

ZÁPADOČESKÁ UNIVERZITA V PLZNI

FAKULTA EKONOMICKÁ

Bakalářská práce

Souvislost doby odbavení s typem vozidla MHD

Relation of bus dwell time with the type of vehicle

Jan Čermák

Plzeň 2024

Čestné prohlášení

Prohlašuji, že jsem bakalářskou práci na téma

„Souvislost doby odbavení s typem vozidla MHD“

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

Plzeň dne

v. r. *Jan Čermák*

Zásady pro vypracování práce

Zpracujte teoretický úvod do zkoumané problematiky.

Představte vybraného provozovatele MHD u uveďte relevantní informace o jím nasazovaných vozidlech.

Na vybrané části sítě MHD za zvolené časové období analyzujte souvislost doby odbavení v zastávce s typem nasazovaných vozidel.

Zformulujte závěr a navrhněte případná doporučení.

Poděkování

Rád bych vyjádřil svou upřímnou vděčnost vedoucímu mé bakalářské práce, Ing Zdeňku Kresovi za jeho skvělé vedení, podporu a odborné rady během celého procesu tvorby této práce. Jeho odborné znalosti a ochota byly pro mě neocenitelné.

Dále bych chtěl poděkovat Ing. Adamu Šťastnému z PMDP, který mi pomohl získat potřebné informace pro vypracování této práce.

Obsah

1. Úvod.....	6
2. Městská hromadná doprava	7
2.1 Faktory ovlivňující dobu odbavení.....	7
2.2 Souvislost doby výměny cestujících s kvalitou služeb veřejné dopravy	9
2.3 Opatření pro zlepšení efektivity nastupování v MHD	10
3. PMDP.....	12
3.1 Historie PMDP.....	12
3.2 Současná situace v PMDP.....	13
4. Big data.....	15
4.1 Výzvy při sběru a analýze big dat.....	16
4.2 Ochrana soukromí.....	16
4.3 Využití big data v MHD	17
5. Testování hypotéz.....	19
5.1 Nulová a alternativní hypotéza	19
5.2 Statistické testy	20
5.3 Postup při testování hypotéz	21
5.4 Omezení při testování hypotéz	22
6. Analýza dob odbavování.....	23
6.1 Analyzovaná autobusová linka a nasazovaná vozidla.....	23
6.2 Systém automatického počítání cestujících....	27
6.3 Sběr a příprava dat	28
6.4 Vyhodnocení dat	34
7. Závěr	43
Seznam použitých zdrojů.....	45
Seznam tabulek.....	49
Seznam obrázků	50
Seznam příloh	51

Abstrakt.....	52
Abstract	53

1 Úvod

Městská hromadná doprava (dále MHD) hraje v moderním městském prostředí významnou roli v uspokojování dopravních potřeb obyvatel a přispívá k udržitelnému rozvoji měst. Různé druhy vozidel, jako jsou autobusy, tramvaje a trolejbusy, mají významný vliv na průběh dopravního provozu a zážitek cestujících. Jedním z faktorů, který významně ovlivňuje zkušenost cestujících je zejména doba cestování (Kresa et al., 2023), jež je mimo jiné tvořena délkou odbavení – tedy dobou, kterou cestující stráví na zastávce při nástupu a výstupu dalších cestujících.

Důležitost rychlosti a efektivity odbavení ve veřejné dopravě není možné přehlížet, neboť má vliv na plynulost provozu, časové potřeby cestujících a celkový komfort dopravy. Dlouhé doby odbavení mohou vést k zahlcení zastávek, zpožděním spojů a zvýšené frustraci cestujících. Téma tak může být relevantním pro města a dopravní operátory.

Cílem této bakalářské práce je zkoumat souvislost mezi délkou doby odbavení a typem konkrétního vozidla MHD. Konkrétně bude sledováno, zdali má počet dveří vozidla vliv na dobu odbavení v zastávce. Pro tento účel budou shromážděna a analyzována data z provozu autobusů SOR NB 12 a SOR NS 12 (lišících se počtem dveří při jinak srovnatelné délce vozidla) na lince 33 systému MHD v Plzni (tj. na nejvytíženější lince na kterou jsou autobusy této kategorie nasazovány). Tato data budou získána díky realizovanému experimentu ve spolupráci s Plzeňskými městskými dopravními podniky. Získaná data budou následně očištěna, analyzována a statisticky vyhodnocována. Po zkoumání dat se předpokládá navržení možných opatření pro poskytování atraktivnějších služeb hromadné dopravy či pro vhodnější využívání big dat pro vyhodnocování provozu MHD.

Tato práce si klade za cíl přispět k lepšímu porozumění významu délky odbavení ve veřejné dopravě a poskytnout základ pro budoucí optimalizace provozu MHD s ohledem na různé typy vozidel. Dále se v obecnější rovině věnuje oblasti problematice využívání big dat pro vyhodnocování provozu MHD.

2 Městská hromadná doprava

Vzhledem k zaměření práce je nejprve vhodné charakterizovat městskou hromadnou dopravu s důrazem na faktory a efekty doby odbavení vozidel MHD. Následně bude v další kapitole podrobněji přiblížena MHD v městě Plzni, a to s ohledem na praktickou část.

Městská hromadná (či též veřejná) doprava sehrává významnou úlohu v každodenním životě obyvatel měst, usnadňující efektivní přesun většího počtu osob mezi různými částmi města. Cílem městské hromadné dopravy je poskytovat pohodlný a cenově dostupný způsob cestování, a to při minimalizaci negativních dopadů na životní prostředí. (Gao & Zhu, 2022)

Každý dopravní prostředek v rámci veřejné dopravy sleduje přesně stanovenou trasu s definovanými zastávkami, kde mohou cestující nastupovat a vystupovat. Tato struktura poskytuje lidem jasný a spolehlivý rámec pro jejich každodenní přepravu. Veřejná doprava hraje nejen roli v odlehčování dopravních zácp a omezení znečištění ovzduší, ale také šetří finanční prostředky obyvatel. (Gao & Zhu, 2022)

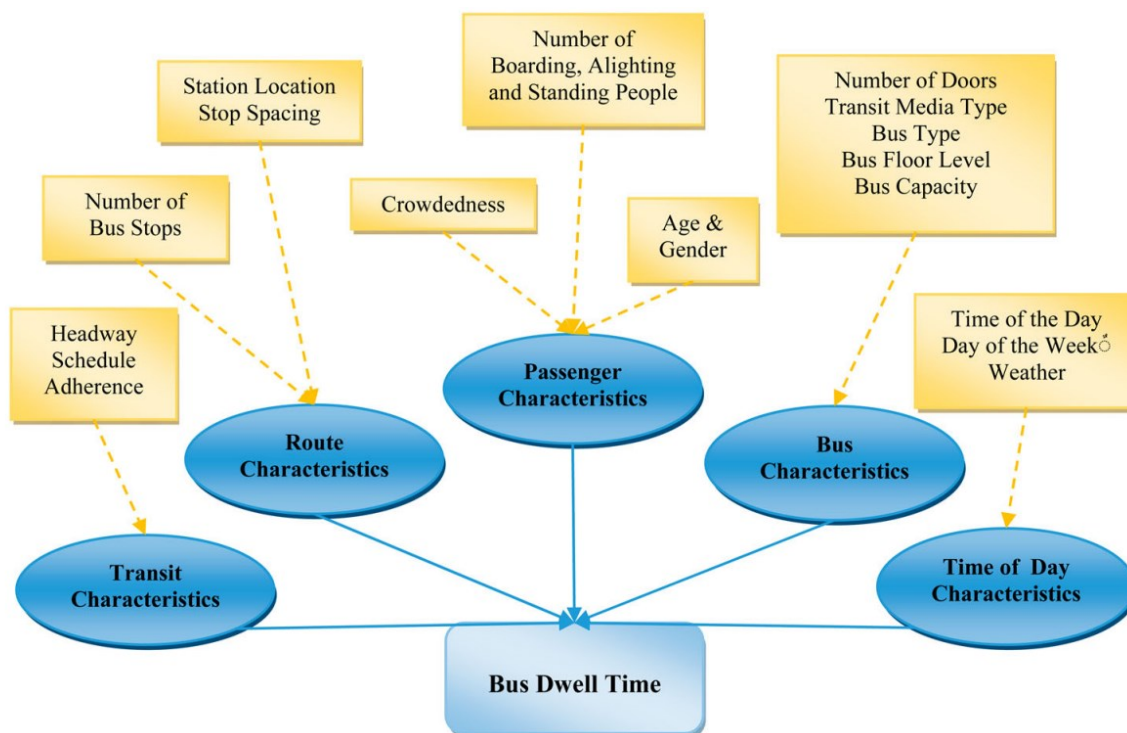
Je důležité zdůraznit, že veřejná doprava má v dnešní době významný vliv na snižování uhlíkové stopy a přispívání k udržitelnějšímu prostředí. S narůstající urbanizací a růstem obyvatelstva ve městech se poptávka po efektivních systémech veřejné dopravy stále zvyšuje. Proto je nezbytné poskytovat moderní, spolehlivé a ekologicky šetrné možnosti dopravy, které plně odpovídají potřebám a očekáváním obyvatel městských oblastí. Například dle opatření Strategického rámce ČR 2030 (2024) je třeba podporovat energeticky efektivní veřejnou hromadnou dopravu.

V městských aglomeracích jsou k dispozici různé druhy veřejné dopravy, od autobusů, trolejbusů a tramvají až po metro.

2.1 Faktory ovlivňující dobu odbavení

Jak už bylo uvedeno, jedním z důležitých faktorů atraktivity MHD je i doba strávená cestováním, a tedy i doba odbavování vozidla v zastávce (či též doba stanicování). Faktory, které dobu odbavování ovlivňují, přiblíží následující podkapitola a obrázek č. 1.

Obrázek č. 1: Faktory ovlivňující dobu nástupu



Zdroj: Rashidi et al., 2022

Jak je z obrázku patrné, těchto faktorů je skutečně celá řada, z nichž jen některé jsou ovlivnitelné dopravcem (např. vlastnosti vozidla, charakteristika linky, či jízdní řád), zatímco jiné jsou zcela mimo jeho moc (počasí, struktura přepravovaných cestujících, intenzita okolní dopravy). Některé z těchto faktorů budou okomentovány podrobněji.

Dynamika procesu nastupování a vystupování cestujících (ovlivňující dobu stanicování) závisí na různých vlastnostech vozidla. Sum et al. (2014) jako hlavní faktory v této souvislosti uvádějí počet a šířku dveří, existence schůdků pro nastupování a vystupování, typ autobusu (jednopodlažní/dvoupodlažní, sólo/kloubový), počet sedadel a místa pro stání (tj. uspořádání interiéru) a způsob výběru jízdného (tj. například cestující musí nastupovat pouze předními dveřmi a odbavovat se u řidiče apod.).

Doba odbavení se dále liší dle množství nastupujících a vystupujících lidí. Zhuang et al. (2022) uvádějí, že v ranních a odpoledních špičkách (7:00–9:00, 16:00–20:00) je doba odbavení delší. Je to dáno jednak větší intenzitou počtu přepravovaných cestujících, tak také větší obsazeností vozidla, což způsobuje horší průchodnost interiéru.

Dalším z faktorů je počasí. Například v zimním období sníh zpomaluje dobu nástupu cestujících obzvláště v období špičky a celkově snižuje spolehlivost autobusové (resp. obecněji povrchové) dopravy. (Zhuang et al., 2022)

Dobu odbavování ovlivňují i cestující, zatímco jejich pohlaví nemá vliv, tak věk cestujících ano. Mladší cestující jsou schopni nastupovat a vystupovat rychleji než starší cestující. (Rashidi et al. 2022)

Úroveň obsazenosti vozidla je dalším kritickým faktorem (jež částečně souvisí i s předchozími), jak se vozidlo blíží ke kritické obsazenosti, může se proces nastupování zpomalit kvůli horší průchodnosti interiéru (tj. omezenému prostoru pro pohyb a usazení cestujících). (Sun et al., 2014) V případě vyšší obsazenosti je dalším faktorem chování cestujících, jelikož pasažéři, kteří cestují delší vzdálenost pravděpodobněji budou setrvávat v interiéru dále od dveří vozidla, než cestující jedoucí – ti naopak mohou setrvávat u prostoru dveří, což může zhoršovat průchod pro další cestující a prodlužovat tak dobu odbavení (Milkovits, 2008). V tuzemsku je v tomto ohledu zajímavý model Solaris Urbino 15 (příp. odvozený Škoda 28 Tr), který se při délce 15 metrů vyznačuje poměrně dlouhým prostorem pro sezení s úzkou uličkou, mezi druhými a třetími dveřmi. Toto uspořádání interiéru (řady sedadel s úzkou uličkou) je typické spíše pro příměstské či dálkové autobusy. Obecně problémy s chováním cestujících řeší také nálepky s výzvou „postupujte dále do vozidla“, které jsou typické zejména v provozech, kde dochází k nástupu do vozidla MHD jen předními dveřmi.

2.2 Souvislost doby výměny cestujících s kvalitou služeb veřejné dopravy

Dopady času stanicování na kvalitu veřejné dopravy nejsou omezeny pouze na provozní efektivitu (např. počet vozidel nutných na obsluhu linku), ale přímo ovlivňují i to, jak cestující vnímají spolehlivost MHD (plnění jízdního řádu) což je dáno zejména tím, jak dlouho čekají na svůj spoj.

Ti cestující, kteří nastupují na spoj blíže jeho výchozí stanici, mají větší šanci na včasný příjezd spoje. To je způsobeno tím, že v průběhu trasy může docházet vlivem výše uvedených faktorů k větším odchylkám příjezdu spoje od stanoveného jízdního řádu. (Zhuang et al., 2022) Dopravci se tento vliv snaží minimalizovat např. různou délkou cestovních časů (tzv. chronometráže) pro jednotlivé části dne a týdne (Drdla, 2021).

Důležitým aspektem je také doba čekání cestujících na zastávce, která je přímo ovlivněna dobou odbavování. Pokud se tato doba prodlužuje kvůli vysoké obsazenosti nebo neefektivním procesům výměny cestujících, může to způsobit frustraci a vnímání nižší kvality poskytovaných služeb. (Zhuang et al., 2022) Jedním ze zdržujících nástupních procesů může být organizace nástupu pouze předními dveřmi (s odbavením u řidiče), který využívají v městské dopravě celodenně především menší města, jako Mariánské Lázně, Opava, nebo Cheb (ve větších městech jako např. Pardubice je pak využíván jen v nočních hodinách). Jako zajímavost lze uvést, že celodenní nástup předními dveřmi je nově zaveden i na většině zastávek v rámci relativně velkého města Ústí nad Labem, a to z důvodu bezpečnosti (DPMÚL, 2023). Snaha o optimalizaci dob výměny cestujících je tedy významná nejen pro sledování provozních ukazatelů, ale i pro celkové zlepšení kvality veřejné dopravy a zvyšování spokojenosti jejich uživatelů. Je nezbytné zaměřit se na procesy nástupu nejen z hlediska efektivity provozu, ale také jako důležitý prvek ovlivňující pohodlí a spokojenost cestujících s veřejnou dopravou. Kuriozním příkladem jsou takzvaní „tlačiči“ (pushers) cestující. Ty lze nalézt například v nejméně frekventovaných stanicích japonských železničních systémů, kde v době přepravních špiček tlačí nastupující cestující do spoje, čímž se snaží urychlit proces nástupu cestujících (a snížit tím případné zpoždění vozidla), ale výsledný dojem a pohodlí cestujících je sniženo (Baseel, 2017).

2.3 Opatření pro zlepšení efektivity nastupování v MHD

Zefektivnění procesu výměny cestujících v městské hromadné dopravě mohou pomoci opatření v oblasti infrastruktury. Jednou z možností je implementace samostatných nástupišť pro vstup a výstup cestujících, což má doložený vliv na výrazné zlepšení plynulosti pohybu pasažérů. Tato opatření oddělují proud cestujících a minimalizují interakce, které by mohly vést ke kolizím či přetížení, a tím zvyšují celkovou efektivitu. Je třeba však vzít v úvahu, že výstavba samostatných nástupišť může být často finanční překážkou, zejména v kontextu omezených rozpočtů městských dopravních systémů. Tyto nástupišťe jsou obvykle rezervována pro nejméně frekventovaná místa, kde objem cestujících opravňuje finanční investice. (Harris et al., 2022) Jako vhodný příklad lze uvést pražský autobusový terminál Nádraží Veveřská, kde je výstupní zastávka autobusů hned u vchodu do metra, a cestující z metra zase vychází hned k nástupní zastávce linek a na povrchu se tak toky cestujících nekříží. Celkově design zastávek

a terminálů má vliv nejen na dobu odbavení, ale i na celkový tok ostatní individuální dopravy (Rowe, 2022).

Kromě toho hraje významnou roli v efektivitě procesu výměny cestujících i fyzický design vozidel. Rozšíření šířky uliček uvnitř vozidel může poskytnout více prostoru pro pohyb cestujících a bylo identifikováno jako faktor, který může potenciálně zlepšit efektivitu jak při výstupu, tak při nastupování. Vytvoření většího prostoru v salonech a za dveřmi vozidel může též vést k zrychlení pohybu cestujících. (Harris et al., 2022) Je proto nutné pečlivě zvážit úpravy platformy a konstrukce vozidel s ohledem na náklady a účinnost.

3 PMDP

Pro účely této práce byla navázána spolupráce s jedním z největších českých provozovatelů MHD, a sice společností Plzeňské městské dopravní podniky, a.s., která zajišťuje městskou hromadnou dopravu na území města Plzeň. Zdejší systém veřejné dopravy je tvořen třemi trakcemi, a to tramvaje, trolejbusy a autobusy. Páteří hromadné dopravy jsou tři tramvajové linky, které spojují největší předměstí s centrem města, vlakovým a autobusovým nádražím. Tramvaje jsou doplněny trolejbusy, které s výjimkou Severního Předměstí spojují ostatní velká předměstí. Autobusové linky poté zajišťují propojení mezi předměstími (tangenciální linky) a zajišťují dopravu do menších a okrajových městských oblastí. (PMDP, 2024)

Vozidla každé z tří trakcí jsou v Plzni odlišena barevně a ve spojení s bílou barvou vytvářejí charakteristický čtyřlístek plzeňských městských barev: žlutou pro tramvaje, zelenou pro trolejbusy a červenou pro autobusy. Veřejná doprava v Plzni patří k nejekologičtějším městským dopravním systémům v Evropě, přičemž dvě třetiny dopravních výkonů jsou zajišťovány právě tramvajemi a trolejbusy, které minimalizují zátěž pro životní prostředí. Díky viditelnosti kolejové a trolejové sítě může každý návštěvník města snadno sledovat hlavní trasy linek a bez problémů se dostat k nejbližší zastávce. (PMDP, 2024)

3.1 Historie PMDP

Plzeňské městské dopravní podniky, a.s., jsou firmou s bohatou historií, sahající až do 19. století. Klíčovou postavou při vzniku této dopravní společnosti byl významný český elektrotechnik, doktor František Křížík. V letech 1896 až 1899 zrealizoval projekt vytvoření elektrické dráhy a přidružených elektráren pro pohon, a to na základě koncese udělené městem Plzní a schválené tehdejšími císařským a královským ministerstvem železnic. (PMDP, 2024)

Původně nazvané "Elektrické dráhy v Plzni a okolí," společnost měla tři klíčové cíle: čistě elektrickou tramvajovou dráhu, výhradní orientaci na provoz elektrické dráhy a rozšíření elektrické dráhy za hranice města. Přestože původní nápady zahrnovaly stavbu nákladní tramvaje až do Štěnovic, rapidní rozvoj elektrotechniky na přelomu 19. a 20. století změnil směr. V roce 1902 byla potřeba zabezpečit dodávky elektřiny pro nové městské divadlo, což vedlo k exponenciálnímu růstu poptávky po elektrické energii. Elektrárna v Cukrovarské ulici a vodní elektrárna v Panském mlýně byly později doplněny o moderní vodní elektrárnu na Denisově nábřeží, která slouží dodnes. (PMDP, 2024)

Ve 30. letech 20. století, po změně poptávky a rozvoji trolejbusové dopravy v roce 1941, se společnost přejmenovala na „Elektrické podniky města Plzně“. Válečná a poválečná léta byla spojena se zavedením a rozvojem trolejbusové dopravy v roce 1941. V roce 1946 byla od elektrárenské činnosti oddělena dopravní činnost. Společnost, nyní nazývaná „Dopravní podniky statutárního města Plzně“, se zaměřovala hlavně na městskou veřejnou dopravu v Plzni. V dobách normalizace se charakter společnosti změnil na státní podnik, ale od 1. května 1998 se vrátila pod plnou kontrolu města Plzně a získala název „Plzeňské městské dopravní podniky, a.s.“ (PMDP, 2024)

Hlavním posláním společnosti zůstává provoz městské veřejné dopravy, avšak provázela ji i řada dalších obchodních aktivit, včetně servisu a mytí vozidel, provozu autoškoly a výroby elektrické energie v historické elektrárně „Hydro“ na Denisově nábřeží, kde sídlí společnost dodnes. (PMDP, 2024)

3.2 Současná situace v PMDP

K roku 2024, ve kterém si Plzeňské městské dopravní podniky připomínají již 125 let od svého založení, ročně podniky přepraví více než 110 milionů cestujících. Vozový park je pravidelně obnovován moderními vozidly s klimatizací, přičemž 100 % autobusů a trolejbusů je nízkopodlažních, a u tramvají dosahuje nízkopodlažnost dvou třetin a její podíl nadále roste. (PMDP, 2024)

Kvalita městské dopravy v Plzni je nezpochybnitelná a v mnoha technologiích jsou PMDP průkopníky v celé České republice. Například společnost jako první město umožnila cestujícím zakoupit si jízdenku všemi vozy městské veřejné dopravy pomocí bezkontaktní bankovní karty. (Fintech, 2015) Další inovativní možností odbavení je mobilní aplikace Virtuální karta, která umožňuje nahrát předplatný kupón do mobilního telefonu nebo na bankovní kartu.

PMDP se obdobně jako v minulosti nezaměřuje pouze na provoz městské hromadné dopravy, ale dále také zajišťuje provoz městského parkovacího systému, veřejného osvětlení a světelné signalizace. S cílem zlepšit dopravu ve městě a rozšířit dopravní služby pro obyvatele byla v roce 2018 zahájena služba sdílených aut, carsharing Karkulka. Kromě toho PMDP od roku 2014 poskytuje službu Senior Expres, určenou pro občany starší 70 let nebo pro držitele průkazu ZTP, což představuje další krok k podpoře mobility a dostupnosti dopravních služeb pro všechny obyvatele města.

PMDP patří v rámci České republiky k těm nejaktivnějším i z pohledu sběru dat o provozu MHD. Tato data lze obecně hodnotit jako tzv. big data, o kterých bude z teoretického pohledu více pojednávat následující kapitola.

4 Big data

„Big data jsou objemná sada strukturovaných, nestrukturovaných a polostrukturovaných datových sad, jejichž správa pomocí tradičních nástrojů pro zpracování dat je náročná. Vyžaduje další infrastrukturu pro řízení, analýzu a přeměnu na statistiky.“ (BasuMallick, 2022)

Objem, rozmanitost a rychlost jsou často označovány jako tři klíčové prvky big data, známé též jako „tři V“ z anglických pojmů volume, velocity a variety.

- **Objem:** Jak již název napovídá, nejčastější charakteristikou velkých dat je jejich vysoký objem. To popisuje obrovské množství dat, která jsou k dispozici pro sběr a průběžně vytvářena z různých zdrojů a zařízení. (Google, 2024)
- **Rychlost:** Dnes jsou data vytvářena v reálném čase, a proto musí být také zpracovávána, zpřístupňována a analyzována stejnou rychlostí, aby měla nějaký smysluplný dopad. (Google, 2024)
- **Rozmanitost:** Označuje různé typy dat a vzhledem k tomu, že má dopad na výkon, je to jeden z hlavních problémů, se kterými se sektor big dat nyní potýká. (BasuMallick, 2022)
Je důležité uspořádat svá data, abyste mohli efektivně spravovat jejich rozmanitost.

Tyto charakteristiky byly poprvé identifikovány v roce 2001 Dougem Laneyem, analytikem v poradenské firmě Meta Group Inc. a později společnost Gartner tuto definici rozšířila. Kromě tří V byly do některých popisů big dat přidány další V, včetně pravdivosti (veracity), hodnoty (value) a variability. (Botelho & Bigelow, 2022) Tato data často pocházejí z procesu dolování dat a přicházejí v různých formátech, od strukturovaných, polostrukturovaných až po nestrukturovaná data. Každá z forem dat vyžaduje určitá specifika.

- **Strukturovaná data:** Tento typ dat je přesně definován a organizován, jak název naznačuje. Obsahuje jasnou strukturu, kterou lze snadno interpretovat jak počítačem, tak člověkem. Informace jsou systematicky uspořádány a umožňují rychlé a jednoduché uložení do databází a snadné zpřístupnění pomocí jednoduchých metod. Díky předem známému formátu dat je správa tohoto typu dat relativně jednoduchá. Příkladem strukturovaných dat jsou informace uchovávané v podnikových databázích, jako jsou tabulky a relační databáze.
- **Polostrukturovaná data:** Kombinují prvky strukturovaných a nestrukturovaných dat. Obsahují informace, které nejsou plně integrovány do konkrétní databáze, ale stále

obsahují klíčové značky, které odlišují různé části v rámci téže databáze. I když nejsou zcela systematická, mají určitou formu organizace. Na první pohled mohou působit jako nestrukturovaná, ale mají v sobě nějaký řád. Příkladem jsou například NoSQL texty, které mohou být zpracovány pomocí klíčových slov, nebo soubory CSV, jež jsou považovány za polostrukturovaná data. (BasuMallick, 2022)

- Nestrukturovaná data: Jsou to informace bez jakékoliv předem definované struktury. Jejich rozsah a různorodost jsou výrazně širší než u strukturovaných dat. Tato data jsou chaotická a náročná na zpracování, porozumění a interpretaci, jelikož jim chybí jasná organizace. Nejsou omezena pevnou strukturou a mohou se měnit v průběhu času. Většina velkých dat spadá do této kategorie, zahrnující komentáře na sociálních sítích, tweety, sdílené příspěvky, videa na YouTube nebo textové zprávy WhatsApp. (BasuMallick, 2022)

4.1 Výzvy při sběru a analýze big dat

Sběr big dat představuje zásadní krok v procesu získávání a využívání dat pro informační analýzy a rozhodovací procesy. Tento proces zahrnuje shromažďování rozmanitých dat z různých zdrojů, jako jsou webové stránky, senzory, transakce, sociální média a podnikové databáze. Po shromáždění těchto dat následuje analýza, která spočívá v důkladném zkoumání a interpretaci informací za účelem odhalení trendů, vzorců a cenných poznatků.

Během těchto procesů však nastává mnoho výzev, kterým je třeba čelit, jako nedostatek talentu a dovedností v oblasti dat. Datoví analytici a datoví inženýři jsou nedostatečně dostupní a patří mezi nejhledanější (a nejlépe placené) profesionály v IT průmyslu. (Google, 2024) Objem dat je též jeden z problémů a společnosti musí určit, která data představují signály ve srovnání se šumem (Segal, 2022). Klíčovým faktorem se stává rozhodnutí, proč jsou data relevantní. Mezi další problémy může patřit rychlost růstu dat, problémy s kvalitou, ochrana soukromí a bezpečnost (Google, 2024).

4.2 Ochrana soukromí

Ochrana soukromí v dnešním světě nabývá na důležitosti. Vzhledem k množství osobních údajů, které jsou dnes o jednotlivcích k dispozici, je zásadní, aby společnosti podnikly kroky k ochraně těchto údajů. (Segal, 2022)

Proto Evropská unie (dále EU) v květnu roku 2018 schválila Obecné nařízení o ochraně osobních údajů (dále GDPR) k řešení problémů s ochranou osobních údajů, které musí podniky dodržovat. GDPR obsahuje například ustanovení o právu být zapomenut, které umožňuje obyvatelům EU požádat o vymazání svých údajů, a omezuje typy údajů, které mohou organizace shromažďovat. Organizace musí nyní získat souhlas od jednotlivců, nebo splnit konkrétní kritéria pro shromažďování osobních údajů. (Botelho & Bigelow, 2022)

4.3 Využití big data v MHD

Vznik big dat otevřel nové cesty pro plánování a řízení městské dopravy. Tato masivní a různorodá data, rychle generovaná a přenášená z mnoha zdrojů, umožňují monitorování a analýzu dopravních vzorů v reálném čase, což vede k lepšímu prediktivnímu modelování a rozhodovacím procesům v sektoru veřejné dopravy (McAfee & Brynjolfsson, 2012). Využití datové analýzy má potenciál způsobit revoluci v systému veřejné dopravy, což povede ke zvýšení efektivity, snížení nákladů a celkově lepší zkušenosti cestujících. (Utilities One, 2023)

Díky analýze big dat je možné optimalizovat trasy a jízdní řády tak, aby vyhovovaly vyvíjejícím se potřebám cestujících (Schroer, 2024). Analýza big dat dále přispívá ke snížení dopravních zácp, efektivnějšímu využití zdrojů dopravních podniků a lepšímu zážitku cestujících. Kromě toho mohou být cestujícím poskytovány informace v reálném čase díky rychlosti zpracování dat, což zlepšuje spolehlivost a pohodlí veřejné dopravy. (Utilities One, 2023)

Analýza big dat může navíc pomoci zlepšit efektivitu stávající infrastruktury v systémech městské dopravy. Použití algoritmů strojového učení ve spojení s big daty umožňuje plánovačům identifikovat trendy a vzorce, které dříve nebylo možné odhalit, což může vést k přesnějšímu rozhodování v městské hromadné dopravě. (Utilities One, 2023)

Tradiční metody sběru dat zahrnují umístění pozorovatelů na zastávkách k měření času obsluhy cestujících a sledování jízd vozidel. Sběr dat probíhá podle určených kroků, včetně zaznamenání času zastavení, otevření a zavření dveří, počtu nastupujících a vystupujících cestujících a dalších speciálních okolností. Díky sběru big dat pomocí technologií jako APC (automatické počítání cestujících) je možné sbírat větší množství dat a následně danou linku lépe analyzovat. Jedinou nevýhodou je, že často z měření nelze vynechat neobvyklé situace, jako například přeprava vozíčkářů. Proto tyto nové postupy vyžadují důkladnou úpravu při sběru a interpretaci dat. (Dueker et al., 2004)

Obecně soubory big dat umožňují získávat poměrně komplexní představu o sledované problematice. V závislosti na charakteru sledovaného jevu i počtu dostupných dat může být i pro menší vzorky big dat aktuální přístup testování hypotéz. V rámci snahy o triangulaci výzkumu realizovaného v praktické části práce bude kromě popisné statistiky právě i testování hypotéz využito. Tento tradiční přístup bude podrobněji popsán v následující kapitole.

5 Testování hypotéz

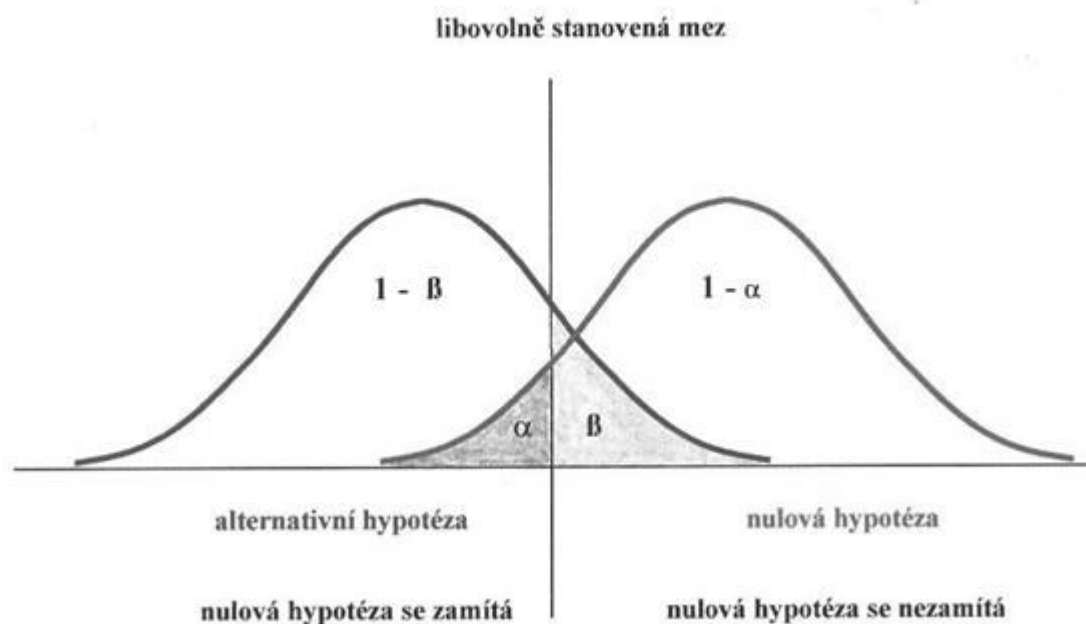
Testování hypotéz je formální postup, který pomáhá zkoumat představy o světě za pomoci statistiky. Často ho využívají vědci k ověření konkrétních „předpovědí“, formulovaných jako hypotézy, jež vyplynuly z teoretických rámců. Tento postup umožňuje systematicky analyzovat data a rozhodovat na základě statistických důkazů o tom, zda jsou určité představy podloženy relevantními a významnými daty. (Bevans, 2023) Testování hypotéz také poskytuje rámec pro rozhodování založené na datech místo osobních názorů nebo zkreslení. S využitím statistické analýzy pomáhá snižovat vliv náhody a rušivých proměnných, poskytující solidní základ pro fundované závěry. (Majaski, 2023) Obecně je tento přístup založen na využívání data z výběru k odvozování závěrů o parametru populace nebo pravděpodobnostní distribuci populace.

5.1 Nulová a alternativní hypotéza

Nulová a alternativní hypotéza jsou nezbytností při každém testování. Na základě předběžných předpokladů o parametru nebo distribuci se vytvoří nulová hypotéza H_0 a alternativní hypotéza H_a je opakem nulové. Obecně H_a představuje počáteční předpoklad, který předpovídá vztah mezi proměnnými a H_0 je předpověď, že mezi zkoumanými proměnnými neexistuje žádný vztah (Hendl, 2015). Postup testování hypotéz zahrnuje použití dat z výběru k určení, zda lze nebo nelze zamítnout H_0 . Pokud je H_0 zamítnuta, statistický závěr je, že alternativní hypotéza H_a je nezamítnuta, ale nikoliv však potvrzena. (Anderson et. al., 2024)

Ideálně by měl postup testování hypotéz vést k nezamítnutí H_0 , když H_0 platí, a k zamítnutí H_0 , když H_0 neplatí. Bohužel, protože testy hypotéz jsou založeny na datech z výběru, musíme vzít v úvahu možnost chyb. Chyba typu I odpovídá zamítnutí H_0 , když H_0 je skutečně pravdivá, a chyba typu II odpovídá přijetí H_0 , když H_0 je nepravdivá. Pravděpodobnost chyby typu I je označována jako α , též označována hladina významnosti, a pravděpodobnost chyby typu II je označována jako β , která nám po odečtení od jedné udává takzvaně sílu testu. (Anderson et. al., 2024; Bedáňová & Linhart, 2024) Tyto chyby lze vidět na obrázku č. 2, který zobrazuje, že čím menší je α , tak tím větší je β a naopak.

Obrázek č. 2: Nulová a alternativní hypotéza vůči chybám



Zdroj: Bedáňová & Linhart, 2024

5.2 Statistické testy

Statistické testy pomáhají výzkumníkům rozhodnout, zda nulovou hypotézu zamítne, či nikoliv a na základě toho vyvodit z dat závěry. K dispozici je celá řada statistických testů, ale všechny jsou založeny na srovnání rozptylu v rámci skupiny (jak jsou data rozprostřena v rámci kategorie) a rozptylu mezi skupinami (jak se jednotlivé kategorie navzájem liší). (Bevans, 2023) Lze tyto testy rozdělit do dvou kategorií, a to na parametrické a neparametrické. Dále lze testy rozdělit na jednovýběrové a dvouvýběrové, kdy se testuje buď jeden soubor dat, či se porovnávají dva nezávislé.

Parametrické testy jako předpoklad vyžadují konkrétní rozdělení, obvykle normální. Tyto testy jsou obecně přesnější než neparametrické, ale jsou náchylnější k chybám vzhledem ke kladeným předpokladům. Příkladem parametrických testů může být například jednovýběrový t-test, test o rozptylu normálního rozdělení a párový t-test. Párový t-test je aplikován v případech, kdy máme dvě měření na každém z n objektů. Ačkoliv jsou samotné objekty obvykle považovány za nezávislé, samotná měření na stejném objektu nejsou nezávislá. Párový

t-test je vhodný například v situaci, kdy je testována účinnost léku na n pacientech. Zde jsou hodnoty naměřené po podání léku. (Mrkvička & Petrášková, 2006; Pavlík, 2024)

Neparametrické testy bývají z pohledu výzkumníka pro určení, zda test provést či ne, jednodušší, jelikož nemají žádné předpoklady rozdělení souboru dat a jsou tedy robustnější vůči odchýlkám od normálního rozdělení. Daní za to je jejich obecně menší citlivost. To znamená, že mohou být méně schopné detekovat rozdíly nebo vztahy ve datech ve srovnání s parametrickými testy, pokud jsou data normální. Vhodnými příklady jsou například Mann-Whitneyho test, Kruskal-Wallisův test a Wilcoxonův test, což je opět párový test. (Mrkvička & Petrášková, 2006)

5.3 Postup při testování hypotéz

Postup začíná formulací výzkumné otázky, kterou chce výzkumník zkoumat, a odvozením nulové a alternativní statistické hypotézy (Hendl, 2015).

Následuje shromažďování dat, což je důležitým krokem, a aby byl statistický test platný, je zásadní provádět vzorkování a shromažďovat data způsobem, který je určen k testování naší hypotézy (Bevans, 2023). Nelze dělat statistické závěry o populaci, která nás zajímá, pokud data nejsou reprezentativní (Biswal, 2023).

V dalším kroku se provede statistický test, při němž se, jak bylo zmíněno výše, porovnává rozptyl. Když je rozptyl mezi skupinami dostatečně výrazný a dochází k minimálnímu nebo žádnému překrývání mezi nimi, statistický test ukáže nízkou p -hodnotu, což je nejmenší hladina významnosti, při které je ještě zamítána nulová hypotéza (Pavlík, 2024). To naznačuje, že je malá pravděpodobnost, že by rozdíly mezi skupinami vznikly náhodou. Naopak, pokud existuje vysoký rozptyl v rámci skupiny a nízký rozptyl mezi skupinami, statistický test reflektuje tuto situaci s vysokou hodnotou p . To indikuje, že jakýkoli rozdíl, který pozorujeme mezi skupinami, je pravděpodobně způsoben náhodou. Výběr konkrétního statistického testu závisí na typech proměnných a úrovni měření dat.

Výsledky statistického testu musí určit, zda má být nulová hypotéza zamítnuta, nebo ne. Ve většině případů se úsudek zakládá na p -hodnotě poskytnuté statistickým testem. Většinou bývá přednastavená hladina významnosti pro zamítnutí nulové hypotézy 0,05 – to znamená, když existuje méně než 5% pravděpodobnost, že by byla nulová hypotéza pravdivá. V jiných oborech, typicky v lékařství výzkumníci používají nižší hladinu významnosti, například 0,01 (1 %). To

snižuje možnost nesprávného zamítnutí nulové hypotézy. (Biswal, 2023) Druhým postupem vyhodnocení statistického testu je porovnání testovacího kritéria, které je testem vypočteno s kritickou hodnotou, která svědčí ve prospěch alternativní hypotézy a určuje se v závislosti na zvolené hladině významnosti. „Jestliže hodnota vypočtené testovací statistiky překročí kritickou hodnotu, znamená to, že existuje evidence pro zamítnutí nulové hypotézy (tzn. „že byl potvrzen rozdíl“). Naopak, pokud se vypočtená testovací statistika ocitne uvnitř oboru přijetí H_0 , nelze zamítnout nulovou hypotézu, a předpokladem tedy je, že platí.“ (Bedáňová & Linhart, 2024)

5.4 Omezení při testování hypotéz

Omezení při testování hypotéz spočívá například v tom, že přesnost výsledků závisí na kvalitě dostupných dat a použitých statistických metod. Nesprávná data nebo nevhodné formování hypotéz mohou vést k chybným závěrům nebo neúspěšným testům. Důležité je také respektovat určení jednotlivých testů a jejich základní předpoklady. Testování hypotéz může také vést k chybám, jako je přijetí nebo zamítnutí nulové hypotézy, když by nemělo dojít k žádnému z těchto rozhodnutí. (Majaski, 2023) Tyto chyby mohou mít za následek falešné závěry nebo přehlédnutí významných vzorů či vztahů v datech. Dále je třeba zamyslet se, zda je výsledek prakticky důležitý, a ne pouze statisticky. Taková situace se nazývá takzvaně „efekt velikosti“. Též kontext zkoumání je důležitý, jelikož některá odvětví potřebují větší vzorky k nalezení významných rozdílů, zatímco jiná ne. (24X7Editing.com, 2023) Omezení je mnoho a pro výzkum je nezbytné brát je v úvahu. Omezením přístupu testování hypotéz (či obecně kvantitativního designu výzkumu) může být i to, že se spoléhá výhradně na data, což nemusí zajišťovat komplexní pochopení zkoumaného tématu.

6 Analýza dob odbavování

Po představení teoretických východisek je možné začít zkoumat big data z provozu MHD prakticky. Pro tyto účel byla navázána spolupráce s Plzeňskými městskými dopravními podniky, které už podrobněji představila kapitola č. 3. S ohledem na zaměření práce na analýzu souvislosti doby odbavení s typem vozidla na vybrané části sítě MHD za zvolené časové období. Bude v této kapitole nejprve přiblížena zkoumaná autobusová linka a na ní nasazovaná vozidla. Následně bude specifikován systém sběru dat, následná práce s daty (očišťování), a dále pak prováděné statistické postupy a testy, včetně následného okomentování závěrů.

6.1 Analyzovaná autobusová linka a nasazovaná vozidla

V rámci této práce bude zkoumána a detailně analyzována doba odbavování vozidel na autobusové lince číslo 33, která spojuje centrum města s největšími sídlištními celky na Severním Předměstí stejně jako s nemocničním areálem na Lochotíně. Tato linka hraje významnou roli v propojení centrálního města s jeho severní částí a představuje u PMDP nejfrekventovanější linku zajišťovanou dvanáctimetrovými (nekloubovými) autobusy. Z převážné většiny se jedná o autobusy výrobce SOR, jejichž specifikace přibližuje tabulka č. 1 níže. Autobusy na této lince operují celotýdenně v časovém rozmezí od 4:30 do 23:00 a zastavují na následujících zastávkách:

Goethova,	Úřad Lochotín,
Náměstí Republiky,	U Družby,
Otýlie Beníškové,	Sokolovská,
Rondel,	Severka,
Lékařská fakulta,	Žlutická,
K Nemocnici,	Krašovská,
Nemocnice Lochotín,	Sídliště Košutka

Pro linku 33 je významná ranní špička od 5 do 7 hodin, kdy jsou odjezdy ze zastávek v intervalech 6-7 minut (návozy k lochotínské nemocnici). Minima dosahují časy mezi spoji v odpolední špičce, po 14. hodině, kdy je nastaven interval od lochotínské nemocnice do centra 5 minut (činěno je tak i zkrácenými spoji). V sedle mezi špičkami i následně se intervaly postupně prodlužují až do večerních hodin, kdy jezdí každých půl hodiny. Linka 33 je v období

přepravních špiček v prokladu s linkou číslo 40, která představuje rychlejší variantu (bez zavlaku do zastávek K Nemocnici a Nemocnice Lochotín). Linka 40 je v provozu pouze v pracovní dny v časech cca 4:00-5:30 a 14:30-18:00.

Tabulka č. 1: Obecná specifikace vozů SOR

PMDP	SOR NB 12 (čtyřdveřový)	SOR NS 12 (třídveřový)
Uvádění do provozu	Od 2010	Od 2019
Šířka dveří	120 cm (druhé až čtvrté) 86 cm (první)	120 cm
Počet sedících osob	27	27
Celková obsaditelnost	102 osob	108 osob
Komfortní obsaditelnost	56 osob	56 osob
Nástupní výška	32,5 cm	34 cm

Zdroj: Vlastní zpracování, 2024 (SOR, 2017; SOR, 2018; Interní materiály PMDP)

Oba typy vozidel jsou koncipovány jako nízkopodlažní s velmi podobnými technickými specifikacemi. Hlavní rozdíl spočívá v jiném počtu dveří, jež mají cestující k dispozici (bližší specifikuje obrázek č. 3 a 4). Celková obsaditelnost u obou vozidel je srovnatelná, nicméně při stanovování maximální obsaditelnosti velmi záleží na zvolené hustotě osob na jednotku plochy vozidla. Město Plzeň jako objednavatel MHD si proto dále stanovuje tzv. komfortní obsaditelnost, která je nižší než výrobcem stanovená maximální kapacita (kalkuluje se nižší hustota obsazení). Objednavatel a dopravce při určování tohoto ukazatele vychází empiricky z pocitové přeplněnosti spoje vnímané cestujícími. Tato obsaditelnost byla u vozidel stanovena zhruba jako poloviční vůči kapacitě udávané výrobcem.

Obrázek č. 3: Čtyřdveřový autobus SOR NB 12



Zdroj: Archiv Zdeněk Kresa, 2024

Obrázek č. 4: Třídveřový autobus SOR NS 12



Zdroj: Archiv Zdeněk Kresa, 2024

Od února 2024 PMDP provozují další nové autobusy SOR NS 12 vyznačující se mírně odlišným uspořádáním interiéru vyplývajícím ze skutečnosti, že druhé dveře jsou provedeny jako vně předsvunuté (pro zajištění lepší průchodnosti vozidlem). V rámci této práce jsou uvažovány dosud převažující starší provedení SOR NB 12 a NS 12 vyznačující se všemi dveřmi otočnými s otevíráním dovnitř. (PMDP, 2024) Plány interiérů blíže zobrazí obrázky č. 5 a 6 níže. Plánky zachycují základní nabízené provedení vozidel, které se v případě PMDP drobně odlišuje zejména v prostoru pro vozíčkáře (u plzeňských vozů NS 12 absentují první dvě sedačky ve směru jízdy u levé bočnice, u vozů NB 12 jsou v levé bočnici tři sklopné sedačky namísto zobrazených dvou), jinak jsou všechny dispozice vozidla korektní.

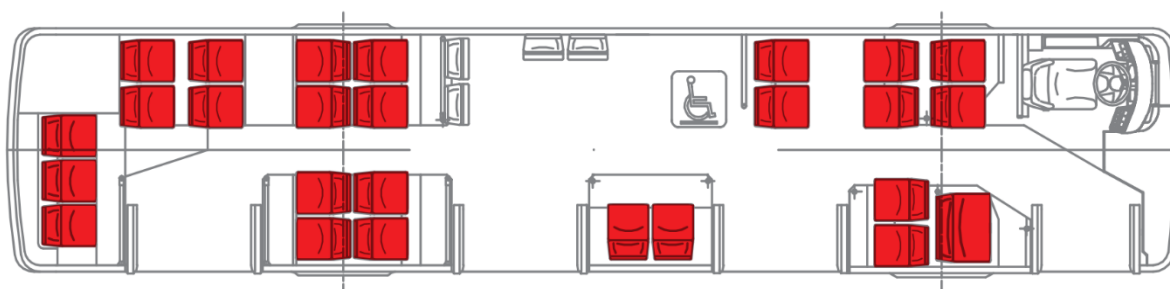
Obrázek č. 5: Plán interiéru vozu SOR NS 12



Zdroj: SOR, 2018

Obrázek č. 6: Plán interiéru vozu SOR NB 12

NB 12 (4dveřový)



Zdroj: SOR, 2017

Ve výbavě části vozidel obou typů představených autobusů nechybí ani systémy automatického počítání cestujících, které jsou stěžejním zdrojem dat pro tuto práci.


6.2 Systém automatického počítání cestujících

Systémy automatického počítání cestujících (dále APC) se v PMDP využívá od roku 2016 na všech třech trakcích a v současnosti je jím vybaveno 133 vozidel. Přesněji 43 tramvají (39% celého provozního stavu), 54 trolejbusů (51%) a 36 autobusů (32%). Tyto systémy lze rozdělit na dvě hlavní části, a to vozidlovou a aplikační.

Vozidlovou část tvoří senzory nad každými dveřmi. Postupným vývojem se ve vozidlech PMDP lze setkat se třemi typy senzorů. Nejrozšířenější a zároveň nejstarší jsou infrsenzory od společnosti DILAX. Jsou méně přesné a fungují na principu přerušování vysílaného infrapaprsku cestujícím. Dále jsou využívány stereokamery od téže firmy, a to pouze v tramvajích typu EVO2. Nejpresnější a nejnovější jsou stereokamery od společnosti VIVOTEK. Používají se od roku 2021 v nových vozidlech (vyjma tramvají 40T ve kterých jsou ještě infrsenzory). Tyto kamery pracují na principu sledování hlav cestujících vůči žluté hraně schodů a využívají se též jako kamery pro snímání dveřního prostoru pro řidiče. Všechny senzory jsou propojeny s jednotkou systému APC, která sbírá data jak z nich, tak z palubního počítače (od společnosti Herman). Z palubního počítače si jednotka bere především časové a polohové údaje (GPS), zastávky a údaje o vozidle. Samotný sběr dat probíhá off-line přímo ve vozidle. Tento proces je velmi robustní a v případě výpadků, nebo jiných problémů je algoritmus schopný rozhodnout, zda nasbíraná data do sčítání zařadit, či ne. Poslední proces vozidlové části systému je přenos dat mezi vozidlem a serverem, který probíhá přes Wi-fi síť až ve vozovně po nastartování vozidla.

Aplikační část systému APC začíná zpracováním dat z vozidel, kdy se získaná data spárují s nahranými jízdními řády a přiřadí se ke konkrétním spojům. Následně se data zpracují do aplikace Abirun ve které jsou ve velmi pěkné podobě prezentována uživateli jak lze vidět níže na obrázku č. 7. Aplikace mimo jiné dokáže dopočítávat cestující například v situacích, kdy se senzoru nepodaří správně zjistit počet nastupujících, či vystupujících, což se projeví například tím, že v konečné stanici počet cestujících ve vozidle není roven nule. Dále program nabízí i řadu funkcí, jako jsou souhrnné statistiky, přehledy, teplotní mapy a mnohé další. Tyto přehledy nebyly pro účel této práce zcela vhodné, proto byla data exportována ve formátu xlxs (tj. pro zpracování kalkulaátorem Excel). Určitou nevýhodou programu byla nutnost exportování dat manuálně po každém spoji.

Obrázek č. 7: Systém Abirun

ABIRUN APC Vozidla automatické počítání cestujících ve vozidlech 

Spoje Ruční sčítání Referenční data Statistiky Diagnostiky Uživatelé Nastavení čeština Adam Štátný

Souhrnný přehled Typ dat APC Typ linky (nevybráno) Datum 02.09.2021 Exportovat ve formátu XLSX

Datum	Vozidla s daty		Vozidla v provozu		Spoje						Linky		Přepr. osoby	Diag.
	Vytížené	Nevytížené	Zpožděné	S náskokem	Celkem	Linky	Přepr. osoby							
čt 2. 9. 2021	47 45%	105	88 84%	105	5 1%	267 50%	321 60%	42 8%	533 14%	3 852	24 62%	39	18 009	162
st 1. 9. 2021	60 57%	105	91 87%	105	10 1%	274 41%	369 55%	68 10%	668 17%	3 852	29 74%	39	23 072	204
út 31. 8. 2021	66 63%	105	71 68%	105	12 1%	490 36%	103 7%	94 7%	1 380 36%	3 852	25 64%	39	58 643	39
po 30. 8. 2021	69 66%	105	71 68%	105	12 1%	500 36%	131 9%	150 11%	1 402 36%	3 849	28 72%	39	63 037	18
ne 29. 8. 2021	44 42%	105	42 40%	105	3 0%	446 44%	36 4%	123 12%	1 017 41%	2 452	20 61%	33	32 375	18
so 28. 8. 2021	44 42%	105	48 46%	105	1 0%	479 40%	53 4%	105 9%	1 196 48%	2 506	22 65%	34	39 293	23
pá 27. 8. 2021	68 65%	105	74 70%	105	4 0%	615 48%	106 7%	140 10%	1 431 37%	3 852	29 74%	39	57 278	45
čt 26. 8. 2021	75 71%	105	75 71%	105	8 0%	586 36%	140 9%	134 8%	1 607 42%	3 852	30 77%	39	65 692	27
st 25. 8. 2021	65 62%	105	69 66%	105	9 1%	422 31%	77 6%	115 8%	1 357 35%	3 852	28 72%	39	60 523	16
út 24. 8. 2021	63 60%	105	64 61%	105	9 1%	463 35%	124 9%	141 11%	1 324 34%	3 852	26 67%	39	57 859	18
po 23. 8. 2021	69 66%	105	69 66%	105	11 1%	455 37%	131 11%	133 11%	1 225 32%	3 849	27 69%	39	53 300	20
ne 22. 8. 2021	38 36%	105	38 36%	105	4 0%	498 50%	53 5%	97 10%	987 40%	2 452	20 61%	33	29 860	15
so 21. 8. 2021	37 35%	105	40 38%	105	2 0%	331 34%	38 4%	91 9%	968 39%	2 506	21 62%	34	36 646	20
pá 20. 8. 2021	69 66%	105	69 66%	105	6 0%	499 36%	114 8%	132 10%	1 388 36%	3 847	28 72%	39	56 873	45
čt 19. 8. 2021	69 66%	105	70 67%	105	5 0%	515 37%	79 6%	150 11%	1 403 39%	3 573	30 79%	38	57 129	37
st 18. 8. 2021	60 57%	105	61 58%	105	5 0%	466 38%	109 9%	119 10%	1 213 34%	3 573	24 63%	38	50 414	22
út 17. 8. 2021	66 63%	105	68 65%	105	7 1%	513 41%	89 7%	117 9%	1 237 35%	3 573	25 66%	38	49 471	48

Zdroj: Interní materiály PMDP, 2024

Kromě APC využívá PMDP i systém Sprinter, který se používá především k dispečerskému monitorování, s primárním využitím dispečinku PMDP pro sledování polohy vozidel pomocí GPS, zaznamenávání zpoždění a řešení mimořádných událostí. Tento systém rovněž uchovává informace o vozidlech, jako jsou časy příjezdů a odjezdů ze zastávek.

6.3 Sběr a příprava dat

Následující část bude věnována procesu sběru dat a jejich následné analýze. Pro účely zkoumání provozních dat bylo vybráno období první poloviny prosince, což obecně bývá spojeno s vyšším provozem a frekvencí cestujících. Zároveň je toto období bez výluk způsobených stavební činností, stejně jako bez státních svátků, dnů volna ve vzdělávacích institucích či bez významných kulturních událostí, které by mohly změnit, či narušit běžný provoz této linky.

Díky tomu jsou tak data z tohoto období velmi dobře navzájem srovnatelná. Sledovaným obdobím bylo 10 pracovních dnů.

Cílem práce je zkoumat rozdíly v době odbavování na základě typu nasazeného vozidla, konkrétně tedy mezi autobusy stejné délky, lišícími se počtem dveří. Na základě domluvy s PMDP byl proveden experiment, kdy první týden (pět pracovních dnů) docházelo k vypravování autobusů typu SOR NB 12, dalších pět dnů pak SOR NS 12. Pro eliminaci zkreslení docházelo k vypravování autobusů na stejnou směnu (33/2), naopak nasazované konkrétní vozy daných typů byly různé, stejně jako jejich řidiči. V období od 4.12.2023 do 8.12.2023 byly nasazovány autobusy typu SOR NB 12 (čtyřdveřové) evidenčních čísel 3578, 3581, 3586 a v období od 11.12.2023 do 15.12.2023 pak vozy typu SOR NS 12 (třídveřové) evidenčních čísel 3632, 3636 a 3637. Sledovány byly všechny spoje směny 33/2 v rámci jednoho dne, což činilo 30 spojů linky 33. Komplikací při realizování experimentu byla porucha nasazeného autobusu dne 8.12.2023 v poledne. Sledovaný autobus byl stažen a nahrazen autobusem bez systému APC, tudíž zhruba polovina spojů tohoto dne chybí.

Po ukončení této provozní části přišlo na řadu získání dat. Data byla ze systému Abirun exportována do programu MS Excel, kde se s nimi dále pracovalo. Při tomto procesu nebylo možné exportovat kvůli chybě systému čtyři spoje, a to nejčastěji během času střídání řidičů (kde se tedy pravděpodobně může vyskytovat jakási chyba aplikace). Strukturu dat, která byla exportována ze dne 11.12.2023 a 13.12.2023, lze vidět na obrázku č. 8 níže. Struktura z ostatních dní se mírně odlišovala tak, že kromě tarifní vzdálenosti zobrazovala i skutečnou.

Obrázek č. 8: Data exportovaná ze systému Abirun

Zastávka	Tarifní vzdálenost	Příjezd	Pobyt	Odjezd	Jízdní řád	Časová odchylka	Nástup	Frekvence cestujících	Výstup	Obsazenost	Kvalita přepravy	Součinitel využití
----------	--------------------	---------	-------	--------	------------	-----------------	--------	-----------------------	--------	------------	------------------	--------------------

Zdroj: Vlastní zpracování, 2024

Jak již bylo řečeno výše, data byla zpracovávána v programu MS Excel, kde byly nejdříve utříděna po jednotlivých dnech. Při prvním pohledu a práci s daty bylo patrné, že data bude nutné očistit od zavádějících hodnot. V první řadě nedávala smysl data z počátečních a konečných stanic, kdy doba pobytu (tj. údaj představující dobu stanicování v zastávce) přesahovala i třicet minut (jak dokumentuje obrázek č. 9). Toto bylo způsobeno pravděpodobně započítáním bezpečnostních přestávek řidičů do doby odbavení v zastávce (např. chybným ovládním palubního počítače).

Obrázek č. 9: Spoj v počáteční stanici s podezřele velkou dobou pobytu

Zastávka	Příjezd	Pobyt	Odjezd	Jízdní řád	Časová odchylna	Nástup	Frekvence cestujících	Výstup	Obsazenost	Kvalita přepravy	Součinitel využití
Goethova	20:04:02	32:16	20:36:18	20:02	02:02	0	4	4	0		

Zdroj: Vlastní zpracování, 2024

Po očištění těchto extrémních maximálních dob odbavování byla sledována minima. Nejprve došlo ke smazání údajů, kdy autobus v zastávce zastavil, žádný cestující nenastupoval ani nevystupoval a vůz se opět rozjel. Následně minimální hodnoty pobytu, při kterých docházelo k výměně cestujících. Jak dokumentuje tabulka č. 2, při pohledu na jednotlivá denní minima se stala otázkou konzistentnost srovnávaných dat. Konkrétně u typu vozidla SOR NS 12, kde se minimální doba mezi jednotlivými dny lišila až o 12 vteřin. Při bližším zkoumání této skutečnosti bylo zjištěno, že výrazně nižší minimální doba odbavení nastala v pondělí 11.12.2023 a ve středu 13.12.2023. Tyto dny jezdily novější typy vozidel SOR NS 12 evidenčních čísel 3636 a 3637 (vyrobené a dodané v roce 2021), oproti vozu č. 3632 (z roku 2019), který jezdil ve zbytku týdne. PMDP na následný dotaz uvedly, že je snaha data ze systému APC zpřesňovat, ale přesné kritické okamžiky u jednotlivých typů vozidel jsou známé programátorům dodavatele. Systém APC vyhodnocuje pobyt vozidla v zastávce nikoliv např. podle otevření či zavření dveří vozu, ale v závislosti na rozjezdu vozidla ze zastávky (ujetí stanoveného počtu metrů či uplynutí stanoveného času). Dále bylo uvedeno, že u novějších vozidel PMDP usilovalo o zpřísnění těchto kritérií pro zpřesnění údajů. Data o době pobytu vozu MHD ze systému automatického počítání se tak ukázala pro účely práce jako zcela zavádějící, ač na první pohled systém nedával sebemenší pochyby o konzistentnosti a srovnatelnosti dat (všechna data byla ve stejné struktuře).

Tabulka č. 2: Minimální doby odbavení ve vteřinách dle APC pro jednotlivé dny vozu SOR NS 12

SOR NS 12					
Den	11.12.	12.12.	13.12.	14.12.	15.12.
Min. doba odbavení	0:06	0:18	0:06	0:17	0:17

Zdroj: Vlastní zpracování, 2024

Na základě provedené konzultace s dopravcem (který byl v kontaktu i s dodavatelem systému APC) se pro účely práce jevílo vhodnější využít data z dispečerského systému Sprinter. Na několika spojích dopravce ověřil, že systém Sprinter vykazuje hodnoty příjezdu a odjezdu ze

zastávek konzistentně pro všechny typy vozidel. Tyto hodnoty (ze systému Sprinter) byly srovnatelné s daty o příjezdu a odjezdu ze systému APC pro nejnovější vozy SOR NS 12 (kde docházelo jen k naprosto zanedbatelným odchylkám), zatímco pro starší vozy se odchylky ukázaly větší, a to v průměrné výši 13 sekund pro starší SOR NS 12 a 10 vteřin pro SOR NB 12. Jako nejvhodnější postup tedy bylo zvoleno využít data ze systému APC pro počty cestujících a informace o pravidelném jízdním řádu spoje, zatímco údaje ze systému Sprinter posloužily pro přesné informace o skutečné době stanicování, času příjezdu a odjezdu ze zastávky.

Data ze systému APC a Sprinter tedy bylo nutné napárovat. Tuto činnost významně komplikovaly zejména některé chybějící údaje v systému Sprinter (pokud údaj ze zastávky chybí, systém ji při exportu úplně vynechá, na rozdíl od systému APC, kde je zobrazen prázdný řádek). Naopak v situaci opakovaného odbavení v jedné zastávce systém APC ukáže jen jednu hodnotu, zatímco systém Sprinter ukáže údajů více (prostá kontrola řádků na jeden spoj tak není dostatečným kontrolním mechanismem). Proto byly vyvinuty další kontrolní mechanismy, zajišťující, že budou data napárovaná korektně. Sledován byl počet unikátních zastávek pro každý spoj, a dále pak byly srovnávány času příjezdu a odjezdu z každé zastávky podle obou systémů (pro zamezení posunu dat). U většiny spárovaných dat byly pozorovány očekávané odchylky, ale výjimečně došlo k nevysvětlitelným jevům jako například údaje, kde se čas příjezdu napříč systémy značně lišil. Dle porovnávání s časem odjezdu byly zjištěny dva nejčastější závěry, a to, že buď systém APC určil čas příjezdu předčasně, když ještě vozidlo ve skutečnosti nebylo v zastávce, a tudíž mezi časy odjezdu byla minimální odchylka, nebo se situace jevila jako posun zastávky. Bohužel nelze vyhodnotit, zda k posunu o zastávku napříč systémy opravdu došlo a nelze ani případně vyhodnotit, které časy jsou ve skutečnosti správné. Proto byly tyto zavádějící hodnoty z pozorování vyloučeny. Dalším potenciálně nepřesným jevem byly extrémně krátké doby odbavení, konkrétně 0 nebo 3 vteřiny v situacích, kdy frekvence cestujících nebyla nulová. Opět se jednalo o poměrně vzácnou situaci, a tak nebyl problém tyto hodnoty z výzkumu vyřadit.

Dále bylo nutné ošetřit situace, kdy vozidlo přijelo do zastávky s předstihem a řidič čekal na pravidelný odjezd dle jízdního řádu (předčasný odjezd ze zastávky není tolerován). V zastávce tedy pobýval déle, než vyžadovala samotná výměna cestujících (jak zachycuje tabulka č. 3). Pro eliminaci tohoto čekání byly vyřazeny zastávky, na které autobus přijel s předstihem více jak 5 vteřin. Tento údaj byl zvolen na základě pozorované minimální doby odbavení v zastávce

(s výměnou alespoň jednoho cestujícího). Tato doba se napříč všemi dny pohybovala v rozmezí 5-8 vteřin (dle očištěných dat systému Sprinter).

Tabulka č. 3: Spoj přijíždějící do zastávky s předstihem a čekající na odjezd dle jízdního řádu

Zastávka	Příjezd	Odjezd	Pobyt	Jízdní řád	Nástup	Frekvence cestujících	Výstup	Obsazenost
Rondel	21:25:35	21:26:11	0:00:36	21:26	5	5	0	7

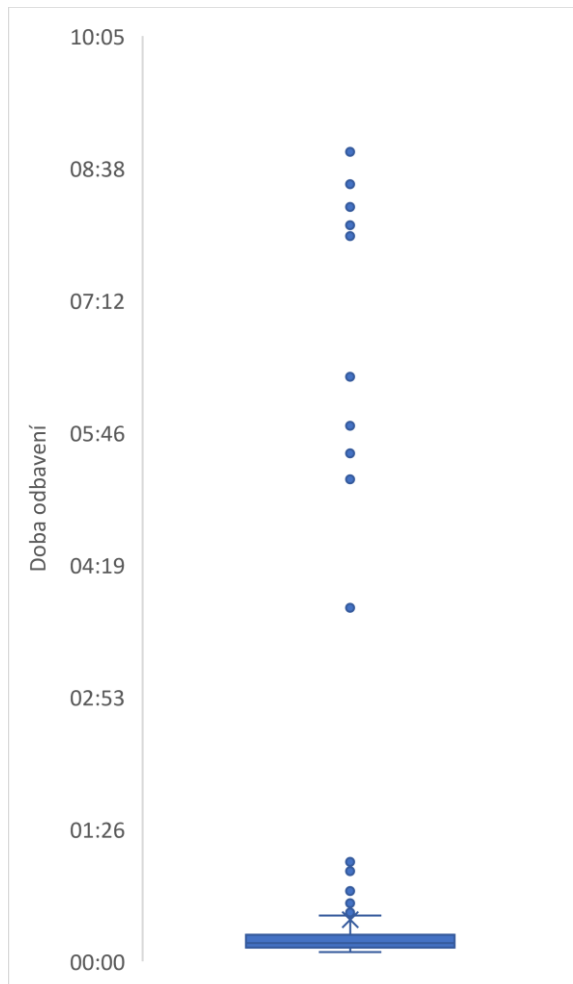
Zdroj: Vlastní zpracování, 2024

Jako poslední byly z další analýzy vyřazeny dva celé spoje ze dne 8.12.2023, kdy po těchto konkrétních spojích nastala výše zmíněná porucha vyžadující výměnu vozidla. Vyřazeny byly poslední dva spoje (před výměnou vozidla), neboť u nich byly pozorovány delší doby odbavení a zpoždění, vyplývající pravděpodobně z projevení samotné, blíže nespecifikované, závady.

Potenciálně zavádějícím údajem může být i nadměrně dlouhá doba odbavení, například v případě, kdy frekvence cestujících je nízká. Extra dlouhé pobyty v zastávkách mohou být způsobené mnoha faktory, jako nástupem vozíčkářů, či lidmi s kočárky, poruchou dveří vozidla, externími vlivy v okolí apod. Jelikož např. nástup rodičů s kočárky či vozíčkářů systém APC, či Sprinter v implementaci pro PMDP zatím rozlišit nedokáže, lze je vnímat jako náhodné výkyvy, které nejsou s ohledem na celkové množství analyzovaných dat zásadní (těmto údajům by dodalo kontext skutečné pozorování reality). Nutno dodat, že se jednalo o relativně malé množství hodnot.

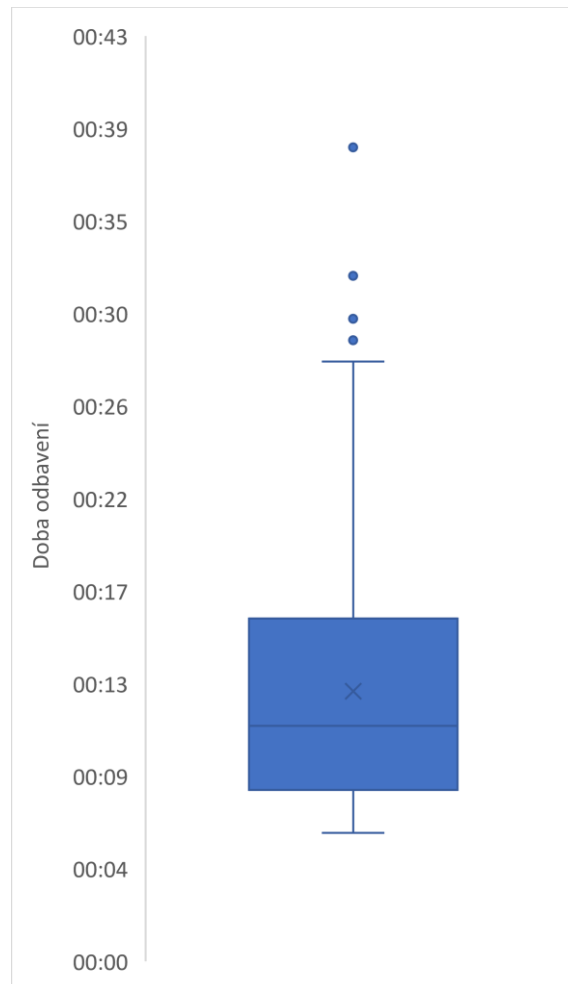
Provedením všech úprav se dataset zmenšil z celkového počtu údajů o stanicování z 3892 na 2397. Rozložení dat se díky provedeným úpravám velmi změnilo. To budou ilustrovat data z 11. prosince 2023: Na obrázku č. 10 lze vidět krabicový diagram surových dat před úpravami a na obrázku č. 11 pak krabicový diagram po všech výše realizovaných úpravách (u druhého grafu byl pro lepší vypovídající schopnost upraven rozsah vertikální osy).

Obrázek č. 10: Krabicový diagram doby odbavení před očištěním ze systému Sprinter dne 11.12.2023



Zdroj: Vlastní zpracování, 2024

Obrázek č. 11: Krabicový diagram doby pobytu po očištění ze systému Sprinter dne 11.12.2023



Zdroj: Vlastní zpracování, 2024

Obecně jsou chybějící a „podivná“ či „zavádějící“ data dle vyjádření Specialisty dopravně provozního úseku (Ing. A. Šťastný, osobní komunikace, 2.4.2024) způsobena výpadky GPS signálu, který byl v případě dat linky 33 pozorován zejména v centru města a zastávky Krašovská. Ač systém APC i Sprinter využívají stejných zdrojů dat, každý z nich se při výpadku chová jinak. Systém Sprinter je vůči výpadkům více imunní a ukazuje stále spolehlivá data pobytu, zatímco systém APC může mít tendence přiřazovat údaje o odbavení k jiné zastávce nebo ukazovat velmi dlouhé doby pobytu.

6.4 Vyhodnocení dat

Nyní jsou očištěná data připravena k samotnému zkoumání. Lze na ně nahlížet ze dvou pohledů, a sice zdali údaje za necelých 2x 5 dnů provozu představují výběrové a nebo základní šetření z pohledu dlouhodobého provozu MHD, a tedy zdali s daty pracovat jako se základním, či s výběrovým souborem. Od toho se odvíjí i zvolené statistické postupy (pro větší vypovídající schopnost je na data nahlíženo postupně oběma těmito způsoby).

Popisná statistika

Prvním pohledem lze říci, že kompletní data o provozu dvou autobusů za dvakrát zhruba 4-5 dnů představují kompletní pozorování reality. Pracuje se s velkým objemem dat, které tedy tvoří vyčerpávající šetření popisující skutečnost. Tyto data nejlépe popíše popisná statistika a pro výpočet směrodatné odchylky a rozptylu bude uvažována populační verze.

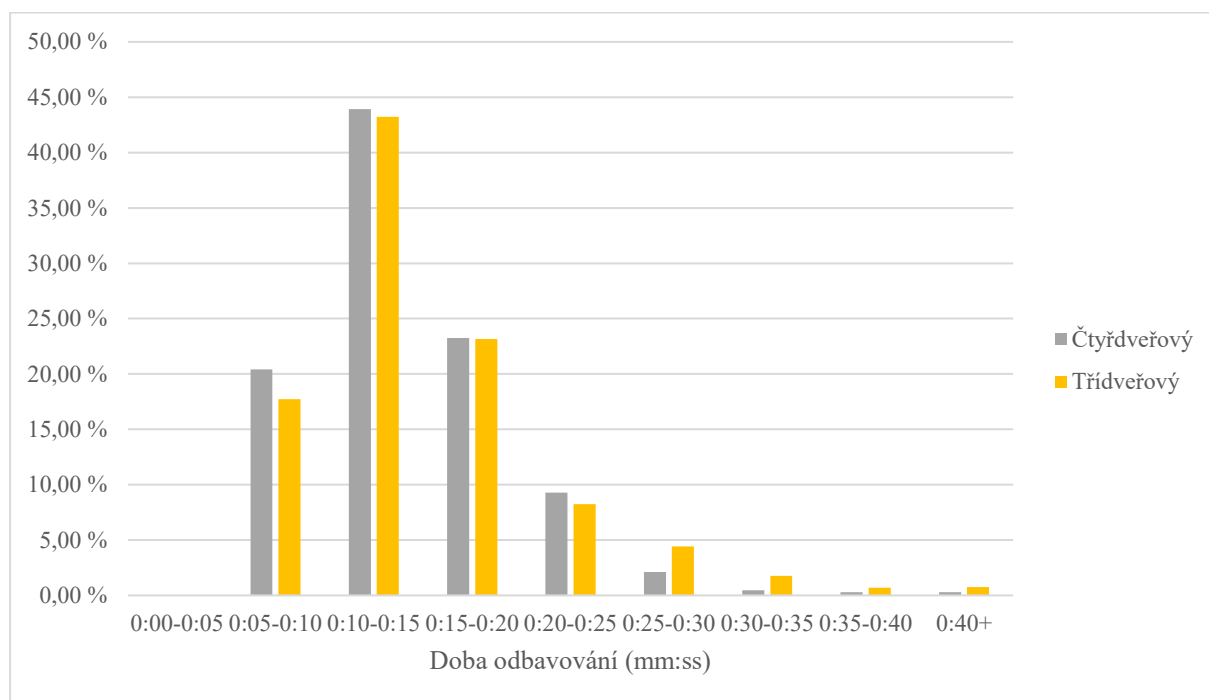
Data z každého pracovního týdne byla nejprve sjednocena a následně, pro ně byly vypočteny základní statistické charakteristiky, které přibližuje tabulka č. 4 a obrázek č. 12. Dále byl pro tyto údaje vytvořen histogram relativní četnosti. Vůbec nejnižší pozorovanou délkou doby pobytu s výměnou cestujících bylo 5 vteřin pro NB i NS 12 a nejvyšší bylo 68 sekund pro NB a 73 vteřin pro NS 12.

Tabulka č. 4: Průměr, směrodatná odchylka a medián pro oba typy autobusů (vše ve formátu minuta:sekunda)

	Průměr	Směrodatná odchylka	Medián
Čtyřdveřový	00:14	00:05	00:12
Třídveřový	00:15	00:07	00:13

Zdroj: Vlastní zpracování, 2024

Obrázek č. 12: Histogram relativní četnosti doby odbavení dle typu vozidla



Zdroj: Vlastní zpracování, 2024

Čtyřdveřové vozidlo, tedy SOR NB 12 se jeví jako mírně rychlejší a jedny dveře navíc v průměru za sledované období urychlily odbavení vozu o 1 vteřinu. I z histogramu je patrné, že vyšší doba odbavení je četnější u třídveřového modelu NS. Kromě základního souhrnného pohledu za celé sledované období byly s využitím napárovaných dat z APC dále analyzovány doby odbavení v závislosti na obratu cestujících (tj. součtu nastupujících a vystupujících cestujících v jedné stanici). Data byla rozdělena do skupin dle velikosti obratu, a pro každou skupinu byly spočteny opět základní charakteristiky. Základní charakteristiku přibližuje nejprve tabulka č. 5. Ještě zajímavější kontext datům dodává tabulka č. 6, která udává, jak často nastala příslušná velikost obratu v jedné zastávce s ohledem na celé období pozorování (v očištěném datasetu se pro čtyřdveřový jednalo o 1088 údajů o odbavení a 1309 pro třídveřový). Pro vzájemné srovnání dat stanicování je předpokládáno, že mezi oběma týdny nedošlo k zásadním výkyvům frekvence cestujících na dané lince. Průměrný obrat osob v jedné zastávce za první týden (čtyřdveřový autobus) činil 6,80 osoby (se směrodatnou odchylkou 6,37), v druhém týdnu pak 6,44 (s odchylkou 6,12). Je patrné, že tedy druhý týden bylo cestujících přepraveno nepatrně méně, ale celkově lze oba týdny považovat za srovnatelné (srovnání dat navíc

komplikují chybějící údaje za půl den provozu v prvním týdnu, které by mohly průměr ještě ovlivnit).

Tabulka č. 5: Doba odbavení podle typů autobusu a dle obratu cestujících

	Typ vozidla		Obrat
	Čtyřdveřový	Třídveřový	
Průměr	00:00:23	00:00:28	30 a více osob
Sm. odch.	00:00:05	00:00:05	
Průměr	00:00:23	00:00:26	20-29 osob
Sm. odch.	00:00:08	00:00:06	
Průměr	00:00:18	00:00:20	10-19 osob
Sm. odch.	00:00:05	00:00:05	
Průměr	00:00:12	00:00:13	9 a méně osob
Sm. odch.	00:00:04	00:00:05	

Zdroj: Vlastní zpracování, 2024

Tabulka č. 6: Relativní četnost kategorií obratu cestujících v jedné zastávce za pozorované období provozu linky 33

Obrat cestujících v 1 zastávce	Typ vozidla		Rozdíl průměrů (hh:mm:ss)
	Čtyřdveřový	Třídveřový	
30 a více osob	0,74 %	0,69 %	00:00:05
20 až 29 osob	4,96 %	4,13 %	00:00:03
10 až 19 osob	18,66 %	17,95 %	00:00:02
9 a méně osob	75,64 %	77,23 %	00:00:01

Zdroj: Vlastní zpracování, 2024

Data potvrzují základní intuitivní předpoklad, že větší množství dveří se projeví na kratší době pobytu zejména v situaci velké výměny cestujících, kdy rozdíl mezi vozidly činí celou směrodatnou odchylku (5 vteřin). Dále je jednoznačně patrné, že při snižující se frekvenci cestujících se i rozdíl mezi jednotlivými typy vozidel snižuje (stále se kratšími doby pobytu vyznačuje NB, ale rozdíly jsou stále méně významné).

Spolu se zohledněním dat z tabulky 6 vychází najevo, že v reálném provozu linky 33 se výhodnost většího počtu dveří ukazuje zejména v situacích velkého obratu cestujících, kterých však na lince nastává naprosté minimum (méně než 1 % všech případů odbavování, méně než

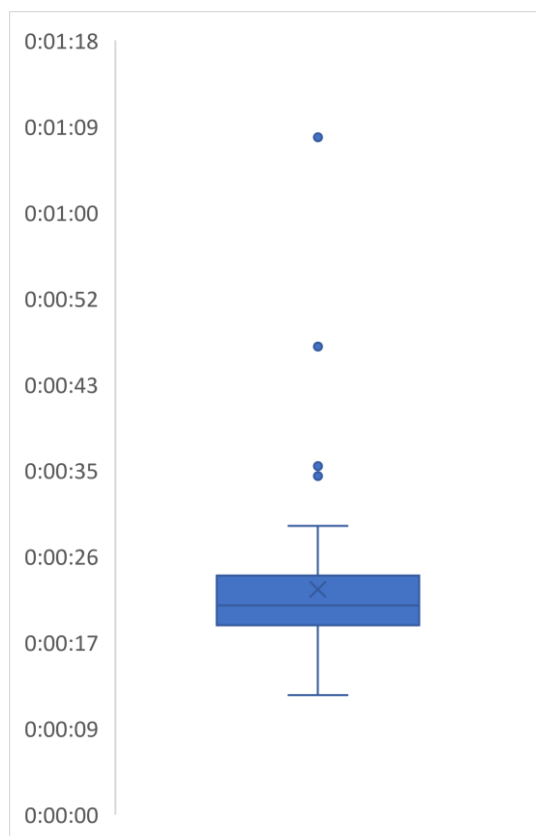
10 stanicování v jednom týdnu). Naopak ve více než 75 % případů je frekvence výměny cestujících v jedné zastávce nízká a průměrný rozdíl mezi odbavením činí pouhou jednu vteřinu. Určení „významného“ a „nevýznamného“ rozdílu mezi dobami odbavování nikoliv z pohledu statistického, ale praktického je diskutabilní i s ohledem na „efekt velikosti“ diskutovaný v sekci 5.4. Je při plánování provozu významný rozdíl vteřin pěti, nebo i tří... dvou?

Nízký podíl případů vysokého obratu cestujících je důležité doplnit o kontext, že situace byla sledována pro nejvytíženější autobusovou linku, na kterou jsou dvanáctimetrové autobusy stabilně u PMDP nasazovány (a v situaci standardní poptávky). S ohledem převládající hodnoty s vteřinovým rozdílem je tedy i průměrný rozdíl za sledovanou směnu za sledované období (a všechny velikosti obratu) roven právě pouze jedné vteřině. Podobnost podílů všech kategorií obratu v tabulce 6 zároveň slouží jako další potvrzení předpokladu, že mezi oběma týdny pozorování nedošlo k zásadním změnám frekvence cestujících. Nepatrně nižší frekvence cestujících v druhém týdnu provozu teoreticky může autobusu NS 12 ve srovnání časů mírně pomoci (doby stanicování při nižší frekvenci budou kratší) a skutečný rozdíl dob odbavení mezi typy tak ještě nepatrně snižovat. Celkově lze ale data o počtu cestujících za oba týdny považovat za srovnatelná.

Rozdílnost dob odbavování mezi typy lze ilustrovat také graficky. Pro tento účel byly obrázky č. 12 a 13 vizualizovány údaje o odbavení v situaci, kdy se v zastávce mění 20 a více cestujících (tj. zhruba kolem 5 % všech stanicování za sledované období).

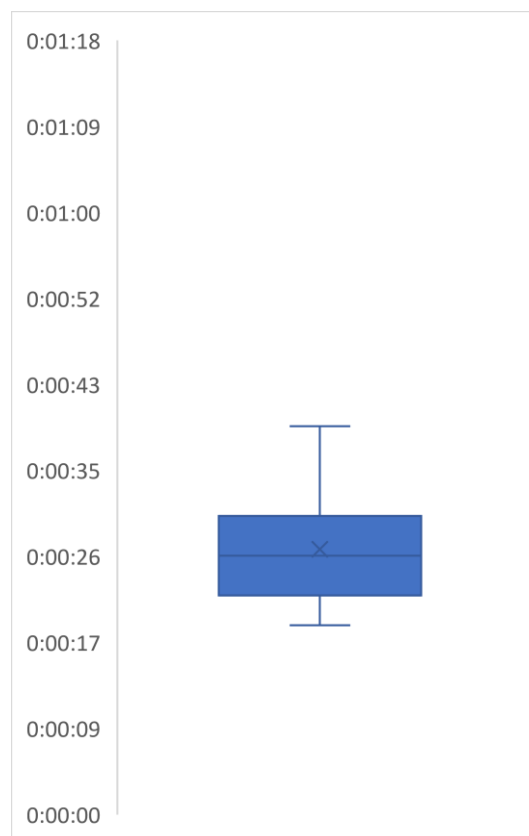
Uváděné podíly stanicování dle obratu cestujících (v tabulce 6) jsou ve skutečnosti ještě nižší, s ohledem na očišťování údajů např. o vynechání údajů o stanicování s nulovou frekvencí. Na méně vytížených linkách tak lze očekávat, že rozdíl mezi dobou pobytu obou typů autobusů bude zcela zanedbatelný. Naopak by čtyřdveřová vozidla nejlépe uplatnila svůj benefit na linkách, které se vyznačují vysokým zatížením s vysokými obraty cestujících. Ač zde by pravděpodobně mohlo být uvažováno spíše o kloubových vozidlech. Nejvyšší obraty nad 30 cestujících totiž představují necelou třetinu celkové maximální obsaditelnosti modelů NB 12 a NS 12 (dle výrobce) a více než polovinu komfortní obsaditelnosti vozidla (jak přibližovala tabulka č. 1).

Obrázek č. 13: Krabicový diagram doby odbavení vozu SOR NB 12 (pro obměnu 20 a více cestujících v jedné stanici; údaje uváděny ve formátu hh:mm:ss)



Zdroj: Vlastní zpracování, 2024

Obrázek č. 14: Krabicový diagram doby odbavení vozu SOR NS 12 (pro obměnu 20 a více cestujících v jedné stanici; údaje uváděny ve formátu hh:mm:ss)



Zdroj: Vlastní zpracování, 2024

Testování hypotéz

Druhým možným pohled je, že nasbíraná data představují jen relativně krátký výběr z možného dlouhodobého pozorování provozu těchto vozidel. Pro tento pohled je možné využít testování hypotéz (jak bylo popsáno v předchozí kapitole). Data za oba týdny pozorování byla napárována do dvojic (doba pobytu a obrat cestujících pro každý typ autobusu v prvním a druhém týdnu pozorování). S ohledem na chybějící údaje za některé spoje je ve výsledku k dispozici 871 párů dat, nad kterými budou statistické testy vykonány. S ohledem na napárování dat do dvojic je vhodné využít párové testy. Vzhledem k bližší nepopsanému rozložení získaných dat je vhodné použít jak parametrický, tak neparametrický test (s ohledem na množství dat lze očekávat, že jsou splněny předpoklady i parametrického typu testu). Jako

parametrický test byl zvolen párový t-test a pro neparametrický pak Wilcoxonův párový test. Nejprve je třeba stanovit nulovou hypotézu H_0 , která tvrdí, že typ vozidla (konkrétně dle počtu dveří) nemá vliv na dobu odbavení, a alternativní hypotézu H_a , která navrhuje, že typ vozidla ovlivňuje dobu odbavení. Nutno podotknout, že parametrický test testuje rozdíl mezi středními hodnotami obou skupin, zatímco neparametrický test porovnává mediány těchto skupin.

Samotné testování dvojic doby pobytu za sledované období by však nebylo dostatečné. Nelze totiž hypoteticky vyloučit situaci, kdy mezi oběma týdny došlo k velkému výkyvu v množství přepravovaných cestujících. Dále je proto nutné otestovat, zdali střední hodnota cestujících, které se ve sledovaném období mění je v mezitýdenním shodná nebo ne. S ohledem na obecnou sílu testů byl tento překlad testován jen parametricky, dalším t-testem (nulová hypotéza předpokládá shodnost hodnot obratu cestujících v zastávkách).

Pro parametrické t testy bylo využito funkce T.TEST v prostředí MS Excel, která vrací p-hodnotu. Při testování hypotézy o rozdílné střední dob odbavení byla kalkulována p-hodnota ve výši 0 (konkrétněji 0,000026), což vede k zamítnutí nulové hypotézy (předpokládající shodu dob) a lze vyslovit závěr, že doba odbavování mezi typy vozidel se za sledované období skutečně liší. Testování hypotézy o shodné střední hodnotě cestujících v mezi týdenním srovnání vedlo ke spočtení p-hodnoty ve výši 0,11. Na všech základních hladinách významnosti tedy nelze zamítnout nulovou hypotézu o shodě obratu cestujících za sledovaném období, což je žádoucí stav. Je tedy patrné, že při obdobné frekvenci cestujících mezi oběma týdny dochází ke statisticky významnému rozdílu v dobách stanicování. Ten tvoří v průměru za celé sledované období 1 vteřinu (tedy ve shodně s popisnou statistikou na nespárovaném vzorku dat prezentovanou v předchozí sekci).

Neparametrický Wilcoxonův test je pro výpočet bez použití specializovaného softwaru nepatrně náročnější. Test byl opět proveden v prostředí MS Excel a nejprve bylo nutné vypočítat rozdíl mezi párovými hodnotami. Následně bylo vypočítáno takzvané signum pro tyto rozdíly, které přiřadí hodnotu 1 pro kladné rozdíly, -1 pro záporné a 0 pro nulové hodnoty (následným vynásobením dojde k odstranění nulových hodnot, které jsou pro výpočet nepodstatné a jsou vyřazovány). V dalším kroku se spočítaly absolutní hodnoty těchto rozdílů a bylo určeno jejich pořadí. Následně se vypočte testové kritérium dle vzorce uvedeného na obrázku č. 17. Výsledkem je testové kritérium, které vyšlo relativně vysoké, konkrétně 4,43 a jelikož hranice kritického oboru pro hladinu významnosti 0,05 je zhruba 1,96 (vycházející z normálního rozdělení), tak lze nulovou hypotézu opět zamítnout. Typ vozidla má tedy i dle

neparametrického testu vliv na dobu odbavení, což potvrzuje závěry všech výše realizovaných statistických pohledů.

Obrázek č. 15: Vzorec pro výpočet Wilcoxonova párového testu

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

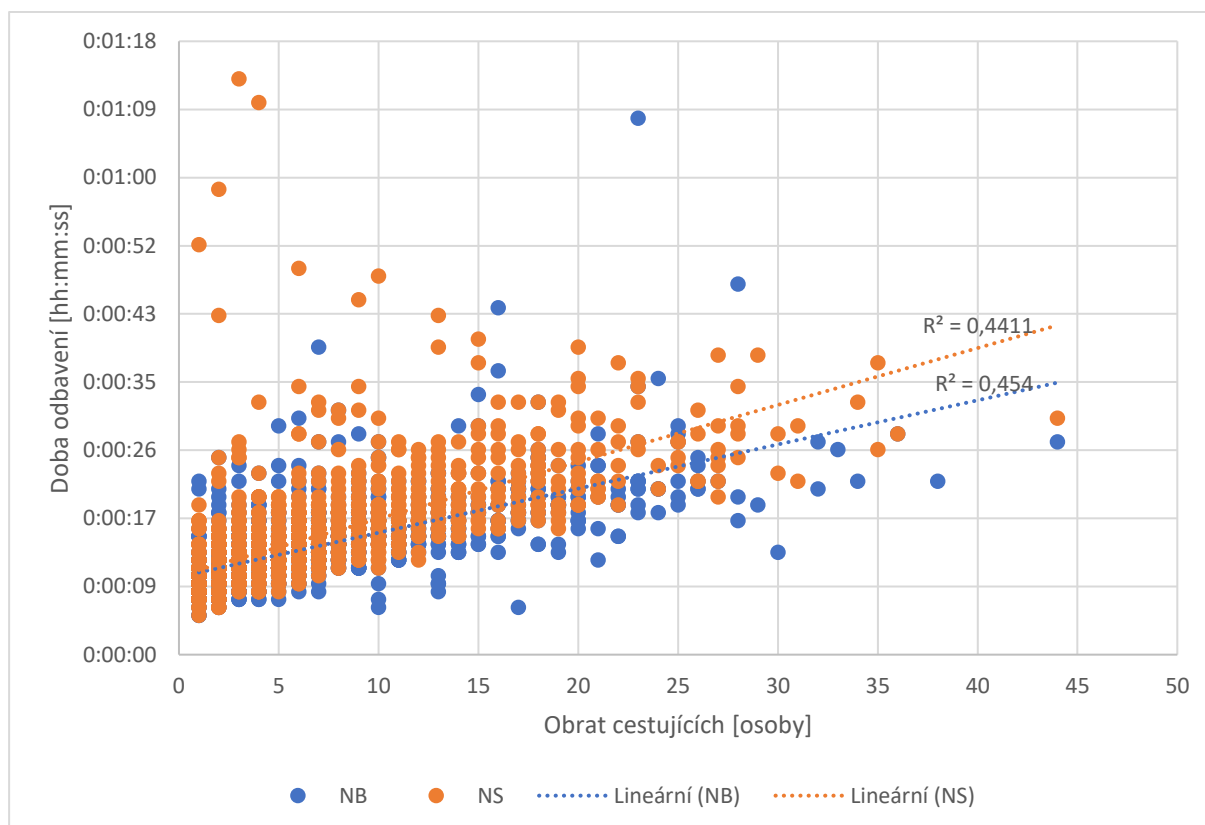
Zdroj: Zichová, 2014

Souvislost doby odbavování a frekvence cestujících

Výše realizované postupy naznačují, že doba odbavení závisí jak na typu vozidla, tak na obratu cestujících (množství osob nastupujících a vystupujících v jedné zastávce). Dále proto byla orientačně zkoumána souvislost doby odbavování s frekvencí cestujících. K tomu posloužil dataset nespárovaných očištěných hodnot. Jako nejvhodnější model byla vyhodnocena lineární regrese, modelující rostoucí dobu odbavení při rostoucím obratu cestujících. Hranice definičního oboru (maximální obrat osob) nebyla v tomto případě stanovena, ale čistě hypoteticky ji tvoří dvojnásobek celkové stanovené kapacity každého z typu vozidel. V praxi se na jedné zastávce vyměňuje cestujících mnohem méně (maximum obratu cestujících na jedné zastávce ve sledovaném období činilo 44 osob). Při lineární regresi je koeficient determinace zhruba kolem 45 % pro oba typy vozidel, a i regresní přímky naznačují trend, že s rostoucím obratem cestujících, je i doba odbavení u třídvěřového vozidla SOR NS 12 delší (souhrnně přibližuje obrázek č. 18). Dále byla statisticky testována významnost provedených regresních modelů (pro každý typ vozidla jeden model). Pro oba vyšla p-hodnota rovna prakticky nule, díky čemuž lze zamítnout statistické nevýznamnosti regresního modelu (blíže specifikováno v příloze A). Lze tak říci, že variabilitu dobu pobytu v zastávce lze u každého typu autobusu zhruba z 45 % vysvětlit obratem cestujících. Zbývající variabilitu může ovlivňovat řada faktorů, zejména „charakteristik cestujících“, jak naznačoval obrázek č. 1. Vliv na dobu odbavování rozhodně mají například nástupy hendikepovaných (řidič musí vystoupit a manipulovat s plošinou pro nástup vozíčkáře) či kočárků. Tyto situace zatím systém APC nedokáže podrobněji rozlišit. Doba stanicování v zastávce může ovlivnit i čekání řidiče na dobíhající

cestujících. Podrobnější analýza regrese nicméně v tuto chvíli přesahuje stanovený cíl této bakalářské práce, a proto představuje spíše námět na další podrobnější šetření.

Obrázek č. 16: Lineární regrese závislosti doby odbavení na obratu cestujících (včetně zobrazení příslušných koeficientů determinace)



Zdroj: Vlastní zpracování, 2024

Závěrem lze uvést rovnice regresních přímk, které mají podobu $y = 0,5596x + 9,8643$ pro vozidlo NB a $y = 0,7213x + 10,021$ pro NS (pro obě platí, že y je doba odbavování ve vteřinách a x obrat cestujících v jedné zastávce).

Shrnutí

Oba testy vzorku párových dat tedy potvrzují stejné závěry jako prostá popisná statistika na celém souboru očištěných dat. Lze tedy konstatovat, že třídvéřové vozidlo SOR NS 12 je za sledované období průměrně pomalejší oproti čtyřdvéřovému SOR NB 12 o jednu vteřinu. Popisná statistika dále ukázala, že počet dveří ovlivňuje dobu odbavení cestujících zejména v situacích velmi vysoké výměny cestujících, kdy průměrný rozdíl činí 5 vteřin. Tyto situace jsou však v běžném provozu poměrně vzácné. Provedená lineární regrese na souboru

očištěných dat rovněž naznačuje, že doba odbavování souvisí s množstvím cestujících, které se v zastávce obměňují a obratem cestujících lze vysvětlit zhruba 45 % variability doby odbavování každého z typů autobusů.

Doporučení

Hlavním zjištěním této práce je, že benefit většího počtu dveří v podobě rychlejší výměny cestujících nastává v praktickém provozu ve velmi malém počtu případů. Za sledované dvoutýdenní období se jednalo zhruba o 5 %, ve kterém dokázaly dveře navíc zkrátit odbavování v zastávce zhruba o 3-5 vteřin (průměrně o 5 vteřin v méně než 1 % případů). V průměru za celé sledované období zkrátily další dveře navíc dobu odbavování v průměru o 1 vteřinu, což lze z praktického pohledu vnímat jako zanedbatelný rozdíl. Nákup čtyřdveřových dvanáctimetrových vozidel by si tak dopravci měli kriticky rozmyslet zejména s uvažováním skutečných frekvencí na linkách a reálné časové úspore. Data jednoznačně ukazují, že benefit více dveří se více uplatní pouze na velmi vyčíslených linkách s vysokými obraty cestujících. Větší počet dveří v bočnici dále vede k odlišnému uspořádání interiéru, které nemusí být pro cestující příliš pohodlné (pro kvalifikovanější popsání názoru cestujících však v této oblasti chybí příslušný výzkum).

Dalším doporučením vyplývajícím z této práce je sjednocení údajů v systému APC dopravce PMDP, kdy bylo zjištěno, že pro novější typy vozidel jsou pravidla příjezdů a dojezdů nastavena nekonzistentně ve srovnání se staršími vozidly a nelze je tudíž navzájem porovnávat. Do budoucna lze také doporučit PMDP větší propojení obou systémů APC a Sprinter, pomocí vlastního vyvinutého řešení, které by odstranilo nutnost ručních exportů a manipulací s daty pro případné rozsáhlejší analýzy.

7 Závěr

Cílem práce bylo zkoumat souvislost mezi délkou doby odbavení a typem konkrétního vozidla MHD. Tento cíl byl splněn díky experimentu provedenému ve spolupráci s PMDP, díky kterému bylo možné získat a vyhodnotit data za bezmála 2x5 dní provozu dvou typů autobusů na stejné lince (respektive stejné směně).

V první části práce byla představena teoretická východiska, včetně představení dopravce PMDP, srovnávaných vozidel a zkoumané linky. Následovala část zaměřená na data z provozu. Konkrétní výzkum byl zaměřen na porovnání třídvéřového a čtyřdvéřového dvanáctimetrového autobusu značky SOR (modely SOR NB 12 a NS 12). Byla nasbírána data pobytů ve stanicích za dva po sobě jdoucí pracovní týdny. Následně byla data očištěna (což se o samo o sobě ukázalo jako výzva, s ohledem na zjištěnou nekonzistentnost dat v jednom z používaných systémů) a analyzována pomocí jednoduché popisné statistiky, testování hypotéz a jednoduché regresní analýzy. Pro statistické testy byl zvolen párový t test jako parametrický a Wilcoxonův párový test jako neparametrický.

Výsledky práce potvrdily intuitivní předpoklad, že vliv počtu dveří se projevuje především při vyšší výměně cestujících. Čtyřdvéřový autobus prokázal v těchto situacích vyšší rychlost výměny cestujících, resp. kratší dobu odbavování. Frekvence výměny 30 a více osob, při které vyniknou největší rozdíly v době odbavení (průměrný rozdíl 5 vteřin) však tvoří méně než 1 % pozorování, a naopak ve více než 75 % případů stanicování je obrat cestujících malý, při kterém benefit většího počtu dveří není příliš výrazný (průměrný rozdíl mezi dobou odbavení činil 1 vteřinu). Prakticky je tak vliv vyššího počtu dveří na dobu odbavování poměrně malý, neboť situace velmi vysokých frekvencí je na linkách poměrně vzácná. Toto zjištění může podniku umožnit lepší plánování nákupu vozidel s ohledem na charakter linek, provozní ukazatele i zákaznickou spokojenost.

Výzkum však není bez omezení. Byl sledován pouze provoz na jedné lince sítě MHD v relativně krátkém časovém období. Data mohla být poznamenána půldenním výpadkem měření dne 8.12.2023 (z důvodu závady na vozidle) a dále s výpadky spojů způsobené ztrátou GPS signálu či chybami jednotlivých systémů.

Celkově lze konstatovat, že problematiku doby odbavení je třeba vnímat komplexně. Práce poukazuje na výzvy spojené s vyhodnocením provozních dat z MHD a zdůrazňuje významnost důkladného očištění a konzistence dat pro různé typy vozidel, ač se data na první pohled mohou

jevit srovnatelně. Tato, zatím spíše přehlížená problematika, se s rozvojem systémů big dat stává naprosto nezbytnou zejména pro správnou interpretaci provozních dat a případné manažerské implikace.

Výsledky této práce mohou podnítit diskusi o problematice doby odbavení a její roli při zlepšování veřejných služeb a prospěšné mohou být také při vývoji dalších typů autobusů. Současně mohou poskytnout užitečné informace pro odborníky zabývající se městskou dopravou, urbanismem a udržitelností v městských oblastech.

Seznam použitých zdrojů

- Anderson, D., Sweeney, D., & Williams, T. (2024, February). *Experimental design*. *Encyclopædia Britannica*. <https://www.britannica.com/science/statistics/Experimental-design>
- Baseel, C. (2017, December 4). *One passenger pusher isn't enough for Tokyo Area Train that needs three-man team*. SoraNews24. <https://soraneews24.com/2017/12/04/one-passenger-pusher-isnt-enough-for-tokyo-area-train-that-needs-three-man-team%E3%80%90vi-deo%E3%80%91/>
- BasuMallick, C. (2022, August 8). *What is Big Data? definition and best practices*. Spiceworks. <http://www.spiceworks.com/tech/big-data/articles/what-is-big-data/>
- Bedáňová, I., & Linhart, P. (2024). *Testování hypotéz ve statistice*. <https://cit.vfu.cz/stat-potr/POTR/Teorie/Predn3/hypotezy.htm>
- Biswal, A. (2023, October 11). *What is hypothesis testing in statistics? types and examples: Simplilearn*. Simplilearn.com. <https://www.simplilearn.com/tutorials/statistics-tutorial/hypothesis-testing-in-statistics>
- Botelho, B., & Bigelow, S. J. (2022, January 5). *What is Big Data and why is it important?*. Data Management. <https://www.techtarget.com/searchdatamanagement/definition/big-data>
- Clifford, T. (2021, July 8). *3 challenges with transportation data collection and how to solve them*. StreetLight Data. <https://www.streetlightdata.com/3-challenges-with-transportation-data-collection-and-how-to-solve-them/>
- Clifford, T. (2023, February 7). *5 urban transportation challenges that big data analytics can help you solve*. StreetLight Data. <https://www.streetlightdata.com/5-urban-transportation-challenges-that-big-data-can-help-you-solve/>
- Česká republika. (2024). *Strategický rámeček ČR 2030*. Akademie městské mobility. <https://www.akademiemobility.cz/ceska-republika-2030>
- DPMÚL. (2023, November 7). *Tisková zpráva*. Dopravní podnik města Ústí nad Labem. <https://dpmul.cz/o-nas/aktuality/tiskova-zprava11>
- Drdla, P. (2021). *2. Technologie, řízení a LINKOTVORBA v MHD 2.1 ... doc*. Ing. Pavel Drdla, Ph.D. <http://drdla.wz.cz/skripta/2.pdf>
- Dueker, K., Kimpel, T., Strathman, J., & Callas, S. (2004). Determinants of bus dwell time. *Journal of Public Transportation*, 7(1), 21–40. <https://doi.org/10.5038/2375-0901.7.1.2>

Fintech. (2015, June 17). *V plzeňské MHD Mohou Lidé platit Bezkontaktní Bankovní kartou.* PMDP a platba kartou. <https://www.bankovnipoplatky.cz/v-plzenske-mhd-mohou-lide-platit-bezkontaktni-bankovni-kartou-28962>

Gao, Y., & Zhu, J. (2022, June 15). *Characteristics, impacts and trends of Urban Transportation.* MDPI. <https://www.mdpi.com/2673-8392/2/2/78>

Google. (2024). *Big data defined: Examples and benefits.* <https://cloud.google.com/learn/what-is-big-data>

Harris, N. G., Simone, F. de, & Condry, B. (2022, March 7). *A comprehensive analysis of passenger alighting and boarding rates - urban rail transit.* SpringerLink. <https://link.springer.com/article/10.1007/s40864-021-00161-8>

Hendl, J. (2015). *Přehled statistických metod: Analýza a Metaanalýza Dat.* Portál.

Intellias. (2023, August 28). *Big Data and transportation: Use cases for urban planning.* <https://intellias.com/big-data-and-transportation-use-cases-urban-planning/>

IOSR Journals. (2021, May 24). *The challenges of using Big Data Analytic in smart transportation.* Academia.edu. https://www.academia.edu/49040326/The_Challenges_of_Using_Big_Data_Analytic_in_Smart_Transportation

Kresa, Z., Tluchor, J., & Heyes, A. (2023, January 1). *Attractiveness of urban public transport from the point of view of young passengers: Experience from the Czech Republic.* IGI Global. <https://www.igi-global.com/chapter/attractiveness-of-urban-public-transport-from-the-point-of-view-of-young-passengers/317273>

Majaski, C. (2023, December 19). Hypothesis to be tested: Definition and 4 steps for testing with example. Investopedia. <https://www.investopedia.com/terms/h/hypothesistesting.asp>

McAfee, A., & Brynjolfsson, E. (2012, October). *Big Data: The management revolution.* Harvard Business Review. <https://hbr.org/2012/10/big-data-the-management-revolution>

Milkovits, M. N. (2008). Modeling the factors affecting bus stop dwell time. *Transportation Research Record: Journal of the Transportation Research Board*, 2072(1), 125–130. <https://doi.org/10.3141/2072-13>

Mrkvička, T., & Petrášková, V. (2006). Úvod do statistiky. <http://home.ef.jcu.cz/~mrkvicka/vyuka/Statistika2Vydani.pdf>

- Pavlík, T. (2024). *Biostatistika Pro Matematickou Biologii*. Matematická biologie učebnice: Biostatistika pro matematickou biologii. <https://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickyh-a-biologickyh-dat--biostatistika-pro-matematickou-biologii>
- PMDP. (2024, January 1). *Plzeňské městské dopravní podniky, A.S.* <https://www.pmdp.cz/>
- Segal, T. (2022, November 29). *What is Big Data? definition, how it works, and uses.* Investopedia. <http://www.investopedia.com/terms/b/big-data.asp>
- Rashidi, S., Ataeian, S., & Ranjitkar, P. (2022). Estimating bus dwell time: A review of the literature. *Transport Reviews*, 43(1), 32–61. <https://doi.org/10.1080/01441647.2021.2023692>
- Rowe, J. (2022). *Public transportation: Planning, Operations and Management*. NY Research Press.
- Schroer, A. (2024, March 29). *Big Data*. Big Data: What It Is, Why It Matters, How It Works | Built In. <https://builtin.com/big-data>
- SOR. (2017, September). SOR NB 12 / NB 18. https://www.sor.cz/wp-content/uploads/2017/09/NB_12-18_ANG_LoRes.pdf
- SOR. (2018, November). *NS 12 Diesel CZ NL - Sor*. SOR NS 12. https://www.sor.cz/wp-content/uploads/2018/11/NS_12_diesel_CZ_NEW_LoRes.pdf
- Sun, L., Tirachini, A., Axhausen, K. W., Erath, A., & Lee, D.-H. (2014). *Models of bus boarding and alighting dynamics*. World Transit Research. <https://www.worldtransitresearch.info/research/5406/>
- Torre-Bastida, A. I. (2018, July 17). *Big Data for transportation and mobility: Recent ...* - wiley online library. Big Data for transportation and mobility: recent advances, trends and challenges. <https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-its.2018.5188>
- Utilities One. (2023, October 25). *Utilities One*. The Role of Data Analytics in Urban Transportation Planning. <https://utilitiesone.com/the-role-of-data-analytics-in-urban-transportation-planning>
- Zhuang, Z., Cheng, Z., Yao, J., Wang, J., & An, S. (2022, September 23). *Bus travel time reliability incorporating stop waiting time and in-vehicle travel time with AVL data - International Journal of Coal Science & Technology*. SpringerLink. <https://link.springer.com/article/10.1007/s40789-022-00544-7>

Zichová, J. (2014). *Neparametrické testy*. Matematická statistika. https://www.kalin.mff.cuni.cz/~zichova/PRFUK/MatStat/09_wilcox_testy_tisk.pdf

24X7Editing.com (2023, October 27). *Dissertation editing, thesis editing*. 24x7 Editing. <https://www.24x7editing.com/exploring-the-limitations-of-hypothesis-testing-with-solutions/>

Seznam tabulek

Tabulka č. 1: Obecná specifikace vozů SOR.....	24
Tabulka č. 2: Minimální doby odbavení ve vteřinách dle APC pro jednotlivé dny vozu SOR NS 12.....	30
Tabulka č. 3: Spoj přijíždějící do zastávky s předstihem a čekající na odjezd dle jízdního řádu.....	32
Tabulka č. 4: Průměr, směrodatná odchylka a medián pro oba typy autobusů (vše ve formátu minuta:sekunda).....	34
Tabulka č. 5: Doba odbavení podle typů autobusu a dle obratu cestujících.....	36
Tabulka č. 6: Relativní četnost kategorií obratu cestujících v jedné zastávce za pozorované období provozu linky 33.....	36

Seznam obrázků

Obrázek č. 1: Faktory ovlivňující dobu nástupu.....	8
Obrázek č. 2: Nulová a alternativní hypotéza vůči chybám.....	20
Obrázek č. 3: Čtyřdveřový autobus SOR NB 12	25
Obrázek č. 4: Třídveřový autobus SOR NS 12.....	25
Obrázek č. 5: Plán interiéru vozu SOR NS 12.....	26
Obrázek č. 6: Plán interiéru vozu SOR NB 12.....	26
Obrázek č. 7: Systém Abirun.....	28
Obrázek č. 8: Data exportovaná ze systému Abirun.....	29
Obrázek č. 9: Spoj v počáteční stanici s podezřele velkou dobou pobytu.....	30
Obrázek č. 10: Krabicový diagram doby pobytu před očištěním ze systému Sprinter dne 11.12.2023.....	33
Obrázek č. 11: Krabicový diagram doby pobytu po očištění ze systému Sprinter dne 11.12.2023.....	33
Obrázek č. 12: Histogram relativní četnosti doby odbavení dle typu vozidla.....	35
Obrázek č. 13: Krabicový diagram doby odbavení vozu SOR NB 12 (pro obměnu 20 a více cestujících v jedné stanici; údaje uváděny ve formátu hh:mm:ss).....	38
Obrázek č. 14: Krabicový diagram doby odbavení vozu SOR NS 12 (pro obměnu 20 a více cestujících v jedné stanici; údaje uváděny ve formátu hh:mm:ss).....	38
Obrázek č. 15: Vzorec pro výpočet Wilcoxonova párového testu.....	40
Obrázek č. 16: Lineární regrese závislosti doby odbavení na obratu cestujících (včetně zobrazení příslušných koeficientů determinace).....	41

Seznam příloh

Příloha A: Jednoduchá regresní analýza

Abstrakt

APA 7: Čermák, J. (2024). Souvislost doby odbavení s typem vozidla MHD [Bakalářská práce, Západočeská univerzita v Plzni].

Klíčová slova: Autobus, big data, doba odbavení, městská hromadná doprava, statistika

Tato práce se zabývá analýzou souvislostí mezi dobou odbavení a typem vozidla městské hromadné dopravy (MHD) na příkladu autobusové linky 33 Plzeňských městských dopravních podniků. Pro účely práce byla získána, očištěna, analyzována a interpretována provozní data o době stanicování a obratu cestujících pro dva typy autobusů stejné délkové kategorie lišící se počtem dveří (SOR NB 12 a SOR NS 12). V teoretické části jsou uvedeny faktory ovlivňující dobu odbavení, představen dopravce, vozidla i zkoumaná linka, a dále využití big dat a základy testování hypotéz. Praktická část je pak zaměřena na analýzu doby odbavení dvou autobusů s rozdílným počtem dveří. Využito je základní popisné statistiky, testování hypotéz a jednoduché regresní analýzy. Ukazuje se, že vyšší počet dveří přispívá k rychlejší výměně cestujících, ale že na nejfrekventovanější plzeňské lince pro danou kategorii vozidel činí průměrný rozdíl v dobách odbavování jednu vteřinu (největší průměrný rozdíl – 5 vteřin – nastává při frekvenci odpovídající méně než 1 % pozorování). Kromě samotných výsledků práce přináší zajímavé poznatky i v oblasti zpracování provozních big dat, neboť byly v procesu přípravy identifikovány nekonzistentní údaje, které by jinak snadno vedly ke zkresleným výsledkům. Na základě výsledků práce jsou vyvozeny závěry a navrhována doporučení pro dopravce, diskutována jsou také omezení výzkumu.

Abstract

APA 7: Čermák, J. (2024). Relation of bus dwell time with the type of vehicle [Bachelor Thesis, University of West Bohemia].

Key words: Bus, big data, dwell time, public transportation, statistics

This work deals with the analysis of correlations between dwell time and the type of vehicles in urban public transportation (PT) using the example of bus line 33 of the Pilsen city transportation company. For the purposes of this work, operational data on dwell time and passenger turnover were obtained, cleaned, analyzed, and interpreted for two types of buses of the same length category but differing in the number of doors (SOR NB 12 and SOR NS 12). The theoretical part introduces factors influencing dwell time, presents the carrier, vehicles, and the studied route, as well as the use of big data and basics of hypothesis testing. The practical part focuses on analyzing the dwell time of two buses with different numbers of doors. Basic descriptive statistics, hypothesis testing, and simple regression analysis are utilized. It is shown that a higher number of doors contributes to faster passenger turnover, but on the busiest Pilsen route for the given vehicle category, the average difference in boarding times is one second (the largest average difference - 5 seconds - occurs at a frequency corresponding to less than 1% of observations). In addition to the actual results of the work, interesting findings are also provided in the area of operational big data processing, as inconsistent data were identified in the preparation process, which could otherwise easily lead to distorted results. Based on the results of the work, conclusions are drawn and recommendations for the carrier are proposed, with discussions on research limitations also included.