 dochází k největšímu zdržení v osobohodinách v úsecích od Náměstí Republiky a od Hlavního nádraží, obě



okolo 60 osobohodin. A stejně jako u linky 1 pak dochází k největšímu dorovnání zpoždění, cca 80 osobohodin, na úseku od Mikulášského náměstí.

Linka č. 4 ve směru na Lochotín má celkově naakumulované zdržení 158,9 hodin a představuje tak značně nejhorší linku ze všech tramvají v této metrice. Nejhoršími úseky jsou U Práce → Sady Pětatřicátníků, Borský park → Dvořákova, Dobrovského → Chodské náměstí a Sady Pětatřicátníku → Pod záhorskem, kde dochází ke zdržení v rozmezí 50 - 75 osobohodin. K velké redukci ztrát dochází na úsecích od Masarykovo náměstí (skoro 100 osobohodin) a od zast. Pod Záhorskem (cca 75 osobohodin). Velká ztráta na lince je způsobena i přes to, že na žádné zastávce zdržení nepřesáhne 30 sekund a je tak způsobeno hlavně vysokou kapacitou, kterou linka přepravuje. I přes vysokou kapacitu v opačném směru dosahuje akumulované hodnoty pouhých 5,9 osobohodin a je tak značně nejlepší linkou v této metrice. Velký rozdíl je způsoben hlavně rozdílným zdržením na zmiňovaných zastávkách v blízkosti světelně řízené dopravy, kde například zastávka Borský park ve směru na Bory způsobuje o 80 osobohodin nižší zdržení než v opačném směru. Nejhorší úsek této linky je od Sadů Pětatřicátníků, kde zdržení dosahuje cca 100 osobohodin. Na všech dalších úsecích však dochází již k relativně malým zdržením a na šesti dochází k dorovnání zpoždění.

Tyto údaje také potvrzují tvrzení, že je důležité brát v ohledu směr jízdy, kde metriky zdržení v osobohodinách jsou značně jiné pro každý směr. U linky č. 1 je akumulované zdržení skoro 2krát větší v opačném směru, u linky č. 4 je větší skoro 27krát.

Grafy s relevantními informacemi ohledně tramvají lze nalézt v příloze C.

K další analýze a vizualizaci těchto údajů lze využít zabudovaného mapového nástroje.

Nejprve lze vizualizovat všechny segmenty, které nedosahují známky 1 podle interního známkování PMDP. Viditelné na obrázku č. 41.

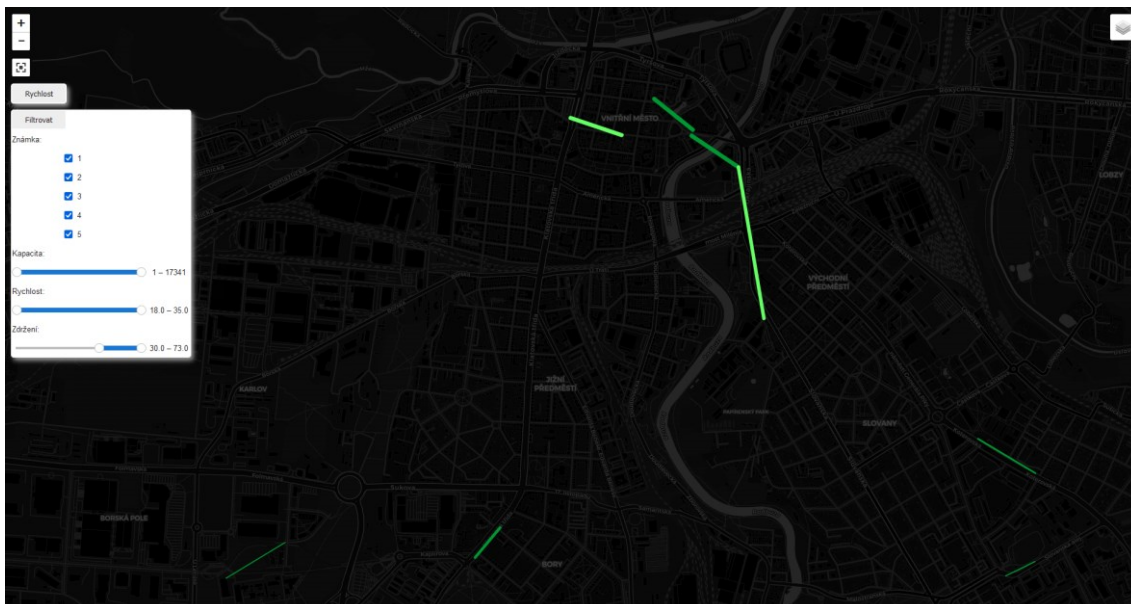
Obrázek č. 41: Tramvajové úseky nedosahující známky 1



Zdroj: Vlastní zpracování, 2023

Dále pak můžeme vyobrazit segmenty, kde dochází k velkému zdržení, tj.  $>30\text{sec}$ . Segmenty jsou znázorněny na obrázku č. 42. Zde je zajímavé, že k relativně velkým zdržením nedochází pouze v centru s hustou dopravou, ale i na místech blízko konečným stanicím.

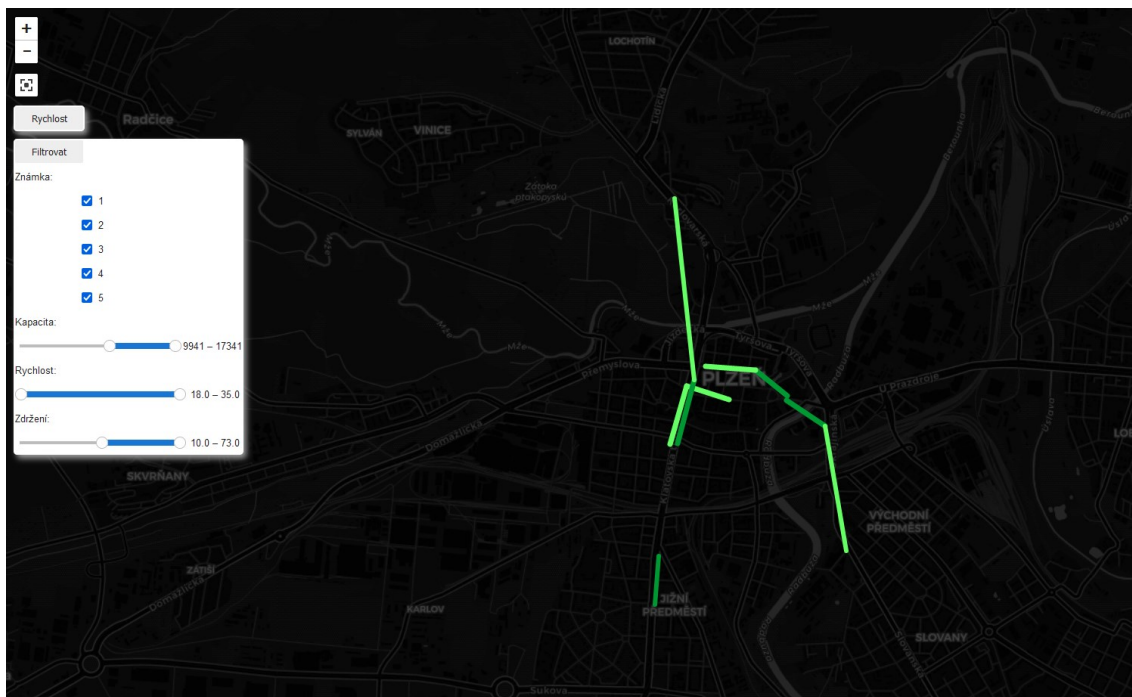
Obrázek č. 42: Tramvajové úseky se zdržením větším než 30 sekund



Zdroj: Vlastní zpracování, 2023

Jak již bylo zmíněno, pro PMDP jsou nejvíce problematické úseky, kde je zdržení a zároveň velká kapacita, tyto úseky lze vidět na obrázku č. 43. Tyto úseky jsou již dle očekávání blízko centra a jedná se o úseky linky 4 a společné úseky linek 1 a 2.

Obrázek č. 43: Tramvajové úseky s vysokou obsazeností a zdržením



Zdroj: Vlastní zpracování, 2023

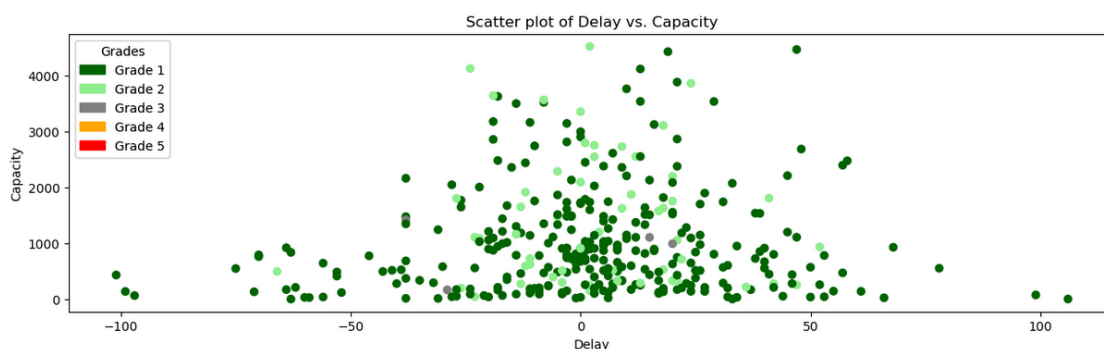
Obecně však lze říci, že tramvaje představují velice spolehlivý způsob dopravy, kde nejlepší známky podle interního hodnocení nedosahuje pouze 12 ze 116 úseků a úseků s vysokou kapacitou (>9 500) a zdržení přesahující 10 vteřin je na pouhých devíti úsecích. Na 4 z těchto úseků dosahuje zdržení 20 a více vteřin a jsou to tedy úseky, kterým by se měla věnovat pozornost. Těmito úseky jsou některé společné úseky linek 1 a 2: Anglické nábřeží → Hlavní nádraží, Hlavní nádraží → Mikulášské náměstí. A úsek linky č. 4 Sady Pětatřicátníků ↔ U Práce. Tento úsek vede přes tři světelně řízené křižovatky, kde případné řešení by se mohlo nacházet v úpravě intervalů světelné signalizace.

#### 4.5.2 Trolejbusy

Stejně jako u tramvají, je prvním krokem analýzy této traktace nalezení extrémních hodnot. Všechny boxové grafy jsou dostupné v příloze D. Na rozdíl od tramvají, kde k extrémním hodnotám docházelo jen u zdržení, u trolejbusů lze nalézt extrémní hodnoty u všech 3 měřených metrik. Ke všem těmto extrémním hodnotám je nutno brát kontext konkrétní linky. U linky č. 11 jsou extrémně nízké hodnoty obsazenosti způsobeny zastávkami

v Křimicích a Malesicích, kde není velké množství lidí. Trolejbus pak pokračuje rovnou do centra, kde obsluhuje velké množství. Po prozkoumání jednotlivých extrémních hodnot je lze považovat za správná měření. Stejně jako u tramvají budou z analýzy pro další metriky vyřazeny koncové zastávky. Na grafech lze sledovat různé rozptyly obsazenosti, které jsou v porovnání s tramvaji řádově nižší. Toto potvrzuje graf na obrázku č. 44. Graf dále ukazuje korelaci interního známkování se zdržením, kde lze vidět že většina segmentů má opět známku 1. Na grafu lze také zpozorovat několik segmentů se známkou 2 a pár segmentů se známkou 3. Většina zdržení se nachází v intervalu (-50s, 50s). Dále lze také pozorovat několik úseků, na kterých dochází k velice vysokému zdržení, u většiny z nich se však jedná o úseky s relativně malou obsazeností.

Obrázek č. 44: Bodový graf zdržení a kapacity trolejbusů.



Zdroj: Vlastní zpracování, 2023

Korelace obsazenosti v obou směrech linek je nižší než u tramvají, ale stále se pohybuje relativně vysoko, v rozmezí 0,82 – 1 s výjimkou linky č. 18, která má pouze 0,64, tato linka obsluhuje trasu CAN – Borská pole, kde k rozdílu dochází v zastávkách na Borských polích. Směrem z Borských polích obsazenost roste mnohem rychleji. To lze odůvodnit pracovníky z tamních podniků, kteří zvolí při odvozu domu prostředek, který přijede nejdříve, i za cenu větší vzdálenosti domu. Směrem do práce pak budou volit spoj, který je bydlišti nejbližší. Data s korelací trolejbusů jsou dostupná v příloze E.

Korelace rychlosti a zdržení, které lze vidět na grafech v příloze F, opět poukazují na skutečnost, že při plánování dopravy není nutné brát v potaz pouze linku, ale také směr jízdy, kde u některých linek je korelace v průměrné rychlosti relativně vysoká, u jiných je korelace skoro nulová. U zdržení pak také záleží na konkrétní lince, v průměru je závislost zdržení však značně nižší než závislost průměrné rychlosti. Největší rozdíly v průměrné rychlosti lze vidět u trolejbusů č. 10, 13 a 14 na trase mezi Doudlevcí ETZ a U Plynárny. Toto je způsobeno jak příznivější světelnou signalizací pro směr Černice /

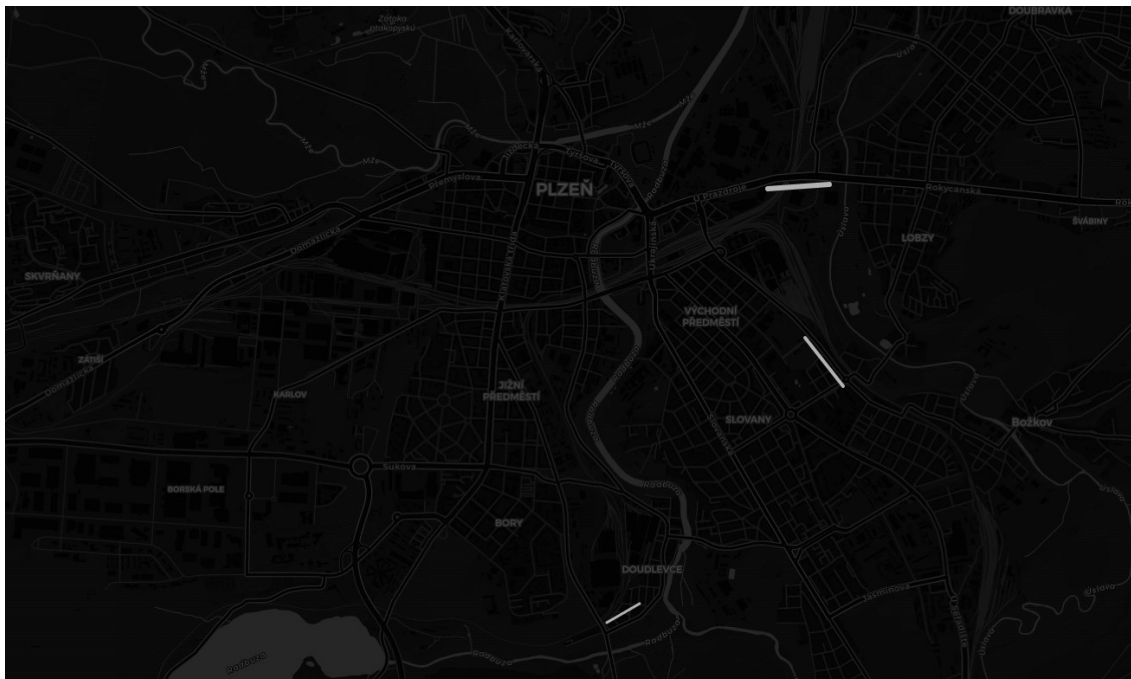
Bory, tak zastávkou navíc v opačném směru, kde je trolejbus nucen zastavit a tím snížit průměrnou rychlost. Mezizastávka (Samaritská) není na korelačních grafech zobrazena, protože nemůže být přidružena zastávce v druhém směru. Z grafů jde dále vyčíst, že na okrajích Plzně a v příměstských oblastech je korelace vysoká, lze vidět například u linky 12 u zastávek v Letkově, následně po vjezdu trolejbusu do města se korelace snižuje a pak zase s výjezdem města se zvyšuje. Tyto poznatky poukazují na fakt, že velkým faktorem na nízké korelace je rozdílná hustota dopravy ve městě v závislosti na směru. Příměstské oblasti s nižší dopravou pak nejsou tolik ovlivněny. Naopak u linek 17 a 18 lze pozorovat vysokou korelaci rychlosti na délce celého spoje i přes to, že většina projíždí skrz město. Lze stanovit, že rychlost pro každý směr na stejném úseku není ovlivněna pouze jedním faktorem jako je hustota dopravy a při plánování je nutné brát v potaz konkrétní specifika daných úseků a jednotlivě zkoumat, jaké faktory v daném místě hrají roli.

Grafy zobrazující kapacitu a zdržení jsou dostupné v příloze G. Rozdíly akumulovaného zdržení pro opačné směry jedné linky pak vychází podobně jako u tramvají velice rozdílně v závislosti na směr.

Důležitost měření v osobohodinách a zaměření se na úseky, kde je velké zdržení i kapacita, lze vidět například u linky č. 10 ve směru B (Tylova). Z prvního grafu lze vyčíst, že zdržení na úseku K Plzenci → Generála Lišky bylo velice podobné jako na úseku Zimní stadion → U Radbuzy. V následujícím grafu (v příloze) je pak vidět dopad jednotlivých zastávek na celkové zdržení, kde úsek od Zimního stadionu způsobuje řádově větší „škody“.

Na identifikaci problematických segmentů všech linek a jejich zobrazení je využit nástroj interaktivní mapy. Na obrázku č. 45 jsou zobrazeny všechny segmenty se známkou 3. Jedná se o 3 různé segmenty. Prvním je Révová → Tyršův most na lince č. 14. Druhým je Stadion Lokomotivy → Nádraží Slovany na lince č. 12. Posledním segmentem je pak Gambrinus → Divadlo Alfa, společný segment pro linky č. 11 a 15. Tento segment je značně nejproblematičtější, protože ho využívají 2 linky a celková obsazenost na segmentu je přes 2 100 lidí. Problémový segment linky 12 má obsazenost 1 415, což je u trolejbusů relativně velké číslo. U těchto 2 segmentů by tedy bylo vhodné zvážit možné kroky k napravení problému. Problematický segment linky č. 14 by měl být také řešen, vzhledem k obsazenosti 164 však není problém akutní.

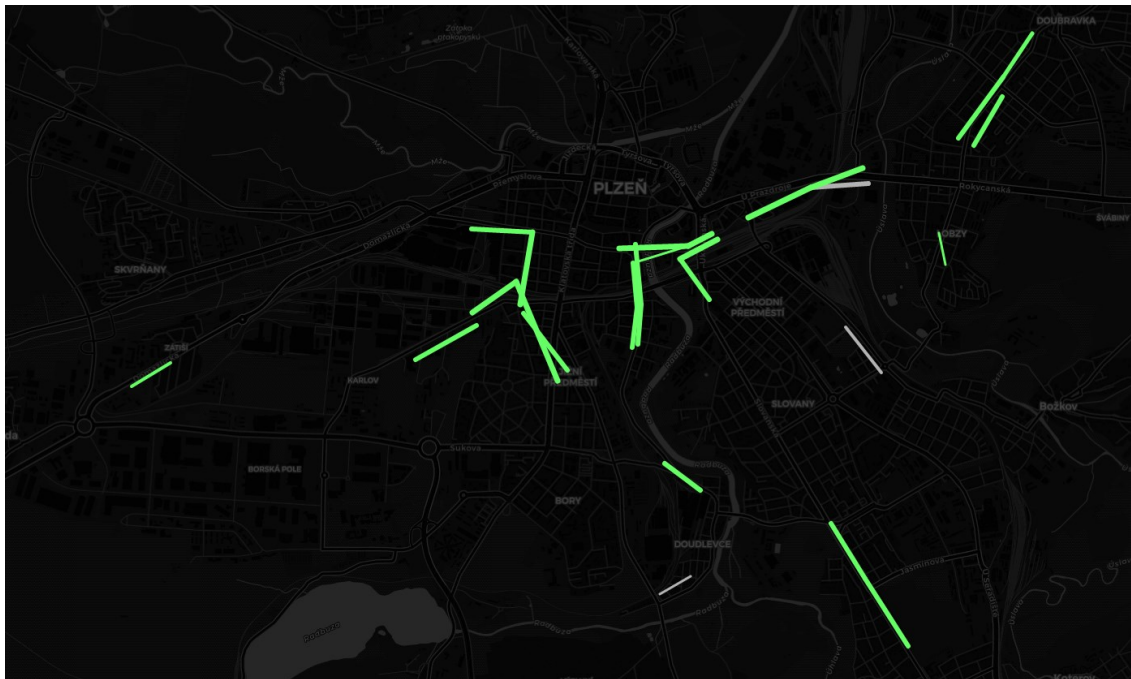
Obrázek č. 45: Trolejbusové úseky se známkou 3



Zdroj: Vlastní zpracování, 2023

Úseků se známkou horší než 1 je již více, zobrazené jsou na obrázku č. 46. Většina těchto segmentů se nachází v centru a až na pár výjimek jsou segmenty sdíleny více linkami.

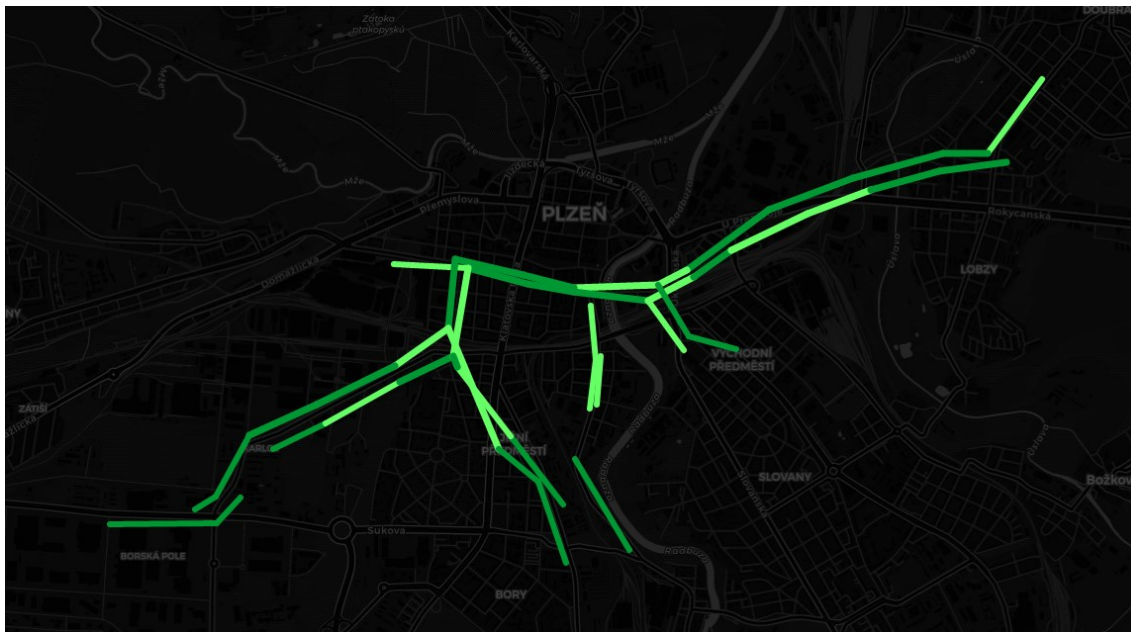
Obrázek č. 46: Trolejbusové úseky se známkou horší než 1



Zdroj: Vlastní zpracování, 2023

Vyfiltrováním úseků s velkou kapacitou pak dostaneme mapu na obrázku č. 47, která ukazuje části města, kde je trolejbusové doprava páteří. Jedná se o centrum města a napojení na Doubravku, Doudlevec a Bory.

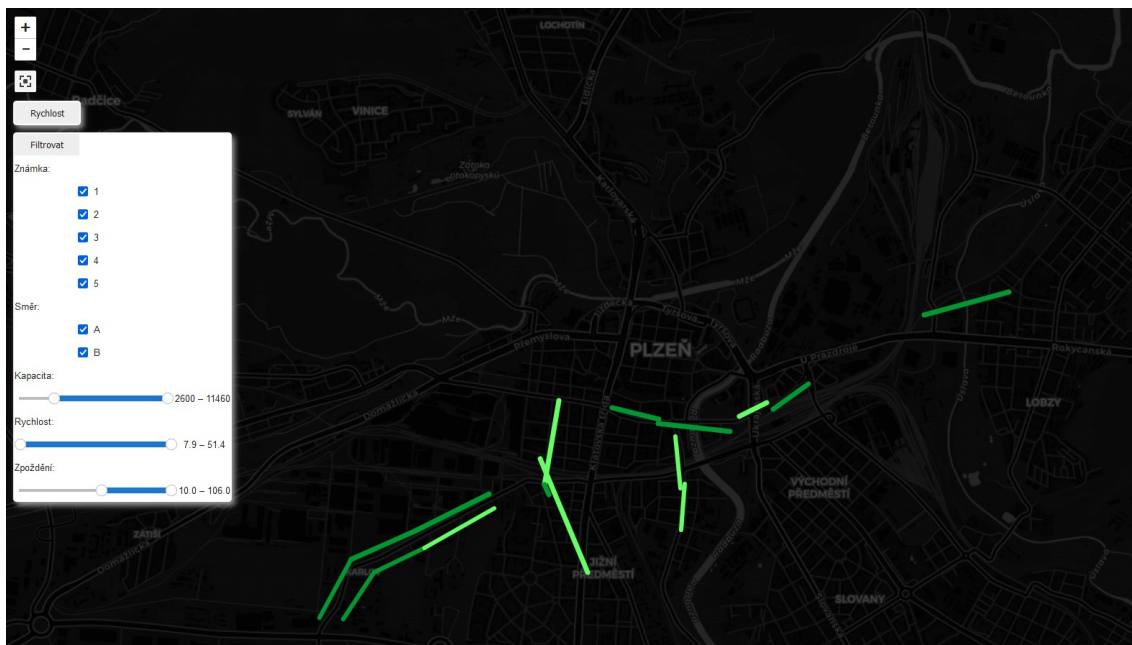
Obrázek č. 47: Trolejbusové úseky s vysokou obsazeností.



Zdroj: Vlastní zpracování, 2023

Při přidání zdržení  $>10$  sekund do filtru pak zobrazí úseky na které by bylo vhodné se zaměřit. Tyto úseky jsou vidět na obrázku č. 48. Problematické úseky se opět soustřeďují v centru. Tyto úseky využívají již velké množství různých linek a optimalizace ze strany PMDP by v těchto místech nejspíš neměla moc velký dopad, protože hlavním úzkým místem se zdá být hustota dopravy v těchto místech. K řešení těchto úseků by se tak mělo přistupovat na celoměstské úrovni a řešit celkovou dopravní situaci na těchto místech. Jedním z těchto problémových úseků je Americká ulice, kde došlo k uzavírcce její střední části pro všechna vozidla mimo veřejné dopravy a zásobování. Zdržení na Americké ulici se pohybuje pod 10 sekund ve směru U Práce  $\rightarrow$  Mrakodrap, v opačném směru pak okolo 20 sekund. Zdržení je zde i přes uzavření vysoké, ale je značně menší než úsek Mrakodrap  $\rightarrow$  Pařížská, kde je zdržení přes 45 sekund u všech linek. Tento rozdíl by mohl být způsoben právě částečným uzavřením Americké ulice, pro potvrzení je však nutné provést porovnání a hlubší analýzu s daty před uzavřením, kterou tento nástroj umožňuje. Jedním z takových řešení by mohl být také právě otevíraný městský okruh, kde bude důležité sledovat změny v dopravě, které způsobí.

Obrázek č. 48: Trolejbusové úseky s velkou obsazeností a zdržením



Zdroj: Vlastní zpracování, 2023

Vygenerovaná tabulka pro tramvaje a trolejbusy a její import do Excelu pak poskytuje srovnání jednotlivých linek a vozidel (vizte obrázek č. 49). Data ukazují, že tramvaje se na dopravě v Plzni podílejí větší částí než trolejbusy, i přes menší počet linek. Ve zdržení na osobu a počet přepravených osob na zastávku vycházejí nejlépe tramvaje. V průměrném zdržení na zastávku pak vychází lépe trolejbusový subsystém (počet přepravených osob je počítán ze zastávky na zastávku, tedy osoba jedoucí 3 zastávky se v datech promítne jako +3). Tramvaje jsou však více závislé na infrastruktuře, která je nákladná na výstavbu a údržbu. Trolejbusy disponují větší flexibilitou a mohou tak obsluhovat oblasti mimo centrum, nebo lokace kde tramvajová infrastruktura nemůže být vystavěna. Pro zvolení správného dopravního prostředku je vždy nutno brát v potaz konkrétní požadavky a možnosti daného místa.



Obrázek č. 49: Import dat z vygenerované tabulky „performance\_table.csv“

	A	B	C	D	E	F	G	H
1	Linka č.	Směr	Průměrná rychlost[km/h]	Celkové zdržení[osobohod]	Přepravených osob	Zdržení na osobu[sec]	Zdržení na zastávku[osobohod]	Osob na zastávku
2	4 A		27	158,9	160664	3,56	8,83	8928,0
3	4 B		26,89	29,99	154755	0,7	1,67	8617,6
4	2 B		26,95	71,37	117251	2,19	3,4	5585,7
5	2 A		27,33	55,08	114851	1,73	2,62	5463,1
6	1 B		26,67	20,95	67468	1,12	1,16	3735,7
7	1 A		25,56	42,73	62128	2,48	2,37	3445,9
8	16 B		21,79	29,58	57238	1,86	1,48	2863,8
9	16 A		22,69	54,84	52273	3,78	3,05	2907,2
10	15 A		26,64	60,67	35369	6,18	2,89	1684,8
11	13 A		25,8	9,98	33301	1,08	0,36	1201,2
12	12 B		27,17	18,91	32807	2,08	0,76	1318,5
13	12 A		28,03	60,84	32341	6,77	2,43	1291,7
14	15 B		25,77	38,55	30880	4,49	1,75	1401,8
15	13 B		22,1	27,91	28367	3,54	1,07	1087,5
16	11 B		26,39	24,55	17247	5,13	1,23	864,1
17	11 A		28,42	3,6	16863	0,77	0,17	796,3
18	10 A		26,05	13,22	9355	5,09	0,7	495,3
19	10 B		19,74	4,74	7988	2,14	0,26	438,2
20	17 B		25,36	21,07	7589	10	0,92	331,4
21	17 A		26,11	15,48	6163	9,04	0,74	294,6
22	14 B		22,26	1,58	2917	1,95	0,12	221,5
23	14 A		25,22	2,6	2506	3,73	0,2	192,8
24	18 B		27,61	4,29	2291	6,75	0,43	229,6
25	18 A		27,45	7,36	2213	11,97	0,67	201,5
26	Tram	Celk	26,7	63,2	677117,0	2,0	3,3	5962,7
27	T bus	Celk	25,3	22,2	377708,0	4,8	1,1	990,1

Zdroj: Vlastní zpracování, 2023

Nejvíce problematické úseky lze identifikovat na základě několika metrik. Jedním z hlavních cílů vytvořeného nástroje byla identifikace úseků s vysokou kapacitou a zároveň s velkým zdržením. Tyto údaje vyjadřuje zdržení v osobohodinách a nejhorší segmenty z pohledu této metriky jsou zobrazeny v tabulce č. 1. Většina těchto úseků se nachází v blízkosti centra města a vyznačují se vysokou obsazeností. Společný faktor u všech úseků je jedna či více světelně řízených křižovatek, kde ke zlepšení plynulosti provozu mělo být přistupováno společnými aktivitami městských správ a dopravních podniků. Jistá míra zdržení je však neodstranitelná, a to zejména v centrálních, vysoce zatížených oblastech. Typicky v křižovatkách, do kterých spoje MHD přijíždí z více směrů, a jejich absolutní preference je tak prakticky neproveditelná (např. tramvajový uzel Sady Pětatřicátníků, křižovatka U Práce, křižovatka Americká x Mikulášská x Ukrajinská x Šumavská).

Tabulka č. 1: Úseky s největším zdržením v osobohodinách

Úsek	Trakce	Linky	Zdržení[osobohod]	Potencionální příčina
Mrakodrap → Pařížská	Trolejbus	11, 12, 15, 16, 17	146,6	Úsek se nachází v centru s velkou hustotou dopravy. Na úseku je dále také kruhový objezd, světelně řízená křižovatka a dále přechody pro chodce.

Sady Pětaticátník ů → U Práce	Tramvaj	4	104,2	Tramvaj projíždí třemi světelně řízenými křižovatkami. Na první křižovatce tramvaj také sdílí kolej s linkami č. 1 a 2, které na křižovatce čekají na levé odbočení.
Anglické nábřeží → Hlavní nádraží	Tramvaj	1, 2	91,4	Část úseku tramvaje sdílí s běžnou dopravou a jsou tak ovlivněny hustotou dopravy. Na úseku se také nachází světelně řízené křižovatka, do které vjíždí z vedlejší ulice (nicméně tramvaje jsou zde preferovány).
Hlavní nádraží → Mikulášské náměstí	Tramvaj	1, 2	80,7	Na tomto úseku se nachází tři světelně řízené křižovatky, na všech je však tramvaj na hlavní komunikaci a s preferencí světelných signálů. Na tomto úseku také dochází k překřížení kolejí a běžné silnice. Rizikovým místem může být Mikulášské náměstí, kde dochází k častým dopravním nehodám při levém odbočování automobilů přes tramvajovou trať a řidiči tramvají tak zde mohou podvědomě jezdit opatrněji.
U Práce → Sady 35	Tramvaj	4	73,3	Stejně jako v opačném směru se tramvaj setkává s třemi světelně řízenými křižovatkami. V tomto směru nedochází ke sdílení nástupní hrany s linkami č. 1 a 2, což by mohlo vysvětlovat menší celkové zdržení (nicméně zdržení může nastávat s ohledem na křížení tras při levém odbočování linek 1 a 2).

Zdroj: Vlastní zpracování, 2023

Analýza byla provedena na datech za jeden měsíc a nelze tak vyloučit vliv krátkodobých uzavírek a omezení provozu. Pro podrobnější analýzy lze provádět analýzu na datech za dlouhé časové období, kterou vyvinutý nástroj bez problémů umožňuje.

## 5 Závěr

Cílem práce bylo poskytnout přehled o sběru a využití dat v oblasti městské hromadné dopravy, analyzovat provozní data, identifikovat možnosti zlepšení dopravního systému a též demonstrovat, jak lze efektivní zpracování a využití dat přispět k lepšímu pochopení provozu hromadné dopravy.

Teoretická část práce se zabývala základními pojmy týkajícími se dat, ochranou soukromí a využitím dat v kontextu městské hromadné dopravy. Ve světle těchto poznatků byl v praktické části vytvořen vlastní nástroj pro analýzu a vizualizaci dat, který umožňuje dopravním podnikům nejen lépe pochopit a optimalizovat své služby, ale také předat zjištěné poznatky díky vizuální prezentaci. S využitím tohoto nástroje byla následně analyzována data o provozu tramvají a trolejbusů za březen 2023. Dále bylo diskutováno o možných příčinách identifikovaných výsledků.

Analýza dat pomocí tohoto nástroje přinesla několik klíčových zjištění. Především bylo zjištěno, že směr jízdy má značný vliv na různé metriky linek, což by mělo být zohledněno při plánování a optimalizaci dopravních tras i tvorbě jízdních řádů.

Také byly identifikovány nejvíce problémové úseky trolejbusových linek. Většina se nachází v blízkosti centra města a na větších dopravních uzlech. Tato místa jsou charakteristická tím, že jimi projíždí různé linky a všechny zde čelí významným problémům. Zjištění ze studie naznačují, že pouhé nasazení většího množství trolejbusů nebo jiné optimalizace v těchto oblastech ze strany městského dopravního podniku (PMDP) nebudou tak účinné, jak by mohlo být očekáváno. Namísto toho by měly být zváženy změny na městské úrovni, které ovlivní dopravu jako celek.

Jednou z limitací této práce je analýza na datech, které představují průměrné hodnoty za celé dny a nezachycují tak závislost dopravy na denní době. Budoucí analýzu s daty v určitém časovém období však autorem vyvinutý nástroj (popsaný v práci) bez problémů umožňuje.

Závěrem lze konstatovat, že řešení problémových linek a dopravních uzlů ve městech vyžaduje komplexní přístup, který zahrnuje změny na městské úrovni a úzkou spolupráci mezi různými aktéry. Využití dat a jejich analýza může hrát klíčovou roli při identifikaci největších problémů a nalezení efektivních řešení. Spolupráce mezi městskými správami,

dopravními podniky a dalšími zainteresovanými stranami je nezbytná pro dosažení cílů v oblasti udržitelnosti, efektivity a komfortu veřejné dopravy.

Výsledky této práce by mohly přispět k diskusi o využití dat v městském prostředí a jejich roli při zlepšování veřejných služeb. Zároveň by mohly poskytnout užitečné informace pro odborníky, kteří se zabývají městskou dopravou, urbanismem a udržitelností ve městech.

## Seznam použitých zdrojů

- ACSI. (2023) [online] Dostupné z <https://www.theacsi.org/>.
- Ahmad, F., Saad Alam, M., Saad Alsaidan, I., & Shariff, S. M. (2020). Battery swapping station for electric vehicles: opportunities and challenges. *IET Smart Grid*, 3(3), 280-286.
- Ashmore, D. P., Pojani, D., Thoreau, R., Christie, N., & Tyler, N. A. (2019). Gauging differences in public transport symbolism across national cultures: Implications for policy development and transfer. *Journal of Transport Geography*, 77, 26–38.
- Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. *New Directions for Evaluation*, 2013(139), 7-32.
- Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M., & Puchinger, J. (2019). Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 101, 254–275.
- Becker, J. U., & Albers, S. (2016). The limits of analyzing service quality data in public transport. *Transportation*, 43(5), 823-842.
- Burnay, C., Dargam, F., & Zarate, P. (2019). Data visualization for decision-making: an important issue. *Operational Research*, 19, 853-855.
- Bitencourt, V. N., Crestani, F., Peuckert, M. Z., Andrades, G. R. H., Krauzer, J. R. M., de Cassia Cintra, C., ... & Garcia, P. C. R. (2023). Net Promoter Score (NPS) as a tool to assess parental satisfaction in pediatric intensive care units. *Jornal de Pediatria*.
- Clairand, J. M., Guerra-Terán, P., Serrano-Guerrero, X., González-Rodríguez, M., & Escrivá-Escrivá, G. (2019). Electric vehicles for public transportation in power systems: A review of methodologies. *Energies*, 12(16), 3114.
- Cui, Y., Xu, H., Wu, J., Sun, Y., & Zhao, J. (2019). Automatic vehicle tracking with roadside LiDAR data for the connected-vehicles system. *IEEE Intelligent Systems*, 34(3), 44-51.
- Crist, P., Greer, E., Ratti, C., Humanes, P., Konzett, G., Tijink, J., Figuero, D., & Lax, R. (2015). Big data and transport: Understanding and assessing options. *International Transport Forum Data Base*.
- Český Statistický Úřad. (2020). Informační společnost v číslech 2020. [online] Dostupné z <https://www.czso.cz/csu/czso/informacni-spolecnost-v-cislech-2020>.
- Český Statistický Úřad. (2022). ICT infrastruktura. [online] Dostupné z [https://www.czso.cz/csu/czso/telekomunikacni\\_a\\_internetova\\_infrastruktura](https://www.czso.cz/csu/czso/telekomunikacni_a_internetova_infrastruktura).
- Daniels, R., & Mulley, C. (2013). Explaining walking distance to public transport: The dominance of public transport supply. *Journal of Transport and Land Use*, 6(2), 5–20.
- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library review*.
- Dixon, W. J. (1953). *Processing data for outliers*. *Biometrics*, 9(1), 74-89.
- Eskildsen, J. K., & Kristensen, K. (2011, June). The accuracy of the net promoter score under different distributional assumptions. In *2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering* (pp. 964-969). IEEE.

- Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, 35(1), 138-151.
- Ge, L., Sarhani, M., Voß, S., & Xie, L. (2021). Review of transit data sources: Potentials, challenges and complementarity. *Sustainability*, 13(20), 11450.
- Godfrid, J., Radnic, P., Vaisman, A., & Zimányi, E. (2022). Analyzing public transport in the city of Buenos Aires with MobilityDB. *Public Transport*, 14(2), 287-321.
- Gordon, J. B., Koutsopoulos, H. N., Wilson, N. H., & Attanucci, J. P. (2013). Automated inference of linked transit journeys in london using fare-transaction and vehicle location data. *Transportation research record*, 2343(1), 17–24.
- Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2014). Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 26(9), 2250-2267. doi: 10.1109/TKDE.2014.2300803.
- Hsu, Y. W., Wang, T. Y., & Perng, J. W. (2020). Passenger flow counting in buses based on deep learning using surveillance video. *Optik*, 202, 163675.
- Ilango, V., Subramanian, R., & Vasudevan, V. (2012). A five step procedure for outlier analysis in data mining. *European Journal of Scientific Research*, 75(3), 327-339.
- Jupyter. (2023). Project Jupyter. <https://jupyter.org/>
- Kalghatgi, G. (2018). Is it really the end of internal combustion engines and petroleum in transport?. *Applied energy*, 225, 965-974.
- Kirubakaran, A., Jain, S., & Nema, R. K. (2009). A review on fuel cell technologies and power electronic interface. *Renewable and sustainable energy reviews*, 13(9), 2430-2440.
- Knowles, R. D., Ferbrache, F., & Nikitas, A. (2020). Transport's historical, contemporary and future role in shaping urban development: Re-evaluating transit oriented development. *Cities*, 99, 102607.
- Kupaysinov, I. (2022). PUBLIC TRANSPORTATION: A GLANCE AT THE HISTORY OF THE TRAM. *Eurasian Journal of Academic Research*, 2(12), 879-880.
- Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. *Korean Journal of Anesthesiology*, 70(4), 407-411. <https://doi.org/10.4097/kjae.2017.70.4.407>
- Litman, T. (2015). *Evaluating public transit benefits and costs*. Victoria Transport Policy Institute Victoria, BC, Canada.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons.
- Manoharan, Y., Hosseini, S. E., Butler, B., Alzahrani, H., Senior, B. T. F., Ashuri, T., & Krohn, J. (2019). Hydrogen fuel cell vehicles; current status and future prospect. *Applied Sciences*, 9(11), 2296.
- Mantelero, A. (2021). *The future of data protection: Gold standard vs. global standard [Preprint]*. *Computer Law & Security Review*, 40. <https://doi.org/10.1016/j.clsr.2020.105500>

- Michta, A., & Haniszewski, T. (2018). Traffic noise experienced on buses, trams and cars in the urban agglomeration of the city of Katowice. *Zeszyty Naukowe. Transport/Politechnika Śląska*.
- Miranda-Segundo, C. A. (n.d.). Data privacy and data protection. *Journal of Cybersecurity Research*, 1(1), 1-10. <https://doi.org/10.35429/JITC.2020.11.4.12.16>
- Morlion P, H. S. (2021). *Eib technical note on data sharing in transport*. European Investment Bank.
- Moutassem, B., Djafri, L., & Gafour, A.-K. (2022). Big data veracity: Methods and challenges benabderrahmane moutassem · laouni djafri · abdel-kader gaafour.
- Myrvoll, T. A., Håkegård, J. E., Matsui, T., & Septier, F. (2017, October). Counting public transport passenger using WiFi signatures of mobile devices. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1-6). IEEE.
- Nishiuchi, H., Todoroki, T., & Kishi, Y. (2015). A fundamental study on evaluation of public transport transfer nodes by data envelop analysis approach using smart card data. *Transportation Research Procedia*, 6, 391-401.
- Novotný, V. (2017). Kvalita provozu veřejné dopravy vyjádřena metodou Level of Service. [online] Dostupné z <http://preferencvhd.info/index.php/2017/10/24/kvalita-provozu-verejne-dopravy-vyjadrena-metodou-level-of-service/>.
- OECD. (2022). Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (OECD/LEGAL/0188).
- Olivo, A., Maternini, G., & Barabino, B. (2019). Empirical study on the accuracy and precision of automatic passenger counting in European bus services. *The Open Transportation Journal*, 13(1).
- Perk, V., Kamp, N., et al. (2003). *Handbook of automated data collection methods for the national transit database* (tech. rep.). National Center for Transit Research (US).
- Porambage, P., Ylianttila, M., Schmitt, C., Kumar, P., Gurtov, A., & Vasilakos, A. V. (2016). The quest for privacy in the internet of things. *IEEE Cloud Computing*, 3(2), 36–45.
- PMDP. (2021). Výroční zpráva PMDP 2021 [Online]. Získáno z <https://www.pmdp.cz/onas/povinne-udaje/vyrocní-zpravy/>.
- PMDP. (2023a). Získáno z <https://www.pmdp.cz>.
- PMDP. (2023b). Kartogramy - listopad 2022.
- Regulation (EU) 2016/679 of the European Parliament and of the Council. (2016). Official Journal of the European Union, L 119/1. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>
- Ritson, M. (2006). Net promoter scores australia 2006. *Melbourne Business School*.
- Rocks, B. (2016). Interval estimation for the “net promoter score”. *The American Statistician*, 70(4), 365-372.
- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163-180.
- Rydning, D. R.-J. G.-J., Reinsel, J., & Gantz, J. (2018). The digitization of the world from edge to core. Framingham: *International Data Corporation*, 16.



- Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences*, 3, 153-160
- Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: The real-world use of big data. *IBM Global Business Services*, 12(2012), 1–20.
- Sharda, R., Delen, D., & Turban, E. (2021). *Analytics, data science, & artificial intelligence: Systems for decision support*. Pearson Education Limited.
- Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2000). *Research methods in psychology*. McGraw-Hill.
- Seaborn, C., Attanucci, J., & Wilson, N. H. (2009). Analyzing multimodal public transport journeys in London with smart card fare payment data. *Transportation research record*, 2121(1), 55-62.
- Turban, E., Sharda, R., & Delen, D. (2010). Decision support and business intelligence systems (required). *Google Scholar*.
- Turner, V., Gantz, J. F., Reinsel, D., & Minton, S. (2014). The digital universe of opportunities: Rich data and the increasing value of the internet of things. *IDC Analyze the Future*, 16, 13–19.
- Tzeng, G. H., Lin, C. W., & Opricovic, S. (2005). Multi-criteria analysis of alternative-fuel buses for public transportation. *Energy policy*, 33(11), 1373-1383.
- Unwin, A. (2020). Why is data visualization important? what is important in data visualization?. *Harvard Data Science Review*, 2(1), 1.
- Van Lierop, D., Badami, M. G., & El-Geneidy, A. M. (2018). What influences satisfaction and loyalty in public transport? a review of the literature. *Transport Reviews*, 38(1), 52–72.
- Vuchic, V. R. (2007). *Urban transit systems and technology*. John Wiley & Sons.
- Wang, D., Coignard, J., Zeng, T., Zhang, C., & Saxena, S. (2016). Quantifying electric vehicle battery degradation from driving vs. vehicle-to-grid services. *Journal of Power Sources*, 332, 193-203.
- Zannat, K. E., & Choudhury, C. F. (2019). Emerging big data sources for public transport planning: A systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science*, 99(4), 601–619.
- Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.
- Zhang, C., Liu, Y., Lu, W., & Xiao, G. (2019). Evaluating passenger satisfaction index based on PLS-SEM model: Evidence from Chinese public transport service. *Transportation Research Part A: Policy and Practice*, 120, 149-164.
- Zhang, T., Ramakrishnan, R., & Livny, M. (1996). Efficient algorithms for mining outliers from large data sets. *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, 427-438
- Zheng, L., Zhu, Y., Xue, B., Liu, M., & Fan, R. (2019). Low-cost gps-aided lidar state estimation and map building. In *2019 IEEE International Conference on Imaging Systems and Techniques (IST)* (pp. 1-6). IEEE.

## **Seznam tabulek**

Tabulka č. 1: Úseky s největším zdržením v osobohodinách.....	65
---	----

## Seznam obrázků

Obrázek č. 1: Růst velikosti digitálního vesmíru .....	10
Obrázek č. 2: Účastníci s mobilním přístupem k internetu v Česku .....	26
Obrázek č. 3: Reprezentace kapacity segmentů v programu Power Point (tramvaje) ....	33
Obrázek č. 4: Vizuální reprezentace kapacity segmentů pro trolejbusy .....	33
Obrázek č. 5: Vizuální reprezentace kapacity segmentů pro autobusy .....	34
Obrázek č. 6: Datová struktura pro nahrání segmentů .....	36
Obrázek č. 7: Datová struktura pro definici linek .....	36
Obrázek č. 8: Vzhled Jupyter Lab, tlačítko na nahrání, spuštění a nahrané soubory .....	36
Obrázek č. 9: Soubor „line.py“ s třídou „Line“ a funkcí „load_lines“ .....	37
Obrázek č. 10: Soubor „segment.py“ s třídou „Segment“ a funkcí „load_segments“ ....	38
Obrázek č. 11: Soubor „map_fnc.py“ s funkcí „create_map“ .....	39
Obrázek č. 12: Mapa s tmavou vrstvou .....	39
Obrázek č. 13: Výpočet statistik .....	40
Obrázek č. 14: Soubor „draw_segment.py“ s funkcí „ant“ .....	40
Obrázek č. 15: Funkce „line_exists“ .....	41
Obrázek č. 16: Funkce „assign_color“ .....	41
Obrázek č. 17: Funkce „assign_weight“ .....	42
Obrázek č. 18: Funkce „calc_delay“ .....	42
Obrázek č. 19: Funkce „draw_ant“ .....	42
Obrázek č. 20: Zobrazení mapy po proběhnutí funkce „ant“ .....	43
Obrázek č. 21: Funkce „poly“ .....	43
Obrázek č. 22: Funkce „update_popup“ .....	44
Obrázek č. 23: Vykreslení čar pomocí funkce „poly“ a zobrazení štítku s informacemi	45
Obrázek č. 24: Funkce „toggle_poly_group“ a vytvoření tlačítka „Rychlost“ .....	45
Obrázek č. 25: Vytvoření filtru podle známky, směru a kapacity .....	46

Obrázek č. 26: Vytvoření filtru podle rychlosti a zdržení .....	46
Obrázek č. 27: Funkce „update_opacity“ .....	47
Obrázek č. 28: Mapa s využitím filtrovací funkce.....	47
Obrázek č. 29: Vygenerovaný „performance_table.csv“ .....	48
Obrázek č. 30: Krabicové grafy linky č. 1.....	50
Obrázek č. 31: Krabicové grafy linky č. 2.....	50
Obrázek č. 32: Krabicové grafy linky č. 4.....	51
Obrázek č. 33: Bodový graf zobrazující kapacitu a zpoždění jednotlivých segmentů ...	51
Obrázek č. 34: Závislost obsazenosti linky č. 1 na směru.....	52
Obrázek č. 35: Závislost obsazenosti linky č. 2 na směru.....	52
Obrázek č. 36: Závislost obsazenosti linky č. 4 na směru.....	53
Obrázek č. 37: Závislost zdržení linky č. 1 na směru .....	54
Obrázek č. 38: Závislost zdržení linky č. 2 na směru .....	54
Obrázek č. 39: Závislost zdržení linky č. 4 na směru.....	55
Obrázek č. 40: Kapacita a zdržení linky č. 1, směr A (Bolevec).....	55
Obrázek č. 41: Tramvajové úseky nedosahující známky 1 .....	58
Obrázek č. 42: Tramvajové úseky se zdržením větším než 30 sekund.....	58
Obrázek č. 43: Tramvajové úseky s vysokou obsazeností a zdržením .....	59
Obrázek č. 44: Bodový graf zdržení a kapacity trolejbusů.....	60
Obrázek č. 45: Trolejbusové úseky se známkou 3.....	62
Obrázek č. 46: Trolejbusové úseky se známkou horší než 1 .....	62
Obrázek č. 47: Trolejbusové úseky s vysokou obsazeností.....	63
Obrázek č. 48: Trolejbusové úseky s velkou obsazeností a zdržením .....	64
Obrázek č. 49: Import dat z vygenerované tabulky „performance_table.csv“ .....	65

## **Seznam příloh**

**Příloha A: Tabulka dalších benefitů hromadné dopravy**

**Příloha B: Soubor „line\_graph.py“ s příslušným kódem a vygenerovanou tabulkou**

**Příloha C: Grafy kapacity a zdržení u tramvajových linek**

**Příloha D: Grafy extrémních hodnot u trolejbusových linek**

**Příloha E: Grafy s korelací obsazenosti u trolejbusových linek**

**Příloha F: Grafy s korelací zdržení a průměrné rychlosti u trolejbusových linek**

**Příloha G: Grafy kapacity a zdržení u trolejbusových linek**

## **Abstrakt**

Tato práce se zabývá využitím dat v městské hromadné dopravě, a to na příkladu tramvajové a trolejbusové dopravy v Plzni. Cílem práce je poskytnout přehled o sběru a využití dat v MHD, analyzovat provozní data a identifikovat možnosti zlepšení dopravního systému a demonstrovat, jak efektivní zpracování a využití dat může přispět k lepšímu pochopení provozu hromadné dopravy. V teoretické části jsou představeny různé zdroje dat, metody sběru a nástroje pro analýzu dat v oblasti dopravy. Praktická část se zaměřuje na konkrétní analýzu městské dopravy v Plzni a vytvoření nástroje, který tuto analýzu umožňuje. Jsou zde hodnoceny parametry jako kapacita, zdržení a rychlost, a to pomocí mnoha detailních grafů a vizualizací. Práce také obsahuje srovnání jednotlivých linek a identifikuje problematické segmenty. Na základě výsledků analýzy jsou navrhovány doporučení pro plánování dopravy.

**APA 7:** Nedvěd, O. (2023). *Využití dat pro řízení městské dopravy* [Bakalářská práce, Západočeská univerzita v Plzni].

**Klíčová slova:** Identifikace úzkých míst v hromadné dopravě, plánování hromadné dopravy, vizualizace dopravní sítě, využití dat v městské dopravě

## **Abstract**

The thesis examines data utilization in urban traffic management, using the example of tram and trolleybus transport in Pilsen. The thesis aims to provide a comprehensive overview of data gathering and application in public transport, analyze operational data and identify opportunities for improving the transport system, and demonstrate how effective data processing and use can contribute to a better understanding of public transport operations. The theoretical part introduces different data sources, collection methods and tools for data analysis in the transport sector. The practical part focuses on a specific analysis of urban traffic in Pilsen and the creation of a tool to facilitate this analysis. Factors such as capacity, delay and speed are evaluated using many detailed graphs and visualizations. The work also includes a comparison of individual lines and identifies challenging segments. Based on the findings, suggestions for transport planning are put forth.

**APA 7:** Nedvěd, O. (2023). *Data usage for urban traffic management* [Bachelor Thesis, University of West Bohemia].

**Key words:** Bottleneck identification in urban traffic, transport network visualisation, urban traffic management, urban traffic planning